Service Rate Control in Mn/Gn/1 Queue

Koichi Nakade, Shunta Nishimura

HAL Id: hal-02164849

https://inria.hal.science/hal-02164849

Submitted on 25 Jun 2019

Service Rate Control in Mn/Gn/1 Queue

Koichi Nakade[1][0000-0002-0933-6614] and Shunta Nishimura[1]

[1] Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan
nakade@nitech.ac.jp

Abstract. Service rate control in queues is an important problem in practice. For example, when the number of customers waiting for finished products is small production has a normal speed, but if it is greater, the production speed should be faster to meet the demand. For an exponential service time distribution, the optimality of the threshold type policy has been proved in literature. On the other hand, in production systems, production time follows a general distribution in general. In this paper, service control of speed depending on the number of customers is discussed. The analytical results of an M/G/1 queue with arrival and service rates depending on the number of customers in the system, which is called an Mn/Gn/1 queue, are utilized for computing performance measure of service rate control. Constant, uniform and exponential distributions on the service time are considered through numerical experiments. The results show that the optimal threshold depends on the type of the distributions even when the mean of service time is the same.

Keywords: Mn/Gn/1 queue, Service Control, Threshold policy, Admission

1   Introduction

Service rate control problems are widely found in practical situations. In a supermarket, if the queue of customers for payment is long, a cashier will be helped by someone, and in a communication network when the congestion occurs the faster line will be used. In a make-to-order production system, when the number of customers waiting for the finished products is small, production has a normal speed, but if it is greater, the production speed should be faster to meet the demand. In addition, the arrival rate of customers may depend on the number of waiting customers because some of the customers may be impatient. Such a system is represented as a queue with state-dependent arrival and service rates.

For the exponential service time distribution with a constant arrival rate, George and Harrison [1] discuss a service control problem and develop an asymptotic method for computing the optimal policy, and optimality of a threshold type policy has been proved in Dimitrakopoulos and Burnetas [2], in which optimal admission control is also discussed. In general, however, as production systems, service time follows a general distribution.

In this paper, control problems of service speed depending on the number of customers under various general service time distributions and Poisson arrivals of cus-

tomers, which is called an Mn/Gn/1 queue, are formulated, and the effects of the types of distributions on the performance measure like a profit are discussed through numerical experiments. To compute performance measure under given service rate control, the analytical results of an Mn/Gn/1 queue by Abouee-Mehrizi and Baron [3] are utilized, which has a single server with arrival and service rates depending on the number of customers in the system. Constant, uniform and exponential distributions on service time are considered through numerical experiments. Optimal threshold policies are compared among these distributions. Admission control is also discussed by changing the upper bound of the queue length.

2　　Model Description

We consider an Mn/Gn/1 queue, where Mn means a Poisson arrival process with a state-dependent arrival rate, and Gn means a general service time distribution with a state-dependent service rate. This service rate is determined by a decision maker. When the number of waiting customers including a customer in service, which is called a state of the system, is i, the arrival rate is λ_i. For example, when customers arrive in a Poisson process with rate λ and an arriving customer observes state i, he/she is assumed to enter the system with probability q_i. Then the actual arrival process in the system is a state-dependent Poisson process with rate $\lambda_i = \lambda q_i$ in state i (see Fig. 1). In addition, we assume that there is a finite integer K which is a minimal value i satisfying $\lambda_i = 0$, that is $q_i = 0$. It happens because if the queue length is too long, any customer cannot wait for service. On the other hand, if the manager wants to reduce the waiting time of customers because long waiting time leads to less customers' satisfaction or the system has to pay costs for the long delay, the system sets the threshold for queue length and refuses the further demand during the time in which the length attains this threshold. If this threshold is K, then $\lambda_K = 0$.

　　We explain the service control. There are J possible service rates $\mu^1, \mu^2, \dots, \mu^J$, where $\mu^j < \mu^{j+1}$ for $j \in \{1,2,\dots,J-1\}$. When the state is i, the service rate decision is defined as a_i for each $i \in \{1,2,\dots,K\}$. Here, if $a_i = j$, the service rate is μ^j. The notation μ_i is the service rate in state i, and thus if $a_i = j$, the service rate is $\mu_i = \mu^j$. For each service rate μ^j, the service cost rate is c^j, where $c^j < c^{j+1}$ ($j \geq 1$). The sequence $\{a_i, \ i \in \{1,2,\dots,K\}\}$ is a service control policy in this model.

