Modelling Load Balancing Mechanisms in Self-Optimising 4G Mobile Networks

Mariusz Głąbowski
Poznan University of Technology
Poznan, Poland
mariusz.glabowski@put.poznanan.pl

Sławomir Hanczewski
Poznan University of Technology
Poznan, Poland
slawomir.hanczewski@put.poznan.pl

Maciej Stasiak
Poznan University of Technology
Poznan, Poland
maciej.stasiak@put.poznan.pl

***Abstract*—This article describes an approximate model of a group of cells in the LTE network with implemented load balancing mechanism. An appropriately modified model of Erlang's Ideal Grading is used to model this group of cells. The model makes it possible to take into account limited availability of resources of individual cells to multi-rate traffic streams. The developed solution allows the basic traffic characteristics in the considered system to be determined, i.e. the occupancy distribution and the blocking/loss probability. Because of the approximate nature of the proposed model, the results obtained based on the model were compared with the results of a digital simulation. The present study validates the proposed model both in terms of the adopted assumptions and in terms of the simulation techniques.**

I. Introduction

In order to effectively use resources of mobile networks, including 4G multi-service networks in particular, it is necessary to develop and implement specific mechanisms designed to make optimisation of the distribution of network traffic between neighbouring cells possible. As a result of the research studies carried out within the 3rd Generation Partnership Project (3GPP) [1] and Next Generation Mobile Networks [2], among others, a concept of self-organising (self-optimizing, self-configuring) networks (SON) has been introduced in which the key element is load balancing in individual cells of a mobile network. Thus developed concept of load balancing makes it possible to take into account load variations in time for individual cells that result from, for example, mobility of users. The concept is based on connection handover from cells with high load to neighbouring cells (though not always – [3]) with lower load. The distribution of network traffic optimised in this way makes it possible to not only use resources of a network more effectively, but also to improve the QoS and QoE parameters (quality of service, quality of experience) of multi-rate traffic streams generated by end users. Effective operation of the load balancing mechanism requires the process of dynamic change of node parameters of the network to be fully automated. These would include, for example, a modification to the coverage provided by individual cells or, also viewed as more effective, modifications to handover mechanisms [1]. The operation of handover procedure is usually initiated without any participation of a terminal (i.e. a mobile station has no influence whatsoever on the decision to initiate the procedure). Connection handover can occur between cells or base stations of the same network (Intra-LTE Handover), or between different networks (Inter-RAT Handover) [1].

Regardless of the load balancing mechanism/algorithm involved, the area covered by a given group of cells can be treated – from the point of view of traffic engineering – as a multi-service system in which the service of multi-rate calls is performed with the use of "separated" resources (cells) that are component parts of a given network. An analysis of the current state of research on the analytical modelling of mobile systems indicates that multi-rate models [4], [5], [6] are typically used for a determination of traffic characteristics of multi-service systems. In most articles addressing this issue only single cells are modelled, while the basic multi-rate model, i.e. the full-availability group model with multi-rate traffic, is used in the process. However, in the case of a group of cells between which connection handover is feasible, it is necessary to employ more complex multi-rate models. This article proposes an approximate method for a determination of traffic characteristics in a group of cells that service multi-rate traffic stream. To determine these characteristics, a generalized model of Erlang's Ideal Grading with multi-rate traffic will be used [7], [8].

The remainder of the paper is structured as follows. Section II gives a description of a general model of the system with limited availability. Section III presents the proposed analytical method for a determination of traffic characteristics in mobile systems that would take into account the phenomenon of connection handover between cells to equalize cell load. In Section IV, the results of the calculations for a number of selected groups are compared with the data provided by a simulation. Finally, the summary and conclusion are provided in Section V.

II. Generalized model of a system with limited availability

A. Resource allocation units

The application of multi-rate analytical models to analyse present-day broadband telecommunications systems is possible with a proper bandwidth discretisation. Discretisation allows the capacity of a system and bit rates demanded by calls to be expressed in the so-called basic bandwidth units (BBU). The discretisation process itself consists in replacing a variable bit rate of call streams with a specific fixed bit rate called the equivalent bandwidth. Individual and specific methods of determining the equivalent bandwidth rely on various parameters, including the admissible packet-loss ratio, admissible delay, link flow capacity (bit rate), average and peak bit rates, packet stream type (e.g., selfsimilar streams) and network type. The algorithms for finding the equivalent bandwidth for given network types and services are presented, for example, in [9]–[14]. The discretisation is widely used in teletraffic engineering by many authors, for example [15]–[17].

