A Contribution to Modeling Two-Way Communication with Retrial Queueing Systems

Attila Kuki[✉], János Sztrik, Ádám Tóth, and Tamás Bérczes

University of Debrecen, Debrecen, Hungary
{attila.kuki,janos.sztrik,adam.toth,tamas.berczes}@inf.unideb.hu

Abstract. The goal of this paper is to investigate the two-way communication by the help of finite source retrial queuing systems. Incoming calls from sources (primary calls) arrive to the server according to a Poisson process. If an incoming call finds the server idle, its service starts. Otherwise, if the server is busy, an arriving (primary or secondary - from the orbit) call moves into the orbit and after some exponentially distributed time it retries to enter to the server. When the server is idle it generates an outgoing call after an exponentially distributed time with different parameters to the calls in the orbit and in the sources, respectively. The service time of the incoming and outgoing calls are exponentially distributed with different rates. Results on two-way communication assume, that after the service an outgoing call (primary or secondary) is sent back to the source. The novelty of this paper is investigating two cases. In Case 1 the secondary outgoing call is sent back to the orbit, thus the pending incoming call will not be lost. In Case 2 after service of the secondary outgoing call its incoming service request will be started immediately. This means a two-phase service. The balance equations are solved by the help of MOSEL-2 tool. Graphical results and comparisons of the cases are presented.

Keywords: Finite-source queuing system · Retrial queues
Call centers · Two-way communication

1 Introduction

This paper deals with investigations on systems with two-way communication. These systems can be modeled effectively by the help of retrial queueing systems. The research on two-way communications has been becoming more and more popular topic of investigations for the last years. The main reason is that there are a many application fields which can be modeled by this type of systems. For example, in business organizations, e.g. in call centers where the agents could perform outgoing calls to sell, advertise and promote products and services of the business. It is very important to increase the utilization the server, see for example [1,2,8,13,17,20]. The most important characteristics of two-way

communication is that and idle server can look for calls inside and outside of the system. In other words, it can perform outgoing calls to the source (primary calls) or to the orbit (secondary calls). The first results on infinite source retrial queueing systems with two-way communication was published by Falin [12], followed by some recent ones, see for example [3, 6, 9, 10, 14–16, 18, 19].

Authors has been investigated the case, when a secondary outgoing call after servicing is sent back to the source [11]. The novelty of this paper is, that a more realistic case is considered regarding secondary outgoing calls from the orbit. A call being in the orbit implies that the call still has an unserved incoming request. So far, the server makes a secondary outgoing call from the orbit, serves the request, and sends back the call to the source. In this case the original incoming request of this call remains unserved. In the model presented here the served secondary outgoing call (an outgoing call from the orbit) is sent back to the orbit again, where the call is able to retry his request for servicing the original incoming call. In addition, in this model an other operational mode is investigated. When a secondary outgoing call from the orbit arrives to the server, after serving the outgoing call, the pending incoming request will be served immediately, as well. When this two-phase service is finished, the call is sent back to the source.

The rest of the paper is organized as follows. In Sect. 2 description of the model is given, the corresponding 2-dimensional Markov process is defined. In Sect. 3 the most interesting results obtained by MOSEL-2 tool are presented. Finally, the paper ends with a Conclusion.

2   Model Description and Notations

This paper deals with a finite source retrial queueing model with one server. The work low of the model can be seen on Fig. 1.

In the source there are N calls. Each call can make a primary incoming call (incoming calls in the system) towards the server. The inter-request times are exponentially distributed with parameter λ_1. When the server is idle, it starts serving the call immediately with an exponentially distributed service time with parameter μ_1. After the service is finished, the call goes back to the source. When the incoming call finds the server busy, it is forwarded to the orbit. This secondary incoming jobs from the orbit may retry their requests for service after a random waiting time. The distribution of this period is exponential with parameter ν_1. In the other hand, the idle server after some exponentially distributed period can make an outgoing calls towards the sources (outgoing calls in the system). Two types of outgoing calls are distinguished:

