

Performance Analysis of M/M/1/N Queue System with Delayed Startup Time and Partial Failure

Huijun Lu, Yang Song

College of Science, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu
Email: luhuijun@nuaa.edu.cn

Received: Aug. 3rd, 2020; accepted: Aug. 14th, 2020; published: Aug. 21st, 2020

Abstract

This paper focuses on an M/M/1/N queuing system with delayed startup time and partial failure. In such a queuing system, the server needs to experience a random time of start-up from the shut-down period to the normal working period which follows an exponential distribution. Partial failure is not to immediately stop service for repair. It means that the system no longer admits any customer from outside, while it conducts service for the existing customers in the system with a lower rate. After all customers finishing the service and leaving, the system enters into the repair period immediately. By analyzing the two dimensional continuous-time Markov chain in this paper, the Quasi Birth and Death (QBD) process for the system is established. By using the censoring technique, the stable distribution of the system is obtained according to the matrix geometric solution method. Therefore, the sensitivity of the system to parameters is further analyzed by assigning different values to the single parameter.

Keywords

M/M/1/N, Delayed Startup Time, Partial Failure, Censoring Technique, Matrix-Geometric Solution

带有延迟启动时间和部分故障的M/M/1/N排队系统的性能分析

卢会军, 宋 旸

南京航空航天大学理学院, 江苏 南京
Email: luhuijun@nuaa.edu.cn

收稿日期: 2020年8月3日; 录用日期: 2020年8月14日; 发布日期: 2020年8月21日

摘要

本文研究的是一个带有延迟启动时间和部分故障的M/M/1/N排队系统。在此排队系统中, 服务台从关闭期到正常工作期间, 需要经过一段随机时长服从指数分布的启动期。本文中所研究的部分故障指的是, 系统发生故障后并不会立即停止服务, 而是以一个较低的速率服务系统中现存的顾客, 同时拒绝新到达的顾客进入系统中, 当所有现存顾客服务完成并离开后系统立即进入修复期。易知本文中的二维连续马尔可夫过程为拟生灭过程(QBD), 利用删失技巧及矩阵分析方法便可得到该系统的平稳分布, 如此, 赋予各参数不同的值, 便可进一步分析系统对各参数的敏感度。

关键词

M/M/1/N, 延迟启动时间, 部分故障, 删失技巧, 矩阵分析方法

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

排队论是对排队等候过程中出现的拥挤和延误进行的数学研究, 其研究内容涉及建模、性能、控制、仿真、计算、推理和优化等各个方面。排队论研究等待服务的每一个组成部分, 包括到达过程、服务过程、服务台数量、等待空间的容量、排队规则和顾客数量(可能是人、数据包、汽车等)。

丹麦数学家 Erlang [1]在一个世纪以前第一次在他的论文中介绍了排队论, 用来研究电话通话问题, 从此开创了这门应用数学学科。Gnedenko 和 Kovalenko [2]讨论了带有部分故障的排队系统并且研究了部分故障在 M/G/1 系统中的影响。部分故障意味着系统在正常工作期间的任意时间点都可能会发生故障, 但是, 当系统发生故障时, 服务不会立即完全停止, 而是继续以一个较低的速率继续工作, 并且不再接受其他外来的顾客。Avi-Itzhak 和 Naor [3]在他们的论文中关注的是两种带有故障的排队模型, 其中一个模型是故障只会发生在服务台正在进行服务时, 另外一个模型是故障可能会发生在系统空闲时。White 和 Christie [4]在他们的论文中研究的是带有故障和带有优先权的排队系统的相似性, 并且讨论了这两种效果。Sridharan 和 Jayashree [5]研究的是具有一般故障、部分故障和完全故障的有限队列的一些特性, 此处的部分故障是服务器发生故障时, 允许顾客进入系统。Wang 等人[6]研究的是带有部分故障的马尔可夫单服务台队列的均衡问题。Economou 和 Kanta [7]研究的却是具有开关周期交替的马尔可夫单服务台排队系统。文章[8] [9] [10]考虑的都是具有工作故障的不同的排队模型。Xu 等人[11]分析的是 M/M/1 排队系统中顾客的均衡策略行为, 该排队系统是带有部分故障和可修复的条件。文章[12] [13] [14]讨论的都是带有延迟启动的排队系统。延迟启动的排队系统指的是, 系统从关闭期到正常工作需要经过一段时间的启动期才能开始正常服务。杨喜娟等人[15]研究的是带有启动时间和工作故障的 M/M/1/N 排队系统。

