A DUAL TANDEM QUEUEING SYSTEM WITH GI SERVICE TIME AT THE FIRST QUEUE

ZSOLT SAFFER
Department of Telecommunications
Budapest University of Technology and Economics, Budapest, Hungary

WUYI YUE
Department of Intelligence and Informatics
Konan University, Kobe 658-8501, Japan

Abstract. In this paper we consider the analysis of a tandem queueing model $M/G/1 \rightarrow /M/1$. In contrast to the vast majority of the previous literature on tandem queuing models we consider the case with GI service time at the first queue and with infinite buffers. The system can be described by an $M/G/1$-type Markov process at the departure epochs of the first queue. The main result of the paper is the steady-state vector generating function at the embedded epochs, which characterizes the joint distribution of the number of customers at both queues. The steady-state Laplace-Stieljes transform and the mean of the sojourn time of the customers in the system are also obtained.

We provide numerical examples and discuss the dependency of the steady-state mean of the sojourn time of the customers on several basic system parameters. Utilizing the structural characteristics of the model we discuss the interpretation of the results. This gives an insight into the behavior of this tandem queueing model and can be a base for developing approximations for it.

1. Introduction. Tandem queueing model consists of sequentially interconnected queues and it is a special case of open queueing networks. This model became great attention in the last decades, since it has several application areas. For example it is commonly used in telecommunication networks (see e.g. in [1]).

Tandem queueing models with infinite and finite buffers has huge literature. The vast majority of the literature deals with the case of Markovian service times, like e.g. the work [6] in which a tandem queueing model with Discrete-Time Markovian Arrival Process (D-MAP) and Phase Type (PH) service times is investigated. Another example is [8], which deals with a multi-server model with finite number of queues, MAP arrival and exponential service times. This model has no buffer at all. Numerous approximation methods have been developed for tandem queueing model with Markovian service times. One method is based on the decomposition of the system into single-server finite buffer stations ([5], [16]). Another approach is the product-form approximation of model with infinite buffers ([3]). Tandem queueing model with deterministic service times is analyzed by Boxma and Resing in [2].

2010 Mathematics Subject Classification. Primary: 60K25, 68M20; Secondary: 90B22.
Key words and phrases. Queueing theory, tandem queueing model, $M/G/1$-type Markov process, $G/M/1$-type Markov process.

The reviewing process of the paper was handled by Yutaka Takahashi as Guest Editor.
very general model with GI service time at the first server and with finite intermediate buffer is investigated and proposed as model of a call center by Dudin et al. in [4].

All of these works consider either specific (Markovian or deterministic) service time at the first queue or at least one finite buffer in the model.

In contrast to the above references, in this paper we consider the case with GI service time at the first queue and with infinite buffers. Analysis of infinite buffer system is somewhat simpler from theoretical point of view comparing to the finite buffer counterpart of the model. This rise the chance to get at least partly close-form results, whose interpretation can contribute to the deeper understanding of the behavior of the model. The considered system can be described by an M/G/1-type Markov chain embedded at the customer departure epochs of the first queue. We apply standard probability-generating function (PGF) techniques and matrix analytic methods for the analysis of the system.

The contribution of this paper is the analysis of the tandem queueing model $M/G/1 \rightarrow .M/1$. We derive the governing equation of the system in terms of vector generating functions (vector GF) and matrix generating functions. This relation characterizes the evolution of the system at the embedded epochs. The main result of the paper is the steady-state vector GF, which characterizes the joint distribution of the number of customers at both stations. Based on it we obtain also the steady-state probability vector of the number of customers at the second station and the vector factorial moments of the steady-state number of customers at the first station, both at the embedded epochs. The steady-state Laplace-Stieljes transform (LST) and the steady-state mean of the sojourn time of the customers in the system are also given.

We discuss the computational procedure and we provide numerical examples to demonstrate it and to have a short look into the dependency of the steady-state mean of the sojourn time of the customers on several basic system parameters. The interpretation of the results is also discussed. It turns out that at the embedded epochs the behavior of the second queue is closely related to two G/M/1-type Markov processes. Utilizing the specific structure of one of these Markov processes enables to interpret these results. This gives an insight into the behavior of this tandem queueing model, which can serve as a base for developing further approximations.

Recently the tandem queueing model was also applied as a model in the area of wireless networks ([14], [10]). Another application area is manufacturing systems and their optimal control ([15], [13]).

From the application point of view the importance of the current work lies in the ability of modeling more general service time of the first station and the deeper understanding of the behavior of this tandem queueing model due to the interpretation of the closed-form results.

The rest of the paper is organized as follows. In section 2 we introduce and explain the model. In section 3 we deal with the structural characteristics of the model. The steady-state number of customers at the embedded epochs is investigated in section 4. In section 5 we derive the steady-state sojourn time results. It is followed by the description of the computational procedure including also numerical examples in section 6. The discussion of the results and possible further step closes the paper in section 7.
2. Model description. We consider a tandem queueing model consisting of two stations. After getting service at first station each customer moves to the second one. Each of the stations has its own queue and one server unit. Both queues have infinite buffers, i.e. all customers arriving to the system will be served at both stations. The customers arriving to the first station according to Poisson process with rate $\lambda$. The customer service times at the first station are independent and identically distributed. $B, \tilde{B}(s), b, b^{(2)}$ denote the service time r.v., its LST and its first two moments, respectively. Each customer departure from the server unit of the first station implies an immediate customer arrival to the second station. The distribution of the customer service time at second station is exponential with parameter $\mu$. The customers during their staying at first station, i.e. during their arrival, waiting, service or departure, are called 1-customers. Similarly, the customers during their staying at second station, i.e. during their arrival, waiting, service or departure, are called 2-customers. According to this the naming of the customers moving from the first station to the second one changes from 1-customers to 2-customers.

We define D-epoch as the epoch just after the customer departure from the first station and just before the arrival of the same customer to the second station. We remark here that at the D-epoch the customer moving from the first station to the second one is neither 1-customer nor 2-customer. The different epochs and the naming of the customers are shown in Fig. 1.

We also define the 1-exhaustion epoch as the D-epoch at the start of the idle period at first station. The server utilization at the first station is $\rho_1 = \lambda b$. We also use the notations $\rho_2 = \frac{\lambda}{\mu}$ and $\rho_2^{\text{max}} = \frac{1}{b\mu}$. We call this tandem queueing model also as $M/G/1\rightarrow ./M/1$ tandem queueing model. The model is illustrated in Fig. 2.

On this tandem queueing model we impose the following assumptions:

A.1 The arrival rate and the mean customer service times at both stations are positive and finite, i.e. $0 < \lambda < \infty$, $0 < b < \infty$ and $0 < 1/\mu < \infty$.
A.2 The arrival process and each of the customer service times are mutually independent.
A.3 The service during the service period is work conserving at both stations.
A.4 The service during the service period is non-preemptive at both stations.
A.5 The customers are served in First-In-First-Out (FIFO) order.
A.6 The model is stable.

**Figure 1.** The different epochs and the naming of the customers
The stability condition of this tandem queueing model is well-known and it is given as

\[
\rho_1 < 1 \quad \text{and} \quad \rho_2 < 1. \tag{1}
\]

This can be explained as follows. The whole system is stable if and only if both stations are stable. The first station is an M/G/1 queue, therefore it is stable if and only if \( \rho_1 < 1 \) (see e.g. in [7]). The second station is stable if and only if the average arrival rate of the 2-customers, which is the same as the average output rate of the 1-customers, does not exceed the constant service rate at second station. However the average output rate of the 1-customers equals the arrival rate to the first station due to the stability of the first station. Thus the stability condition of the second station is given as \( \lambda < \mu \), which is equivalent to \( \rho_2 < 1 \).

When \( Y \) denotes a matrix, then \([Y]_{j,l}\) stands for its \( j,l \)-th element. Similarly \([y]_j\) denotes the \( j \)-th element of vector \( y \). When \( \hat{Y}(z) \), \( |z| \leq 1 \) is a matrix generating function, \( Y \) denotes its value at \( z = 1 \), i.e., \( Y = \hat{Y}(1) \). When \( \hat{y}(z) \), \( |z| \leq 1 \) is a vector GF, \( y^{(k)} \) denotes its \( k \)-th \((k \geq 1) \) factorial moment, i.e., \( y^{(k)} = \frac{d^k}{dz^k}\hat{y}(z)|_{z=1} \) and \( y \) denotes its value at \( z = 1 \), i.e., \( y = \hat{y}(1) \). We also use notation \( y^{(k)} = y^{(k)}e \). The notation \( eig(Y) \) stands for the eigenvalues of matrix \( Y \).

3. Structural characteristics of the system. In the first part of this section we give the description of the system by means of a bivariate Markov chain embedded at D-epochs. We introduce several matrices describing the evolution of the number of 2-customers during different intervals and several partial matrix PGFs. Some of these matrices serve as a building block of the probability transition matrix of the embedded Markov chain. This transition matrix has specific structure, since it is an M/G/1-type on block level and its block matrices are G/M/1-type. In the second part of the section we characterize the evolution of the system in terms of vector GFs and matrix generating functions. This leads to the governing equation of the system for the steady-state vector GF of the number of customers at both stations.