– After an exponentially distributed idle period with parameter λ_2 the server may call a call from the source to be served (primary outgoing call),
– The server is able to make a call from the orbit, as well (secondary outgoing call). It is performed after an exponentially distributed idle period with parameter ν_2.
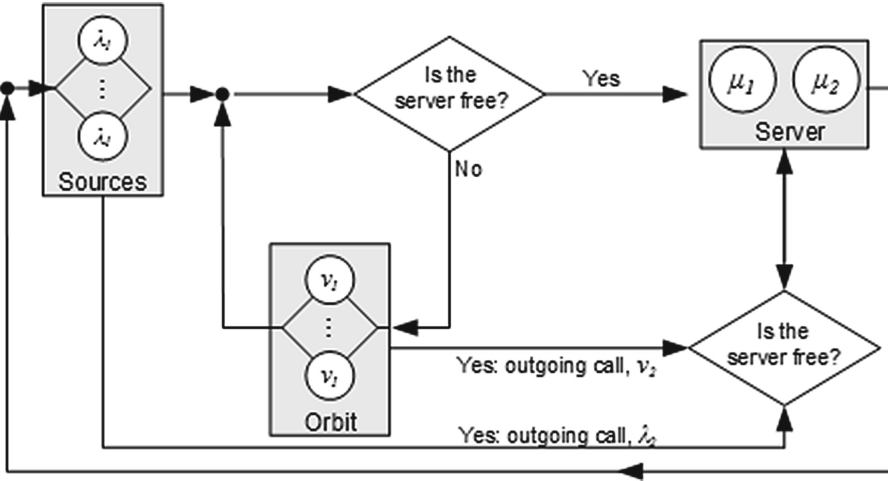
Fig. 1. A retrial queue with components

The outgoing calls (primary and secondary) are served at the server in an exponentially distributed service period with parameter μ_2. A primary outgoing call (called from the source) goes back to the source after the service. When an outgoing call comes from the orbit (secondary outgoing call), two cases can be considered.

– Case 1. The call came from the orbit, which means this call has an unserved incoming request. After the outgoing call is served, this incoming request remains unserved. This call is sent back to the orbit after the outgoing service is finished, thus the call will be able to retry its incoming call,
– Case 2. As in the previous case, after the outgoing call is processed, it has an unserved primary call. In this case the server is able to serve the incoming request immediately after the outgoing job was finished. That means a two-phase service. First the outgoing call is served, after the incoming one. When both of the service phases has been finished, the call is sent back to the source.

It is assumed that the arrivals of primary incoming calls, retrial intervals of secondary incoming calls, service times of incoming and outgoing calls, and the time to make outgoing calls are mutually independent.

We denote the number of calls in orbit and the server state at time t by O(t) and S(t), respectively.

Obviously, when the server is busy the number of calls in the orbit cannot be equal to N, i.e. $O(t) < N$. As stated above in Case 2, after the service both incoming and the two-phase outgoing calls go to a free state. This means that when the server is idle, there will be at least one call in free state, i.e. again $O(t) < N$. Thus, the state space of the process $(S(t), O(t))$ is the set of $\{0, 1, 2\} \times \{0, 1, 2, ..., N-1\}$. In Case 1 after the service the secondary outgoing call goes back to the orbit. This means that when the server is idle, the source

can be empty. Thus, the state space of the process $(S(t), O(t))$ is the set of $\{0, 1, 2\} \times \{0, 1, 2, ..., N\}$.

Because of the finite state space these two-dimensional Markov processes are always stable.

Let us define the state of the server by $S(t)$, that is

$$S(t) = \begin{cases} 0, & \text{when the server is idle} \\ 1, & \text{when an first order request is in service} \\ 2, & \text{when a second order request from source is in service} \\ 3, & \text{when a second order request from orbit is in service} \end{cases}.$$

The used numerical values of the parameters can be seen in Table 1. Some special values of the parameters in the model described above give back models which have been investigated earlier by authors.

– $\lambda_2 = \nu_2 = 0$ provides a classical single server retrial queue studied by e.g. [4,5].
– $\lambda_2 = 0, \mu_2 = \mu_1$ provides a single server retrial queue with two-way communication with search of the customers from the orbit. The reason of the outgoing calls is to shorten the idle period of the server.
– $\mu_2 = \mu_1$ provides a single server retrial queue with two-way communication with search of the customers. The reason of the outgoing calls is again to shorten the idle period of the server.

