PriMera Scientific Engineering
Volume 5 Issue 5 November 2024
ISSN: 2834-2550

Control Theory for Queuing Systems with Moving Servers and Some Open Problems

Type: Review Article
Received: October 18, 2024
Published: October 30, 2024

Citation:
Asaf Hajiyev. "Control Theory for Queuing Systems with Moving Servers and Some Open Problems". PriMera Scientific Engineering 5.5 (2024): 15-23.

Asaf Hajiyev*

Institute of Control Systems, Ministry of Science and Education, Azerbaijan Republic

***Corresponding Author:** Asaf Hajiyev, Institute of Control Systems, Ministry of Science and Education, Azerbaijan Republic.

Abstract

For queuing systems with moving servers, the control policy which means delays of a beginning service is introduced. In the capacity of efficiency index of systems is taken a customer's average waiting time before service. Although it seems that it is a paradoxical idea to introduce delays of beginning service, it is shown that for some systems it gives a gain in a customer's average waiting time before service. The class of queuing systems for which it is advisable to introduce delays is described. The form of an optimal function minimizing the efficiency index is found.

It is shown that if the intervals between neighbor services have exponential distribution, then the gain in a customer's average waiting time before service equals 10% and independent of parameter of exponential distribution. For uniform distribution such gain equals 3.5% and also independent of parameter of uniform distribution. The criterion to define for which systems the gain is greater are given. Some open problems and numerical examples demonstrating theoretical results are given.

Keywords: queues with moving servers; a customer's average waiting time; delay of beginning service; optimal function

Introduction

At the beginning of the XX century the solving of many practical problems for systems with complicated structures led to the appearance of a new field of investigations, which came to be known as queuing theory or queuing systems. The main objects for queuing systems are customers (which arrive at a system and should be served) and servers (which should give a service for customers). Pioneer papers of queuing systems belong to the worker of Copenhagen telephone company, Agner Erlang in the beginning of XX-th century. Later, Agner Erlang became well known scientist who can be considered as one of the founders of queuing theory. Essential contribution to queuing theory have been made by professors of Lomonosov Moscow State University B.V.Gnedenko, A.Ya.Khinchin Yu.K. Belyaev (USSR, Russia), D.Cox, W.Smith (Great Britain), L.Kleinrock, T.Saaty, S.Ross, G.Newell (USA) and others.

For optimization (minimization or maximization) of the chosen efficiency indices - queue length, a customer average waiting time, and others it is necessary to introduce various control strategies which play an important role from theoretical as well from practical point of view. From practical point of view such control strategy must be easy implemented in practice and give a gain in chosen efficiency index.

In this field of investigations, a lot of books [1, 2, 5, 9-11, 12, 23, 25] and papers [3, 4, 6, 17-22, 24, 26, 28], even the special journals (*Queuing systems, Transportation Science, Transportation and Communication, Operation Research, Mathematics of Operation Research and others*) are published today.

As a matter of development of queuing theory in the second part of the XX-th century the methods and approaches of queuing theory have been applied in investigating of traffic flows, airports, control of traffic flows in tunnels, shipping, transportation, communication systems, elevators and escalators systems, network of computers, vertical (elevators and escalators) transportation and others. The investigation of such systems was outside of the framework of methods and approaches of the classic queuing theory and led to an appearance of new branches of investigation, such as the system with moving servers. Although these systems have a complicated structure all of them are unified in one common idea - systems with moving servers.

Nowadays, systems with moving servers are not investigated so widely, because they have rather complicated and a random structure as customers arriving epochs, service beginning instants, service time are values of random processes, hence, it is necessary to develop the new methods and approaches for their investigation. Usage the methods and approaches of probability theory, stochastic processes, mathematical statistics, simulation, numerical analysis and other branches of mathematics allow to create theory of queuing systems with moving servers, which can help to calculate different characteristics of the systems with moving servers, investigate various disciplines of service, create a control theory and take necessary decisions and recommendations for practical applications.

