

Analysis of IMS/NGN Call Processing Performance Using Phase-Type Distributions Based on Experimental Histograms

Sylwester Kaczmarek^(✉) and Maciej Sac

Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, 11/12 Narutowicza Street, 80-233 Gdańsk, Poland
{kasy1, msac}@eti.pg.gda.pl

Abstract. The paper describes our further research done with the proposed analytical and simulation traffic models of the Next Generation Network (NGN), which is standardized for delivering multimedia services with strict quality and includes elements of the IP Multimedia Subsystem (IMS). The aim of our models of a single IMS/NGN domain is to evaluate two standardized call processing performance parameters, which appropriate values are very important for satisfaction of users and overall success of the IMS/NGN concept. These parameters are mean Call Set-up Delay $E(CSD)$ and mean Call Disengagement Delay $E(CDD)$. Our latest investigations concern improving the conformity of the analytical results and experimental results obtained using the simulation model, which implements the operation of real network elements according to current standards and research. In this paper the results of calculations using PH/PH/1 queuing systems are presented, in which arrival and service distributions are phase-type distributions computed using maximum likelihood and distance minimization fitting algorithms based on experimental histograms. Presented latest results are compared to these obtained using other, previously investigated queuing systems. Additionally, computational complexity of all examined queuing systems is analyzed. As a result, some general remarks concerning all tested queuing systems and their applicability to NGN are provided.

Keywords: IMS · NGN · Call processing performance · Phase-type distributions · PH/PH/1 · Traffic model

1 Introduction

Our research concerns the Next Generation Network (NGN) [1], which is a standardized network architecture based on the IP Multimedia Subsystem (IMS) [2] (hence the names “IMS-based NGN” and “IMS/NGN” are commonly used), proposed for fulfilling the needs of current and future information society concerning delivering multimedia services. For a commercial success this new concept of telecommunication network should be properly designed, to guarantee the values of quality parameters, which are satisfactory for users (e.g. standardized call processing performance parameters [3, 4], which include mean Call Set-up Delay $E(CSD)$ mean Call

Disengagement Delay $E(CDD)$). To achieve this aim, proposition and application of appropriate traffic models is necessary to investigate the relations between network parameters and guaranteed quality parameters.

Performed review indicated that research on these aspects is in early phase and the majority of available traffic models do not include the operation of all standardized IMS/NGN elements (such as RACF, Resource and Admission Control Functions, responsible for resource control) or service scenarios and do not allow assessment of call processing performance parameters. The fact that IMS/NGN is now in the implementation stage and the problem of ensuring quality becomes more and more important led us to start investigations in this area.

In the paper we continue our research concerning the formerly proposed simulation [5] and analytical [6] traffic model of a single domain of IMS/NGN, which allow evaluation of $E(CSD)$ and $E(CDD)$. Our latest investigations focus on finding the best queuing system for the analytical model to achieve the best conformity with the experimental results provided by the simulation model. In the simulation model not theoretical queuing systems but the operation of real network elements and standardized call scenarios are accurately implemented (the simulated elements process SIP and Diameter messages according to standards). Consequently, the simulation model reflects the phenomena taking place in real IMS/NGN network and can be considered as a reference for evaluation of quality of the analytical results. In the previous stages of our research we examined the following queuing systems in our analytical model (as IMS/NGN elements are servers with lots of memory available for message queues and in all examined cases high queue lengths are very improbable – due to constraints on $E(CSD)$ and $E(CDD)$ – the analysis of the system can be readily done using queuing models with infinite waiting room capacity):

- M/G/1 based on two moments of service distribution – used in our first approach [6] although were aware of the fact that intervals between messages at the inputs of IMS/NGN elements are generally not exponential; the results were under many conditions acceptable but poor conformity of calculations and simulations could be observed under high load and also when IMS/NGN elements were connected using links with relatively low bandwidth,
- G/G/1 based on two or three moments of arrival distribution and two moments of service distribution [7] – used in the next step of our research; they did not significantly improve the results,
- PH/PH/1 with arrival and service distributions represented by phase-type distributions [8–11] obtained using moment matching algorithms [12]; good conformity between analytical and simulation results were observable but we still did not find one universal queuing system for all considered data sets.

The completion of the above mentioned research, introduced in this paper, are investigations with PH/PH/1 queuing systems obtained using maximum likelihood and distance minimization fitting algorithms, which base on whole experimental histograms (not only on their moments) and due to this fact can lead to better results than for the previously examined queuing systems. In this paper the results for PH/PH/1 queuing systems obtained using maximum likelihood and distance minimization fitting algorithms are compared to the best results achieved in the previous stages. Moreover, we

present the analysis of computational complexity for the queuing systems used in all stages of our research (M/G/1, G/G/1, PH/PH/1 with all investigated fitting algorithms), which has not been evaluated before. The rest of the text is organized as follows. Basic details about the proposed IMS/NGN traffic models are presented in Sect. 2. Elementary information about phase-type distributions, algorithms for fitting this type of distributions to arrival and service distributions in IMS/NGN and analyzing PH/PH/1 queuing systems is provided in Sect. 3. The results of the performed investigations are presented and discussed in Sect. 4. Summary and future work are described in Sect. 5.

2 Traffic Model of IMS/NGN

Our research concerns the ITU-T NGN architecture, which is the most advanced of all available NGN solutions [13]. Based on current standards and scientific work we have proposed an analytical [6] and simulation [5] model of a single ITU-T NGN domain, which contain the elements depicted in Fig. 1, performing the following roles [6, 7, 12]:

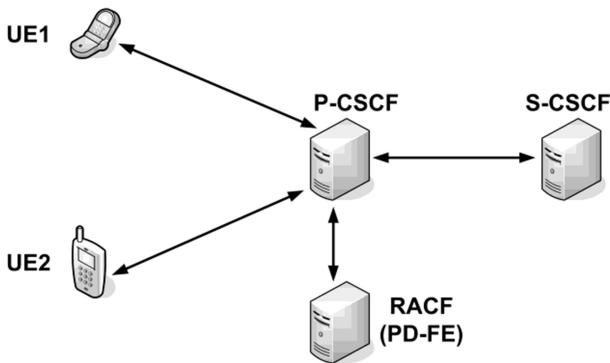


Fig. 1. Model of a single domain of IMS/NGN [6, 14, 15].

- User Equipments (UEs): user terminals generating call set-up and disengagement requests, registered in the domain,
- P-CSCF (Proxy – Call Session Control Function): the server responsible for receiving all messages from UEs and forwarding them to the S-CSCF server,
- S-CSCF (Serving – Call Session Control Function): the main server handling all calls,
- RACF (Resource and Admission Control Functions): the unit of the transport stratum, which allocates resources for a new call and releases resources during call disengagement.

