mathematics

MDPI

Article

Analysis of a Batch Arrival, Batch Service Queuing-Inventory System with Processing of Inventory While on Vacation

Achyutha Krishnamoorthy [1], Anu Nuthan Joshua [2,†] and Dmitry Kozyrev [3,4,*]

[1] Centre for Research in Mathematics, C.M.S. College, Kottayam 686001, India; achyuthacusat@gmail.com
[2] Department of Mathematics, Union Christian College, Aluva 683102, India; anunuthanjoshua@gmail.com
[3] Applied Probability and Informatics Department, Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya St, 117198 Moscow, Russia
[4] V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, 65 Profsoyuznaya Street, 117997 Moscow, Russia
* Correspondence: kozyrev-dv@rudn.ru
† Working for Doctoral Degree at Department of Mathematics, Cochin University of Science and Technology, Cochin-22.

Abstract: A single-server queuing-inventory system in which arrivals are governed by a batch Markovian arrival process and successive arrival batch sizes form a finite first-order Markov chain is considered in this paper. Service is provided in batches according to a batch Markovian service process, with consecutive service batch sizes forming a finite first-order Markov chain. A service starts for the next batch on completion of the current service, provided that inventory is available at that epoch; otherwise, there will be a delay in starting the next service. When the service of a batch is completed, the inventory decreases by 1 unit, irrespective of batch size. A control policy in which the server goes on vacation when a service process is frozen until a quorum can initiate the next batch service is proposed to ensure idle-time utilization. During the vacation, the server produces inventory (items) for future services until it hits a specified level L or until the number of customers in the system reaches a maximum service batch size N, with whichever occurring first. In the former case, a server stays idle once the processed inventory level reaches L until the number of customers reaches (or even exceeds because of batch arrival) a maximum service batch size N. The time required for processing one unit of inventory follows a phase-type distribution. In this paper, the steady-state probability vector of this infinite system is computed. The distributions of inventory processing time in a vacation cycle, idle time in a vacation cycle, and vacation cycle length are found. The effect of correlation in successive inter-arrival times and service times on performance measures for such a queuing system is illustrated with a numerical example. An optimization problem is considered. The proposed system is then compared with a queuing-inventory system without the Markov-dependent assumption on successive arrivals as well as service batch sizes using numerical examples.

Keywords: queuing-inventory system; batch Markovian arrival process; batch Markovian service process; Markov-dependent arrival and service batches; vacation; $N-$policy

check for updates

Citation: Krishnamoorthy, A.; Joshua, A.N.; Kozyrev, D. Analysis of a Batch Arrival, Batch Service Queuing-Inventory System with Processing of Inventory While on Vacation. *Mathematics* **2021**, *9*, 419. https://doi.org/10.3390/math9040419

Academic Editor: Elvira Di Nardo

Received: 15 December 2020
Accepted: 31 January 2021
Published: 20 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Bulk arrival and bulk service queues have been extensively analyzed in the literature (for example, see Chaudhry and Templeton [1] for an in-depth study on bulk queues). The earliest work considered arrival and service processes to be mutually independent. Furthermore, inter-arrival times and successive service times were assumed to be independent. The next stage of development had a relaxed assumption of independence between successive inter-arrival times and/or successive service times. One such extension is the Markovian arrival process (MAP) or Markovian service process (MSP) (single or multi-server queues), in which successive inter-arrival times or successive service times are correlated through the respective semi-Markov processes. Its extension to the batch Markovian arrival

process ($BMAP$) and/or the batch Markovian service process ($BMSP$) considers batch arrival and/or batch service. (Refer to [2–4] for more details and to [5–7] for reviews on $BMAP$.) Successive arrival batches are assumed to be mutually independent and "within independent" in the sense that successive arrival batch sizes are independent. This is also true for successive service batches. However, there is only one published paper [8] wherein arrivals are in batches ($BMAP$) and service is also in batches ($BMSP$), with arrival and service batch sizes forming two distinct Markov chains. Thus, successive arrival batch sizes and successive service batch sizes are determined by two distinct Markov chains. The purpose of this paper is to extend the work of Krishnamoorthy A. and Anu Nuthan Joshua [8] for queuing inventory, with items that are to be served to customers or to be used for serving customers processed by the server while idle.

This work could also be considered an extension of the queuing-inventory problem considered in Divya et al. [9]. In [9], the authors considered a single server queuing-inventory system in which customer arrival is governed by a Markovian arrival process (MAP). The service process with as well as without inventory in stock follows two distinct phase-type distributions (one in which the processed item is available and the other in which the processed item is not available at a service commencement epoch). This assumption of the service process is made based on the observation that, with the availability of additional items at the service commencement epoch, the service time of customers becomes shorter as the item does not need to be processed before initiation of the service. The server goes on vacation when the system is empty and produces inventory for future use. The server returns from vacation once the number of customers in the system reaches a certain prescribed limit N. It is assumed that customers join the queue with probability p and, after spending a random time period in the queue (which is exponentially distributed), become impatient and renege. The impacts of customer behavior on individual optimal strategies, revenue to the server, and social optimal strategy are analyzed extensively in that paper using numerical experiments. Earlier works, with some connection to the present work, in terms of processing of items to be delivered to the customers, are Kazirmsky [10] (service time depends on the number of items processed and customer arrivals following $BMAP$), Hanukov et al. [11], and Divya et al. [12], in addition to Divya et al. [9]. There are other works in which the service requires an additional item (see [13–15]). In the models analyzed by Baek et al. [13] and Dhanya et al. [15], additional items required for services arrive according to the Markovian arrival process (MAP). All of the works mentioned above are related to queuing systems with single arrival and/or single service at a time. The literature on vacation queuing systems is also quite extensive. (The concept was introduced by Levy and Yechiali [16] and reviews of the literature can be found in [17–19]).

However, it is more realistic to consider systems in which both customer arrivals and services occur in batches. Markov dependence on successive arrival and service batch sizes is observable in many real-life situations. It is a useful strategy in optimizing performance and in balancing workload by suitably assigning values to the transition probabilities in the Markov chain that decides successive service batch sizes (see [8] for details). For example, many production plants ensure that a minimum number of machines (service batch size) are put to use (which varies from time to time based on demand) to ensure optimum production. The demands (arrival batch size) are accommodated based on previous experience of successful levels of production. The Markov-dependent assumption on successive service batch sizes has the following disadvantage: whenever the server does not find the required numbers of customers to initiate the next batch service, it stays idle. However, in the model that is studied in this paper, the server effectively utilizes its idle time to further reduce the waiting times of customers by engaging itself in producing items for service (by going on vacation). The vacation expires when N customers (the maximum service batch size) accumulate in the system. Thus, in this paper, a model that extends [9] and captures the following dependencies is considered:

- Successive inter-arrival times and successive service times are correlated.

- Consecutive arrival and service batch sizes form two distinct first-order Markov chains.
- Service process is governed by a *BMSP* but has transition rates depending on whether there are processed items available/not available for service commencement of the batch now being served; if quorum for the next batch, as determined by the Markov chain rule, is not available at a service completion epoch, the service process is frozen until the number of customers in the system reaches N.

This model (Model I) is then compared with another (Model II) in which the successive arrival and service batch sizes are not Markov-dependent. In Model II, the server goes on vacation only when the system is empty. The working of the models analyzed in this paper is illustrated using the flowcharts given in Figures 1 and 2.

