Paper

Approximation of Message Inter-Arrival and Inter-Departure Time Distributions in IMS/NGN Architecture Using Phase-Type Distributions

Sylwester Kaczmarek and Maciej Sac

Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gdańsk, Poland

Abstract—Currently it is assumed that requirements of the information society for delivering multimedia services will be satisfied by the Next Generation Network (NGN) architecture, which includes elements of the IP Multimedia Subsystem (IMS) solution. In order to guarantee Quality of Service (QoS), NGN has to be appropriately designed and dimensioned. Therefore, proper traffic models should be proposed and applied. This requires determination of queuing models adequate to message inter-arrival and inter-departure time distributions in the network. In the paper the above mentioned distributions in different points of a single domain of NGN are investigated, using a simulation model developed according to the latest standards and research. Relations between network parameters and obtained message inter-arrival as well as inter-departure time distributions are indicated. Moreover, possibility of approximating the above mentioned distributions using phase-type distributions is investigated, which can be helpful in identifying proper queuing models and constructing an analytical model suitable for NGN.

Keywords—call processing performance, IMS, message inter-arrival time distributions, message inter-departure time distributions, NGN, phase-type distributions, traffic model.

1. Introduction

Next Generation Network (NGN) [1] is a proposition for a standardized, packet-based telecommunication network architecture that delivers multimedia services with guaranteed quality. NGN consist of two stratums: transport stratum containing elements specific for particular transport technology and transport independent service stratum, which includes IP Multimedia Subsystem (IMS) [2] elements and uses mainly SIP (Session Initiation Protocol) [3] as well as Diameter [4] communication protocols.

In order to operate properly, both NGN stratums must be correctly designed. For this reason proposition and application of appropriate traffic models that evaluate among others call processing performance metrics [5], [6] are required. These metrics were formerly known as Grade of Service (GoS) parameters and include Call Set-up Delay (CSD) as well as Call Disengagement Delay (CDD).

As a result of the performed review [7], [8], the authors found out that standards organizations do not consider traffic engineering in their work and a majority of current research does not provide traffic models fully compatible with IMS-based NGN architecture (also abbreviated in the next part of the paper as IMS/NGN). Therefore, we decided to propose our own simulation [9] as well as analytical [10] model and use it to evaluate mean CSD and mean CDD in a single IMS/NGN domain.

During our investigations [10] some differences between analytical and simulation results were obtained, which are noticeable for high load. This leads to the research concerning message inter-arrival and inter-departure time distributions in a single domain of IMS/NGN performed using the simulation model. This would allow improving the precision of the analytical model by replacing simple M/G/1 queuing systems approximating the operation of servers and links with queuing models more adequate for the characteristics of IMS/NGN.

The results of this research were initially presented at a conference [11]. After that, they were supplemented by approximations of obtained message inter-arrival and inter-departure time histograms using phase-type distributions [12]–[16] and assessment of these approximations. All these aspects are presented in this paper, which is organized as follows. In Section 2 a description of the IMS/NGN network model and call scenario is provided. The simulation model used to measure message inter-arrival and inter-departure time distributions is presented along with the assumed measurements methodology. Histograms of message inter-arrival and inter-departure time obtained in different points of the network are described and commented in Section 3. Use of phase-type distributions to approximate the above mentioned histograms is examined in Section 4. Summary and future work are presented in Section 5.

2. Traffic Model of IMS/NGN

Network model [17], [18] and basic call scenario with two-stage resource reservation [18]–[21] in a single domain of ITU-T NGN architecture (the most advanced of all available NGN solutions [7], [8], [10]) are presented in Fig. 1 and Fig. 2 respectively. The model and the scenario are based

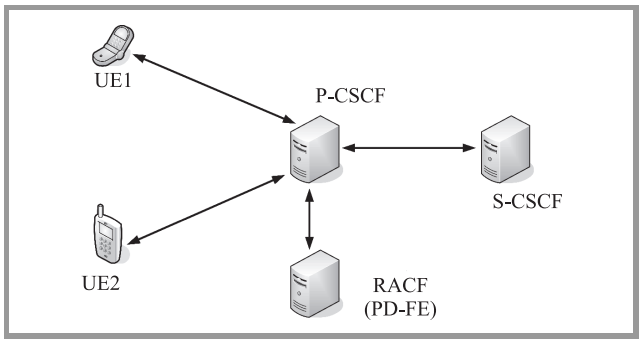


Fig. 1. Model of a single domain of IMS/NGN [17], [18].

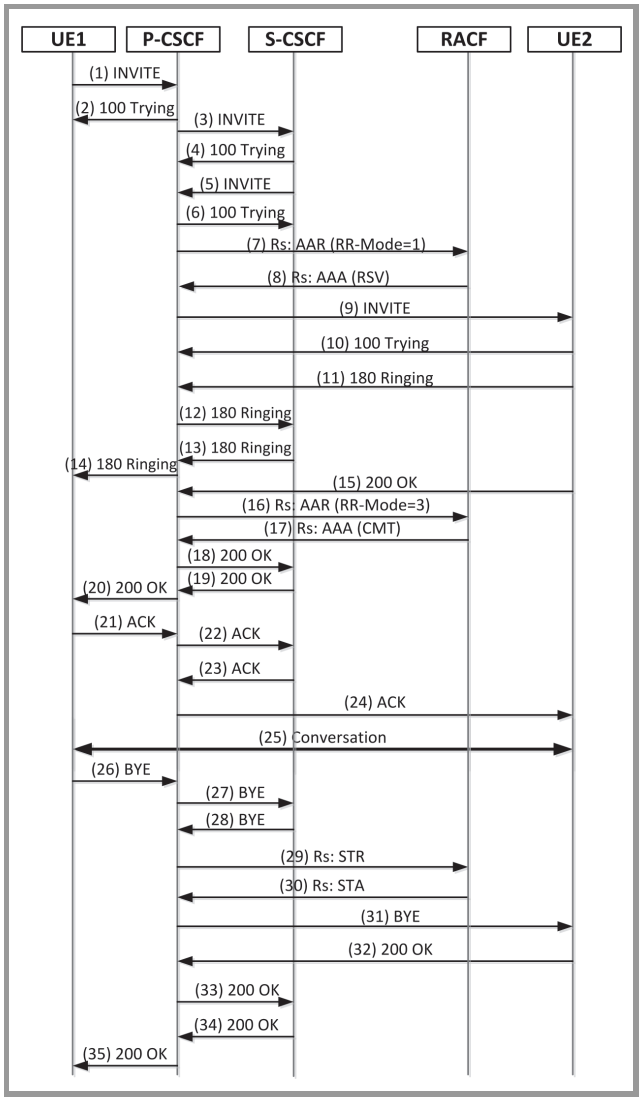


Fig. 2. Call set-up (messages 1–24) and call disengagement (messages 26–35) scenario in a single domain of IMS/NGN.

on the assumption that standard voice calls are requested by users and thus application servers are not used. Moreover, we assume that codec sets in User Equipments (UEs) are compatible and no audio announcements are played during the call.