Systems with moving servers can be considered also as multi-staged or multi-class queuing systems and for their investigation the models of these queues can be applied. Investigation of queues with moving servers leads to the construction and research of the new mathematical models and in the frames of such mathematical models the new methods of investigations are forming.

A particular interest in this division are mathematical models of moving particles, which describe a behavior of a wide class of complicated queuing systems with moving servers and could be successfully applied for traffic, public transportation systems, networks of computers, communication nets, medical, biological and other systems.

Construction of mathematical model

Consider queuing system for which $t_1, t_2,\dots,t_n$ is the sequence of service starts instants. A stationary flow of customers with finite intensity arrives to service and this flow independent of t_i, $i=1,2,\dots,n,$.. At the instant t_i all customers that arrived over the interval $[t_{i-1}, t_i)$ are served instantaneously, i.e. it is assumed that server has infinite volume. In the capacity of an efficiency index, we take a customer's average waiting time (*CWT*) before service which will be denoted w. Customer waiting time before service is defined as the time elapsed from the arrival of customer into the system until the next current service start. Similar models have been introduced in [13, 14]. Let us assume that the sequence $t_1, t_2,\dots,t_n$ is a stationary renewal process. We introduce control the service start instants; for this purpose we turn from the sequence $t_1, t_2,\dots,t_n$ to the new sequence $t_1^*, t_2^*,\dots, t_n^*$ using the following rule.

Denote $\eta_1= t_1$, $\eta_2=t_2 -t_1,\dots,\eta_n =t_n - t_{n-1}$; $t_i^* = t_i -g(\eta_i)$, where g€G is the class of measurable and nonnegative functions. Further we can define the intervals between services in the controlled system as.

$$\eta_1^*= \eta_1 +g(\eta_1),\ \eta_2^*=\eta_2+g(\eta_2),\dots,\ \eta_n^*=\eta_n +g(\eta_n)$$

Remark

With an idea to introduce delays, first time we faced during our work in Lomonosov Moscow State University. The cleaning service women, who improved (reduced a passenger average waiting time) service in fact introduced the delay control policy but they could not explain why. If two elevators came down to the first floor almost at the same time, the women delayed one of the elevators for cleaning, and only afterward they were done it was ready for use. The great A.N.Kolmogorov paid attention that the cleaning women do not know how they are improving service.

Assume that the sequence $t_1, t_2,\dots,t_n$ is a renewal process ($\eta_1, \eta_2,\dots,\eta_n$ is identical and independent random variables) then according to [7] we have.

$$w=\frac{E\eta^2}{2E\eta},\ \sigma^2=\left(\frac{E\eta^3}{3E\eta}\right)-\left(\frac{E\eta^2}{2E\eta}\right)^2 \quad (1)$$

Now it will be shown that formula (1) is true for any identically distributed (perhaps dependent) random variables.

Theorem 1

Let $\eta_1, \eta_2,\dots,\eta_n$ be identically distributed random variables (may be dependent). Then for this model (described above) the expression (1) is true.

Proof

Let S_i be the epoch of the arrival of the customer i, $\Psi(S_i)$ be a number of customers, who arrived at the epoch S_i, $\Phi_i=\{S: \Psi(s)>0, S \in t_{i-1},t_i)\}$, V_i be the waiting time of all customers, who arrived during the time interval $[0, t]$.

Then according to Campbell's formula [8].

$$E\left(\frac{v_i}{\eta_i}\right)=E\left\{\sum_{S\in\Phi}\frac{(t_i-s)\Psi(s)}{\eta_i}\right\}=\int(t_i-s)\mu ds=\mu\frac{\eta_i^2}{2}.$$

Denote by w_N the conditional an average waiting time of all customers, for given $\eta_1, \eta_2,\dots,\eta_n$.

