

Approach to Estimation of Performance Measures for SIP Server Model with Batch Arrivals

Yurii Orlov¹, Yuliya Gaidamaka², and Elvira Zaripova²(✉)

¹ Keldysh Institute of Applied Mathematics Russian Academy of Sciences,
Miusskaya sq. 4, 125047 Moscow, Russia
ov31509f@yandex.ru

² Peoples' Friendship University of Russia, Miklukho-Maklaya street 6,
117198 Moscow, Russia
{ygaidamaka, ezarip}@sci.pfu.edu.ru
<http://api.sci.pfu.edu.ru>

Abstract. In this paper an approach to analysis of dependence of Session Initiation Protocol server model with batch arrivals performance measures on batch size distribution is considered. Proposed approach employs non-parametric methods of statistical analysis. It is shown that there is statistical reliable dependence of performance measures, taken for signaling traffic analysis, on distance between distributions in definite norm. On the basis of proposed analysis elasticity coefficients were evaluated depending on distance between batch size distributions. This approach enables to get correction factors for estimation of these parameters in case distribution functions differ from uniform.

Keywords: Optimization · SIP mathematical model · Distribution function · Norm · Performance measure · Queuing system · Parameter sensibility · Sample · Batch arrivals

1 Introduction

Developing telecommunication services are successfully provided via IP-based Multimedia Subsystem (IMS), where Session Initiation Protocol (SIP) is the main signaling protocol [1]. Signaling traffic load is auxiliary for communication nodes and is used for providing communications services to users. Signaling messages have so-called *life time*. When life time is over, information becomes not actual, signaling messages are retransmitted and overload the SIP server that processes them. Among the other problems with Session Initiation Protocol the overloading problem was stated by IETF in [2], and solution requirements that address the problems were formulated in [3]. Providing telecommunication

This work was supported in part by the Russian Foundation for Basic Research, projects No. 15-07-03051, 15-07-03608.

services of required quality results in necessity of detailed research of communication system structure, statistic data analysis, implementation of processing algorithm for different types of signaling messages in order to increase number of successfully initiated sessions and decrease of service and sojourn time at a SIP server [6]. Many papers from the IETF documents [3–5] to a number of scientific papers [11–16, 18–24] were devoted to developing and investigating mechanisms of SIP-server overload control. For example, signaling messages service process was presented as asymmetric polling system [9], as queueing model with threshold control [10]. Sometimes you can see research even with unreliable servers [17].

In this paper were chosen the following parameters for the investigation: average queue length and average waiting time. The first one enables estimation of SIP server's buffer capacity during the busy hour, the second when summed with service time enables estimation of SIP server sojourn time and is comparable to the message life time.

Process of signaling messages arrival and processing at a SIP server is performed via a single-server queue with batch arrivals and vacations. The vacation models the time interval when server processes messages that differ from signaling ones. In a model the batch arrival of customers corresponds to the simultaneous requests of the group of online subscribers.

In [6–8] there is an estimation of average value of queue length and average waiting time for general distribution of batch size. In [8] the analysis for the four batch size distribution functions has been done: Zipf, geometric, logarithmic and uniform distribution. In [8] a statistical dependence of investigating parameters on batch size distribution in case of Poisson process has been analyzed. In contrast with the papers where an investigation was performed only for geometrical distribution, in [8] it is recommended to use a uniform distribution of batch size that essentially simplifies formulas for calculations. In this paper an approach to performance measures analysis is proposed, sensitivity of model parameters to batch size distribution was estimated.

Distribution function variation was evaluated in [8] under different norms. Those with highest reliability of estimated statistical dependence were chosen. It turned out that analysis of numeric evaluation of SIP server performance parameters sensitivity to batch size distribution variation can be carried out with high reliability (0.99).