The above mentioned network elements participate in a standardized voice call scenario [14–18] depicted in Fig. 2. For resource operations (allocation, release) Diameter protocol [19] messages are exchanged between P-CSCF and RACF. Other

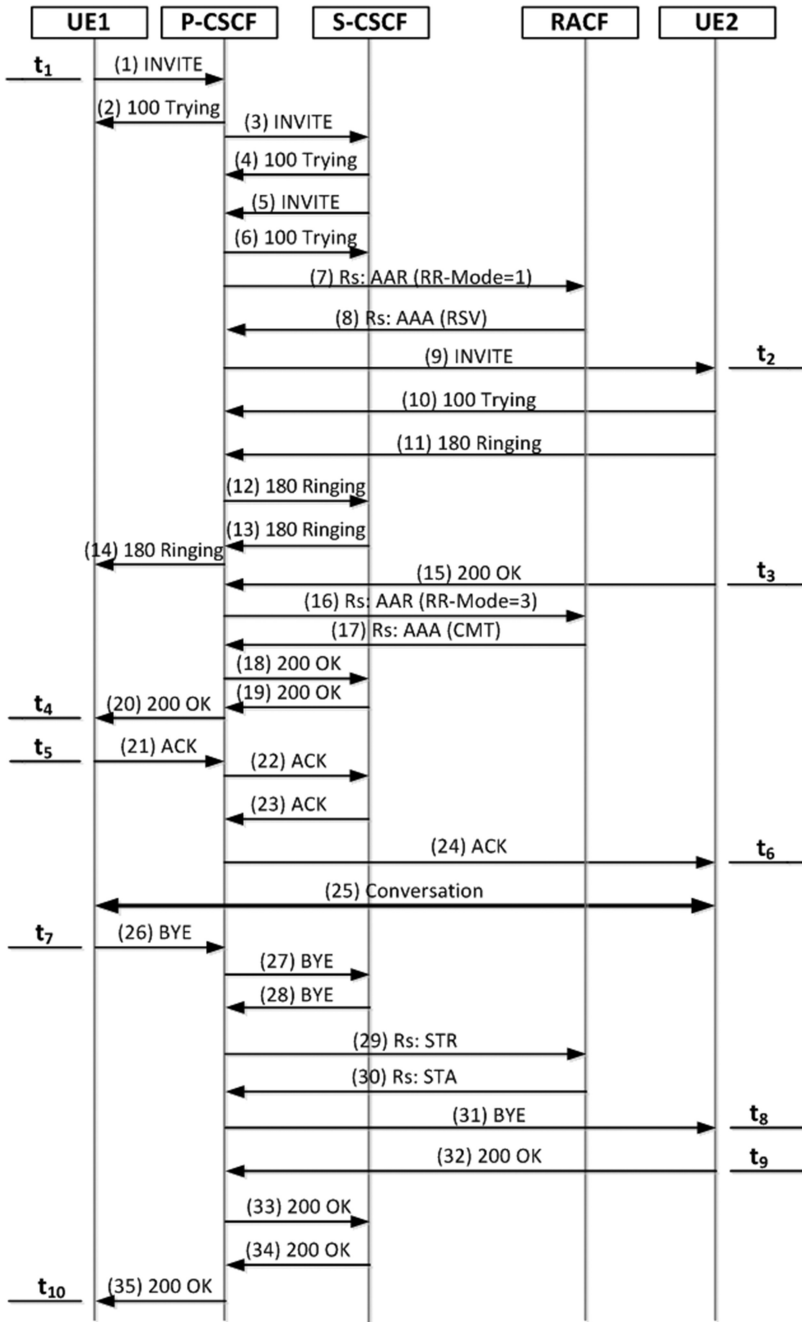


Fig. 2. Call set-up (messages 1–24) and call disengagement (messages 26–35) scenario in a single domain of IMS/NGN [6].

elements communicate using SIP protocol [20]. More details regarding the call scenario are out of the scope of this paper and can be found in [6].

Based on the t_1-t_{10} times illustrated in Fig. 2 and the definitions given by the ITU-T [3, 4], Call Set-up Delay (*CSD*) and Call Disengagement Delay (*CDD*) for a single call can be calculated as follows:

$$CSD = (t_2-t_1) + (t_4-t_3) + (t_6-t_5) \quad (1)$$

$$CDD = (t_8-t_7) + (t_{10}-t_9) \quad (2)$$

Formulas (1) and (2) are used to obtain the output variables evaluated by our traffic models, which are mean *CSD*, $E(CSD)$, and mean *CDD*, $E(CDD)$ [3, 4]. The analytical and simulation models represent different approaches to assess $E(CSD)$ and $E(CDD)$ (more details will be provided later). However, both of them use a similar set of the input variables, from which the most vital are [5, 6]:

- λ_{INV} : the intensity of aggregated call set-up requests (SIP INVITE messages from many terminals) generated by UE1 with exponential intervals,
- T_{INV1} and T_{INV2} : the times of processing SIP INVITE messages by P-CSCF and S-CSCF correspondingly,
- a_k ($k = 1, 2, \dots, 8$): the factors determining times of processing other SIP and Diameter messages by CSCF servers according to Table 1,
- T_X : the time of processing messages by RACF,
- lengths and bandwidths of optical links, lengths of transmitted messages: values necessary to calculate communication times between network elements.

Table 1. Message processing times of CSCF servers.

Message	P-CSCF processing times	S-CSCF processing times
SIP INVITE	T_{INV1}	T_{INV2}
SIP 100 Trying	$T_{TR1} = a_1 \cdot T_{INV1}$	$T_{TR2} = a_1 \cdot T_{INV2}$
SIP 180 Ringing	$T_{RING1} = a_2 \cdot T_{INV1}$	$T_{RING2} = a_2 \cdot T_{INV2}$
SIP 200 OK (answer to INVITE)	$T_{OK1} = a_3 \cdot T_{INV1}$	$T_{OK2} = a_3 \cdot T_{INV2}$
SIP ACK	$T_{ACK1} = a_4 \cdot T_{INV1}$	$T_{ACK2} = a_4 \cdot T_{INV2}$
SIP BYE	$T_{BYE1} = a_5 \cdot T_{INV1}$	$T_{BYE2} = a_5 \cdot T_{INV2}$
SIP 200 OK (answer to BYE)	$T_{OKBYE1} = a_6 \cdot T_{INV1}$	$T_{OKBYE2} = a_6 \cdot T_{INV2}$
Diameter AAA	$T_{AAA1} = a_7 \cdot T_{INV1}$	$T_{AAA2} = a_7 \cdot T_{INV2}$
Diameter STA	$T_{STA1} = a_8 \cdot T_{INV1}$	$T_{STA2} = a_8 \cdot T_{INV2}$

The simulation model [5, 6], developed in the OMNeT++ framework [21], implements the operation of the network elements and call scenario depicted in Figs. 1 and 2. CSCF servers in the model include Central Processing Units (CPUs), which are responsible for handling messages incoming from 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 links when links are busy. During simulation the t_I-t_{I0} times (Fig. 2) are stored for all performed call scenarios and used to obtain *CSD* (1) and *CDD* (2). At the end of simulation $E(CSD)$ and $E(CDD)$ are calculated along with corresponding confidence intervals for a defined confidence level [5, 6].

In the analytical model mean values of theoretical delays introduced by the elements depicted in Fig. 1 are calculated and appropriately summed to obtain mean *CSD* and mean *CDD* [6, 7, 12]. These component delays include:

- mean message waiting times in communication queues; for calculation of these times mathematical models of queuing systems are used,
- message transmission times (message lengths divided by links bandwidths),
- propagation times (equal to 5 $\mu\text{s}/\text{km}$ – optical links are assumed),
- mean message waiting times in CSCF servers CPU queues; for calculation of these times mathematical models of queuing systems are used,
- message processing times by CSCF servers CPUs (Table 1) and RACF (T_X input variable).

