Queue length analysis of a Markov-modulated vacation queue with dependent arrival and service processes and exhaustive service policy[☆]

Gábor Horváth[a,b], Zsolt Saffer[a], and Miklós Telek[b]

[a]*Budapest University of Technology and Economics, Dept. of Networked Systems and Services, Magyar Tudósok Körútja 2, 1117 Budapest, Hungary*
[b]*MTA-BME Information systems research group, Magyar Tudósok Körútja 2, 1117 Budapest, Hungary*

Abstract

The paper introduces a class of vacation queues where the arrival and service processes are modulated by the same Markov process, hence they can be dependent. The main result of the paper is the probability generating function for the number of jobs in the system. The analysis follows a matrix-analytic approach. A step of the analysis requires the evaluation of the busy period of a quasi birth death process with arbitrary initial level. This element can be useful in the analysis of other queueing models as well. We also discuss several special cases of the general model. We show that these special settings lead to simplification of the solution.

Keywords: Vacation queue, MAP dependent arrival and service process, QBD, matrix analytic methods, stationary analysis

1. Introduction

The importance of vacation queues comes from their diverse application fields: modeling various computer systems, telecommunication protocols, manufacturing, logistics, etc. A large number of vacation queue variants have been analyzed in the literature, with different combinations of arrival processes, service time distributions, vacation time distributions and service policies, see [8] and [17] for recent surveys. The M/G/1 vacation model has already been extensively studied in the 1980's. For details on analysis works on classical vacation models the reader is referred to the survey of Doshi [6] or the book of Takagi [16].

Later on another vacation queue variants have been analyzed in the literature, like the GI/M/1 vacation queue in the work of Baba [3] or the discrete-time GI/Geom/1 vacation queue in the paper of J. H. Li et al. [10].

For more suitable modeling of manufacturing systems, real-life scenarios and specific modern telecommunication networks the basic vacation models have been generalized in several directions, like multi-server vacation models, working vacation

---

models and application specific extended vacation models. For results on such models the reader is referred to [18], [14] and to surveys [8] and [17].

Due to the versatility of the Markovian Arrival Process (MAP) [11], vacation queues with MAP input and general service times have also been investigated in several past papers [5, 12, 13]. As a special variant of the MAP/G/1 vacation model, the MAP/PH/1 vacation model have been analyzed by A. S. Alpha for exhaustive time-limited and gated time-limited service disciplines in [1, 2] and by C. Goswami and N. Selvaraju in [4]. All of these models assume independent service times. Up to the best knowledge of the authors the MAP/MAP/1 vacation queue, the vacation model where both the arrival and service processes are MAPs (hence, raising the opportunity of considering correlated inter-arrival and service times), has not been considered yet in the literature.

In this paper we consider a more general class of exhaustive vacation queues with dependent Markov modulated arrival and service processes (which we shortly refer to as *Markov modulated vacation queue*). The analysis of this model class requires the introduction of a new analysis element, the evaluation of the busy period of quasi birth death (QBD) processes with arbitrary initial level. As the main result of the paper we provide a computationally efficient expression of the probability-generating function (PGF) of the number of jobs in the general system. Additionally, we investigate the simplifications applicable for special model variants (e.g., for MAP/MAP/1 vacation model). In the last part of the paper we provide numerical examples and investigate the effects of different vacation distributions on the mean number of jobs.

The rest of the paper is organized as follows. In section 2 we introduce the general exhaustive vacation model with dependent Markov modulated arrival and service processes and discuss several special model variants. The analysis part of the work is presented in section 3. After a step-by-step establishment of the relations for characterizing the transient behavior of the model we derive the expression of the PGF of the number of jobs in the system. In section 4 the analysis of some special cases of the general model are provided. Numerical examples close the paper in section 5.

2. Model description

We consider the *Markov modulated vacation queue*. This model falls in the class of single server FCFS queues with multiple vacations, exhaustive discipline [17] and dependent Markov modulated arrival and service processes. According to the rule of the exhaustive service discipline the server serves the jobs in the queue until it gets idle, then the server leaves for vacation for an independent and identically distributed random amount of time. If the queue is idle at the end of the vacation the server leaves for a new vacation, otherwise it starts serving the jobs in the queue. The random vacation time, its probability density function (pdf) and its Laplace transform (LT) are denoted by $\tilde{\sigma}, \sigma(t)$ and $\sigma^*(s) = E(e^{-s\tilde{\sigma}})$, respectively.

The arrivals and services are characterized by seven matrices: $\mathbf{L}_v$, $\mathbf{F}_v$, $\mathbf{B}_s$, $\mathbf{L}_s$, $\mathbf{F}_s$, $\mathbf{\Pi}_{vs}$ and $\mathbf{\Pi}_{sv}$. Their interpretations are as follows.

- During the vacations the arrivals are given by a MAP, where the entries of $\mathbf{L}_v$ are the rates of transitions without a job arrival, and the entries of $\mathbf{F}_v$ are the rates of transitions that are accompanied by a job arrival. Matrix $\mathbf{L}_v + \mathbf{F}_v$ is

therefore the generator of the continuous time Markov chain (CTMC) with N_v states which modulates the arrivals during the vacation.

- When the server is back from the vacation and serves jobs, the queue behaves like a quasi birth-death (QBD) [9] process, where the matrices $\mathbf{B}_s, \mathbf{L}_s$ and $\mathbf{F}_s$ contain the transition rates associated with a service completion, without service completion and job arrival, with a job arrival, respectively. In this case the generator of the modulating CTMC is $\mathbf{B}_s + \mathbf{L}_s + \mathbf{F}_s$, and it has N_s states.

- The transition probabilities between the vacation and service periods are given by $N_v \times N_s$ stochastic matrix $\mathbf{\Pi}_{vs}$, whose entries are the probabilities of the state transitions occurring at the end of the vacation period. The probabilities of (extra) transitions (additional to those of $\mathbf{B}_s$) between the service and vacation periods are given by the stochastic matrix $\mathbf{\Pi}_{sv}$ of size $N_s \times N_v$.

We impose the following assumptions on the model:

A.1 The phase process of the QBD characterized by matrix $\mathbf{B}_s + \mathbf{L}_s + \mathbf{F}_s$ and the phase process of the MAP characterized by matrix $\mathbf{L}_v + \mathbf{F}_v$ are irreducible.

A.2 The rates of transitions in the QBD and in the MAP during the vacation are finite as well as the mean vacation time is positive and finite.

A.3 The arrival process during the vacation and the sequence of the vacation periods are mutually independent.

A.4 The model is stable.

The QBD describing the service period is stable if and only if the stationary drift of its level process is negative [9]. Hence the necessary and sufficient condition of the stability of this vacation model is

$$\alpha_s \mathbf{F}_s \mathbb{1} - \alpha_s \mathbf{B}_s \mathbb{1} < 0, \tag{1}$$

where $\mathbb{1}$ is the column vector of ones and α_s is the stationary distribution of the phase process of the QBD and can be obtained as the unique solution of the system of linear equations $\alpha_s (\mathbf{B}_s + \mathbf{L}_s + \mathbf{F}_s) = 0$ and $\alpha_s \mathbb{1} = 1$.

During the vacation period the queue length is monotone increasing, but due to the fact that the mean vacation time is finite the mean number of customers arrived during the vacation period is finite as well.

If the QBD describing the service period is positive recurrent the customers which arrived during the vacation period are served during the consecutive service period in finite time and this way the overall vacation model remains positive recurrent if the service period is positive recurrent.

2.1. Special model variants

The introduced *Markov modulated vacation queue* is a general model which covers a number of special cases. Some of those special cases are already discussed in the literature, e.g., MAP/PH/1 vacation queue and some have not been considered before, e.g., the QBD and the MAP/MAP/1 vacation queues. Below we list some special cases and define the corresponding model parameters.

- *Markov modulated vacation queue without special phase change*: In this case no additional phase transitions occur at the start and at the end of the vacation periods, i.e., $\mathbf{\Pi}_{vs} = \mathbf{\Pi}_{sv} = \mathbf{I}$. As a consequence the matrices $\mathbf{F}_v$, $\mathbf{L}_v$, $\mathbf{F}_s$, $\mathbf{L}_s$ and $\mathbf{B}_s$ have the same size, $N_s = N_v$.

- *QBD vacation queue*: A special case of the *Markov modulated vacation queue* is the QBD queue with vacations, which is similar to an ordinary QBD queue, but the transitions representing a service event in the service period are not accompanied by a service event in the vacation periods. The matrices characterizing the system are $\mathbf{F}_v = \mathbf{F}_s = \mathbf{F}$, $\mathbf{L}_s = \mathbf{L}$, $\mathbf{B}_s = \mathbf{B}$, $\mathbf{L}_v = \mathbf{B} + \mathbf{L}$ and $\mathbf{\Pi}_{vs} = \mathbf{\Pi}_{sv} = \mathbf{I}$. Observe that the service times and the arrival times can be (cross) dependent in this case.

