6^{th} *International Conference on Applied Informatics*
Eger, Hungary, January 27–31, 2004.

Multiserver Retrial Queues with Finite Number of Heterogeneous Sources*

János Roszik, Béla Almási, János Sztrik

Department of Informatics Systems and Networks, University of Debrecen
e-mail: {jroszik,almasi,jsztrik}@inf.unideb.hu

Abstract

The aim of this paper is to investigate a multiserver retrial queueing system with a finite number of heterogeneous sources of calls. The novelty of the investigation is the heterogeneity of the sources, that is every source is characterized by different arrival, retrial and service rates. All random variables involved in the model construction are assumed to be exponentially distributed and independent of each other. The present paper generalizes the results for the same type of homogeneous retrial queueing system.

The MOSEL (Modeling, Specification and Evaluation Language) language and software package, which was developed at the University of Erlangen, Germany, was used to formulate the model and to calculate the derived main performance measures. The effects of various system parameters on the system measures are displayed and analyzed graphically.

Categories and Subject Descriptors: C.4 [Performance of Systems]: Modeling Techniques, Performance Attributes; G.3 [Probability and Statistics]: Queueing Theory, Stochastic Processes; I.6.4 [Model Validation and Analysis];

Key Words and Phrases: Retrial queueing systems, finite number of heterogeneous sources, multiserver queues, performance tool, performance measures

1. Introduction

Retrial queues (or queues with returning customers) have been often used to model computer systems and communication networks, such as magnetic disk memory systems [8], cellular mobile networks [9] and local-area networks with

*Research is partially supported by Hungarian Scientific Research Fund OTKA T0-34280/2000 and FKFP grant 0191/2001.

CSMA/CD protocol [7]. For some fundamental methods and results on retrial queues see [2, 3, 5, 6], and references therein.

The purpose of the present paper is to give the main stationary performance measures of the heterogeneous finite-source multiserver model described in the next section, and to analyze graphically the effects of changing various system parameters on these measures. This paper generalizes the results of [6], where homogeneous finite-source multiserver systems were dealt with. The performance modeling tool MOSEL (Modeling, Specification and Evaluation Language), see [4], is used to formulate the model and to obtain the system performance measures. The analytical results can be graphically displayed using IGL (Intermediate Graphical Language) which belongs to the MOSEL tool.

The paper is organized as follows. In Section 2, the accurate description of the model is given, and the main performance measures of the system are derived that can be obtained using the MOSEL tool. In Section 3, some numerical examples are treated. Conclusions and directives for the future work are given in Section 4.

2. Model description

Consider a finite-source multiserver queueing system with c servers, where the primary calls are generated by K, $c < K < \infty$ heterogeneous sources, that is every source has different parameters. The servers can be in two states: idle and busy. If a server is idle, it can serve the calls of the sources. Each of the sources can be in three states: free, sending repeated calls and under service. If the i-th source is free at time t it can generate a primary call during interval $(t, t + dt)$ with probability $\lambda_i dt + o(dt)$. If there is a free server at the time of arrival of a call then the call starts to be served, that is the source moves into the under service state and the corresponding server moves into busy state. The service is finished during the interval $(t, t + dt)$ with probability $\mu_i dt + o(dt)$. If all the servers are busy on arrival, then the source starts generation of a Poisson flow of repeated calls with rate ν_i until it finds a server free. After service the source becomes free, and it can generate a new primary call, and the server becomes idle so it can serve a new call. All the times involved in the model are assumed to be mutually independent of each other.

2.1. The underlying Markov chain

The state of the system at time t can be described with the stochastic process $X(t) = (\alpha_1, ..., \alpha_c, \beta_1, ..., \beta_{N(t)})$, where c is the number of servers and $N(t)$ is the number of sources of repeated calls at time t. Because of the heterogeneity of the sources we need to identify their indices at the servers, they are denoted by α_i, i=$1, ..., c$, if there is a customer under service at the i-th server, otherwise this value is 0. Furthermore, we have to identify the sources in the sending repeated calls state, so we denote their indices by β_j, j=1, ..., $N(t)$, if there is a customer in this state, otherwise this last component is 0.