Call set-up and disengagement requests are sent by UEs registered in the domain to the P-CSCF (Proxy – Call Session Control Function) server, which forwards them to the S-CSCF (Serving – Call Session Control Function) element, the main server handling all calls. The RACF (Resource and Admission Control Functions) unit representing the transport stratum allocates resources for a new call and releases resources associated with a disengaged call.

All elements of the network communicate using SIP protocol [3]. An exception is the communication of P-CSCF with RACF, for which Diameter protocol [4] is applied. Due to limited space available in the paper more detailed information about the network model and call scenario are not provided. This information can be found in [8], [10].

Based on the presented network model (Fig. 1) and assumed signaling messages exchange (Fig. 2), a simulation model of a single domain of IMS/NGN was proposed [9], which precisely implements the operation (algorithms) of all network elements as well as call set-up and disengagement scenarios. The aim of the model is to evaluate mean Call Set-up Delay E(CSD) and mean Call Disengagement Delay E(CDD) [5], [6]. Additionally, during simulations times of message arrival and departure at all network elements and links can be gathered. This allows calculation of histograms estimating message inter-arrival and inter-departure time distributions.

Logical structure of the simulation model is presented in Fig. 3. Details regarding the implementation of the model in the OMNeT++ [22] simulation environment are out of the scope of this paper and can be found in [9]. CSCF servers include Central Processing Units (CPUs), which are responsible for handling messages incoming from

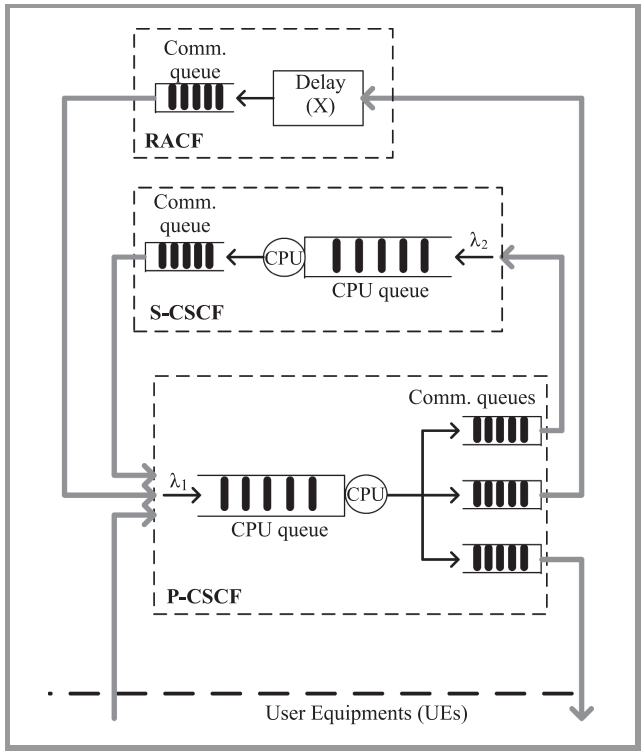


Fig. 3. Logical structure of the proposed traffic model.

other elements according to the assumed call scenario (Fig. 2). When CPUs are busy incoming messages are stored in CPU queues. Other network elements in the model respond with a particular delay (RACF responsible for resource allocation and release as well as UEs representing many user terminals). Each element of the model includes communication queues (one for each outgoing link), which buffer messages before sending them through busy links.

For proper operation of the model the following assumptions are taken [9]:

- message loss probability is negligible and thus there are no message retransmissions and each request is properly handled,
- UE1 generates aggregated call set-up requests (SIP INVITE messages) from many terminals with exponential intervals and given intensity, λ_{INV},
- call duration time is determined by an exponential distribution with definable mean (default 180 s),
- times of processing SIP INVITE messages by P-CSCF and S-CSCF are given by random variables T_{INV1} and T_{INV2} correspondingly,
- a_k factors (k = 1, 2, ... 8) determine times of processing other SIP and Diameter messages by CSCF servers

$$\begin{gathered} T_{TRi} = a_1 \cdot T_{INVi},\ T_{RINGi} = a_2 \cdot T_{INVi}, \\ T_{OKi} = a_3 \cdot T_{INVi},\ T_{ACKi} = a_4 \cdot T_{INVi}, \\ T_{BYEi} = a_5 \cdot T_{INVi},\ T_{OKBYEi} = a_6 \cdot T_{INVi}, \\ T_{AAAi} = a_7 \cdot T_{INVi},\ T_{STAi} = a_8 \cdot T_{INVi}, \\ i = 1 \text{ for P-CSCF, 2 for S-CSCF}, \end{gathered} \tag{1}$$

- time of processing messages by RACF is described by a random variable, T_X,
- UE1 and UE2 represent many user terminals processing messages in nonzero time; SIP INVITE message processing time in UE1 and UE2 is described by a given distribution (default: uniform distribution with values from 1 to 5 ms); times of processing other messages are related to this time as in Eq. (1),
- communication times between elements of the network depend on definable lengths of optical links, bandwidth and lengths of transmitted messages,
- network elements communicate over dedicated interfaces; UE1 and UE2 represent many terminals connected to P-CSCF through a switch/router, communication times between the switches/routers and particular terminals are included in message processing times of UE1 and UE2,
- simulation is finished when the first of the following conditions occurs: total simulation time is exceeded, maximum number of generated calls is exceeded, confidence intervals for mean CSD and mean CDD are not greater than assumed maximum values.