Since only sampling distribution functions of parameters are available for preliminary analysis, so we want to evaluate model parameters with statistical fluctuation of sampling distribution function when the distribution is not converge to any general population because of non-stationary behavior. We denote, that the batch size is a measured on practice, whereas average waiting time and average queue length depend on messages service time and processor vacation time. Distribution functions of the last two parameters are assumed to be known, because they depend on the SIP server hardware implementation. However, confidence interval for average queue length may not be got from corresponding empirical distribution, because we do not know this distribution function.

We can evaluate these parameters by means of the simplified models only. In [8] the sensitivity analysis for one of models was carried out. Analysis was later applied to empirical distribution function of batch size.

2 SIP Server Model as a Queue with Batch Arrivals and Vacations

A mathematical model of signaling messages processing at a SIP server is investigated as a queuing system with batch arrivals and vacations. According to Basharin-Kendall notation this system is denoted as $M^{[X]}|G|1|\infty$. Let suppose that a batch of customers arrives according to Poisson process with rate λ . Customer service time is a random variable with a distribution function $B(t)$, where b_1 is the mean value and b_2 is the finite second moment. If there are no customers in the queue, the server goes for a vacation. The vacation time is a random variable with a distribution function $V(t)$ with finite first and second moments v_1 and v_2 .

Let $f(k)$ be probability that batch size is equal to $k \geq 1$. We denote the corresponding distribution function $F(k)$. For the queuing system $M^{[X]}|G|1|\infty$ with vacations we can find average queue length and average waiting time depending on offered load ρ which is estimated as follows:

$$\rho = \lambda b_1 l^{(1)}, \tag{1}$$

where $l^{(1)}$ is average batch size with distribution function $F(k)$.

In [6] a generating function was obtained for the queue length distribution for a single server model with vacations. The result is expressed as follows:

$$P(z) = \frac{1 - \rho}{\lambda v_1} \cdot \frac{1 - z}{1 - L(z)} \cdot \frac{1 - \phi(\lambda, z)}{\beta(\lambda, z) - z}, \tag{2}$$

where $L(z)$ is a generating function for batch size with the distribution function $F(k)$: $L(z) = \sum_{k=0}^{\infty} f(k)z^k$. Other functions mentioned in (2) are expressed with

the following equations: $v_1 = \int_0^{\infty} t dV(t)$, $\phi(\lambda, z) = \int_0^{\infty} e^{-\lambda t(1-L(z))} dV(t)$, and $\beta(\lambda, z) = \int_0^{\infty} e^{-\lambda t(1-L(z))} dB(t)$. According to [8] average queue length depending on load ρ is obtained by the generating function (2) as follows:

$$N = \lim_{z \rightarrow 1} P'(z) = \frac{v_2}{2v_1 b_1} \rho + \frac{l^{(2)} - l^{(1)}}{2l^{(1)}} \frac{\rho}{1 - \rho} + \frac{b_2}{2b_1^2} \frac{\rho^2}{1 - \rho}. \tag{3}$$

Taking into account the notation $v_s = \int_0^{\infty} t^s dV(t)$, $b_s = \int_0^{\infty} t^s dB(t)$ and $l^{(s)} = \sum_k k^s f(k)$, $s = 1; 2$, the average waiting time is obtained as follows:

$$\tau = \frac{N b_1}{\rho} = \frac{v_2}{2v_1} + \frac{l^{(2)} - l^{(1)}}{2l^{(1)}} \frac{b_1}{1 - \rho} + \frac{b_2}{2b_1} \frac{\rho}{1 - \rho}. \tag{4}$$

You can see the average queue length (Fig. 1) and average waiting time (Fig. 2) for 4 batch size distributions for exponential service time and vacation time with the following initial data $b_1 = 15$ ms – average service time, $v_1 = 15l_1$ ms – average vacation time, $l_1 = 3$ – average batch size. This is the starting point of the reseaches. The plots (Figs. 1 and 2) have similar behavior. We investigate the sensitivity of the model parameters to probability variation of batch size.