Table 1. Numerical values of model parameters

Parameter	Symbol	Value
Number of calls in source	N	10
Incoming generation rate	λ_1	[0.1..5.1]
Outgoing generation rate	λ_2	0.2
Incoming retrial rate	ν_1	0.1
Outgoing generation rate from orbit	ν_2	0.2
Incoming service rate	μ_1	1
Outgoing service rate	μ_2	1

It is not difficult to see that the system of balance equations for the stationary probabilities in Case 1 are

$$p_{i,j} = \lim_{t \to \infty} P(S(t) = i, O(t) = j), i = 0, 1, 2, 3 \text{ and } j = 0, 1, ..N$$

$$[(N-j)(\lambda_1 + \lambda_2) + j(\nu_1 + \nu_2)] p_{0,j} = \mu_1 p_{1,j} + \mu_2 p_{2,j} + \mu_2 p_{3,j-1}$$

$$[(N-j-1)\lambda_1 + \mu_1] p_{1,j} = (N-j)\lambda_1 p_{0,j} + (j+1)\nu_1 p_{0,j+1} + (N-j)\lambda_1 p_{1,j-1}$$

$$[(N-j-1)\lambda_1 + \mu_2] p_{2,j} = (N-j)\lambda_2 p_{0,j} + (N-j)\lambda_1 p_{2,j-1}$$

$$[(N - j - 1)\lambda_1 + \mu_2]\, p_{3,j} = (j + 1)\nu_2 p_{0,j+1} + (N - j)\lambda_1 p_{3,j-1}$$

with $p_{1,-1} = p_{2,-1} = p_{3,-1} = 0$.

Similarly, the system of balance equations for the stationary probabilities in Case 2 can be written as

$$p_{i,j} = \lim_{t \to \infty} P(S(t) = i, O(t) = j), i = 0, 1, 2, 3, 4 \ \text{ and } \ j = 0, 1, ..N - 1$$

$$[(N - j)(\lambda_1 + \lambda_2) + j(\nu_1 + \nu_2)]\, p_{0,j} = \mu_1 p_{1,j} + \mu_2 p_{2,j}$$

$$[(N - j - 1)\lambda_1 + \mu_1]\, p_{1,j} =$$
$$= (N - j)\lambda_1 p_{0,j} + (j + 1)\nu_1 p_{0,j+1} + (N - j)\lambda_1 p_{1,j-1} + \mu_2 p_{3,j}$$

$$[(N - j - 1)\lambda_1 + \mu_2]\, p_{2,j} = (N - j)\lambda_2 p_{0,j} + (N - j)\lambda_1 p_{2,j-1}$$

$$[(N - j - 1)\lambda_1 + \mu_2]\, p_{3,j} = (j + 1)\nu_2 p_{0,j+1} + (N - j)\lambda_1 p_{3,j-1}$$

with $p_{0,N} = p_{1,-1} = p_{2,-1} = p_{3,-1} = 0$.

As soon as we have calculated the distributions defined above (by the help of MOSEL-2 tool, see the next section), the most important steady-state system performance measures can be obtained in the following way.

– *Utilization 1*

$$U_1 = \sum_{o=0}^{N} P(1, o)$$

– *Utilization 2*

$$U_2 = \sum_{s=2}^{3} \sum_{o=0}^{N} P(s, o)$$

– *Average number of jobs in the orbit*

$$\overline{O} = \sum_{s=0}^{3} \sum_{o=0}^{N} o P(s, o)$$

– *Average number of active primary users*

$$\overline{M} = N - \overline{O} - U_1 - U_2$$

– *Average generation rate of primary users*

$$\overline{\lambda_1} = \lambda_1 \overline{M}$$

– *Mean time spent in orbit by using Little-formula*

$$\overline{W} = \frac{\overline{O}}{\overline{\lambda_1}}$$

3   Numerical Results

Investigating the functionality and the behavior of the system several numerical calculations were performed. Solving the system balance equations described above the MOSEL-2 tool was used. For Markov-processes it is a very efficient tool. MOSEL-2 is a model description language and are equipped with several model translators. Using these translators third-party performance evaluation tools can be used. For obtaining the stationary system probabilities, here the SPNP tool is used. SPNP performs numerical calculations instead of simulation (see in [7]). From the probabilities the well known system characteristics are also be calculated. The most interesting performance characteristics obtained by these tools are graphically presented in this section. On the figures the lines represent different working assumptions or cases. The applied values for the parameters are listed in Table 2.

