A TWO-PRIORITY SINGLE SERVER RETRIAL QUEUE WITH ADDITIONAL ITEMS

1DHANYA SHAJIN

1 Department of Mathematics
S. N. College, Chempazhanthy
Trivandrum, Kerala-695587, India

2,3 A. N. Dudin*, 2,3 Olga Dudina and 4A. Krishnamoorthy

2 Belarusian State University, 4, Nezavisimosti Ave., Minsk, 220030, Belarus
3 Peoples Friendship University of Russia, 6 Miklukho-Maklaya St, Moscow, 117198, Russia
4 Department of Mathematics, CMS College, Kottayam-686001, India

(Communicated by Wuyi Yue)

Abstract. In this paper, we study a priority queueing-inventory problem with two types of customers. Arrival of customers follows Marked Markovian arrival process and service times have phase-type distribution with parameters depending on the type of customer in service. For service of each type of customer, a certain number of additional items are needed. High priority customers do not have waiting space and so leave the system when on their arrival a priority 1 customer is in service or the number of available additional items is less than the required threshold. Preemptive priority is assumed. Type 2 customers, encountering a busy server or idle with the number of available additional items less than a threshold, go to an orbit of infinite capacity to retry for service. The customers in orbit are non-persistent: if on retrial the server is found to be busy/idle with the number of additional items less than the threshold, this customer abandons the system with certain probability. Such a system represents an accurate enough model of many real-world systems, including wireless sensor networks and system of cognitive radio with energy harvesting and healthcare systems. The probability distribution of the system states is computed, using which several of the characteristics are derived. A detailed numerical study of the system, including the analysis of the influence of the threshold, is performed.

1. Introduction. In classical retrial queue service is rendered to customers without regard to factors such as items required for providing service. This is equivalent to assuming an abundance of availability of such items at the service facility. However, if the availability of the additional items is not always guaranteed, then the

2020 Mathematics Subject Classification. Primary: 60K25, 90B05; Secondary: 68M20, 90B22.
Key words and phrases. Markovian arrival process (MAP), marked Markovian arrival process (MMAP), phase-type distribution, additional items, retrial queue.

The first author is supported by Kerala State Council for Science, Technology & Environment: 001-07/PDF/2016/KSCSTE in Department of Mathematics, CMS College, Kottayam-686001, India.
The second and third authors are supported by “RUDN University Program 5-100”.
The fourth author is supported by UGC No.F.6-6/2017-18/EMERITUS-2017-18-GEN-10822 (SA-II) and DST project INT/RUS/RSF/P-15.
* Corresponding author: A. N. Dudin.
customers will have to either wait or quit the system. This situation prevails in service facilities such as hospitals, automobile service stations, wholesale and retail shops and so on.

Motivated by the above-indicated real-life situations, in this paper we consider a retrial queue with a single server. High priority arriving customers get into service preempting the service of the low priority customer, provided the one in service is of low priority and the minimum required number of additional items for the service of the former is available. The preempted low priority customer is lost to the system. Additional items used for the service of a customer cannot be reused. They are discarded. These additional items are supplied to the system according to a Markovian arrival process ($MAP$). The arrival of customers to the system is according to a marked Markovian arrival process ($MMAP$). Low priority customers in the external arrival stream will join for service immediately provided the server is idle and a minimum level of additional items is available in stock; else such customers are directed to an orbit of infinite capacity from which they try to access the server with inter-retrial time exponentially distributed. Such retrial customers have a tendency to abandon the system with certain probability whenever the server is found to be busy/server idle with less than the minimum level of additional items. With complimentary probability, the customer gets back to the orbit. This process of abandonment of the system by low priority customers on unsuccessful retrial, coupled with linear retrial makes the system stable if the mentioned probability is greater than zero.

We give three real-life examples of the problem being investigated.

1. **Wireless sensors with energy harvesting:** Let us consider a sensor that monitors some area or the state of some equipment. Modern sensors have an opportunity to harvest energy from the outside (e.g., using solar, radio-frequency or wind energy) and store it. It is assumed that such stored energy is divided into energy units (that correspond to additional items in our model). Several energy units are required for transmission of one information unit. We assume that the sensor transmits to the central node two types of information units. Type 1 information unit (type 1 customer) has primary importance. Such information corresponds to the cases when someone invades into the secured area or the critically important equipment encounter a breakdown. Type 2 information unit (type 2 customer) corresponds to non-important routine information about the object parameters. If type 1 information arrives, the sensor immediately stops the transmission of routine information and starts the transmission of important information. Thus, type 1 customers have an absolute (preemptive) priority over type 2 customers. We assume that type 2 information that cannot be transmitted upon arrival may be regenerated (customers make retrials). To account the possibility of obsolescence of information we assume that type 2 customers are non-persistent. The queueing system considered in our paper can help to optimally organize the operation of the sensor. For more information and the state-of-art in investigations of systems with energy harvesting see, e.g., [1, 8, 25].

2. **Cognitive radio systems with energy harvesting:** In cognitive radio systems, primary users have the preemptive priority over the secondary users. Arrivals of the primary user cause immediate interruption to the service of secondary user. The secondary user, which encounters a busy server, retry for service. For more information about application of queueing theory to
analysis of cognitive radio systems see, e.g., [11, 28]. If this radio system is autonomous and harvests the energy for the surrounding environment, then the operation of the system is described by the model considered in this paper and the additional items are the energy units. A possibility of service of the secondary users increases the efficiency of the use of the equipment and profit of the service provider. However, it is necessary to elaborate on the optimal strategy of the secondary users admission, taking into account the amount of available harvested energy in particular.

3. **Applications for modelling healthcare problems, in particular, planning and scheduling of operation theater**: We model the work of an operation theater in a hospital. This room and the team of surgeons can be considered as the servers for patients service. There are two kinds of arriving patients (customers): a) emergency or non-elective patients for whom surgery is unexpected and hence needs to be performed urgently; b) elective patients, service of which can be planned well in advance and be temporarily postponed, see, e.g., [5]. The arrival of a non-elective patient implies a break of treatment of the elective patients, if it is currently provided. In the case if the service is provided to a non-elective patient, the new non-elective patient is rejected (he/she is urgently delivered to another hospital). As the additional items required for service provisioning, the following ones can be considered: intensive care unit (ICU) staff, ICU beds, stents, blood, etc. The number of the additional items required can be different for emergency and elective patients and there is a necessity to have some stock of the items dedicated for emergency cases. The elective customers whose ongoing service is canceled due to the emergency patient’s arrival, treating can retry to arrange a new appointment later on.

The problem described above with orbit replaced by waiting room of infinite capacity for type 2 customers and abandonment of type 2 customers, is discussed in [3].

Retrial is a typical feature of many real-world systems, see, e.g. surveys [2, 12, 18, 26]. Retrial queues with heterogeneous customers and priority are considered, e.g., in [4, 7, 9, 13, 14, 23, 27]. The most popular assumption in these papers is that the priority customers, which arrive when the server is busy, are queued in a finite or infinite buffer, while the non-priority customers in such a situation move to the orbit. But in [14] it is assumed that there is no buffer for high priority customers. The high priority customer, which arrives when the server is busy, interrupts the current service if the service is provided to a low priority customer, or rejected otherwise. With motivation stemming from application to the design of wireless sensor networks and healthcare system presented above we proceed with the assumptions indicated there. Possibility of the use of additional items for providing service is not assumed in the papers listed above.