Then

$$w_N=E\left\{E\left(\frac{w_n}{\eta_1,\eta_2,\dots,\eta_N}\right)\right\}=\sum_{i=1}^{N}E\left\{E\left(\frac{v_i}{\eta_1,\eta_2,\dots,\eta_N}\right)\right\}=$$
$$=\sum_{i=1}^{N}E\left(E\frac{v_i}{\eta_i}\right)=\mu\frac{\sum_{i=1}^{N}E\eta_i^2}{2}$$
$$w=\lim_{N\to\infty}\frac{w_N}{R(t)}=\frac{E\eta^2}{2E\eta}$$

Where η is a random variable with distribution function $F(x)$.

$$\sigma^2=\frac{E\eta^3}{3E\eta}-\left(\frac{E\eta^2}{2E\eta}\right)^2 \qquad (2)$$

Similarly, for the variance we have.

Generally, formulas (1) and (2) are true for any identically distributed random variables η_i for which $\eta_i \geq 0$, $E\eta_i^3 < \infty$ for all i.

Now we shall take delays into account, i.e. from the random variables $\eta_1, \eta_2,\dots,\eta_n,\dots$ we pass to random variables $\eta_1^*, \eta_2^*,\dots,\eta_n^*,\dots$ where

$$\eta_i^* = \eta_i + g(\eta_i),\ i = 1,2,\dots,\ g \in G.$$

Here G is a class of measurable and non-negative functions. Let w^*, σ^{2*} be the expectations and variance of the delayed waiting times until service. Our interest concerns the following problem. For which systems can the introduction of delays diminish w^* (i.e.

give better service). Similarly to (1) and (2) we have.

$$w^* = \frac{E(\eta^*)^2}{2E\eta^*}, (\sigma^2)^* = \frac{E(\eta^*)^3}{3E\eta^*} - \left(\frac{E(\eta^*)^2}{2E\eta^*}\right)^2.$$

Denote w(g) a customer's average waiting time in a system with control function g, $w(0)=w$. $M(g)=w(g)-w$, $c = E\eta^2 /2E\eta$. Following [13, 14] we introduce the following conceptions.

Definition 1.

Service is improvable if there exists a function $g(x)$€G such that M(g)<0.

Theorem 1

Service can be improved if there exist $x_0 < c$ such that $F(x_0) > 0$.

Definition 2

The function $g^*(x)$ is called an optimal if.

$$\min_{g€G} M(g(x)) = M(g^*(x))$$

Theorem 2

Under the conditions of Theorem 1 the optimal function has the following form.

$$g^*(x) = max(0, c_1-x) = (c_1-x)^+$$

Where c_1 is the unique solution of the equation.

$$c_1^2 = \int_{c_1}^{\infty} (x - c_1)^2 dF(x)$$

The proofs of Theorem 1 and 2 can be found in [12].

Example 1

Let us put $F(x) = 1-exp(-x)$, $x \geq 0$. The stationary flow of customers with finite intensity arrives to service. Numerical computations yield w = 1.0; w*=0.90; $\sigma^2 = 1$, $\sigma^{*2} = 0.67$; The gain in *CWT* is 10% but in variance 33%.

Let us try to change the intensity of the intervals between neighbor instants of services. Below in Table 1 there are the results of simulation of this system for various values of intensity of input flow. Simulation shows that the gain in *CWT* is changing around 10%.

Below in Fig. 1 the behavior of *CWT* depending on delays is given. The graph shows that there exists some optimal delay function where *CWT* attains its minimum value and afterward it increases.

λ	*gain (%) (w-w*)/w*	*λ*	*gain (%) (w-w*)/w*
0.20	9.91%	0.80	9.94
0.25	9.96%	0.85	10.0
0.30	9.97%	0.90	9.99
0.35	9.88%	0.95	9.81
0.40	9.99%	1.00	10.01
0.45	9.83%	1.05	9.86
0.50	9.89%	1.10	9.91
0.55	9.89%	1.15	9.83
0.60	9.92%	1.20	9.81
0.65	9.86%	1.25	9.82
0.70	9.96%	1.30	10.0

Table 1: (Gain in a customer's average waiting time for system with exponential distribution of intervals between services).