3.1. Description of the system at D-epochs. We describe the system and its evolution at D-epochs. Let \( N_1(\ell) \) and \( N_2(\ell) \) be the number of customers at the first and at the second station at the \( \ell \)-th D-epoch, for \( \ell > 0 \), respectively. The sequence \( \{(N_1(\ell),N_2(\ell)), \ell > 0\} \) is a bivariate homogeneous embedded Markov chain on the state space \( \{(0,1,\ldots),(0,1,\ldots)\} \). We say that the chain is in state \((i,k)\) when \( N_1(\ell) = i \) and \( N_2(\ell) = k \). We also say that the chain is in level \( i \) when \( N_1(\ell) = i \). Let \( p_{i,k}(j,n) \) denote the probability of transition from state \((i,k)\) to state \((j,n)\) in
this Markov chain, i.e.

\[ p_{i,k}(j,n) = P\{N_1(\ell + 1) = j, N_2(\ell + 1) = n \mid N_1(\ell) = i, N_2(\ell) = k\}, \]

\[ \ell \geq 1, \quad i, k, j, n \geq 0. \]

In the following we consider the structure of the transition probability matrix of the above defined Markov chain. We define \( \gamma_u \) as the probability of \( u \) 2-customer departures during the idle period of the first station. Due to the Poisson arrival process this idle period is exponentially distributed. Using it the probability \( \gamma_u \) can be expressed as

\[ \gamma_u = \int_0^\infty \frac{(\mu t)^u}{u!} e^{-\mu t} \lambda e^{-\lambda t} dt \quad u \geq 0. \]

Furthermore we define the probability \( \gamma^*_u \) as

\[ \gamma^*_u = \sum_{k=u}^{\infty} \gamma_k, \quad u \geq 0. \]

We define the \( \infty \times \infty \) matrix \( \Pi^- \) as

\[ \Pi^- = \begin{pmatrix}
\gamma^*_1 & \gamma_0 & 0 & 0 & \ldots \\
\gamma^*_2 & \gamma_1 & \gamma_0 & 0 & \ldots \\
\gamma^*_3 & \gamma_2 & \gamma_1 & \gamma_0 & \ldots \\
\gamma^*_4 & \gamma_3 & \gamma_2 & \gamma_1 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}. \]

Matrix \( \Pi^- \) describes the transition in number of 2-customers during the idle period of the first station.

It follows from the definitions of \( \gamma^*_u \) and \( \gamma_u \) that matrix \( \Pi^- \) is stochastic, since it is nonnegative and

\[ \Pi^- e = e, \tag{2} \]

where \( e \) is an \( \infty \times 1 \) column vector having all elements equal to one.

We define \( \beta_{r,u} \) as the joint probability of \( r \) 1-customer arrivals and \( u \) 2-customer departures during a 1-customer service time. The joint probability \( \beta_{r,u} \) can be expressed as

\[ \beta_{r,u} = \int_0^\infty \frac{(\lambda t)^r}{r!} e^{-\lambda t} \frac{(\mu t)^u}{u!} e^{-\mu t} dB(t) \quad r, u \geq 0. \]

Moreover we also define the joint probability \( \beta^*_r \) as

\[ \beta^*_r = \sum_{k=u}^{\infty} \beta_{r,k}, \quad r, u \geq 0. \]

We define the \( \infty \times \infty \) matrix \( \Pi^d \) as

\[ \Pi^d = \begin{pmatrix}
\beta^*_{r,0} & 0 & 0 & 0 & \ldots \\
\beta^*_{r,1} & \beta_{r,0} & 0 & 0 & \ldots \\
\beta^*_{r,2} & \beta_{r,1} & \beta_{r,0} & 0 & \ldots \\
\beta^*_{r,3} & \beta_{r,2} & \beta_{r,1} & \beta_{r,0} & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}. \]
The \((k, n)\)-th element of matrix \(\Pi_r^d\) describes the joint probabilities of \(r\) 1-customer arrival and the \(k \Rightarrow n\) transition in number of 2-customers during the first 1-customer service time after the idle period of the first station.

Let \(a_r\) denote the probability of \(r\) 1-customer arrivals during a 1-customer service time for \(r \geq 0\). Thus \(a_r\) is given by

\[
a_r = \int_0^\infty \frac{(\lambda t)^r}{r!} e^{-\lambda t} dB(t), \quad r \geq 0,
\]

and it can be expressed as

\[
a_r = \sum_{u=0}^\infty \beta^*_r,u = \beta^*_r,0, \quad r \geq 0. \tag{3}
\]

The definition of \(\beta^*_r,u\) and (3) imply that

\[
\Pi^d_r e = \sum_{u=0}^\infty \beta_r,u e = a_r e, \quad r \geq 0. \tag{4}
\]

Next we define the \(\infty \times \infty\) matrix \(\Pi_r^{(0)}\) as

\[
\Pi_r^{(0)} = \Pi - \Pi_r^d. \tag{5}
\]

The \((k, n)\)-th element of matrix \(\Pi_r^{(0)}\) describes the joint probabilities of the \(k \Rightarrow n\) transition in number of 2-customers from the 1-exhaustion epoch to the D-epoch at the end of the next 1-customer service and \(r\) 1-customer arrival during that 1-customer service. It can be seen from their definitions that matrix \(\Pi^d\) has G/M/1-type structure, while matrix \(\Pi^d_r\) is lower triangular. It follows from (5) that \(\Pi_r^{(0)}\) has also G/M/1-type structure.

Using (2), (4) and (5) yields

\[
\Pi_r^{(0)} e = a_r e, \quad r \geq 0.
\]

We also define the \(\infty \times \infty\) matrix \(\Pi_r^+\) as

\[
\Pi_r^+ = \begin{pmatrix}
\beta^*_r,1 & \beta^*_r,0 & 0 & 0 & \cdots \\
\beta^*_r,2 & \beta_r,1 & \beta^*_r,0 & 0 & \cdots \\
\beta^*_r,3 & \beta^*_r,2 & \beta_r,1 & \beta^*_r,0 & \cdots \\
\beta^*_r,4 & \beta^*_r,3 & \beta^*_r,2 & \beta_r,1 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
\end{pmatrix}.
\]

The \((k, n)\)-th element of matrix \(\Pi_r^+\) describes the joint probabilities of the \(k \Rightarrow n\) transition in number of 2-customers during a 1-customer service time and \(r\) 1-customer arrival during that 1-customer service. It can be seen from the structure of matrix \(\Pi^+_r\) that it has G/M/1-type structure.

By using the definition of \(\beta^*_r,u\) and (3) the row sums of \(\Pi_r^+\) can be expressed as

\[
\Pi_r^+ e = \sum_{u=0}^\infty \beta_r,u e = a_r e, \quad r \geq 0.
\]
We define the special Toeplitz matrix $\Omega_1$ as

$$
\Omega_1 = \begin{pmatrix}
0 & 0 & 0 & 0 & \ldots \\
1 & 0 & 0 & 0 & \ldots \\
0 & 1 & 0 & 0 & \ldots \\
0 & 0 & 1 & 0 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix},
$$

and the matrix $E_0$ as

$$
E_0 = \begin{pmatrix}
1 & 0 & 0 & 0 & \ldots \\
0 & 0 & 0 & 0 & \ldots \\
0 & 0 & 0 & 0 & \ldots \\
0 & 0 & 0 & 0 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}.
$$

Observe that, using also (3), these special matrices connect the matrices $\Pi_{r}^d$ and $\Pi_{r}^+$ as

$$
\Pi_{r}^d = \Omega_1 \Pi_{r}^+ + a_r E_0, \quad r \geq 0.
$$

Furthermore (5) and (6) imply that

$$
\Pi_{r}^{(0)} = \Pi^{-} \Omega_1 \Pi_{r}^{+} + a_r \Pi^{-} E_0, \quad r \geq 0.
$$

Let $\Pi$ be the transition probability matrix of the above defined Markov chain in terms of block matrices, in which the $(k,n)$-th element of the $(i,j)$-th block matrix equals $p_{i,k}(j,n)$ for $i, j, k, n \geq 0$. It follows from the interpretation of matrices $\Pi_{r}^{(0)}$ that the $(0,j)$-th block matrix of $\Pi$, for $j \geq 0$, equals $\Pi_{j}^{(0)}$.

The $(i,j)$-th block matrix of $\Pi$, for $i \geq 1$ and $j \geq 0$, describes the transitions, in which $j - i + 1$ customers arrive during a 1-customer service time. Hence the $(i,j)$-th block matrix of $\Pi$, for $i \geq 1$ and $j \geq 0$, equals $\Pi_{j-i+1}^{+}$. Therefore the matrix $\Pi$ is given as

$$
\Pi = \begin{pmatrix}
\Pi_{0}^{(0)} & \Pi_{1}^{(0)} & \Pi_{2}^{(0)} & \Pi_{3}^{(0)} & \ldots \\
\Pi_{0}^{+} & \Pi_{1}^{+} & \Pi_{2}^{+} & \Pi_{3}^{+} & \ldots \\
0 & \Pi_{0}^{+} & \Pi_{1}^{+} & \Pi_{2}^{+} & \ldots \\
0 & 0 & \Pi_{0}^{+} & \Pi_{1}^{+} & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}.
$$

**Remark 1.** The transition matrix of the Markov chain embedded at D-epochs, $\Pi$, has $M/G/1$-type structure on block level and its every block matrix has $G/M/1$-type structure.