The described simulation model was used to obtain histograms estimating message inter-arrival and inter-departure time distributions. For this reason message arrival and departure times were measured at the following points of all network elements (Fig. 4):

- CPU_i – the time of message arrival (the whole message, all bits are received) to the input of the P-CSCF or S-CSCF CPU queue,
- CPU_o – the time of message departure from the output of the P-CSCF or S-CSCF CPU,
- L_{ik} – the time of message arrival to the input of link communication queue for k-th link (if the link is not busy, then this is the time of sending the first bit of the message),
- L_{ok} – the time of sending the last bit of the message through k-th link.

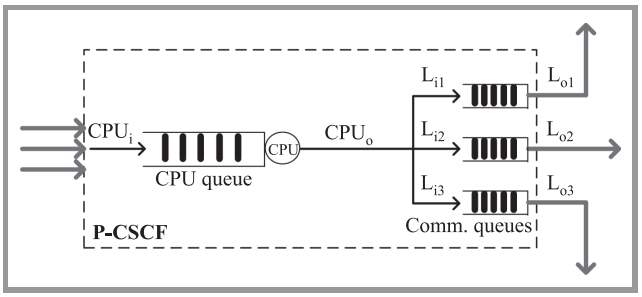


Fig. 4. Points of message arrival and departure times measurements based on the example of the P-CSCF server.

Acquired results were further processed in the MATLAB environment [23] to obtain histograms of message inter-arrival and inter-departure time at different points of the network. Additionally, some statistical values regarding the retrieved data were computed, including mean value and variance of message inter-arrival and inter-departure time. Also values of the c^2 variation coefficient were calculated, which is the ratio of variance of message inter-arrival or inter-departure time to squared mean value of this time.

3. Results

The simulation model described in Section 2 was used to investigate message inter-arrival time distributions at CPU_i and L_{ik} points as well as message inter-departure time distributions at CPU_o and L_{ok} points of all elements of the IMS/NGN domain (Figs. 1 and 4). During the investigations data sets presented in Table 1 and message lengths

presented in Table 2 were used. Additionally, the following assumptions were made:

- warm-up period 1500 s,
- 5 measurement periods,
- 0.95 confidence level,
- $a_1 = 0.2$, $a_2 = 0.2$, $a_3 = 0.6$, $a_4 = 0.3$, $a_5 = 0.6$, $a_6 = 0.3$, $a_7 = 0.6$, $a_8 = 0.6$,
- T_{INV1}, T_{INV2} and T_X input parameters are taken as constant values representing the maximum INVITE processing time by P-CSCF, the maximum INVITE processing time by S-CSCF, and the maximum message processing time by RACF respectively,
- simulation is finished when confidence intervals for E(CSD) and E(CDD) do not exceed 5% of mean Call Set-up Delay and mean Call Disengagement Delay or when total simulation time exceeds 10000 s; with such stop conditions at least 10000 message inter-arrival or inter-departure times were obtained during each simulation.

Table 1
Input data sets

Data set	λ_{INV} [1/s]	T_{INV1} [ms]	T_{INV2} [ms]	T_X [ms]	Links
1	100, 190, 220	0.5	0.5	3	No links
2	100, 190, 220	0.5	0.5	3	300 km 10 Mbit/s
3	100, 190, 220	0.5	0.5	3	300 km 100 Mbit/s

Table 2
Message lengths [24]

Message	Length [bytes]
SIP INVITE	930
SIP 100 Trying	450
SIP 180 Ringing	450
SIP 200 OK (answer to INVITE)	990
SIP ACK	630
SIP BYE	510
SIP 200 OK (answer to BYE)	500
Diameter messages	750

Although measurements in the simulation environment were performed at all available CPU_i, L_{ik}, CPU_o, L_{ok} points and for all data sets (Table 1), due to limited space we demonstrate only selected results. It is important to mention that under tested conditions offered loads to the P-CSCF CPU (ρ_1) and S-CSCF CPU (ρ_2) were as follows:

- $\rho_1 = 0.41$ and $\rho_2 = 0.16$ for $\lambda_{INV} = 100$ [1/s],
- $\rho_1 = 0.78$ and $\rho_2 = 0.30$ for $\lambda_{INV} = 190$ [1/s],
- $\rho_1 = 0.90$ and $\rho_2 = 0.35$ for $\lambda_{INV} = 220$ [1/s].

Performed research (Figs. 5–12) demonstrated that message inter-arrival and inter-departure time distributions in a single domain of IMS/NGN quite significantly differed

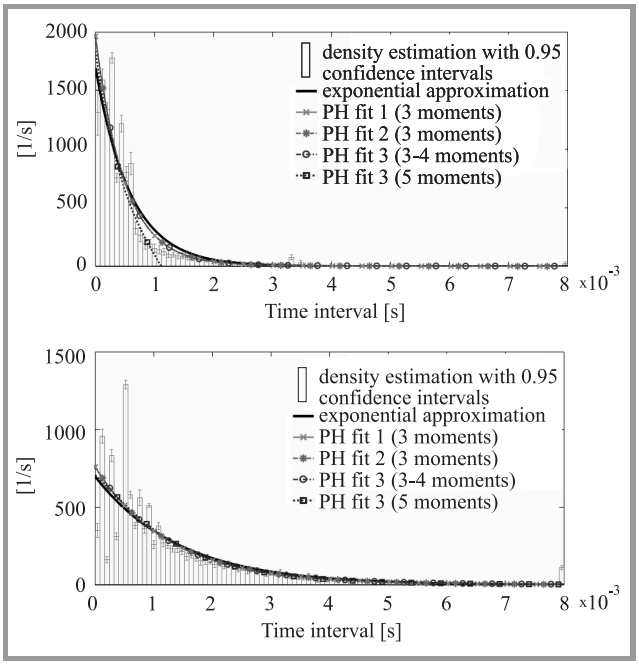


Fig. 5. Histograms of message inter-arrival times at the P-CSCF (top, $c^2 = 2.19$) and S-CSCF (bottom, $c^2 = 1.33$) CPU_i (data set 1, $\lambda_{INV} = 100$ [1/s]).