Because of the exponentiality of the involved random variables this process is a Markov chain with a finite state space. Since the state space of the process $(X(t), t \geq 0)$ is finite, the process is ergodic for all reasonable values of the rates involved in the model construction. We assume that the system is in the steady state, and we define the stationary probabilities as follows.

$$P(i_1, ..., i_c, 0) = \lim_{t \to \infty} \mathrm{P}\{\alpha_1 = i_1, ..., \alpha_c = i_c, N(t) = 0\}$$

$$P(i_1, ..., i_c, j_1, ..., j_k) = \lim_{t \to \infty} \mathrm{P}\{\alpha_1 = i_1, ..., \alpha_c = i_c, \beta_1 = j_1, ..., \beta_k = j_k\},$$

$$k = 1, ..., K - c$$

Because of the fact that the state space of the describing Markov chain is very large, it is very difficult to get the steady state probabilities by solving the system of steady state equations. To simplify this procedure we used the efficient software tool MOSEL to formulate the model and to calculate these probabilities.

The results of the tool are in agreement with the results of [6] in the homogeneous case. In the heterogeneous case, the results were validated by the corresponding FIFO queueing model, since with very high retrial rates and few sources the difference between the two models is negligible.

Let us denote the number of busy servers at time t by $C(t)$, and denote by $p_{kl} = \lim_{t \to \infty} \mathrm{P}\{C(t) = k, N(t) = l\}$ the joint distribution of the number of busy servers and the number of sources that are sending repeated calls.

Knowing the steady state probabilities the main performance measures can be obtained as follows.

- *The probability of all servers are busy*

$$p_c = \mathrm{P}\{\alpha_1 > 0, ..., \alpha_c > 0\} = \sum_{l=0}^{K-c} p_{cl}.$$

- *The mean number of sources of repeated calls*

$$\overline{N} = \mathrm{E}N(t) = \sum_{k=0}^{c} \sum_{l=1}^{K-c} l p_{kl}.$$

- *The probability of the i-th source is sending repeated calls*

$$N_i = \sum_{i_1, ..., i_c} \sum_{k=1}^{K-c} \sum_{\substack{j_1, ..., j_k \\ i \in \{j_1, ..., j_k\}}} P(i_1, ..., i_c, j_1, ..., j_k), \quad i = 1, ..., K.$$

- *The mean number of busy servers*

$$\overline{Y} = \mathrm{E}C(t) = \sum_{k=1}^{c} \sum_{l=0}^{K-c} k p_{kl}.$$

- *The probability of the i-th source is under service*

$$Y_i = \sum_{\substack{i_1,\dots,i_c \\ i \in \{i_1,\dots,i_c\}}} P(i_1, ..., i_c, 0) +$$

$$\sum_{\substack{i_1,\dots,i_c \\ i \in \{i_1,\dots,i_c\}}} \sum_{k=1}^{K-c} \sum_{j_1,\dots,j_k} P(i_1, ..., i_c, j_1, ..., j_k), \quad i = 1, ..., K.$$

- *The utilization of the i-th source*

$$U_i = 1 - N_i - Y_i, \quad i = 1, ..., K.$$

- *The mean rate of generation of primary calls of the i-th source*

$$\overline{\lambda_i} = \lambda_i (1 - Y_i - N_i) = \lambda_i U_i, \quad i = 1, ..., K.$$

- *The mean response time of the i-th source* (based on [1])

$$\overline{T_i} = \frac{N_i + Y_i}{\overline{\lambda_i}}, \quad i = 1, ..., K.$$

- *The mean waiting time of the i-th source*

$$\overline{W_i} = \overline{T_i} - \frac{1}{\mu_i}, \quad i = 1, ..., K.$$

3. Numerical examples

The results of the tool can be displayed graphically with the help of IGL, which belongs to MOSEL. The system parameters of the following figures are given in Table 1.

In Figure 1, the utilizations of the sources are displayed versus the primary request generation rate. As it was expected, the utilizations of the sources decrease as they send their requests more often. With very short interarrival times the difference between the sources is getting very small, so in this case, it is worth to approach with the homogeneous model in order to make the state space much smaller.