We define the partial matrix PGF $\hat{\Pi}_{r}^{d}(z)$ as

$$
\hat{\Pi}_{r}^{d}(z) = \sum_{r=0}^{\infty} \Pi_{r}^{d} z^r, \quad |z| \leq 1.
$$
The definition of $\Pi^d$ implies that the matrix $\hat{\Pi}^d(z)$ can be expressed as

$$\hat{\Pi}^d(z) = \begin{pmatrix}
\hat{\beta}_0^*(z) & 0 & 0 & 0 & \ldots
\hat{\beta}_1^*(z) & \hat{\beta}_0(z) & 0 & 0 & \ldots
\hat{\beta}_2^*(z) & \hat{\beta}_1(z) & \hat{\beta}_0(z) & 0 & \ldots
\hat{\beta}_3^*(z) & \hat{\beta}_2(z) & \hat{\beta}_1(z) & \hat{\beta}_0(z) & \ldots
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix},$$

where

$$\hat{\beta}_u(z) = \sum_{r=0}^{\infty} \beta_{r,u} z^r, \quad u \geq 0, \ |z| \leq 1.$$ 

$$\hat{\beta}_u^*(z) = \sum_{r=0}^{\infty} \beta_{r,u}^* z^r, \quad u \geq 0, \ |z| \leq 1.$$ 

The partial PGF matrix $\hat{\Pi}^{(0)}(z)$ is defined as

$$\hat{\Pi}^{(0)}(z) = \Pi - \hat{\Pi}^d(z), \quad |z| \leq 1.$$ (8)

Using (8), (3) as well as several definitions gives

$$\hat{\Pi}^{(0)}(z)e = \Pi \cdot \hat{\Pi}^d(z)e = \Pi \cdot \sum_{u=0}^{\infty} \hat{\beta}_u(z)e$$

$$= \sum_{r=0}^{\infty} a_r z^r e = \tilde{B}(\lambda - \lambda z)e.$$ (9)

Setting $z = 1$ in (8) and (9) shows that matrix $\Pi^{(0)}$ is stochastic, since it is nonnegative and we have

$$\Pi^{(0)}e = e.$$ 

Furthermore we define the partial matrix PGF $\hat{\Pi}^+(z)$ as

$$\hat{\Pi}^+(z) = \sum_{r=0}^{\infty} \Pi^+_r z^r, \quad |z| \leq 1.$$ (10)

The matrix $\hat{\Pi}^+(z)$ can be expressed as

$$\hat{\Pi}^+(z) = \begin{pmatrix}
\hat{\beta}_1^*(z) & \hat{\beta}_0(z) & 0 & 0 & \ldots
\hat{\beta}_1^*(z) & \hat{\beta}_0(z) & 0 & 0 & \ldots
\hat{\beta}_2^*(z) & \hat{\beta}_1(z) & \hat{\beta}_0(z) & 0 & \ldots
\hat{\beta}_2^*(z) & \hat{\beta}_1(z) & \hat{\beta}_0(z) & 0 & \ldots
\hat{\beta}_3^*(z) & \hat{\beta}_2(z) & \hat{\beta}_1(z) & \hat{\beta}_0(z) & \ldots
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}.$$ 

Using (3) as well as several definitions yields

$$\hat{\Pi}^+(z)e = \sum_{u=0}^{\infty} \hat{\beta}_u(z)e = \sum_{r=0}^{\infty} a_r z^r e = \tilde{B}(\lambda - \lambda z)e.$$ (11)
Setting $z = 1$ in $\hat{\Pi}^+$ gives matrix $\Pi^+$. It follows from the definitions of $\Pi^+_r$, $\hat{\beta}_u(z)$ and $\hat{\beta}_u^*(z)$ that the matrix $\Pi^+$ can be expressed as

$$
\Pi^+ = \begin{pmatrix}
\beta_1^* & \beta_0 & 0 & 0 & \ldots \\
\beta_2^* & \beta_1 & \beta_0 & 0 & \ldots \\
\beta_3^* & \beta_2 & \beta_1 & \beta_0 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}.
$$

Setting $z = 1$ in (11) shows that matrix $\Pi^+$ is stochastic, i.e. its every element can be interpreted as probability and $\Pi^+ e = e$.

The definitions of $\hat{\Pi}^d(z)$, $\hat{\Pi}^+(z)$, $\Omega_1$ and $E_0$ imply that

$$
\hat{\Pi}^d(z) = \Omega_1 \hat{\Pi}^+(z) + \hat{\beta}_0^*(z) E_0.
$$

Using (8) leads to

$$
\hat{\Pi}^{(0)}(z) = \Pi^- \Omega_1 \hat{\Pi}^+(z) + \hat{\beta}_0^*(z) \Pi^- E_0.
$$

Finally setting $z = 1$ in (12) yields

$$
\Pi^{(0)} = \Pi^- \Omega_1 \Pi^+ + \Pi^- E_0.
$$

3.2. The evolution of the system. We define the $1 \times \infty$ row vector $q_i$, for $i \geq 0$, by its $k$-th element as

$$
[q_i]_k = \lim_{\ell \to \infty} P\{N_1(\ell) = i, N_2(\ell) = k\}, \text{ for } i, k \geq 0.
$$

The vector $q_i$ represents the steady-state joint probabilities of the number of customers at both stations for the case when the number of 1-customers in the system is $i$.

We also define the corresponding steady-state vector GF $q(z)$ as

$$
q(z) = \sum_{i=0}^{\infty} q_i z^i, \quad |z| \leq 1.
$$

We also introduce a notation for the corresponding steady-state PGF, which is the PGF of the number of 1-customers, as

$$
\tilde{q}(z) = q(z) e.
$$

Furthermore we define the $1 \times \infty$ row vector $m$ as the conditional steady-state probability vector of the number of customers at the second station at 1-exhaustion epoch, i.e. given that the number of 1-customers is 0.

This definition implies that the joint probability $[q_0]_k$ equals $[m]_k P\{\text{the first station is idle at a D-epoch}\}$ for $k \geq 0$. This probability equals the reciprocal of the mean number of customers served during a service period at station 1. This is $1/ \rho_1$, since the first station has an M/G/1 queue and its stochastic behavior is independent of the one of the second station. Using it we get the following relationship between $q_0$ and $m$

$$
q_0 = (1 - \rho_1) m.
$$
Theorem 3.1. The governing equation of the stable \( M/G/1 > ./M/1 \) tandem queueing model satisfying assumptions \( A.1 - A.6 \) is given in terms of \( \hat{q}(z) \) and \( m \) as

\[
\hat{q}(z)(zI - \hat{\Pi}^+(z)) = (1 - \rho_1)m(z\Pi^0 - I)\hat{\Pi}^+(z) + (1 - \rho_1)m^0(z)\Pi^0 \cdot E_0,
\]

where \( I \) denotes the \( \infty \times \infty \) identity matrix.

Proof. Writing the equilibrium equation of the Markov chain in partitioned form yields

\[
q_i = q_0\Pi^{(0)}_i + \sum_{k=1}^{i+1} q_k\Pi^+_i - k, \quad i \geq 0. \tag{16}
\]

Multiplying both sides of (16) by \( z^i \), summing over \( i \geq 0 \) and rearrangement leads to

\[
\sum_{i=0}^{\infty} q_i z^i = \sum_{i=0}^{\infty} q_0\Pi^{(0)}_i z^i + \sum_{i=0}^{\infty} z^i \sum_{k=1}^{i+1} q_k\Pi^+_i - k = q_0\Pi^{(0)}(z) + \frac{1}{z} \sum_{k=1}^{\infty} z^k q_k \sum_{i=k}^{\infty} z^{i-k} \Pi^+_i - k. \tag{17}
\]

Using the definitions of \( \hat{q}(z) \) and \( \hat{\Pi}^+(z) \) in (17) gives

\[
\hat{q}(z) = q_0\hat{\Pi}^{(0)}(z) + \frac{1}{z}(\hat{q}(z) - q_0\hat{\Pi}^+(z)). \tag{18}
\]

Applying (14) in (18) and rearranging it leads to

\[
\hat{q}(z)(zI - \hat{\Pi}^+(z)) = (1 - \rho_1)m(z\Pi^{(0)}(z) - \hat{\Pi}^+(z)). \tag{19}
\]

The statement of the theorem comes by applying (12) in (19) and rearranging it. \( \square \)

Remark 2. The structure of relation (19) is similar to that of the relation of the vector GF of the number of customers at customer departure epoch in the BMAP/G/1 queue (see e.g. in [11]). In the \( M/G/1 > ./M/1 \) tandem queueing model the number of 2-customers plays similar role as the phase of BMAP in the BMAP/G/1 queue.

4. The steady-state number of customers at D-epoch. In this section we derive several formulas starting from the governing equation of the model. The form of (15) implies that all of these formulas can be given in terms of \( m \). Therefore first we give a method for determining the vector \( m \). It enables the derivation of the steady-state vector GF of the number of customers at both stations, \( \hat{q}(z) \), which is the main result of the paper. Then the derivation of the steady-state probability vector of the number of 2-customers at D-epochs, \( q \), follows. In the last part of the section we provide also the expression of the vector factorial moments of the steady-state number of 1-customers at D-epochs, \( q^{(n)} \) for \( n \geq 1 \).
4.1. Determination of vector $\mathbf{m}$. We define matrix $\mathbf{G}$, whose $(k, \ell)$-th element is given as the probability that starting from state $(n+1, k)$ in the Markov chain defined in subsection 3.1 the first state visited in level $n$ is $(n, \ell)$, $n \in \{0, 1, 2, \ldots\}$, $k, \ell \geq 0$.