Due to the fact that the simulation model implements real IMS/NGN elements and call scenario, it reflects the phenomena occurring in the real network (message transmission, queuing, handling, etc.) and can be used to assess the quality of the analytical model based on theoretical calculations. More details about both models cannot be provided in this paper due to lack of space and can be found in [5–7, 12].

3 Phase-Type Distributions

Phase-type distributions [8–11] are 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 [8–11]: exponential distribution, Erlang distribution, Coxian distribution, Hyperexponential distribution (also called a mixture of exponential) and Hypoexponential distribution.

Our interest in phase-type distributions is related to their very important feature. As the set of phase-type distributions is dense in the field of all positive-valued distributions [8–11], they can represent or approximate (with any accuracy) any positive valued distribution. There are several algorithms for fitting different subsets of phase-type distributions to experimental data based on a specified number of first moments [8–11] or whole experimental histograms [11, 22–24].

Using such algorithms phase-type distributions can be fitted to all arrival and service distributions in the network [25]. After that, some characteristics of queues, like mean message waiting times (required in our research), can be calculated through analysis of PH/PH/1 queuing systems. Such analysis can be done by solving quasi-birth-and-death (QBD) Markov chains using matrix-analytic methods [26, 27]. Alternatively, formulas for analyzing queues with MAP (Markovian Arrival Process) or BMAP (Batch Markovian Arrival Process) distributions [28, 29] can be used as

these types of distributions are supersets of phase-type distributions. Several solvers are available and we have tested two [30, 31] designed for the MATLAB environment used in our research. Both of the tested solvers provided very similar results, however the second one [31] was chosen for further use due to faster and direct calculation of mean waiting time [12].

As already mentioned, in this paper we perform investigations with PH/PH/1 queuing systems, in which phase-type distributions are fitted using maximum likelihood and distance minimization algorithms based on whole experimental histograms. The set of investigated algorithms included (the first three belong to the family of maximum likelihood algorithms, while the last two are distance minimization algorithms):

- EMpht [11] – fitting general phase-type distributions using Expectation-Maximisation (EM) algorithm; in our research distributions of order 4 were fitted with the number of iterations equal to 6000; due to large computation times the number of samples of each experimental histogram was limited to 100000,
- GFIT1 [22] – fitting Hyper-Erlang phase-type distributions using Expectation-Maximisation (EM) algorithm; the order of fitted distributions was set to 4; the number of samples of each experimental histogram was limited to 1000000,
- GFIT2 [22] – the same algorithm as in GFIT1, however, here distributions of order 10 were fitted and the number of samples of each experimental histogram was limited to 700000,
- PhFit1 [24] – fitting acyclic phase-type distributions using distance minimization algorithm; distributions of order 4 were fitted with total number of iterations equal to 3000 (3 rounds with 1000 iterations); the investigated distance metric was relative entropy; the number of samples of each experimental histogram was limited to 200000,
- PhFit3 [24] – the same algorithm as in PhFit1, however, here distributions of order 10 were fitted and the number of samples of each experimental histogram was limited to 100000.

The number of iterations was chosen experimentally for each of the algorithms according to its convergence speed.

To apply the above mentioned fitting algorithms in the analytical model, a full description of arrival and service distributions for all IMS/NGN elements was necessary. This was not a problem for service distributions as they result from times of processing individual messages by network elements (defined by the input variables; for links they are determined by message lengths and link bandwidths) and the set of messages handled by each element (given by the assumed call scenario).

When it comes to arrival distributions, only theoretical intensity of messages handled by all networks elements is known from the call scenario and λ_{INV} . Therefore, in this case phase-type distributions were fitted to experimental histograms recorded using the simulation model [25]. The results of the performed investigations are described in the next section.

4 Results

In order to compare the latest $E(CSD)$ and $E(CDD)$ results with these obtained in our previous papers [6, 7, 12], the same input data sets (presented in Table 2) were used. We also assumed identical message lengths (Table 3) as well as the a_k factors: $a_1 = 0.2, a_2 = 0.2, a_3 = 0.6, a_4 = 0.3, a_5 = 0.6, a_6 = 0.3, a_7 = 0.6, a_8 = 0.6$ (Table 1).

Table 2. Input data sets.

Data set	λ_{INV} [1/s]	T_{INV1} [ms]	T_{INV2} [ms]	T_X [ms]	Links
1a	5–250	0.5	0.5	3	No links
1b	5–250	0.5	0.5	3	300 km, 10 Mbit/s
1c	5–250	0.5	0.5	3	300 km, 100 Mbit/s

Table 3. Message lengths [32].

Message	Length in 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

The values of the input variables were taken according to current research (e.g. message lengths, which were measured experimentally in paper [32]) and our experience regarding telecommunication networks (e.g. the length of 300 km assumed for all links represents a typical distance between larger cities in Poland). Due to the fact that in our traffic models quite a large number of the input variables is used (Sect. 2), in the research presented in this paper we focused on call set-up request intensity, which is a typical and very commonly investigated parameter for telecommunication networks. Additionally, we examined different parameters of links (length and bandwidth), which have a significant impact on the final results.

As already mentioned in Sect. 3, in the research described in this paper phase-type distributions were fitted to message arrival and service distributions for all network elements, using maximum likelihood and distance minimization algorithms. In the next step mean message waiting times were calculated through analysis of PH/PH/1 queuing systems. The resultant mean waiting times were applied to compute the considered total system responses – $E(CSD)$ and $E(CDD)$.

Apart from fitting phase-type distributions to both arrival and service distributions of all elements (PH/PH/1 queuing systems), we also assumed that either arrival or service distributions are exponential (special cases of phase-type distributions), which resulted in M/PH/1 and PH/M/1 queuing systems correspondingly.

The above mentioned computations required full description of arrival and service distributions for all network elements. As arrival distributions are not fully described by the input variables (Sect. 3), they were obtained experimentally, using our simulation model. The simulations were performed with the following parameters [7, 12]:

- warm-up period: 1500 s,
- 5 measurement periods,
- 0.95 confidence level,
- 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 times were obtained during each simulation.