- *QBD vacation queue with independent arrival and service processes*: A special case of the *QBD vacation queue* is when two independent Markov chains modulate the arrival and the service processes. In this case the individual processes, i.e. the arrival process and the service process can be correlated, but the arrival and the service processes are independent. Let the arrivals be generated by a MAP with matrices $(\mathbf{D}_0, \mathbf{D}_1)$ and the service times be generated by a MAP with matrices $(\mathbf{S}_0, \mathbf{S}_1)$. In this case the characterizing matrices are

$$\mathbf{F}_v = \mathbf{F}_s = \mathbf{D}_1 \otimes \mathbf{I}, \mathbf{L}_v = \mathbf{D}_0 \oplus (\mathbf{S}_0 + \mathbf{S}_1), \mathbf{L}_s = \mathbf{D}_0 \oplus \mathbf{S}_0,$$
$$\mathbf{B}_s = \mathbf{I} \otimes \mathbf{S}_1, \mathbf{\Pi}_{vs} = \mathbf{\Pi}_{sv} = \mathbf{I}.$$

- *MAP/MAP/1 vacation queue*: This model is very similar to the previous one. A MAP with matrices $(\mathbf{D}_0, \mathbf{D}_1)$ generates the arrivals and a MAP with matrices $(\mathbf{S}_0, \mathbf{S}_1)$ generates the service times. The only difference compared to the previous model is the behavior during the vacation time: in the *MAP/MAP/1 vacation queue* the phase of the service process is frozen during the vacation. Consequently, the matrices characterizing the queue are

$$\mathbf{F}_v = \mathbf{F}_s = \mathbf{D}_1 \otimes \mathbf{I}, \mathbf{L}_v = \mathbf{D}_0 \otimes \mathbf{I}, \mathbf{L}_s = \mathbf{D}_0 \oplus \mathbf{S}_0,$$
$$\mathbf{B}_s = \mathbf{I} \otimes \mathbf{S}_1, \mathbf{\Pi}_{vs} = \mathbf{\Pi}_{sv} = \mathbf{I}.$$

  Similar to the previous case, both the inter-arrival and service times can be correlated, but the arrival and service processes do not dependent on each other. In contrast with the *QBD vacation queue with independent arrival and service processes* the *MAP/MAP/1 vacation queue* is not a special case of the *QBD vacation queue*, but it is a special case of the *Markov modulated vacation queue without special phase change*.

- *MAP/PH/1 vacation queue*: When the arrivals are generated by a MAP with matrices $(\mathbf{D}_0, \mathbf{D}_1)$ and the service time is phase type distributed with with initial vector α and generator matrix $\mathbf{A}$, the characterizing matrices are

$$\mathbf{F}_v = \mathbf{F}_s = \mathbf{D}_1 \otimes \mathbf{I}, \ \mathbf{L}_v = \mathbf{D}_0 \otimes \mathbf{I}, \ \mathbf{L}_s = \mathbf{D}_0 \oplus \mathbf{A},$$
$$\mathbf{B}_s = -\mathbf{I} \otimes \mathbf{A}\mathbb{1}\alpha, \ \mathbf{\Pi}_{vs} = \mathbf{\Pi}_{sv} = \mathbf{I}.$$

  In this case the service times are independent and identically distributed.

The subset relation of the listed special model sets is depicted in Figure 1. Apart of these standard cases the flexibility of the general model is exemplified by the following special models.
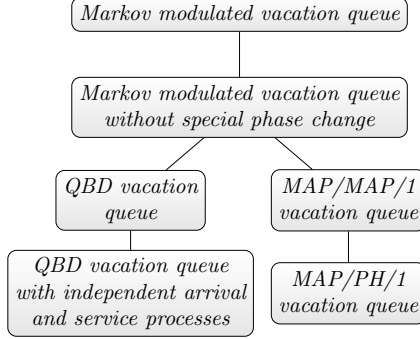
Figure 1: Subset relations of the considered special vacation queue models

- *M/PH/1 queue in a random environment*: Assume that the environment is given by a two-state CTMC with transition rates ν_1, ν_2. In state 1 the arrival rate is λ_1, and the service times are phase-type (PH) distributed with initial vector α and generator matrix $c_1\mathbf{A}$. In state 2 the arrival rate is λ_2 and the service times are given by $(\alpha, c_2\mathbf{A})$. The matrices characterizing the vacation queue are

$$\mathbf{F}_s = \begin{bmatrix} \lambda_1\mathbf{I} & \\ & \lambda_2\mathbf{I} \end{bmatrix}, \quad \mathbf{L}_s = \begin{bmatrix} c_1\mathbf{A} - \lambda_1 - \nu_1 & \nu_1\mathbf{I} \\ \nu_2\mathbf{I} & c_2\mathbf{A} - \nu_2 - \lambda_2 \end{bmatrix},$$

$$\mathbf{B}_s = \begin{bmatrix} -c_1\mathbf{A}\mathbb{1}\alpha & \\ & -c_2\mathbf{A}\mathbb{1}\alpha \end{bmatrix}, \quad \mathbf{F}_v = \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix},$$

$$\mathbf{L}_v = \begin{bmatrix} -\nu_1 - \lambda_1 & \nu_1 \\ \nu_2 & -\nu_2 - \lambda_2 \end{bmatrix}, \quad \mathbf{\Pi}_{vs} = \begin{bmatrix} \alpha & \\ & \alpha \end{bmatrix}, \quad \mathbf{\Pi}_{sv} = \begin{bmatrix} \mathbb{1} & \\ & \mathbb{1} \end{bmatrix},$$

where $\mathbb{1}$ is the column vector of ones of compatible size.

- *MAP/MAP/1 vacation queue with discouraged customers*: In this vacation queue the arrival process is different in the vacation and the service periods, due to the fact that discouraged customers do not attend the system during the vacation period. If the ratio of discouraged customers is ϱ then the characterizing matrices are

$$\mathbf{F}_s = \mathbf{D}_1 \otimes \mathbf{I}, \mathbf{F}_v = (1 - \varrho)\mathbf{D}_1 \otimes \mathbf{I}, \mathbf{L}_v = (\mathbf{D}_0 + \varrho\mathbf{D}_1) \otimes \mathbf{I}, \mathbf{L}_s = \mathbf{D}_0 \oplus \mathbf{S}_0,$$

$$\mathbf{B}_s = \mathbf{I} \otimes \mathbf{S}_1, \mathbf{\Pi}_{vs} = \mathbf{\Pi}_{sv} = \mathbf{I}.$$

The following section provides the analysis of the general vacation model. Some of the special cases are considered in section 4.

3. The number of jobs in the system

To characterize the number of jobs in the system, let us introduce the two dimensional process $\mathcal{X}(t) = \{\mathcal{N}(t), \mathcal{J}(t), t \geq 0\}$, where $\mathcal{N}(t)$ denotes the number of jobs (also referred to as *levels*) and $\mathcal{J}(t)$ denotes the state of the modulating CTMC (also referred to as *phase*) at time t. For the analysis of $\mathcal{X}(t)$ the evolution of the queue is divided to *cycles*, as shown in Figure 2. Each cycle starts with a vacation
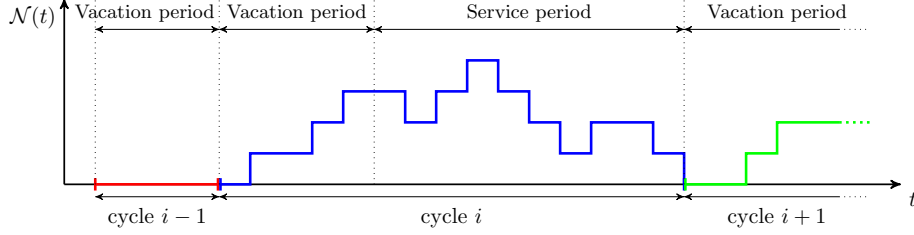
Figure 2: Cycles in the evolution of the queue

period, which is followed by a service period, and the cycle ends when the last job leaves the system. Note that a cycle can also be degenerate, if no jobs arrive during the vacation period, there is no service period (see cycle $i-1$ in Figure 2).

The stationary probability that there are ℓ jobs in the system is closely related to M_ℓ, the mean time spent at level ℓ in a stationary cycle.

$$M_\ell = \underbrace{\int_{u=0}^\infty \sigma(u)\,\beta\mathbf{\Pi}_{sv}\!\int_{t=0}^u \mathbf{P}_\ell^{(\mathbf{v})}(t)\mathbb{1}\,dt\,du}_{M_\ell^{(v)}} + \underbrace{\int_{u=0}^\infty \sigma(u)\sum_{m=1}^\infty \beta\mathbf{\Pi}_{sv}\,\mathbf{P}_m^{(\mathbf{v})}(u)\mathbf{\Pi}_{vs}\mathbf{H}_{m,\ell}\mathbb{1}\,du}_{M_\ell^{(s)}}, \quad (2)$$

where row vector β of size N_s is the stationary phase distribution at the end of a service period (hence, $\beta\mathbf{\Pi}_{sv}$ is the one at the beginning of a vacation period), matrix $\mathbf{P}_\ell^{(\mathbf{v})}(t)$ characterizes the number of arrivals up to time t during the vacation period, defined as

$$[\mathbf{P}_\ell^{(\mathbf{v})}(t)]_{i,j} = P(\mathcal{N}(t)=\ell, \mathcal{J}(t)=j | \mathcal{N}(0)=0, \mathcal{J}(0)=i),$$

and $[\mathbf{H}_{m,\ell}]_{i,j}$ is the mean times spent in level ℓ and phase j in the service period starting from level m and phase i. The first term of (2), $M_\ell^{(v)}$, corresponds to the vacation period, and the second term, $M_\ell^{(s)}$, to the service period. From M_ℓ the stationary distribution of $\mathcal{N}(t)$ is obtained by normalization, $q_\ell = \lim_{t\to\infty} P(\mathcal{N}(t)=\ell) = M_\ell / \sum_k M_k$.

