ALGORITHMIC COMPUTATION OF MAP/PH/1 QUEUE WITH
FINITE SYSTEM CAPACITY AND TWO-STAGE VACATIONS

QINGQING YE∗
School of Mathematics and Statistics
Nanjing University of Information Science and Technology
Nanjing, Jiangsu 210044, China

(Communicated by the Wuyi Yue)

Abstract. In this article, we study a discrete-time MAP/PH/1 queue with
finite system capacity and two-stage vacations. The two-stage vacations pol-
icy which comprises single working vacation and multiple vacations is featured
by that once the system is empty during the regular busy period, the system
first takes the working vacation during which the server can still provide the
service but at a lower service rate. After this working vacation, if the sys-
tem is empty, the server will take a vacation during which the server stops its
service completely, otherwise, the server resumes to the normal service rate.
For this queue, using the matrix-geometric combination solution method, we
obtain the stationary probability vectors when the traffic intensity is not equal
to one. In addition, we discuss the spectrum properties of the key matrices
and give their decomposition results that can be used to reduce the computa-
tion loads. Further, waiting time is derived by constructing an absorbing
Markov chain. Various performance measures are obtained. At last, some nu-
merical examples are presented to show the impacts of system parameters on
performance measures.

1. Introduction. Queueing theory is the well recognized mathematical tool for
solving the problems of design, performance evaluation, capacity planning and op-
timization of various real-world systems, and has extensive applications in the areas
of computer networks, production management, communication and so forth. As a
branch of the study of queueing theory, vacation queues have been investigated by
many authors, the readers may refer to surveys of Doshi [11] and the book by Tian
and Zhang [31], as well as references therein.

In the vacation queues, there is an underlying assumption that the server com-
pletely stops its service during the vacation period. In 2002, Servi and Finn [29]
introduced a working vacation queue model with the idea of offering services at a
lower rate whenever the server is on vacation. They first studied an M/M/1 queue
with working vacations and applied their results to analyze a WDM optical access
network using multiple wavelengths which can be reconfigured. Their model was

2020 Mathematics Subject Classification. Primary: 60K25; Secondary: 68M20.
Key words and phrases. MAP/PH/1 queue, two-stage vacations, finite capacity, matrix-
geometric combination solution, loss probability, waiting time.

The first author is supported by Natural Science Foundation of Jiangsu Province (grant Nos.
BK20180783, 18KJB110021), and The Startup Foundation for Introducing Talent of Nanjing Uni-
versity of Information Science and Technology(grant No. 2017r082).
∗ Corresponding author: Qingqing Ye.
generalized to the case of $M/G/1$ type queue in (Kim et al. [17], Wu and Takagi [35], Li et al. [21]), and to the case of $GI/M/1$ type queue in (Baba [5], Li and Tian [21]). Parallel to the results of the continuous time queue with working vacations, Tian et al. [32] discussed the discrete time $Geom/Geom/1$ queue with working vacations. Subsequently, Li et al. [19] and Li and Tian [20] analyzed the discrete-time $GI/Geo/1$ type queue with multiple working vacations and the discrete-time $GI/Geo/1$ type queue with multiple working vacations and vacation interruption, respectively. Extension to the batch arrival discrete-time $Geo/GI/1$ queue with working vacations was studied by Li et al. [22]. Recently, Gao et al. [13], Luo et al. [25], Yang and Wu [36] and others considered the working vacation queueing systems with various features. More details and recent works related to working vacation, the readers can refer to the surveys of Tian et al. [33] and Chandrasekaran et al. [9].

There is a wide variety of literature regarding the vacation queues or working vacation queues, however, a very few works focused on the combination of vacation policy and working vacation policy. The recent work Ye and Liu [37] presents a two stage vacations policy which comprises single working vacation and multiple vacations. The feature of this vacation policy is that once the system becomes empty during the regular busy period, the server first takes the working vacation for the possible arriving customers at a lower service rate. After this working vacation, if the system is still empty, the server will take a vacation during which the server stops its service completely, otherwise, the server will resume to the normal service rate immediately. From the theoretical perspective, the two stage vacations policy is the extension of classical working vacation policy and vacation policy, in fact, when the length of duration of working vacation degenerates to zero, the queue with two stage vacations policy becomes to queue with multiple vacations, on the other hand, when the length of duration of vacation degenerates to zero, the queue with two stages vacation policy becomes to queue with single working vacation. From the application perspective, this vacation policy has a potential application in service systems with variable service rate, for example, the bank windows usually can be divided into personal service windows and corporation business windows, when the customers with corporation business are relatively few, the bank usually keep one corporation business window in operation for potential customers. This Min Speed period can be viewed as working vacation. In addition, if the customers with personal service grow too large and there is no corporation business to be deal with, the corporation business windows may temporarily stop their service and start to serve for the personal service. This period can be viewed as vacation. In above example, we can regard that the corporation business windows take the two-stages vacations policy. For the queue with two-stages vacations policy, Ye and Liu [37] first analyzed the $M/M/1$ type queue with this two stage vacation policy. Subsequently, Ye and Liu [38] generalized this model to the case of $GI/M/1$ queue. Ye and Liu [39] generalized this model to the case of discrete time $Geom/Geom/1$ queue, Ye [40] generalized this model to the case of batch arrival queue.

The discrete-time Markovian arrival process (D-MAP), which is a good representation of bursty and correlated traffic arising in telecommunication networks based on the Asynchronous Transfer Mode (ATM) environment, can represent a variety of arrival processes which include the Bernoulli arrival process, discrete-time PH-renewal process, and Markov modulated Bernoulli process. For the details on D-MAP and related topics, the readers can refer to Neuts [27] and Neuts [28]. For the queue with MAP and the vacation policy, Alfa [1] first analyzed a discrete-time
MAP/PH/1 queue with exhaustive and non-exhaustive service using the matrix-
geometric procedure. Subsequently, Alfa [2] studied the discrete MAP/PH/1 queue
with gated time-limited service. In Goswami and Selvaraju [15], the discrete time
MAP/PH/1 queue with multiple working vacations is analyzed. Subsequently, Sreeni-
vasan et al. [30] analyzed the MAP/PH/1 queue with working vacation, vacation
interruption and N policy. Baba [6] analyzed the M/PH/1 queue with working va-
ciations and vacation interruption. Recently, Chakravarthy and Ozkar [10] studied
the MAP/PH/1 queue with working vacation and crowdsourcing. Vadivu and Aru-
maganathan [34] analyzed an MAP/G/1/N queue with two phases of service under
single (multiple) vacation(s), Banik [7] studied a BMAP/R/1 queue with R-type
multiple working vacations.

In this paper, we generate the two-stage vacations policy studied in Ye and Liu
[37, Ye and Liu [38], Ye and Liu [39] and Ye [40] to the MAP/PH/1 type queue, in
addition, we assume the system capacity is finite and any customers who meet a full
capacity is rejected and lost, which is more accordant with the real-life situation.
The rest of this paper is organized as follows. In Section 2, we provide a description
of MAP/PH/1/N queue with two-stage vacations policy. The steady-state analysis
of the model is presented in Section 3. Specifically, in Section 3.1, we formulate
the model as a finite quasi-birth-death (QBD) process; In Section 3.2, we discuss some
properties of the key rate matrices; In Section 3.3, the stationary probability vectors
are given by the matrix-geometric combination method. Section 4 obtains the loss
probability and its limit when the system capacity goes to infinite. The waiting
time distribution is analyzed in Section 5. Section 6 presents several important
performance measures. Several numerical examples are presented in Section 7. The
last is the conclusion.

2. Model description. We consider a discrete-time MAP/PH/1 queue with two-
stage vacation policy and finite system capacity which can be defined explicitly as
follows.