The results obtained for data sets 1a–1c (Table 2) are presented in Figs. 3, 4 and 5 as $E(CSD)$ and $E(CDD)$ versus call set-up request intensity curves. In each subfigure (Figs. 3, 4 and 5 contain six subfigures) we present results of simulations, calculations

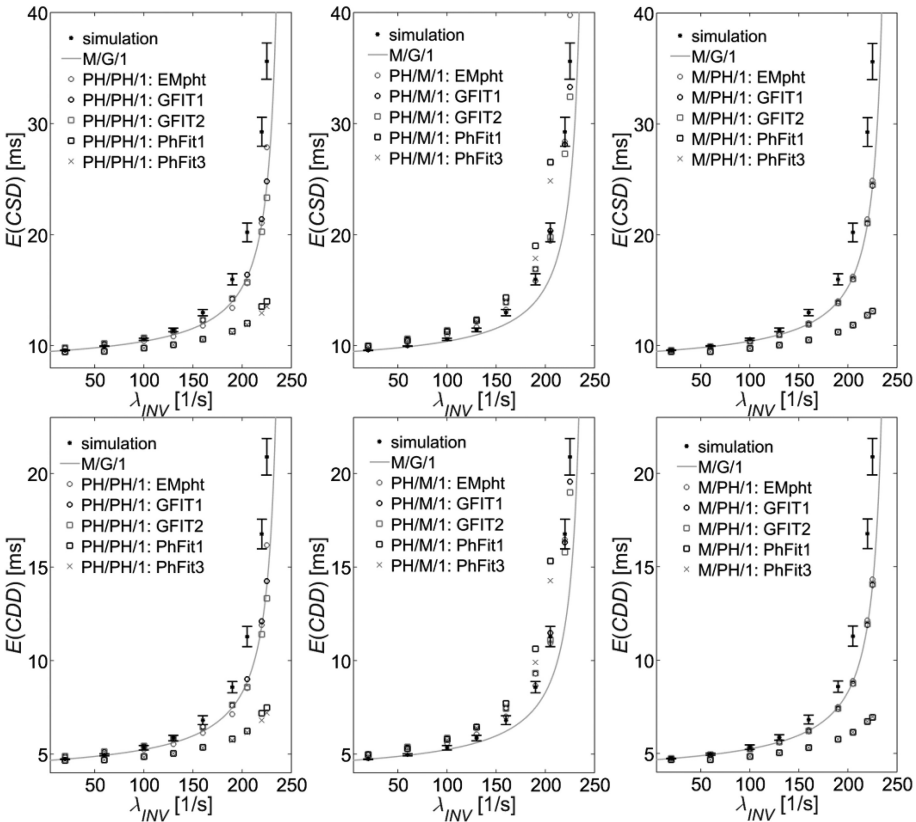


Fig. 3. Results for data set 1a (no links); left: PH/PH/1 queuing systems, center: PH/M/1 queuing systems, right: M/PH/1 queuing systems; results at the top refer to $E(CSD)$ and at the bottom – to $E(CDD)$.

using M/G/1 queuing systems (our initial approach [6]) and analytical approaches based on phase type distributions (PH/PH/1, PH/M/1, M/PH/1) obtained using maximum likelihood and distance minimization algorithms (Sect. 3). The names of the fitting algorithms are included in legends (after colons).

Generally, it can be noticed that conformity of calculations and simulations of call processing performance parameters is dependent on many factors, including queuing systems used in the analytical model (Figs. 3, 4 and 5). Very important is also call set-up request intensity as well as available link bandwidth (links involve communication queues, which load is related to message intensity and link bandwidth; in data set 1a network elements are connected directly without links, thus, there are no communication queues).

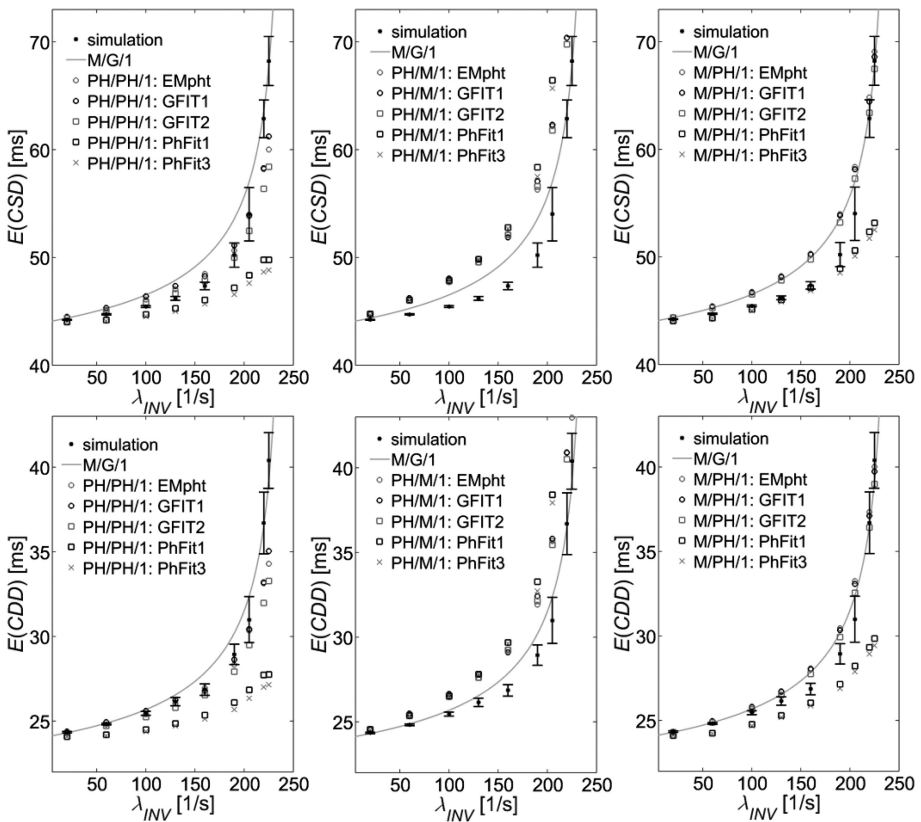


Fig. 4. Results for data set 1b (link length 300 km and bandwidth 10 Mbit/s); left: PH/PH/1 queuing systems, center: PH/M/1 queuing systems, right: M/PH/1 queuing systems; results at the top refer to $E(CSD)$ and at the bottom – to $E(CDD)$.

We can observe that the relations between analytical and simulation results are similar for $E(CSD)$ and $E(CDD)$, when the same queuing systems are investigated. However, $E(CSD)$ values are always larger than $E(CDD)$, since the process of call set-up is much more complicated, comparing to call disengagement, and involves more messages. For each analyzed data set (Table 2) PhFit1 and PhFit3 algorithms give analytical results very far from simulations and other calculations. This rule applies especially for PH/PH/1 and M/PH/1 queuing systems with PhFit1 and PhFit3 algorithms. As a result, these algorithms are not useful in practice.

Based on Figs. 3, 4 and 5, it can be also noticed that in the majority of analyzes cases PH/M/1 queuing systems overestimate $E(CSD)$ and $E(CDD)$, while M/PH/1 queuing systems give results very similar to M/G/1 (this does not apply to PhFit1 and PhFit3 algorithms).