Over the next subsections closed form formulas are provided for all ingredients of the solution, thus

- for the PGF of $\mathbf{P}_\ell^{(\mathbf{v})}(t)$ in terms of ℓ in Section 3.1,

- for β in Section 3.2,

- for $\mathbf{H}_{m,\ell}$ in Section 3.3,

- for the PGFs of $M_\ell^{(v)}$ and $M_\ell^{(s)}$ in terms of ℓ in Section 3.4,

finally, the pieces are put together in Section 3.5.

3.1. The evolution of the number of jobs during the vacation period

The evolution of the number of jobs during the vacation period resembles to the counting process of a MAP given by matrices $\mathbf{L}_v, \mathbf{F}_v$. Thus, for matrices $\mathbf{P}_\ell^{(\mathbf{v})}(t)$ we

have

$$\frac{d}{dt}\mathbf{P}_\ell^{(\mathbf{v})}(t) = \mathbf{P}_{\ell-1}^{(\mathbf{v})}(t)\mathbf{F}_v + \mathbf{P}_\ell^{(\mathbf{v})}(t)\mathbf{L}_v, \quad \text{for } \ell > 0, \tag{3}$$

$$\frac{d}{dt}\mathbf{P}_0^{(\mathbf{v})}(t) = \mathbf{P}_0^{(\mathbf{v})}(t)\mathbf{L}_v, \tag{4}$$

with initial condition $\mathbf{P}_\ell^{(\mathbf{v})}(0) = \delta_{0,\ell}\mathbf{I}$, where δ denotes the Kronecker delta. Similar to [9, Sec. 3], multiplying the ℓth equation by z^ℓ, summing up and solving the differential equation gives the generating function

$$\mathbf{P}^{(v)*}(z,t) = \sum_{\ell=0}^{\infty} z^\ell \mathbf{P}_\ell^{(\mathbf{v})}(t) = e^{(\mathbf{L}_v + z\mathbf{F}_v)t}. \tag{5}$$

3.2. The phase distribution at the end of the service period

The phase at the end of the cycles, $\mathcal{J}(t)$, forms an embedded discrete time Markov chain (DTMC).

The probability matrices characterizing the number of arriving jobs and the phase transitions during the vacation period are $\int_0^\infty \sigma(x)\mathbf{P}_m^{(\mathbf{v})}(x)\,dx$. If m jobs are in the queue when the system enters the service period, the phase transitions over the busy period are given by $\mathbf{G}^m$, where matrix $\mathbf{G}$ is the minimal non-negative solution to the matrix-quadratic equation $\mathbf{0} = \mathbf{B}_s + \mathbf{L}_s\mathbf{G} + \mathbf{F}_s\mathbf{G}^2$ ([9]). Thus, the transition probability matrix of the DTMC, denoted by $\mathbf{Q}$, is expressed by

$$\mathbf{Q} = \int_0^\infty \sigma(x) \sum_{m=0}^{\infty} \mathbf{\Pi}_{sv}\mathbf{P}_m^{(\mathbf{v})}(x)\mathbf{\Pi}_{vs}\mathbf{G}^m \, dx. \tag{6}$$

The stationary distribution of $\mathbf{Q}$, denoted by β, is determined by the linear system $\beta\mathbf{Q} = \beta, \beta\mathbb{1} = 1$.

Theorem 3.1. *Vector β is the solution to the set of linear equations*

$$vec^T\langle\mathbf{\Pi}_{vs}^T\rangle\sigma^*(\mathbf{L}_v^T \otimes \mathbf{I} + \mathbf{F}_v^T \otimes \mathbf{G})(\mathbf{\Pi}_{sv}^T\beta^T \otimes \mathbf{I}) = \beta, \quad \beta\mathbb{1} = 1. \tag{7}$$

PROOF. In the first step the matrix-exponential representation for $\beta\sum_{m=0}^{\infty}\mathbf{\Pi}_{sv}\mathbf{P}_m^{(\mathbf{v})}(x)\mathbf{\Pi}_{vs}\mathbf{G}^m$ is derived. Applying the column stacking $\text{vec}\langle\rangle$ operator and the identity $\text{vec}\langle\mathbf{A}\mathbf{X}\mathbf{B}\rangle = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}\langle\mathbf{X}\rangle$ [15] yields

$$\sum_{m=0}^{\infty} \beta\mathbf{\Pi}_{sv}\mathbf{P}_m^{(\mathbf{v})}(t)\mathbf{\Pi}_{vs}\mathbf{G}^m = vec^T\langle\sum_{m=0}^{\infty} \mathbf{G}^{Tm}\mathbf{\Pi}_{vs}^T\mathbf{P}_m^{(\mathbf{v})^T}(t)\mathbf{\Pi}_{sv}^T\beta^T\rangle$$

$$= \left((\beta\mathbf{\Pi}_{sv} \otimes \mathbf{I}) \sum_{m=0}^{\infty} \left(\mathbf{P}_m^{(\mathbf{v})}(t) \otimes \mathbf{G}^{Tm}\right) \text{vec}\langle\mathbf{\Pi}_{vs}^T\rangle\right)^T \tag{8}$$

$$= vec^T\langle\mathbf{\Pi}_{vs}^T\rangle \sum_{m=0}^{\infty} \left(\mathbf{P}_m^{(\mathbf{v})^T}(t) \otimes \mathbf{G}^m\right) (\mathbf{\Pi}_{sv}^T\beta^T \otimes \mathbf{I}).$$

Transposing (3), Kronecker multiplying both sides by $\mathbf{G}^m$, summing from $m = 1$ to ∞ and adding (4) gives

$$\frac{d}{dt} \sum_{m=0}^{\infty} \left(\mathbf{P}_m^{(\mathbf{v})^T}(t) \otimes \mathbf{G}^m \right)$$

$$= (\mathbf{F}_v^T \otimes \mathbf{G}) \sum_{m=0}^{\infty} \left(\mathbf{P}_m^{(\mathbf{v})^T}(t) \otimes \mathbf{G}^m \right) + (\mathbf{L}_v^T \otimes \mathbf{I}) \sum_{m=0}^{\infty} \left(\mathbf{P}_m^{(\mathbf{v})^T}(t) \otimes \mathbf{G}^m \right),$$

from which $\sum_{m=0}^{\infty} \left(\mathbf{P}_m^{(\mathbf{v})^T}(t) \otimes \mathbf{G}^m \right) = e^{(\mathbf{L}_v^T \otimes \mathbf{I} + \mathbf{F}_v^T \otimes \mathbf{G})t}$ follows. Inserting this solution back to (8), multiplying it by $\sigma(x)$ and taking the integral provides the theorem.

3.3. The mean time spent in different levels during the service period

As a new contributions of the paper we derive matrix $\mathbf{H}_{m,\ell}$, thus the mean time spent in various phases of level ℓ starting from level m in a QBD characterized by matrices $\mathbf{B}_s, \mathbf{L}_s$ and $\mathbf{F}_s$. It is known that QBDs have a matrix-geometric stationary distribution with matrix $\mathbf{R}$, and that the mean time spent at different phases of level ℓ starting from level 0 before returning to level 0 is given by $\mathbf{R}^\ell$ [9].

However, in our vacation queue the starting level after a vacation is not 0, but the number of arrivals accumulated during the vacation, which is denoted by m. According to our best knowledge, this measure has not been investigated yet.

For $m > 0$, we define matrix $\mathbf{P}_{m,\ell}^{(\mathbf{s})}$ corresponding to the service period as

$$[\mathbf{P}_{m,\ell}^{(\mathbf{s})}(t)]_{i,j} = P(\Theta > u+t, \mathcal{N}(u+t) = \ell, \mathcal{J}(u+t) = j | \mathcal{N}(u) = m, \mathcal{J}(u) = i, \tilde{\sigma} = u),$$

where u marks the beginning and Θ marks the end of the service period, thus $\Theta = \min\{t : \mathcal{N}(u+t) = 0\}$. For $\ell > 1$ matrix $\mathbf{P}_{m,\ell}^{(\mathbf{s})}(t)$ satisfies

$$\frac{d}{dt} \mathbf{P}_{m,\ell}^{(\mathbf{s})}(t) = \mathbf{P}_{m,\ell-1}^{(\mathbf{s})}(t)\mathbf{F}_s + \mathbf{P}_{m,\ell}^{(\mathbf{s})}(t)\mathbf{L}_s + \mathbf{P}_{m,\ell+1}^{(\mathbf{s})}(t)\mathbf{B}_s, \tag{9}$$

and for $\ell = 1$ we have

$$\frac{d}{dt} \mathbf{P}_{m,1}^{(\mathbf{s})}(t) = \mathbf{P}_{m,1}^{(\mathbf{s})}(t)\mathbf{L}_s + \mathbf{P}_{m,2}^{(\mathbf{s})}(t)\mathbf{B}_s, \tag{10}$$

with initial values $\mathbf{P}_{m,\ell}^{(\mathbf{s})}(0) = \delta_{m,\ell}\mathbf{I}$. We are interested in the mean time spent in different states during the busy period, that is $\mathbf{H}_{m,\ell} = \int_{t=0}^{\infty} \mathbf{P}_{m,\ell}^{(\mathbf{s})}(t) \, dt$.