Model I

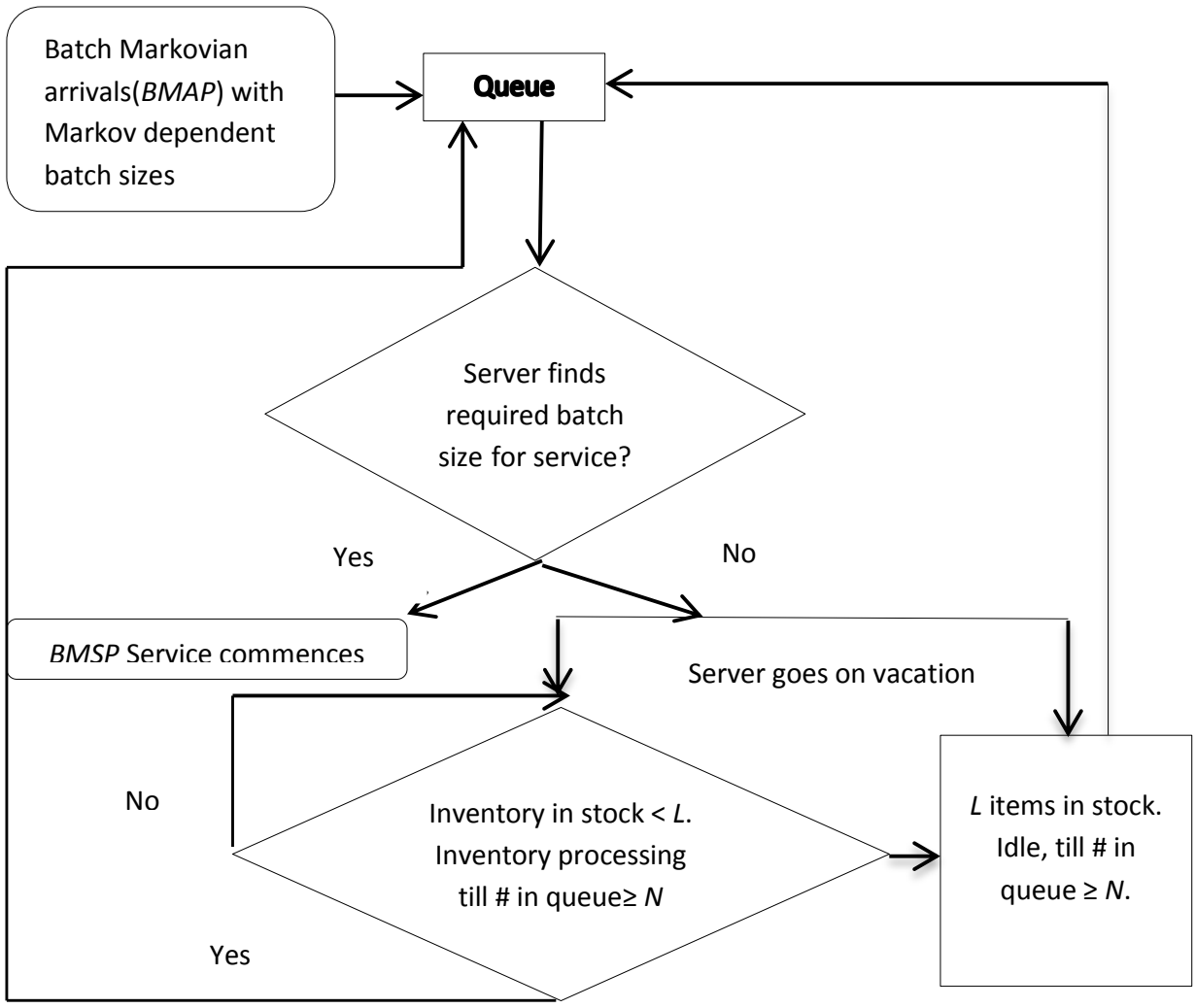


Figure 1. Flowchart indicating the functioning of a queuing-inventory system with Markov-dependent assumption on successive arrival as well as service batch sizes: Model I.

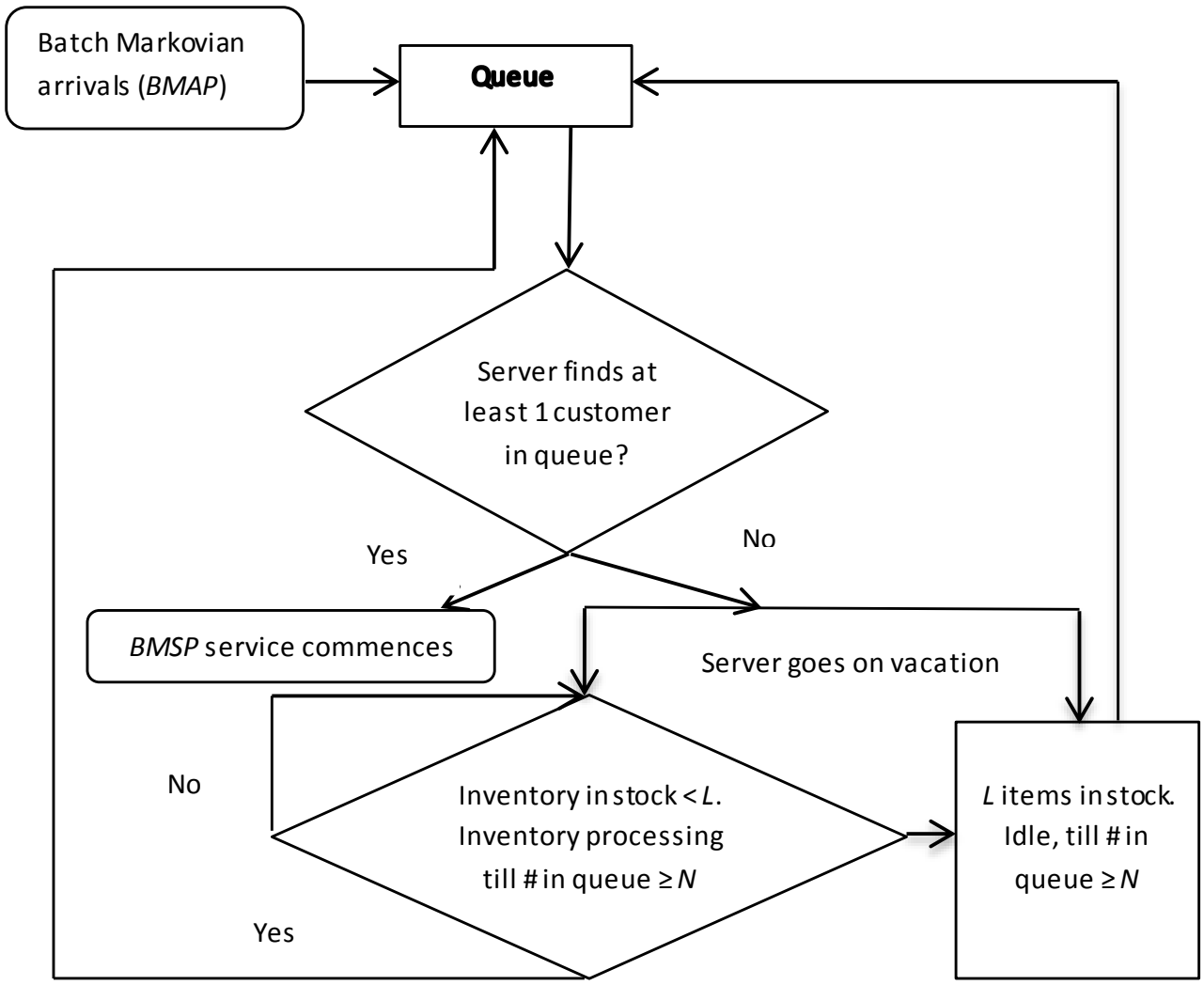


Figure 2. Flowchart indicating the functioning of a queuing-inventory system without Markov-dependent assumption on successive arrival as well as service batch sizes: Model II.

This paper discusses a very general model. However, it is very complex. The main problem is in the computation. The dimension is very high, and thus, one may face tractability.

This paper is organized as follows. The mathematical formulation of Model I is described in Section 2. Section 3 deals with a steady-state analysis of this queuing system. In Section 4, certain distributions associated with vacation are derived and performance measures for the queuing system under consideration are analyzed. A numerical example is provided in Section 5. A cost function based on performance measures constructed and the optimal value of L that minimizes the cost is computed in Section 6. The description and formulation of Model II are given in Section 7. Its steady-state analysis and system characteristics are presented in Sections 8 and 9, respectively. A numerical comparison between the two models is presented in Section 10. The conclusions that are drawn from the study of the proposed problems are briefly sketched in Section 11.

2. Model Description and Formulation of Model I

Consider a single server queuing-inventory system with customers arriving according to a batch Markovian arrival process $(BMAP)$ with maximum arrival batch size a. Successive arrival batch sizes form a first-order MC $\{X_n; n \geq 1\}$ with tpm $P = [p_{ij}]$ on state-space $\{1, 2, 3...a\}$. The service time duration is based on whether items are available at a service initiation epoch. The service process follows a batch Markovian service process $(BMSP)$ with successive service batch sizes forming a first-order MC $\{Y_n; n \geq 1\}$ with tpm $Q = [q_{ij}]$ on state-space $\{1, 2, 3...N\}$. Additional items are required for providing service. If an item is not available for service, the server has to process it before the start of service and this increases the service time duration. If there is an item available at a service commencement epoch, service will be provided at a rate μ; otherwise, service is provided at a rate $\theta\mu; 0 < \theta < 1$. This includes the processing time of the item and the time for serving the present batch of customers. The server goes on vacation when the service process is frozen due to a lack of quorum to initiate the next batch service, as per MC determining the service batch size. Let $0(k)$ denote the status of the server when the service process is frozen until k customers are reached to initiate the next batch service. During vacation, the server processes inventory until the inventory level becomes L or until the number of customers in the system is equal to or exceeds N, whichever occurs first. The server becomes idle once the inventory level reaches L and waits for N customers to accumulate to initiate the next service, provided that the former precedes the latter (the inventory level reaches L first). The inventory processing time follows phase-type distribution $PH(\alpha, T)$ of order t_1. Only one item is provided to each batch of customers undergoing service, irrespective of batch size. The additional item can be regarded as an essential item for providing service to customers, irrespective of the size of the batch to undergo service. Exactly one item is required to provide service to each batch. This item cannot be reused; that is, it belongs to the "consumable class".

The arrival process is defined using two stochastic matrices, D_0 and D_1, of order m. Here, entries of D_0 denote transition rates of an underlying MC of $BMAP$ without arrivals and entries of D_1 denote transition rates of an underlying MC of $BMAP$ with batch arrivals. Each arrival batch size is determined by the MC rule with tpm P. Hence, if the last arrival batch is of size i, the next arrival batch size is j with probability p_{ij}, i.e., the transition rates of an underlying MC of $BMAP$ if the last arrival batch size is i and the next arrival of batch size is j are specified by matrix $p_{ij}D_1$. Similarly, the size of the next batch to be served is determined by the MC with tpm Q. The service process is defined using two matrices S_0 and S_1 of order t_2, where entries of S_0 denote transition rates of an underlying MC of $BMSP$ without departures and S_1 denote transition rates of an underlying MC of $BMSP$ with departures, i.e., the transition rates of $BMSP$ if the current service batch size is i and the next batch to be served is of size j are specified by the matrix $q_{ij}S_1$.