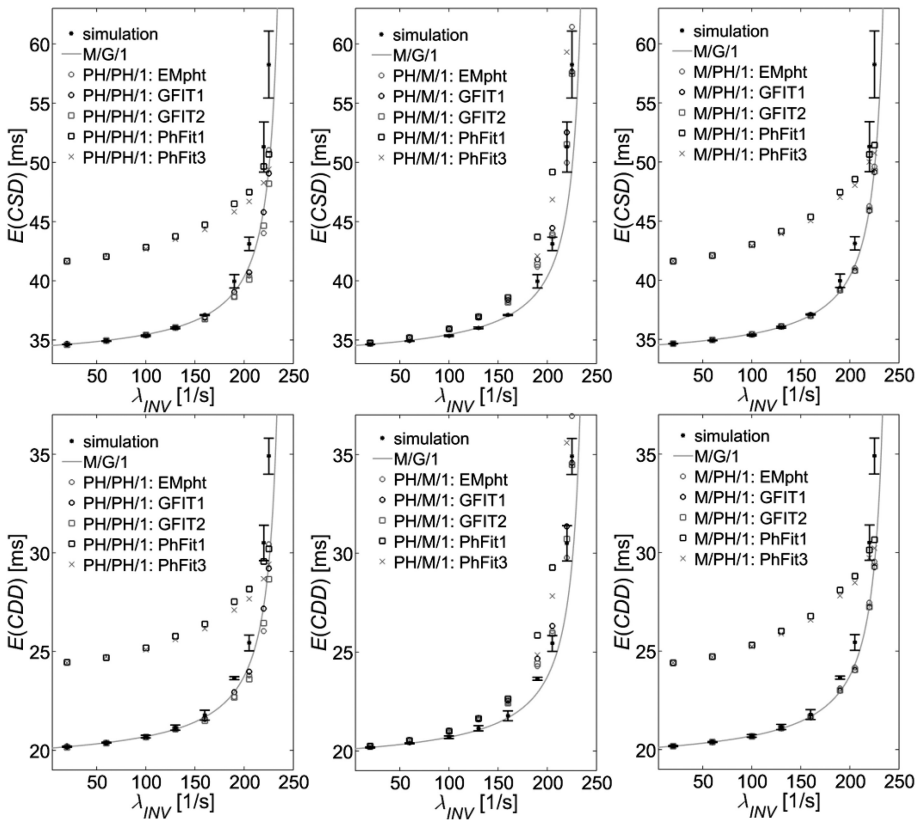


Fig. 5. Results for data set 1c (link length 300 km and bandwidth 100 Mbit/s); left: PH/PH/1 queuing systems, center: PH/M/1 queuing systems, right: M/PH/1 queuing systems; results at the top refer to $E(CSD)$ and at the bottom – to $E(CDD)$.

In order to thoroughly examine the subject of the paper, we have also done some more extended research with the root-mean-square error (RMSE) [7, 12] defined as follows:

$$\text{RMSE} = \sqrt{\frac{E}{\lambda_{INV} \in \Lambda} (Y_{\text{simulation}} - Y_{\text{analytical}})^2}, \tag{3}$$

where Y is either $E(CSD)$ or $E(CDD)$ and $E()$ is the averaging operator over a particular set of call set-up request intensities $\lambda_{INV} \in \Lambda$. To investigate different network conditions, for all data sets the following sets of λ_{INV} were assumed [7, 12]:

- “green” – IMS/NGN elements are low loaded and $E(CSD)$, $E(CDD)$ change very little with call set-up request intensity ($\lambda_{INV} = 20, 60, 100$ [1/s]),
- “yellow” – IMS/NGN elements are quite highly loaded and $E(CSD)$, $E(CDD)$ start noticeably increasing with call set-up request intensity ($\lambda_{INV} = 130, 160, 190$ [1/s]),
- “red” – IMS/NGN elements are overloaded and $E(CSD)$, $E(CDD)$ start going to infinity ($\lambda_{INV} = 205, 220, 225$ [1/s]),
- “gr-yel” – the set including all call set-up request intensities from the “green” set and almost all call set-up request intensities from the “yellow” set ($\lambda_{INV} = 20, 60, 100, 130, 160$ [1/s]),
- “all” – the set containing all call set-up request intensities from the “green”, “yellow” and “red” sets ($\lambda_{INV} = 20, 60, 100, 130, 160, 190, 205, 220, 225$ [1/s]).

The results of RMSE computations are presented in Tables 4, 5 and 6. The RMSE values were calculated for the analytical results presented in Figs. 3, 4 and 5 as well as for other queuing systems investigated in our previous papers – G/G/1 based on moments [7] (“G/G/1” in Tables 4, 5 and 6), PH/PH/1 with phase-type distributions obtained using moment matching algorithms [12] (“PH/PH/1 moments” in Tables 4, 5 and 6). Two best results for each of these previous queuing systems are included along with the names of G/G/1 formulas and PH/PH/1 fitting algorithms based on moments. The names are provided in the brackets below the results. In each column of Tables 4, 5 and 6 we marked two best (bold and underlined font) and two worst RMSE results (bold and italic font with gray background).

The obtained RMSE values demonstrate that for various IMS/NGN parameters the best analytical $E(CSD)$ and $E(CDD)$ results can be achieved using different queuing systems (PH/PH/1 with phase-type distributions based on whole experimental histograms or only on their moments; G/G/1; M/G/1). Moreover, particular fitting algorithms for PH/PH/1 and G/G/1 formulas offer different efficiency. It is also important that often several of the above mentioned approaches offer very good performance (RMSE about 0.5 ms or even smaller), however, our aim is to find the best one.

In the majority of cases PH/PH/1 queuing systems (and their special cases – PH/M/1, M/PH/1) with proper fitting algorithms based on whole experimental histograms (EMpht, GFIT1, GFIT2) are the best or almost the best in terms of RMSE. However, the differences to the other queuing systems are often minimal. To decide which queuing system is the most suitable for IMS/NGN, we also investigated computational complexity (the time required to obtain analytical results for particular queuing systems). It is especially important in engineering applications, in which both

Table 4. RMSE for data set 1a (no links).

	RMSE for $E(CSD)$, [ms]					RMSE for $E(CDD)$, [ms]				
	<i>green</i>	<i>yellow</i>	<i>red</i>	<i>gr-yel</i>	<i>all</i>	<i>green</i>	<i>yellow</i>	<i>red</i>	<i>gr-yel</i>	<i>all</i>
M/G/1	0,110	1,317	8,399	0,506	4,909	0,056	0,727	5,089	0,290	2,968
PH/PH/1: EMpht	0,134	1,656	7,031	0,592	4,171	0,082	0,941	4,225	0,346	2,499
PH/PH/1: GFIT1	0,192	1,099	8,041	0,329	4,687	0,130	0,593	4,871	0,191	2,834
PH/PH/1: GFIT2	0,214	1,081	9,179	0,343	5,337	0,145	0,581	5,580	0,201	3,240
PH/PH/1: PhFit1	0,644	3,203	16,209	1,375	9,547	0,396	1,916	9,983	0,839	5,873
PH/PH/1: PhFit3	0,661	3,263	16,622	1,386	9,788	0,406	1,953	10,240	0,846	6,023
PH/M/1: EMpht	0,401	0,224	2,482	0,351	1,457	0,264	0,172	1,588	0,241	0,935
PH/M/1: GFIT1	0,622	0,942	1,507	0,770	1,087	0,403	0,672	0,821	0,509	0,655
PH/M/1: GFIT2	0,653	0,926	2,203	0,771	1,430	0,422	0,661	1,232	0,508	0,843
PH/M/1: PhFit1	0,542	1,982	201,18	0,835	116,15	0,353	1,336	125,84	0,550	72,658
PH/M/1: PhFit3	0,508	1,362	71,779	0,748	41,450	0,332	0,945	44,967	0,495	25,968
M/PH/1: EMpht	0,102	1,301	8,020	0,481	4,691	0,051	0,717	4,852	0,275	2,832
M/PH/1: GFIT1	0,126	1,411	8,395	0,533	4,915	0,067	0,787	5,087	0,307	2,972
M/PH/1: GFIT2	0,118	1,388	8,330	0,520	4,876	0,061	0,772	5,045	0,299	2,947
M/PH/1: PhFit1	0,693	3,221	16,832	1,403	9,903	0,426	1,926	10,370	0,856	6,094
M/PH/1: PhFit3	0,688	3,218	16,827	1,399	9,899	0,423	1,924	10,366	0,854	6,092
G/G/1 1st [7] (name of formula in brackets)	0,113	0,802	6,656	0,265	5,119	0,080	0,439	3,986	0,172	3,106
G/G/1 2nd [7] (name of formula in brackets)	0,266	0,925	7,029	0,306	5,310	0,181	0,489	4,227	0,188	3,225
PH/PH/1 moments 1st [12] (name of algorithm in brackets)	0,111	0,700	2,519	0,343	1,542	0,057	0,515	1,431	0,192	0,902
	(M/PH/1: PH2 (3))	(PH/M/1: PH2 (3))	(PH/M/1: PH2 (5))	(PH/PH/1: ME (5))	(PH/M/1: PH2 (5))	(M/PH/1: PH2 (3))	(PH/M/1: PH2 (3))	(PH/M/1: PH2 (5))	(PH/PH/1: ME (5))	(PH/M/1: PH2 (5))
PH/PH/1 moments 2nd [12] (name of algorithm in brackets)	0,111	0,719	2,529	0,357	1,545	0,057	0,526	1,437	0,204	0,903
	(M/PH/1: PH2 (5))	(PH/M/1: PH2 (5))	(PH/M/1: PH2 (3))	(PH/PH/1: PH2 (5))	(PH/M/1: PH2 (3))	(M/PH/1: PH2 (5))	(PH/M/1: PH2 (5))	(PH/M/1: PH2 (3))	(PH/PH/1: PH2 (5))	(PH/M/1: PH2 (3))