Customers arrives at the system according to a discrete time Markovian arrival
process, which can be described by two $n$ dimensional substochastic matrices $D_0$
and $D_1$. $D_0$ represents the transition probability from phase $i$ to $j$ without an
arrival, $D_1$ represents the transition probability from phase $i$ to $j$ with an arrival.
We assume that the matrix $D = D_0 + D_1$ is irreducible stochastic and $\theta$ is the
stationary distribution vector of $D$, i.e., $\theta D = \theta$ and $\theta e = 1$, where $e$ is the column
vector of ones of appropriate dimension, then the stationary arrive rate can be given
by $\lambda = \theta D e$.

The two stage vacation policy which comprises single working vacation and mul-
tiple vacations is described as follows. During the normal service period, the cus-
tomers are served according to a phase type distribution with representation $(\beta, S_1)$
of order $m_1$. Let $S_0^0 = e - S_1 e$ and we assume that $S_1 + S_0^0 \beta$ is irreducible and let $\xi_1$
be the row vector satisfying $\xi_1 (S_1 + S_0^0 \beta) = \xi_1$ and $\xi_1 e = 1$, then the service rate
during the regular service period can be expressed as $\mu_b = \xi_1 (S_0^0 \beta) e$. When the
system becomes empty, a working vacation that follows a phase type distribution
with representation $(\alpha, T_1)$ of order $r_1$ is taken. During the working vacation period,
the server can still provide the service to the potential customers, but the service
time follows a PH-distribution with representation $(\delta, S_2)$ of order $m_2$. We also let
$S_0^2 = e - S_2 e$ and assume $S_2 + S_0^2 \delta$ is irreducible, and the service rate during the
working vacation period can be expressed as $\mu_v = \xi_2 (S_0^2 \delta) e$, in which $\xi_2$ satisfies
\( \xi_2 (S_2 + S_0^\delta) = \xi_2 \) and \( \xi_2 e = 1 \). We assume \( \mu_v < \mu_b \). When the working vacation ends, if there are customers staying in the system, the server switches to regular busy period immediately. If there is no customers after the working vacation, the server will take a vacation, which follows a PH-distribution with \((\gamma, T_2)\) of order \( r_2 \). During the vacation period, the server stops the service completely. After the vacation, the server will switch to regular busy period if there are customers arriving during the vacation, otherwise, another vacation will go on. To describe this two-stage vacation policy clearly, we represent schematically this system in Fig. 1.

![Figure 1. Schematic representation](image)

We assume the system has a finite system capacity \( N \), any customers who meets a full capacity is rejected and lost. Further, the customers are served according to the arrival order i.e., FCFS. We assume the inter-arrival times, service times and working vacation are mutually independent.

3. Steady-state analysis.

3.1. The finite quasi-birth-and-death process. In a discrete-time queuing system, the arrivals, the departures and the ends of vacations may happen at the same time. We assume that the time axis is allotted into intervals of queue length with the length of a slot being unity, to be more specific, we can let the time axis be marked by \( 0, 1, 2, \ldots, t \). In this paper we analyze the model for early arrival system (EAS), therefore, the potential arrival occurs in \((t, t^+)\), where \( t^+ \) is the moment immediately after \( t \), and the potential departure takes place in \((t^-, t)\), where \( t^- \) represents the moment immediately before \( t \), in addition, the beginning and ending of the working vacations also take places at the instant \( t^+ \).

Let \( N_t \) be the number of customers in the system at time \( t^+ \), \( J_t \) represents the phase of arrival at time \( t^+ \), \( Q_t \) is the state of server at time \( t^+ \) (\( Q_t = 0 \) represents the server is in normal service period, \( Q_t = 1 \) represents the server is in the working vacation and \( Q_t = 2 \) represents the server is in the vacation). If \( Q_t = 0 \), \( K_t^b \) gives the phase of service in normal service period. If \( Q_t = 1 \), \( V_t^w \) is the phase of working vacation duration, and \( K_t^w \) give the phase of service in the working vacation. If \( Q_t = 2 \), \( V_t^v \) is the phase of vacation duration. Then we can know that \( \Delta = \{(N_t, (Q_t, J_t, V_t^w, K_t^w) \cup (Q_t, J_t, V_t^v) \cup (Q_t, J_t, K_t^b)) \mid t = 0, 1, 2, \ldots\} \), is a Markov chain with state space \( \Omega = \{(0) \times \{1, j, v_1\} \cup \{0\} \times \{2, j, v_2\} \cup \{k\} \times \{0, j, v_1, k_1\} \cup \{1, j, v_2\} \cup \{2, j, k_2\}\} \), where \( 1 \leq k \leq K \), \( 1 \leq j \leq n \), \( 1 \leq v_1 \leq r \), \( 1 \leq v_2 \leq r_2 \), \( 1 \leq k_1 \leq m_1 \), \( 1 \leq k_2 \leq m_2 \).
Using the lexicographical order of the states, we get the transition probability matrix $P$ of the finite quasi-birth-and-death (QBD) process as follows.

\[
P = \begin{bmatrix}
0 & B_{00} & B_{01} & B_{10} & A_1 & A_0 \\
1 & A_0 & A_1 & \cdots & \cdots & A_0 \\
2 & \cdots & \cdots & \cdots & A_1 & A_0 \\
\vdots & \ddots & \ddots & \ddots & \cdots & \cdots \\
N-1 & \cdots & \cdots & \cdots & A_1 & A_0 \\
N & A_0 & A_1 & \cdots & \cdots & A_0
\end{bmatrix}, \quad (1)
\]

where

\[
B_{00} = \begin{bmatrix}
D_0 \otimes T_1 & D_0 \otimes T_1^0 e^\gamma \\
0 & D_0 \otimes (T_2 + T_2^0 e^\gamma)
\end{bmatrix},
\]

\[
B_{01} = \begin{bmatrix}
D_1 \otimes T_1 & D_1 \otimes T_1^0 e^\gamma \\
0 & D_1 \otimes T_2^0 e^\gamma
\end{bmatrix},
\]

\[
B_{10} = \begin{bmatrix}
D_0 \otimes T_1 \otimes S_2^0 & D_0 \otimes T_1^0 \otimes S_2^0 e^\beta \\
0 & 0
\end{bmatrix},
\]

\[
A_0 = \begin{bmatrix}
D_0 \otimes T_1 \otimes S_2^0 & D_0 \otimes T_1^0 \otimes S_2^0 e^\beta \\
D_1 \otimes T_2 & D_1 \otimes T_2^0 e^\beta
\end{bmatrix},
\]

\[
A_1 = \begin{bmatrix}
D_0 \otimes T_1 \otimes S_2^0 & D_0 \otimes T_1^0 \otimes S_2^0 e^\beta \\
D_0 \otimes T_2 & D_0 \otimes T_2^0 e^\beta
\end{bmatrix},
\]

\[
A_2 = \begin{bmatrix}
D_0 \otimes T_1 \otimes S_2^0 & D_0 \otimes T_1^0 \otimes S_2^0 e^\beta \\
0 & 0
\end{bmatrix},
\]

\[
C = \begin{bmatrix}
D \otimes T_1 \otimes S_2^0 & D \otimes T_1^0 \otimes S_2^0 e^\beta \\
D \otimes T_2 & D \otimes T_2^0 e^\beta
\end{bmatrix}.
\]

From the structure of $P$, we observe that the Markov chain is a finite quasi birth and death process (QBD).