For papers devoted to the analysis of the priority systems with the use of additional items for service of heterogeneous customers we refer to [24] and [29]. Effect of retrials is not considered in these papers. In [24], service time of a customer is assumed to be equal to zero.

For $\text{MAP}$ and $\text{MMAP}$ one may refer to the review article by Chakravarthy [6]. Kim et al. [16] and Kim et al. [17] interpret additional items required for service as a token needed to start service. Similar queueing models arise in the analysis of so-called queueing/inventory models, see [20], [21], [22].
This paper is arranged as follows. In Section 2, the model under study is described; also we investigate the stability of the system. For the stable system, we derive the system state distribution. Performance indices of the system are given in Section 3. Performance measures are numerically illustrated in Section 4. Finally, we conclude the paper indicating a plan of future extension of the model described.

2. Mathematical formulation. We consider a single-server queueing system with two types of customers, no buffer for type 1 customers, an infinite orbit for type 2 customers and a stock of capacity $K$ for additional items. The structure of the system under study is given in Figure A.

Customers arrive at the system according to the marked Markovian arrival process (MMAP) (see [15] for its definition). The customers in the MMAP are heterogeneous and so of different types. The arrival of customers is directed by the stochastic process $\nu(t), t \geq 0$ which is an irreducible continuous-time Markov chain with the state space $\{1, 2, ..., w\}$. The sojourn time of this chain in the state $\nu$ is exponentially distributed with parameter $\lambda(\nu)$. When the sojourn time in the state $\nu$ expires, with probability $p^{(0)}(\nu, \nu)$, the process jumps to the state $\nu$ without generation of a customer, $\nu, \nu' \in \{1, 2, ..., w\}, \nu \neq \nu'$ and with probability $p^{(\ell)}(\nu, \nu)$ the process $\nu(t), t \geq 0$ jumps to the state $\nu'$ with the generation of type $\ell$ customer, $\ell = 1, 2; \nu, \nu' \in \{1, 2, ..., w\}$.

The behaviour of the MMAP is completely characterized by the matrices $D_\ell, \ell = 0, 1, 2$, defined by

$$
D_0 = \begin{bmatrix}
-\lambda^{(1)} & \lambda^{(1)} & \lambda^{(1)} & \cdots & \lambda^{(1)} \\
\lambda^{(2)} & -\lambda^{(2)} & \lambda^{(2)} & \cdots & \lambda^{(2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\lambda^{(w)} & \lambda^{(w)} & \lambda^{(w)} & \cdots & -\lambda^{(w)}
\end{bmatrix},
D_1 = \begin{bmatrix}
\lambda^{(1)} & \lambda^{(1)} & \lambda^{(1)} & \cdots & \lambda^{(1)} \\
\lambda^{(2)} & \lambda^{(2)} & \lambda^{(2)} & \cdots & \lambda^{(2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\lambda^{(w)} & \lambda^{(w)} & \lambda^{(w)} & \cdots & \lambda^{(w)}
\end{bmatrix},
D_2 = \begin{bmatrix}
\lambda^{(1)} & \lambda^{(1)} & \lambda^{(1)} & \cdots & \lambda^{(1)} \\
\lambda^{(2)} & \lambda^{(2)} & \lambda^{(2)} & \cdots & \lambda^{(2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\lambda^{(w)} & \lambda^{(w)} & \lambda^{(w)} & \cdots & \lambda^{(w)}
\end{bmatrix}.
$$

The matrix $D = D_0 + D_1 + D_2$ represents the generator of the process $\nu(t), t \geq 0$.

The average total arrival intensity $\lambda$ is defined by $\lambda = \theta(D_1 + D_2)e$ where $\theta$ is the invariant vector of the stationary distribution of the Markov chain $\nu(t), t \geq 0$. The vector $\theta$ is the unique solution to the system $\theta D = 0$ and $\theta e = 1$. The average
arrival intensity $\lambda_\ell$ of type $\ell$ customers is defined by $\lambda_\ell = \theta D_\ell \mathbf{e}$, $\ell = 1, 2$. Here, $\mathbf{e}$ is a column vector of 1's of appropriate order.

The squared overall (without differentiating the types of customers) coefficient of variation of intervals between successive arrivals is given as

$$c_{\text{var}} = 2\lambda \theta (-D_0)^{-1} \mathbf{e} - 1.$$  

The squared coefficient of variation of inter-arrival times of type $\ell$ customers is given as

$$c^{(\ell)}_{\text{var}} = 2\lambda_\ell \theta \left(-D_0 - D_1^{(7)}\right)^{-1} \mathbf{e} - 1, \ell \neq \ell, \ell = 1, 2.$$  

The coefficient of correlation of two successive intervals between arrivals is given as

$$c_{\text{cor}} = \left[\lambda \theta (-D_0)^{-1}(D - D_0)(-D_0)^{-1} \mathbf{e} - 1\right] / c_{\text{var}}.$$  

We assume that $K_\ell$ additional items are required for service of each type $\ell$ customer, $\ell = 1, 2$. So, at the moment when type $\ell$ customer is chosen for service, the number of additional items in the stock decreases by $K_\ell$, $\ell = 1, 2$.

Type 1 customers are assumed to be priority customers and have limited preemptive priority over type 2 customers. Arriving type 1 customer is not accepted for service (is lost) only if the server provides service to another type 1 customer or the number of additional items in the stock is less than $K_1$. If the server provides service to type 2 customer at the moment of arrival of type 1 customer, service of type 2 customer is immediately terminated, it leaves the system permanently (is lost) and the engaged into service additional items are lost as well provided there are at least $K_1$ additional items available.

**Different scenarios of the system behaviour during an arbitrary type 1 customer arrival epoch are:**

1. the server is busy with type 1 customer, the arriving type 1 customer is lost.
2. the number of additional items in the stock is less than $K_1$, the arriving type 1 customer is lost. If at this moment the server is providing service to type 2 customer, the service is not terminated.
3. the number of additional items in the stock is greater than or equal to $K_1$ and the server is free, the arriving type 1 customer occupies the server.
4. the number of additional items in the stock is greater than or equal to $K_1$ and the server is busy with type 2 customer, the arriving type 1 customer occupies the server. Type 2 customer, whose service was terminated, is lost.

Arriving type 2 customer can start service only if the server is idle and the number of additional items in the stock is greater than or equal to some preassigned threshold $N$, $N \geq K_2$. Otherwise, this customer joins the orbit of infinite capacity from where he makes repeated attempts for service according to an exponentially distributed time with parameter $n\gamma$ when $n$ customers are in the orbit. On successful retrial (server found idle and the number of available additional items is not less than $N$) he joins for service. If the retrial is a failure then with probability $q$, he leaves the system forever, and with complimentary probability he returns to orbit. The threshold $N$ is an important control parameter. If $N = K_2$, then the arriving type 2 customer starts service when the server is idle. If $N > K_2$, then the type 2 customer cannot start service even when the server is idle and there are additional items for its service but the number of additional items is less than $N$. The idea of introducing this control parameter $N$ is to have some reserve of items for the case when type 1 customer arrives. The choice of $N$ is not trivial. If $N$ is too large, the server is not effectively used. Frequently it does not work even when there is
a lot of type 2 customers in the orbit. Also, a large value of $N$ implies the higher probability that the stock for additional items will be overloaded and arriving items will be lost. If $N$ is too small (we make insufficient reservation of additional items in the stock), then the probability that the priority type 1 customer cannot start service due to the lack of additional items in the stock is large. One of the goals of the analysis presented in this paper is to highlight the influence of the threshold $N$.