accuracy and complexity count. In the next part we present some details about calculation times in the analytical model for one data set (Intel Core i7-2670QM CPU 8×2.20 GHz, 8 GB RAM, Windows 7 Professional SP1 64-bit):

- for M/G/1 [6] computations are the fastest of all (several seconds) and require only the input variables (there is no need to use experimental histograms),
- for G/G/1 [7] computations last about one minute (experimental histograms are needed to obtain second and third moments of arrival distributions, which are used in calculations),
- for PH/PH/1 (and their special cases – PH/M/1, M/PH/1) with moment matching algorithms [12] computations take from one to several minutes (experimental histograms are needed to obtain second, third, fourth and fifth moments of arrival distributions, which are used in calculations),

Table 5. RMSE for data set 1b (link length 300 km and bandwidth 10 Mbit/s).

	RMSE for $E(CSD)$, [ms]					RMSE for $E(CDD)$, [ms]				
	<i>green</i>	<i>yellow</i>	<i>red</i>	<i>gr-yl</i>	<i>all</i>	<i>green</i>	<i>yellow</i>	<i>red</i>	<i>gr-yl</i>	<i>all</i>
M/G/1	0,705	2,637	1,805	1,477	1,889	0,133	0,954	1,059	0,489	0,826
PH/PH/1: EMpht	0,460	0,852	5,450	0,724	3,196	0,031	0,356	4,095	0,068	2,373
PH/PH/1: GFIT1	0,666	0,990	4,834	0,834	2,875	0,111	0,187	3,702	0,099	2,141
PH/PH/1: GFIT2	0,290	0,410	6,839	0,374	3,959	0,138	0,658	5,011	0,242	2,919
PH/PH/1: PhFit1	1,201	2,274	13,657	1,341	8,024	1,104	2,216	9,396	1,355	5,610
PH/PH/1: PhFit3	1,311	2,744	14,599	1,547	8,610	1,177	2,511	9,999	1,488	5,991
PH/M/1: EMpht	1,533	4,848	7,132	2,863	5,057	0,655	2,368	3,929	1,359	2,675
PH/M/1: GFIT1	1,759	5,178	7,782	2,934	5,492	0,805	2,580	4,315	1,398	2,940
PH/M/1: GFIT2	1,565	4,974	6,860	2,862	4,975	0,675	2,450	3,745	1,357	2,613
PH/M/1: PhFit1	1,622	6,031	19,388	3,175	11,76	0,714	3,129	11,619	1,563	6,959
PH/M/1: PhFit3	1,591	5,618	16,077	3,131	9,875	0,694	2,869	9,521	1,538	5,755
M/PH/1: EMpht	0,864	3,001	2,807	1,727	2,424	0,226	1,170	1,364	0,631	1,046
M/PH/1: GFIT1	0,839	2,894	2,566	1,674	2,285	0,210	1,102	1,292	0,598	0,988
M/PH/1: GFIT2	0,705	2,414	1,976	1,415	1,846	0,128	0,796	1,217	0,438	0,843
M/PH/1: PhFit1	0,979	0,870	10,786	0,812	6,273	0,949	1,343	7,589	0,974	4,483
M/PH/1: PhFit3	1,013	1,163	11,352	0,896	6,614	0,979	1,535	7,957	1,054	4,713
G/G/1 1st [7] (name of formula in brackets)	0,550 (Diff22)	1,042 (Diff22)	4,729 (Daley)	0,660 (Diff22)	3,229 (My2)	0,123 (KLB)	0,300 (My1)	2,625 (Daley)	0,206 (My1)	2,223 (My2)
G/G/1 2nd [7] (name of formula in brackets)	0,698 (KLB)	1,075 (Diff21)	4,867 (My2)	1,028 (My1)	3,307 (King2)	0,166 (King2)	0,331 (Diff21)	2,742 (Page)	0,280 (KLB)	2,381 (King2)
PH/PH/1 moments 1st [12] (name of algorithm in brackets)	0,405 (PH/PH/1: APH2 (3))	0,515 (PH/PH/1: APH2 (3))	2,025 (M/PH/1: PH2 (3))	0,502 (PH/PH/1: APH2 (3))	1,874 (M/PH/1: PH2 (3))	0,038 (PH/PH/1: PH2 (5))	0,428 (PH/PH/1: ME (3-4))	1,181 (M/PH/1: PH2 (3))	0,069 (PH/PH/1: PH2 (5))	0,834 (M/PH/1: PH2 (3))
PH/PH/1 moments 2nd [12] (name of algorithm in brackets)	0,500 (PH/PH/1: ME (5))	0,613 (PH/PH/1: ME (5))	2,025 (M/PH/1: PH2 (5))	0,602 (PH/PH/1: ME (5))	1,874 (M/PH/1: PH2 (5))	0,060 (PH/PH/1: APH1 (3))	0,474 (PH/PH/1: APH1 (3), PH2 (3))	1,181 (M/PH/1: PH2 (5))	0,071 (PH/PH/1: APH1 (3), PH2 (3))	0,834 (M/PH/1: PH2 (5))

- PH/PH/1 (and their special cases – PH/M/1, M/PH/1) with maximum likelihood and distance minimization fitting algorithms are the worst case; fitting phase-type distributions to whole experimental histograms lasts from 30 min to several hours with parameters of the algorithms described in the previous section; therefore, fitting was performed offline and its results were saved to disk for further use.