Define the traffic intensity

\[
\rho = \lambda / \mu_b. \quad (2)
\]

It is easy to verify that $\rho < 1$ makes the variant of (1) with infinite system capacity ($N = \infty$) positive recurrent. In fact, since $A_i$, $i = 0, 1, 2$, is upper triangular so that the matrix $A = A_0 + A_1 + A_2$ is a reducible and stochastic matrix. After some manipulations, we can obtain that

\[
A = \begin{bmatrix}
D \otimes T_1 \otimes (S_2 + S_2^0 e^\beta) & 0 \\
D \otimes T_2 & 0 \\
D \otimes T_2 & 0 \\
D \otimes (S_1 + S_2^0 e^\beta)
\end{bmatrix}.
\]

(3)

For simplicity and ease of notations, we write matrix $A$ and $A_k$, $k = 0, 1, 2$, as follows.

\[
A = \begin{bmatrix}
A_{ww} & 0 & A_{ws} \\
A_{uw} & A_{us} & A_{ss}
\end{bmatrix}, \quad A_k = \begin{bmatrix}
A_{kw} & 0 & A_{kw} \\
A_{ku} & A_{ku} & A_{ks} \\
A_{kw} & A_{kw} & A_{ks}
\end{bmatrix}, \quad k = 0, 1, 2.
\]

(4)
Based on theorem 7.3.1 in Latouche and Ramaswami [18], we know the corresponding infinite QBD process is positive recurrent if and only if
\[ \hat{\lambda} A^{ss} e > \hat{\mu} A \]
where \( \hat{\lambda} = \theta \otimes \xi_1 \) such that \( \hat{\lambda} A^{ss} = \hat{\lambda} \) and \( \hat{\mu} e = 1 \). From (5), we know
\[ (\theta \otimes \xi_1) (D_0 \otimes S_0^{0} \beta) e > (\theta \otimes \xi_1) (D_1 \otimes S_1) e, \]
then
\[ (\theta D e) \otimes (\xi_1 S_0^{0} \beta e) > (\theta D_1 e) \otimes (\xi_1 (S_1 + S_0^{0} \beta) e) , \]
which implies \( \mu_0 > \lambda \), then \( \rho < 1 \).

We should note that although the assumption \( \rho < 1 \) is not needed for the queue with finite buffer, there are still differences in the solution procedure between \( \rho < 1 \) and \( \rho > 1 \).

3.2. The rate matrices. Since the Markov chain is a finite QBD process, there are many approaches that can be used to analyze this finite QBD process, such as the methods developed in Naumov [26] and Gaver et al. [14]. Here, we analyze it by the matrix-geometric combination method (see Latouche and Ramaswami [18]). To this end, it is necessary to solve the rate matrices \( R_1 \) and \( R_2 \), where \( R_1 \) is minimum non-negative solution to the matrix equation
\[ A_0 + R A_1 + R^2 A_2 = R, \]
and \( R_2 \) is the minimum non-negative solution to the matrix equation
\[ A_2 + R A_1 + R^2 A_0 = R. \]
The computation of \( R_1 \) and \( R_2 \) can be carried out by a number of well-known algorithms (see Chapter 8 in Latouche and Ramaswami [17]). For example, a simple approach for computing \( R_1 \) is as follows. Let \( R_1^{(n)} \) be the value of \( R_1 \) at the \( n \)th iteration, where \( R_1^{(0)} = 0 \) and \( R_1^{(n+1)} = A_0 + R_1^{(n)} A_1 + \left( R_1^{(n)} \right)^2 A_2 \), this iteration procedure is continued until \( \| R_1^{(n+1)} - R_1^{(n)} \| < \varepsilon \), where \( \varepsilon \) is a very small positive number and \( |A| = \max_{1 \leq j \leq n} \left\{ \sum_{i=1}^{n} |a_{i,j}| \right\} \) for a matrix \( A = (a_{i,j})_{n \times n} \).

It is easy to find that computing the rate matrices \( R_1 \) and \( R_2 \) of order \( (nr_1 m_2 + nr_2 + nm_2) \) are usually the heaviest computation load in the process of computing the stationary probability vectors. In order to improve the computational efficiency, we apply a decomposition-based method for the rate matrices \( R_1 \) and \( R_2 \) that is proposed in Alfa [3]. Since \( R_1 \) and \( R_2 \) have the similar decomposition structures, here, we just give the decomposition results for \( R_1 \).

From the structures of \( A_i \), \( i = 0, 1, 2 \), we immediately see that \( R_1 \) satisfies the form of
\[ R_1 = \begin{bmatrix} R_{1w} & 0 & R_{1ws} \\ R_{1w} & R_{1s} & R_{1ws} \\ R_{1s} & R_{1ws} & R_{1ss} \end{bmatrix}, \]
and
\[ R_2 = \begin{bmatrix} (R_{1w})^2 & 0 & R_{1ws} R_{1ws} + R_{1ws} R_{1ss} \\ (R_{1w})^2 & (R_{1w})^2 & R_{1ws} R_{1s} + R_{1ws} R_{1s} \\ (R_{1s})^2 & (R_{1s})^2 & (R_{1ss})^2 \end{bmatrix} \]
Substituting (9) in (6) leads to the following equations.
\[ A_0^{ww} + R_1^{ww} A_1^{ww} + (R_1^{ww})^2 A_2^{ww} = R_1^{ww}, \]
\[ A_0^{ws} + R_1^{ws} A_1^{ws} + R_1^{ws} A_2^{ws} + (R_1^{ws})^2 A_2^{ws} = R_1^{ws}, \]
\[ A_0^{uv} + R_1^{uv} A_1^{uv} + (R_1^{uv})^2 A_2^{uv} = R_1^{uv}, \]
\[ A_0^{us} + R_1^{us} A_1^{us} + R_1^{us} A_2^{us} + (R_1^{us})^2 A_2^{us} = R_1^{us}, \]
\[ A_0^{ss} + R_1^{ss} A_1^{ss} + (R_1^{ss})^2 A_2^{ss} = R_1^{ss}. \]  

From equations (10)-(14), we can find that \( R_1^{ss} \) is the rate matrix associated with corresponding MAP/PH/1 queue without vacations, \( R_1^{ws} \) and \( R_1^{uv} \) correspond to the rate matrices during the working vacation period and vacation period, respectively. Moreover, by the iterative algorithm, \( R_1^{ws} \), \( R_1^{uv} \) and \( R_1^{ss} \) can be easily computable than computing the matrix \( R_1 \) since their dimensions are smaller than \( R_1 \), the remaining non-zero block matrices are \( R_1^{ws} \) that connects matrices \( R_1^{ws} \) and \( R_1^{us} \), and \( R_1^{ss} \) that connects matrices \( R_1^{ss} \) and \( R_1^{us} \). In order to calculate the \( R_1^{ws} \) and \( R_1^{us} \), we adopt the similar method in Alfa [3]. We first rewrite equations (11) and (13) as

\[ R_1^{ws} = U_2 + R_1^{ws} U_1 = U_0 R_1^{ws} A_2^{ws}, \]
\[ R_1^{us} = V_2 + R_1^{us} V_1 = V_0 R_1^{us} A_2^{us}, \]

where

\[ U_0 = R_1^{uw}, U_1 = A_1^{ws} + R_1^{ws} A_2^{ws}, U_2 = A_0^{ws} + R_1^{uw} A_1^{ws} + (R_1^{uw})^2 A_2^{ws}, \]

and

\[ V_0 = R_1^{uv}, V_1 = A_1^{us} + R_1^{us} A_2^{us}, V_2 = A_0^{us} + R_1^{uv} A_1^{us} + (R_1^{uv})^2 A_2^{us}. \]

Further, if we let

\[ \tilde{U}_1 = -(I - U_1)(A_2^{ws})^{-1}, \tilde{U}_2 = U_2 (A_2^{ws})^{-1}, \]

and

\[ \tilde{V}_1 = -(I - V_1)(A_2^{us})^{-1}, \tilde{V}_2 = V_2 (A_2^{us})^{-1}, \]

where \( I \) refers to an identity matrix, then, under the assumption that the inverse of \( A_2^{ws} \) exists, we have the following equations

\[ R_1^{ws} \tilde{U}_1 + U_0 R_1^{ws} = \tilde{U}_2, \]
\[ R_1^{us} \tilde{V}_1 + V_0 R_1^{us} = \tilde{V}_2, \]

(17) and (18) are Sylvester functions, numerous of algorithms can be used to solve above functions, such as Bartels-Stewarts algorithm (Bartel and Stewart [5]). It should be note that the inverse of \( A_2^{ws} \) is required to be existed when we apply Bartels-Stewarts algorithm to carried out some operations. In order to solve the problem, an alternative approach using vectorization idea proposed in Alfa [3] can be taken.