**Different scenarios of the system behaviour at an arbitrary type 2 customer arrival epoch are:**

1. the number of additional items in the stock is greater than or equal to $N$, $K_2 - 1 < N \leq K$, and the server is free, the arriving customer occupies the server and the number of additional items decreases by $K_2$.
2. the number of additional items in the stock is less than $N$ or the server is busy, the arriving type 2 customer proceeds to the orbit.

Additional items arrive at the system according to the Markovian arrival process (MAP). Arrival in the MAP (see Chakravarthy [6]) is a special class of semi-Markov process with underlying continuous-time Markov chain (CTMC), $\{\zeta(t), t \geq 0\}$, on the state space $\{1, 2, \ldots, m\}$ with generator $H = H_0 + H_1$ such that $H_0$ governs transitions corresponding to no arrivals and $H_1$ accounts for transitions corresponding to arrivals. These matrices, $H_0$ and $H_1$, are of the form given by

\[
H_0 = \begin{pmatrix}
    h_{11}^{(0)} & h_{12}^{(0)} & \cdots & h_{1m}^{(0)} \\
    h_{21}^{(0)} & h_{22}^{(0)} & \cdots & h_{2m}^{(0)} \\
    \vdots & \vdots & \ddots & \vdots \\
    h_{m1}^{(0)} & h_{m2}^{(0)} & \cdots & h_{mm}^{(0)}
\end{pmatrix},
H_1 = \begin{pmatrix}
    h_{11}^{(1)} & h_{12}^{(1)} & \cdots & h_{1m}^{(1)} \\
    h_{21}^{(1)} & h_{22}^{(1)} & \cdots & h_{2m}^{(1)} \\
    \vdots & \vdots & \ddots & \vdots \\
    h_{m1}^{(1)} & h_{m2}^{(1)} & \cdots & h_{mm}^{(1)}
\end{pmatrix},
\]

where $h_{ji}^{(0)} = -\left(\sum_{j=1, j\neq i}^{m} h_{ij}^{(0)} + \sum_{j=1}^{m} h_{ij}^{(1)}\right)$. Thus, $h_{ij}^{(1)}$, $1 \leq i, j \leq m$, represents the rate of transition from $i$ to $j$ through an arrival, while $h_{ij}^{(0)}$, $1 \leq i, j \leq m, i \neq j$, represents the rate of transition from $i$ to $j$ without an arrival. Note that transition from $i$ to $i$ is possible only through an arrival and not otherwise. Let $\eta$ be the steady-state probability vector of $H$. Then $\eta H = 0$ and $\eta e = 1$. The fundamental rate $\lambda_R$ of this MAP is given by $\lambda_R = \eta H_1 e$ which gives the expected number of arrivals per unit of time. The coefficient of variation $c_{var}$ of intervals between arrivals is calculated as $c_{var} = 2\lambda_R \eta (-H_0)^{-1} e - 1$ and the coefficient of correlation $c_{cor}$ of intervals between successive arrivals is given as $c_{cor} = (\lambda_R \eta (-H_0)^{-1} H_1 (-H_0)^{-1} e - 1)/c_{var}$.

If at the arrival epoch of additional item the stock of items is full, then the arriving item is lost.

The service time of an arbitrary type $\ell$ customer has phase-type distribution with an irreducible representation $(\beta_{\ell}, S_{\ell})$, $\ell = 1, 2$. This service time can be interpreted as the time until the underlying Markov chain $\{\eta^{(\ell)}(t), t \geq 0\}$ with state space $\{1, 2, \ldots, M_{\ell}, M_{\ell} + 1\}$ reaches the single absorbing state $M_{\ell} + 1$ conditioned on the fact that the initial state of this process is selected among the states $\{1, 2, \ldots, M_{\ell}\}$ according to the initial probability vector $\beta_{\ell} = (\beta_{\ell}^{(1)}, \beta_{\ell}^{(2)}, \ldots, \beta_{\ell}^{(M_{\ell})})$. The transition rates of the process $\{\eta^{(\ell)}(t), t \geq 0\}$ within the set $\{1, 2, \ldots, M_{\ell}\}$ are defined by the sub-generator $S_{\ell}$, and the transition rates into the absorbing state are given by the entries of the column vector $S^{(0)}_{\ell} = -S_{\ell} e$. The mean service time of type $\ell$ customer
is calculated by
\[ b_1^{(\ell)} = \beta_\ell (-S_\ell)^{-1} \mathbf{e}, \quad \ell = 1, 2. \]

The squared coefficient of variation is given by
\[ \sigma_{var}^{(\ell)} = b_2^{(\ell)} / \left( b_1^{(\ell)} \right)^2 - 1 \text{ where } b_2^{(\ell)} = 2\beta_\ell (-S_\ell)^{-2} \mathbf{e}. \]

Under the above assumption we consider continuous-time Markov chain \( \{\xi(t), t \geq 0\} = \{(N(t), S(t), I(t), \nu(t), \zeta(t), \eta(t)), t \geq 0\} \) where
\[ N(t) \text{ is the number of type 2 customers in the orbit;} \]
\[ S(t) \text{ is the status of server:} \]
\[ S(t) = \begin{cases} 0 & \text{if server idle,} \\
1 & \text{if server busy with type 1 customer,} \\
2 & \text{if server busy with type 2 customer;} \end{cases} \]
\[ I(t) \text{ is the number of additional items;} \]
\[ \nu(t) \text{ is the state of the underlying MMAP;} \]
\[ \zeta(t) \text{ is the state of the underlying MAP;} \]
\[ \eta(t) \text{ is state of the phase-type service time} \]
\[ \text{at the moment } t, \ t \geq 0. \]

The Markov chain \( \{\xi(t), t \geq 0\} \) has the following state space
\[ \{(n, 0, i, j, k), n \geq 0; 0 \leq i \leq K; 1 \leq j \leq w; 1 \leq k \leq m\} \bigcup \]
\[ \{(n, \ell, i, j, k, h), n \geq 0; \ell = 1, 2; 0 \leq i \leq K; 1 \leq j \leq w; 1 \leq k \leq m; 1 \leq h \leq M_\ell\}. \]

In the sequel we need the following notations and abbreviations:
- \( \mathbf{e}_j = \) Column vector of 1's of order \( j \).
- \( I = \) Identity matrix of appropriate order.
- \( I_j = \) Identity matrix of order \( j \).
- \( \otimes = \) Kronecker product of matrices.
- \( \oplus = \) Kronecker sum of matrices.
- \( \delta_{\text{(condition)}} = \begin{cases} 1 & \text{if the condition holds true,} \\
0 & \text{otherwise.} \end{cases} \)
- \( I_{j,k} = \) Matrix of order \( K + 1 \) with all entries zero except \( (I_{j,k})_{i,i} = 1 \) for \( j \leq i \leq k \).
- \( E_{j,k}^{(\ell)} = \) Matrix of order \( K + 1 \) with all entries zero except \( (E_{j,k}^{(\ell)})_{i,i-K_\ell} = 1 \) for \( j \leq i \leq k; \ell = 1, 2 \).
- \( E^+ = \) Matrix of order \( K + 1 \) with all entries zero except \( (E^+)_{i,i+1} = 1 \) for \( 1 \leq i \leq K \).
- \( O = \) Zero matrix of appropriate order.
- \( AQTM C : \) Asymptotically quasi-Toeplitz Markov chain.