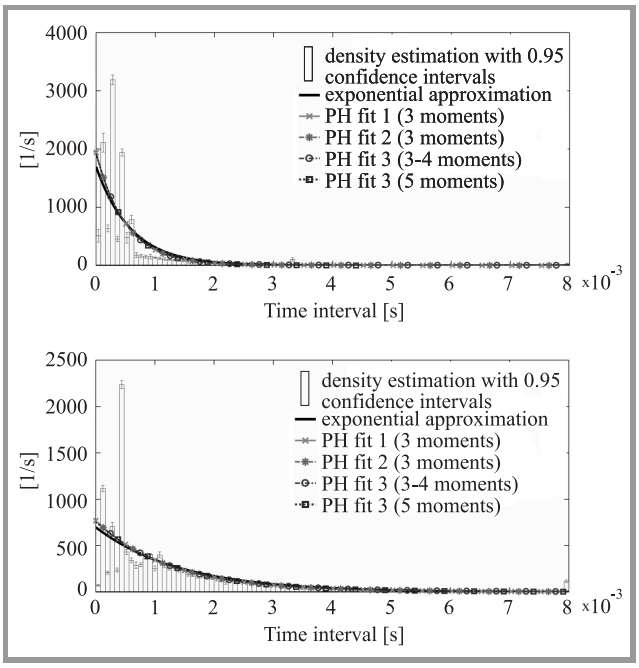


Fig. 6. Histograms of message inter-departure times at the P-CSCF (top, $c^2 = 2.16$) and S-CSCF (bottom, $c^2 = 1.36$) CPU_o (data set 1, $\lambda_{INV} = 100$ [1/s]).

from exponential distributions and depended on many parameters (in this section only conformity of the obtained histograms and their exponential approximations is discussed; all approximations will be considered and evaluated mathematically in the next section). This fact can be also confirmed by analyzing the c^2 coefficient, which for the most obtained data is far from the value 1.

Message inter-arrival and inter-departure time distributions are dependent among others on the offered load to CSCF servers CPUs. For low loads (low SIP INVITE message intensities λ_{INV}, Figs. 5–6) obtained histograms are to a certain extent similar to exponential distributions, especially at CPU$_i$ points (Fig. 5). Higher load results in more multimodal histograms with lower c^2 values, which is visible

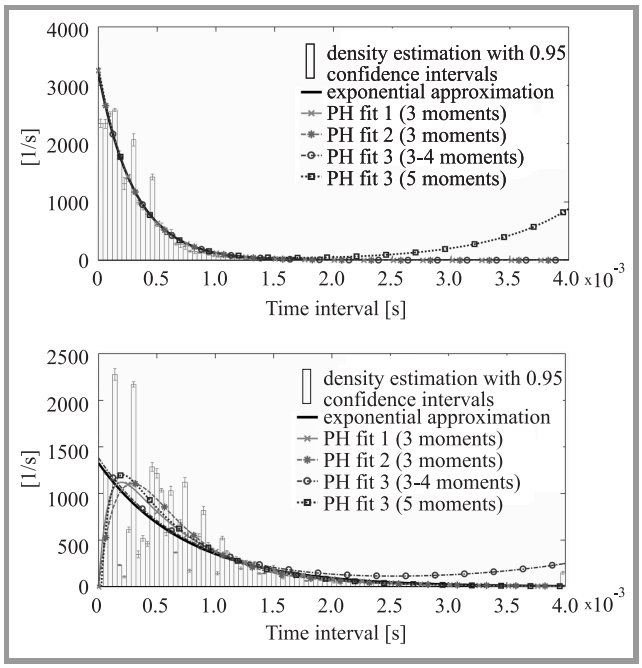


Fig. 7. Histograms of message inter-arrival times at the P-CSCF (top, $c^2 = 1.12$) and S-CSCF (bottom, $c^2 = 0.87$) CPU$_i$ (data set 1, $\lambda_{INV} = 190$ [1/s]).

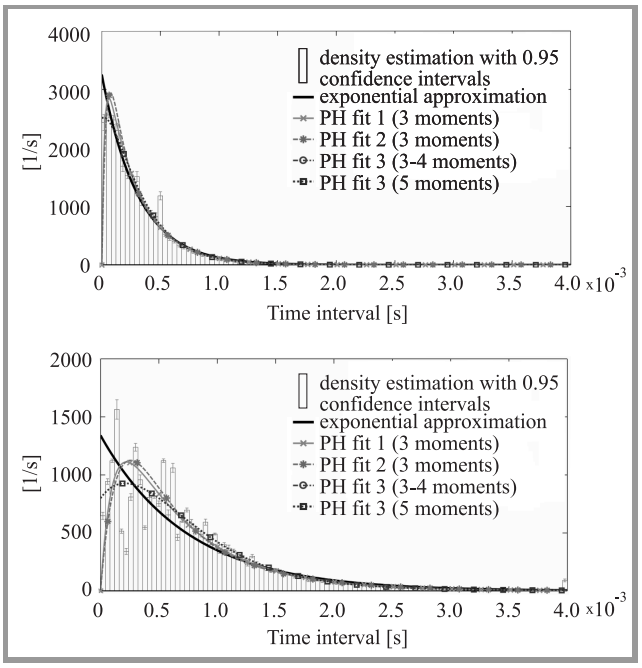


Fig. 9. Histograms of message inter-arrival times at the P-CSCF (top, $c^2 = 0.90$) and S-CSCF (bottom, $c^2 = 0.81$) CPU$_i$ (data set 3, $\lambda_{INV} = 190$ [1/s]).

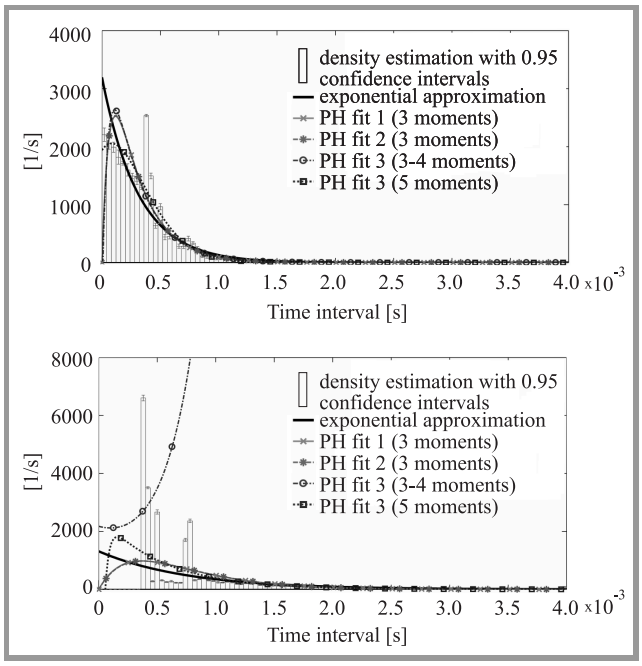


Fig. 8. Histograms of message inter-arrival times at the P-CSCF (top, $c^2 = 0.69$) and S-CSCF (bottom, $c^2 = 0.59$) CPU$_i$ (data set 2, $\lambda_{INV} = 190$ [1/s]).