For an \( n \times m \) matrix \( B \), we define

\[ vecB = [B_{1,1}, ..., B_{n,1}, B_{1,2}, ..., B_{n,2}, ..., B_{1,m}, ..., B_{n,m}]^T \]

be termed the vectorized version of \( B \), where \( B_{i,j} \) is the \((i, j)\)th element of matrix \( B \), and \( T \) is reserved for transpose of a matrix. Note that if \( Z = WXY \), then \( vecZ = (Y^T \otimes W) vecX \) (Graham [16]), then, applying the vectorization method for equations (15) and (16), we can obtain

\[ vecR_1^{ws} = \left[ I - U_1^T \otimes I - (A_2^{ws})^T \otimes U_0 \right]^{-1} vecU_2, \]
\[ vecR_1^{us} = \left[ I - V_1^T \otimes I - (A_2^{us})^T \otimes V_0 \right]^{-1} vecV_2. \]
3.3. The stationary probability vectors. Let $\pi$ be the stationary probability vector associated with transition matrix $P$ such that

$$\pi P = \pi, \quad \pi e = 1. \quad (22)$$

We partition the vector $\pi$ as $\pi = [\pi_0, \pi_1, \pi_2, ... \pi_N]$, in which $\pi_0$ is a $(nr_1m_2+nr_2)$ dimensional vector representing the situation that there is no customer in the system, $\pi_k (k = 1, 2, ..., N)$ is a $(nr_1m_2+nr_2+nm_1)$ dimensional vector representing the situation that there are $k$ customers in the system. Based on the results in Akar et al.[4], we have that if $\rho \neq 1$, the expression of stationary probabilities can be given by

$$\pi_k = v_1 R_1^{k-1} + v_2 R_2^{N-k}, \quad 1 \leq k \leq N, \quad (23)$$

and $[\pi_0, v_1, v_2]$ is the left invariant eigenvector of $B[R_1, R_2]$ satisfying

$$[\pi_0, v_1, v_2] B[R_1, R_2] = [\pi_0, v_1, v_2], \quad (24)$$

normalized so that

$$\pi_0 e + (v_1 S_1' + v_2 S_2') e = 1, \quad (25)$$

where

$$B[R_1, R_2] = \begin{bmatrix} B_{00} & B_{01} & 0 \\ B_{10} & A_1 + R_1 A_2 & R_1^{N-2} (A_0 + R_1 e - R_1) \\ R_2^{N-1} B_{10} & R_2^{N-2} (A_2 + R_2 A_1 - R_2) & R_2 A_0 + C \end{bmatrix}, \quad (26)$$

and

$$S_1' = \sum_{k=0}^{N-1} R_1^k, \quad S_2' = \sum_{k=0}^{N-1} R_2^k. \quad (27)$$

In order to simplify the computation procedure for computing the stationary probability vectors of our model, we first follow the analysis in Akar et al.[4] to discuss some spectrum properties for $R_1$ and $R_2$ under the condition $\rho < 1$ and $\rho > 1$, respectively.

Based on the results in Gail et al.[12], If $\rho < 1$, it can be shown that $sp (R_1) < 1$ and $sp (R_2) = 1$, where $sp (R_1)$ and $sp (R_2)$ denote the spectrum of matrix $R_1$ and $R_2$, i.e., the largest modulus of all eigenvalues of $R_1$ and $R_2$. Therefore, we can know that $(I - R_1)$ is nonsingular, then the expression of $S_1'$ can be simplified as follows.

$$S_1' = \sum_{k=0}^{N-1} R_1^k \quad (I - R_1)^{-1}. \quad (28)$$

However, since one of eigenvalues of $R_2$ is one, so a simplification of form (28) for $S_2'$ is not possible. Let

$$y_1 = (A_2 - R_2 A_0) e, \quad (29)$$

and note that

$$R_2 y_1 = (R_2 A_2 - R_2^2 A_0) e$$
$$= (R_2 A_2 - R_2 + A_2 + R_2 A_1) e$$
$$= (R_2 (A_2 + A_1 - I) + A_2) e$$
$$= (A_2 - R_2 A_0) e$$
$$= y_1,$$
it shows that $y_1$ is a right invariant eigenvector of $R_2$. In addition, the left invariant vector $x$ of $A (A = A_0 + A_1 + A_2)$ is also a left invariant eigenvector of $R_2$, i.e., $x R_2 = x$, we define the following matrix

$$M_1 = \frac{y_1 x}{x y_1},$$

which is of rank one, then we can know that the matrix $(I - \hat{R}_2)$ is invertible, where $\hat{R}_2 = R_2 - M_1$. By induction, we can obtain

$$R_2^k = \begin{cases} \hat{R}_2^k + M, & k \geq 1, \\ I, & k = 0. \end{cases}$$

Then $S_2'$ can be explicitly expressed as follows.

$$S'_2 = \sum_{k=0}^{N-1} \hat{R}_2^k + (N - 1) M_1 = (I - \hat{R}_2) \left(I - \hat{R}_2\right)^{-1} + (N - 1) M_1. \quad (32)$$

If the traffic intensity $\rho > 1$, the form of the stationary probability vectors are still same as in (23), but the spectrum properties of the matrix geometric factors $R_1$ and $R_2$ are swapped, that is $sp (R_1) = 1$ and $sp (R_2) < 1$. Therefore, the expression (28) is not valid any more since $(I - R_1)$ is not invertible. Define $y_2$ as

$$y_2 = (A_0 - R_1 A_2) e,$$

and

$$M_2 = \frac{y_2 x}{x y_2},$$

where $x$ is the left invariant vector of matrix $A (A = A_0 + A_1 + A_2)$, it can show that $x$ and $y_2$ are the left and right invariant eigenvectors of $R_1$, respectively, and the following holds:

$$R_1^k = \begin{cases} \hat{R}_1^k + M_2, & k \geq 1, \\ I, & k = 0. \end{cases}$$

where $\hat{R}_1 = R_1 - M_2$. Furthermore, we can write $S_1'$ as follows:

$$S'_1 = \sum_{k=0}^{N-1} \hat{R}_1^k + (N - 1) M_2 = (I - \hat{R}_1) \left(I - \hat{R}_1\right)^{-1} + (N - 1) M_2 \quad (36)$$

On the other hand, since $R_2$ is invertible, we can know

$$S'_2 = \sum_{k=0}^{N-1} R_2^k = (I - R_2^N) (I - R_2)^{-1} \quad (37)$$

4. Loss probability. In this section, we consider the value of probability $P_{\text{loss}}$ that represents the loss probability due to the full capacity of system. If we partition the vector $\pi_N$ as $(\pi_{N,0}, \pi_{N,1}, \pi_{N,2})$, where $\pi_{N,0}$, $\pi_{N,1}$ and $\pi_{N,2}$ represent the probability vector that there are $N$ customers in the system and the server is in working vacation period, vacation period and regular busy period, respectively. Then, we can get the following proposition.