Let us assume that we identify the equivalent bandwidths for all packet stream types offered in a given multi-service system. In such a case, the BBU can be calculated as the greatest common divisor of the equivalent bandwidths determined for all call streams:

$$R_{\text{BBU}} = GCD(R_1, R_2, ..., R_M), \tag{1}$$

where R_i is the equivalent bandwidth for calls of class i.
The number of allocation units demanded by calls of class i is defined by the following dependence:

$$t_i = \left\lceil \frac{R_i}{R_{\text{BBU}}} \right\rceil, \tag{2}$$

The capacity of the system expressed in allocation units is equal to:

$$V = \left\lfloor \frac{C}{R_{\text{BBU}}} \right\rfloor, \tag{3}$$

where C is the capacity of the telecommunications system under consideration.

B. Erlang's Ideal Grading model

Let us consider the multi-rate model of EIG group (Erlang's Ideal Grading) with differentiated availabilities for particular traffic classes [7], [8].

The presented group is offered M independent classes of Poisson traffic streams of the following intensities: λ_1, λ_2, ..., λ_M. A call of class i needs t_i BBU for a connection to be initiated. The holding time of calls of particular classes is charakterised by an exponential distribution with the parameters: μ_1,μ_2,...,μ_M. As a result, the mean traffic introduced to the system by a traffic stream of class i equals:

$$A_i = \lambda_i / \mu_i. \tag{4}$$

The EIG group is described by three parameters: capacity V BBUs, availability d_i BBUs and the number of load groups g_i for class i calls. Availability determines the number of BBUs in a group to which a given traffic source of class i has access. Sources that have access to the same resources of a group create a load group. The assumption is that traffic of class i offered to a group is distributed evenly throughout all load groups of class i. The number of load groups of class i in the EIG model is equal to the number of possible combinations of d_i BBUs from among all V BBUs:

$$g_i = \begin{pmatrix} V \\ d_i \end{pmatrix}. \tag{5}$$

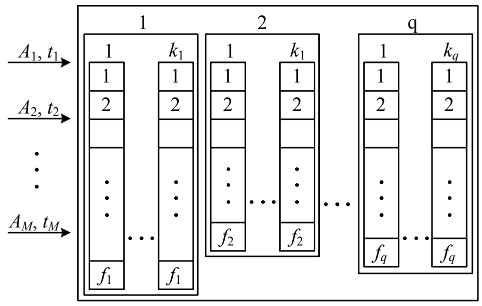

Fig. 1. System with limited availability

If the values for availability of $d_1, d_2, ..., d_m$ BBUs for all call classes are known, we can determine then the occupancy distribution in the EIG with differentiated availability on the basis of the following formula [7]:

$$n\,[P_n]_V = \sum_{i=1}^{m} A_i t_i [1 - \sigma_i(n - t_i)]\,[P_{n-t_i}]_V\,, \tag{6}$$

where A_i is traffic offered to the group by calls of class i (Formula (4)), whereas $\sigma_i(n)$ is the conditional transition probability that takes into account the availability parameters for calls of individual classes:

$$\sigma_i(n) = 1 - \sum_{d-t_i+1}^{k} \begin{pmatrix} d_i \\ x \end{pmatrix} \begin{pmatrix} V-d_i \\ n-x \end{pmatrix} \Big/ \begin{pmatrix} V \\ n \end{pmatrix}, \tag{7}$$

where:
$k = n - t_i$, if $(d_i - t_i + 1) \leq (n - t_i) < d_i$,
$k = d_i$, if $(n - t_i) \geq d_i$.
The blocking probability for calls of class i in considered EIG can be determined on the basis of the following formula:

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

Distribution (6) is an approximate distribution. Nevertheless, numerous studies and analyses carried out by the authors have indicated high accuracy of the EIG model [8], [18]–[20].

C. System with limited avialability

Consider now a generalised model of a system with limited availability, i.e. a system composed of a number of separated resources with different capacities. Another assumption is that the system is consist of q types of resources. Each type of resources is unambiguously defined by two parameters: k_s – the number of resources of type s ($1 \leq s \leq q$); f_s – the capacity of the separated resource of type s ($1 \leq s \leq q$). The example of tis kind of system is presented on Figure 1. The total capacity V of the system is then [21], [22]:

$$V = \sum_{s=1}^{q} k_s f_s. \tag{9}$$

The considered system services a call only when this call can be entirely carried by an arbitrary single resource.