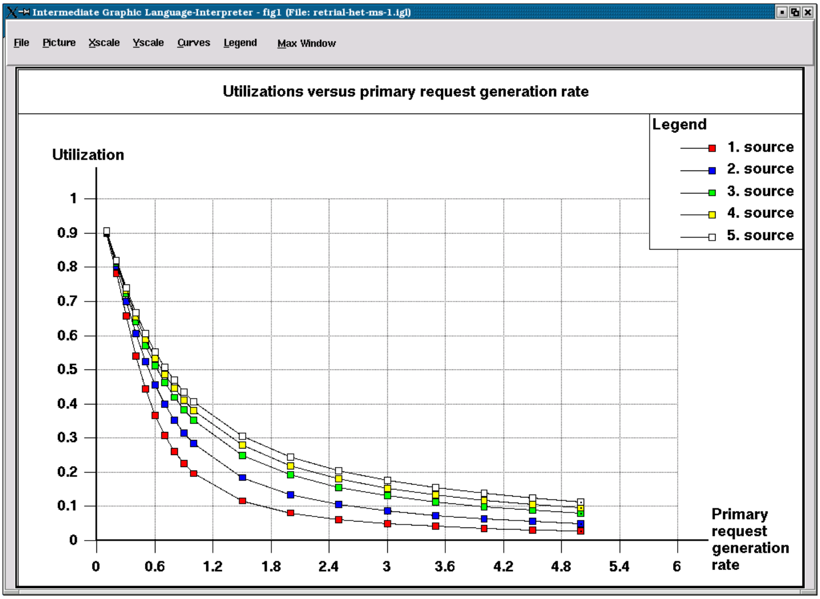

Figure 1: Utilizations versus primary request generation rate

In Figure 2, the mean response times of the sources are represented versus the primary request generation rate. As it can be seen, the first source, which has the smallest retrial rate, has much higher mean response time.

In Figure 3, we can see the mean response times of the customers depending on the retrial rate. If the retrial rate is much higher than the primary request generation rate, the differences between the mean response times are not considerable.

Finally, in Figure 4, the mean number of retrying customers is displayed versus the retrial rate. It is clearly shown, how long the increase in the retrial rate has substantial influence on the mean number of waiting customers.

	c	K	λ	μ	ν
Figure 1,2	2	5	x axis	1,1,1,1,1	0.5,1,2,3,5
Figure 3,4	2	5	0.1,0.2,0.5,0.8,1	1,1,1,1,1	x axis

Table 1: Input system parameters

4. Conclusion and future work

In this paper, we studied a heterogeneous finite-source multiserver queueing system with returning customers. The novelty of the investigation is the heterogeneity of the sources, that is every source is characterized by different arrival,