**Proposition 1.** In the stable $M/G/1 \rightarrow .M/1$ tandem queueing model satisfying assumptions A.1 - A.6 the vector $\mathbf{m}$ is uniquely determined by the following system of linear equations:

$$m \Pi^- \left( \Omega_1 \mathbf{G} + \mathbf{E}_0 \sum_{r=0}^{\infty} a_r \mathbf{G}^r \right) = m \quad \text{and} \quad \mathbf{me} = 1, \quad (20)$$

where $\mathbf{G}$ can be computed from the following matrix equation:

$$\mathbf{G} = \sum_{r=0}^{\infty} \Pi_r^+ \mathbf{G}^r. \quad (21)$$

**Proof.** The $M/G/1$-type structure of matrix $\Pi$ ensures that matrix $\mathbf{G}$ can be computed by applying standard matrix-analytic method (see e.g. in [9]). This leads to

$$\Pi_0^+ + \Pi_1^+ \mathbf{G} + \sum_{r=1}^{\infty} \Pi_{r+1}^+ \mathbf{G}^{r+1} = \mathbf{G}. \quad (22)$$

Rearranging (22) results the second statement of the theorem. The sum of elements of vector $\mathbf{m}$ is 1, since it is a conditional probability vector, i.e.

$$\mathbf{me} = 1. \quad (23)$$

In order to establish a relation for $\mathbf{m}$ we use again a standard matrix-analytic argument. The transition from a 1-exhaustion epoch to the next 1-exhaustion epoch brings vector $\mathbf{m}$ to itself. Let us consider the transitions in the number of customers at the second station during the idle period and the succeeding first customer service time at the first station. If the number of 1-customer arrivals during that first customer service time is $r$ then the above transition can be described by matrix $\Pi_r^{(0)}$. At the end of this transition there are $r$ number of customers at the first station, i.e. the Markov chain is in level $r$. Getting back from level $r$ to level 0, i.e. to idle state of the first station, requires $r$ consecutive level decreases. Each of these level decreases can be described by matrix $\mathbf{G}$. Counting for all possible values of $r$ results in a relation for vector $\mathbf{m}$ as

$$\mathbf{m} \sum_{r=0}^{\infty} \Pi_r^{(0)} \mathbf{G}^r = \mathbf{m}. \quad (24)$$

Applying (7) in (24) yields

$$m \Pi^- \left( \Omega_1 \sum_{r=0}^{\infty} \Pi_r^+ \mathbf{G}^r + \mathbf{E}_0 \sum_{r=0}^{\infty} a_r \mathbf{G}^r \right) = m. \quad (25)$$

Taking into consideration (23) and using (21) in (25) gives the first statement of the theorem. \qed
4.2. Steady-state vector GF of the number of customers.

Lemma 4.1. Let us consider the stochastic matrix $Y$ and a Markov chain, whose transition probabilities are given by $Y$. If the Markov chain is irreducible and aperiodic then the matrix $(I - Y + ey)$ is nonsingular, where $y$ is the steady-state probability vector of that Markov chain.

Proof. If the Markov chain is irreducible and aperiodic then the probability of each state of the Markov chain after $n$-step transition always converge to a limit as $n \to \infty$, i.e. the steady-state probability vector exists and it is unique. Furthermore the steady-state probability vector is a solution of the system of linear equations $yY = y$.

First we consider the case, when the Markov chain is positive recurrent. We apply an indirect argument. In this case $y$ is a distribution and it is uniquely determined by the system of linear equations $y(Y - I) = 0$ and $ye = 1$, where 0 is the row vector having all elements equal zero. Let us assume that matrix $(I - Y + ey)$ is nonsingular, which fulfills the homogenous system of linear equations $x(I - Y + ey) = 0$. Hence $x(I - Y + ey)e = 0$ also holds which together with $y \neq 0$ implies that $xe = 0$, i.e. the row sum of $x$ is zero. Using this it follows from the above homogenous system of linear equations that $xe = 0$, i.e. the row sum of $x$ is zero. However this a contradiction as it was shown that the row sum of $x$ is zero, therefore matrix $(I - Y + ey)$ must be nonsingular.

Now consider the other case when the Markov chain is null recurrent or transient. Solving the system of linear equations $yY = y$ by applying recursive substitution leads to $y = y \lim_{n \to \infty}(Y)^n$. If the Markov chain is null recurrent or transient then the $n$-step transition probabilities to each state of the chain always converge to zero as $n \to \infty$, i.e. $\lim_{n \to \infty}(Y)^n = 0$, where 0 is the zero matrix. It follows that the only solution of the homogenous system of linear equations $y(Y - I) = 0$ is $y = 0$ and thus the matrix $(Y - I)$ must be nonsingular. Due to $y = 0$ the matrices $(I - Y + ey)$ are the same, which implies the statement for this case.

We define matrix $\Pi^*(z)$ for $|z| \leq 1$ as

$$\Pi^*(z) = \frac{\hat{\Pi}^+(z)}{B(\lambda - \lambda z)}.$$  \hfill (26)

It follows from the definition of $\hat{\Pi}^+(z)$ and from (11) that $\frac{\hat{\Pi}^+(z)}{B(\lambda - \lambda z)}$ is stochastic for real $z$ in the region $0 < z \leq 1$.

Let $\pi^*(z)$ be the steady-state probability vector of the Markov chain whose transition matrix is stochastic matrix $\Pi^*(z)$. This steady-state probability vector exists and it is unique, since the above Markov chain has one irreducible class of aperiodic states. This can be seen from the structure of matrix $\Pi^*(z)$. The steady-state probability vector, $\pi^*(z)$, can be uniquely determined by

$$\pi^*(z)\Pi^*(z) = \pi^*(z) \quad \text{and} \quad \pi^*(z)e = 1.$$  \hfill (27)
We define also the matrices $\Psi(z)$ and $\Phi(z)$ as
\[
\Psi(z) = \frac{\bar{B}(\lambda - \lambda z) - z}{B(\lambda - \lambda z)} (I - \Pi^+(z) + e\pi^+(z))^{-1}
\]
\[
\Phi(z) = \frac{1}{B(\lambda - \lambda z)} (I - \Pi^+(z) + e\pi^+(z))^{-1}.
\] (28)

**Theorem 4.2.** In the stable $M/G/1 > .M/1$ tandem queueing model satisfying assumptions A.1 - A.6 the expression of the steady-state vector GF $\hat{q}(z)$ can be characterized as follows.

1. The steady-state vector GF $\hat{q}(z)$ is given as a sum of two terms.
   - The first term consists of multiplicative terms, where
     - the first multiplicative term, $\hat{q}(z)$, is the steady-state PGF of the number of 1-customers and
     - the second multiplicative term, $\pi^+(z)$, depends only on matrix GF $\Pi^+(z)$, which is the steady-state matrix GF of the number of customers at the customer arrival epochs of the virtual $G/M/1$ queue with interarrival time $B$ and mean service time $\frac{1}{\mu}$ (characterized by matrix $\Pi^+$).
   - The second term represents the dependency on two $G/M/1$-type structure matrix GFs, $\Pi^{(0)}(z)$ and $\Pi^+(z)$.

2. Furthermore the expression of $\hat{q}(z)$ is given as
\[
\hat{q}(z) = \hat{q}(z)\pi^+(z)(I - \Psi(z))^{-1} + (1 - \rho_1)(z\Pi^{(0)}(z) - \Pi^+(z))\Phi(z)(I - \Psi(z))^{-1},
\] (29)
where $\hat{q}(z)$ is the PGF of the number of customers in an $M/G/1$ queue, i.e. it is the Pollaczek-Khinchine transform formula.

**Proof.** We start from the governing equation of the system. Multiplying the (19) by $\frac{1}{B(\lambda - \lambda z)}$, adding $\hat{q}(z)(1 - \frac{z}{B(\lambda - \lambda z)}I)$ to both sides of it and using (26) leads to
\[
\hat{q}(z)(I - \Pi^+(z)) = \frac{\bar{B}(\lambda - \lambda z) - z}{B(\lambda - \lambda z)} \hat{q}(z)
\]
\[
+ \frac{1}{B(\lambda - \lambda z)} (1 - \rho_1)(z\Pi^{(0)}(z) - \Pi^+(z)).
\] (30)

We define the vector $\mathbf{r}(z)$ for the r.h.s. of (30) as
\[
\mathbf{r}(z) = \frac{-\bar{B}(\lambda - \lambda z) - z}{B(\lambda - \lambda z)} \hat{q}(z)
\]
\[
+ \frac{1}{B(\lambda - \lambda z)} (1 - \rho_1)(z\Pi^{(0)}(z) - \Pi^+(z)).
\] (31)

Applying (31) in (30) gives
\[
\hat{q}(z)(I - \Pi^+(z)) = \mathbf{r}(z).
\] (32)