Let $N_1(t)$ be the number of customers in the queue at time t. We write $N_1(t) = n$ as (l, p) if $n = lq + p; l \geq 0; 0 \leq p \leq q - 1$, where $q = max\{a, N\}$. The purpose of redefining the level using maximum arrival or service batch size is to obtain an $LIQBD$ structure for the generator matrix.

Let $N_2(t)$ be the number of processed inventory available at t.

Let $J(t)$ be the status of the server at t.

$$J(t) = \begin{cases} 0, & if\ the\ server\ is\ on\ vacation \\ 1, & if\ the\ server\ is\ busy \end{cases}$$

$A(t)$ is the size of the last arrival batch before time t.

$B(t)$ is the size of the service batch at time t.

$K_1(t)$ is the phase of inventory processing.

$K_2(t)$ is the state of an underlying MC of $BMSP$.

$M(t)$ is the state of an underlying MC of $BMAP$.

The above model can be studied using a $CTMC$, $\{(N_1(t), N_2(t), J(t), A(t), B(t), K_1(t), K_2(t), M(t)) : t \geq 0\}$ on state-space $\Omega_0 \cup_{l \geq 1} \Omega_l$.

Here, $1 \leq n' \leq L, 1 \leq n_1 \leq a, 1 \leq n_2 \leq N, 1 \leq k_1 \leq t_1, 1 \leq k_2 \leq t_2, 1 \leq k_3 \leq m$. The states in Table 1 correspond to level 0, i.e., to the states with $0, 1, 2...q-1$ customer(s) in the queue. Table 2 corresponds to level l, i.e., to the states with $lq, lq+1, lq+2...(l+1)q-1$ customer(s) in the queue. By redefining the level as described above, the infinitesimal generator of this $CTMC$ can be brought to the form of an $LIQBD$ with a generator matrix:

$$\mathcal{Q}_1 = \begin{bmatrix} B_{00} & B_{01} & & & & \\ B_{10} & B_1 & B_0 & & & \\ & B_2 & B_1 & B_0 & & \\ & & B_2 & B_1 & B_0 & \\ & & & \ddots & \ddots & \ddots \end{bmatrix} \tag{1}$$

$$B_2 = \begin{bmatrix} C_q & C_{q-1} & C_{q-2} & \cdots & C_1 \\ & C_q & C_{q-1} & \cdots & C_2 \\ & & & \ddots & \vdots \\ & & & & C_q \end{bmatrix} \tag{2}$$

in which, for $i = 1, 2, ..q$,

$$C_i = \begin{bmatrix} C_{00}^i & & & & \mathbf{0} \\ C_{10}^i & & & & \mathbf{0} \\ & C_{21}^i & & & \mathbf{0} \\ & & \ddots & & \vdots \\ & & & C_{LL-1}^i & \mathbf{0} \end{bmatrix} \tag{3}$$

$$\begin{aligned} C_{00}^i &= I_a \otimes Q.E_{ii}(N) \otimes \theta S_1 \otimes I_m \\ C_{jj-1}^i &= I_a \otimes Q.E_{ii}(N) \otimes S_1 \otimes I_m; for j = 1, 2, ...L. \\ E_{ii}(N) &= e_i(N).e_i^{'}(N). \\ C_i &= \mathbf{0}; i > N. \end{aligned} \tag{4}$$

$$B_0 = \begin{bmatrix} A_q & & & \\ A_{q-1} & A_q & & \\ \vdots & \vdots & \ddots & \\ A_1 & A_2 & \cdots & A_q \end{bmatrix} \tag{5}$$

For $i = 1, 2, ..q$,

$$\begin{aligned} A_i &= I_{L+1} \otimes P.E_{ii}(a) \otimes I_N \otimes I_{t_2} \otimes D_1; \\ A_i &= \mathbf{0}, i > a. \end{aligned} \tag{6}$$

$$B_1 = \begin{bmatrix} F & A_1 & A_2 & \cdots & A_{q-1} \\ C_1 & F & A_1 & \cdots & A_{q-2} \\ \vdots & \vdots & \vdots & \vdots & \\ C_{q-1} & C_{q-2} & C_{q-3} & \cdots & F \end{bmatrix} \tag{7}$$

in which

$$F = \begin{bmatrix} F_0 & & & \\ & F_1 & & \\ & & \ddots & \\ & & & F_1 \end{bmatrix} \tag{8}$$

$$\begin{aligned} F_0 &= I_{aN} \otimes (\theta S_0 \oplus D_0); \\ F_1 &= I_{aN} \otimes (S_0 \oplus D_0). \end{aligned} \tag{9}$$

Table 1. States in Ω_0 and their descriptions.

Sl. No	State	Description
1	$(0, p, n', 0, n_1, 0(n_2), k_1, k_2, k_3)$ $0 \le p < N; 0 \le n' \le L-1$	Service process is frozen with the server on vacation producing inventory for future use
2	$(0, p, L, 0, n_1, 0(n_2), k_2, k_3)$ $0 \le p < N$	Server is idle as the maximum inventory level is reached and the system is on vacation
3	$(0, p, 0, 1, n_1, n_2, k_2, k_3)$ $0 \le p \le q-1$	Service process without inventory (at the commencement epoch of current service) is ongoing
4	$(0, p, n', 1, n_1, n_2, k_2, k_3)$	Service process with inventory is ongoing
5	$(0, p, L, 1, n_1, n_2, k_2, k_3)$ $0 \le p \le q-1; p+n_2 \ge N; 1 \le n_1 \le p+n_2$	Service process with inventory is activated with customer arrivals on expiry of vacation

Table 2. States in $\Omega_l; l \ge 1$ and their descriptions.

Sl. No	State	Description
1	$(l, p, 0, 1, n_1, n_2, k_2, k_3)$ $0 \le p \le q-1$	Service process without inventory (at commencement the epoch of current service) is ongoing
2	$(l, p, n', 1, n_1, n_2, k_2, k_3)$ $0 \le p \le q-1$	Service process with inventory is ongoing

The transition rate submatrices amongst the various levels are provided in Tables 3–7.

Table 3. Transition rate submatrices from level 0 to itself.

From	To	Rate Matrix
$(0, p, n', 0, n_1, 0(n_2))$	$(0, p, n', 0, n_1, 0(n_2))$	$T \oplus (I_{t_2} \otimes D_0)$
$(0, p, L, 0, n_1, 0(n_2))$	$(0, p, L, 0, n_1, 0(n_2))$	$I_{t_2} \otimes D_0$
$(0, p, 0, 1, n_1, n_2)$	$(0, p, 0, 1, n_1, n_2)$	$\theta S_0 \oplus D_0$
$(0, p, n', 1, n_1, n_2)$	$(0, p, n', 1, n_1, n_2)$	$S_0 \oplus D_0$
$(0, p, n', 0, n_1, 0(n_2))$	$(0, p, n'+1, 0, n_1, 0(n_2))$	$T^0 \alpha \otimes I_{t_2 m}$
$(0, p, L-1, 0, n_1, 0(n_2))$	$(0, p, L, 0, n_1, 0(n_2))$	$T^0 \otimes I_{t_2 m}$
$(0, p, n', 0, n_1, 0(n_2))$	$(0, p+n_1', n', 0, n_1', 0(n_2))$	$I_{t_1 t_2} \otimes p_{n_1 n_1'} D_1$
$(0, p, 0, 0, n_1, 0(n_2))$	$(0, p+n_1'-n_2, 0, 1, n_1', n_2)$	$e(t_1) \otimes I_{t_2} \otimes p_{n_1 n_1'} D_1$
$(0, p, L, 0, n_1, 0(n_2))$	$(0, p+n_1', L, 0, n_1', 0(n_2))$	$I_{t_2} \otimes p_{n_1 n_1'} D_1$
$(0, p, L, 0, n_1, 0(n_2))$	$(0, p+n_1'-n_2, L, 1, n_1', n_2)$	$I_{t_2} \otimes p_{n_1 n_1'} D_1$
$(0, p, 0, 1, n_1, n_2)$	$(0, p+n_1', 0, 1, n_1', n_2)$	$I_{t_2} \otimes p_{n_1 n_1'} D_1$
$(0, p, n', 1, n_1, n_2)$	$(0, p+n_1', n', 1, n_1, n_2)$	$I_{t_2} \otimes p_{n_1 n_1'} D_1$
$(0, p, 0, 1, n_1, n_2)$	$(0, p-n_2', 0, 1, n_1, n_2')$	$q_{n_2 n_2'} \theta S_1 \otimes I_m$
$(0, p, 0, 1, n_1, n_2)$	$(0, p, 0, 0, n_1, 0(n_2'))$	$\alpha \otimes q_{n_2 n_2'} \theta S_1 \otimes I_m$
$(0, p, n', 1, n_1, n_2)$	$(0, p-n_2', n'-1, 1, n_1, n_2')$	$q_{n_2 n_2'} S_1 \otimes I_m$
$(0, p, n', 1, n_1, n_2)$	$(0, p, n'-1, 0, n_1, 0(n_2'))$	$\alpha \otimes q_{n_2 n_2'} S_1 \otimes I_m$

Table 4. Transition rate submatrices from level 0 to 1.