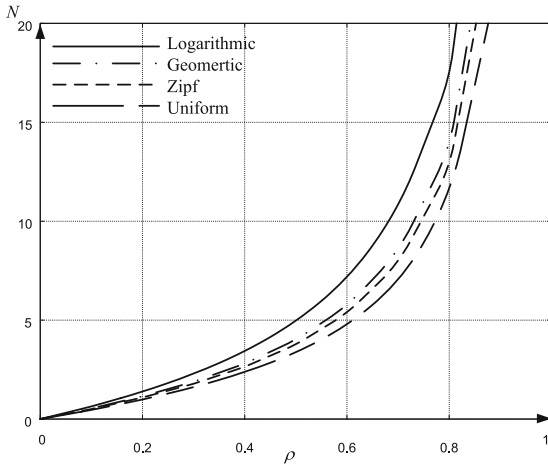


Fig. 1. Average queue length depending on load.

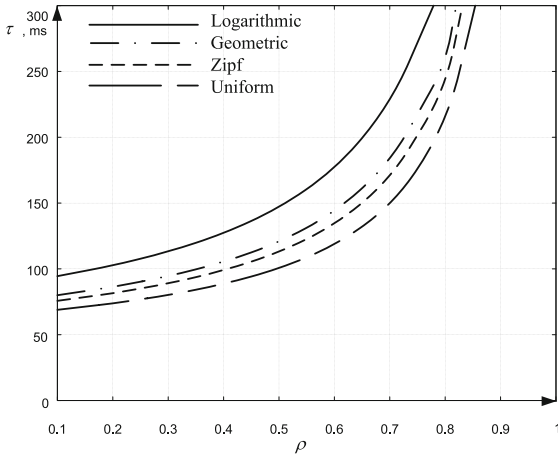


Fig. 2. Average waiting time depending on load.

3 Sensitivity of the Model Parameters to Probability Variation

As a base of our analysis we use 4 batch size distributions with following distribution functions (Table 1), where $f_i(k)$ is the probability, that batch size is equal k . We assume, that maximum batch size is equal to 8. The distribution functions can be found by formula $F_i(n) = \sum_{k=1}^n f_i(k)$.

We consider 5 types of norms and define the distance between two DFs by the formulas from the Table 2, where the distance from i to j distribution functions is ρ_{ij}^{norm} . The first step of our investigation was to compare distances between DFs in several norms.

Table 3 shows the distances between batch size DFs in considered norms. For example, the first row shows the distance between Logarithmic and Geometric batch size distributions, the last row – the distance between Zipf and Uniform distributions. Statistics show that the distances are dependent from each other. So, for the next step of our investigation we can use any appropriate norm.

Table 1. Distribution functions

Logarithmic DF	$f_1(k) = \frac{1}{Z_1} \frac{(0,85)^k}{k}, \quad Z_1 = \sum_{k=1}^8 \frac{(0,85)^k}{k}$
Geometric DF	$f_2(k) = \frac{1}{Z_2} (0,67)^k, \quad Z_2 = \sum_{k=1}^8 (0,67)^k$
Zipf DF	$f_3(k) = \frac{1}{Z_3^k}, \quad Z_3 = \sum_{k=1}^8 1/k$
Uniform DF	$f_4(k) = 1/8$

Table 2. Definition of distance in norms

C norm	$\rho_{ij}^C = \max_k f_i(k) - f_j(k) $
$L1$ norm	$\rho_{ij}^{L1} = \sum_{k=1}^8 f_i(k) - f_j(k) $
Hellinger distance	$\rho_{ij}^{He} = \sum_{k=1}^8 \left(\sqrt{f_i(k)} - \sqrt{f_j(k)} \right)^2$
Kullback-Leibler distance	$\rho_{ij}^{KL} = \sum_{k=1}^8 f_i(k) \ln (f_i(k)/f_j(k))$
Supplement for total area S	$\rho_{ij}^S = 1 - \sum_{k=1}^8 \min (f_i(k), f_j(k))$