　　Assume that policy $\{a_i, \ i \in \{1,2,\dots,K\}\}$ is given. When the state becomes i just after some service is finished, new service starts and service time follows a general distribution function $B^j(x)$ when $a_i = j$. If during service a new customer arrives and the state becomes $i+1$, then service rate becomes μ_{i+1}. In this case, if the remaining service time is η just before the arrival, after the arrival it is changed to $\frac{\eta \mu_i}{\mu_{i+1}} = \eta/\alpha_{i+1}$, where $\alpha_i = \frac{\mu_i}{\mu_{i-1}}$, $i \geq 1$. Here α_i denotes the ratio of service rates for state i and $i-1$.

　　When policy $\{a_i, \ i \in \{1,2,\dots,K\}\}$ is fixed, this system follows an Mn/Gn/1 queue which is defined in Abouee-Mehrizi and Baron[2]. Here we denote the service time distribution, its density function and the remaining service time density function in a

steady state by $B_i(x), b_i(x), h_i(x)$, respectively. As shown, $B_i(x) = B^j(x)$ when $a_i = j$. For computational convenience it is assumed that $c_0 = 0$ and $\alpha_1 = 1$.

The reward and cost consist of the followings: the waiting cost rate for each customer is w, the reward for each customer is r, and the service cost rate is c^j when the service rate is μ^j as shown above.

Let Z denote the average profit. For $i \in \{0,1,\dots,K\}$, $P_F(i)$ is a steady state distribution under the defined Mn/Gn/1/K queue with a given policy $\{a_i, \ i \in \{1,2,\dots,K\}\}$. Then Z is given as $Z=R\text{-}W\text{-}C$, where

$$R = r \sum_{i=0}^{K-1} P_F(i)\lambda_i, \qquad W = w \sum_{i=1}^{K} i P_F(i), \qquad C = \sum_{i=1}^{K} P_F(i)c_{a_i} \qquad (1)$$

The objective in the model is to find an optimal service rate control policy $\{a_i, \ i \in \{1,2,\dots,K\}\}$ which maximizes the average profit Z.

The following notations are also used.

$\tilde{h}_i(s)$: Laplace transform of $h_i(x)$,

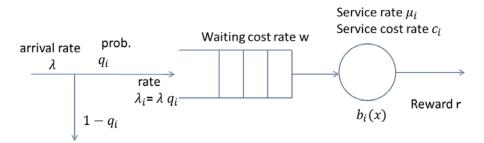$\tilde{b}_i(s)$: Laplace transform of $b_i(x)$.

Fig. 1. Mn/Gn/1 queue

3 Analytical Results of Mn/Gn/1 Queue

In Abouee-Mehrizi and Baron [3], an Mn/Gn/1 queue is analyzed theoretically (In [3], some typos in equations are found and they are corrected in the following). It is an extensive result of Kerner [4], which analyzes an Mn/G/1 queue.

For the Mn/Gn/1 queue with an infinite buffer, under the condition that

$$\sum_{i=1}^{\infty} \frac{\lambda_0}{\lambda_i} \prod_{j=0}^{i-1} \frac{1 - \tilde{h}_j\left(\frac{\lambda_{j+1}}{\alpha_{j+1}}\right)}{\tilde{b}_{j+1}(\lambda_{j+1})} < \infty,$$

the steady state distribution is given by

$$P(i) = \frac{\lambda_0 P(0)}{\lambda_i} \prod_{j=0}^{i-1} \frac{1 - \tilde{h}_j\left(\frac{\lambda_{j+1}}{\alpha_{j+1}}\right)}{\tilde{b}_{j+1}(\lambda_{j+1})}, \qquad i = 1,2,\dots$$

$$P(0) = \cfrac{1}{1 + \sum_{i=1}^{\infty} \cfrac{\lambda_0}{\lambda_i} \prod_{j=0}^{i-1} \cfrac{1 - \tilde{h}_j\left(\frac{\lambda_{j+1}}{\alpha_{j+1}}\right)}{\tilde{b}_{j+1}(\lambda_{j+1})}}$$

Here $\tilde{h}_0(s) = \tilde{b}_1(s)$, and $\tilde{h}_i(s)$ satisfies the following recursive equations.

$$\tilde{h}_i(s) = \frac{\lambda_i}{s - \lambda_i}\left[\frac{\tilde{b}(\lambda_i)\left(1 - \tilde{h}_{i-1}\left(\frac{s}{\alpha_i}\right)\right)}{1 - \tilde{h}_{i-1}\left(\frac{\lambda_i}{\alpha_i}\right)} - \tilde{b}_i(s) \right].$$

