Applied Mathematical Sciences, Vol. 12, 2018, no. 25, 1239 - 1249
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2018.88123

A Queueing Model for Sleep as a Vacation

Nian Liu

School of Mathematics and Statistics
Central South University
Changsha 410083, Hunan, China

Myron Hlynka

Department of Mathematics and Statistics
University of Windsor
Windsor, Ontario, Canada N9B 3P4

Copyright © 2018 Nian Liu and Myron Hlynka. This article is distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Vacation queueing systems are widely used as an extension of the classical queueing theory. We consider both working vacations and regular vacations in this paper, and compare systems with vacations to the regular $M/M/1$ system via mean service rates and expected numbers of customers, using matrix-analytic methods.

Mathematics Subject Classification: 60K25, 90B22, 60G20

Keywords: vacation queues, working vacation, matrix-analytic methods, quasi birth and death processes

1 Introduction

In an article in the journal Science in 2013, Xie et al. ([9]) stated "the restorative function of sleep may be a consequence of the enhanced removal of potentially neurotoxic waste products that accumulate in the awake central nervous system" indicating the value of sleep in changing the parameters of the brain's functioning. We can choose to consider the brain as a server in a queueing

system which decreases its service rate over time but recovers after it has a rest (vacation).

Vacation queueing systems have been studied by many authors with different models ([1], [3], [5], [8], [10]). Working vacations, introduced by Servi and Finn(2002)[7], refer to a time period, during which the service slows but does not stop. Servers would gradually get exhausted during continuous work, but the service rate could increase after a vacation of the server. We include two types of systems in this paper. The first kind of system is the regular $M/M/1$ system, in which the server works without vacations, and the service rate is a constant with a relatively low value [2]. The other kind of system also has exponential interarrival and service times. However, the service rate changes after each state transition. When the service rate decreases to a certain value, the server stops working and has a vacation, after which the service rate would return to the highest level.

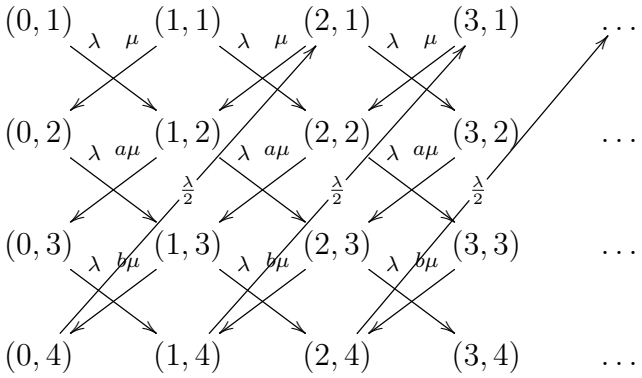
To compare the performances of different queueing systems, two commonly used measures are the expected waiting time of a customer, $E(W)$), and the expected number of customers $E(L)$ in the system. These are related via Little's formula [2]. In this paper, we only use $E(L)$ to measure performance of different queueing systems. Values of $E(L)$ are obtained using matrix-analytic methods ([4], [6]).

We show that the system with vacations performs better than the regular $M/M/1$ system under certain conditions.

2 Quasi Birth and Death Processes; 4 Phases

In this section, we compare a queueing system with working vacations with a regular $M/M/1$ system having a constant service rate.

Consider the decrease of service rate over time as working vacations ([7]), during which the server works with lower efficiency. The number of customers in the system and states of the server form a continuous time Markov process $\{(X(t), Y(t)), t \geq 0\}$, where X is the level variable (number of customers) and Y is the phase variable (server efficiency with low values indicating a high efficiency). Each state (n, i) with $X(t) = n > 0$ and $Y(t) = i < 4$ moves to $(n-1, i+1)$ with rate μ_i, or to $(n+1, i+1)$ with rate λ. For all $n \in \mathbb{N}$, state $(n, 4)$ will always go to $(n+2, 1)$ with rate $\lambda/2$. State $(0, i)$ will always go to $(1, i+1)$ with rate λ, $i = 1, 2, 3$. Set $\mu_1 = \mu$, $\mu_2 = a\mu$ and $\mu_3 = b\mu$ $(0 < b < a < 1)$. The process can be shown by the following network. The system takes a working vacations when $Y = 2$ or 3, having a regular vacation with interval $\sim Exp(\lambda/2)$ when $Y = 4$.