**Proposition 1.** If $\rho \neq 1$,

$$P_{\text{loss}} = \lambda^{-1} \pi_{N,0} (D_1 \otimes T_1 \otimes S_2) e + \lambda^{-1} \pi_{N,1} (D_1 \otimes T_2) e + \lambda^{-1} \pi_{N,2} (D_1 \otimes S_1) + \lambda^{-1} \pi_{N,0} (D_1 \otimes T_1^0 \otimes S_2) e + \lambda^{-1} \pi_{N,1} (D_1 \otimes T_2^0) e,$$

$$\quad (38)$$
and

$$\lim_{N \to \infty} P_{\text{loss}} = \max \left\{ 0, 1 - \frac{1}{\rho} \right\}.$$  \hfill (39)

**Proof.** Since the system capacity is $N$, when the customer finds $N$ customers staying in the system when it arrives, it will leave the system at its arrival without the service, thus, the loss probability $P_{\text{loss}} = \lambda^{-1} \pi_N A_0 e$ and we can obtain the expression of (38) directly. Intuitively, the first item of right-hand size of (38) is the probability that the new customer finds that the system in working vacation period and there are $N$ customers stayed in the system when it arrives, the second item is the probability that the new customer finds that the system is in vacation period and there are $N$ customers in the system when it arrives and the last item is the probability that the new customer finds that the system is in regular busy period and there are $N$ customers in the system. For the limit of $P_{\text{loss}}$, note that if $\rho < 1$, $\pi_N \to 0$, when $N \to \infty$, by $P_{\text{loss}} = \lambda^{-1} \pi_N A_0 e$, we can directly derive that $\lim_{N \to \infty} P_{\text{loss}} = 0$. On the other hand, applying Little’s law to the server, we can know

$$1 - P_{\text{loss}} = \frac{1 - \pi_0\epsilon}{\rho}$$  \hfill (40)

the right-hand side of (40) represents the ratio of the amount per time unit of processed workload $1 - \pi_0$ and that of offered workload $\rho$. If $\rho > 1$, we know that $\pi_0 \to 0$ when $N \to \infty$, from (40), $\lim_{N \to \infty} P_{\text{loss}} = 1 - 1/\rho$. \hfill $\square$

5. **Waiting time analysis.** In this section, we consider the waiting time of the successful customers, i.e. the waiting time of customers who are not lost due to the full system capacity. We define $z_i, (0 \leq i \leq N - 1)$ as the stationary vector corresponding to state of finding $i$ customers in the system by a successful customer when it arrives. The vector $z_0$ can be partitioned into $z_0 = [z_{0,0}, z_{0,1}]$, and $z_i$, for $1 \leq i \leq N - 1$, can be partitioned into $z_i = [z_{i,0}, z_{i,1}, z_{i,2}]$, where $z_{i,0}, 0 \leq i \leq N - 1$, represents the probability vector that a successful customer observe $i$ customers in the system and the system is in working vacation period at the arrival epoch; $z_{i,1}, 0 \leq i \leq N - 1$, represents the probability vector that a successful customer observe $i$ customers in the system and the system is in vacation at the arrival epoch; $z_{i,2}, 1 \leq i \leq N - 1$, represents the probability vector that a successful customer observe $i$ customers in the system and the system is in regular busy period at the arrival epoch. We have

$$z_{0,0} = \frac{\pi_{0,0} (D_1 \otimes T_1) + \pi_{1,0} (D_1 \otimes T_1 \otimes S_0^0) + \pi_{1,2} (D_1 \otimes S_1^0 \alpha)}{\lambda (1 - P_{\text{loss}})},$$  \hfill (41)

$$z_{0,1} = \frac{\pi_{01} (D_1 \otimes (T_2 + T_0 S_1^0 \gamma))}{\lambda (1 - P_{\text{loss}})},$$  \hfill (42)

for $1 \leq i \leq N - 1$

$$z_{i,0} = \frac{\pi_{i,0} (D_1 \otimes T_1 \otimes S_2) + \pi_{i+1,0} (D_1 \otimes T_1 \otimes S_0^0 \delta)}{\lambda (1 - P_{\text{loss}})},$$  \hfill (43)

$$z_{i,1} = \frac{\pi_{i,1} (D_1 \otimes T_2)}{\lambda (1 - P_{\text{loss}})},$$  \hfill (44)

$$z_{i,2} = \frac{\left\{ \pi_{i,2} (D_1 \otimes S_1) + \pi_{i+1,2} (D_1 \otimes S_0^0 \beta) + \pi_{i,0} (D_1 \otimes T_1^0 \otimes S_0^0 \delta) \right\} + \pi_{i+1,0} (D_1 \otimes T_0^0 \otimes S_2^0 \beta) + \pi_{i,1} (D_1 \otimes T_0^0 \beta)}{\lambda (1 - P_{\text{loss}})}. \hfill (45)$$
In order to analyze the waiting time, we define another transition probability matrix

\[
\hat{P} = \begin{bmatrix}
0 & 1 & 2 & \cdots & N-2 & N-1 \\
\end{bmatrix}
\begin{bmatrix}
\hat{B}_{00} & \hat{B}_{01} & \hat{B}_{02} & \cdots & \hat{B}_{0,N-2} & \hat{B}_{0,N-1} \\
\end{bmatrix}
\begin{bmatrix}
\hat{A}_1 & \hat{A}_2 & \cdots & \hat{A}_{N-2} & \hat{A}_{N-1} \\
\end{bmatrix},
\]

(46)

where

\[
\hat{B}_{00} = \begin{bmatrix}
T_1 & T_1^0 \gamma & 0 \\
0 & T_2 + T_2^0 \gamma & T_2^0 \gamma \otimes S_2^0 \\
\end{bmatrix},
\]

\[
\hat{B}_{01} = \begin{bmatrix}
T_1 \otimes S_2^0 & 0 & 0 \\
0 & S_2^0 \alpha & 0 \\
\end{bmatrix},
\]

\[
\hat{A}_1 = \begin{bmatrix}
T_1 \otimes S_2^0 & 0 & T_1^0 \otimes S_2^0 \beta \\
T_2 & T_2^0 \beta & S_1 \\
\end{bmatrix},
\]

\[
\hat{A}_2 = \begin{bmatrix}
T_1 \otimes S_2^0 \delta & 0 & T_1^0 \otimes S_2^0 \beta \\
0 & S_1^0 \beta & 0 \\
\end{bmatrix}.
\]

Define

\[
z^n = \left[ z^n_0, z^n_1, z^n_2, \ldots, z^n_{N-1} \right],
\]

where \( z^n_0 = [z^n_0, z^n_0, z^n_0, \ldots, z^n_{N-1}] \), in which \( z^n_0 = z_0,0 = \pi_0 \), \( z^n_1 = z_{1},0 = \pi_1 \), \( z^n_2 = z_{2},0 = \pi_2 \), and \( z^n_i = [z^n_{i,0}, z^n_{i,1}, z^n_{i,2}] \) for \( i = 1, 2, \ldots, N-1 \), in which, \( z^n_{i,0} = z_{i,0} \), \( z^n_{i,1} = z_{i,1} \), \( z^n_{i,2} = z_{i,2} \), where \( I (r) \) refers to an identity matrix of dimension \( r \), and \( e (n) \) refer to the column vector of ones of \( n \) dimension. Let

\[
z^{n+1} = z^n \hat{P}, \quad n \geq 0
\]