In the Mn/Gn/1/K queue, where the number of customers in the system is limited to K, the steady state distribution function $P_F(i)$ for $i = 0, \dots, K - 1$ is given by

$$P_F(i) = \frac{\lambda_0 P_F(0)}{\lambda_i} \prod_{j=0}^{i-1} \frac{1 - \tilde{h}_j\left(\frac{\lambda_{j+1}}{\alpha_{j+1}}\right)}{\tilde{b}_{j+1}(\lambda_{j+1})}, \quad i = 1, \dots, K - 1,$$

$$P_F(0) = \cfrac{P_F^A(0)}{\cfrac{\lambda_0}{\lambda_{K-1}} \prod_{j=0}^{K-2} \cfrac{1 - \tilde{h}_j\left(\frac{\lambda_{j+1}}{\alpha_{j+1}}\right)}{\tilde{b}_{j+1}(\lambda_{j+1})} + P_F^A(0)\left(1 + \sum_{i=1}^{K-2} \cfrac{\lambda_0}{\lambda_1} \prod_{j=0}^{i-1} \cfrac{1 - \tilde{h}_j\left(\frac{\lambda_{j+1}}{\alpha_{j+1}}\right)}{\tilde{b}_{j+1}(\lambda_{j+1})} \right)},$$

where

$$P_F^A(0) = \frac{\mu_b^F}{\lambda_{K-1} + \alpha_K \mu_b^F}, \quad P_F^A(1) = \frac{\lambda_{K-1}}{\lambda_{K-1} + \alpha_K \mu_b^F},$$

$$\frac{1}{\mu_b^F} = \frac{1}{\mu_{K-1}} - \frac{1}{\lambda_{K-1}} + \sum_{j=i}^{K-2}\left(\frac{1}{\mu_j} - \frac{1}{\lambda_j}\right) \prod_{i=j}^{K-2} \frac{1}{\alpha_{i+1}} \frac{\tilde{b}_{i+1}(\lambda_{i+1})}{1 - \tilde{h}_i\left(\frac{\lambda_{i+1}}{\alpha_{i+1}}\right)}$$

$$+ \frac{1}{\mu_1} \frac{\tilde{b}_1(\lambda_1)}{1 - \tilde{h}_0\left(\frac{\lambda_1}{\alpha_1}\right)} \prod_{i=1}^{K-2} \frac{1}{\alpha_{i+1}} \frac{\tilde{b}_{i+1}(\lambda_{i+1})}{1 - \tilde{h}_i\left(\frac{\lambda_{i+1}}{\alpha_{i+1}}\right)}$$

Using $P_F^A(0)$ and $P_F^A(1)$, we have

$$P_F(K - 1) = \left(1 - F_F(K - 2)\right)P_F^A(0), \quad P_F(K) = \left(1 - F_F(K - 2)\right)P_F^A(1)$$

where $F_F(i) = \sum_{j=0}^{i} P_F(j)$.

When the policy $\{a_i, \ i \in \{1, 2, \dots, K\}$ is given, by deriving steady state probabilities and substituting them to equation (1), the average profit can be obtained.

4   Numerical Experiments

4.1   Parameter settings

In the following, the constant, uniform and exponential distributions are applied as service distributions. For given i, the mean of service time is fixed as $\frac{1}{\mu_i}$.

(a) Constant distribution

$$B_i(x) = \begin{cases} 0 & (0 \leq x < 1/\mu_i) \\ 1 & (1/\mu_i \leq x) \end{cases}, \tilde{b}_i(s) = e^{-s/\mu_i}$$

(b) Uniform distribution on $[0, \frac{2}{\mu_i}]$

$$B_i(x) = \begin{cases} 0 & (x < 0) \\ \frac{\mu_i x}{2} & \left(0 \leq x \leq \frac{2}{\mu_i}\right) \\ 1 & \left(\frac{2}{\mu_i} < x\right) \end{cases}, \tilde{b}_i(s) = \frac{\mu_i}{2s}\left(1 - e^{-s\frac{2}{\mu_i}}\right)$$

(c) Exponential distribution

$$B_i(x) = 1 - e^{-\mu_i x} \quad (x \geq 0), \tilde{b}_i(s) = \frac{\mu_i}{s+\mu_i}$$

Set $K = 5, \lambda = 15, q_n = 1 - \frac{n}{150}$. Two types of service rates exist as $\mu^1 = 10$ and $\mu^2 = 20$. Service cost rates are set as $c^1 = 50$ and $c^2 = 150$, a waiting cost rate and reward for each service are $w = 2$ and $r = 10$, respectively.