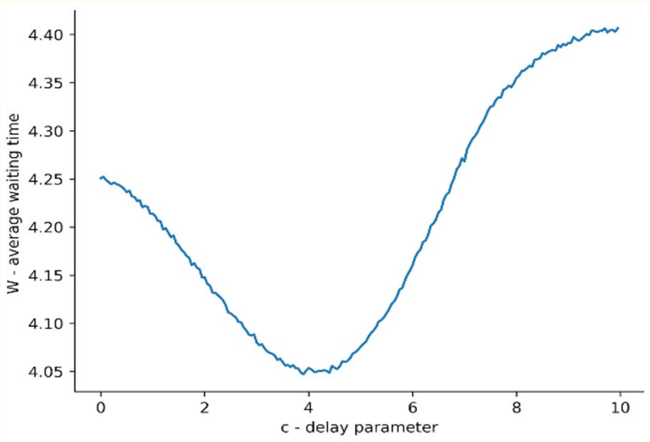

Figure 1

The following question arises. It seems that for this system the gain independent of the intensity of the intervals between services.

Theorem 3

If the intervals between neighbor services have exponential distribution, then under the conditions of Theorem 1.

$$\frac{w - w^{\cdot}}{w} = 10\%$$

Proof. . Routine transformation and calculations show that

$$w(c) = \left(\frac{\left(a^2c^2 + 2(ac)\,exp(-ac) \right)}{2a(ac + exp(-ac))} \right) \quad (3)$$

and $w = w(0) = E\eta^2/2\,E\eta = 1/\lambda$. It follows from Theorem 2 that at the point c1 a customer average waiting time $w(c_1)$ attains its minimum value then.

$$\frac{\partial w(c)}{\partial c} = 0, \; if \;\; c = c_1 \quad (4)$$

Combining (3) and (4) we have.

$$\alpha^2 c^{*2} - 2\exp(-\alpha c^*) = 0 \quad (5)$$

As the equation (5) has a unique solution, hence $\alpha c_1 = constant$.

The simple calculations yield that.

$$\frac{w - w^*}{w} = \frac{\left(\frac{1}{\alpha} - c_1\right)}{\left(\frac{1}{\alpha}\right)} = 1 - \alpha c_1 = 0.90$$

Thus, the gain in *CWT* for the model where the intervals between neighbor instants of services have exponential distribution is 10%.

Consider the model where intervals between neighbor services are independent and have uniform distribution at the interval [*0*, *b*]. Let $t_1, t_2, \ldots, t_n$ be the instants when service starts in system. We assume $t_i = i_d + x_i$, where x_i independent and identically distributed random variables having uniform distribution at the interval [-*1/2*,+*1/2*], *d* = 1.

Such scheme describes, for example, the behavior of a public transport where *id* is an ideal schedule $t_1, t_2, \ldots, t_n$ is real timetable. Here we have.

$$\eta_i = t_i - t_{i-1} = = id + x_i - (i-1)d - x_{i-1} = \text{d} + (x_i - x_{i-1}).$$

Example 2

Let *d*=1 and x_i have a uniform distribution at the interval [-1/2.+1/2]. Simple calculations yield .

w = 0.583; σ2= 0.160; and w* = 0.5626, σ2*= 0.156;

The gain in *CWT* is 3.49%. Below in Table 2 there are the results of simulation for various values of intensity of input flow for systems where intervals between services have uniform distribution.

Theorem 4

If the intervals between neighbor services have uniform distribution at the interval [*0, b*] then.

$$\frac{w - w^*}{w} = 3.5\%$$

i.e. the gain in *CWT* is 3.5% and independent of *b*.

λ	***gain, (w-w*)/w***	λ	***gain, (w-w*)/w***
1.1	3.46%	2.6	3.60%
1.2	3.55%	2.7	3.49%
1.3	3.49%	2.8	3.59%
1.4	3.60%	2.9	3.54%
1.5	3.46%	3.0	3.50%
1.6	3.46%	3.1	3.54%
1.7	3.41%	3.2	3.50%
1.8	3.50%	3.3	3.47%
1.9	3.34%	3.4	3.51%
2.0	3.46%	3.5	3.47%
2.1	3.48%	3.6	3.51%
2.2	3.48%	3.7	3.49%
2.3	3.48%	3.8	3.48%
2.4	3.50%	3.9	3.57%
2.5	3.47%	4.0	3.61%

Table 2: (Gain in a customer's average waiting time for system with exponential distribution of intervals between services).