(47)

further partition \( z^n \) as \( z^n = [z^n_0, z^n_1, z^n_2, \ldots, z^n_{N-1}] \), let \( W_j \) be the probability that a successful customers waiting time is less than or equal to \( j \) units, then

\[
W_j = z^n_0 e, \quad j \geq 0
\]

(48)

and \( z^n_{0,0} \) can be calculated recursively as follows

\[
z^n_{0,0} = z_{0,0} \gamma_1 + z_{0,0} \gamma_2 \otimes S_2^0 + z_{0,1} \gamma_3 \otimes S_2^0 \alpha,
\]

\[
z^n_{0,1} = z_{0,0} \gamma_4 + z_{0,1} \gamma_5 \otimes S_2^0 \gamma + z_{0,2} \gamma_6 \otimes S_2^0 \gamma \otimes S_2^0,
\]

and for \( 1 \leq i \leq N-1 \)

\[
z^n_{i,0} = z^n_{i,0} \gamma_1 \gamma_7 \otimes S_2^0 + z^n_{i+1,0} \gamma_8 \gamma_7 \otimes S_2^0 \delta,
\]

\[
z^n_{i,1} = z^n_{i,1} \gamma_7 T_2,
\]

\[
z^n_{i,2} = z^n_{i,2} \gamma_7 T_2 \gamma + z^n_{i,1} \gamma_8 \gamma_7 \otimes S_2^0 \beta + z^n_{i,2} \gamma_9 \gamma_7 \otimes S_2^0 \delta + z^n_{i+1,0} \gamma_8 \gamma_7 \otimes S_2^0 \delta + z^n_{i+1,2} \gamma_7 \otimes S_2^0 \beta.
\]

6. Performance measures. The probabilities that the queue is empty and full are given by

\[
P_{\text{empty}} = \pi_0 e,
\]

(49)

\[
P_{\text{full}} = \pi_N e = (v_2 + v_1 R_1^{N-1}) e.
\]

(50)

In order to calculate the mean queue length, one can use the following identities

\[
\sum_{k=0}^{K} k B^k = (I - B^{K+1}) (I - B)^{-2} - (I + KB^{K+1}) (I - B)^{-1},
\]

(51)
and
\[
\sum_{k=1}^{K} kB^{k-1} = (I - B^K) (I - B)^{-2} - KB^K (I - B)^{-1},
\]  
(52)
which hold for an arbitrary matrix \( B \) when \( (I - B) \) is invertible, then, the mean queue length \( E(L) \) can be given from two aspects: \( \rho < 1 \) and \( \rho > 1 \).

i) If \( \rho < 1 \), according to the analysis in Section 3.3, we can know \( (I - R_1) \) is invertible, but \( (I - R_2) \) is not invertible, thus
\[
E(L) = \sum_{k=0}^{N} k \pi_k e = \sum_{k=1}^{N} k \left( v_1 R_1^{k-1} + v_2 R_2^{N-k} \right) e
\]
\[
= v_1 \sum_{k=1}^{N} k R_1^{k-1} e + v_2 \sum_{k=1}^{N} k R_2^{N-k} e
\]
\[
= v_1 \sum_{k=1}^{N} k R_1^{k-1} e + v_2 \sum_{k=1}^{N} k \left( N - n \right) \left( \hat{R}_2^n + M_1 \right) e
\]
\[
= v_1 \sum_{k=1}^{N} k R_1^{k-1} e + v_2 \sum_{k=1}^{N} \left( N \sum_{n=0}^{N-1} \hat{R}_2^n + (N - 1) M_1 - \sum_{n=0}^{N-1} n \hat{R}_2^n - \frac{N(N-1)}{2} M_1 \right) e
\]
\[
= v_1 \left( (I - \hat{R}_1^N) (I - R_1)^{-2} - N \hat{R}_1^N (I - R_1)^{-1} \right) e + v_2 \left( N \left( I - \hat{R}_2^N \right) (I - \hat{R}_2)^{-1} \right) e
\]
\[
+ N^2 M_1 - \left( I - \hat{R}_2^N \right) (I - \hat{R}_2)^{-2} + N \hat{R}_2^N \left( I - \hat{R}_2 \right)^{-1} - \frac{N(N-1)}{2} M_1 \right) e.
\]
(53)

2) If \( \rho > 1 \), note that \( (I - R_1) \) is not invertible, but \( (I - R_2) \) is invertible, we can obtain the mean queue length as follows.

\[
E(L) = \sum_{k=0}^{N} k \pi_k = \sum_{k=1}^{N} k \left( v_1 R_1^{k-1} + v_2 R_2^{N-k} \right) e
\]
\[
= v_1 \sum_{k=1}^{N} k R_1^{k-1} e + v_2 \sum_{k=1}^{N} k R_2^{N-k} e
\]
\[
= v_1 \sum_{k=1}^{N} k \left( \hat{R}_1^{k-1} + M_2 \right) e + v_2 \sum_{k=1}^{N} \left( N - n \right) \hat{R}_2^n e
\]
\[
= v_1 \left( (I - \hat{R}_1^N) (I - \hat{R}_1)^{-2} - N \hat{R}_1^N (I - \hat{R}_1)^{-1} + \frac{N(N+1)M_2}{2} \right) e
\]
\[
+ v_2 \left( N (I - \hat{R}_2^N) (I - \hat{R}_2)^{-1} - (I - \hat{R}_2^N) (I - \hat{R}_2)^{-2} + (I + (N - 1) \hat{R}_2^N) (I - \hat{R}_2)^{-1} \right) e.
\]
(54)

Partition \( \pi_0 \) into \([\pi_{0,0}, \pi_{0,1}]\) and \( \pi_i \) into \([\pi_{i,0}, \pi_{i,1}, \pi_{i,2}]\), for \( 1 \leq i \leq N \), where \( \pi_{i,j} (i = 0, 1, 2, ..., j = 0, 1, 2) \) represents the probability vector that there are \( i \) customers in the system with the server is in state \( j \) = 0 for the system on working vacation period and \( j = 1 \) for the system on vacation period and \( j = 2 \) for the system on normal service period. Then the state probabilities of the server in steady state are shown as follows.

The probability that the server is in working vacation is given by
\[
P_W = \sum_{i=0}^{N} \pi_{i,0} e.
\]
(55)

The probability that the server is in vacation is given by
\[
P_V = \sum_{i=0}^{N} \pi_{i,1} e.
\]
(56)
The probability that the server is in regular busy period is given by
\[ P_B = \sum_{i=1}^{N} \pi_{i,2} e. \]  
(57)

The mean waiting time can be obtained by Little law
\[ E(W) = E(N)/\lambda (1 - P_{\text{loss}}). \]  
(58)

7. Numerical examples. Through the aforementioned analysis, some system characteristics in steady state are derived, in this section, we aim to explore the relationships between system capacity and the key performance measures.

Example 1. The goal of the first example is to demonstrate the effect of system capacity \( N \) on various performance measures under the condition that the traffic intensity \( \rho < 1 \). The system parameters are defined as follows.