**Lemma 2.1.** The infinitesimal generator \( Q \) of the Markov chain \( \{\xi(t), t \geq 0\} \) describing the model, has a block-tridiagonal structure which is of the form
\[
Q = \begin{pmatrix}
Q_{00} & Q_0 & Q_0 & Q_0 \\
Q_{10} & Q_{11} & Q_0 & Q_0 \\
Q_{21} & Q_{22} & Q_0 & Q_0 \\
& & \ddots & \ddots & \ddots
\end{pmatrix}
\]
where each sub-matrix is of order \(wm(K + 1)(1 + M_1 + M_2)\).

\[
Q_0 = \begin{pmatrix}
Q^{(0,0)}_0 & O & O \\
O & Q^{(1,1)}_0 & O \\
O & O & Q^{(2,2)}_0
\end{pmatrix},
Q_{n,n-1} = \begin{pmatrix}
Q^{(0,0)}_{n,n-1} & O & O \\
O & Q^{(1,1)}_{n,n-1} & O \\
O & O & Q^{(2,2)}_{n,n-1}
\end{pmatrix}, \quad n \geq 1,
\]

\[
Q_{n,n} = \begin{pmatrix}
Q^{(0,0)}_{n,n} & Q^{(1,0)}_{n,n} & Q^{(1,1)}_{n,n} & Q^{(1,2)}_{n,n} \\
Q^{(2,0)}_{n,n} & O & Q^{(2,1)}_{n,n} & Q^{(2,2)}_{n,n}
\end{pmatrix}, \quad n \geq 0,
\]

with

\[Q^{(0,0)}_0 = I_{1,N} \otimes D_2 \otimes I_m,\]
\[Q^{(0,0)}_{n,n-1} = I_{1,N} \otimes \Delta q \gamma_I_{wm} \quad \text{for } \ell = 1, 2,\]
\[Q^{(2,1)}_0 = I_{K+1} \otimes D_2 \otimes I_m \otimes I_{mM_1} \otimes \beta_1,\]
\[Q^{(2,1)}_{n,n-1} = I_{K+1} \otimes \Delta q \gamma I_{wmM_1} \quad \text{for } \ell = 1, \ell = 1, 2,\]
\[Q^{(2,2)}_0 = I_{K+1} \otimes \Delta q \gamma I_{wmM_2} \quad \text{for } \ell = 1, 2,\]
\[Q^{(0,0)}_{n,n} = F_0 - \delta_{(n>0)} [I_{1,N} \otimes \Delta q \gamma I_{wm} + I_{N+1,K+1} \otimes \Delta q \gamma I_{wm}] \quad \text{for } n \geq 0,\]
\[Q^{(2,2)}_{n,n} = F_2 - \delta_{(n>0)} I_{K+1} \otimes \Delta q \gamma I_{wmM_2} \quad \text{for } n \geq 0,\]

where

\[F_0 = I_{1,K} \otimes (D_0 \otimes H_0) + E^+ \otimes I_w \otimes H_1 + I_{K+1,K+1} \otimes (D_0 \otimes H) + I_{1,K} \otimes D_1 \otimes I_m,\]
\[F_1 = I_{1,K} \otimes (D_0 \otimes H_0 \otimes S_1) + E^+ \otimes I_w \otimes H_1 \otimes I_{mM_1} + I_{K+1,K+1} \otimes (D_0 \otimes H \otimes S_1),\]
\[F_2 = I_{1,K} \otimes (D_0 \otimes H_0 \otimes S_2) + E^+ \otimes I_w \otimes H_1 \otimes I_{mM_2} + I_{1,K} \otimes D_1 \otimes I_{mM_2} + I_{K+1,K+1} \otimes (D_0 \otimes H \otimes S_2).\]

**Proof.** The proof is implemented by the analysis of all transitions of the Markov chain \(\{\xi(t), t \geq 0\}\) during an infinitesimal interval. Then we rewrite the intensities of these transitions in block matrix form.

The infinitesimal generator \(Q\) has a block-tridiagonal structure because no more than one arrival can occur and no more than one service can be completed in the system during the small interval of negligible size. That is, \(Q_{n_1,n_2} = 0\), if \(|n_1 - n_2| > 1\).
The entries of the blocks $Q_{n,n-1}$ for $n \geq 1$, define the intensities of the events that decrease the number of type 2 customers in the orbit by one.

The non-diagonal entries of the blocks $Q_{n,n}$ for $n \geq 0$, define the intensities of the events of the Markov chain $\{\xi(t), t \geq 0\}$ that do not change the number of type 2 customers in the orbit.

The entries of the matrix $Q_0$ define the intensities of the transitions that increase the number of type 2 customers in the orbit by one.

The Markov chain $\{\xi(t), t \geq 0\}$ belongs to the class of continuous-time AQTMC.

From [19], the sufficient condition for the ergodicity of AQTMC $\{\xi(t), t \geq 0\}$ is expressed in terms of the matrices as follows:

$$Y_0 = \lim_{n \to \infty} R_n^{-1} Q_{n,n-1}, \quad Y_1 = \lim_{n \to \infty} R_n^{-1} Q_{n,n} + I, \quad Y_2 = \lim_{n \to \infty} R_n^{-1} Q_0$$

where $R_n$ is a diagonal matrix with diagonal entries defined as the moduli of the corresponding diagonal entries of the matrix $Q_{n,n}$, $n \geq 0$.

Here, $R_n$ is the block diagonal matrix with the diagonal blocks $R_n^{(\ell)}$, $n \geq 0$, $\ell = 0, 1, 2$, defined as follows:

$$R_n^{(\ell)} = \begin{cases} \hat{F}_0 + \delta_{(n>0)} [I_{N+1,K+1} \otimes n q_{wm} I_{wm} + I_{1,N} \otimes n q_{wm} I_{wm}] & \text{for } \ell = 0, \\ \hat{F}_1 + (\delta_{(n>0)} I_{1,K+1} \otimes n q_{wm} I_{wm,M_1}) & \text{for } \ell = 1, \\ \hat{F}_2 + (\delta_{(n>0)} I_{1,K+1} \otimes n q_{wm} I_{wm,M_2}) & \text{for } \ell = 2, \end{cases}$$

where the matrices $\hat{F}_l$, $l = 0, 1, 2$, are the diagonal matrices the diagonal entries of which coincide with the diagonal entries of the matrices $F_l$, $l = 0, 1, 2$, correspondingly:

$$\hat{F}_0 = I_{1,K} \otimes (\Lambda_0 \otimes \Sigma_0) + I_{K+1,K+1} \otimes (\Lambda_0 \otimes \Sigma) - I_{1,K} \otimes \Lambda_1 \otimes I_m,$$

$$\hat{F}_1 = I_{1,K} \otimes (\Lambda_0 \otimes \Sigma_0 \otimes Z_1) + I_{K+1,K+1} \otimes (\Lambda_0 \otimes \Sigma \otimes Z_1) - I_{1,K} \otimes \Lambda_1 \otimes I_{mM_1},$$

$$\hat{F}_2 = I_{1,K} \otimes (\Lambda_0 \otimes \Sigma_0 \otimes Z_2) + I_{K+1,K+1} \otimes (\Lambda_0 \otimes \Sigma \otimes Z_2) - I_{1,K} \otimes \Lambda_1 \otimes I_{mM_2},$$

$\Sigma, \Sigma_0, \Lambda_0, \Lambda_1, Z_1, Z_2$ are diagonal matrices with diagonal entries defined by the diagonal entries of the matrices $-H_1, -D_0, -D_1, -S_1, -S_2$ respectively.