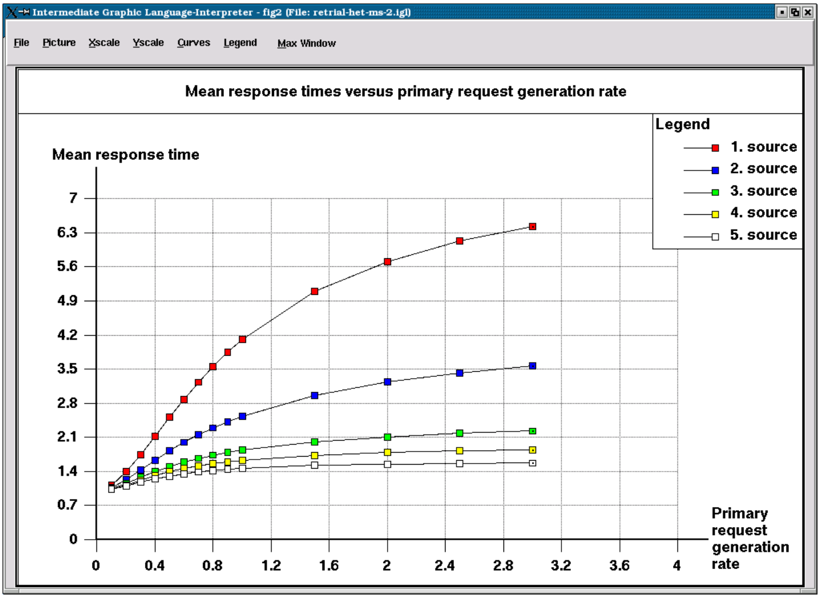

Figure 2: Mean response times versus primary request generation rate

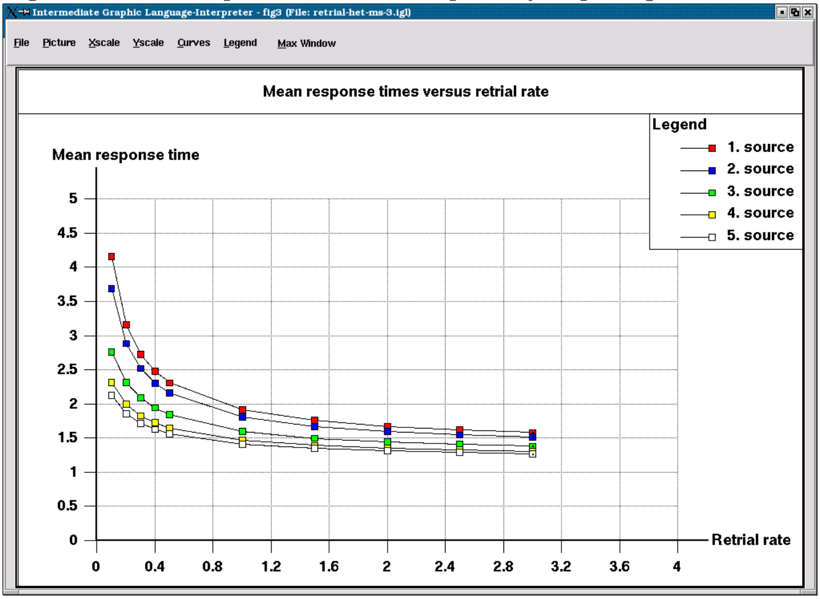

Figure 3: Mean response times versus retrial rate

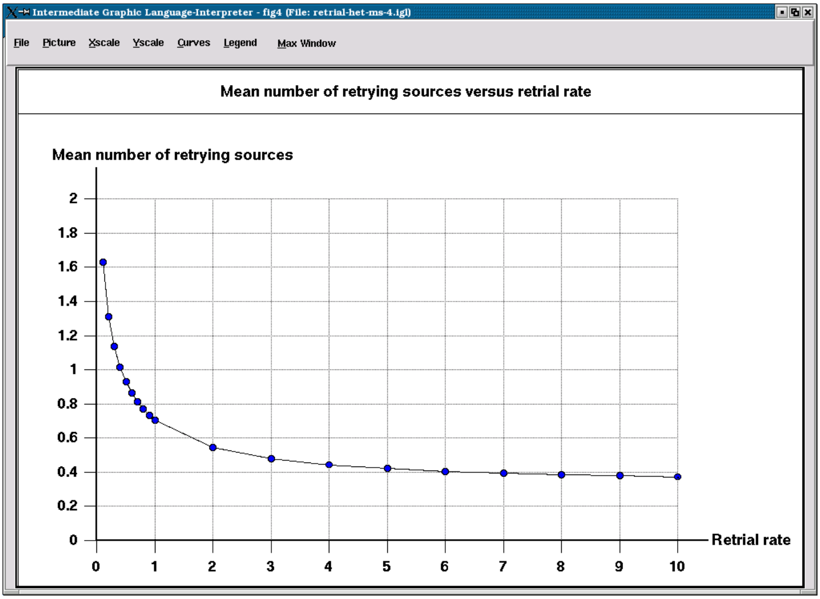

Figure 4: Mean number of retrying sources versus retrial rate

retrial and service rates. The performance modeling tool MOSEL was used to formulate the model and to obtain the system performance measures, and the effects of the changing of the primary request generation and retrial rates were analyzed graphically.

The current study can be considered as an initial step towards the analysis of heterogeneous multiserver retrial queues with non-reliable components.

References

[1] Almási B.: Response Time for Finite Heterogeneous Nonreliable Queueing Systems, *Computers and Mathematics with Applications*, Vol. 31 (1996), 55-59.

[2] Artalejo J.R.: Retrial queues with a finite number of sources, *J. Korean Math. Soc.*, Vol. 35 (1998), 503-525.

[3] Artalejo J.R.: Accessible bibliography on retrial queues, *Math. Comput. Modeling*, Vol. 30 (1999), 1-6.

[4] Begain K., Bolch G. and Herold H.: *Practical performance modeling, application of the MOSEL language*, Kluwer Academic Publisher, Boston, 2001.

[5] Falin G.I.: A survey of retrial queues, *Queueing Systems*, Vol. 7 (1990), 127-168.

[6] Falin G.I. and Templeton J.G.C.: *Retrial queues*, Chapman and Hall, London, 1997.

[7] Li H. and Yang T.: A single server retrial queue with server vacations and a finite number of input sources, *European Journal of Operational Research*, Vol. 85 (1995), 149-160.

[8] Ohmura H. and Takahashi Y.: An analysis of repeated call model with a finite number of sources, *Electronics and Communications in Japan*, Vol. 68 (1985), 112-121.

[9] Tran-Gia P. and Mandjes M.: Modeling of customer retrial phenomenon in cellular mobile networks, *IEEE Journal of Selected Areas in Communications*, Vol. 15 (1997), 1406-1414.

Postal addresses

J. Roszik, B. Almási, J. Sztrik
Department of Informatics Systems and
Networks, University of Debrecen
P.O. Box 12, H-4010 Debrecen
Hungary