Table 2. Numerical values of model parameters

Case studies											
No.	N	λ_1	λ_2	ν_1	ν_2	μ_1	μ_2	C_W	C_1	C_2	C_P
Figure 2	10	0.1..5.1	0.2	0.1	0.2	1	1				
Figure 3	10	0.1..5.1	0.2	0.1	0.2	1	1				
Figure 4	10	0.1..5.1	0.2	0.1	0.2	1	1				
Figure 5	10	0.1..5.1	0.2	0.1	0.2	1	1				
Figure 6	10	0.1..5.1	0.2	0.1	0.2	1	1	100	10	10	1
Figure 7	10	0.1..5.1	0.2	0.1	0.2	1	1	100	10	10	1

On Figs. 2, 3, and 4 the mean waiting time of the calls are represented in function of the incoming generation rate for Case 1, Case 2 and comparing the two cases, respectively.

On the first two figures four cases are displayed. "No outgoing" means, that there are only incoming calls in the system. This is a common finite source retrial system. "Outgoing - Only from source" is for the case, when only primary outgoing calls are performed. The line "Outgoing - Only from orbit" is for secondary outgoing calls only. The fourth line represents the investigated Case 1. The similar lines are on the figure for Case 2. Note that, the "No outgoing" lines are the same. The reason of the virtual deviation is the different scale of axes y.

For these values of parameters except the "Outgoing - Only from source" case an interesting maximum value of the mean waiting time can be observed. When we consider a simple retrial queueing system, it can be found a parameter setting, where this maximum feature can be observed. This is a general characteristics of the retrial queues. With slightly modifications of the parameters the same maximum effect also appears here. On Fig. 4 the Case 1 and Case 2 are compared
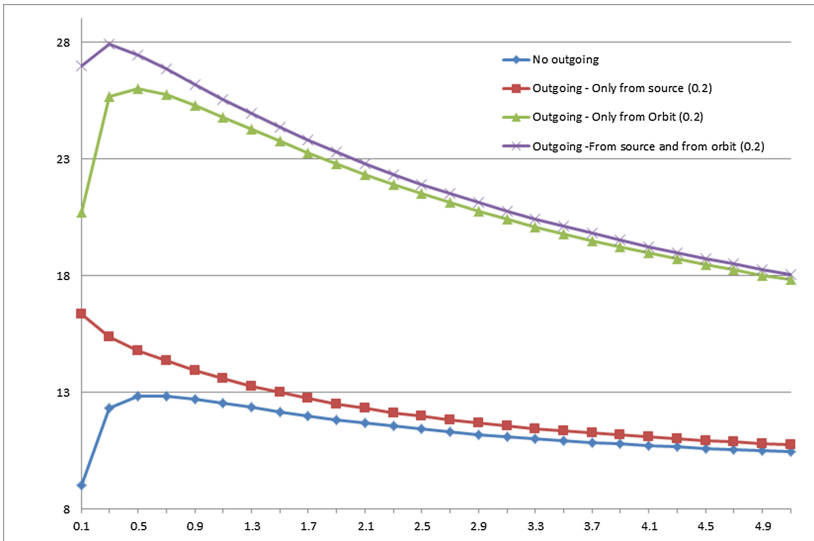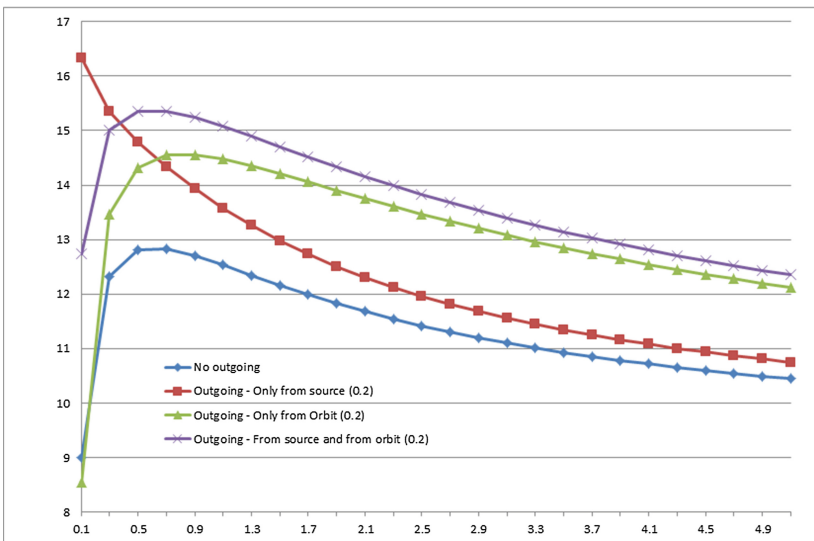
Fig. 2. Mean waiting time (Case 1) vs. λ_1