Adding $\hat{q}(z)e\pi^+(z)$ to both sides of (32) yields
\[
\hat{q}(z)(I - \Pi^+(z) + e\pi^+(z)) = \hat{q}(z)e\pi^+(z) + \mathbf{r}(z).
\] (33)
Utilizing the irreducibility and aperiodicity of the Markov chain whose transition matrix is stochastic matrix \( \Pi^*(z) \), Lemma 4.1 can be applied. Hence the matrix 

\[
(I - \Pi^*(z) + e\pi^*(z))
\]

is nonsingular for real \( z \) in the region \( 0 < z \leq 1 \). Utilizing it we multiply both sides of (33) by 

\[
(I - \Pi^*(z) + e\pi^*(z))^{-1}
\]

to obtain

\[
\hat{q}(z) = \hat{q}(z)e\pi^*(z)(I - \Pi^*(z) + e\pi^*(z))^{-1} + \tilde{r}(z)(I - \Pi^*(z) + e\pi^*(z))^{-1}.
\]

Using \( \pi^*(z)(I - \Pi^*(z) + e\pi^*(z))^{-1} = \pi^*(z) \) and (13) in (34) leads to

\[
\hat{q}(z) = \hat{q}(z)e\pi^*(z) + \tilde{r}(z)(I - \Pi^*(z) + e\pi^*(z))^{-1}.
\]

We remark here that \( \hat{q}(z) \) is the PGF of the number of 1-customers in the system at D-epochs, which is the PGF of the number of customers in the \( M/G/1 \) queue due to the stochastic independency of the first station. Applying (31) in (35) and rearrangement results in

\[
\hat{q}(z) = \frac{\tilde{B}(\lambda - \lambda z) - z}{B(\lambda - \lambda z)}\hat{q}(z)(I - \Pi^*(z) + e\pi^*(z))^{-1}
\]

\[
+ \frac{1}{B(\lambda - \lambda z)}(1 - \rho_1)m(z\tilde{\Pi}^0_0(z) - \tilde{\Pi}^+_1(z)) (I - \Pi^*(z) + e\pi^*(z))^{-1}.
\]

This can be rewritten by using the matrices \( \Psi(z) \) and \( \Phi(z) \) as

\[
\hat{q}(z) = \hat{q}(z)(z\Psi(z)) + \hat{q}(z)e\pi^*(z) + (1 - \rho_1)m(z\tilde{\Pi}^0_0(z) - \tilde{\Pi}^+_1(z))\Phi(z).
\]

Applying recursive substitution of the above expression of \( \hat{q}(z) \) into the r.h.s. of (37) yields

\[
\hat{q}(z) = \hat{q}(z)\lim_{k\to\infty} (\Psi(z))^k
\]

\[
+ \sum_{k=0}^{\infty} \hat{q}(z)e\pi^*(z)(\Psi(z))^k
\]

\[
+ \sum_{k=0}^{\infty} (1 - \rho_1)m(z\tilde{\Pi}^0_0(z) - \tilde{\Pi}^+_1(z))\Phi(z)(\Psi(z))^k.
\]

In order to investigate the convergence of the infinite matrix sum \( \sum_{k=0}^{\infty} (\Psi(z))^k \) now we consider the eigenvalue of matrix \( \Psi(z) \) with the highest absolute value on a specific real interval. It follows from the non-singularity of matrix 

\[
(I - \Pi^*(z) + e\pi^*(z)) \text{ the homogenous equation } x(I - \Pi^*(z) + e\pi^*(z)) = 0 \text{ has no non-trivial solution. Hence the value 0 is not an eigenvalue of matrix } (I - \Pi^*(z) + e\pi^*(z)).
\]

Let \( \kappa^* \) denote the absolute value of the eigenvalue of matrix 

\[
(I - \Pi^*(z) + e\pi^*(z))^{-1}
\]

with the smallest absolute value. It follows that the absolute value of the eigenvalue of matrix 

\[
(I - \Pi^*(z) + e\pi^*(z))^{-1}
\]

with the highest absolute value is \( \frac{1}{\kappa^*} \). On the other hand \( \tilde{B}(\lambda - \lambda z) \) goes to 1 as real-valued \( 0 < z \leq 1 \) goes to 1. It implies that in a properly chosen \( (z_0, 1) \) interval the expression 

\[
\frac{\tilde{B}(\lambda - \lambda z) - z}{B(\lambda - \lambda z)}
\]

can take any small value. Therefore for a given positive real number \( \kappa^* \) there exists a real \( z_0 < 1 \) so that for every \( z_0 < z < 1 \) the relation 

\[
\frac{\tilde{B}(\lambda - \lambda z) - z}{B(\lambda - \lambda z)} < \kappa^*
\]

holds. Taking all these into account and using the definition of \( \Psi(z) \) we conclude that there exists a real \( z_0 < 1 \) so that
for every $z_0 < z < 1$ the absolute value of the eigenvalue of matrix $\Psi(z)$ with the highest absolute value is less than $\kappa^* \frac{1}{\kappa^*} = 1$. In other words

$$\exists z_0 < 1 \text{ for which } |\text{eig}(\Psi(z))| < 1, \quad z_0 < z < 1, \quad z_0, z \in \mathbb{R}.$$ 

It follows that for $z_0 < z < 1$ the infinite matrix sum $\sum_{k=0}^{\infty} (\Psi(z))^k$ converges to a finite matrix and thus the following relations hold:

$$\sum_{k=0}^{\infty} (\Psi(z))^k = (I - \Psi(z))^{-1},$$

$$\lim_{k \to \infty} (\Psi(z))^k = 0. \quad (39)$$

Applying (39) in (38) results in the expression of the statement, (29).

So far we showed the validity of the expression (29) for real $z$ in an open real interval $(z_0, 1)$. Now we apply the concept of the analytic continuation in the complex analysis to extend the validity scope of (29) to $|z| \leq 1$. Any PGF is an analytic complex function thus also $\hat{q}(z)$ is analytic. Moreover also the expression on the r.h.s. of (29) is an analytic complex function. It follows that they are the same in the whole convergence region of the vector GF $\hat{q}(z)$, hence the formula (29) is valid in the whole $|z| \leq 1$ region. \hfill $\Box$

4.3. Steady-state probability vector of the number of 2-customers. Let $\pi^+$ be the steady-state probability vector of the Markov chain, whose transition matrix is stochastic matrix $\Pi^+$. It can be uniquely determined by

$$\pi^+ \Pi^+ = \pi^+ \text{ and } \pi^+ e = 1.$$ 

**Corollary 1.** In the stable $M/G/1- > .M/1$ tandem queueing model satisfying assumptions A.1 - A.6 the expression of the steady-state probability vector $q$ can be characterized as follows.

1. The steady-state probability vector $q$ is given as a sum of two terms, in which
   - the first term is the steady-state probability vector of the number of customers at the customer arrival epochs of the virtual $G/M/1$ queue with interarrival time $B$ and mean service time $\frac{1}{\mu}$ (characterized by matrix $\Pi^+$) and
   - the second term represents the dependency on the difference of two $G/M/1$-type structure matrices, $(\Pi^{(0)} - \Pi^+)$. 

2. The steady-state probability vector of the number of customers at the customer arrival epochs of the above virtual $G/M/1$ queue, $\pi^+$, can be given in dependency of the stability of that virtual $G/M/1$ queue as
   - If $\rho_2^{\max} < 1$ then the above virtual $G/M/1$ queue is stable and the probability vector $\pi^+$ represents a geometrical distribution, for which
     $$[\pi^+]_k = (1 - \sigma_+) \sigma_+^k, \quad k \geq 0, \quad (40)$$
     where $\sigma_+$ is the only root of the equation
     $$\sigma_+ = \tilde{B} (\mu - \mu \sigma_+) \quad (41)$$
     in the region $0 < \sigma_+ < 1$. 

If $\rho_{2}^{\text{max}} \geq 1$ then the above virtual $G/M/1$ queue is unstable and the probability vector $\pi^{+}$ is given as

$$\pi^{+} = 0.$$  \hfill (42)

3. Furthermore the expression of $q$ is given by

$$q = \pi^{+} + (1 - \rho_{1})m(\Pi^{(0)} - \Pi^{+})(I - \Pi^{+} + e\pi^{+})^{-1}.$$  \hfill (43)

\textbf{Proof.} The third statement of the corollary comes by setting $z = 1$ in (29). The steady-state probability vector of the number of customers at the customer arrival epochs of the above virtual $G/M/1$ queue is $\pi^{+}$, since the transition matrix of the characterizing Markov chain is $\Pi^{+}$. The first statement of the theorem comes from the structure of (43) and from the above consideration. The second statement is well-known from the theory of standard $G/M/1$ queue (see e.g. in [7]). If $\rho_{2}^{\text{max}} \geq 1$, then the Markov chain whose transition matrix is stochastic matrix $\Pi^{+}$ becomes instable, i.e. it becomes either null recurrent or transient, which can be checked by applying Foster’s criterion on matrix $\Pi^{+}$.

\textbf{Remark 3.} As the mean length of the idle period of station 1 tends to 0 then matrix $\Pi^{(0)}$ becomes equal to matrix $\Pi^{+}$. As a consequence of it the second term in (43) representing the dependency on the difference of these matrices vanishes.

4.4. The vector factorial moments of the steady-state number of 1-customers. Let $q^{(0)}$ denote the value of $\widehat{q}(z)$ at $z = 1$, i.e. $q^{(0)} = \widehat{q}(1)$.