Theorem 3.2. *For $m \geq 1, \ell \geq 1$, matrices $\mathbf{H}_{m,\ell}$ are expressed by*

$$\begin{aligned} \mathbf{H}_{m,\ell} &= -\boldsymbol{\Psi}\mathbf{S}^m\mathbf{R}^\ell + \boldsymbol{\Psi}\mathbf{R}^{\ell-m}, &\quad \text{for } 1 \leq m \leq \ell, \\ \mathbf{H}_{m,\ell} &= -\boldsymbol{\Psi}\mathbf{S}^m\mathbf{R}^\ell + \boldsymbol{\Psi}\mathbf{S}^{m-\ell}, &\quad \text{for } 1 \leq \ell < m, \end{aligned} \tag{11}$$

where matrix $\boldsymbol{\Psi}$ is given as

$$\boldsymbol{\Psi} = (-\mathbf{S}\mathbf{F}_s - \mathbf{L}_s - \mathbf{R}\mathbf{B}_s)^{-1}, \tag{12}$$

8

and matrices $\mathbf{R}$ and $\mathbf{S}$ are determined as the minimal non-negative solutions to the quadratic equations

$$0 = \mathbf{F}_s + \mathbf{R}\mathbf{L}_s + \mathbf{R}^2\mathbf{B}_s, \quad 0 = \mathbf{B}_s + \mathbf{S}\mathbf{L}_s + \mathbf{S}^2\mathbf{F}_s. \tag{13}$$

PROOF. Integrating the differential equation (9) and (10) from $t = 0$ to ∞ we get

$$\mathbf{P}^{(\mathbf{s})}_{m,\ell}(\infty) - \mathbf{P}^{(\mathbf{s})}_{m,\ell}(0) = \mathbf{H}_{m,\ell-1}\mathbf{F}_s + \mathbf{H}_{m,\ell}\mathbf{L}_s + \mathbf{H}_{m,\ell+1}\mathbf{B}_s, \quad \text{for } \ell > 1,$$

$$\mathbf{P}^{(\mathbf{s})}_{m,1}(\infty) - \mathbf{P}^{(\mathbf{s})}_{m,1}(0) = \mathbf{H}_{m,1}\mathbf{L}_s + \mathbf{H}_{m,2}\mathbf{B}_s.$$

These two equations and the initial value $\mathbf{P}^{(\mathbf{s})}_{m,\ell}(0)$ lead to four different cases, in general: (1) when $\ell = 1$, (2) when $1 < \ell < m$, (3) when $1 < \ell = m$, finally, (4) when $\ell > m$. The corresponding equations are

$$-\delta_{m,1}\mathbf{I} = \mathbf{H}_{m,1}\mathbf{L}_s + \mathbf{H}_{m,2}\mathbf{B}_s, \tag{14}$$

$$0 = \mathbf{H}_{m,\ell-1}\mathbf{F}_s + \mathbf{H}_{m,\ell}\mathbf{L}_s + \mathbf{H}_{m,\ell+1}\mathbf{B}_s, \qquad \text{for } 1 < \ell < m, \tag{15}$$

$$-\mathbf{I} = \mathbf{H}_{m,m-1}\mathbf{F}_s + \mathbf{H}_{m,m}\mathbf{L}_s + \mathbf{H}_{m,m+1}\mathbf{B}_s, \qquad \text{for } m > 1, \tag{16}$$

$$0 = \mathbf{H}_{m,\ell-1}\mathbf{F}_s + \mathbf{H}_{m,\ell}\mathbf{L}_s + \mathbf{H}_{m,\ell+1}\mathbf{B}_s, \qquad \text{for } \ell > m. \tag{17}$$

The solution of (14)-(17) is given by a matrix-geometric combination

$$\mathbf{H}_{m,\ell} = \boldsymbol{\Phi}\mathbf{R}^{\ell-1} + \boldsymbol{\Psi}\mathbf{S}^{m-\ell}, \qquad \text{for } 1 \le \ell \le m, \tag{18}$$

$$\mathbf{H}_{m,\ell} = \mathbf{H}_{m,m}\mathbf{R}^{\ell-m}, \qquad \text{for } 1 \le m < \ell, \tag{19}$$

where matrices $\mathbf{R}$ and $\mathbf{S}$ are obtained such that the regular equations (15) and (17) are satisfied for any $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$. $\mathbf{R}$ and $\mathbf{S}$ are the minimal non-negative solutions to the quadratic equations (13) [9, Sec. 10].

Matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are obtained from the solution of the irregular equations (14) and (16) as

$$0 = \boldsymbol{\Phi}(\mathbf{L}_s + \mathbf{R}\mathbf{B}_s) + \boldsymbol{\Psi}(\mathbf{S}^{m-1}\mathbf{L}_s + \mathbf{S}^{m-2}\mathbf{B}_s),$$

$$-\mathbf{I} = \boldsymbol{\Phi}\mathbf{R}^{m-2}\underbrace{(\mathbf{F}_s + \mathbf{R}\mathbf{L}_s + \mathbf{R}^2\mathbf{B}_s)}_{\mathbf{0}} + \boldsymbol{\Psi}(\mathbf{S}\mathbf{F}_s + \mathbf{L}_s + \mathbf{R}\mathbf{B}_s).$$

The solution of $\boldsymbol{\Psi}$ is given by (12) and the solution of $\boldsymbol{\Phi}$ is

$$\boldsymbol{\Phi} = \boldsymbol{\Psi}(\mathbf{S}^{m-1}\mathbf{L}_s + \mathbf{S}^{m-2}\mathbf{B}_s)(\mathbf{L}_s + \mathbf{R}\mathbf{B}_s)^{-1}$$

$$= -\boldsymbol{\Psi}\mathbf{S}^m\mathbf{F}_s(\mathbf{L}_s + \mathbf{R}\mathbf{B}_s)^{-1} = -\boldsymbol{\Psi}\mathbf{S}^m\mathbf{R}, \tag{20}$$

where we exploited various identities of the fundamental matrices of QBDs.

Finally using the expressions of $\mathbf{H}_{m,\ell}$ from (19) and (18) as well as (12) and (20) establish the theorem.

Theorem 3.2 reflects a quantitative behavior which is present also in the piece-wise homogeneous QBD precess with a single point of inhomogeneity. Above the point of inhomogeneity ($m < \ell$ in our case) a matrix geometric behavior characterizes $\mathbf{H}_{m,\ell}$ with initial matrix $\boldsymbol{\Psi}(\mathbf{I} - \mathbf{S}^m\mathbf{R}^m)$ and matrix geometric term $\mathbf{R}^{\ell-m}$. While below the point of inhomogeneity ($\ell < m$ in our case) a combination of two matrix geometric series characterizes $\mathbf{H}_{m,\ell}$ with matrix geometric terms $\mathbf{R}^\ell$ and $\mathbf{S}^{m-\ell}$.

Corollary 1. *The generating function of $\mathbf{H}_{m,\ell}$ with respect to ℓ is expressed by*

$$\mathbf{H}_m^*(z) = \sum_{\ell=1}^{\infty} z^{\ell} \mathbf{H}_{m,\ell} = \mathbf{\Psi}(z^m \mathbf{I} - \mathbf{S}^m)\left((\mathbf{I} - z\mathbf{R})^{-1} + (z\mathbf{I} - \mathbf{S})^{-1}\mathbf{S}\right). \qquad (21)$$

PROOF. The generator function is derived by routine calculations as

$$\mathbf{H}_m^*(z) = -\mathbf{\Psi}\mathbf{S}^m \sum_{\ell=1}^{\infty} z^{\ell}\mathbf{R}^{\ell} + \mathbf{\Psi}\sum_{\ell=1}^{m} z^{\ell}\mathbf{S}^{m-\ell} + \mathbf{\Psi}\sum_{\ell=m+1}^{\infty} z^{\ell}\mathbf{R}^{\ell-m}$$

$$= -\mathbf{\Psi}\mathbf{S}^m \sum_{\ell=1}^{\infty} z^{\ell}\mathbf{R}^{\ell} + \mathbf{\Psi}z^m \sum_{\ell=1}^{m}(\mathbf{S}/z)^{m-\ell} + \mathbf{\Psi}z^m \sum_{\ell=1}^{\infty} z^{\ell}\mathbf{R}^{\ell}$$

$$= \mathbf{\Psi}(z^m\mathbf{I} - \mathbf{S}^m)\left((\mathbf{I} - z\mathbf{R})^{-1} - \mathbf{I}\right) + \mathbf{\Psi}z^m \sum_{\ell=0}^{m-1}(\mathbf{S}/z)^{\ell}$$

$$= \mathbf{\Psi}(z^m\mathbf{I} - \mathbf{S}^m)\left((\mathbf{I} - z\mathbf{R})^{-1} - \mathbf{I} + (\mathbf{I} - \mathbf{S}/z)^{-1}\right),$$

from which the corollary follows since $(\mathbf{I} - \mathbf{S}/z)^{-1} - \mathbf{I} = (\mathbf{I} - \mathbf{S}/z)^{-1}\mathbf{S}/z = (z\mathbf{I} - \mathbf{S})^{-1}\mathbf{S}$.