Proof

Assume that the intervals between neighbor instants of services have uniform distribution at the interval [*0, b*]. Routine transformations and calculations give.

$$w(c) = \frac{\left(b^3 + 2c^3\right)}{3\left(b^2 + c^2\right)}, \; w = w(0) = \frac{b}{3}$$

According to Theorem 2 at the point c_1 a customer average waiting time $w(c)$ attains its minimum value then.

$$\frac{\partial w(c)}{\partial c} = 0, \quad if \;\; c = c_1 \qquad (4)$$

Simple calculations yield that from (4) it follows that.

$$c^{*3} + 3c^* b^2 - b^3 = 0$$

Hence,

$$(w(0) - w^*)/w = (b^{c*2} - 2^{c*3})/(b(b^2 + c^{*2}) = const.{=}0.35$$

Let us now compare examples 1 and 2. In example 1 we have a gain in customer average waiting time before service is around 10%, while in example 2 is only 3.5%.

The following question arises. For which distributions the gain in *CWT* is greater?

Definition 3

Introduce $k = E\eta^2 / (E\,\eta)^2$ and call it *strong coefficient of variation*.

Remind that $Var(\eta)/(E\eta)^2$ is called coefficient of variation [23].

THEOREM 5

If for two systems with service distribution functions $F_1(x)$ and $F_2(x)$ and strong coefficient of variations k_1 and k_2 and the following expression is held.

$$1/k_1^{1/2} \leq [(k_2+1)^{1/2}-1]/k_2$$

Then $w_1{}^* = w_1(g^*))/w_1 \leq w_2{}^* = w2(g^*))/w_2$ and hence $(w_2 - w_2{}^*)/w \leq (w_1 - w_1{}^*)/w$ i.e. service for the second system can be improved better.

Example 3

Let us put $F_1(x)$=1-$exp(-x)$ and $F_2(x)= x/2, x€[0, 2]$ i.e. the distribution. Numerical calculations yield:

$w_1=1, g_1{}^*(x) = (1-x)^+, k_1=2, w_1(g_1{}^*)=1/2;$

$w_2=2/3, g_2{}^*(x) = (1-x)^+, k_2=4/3, w_2(g_2{}^*)=1/2$

Since $k_1=2$, $1/ k_1^{1/2} \sim 0.71$; and $k_2 = 4/3$; $[(k_2+1)^{1/2}- 1]/k_2 \sim 0.97$;

Then $w_1{}^*/ w_1=1/2 < w_2{}^* /w_2= 1/3$ and hence $(w_1 - w_1{}^*)/w_1 = 1/2 > (w_2 - w_2{}^*)/w_2 =1/4$.

Open Problems 1

For any one parametrical distribution the gain in *CWT* is fixed depending only on the type of distribution but independent of the parameters of distribution.

It is obviously that if for two distribution functions F(x) and G(x) is held.

$$F(x) \geq G(x) \text{ or } G(x) \geq F(x) \text{ for all } x \quad (*)$$

and

$$\int_{c_1}^{\infty}(x-c_1)^2\,dF(x) = \int_{c_1}^{\infty}(x-c_1)^2\,dG(x) = c_1{}^2 \; (**)$$

Then $F(x) = G(x)$ for any x.

Open Problem 2

The question is: If the condition (*) is not held but (**) is true what we can say about $F(x)$ and $G(x)$ or what is relation between $F(x)$ and $G(x)$?