The motivation for the model is that each arrival or service completion takes time and reduces the server's efficiency so Y (phase) increases at each step ($i = 1, 2, 3$). For $i = 4$, the server is exhausted and even though there may be customers to be served, the server takes a vacation long enough for 2 more customers to arrive, and then begins service with renewed vigor and first level of efficiency. Setting up the model in this way keeps the transition between states exponential at all times. The interarrival rate for customers is λ so the expected time between until the next customer is $1/\lambda$. The expected time for two customers to arrive is $2/\lambda$ so we take the arrival rate to be $\lambda/2$ to move from state $(n, 4)$ to state $(n + 2, 1)$ (vacation time). Another approach could have used the sum of two exponentials (each with rate λ) but we can keep our model simpler by using rate $\lambda/2$ to have 2 customers arrive. The two approaches are not identical, though the mean times are the same, but we keep our state space more tractable using our approach.

Theorem 2.1. *The system is stable if* $\lambda < \dfrac{\frac{\mu}{\lambda+\mu} + \frac{a\mu}{\lambda+a\mu} + \frac{b\mu}{\lambda+b\mu} + 0 \cdot \frac{2}{\lambda}}{\frac{1}{\lambda+\mu} + \frac{1}{\lambda+a\mu} + \frac{1}{\lambda+b\mu} + \frac{2}{\lambda}}$

Proof. It is sufficient to consider the situation when the level is large as that determines the stability condition. For states with phase variable $Y = i$ ($i = 1, 2, 3, 4$), and level X large, let v_i be the state transition rate, and let w_i be proportion of sojourn time in those states.

$$w_i = \frac{\frac{1}{v_i}}{\sum_{i=1}^{4} \frac{1}{v_i}} \tag{1}$$

where $v_1 = \lambda + \mu$, $v_2 = \lambda + a\mu$, $v_3 = \lambda + b\mu$, $v_4 = \lambda/2$.

The average service rate of the system (for large level X) should be calculated as a weighted average.

$$\begin{aligned}\bar{\mu} &= \sum_{i=1}^{4} w_i \mu_i \\ &= \frac{\frac{\mu}{\lambda+\mu} + \frac{a\mu}{\lambda+a\mu} + \frac{b\mu}{\lambda+b\mu} + 0 \cdot \frac{2}{\lambda}}{\frac{1}{\lambda+\mu} + \frac{1}{\lambda+a\mu} + \frac{1}{\lambda+b\mu} + \frac{2}{\lambda}} \end{aligned} \tag{2}$$

The system is stable if $\lambda < \bar{\mu}$. The result follows. □

We note in the previous proof that $\bar{\mu}$ is a function of λ. To emphasize this, we define

$$g(\lambda) \triangleq \frac{\frac{\mu}{\lambda+\mu} + \frac{a\mu}{\lambda+a\mu} + \frac{b\mu}{\lambda+b\mu} + 0 \cdot \frac{2}{\lambda}}{\frac{1}{\lambda+\mu} + \frac{1}{\lambda+a\mu} + \frac{1}{\lambda+b\mu} + \frac{2}{\lambda}}.$$

Unfortunately, for our 4 phase model, it turns out that regardless of λ, μ, a, b, the expected number of customers will be shorter under a regular $M/M/1$ model with service rate $b\mu$ than under our model that allows for a vacation, at the cost of two customers arriving. We prove this as follows.

Theorem 2.2. *For the 4 phase model which is stable (i.e. $g(\lambda) > \lambda$), $g(\lambda)$ is always smaller than $b\mu$.*

Proof. First note that in our 4 phase model, states (0,1), (0,2) and (1,1) are not recurrent. Further, the average service rate that appears for large level X is an upper bound on the rate for small levels (like 1). So we will work with the service rate for large levels. Now

$$g(\lambda) - b\mu = -\mu \cdot \frac{\lambda^3(4b-a-1) + 2\lambda^2\mu(ab+b-a+2b^2) + 3\lambda\mu^2b^2(a+1) + 2ab^2\mu^3}{3\lambda\mu^2(a+b+ab) + 4\lambda^2\mu(a+b+1) + 5\lambda^3 + 2ab\mu^3}$$