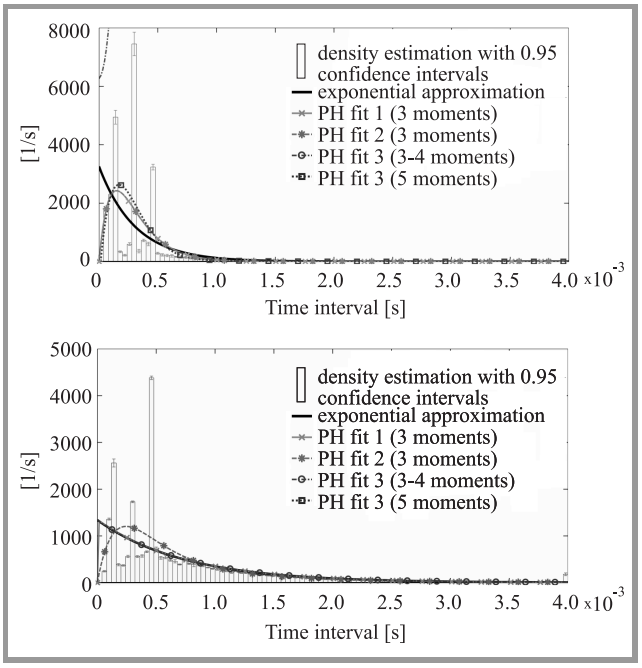


Fig. 10. Histograms of message inter-departure times at the P-CSCF (top, $c^2 = 0.58$) and S-CSCF (bottom, $c^2 = 1.00$) CPU$_o$ (data set 3, $\lambda_{INV} = 190$ [1/s]).

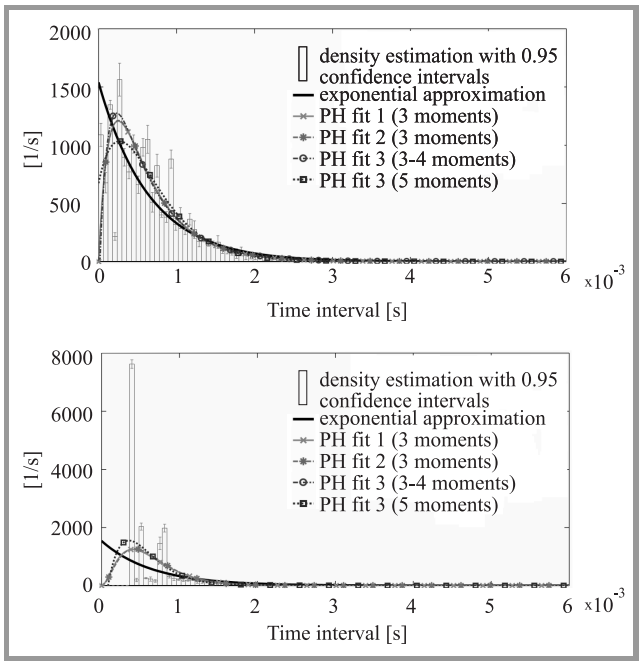


Fig. 11. Histogram of message inter-arrival times at the input of the P-CSCF→S-CSCF link (top, c^2 = 0.67) and histogram of message inter-departure times at the output of the P-CSCF→S-CSCF link (bottom, c^2 = 0.42) (data set 2, λ_{INV} = 220 [1/s]).

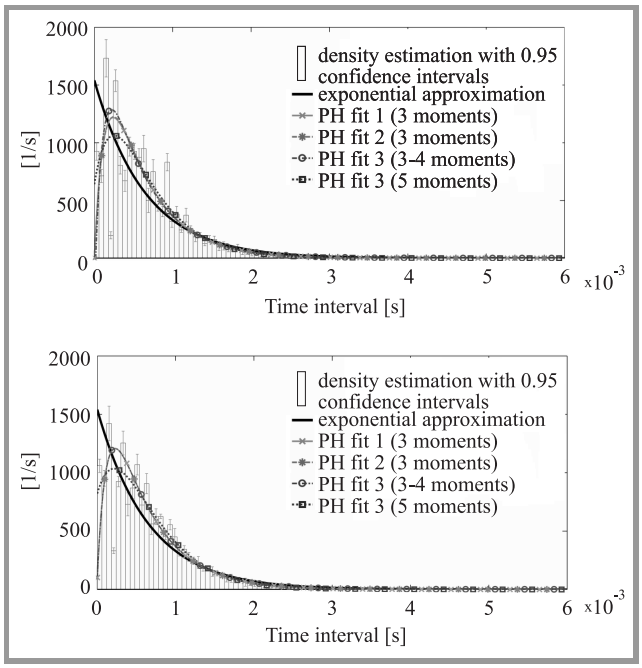


Fig. 12. Histogram of message inter-arrival times at the input of the P-CSCF→S-CSCF link (top, c^2 = 0.69) and histogram of message inter-departure times at the output of the P-CSCF→S-CSCF link (bottom, c^2 = 0.71) (data set 3, λ_{INV} = 220 [1/s]).

during analysis of the presented results (Figs. 7–10), particularly at CPU$_o$ points (Fig. 10).

Such character of message inter-departure time distributions at the outputs of CSCF servers CPUs (Fig. 10) can be explained by the fact that under high load CPU queues are in most cases nonempty and one message is processed just after handling another. Therefore, message inter-departure times are usually very close to message processing times (Eq. (1)), which are a finite set of values. This results in multimodal message inter-departure time distributions.

Due to different set of messages processed by P-CSCF and S-CSCF these two servers have, however, slightly different inter-departure time distributions at CPU$_o$ points (Fig. 10). Although in both elements these distributions are multimodal, for S-CSCF server there is in most cases one dominant peak at 0.5 ms and several smaller peaks, while for P-CSCF unit the position of the dominant peak is more dependent on the input parameters.