The explicit forms of the square matrices $Y_\ell$, $\ell = 0, 1, 2$ are as follows:

- If $q \neq 0$, then

$$Y_0 = \begin{pmatrix} Y_0^{(0,0)} & O & Y_0^{(0,2)} \\ O & Y_0^{(1,1)} & O \\ O & O & Y_0^{(2,2)} \end{pmatrix}, \quad Y_1 = Y_2 = O$$

where $Y_0^{(\ell,\ell)} = I_{(K+1)wmM_\ell}$, $\ell = 1, 2$,

$$(Y_0^{(0,0)})_{ij} = \begin{cases} I_{wm}, & j = i, \ 1 \leq i \leq N, \\ O, & \text{otherwise,} \end{cases}$$

$$(Y_0^{(0,2)})_{ij} = \begin{cases} I_{wm} \otimes \beta_2, & j = i - K_2, \ N + 1 \leq i \leq K + 1, \\ O, & \text{otherwise.} \end{cases}$$

- If $q = 0$, then
As proved in [19], the stationary state distribution of the Markov chain \( y \) exists if the inequality

\[
y Y_2 \mathbf{e} < y Y_0 \mathbf{e}
\]

holds true, where the row vector \( y \) is the unique solution to the following system of linear algebraic equations

\[
y (Y_0 + Y_1 + Y_2) = yY = y, \; ye = 1
\]

where

\[
Y = \begin{pmatrix}
Y^{(0,0)} & Y^{(0,1)} & Y^{(0,2)} \\
Y^{(1,0)} & Y^{(1,1)} & 0 \\
Y^{(2,0)} & Y^{(2,1)} & Y^{(2,2)}
\end{pmatrix}
\]

If \( q \neq 0 \), then the condition (2) can be rewritten as

\[
y Y_0 \mathbf{e} > 0
\]

and it always holds true if the vector \( y \) satisfies system (3) rewritten as

\[
y Y_0 = y, \; ye = 1.
\]

If \( q = 0 \), by direct substitution into (3), it can be verified that the solution to system (3) has the following form:

\[
y = (y^{(0)}, y^{(1)}, y^{(2)})
\]

where \( y \) satisfy the following relations:

\[
y^{(1)} = y^{(0)} U_1, \; y^{(2)} = y^{(0)} U_2.
\]

Hence we have

\[
(y^{(0)}, y^{(0)} U_1, y^{(0)} U_2) Y = (y^{(0)}, y^{(0)} U_1, y^{(0)} U_2).
\]

From the normalizing condition

\[
y^{(0)} [I + U_1 + U_2] \mathbf{e} = 1
\]

with

\[
U_\ell = \begin{cases}
  \left[ Y^{(0,1)} + U_2 Y^{(2,1)} \right] \left[ I - Y^{(1,1)} \right]^{-1}, & \ell = 1, \\
  Y^{(0,2)} \left[ I - Y^{(2,2)} \right]^{-1}, & \ell = 2.
\end{cases}
\]

From (2) we have

\[
y^{(0)} \left[ Y^{(0,0)} + U_1 Y^{(1,1)} + U_2 Y^{(2,2)} \right] \mathbf{e} < y^{(0)} Y^{(0,2)} \mathbf{e}.
\]
Thus we have proved the following theorem

**Theorem 2.2.** If \( q \neq 0 \), then the Markov chain \( \{\xi(t), t \geq 0\} \) is ergodic for any choice of system parameters. In case \( q = 0 \), the Markov chain \( \{\xi(t), t \geq 0\} \) is ergodic if (7) holds true.

Suppose that (7) is fulfilled if \( q = 0 \). Denote the stationary probabilities of system states as follows:

\[
x(n, \ell, i, j, k, h) = \lim_{t \to \infty} P\{N(t) = n, S(t) = \ell, I(t) = i, \nu(t) = j, \zeta(t) = k, \eta(t) = h\}.
\]

Let us form the row vectors

\[
x(n, \ell, i, j, k, h), n \geq 0; 0 \leq i \leq K; 1 \leq j \leq w; 1 \leq k \leq m; 0 \leq h \leq M_\ell.
\]

Write the row vectors \( x_n \) as follows:

\[
x(n, \ell) = (x(n, \ell, 0), x(n, \ell, 1), ..., x(n, \ell, K)), \quad \ell \in \{0, 1, 2\},
\]

\[
x_n = (x(n, 0), x(n, 1), x(n, 2)), \quad n \geq 0.
\]

It is well known that the probability vectors \( x_n, n \geq 0 \), satisfy the following system of linear equations:

\[
xQ = 0 \quad \text{and} \quad x e = 1,
\]

where \( Q \) is the infinitesimal generator of \( \{\xi(t), t \geq 0\} \).

Taking into account that the fact the matrix \( Q \) has a block-tridiagonal structure, to solve the system of equations given in (9), we propose the following numerically stable algorithm (see [19]).

**Theorem 2.3.** The vectors \( x_n, n \geq 0 \) are calculated as

\[
x_n = x_0 F_n, \quad n \geq 1,
\]

where the matrices \( F_n \) are calculated using the recurrent formulas

\[
F_0 = I, \quad F_n = -F_{n-1} Q_0 [Q_{nn} + Q_0 G_n]^{-1}, \quad n \geq 1,
\]

the matrices \( G_n \) are computed using the backward recursion

\[
G_n = -[Q_{n+1,n+1} + Q_0 G_{n+1}]^{-1} Q_{n+1,n}, \quad n \geq 0,
\]

and the vector \( x_0 \) is the unique solution to the system

\[
x_0 (Q_{00} + Q_0 G_0) = 0, \quad x_0 \sum_{n=0}^{\infty} F_n e = 1.
\]

As \( n \to \infty \) the sequence of matrices \( G_n \) tends to the matrix \( G \) which is the minimal non-negative solution to the matrix equation (see Klimenok and Dudin [19])

\[
G = Y_0 + Y_1 G + Y_2 G^2.
\]
3. Performance measures.

1. The probability that at an arbitrary epoch the server is busy with type 1 customers:
\[ p_{1,\text{busy}} = \sum_{n=0}^{\infty} x(n, 1)e. \]

2. The probability that at an arbitrary epoch the server is busy with type 2 customers:
\[ p_{2,\text{busy}} = \sum_{n=0}^{\infty} x(n, 2)e. \]

3. The average number of customers in the orbit:
\[ N_O = \sum_{n=1}^{\infty} nx_ne. \]

4. The average number of additional items in the stock:
\[ N_{\text{item}} = \sum_{n=0}^{\infty} K \sum_{i=1}^{K} \left[ x(n, 0, i) + x(n, 1, i) + x(n, 2, i) \right] e. \]

5. The average intensity of flow of type \( \ell \) customers who receive service:
\[ F_\ell = \sum_{n=0}^{\infty} x(n, \ell) \left( e(K+1)w_m \otimes S_0^{(\ell)} \right), \; \ell = 1, 2. \]

6. The probability that an arbitrary type \( \ell \) customer will be lost:
\[ p_{\ell,\text{loss}} = 1 - \frac{F_\ell}{\lambda_\ell}, \; \ell = 1, 2. \]