The denominator is always positive so we define

$$f(\lambda) \triangleq \lambda^3(4b-a-1) + 2\lambda^2\mu(ab+b-a+2b^2) + 3\lambda\mu^2b^2(a+1) + 2ab^2\mu^3,$$

Note that $g(\lambda) - b\mu < 0 \Leftrightarrow f(\lambda) > 0$.
We will view the situation graphically by considering $f(\lambda)$ which is usually a cubic in λ.
Case 1: $4b - a - 1 = 0$. Then $f(\lambda)$ becomes a quadratic. Also $a = 4b - 1$. The coefficient of λ^2 in $f(\lambda)$ is $2\mu(ab + b - a + 2b^2) = 2\mu(a(b+1) + b + 2b^2)$, which is > 0, as the quadratic is convex. The two real roots of $f(\lambda) = 0$ are $-b\mu$ and $-\frac{b\mu(4b-1)}{6b^2-4b+1}$, which are both negative. So the value of $f(\lambda)$ is positive for value of λ which is greater than the largest root so $f(\lambda) > 0$ for $\lambda > 0$, as desired.
Case 2: When $4b - a - 1 < 0$, $f(\lambda)$ is a cubic with a negative coefficient for the λ^3 term. Let $A = \sqrt{9a^2b^2 - 4a^2b - 14ab^2 + 4a^2 - 4ab + 9b^2}$. The 3 roots

of $f(\lambda) = 0$ are $-b\mu, -\frac{\mu(3ab-2a+3b+A)}{2(4b-a-1)}$ and $-\frac{\mu(3ab-2a+3b-A)}{2(4b-a-1)}$. Two of the three roots of $f(\lambda)$ are negative with the largest root $-\frac{\mu(3ab-2a+3b+A)}{2(4b-a-1)}$. So the cubic $f(\lambda)$ will be positive between the second largest root and the largest root, after which it becomes negative. But for λ greater than the largest root, we have $g(\lambda) > \lambda$ so we are outside the stable region of the system. So our result is still true.

Case 3: When $4b-a-1 > 0$, $f(\lambda)$ is a cubic with a positive coefficient for the λ^3 term. Again, we get 3 roots of $f(\lambda) = 0$. The largest of the three roots is $-\frac{\mu(3ab-2a+3b-A)}{2(4b-a-1)}$. However, the largest root would be a negative number under the following analysis.

$$\begin{aligned} 4b-a-1>0 &\Rightarrow b<a<4b-1 \\ &\Rightarrow b<4b-1 \\ &\Rightarrow b\in(\frac{1}{3},1) \\ 0<a<1 &\Rightarrow \frac{a}{a+1}\in(0,\frac{1}{2}) \\ &\Rightarrow \frac{2a}{3a+3}\in(0,\frac{1}{3}) \\ &\Rightarrow b>\frac{a}{a+1} \\ &\Rightarrow 3ab-2a+3b>0 \end{aligned}$$

Thus, there would be

$$\begin{aligned} &3ab-2a+3b-\sqrt{9a^2b^2-4a^2b-14ab^2+4a^2-4ab+9b^2}<0 \\ &\Leftrightarrow (3ab-2a+3b)^2<9a^2b^2-4a^2b-14ab^2+4a^2-4ab+9b^2 \\ &\Leftrightarrow ab(4b-a-1)<0 \end{aligned}$$

Since $f(\lambda)$ is a cubic with a positive coefficient for λ^3, then $f(\lambda)$ must be positive for all λ larger than the largest root of $f(\lambda) = 0$ so $f(\lambda) > 0$ for all $\lambda > 0$.

The result follows. □

Hence, when the service rate of a regular $M/M/1$ system equals the lowest service rate in the 4 phase system, the 4 phase system will always have a lower overall average service rate than the $M/M/1$ system. This means that the M/M/1 system will have a lower expected number of customers than the 4 phase system and there is no advantage in using the 4 phase system. As a result, we consider a 5 phase system.

3 Quasi-Birth-and-Death Process with 5 States of Service

Add one more phase standing for $c\mu$ $(0 < c < b < a < 1)$ as service rate to the former system, and change the constant service rate in $M/M/1$ to $c\mu$. The proportion of time when the server stops working in the new system would decrease. With fixed a, b, c and μ, there would be a range of λ such that the average service rate in the new system is higher than that in $M/M/1$, and the expected number of customers would be reduced when servers take some time to rest. The statement could be proved more succinctly by numerical methods rather than analytical ones.

