Applied Mathematics and Computation 261 (2015) 284–301

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Threshold properties of the $M/M/1$ queue under T-policy with applications

Xuelu Zhang [a,b], Jinting Wang [a,*], Tien Van Do [c]

[a] Department of Mathematics, Beijing Jiaotong University, 100044, Beijing, China
[b] College of Science, Qilu University of Technology, 250353, Jinan, China
[c] Department of Networked Systems and Services, Budapest University of Technology and Economics, Budapest, Hungary

ARTICLE INFO

Keywords:
Queueing system
Monotonicity
Convexity or concavity
State-dependent service rate
Cost conservation

ABSTRACT

In this paper, we consider a single server queueing system with a threshold control policy and its application for controlling the energy consumption of a computing server. If the number of customers in the system is less than a threshold, the service rate is set in a low value and it also can be switched to a high value once the number of customers reaches the threshold. We study the monotonicity, convexity or concavity properties of the key performance measures of the system such as the steady-state probability distribution, the expected number of customers in the system and in the queue, the expected sojourn time in the system and the waiting time in the queue, with respect to the threshold. Based on these properties, we further study a real-life problem regarding the energy consumption of a computing server. Numerical results show that the state-dependent service policy is a promising technique to balance the energy consumption and the quality of service of a computing server.

1. Introduction

Convexity or concavity is a very important property in optimization models because it can guarantee the unique global optimal solution. As an example, for an $M/M/1$ queueing system with a First-In-First-Out (FIFO) queue discipline, the expected sojourn time is convex with respect to the service rate leads to the the expected steady-state total cost per unit time also is convex with respect to the service rate. Therefore the cost function yields its unique minimum value at a stationary point. Similarly, the expected number in the system is convex with respect to the arrival rate leads to the the expected steady-state net benefit per unit time is concave with respect to the arrival rate. So the net benefit yields the maximum value at the boundary point or the stationary point. In this regard, the convexity or concavity property is exploited in a lot of queueing models and some performance measures, such as the probability that the system is empty, the expected number in the system and in the queue, as well as the expected sojourn time in the system and the waiting time in the queue, are viewed as objective functions to search for the optimal solution including the optimal arrival rate, service rate, the number of server and the traffic intensity. For more information about how to use the convexity or concavity to determine the optimal solution, one may refer to the monograph of Stidham [1].

Not only the $M/M/1$ queueing system, other queueing models also possess similar properties. As a means of control, threshold policy is widely used in various queueing systems. Guo and Hassin [2] studied the Markovian vacation queues. Under certain

* Corresponding author. Tel.: +86 1051688449.
E-mail address: jtwang@bjtu.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.amc.2015.03.109
0096-3003/

circumstance, the expected sojourn time is a convex function and the social welfare function is a concave function. Hence these functions attain the optimal values at the stationary points and these results are crucial to determine the strategies of the customers and the managers. Dimitrakopoulos and Burnetas [3] discussed a similar queueing model with dynamic service policy and in the case of $T = 1$ the authors obtained the convexity of the sojourn time and the concavity of the social welfare function with respect to the arrival rate. But for $T \geq 2$, to the best of the authors' knowledge, there is no related results in the queueing literature.

The new contributions of our paper can be summarized as follows. (1) For the $M/M/1$ queueing system with a general threshold $T \geq 2$, we establish the monotonicity and convexity or concavity of the six performance measures with respect to the threshold T for the first time. (2) Based on the above properties, we define two functions in a practical energy consumption model: the power consumption function and the cost function. Furthermore, we show that these two functions possess some properties including monotonicity, convexity or concavity. (3) We apply the theoretic results obtained in this paper to study the power consumption and the minimum cost of operating a computing node. Through adjusting the clock frequency of a processor and the threshold, a cost of operating server and the quality of service in the term of waiting time can be balanced.

