Performance Analysis and Optimization of a Queueing Model for a Multi-skill Call Center in M-Design

Dequan Yue, Chunyan Li and Wuyi Yue

Abstract This paper studies a queuing model of a multi-skill call center in M-design. In this model, there are two types of customers and three groups of servers who have different skills. Servers in Group 1 can only serve type 1 customers, servers in Group 2 can only serve type 2 customers, and servers in Group 3 can serve both type 1 and type 2 customers. We obtain the state-transition rates by using results from M/M/c/c and M/M/c queueing systems. Then, we establish equations for the steady-state probabilities of the system. Finally, we obtain the computational formula for the service level and we present an optimization of a staffing problem.

Keywords Multi-skill call center · Queuing model · Steady-state probabilities · Service level · Optimization

---

D. Yue(✉)
College of Science, Yanshan University, Qinhuangdao 066004, China
email: ydq@ysu.edu.cn

C. Li
School of Economics and Management, Yanshan University, Qinhuangdao 066004, China

C. Li
Zhijiang College of Zhejiang University of Technology, Hangzhou 310024, China
email: llccyy1980@126.com

W. Yue
Department of Intelligence and Informatics, Konan University, Kobe 658-8501, Japan
email: yue@konan-u.ac.jp

T.V. Do et al. (eds.), *Queueing Theory and Network Applications*,
Advances in Intelligent Systems and Computing 383,
DOI: 10.1007/978-3-319-22267-7_14

1 Introduction

Call centers are becoming increasingly important in the global business environment. Correspondingly, as the importance and complexity of modern call centers grow, there is a proliferation of literature relating to them, typically focusing on queueing models. In a queueing model of a call center, the call agents and calls correspond to servers and customers, respectively. For important related surveys, we refer to Koole and Mandelbaum [1] and Gans et al. [2]. An introduction to staffing problems with relevant bibliographic references can be found in Aksin et al. [3].

Multi-skill call centers have emerged and have recently been studied in the literature. A multi-skill call center handles several types of calls, and each agent has a selected number of skills. The agents are distinguished by the set of call types they can handle. A typical example is an international call center where incoming calls are in different languages, see Gan et al. [2].

Perry and Nilsson [4] considered a multi-skill call center with two classes of calls that are served by a single pool of agents. They determined the required number of agents and an assignment policy to satisfy a target for the expected waiting times of callers. Such multi-skill call centers are referred to as V-models or V-designs. Bhulai and Koole [5] proposed scheduling policies and showed that the policy is optimal for equal service time distributions. Gans and Zhou [6] also studied the same V-design model using a linear programming approach. They obtained results for the case of unequal service rates.

Örmeci [7] studied a dynamic admission control for a multi-skill call center in M-design where there are two classes of calls and three stations: one dedicated to each class, and one shared station. He showed that serving a call in its assigned station, whenever possible, is optimal. In this paper, we study an M-design model for a multi-skill call center by using a queueing model. We focus on the performance analysis and optimization for this M-design model of a multi-skill call center.

The rest of the paper is organized as follows. In Section 2, we describe the M-design model for a multi-skill call center. In Section 3, we obtain the state-transition rates by using results of M/M/c/c and M/M/c queueing systems. Then, we establish equations for the steady-state probabilities of the system. In Section 4, we obtain the computational formula for the service level and present a staffing problem. Section 5 concludes the paper.

2 System Model

In this paper, we study an M-design model for a multi-skill call center where there are two types of calls and three groups of servers.

1. **Arrival Process:** There are two types of calls (or customers). The calls of Type 1 and Type 2 arrive according to a Poisson process with rates λ_1 and λ_2 ,

respectively. Arriving calls are lined in two queues. Queue 1 and Queue 2 consist of calls of Type 1 and Type 2, respectively. There are infinite waiting spaces for both queues.