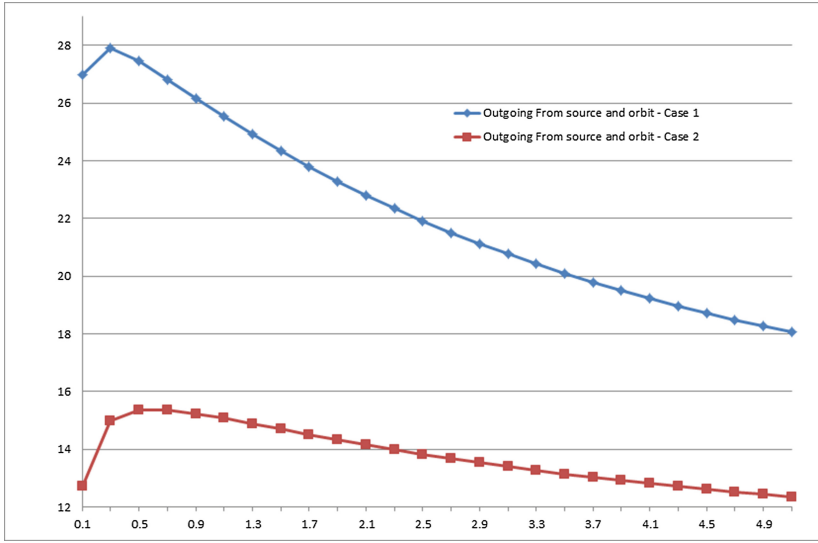**Fig. 3.** Mean waiting time (Case 2) vs. λ_1

Fig. 4. Mean waiting time vs. λ_1

with all of the incoming and outgoing calls. This figure reflects and ensures the expected behaviour of Case 1 and 2.

Figure 5 displays the probability of the busy server in Case 1. That means, the server state can be $S(t) = 1, 2, 3$. The running parameter is the incoming arrival rate, λ_1 again. Pairs of lines can be observed on this figure. One pair is the outgoing and outgoing from orbit only lines, while the other pair is the no outgoing and the outgoing from source only lines. For this set of parameters for smaller values of λ_1 the first pair while for larger values of λ_1 the second pair has larger values, i.e. higher server utilization.

In this type of service or production systems the waiting time of calls and the utilization of the server are singular quantities. They cannot be optimized at the same time. Optimizing the server utilization will increase the waiting time of calls. Some balance or some combined indicator has to be involved. The following expected loss $E(L)$ function enables the "fair" investigation of the system.

$$E(L) = C_w(1 - U_1 - U_2) + C_1\mu_1 U_1 + C_2\mu_2 U_2 + C_P(E(O) + U_1 + U_2).$$