The following lemma provides some relations among the matrices playing an important role in the system, which will be referred several times in the sequel.

Lemma 3.3. *Among matrices $\mathbf{B}_s, \mathbf{L}_s, \mathbf{F}_s$ and matrix $\mathbf{\Psi}$ we have that*

$$\mathbf{B}_s\mathbf{\Psi} = (\mathbf{I} - \mathbf{SR})^{-1}\mathbf{S}, \qquad (22)$$

$$\mathbf{F}_s\mathbf{\Psi} = (\mathbf{I} - \mathbf{RS})^{-1}\mathbf{R}, \qquad (23)$$

$$\mathbf{L}_s\mathbf{\Psi} = \mathbf{I} - (\mathbf{I} - \mathbf{SR})^{-1} - (\mathbf{I} - \mathbf{RS})^{-1}. \qquad (24)$$

PROOF. We note that for a stable vacation queue matrix $\mathbf{I} - \mathbf{SR}$ and $\mathbf{I} - \mathbf{RS}$ are non-singular, due to the fact that the spectral radius of $\mathbf{S}$ and $\mathbf{R}$ are one and less than one, respectively. Multiplying both sides of (12) by $\mathbf{S}(-\mathbf{SF}_s - \mathbf{L}_s - \mathbf{RB}_s)$ from the left and exploiting that $-\mathbf{S}^2\mathbf{F}_s - \mathbf{SL}_s = \mathbf{B}_s$ due to (13) gives

$$\mathbf{B}_s\mathbf{\Psi} - \mathbf{SRB}_s\mathbf{\Psi} = \mathbf{S},$$

which is equivalent to (22). (23) can be proven similarly, by multiplying (12) by $\mathbf{R}(-\mathbf{SF}_s - \mathbf{L}_s - \mathbf{RB}_s)$ from the left.

Proving the third equality is a bit more involved. Making use of relation $\mathbf{U} = \mathbf{L}_s + \mathbf{RB}_s$ (valid for all QBDs, see [9]) and $\mathbf{H}_{1,1} = (-\mathbf{U})^{-1}$ (by stochastic interpretation, i.e., both represent the mean time spent at level 1 before returning to level 0, starting from level 1) leads to

$$\mathbf{L}_s\mathbf{\Psi} = (\mathbf{U} - \mathbf{RB}_s)\mathbf{\Psi} = (-\mathbf{H}_{1,1})^{-1}\mathbf{\Psi} - \mathbf{RB}_s\mathbf{\Psi}$$

$$= -(\mathbf{I} - \mathbf{SR})^{-1} - \mathbf{R}(\mathbf{I} - \mathbf{SR})^{-1}\mathbf{S}$$

$$= -(\mathbf{I} - \mathbf{SR})^{-1} - \mathbf{R}\sum_{\ell=0}^{\infty}(\mathbf{SR})^{\ell}\mathbf{S}$$

$$= -(\mathbf{I} - \mathbf{SR})^{-1} - \sum_{\ell=0}^{\infty}\mathbf{RS}(\mathbf{RS})^{\ell}$$

$$= -(\mathbf{I} - \mathbf{SR})^{-1} - \left((\mathbf{I} - \mathbf{RS})^{-1} - \mathbf{I}\right),$$

since $\mathbf{H}_{1,1} = \mathbf{\Psi}(\mathbf{I} - \mathbf{SR})$ according to (11).

Additional to matrices $\mathbf{R}$ and $\mathbf{S}$, we introduce matrix $\mathbf{G}$ as the minimal non-negative solutions to the quadratic matrix equation [9]

$$0 = \mathbf{B}_s + \mathbf{L}_s\mathbf{G} + \mathbf{F}_s\mathbf{G}^2.$$

The next two propositions are necessary to arrive to short queue length formulas at the end of the section.

Proposition 1. *Matrices* $\mathbf{G}, \mathbf{S}$ *and* $\mathbf{\Psi}$ *are related as*

$$\mathbf{G}^m\mathbf{\Psi} = \mathbf{\Psi}\mathbf{S}^m, \tag{25}$$

for $m \geq 0$.

PROOF. A well known identity in the matrix-analytic solution of QBDs is that $\mathbf{G} = (-\mathbf{U})^{-1}\mathbf{B}_s$, yielding

$$\mathbf{G}\mathbf{\Psi} = (-\mathbf{U})^{-1}\mathbf{B}_s\mathbf{\Psi} = \mathbf{H}_{1,1}(\mathbf{I} - \mathbf{SR})^{-1}\mathbf{S} = \mathbf{\Psi}\mathbf{S},$$

for $m = 1$, where we utilized (22) and that $(-\mathbf{U})^{-1} = \mathbf{H}_{1,1} = \mathbf{\Psi}(\mathbf{I} - \mathbf{SR})$. Applying this relation recursively provides (25) for $m > 1$.

Proposition 2.

$$\mathbf{\Psi}\left((\mathbf{I} - z\mathbf{R})^{-1} + (z\mathbf{I} - \mathbf{S})^{-1}\mathbf{S}\right) = (-\mathbf{B}_s/z - \mathbf{L}_s - \mathbf{F}_s z)^{-1}. \tag{26}$$

PROOF. The proof is based on simple algebraic manipulations. Based on Lemma 3.3, $(-\mathbf{B}_s/z - \mathbf{L}_s - \mathbf{F}_s z)\mathbf{\Psi}$ simplifies to

$$(-\mathbf{B}_s/z - \mathbf{L}_s - \mathbf{F}_s z)\mathbf{\Psi} = -\mathbf{I} + (\mathbf{I} - \mathbf{SR})^{-1}(z\mathbf{I} - \mathbf{S})/z + (\mathbf{I} - \mathbf{RS})^{-1}(\mathbf{I} - z\mathbf{R}),$$

which, multiplied by $\left(((\mathbf{I} - z\mathbf{R})^{-1} + (z\mathbf{I} - \mathbf{S})^{-1}\mathbf{S}\right)$ gives

$$-(\mathbf{I} - \mathbf{SR})^{-1}(\mathbf{I} - \mathbf{SR})(\mathbf{I} - z\mathbf{R})^{-1} + \frac{1}{z}(\mathbf{I} - \mathbf{SR})^{-1}(z\mathbf{I} - \mathbf{S})(\mathbf{I} - z\mathbf{R})^{-1} + (\mathbf{I} - \mathbf{RS})^{-1}$$

$$-(\mathbf{I} - \mathbf{RS})^{-1}(\mathbf{I} - \mathbf{RS})(z\mathbf{I} - \mathbf{S})^{-1}\mathbf{S} + \frac{1}{z}(\mathbf{I} - \mathbf{SR})^{-1}\mathbf{S} + (\mathbf{I} - \mathbf{RS})^{-1}(\mathbf{I} - z\mathbf{R})(z\mathbf{I} - \mathbf{S})^{-1}\mathbf{S}$$

$$= -\frac{1}{z}(\mathbf{I} - \mathbf{SR})^{-1}\mathbf{S} + \frac{1}{z}(\mathbf{I} - \mathbf{SR})^{-1}\mathbf{S} - (\mathbf{I} - \mathbf{RS})^{-1}\mathbf{R}(z\mathbf{I} - \mathbf{S})(z\mathbf{I} - \mathbf{S})^{-1}\mathbf{S} + (\mathbf{I} - \mathbf{RS})^{-1}$$

$$= \mathbf{I}.$$

3.4. The mean time spent at different levels in a stationary cycle

The next theorem gives expressions for the mean time spent at different levels in the individual parts of the stationary cycle.

Theorem 3.4. *The PGF of the mean time spent at different levels in the vacation period of the stationary cycle,* $M^{(v)*}(z) = \sum_{\ell=0}^{\infty} z^\ell M_\ell^{(v)}$ *can be expressed as*

$$M^{(v)*}(z) = \beta\mathbf{\Pi}_{sv}\left(\mathbf{I} - \sigma^*(\mathbf{L}_v + z\mathbf{F}_v)\right)(-\mathbf{L}_v - z\mathbf{F}_v)^{-1}\mathbb{1}, \tag{27}$$

where $\sigma^*(\mathbf{M})$ *with square matrix* $\mathbf{M}$ *is defined by* $\int_{u=0}^{\infty} \sigma(u)\, e^{\mathbf{M}u}du$.