\textbf{Corollary 2.} In the stable $M/G/1-\rightarrow/M/1$ tandem queueing model satisfying assumptions A.1 - A.6 the vector factorial moments of the steady-state number of 1-customers is given as a sum of two terms, from which the first term is the multiplication of

- the $n$-th factorial moment of the number of customers in an $M/G/1$ queue corresponding to the first station of the model and
- the steady-state probability vector of the number of customers at the customer arrival epochs of the virtual $G/M/1$ queue with interarrival time $B$ and mean service time $\frac{1}{\mu}$ (characterized by matrix $\Pi^{+}$), $\pi^{+}$.

Furthermore the expression of $q^{(n)}$, for $n \geq 1$, is given by

$$q^{(n)} = q^{(n)} \pi^{+}$$

\hfill (44)

\begin{align*}
&+ \sum_{l=0}^{n-1} \binom{n}{l} q^{(l)} \pi^{*(n-l)} \\
&+ \sum_{k=0}^{n-1} \binom{n}{k} \sum_{l=0}^{k} \binom{k}{l} q^{(l)} \pi^{*(k-l)} \\
&\times \left( (-\lambda)^{n-k} B^{(n-k)(0)} I - \Pi^{+(n-k)} + \Gamma^{(n-k)} \right) \\
&\times (I - \Pi^{+} + e\pi^{+})^{-1} \\
&- \sum_{k=0}^{n-1} \binom{n}{k} q^{(k)} \left( \left(1_{0 \leq n-k < 2}\right) I - \Pi^{+(n-k)} + \Gamma^{(n-k)} \right) \\
&\times (I - \Pi^{+} + e\pi^{+})^{-1} \\
&+ (1 - \rho_{1}) m(\Pi^{(0)(n)} - \Pi^{+(n)})(I - \Pi^{+} + e\pi^{+})^{-1} \\
&+ n(1 - \rho_{1}) m\Pi^{(0)(n-1)}(I - \Pi^{+} + e\pi^{+})^{-1}.
\end{align*}
where \(1_{(\text{con})}\) denotes the indicator of condition “con” and for \(n \geq 1\)

\[
\Gamma^{(n)} = e \sum_{k=0}^{n} \binom{n}{k} \pi^{(k)}(-\lambda)^{n-k} \bar{B}^{(n-k)(0)}, \tag{45}
\]

\[
\pi^{(n)}(I - \Pi^+) = \sum_{k=0}^{n-1} \binom{n}{k} \pi^{(k)} \left( \Pi^{+(n-k)} - (-\lambda)^{n-k} \bar{B}^{(n-k)(0)} \right),
\]

as well as

\[
q^{(n)} = \left. \frac{d^n \bar{B}(\lambda - \lambda z) \left(1-\rho_1\right)(1-z)}{d z^n} \right|_{z=1}, \tag{46}
\]

i.e. \(q^{(n)}\) is the \(n\)-th derivative of the Pollaczek-Khinchine transform formula at \(z = 1\).

**Proof.** The corollary can be proved by taking the \(n\)-th derivative of (29) and setting \(z = 1\). \(\square\)

5. **Sojourn time.** In this section we derive several results for the steady-state sojourn time of a customer in the system. The section starts by introducing several definitions of different measures related to sojourn time of the customers in the system. This is followed by the derivation of the LST of the steady-state sojourn time of the customers in the system. Based on it also the steady-state mean sojourn time of the customers in the system is derived.

5.1. **Definitions.** The sojourn time of a 1-customer is defined as the time elapsed from its arrival to station 1 until its D-epoch. Let \(W^*_1,1\) denote the sojourn time of a 1-customer that arrives as the \(\ell\)-th into station 1, for \(\ell \geq 1\). We define the cumulative distribution function of the steady-state sojourn time of an 1-customer, \(W^*_1(t)\), as

\[
W^*_1(t) = \lim_{\ell \to \infty} Pr\{W^*_1,1 \leq t\}, \quad t \geq 0.
\]

The LST of the steady-state sojourn time of an 1-customer is defined as

\[
\bar{W}^*_1(s) = \int_{t=0}^{\infty} e^{-st} dW^*_1(t), \quad Re(s) \geq 0.
\]

Furthermore we define the sojourn time of a customer in the system as the time elapsed from its arrival to station 1 until its departure from station 2. Let \(W^*_\ell\) denote the sojourn time of a customer that arrives as the \(\ell\)-th into the system, for \(\ell \geq 1\). We define the cumulative distribution function of the steady-state sojourn time of a customer in the system, \(W^*(t)\), as

\[
W^*(t) = \lim_{\ell \to \infty} Pr\{W^*_\ell \leq t\}, \quad t \geq 0.
\]

The LST of the steady-state sojourn time of a customer in the system is defined as

\[
\bar{W}^*(s) = \int_{t=0}^{\infty} e^{-st} dW^*(t), \quad Re(s) \geq 0.
\]

The sojourn time of a 2-customer is defined as the time elapsed from the D-epoch just before its arrival to station 2 until its departure from station 2. Let \(W^*_\ell,2\) denote the sojourn time of a 2-customer that arrives as the \(\ell\)-th into station 2, for \(\ell \geq 1\).
Using that the customer arriving as the \( \ell \)-th into station 1 arrives to station 2 also as the \( \ell \)-th, \( W^*_\ell \) can be given as
\[
W^*_\ell = W^*_{\ell,1} + W^*_{\ell,2}, \quad \ell \geq 1.
\]

5.2. LST of the sojourn time. We condition the steady-state sojourn times \( W^*_{\ell,1} \) and \( W^*_{\ell,2} \) on the number of 1-customers at the \( \ell \)-th D-epoch. Thus we define the conditional cumulative distribution function of the steady-state sojourn time of a 1-customer, given that the number of 1-customers at the D-epoch of that 1-customer is \( n \) as
\[
W^*_1(t|n) = \lim_{\ell \to \infty} \Pr\{W^*_{\ell,1} \leq t | N_1(\ell) = n\}, \quad t \geq 0, \quad n \geq 0.
\]

The corresponding conditional LST is defined as
\[
\tilde{W}^*_1(s|n) = \int_0^\infty e^{-st}dW^*_1(t|n), \quad Re(s) \geq 0, \quad n \geq 0.
\]

Furthermore we define also the steady-state joint probability \( W^*_1(t,n) \) and its LST as
\[
W^*_1(t,n) = \lim_{\ell \to \infty} \Pr\{W^*_{\ell,1} \leq t, N_1(\ell) = n\}, \quad t \geq 0, \quad n \geq 0,
\]
and
\[
\tilde{W}^*_1(s,n) = \int_0^\infty e^{-st}dW^*_1(t,n), \quad Re(s) \geq 0, \quad n \geq 0.
\]

It follows from the definition of the conditional probability that the LST \( \tilde{W}^*_1(s|n) \) can be expressed as
\[
\tilde{W}^*_1(s|n) = \frac{\tilde{W}^*_1(s,n)}{P_1(n)}, \quad (47)
\]
where \( P_1(n) \) is the steady-state probability that the number of 1-customers at a D-epoch is \( n \) and it is given as
\[
P_1(n) = \lim_{\ell \to \infty} \Pr\{N_1(\ell) = n\}, \quad n \geq 0.
\]

Similarly we define the conditional cumulative distribution function of the steady-state sojourn time of a 2-customer, given that the number of 2-customers at the D-epoch just before the arrival of that 2-customer is \( n \) as
\[
W^*_2(t|n) = \lim_{\ell \to \infty} \Pr\{W^*_{\ell,2} \leq t | N_2(\ell) = n\}, \quad t \geq 0, \quad n \geq 0.
\]

Again the corresponding conditional LST can be defined as
\[
\tilde{W}^*_2(s|n) = \int_0^\infty e^{-st}dW^*_2(t|n), \quad Re(s) \geq 0, \quad n \geq 0.
\]

**Theorem 5.1.** In the stable \( M/G/1-\>.,/M/1 \) tandem queueing model satisfying assumptions A.1 - A.6 the LST of the steady-state sojourn time of the customers in the system is given as
\[
\tilde{W}^*(s) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \left| q_{n_1n_2} \right| \int_0^\infty \frac{(\lambda t)^{n_1}}{n_1!}e^{-\lambda t}e^{-st}dW^*_1(t) \int_0^\infty \frac{(\lambda t)^{n_2}}{n_2!}e^{-\lambda t}dW^*_1(t)
\]
\[
\quad \times \left( \frac{\mu}{s+\mu} \right)^{n_2+1}, \quad (48)
\]
where $W_1^*(t)$ can be obtained by inverse LST transform of $\tilde{W}_1^*(s)$, which is given as
\begin{equation}
\tilde{W}_1^*(s) = \frac{s(1-\rho_1)}{s - \lambda + \lambda B(s)} \tilde{B}(s).
\end{equation}

**Proof.** If both the number of 1-customers and the number of 2-customers at D-epoch are given then both the sojourn time of the 1-customers and the sojourn time of the 2-customers can be determined independently of each other. Let $n_1$ and $n_2$ be the number of 1-customers and the number of 2-customers both at D-epoch, respectively. Then the LST of the sojourn time of the 1-customers and the LST of the sojourn time of the 2-customers is $\tilde{W}_1^*(s|n_1)$ and $\tilde{W}_2^*(s|n_2)$, respectively. Utilizing the independency of these conditional LSTs and unconditioning leads to
\begin{equation}
\tilde{W}^*(s) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} [q_{n_1}]_{n_2} \tilde{W}_1^*(s|n_1) \tilde{W}_2^*(s|n_2).
\end{equation}