The investigated histograms are also influenced by the parameters of the links used to connect network elements, which involve application of communication queues (Fig. 3). This fact can be observed by comparing results presented in Fig. 7, obtained based on the assumption that all network elements are in one place and thus communication queues are not involved, to results demonstrated in Figs. 8–9, where links with particular length and bandwidth are used. In order to simplify simulations, it is assumed that all links have the same length and bandwidth.

The influence of communication links on message inter-arrival and inter-departure time distributions is dependent on the amount of offered load to the links, which is related to the available bandwidth. For low bandwidth (high offered load, Fig. 8) communication links may limit the minimal interval between the messages arriving to the CSCF server, which cannot be smaller than the time of sending the shortest message. This is also clearly visible in Fig. 11, where histograms at the input and output of the link between P-CSCF and S-CSCF are presented.

Links with relatively high available bandwidth (low offered load, Fig. 9 and 12) do not have a strong impact on message inter-arrival and inter-departure time distributions at CSCF servers. It can be noticed that for such links message inter-arrival time distributions at CPU$_i$ points are slightly closer to exponential (less multimodal, Fig. 9), even comparing to the data set 1 (Fig. 7), where all network elements are connected directly to each other. It is important that the same experiments were performed for link bandwidths of 100 Mbit/s and 1 Gbit/s, however, no visible differences were noticed. Therefore, results obtained for 1 Gbit/s link bandwidth are not presented in the paper.

4. Approximations of Obtained Histograms Using Phase-Type Distributions

This section is dedicated to approximations of the message inter-arrival and inter-departure time histograms in a single domain of IMS/NGN using phase-type distributions [12]–[16] (Figs. 5–12). This term refers to the set of probability distributions that result from a system of one or more inter-related Poisson processes occurring in sequence,

or phases. Special cases of continuous phase-type distributions are [12]–[16], [25]:

- degenerate distribution (point mass at zero or the empty phase-type distribution) – 0 phases,
- exponential distribution - 1 phase,
- Erlang distribution – 2 or more identical phases in sequence,
- deterministic distribution (or constant) – the case of an Erlang distribution with infinite number of phases,
- Coxian distribution – 2 or more phases in sequence with a probability of reaching the terminating state after each phase,
- Hyperexponential distribution (also called a mixture of exponential) – 2 or more non-identical parallel phases, each of which has its own probability of occurring,
- Hypoexponential distribution – 2 or more (not necessarily identical) phases in sequence, a generalization of an Erlang distribution (in which phases are identical).

A very important feature of the set of phase-type distributions is that it is dense in the field of all positive-valued distributions [12]–[16], [25]. Therefore, phase-type distributions can represent or approximate (with any accuracy) any positive valued distribution.
Several algorithms for fitting different subsets of phase-type distributions to experimental data with respect to specified number of first moments have been proposed [12]–[16], [25]. The following algorithms are considered in this paper (Figs. 5–12):

- PH fit 1 [12], [26] – fitting acyclic Erlang-Coxian phase-type distributions with respect to 3 moments of experimental data,
- PH fit 2 [13], [27] – fitting minimal order acyclic phase-type distributions with respect to 3 moments of experimental data,
- PH fit 3 [14], [15], [27] – fitting phase-type distributions with respect to any number of moments of experimental data; in the paper we consider two cases: 3–4 moments (resultant phase-type distributions are the same for 3 and 4 moments) as well as 5 moments.

Apart from the above mentioned algorithms, we also approximated the obtained histograms using an exponential distribution (a special case of phase-type distributions) with λ parameter taken as the inverse of mean interval between messages. All calculations were performed in the MATLAB [23] environment.
Examples of fitting phase-type distributions to the obtained histograms are presented in Figs. 5–12. As can be observed, fitted distributions are much more smoother than the message inter-arrival and inter-departure time histograms. It is also very important that the PH fit 3 algorithm is sensitive to the input data (moments of intervals between messages) and sometimes produces results which are not acceptable (i.e., PH fit 3 with respect to 3–4 moments for S-CSCF CPU_i in Fig. 8). Such results were discarded during our research. This problem does not occur for other approximations (PH fit 1, PH fit 2, exponential).
For all histograms of intervals between messages at inputs and outputs of all network elements (Fig. 1 and 4) we applied phase-type distributions fitting, which was described earlier in this section. A more extensive set of call set-up request intensities was considered comparing to Table 1 (9 values ranging from 20 to 225 [1/s]), which resulted in 2700 fitted distributions. As mentioned before, some number of the PH fit 3 results were unacceptable and had to be rejected. For all accepted results chi-square tests [28] were performed, in order to assess goodness of fitting phase-type distributions to the obtained histograms. The following test parameters were assumed:

- message inter-arrival or inter-departure times were divided into 50 bins; bins were concatenated when necessary, to guarantee theoretical (expected) frequency not less than 5,
- significance level 0.05,
- number of degrees of freedom included the number of parameters determined for each fitted phase-type distribution.

The performed chi-square tests indicated that in the majority of cases histograms and fitted phase-type distributions significantly differ from each other (with respect to the assumed significance level). Therefore, detailed test results are not presented in the paper. Hypotheses that histograms follow particular phase-type distributions were not rejected only for some distributions at very low call set-up request intensities of 20 [1/s] and only at some measurement points in the network (inputs and outputs of the P-CSCF→RACF, P-CSCF→UE2 and RACF→P-CSCF links). Only at the input and output of the UE1→P-CSCF link there was a good conformity of histograms and fitted distributions for the range of 20–160 call set-up requests per second.
In order to check which of the applied approximations is the closest to the histograms obtained in particular measurement points, the following quality measure was assumed: mean difference between chi-square test statistics and critical chi-square values for the calculated number of degrees of freedom and assumed significance level. The averaging included all results for a particular approximation and measurement point (all call set-up request intensities and data sets).
Results of these investigations are presented in Table 3. It can be noticed that generally the best results are given by the PH fit 3 algorithm (variant with 5 moments fitted).