Let us take a closer look now at the method for a determination of the availability parameter in the system with limited availability. According to the definition of the system with limited availability, a new call of class i (that demands t_i BBUs for service) can be admitted for service if it can be entirely carried by one of its resources. Consider then a single resource of any type s with the capacity f_s BBUs. As seen from the perspective of calls of class i, the availability in the equivalent group that corresponds to the considered resources of the system with limited availability is equal to the number of BBUs that can be occupied by calls of class i in the considered resources:

$$d_{i,l} = \left\lfloor \frac{f_s}{t_i} \right\rfloor t_i, \tag{10}$$

where l is the successive number of the resource of type s ($1 \leq l \leq k_s$).

Availability for calls of class i that corresponds to the resources of one type s is equal to:

$$d_i^s = \sum_{l=1}^{k_s} d_{i,l} = k_s \left\lfloor \frac{f_s}{t_i} \right\rfloor t_i. \tag{11}$$

After all types of resources in the considered system have taken into account, we get the total availability that corresponds to (and reflects) the structure of the system with limited availability:

$$d_i = \sum_{s=1}^{q} d_i^s = t_i \sum_{s=1}^{q} k_s \left\lfloor \frac{f_s}{t_i} \right\rfloor. \tag{12}$$

After determination of the parameters d for all traffic classes, we are in position to approximate the occupancy distribution and blocking probability for particular traffic classes by EIG model (II-B).

III. Modelling of a group of cells with handover between cells

A. Assumptions for the model of a system with handover between cells

Consider the group of cells presented in Figure 2 where each cell can have a different capacity. Let us adopt that the term "a group of cells” denotes a set of all cells in a given considered area, while the term "a cluster of cells” will denote a set of cells that are directly adjacent to each other. Another assumption is that the group of cells is offered m independent Poisson call streams with the intensities: λ_1, λ_2, …, λ_m. As earlier, the BBUs demanded for calls of individual classes will be denoted by the symbols: t_1, t_2, …, t_M. Holding times for calls of all classes have exponential distributions with the parameters: μ_1, μ_2, …, μ_M.

Yet another assumption is that in the considered group of cells the handover mechanism is used to optimise – from the point of view of load equalisation – the actual usage of available resources. To illustrate the operation of the load balancing mechanism for loads generated as a result of handover between cells, let us consider the group of 10 selected cells presented in Figure 2. Each of the cells shows a cylindrical vessel filled with fluid that represent the load level of each cell. As it can be easily observed, cells 3, 6, 7 and 10 are those that are most heavily loaded. Handover mechanisms make it possible for those connections that cannot be serviced in cells 3, 6, 7 and 10 to be transferred to neighbouring cells, for example, from cell 3 to cells 1, 2, 4 and 5 and from cell 10 to cells 8 and 9.

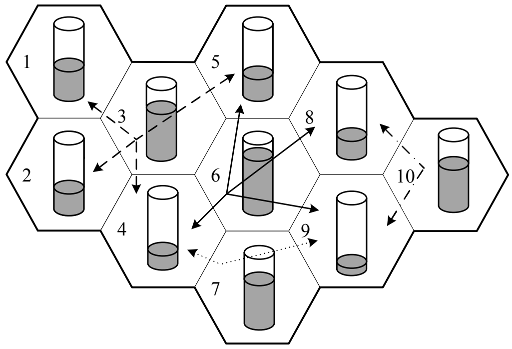

Fig. 2. Connection handover in the group of cells

A group of cells services a call only when it can be carried exclusively by one of the available cells. Observe that calls that arrive to a given cell will be admitted when this cell, or any other cells that are in its immediate vicinity (i.e. within a given cluster of cells) have a sufficient number of resources.

B. Characteristics of a group of cells

On the basis of the above dependencies we can now determine traffic characteristics for the group of cells with the handover mechanism taken into consideration. These characteristics result from the generalised model of the system with limited availability (Section II) that can be subsequently used to model the considered mobile system.