Table 3. Distances between batch size distributions

Comparison	C	$L1$	He	KL	S
Logarithmic, Geometric	0,13	0,27	0,02	0,04	0,134
Logarithmic, Zipf	0,11	0,26	0,03	0,05	0,129
Logarithmic, Uniform	0,35	0,86	0,26	0,52	0,431
Geometric, Zipf	0,05	0,18	0,01	0,03	0,090
Geometric, Uniform	0,22	0,71	0,18	0,33	0,354
Zipf, Uniform	0,24	0,60	0,13	0,26	0,302

Let find out how to change the values N and τ in formulas (3)–(4), if probability variation of $f(k)$ is low. Let a new distribution can be expressed as $\tilde{f}(k) = f(k) + \varepsilon(k)f(k)$, moreover, under the same normaling the equality $\sum_k \varepsilon(k)f(k) = 0$ is fulfilled. Let introduce the variable $E = \sup_k |\varepsilon(k)|$. Then we get the following estimation of distance between distributions in $L1$ norm:

$$\varepsilon = \left\| f - \tilde{f} \right\|_{L1} = \sum_k \left| f(k) - \tilde{f}(k) \right| = \sum_k |\varepsilon(k)| f(k) \leq E. \tag{5}$$

Let use (5) to get estimation of some differentiable function variation depending on average batch size in case $E \ll 1$:

$$\begin{aligned} \left| \delta l^{(1)} \right| &= \left| \sum_k k (f(k) - \tilde{f}(k)) \right| < \sum_k k \left| f(k) - \tilde{f}(k) \right| = \sum_k k |\varepsilon(k)| f(k) \leq E l^{(1)}; \\ \left| \delta u(l^{(1)}) \right| &= \left| \frac{\partial u(l^{(1)})}{\partial l^{(1)}} \right| \cdot \left| \delta l^{(1)} \right| \leq E u \left| \frac{\partial \ln u}{\partial \ln l} \right|_{l=l^{(1)}} + o(E). \end{aligned} \tag{6}$$

Logarithmic derivative of function with respect to parameter’s logarithm is called function sensitivity to parameter variation. Then, from (6) comes that variation of some function from the mean value, obtained due to distribution variation, at linear approximation on E does not exceed this function multiplied by the supremum of distribution density variation and by modulus of specified sensitivity. In our case estimations of average queue length (3) and average waiting time (4), that are linear for E , are expressed as follows:

$$\begin{aligned} \left| \delta N \right| &\leq \frac{E\rho}{2} \cdot \left| \frac{v_2}{v_1 b_1} + \frac{l^{(2)} - l^{(1)}}{l^{(1)}} \frac{b_1}{(1 - \rho)^2} + \frac{b_2}{b_1^2} (2 - \rho) \frac{\rho}{(1 - \rho)^2} \right|; \\ \left| \delta \tau \right| &\leq \frac{E\rho}{2(1 - \rho)^2} \cdot \left| \frac{l^{(2)} - l^{(1)}}{l^{(1)}} b_1 + \frac{b_2}{b_1} \right|. \end{aligned} \tag{7}$$

However, theoretical estimations (7) do not possess adequate accuracy as they appear to be too excessive. Despite this fact they cannot be improved within of functions to be chosen for empirical distribution functions. Unimprovability comes out of existence of a variation when $\varepsilon(k) = \{0; \pm E\}$. Inadequacy of accuracy comes out of condition (5): as far as $\varepsilon \leq E$, then for heavily nonuniform distributions the variation norm in the form of $\sup_k |\varepsilon(k)|$ is too crude estimate, since for such kind of distributions the distance between distributions can be significantly less than E . That is why in [4] an analysis of sensitivity of these parameters based on the numerical results for various functions $f(k)$ was carried out. It was found out that four types of norms – L1 and C for $F(k)$, L1 for $f(k)$ and similar to them the forth norm that is a supplement for total area S to the unity of two densities – determine rather exactly the variation of values N and τ under variation of densities.