- The arrival is a discrete the Markov arrival process which has representation as
  \[ D_0 = \begin{bmatrix} 0.25 & 0.25 \\ 0.15 & 0.55 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0.3 & 0.2 \\ 0.1 & 0.2 \end{bmatrix}. \]

- The working vacation time follows a phase-type distribution \((\alpha, T_1)\), where
  \[ \alpha = [0.1, 0.9], \quad T = \begin{bmatrix} 0.2 & 0.6 \\ 0.7 & 0.1 \end{bmatrix}. \]

- The vacation time follows a phase-type distribution \((\gamma, T_2)\), where
  \[ \gamma = [0.2, 0.8], \quad T = \begin{bmatrix} 0.2 & 0.6 \\ 0.6 & 0.1 \end{bmatrix}. \]

- The service time during the normal service period follows a phase-type distribution \((\beta, S_1)\), where
  \[ \beta = [0.3, 0.7], \quad S_1 = \begin{bmatrix} 0.1 & 0.3 \\ 0.5 & 0.2 \end{bmatrix}. \]

- The service time during the working vacation period follows a phase-type distribution \((\delta, S_2)\), where
  \[ \delta = [0.5, 0.5], \quad S_2 = \begin{bmatrix} 0.6 & 0.1 \\ 0.3 & 0.2 \end{bmatrix}. \]

Based on the assumptions above, we can calculate the arrive rate \( \lambda = 0.3714 \) and the service rate in normal service period \( \mu_b = 0.4351 \), then the traffic intensity \( \rho = \lambda/\mu_b = 0.8536 < 1 \). For this queue, we plot the trends of \( E(L) \), \( P_{\text{empty}} \), \( P_{\text{full}} \), \( P(W) \), \( P(V) \) and \( P(B) \) with the change of \( N \) in Figs.2-7. From Fig.2, it is seen that mean queue length \( E(L) \) first significantly increases when \( N \) grows and, then, stabilizes at some constant which is obviously the corresponding the mean queue length of MAP/PH/1 queue with two stage vacations and infinite system capacity. As it is seen from Fig.3 and Fig.4, both \( P_{\text{empty}} \) and \( P_{\text{full}} \) decrease with the increase of system capacity \( N \), the difference is that \( P_{\text{empty}} \) converges to a non-zero constant, \( P_{\text{full}} \) converges to zero. In Figs.5-7, we compare the effects of system capacity \( N \) on the probabilities \( P(W) \), \( P(V) \) and \( P(B) \), evidently, we observe that \( P(W) \) and \( P(V) \) have the similar change trend with the increase of \( N \), that is, both decreases and converges to a non-zero constant when \( N \) goes to infinite, however, \( P(B) \) first increases and, then, stabilizes at some level. For the limit of \( P_{\text{loss}} \), from Fig. 8, we observe that \( P_{\text{loss}} \) decreases and approaches to zero, further, we plot the trend of \( \log_{10}(P_{\text{loss}}) \) with \( N \) varying in Fig. 9, obviously, we observe that \( \log_{10}(P_{\text{loss}}) \)
behaves in a straight line with $N$ varying, that is, $\log_{10}(P_{\text{loss}}) \sim \xi N$, where $\xi$ is a negative constant, therefore, we conjecture that $P_{\text{loss}}$ is asymptotically exponential, that is, $P_{\text{loss}} \sim we^{\xi N}$, here $w$ is a positive constant.
Example 2. In this example, we investigate the impact of the system capacity \( N \) on various performance measures under the condition that the traffic intensity \( \rho > 1 \). We assume that:

- The arrival is a discrete Markov arrival process which has representation as
  \[
  D_0 = \begin{bmatrix} 0.25 & 0.25 \\ 0.15 & 0.55 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0.3 & 0.2 \\ 0.1 & 0.2 \end{bmatrix}
  \]

- The working vacation time follows a phase-type distribution \((\alpha, T_1)\), where
  \[\alpha = [0.1, 0.9], \quad T = \begin{bmatrix} 0.2 & 0.6 \\ 0.7 & 0.1 \end{bmatrix}.\]

- The vacation time follows a phase-type distribution \((\gamma, T_2)\), where
  \[\gamma = [0.2, 0.8], \quad T = \begin{bmatrix} 0.2 & 0.6 \\ 0.6 & 0.1 \end{bmatrix}.\]

- The service time during the normal service period follows a phase-type distribution \((\beta, S_1)\), where
  \[\beta = [0.3, 0.7], \quad S_1 = \begin{bmatrix} 0.1 & 0.3 \\ 0.5 & 0.2 \end{bmatrix}.\]

- The service time during the working vacation period follows a phase-type distribution \((\delta, S_2)\), where
  \[\delta = [0.5, 0.5], \quad S_2 = \begin{bmatrix} 0.6 & 0.1 \\ 0.3 & 0.2 \end{bmatrix}.\]

According to the above assumptions, we can calculate the arrive rate \( \lambda = 0.6286 \) and the service rate in normal service period \( \mu_b = 0.4351 \), then the traffic intensity \( \rho = \lambda / \mu_b = 1.4446 > 1 \). For this queue, similar with example 1, we plot the trends of \( E(L), P_{\text{empty}}, P_{\text{full}}, P(W), P(V) \) and \( P(B) \) with the change of \( N \) in Figs.10-15. As is seen that in Fig.10, the mean queue length increases linearly as \( N \) increases, which agrees with the intuitive expectation, in fact, \( \rho > 1 \) indicates the corresponding MAP/PH/1/queue is not positive recurrent, then the \( E(L) \) approaches to infinite as \( N \) increases can be expected. Further, from Figs. 11-13, we can observe that probabilities \( P(W) \) and \( P(V) \) approach to zero and \( P(B) \) approaches to one as \( N \) increases, thus, we can conclude that, under the condition that \( \rho > 1 \), the server will always stay in the regular busy period if the system capacity \( N \) is large enough.

The behaviors of loss probability \( P_{\text{loss}} \) are depicted in Fig. 16 and Fig. 17, it is expectable that decreases as \( N \) grows, but, stabilizes at some non-zero constant, to explore the exact value of the limit of \( P_{\text{loss}} \), we compare the loss probabilities under different \( N \) and in Table 1, as it seen from Table 1, when \( \rho < 1 \), the loss probability \( P_{\text{loss}} \) tends to zero as \( N \) goes to infinite, however, when \( \rho > 1 \), we can find that loss probability \( P_{\text{loss}} \) tends to \( 1 - 1/\rho \).

Example 3. In this example, we compare behaviours between the MAP/PH/1/N queue with two-stage vacations and the corresponding the MAP/PH/1/N queue with working vacations. With the same system parameters, we plot the mean queue length \( E(L) \) with the increase of system capacity \( N \) for the conditions \( \rho < 1 \) and \( \rho > 1 \) in the Figure 18 and Figure 19, respectively. From Figure 18, we can observe that mean queue length in the queue with two-stage vacations is larger than the corresponding ordinary queue with working vacations under the condition that \( \rho < 1 \), which is consistent with our intuition. However, from Figure 19, we can readily find that the mean queue length \( E(L) \) in the queue with two-stage vacations
Figure 10. $E(L)$ versus $N$

Figure 11. $P_{\text{empty}}$ versus $N$

Figure 12. $P_{\text{full}}$ versus $N$

Figure 13. $P_W$ versus $N$

Figure 14. $P_V$ versus $N$

Figure 15. $P_B$ versus $N$

Figure 16. $P_{\text{loss}}$ versus $N$

Figure 17. $\log_{10}(P_{\text{loss}})$ versus $N$
is identical to the ordinary queue with working vacations under the condition that \( \rho > 1 \). To explain this phenomenon, we note that the traffic intensity \( \rho > 1 \) means that arrival rate \( \lambda \) is greater than service rate \( \mu_b \), then, most of time, the system runs at the full capacity. Thus, we believe that the stationary mean queue length is close to the system capacity \( N \) under the condition that \( \rho > 1 \). Therefore, the trends in Figure 19 can be expected.