2. **Service Process:** There are three groups of servers (agents). Group 1, Group 2 and Group 3 consist of N_1 servers, N_2 servers and N_3 servers, respectively. Group 1 and Group 2 are specialized servers who can only serve customers of Type 1 and Type 2 calls, respectively. Group 3 is made up of flexible servers who can serve customers of both Type 1 and Type 2. The service times of servers in Group 1, Group 2 and Group 3 are all exponentially distributed with parameters μ_1, μ_2 and μ_3, respectively.
3. **Routing Policy:** A arriving customer of Type 1 (or Type 2) have priority to be served by a server in Group 1 (or Group 2) if there are free servers in Group 1 (or Group 2) and free servers in Group 3. If all servers in Group 1 (Group 2) are busy, the customer will be serviced by a free server in Group 3. If all servers are busy in Group 3, the customer must wait in Queue 1 or Queue 2.
4. **Queuing Discipline:** A free server in Group 1 (Group 2) serve the waiting customers in Queue 1 (Queue 2) according to a First-come First-served (FCFS) discipline, and a free server in Group 3 serve the waiting customers in Queue 1 or Queue 2 according to FCFS discipline. If all servers in Group 1 and Group 2 are busy and there are waiting customers both in Queue 1 and Queue 2, a free server in Group 3 will select randomly (with equal probability) a customer of Type 1 and Type 2 for service.

3 The Steady-State Probabilities

In this section, we firstly define the states of the model. Then, we derive the state-transition rates by using results of M/M/c/c and M/M/c queueing systems. Finally, we obtain equations for the steady-state probabilities of the system. It is assumed that the system is stationary.

3.1 The State Space

The M-design model has three skill groups. Each group has three states: an idle state (denoted by 1), that being the case where at least one agent is idle; a busy state (denoted by 2), the case where all agents in the group are busy and there no calls waiting for service rendered by this group; and an overload state (denoted by 3), the case where all agents in the group are busy and there is at least one call waiting for service by this group. Theoretically, the system has 27 states. However, due to the routing policy assumed in Section 2, the 15 states above marked in boldface do not exist. Therefore, the state space actually consists of 12 states, which are given by

$$E = \{(111),(121),(122),(132),(211),(212),(222),(221),(232),(312),(322),(332)\}$$

Let $n_1(n_2)$ be the number of customers waiting for service including those being serviced by servers of Group 1 (2), and n_3 be the number of customers being serviced by servers of Group 3. The i th state in E is denoted by $S_i, i=1,2,...,12$.

3.2 The Calculation of the State-Transition Rates

The transition of states occurs due to either the arrival of a call or the completion of a service.

1. The state-transition due to the arrival of calls.

Consider the state S_1. The trigger for the transfer from state S_1 to state S_3 is a call of Type 1. The transition rate q_{1-3} from state S_1 to state S_3 is given as follows:

$$q_{1-3} = \lambda_1 P(n_1 = N_1 - 1) \tag{1}$$

where $P(n_1 = N_1 - 1)$ is the probability that there are $N_1 - 1$ calls of Type 1 needing to be serviced by the agents in Group 1.

Note that if the process is in the state S_1 then the number of calls of either Type 1 or Type 2 is less than the number of the agents either in Group 1 or Group 2, i.e., $n_1 < N_1$ and $n_2 < N_2$. In this case, each queue behaves like an M/M/c/c loss queuing system. Thus, using the results of the M/M/c/c loss queuing system, we have

$$P(n_1 = N_1 - 1) = \frac{\rho_1^{N_1-1}}{(N_1-1)!\sum_{j=0}^{N_1}\frac{\rho_1^j}{j!}} \tag{2}$$

where $\rho_1 = \frac{\lambda_1}{\mu_1}$. Similarly, the other transition rates q_{i-j} caused by the arrival of calls are given as follows:

$$q_{1-2} = q_{3-5} = q_{6-8} = q_{9-11} = \lambda_2 P(n_2 = N_2 - 1),$$

$$q_{1-3} = q_{2-5} = q_{4-8} = q_{7-10} = \lambda_1 P(n_1 = N_1 - 1),$$

$$q_{2-4} = \lambda_2 P^1(n_3 = N_3 - 1),\ q_{3-6} = \lambda_1 P^2(n_3 = N_3 - 1),$$