Here, the threshold policy is applied. The threshold which is the minimal value satisfying $\mu_i = \mu^2$ is denoted as i_B $(1 \leq i_B \leq K + 1)$, and under the threshold policy with i_B, $a_i = 1$ for $i = 1,2, ..., i_B - 1$, and $a_i = 2$ for $i = i_B, i_B + 1, ..., K$. Here, $i_B = K + 1$ means $\mu_i = \mu^1$ for all $i = 1,2, ..., K$, and $i_B = 1$ means $\mu_i = \mu^2$ for all $i = 1,2, ..., K$.

4.2 Effect of distributions

Table 1 shows probabilities, costs, and rewards for each i_B and each service time distribution. In this table, $p[i]$ shows the steady state probability $P_F(i)$. For example, when service time is constant, $i_B = 3$ is the best threshold and thus it is optimal among threshold type policies that the slower service is applied in states 1,2,3 and the faster service is used in states 4 and 5.

For all distributions, as i_B increases, R and C decrease and W increases. When i_B is large, the low service rate is applied more and $P_F(K)$ is large, and thus less customers are permitted to receive service. As a result, the total arrival rate and the average service cost decrease. On the other hand, the low service rate leads to the long waiting time regardless of the less arrival rate of entering customers.

As the variance of service time increases, R and C decrease. $P_F(K)$ increases as the variance increases, and thus in the similar way as above the increase of variance leads to less amount of entering customers and less service cost. For constant, uniform and exponential distributions, the values of optimal i_B are 3, 4 and 5, respectively. Thus, the type of service time distribution affects the optimal threshold.

In the case that $\lambda = 15, q_n = 1 - \frac{n}{150}$, as shown in Table 1, as the variance increases, the profit Z decreases for each i_B. Table 2 shows probabilities and profits, in the case that $\lambda = 20, q_n = 1 - \frac{n}{200}$, under several types of service time distributions for i_B=5, 6. In this case, it is found that the uniformly distributed service time achieves more profits compared with the case of constant service time distribution with the same mean. When i_B is high, reward and service cost are almost the same, and the difference of waiting time cost becomes more important to profit Z.

4.3   Admission of Arrivals

As shown in section 2, parameter K can be considered as an admission control parameter. Table 3 shows the values of profits Z for various K and threshold i_B, under uniform service time distribution, when $c^2 = 130$ and 150. For $c^2 = 130$ the profit is greater as K is larger. This is because the greater K leads to more customers receiving service, which gives more profits. When cost parameter c^2 is 150, however, $K=4$ maximizes the profit when threshold i_B is 3 or 4, and for $K=5$ is the best when i_B is 2. The large K implies that the more customers receive the faster service, and when the service cost parameters are high, the service cost has more effect on Z compared with admission reward. For $c^2 = 150$, the combination of $i_B = 4$ and $K=4$ gives the best profit.

5   Conclusion

This paper has formulated a mathematical model of service rate control in a state-dependent M/G/1 queue. Throughout numerical experiments, it is found that under threshold policies, the value of optimal threshold depends on the type of service time distributions. Usually, as the variance of service distribution increases, the reward from by customer arrivals and the holding cost decreases and the profit also decreases. In some case of parameter sets, however, the total profit is greater under the uniform distribution of service time compared with the case of the constant distribution. In addition, the combination of optimal threshold of service and admission of arrivals is observed to depend on cost parameters.

The effects of various parameters such as the arrival rate and q_n into performance need to be investigated. Theoretical discussion for the optimality of threshold policies under general distributions are also important. They are left for further research.

References

1. George, J. M., Harrison, J. M. : Dynamic Control of a Queue with Adjustable Service Rates. Operations Research, 49(5), 720-731 (2001)
2. Dimitrakopoulos, Y., Burnetas, A. : The Value of Service Rate Flexibility in an M/M/1 Queue with Admission Control. IISE Transactions 49(6), 603-621 (2017)
3. Abouee-Mehrizi, H., Baron, O. : State-Dependent M/G/1 Queueing Systems. Queueing Systems, 82(1-2), 121-148 (2016)
4. Kerner, Y. : The conditional distribution of the residual service time in the Mn/G/1 queue. Stochastic Models, 24(3), 364-375 (2008)

Table 1. Probabilities and Profits ($\lambda = 15, q_n = 1 - \frac{n}{150}$)