基于上述学者的研究, 本文考虑的是带有延迟启动时间和部分故障 M/M/1/N 排队系统的性能。QBD 的特点和矩阵分析方法在解决二维马尔可夫过程中有非常重要的应用, 相关问题在论文[16] [17] [18] [19]中有提及到。删失技巧在有限空间队列和无穷空间队列的排队模型中起着重要的作用, 它可以使得计算更加的简便。

本文的排队系统可以模拟生活中的一些问题。假设有一台正常工作的机器在某一时刻发生了故障, 它会立即被一台工作速率更低的备用机器所代替, 以保障持续的工作。在这种情况下, 备用机器仅完成系统内未完成的任务, 当所有现存任务完成后, 发生故障的机器才会被送去修理。

本文的组织结构如下: 第 2 节描述了本文的模型; 第 3 节中运用删失技巧和矩阵分析方法, 得到了排队系统的平稳分布; 第 4 节给出了一些数值算例; 第 5 节总结。

2. 模型描述

本文考虑的是一个带有延迟启动时间和部分故障的 M/M/1/N 排队模型, 其模型描述如下:

1) 系统中只有一个服务台, 服务台每次只能服务一个顾客, 队列最大容量为 N 。

2) 顾客的到达是一个泊松过程, 参数为 λ 。

3) 当服务台正常服务完系统中所有的顾客后, 立即进入一段随机长度为 L 的休眠期, L 服从参数为 θ 的指数分布。若休眠期结束时, 队列中有顾客, 则服务台立即进入正常忙期, 正常忙期内顾客的服务时长服从参数为 μ 的指数分布。否则, 队列为空, 系统立即进入关闭期。

4) 若服务台在正常工作过程中发生故障, 则系统不再接受新到达的顾客, 并且服务台将以较低的速率 μ_w 进行服务, 直到系统为空, 此时系统立即进入修复期进行维修。维修完成后系统进入休眠期。假设系统发生故障的过程服从参数为 α 的泊松过程, 维修时长服从参数为 β 的指数分布。

5) 在系统关闭期内, 若有外部顾客到达, 则关闭期结束, 经过一段随机长度为 T 的启动期, 系统进入正常忙期, T 是服从参数为 s 的指数分布。

6) 上述的随机过程都是相互独立的, 且在本文中规定 $\lambda < \mu_w < \mu$, 以保证系统最终是平稳的。服务台的服务遵循先到先服务(FCFS)原则。

3. 系统的平稳分布

令 $N(t)$ 表示 t 时刻系统队列中的顾客数, $I(t)$ 表示 t 时刻系统的状态。根据第 2 节的模型描述, 有:

$$\begin{cases} I(t) = 0, \text{ 启动或关闭状态} \\ I(t) = 1, \text{ 休眠状态} \\ I(t) = 2, \text{ 正常服务状态} \\ I(t) = 3, \text{ 故障状态} \\ I(t) = 4, \text{ 维修状态} \end{cases}$$

由此可知, $\{(N(t), I(t)), t \geq 0\}$ 是一个二维连续时间马尔可夫过程, 状态空间为

$$\Omega = \{(0,0), (0,1), (0,4)\} \cup \{(k, j) : 1 \leq k \leq N, j = 0, 1, 2, 3\}.$$

特别地, $(0,0)$ 表示的是系统关闭期; $(0,1)$ 表示的是系统休眠期; $(0,4)$ 表示系统维修期; (k, j) 表示系统当前状态为 j , 同时系统中有 k 个顾客。系统的状态转移图如图 1 所示。从图 1 中, 可以看到上述二维连续时间马尔可夫过程是不可约的, 根据[20]文中的定理 6.2.16 和推论 6.2.6, 可知本文系统的平稳分布是存在且唯一的。