7. The probability of loss of an arbitrary type 1 customer because the server is busy with type 1 customer:
\[ p_{1,\text{busy loss}} = \frac{1}{\lambda_1} \sum_{n=0}^{\infty} x(n, 1) \left( I_{K+1} \otimes D_1 \otimes I_{mM_1} \right) e. \]

8. The probability of loss of an arbitrary arriving type 1 customer due to lack of additional items:
\[ p_{1,\text{lack loss}} = \frac{1}{\lambda_1} \sum_{n=0}^{\infty} \sum_{i=1}^{K} \left[ x(n, 0, i) (D_1 \otimes I_m) + x(n, 2, i) (D_1 \otimes I_{mM_2}) \right] e. \]

9. The probability that service of an arbitrary type 2 customer is terminated by an arriving type 1 customer:
\[ p_{2,\text{terminate}} = \frac{1}{\lambda_2} \sum_{n=0}^{\infty} \sum_{i=K_1}^{K} x(n, 2, i) \left( D_1 \otimes I_{mM_2} \right) e. \]

10. The probability that an arbitrary additional item will be lost:
\[ p_{\text{item}} = \frac{1}{\lambda_R} \sum_{n=0}^{\infty} \sum_{\ell=1}^{2} \sum_{K=0}^{K} \left[ x(n, \ell, K) \left( I_w \otimes H_1 \otimes I_{mM_1} \right) + x(n, 0, K) \left( I_w \otimes H_1 \right) \right] e. \]

11. The probability of an arbitrary type 1 arrival occupies the server upon arrival:
\[ p_{1,\text{take}} = \frac{1}{\lambda_1} \sum_{n=0}^{\infty} \sum_{i=K_1}^{K} \left[ x(n, 0, i) \left( D_1 \otimes I_m \right) + x(n, 2, i) \left( D_1 \otimes I_{mM_2} \right) \right] e. \]
12. The probability of an arbitrary arriving type 2 customer occupies the server upon arrival:

\[
p_2^{\text{take}} = \frac{1}{\lambda_2} \sum_{n=0}^{\infty} \sum_{i=N}^{K} x(n, 0, i) (D_2 \otimes I_m) e.
\]

13. The probability that the server is idle:

\[
p_{\text{idle}} = \sum_{n=0}^{\infty} x(n, 0) e.
\]

14. The probability that the server is idle with customers in the orbit:

\[
p_{\text{idle}}(c) = p_{\text{idle}} - x(0, 0) e.
\]

15. Effective rate at which the customers leave the system after unsuccessful retrial:

\[
E_{\text{leave}} = q\gamma \sum_{n=1}^{\infty} \left[ \sum_{\ell=1}^{n} n x(n, \ell) e + \sum_{i=0}^{N-1} n x(n, 0, i) e \right].
\]

16. Effective rate at which the customers enter into service after successful retrial:

\[
E_{\text{enter}} = \gamma \sum_{n=1}^{\infty} \sum_{i=N}^{K} n x(n, 0, i) e.
\]

Next we provide a few computational results that provide in-depth understanding of the system.

4. Numerical illustration. Numerical experiments have essentially 4 goals. These are to illustrate:

- Effect of variation of \( N \).
- Effect of variation of \( q \).
- Effect of correlation in the arrival process of customers and additional items.
- Effect of retrial rate \( \gamma \) and loss probability \( q \).

We assume that the stock of the additional items has capacity \( K = 20 \), the number of additional items required for service of type 1 customer is \( K_1 = 5 \) and the number of additional items required for service of type 2 customer is \( K_2 = 2 \). The retrial rate \( \gamma = 2 \).

PH service process of type 1 customers are characterized by

\[
\beta_1 = \begin{pmatrix} 1 & 0 \end{pmatrix} \quad \text{and} \quad S_1 = \begin{pmatrix} -5 & 5 \\ 0 & -5 \end{pmatrix}
\]

for which the mean service time \( b_1^{(1)} = 0.4 \) and the coefficient of variation \( c^{(1)}_{\text{var}} = 0.5 \).

PH service process of Type 2 customers are characterized by

\[
\beta_2 = \begin{pmatrix} 1 & 0 \end{pmatrix} \quad \text{and} \quad S_2 = \begin{pmatrix} -3 & 3 \\ 0 & -3 \end{pmatrix}
\]

In this case the mean service time \( b_2^{(2)} = 0.67 \) and \( c^{(2)}_{\text{var}} = 0.5 \).

In order to demonstrate the effect of correlation, we introduce two \( \text{MMAP} \) arrival flows of customers and two \( \text{MAP} \) arrival flows of additional items. Fix two \( \text{MMAPs} \) coded as \( \text{MMAP}^0 \) and \( \text{MMAP}^{0.4} \). Both have the same average total arrival rate \( \lambda = 1 \). The average intensity of Type 1 and Type 2 customers are \( \lambda_1 = 0.1 \) and \( \lambda_2 = 0.9 \) respectively.
\[ MMAP^0 \] is defined by
\[ D_0 = -1, D_1 = 0.1, D_2 = 0.9. \]

The coefficient of correlation \( c_{\text{cor}} = 0 \) and the coefficient of variation \( c_{\text{var}} = 1.\)

\[ MMAP^{0.4} \] is defined by
\[ D_0 = \begin{pmatrix} -3.3977 & 0 \\ 0.001 & -0.1102 \end{pmatrix}, D_1 = \begin{pmatrix} 0.3362 & 0.0035 \\ 0.0012 & 0.0097 \end{pmatrix}, D_2 = \begin{pmatrix} 3.026 & 0.032 \\ 0.0109 & 0.0874 \end{pmatrix}. \]

The coefficient of correlation \( c_{\text{cor}} = 0.4 \) and the coefficient of variation \( c_{\text{var}} = 12.39. \)

Next fix two \( MAP \) arrival flow of additional items coded as \( MAP^0 \) and \( MAP^{0.4}. \)

Both have the same average total arrival rate \( \lambda_R = 2.5. \)

\[ MAP^0 \] is defined by
\[ H_0 = -2.5, H_1 = 2.5. \]

The coefficient of correlation \( c_{\text{cor}} = 0 \) and the coefficient of variation \( c_{\text{var}} = 1.\)

\[ MAP^{0.4} \] is defined by
\[ H_0 = \begin{pmatrix} -8.4942 & 0 \\ 0.0025 & -0.2755 \end{pmatrix}, H_1 = \begin{pmatrix} 8.4057 & 0.0885 \\ 0.0303 & 0.2427 \end{pmatrix}. \]

The coefficient of correlation \( c_{\text{cor}} = 0.4 \) and the coefficient of variation \( c_{\text{var}} = 12.39. \)

Table 1 and Figure 1 provide dependence of the average number \( N_O \) of type 2 customers in the orbit and the average number \( N_{\text{item}} \) of additional items in the system for distinct combinations of arrival processes of additional items and that of customers for various values of the minimum number \( N \) of required additional items. These two characteristics increase with the value of \( N. \) This is easily explained. The growth of \( N \) implies the decrease of the chances of type 2 customer’s entering the service upon arrival. This leads to the increase in the average number of type 2 customers in the orbit, and, hence, the probability of these customers loss due to non-persistency. Thus, fewer items are used for the service of type 2 customers. In turn, this implies the increase in the average number of additional items in the system. Note, that the difference between the highest and lowest values of \( N_{\text{item}} \) is maximum for the combination \( MAP^{0.4} + MMAP^0. \)

Figure 2 along with Table 2 provide the graphical and numerical illustration of behaviour of type 1 customer loss probability, \( p_{\text{loss}}^1, \) and that of an arriving additional item being lost, \( p_{\text{loss}}^{\text{item}}. \) The growth of \( N \) evidently creates better conditions for type 1 customers that results in an increase of \( p_{\text{loss}}^1 \). The increase of \( p_{\text{loss}}^{\text{item}} \) stems from the increase of the average number of items in the stock that was explained above.