Similarly the PGF of the mean time spent at different levels in the service period of the stationary cycle, $M^{(s)*}(z) = \sum_{\ell=0}^{\infty} z^\ell M_\ell^{(s)}$ *can be expressed as*

$$M^{(s)*}(z) = \beta\left(\mathbf{I} - \mathbf{\Pi}_{sv}\sigma^*(\mathbf{L}_v + z\mathbf{F}_v)\mathbf{\Pi}_{vs}\right)(\mathbf{B}_s/z + \mathbf{L}_s + \mathbf{F}_s z)^{-1}\mathbb{1}. \tag{28}$$

PROOF. By applying (5) in the first term of (2) the generating function $M^{(v)*}(z)$ can be expressed as

$$
\begin{aligned}
M^{(v)*}(z) &= \int_{u=0}^{\infty} \sigma(u)\,\beta\mathbf{\Pi}_{sv} \int_{t=0}^{u} \sum_{\ell=0}^{\infty} z^{\ell} \mathbf{P}_{\ell}^{(\mathbf{v})}(t)\mathbb{1}\,dt\,du \\
&= \beta\mathbf{\Pi}_{sv} \int_{u=0}^{\infty} \sigma(u) \int_{t=0}^{u} e^{(\mathbf{L}_v + z\mathbf{F}_v)t}\mathbb{1}\,dt\,du \\
&= \beta\mathbf{\Pi}_{sv} \int_{u=0}^{\infty} \sigma(u)\left(\mathbf{I} - e^{(\mathbf{L}_v + z\mathbf{F}_v)u}\right)(-\mathbf{L}_v - z\mathbf{F}_v)^{-1}\mathbb{1}\,du.
\end{aligned}
\tag{29}
$$

The expression (27) comes by applying the shorthand notation $\sigma^*(\mathbf{M})$ in (29).

$M^{(s)*}(z)$ is obtained by substituting $\mathbf{H}_m^*(z)$ provided by Corollary 1 into the definition (2). We get

$$
\begin{aligned}
M^{(s)*}(z) &= \sum_{\ell=0}^{\infty} z^{\ell} M_{\ell}^{(s)} = \int_{u=0}^{\infty} \sigma(u)\beta\mathbf{\Pi}_{sv} \sum_{m=0}^{\infty} \mathbf{P}_m^{(\mathbf{v})}(u)\mathbf{\Pi}_{vs}\mathbf{H}_m^*(z)\mathbb{1}\,du \\
&= \int_{u=0}^{\infty} \sigma(u)\beta\mathbf{\Pi}_{sv} \sum_{m=0}^{\infty} \mathbf{P}_m^{(\mathbf{v})}(u)z^m\mathbf{\Pi}_{vs}\mathbf{\Psi}\left((\mathbf{I} - z\mathbf{R})^{-1} + (z\mathbf{I} - \mathbf{S})^{-1}\mathbf{S}\right)\mathbb{1}\,du \\
&\quad - \beta \int_{u=0}^{\infty} \sigma(u) \sum_{m=0}^{\infty} \mathbf{\Pi}_{sv}\mathbf{P}_m^{(\mathbf{v})}(u)\mathbf{\Pi}_{vs} \underbrace{\mathbf{\Psi}\mathbf{S}^m}_{\mathbf{G}^m\mathbf{\Psi}}\left((\mathbf{I} - z\mathbf{R})^{-1} + (z\mathbf{I} - \mathbf{S})^{-1}\mathbf{S}\right)\mathbb{1}\,du,
\end{aligned}
$$

where we made use of Proposition 1 to write $\mathbf{G}^m\mathbf{\Psi}$ instead of $\mathbf{\Psi}\mathbf{S}^m$. After this swap, however, it can be recognized that the integral is equal to matrix $\mathbf{Q}$ (introduced by (6)), for which $\beta\mathbf{Q} = \beta$ holds. This observation simplifies $M^{(s)*}(z)$ to

$$
\begin{aligned}
M^{(s)*}(z) &= \beta\mathbf{\Pi}_{sv}\sigma^*(\mathbf{L}_v + z\mathbf{F}_v)\mathbf{\Pi}_{vs}\mathbf{\Psi}\left((\mathbf{I} - z\mathbf{R})^{-1} + (z\mathbf{I} - \mathbf{S})^{-1}\mathbf{S}\right)\mathbb{1} \\
&\quad - \beta\mathbf{\Psi}\left((\mathbf{I} - z\mathbf{R})^{-1} + (z\mathbf{I} - \mathbf{S})^{-1}\mathbf{S}\right)\mathbb{1} \\
&= -\beta\left(\mathbf{I} - \mathbf{\Pi}_{sv}\sigma^*(\mathbf{L}_v + z\mathbf{F}_v)\mathbf{\Pi}_{vs}\right)\mathbf{\Psi}\left((\mathbf{I} - z\mathbf{R})^{-1} + (z\mathbf{I} - \mathbf{S})^{-1}\mathbf{S}\right)\mathbb{1}.
\end{aligned}
\tag{30}
$$

The final expression (28) is obtained by applying Proposition 2 in (30).

We note that the relative simplicity of (28) is due to the matrix swapping relation of Proposition 1.

3.5. The generating function of the number of jobs in the system

The previous results lead to the following main theorem. Let q_i be the stationary probability that there are i jobs in the system for $i \geq 0$.

Theorem 3.5. *The generating function of the number of jobs in the system,* $q^*(z) = \sum_{i=0}^{\infty} z^i q_i$, *is given by*

$$
\begin{aligned}
q^*(z) = \frac{1}{c}\beta\Bigg(&\left(\mathbf{I} - \mathbf{\Pi}_{sv}\sigma^*(\mathbf{L}_v + z\mathbf{F}_v)\mathbf{\Pi}_{vs}\right)(\mathbf{B}_s/z + \mathbf{L}_s + \mathbf{F}_s z)^{-1} \\
&- \left(\mathbf{\Pi}_{sv} - \mathbf{\Pi}_{sv}\sigma^*(\mathbf{L}_v + z\mathbf{F}_v)\right)(\mathbf{L}_v + z\mathbf{F}_v)^{-1}\Bigg)\mathbb{1},
\end{aligned}
\tag{31}
$$

where β is determined by (7) and constant c satisfies $\lim_{z\to 1} q^*(z) = 1$.

12

PROOF. The PGF of the mean time spent at different levels in the stationary cycle, $M^*(z) = \sum_{\ell=0}^{\infty} z^\ell M_\ell$. The probability q_i is proportional to M_i, therefore $q^*(z)$ can be expressed as

$$q^*(z) = \frac{1}{c} M^*(z) = \frac{1}{c} (M^{(v)*}(z) + M^{(s)*}(z)), \tag{32}$$

where c is an appropriate normalization constant. Inserting (27) and (28) into (32) provides the theorem.

Taking the derivatives of $q^*(z)$ at $z \to 1$ gives the factorial moments of the number of jobs in the queue.

4. Analysis of special model variants

In this section we present simpler analysis results for special sub-classes of the general *Markov modulated vacation queue*.

4.1. Stationary distribution at service completion

For the general *Markov modulated vacation queue* the phase distribution at the end of service, β, can be computed according to (7). Unfortunately, the numerical analysis based on (7) is a computationally demanding task, due to the Kronecker product in the evaluation of $\sigma^*(\mathbf{L}_v^T \otimes \mathbf{I} + \mathbf{F}_v^T \otimes \mathbf{G})$.

The following theorem presents a much simpler computation of β for a subset of *Markov modulated vacation queue*s.

Theorem 4.1. *In the special case of Markov modulated vacation queues, where* $\mathbf{G}$ *commutes with* $\mathbf{\Pi}_{vs}$ *and* $\mathbf{G}$ *commutes with* $(\mathbf{L}_v + \mathbf{F}_v \mathbf{G})$, *the stationary distribution at service completion,* β, *is the unique solution of*

$$\beta \mathbf{\Pi}_{sv} \sigma^* (\mathbf{L}_v + \mathbf{F}_v \mathbf{G}) \mathbf{\Pi}_{vs} = \beta, \quad \beta \mathbb{1} = 1. \tag{33}$$

PROOF. The stationary distribution at service completion, β, satisfies the equation $\beta \mathbf{Q} = \beta$. Starting from this and using (6) this equation for β can be rearranged as

$$\begin{aligned}
\beta &= \beta \int_0^\infty \sigma(x) \sum_{m=0}^\infty \mathbf{\Pi}_{sv} \mathbf{P}_m^{(\mathbf{v})}(x) \mathbf{\Pi}_{vs} \mathbf{G}^m \, dx \\
&= \beta \mathbf{\Pi}_{sv} \int_0^\infty \sigma(x) \sum_{m=0}^\infty \mathbf{P}_m^{(\mathbf{v})}(x) \mathbf{\Pi}_{vs} \mathbf{G}^m \, dx \\
&= \beta \mathbf{\Pi}_{sv} \int_0^\infty \sigma(x) \sum_{m=0}^\infty \mathbf{P}_m^{(\mathbf{v})}(x) \mathbf{G}^m \, dx \mathbf{\Pi}_{vs},
\end{aligned} \tag{34}$$

where in the last step we made use of the commutativity of $\mathbf{G}$ and $\mathbf{\Pi}_{vs}$.