Taking into account all the aspects presented in the previous part of this section (accuracy and complexity), we propose to use the following queuing systems in the analytical model of a single domain of IMS/NGN. For data sets 1a (no links; Table 4) and 1c (links with relatively high bandwidth; Table 6) we suggest using simple M/G/1

Table 6. RMSE for data set 1c (link length 300 km and bandwidth 100 Mbit/s).

	RMSE for $E(CSD)$, [ms]					RMSE for $E(CDD)$, [ms]				
	<i>green</i>	<i>yellow</i>	<i>red</i>	<i>gr-yel</i>	<i>all</i>	<i>green</i>	<i>yellow</i>	<i>red</i>	<i>gr-yel</i>	<i>all</i>
M/G/1	0,069	0,518	6,014	0,064	3,485	0,003	0,410	3,701	0,058	2,150
PH/PH/1: EMpht	0,020	0,758	6,114	0,120	3,557	0,031	0,564	3,758	0,112	2,194
PH/PH/1: GFIT1	0,053	0,529	6,329	0,093	3,667	0,016	0,421	3,900	0,085	2,265
PH/PH/1: GFIT2	0,040	0,783	7,166	0,163	4,162	0,021	0,579	4,421	0,129	2,574
PH/PH/1: PhFit1	7,045	7,232	5,149	7,265	6,543	4,248	4,343	3,188	4,380	3,961
PH/PH/1: PhFit3	7,004	6,826	5,783	7,115	6,560	4,223	4,086	3,584	4,286	3,974
PH/M/1: EMpht	0,347	1,083	2,071	0,689	1,364	0,176	0,584	1,311	0,380	0,835
PH/M/1: GFIT1	0,407	1,415	1,101	0,783	1,061	0,215	0,788	0,738	0,439	0,636
PH/M/1: GFIT2	0,384	1,171	0,605	0,712	0,793	0,200	0,636	0,395	0,392	0,447
PH/M/1: PhFit1	0,386	2,381	131,94	0,835	76,188	0,201	1,394	82,503	0,473	47,640
PH/M/1: PhFit3	0,391	1,594	33,468	0,836	19,346	0,204	0,900	20,958	0,473	12,112
M/PH/1: EMpht	0,069	0,390	5,903	0,076	3,416	0,002	0,324	3,630	0,035	2,104
M/PH/1: GFIT1	0,045	0,486	6,266	0,073	3,629	0,014	0,390	3,858	0,067	2,239
M/PH/1: GFIT2	0,052	0,469	6,206	0,072	3,593	0,009	0,378	3,819	0,060	2,216
M/PH/1: PhFit1	7,123	7,905	5,039	7,539	6,798	4,289	4,741	3,132	4,539	4,110
M/PH/1: PhFit3	7,079	7,609	5,221	7,411	6,715	4,259	4,551	3,248	4,456	4,058
G/G/1 1st [7] (name of formula in brackets)	0,111	0,584	5,089	0,121	4,132	0,035	0,419	3,120	0,098	2,548
G/G/1 2nd [7] (name of formula in brackets)	0,150	0,774	5,173	0,156	4,341	0,055	0,573	3,174	0,110	2,686
PH/PH/1 moments 1st [12] (name of algorithm in brackets)	0,051	0,426	1,556	0,070	1,137	0,005	0,350	0,947	0,041	0,664
	(PH/PH/1: APH2 (3))	(M/PH/1: PH2 (3))	(PH/M/1: PH2 (3))	(M/PH/1: PH2 (3))	(PH/M/1: PH2 (5))	(M/PH/1: PH2 (3))	(M/PH/1: PH2 (3))	(PH/M/1: PH2 (3))	(PH/PH/1: ME (3-4))	(PH/M/1: PH2 (5))
PH/PH/1 moments 2nd [12] (name of algorithm in brackets)	0,058	0,426	1,556	0,070	1,138	0,005	0,350	0,947	0,049	0,664
	(M/PH/1: PH2 (5))	(M/PH/1: PH2 (5))	(PH/M/1: PH2 (5))	(M/PH/1: PH2 (5))	(PH/M/1: PH2 (3))	(M/PH/1: PH2 (5))	(M/PH/1: PH2 (5))	(PH/M/1: PH2 (5))	(M/PH/1: PH2 (3))	(PH/M/1: PH2 (3))

queuing systems, which give acceptable results (comparable to the best results), when “green” and “yellow” sets of λ_{INV} are considered. For “red” and “all” sets of call set-up request intensity PH/M/1 queuing systems are the best.

When links with relatively low bandwidth are used (data set 1b; Table 5) the situation is different. PH/PH/1 queuing systems give the best results for “green” and “yellow” sets of λ_{INV} . For “red” and “all” the smallest and very comparable RMSE values are offered by M/G/1 and M/PH/1 queuing systems. Due to much simpler calculations, we propose to use M/G/1.

For the above mentioned PH/PH/1 and PH/M/1 queuing systems RMSE results obtained using maximum likelihood and distance minimization algorithms operating on whole experimental histograms are in most cases not significantly better than for

algorithms based only on moments [12]. As only slightly better results do not compensate for the high computational complexity (many hours of calculations for maximum likelihood and distance minimization algorithms), for engineering purposes more applicable are moment matching algorithms.

The necessity to choose different queuing systems in the analytical traffic model for different data sets and call set-up request intensities results from the complexity of IMS/NGN elements and call scenario. In the network there are many instances of call scenarios performed simultaneously, each of them concern sending and processing a set of signaling messages (Fig. 2). During a single instance most of the network elements is used several times. Our previous research [25] indicated that in many cases even a small change in load of one IMS/NGN element results in modification of types of message inter-arrival time distributions in the whole system. Consequently, it is very difficult to select a single queuing system, which would be efficient for all sets of investigated input variables.

5 Conclusions and Future Work

In the paper we continue our work on analytical and simulation traffic model of a single domain of IMS/NGN. Our aim is to find the best queuing system for the analytical model, to achieve possibly the best conformity with experimental results. These are provided by our simulation model, which does not implement theoretical queuing systems, but reflects the operation of real IMS/NGN network elements and standardized call scenarios.

In this paper investigations using PH/PH/1 queuing systems (and their special cases – PH/M/1, M/PH/1) are performed, in which arrival and service distributions are represented by phase-type distributions obtained using maximum likelihood and distance minimization algorithms. The obtained results are compared to these calculated using previously investigated queuing systems (M/G/1; G/G/1 based on two or three moments of arrival distribution and two moments of service distribution; PH/PH/1 with phase-type distributions obtained using moment matching algorithms). The compared parameters are total system responses – mean Call Set-up Delay $E(CSD)$ as well as mean Call Disengagement Delay $E(CDD)$, standardized by the ITU-T.

The performed research demonstrates that performance of the tested queuing systems in IMS/NGN depends on several network parameters, including load of network elements and links (for links load results from message intensity and available bandwidth). Often for the assumed set of parameters more than one of the investigated queuing systems offer accuracy, which can be considered as good or even very good. However, due to specific nature of IMS/NGN elements and call scenario, it is very difficult to point one universal queuing system, appropriate for all sets of network parameters.