$$q_{5-8} = (\lambda_1 + \lambda_2) P^3(n_3 = N_3 - 1),$$

$$q_{4-7} = q_{8-10} = q_{11-12} = \lambda_2,\ q_{6-9} = q_{8-11} = q_{10-12} = \lambda_1$$

where

$$P(n_2 = N_2 - 1) = \frac{\rho_2^{N_2-1}}{(N_2-1)!\sum_{j=0}^{N_2}\frac{\rho_2^j}{j!}}, \tag{3}$$

$$P^1(n_3 = N_3 - 1) = \frac{\rho_4^{N_3-1}}{(N_3-1)!\sum_{j=0}^{N_3}\frac{\rho_4^j}{j!}}, \tag{4}$$

$$P^2(n_3 = N_3 - 1) = \frac{\rho_3^{N_3-1}}{(N_3-1)!\sum_{j=0}^{N_3}\frac{\rho_3^j}{j!}}, \tag{5}$$

$$P^3(n_3 = N_3 - 1) = \frac{(\rho_3+\rho_4)^{N_3-1}}{(N_3-1)!\sum_{j=0}^{N_3}\frac{(\rho_3+\rho_4)^j}{j!}} \tag{6}$$

where $\rho_2 = \frac{\lambda_2}{\mu_2}$, $\rho_3 = \frac{\lambda_1}{\mu_3}$, $\rho_4 = \frac{\lambda_2}{\mu_3}$.

5. The state-transition due to the completion of a service of a call.

Consider the state S_2. The trigger of the transfer from state S_2 to state S_1 is due to a service completion of a call of Type 2. If the state process is in state S_2, then all N_2 agents are busy. Thus, the transition rate from state S_2 to state S_1 is $q_{2-1} = N_2\mu_2$. Similar analysis gives the other transition rates q_{i-j} caused by a service completion which are given as follows:

$q_{2-1} = q_{5-3} = q_{8-6} = q_{11-9} = N_2\mu_2$, $q_{3-1} = q_{5-2} = q_{8-4} = q_{10-7} = N_1\mu_1$,

$q_{4-2} = q_{6-3} = q_{8-5} = N_3\mu_3$, $q_{7-4} = q_{10-8} = (N_2\mu_2 + N_3\mu_3)P(n_2 = N_2 + 1)$,

$q_{9-6} = q_{11-8} = (N_1\mu_1 + N_3\mu_3)P(n_1 = N_1 + 1)$,

$$q_{12-11} = (N_2\mu_2 + \frac{1}{2}N_3\mu_3)P(n_2 = N_2 + 1),$$

$$q_{12-10} = (N_1\mu_1 + \frac{1}{2}N_3\mu_3)P(n_1 = N_1 + 1)$$

where the probabilities of $P(n_1 = N_1 + 1)$ and $P(n_2 = N_2 + 1)$ can be obtained by using the results of the M/M/c queuing system which are given as follows:

$$P(n_1 = N_1 + 1) = \frac{\rho_1^{N_1+1}}{N_1(N_1)!}P_0^1, \tag{7}$$

$$P(n_2 = N_2 + 1) = \frac{\rho_2^{N_2+1}}{N_2(N_2)!}P_0^2 \tag{8}$$

where

$$P_0^1 = \left[\sum_{j=0}^{N_1-1}\frac{\rho_1^j}{j!} + \frac{N_1\rho_1^{N_1}}{N_1!(N_1-\rho_1)}\right]^{-1}, \tag{9}$$

$$P_0^2 = \left[\sum_{j=0}^{N_2-1}\frac{\rho_2^j}{j!} + \frac{N_2\rho_2^{N_2}}{N_2!(N_2-\rho_2)}\right]^{-1}. \tag{10}$$