Next we derive an explicit form for $\sum_{m=0}^{\infty} \mathbf{P}_m^{(\mathbf{v})}(x)\mathbf{G}^m$ for the case when the commutativity of $\mathbf{G}$ and $(\mathbf{L}_v + \mathbf{F}_v\mathbf{G})$ holds. Multiplying differential equations (3) and (4) by $\mathbf{G}^m$ from the right and summing them with regards to m leads to

$$
\begin{aligned}
\frac{d}{dx} \sum_{m=0}^{\infty} \mathbf{P}_m^{(\mathbf{v})}(x)\mathbf{G}^m &= \sum_{m=0}^{\infty} \mathbf{P}_m^{(\mathbf{v})}(x)\mathbf{L}_v\mathbf{G}^m + \sum_{m=0}^{\infty} \mathbf{P}_m^{(\mathbf{v})}(x)\mathbf{F}_v\mathbf{G}^{m+1} \\
&= \sum_{m=0}^{\infty} \mathbf{P}_m^{(\mathbf{v})}(x)(\mathbf{L}_v + \mathbf{F}_v\mathbf{G})\mathbf{G}^m \qquad (35)\\
&= \sum_{m=0}^{\infty} \mathbf{P}_m^{(\mathbf{v})}(x)\mathbf{G}^m(\mathbf{L}_v + \mathbf{F}_v\mathbf{G}),
\end{aligned}
$$

where the last step makes use of the commutativity of $\mathbf{G}$ and $(\mathbf{L}_v + \mathbf{F}_v\mathbf{G})$. (35) together with $\sum_{m=0}^{\infty} \mathbf{P}_m^{(\mathbf{v})}(0)\mathbf{G}^m = \mathbf{I}$ implies that

$$
\sum_{m=0}^{\infty} \mathbf{P}_m^{(\mathbf{v})}(x)\mathbf{G}^m = e^{(\mathbf{L}_v + \mathbf{F}_v\mathbf{G})x}. \qquad (36)
$$

The theorem is proven by applying (36) in (34) and using the shorthand notation for the matrix LT.

The result of Theorem 4.1 makes a more efficient computation of β possible, as the matrices involved are much smaller due to the lack of Kronecker operations. For the special case when even the supplementary condition $\mathbf{\Pi}_{sv} = \mathbf{\Pi}_{vs} = \mathbf{I}$ holds, β can be obtained even simpler.

Theorem 4.2. *In the special case of Markov modulated vacation queues, where $\mathbf{\Pi}_{sv} = \mathbf{\Pi}_{vs} = \mathbf{I}$ and $\mathbf{G}$ commutes with $(\mathbf{L}_v + \mathbf{F}_v\mathbf{G})$, the stationary distribution at service completion, β, is the unique solution of*

$$
\beta\,\mathbf{G} = \beta, \quad \beta\mathbb{1} = 1. \qquad (37)
$$

PROOF. Applying $\mathbf{\Pi}_{sv} = \mathbf{\Pi}_{vs} = \mathbf{I}$ to (33) leads to a simplified equation for β as $\beta\sigma^*(\mathbf{L}_v + \mathbf{F}_v\mathbf{G}) = \beta$. It can be seen from the Taylor expansion of the exponential term in the matrix LT that β fulfills the equation $\beta(\mathbf{L}_v + \mathbf{F}_v\mathbf{G}) = 0$. Making use of the commutativity of $\mathbf{G}$ and $(\mathbf{L}_v + \mathbf{F}_v\mathbf{G})$ we can write

$$
0 = \beta(\mathbf{L}_v + \mathbf{F}_v\mathbf{G})\mathbf{G} = \beta\mathbf{G}(\mathbf{L}_v + \mathbf{F}_v\mathbf{G}). \qquad (38)
$$

The matrix $(\mathbf{L}_v + \mathbf{F}_v\mathbf{G})$ is a proper generator matrix, since it can contain negative elements only in the diagonal and $(\mathbf{L}_v + \mathbf{F}_v\mathbf{G})\mathbb{1} = \mathbf{L}_v\mathbb{1} + \mathbf{F}_v\mathbb{1} = 0$, since $\mathbf{G}$ is stochastic. Therefore the homogenous equation $\beta(\mathbf{L}_v + \mathbf{F}_v\mathbf{G}) = 0$ determines β (up to a normalization constant) uniquely, which together with (38) implies that $\beta\mathbf{G} = \beta$.

It is an interesting qualitative property in Theorem 4.2 that the stationary phase distribution at the end of service is independent of the vacation time distribution.

14

This way Theorem 4.2 provides an easy to check condition (the commutativity of 2 matrices) for the stationary distribution to be independent of the vacation time and to be easy to compute.

The practical importance of Theorem 4.2 comes from the fact that the required commutativity holds in a number of practically important cases. E.g., in case of a *MAP/MAP/1 vacation queue* matrix $(\mathbf{L}_v + \mathbf{F}_v\mathbf{G}) = (\mathbf{D}_0 \otimes \mathbf{I}) + (\mathbf{D}_1 \otimes \mathbf{I})\mathbf{G}$ and matrix $\mathbf{G}$ do commute according to [7, Theorem 9], and the same relation applies for the *MAP/PH/1 vacation queue* as well. On the contrary, the commutativity relation of Theorem 4.2 does not hold in general for the *QBD vacation queue* and for the *QBD vacation queue with independent arrival and service processes*.

4.2. Special formulas for the number of jobs

In the special case of *Markov modulated vacation queue without special phase change* (where $\mathbf{\Pi}_{sv} = \mathbf{\Pi}_{vs} = \mathbf{I}$) Theorem 3.5 simplifies to the following corollary.

Corollary 2. *When* $\mathbf{\Pi}_{sv} = \mathbf{\Pi}_{vs} = \mathbf{I}$ *the generating function of the number of jobs in the system is*

$$q^*(z) = \frac{1}{c}\beta\Big(\mathbf{I} - \sigma^*(\mathbf{L}_v + z\mathbf{F}_v)\Big)\Big((\mathbf{B}_s/z + \mathbf{L}_s + \mathbf{F}_s z)^{-1} - (\mathbf{L}_v + z\mathbf{F}_v)^{-1}\Big)\mathbb{1} \quad (39)$$

Further more, in case of the *QBD vacation queue* Theorem 3.5 further simplifies.

Corollary 3. *In the QBD vacation queue, where* $\mathbf{\Pi}_{vs} = \mathbf{\Pi}_{sv} = \mathbf{I}$ *and* $\mathbf{F}_v = \mathbf{F}_s = \mathbf{F}$, $\mathbf{L}_s = \mathbf{L}$, $\mathbf{B}_s = \mathbf{B}$, $\mathbf{L}_v = \mathbf{B} + \mathbf{L}$, *the generating function of the number of jobs in the system can be expressed as*

$$
\begin{aligned}
q^*(z)^{(QBD)} &= \frac{1}{c}\beta\big(\mathbf{I} - \sigma^*(\mathbf{B} + \mathbf{L} + z\mathbf{F})\big)\left((\mathbf{B}/z + \mathbf{L} + \mathbf{F}z)^{-1} - (\mathbf{B} + \mathbf{L} + z\mathbf{F})^{-1}\right)\mathbb{1} \\
&= \frac{1}{c}\beta\big(\mathbf{I} - \sigma^*(\mathbf{B} + \mathbf{L} + z\mathbf{F})\big)(\mathbf{B} + \mathbf{L} + z\mathbf{F})^{-1} \\
&\quad \times \left((\mathbf{B} + \mathbf{L} + z\mathbf{F} + \mathbf{B}/z - \mathbf{B}/z)(\mathbf{B}/z + \mathbf{L} + \mathbf{F}z)^{-1} - \mathbf{I}\right)\mathbb{1} \\
&= \frac{1}{c}(1 - z)\beta\big(\sigma^*(\mathbf{B} + \mathbf{L} + z\mathbf{F}) - \mathbf{I}\big)(\mathbf{B} + \mathbf{L} + z\mathbf{F})^{-1}\mathbf{B}(\mathbf{B} + z\mathbf{L} + z^2\mathbf{F})^{-1}\mathbb{1}.
\end{aligned}
$$
$$(40)$$

Both of the above expressions are products of transform domain expressions, which can be seen as factorization property. [5]. The second term of the product has a nice stochastic interpretation, because the vector generating function of the stationary queue length distribution of a QBD (without vacations) with regular matrices $\mathbf{B}$, $\mathbf{L}$, $\mathbf{F}$ and matrix $\mathbf{L}_0 = \mathbf{L} + \mathbf{B}$ at level zero is $\pi^*(z) = (1 - z)\mathbf{B}(\mathbf{B} + z\mathbf{L} + z^2\mathbf{F})^{-1}\mathbb{1}$.

Considering the subset relations of the set of special models depicted in Figure 1, we can see that the 2 branches have got different benefits. The *MAP/MAP/1 vacation queue* and the *MAP/PH/1 vacation queue* fulfills the commutativity condition of Theorem 4.2 and consequently the computation of vector β is computationally easy, while the *QBD vacation queue* and the *QBD vacation queue with independent arrival and service processes* do not fulfills the commutativity condition of Theorem 4.2 in general, but exhibit the factorization property.