令 $P(k, j) = \lim_{t \rightarrow \infty} P\{N(t) = k, I(t) = j\} (0 \leq k \leq N, j = 0, 1, 2, 3, 4)$, 为系统的平稳分布, 则平衡方程如下

$$\lambda P(0,0) = \theta P(0,1), \tag{1}$$

$$(\lambda + \theta) P(0,1) = \mu P(1,2) + \beta P(0,4), \tag{2}$$

$$\beta P(0,4) = \mu_w P(1,3), \tag{3}$$

$$(\lambda + s)P(1,0) = \lambda P(0,0), \tag{4}$$

$$(\lambda + \theta)P(1,1) = \lambda P(0,1). \tag{5}$$

由归一化条件知

$$P(0,0) + P(0,1) + P(0,4) + \sum_{k=1}^N \sum_{j=0}^3 P(k,j) = 1 \tag{6}$$

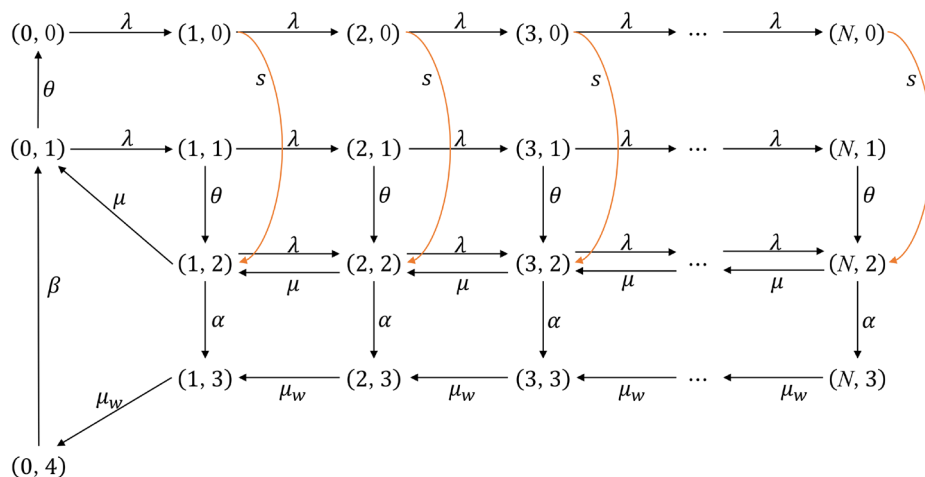


Figure 1. State transition diagram (a)
图 1. 状态转移图(a)

根据删失技巧, 可以得到删失后的马尔可夫过程的状态转移图, 如图 2 所示, 然后通过一些简单的计算, 就可以得到本文模型的平稳分布。

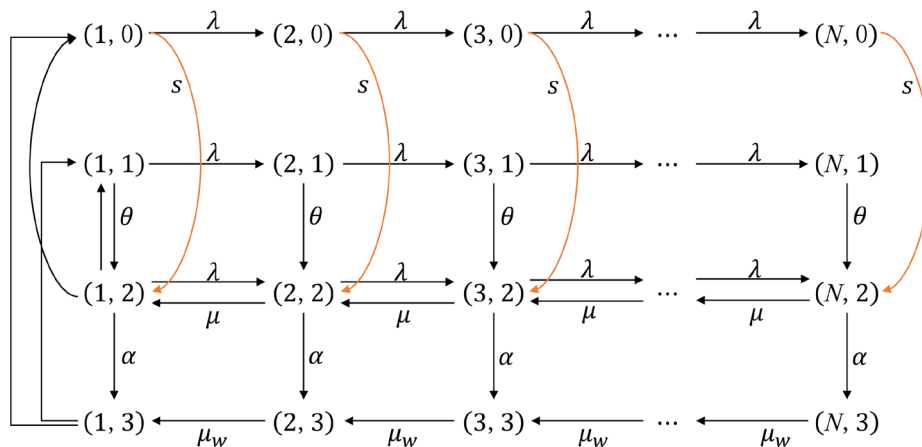


Figure 2. State transition diagram (b)
图 2. 状态转移图(b)