From	To	Rate Matrix
$(0, p, n', 0, n_1, 0(n_2))$	$(1, p + n_1' - (n_2 + q), n', 1, n_1', n_2)$	$e(t_1) \otimes I_{t_2} \otimes p_{n_1 n_1'} D_1$
$(0, p, L, 0, n_1, 0(n_2))$	$(1, p + n_1' - (n_2 + q), L, 1, n_1', n_2)$	$I_{t_2} \otimes p_{n_1 n_1'} D_1$
$(0, p, n', 1, n_1, n_2)$	$(1, p + n_1' - q, n', 1, n_1', n_2)$	$I_{t_2} \otimes p_{n_1 n_1'} D_1$

Table 5. Transition rate submatrices from level $l \geq 1$ to $l - 1$.

From	To	Rate Matrix
$(l, p, 0, 1, n_1, n_2)$	$(l - 1, p + q - n_2', 0, 1, n_1, n_2')$	$q_{n_2 n_2'} \theta S_1 \otimes I_m$
$(l, p, n', 1, n_1, n_2)$	$(l - 1, p + q - n_2', n' - 1, 1, n_1, n_2')$	$q_{n_2 n_2'} S_1 \otimes I_m$

Table 6. Transition rate submatrices from level l to itself.

From	To	Rate Matrix
$(l, p, 0, 1, n_1, n_2)$	$(l, p, 0, 1, n_1, n_2)$	$\theta S_0 \oplus D_0$
$(l, p, n', 1, n_1, n_2)$	$(l, p, n', 1, n_1, n_2)$	$S_0 \oplus D_0$
$(l, p, 0, 1, n_1, n_2)$	$(l, p + n_1', 0, 1, n_1', n_2)$	$I_{t_2} \otimes p_{n_1 n_1'} D_1$
$(l, p, n', 1, n_1, n_2)$	$(l, p + n_1', n', 1, n_1', n_2)$	$I_{t_2} \otimes p_{n_1 n_1'} D_1$
$(l, p, 0, 1, n_1, n_2)$	$(l, p - n_2', 0, 1, n_1, n_2')$	$q_{n_2 n_2'} \theta S_1 \otimes I_m$
$(l, p, n', 1, n_1, n_2)$	$(l, p - n_2', n' - 1, 1, n_1, n_2')$	$q_{n_2 n_2'} S_1 \otimes I_m$

Table 7. Transition rate submatrices from level $l \geq 1$ to $l + 1$.

From	To	Rate Matrix
$(l, p, 0, 1, n_1, n_2)$	$(l + 1, p + n_1' - q, 0, 1, n_1', n_2)$	$I_{t_2} \otimes p_{n_1 n_1'} D_1$
$(l, p, n', 1, n_1, n_2)$	$(l + 1, p + n_1' - q, n', 1, n_1', n_2)$	$I_{t_2} \otimes p_{n_1 n_1'} D_1$

About the transitions:

- $(0, p, n', 0, n_1, 0(n_2)) \rightarrow (0, p + n_1', n', 0, n_1', 0(n_2))$ denotes the transition associated with the arrival of n_1' customers to the system. However, the server is still on vacation as there are not N customers in queue.
- $(0, p, 0, 0, n_1, 0(n_2)) \rightarrow (0, p + n_1' - n_2, 0, 1, n_1', n_2)$ denotes the transition associated with the arrival of n_1' customers to the system, which activates the service of a batch of n_2 customers without providing inventory.
- $(0, p, n', 1, n_1, n_2) \rightarrow (0, p, n' - 1, 0, n_1, 0(n_2'))$ denotes the transition associated with service completion of a batch of n_2 customers. As there are not n_2' customers specified by the MC rule, the server goes on vacation.
- $(0, p, n', 0, n_1, 0(n_2)) \rightarrow (0, p, n' + 1, 0, n_1, 0(n_2))$ denotes the transition associated with processing of a unit item while on vacation.
- $(l, p, n', 1, n_1, n_2) \rightarrow (l - 1, p + q - n_2', n' - 1, 1, n_1, n_2')$ denotes the transition associated with service completion of a batch of n_2 customers and the initiation of service to a batch of n_2' customers specified by the MC rule (which decreases the number in queue from $lq + p$ to $(l - 1)q + p + q - n_2' = lq + p - n_2'$.)
- $(l, p, n', 1, n_1, n_2) \rightarrow (l, p, n', 1, n_1, n_2)$ denotes the transition without service completion or arrival.

3. Steady-State Analysis for Model I

In this section, the queuing inventory system considered in Model I is analyzed in the steady-state. The condition for ergodicity for such a queuing inventory system is found and steady-state probability vectors of the system states are derived.

3.1. Ergodicity Condition

$$\mathcal{B} = \begin{bmatrix} F+C_q+A_q & C_{q-1}+A_1 & C_{q-2}+A_2 & \cdots & C_1+A_{q-1} \\ C_1+A_{q-1} & F+C_q+A_q & C_{q-1}+A_1 & \cdots & C_2+A_{q-2} \\ \vdots & \vdots & \vdots & \vdots & \\ C_{q-1}+A_1 & C_{q-2}+A_2 & C_{q-3}+A_3 & \cdots & F+C_q+A_q \end{bmatrix} \tag{10}$$

Let $\mathbf{y} = (\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_q)$ denote the steady-state probability vector of generator $\mathcal{B} = B_0 + B_1 + B_2$. The matrix is block-circulant, and hence, the solution to equations,

$$\mathbf{y}\mathcal{B} = \mathbf{0}, \mathbf{ye} = 1 \tag{11}$$

is given by

$$\mathbf{y} = \frac{1}{q}(\mathbf{e}'(q) \otimes \mathbf{v}), \tag{12}$$

where $\mathbf{v}$ is a solution to the equation $\mathbf{v}(F + C_q + A_q + C_1 + A_{q-1} + .. + C_{q-1} + A_1) = \mathbf{0}$.

The *LIQBD* description of the model indicates that the queuing system is stable if and only if the left drift rate exceeds that of the right drift [20]. That is, $\mathbf{y}B_0e < \mathbf{y}B_2e$.

Therefore, the given system is stable if and only if

$$\mathbf{v}.\sum_{i=1}^{q}(iA_i.\mathbf{e}) < \mathbf{v}.\sum_{i=1}^{q}(iC_i.\mathbf{e}). \tag{13}$$

3.2. Steady-State Probability Vector

Let $\mathbf{x}$ be the steady-state probability vector of $\mathcal{Q}_1$. We partition this vector as

$$\mathbf{x} = (\mathbf{x_0}, \mathbf{x_1}, \mathbf{x_2}, \ldots).$$

Under the stability condition, we have

$$\mathbf{x_i} = \mathbf{x_1}R^{i-1}, i \geq 2 \tag{14}$$

where the matrix R is the minimal nonnegative solution to the matrix quadratic equation,

$$R^2B_2 + RB_1 + B_0 = \mathbf{0}. \tag{15}$$

The vectors $\mathbf{x}_0$ and $\mathbf{x}_1$ are obtained by solving the equations

$$\begin{aligned} \mathbf{x}_0B_{00} + \mathbf{x}_1B_{10} &= \mathbf{0}, \\ \mathbf{x}_0B_{01} + \mathbf{x}_1(B_1 + RB_2) &= \mathbf{0}, \end{aligned} \tag{16}$$

subject to the normalizing condition:

$$\mathbf{x}_0\mathbf{e} + \mathbf{x}_1(I - R)^{-1}\mathbf{e} = 1. \tag{17}$$

4. System Characteristics for Model I

In this section, a few distribution functions governing the system as well as some of the performance measures are computed. These are of importance as they throw light on the system performance. For the queuing-inventory model under consideration, a vacation cycle refers to the time period starting from the instant at which the service process is frozen due to a lack of quorum to initiate the next batch service as per the *MC* rule for service batch sizes (vacation begins) to the instant when the number of customers exceeds N or the next service is initiated.

4.1. Distribution of Inventory Processing Time in a Vacation Cycle

Once the service process is frozen due to a lack of quorum for initiating the next batch service (as service batch sizes are specified by the *MC* with *tpm* Q), vacation begins.

The server processes inventory until the number of customers in the system exceeds N or the inventory level reaches L. Thus, the inventory processing time in a vacation cycle is the time until the Markov process $(N_1(t), N_2(t), A(t), K_1(t), M(t) : t \geq 0)$ on state space, $\{(n, n', n_1, k_1, k_3) : 0 \leq n \leq N-1, 0 \leq n' \leq L-1, 1 \leq n_1 \leq a, 1 \leq k_1 \leq t_1, 1 \leq k_3 \leq m\}$ is absorbed in state $\{*_1\}$, indicating that the number of customers in the system exceeds N or state $\{*_2\}$, indicating that the inventory level hits L. The trnsition rate submatrices are given in Table 8.