Let us denote variation of $f(k)$ by δf , which is denominated in norm C for distribution function. We denote by $\delta N(\rho)$ variation of average queue length and by $\delta\tau(\rho)$ variation of average waiting time for a given value of load ρ . Data analysis showed that under variation of uniform distribution, which determined on the interval from 1 to the maximum batch size (in our case maximum is equal to 8), there is a relationship between $\delta N(\rho)$ and δf , also $\delta\tau(\rho)$ and δf with determination 0.99:

$$\begin{aligned} \left| \frac{\delta N(\rho)}{N(\rho)} \right| &= 0,2(1 - \ln \rho) (\delta f)^{0,27}; \\ \left| \frac{\delta\tau(\rho)}{\tau(\rho)} \right| &= \frac{0,106 + 0,041 \ln \rho}{\rho} (\delta f)^{0,27}; \\ \rho &\in [0,1; 0,9]. \end{aligned} \tag{8}$$

4 Algorithm for Average Queue Length Calculation

Practical application of the described method is the algorithm for average queue length calculation.

Step 1. Choose the basic predicted batch size distribution of $f(k)$ on the interval $1 \leq k \leq M$, for example, uniform distribution: $f(k) = 1/8$.

Step 2. Calculate the average queue length N_0 for the basic distribution:

$$N_0 = \frac{v_2}{2v_1b_1}\rho + \frac{M-1}{3} \frac{\rho}{1-\rho} + \frac{b_2}{2b_1^2} \frac{\rho^2}{1-\rho}, \tag{9}$$

For basic distribution function Eq. (3) is considered as null approximation.

Step 3. Get the empirical batch size distribution function $F(k)$ from measurements. Let consider that empirical distribution, being investigated, is stationary, $F(k)$ is its distribution function.

Step 4. Calculate the distance between empirical and uniform distributions:

$$\delta f = \sup_k |F(k) - k/M|. \tag{10}$$

Step 5. Substituting the result of (10) in (8) we get estimation of average queue length that corresponds to the following empirical distribution:

$$N/N_0 \approx 1 + 0,2(1 - \ln \rho) (\delta f)^{0,27}, \quad \rho \in [0,1; 0,9] \tag{11}$$

Estimation for average waiting time variation is expressed in the same way.

Equation (11) is computationally much simpler than calculation of the generating function in accordance with (2), where $L(z)$ is calculated through empirical distribution of $F(k)$. That is, firstly, rather difficult and, secondly, leads to calculation errors that may exceed approximation inaccuracy for (11).

5 Average Queue Length for Geometric Batch Size Distribution

For example, let consider that an empirical distribution of $f_n(k)$ is taken from a general population $f(k)$ that has geometric distribution with a parameter q on the interval $1 \leq k \leq M$. That means $f(k) = \frac{1-q}{1-q^M} q^{k-1}$.

We show you algorithm for average queue length calculation with geometric distribution function.

Step 1. Basic batch size distribution:

$$f(k) = 1/8, 1 \leq k \leq M - \text{uniform distribution.}$$

Step 2. Average queue length N_0 for the uniform distribution:

$$N_0(\rho), \quad \rho \in [0, 1; 0, 9].$$

Step 3. Empirical batch size distribution is geometric distribution with following parameters: $q = 0,67, M = 8$ [8].