3.3 The Equations for the Steady-State Probabilities

Let $P_i, i = 1, 2, ..., 12$ be the steady-state probabilities of the state process. Then, we can obtain the equations for the steady-state probabilities of the system as follows:

$$P_1(q_{1-2} + q_{1-3}) = P_2q_{2-1} + P_3q_{3-1},$$

$$P_2(q_{2-1} + q_{2-4} + q_{2-5}) = P_1q_{1-2} + P_4q_{4-2} + P_5q_{5-2},$$

$$P_3(q_{3-1} + q_{3-5} + q_{3-6}) = P_1q_{1-3} + P_5q_{5-3} + P_6q_{6-3},$$

$$P_4(q_{4-2} + q_{4-7} + q_{4-8}) = P_2q_{2-4} + P_7q_{7-4} + P_8q_{8-4},$$

$$P_5(q_{5-2} + q_{5-3} + q_{5-8}) = P_2q_{2-5} + P_3q_{3-5} + P_8q_{8-5},$$

$$P_6(q_{6-3} + q_{6-8} + q_{6-9}) = P_3q_{3-6} + P_8q_{8-6} + P_9q_{9-6},$$

$$P_7(q_{7-4} + q_{7-10}) = P_4q_{4-7} + P_{10}q_{10-7},$$

$$P_8(q_{8-4}+q_{8-5}+q_{8-6}+q_{8-10}+q_{8-11})=P_4q_{4-8}+P_5q_{5-8}+P_6q_{6-8}+P_{10}q_{10-8}+P_{11}q_{11-8},$$

$$P_9(q_{9-6}+q_{9-11})=P_6q_{6-9}+P_{11}q_{11-9},$$

$$P_{10}(q_{10-7}+q_{10-8}+q_{10-12})=P_7q_{7-10}+P_8q_{8-10}+P_{12}q_{12-10},$$

$$P_{11}(q_{11-8}+q_{11-9}+q_{11-12})=P_8q_{8-11}+P_9q_{9-11}+P_{12}q_{12-11},$$

$$P_{12}(q_{12-10}+q_{12-11})=P_{10}q_{10-12}+P_{11}q_{11-12},$$

$$\sum_{i=1}^{12}P_i=1\cdot$$

All the steady-state probabilities can be obtained by solving these equations. However, the calculation of these probabilities are very cumbersome. In next section, these probabilities are calculated numerically by using Matlab software.

4 Optimization Problem

In this section, we first consider the calculation of the service level. Then, we consider the staffing problem.

4.1 The Calculation of the Service Level

The service level is defined as the percentage of the serviced calls in a given fixed waiting time. Actually, the 80/20 principle is a general rule, that is to say at least 80 percent of the calls should be serviced within a 20 second waiting time. We can derive the service level using the steady-state probabilities.

Let P_{sl}^1 and P_{sl}^2 be the probabilities that the call of Type 1and Type 2 is serviced in a fixed time T_1 and T_2, respectively. Let $P_{ns}^1=1-P_{sl}^1$ and $P_{ns}^2=1-P_{sl}^2$.

Consider a call of Type 1. Calls of Type 1 form a queue when the process is in states S_9, S_{11}, S_{12}. It can be seen that the service rate for calls of Type 1 in each state of S_9 and S_{11} is $N_1\mu_1+N_3\mu_3$, and the service rate in state S_{12} is $N_1\mu_1+\frac{1}{2}N_3\mu_3$, Thus, we get the probability P_{ns}^1 that a call of Type 1 can not be serviced in time T_1 as follows:

$$P_{ns}^1=P_9\sum_{i=k_1}^{\infty}P(n_1=i)+P_{11}\sum_{i=k_1}^{\infty}P(n_1=i)+P_{12}\sum_{i=k_2}^{\infty}P(n_1=i) \quad (11)$$

where

$$k_1 = N_1 + N_3 + T_1(N_1\mu_1 + N_3\mu_3), \tag{12}$$

$$k_2 = N_1 + \frac{1}{2}N_3 + T_1(N_1\mu_1 + \frac{1}{2}N_3\mu_3). \tag{13}$$

Similarly, we get the probability P_{ns}^2 that a call of Type 2 can not be served in time T_2 as follows:

$$P_{ns}^2 = P_7\sum_{i=k_3}^{\infty}P(n_2 = i) + P_{10}\sum_{i=k_3}^{\infty}P(n_2 = i) + P_{12}\sum_{i=k_4}^{\infty}P(n_2 = i) \tag{14}$$