Conclusion

Mathematical models describing the behavior of wide class of queues with moving servers are constructed. The prototypes of such models can be public transportation, traffic, communication systems airport facilities and others. The control function which means delays of beginning service is introduced. The class of systems for which it is advisable to introduce delays is described. The form of optimal control function minimizing a customer's average waiting time is derived. It is shown that the optimal delay function has linear form. Numerical examples demonstrating theoretical results are given. It is shown that for systems where intervals between services have exponential or uniform distribution the gain in a customers' average waiting time independent of the parameters of these distributions. The criterion which allows to define for which systems the gain in a customers' average waiting time is greater has been derived.

References

1. Asmussen S. "Applied Probab. and Queues". Springer-Verlag (2020).
2. Baccelli F and Bremaud P. "Elements of Queueing Theory, 2003". Springer-Verlag, Berlin (2003).
3. Belyaev YuK., et al. "Markov approximation of stochastic model of motion on a two-road lane". M., МАДИ (2002): 32 (in Russian).
4. Blank M. "Ergodic properties of a simple deterministic traffic flow model". J. Stat. Phys 111 (2003): 903-930.
5. Borovkov AA. "Asymptotic methods in queuing theory". John Willey Son, NY (1984).
6. Bozejko W and Bocewicz G. "Modelling and Performance Analysis of Cyclic Systems". Springer, Studies in Systems, Decision and Control 241 (2020).
7. Cox D. "Renewal theory". Chapman and Hall (1967).
8. Franken P., et al. "Queues and Point Processes". Wiley, Chichester (1982).
9. Gazis DC. "Traffic Theory". Berlin: Springer (2002).
10. Gnedenko BV and Kovalenko IN. "Introduction to Queuing Theory". Birkhauser (1989).
11. Haight F. "Mathematical Theory of Traffic Flow". Academic Press (1968).
12. Hajiyev AH. Mammadov TSh. "Mathematical models of moving particles and their Applic". Lambert, Academic Publish, Germany (2013): 134.
13. Hajiyev AH and Mammadov TSh. "Cyclic queues with delays". RAS, Doklady, Mathem 1 (2009).
14. Hajiyev AH and Mammadov TSh. "Mathematical models of moving particles and its application". Theory of Probab. Appl 56.4 (2011): 1-14.
15. Kerner BS. "Introduction to Modern Traffic Flow Theory and Control". Berlin: Springer (2009).
16. Khinchin AYa. "Mathematical theory of queues". M (1963) (in Russian).
17. Kleinrock L. "Queuing systems". John Wiley Sons 1.2 (2008).
18. Long Z., et al. "Dynamic Scheduling of Multiclass Many-Server Queues with Abandonment: The Generalized cμ/h Rule". Oper. Res 68 (2020): 1218-1230.
19. Lee H-S and Srinivasan MM. "The shuttle dispatch problem with compound Poisson arrivals: controls at two terminals". Queuing Systems 6 (1990): 207-222.
20. Moeschlin O and Poppinga C. "Controlling traffic lights at a bottleneck with renewal arrival processes". Proc. Inst. Math. And Mech. Azerb. National Acad. Sci 14 (2001): 187-194.
21. Nagatani T. "The physics of traffic jams". Rep. Prog. Phys 65 (2002): 1311-1356.
22. Nagel K, Wagner R and Woesler R. "Still flowing: Approaches to traffic flow and traffic jam modelling". Oper. Research 51.5 (2003): 681-710.
23. Newell GF. "A simplified theory of kinematic waves in highway traffic, part I: General theory". Transp. Research Part B. Methodological (1993).
24. Newell GF. "Applications of queuing theory". London: Chapman and Hall (1982).
25. Renyi A. "On two mathematical models of the traffic on a divided highway". J. Appl. Probab 1 (1964): 311-320.
26. Ross SM. "Average delay in queues with non-stationary Poisson arrivals". J. Appl. Probab 15 (1978): 602-609.
27. Saati TL. "Elements of queueing theory, with applications". McGrow Hill Book Comp (1961).
28. Shiryaev AN. "Probability". Springer-Verlag, New York- Berlin-Heidelberg (2013): 580.
29. Zhou W, Huang W and Zhang R. "A two-stage queueing network on form postponement supply chain with correlated demands". Appl. Math. Model 38 (2014): 2734-2743.