Apart from accuracy, in engineering applications very important is also computational complexity (the time necessary to obtain analytical results for particular queuing system), which has not been analyzed in our previous papers. Considering both accuracy and complexity, and assuming that network is not overloaded (“green” and “yellow” sets of call set-up request intensity; Sect. 4), we suggest using M/G/1 and

PH/PH/1 queuing systems in the analytical model of a single domain of IMS/NGN. The first type of queuing systems should be applied when IMS/NGN elements are connected directly or when links with relatively high bandwidth are used (data sets 1a, 1c; Table 2). The second one performs the best for links with relatively low bandwidth (data set 1b; Table 2). Due to smaller time necessary to fit phase-type distributions, for PH/PH/1 queuing systems we propose to use moment matching algorithms, which offer final results only slightly worse than maximum likelihood and distance minimization algorithms operating on whole experimental histograms.

In the next step we are going to perform similar research to that described in this paper, but in a multidomain IMS/NGN environment (appropriate analytical and simulation traffic models are under development [33–35]).

References

1. General overview of NGN, ITU-T Recommendation Y.2001, December 2004
2. IP Multimedia Subsystem (IMS); Stage 2 (Release 14), 3GPP TS 23.228 v14.3.0, March 2017
3. Call processing performance for voice service in hybrid IP networks, ITU-T Recommendation Y.1530, November 2007
4. SIP-based call processing performance, ITU-T Recommendation Y.1531, November 2007
5. Kaczmarek, S., Kaszuba, M., Sac, M.: Simulation model of IMS/NGN call processing performance. *Gdańsk Univ. Technol. Faculty ETI Ann.* **20**, 25–36 (2012)
6. Kaczmarek, S., Sac, M.: Traffic model for evaluation of call processing performance parameters in IMS-based NGN. In: Grzech, A., et al. (eds.) *Information systems architecture and technology: networks design and analysis*, pp. 85–100. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław (2012)
7. Kaczmarek, S., Sac, M.: Analysis of IMS/NGN call processing performance using G/G/1 queuing systems approximations. *Telecommun. Rev. Telecommun. News* (8–9), 702–710 (2012)
8. Osogami, T., Harchol-Balter, M.: Closed form solutions for mapping general distributions to quasi-minimal PH distributions. *Perform. Evaluat.* **63**(6), 524–555 (2006)
9. Bobbio, A., Horvath, A., Telek, M.: Matching three moments with minimal acyclic phase type distributions. *Stochastic Models* **21**(2–3), 303–326 (2005)
10. Telek, M., Horvath, G.: A minimal representation of Markov arrival processes and a moments matching method. *Perform. Evaluat.* **64**(9–12), 1153–1168 (2007)
11. Asmussen, S., Nerman, O., Olsson, M.: Fitting phase-type distributions via the EM algorithm. *Scand. J. Stat.* **23**(4), 419–441 (1996)
12. Kaczmarek, S., Sac, M.: Analysis of IMS/NGN call processing performance using phase-type distributions. In: Grzech, A., et al. (eds.) *Information systems architecture and technology: network architecture and applications*, pp. 23–39. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław (2013)
13. Kaczmarek, S., Sac, M.: Traffic modeling in IMS-based NGN networks. *Gdańsk Univ. Technol. Faculty ETI Ann.* **1**(9), 457–464 (2011)
14. Functional requirements and architecture of next generation networks, ITU-T Recommendation Y.2012, April 2010
15. IMS for next generation networks, ITU-T Recommendation Y.2021, September 2006
16. Resource and admission control functions in next generation networks, ITU-T Recommendation Y.2111, November 2011

17. Resource control protocol no. 1, version 3 – Protocol at the Rs interface between service control entities and the policy decision physical entity, ITU-T Recommendation Q.3301.1, August 2013
18. Pirhadi, M., Safavi Hemami, SM., Khademzadeh, A.: Resource and admission control architecture and QoS signaling scenarios in next generation networks. *World Appl. Sci. J. 7* (Special Issue of Computer & IT), 87–97 (2009)
19. Fajardo, V., et al.; Diameter Base Protocol, IETF RFC 6733, October 2012
20. Rosenberg, J., et al.: SIP: Session Initiation Protocol, IETF RFC 3261, June 2002
21. OMNeT++ Discrete Event Simulator – Home. <http://www.omnetpp.org>. Accessed 31 May 2017
22. Thümmler, A., Buchholz, P., Telek, M.: A novel approach for fitting probability distributions to real trace data with the EM algorithm. In: *Proceedings of the 6th IEEE International Conference on Dependable Systems and Networks*, Yokohama, Japan, pp. 712–721 (2005)
23. Wang, J., Liu, J., She, C.: Segment-based adaptive hyper-erlang model for long-tailed network traffic approximation. *J. Supercomput.* **45**(3), 296–312 (2008)
24. Horvath, A., Telek, M.: PhFit: a general phase-type fitting tool. In: *Proceedings of the 12th International Conference on Computer Performance Evaluation, Modelling Techniques and Tools, TOOLS 2002*, London, United Kingdom, pp. 82–91 (2002)
25. Kaczmarek, S., Sac, M.: Approximation of message inter-arrival and inter-departure time distributions in IMS/NGN architecture using phase-type distributions. *J. Telecommun. Inform. Technol.* (3), 9–18 (2013)
26. Pérez, J.F., van Velthoven, J., van Houdt, B.: Q-MAM: a tool for solving infinite queues using matrix-analytic methods. In: *Proceedings of the SMCTools 2008*, Athens, Greece (2008)
27. Gross, D., Shortle, J., Thompson, J., Harris, C.: *Fundamentals of Queueing Theory*, 4th edn. Wiley, New York (2008)
28. Lucantoni, D.M., Meier-Hellstern, K.S., Neuts, M.F.: A single-server queue with server vacations and a class of nonrenewal arrival processes. *Adv. Appl. Probab.* **22**(3), 676–705 (1990)
29. Lucantoni, D.M.: New results on the single server queue with a batch Markovian arrival process. *Commun. Stat. Stochast. Models* **7**(1), 1–46 (1991)
30. PATS: Performance analysis of telecommunication systems. <http://win.ua.ac.be/~vanhoudt/>. Accessed 31 May 2017
31. TELCOM 2825 Information systems and network infrastructure protection. http://www.sis.piit.edu/~dtipper/2130_notes.html. Accessed 31 May 2017
32. Abhayawardhana, V.S., Babbage, R.: A traffic model for the IP Multimedia Subsystem (IMS). In: *IEEE 65th Vehicular Technology Conference, VTC2007*, Dublin, Ireland (2007)
33. Kaczmarek, S., Sac, M.: Traffic model of a multidomain IMS/NGN. *Telecommun. Rev. Telecommun. News* (8–9), 1030–1038 (2014)
34. Kaczmarek, S., Sac, M.: Call processing performance in a multidomain IMS/NGN with asymmetric traffic. In: Grzech, A., et al. (eds.) *Information Systems Architecture and Technology: Selected Aspects of Communication and Computational Systems*, pp. 11–28. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław (2014)
35. Kaczmarek, S., Sac, M.: Verification of the analytical traffic model of a multidomain IMS/NGN using the simulation model. In: Grzech, A., et al. (eds.) *Information Systems Architecture and Technology: Proceedings of 36th International Conference on Information Systems Architecture and Technology – ISAT 2015 – Part II. Advances in Intelligent Systems and Computing*, vol. 430, pp. 109–130. Springer, Cham (2016)