3.1 Matrix-Analytic Methods for Calculating the Expected Number of Customers

For the 5-phase system, let the states be
$(0,1),(0,2),(0,3),(0,4),(0,5),(1,1),(1,2),(1,3),(1,4),(1,5),(2,1),(2,2),\ldots$. The Q-matrix (infinitesimal matrix) of the system with 5 states of service is

$$Q1 = \begin{pmatrix} A_{00} & A_{01} & A_{02} & & & & \\ A_{10} & A_{11} & A_{01} & A_{02} & & & \\ & A_{10} & A_{11} & A_{01} & A_{02} & & \\ & & A_{10} & A_{11} & A_{01} & A_{02} & \\ & & & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

where

$$A_{00} = \begin{pmatrix} -\lambda & & & & \\ & -\lambda & & & \\ & & -\lambda & & \\ & & & -\lambda & \\ & & & & -\lambda/2 \end{pmatrix}_{5\times 5}, A_{01} = \begin{pmatrix} 0 & \lambda & & & \\ & 0 & \lambda & & \\ & & 0 & \lambda & \\ & & & 0 & \lambda \\ & & & & 0 \end{pmatrix}_{5\times 5}$$

$$A_{02} = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \\ \lambda/2 & 0 & \ldots & 0 \end{pmatrix}_{5\times 5}, A_{10} = \begin{pmatrix} 0 & \mu & & & \\ & 0 & a\mu & & \\ & & 0 & b\mu & \\ & & & 0 & c\mu \\ & & & & 0 \end{pmatrix}_{5\times 5}$$

$$A_{11} = \begin{pmatrix} -(\lambda+\mu) & & & & \\ & -(\lambda+a\mu) & & & \\ & & -(\lambda+b\mu) & & \\ & & & -(\lambda+c\mu) & \\ & & & & -\lambda/2 \end{pmatrix}_{5\times 5}$$

Since states $(0,1)$, $(0,2)$ and $(1,1)$ are not positive recurrent, we delete the corresponding rows and columns from Q1. Let

$$A_0 = \begin{pmatrix} A_{02} & 0 \\ A_{01} & A_{02} \end{pmatrix}, A_1 = \begin{pmatrix} A_{11} & A_{01} \\ A_{10} & A_{11} \end{pmatrix}, A_2 = \begin{pmatrix} 0 & A_{10} \\ 0 & 0 \end{pmatrix}$$

After that, the Q matrix could be written as

$$Q = \begin{pmatrix} B_{11} & B_{12} & & & & \\ B_{21} & A_1 & A_0 & & & \\ & A_2 & A_1 & A_0 & & \\ & & A_2 & A_1 & A_0 & \\ & & & \ddots & \ddots & \ddots \end{pmatrix} \tag{3}$$

where

$$B_{11} = \begin{pmatrix} -\lambda & 0 & 0 & 0 & 0 & -\lambda & 0 \\ 0 & -\lambda & 0 & 0 & 0 & 0 & -\lambda \\ 0 & 0 & -\lambda/2 & 0 & 0 & 0 & 0 \\ a\mu & 0 & 0 & -(\lambda + a\mu) & 0 & 0 & 0 \\ 0 & b\mu & 0 & 0 & -(\lambda + b\mu) & 0 & 0 \\ 0 & 0 & c\mu & 0 & 0 & -(\lambda + c\mu) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\lambda/2 \end{pmatrix}_{7\times 7}$$

$$B_{12} = \begin{pmatrix} 0 & & & & & & \\ 0 & 0 & & & & & \\ \lambda/2 & 0 & 0 & & & & \\ 0 & 0 & \lambda & 0 & & & \\ 0 & 0 & 0 & \lambda & 0 & & \\ 0 & 0 & 0 & 0 & \lambda & 0 & \\ 0 & 0 & 0 & 0 & 0 & -\lambda/2 & 0 \end{pmatrix}_{7\times 10}, \quad B_{21} = \begin{pmatrix} \mu & 0 & 0 & 0 \\ & a\mu & 0 & 0 \\ & & b\mu & 0 \\ & & & c\mu \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \end{pmatrix}_{10\times 7}$$

Note that Q has the form of a quasi birth and death process while $Q1$ did not.