The present paper proposes the following method for modelling a group of cells that jointly service multi-rate traffic streams (with the introduction of the handover mechanism). First, we determine the neighbouring cells for each of the cells within the group, and this cluster of cells will be then modelled by the EIG group. For example, for the group of cells presented in Figure 3,we will focus on the seven clusters of neighbouring cells:

— cluster 1: 1, 2, 3, 4, 5, 6, 7;
— cluster 2: 1, 2, 3, 7;
— cluster 3: 1, 2, 3, 4;
— cluster 4: 1, 3, 4, 5;
— cluster 5: 1, 4, 5, 6;
— cluster 6: 1, 5, 6, 7;
— cluster 7: 1, 2, 6, 7.

For each of the clusters of cells we define its structure (capacities of individual cells) as well as the value of offered traffic that is a component part of the total traffic offered to the group of cells. Then, for each of the clusters, we determine the occupancy distribution and the blocking (loss) probability Thus obtained values for the blocking probability for each of the clusters make input data necessary for a determination of the blocking probability in the whole area.

This article uses the approach, developed in [23], that makes a determination of the blocking probability in the whole area (group of cells) possible on the basis of a determination of the blocking probability in individual clusters of cells. This probability is determined as the geometric mean of the probabilities of individual clusters of cells.

The complete algorithm for determining traffic characteristics of the considered system is presented below:

— Determination of the capacity of the system V – Formula (9).
— Determination fo the limited availability d_i for calls of class i, expressed in BBU (Formula (12)):
— Calculation of traffic of class i offered to the group of cells – Formula (4).
— Determination of the occupancy distribution in the cluster of cells No. j on the basis of Formula (6), which can be rewritten in the following way:

$$n\,[P_n]_{V_j} = \sum_{i=1}^{m} A_{ij} t_i \sigma_{ij}(n-t_i)\,[P_{n-t_i}]_{V_j}, \qquad (13)$$

where V_j denotes the capacity of the cluster of cells no. j, A_{ij} – the intensity of traffic of class i offered to the cluster j (the assumption is that traffic is distributed evenly to all cells), σ_{ij} – conditional transition probability for calls of class i in cluster j.
— Calculation of the blocking (loss) probability of calls of class i in the cluster of cells No. j on the basis of Formula (8) which can be rewritten in the following way:

$$E_{ij} = \sum_{n=0}^{V_j} [P_n]_{V_j} [1 - \sigma_{ij}(n)]. \qquad (14)$$

– Calculation of the blocking probability for calls of class i in the group of cells:

$$E_i = \sqrt[J]{E_{i1} \cdot E_{i2} \cdots E_{iJ}}, \qquad (15)$$

where J denotes the number of clusters of cells in a given group of cells (within the area under consideration).

IV. A Comparison of the Analytical Results with the Simulation Results

The presented method for a determination of traffic characteristics of a group of cells servicing jointly the offered mixture of traffic streams is an approximate method. In order to evaluate the accuracy of the proposed solution the results of the calculations were compared with the data obtained during a simulation [24]. The study was conducted for a typical system of 7 cells, presented in Figure. 3. The assumption was that a call appearing in a given cluster of neighbouring cells could be admitted for service only if this cluster (i.e. any randomly chosen cell of the cluster) had a sufficient number of free resources.

The study was conducted for the following structure of the group of cells:

— **Group 1**

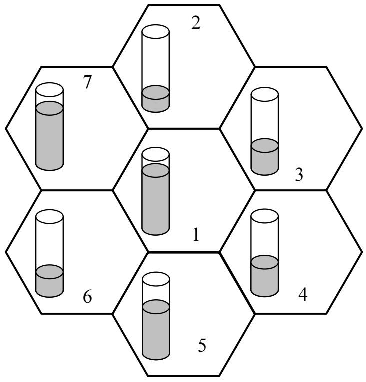

Fig. 3. Structure of the studied mobile system

— Capacity of individual cells, expressed in BBU (Figure (3)): $f_1 = 30$, $f_2 = 45$, $f_3 = 45$, $f_4 = 45$, $f_5 = 35$, $f_1 = 35$, $f_7 = 35$.
— Structure of offered traffic: $t_1 = 1$ BBU, $t_2 = 2$ BBU, $t_3 = 5$ BBU, $A_1t_1 : A_2t_2 : A_3t_3 = 1 : 1 : 1$.