The first component is the loss on idle state of the system. The second and the third components are the cost of servicing incoming and outgoing calls, respectively. Here the speed of the service has to be taken into consideration, thus beside the cost weights the service rates are present as multiplicative factors. The last component states the loss of the system not in production from the point of view of the calls, i.e. it is the sojourn time of the call: it is under service or it is in the virtual waiting facility (orbit). For a given set of parameters (listed in Table 2) the shape of loss functions can be seen on Figs. 6 and 7.
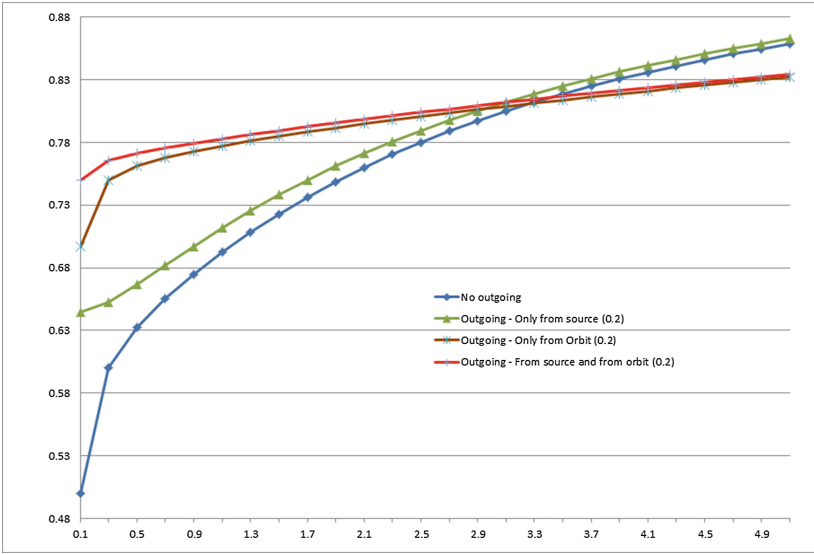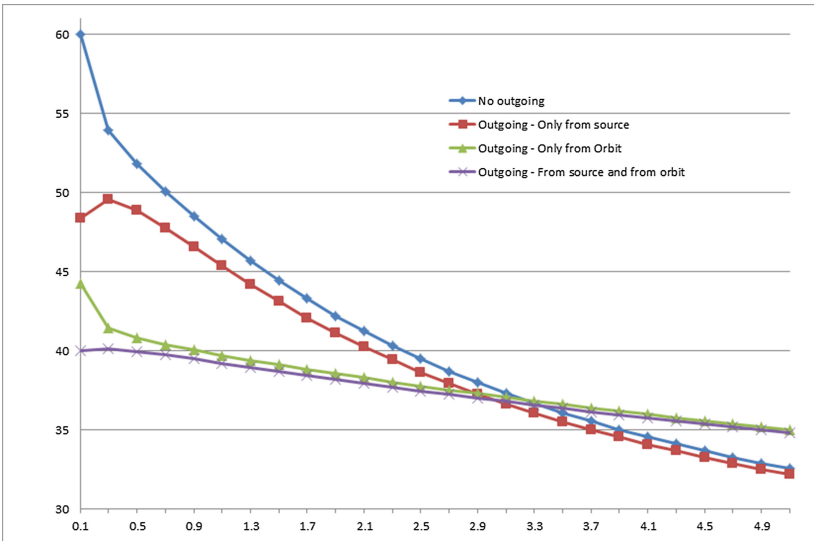
Fig. 5. Probability of server is busy vs. λ_1

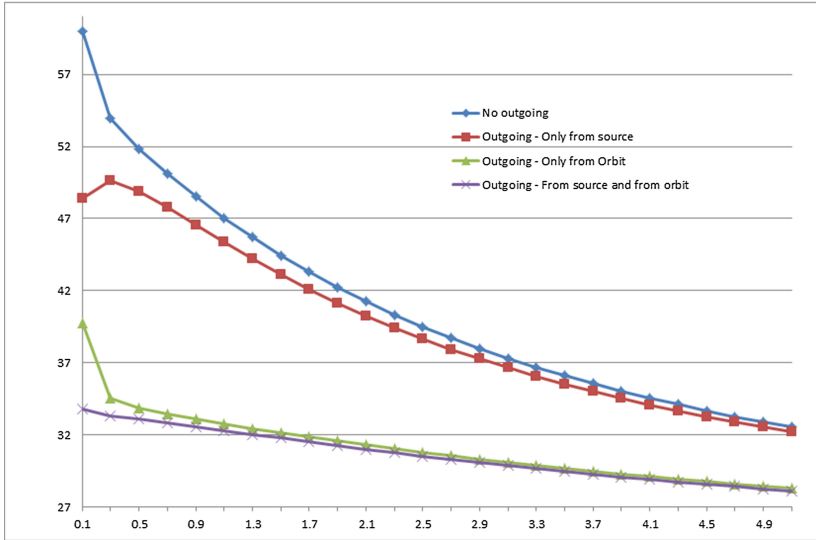**Fig. 6.** The loss function in Case 1 vs. λ_1

Fig. 7. The loss function in Case 2 vs. λ_1

In Case 1 and Case 2 it can be observed that E(L) values are much lower in the two-way cases, especially when the orbit is involved. In Case 1 these values are higher than in Case 2. It is the effect of the two-phase service of the calls. For the one-phase service (Case 1) it is interesting, that large values of the incoming rate implies lower values of loss functions for incoming only and outgoing from source only cases (see the line intersections on Fig. 6). The same effect can be seen on Fig. 5, as well. The reason of this similar behavior is that the first three components of the loss function contain the effect of the utilizations, and these components have large weights.