The infinitesimal generator for this $CTMC$ is

$$\mathcal{Q}_2 = \begin{bmatrix} O & O_1 & O_2 \\ \mathbf{0} & 0 & 0 \end{bmatrix} \tag{18}$$

where

$$O = \begin{bmatrix} O' & J_1 & J_2 & \cdots & J_{N-1} \\ & O' & J_1 & \cdots & J_{N-2} \\ & & \ddots & & \vdots \\ & & & & O' \end{bmatrix} \tag{19}$$

$$O_1 = \begin{bmatrix} e(L) \otimes G_N \\ e(L) \otimes (G_N + G_{N-1}) \\ \vdots \\ e(L) \otimes (G_N + G_{N-1} + ...G_1) \end{bmatrix} \tag{20}$$

$$O_2 = \begin{bmatrix} e_L(L) \otimes e(a) \otimes T^0 \otimes e(m) \\ e_L(L) \otimes e(a) \otimes T^0 \otimes e(m) \\ \vdots \\ e_L(L) \otimes e(a) \otimes T^0 \otimes e(m) \end{bmatrix} \tag{21}$$

$$O' = \begin{bmatrix} I_a \otimes (T \oplus D_0) & I_a \otimes (T^0\alpha \otimes I_m) & & \\ & I_a \otimes (T \oplus D_0) & I_a \otimes (T^0\alpha \otimes I_m) & \\ & \ddots & \ddots & I_a \otimes (T^0\alpha \otimes I_m) \\ & & & I_a \otimes (T \oplus D_0) \end{bmatrix} \tag{22}$$

$$\begin{aligned} J_i &= I_L \otimes P.E_{ii}(a) \otimes I_{t_1} \otimes D_1, \\ J_i &= 0; i > a. \end{aligned} \tag{23}$$

$$G_j = \begin{bmatrix} e(t_1) \otimes \delta_{1j} \\ e(t_1) \otimes \delta_{2j} \\ \vdots \\ e(t_1) \otimes \delta_{aj} \end{bmatrix} \tag{24}$$

$$G_j = 0; j > a \tag{25}$$

and δ_{ij} is the column vector with entries as the sum of rows of $p_{ij}D_1$.

Table 8. Transition rate submatrices.

From	To	Rate Matrix
(n, n', n_1)	$(n + n_1', n', n_1')$	$I_{t_1} \otimes p_{n_1 n_1'} D_1$
(n, n', n_1)	(n, n', n_1)	$T \oplus D_0$
(n, n', n_1)	$(n, n' + 1, n_1)$	$T^0 \alpha \otimes I_m$
$(n, L-1, n_1)$	$\{*_2\}$	$T^0 \otimes e_m$
$(N-j, n', n_1)$	$\{*_1\}$	$(\sum_{N-j}^{N} G_j)$

The initial probability vector of this infinitesimal generator is $\beta_1 = \frac{1}{h_1}(\sum_{n_2,k_2} x_{000010(n_2)1k_21}, \cdots \sum_{n_2,k_2} x_{0000a0(n_2))t_1k_2m}, \cdots \sum_{n_2,k_2} x_{00L-10a0(n_2)t_1k_2m}, \cdots \sum_{n_2,k_2} x_{0N-1L-1010(n_2)1k_2m} \cdots \sum_{n_2,k_2} x_{0N-1L-10a0(n_2)1k_2m} \cdots \sum_{n_2,k_2} x_{0N-1L-10a0(n_2)t_1k_2m})$, with $h_1 = \sum_{q,n',n_1,n_2,k_1,k_2,k_3} x_{0qn'0n_10(n_2)k_1k_2k_3}$.

Lemma 1. *If the inventory level reaches L before the expiry of vacation, the expected inventory processing time in a vacation cycle is* $\beta_1(-O)^{-2}O_2$. *Otherwise, the expected inventory processing time in a vacation cycle is* $\beta_1(-O)^{-2}O_1$.

4.2. Distribution of Idle Time in a Vacation Cycle

The idle time of the server is 0 if the number of customers in the system exceeds N before the inventory level hits L. To study the distribution of idle time in a vacation cycle (i.e., time until the number of customers in the system exceeds N after L items are processed), consider the $MC, (N_1(t), L, A(t), M(t) : t \geq 0)$ on state space $\{(n, L, n_1, k_3) : 0 \leq n \leq N-1, 1 \leq n_1 \leq a, 1 \leq k_3 \leq m\} \cup \{*_1\}$, where $\{*_1\}$ denotes the absorbing state and the number of customers in the system exceeds N. The transition rate submatrices are indicated in Table 9.

The infinitesimal generator of this Markov chain is

$$\mathcal{Q}_3 = \begin{bmatrix} U & U^0 \\ \mathbf{0} & 0 \end{bmatrix} \tag{26}$$

where

$$U = \begin{bmatrix} I_a \otimes D_0 & L_1 & L_2 & \cdots & L_{N-1} \\ & I_a \otimes D_0 & L_1 & \cdots & L_{N-2} \\ & & \ddots & & \vdots \\ & & & & I_a \otimes D_0 \end{bmatrix} \tag{27}$$

$$U^0 = \begin{bmatrix} \gamma_N \\ \gamma_N + \gamma_{N-1} \\ \vdots \\ \gamma_N + \gamma_{N-1} + \ldots \gamma_1 \end{bmatrix} \tag{28}$$

$$\gamma_j = \begin{bmatrix} \delta_{1j} \\ \delta_{2j} \\ \vdots \\ \delta_{aj} \end{bmatrix} \tag{29}$$

$$\begin{aligned} L_i &= P.E_{ii}(a) \otimes D_1, \\ L_i &= 0; i > a \end{aligned} \tag{30}$$

Table 9. Transition rate submatrices.

From	To	Rate Matrix
(n, L, n_1)	$(n + n_1', L, n_1')$	$p_{n_1 n_1'} D_1$
(n, L, n_1)	(n, L, n_1)	$I_a \otimes D_0$
$(N - j, L, n_1)$	$\{*_1\}$	$\sum_{N-j}^{N} \gamma_j$

The conditional distribution of idle time in a vacation cycle follows $PH(\beta_2, U)$ distribution with initial probability vector $\beta_2 = \frac{1}{h_2}(\sum_{n_2,k_2} x_{00L010(n_2)k_21}, \cdots \sum_{n_2,k_2} x_{00L010(n_2)k_2m}, \cdots \sum_{n_2,k_2} x_{00L0a0(n_2)k_2m}, \cdots \sum_{n_2,k_2} x_{0N-1L0a0(n_2)k_21} \cdots \sum_{n_2,k_2} x_{0N-1L0a0(n_2)k_2m})$, with $h_2 = \sum_{q,n_1,n_2,k_2,k_3} x_{0qL0n_10(n_2)k_2k_3}$.

Lemma 2. *The conditional expectation of idle time in a vacation cycle, given the inventory level, reaches L only after the number of customers in system exceeds N and is* $\beta_2(-U)^{-1}e$. *The expected idle time in a vacation cycle is* $\beta_2(-U)^{-1}e \times (\int_0^\infty \beta_1 e^{Ot} O_2 dt)$.

4.3. Distribution of Vacation Cycle Length

The vacation cycle length can be studied as an MC $(N_1(t), N_2(t), A(t), K_1(t), M(t) : t \geq 0)$ on state-space $\{(n, n', n_1, k_1, k_3) : 0 \leq n \leq N-1, 0 \leq n' \leq L-1, 1 \leq n_1 \leq a, 1 \leq k_1 \leq t_1, 1 \leq k_3 \leq m\} \cup \{(n, L, n_1, k_3) : 0 \leq n \leq N-1, 1 \leq n_1 \leq a, 1 \leq k_4 \leq m\} \cup \{*_1\}$, where $\{*_1\}$ denotes the absorbing state and the number of customers in the system exceeds N. The transition rate submatrices are indicated in Table 10.

The infinitesimal generator of this Markov chain is

$$\mathcal{Q}_4 = \begin{bmatrix} W & W^0 \\ \mathbf{0} & 0 \end{bmatrix} \tag{31}$$

where,

$$W = \begin{bmatrix} H & V_1 & V_2 & \cdots & V_{N-1} \\ & H & V_1 & \cdots & V_{N-2} \\ & & \ddots & & \vdots \\ & & & & H \end{bmatrix} \tag{32}$$

$$W^0 = \begin{bmatrix} W_N \\ W_{N-1} \\ \vdots \\ W_1 \end{bmatrix} \tag{33}$$

$$V_i = \begin{bmatrix} I_L \otimes P.E_{ii}(a) \otimes I_{t_1} \otimes D_1 & \mathbf{0} \\ \mathbf{0} & P.E_{ii}(a) \otimes D_1 \end{bmatrix} \tag{34}$$

$$H = \begin{bmatrix} I_a \otimes (T \oplus D_0) & I_a \otimes (T^0\alpha \otimes I_m) & & & \\ & I_a \otimes (T \oplus D_0) & I_a \otimes (T^0\alpha \otimes I_m) & & \\ & & \ddots & \ddots & \\ & & & I_a \otimes (T \oplus D_0) & I_a \otimes (T^0 \otimes I_m) \\ & & & & I_a \otimes D_0 \end{bmatrix} \tag{35}$$

$$W_i = \begin{bmatrix} e(L) \otimes (G_N + G_{N-1} + ...G_{N-(i-1)}) \\ \gamma_N + \gamma_{N-1} + ...\gamma_{N-(i-1)} \end{bmatrix} \tag{36}$$

Table 10. Transition rate submatrices.