$$F(k) = \frac{1-q}{1-q^M} q^{k-1}, \quad 1 \leq k \leq M.$$

Step 4. Distance between empirical and basic distributions:

$$\delta f = 0,35.$$

Step 5. Estimation of average queue length:

$$N \approx 0,14(1 - \ln \rho) N_0(\rho), \quad \rho \in [0, 1; 0, 9].$$

6 Conclusion

This paper presents the approach to estimation of the performance measures for SIP server model with batch arrivals and vacations depending on the batch size distribution. Investigation of this particular dependence was motivated by the fact that the batch size distribution is not known as a general population and, moreover, cannot be recognized as far as empirical evaluations of this population are non-stationary. That is why approximate evaluation methods, that are not associated with a specified functional class of mentioned distributions, are of great significance and actuality. Therefore, the method that considers coefficients of the model parameters sensitivity to adjustment of the distance between distribution functions seems to be efficient among nonparametric techniques. This method may be used for non-stationery distributions when non-stationary behavior is interpreted as definite variation of some basic distribution (for example, uniform). This approach enables to circumvent technical difficulty coming from absence of convergence theorem both for probability and the norm for random variables being investigated.

Proposed approach to evaluation of performance measures of a Session Initiation Protocol server model and given analysis of parameters sensitivity leads to recommendations for engineers to use simple formulas for preliminary evaluation of presence signaling messages service.

Acknowledgment. This work was supported in part by the Russian Foundation for Basic Research, projects No. 15-07-03051, 15-07-03608.

We thank Professor Konstantin Samouylov from Peoples' Friendship University of Russia for comments that greatly improved the paper.

References

1. Rosenberg, J., et al.: IETF RFC 3261 SIP: Session Initiation Protocol (2002). <http://www.ietf.org/rfc/rfc3261.txt>
2. Rosenberg, J.: IETF RFC 5390 Requirements for Management of Overload in the Session Initiation Protocol (2008). <https://tools.ietf.org/html/rfc5390>
3. Hilt, V., Noel, E., Shen, C., Abdelal, A.: IETF RFC 6357 Design Considerations for Session Initiation Protocol (SIP) Overload, Control (2011). <https://tools.ietf.org/html/rfc6357>
4. Gurbani, V., Hilt, V., Schulzrinne, H.: IETF RFC 7339 Session Initiation Protocol (SIP) Overload Control (LBOC) (2014). <https://tools.ietf.org/html/rfc7339>
5. Noel, E., Williams, P.M.: IETF RFC 7415 Session Initiation Protocol (SIP) Rate Control (RBOC) (2015). <https://tools.ietf.org/html/rfc7415>
6. Samouylov, K.E., Sopin, E.S.: On analysis of M[X]|G|1|r queueing system. Bulletin of Peoples' Friendship University of Russia. Series Mathematics, Information Sciences, Physics, no. 1, pp. 91–97 (2011)
7. Gaidamaka, Y., Pechinkin, A., Razumchik, R., Samouylov, K., Sopin, E.: Analysis of M|G|1|R queue with batch arrivals and two hysteretic overload control policies. Int. J. Appl. Math. Comput. Sci. **24**(3), 519–534 (2014)
8. Gaidamaka, Y.V., Zaripova, E.R., Orlov, Y.N.: Analysis of the impact the batch size distribution on parameters of the SIP-server queueing model with batch arrivals. KIAM Preprint no. 27, pp. 1–16, Moscow (2015). <http://library.keldysh.ru/preprint.asp?id=2015-27>
9. Gaidamaka, Y., Zaripova, E.: Comparison of polling disciplines when analyzing waiting time for signaling message processing at SIP-server. In: Dudin, A., et al. (eds.) ITMM 2015. CCIS, vol. 564, pp. 358–372. Springer International Publishing, Switzerland (2015)
10. Gaidamaka, Y.V.: Model with threshold control for analyzing a server with an SIP protocol in the overload mode. Autom. Control Comput. Sci. **47**(4), 211–218 (2013)
11. Abaev, P., Gaidamaka, Y., Samouylov, K.E.: Modeling of hysteretic signaling load control in next generation networks. In: Andreev, S., Balandin, S., Koucheryavy, Y. (eds.) NEW2AN/ruSMART 2012. LNCS, vol. 7469, pp. 440–452. Springer, Heidelberg (2012)
12. Abaev, P., Gaidamaka, Y., Samouylov, K.E.: Queueing model for loss-based overload control in a SIP server using a hysteretic technique. In: Andreev, S., Balandin, S., Koucheryavy, Y. (eds.) NEW2AN/ruSMART 2012. LNCS, vol. 7469, pp. 371–378. Springer, Heidelberg (2012)