4   Conclusion

This paper gives a contribution to the model described in [11]. The original model stated, that a secondary outgoing call is sent back to the source after service. Let's consider a bank, where the calls are called to give some signature sample (outgoing calls). These calls can be outside the bank (in free state) or inside the bank, waiting for some transaction (incoming calls in the orbit). When the call connected from the orbit for the outgoing call, it is quite natural not send it outside the bank but keep it inside (Case 1) or after the signature perform its original transaction request (Case 2). The numerical results proof that in Case 2 the most important performance measures (waiting time, utilization etc.) are better than in Case 1. A loss function keeping balance between utilization and waiting times has been also introduced. In the future it would be interesting to investigate the sensitivity of the loss function to the parameter changing.

References

1. Aguir, S., Karaesmen, F., Akşin, O.Z., Chauvet, F.: The impact of retrials on call center performance. OR Spectr. **26**(3), 353–376 (2004)
2. Aksin, Z., Armony, M., Mehrotra, V.: The modern call center: a multi-disciplinary perspective on operations management research. Prod. Oper. Manag. **16**(6), 665–688 (2007)
3. Artalejo, J.R., Phung-Duc, T.: Markovian retrial queues with two way communication. J. Ind. Manag. Optim. **8**(4), 781–806 (2012)
4. Artalejo, J.: Retrial queues with a finite number of sources. J. Korean Math. Soc. **35**, 503–525 (1998)
5. Artalejo, J., Corral, A.G.: Retrial Queueing Systems: A Computational Approach. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78725-9
6. Artalejo, J., Phung-Duc, T.: Single server retrial queues with two way communication. Appl. Math. Model. **37**(4), 1811–1822 (2013)
7. Begain, K., Bolch, G., Herold, H.: Practical Performance Modeling. Application of the MOSEL Language. Kluwer Academic Publisher, Boston (2001)
8. Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., Zhao, L.: Statistical analysis of a telephone call center: a queueing-science perspective. J. Am. Stat. Assoc. **100**(469), 36–50 (2005)
9. Dimitriou, I.: A retrial queue to model a two-relay cooperative wireless system with simultaneous packet reception. In: Wittevrongel, S., Phung-Duc, T. (eds.) ASMTA 2016. LNCS, vol. 9845, pp. 123–139. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-43904-4_9
10. Dragieva, V., Phung-Duc, T.: Two-way communication M/M/1 retrial queue with server-orbit interaction. In: Proceedings of the 11th International Conference on Queueing Theory and Network Applications, p. 11. ACM (2016)
11. Dragieva, V., Phung-Duc, T.: Two-way communication M/M/1//N retrial queue. In: Thomas, N., Forshaw, M. (eds.) ASMTA 2017. LNCS, vol. 10378, pp. 81–94. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-61428-1_6
12. Falin, G.: Model of coupled switching in presence of recurrent calls. Eng. Cybern. **17**(1), 53–59 (1979)
13. Gans, N., Koole, G., Mandelbaum, A.: Telephone call centers: tutorial, review, and research prospects. Manuf. Serv. Oper. Manag. **5**(2), 79–141 (2003)
14. Nazarov, A., Phung-Duc, T., Paul, S.: Heavy outgoing call asymptotics for $MMPP/M/1/1$ retrial queue with two-way communication. In: Dudin, A., Nazarov, A., Kirpichnikov, A. (eds.) ITMM 2017. CCIS, vol. 800, pp. 28–41. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68069-9_3
15. Nazarov, A.A., Paul, S., Gudkova, I., et al.: Asymptotic analysis of Markovian retrial queue with two-way communication under low rate of retrials condition. In: Proceedings 31st European Conference on Modelling and Simulation (2017)

16. Phung-Duc, T., Rogiest, W.: Two way communication retrial queues with balanced call blending. In: Al-Begain, K., Fiems, D., Vincent, J.-M. (eds.) ASMTA 2012. LNCS, vol. 7314, pp. 16–31. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30782-9_2
17. Pustova, S.: Investigation of call centers as retrial queuing systems. Cybern. Syst. Anal. **46**(3), 494–499 (2010)
18. Sakurai, H., Phung-Duc, T.: Two-way communication retrial queues with multiple types of outgoing calls. Top **23**(2), 466–492 (2015)
19. Sakurai, H., Phung-Duc, T.: Scaling limits for single server retrial queues with two-way communication. Ann. Oper. Res. **247**(1), 229–256 (2016)
20. Wolf, T.: System and method for improving call center communications, US Patent App. 15/604,068, 30 November 2017