— **Group 2**

— Capacity of individual cells, expressed in BBU (Figure (3)): $f_1 = 90$, $f_2 = 45$, $f_3 = 45$, $f_4 = 90$, $f_5 = 60$, $f_1 = 60$, $f_7 = 70$.
— Structure of offered traffic: $t_1 = 1$ BBU, $t_2 = 3$ BBU, $t_3 = 6$ BBU, $A_1t_1 : A_2t_2 : A_3t_3 = 1 : 1 : 1$.

Figures 4 and 5 shows the results of the calculations and the simulation of the blocking probability for individual classes of call streams in the considered group with a handover mechanism composed of 7 cells. The results are presented in relation to the average value of traffic a offered to one bandwidth unit in the group of cells:

$$a = \frac{\sum_{i=1}^{M} A_i t_i}{V} \qquad (16)$$

The study was conducted for the value of the intensity of traffic offered to a bandwidth unit within the range $0.5 \div 1.2$ Erl. The results of the simulation are presented in Figures 4 and 5 as relevant points in the diagram and with the 95-per cent confidence interval calculated after the t-Student distribution for five series, 1 000 000 calls each (of the class that generated the lowest number of calls) in each of the series. For every point of the simulation the confidence interval is at least two orders of magnitude lower than the results of the simulation.

Fig. 4 and 5 present the results for the modelling of the group of cells in which the possibility to equalise load as a result of connection handover to neighbouring cells was taken into consideration. It is observable that the proposed method for a determination of the blocking probability ensures high accuracy of calculations.

V. Conclusions

This article proposes a new approximate method for calculations of the blocking probability in a group of cells in which the handover mechanism is introduced to equalise load. The proposed method uses an analytical model of EIG

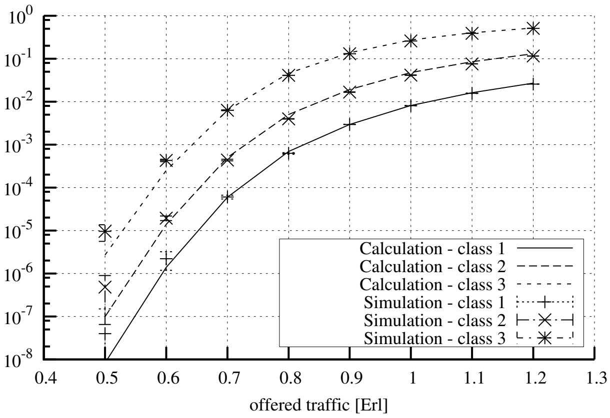

Fig. 4. Loss probability in a group of cells with handover mechanism limited to neighbouring cells (group 1)

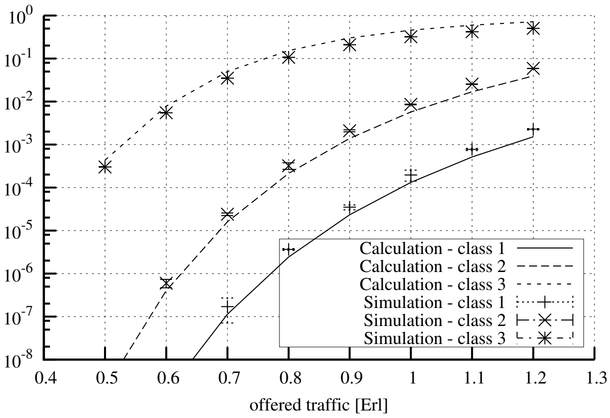

Fig. 5. Loss probability in a group of cells with handover mechanism limited to neighbouring cells (group 2)

with multi-rate traffic streams and differentiated availabilities. The article considers a mechanism for equalising load in which connection handover is possible only between physically neighbouring cells. The relevant literature of the subject addresses also those handover algorithms that make use of cells that are not in close and immediate vicinity [3]. It should be emphasized though that from the point of view of the proposed model the assumption of neighbourhood influences only the structure of the so-called cluster of cells. The results of the analytical calculations of the considered mobile system are compared with the data provided by the simulation. The comparison confirms high accuracy of the proposed method.

REFERENCES

[1] 3GPP, "Self-configuring and self-optimizing network (son) use cases and solutions (release 9)," 3GPP, Tech. Rep. TR 36.902, Sepember 2010.

[2] NGMN, "Next generation mobile networks beyond hspa & evdo (white paper)," NGMN, Tech. Rep. v3.0, December 2006.

[3] N. Zia and A. Mitschele-Thiel, "Self-organized neighborhood mobility load balancing for lte networks," in *2013 IFIP Wireless Days (WD)*, Nov 2013, pp. 1–6.