13. Gaidamaka, Y., Pechinkin, A., Razumchik, R., Samouylov, K., Sopin, E.: Analysis of M|G|1|R queue with batch arrivals and two hysteretic overload control policies. *Int. J. Appl. Math. Comput. Sci.* **24**(3), 519–534 (2014)
14. Abaev, P., Gaidamaka, Y., Pechinkin, A., Razumchik, R., Samouylov, K., Shorgin, S.: Hysteretic control technique for overload problem solution in network of SIP servers. *Comput. Inform.* **33**(1), 218–236 (2014)
15. Shorgin, S., Samouylov, K., Gaidamaka, Y., Etezov, S.: Polling system with threshold control for modeling of SIP server under overload. In: *Advances in Intelligent Systems and Computing*, vol. 240, pp. 97–107 (2014)
16. Abhayawardhana, V.S., Babbage, R.: A traffic model for the IP multimedia subsystem (IMS). In: *Proceedings of 65th Vehicular Technology Conference*, pp. 783–787 (2007)
17. Vishnevsky, V., Kozyrev, D., Semenova, O.: Redundant queuing system with unreliable servers. In: *2014 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pp. 283–286 (2014)
18. Abaev, P.O., Gaidamaka, Y.V., Pechinkin, A.V., Razumchik, R.V., Shorgin, S.Y.: Simulation of overload control in SIP server networks. In: *Proceedings of the 26th European Conference on Modelling and Simulation, ECMS 2012, Koblenz, Germany*, pp. 533–539 (2012)
19. Abaev, P.O., Pechinkin, A.V., Razumchik, R.V.: On analytical model for optimal SIP server hop-by-hop overload control. In: *Proceedings of the 4th International Congress on Ultra Modern Telecommunications and Control Systems (IEEE ICUMT-2012), Saint-Petersburg, Russia*, pp. 299–304 (2012)
20. Abaev, P., Gaidamaka, Y., Samouylov, K., Shorgin, S.: Design and software architecture of SIP server for overload control simulation. In: *Proceedings of the 27th European Conference on Modelling and Simulation (ECMS 2013), 27–30 May, Aalesund, Norway*, pp. 580–586 (2013)
21. Samouylov, K., Abaev, P., Gaidamaka, Y., Pechinkin, A., Razumchik, R.: Analytical modelling and simulation for performance evaluation of SIP server with hysteretic overload control. In: *Proceedings of the 28th European Conference on Modelling and Simulation (ECMS 2014), 27–30 May, Brescia, Italy*, pp. 603–609 (2014)
22. Gaidamaka, Y., Pechinkin, A., Razumchik, R.: Time-related stationary characteristics in queueing system with constant service time under hysteretic policy. In: *Proceedings of the 6th International Congress on Ultra Modern Telecommunications and Control Systems (ICUMT 2014), Russia*, pp. 634–640 (2014)
23. Samouylov, K., Gaidamaka, Y., Talanova, M., Pavlotsky, O.: Simulation of SIP-server with hysteretic input and loss-based overload control scheme. In: *Proceedings of the 6th International Congress on Ultra Modern Telecommunications and Control Systems (ICUMT 2014), St. Petersburg, Russia*, pp. 589–594 (2014)
24. Samouylov, K., Gaidamaka, Y., Abaev, P., Talanova, M., Pavlotsky, O.: Analytical modeling of rate-based overload control with token bucket traffic shaping on client side. In: *Proceedings of the 29th European Conference on Modelling and Simulation (ECMS 2015). Digitaldruck Pirrot GmbH, Alben, pp. 669–674 (2015)*