Table 3
Quality of tested approximations (numbers of moments for particular approximartions are given in brackets)

Point of measurement	Best approximation	...	...	...	Worst approximation
P-CSCF CPU$_i$	PH fit 1 (3)	PH fit 3 (3–4)	PH fit 2 (3)	Exponential	PH fit 3 (5)
P-CSCF CPU$_o$	PH fit 2 (3)	PH fit 1 (3)	PH fit 3 (5)	Exponential	PH fit 3 (3–4)
P-CSCF→RACF link in	PH fit 2 (3)	PH fit 1 (3)	Exponential	PH fit 3 (5)	PH fit 3 (3–4)
P-CSCF→RACF link out	PH fit 2 (3)	PH fit 1 (3)	Exponential	PH fit 3 (5)	PH fit 3 (3–4)
P-CSCF→S-CSCF link in	PH fit 1 (3)	PH fit 3 (5)	PH fit 2 (3)	Exponential	PH fit 3 (3–4)
P-CSCF→S-CSCF link out	PH fit 3 (5)	PH fit 2 (3)	PH fit 1 (3)	Exponential	PH fit 3 (3–4)
P-CSCF→UE1 link in	PH fit 3 (5)	PH fit 1 (3)	PH fit 2 (3)	PH fit 3 (3–4)	Exponential
P-CSCF→UE1 link out	PH fit 3 (3–4)	PH fit 3 (5)	PH fit 2 (3)	PH fit 1 (3)	Exponential
P-CSCF→UE2 link in	PH fit 1 (3)	PH fit 2 (3)	Exponential	PH fit 3 (5)	PH fit 3 (3–4)
P-CSCF→UE2 link out	PH fit 1 (3)	PH fit 2 (3)	Exponential	PH fit 3 (3–4)	PH fit 3 (5)
RACF→P-CSCF link in	PH fit 2 (3)	PH fit 1 (3)	Exponential	PH fit 3 (5)	PH fit 3 (3–4)
RACF→P-CSCF link out	PH fit 2 (3)	PH fit 1 (3)	Exponential	PH fit 3 (5)	PH fit 3 (3–4)
S-CSCF CPU$_i$	PH fit 3 (5)	PH fit 2 (3)	PH fit 1 (3)	Exponential	PH fit 3 (3–4)
S-CSCF CPU$_o$	PH fit 3 (5)	PH fit 2 (3)	PH fit 1 (3)	Exponential	PH fit 3 (3–4)
S-CSCF→P-CSCF link in	PH fit 3 (5)	PH fit 2 (3)	PH fit 1 (3)	Exponential	PH fit 3 (3–4)
S-CSCF→P-CSCF link out	PH fit 3 (5)	PH fit 2 (3)	PH fit 1 (3)	Exponential	PH fit 3 (3–4)
UE1→P-CSCF link in	PH fit 1 (3)	Exponential	PH fit 2 (3)	PH fit 3 (3–4)	PH fit 3 (5)
UE1→P-CSCF link out	PH fit 2 (3)	PH fit 1 (3)	Exponential	PH fit 3 (5)	PH fit 3 (3–4)
UE2→P-CSCF link in	PH fit 3 (5)	Exponential	PH fit 1 (3)	PH fit 2 (3)	PH fit 3 (3–4)
UE2→P-CSCF link out	PH fit 3 (5)	PH fit 1 (3)	Exponential	PH fit 2 (3)	PH fit 3 (3–4)

Only slightly worse are the PH fit 1 and PH fit 2 algorithms, which operate on 3 moments of experimental data. Exponential approximations of the obtained histograms offer even poorer quality, which is, however, better than that of the PH fit 3 algorithm (3 moments).

5. Conclusions and Future Work

The aim of the work presented in the paper was to investigate message inter-arrival and inter-departure time distributions in a single domain of IMS/NGN architecture. The investigations also concerned a possibility of approximation of the above mentioned distributions using phase-type distributions. For these reasons the developed simulation model, which conforms to the latest standards and research, was used to gather data in different points of the network. Obtained results indicate that message inter-arrival and inter-departure time distributions in the system are influenced by many parameters (including offered load to CSCF servers and links) and are not exponential. This is especially visible for higher loads, for which histograms of time intervals between messages are multimodal, particularly at the output of CSCF servers CPUs.

To all obtained inter-arrival and inter-departure time histograms several phase-type distributions were fitted using available algorithms. Performed chi-square tests demonstrated that the acquired histograms in most cases do not follow the fitted distributions. Therefore, a metric, to examine which distributions are the most similar to the histograms was proposed. The best phase-type approximations are generally PH fit 3, based on 5 moments of intervals between messages. Only slightly worse results can be achieved by applying the PH fit 1 and PH fit 2 algorithms using 3 moments of experimental data.

The research described in this paper demonstrates that the problem of constructing an accurate analytical model of IMS/NGN is very complicated. According to the presented results, the obtained message inter-arrival as well as inter-departure time distributions and their exponential approximations generally significantly differ from each other. However, our experience indicates that in many cases using simple M/G/1 queues (with exponential inter-arrival times) to describe the operation of IMS/NGN servers and links gives satisfactory results when we consider the response of the whole system. Call processing performance (mean Call Set-up Delay as well as mean Call Disengagement Delay) results achieved with M/G/1 queues are very close to simulations and are comparable to the results obtained using commonly known G/G/1 approximations [10], [29].

Taking all mentioned facts into consideration, the authors are going to continue work on determination of proper queuing models for CSCF servers CPUs and optical links, in order to achieve better conformity of calculations and simulations for the whole investigated IMS/NGN architec-

ture. Apart from that, development of the traffic model is being planned in order to carry out research in a multi-domain IMS/NGN architecture, including also the elements specific for MPLS, Ethernet and FSA transport technologies [30]–[32].

Acknowledgements

This research work was partially supported by the system project "InnoDoktorant – Scholarships for PhD students, Vth edition" co-financed by the European Union in the frame of the European Social Fund.

References

[1] "General overview of NGN", ITU-T Rec. Y.2001, Dec. 2004.

[2] "IP Multimedia Subsystem (IMS); Stage 2 (Release 11)", 3GPP TS 23.228 v11.0.0, Mar. 2011.

[3] J. Rosenberg *et al.*, "SIP: Session Initiation Protocol", IETF RFC 3261, Jun. 2002.

[4] P. Calhoun *et al.*, "Diameter Base Protocol", IETF RFC 3588, Sept. 2003.

[5] "Call processing performance for voice service in hybrid IP networks", ITU-T Rec. Y.1530, Nov. 2007.

[6] "SIP-based call processing performance", ITU-T Rec. Y.1531, Nov. 2007.

[7] S. Kaczmarek and M. Sac, "Traffic modeling in IMS-based NGN networks", *Gdańsk University of Technology Faculty of ETI Annals*, vol. 1, no 9, pp. 457–464, 2011.