If the number of 2-customers at D-epoch just before the arrival of the tagged 2-customer is $n_2$ then the sojourn time of the tagged 2-customer equals the sum of the service times of the $n_2$ 2-customers in the second station and the service time requirements of the tagged 2-customer. Note that due to the exponential service time at the second station also the 2-customer under service needs a full service time. Thus the conditional LST of the sojourn time of the 2-customers, given that the number of 2-customers at D-epoch just before the arrival of the tagged 2-customer is $n_2$, can be given as
\begin{equation}
\tilde{W}_2^*(s|n_2) = \left(\frac{\mu}{s+\mu}\right)^{n_2+1}.
\end{equation}

Using (51) and (47) in (50) yields
\begin{equation}
\tilde{W}^*(s) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} [q_{n_1}]_{n_2} \tilde{W}_1^*(s, n_1) P_1(n_1) \left(\frac{\mu}{s+\mu}\right)^{n_2+1}.
\end{equation}

Utilizing that the sojourn time of a 1-customer departing from station 1 is independent of the number of 1-customers arriving during it, $\tilde{W}_1^*(s, n)$ can be given by
\begin{equation}
\tilde{W}_1^*(s, n_1) = \int_{t=0}^{\infty} \frac{(\lambda t)^{n_1}}{n_1!} e^{-\lambda t} e^{-st} dW_1^*(t).
\end{equation}

Here $W_1^*(t)$ is the inverse LST transform of $\tilde{W}_1^*(s)$, which is the LST of the M/G/1 queue, since the stochastic behavior of the first station is independent of the one of the second one. This can be given by means of the Pollaczek-Khinchine transform formula (see e.g. in [7]), which results in (49).

Setting $s = 0$ in (53) gives $P_1(n_1)$ as
\begin{equation}
P_1(n_1) = \int_{t=0}^{\infty} \frac{(\lambda t)^{n_1}}{n_1!} e^{-\lambda t} dW_1^*(t).
\end{equation}

The statement of the theorem comes by applying (53) and (54) in (52). \qed
5.3. Mean sojourn time. We define the steady-state mean sojourn times of the 1-customers \( E[W^*_1] \), of the 2-customers \( E[W^*_2] \) and the customers in the system \( E[W^*] \) respectively. Furthermore we define the conditional steady-state mean sojourn time of a 2-customer, given that the number of 2-customers at the D-epoch just before the arrival of that 2-customer is \( n \), as

\[
E[W^*_2|n] = \lim_{\ell \to \infty} E[W^*_2|N_2(\ell) = n], \quad n \geq 0.
\]

**Corollary 3.** In the stable \( M/G/1− > ./M/1 \) tandem queueing model satisfying assumptions A.1 - A.6 the steady-state mean sojourn time of the customers in the system is given as

\[
E[W^*] = b + \frac{\lambda b^{(2)}}{2(1 - \rho_1)} + \frac{1}{\mu} \left( 1 + \sum_{n_2=0}^{\infty} n_2[q] n_2 \right).
\]

**Proof.** Taking the first derivative of (50) at \( s = 0 \) yields

\[
E[W^*] = E[W^*_1] + \sum_{n_2=0}^{\infty} [q] n_2 E[W^*_2|n_2].
\]

The the steady-state mean sojourn time of the 1-customers is the steady-state mean sojourn time in the standard M/G/1 queue, which is given (see e.g. [7]) as

\[
E[W^*_1] = b + \frac{\lambda b^{(2)}}{2(1 - \rho_1)}.
\]

Given that the number of 2-customers at a D-epoch is \( n_2 \geq 0 \), the conditional mean sojourn time in the second station is the sum of \( n_2 \) times the mean service time of a 2-customer and the service time of the actually arriving 2-customer. Therefore \( E[W^*_2|n_2] \) can be given by

\[
E[W^*_2|n_2] = \frac{n_2 + 1}{\mu}.
\]

The corollary comes by applying (57) and (58) in (56) and rearranging it. \( \Box \)

**Remark 4.** Applying Little’s law in (55) results in the expression of the steady-state mean total number of customers, i.e. the sum of the mean number of 1-customers and 2-customers, in the system, \( E[Q] \), as

\[
E[Q] = \rho_1 + \frac{\lambda^2 b^{(2)}}{2(1 - \rho_1)} + \rho_2 \left( 1 + \sum_{n_2=1}^{\infty} n_2[q] n_2 \right).
\]

Note that the steady-state total number of customers differs from the steady-state total number of customers at D-epochs, since the number of 2-customers at an arbitrary epoch and at D-epochs are different. This difference is also reflected in the last term of (59).
6. The numerical solution. In this section we deal with the computational procedure. We consider how to keep the numeric computation tractable and provide the steps of the numeric computation of the steady-state mean sojourn time of this model. We describe how we model the GI service time by Coaxian-2 distribution. In the last part of this section we present numerical examples, which demonstrates the computational procedure and show the dependency of the steady-state mean sojourn time on several system parameters.

6.1. Numerical considerations. To keep the computation of the infinite dimensional vectors and matrices tractable we apply an upper limit $X$ on the number of 2-customers in the system, i.e. $k \leq X$, and an upper limit on the number of 1-customer arrivals during a 1-customer service time, i.e. $r \leq R$. This results in finite number of elements in these vectors and matrices as well as finite number of matrices $\Pi^+_r$ and $\Pi_r^{(0)}$, which results in finite number of required operations among others in computing matrix $G$ from (21) and vector $m$ from (20). The proper values of $X$ and $R$ depend on the required precision and can be determined in an iterative manner until the difference of consecutive values of e.g. the probabilities $[q]_k$, for every $k \leq X$, becomes less than the specified error. The detailed investigation of this and other existing truncation methods and the evaluation of their impact to the computational procedure is out of scope of this work and it could be a topic of future research.

6.2. The steps of the computation. Taking into account the above numerical considerations in the computational procedure, the steps of the computation of the steady-state mean sojourn time of this model can be summarized as follows:

1. Determination of $\sigma_+$, the only root of the equation (41) in the region $0 < \sigma_+ < 1$.
2. Calculation of matrix $G$ by means of successive substitution in the nonlinear matrix equation equation (21) starting with $G = 0$.
3. Computation of vector $m$ from the system of linear equations (20).
4. Calculation of the steady-state probability vector $q$ by applying (40) and (43).
5. Calculation of the steady-state mean sojourn time of the customers in the system, $E[W^*]$, by applying (55).

6.3. Modeling GI service time. The distribution of the GI service time is modeled by a Coaxian-2 distribution (see Fig. 3). This distribution model consists of two exponentially distributed phases with parameters $\mu_1$ and $\mu_2$. After expiry of the first phase the model either takes the second phase with probability $p_1$ or the modeling ends with probability $1 - p_1$.

![Figure 3. The Coaxian-2 distribution model](image-url)
To determine the parameters of the Coaxian-2 model from the mean \((b)\) and coefficient of variation \((c)\) of the GI service time we apply the following simple two-moment fit (see [12]):

\[
\mu_1 = \frac{2}{b}, \quad p_1 = \frac{0.5}{c_b}, \quad \mu_2 = \mu_1 p_1. \tag{60}
\]

The above Coaxian-2 model can realize distributions with coefficient of variation in the range \(c_b \geq 0.5\).

We selected the Coaxian-2 distribution for the numerical examples due to its generality and simplicity. However we remark here that the results for the M/G/1−M/1 tandem queueing model presented in this paper are valid for any GI service time distribution.

6.4. Numerical examples. In this subsection we provide examples for the numerical solution. The default values of the system parameters used for these examples are \(\rho_1 = 0.5, \quad \rho_2 = 0.3, \quad b = 1, \quad c_b = 1\). The values of \(\lambda\) and \(\mu\) are implicitly determined by \(\rho_1, b\) and \(\rho_2\).

In Fig. 4 we have plotted the steady-state mean sojourn time of the customers \((E[W^*])\), of the 1-customers \((E[W^*_1])\) and of the 2-customers \((E[W^*_2])\) as a dependency of \(\rho_1\) besides of keeping \(\rho_2\) fixed. This is achieved by simultaneous changing of \(\lambda\) and \(\mu\).

It can be seen on the figure that increasing \(\rho_1\) leads to higher \(E[W^*_1]\), which is clear from the first term of (55). On the other hand higher \(\rho_1\) results in lower \(E[W^*_2]\). One reason for it is that higher \(\rho_1\) is achieved by higher \(\mu\), which causes the second term in (55) \((= E[W^*_2])\) directly to be less, since \(\mu\) arises in its denominator. Another reason is that higher the \(\rho_1\) lower the weight of the second term in (43), \((1 - \rho_1)\), causing less probabilities in \(q\), which also results in less \(E[W^*_2]\) according to its expression. Hence the dominant term in \(E[W^*]\) is \(E[W^*_2]\) for lower values of \(\rho_1\), while for higher values of \(\rho_1\) the \(E[W^*]\) is dominated by \(E[W^*_1]\).
In Fig. 5 we have plotted the steady-state mean sojourn time of the customers \(E(W^{*})\), of the 1-customers \(E(W^{*}_1)\) and of the 2-customers \(E(W^{*}_2)\) as a dependency of \(\rho_2\) besides of keeping \(\rho_1\) fixed. This is achieved by changing of \(\mu\).