(a)   Constant distribution

iB	6	5	4	3	2	1
p[0]	0.0086	0.0102	0.0194	0.0422	0.1027	0.2736
p[1]	0.0297	0.0354	0.0672	0.1460	0.3554	0.3048
p[2]	0.0741	0.0885	0.1677	0.3644	0.2658	0.2090
p[3]	0.1722	0.2055	0.3897	0.2525	0.1584	0.1222
p[4]	0.3914	0.4670	0.2647	0.1464	0.0885	0.0681
p[5]	0.3240	0.1933	0.0914	0.0484	0.0291	0.0224
R	99.142	118.307	133.666	140.515	143.915	145.280
W	7.760	7.327	6.174	4.921	3.724	2.947
C	49.571	68.819	84.637	92.625	99.051	108.960
Z	41.811	42.161	42.855	42.969	41.140	33.373

(b)   Uniform distribution

iB	6	5	4	3	2	1
p[0]	0.0238	0.0286	0.0442	0.0757	0.1414	0.2852
p[1]	0.0513	0.0617	0.0952	0.1631	0.3046	0.2652
p[2]	0.0976	0.1173	0.1811	0.3102	0.2393	0.1968
p[3]	0.1770	0.2127	0.3285	0.2298	0.1630	0.1314
p[4]	0.3146	0.3780	0.2381	0.1521	0.1045	0.0837
p[5]	0.3356	0.2017	0.1129	0.0692	0.0471	0.0377
R	97.617	117.303	130.675	137.538	141.246	142.958
W	7.428	6.910	5.919	4.854	3.851	3.153
C	48.809	68.735	82.886	91.323	98.318	107.219
Z	41.380	41.659	41.870	41.361	39.076	32.587

(c)   Exponential Distribution

iB	6	5	4	3	2	1
p[0]	0.0500	0.0608	0.0816	0.1178	0.1820	0.3079
p[1]	0.0750	0.0912	0.1225	0.1766	0.2729	0.2309
p[2]	0.1117	0.1358	0.1825	0.2632	0.2033	0.1720
p[3]	0.1654	0.2010	0.2701	0.1948	0.1505	0.1273
p[4]	0.2431	0.2955	0.1985	0.1432	0.1106	0.0936
p[5]	0.3549	0.2157	0.1449	0.1045	0.0807	0.0683
R	95.001	115.495	126.174	132.465	136.317	138.423
W	7.082	6.453	5.632	4.765	3.954	3.345
C	47.501	68.533	80.256	88.353	95.414	103.817
Z	40.418	40.508	40.286	39.347	36.948	31.260

Table 2. Probabilities and Profits ($\lambda = 20, q_n = 1 - \frac{n}{200}$)

(a) i_B=5				(b) i_B=6		
	constant	uniform	exponent		constant	uni	exponent
p[0]	0.0008	0.0067	0.0219	p[0]	0.0006	0.0050	0.0164
p[1]	0.0048	0.0204	0.0439	p[1]	0.0036	0.0154	0.0329
p[2]	0.0238	0.0576	0.0873	p[2]	0.0179	0.0434	0.0654
p[3]	0.1134	0.1583	0.1728	p[3]	0.0856	0.1192	0.1296
p[4]	0.5321	0.4292	0.3405	p[4]	0.4015	0.3232	0.2553
p[5]	0.3252	0.3278	0.3336	p[5]	0.4908	0.4937	0.5004
R	132.443	132.115	131.172	R	99.943	99.500	98.356
W	8.294	7.933	7.534	W	8.712	8.443	8.151
C	82.480	82.447	82.268	C	49.972	49.750	49.178
Z	41.669	41.735	41.370	Z	41.259	41.307	41.027

Table 3. Profits under different Parameter K ($\lambda = 15, q_n = 1 - \frac{n}{150}$)

		iB=8	7	6	5	4	3	2	1
c2=130	K=3					41.690	45.895	47.272	44.056
	4				42.096	46.404	49.087	49.278	45.979
	5			41.380	45.962	48.889	50.382	50.154	46.883
	6		40.118	44.400	47.827	49.889	50.928	50.544	47.317
	7	38.566	42.804	46.337	48.678	50.296	51.152	50.712	47.524
150	3					41.690	41.141	38.370	30.828
	4				42.096	42.116	41.436	38.937	32.054
	5			41.380	41.659	41.870	41.361	39.076	32.587
	6		40.118	40.528	41.174	41.555	41.192	39.062	32.813
	7	38.566	39.048	39.926	40.753	41.272	41.022	39.001	32.900