![Figure 18. \( \rho < 1 \)](image1)

![Figure 19. \( \rho > 1 \)](image2)

8. Conclusion. In this paper, we have analyzed the MAP/PH/1 queue with two-stage vacations policy and finite system capacity and have done works in several aspects: By matrix-geometric combination method, we obtained the stationary probability vectors and the spectrum properties and decomposition results on the rate matrix \( R_1 \) and \( R_2 \) are discussed. The loss probability and its limit when the system capacity goes to infinite are derived. By constructing absorbing Markov chain, we analyzed the waiting time. Various performance measures are derived, and some numerical examples are presented to explore the relationship between the system capacity \( N \) and various performance measures.

It should be noted that we didn’t discuss the situation when \( \rho = 1 \), which is complex and cannot be solved by the matrix-geometric combination method, so we will discuss this situation in the future works.

Acknowledgments. The research was supported by Natural Science Foundation of Jiangsu Province (grant BK20180783, 18KJB110021), and The Startup Foundation for Introducing Talent of Nanjing University of Information Science and Technology (grant 2017r082).

REFERENCES

[1] A. Alfa, A discrete MAP/PH/1 queue with vacations and exhaustive time-limited service, *Operations Research Letters*, 18 (1995), 31–40.
[2] A. Alfa, Discrete time analysis of MAP/PH/1 vacation queue with gated time service, *Queueing Systems*, 29 (1998), 35–54.
[3] A. Alfa, Some decomposition results for a class of vacation queue, *Operations Research Letters*, 42 (2014), 140–144.
[4] N. Akar, N. Oguz and K. Sohraby, A Novel Computational Method for Solving Finite QBD Processes, *Stochastic Models*, 16 (2000), 273–311.
[5] Y. Baba, Analysis of a GI/M/1 queue with multiple working vacations, *Operations Research Letters*, 33 (2005), 201–209.
[6] Y. Baba, The M/PH/1 queue with working vacations and vacation interruption, *Journal of Systems Science and Systems Engineering*, 19 (2010), 496–503.
[7] A. Banik, Stationary analysis of a BMAP/R/1 queue with R-type multiple working vacations, Communications in Statistics-Simulation and Computation, 46 (2015), 1035–1061.
[8] R. Bartel and G. Stewart, Solution of the equation AX+XB=C, Communications of the ACM, 15 (1972), 820–826.
[9] V. Chandrasekaran, V. Indhira, M. Saravanarajan and P. Rajadurai, A survey on working vacation queueing models, International Journal of Pure and Applied Mathematics, 106 (2016), 33–41.
[10] S. Chakravarthy and S. Ozkar, MAP/PH/1 queueing model with working vacation and crowdsourcing, Mathematica Applicanda, 44 (2016), 263–294.
[11] B. Doshi, Queueing systems with vacations–A survey, Mathematica Applicanda, 1 (1986), 29–66.
[12] H. Gail, S. Hantler and B. Taylor, Solutions of the basic matrix equation for M/G/1 and G/M/1 type Markov chain, Stochastic Models, 10 (1994), 1–43.
[13] S. Gao, J. Wang and W. Li, An M/G/1 retrial queue with general retrial times, working vacations and vacation interruption, Asia-Pacific Journal of Operational Research, 31 (2014), 1440006.
[14] D. Gaver, P. Jacobs and G. Latouche, Finite birth-and-death models in randomly changing environments, Advances in applied probability, 16 (1984), 715–731.
[15] C. Goswami and N. Selvaraju, The discrete-time MAP/PH/1 queue with multiple working vacations, Applied Mathematical Modelling, 34 (2010), 931–946.
[16] A. Graham, Kronecker Products and Matrix Calculus: With Applications, John-Wiley, New York, 1981.
[17] J. Kim, D. Choi and K. Chae, Analysis of Queue-length Distribution of the M/G/1 Queue with Working Vacation, Hawaii International Conference on Statistics and Related Fields, 2003.
[18] J. Li, N. Tian and W. Liu, Discrete-time GI/Geom/1 queue with multiple working vacations, Queueing systems, 56 (2007), 53–63.
[19] J. Li and N. Tian, The discrete-time GI/Geom/1 queue with working vacations and vacation interruption, Applied Mathematics and Computation, 185 (2007), 1–10.
[20] J. Li and N. Tian, Performance analysis of a G/M/1 Queue with single working vacation, Applied Mathematics and Computation, 217 (2011), 4960–4971.
[21] J. Li, N. Tian, Z. Zhang and H. Luh, Analysis of the M/G/1 Queue with exponentially working vacations–a matrix analytic approach, Queueing Systems, 61 (2009), 139–166.
[22] J. Li, W. Liu and N. Tian, Steady-state analysis of a discrete-time batch arrival queue with working vacations, Performance Evaluation, 67 (2010), 897–912.
[23] T. Li, Z. Liu and Z. Wang, M/M/1 retrial queue with collisions and working vacation interruption under N-policy, RAIRO-Operations Research, 46 (2012), 355–371.
[24] C. Luo, W. Li, K. Yu and C. Ding, The matrix-form solution for GeoK /G/1/N working vacation queue and its application to state-dependent cost control, Computers and Operations Research, 67 (2016), 63–74.
[25] V. Naumov, Matrix-multiplicative approach to quasi-birth-death process analysis, in Matrix-Analytic Methods in Stochastic Models, Marcel Dekker, New York, 1996, 87–106.
[26] M. Neuts, Matrix-geometric Solution in Stochastic Model, Johns Hopkins University Press, Baltimore, MD, 1981.
[27] M. Neuts, Structured Stochastic Matrices of M/G/1 Type and Their Applications, Marcel Dekker, New York, 1989.
[28] L. Servi and S. Finn, M/M/1 queues with working vacation (M/M/1/WV), Performance Evaluation, 50 (2002), 41–52.
[29] C. Sreenivasan, S. Chakravarthy and A. Krishnamoorthy, MAP / PH /1 queue with working vacations, vacation interruptions and N policy, Applied Mathematical Modelling, 37 (2013), 3879–3893.
[30] N. Tian and Z. Zhang, Vacation Queueing Models-Theory and Application, Springer-Verlag, New York, 2006.
[31] N. Tian, Z. Ma and M. Liu, The discrete time Geom/Geom/1 queue with multiple working vacations, Applied Mathematical Modelling, 32 (2008), 2941–2953.
[32] N. Tian, J. Li and Z. Zhang, Matrix-analytic method and working vacation queues-survey, International Journal of Information and Management Sciences, 20 (2009), 603–633.
[34] A. Vadivu and R. Arumuganathan, Analysis of MAP/G/1/N queue with two phases of service under single (multiple) vacation(s), International Journal of Operational Research, 25 (2016), 47–76.

[35] D. Wu and H. Takagi, M/G/1 queue with multiple working vacations, Performance Evaluation, 63 (2006), 654–681.

[36] D. Yang and D. Wu, Cost-minimization analysis of a working vacation queue with N-policy and server breakdowns, Computers and Industrial Engineering, 82 (2015), 151–158.

[37] Q. Ye and L. Liu, The analysis of M/M/1 queue with two vacation policies (M/M/1/SWV+MV), International Journal of Computer Mathematics, 94 (2017), 115–134.

[38] Q. Ye and L Liu, Performance Analysis of the GI/M/1 Queue with Single Working Vacation and Vacations, Methodology and Computing in Applied Probability, 19 (2017), 685–714.

[39] Q. Ye and L Liu, The analysis of discrete time Geom/Geom/1 queue with single working vacation and multiple vacations (Geom/Geom/1/SWV+MV), RAIRO-Operations Research, 52 (2018), 95–117.

[40] Q. Ye, The analysis of $M^X$/M/1 queue with two-stage vacations policy, Communications in Statistics-Theory and Methods, 2018.

Received November 2018; 1st revision February 2019; 2nd revision February 2019.

E-mail address: yeqingzero@gmail.com