Table 3 and Figure 3 show that the probability \( p_{\text{loss}} \) of type 1 customer loss due to the lack of items significantly decreases with grows of \( N. \) This agrees with observations made above that the growth of \( N \) implies larger average number of items in the stock. The probability \( p_{\text{loss}}^{\text{busy}} \) of type 1 customer loss, because this customer arrives when the service is provided to another type 1 customer, increases with the growth of \( N. \) This stems from the decrease of the probability of type 1 customer loss due to lack of items and higher probability that type 1 customer is in service at an arbitrary moment. In all descriptions above \( q \) was fixed at 0.2.
Table 1. Dependence of $N_O$ and $N_{item}$ on $N$ for $q = 0.2$

![Figure 1. Dependence of average number of customers in the orbit $N_O$ and average number of additional items in the stock $N_{item}$ on $N$](image)

It is evidently seen from Figures 1-3 that correlation in arrival processes of customers and additional items has a very significant effect on the values of performance measures of the system.

Next, we vary over $q$ and examine the dependence of $N_O$ and $N_{item}$ on it. With $q$ increasing (see Table 4 and Figure 4) all combinations $MAP + MMAP$ result in reduction in type 2 customers in the orbit, which is highly significant for $MAP^{0.4} + MMAP^{0.4}$ combination. The reneging probability $q$ thus plays a very crucial role in determining the number of orbital customers in the system. It is clear that the increase of $q$ leads to more intensive loss of type 2 customers from the orbit and therefore, to the essential decrease of the average number of customers in orbit, and the increase of the number of items in the stock.

Dependence of the loss probability $p_{loss}^{1\text{type}}$ of type 1 customer and the loss probability $p_{item}$ of an arbitrary additional item are provided in Table 5 and Figure 5.
The loss probability of additional item increases with increase in the value of $q$ (see Table 5b and Figure 5b). This is because additional items are not intensively used by type 2 customers (with increase of $q$ the loss probability of type 2 customers evidently grows). The growth of the number of additional items implies the decrease of the loss probability of type 1 customers.

Table 6 and Figure 6 provide the dependence on $q$ of loss probability of a type 1 customer due to server being busy with a customer of the same type and that due to unavailability of required number of additional items, with increasing value of $q$. It may be noted that $p_{\text{busy loss}}$ increases for all combination of $MAP$ & $MMAP$ (see Table 6a) with increase in value of $q$. Availability of additional items increases, thus, $p_{\text{lack loss}}$ turns out to be decreasing (see Table 6b).

In Table 7 and Figure 7 the combined effect of retrial rate and abandoning probability on the number of customers (type 2) in the orbit is illustrated. We
Table 3. Dependence of $p_1^{\text{busy loss}}$ and $p_1^{\text{lack loss}}$ on $N$ for $q = 0.2$

![Graph A](image1.png) ![Graph B](image2.png)

Figure 3. Dependence of probability of an arbitrary arriving type 1 customer loss because the server is busy with type 1 customer $p_1^{\text{busy loss}}$ and probability of an arbitrary arriving type 1 customer loss due to lack of additional items $p_1^{\text{lack loss}}$ on $N$

adjusted $q$ and $\gamma$ values such that the product of $\gamma$ and $q$ in each row is the same. With $q$ increasing and $\gamma$ decreasing such that product $q\gamma = 0.15$, we see that $N_O$ decreases for all arrival combinations MAP and MMAP of additional items and customers.

Conclusion. We analyzed a priority queueing-inventory problem with two types of customers. Arrivals of customers follow marked Markovian arrival process and service time durations are phase-type distributed with representation $(\beta_1, S_1)$ of order $M_1$ and $(\beta_2, S_2)$ of order $M_2$ respectively for type 1 and type 2 customers, which are mutually independent. For service of each type of customer, a certain minimum number of additional items are needed. High priority customers do not have waiting space and so leave the system when on its arrival a priority 1 is
found in service/number of additional items available for its service is less than the required (when type 2 is in service). Preemption priority is assumed. Type 2 customers, encountering a busy server, go to an orbit of infinite capacity to retry for service; even when the server is idle but the number of additional items is less than the threshold, then again low priority customers go to orbit. If on retrial, the server is found to be busy/idle with the number of additional items less than the threshold, the low priority customer abandons the system with probability \( q \); with complementary probability, it gets back to orbit. The ergodicity condition is derived. The system state probability is computed from which several characteristics of the system are derived. A detailed numerical analysis of the system is performed.

Extending the model discussed to the multi-server case is proposed as future work. This will be quite challenging.

| \( q \) | \( MAP^{0+} \) | \( MAP^{0+} \) | \( MMAP^{0+} \) | \( MMAP^{0+} \) |
|-------|----------------|----------------|----------------|----------------|
| 0.1   | 10.238543      | 10.238543      | 10.238543      | 10.238543      |
| 0.2   | 10.354625      | 10.354625      | 10.354625      | 10.354625      |
| 0.3   | 10.470897      | 10.470897      | 10.470897      | 10.470897      |
| 0.4   | 10.587423      | 10.587423      | 10.587423      | 10.587423      |
| 0.5   | 10.704958      | 10.704958      | 10.704958      | 10.704958      |
| 0.6   | 10.822435      | 10.822435      | 10.822435      | 10.822435      |
| 0.7   | 10.940112      | 10.940112      | 10.940112      | 10.940112      |
| 0.8   | 11.058069      | 11.058069      | 11.058069      | 11.058069      |
| 0.9   | 11.176168      | 11.176168      | 11.176168      | 11.176168      |
| 1.0   | 11.294572      | 11.294572      | 11.294572      | 11.294572      |

Table 4. Dependence of \( N_O \) and \( N_{item} \) on \( q \) for \( N = 4 \)

**Figure 4.** Dependence of average number of customers in the orbit \( N_O \) and average number of additional items in the stock \( N_{item} \) on \( q \)
Table 5. Dependence of $p_{1}^{\text{loss}}$ and $p_{\text{item}}^{\text{loss}}$ on $q$ for $N = 4$

| $q$ | MAP$^{0+}$ | MAP$^{0+}$ | MAP$^{0+}$ | MAP$^{0+}$ |
|-----|-------------|-------------|-------------|-------------|
| 0.1 | 0.060154    | 0.528047    | 0.612127    | 0.574752    |
| 0.2 | 0.068159    | 0.560329    | 0.607792    | 0.566495    |
| 0.3 | 0.075424    | 0.489551    | 0.604719    | 0.560590    |
| 0.4 | 0.081344    | 0.456390    | 0.598905    | 0.549560    |
| 0.5 | 0.085940    | 0.442329    | 0.596454    | 0.545017    |
| 0.6 | 0.088865    | 0.436726    | 0.595472    | 0.543215    |

References

[1] A. Arafa, T. Tong, M. Fu, S. Ulukus and W. Chen, Delay minimal policies in energy harvesting communication systems, IEEE Transactions on Communications, 66 (2018), 2918–2930.