It is not surprising that changing \(\rho_2\) has no effect on \(E(W^{*}_1)\), since it depends only on \(\rho_1\), which is kept fixed. Thus for higher values of \(\rho_2\), as the figure shows, \(E(W^{*})\) is dominated by \(E(W^{*}_2)\). Higher value of \(\rho_2\) is achieved by less \(\mu\), which causes \(E(W^{*}_2)\) directly to be higher, since \(\mu\) arises in the denominator of this term.

In Fig. 6 we have plotted the steady-state mean sojourn time of the customers \(E(W^{*})\), of the 1-customers \(E(W^{*}_1)\) and of the 2-customers \(E(W^{*}_2)\) as a dependency of the coefficient of variation of \(B\).
It can be seen on the figure that the effect of $c_b$ on $E[W^*_2]$ is marginal. Therefore the dependency of $E[W^*_2]$ on $c_b$ is dominated by the dependency of $E[W^*_1]$ on $c_b$. This dependency is clear as $E[W^*_1]$, the first term of (55), directly depends on $c_b$ via the second moment $b^{(2)}$.

7. Conclusion. In this section we discuss the most important results of the paper ((29), (43) and (55)) and identify a potential future research work.

The expression of the steady-state vector GF $\hat{q}(z)$ can be interpreted as follows. The first term in (29), $\hat{q}(z)\pi^+(z)(I - \Psi(z))^{-1}$, represents the dependency on a queuing model consisting of two independent queues, the M/G/1 queue of the first station and the virtual G/M/1 queue characterized by matrix $\Pi^+$. The M/G/1—> ./M/1 tandem queueing model tends to this queuing model when $\rho_1 \to 1$, since in this case the arrival process of the second station becomes a GI process with interarrival time $B$. The second term in (29) incorporates the evolution of number of customers also in the case when the interarrival time to station 2 is eventually stretched due to idle periods of station 1, whose description requires also $\Pi^0$ besides of $\Pi^+$. Thus this term represents the dependency on the bias from the above limiting queueing model. Lower the $\rho_1$ higher the weight of this term, since the probability of idle of station 1, $(1 - \rho_1)$, is also higher. Hence the (29) expression of the steady-state vector GF $\hat{q}(z)$ can be interpreted as characterizing the behavior of the $M/G/1$—> ./M/1 tandem queueing model by a sum of dependencies on a limiting queueing model and on a bias from it.

A similar interpretation can be given to the expression of the steady-state probability vector $q$. The first term, $\pi^+$, and the second term in (43) represents the probability vector in the virtual G/M/1 queue characterized by matrix $\Pi^+$ and the dependency on the bias from this limiting queueing model, respectively. Again lower the $\rho_1$ higher the weight of the second term, since the probability of idle of station 1, $(1 - \rho_1)$, is also higher. Therefore the (43) expression of the steady-state probability vector $q$ can be interpreted as characterizing the behavior of the second station of the $M/G/1$—> ./M/1 tandem queueing model by a sum of dependencies on a limiting queue and on a bias from it. This expression shows explicitly the dependency on the stability of the virtual G/M/1 queue. Besides of the instability ($\rho_2^{\text{max}} \geq 1$) of this queue, the whole model (including also the bias from this limiting queue) remains stable (as far as $\rho_1 < 1$ and $\rho_2 < 1$), since the queue recovers itself during the idle periods of the first station. However in this case due to the instability of the virtual G/M/1 queue its contribution to the number of 2-customers vanishes, since the probability of having any finite number of customers in its queue tends to 0.

The expression of the steady-state mean sojourn time of the customers in the system, $E[W^*]$ can be simply interpreted as the sum of two terms, from which the first one is the steady-state mean sojourn time of the 1-customers ($b + \frac{b^{(2)}}{2(1 - \rho_1)}$) and the second one is the steady-state mean sojourn time of the 2-customers. This is a consequence of the definition that the sojourn time of the customers in the system is the sum of the sojourn time of the 1-customers and the sojourn time of the 2-customers. The (55) expression shows that the second term is weighted by the factor $\frac{1}{\mu}$. Hence in case of high value of $\mu$ the contribution of the second term to $E[W^*]$ tends to vanish. We remark here that numerical solution is necessary only for computing vector $q$ in this second term.
Table 1 gives an overview about the behavior of the $M/G/1 \rightarrow .M/1$ tandem queueing model in limit ranges of the system parameters in terms of the necessary terms to compute $E[W^*]$. 

| Limit ranges | First term in (55) | Second term in (55) |
|--------------|-------------------|-------------------|
| $\mu \gg 1$  | $b + \frac{\lambda b^{(2)}}{\pi (1 - \rho_1)}$ | $\frac{1}{\mu} \left( 1 + \sum_{n_2=0}^{\infty} n_2 [q]_{n_2} \right)$ |
| $\mu \ll 1$ and $\rho_1 \rightarrow 1$ and $\rho_2^{\text{max}} < 1$ | + | - |
| $\mu \ll 1$ and $\rho_1 < 1$ and $\rho_2^{\text{max}} > 1$ | + | + |
| $\mu \ll 1$ and $\rho_1 < 1$ and $\rho_2^{\text{max}} < 1$ | + | + |

**Table 1.** Necessary terms to compute $E[W^*]$ in different limit ranges of the system parameters

The case $\mu \ll 1$ and $\rho_1 \rightarrow 1$ and $\rho_2^{\text{max}} > 1$ is not considered, since it leads to an unstable system. This is because $\rho_1 \rightarrow 1$ and $\rho_2^{\text{max}} > 1$ implies $\rho_2 > 1$, which follows from the following relation:

$$\rho_2^{\text{max}} = \frac{1}{b\mu} = \frac{\lambda}{\lambda b\mu} = \frac{\lambda}{\rho_1 \mu} = \frac{\rho_2}{\rho_1}.$$

Based on the behavior of the $M/G/1 \rightarrow .M/1$ tandem queueing model in limit ranges of the system parameters presented in Table 1 an approximation can be developed to compute $E[W^*]$ with less computational effort. This is a potential future research task.

**REFERENCES**

[1] D. Bertsekas and R. Gallager, “Data Networks,” 2nd Edition, Prentice Hall, 1991.
[2] O. J. Boxma and J. A. C. Resing, *Tandem queues with deterministic service times*, Annals of Operations Research, 49 (1994), 221–239.
[3] G. Casale, P. G. Harrison and M. Vigliotti, *Product-Form Approximation of Tandem Queues via Matrix Geometric Methods*, in “6th International Workshop on the Numerical Solution of Markov Chains (NSMC 2010),” (2010).
[4] A. N. Dudin, C. S. Kim, V. I. Klimenok and O. S. Taramin, *A dual tandem queueing system with a finite intermediate buffer and cross traffic*, in “5th International Conference on Queueing Theory and Network Applications (QTN 2010),” Beijing, (2010), 102–109.
[5] A. Heindl, *Decomposition of general tandem queueing networks with mmpp input*, Performance Evaluation, 44 (2001), 5–23.
[6] B. Van Houdt and A. S. Alfa, *Response time in a tandem queue with blocking, Markovian arrivals and phase-type services*, Operations Research Letters, 33 (2005), 373–381.

[7] L. Kleinrock, “Queuing Systems. Vol I: Theory,” John Wiley, 1975.

[8] V. Klimenok, A. Dudin and V. Vishnevsky, *On the stationary distribution of tandem queue consisting of a finite number of stations*, in “Computer Networks, Communications in Computer and Information Science,” 201 (2012), 383–392.

[9] G. Latouche and V. Ramaswami, “Introduction to Matrix Geometric Methods in Stochastic Modeling, ASA-SIAM Series on Statistics and Applied Probability,” Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; American Statistical Association, Alexandria, VA, 1999.

[10] L. Le and E. Hossain, *Tandem queue models with applications to QoS routing in multihop wireless networks*, IEEE Trans. on Mobile Computing, 7 (2008), 1025–1040.

[11] D. L. Lucantoni, *New results on the single server queue with a batch markovian arrival process*, Comm. Statist. Stochastic Models, 7 (1991), 1–46.

[12] R. A. Marie, *Calculating equilibrium probabilities for \( \lambda(n)/C_k/1/N \) queue*, in “Proceedings Performance’80,” Toronto, (1980), 117–125.

[13] D. G. Pandelis, *Optimal control of flexible servers in two tandem queues with operating costs*, Probability in the Engineering and Informational Sciences, 22 (2008), 107–131.

[14] Y. E. Sagduyu and A. Ephremides, *Network Coding in Wireless Queuing Networks: Tandem Network Case*, in “Proc. IEEE International Symposium on Information Theory,” Seattle, WA, (2006).

[15] M. van Vuuren, “Performance Analysis of Manufacturing Systems: Queueing Approximations and Algorithms,” Ph.D thesis, Eindhoven University of Technology, Department of Mathematics and Computer Science, 2007.

[16] M. van Vuuren, I. J. B. F. Adan and A. E. Resing-Sassenb, *Performance analysis of multi-server tandem queues with finite buffers and blocking*, OR Spektrum, 27 (2005), 315–338.

Received September 2012; 1st revision January 2013; final revision June 2013.

E-mail address: safferzs@hit.bme.hu
E-mail address: yue@konan-u.ac.jp