Let $\vec{\pi}_0 = (\pi_{(0,3)}, \pi_{(0,4)}, \pi_{(0,5)}, \pi_{(1,2)}, \pi_{(1,3)}, \pi_{(1,4)}, \pi_{(1,5)})$.

For $j \geq 1$, let $\vec{\pi}_j = (\pi_{(j+1,1)}, \ldots, \pi_{(j+1,5)}, \pi_{(j+2,1)}, \ldots, \pi_{(j+2,5)})$. Let $\vec{\pi} = (\vec{\pi}_0, \vec{\pi}_1, \ldots)$. From $\vec{\pi}Q = \vec{0}$, we have:

$$\vec{\pi}_0 B_{11} + \vec{\pi}_1 B_{21} = 0 \tag{4}$$

$$\vec{\pi}_0 B_{12} + \vec{\pi}_1 (A_1 + RA_2) = 0 \tag{5}$$

Also,

$$\vec{\pi}_j = \vec{\pi}_1 R^{j-1}, \quad \forall j \geq 1 \tag{6}$$

$$R^2 A_2 + RA_1 + A_0 = 0 \tag{7}$$

where the R matrix (10×10) can be found using iteration.

$$\begin{aligned} R(0) &= [0], \\ R(n+1) &= -\sum_{k=0,k\neq 1}^{\infty} R^k(n) A_k A_1^{-1}, n \geq 0 \\ &= -(A_0 A_1^{-1} + R^2(n) A_2 A_1^{-1}). \end{aligned} \tag{8}$$

Let $\vec{e}$ be a column vector if 1's of various lengths, as appropriate. Using the expression in equation (6), $\vec{\pi}\vec{e} = 1$ implies

$$\vec{\pi}_0 e + \vec{\pi}_1 (I - R)^{-1} e = 1. \tag{9}$$

Using (4), (5) and (9), $\vec{\pi}_0$ and $\vec{\pi}_1$ can be obtained. From these, limiting probabilities for all states are obtained using (6). Next

$$\begin{aligned} E(L) &= \sum_{j=1}^{\infty} \vec{\pi}_1 R^{j-1} (j\vec{e} + (1,1,1,1,1,2,2,2,2,2)^T) \\ &= \vec{\pi}_1 (\sum_{j=1}^{\infty} j R^{j-1} \vec{e} + \sum_{j=1}^{\infty} R^{j-1} (1,1,1,1,1,2,2,2,2,2)^T) \\ &= \vec{\pi}_1 ((I-R)^{-2} \vec{e} + (I-R)^{-1} (1,1,1,1,1,2,2,2,2,2)^T) \end{aligned} \tag{10}$$

3.2 Numerical Example of Comparing Two Systems

Set $a = 0.99$, $b = 0.98$ and $c = 0.1$. Then the expected number in the two systems (5 phase system vs M/M/1 with lowest service rate of the 5 phase system) in terms of λ and μ is shown in Figure 1.

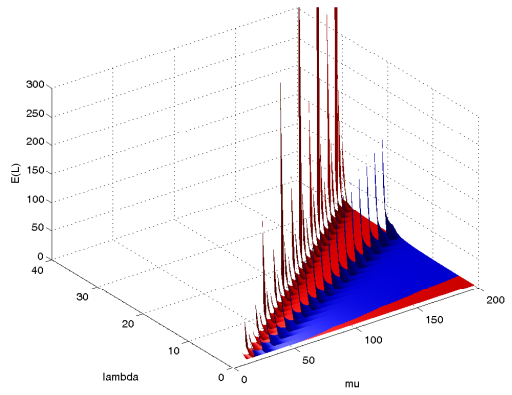


Figure 1: Expected Numbers of Customers Varying with λ and μ

The expected numbers of the 5 phase system are plotted in red, and those of the $M/M/1$ system are plotted in blue. We see Figure 2 that the new system is better than the regular one only when the load $\frac{\lambda}{\mu}$ is within a certain range (k_1, k_2), where $k_2 = c$. The value of λ/μ such that two systems have the same $E(L)$ is k_1.