4.3. The number of jobs in case of phase-type distributed vacations

If the vacations are phase-type (PH) distributed, the analysis of the number of jobs in the system is much simpler. Assuming that the density function of the vacations is $\sigma(x) = \alpha e^{\mathbf{A}x} a$, the queue length process $\mathcal{X}(t)$ is a CTMC with a QBD structured generator $\mathbf{Q}$, which is

$$\mathbf{Q} = \begin{bmatrix} \mathbf{L}_0 & \mathbf{F}_0 & & \\ \mathbf{B}_1 & \mathbf{L} & \mathbf{F} & \\ & \mathbf{B} & \mathbf{L} & \mathbf{F} \\ & & \ddots & \ddots & \ddots \end{bmatrix}.$$

The blocks of the QBD in the regular part (non-zero levels) are defined by

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} & \\ & \mathbf{B}_s \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} \mathbf{L}_v \oplus \mathbf{A} & (\mathbf{I} \otimes a)\mathbf{\Pi}_{vs} \\ & \mathbf{L}_s \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \mathbf{F}_v \otimes \mathbf{I} & \\ & \mathbf{F}_s \end{bmatrix}.$$

The states belonging to level i are partitioned into two groups. The first state group corresponds to the vacation, the second to the service periods. The end of the vacation triggers a transition to the second state group and the phases are changed according to matrix $\mathbf{\Pi}_{vs}$.

Moving from the second group to the first one and initiating a vacation period can occur only when the last departure leaves the system (the QBD reaches level 0). The blocks for the irregular level 0 are

$$\mathbf{B}_1 = \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_s \mathbf{\Pi}_{sv} \otimes \alpha \end{bmatrix}, \quad \mathbf{L}_0 = \mathbf{L}_v \oplus \mathbf{A} + \mathbf{\Pi}_{vs}\mathbf{\Pi}_{sv} \otimes a\,\alpha, \quad \mathbf{F}_0 = \begin{bmatrix} \mathbf{F}_v \otimes \mathbf{I} & \mathbf{0} \end{bmatrix}.$$

The stationary distribution of $\mathcal{X}(t)$ is matrix-geometric

$$p_\ell = p_1 \mathbf{R}^{\ell-1}, \quad \ell \geq 1,$$

where $\mathbf{R}$ is the minimal non-negative solution to the matrix-quadratic equation $\mathbf{0} = \mathbf{F} + \mathbf{R}\mathbf{L} + \mathbf{R}^2\mathbf{B}$, while p_0 and p_1 are obtained by solving the linear system representing the equilibrium equations for levels 0 and 1 and the normalization condition:

$$p_0 \mathbf{L}_0 + p_1 \mathbf{B}_1 = \mathbf{0},$$
$$p_0 \mathbf{F}_0 + p_1 \mathbf{L} + p_1 \mathbf{R} \mathbf{B} = \mathbf{0},$$
$$p_0 \mathbb{1} + p_1 (\mathbf{I} - \mathbf{R})^{-1} \mathbb{1} = 1.$$

5. Numerical example

The presented procedure has been implemented in Mathematica with high precision arithmetic[1]. $\sigma^*(\mathbf{X})$ for any matrix $\mathbf{X}$ is obtained by numerical integration, and the moments of the queue length are calculated by numerical derivation of $q^*(z)$.
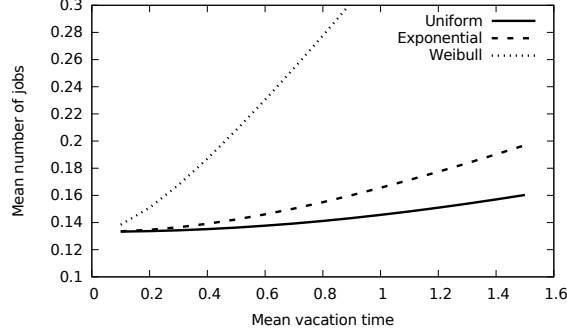
---

[1] The implementation can be downloaded from `http://www.hit.bme.hu/~ghorvath/software`

Figure 3: The mean number of jobs in the system

	Uniform	**Exponential**	**Weibull**
The general model	(0.546,0.109,0.345)	(0.543,0.116,0.341)	(0.539,0.142,0.319)
The MAP/MAP/1 vacation queue	(0.214,0.097,0.091, …0.04,0.382,0.176)	(0.214,0.097,0.091, …0.04,0.382,0.176)	(0.214,0.097,0.091, …0.04,0.382,0.176)
QBD vac. queue	(0.546,0.109,0.345)	(0.543,0.115,0.342)	(0.53,0.14,0.33)
The indep. QBD vacation queue	(0.21,0.101,0.09, …0.041,0.373,0.185)	(0.207,0.104,0.089, …0.042,0.367,0.191)	(0.197,0.113,0.086, …0.046,0.348,0.21)

Table 1: Vector β as a function of the vacation distribution

The first numerical experiment investigates the effect of the mean and the distribution of the vacation time on the mean number of jobs in the system. We consider the following matrices defining a *QBD vacation queue*.

$$\mathbf{B}_s = \begin{bmatrix} 8 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{L}_s = \begin{bmatrix} -8.05 & 0.05 \\ 1 & -3 \end{bmatrix}, \mathbf{F}_s = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix},$$

$$\mathbf{F}_v = \mathbf{F}_s, \mathbf{L}_v = \mathbf{L}_s + \mathbf{B}_s, \mathbf{\Pi}_{sv} = \mathbf{\Pi}_{vs} = \mathbf{I}.$$

In this example the arrival and service processes are dependent as much as possible: a service can occur only in state 1, and an arrival only in state 2.

The computation has been performed for the following type of vacation distributions: Uniform distribution, Exponential distribution, and Weibull distribution with shape parameter $k = 1/2$. The mean number of jobs is depicted in Figure 3. As expected, the number of jobs in the system is the highest when the vacation times are Weibull distributed which has the heaviest tail.

The next experiment demonstrates the effect of Theorem 4.2, i.e., that β, the stationary phase distribution at the end of the service period does not depend on the vacation time distribution in the MAP/MAP/1 case. We have set up simple examples for all the special cases and computed β with various vacation time distributions. The results are shown in Table 1, according to which β does depend on the vacation time distribution in all special cases considered in the paper, except the MAP/MAP/1 one.

17

6. Conclusions

In this paper we considered a general class of vacation queues with dependent arrival and service processes. Several previously studied vacation queues can be obtained as special cases of the general model as it is depicted on Figure 1. Based on the principles of the matrix-analytic approach we presented a computationally efficient description of the stationary queue length distribution in transform domain, which is based on an essential matrix swapping relation provided by Proposition 1. Finally, a numerical example demonstrates the stochastic behavior of the considered class of vacation queues, e.g. the effect of vacation time distribution.

Unfortunately, we do not have nice stochastic interpretations for the majority of compact matrix expressions which were obtained in an essentially algebraic way. It is among our future research plans to find such interpretations which make these results easier to use.

[1] A.-S. Alfa, A discrete MAP/PH/1 queue with vacations and exhaustive time-limited service, *Oper. Res. Lett.*, **18** (1995), 31–40.

[2] A.-S. Alfa, Discrete time analysis of MAP/PH/1 vacation queue with gated time-limited service, *Queueing Systems*, **29** (1998), 35–54.

[3] Y. Baba, Analysis of a GI/M/1 queue with multiple working vacations, *Operation Research Letters*, **33** (2005), 201–209.

[4] N. S. C. Goswami, The discrete-time MAP/PH/1 queue with multiple working vacations, *Applied Mathematical Modelling*, **34** (2010), 931–946.

[5] S. Chang and T. Takine, Factorization and stochastic decomposition properties in bulk queues with generalized vacations, *Queueing Systems*, **50** (2005), 165–183.

[6] B. T. Doshi, Queueing systems with vacations - a survey, *Queueing Systems*, **1** (1986), 29–66.

[7] G. Horváth, B. Van Houdt and M. Telek, Commuting matrices in the queue length and sojourn time analysis of MAP/MAP/1 queues, *Stochastic Models*, **30** (2014), 554–575.

[8] J.-C. Ke, C.-H. Wu and Z. G. Zhang, Recent developments in vacation queueing models: a short survey, *International Journal of Operations Research*, **7** (2010), 3–8.

[9] G. Latouche and V. Ramaswami, *Introduction to matrix analytic methods in stochastic modeling*, vol. 5, Siam, 1999.

[10] J. H. Li, N. S. Tian and W. Liu, Discrete-time GI/Geom/1 queue with multiple working vacations, *Queueing Systems*, **56** (2007), 53–63.

[11] M. F. Neuts, A versatile Markovian point process, *Journal of Applied Probability*, 764–779.

[12] Z. Saffer and M. Telek, Analysis of BMAP/G/1 vacation model of non-M/G/1-type, in *EPEW*, vol. 5261 of LNCS, Springer, Mallorca, Spain, 2008, 212–226.

[13] Z. Saffer and M. Telek, Closed form results for BMAP/G/1 vacation model with binomial type disciplines, *Publ. Math. Debrecen*, **76** (2010), 359–378.

[14] Z. Saffer and W. Yue, M/M/c multiple synchronous vacation model with gated discipline, *Journal of Industrial and Management Optimization (JIMO)*, **8** (2012), 939–968.

[15] W.-H. Steeb, *Matrix calculus and Kronecker product with applications and C++ programs*, World Scientific, 1997.

[16] H. Takagi, *Queueing Analysis - A Foundation of Performance Evaluation, Vacation and Prority Systems, vol.1*, North-Holland, New York, 1991.

[17] N. Tian and Z. G. Zhang, *Vacation queueing models: theory and applications*, vol. 93, Springer Science & Business Media, 2006.

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