[2] J. R. Artalejo and A. Gomez Corral, Retrial Queueing Systems: A Computational Approach, Springer, Berlin, 2008.

[3] J. Baek, O. Dudina and C. Kim, Queueing system with heterogeneous impatient customers and consumable additional items, International Journal of Applied Mathematics and Computer Science, 27 (2017), 367–384.

[4] L. Boutarfa and N. Djellab, On the performance of the $M_{1}M_{2}G_{1}G_{2}/1$ retrial queue with pre-emptive resume policy, Yugoslav Journal of Operations Research, 25 (2015), 153–164.

[5] B. Cardoen, E. Demeulemeester and J. Belien, Operating room planning and scheduling: A literature review, European Journal of Operational Research, 201 (2010), 921–932.
Table 6. Dependence of \( p_{1}^{\text{busy loss}} \) and \( p_{1}^{\text{lack loss}} \) on \( q \) for \( N = 4 \)

![Diagram](attachment:image.png)

Figure 6. Dependence of probability of an arbitrary arriving type 1 customer loss because the server is busy with type 1 customer \( p_{1}^{\text{busy loss}} \) and probability of an arbitrary arriving type 1 customer loss due to lack of additional items \( p_{1}^{\text{lack loss}} \) on \( q \).

[6] S. R. Chakravarthy, The batch Markovian arrival process: A review and future work, *Advances in Probability Theory and Stochastic Process: Notable Publications* (eds. A. Krishnamoorthy, et al.), New Jersey, 2001, 21–49.

[7] B. D. Choi and Y. Chang, Single server retrial queues with priority customers, *Mathematical and Computer Modelling*, 30 (1999), 7–32.

[8] E. De Cuypere, K. De Turck and D. Fiems, A queueing model of an energy harvesting sensor node with data buffering, *Telecommunication Systems*, 67 (2018), 281–295.

[9] I. Dimitrious, Analysis of a priority retrial queue with dependent vacation scheme and application to power saving in wireless communication systems, *The Computer Journal*, 56 (2015), 1363–1380.

[10] A. N. Dudin and V. I. Klimenok, Queueing systems with passive servers, *International Journal of Stochastic Analysis*, 9 (1996), 185–204.
| $q$ | $\gamma$ | $MAP^0 + MMAP^0$ | $MAP^{0.4} + MMAP^{0.4}$ | $MAP^{0.4} + MMAP^{0.4}$ | $MAP^{0.4} + MMAP^{0.4}$ |
|-----|---------|-----------------|-----------------|-----------------|-----------------|
| 0.1 | 1.5     | 0.959861        | 3.82764         | 4.300535        | 3.145578        |
| 0.2 | 0.75    | 1.328719        | 3.583343        | 4.430688        | 3.384015        |
| 0.3 | 0.5     | 1.628854        | 3.746986        | 4.534785        | 3.574155        |
| 0.4 | 0.375   | 1.881123        | 3.887829        | 4.623184        | 3.731486        |
| 0.5 | 0.3     | 2.098026        | 4.012896        | 4.700989        | 3.865063        |
| 0.6 | 0.25    | 2.287694        | 4.126279        | 4.771088        | 3.980639        |
| 0.7 | 0.2143  | 2.455731        | 4.230599        | 4.835292        | 4.082103        |
| 0.8 | 0.1875  | 2.606171        | 4.327642        | 4.89482         | 4.172209        |
| 0.9 | 0.1667  | 2.742017        | 4.418685        | 4.950534        | 4.252983        |
| 1   | 0.1500  | 2.865570        | 4.504674        | 5.003066        | 4.325962        |

Table 7. Dependence of $N_O$ on $\gamma$ and $q$ for $N = 4$

![Figure 7. Dependence of $N_O$ on $\gamma$](image)

[11] A. N. Dudin, M. H. Lee, O. Dudina and S. K. Lee, Analysis of priority retrial queue with many types of customers and servers reservation as a model of cognitive radio system, *IEEE Transactions on Communications*, 65 (2017), 186–199.

[12] G. I. Falin and J. G. C. Templeton, *Retrial Queues*, Chapman& Hall, London, 1997.

[13] G. I. Falin, J. R. Artalejo and M. Martin, On the single server retrial queue with priority customers, *Queueing Systems*, 14 (1993), 439–455.

[14] S. Gao, A preemptive priority retrial queue with two classes of customers and general retrial times, *Operational Research*, 15 (2015), 233–251.

[15] Q. M. He, Queue with marked customers, *Advances in Applied Probability*, (1996), 567–587.

[16] C. Kim, S. Dudin and V. Klimenok, The $MAP/PH/1/N$ queue with flows of customers as model for traffic control in telecommunication networks, *Performance Evaluation*, 66 (2009), 564–579.

[17] C. Kim, S. Dudin and V. Klimenok, Queueing system with batch arrival of customers in sessions, *Computers and Industrial Engineering*, 62 (2012), 890–897.

[18] J. Kim and B. Kim, A survey of retrial queueing systems, *Annals of Operations Research*, 247 (2016), 3–36.

[19] V. I. Klimenok and A. N. Dudin, Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory, *Queueing Systems*, 54 (2006), 245–259.

[20] A. Krishnamoorthy, B. Binitha and D. Shaqin, A revisit to queueing-inventory system with reservation, cancellation and common life time, *OPSEARCH*, 54 (2017), 336–350.
[21] A. Krishnamoorthy, D. Shajin and B. Lakshmy, On a queueing-inventory with reservation, cancellation, common life time and retrial, *Annals of Operations Research*, (2016), 1–25.

[22] A. Krishnamoorthy, D. Shajin and B. Lakshmy, Product form solution for some queueing-inventory supply chain problem, *OPSEARCH*, 53 (2016), 85–102.

[23] C. Langaris and E. Moutzoukis, A retrial queue with structured batch arrivals, priorities and server vacations, *Queueing Systems*, 20 (1995), 341–368.

[24] M. Liu, M. Feng and C. Y. Wong, Flexible service policies for a Markov inventory system with two demand classes, *International Journal of Production Economics*, 151 (2014), 180–185.

[25] K. Patil, K. De Turck and D. Fiems, A two-queue model for optimising the value of information in energy-harvesting sensor networks, *Performance Evaluation*, 119 (2018), 27–42.

[26] T. Phung-Duc, Retrial queueing models: A survey on theory and applications, (to appear) in *Stochastic Operations Research in Business and Industry* (eds. T. Dohi, K. Ano and S. Kasahara), WSP, 2017.

[27] J. Walraevens, D. Claeys and T. Phung-Duc, Asymptotics of queue length distributions in priority retrial queues, *Performance Evaluation*, 127 (2018), 235–252.

[28] Y. Wang, X. Tang and T. Wang, A unified QoS and security provisioning framework for wiretap cognitive radio networks: A statistical queueing analysis approach, *IEEE Transactions on Wireless Communications* 18 (2019), 1548–1565.

[29] N. Zhao and Z. Lian, A queueing-inventory system with two classes of customers, *International Journal of Production Economics*, 129 (2011), 225–231.

Received October 2018; revised March 2019.

E-mail address: dhanya.shajin@gmail.com
E-mail address: dudin@bsu.by
E-mail address: dudina@bsu.by
E-mail address: achyuthacusat@gmail.com