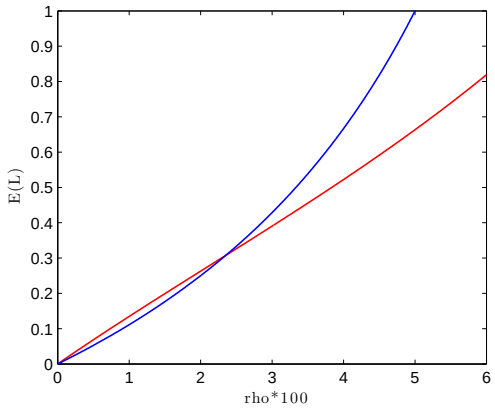


Figure 2: Expected Numbers of Customers Varying with $\rho = \lambda/\mu$

The value of k_1 is estimated using MATLAB. For $a = 0.99$, $b = 0.98$ and $c = 0.1$, k_1 is calculated to be around 0.02358. Thus, with $\lambda/\mu \in (0.02358, 0.10000)$, the 5 phase system performs better than the regular $M/M/1$ system.

4 Conclusion

Through comparisons on mean service rates and expected numbers of customers, we are able to state that, with two kinds of working vacations and one phase for regular rest, a 4 phase queueing system can never outperform the regular $M/M/1$ system with the minimal service rate. However, after we add another phase for the working vacation, it is possible for the queueing system to outperform the regular $M/M/1$ system, but only when the ratio of λ and μ is within a certain range. The boundary of that range depends on the service rate decrease during working vacations. Basically, there is evidence that sleep is a valuable tool in allowing the brain to recuperate to its normal functioning. In a better model of the brain's recovery system, there would be a larger number of phases and the service rate would be large initially and drop off close to zero in the final phase. Our limited 5 phase model indicates that there is a real possibility for improved functioning with a good sleep cycle. The exact parameters of such a cycle would need to be estimated by a large data set,

but the analysis here suggests that such a data collection is a valuable resource.

Acknowledgements. We acknowledge funding and support from MITACS Global Internship program, CSC Scholarship.

References

[1] B. Doshi, Queueing systems with vacations - A survey, *Queueing Systems,* **1** (1986), 29 - 66. https://doi.org/10.1007/bf01149327

[2] D. Gross, J. Shortle, J. Thompson and C. Harris, *Fundamentals of Queueing Theory*, (Fourth ed.), Wiley, New York, 2008. https://doi.org/10.1002/9781118625651

[3] P. Guo and R. Hassin, Strategic behavior and social optimization in Markovian vacation queues: The case of heterogeneous customers, *European Journal of Operational Research,* **222** (2012), 278 - 286. https://doi.org/10.1016/j.ejor.2012.05.026

[4] Q.M. He, *Fundamentals of Matrix-Analytic Methods,* Springer, New York, 2014. https://doi.org/10.1007/978-1-4614-7330-5

[5] O. Isijola-Adakeja and O. Ibe, M/M/1 Multiple Vacation Queueing Systems With Differentiated Vacations and Vacation Interruptions, *IEEE Access,* **2** (2014), 1384 - 1395. https://doi.org/10.1109/access.2014.2372671

[6] G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling,* ASA-SIAM, 1999. https://doi.org/10.1137/1.9780898719734

[7] L. Servi and S. Finn, M/M/1 queues with working vacations (M/M/1/WV), *Performance Evaluation,* **50** (2002), 41 - 52. https://doi.org/10.1016/s0166-5316(02)00057-3

[8] N. Tian, Z.G. Zhang, *Vacation Queueing Models Theory and Applications,* Springer, New York, 2006. https://doi.org/10.1007/978-0-387-33723-4

[9] L. Xie, H. Kang, Q. Xu, M.J. Chen, Y. Liao, M. Thiyagarajan, J. O'Donnell, D.J. Christensen, C. Nicholson, J.J. Iliff, T. Takano, R. Deane, M. Nedergaard, Sleep drives metabolite clearance from the adult brain, *Science,* **342** (2013), 373 - 377. https://doi.org/10.1126/science.1241224

[10] M. Zhang and Z. Hou, Performance analysis of M/G/1 queue with working vacations and vacation interruption, *Journal of Computational and*

Applied Mathematics, **234** (2010), 2977 - 2985.
https://doi.org/10.1016/j.cam.2010.04.010

Received: September 10, 2018; Published: October 15, 2018