From	To	Rate Matrix
(n, n', n_1)	$(n + n_1', n', n_1')$	$I_{t_1} \otimes p_{n_1 n_1'} D_1$
(n, n', n_1)	(n, n', n_1)	$T \oplus D_0$
(n, n', n_1)	$(n, n' + 1, n_1)$	$T^0 \alpha \otimes I_m$
$(n, L-1, n_1)$	(n, L, n_1)	$T^0 \otimes I_m$
(n, L, n_1)	$(n + n_1', L, n_1')$	$p_{n_1 n_1'} D_1$
(n, L, n_1)	(n, L, n_1)	D_0
$(N-j, n', n_1)$	$\{*_1\}$	$\sum_{N-j}^{N} G_j$
$(N-j, L, n_1)$	$\{*_1\}$	$\sum_{N-j}^{N} \gamma_j$

Thus, the distribution of vacation cycle length follows $PH(\beta_3, W)$ distribution with initial probability vector $\beta_3 = \frac{1}{h_3}(\sum_{n_2,k_2} x_{000010(n_2)1k_21}, \cdots \sum_{n_2,k_2} x_{000010(n_2)0(n_2)1k_2m}, \cdots \sum_{n_2,k_2} x_{000010(n_2)t_1k_2m}, \cdots \sum_{n_2,k_2} x_{0000a0(n_2)1k_21} \cdots \sum_{n_2,k_2} x_{0000a0(n_2)0(n_3)t_1k_2m}, \cdots \sum_{n_2,k_2} x_{00L0a0(n_2)t_1k_2m}, \cdots \sum_{n_2,k_2} x_{0N-1L0a0(n_2)1k_2m} \cdots \sum_{n_2,k_2} x_{0N-1L0a0(n_2)t_1k_2m})$ with, $h_3 = \sum_{q,n',n_1,n_2,k_1,k_2,k_3} x_{0qn'0n_10(n_2)k_1k_2k_3}$.

Lemma 3. *The expected vacation cycle length is* $\beta_3(-W)^{-1}e$.

4.4. Other Performance Measures

To study the qualitative behavior of the queuing-inventory system considered in this paper, the formulas for some key performance measures are derived. The usefulness of the system is analyzed or compared using these measures.

1. Expected queue length:

$$E_{QL} = \sum_{l=0}^{\infty} \sum_{p=0}^{q-1} (lq + p) x_{(l,p)} \mathbf{e}. \tag{37}$$

where $n = lq + p$ is the number of customers in the queue and $x_{(l,p)}$ is the probability that the system is found in super-state n.

2. Expected number of inventory available:

$$\begin{aligned} E_I = \sum_{n'=1}^{L-1} n' [& \sum_{p=0}^{q-1} \sum_{n_1=1}^{a} \sum_{n_2=1}^{N} \sum_{k_1=1}^{t_1} \sum_{k_2=1}^{t_2} \sum_{k_3=1}^{m} x_{(0,p,n',0,n_1,0(n_2),k_1,k_2,k_3)} + \sum_{l \geq 0} \sum_{p=0}^{q-1} \sum_{n_1=1}^{a} \sum_{n_2=1}^{N} \sum_{k_2=1}^{t_2} \sum_{k_3=1}^{m} x_{(l,p,n',1,n_1,n_2,k_2,k_3)}] \\ + L [& \sum_{p=0}^{q-1} \sum_{n_1=1}^{a} \sum_{n_2=1}^{N} \sum_{k_2=1}^{t_2} \sum_{k_3=1}^{m} x_{(0,p,L,0,n_1,0(n_2),k_2,k_3)} + \sum_{l \geq 0} \sum_{p=0}^{q-1} \sum_{n_1=1}^{a} \sum_{n_2=1}^{N} \sum_{k_2=1}^{t_2} \sum_{k_3=1}^{m} x_{(l,p,L,1,n_1,n_2,k_2,k_3)}], \end{aligned} \tag{38}$$

where $\sum_{p=0}^{q-1} \sum_{n_1=1}^{a} \sum_{n_2=1}^{N} \sum_{k_1=1}^{t_1} \sum_{k_2=1}^{t_2} \sum_{k_3=1}^{m} x_{(0,p,n',0,n_1,0(n_2),k_1,k_2,k_3)}$ is the probability that the system is found in a state with n' items in stock while on vacation and processing inventory, $\sum_{p=0}^{q-1} \sum_{n_1=1}^{a} \sum_{n_2=1}^{N} \sum_{k_2=1}^{t_2} \sum_{k_3=1}^{m} x_{(0,p,L,0,n_1,0(n_2),k_2,k_3)}$ is the probability that the system is in a state with inventory L while the server is idle, and $\sum_{l \geq 0} \sum_{p=0}^{q-1} \sum_{n_1=1}^{a} \sum_{n_2=1}^{N} \sum_{k_2=1}^{t_2} \sum_{k_3=1}^{m} x_{(l,p,n',1,n_1,n_2,k_2,k_3)}$ is the probability that the system is found in a state with inventory n' while serving customers.

3. Probability that the server is on vacation and processing inventory:

$$P_{SVI} = \sum_{p=0}^{q-1} \sum_{n'=1}^{L-1} \sum_{n_1=1}^{a} \sum_{n_2=1}^{N} \sum_{k_1=1}^{t_1} \sum_{k_2=1}^{t_2} \sum_{k_3=1}^{m} x_{(0,p,n',0,n_1,0(n_2),k_1,k_2,k_3)}, \tag{39}$$

where $x_{(0,p,n',0,n_1,0(n_2),k_1,k_2,k_3)}$ is the probability that the system is in a state with p customers in queue and processing inventory (inventory processing phase k_1).

4. Probability that the server is idle:

$$P_I = \sum_{p=0}^{q-1} \sum_{n_1=1}^{a} \sum_{n_2=1}^{N} \sum_{k_2=1}^{t_2} \sum_{k_3=1}^{m} x_{(0,p,L,0,n_1,0(n_2),k_2,k_3)}, \tag{40}$$

where $x_{(0,p,L,0,n_1,0(n_2),k_2,k_3)}$ is the probability that the system is in a state with p customers in queue and L items in inventory.

5. Probability that the server is on vacation:

$$P_V = P_{SVI} + P_I. \tag{41}$$

6. Fraction of time that the server is busy serving a batch of n_2 customers without having inventory at the commencement of service:

$$TW_{n_2} = \sum_{l=0}^{\infty} \sum_{p=0}^{q-1} \sum_{n_1=1}^{a} \sum_{k_2=1}^{t_2} \sum_{k_3=1}^{m} x_{(l,p,0,1,n_1,n_2,k_2,k_3)}, \tag{42}$$

where $x_{(l,p,0,1,n_1,n_2,k_2,k_3)}$ is the probability that the system is in a state with $n = lq + p$ customers in queue and n_2 customers are served without inventory while in service phase k_2.

7. Fraction of time that the server serves without inventory:

$$T_{WI} = \sum_{n_2=1}^{N} TW_{n_2} \tag{43}$$

8. Fraction of time that the server is busy serving a batch of n_2 customers with inventory available at the beginning of service:

$$TI_{n_2} = \sum_{l=0}^{\infty} \sum_{p=0}^{q-1} \sum_{n'=1}^{L} \sum_{n_1=1}^{a} \sum_{k_2=1}^{t_2} \sum_{k_3=1}^{m} x_{(l,p,n',1,n_1,n_2,k_2,k_3)}, \tag{44}$$

where $x_{(l,p,n',1,n_1,n_2,k_2,k_3)}$ is the probability that the system is in a state with $n = lq + p$ customers in queue and n_2 customers served (when n' items are in stock) while in service phase k_2.

9. Fraction of time that the server serves with inventory:

$$T_I = \sum_{n_2=1}^{N} TI_{n_2} \tag{45}$$

5. Numerical Example for Model I

The applicability of the results derived earlier is illustrated using a few examples. In all these examples, it is assumed that the arrival process is a $BMAP$ with representation (D_0, D_1) and maximum arrival batch size $a = 2$. The successive arrival batch sizes form a Markov chain with tpm, P. The service process is a $BMSP$ with representation (S_0, S_1) and maximum service batch size $N = 3$. The successive service batch sizes form a Markov chain with tpm, Q.

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.8 & 0.2 \end{bmatrix}, Q = \begin{bmatrix} 0.1 & 0.2 & 0.7 \\ 0.3 & 0.1 & 0.6 \\ 0.4 & 0.1 & 0.5 \end{bmatrix},$$

The server goes on vacation if there are not enough customers in queue to initiate the next batch service (for example, suppose only 1 customer is in queue and, if the next service batch size as per the MC rule is 2, the server goes on vacation). During vacation, it processes

inventory until L items are processed (finite storehouse capacity). The vacation expires when N customers accumulate in the queue. The server remains idle once the inventory level hits L before the number in the queue reaches N until the end of the vacation.

Example 1. *Consider a queuing-inventory system where the arrival process is a BMAP with a mean arrival rate of a batch of customers as 1 and with* (D_0, D_1)*:*

1. *BMAP with correlation in successive inter-arrival times (CA):*

$$D_0 = \begin{bmatrix} -2.2444 & 0.0673 \\ 0.0374 & -0.4489 \end{bmatrix}, D_1 = \begin{bmatrix} 2.0948 & 0.0823 \\ 0.0374 & 0.3741 \end{bmatrix}$$

2. *BMAP with uncorrelated successive inter-arrival times (UA):*

$$D_0 = \begin{bmatrix} -2 & 2 \\ 0 & -2 \end{bmatrix}, D_1 = \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix}$$

The first arrival process (CA) has correlated inter-arrival times (the correlation coefficient between successive inter-arrival times is 0.2245*). Successive inter-arrival times are independent for the second arrival process (UA).*

Similarly, the service process BMSP is normalized with the mean service rate of a batch of customers as 1 and (S_0, S_1)*:*

1. *BMSP with correlation in successive service times (CS):*

$$S_0 = \begin{bmatrix} -2.1738 & 0.0072 \\ 0.0072 & -0.4347 \end{bmatrix}, S_1 = \begin{bmatrix} 2.1449 & 0.0217 \\ 0.0072 & 0.4203 \end{bmatrix}$$

2. *BMSP with uncorrelated successive service times (US):*

$$S_0 = \begin{bmatrix} -1 & 0 \\ 0 & -3.5 \end{bmatrix}, S_1 = \begin{bmatrix} 1 & 0 \\ 1.75 & 1.75 \end{bmatrix}$$

The first service process (CS) has correlated successive service times (the correlation coefficient of successive service times is 0.2792*.) For the second process (US), successive service times are independent.*

The inventory processing time follows $PH(\alpha, T), \alpha = [0.6, 0.4]$*,*

$$T = \begin{bmatrix} -2 & 2 \\ 0 & -2 \end{bmatrix}$$

The mean inventory processing time is 0.8.