[8] S. Kaczmarek and M. Sac, "Zagadnienia inżynierii ruchu w sieciach NGN bazujących na IMS" ("Traffic engineering aspects in IMS-based NGN networks"), in *Biblioteka teleinformatyczna, t. 6. Internet 2011* (*Teleinformatics library, vol. 6. Internet 2011*), D. J. Bem *et al.*, Eds. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 2012, pp. 63–115 (in Polish).

[9] S. Kaczmarek, M. Kaszuba and M. Sac, "Simulation model of IMS/NGN call processing performance", *Gdańsk University of Technology Faculty of ETI Annals*, vol. 20, pp. 25–36, 2012.

[10] S. Kaczmarek and M. Sac, "Traffic Model for Evaluation of Call Processing Performance Parameters in IMS-based NGN", in *Information Systems Architecture and Technology: Networks Design and Analysis*, A. Grzech *et al.*, Eds. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 2012, pp. 85–100.

[11] S. Kaczmarek and M. Sac, "Message Inter-Arrival and Inter-Departure Time Distributions in IMS/NGN Architecture", in *Proc. 17th Polish Teletraffic Symp. PTS 2012*, Zakopane, Poland, 2012, pp. 37–43.

[12] T. Osogami and M. Harchol-Balter, "Closed form solutions for mapping general distributions to quasi-minimal PH distributions", *Perform. Eval.*, vol. 63, no. 6, pp. 524–55, 2006.

[13] A. Bobbio, A. Horvath and M. Telek, "Matching three moments with minimal acyclic phase type distributions", *Stoch. Mod.*, vol. 21, no. 2–3, pp. 303–326, 2005.

[14] M. Telek and G. Horvath, "A minimal representation of Markov arrival processes and a moments matching method", *Perform. Eval.*, vol. 64, no. 9–12, pp. 1153–1168, 2007.

[15] A. van de Liefvoort, "The moment problem for continuous distributions", Tech. rep., University of Missouri, WP-CM-1990-02, Kansas City, USA, 1990.

[16] S. Asmussen, O. Nerman and M. Olsson, "Fitting Phase-type distributions via the EM Algorithm", *Scandinavian J. Statist.*, vol. 23, no. 4, pp. 419–441, 1996.

[17] "Functional requirements and architecture of next generation networks", ITU-T Rec. Y.2012, Apr. 2010.

[18] "IMS for next generation networks", ITU-T Rec. Y.2021, Sept. 2006.

[19] "Resource and admission control functions in next generation networks", ITU-T Rec. Y.2111, Nov. 2008.

[20] "Resource control protocol no. 1, version 2 – Protocol at the Rs interface between service control entities and the policy decision physical entity", ITU-T Rec. Q.3301.1, Jun. 2010.

[21] M. Pirhadi, S. M. Safavi Hemami and A. Khademzadeh, "Resource and admission control architecture and QoS signaling scenarios in next generation networks", *World Appl. Sci. J. 7 (Special Issue of Computer & IT)*, pp. 87–97, 2009.

[22] OMNeT++ Network Simulation Framework [Online]. Available: http://www.omnetpp.org

[23] MATLAB – The Language of Technical Computing [Online]. Available: http://www.mathworks.com/products/matlab

[24] V. S. Abhayawardhana and R. Babbage, "A traffic model for the IP Multimedia Subsystem (IMS)", in *Proc. IEEE 65th Veh. Technol. Conf. VTC 2007-Spring*, Dublin, Ireland, 2007.

[25] T. Czachórski, "Modele kolejkowe w ocenie efektywności sieci i systemów komputerowych" ("Queuing models in evaluation of effectiveness of computer networks and systems"). Gliwice: Pracownia Komputerowa Jacka Skalmierskiego, 1999 (in Polish).

[26] Moment Matching Algorithms [Online]. Available: http://www.cs.cmu.edu/ osogami/code/momentmatching/index.html

[27] BuTools Program Packages [Online]. Available: http://webspn.hit.bme.hu/ telek/tools/butools/butools.html

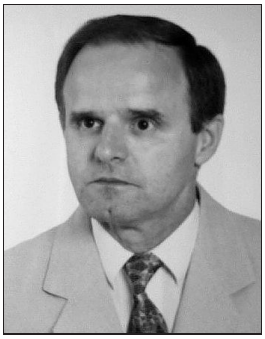
[28] G. W. Corder and D. I. Foreman, *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*. Wiley, 2009.

[29] S. Kaczmarek and M. Sac, "Analysis of IMS/NGN call processing performance using G/G/1 queuing systems approximations", *Przegl. Telekomun. i Wiadom. Telekomun.* (*Telecommun. Rev. & Telecommun. News*), no. 8–9, pp. 702–710, 2013.

[30] "Centralized RACF architecture for MPLS core networks", ITU-T Rec. Y.2175, Nov. 2008.

[31] "Ethernet QoS control for next generation networks", ITU-T Rec. Y.2113, Jan. 2009.

[32] "Requirements for the support of flow state aware transport technology in an NGN", ITU-T Rec. Y.2121, Jan. 2008.



Sylwester Kaczmarek received his M.Sc. in Electronics Engineering, Ph.D. and D.Sc. in Switching and Teletraffic Science from the Gdańsk University of Technology, Gdańsk, Poland, in 1972, 1981 and 1994, respectively. His research interests include: IP QoS and GMPLS networks, switching, QoS routing, teletraffic, multimedia services and quality of services. Currently, his research is focused on developing and applicability of VoIP and IMS/NGN technology. So far he has published more than 200 papers. Now he is the Head of Teleinformation Networks Department at GUT.

E-mail: kasyl@eti.pg.gda.pl
Department of Teleinformation Networks
Faculty of Electronics, Telecommunications and Informatics
Gdańsk University of Technology
Gabriela Narutowicza st 11/12
80-233 Gdańsk, Poland

Maciej Sac received his M.Sc. degree in Telecommunications from Gdańsk University of Technology in 2009. Since 2009 he has been a Ph.D. student at Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics. His research interests are focused on ensuring reliability and Quality of Service in IMS/NGN networks by the means of traffic engineering.

E-mail: Maciej.Sac@eti.pg.gda.pl
Department of Teleinformation Networks
Faculty of Electronics, Telecommunications
and Informatics
Gdańsk University of Technology
Gabriela Narutowicza st 11/12
80-233 Gdańsk, Poland