The organization of this paper is as follows. Section 2 presents a review of the related literature. The model under consideration is described in Section 3 and we give some notations and preliminary results in Section 4. Moreover, we also discuss the relation between this model and the normal $M/M/1$ queue. According to the value of traffic intensity, we study the properties of these performance measures in Section 5 and Section 6, respectively. In Section 7 we give some practical applications in energy consumption problem. Finally, we give some conclusions and future research topics of our research. The properties of these performance measures with respect to the traffic intensity are summarized in the Appendix.

2. Literature review

The work on queueing systems with threshold policy is extensive, dating back to the seminal study of Yadin and Naor [4], in which the service control N-policy was introduced. Other service control queueing systems include the T-policy and the D-policy, which were introduced by Heyman [5] and Balachandran [6], respectively. A recent survey of the topic area can be found in Tadj and Choudhury [7] and Stidham [1]. Of particular note are Tian and Zhang [8] and Ke et al. [9], both of which give updated developments in vacation queueing models that related to the queueing systems considered in this paper closely.

The N-policy introduced by Yadin and Naor [4] assumed that the server starts to work when the number of customers reaches a predetermined threshold N and once the server starts to work, it does not take the next vacation until the system is empty. This policy has been used in queueing system with batch arrival [10], machine repair problem [11], multi-threshold vacation policy [12], wireless sensor networks [13] and strategic behavior of customers [2]. A short survey under N-policy can be seen in Jayachitra and Albert [14].

Heyman [5] introduced the T-policy, assuming that the server will be activated after T time units when the last busy period ends. Zhang et al. [15] established the convexity of the cost function in T based on the service cycle for the $M/G/1$ queue with T-policy. Wang et al. [16] investigated the T-policy of $M/G/1$ queue with server breakdowns and startup times. Because the expected cost function is convex in the decision variable T, they obtained the optimal threshold T. Ke [17] introduced a modified model which takes at most J vacations. If no customers are found by the end of the jth vacation, the server stays in the system with a dormant state. Recent studies on the variants of the T-policy model can be found in [18] and references therein.

In contrast to the N-policy and T-policy, the D-policy considers the total service times of the waiting customers to decide when the server is turned back. The pioneering work in this field is Balachandran [6] who first introduced the D-policy. Recently, Artalejo [19] investigated the steady-state distribution, moments and some variants of $M/G/1$ queue under the D-policy. Later on, a great number of papers on the D-policy were developed, such as Lee et al. [20] investigated the performance measures of $M^X/G/1$ queue; Choudhury [21] dealt with the steady state behavior with two phase service of $M/G/1$ queue and carried out an extensive analysis for the queue size distributions; Lee et al. [22] studied the steady-state queue length and waiting time of the $M/G/1$ queue with multiple vacations and then obtained the mean performance measures; Unlike the references above, Lee et al. [23] obtained the performance measures under the LCFS discipline (Last-Come-First-Served) and the optimal value D under a linear cost function.

Although extensive works have been done in the queueing systems with threshold policy over the past three decades, there is still a growing body of work that studies new problems under different modeling assumptions motivated by real systems, for example, queues with multiple adaptive vacations (MAV), queues with randomized control policy $\langle T, p\rangle$-policy and $\langle p, N\rangle$-policy, etc. We will not review this literature herein, but remark that the tractable analytical framework proposed in our paper has the potential to be applicable in those settings as well.

Regarding the monotonicity, convexity or concavity properties of the key performance measures, Rolfe [24] established that the expected sojourn time is convex and decreasing in c for the $M/D/c$ queue. Then, Dyer and Proll [25] proved Rolfe's conjecture for an $M/M/c$ queue. Grassmann [26] stated that the expected number of customers in the system and the mean number of customers in the queue are convex with respect to the traffic intensity in the $M/M/c$ queue. It is worth mentioning that Lee and Cohen [27] independently got the same results. Remarkably, the sharp bounds and simple approximations are introduced when the closed form of performance measures is unknown or complicated. This technique can reduce the computational burden and lead to a simple formula that describes the property of performance measure. Lots of approximation methods and results have been developed, such as Hokstad [28], Nozaki and Ross [29] and Harel [30,31]. Using the approximation methods, Harel and Zipkin [32] showed that the average sojourn time and the standard deviation are jointly convex in arrival and service rates in