The maximum inventory produced in a vacation cycle is $L = 4$*.*

Assume $\theta = 0.3$*, i.e., the service rate without any inventory is 0.3 times the service rate with inventory.*

First, the effect of ρ*, the traffic intensity or the mean number of arrivals during an average service time (obtained by varying arrival and service rates) on performance measures,*

- *expected queue length,* E_{QL}*;*
- *expected number of inventory available,* E_I*;*
- *probability that the server is on vacation and processing inventory,* P_{SVI}*; and*
- *probability that the server is idle,* P_I

under four circumstances,

1. *correlated arrival and correlated service (CACS),*
2. *uncorrelated arrival and correlated service (UACS),*
3. *uncorrelated arrival and uncorrelated service (UAUS), and*
4. *correlated arrival and uncorrelated service (CAUS)*

is studied using graphs.

The following can be seen from Figure 3:

- As ρ increases, the expected queue length increases (see Figure 3a). Hence, the server goes on vacation less often (see Figure 3c,d) and the expected inventory available decreases (see Figure 3b). The server has to start service without inventory and has to process inventory first before offering services. This slows the service process further and increases queue length. The server's idle time decreases with the increase in ρ, as expected.
- The increase in expected queue length is most remarkable for correlated arrival and service process (indicated in red). In contrast, if both arrival and service processes are uncorrelated (indicated in blue), the queue length increases very slowly, and hence, the server goes on vacation more often and stays idle for a longer time in comparison.

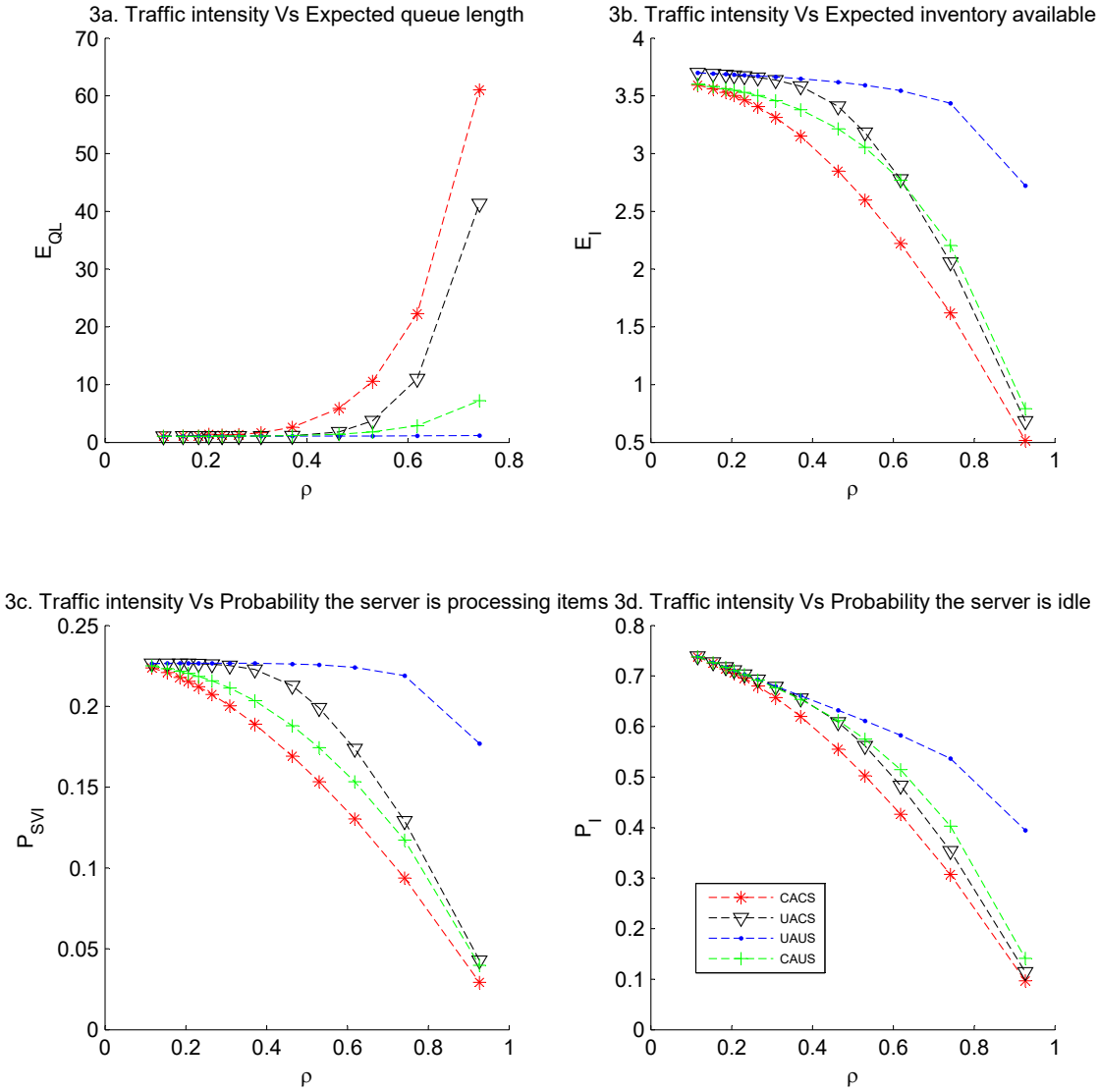


Figure 3. Effect of ρ and traffic intensity on performance measures.

The behavior of the queuing-inventory system considered in this paper depends on the arrival and service process (in particular, the *tpm*'s P and Q, respectively). However, even then, ρ could be effectively used to analyze system behavior, since for varying values of ρ, the values of performance measures increase or decreases almost identically.

6. Cost Analysis for Model I

Based on performance measures, a cost function is constructed for the queuing inventory model under consideration:

$$C = C_{QL} \times E_{QL} + C_{IP} \times E_{IP} \times P_{SVI} + C_{HPI} \times E_I + C_I \times P_I + \sum_{j=1}^{3} TW_j \times CW_j + \sum_{j=1}^{3} TI_j \times C_j \quad (46)$$

where
C_{QL}: holding cost for retaining a customer in queue per unit time
C_{IP}: cost for producing unit inventory per unit time
E_{IP}: expected inventory produced per unit time
C_{HPI}: holding cost per inventoried item per unit time
C_I: cost for remaining idle per unit time
CW_j: cost per unit time for offering services to a batch of j customers without inventory (this includes cost for the production of inventory required for service)
C_j: cost per unit time for offering services to a batch of j customers with inventory at a service commencement epoch.

The objective is to find an L, the maximum number of items that are to be processed during vacation that minimizes the cost function. With the increase in L, the fraction of time that the server serves with inventory increases, considerably decreasing the length of the queue. The increase in L increases the overall cost of processing items as well as the holding cost of processed items. Consider 3 types of costs for offering service with or without inventory while fixing $C_{QL} = 1, C_{IP} = 2, C_{HPI} = 1, C_I = 1$. Higher values are given for CW_j, as the server needs to process inventory (which involves a cost) before the start of service.

1. A linear cost for offering service with or without inventory, $(CW_1 = 8, CW_2 = 16, CW_3 = 24, C_1 = 5, C_2 = 10, C_3 = 15)$
2. A linear cost for offering service with inventory and a nonlinear cost for offering services without inventory, $(CW_1 = 8, CW_2 = 64, CW_3 = 512, C_1 = 5, C_2 = 10, C_3 = 15)$
3. A nonlinear cost for offering service with or without inventory, $(CW_1 = 8, CW_2 = 64, CW_3 = 512, C_1 = 5, C_2 = 25, C_3 = 125)$

For Example 2, the costs are as follows (Table 11).

Table 11. Effect of L on the cost function.

$L \downarrow$	Cost 1	Cost 2	Cost 3
2	**7.6402**	19.8157	24.3381
3	8.2056	18.1748	23.3437
4	8.8102	17.3107	22.9285
5	9.4440	16.8325	**22.7956**
6	10.1039	16.6005	22.8427
7	10.7883	**16.5465**	23.0208
8	11.4952	16.6302	23.3009
9	12.2229	16.8246	23.6638

The minimum cost is indicated by bold font. The value of L, in the case of linear costs for service, is the least. For costs 2 and 3, the value decreases first with the increase in L value and, on reaching a minimum, starts climbing up with further increase in the value of L. Of course, these are input-specific.

7. Description and Formulation of Model II

This model differs from the one discussed in Sections 2–6, in the following respects:

(i) The Markov dependence between two consecutive arriving batch sizes is taken out in Model II and the Markov dependence between two consecutive service batch sizes. (ii) This results in service commencement of the next batch immediately after completion of the current batch service, provided that at least one customer is waiting for service. Otherwise, the server stays idle/starts processing items for future services. Accommodating Markov dependence of the successive arrival batch sizes and between successive service batches introduced enormous complexity in the analysis of Model I. Naturally, the effect of the Markov dependence in the service batch sizes is the increase in idle time of the server. This and other distinctions in the performance of the two models are illustrated through numerical examples.