删掉掉状态 (0,0), (0,1), (0,4), 可知一些状态转移速率变为:

$$(1,3) \xrightarrow{\lambda\mu_w/(\lambda+\theta)} (1,1),$$

$$(1,3) \xrightarrow{\theta\mu_w/(\lambda+\theta)} (1,0),$$

$$\pi_1 = -\pi_2 B A_1^{-1} = \pi_2 R_1, \tag{13}$$

$$\pi_N = -\pi_{N-1} C A_N^{-1} = \pi_{N-1} R_N, \tag{14}$$

其中 $R_1 = -B A_1^{-1}$, $R_N = -C A_N^{-1}$ 。

接下来, 将(14)代入(10)中, 如果 $(A - C A_N^{-1} B)$ 也是非奇异的, 可得

$$\pi_{N-1} = -\pi_{N-2} C (A - C A_N^{-1} B)^{-1} = \pi_{N-2} R_{N-1}, \tag{15}$$

其中 $R_{N-1} = -C (A - C A_N^{-1} B)^{-1}$ 。

最后, 根据(9)和(12), 有

$$\pi_k = -\pi_{k-1} C (A + R_{k+1} B)^{-1} = \pi_{k-1} R_k, \quad k = 2, 3, \dots, N-2, \tag{16}$$

其中 $R_k = -C (A + R_{k+1} B)^{-1}$ 。

显然, R_k 可由(12)~(16)推出。根据 Elhafs 和 Molle [16]文中所述, 可知

$$\pi_k = \pi_2 R_k^*, \quad k = 2, 3, \dots, N,$$

其中 $R_k^* = \prod_{j=2}^k R_j$, $R_1 = -B A_1^{-1}$, $R_2 = I$ 。在(12)中, 令 $k=3$, 可得

$$\pi_3 = \pi_2 R_3. \tag{17}$$

将(13)和(17)代入(8)中,

$$\pi_2 (R_1 C + A + R_3 B) = 0. \tag{18}$$

根据归一化条件

$$\pi_1 e_1 + \sum_{k=2}^N \pi_k e = 1,$$

得

$$\pi_1 e + \pi_2 \sum_{k=2}^N R_k^* e = 1, \tag{19}$$

其中 e 是单位列向量。将(13)代入(19)中, 易得

$$\pi_2 \left(R_1 e + \sum_{k=2}^N R_k^* e \right) = 1. \tag{20}$$

结合(18)和(20), π_2 易知, 再由(13)便可得 π_1 。接着根据(12), 有

$$\pi_k = \pi_{k-1} R_k = \pi_{k-2} R_{k-1} R_k = \dots = \pi_2 R_k^*, \quad 2 \leq k \leq N.$$

由于删失后的马尔可夫过程满足下面的归一化条件

$$\sum_{k=1}^N \sum_{j=0}^3 \pi_{kj} = 1, \tag{21}$$

由(6)和(21), 易知

$$\pi_{kj} = \frac{P(k, j)}{1 - P(0, 0) - P(0, 1) - P(0, 4)}, \quad (1 \leq k \leq N, j = 0, 1, 2, 3).$$

同时, 基于(3)~(5), 可以得到下面的等式:

$$\begin{cases} P(0,0) = \frac{(\lambda + s)\pi_{10}/\lambda}{1 + (\lambda + s)\pi_{10}/\lambda + \mu\pi_{11}/(\lambda + \theta) + \mu_w\pi_{13}/\beta} \\ P(0,1) = \frac{\mu\pi_{11}/(\lambda + \theta)}{1 + (\lambda + s)\pi_{10}/\lambda + \mu\pi_{11}/(\lambda + \theta) + \mu_w\pi_{13}/\beta} \\ P(0,4) = \frac{\mu_w\pi_{13}/\beta}{1 + (\lambda + s)\pi_{10}/\lambda + \mu\pi_{11}/(\lambda + \theta) + \mu_w\pi_{13}/\beta} \end{cases}$$

4. 系统性能分析

设 $m = 1 - P(0,0) - P(0,1) - P(0,4)$ 。系统的稳态性能定义如下:

1) 系统吞吐率

$$TP = \sum_{k=1}^N [P(k,2) \cdot \mu + P(k,3) \cdot \mu_w] = m \sum_{k=1}^N (\pi_{k2}\mu + \pi_{k3}\mu_w)。$$

2) 系统稳态可用度

$$Avai = \sum_{k=1}^N [P(k,2) + P(k,3)] = m \sum_{k=1}^N (\pi_{k2} + \pi_{k3})。$$

3) 系统稳态队长

$$Length = \sum_{k=1}^N \left[k \cdot \sum_{i=0}^3 P(k,i) \right] + 0 \cdot P(0,4) = m \sum_{k=1}^N \left(k \cdot \sum_{i=0}^3 \pi_{ki} \right)。$$