[4] D. Staehle and A. Mäder, "An analytic approximation of the uplink capacity in a UMTS network with heterogeneous traffic," in *18th International Teletraffic Congress (ITC18)*, Berlin, 2003, pp. 81–91.

[5] M. Głąbowski, M. Stasiak, A. Wiśniewski, and P. Zwierzykowski, "Uplink blocking probability calculation for cellular systems with WCDMA radio interface, finite source population and differently loaded neighbouring cells," in *Proceedings of Asia-Pacific Conference on Communications*, Perth, 2005, pp. 138–142, best paper award.

[6] M. Stasiak, J. Wiewiora, P. Zwierzykowski, and D. Parniewicz, "Analytical model of traffic compression in the umts network," in *Computer Performance Engineering*, ser. Lecture Notes in Computer Science, J. Bradley, Ed. Springer Berlin Heidelberg, 2009, vol. 5652, pp. 79–93. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-02924-0_7

[7] M. Stasiak and S. Hanczewski, "A model of WCDMA radio interface with reservation," in *Proceedings of International Symposium on Information Theory and Its Applications, 2008. ISITA 2008.*, Dec 2008, pp. 1–6.

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

[9] F. Kelly, "Notes on effective bandwidth," University of Cambridge, Tech. Rep., 1996.

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

[11] T. Neame, M. Zuckerman, and R. Addie, "A paractical approach for multimedia traffic modeling," in *5th International Conference on Broadband Communications*, Hong Kong, November 1999, pp. 73–82.

[12] N. L. S. Fonseca, G. S. Mayor, and C. A. V. Neto, "On the equivalent bandwidth of self-similar sources," *ACM Trans. Model. Comput. Simul.*, vol. 10, no. 2, pp. 104–124, 2000.

[13] S. Bodamer and J. Charzinski, "Evaluation of effective bandwidth schemes for self-similar traffic," *Proceedings of the 13th ITC Specialist Seminar on IP Measurement, Modeling and Management, Monterey*, no. 2, pp. 21–1–21–10, 2000.

[14] J. Charzinski, "Internet traffic measurment and characterisation results," in *13th Internetional Symposium on Services and Local Access (ISSLS'2000)*, Stockholm, Sweden, June 2000.

[15] M. Głąbowski, S. Hanczewski, and M. Stasiak, "Erlang's Ideal Grading in Diffserv modelling," in *AFRICON, 2011*, Sept 2011, pp. 1–6.

[16] 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." *Annales des Télécommunications*, vol. 68, no. 5-6, pp. 327–344, 2013.

[17] 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, p. 17, 2012, article ID 547909.

[18] S. Hanczewski and M. Stasiak, "Performance modelling of Video-on-Demand systems," in *Proceedings of 17th Asia-Pacific Conference on Communications (APCC)*, Oct. 2011, pp. 784 –788.

[19] M. Stasiak, M. Głąbowski, and S. Hanczewski, "The application of the Erlang's Ideal Grading for modelling of UMTS cells," in *2012 8th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP)*, July 2012, pp. 1–6.

[20] M. Głąbowski, S. Hanczewski, and M. Stasiak, "Modelling of cellular networks with traffic overflow," *Mathematical Problems in Engineering*, vol. 2015, p. 15, 2015, article ID 286490.

[21] M. Głąbowski, "Recurrent calculation of blocking probability in multiservice switching networks," in *Proceedings of Asia-Pacific Conference on Communications*, Busan, 2006, pp. 1–5.

[22] M. Stasiak and M. Głąbowski, "Point-to-point blocking probability in switching networks with reservation," in *Proceedings of 16th International Teletraffic Congress*, vol. 3A. Edinburgh, UK: Elsevier, Jun. 1999, pp. 519–528.

[23] M. Głąbowski, M. Sobieraj, and M. Stasiak, "Evaluation of traffic characteristics of UMTS with bandwidth reservation and handoff mechanism," in *IEEE International Conference on Telecommunications and Malaysia International Conference on Communications, ICT-MICC 2007.*, May 2007, pp. 736–740.

[24] M. Głąbowski and A. Kaliszan, "Simulator of full-availability group with bandwidth reservation and multi-rate Bernoulli-Poisson-Pascal traffic streams," in *EUROCON, 2007. The International Conference on Computer as a Tool*, Sept 2007, pp. 2271–2277.