where

$$k_3 = N_2 + N_3 + T_2(N_2\mu_2 + N_3\mu_3), \tag{15}$$

$$k_4 = N_2 + \frac{1}{2}N_3 + T_2(N_2\mu_2 + \frac{1}{2}N_3\mu_3). \tag{16}$$

Remark 1. The probability $P(n_1 = i)$ [$P(n_2 = i)$] in Eq. (11)[Eq. (14)] is the probability that there are i customers in the M/M/ N_1 [N_2] queuing system with arrival rate λ_1 [λ_2] and service rate μ_1 [μ_2] which is given in [8]. Their expressions are omitted.

Table 1 gives numerical results for the service levels P_{sl}^1 and P_{sl}^2 for λ_1=5, μ_1=0.5, λ_2=4, μ_2=0.3, μ_3=0.2, T_1=20, T_2=30.

Table 1 The numerical results of the service levels P_{sl}^1 and P_{sl}^2

N_1	N_2	N_3	P_{sl}^1	P_{sl}^2
20	20	20	1.0000	1.0000
20	15	25	1.0000	0.9765
15	20	25	1.0000	1.0000
15	20	10	0.9988	0.9995
15	15	15	0.9994	0.9547
20	15	10	1.0000	0.8693
15	15	10	0.9984	0.9135
11	15	10	0.8200	0.8983
11	14	11	0.8028	0.6677
11	16	9	0.8075	0.9551

4.2 *Optimization of a Staffing Problem*

Let C_1, C_2 and C_3 be the costs for each server in Group 1, Group 2 and Group 3, respectively. In order to minimize the cost, we try to find the optimal number of servers N_1, N_2 and N_3 subject to the constraint conditions. The optimization of a staffing problem can be expressed as follows:

$$\begin{aligned} \min \quad & C_1N_1 + C_2N_2 + C_3N_3 \\ \text{s.t.} \quad & P_{sl}^1 \geq \alpha_1, \\ & P_{sl}^2 \geq \alpha_2, \\ & N_1, N_2, N_3 \in Z^+ \end{aligned}$$

where α_1 and α_2 are the given service rate of the call of Type 1 and the call of Type 2, Z^+ denote the set of positive integer.

5 Conclusions

This paper has studied a queuing model of the M-design multi-skill call center. We have obtained the transition rates of states by using results from an M/M/c/c and M/M/c queueing system and then established equations for the steady-state probabilities of the system. We have derived the computational formula for the service level and presented an optimization of a staffing problem. In this work, we studied an exponential model in a multi-skill call center. A further extension for future research would be to study non-exponential models or models with impatient customers.

Acknowledgments This work is supported in part by the National Natural Science Foundation of China (No. 71071133) and the MEXT, Japan.

References

1. Koole, G., Mandelbaum, A.: Queueing models of call centers: an introduction. Annals of Operations Research **113**, 41–59 (2002)
2. Gans, N., Koole, G., Mandelbaum, A.: Telephone call centers: tutorial, review, and research prospects. Manufacturing & Service Operation Management **5**(2), 79–141 (2003)
3. Aksin, Z., Armony, M., Mehrotra, V.: The modern call-center: a multi-disciplinary perspective on operations management research. Production and Operations Management **16**, 655–688 (2007)

4. Perry, M., Nilsson A.: Performance modeling of automatic call distributors: assignable grade of service staffing. In: Proceedings of the 14th International Switching Symposium, pp. 294–298 (1992)
5. Bhulai, S., Koole, G.: A queueing model for call blending in call centers. IEEE Transactions on Automatic Control **48**, 1434–1438 (2003)
6. Gans, N., Zhou, Y.: A call-routing problem with service-level constraints. Operations Research **51**(2), 255–271 (2003)
7. Örmeci, L.E.: Dynamic admission control in a call center with one shared and two dedicated service facilities. IEEE Transactions on Automatic Control **49**(7), 1157–1161 (2004)
8. Gross, D., Harris, C.M.: Fundamentals of Queueing Theory, 2nd edn. Wiley, New York (1985)