4) 系统处于正常忙期的概率

$$P_n = \sum_{k=1}^N P(k,2) = m \sum_{k=1}^N \pi_{k2}。$$

5) 系统处于故障忙期的概率

$$P_b = \sum_{k=1}^N P(k,3) = m \sum_{k=1}^N \pi_{k3}。$$

6) 系统处于修复期的概率

$$P_r = P(0,4)。$$

7) 系统处于关闭期的概率

$$P_{idle} = P(0,0)。$$

其次, 我们给出了一些数值算例来说明一些参数对系统性能的影响情况。

图 3 反映的是 λ 对 TP 的影响, 当其它参数不变时, TP 随着 λ 的增大而增大, 即系统吞吐量随着到达速率的增大而增大。图 4 反映的是 α 对 $Avai$ 的影响和 μ 对 $Length$ 的影响, 当 α 不断增大时 $Avai$ 先减小到某一数值后不断增大, 当 μ 增大时 $Length$ 随之而减小。图 5 反映的是 θ 和 s 对 P_n 的影响, 当 θ 的增大时, P_n 先增大到某一数值后不断减小, 而 P_n 随着 s 的增大而增大直到趋于一个固定的值。图 6 反映的是 μ_w 对 P_b 和 P_r 的影响, 显然, P_b 和 P_r 随着 μ_w 的增大而减小。图 7 反映了 θ 对 P_{idle} 的影响, 其中 P_{idle} 随着 θ 的增大而增大。

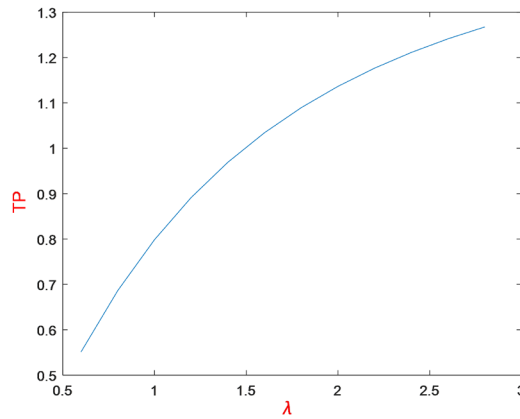


Figure 3. λ versus TP for $(\alpha, \beta, \mu, \mu_w, \theta, s, N) = (2, 2, 4, 3, 2, 3, 15)$

图 3. 当 $(\alpha, \beta, \mu, \mu_w, \theta, s, N) = (2, 2, 4, 3, 2, 3, 15)$ 时, λ 对 TP 的影响

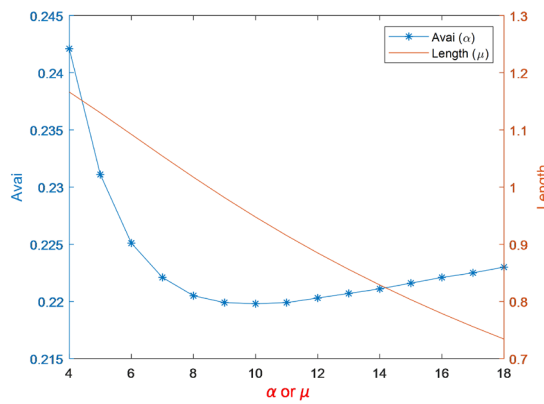


Figure 4. α versus $Avai$ for $(\beta, \lambda, \mu, \mu_w, \theta, s, N) = (2, 2, 4, 3, 2, 3, 15)$;

μ versus $Length$ for $(\alpha, \beta, \lambda, \mu_w, \theta, s, N) = (2, 2, 2, 3, 2, 3, 15)$

图 4. 当 $(\beta, \lambda, \mu, \mu_w, \theta, s, N) = (2, 2, 4, 3, 2, 3, 15)$ 时, α 对 $Avai$ 的影响; 当 $(\alpha, \beta, \lambda, \mu_w, \theta, s, N) = (2, 2, 2, 3, 2, 3, 15)$ 时, μ 对 $Length$ 的影响

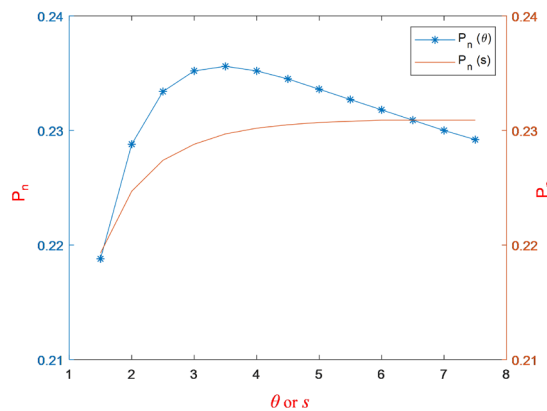


Figure 5. θ and s versus P_n for $(\alpha, \beta, \lambda, \mu, \mu_w, s, N) = (2, 2, 2, 4, 3, 3, 15)$ and $(\alpha, \beta, \lambda, \mu, \mu_w, \theta, N) = (2, 2, 2, 4, 3, 2, 15)$

图 5. 当 $(\alpha, \beta, \lambda, \mu, \mu_w, s, N) = (2, 2, 2, 4, 3, 3, 15)$ 时, θ 对 P_n 的影响; 当 $(\alpha, \beta, \lambda, \mu, \mu_w, \theta, N) = (2, 2, 2, 4, 3, 2, 15)$ 时, s 对 P_n 的影响

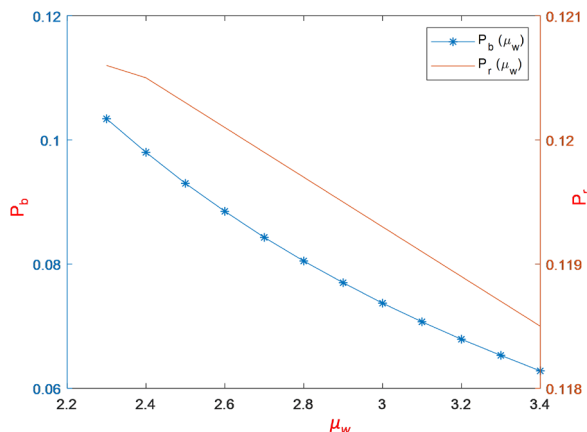


Figure 6. μ_w versus P_b and P_r for $(\alpha, \beta, \lambda, \mu, \theta, s, N) = (2, 2, 2, 4, 2, 3, 15)$

图 6. 当 $(\alpha, \beta, \lambda, \mu, \theta, s, N) = (2, 2, 2, 4, 2, 3, 15)$ 时, μ_w 对 P_b 和 P_r 的影响

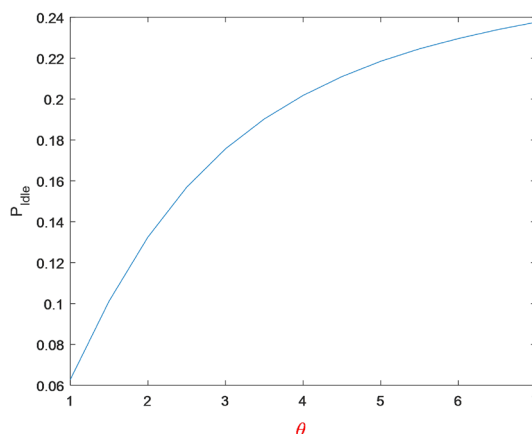


Figure 7. θ versus P_{idle} for $(\alpha, \beta, \lambda, \mu, \mu_w, s, N) = (2, 2, 2, 4, 3, 3, 15)$

图 7. 当 $(\alpha, \beta, \lambda, \mu, \mu_w, s, N) = (2, 2, 2, 4, 3, 3, 15)$ 时, θ 对 P_{idle} 的影响

5. 总结

本文主要研究的是带有延迟启动时间和部分故障的 M/M/1/N 排队系统的性能, 根据删失技巧和矩阵分析方法, 得到了该系统的平稳分布然后分析了系统的性能。作为本文的拓展, 我们将考虑带有无限等待队列的排队系统, 使用同样的技巧, 可以研究排队系统的尾部渐近性。

参考文献

- [1] Erlang, A.K. (1909) The Theory of Probabilities and Telephone Conversations. *Nyt Tidsskrift for Matematik*, **20**, 33-39.
- [2] Gnedenko, B.V. and Kovalenko, I.N. (1989) Introduction to Queueing Theory. Birkhauser Boston Incorporated, Cambridge, MA.
- [3] Avi-Itzhak, B. and Naor, P. (1963) Some Queueing Problems with the Service Station Subject to Breakdown. *Operations Research*, **11**, 303-320. <https://doi.org/10.1287/opre.11.3.303>
- [4] White, H. and Christie, L.S. (1958) Queueing with Preemptive Priorities or with Breakdown. *Operations Research*, **6**, 79-95. <https://doi.org/10.1287/opre.6.1.79>
- [5] Sridharan, V. and Jayashree, P.J. (1996) Some Characteristics on a Finite Queue with Normal, Partial and Total Failures. *Microelectronics and Reliability*, **36**, 265-267. [https://doi.org/10.1016/0026-2714\(95\)00088-J](https://doi.org/10.1016/0026-2714(95)00088-J)

-
- [6] Li, L., Wang, J.T. and Zhang, F. (2013) Equilibrium Customer Strategies in Markovian Queues with Partial Breakdowns. *Computers & Industrial Engineering*, **66**, 751-757. <https://doi.org/10.1016/j.cie.2013.09.023>
- [7] Economou, A. and Kanta, S. (2008) Equilibrium Balking Strategies in the Observable Single-Server Queue with Breakdowns and Repairs. *Operations Research Letters*, **36**, 696-699. <https://doi.org/10.1016/j.orl.2008.06.006>
- [8] Liu, Z. and Song, Y. (2014) The $M^x/M/1$ Queue with Working Breakdown. *RAIRO-Operations Research*, **48**, 399-413. <https://doi.org/10.1051/ro/2014014>
- [9] Yang, D.Y. and Cho, Y.C. (2019) Analysis of the N-Policy GI/M/1/K Queuing Systems with Working Breakdowns and Repairs. *Computer Journal*, **62**, 130-143. <https://doi.org/10.1093/comjnl/bxy051>
- [10] Kalidass, K. and Kasturi, R. (2012) A Queue with Working Breakdowns. *Computers & Industrial Engineering*, **63**, 779-783. <https://doi.org/10.1016/j.cie.2012.04.018>
- [11] Xu, B. and Xu, X. (2018) Equilibrium Strategic Behavior of Customers in the M/M/1 Queue with Partial Failures and Repairs. *Operational Research*, **18**, 273-292. <https://doi.org/10.1007/s12351-016-0264-7>
- [12] 魏瑛源, 唐应辉. 延迟 $\text{Min}(N,D)$ -策略下 M/G/1 排队系统的离去过程[J]. 应用数学, 2018, 31(4): 820-829.
- [13] 潘取玉, 唐应辉. 在延迟 $\text{Min}(N,D)$ -控制策略下修理设备可更换的 M/G/1 可修排队系统及最优控制策略[J]. 数学物理学报, 2018, 38(5): 1014-1031.
- [14] 贺灵悦, 唐应辉. 带启动时间和多重休假的 $\text{Min}(N,V)$ -策略 M/a/1 排队系统[J]. 系统科学与数学, 2017, 37(3): 846-862.
- [15] 杨喜娟, 李忠学, 黎锁平, 武福. 带启动时间和工作故障的 M/M/1/N 排队系统性能分析[J]. 控制理论与应用, 2019, 36(4): 561-569.
- [16] Elhafsi, E.H. and Molle, M. (2007) On the Solution to QBD Processes with Finite State Space. *Stochastic Analysis and Applications*, **25**, 763-779. <https://doi.org/10.1080/07362990701419946>
- [17] 田乃硕, 岳德权. 拟生灭过程与矩阵几何解[M]. 北京: 科学出版社, 2002.
- [18] Neuts, M.F. (1981) Matrix-Geometric Solution in Stochastic Model: an Algorithmic Application. The Johns Hopkins University Press, Baltimore, MD.
- [19] Personè, V.N. and Grassi, V. (1996) Solution of Finite QBD Processes. *Journal of Applied Probability*, **33**, 1003-1010. <https://doi.org/10.2307/3214981>
- [20] 叶尔骅, 张德平. 概率论与随机过程[M]. 北京: 科学出版社, 2005.