Now, we give a detailed description of Model II. Consider a single-server queuing-inventory system with customers arriving according to a batch Markovian arrival process ($BMAP$) with maximum arrival batch size a and representation $\{D_0, D_1 ... D_a\}$. The service time duration is based on whether items are available at a service initiation epoch. The service process follows batch Markovian service process ($BMSP$) with representation $\{S_0, S_1, ... S_N\}$. The server goes on vacation when the number of customers in the system is 0. The other assumptions remain the same as in Model I.

The arrival process is defined using matrices $\{D_0, D_1, ... D_a\}$ of order m. Here, the entries of D_0 denote transition rates of underlying MC of $BMAP$ without arrivals, and the entries of $\{D_i; i > 0\}$ denote transition rates of underlying MC of $BMAP$ with batch arrival of size i. The service process is defined using matrices $\{S_0, S_1, ... S_N\}$ of order t_2, where entries of S_0 denote transition rates of underlying MC of $BMSP$ without departures and $\{S_j; j > 0\}$ denotes transition rates of underlying MC of $BMSP$ with departures of size j. In the formulation of Model II, the number of customers in the system is considered rather than the number in the queue as successive service batch sizes are independent and there is no need to specify service batch size at a given epoch t.

Let $N_1(t)$ be the number of customers in the system at time t. We write $N_1(t) = n$ as $(l; p)$ if $n = lq + p; l \geq 0; 0 \leq p \leq q - 1$, where $q = max\{a, N\}$. Here, we redefined the level to obtain the $LIQBD$ structure for the generator matrix.

Let $N_2(t)$ be the number of processed inventory available at t.

Let $J(t)$ be the status of the server at t.

$$J(t) = \begin{cases} 0, & if\ the\ server\ is\ on\ vacation \\ 1, & if\ the\ server\ is\ busy \end{cases}$$

$K_1(t)$ is the phase of inventory processing.
$K_2(t)$ is the state of an underlying MC of $BMSP$.
$M(t)$ is the state of an underlying MC of $BMAP$.

The above model can be studied using a $CTMC$, $\{(N_1(t), N_2(t), J(t), K_1(t), K_2(t), M(t)) : t \geq 0\}$ on state-space $\Omega_0 \cup_l \Omega_{l \geq 1}$.

Here, $1 \leq n' \leq L, 1 \leq k_1 \leq t_1, 1 \leq k_2 \leq t_2, 1 \leq k_3 \leq m$. The states in Table 12 correspond to level 0, i.e., to the states with $0, 1, 2 ... q - 1$ customer(s) in the system. Table 13 corresponds to level l, i.e., to the states with $lq, lq + 1, lq + 2 ... (l + 1)q - 1$ customer(s) in the system. The infinitesimal generator of this $CTMC$ is an $LIQBD$ when we redefine the level as described above:

$$\mathcal{Q}_5 = \begin{bmatrix} B'_{00} & B'_{01} & & & & \\ B'_{10} & B'_1 & B'_0 & & & \\ & B'_2 & B'_1 & B'_0 & & \\ & & B'_2 & B'_1 & B'_0 & \\ & & & \ddots & \ddots & \ddots \end{bmatrix} \tag{47}$$

$$B_2' = \begin{bmatrix} C_q' & C_{q-1}' & C_{q-2}' & \cdots & C_1' \\ & C_q' & C_{q-1}' & \cdots & C_2' \\ & & & \ddots & \vdots \\ & & & & C_q' \end{bmatrix} \tag{48}$$

in which, for $i = 1, 2, ..q$,

$$C_i' = \begin{bmatrix} \theta S_i \otimes I_m & & & ... & \mathbf{0} \\ S_i \otimes I_m & & & ... & \mathbf{0} \\ & S_i \otimes I_m & & ... & \mathbf{0} \\ & & \ddots & & \vdots \\ & & & S_i \otimes I_m & \mathbf{0} \end{bmatrix} \tag{49}$$

$$C_i' = \mathbf{0}; i > N; \tag{50}$$

$$B_0' = \begin{bmatrix} A_q' & & & & \\ A_{q-1}' & A_q' & & & \\ \vdots & \vdots & & \ddots & \\ A_1' & A_2' & A_3' & \cdots & A_q' \end{bmatrix} \tag{51}$$

For $i = 1, 2, ..q$,

$$\begin{aligned} A_i' &= I_{L+1} \otimes I_{t_2} \otimes D_i; \\ A_i' &= \mathbf{0}, i > a. \end{aligned} \tag{52}$$

$$B_1' = \begin{bmatrix} F' & A_1' & A_2' & \cdots & A_{q-1}' \\ C_1' & F' & A_1' & \cdots & A_{q-2}' \\ \vdots & \vdots & \vdots & \vdots & \\ C_{q-1}' & C_{q-2}' & C_{q-3}' & \cdots & F' \end{bmatrix} \tag{53}$$

in which

$$F' = \begin{bmatrix} (\theta S_0 \oplus D_0) & & & \\ & (S_0 \oplus D_0) & & \\ & & \ddots & \\ & & & (S_0 \oplus D_0) \end{bmatrix} \tag{54}$$

Table 12. States in Ω_0 and their descriptions.

Sl. No	State	Description
1	$(0, p, n', 0, k_1, k_2, k_3)$ $0 \le p < N; 0 \le n' \le L-1$	Service process is frozen with the server on vacation producing inventory for future use
2	$(0, p, L, 0, k_2, k_3)$ $0 \le p < N$	Server is idle as the maximum inventory level is reached and the system is on vacation
3	$(0, p, 0, 1, k_2, k_3)$ $0 \le p \le q-1$	Service process without inventory (at the commencement epoch of the current service) is ongoing
4	$(0, p, n', 1, k_2, k_3)$	Service process with inventory is ongoing

Table 13. States in $\Omega_l; l \geq 1$ and their descriptions.

Sl. No	State	Description
1	$(l, p, 0, 1, k_2, k_3)$ $0 \leq p \leq q-1$	Service process without inventory (at the commencement epoch of the current service) is going on
2	$(l, p, n', 1, k_2, k_3)$ $0 \leq p \leq q-1$	Service process with inventory is ongoing

The transition rate submatrices amongst the various levels are provided in Tables 14–19.

Table 14. Transition rate submatrices from level 0 to itself.

From	To	Rate Matrix
$(0, p, n', 0))$	$(0, p, n', 0)$	$T \oplus (I_{t_2} \otimes D_0)$
$(0, p, L, 0)$	$(0, p, L, 0)$	$I_{t_2} \otimes D_0$
$(0, p, 0, 1)$	$(0, p, 0, 1)$	$\theta S_0 \oplus D_0$
$(0, p, n', 1)$	$(0, p, n', 1)$	$S_0 \oplus D_0$
$(0, p, n', 0)$	$(0, p, n'+1, 0)$	$T^0 \alpha \otimes I_{t_2 m}$
$(0, p, L-1, 0)$	$(0, p, L, 0)$	$T^0 \otimes I_{t_2 m}$
$(0, p, n', 0)$	$(0, p+i, n', 0)$	$I_{t_1 t_2} \otimes D_i$
$(0, p, n', 0)$	$(0, p+i, n', 1)$	$e(t_1) \otimes I_{t_2} \otimes D_i$
$(0, p, L, 0)$	$(0, p+i, L, 0)$	$I_{t_2} \otimes D_i$
$(0, p, L, 0)$	$(0, p+i, L, 1)$	$I_{t_2} \otimes D_i$
$(0, p, n', 1)$	$(0, p+i, n', 1)$	$I_{t_2} \otimes D_i$
$(0, p, 0, 1)$	$(0, p-j, 0, 1)$	$\theta S_j \otimes I_m$
$(0, p, 0, 1)$	$(0, p-j, 0, 0)$	$\alpha \otimes \sum_{k=j}^{N} \theta S_k \otimes I_m$
$(0, p, n', 1)$	$(0, p-j, n'-1, 0)$	$\alpha \otimes \sum_{k=j}^{N} S_k \otimes I_m$
$(0, p, n', 1)$	$(0, p-j, n'-1, 1)$	$S_j \otimes I_m$

Table 15. Transition rate submatrices from level 0 to 1.

From	To	Rate Matrix
$(0, p, n', 0)$	$(1, p+i-q, n', 1)$	$e(t_1) \otimes I_{t_2} \otimes D_i$
$(0, p, L, 0)$	$(1, p+i-q, L, 1)$	$I_{t_2} \otimes D_i$
$(0, p, n', 1)$	$(1, p+i-q, n', 1)$	$I_{t_2} \otimes D_i$

Table 16. Transition rate submatrices from level 1 to 0.

From	To	Rate Matrix
$(1, p, 0, 1)$	$(0, p+q-j, 0, 1)$	$\theta S_j \otimes I_m$
$(1, p, n', 1)$	$(0, p+q-j, n'-1, 1)$	$S_j \otimes I_m$
$(1, p, 0, 1)$	$(0, p+q-j, 0, 0)$	$\alpha \otimes \sum_{k=j}^{N} \theta S_k \otimes I_m$
$(1, p, n', 1)$	$(0, p+q-j, n'-1, 0)$	$\alpha \otimes \sum_{k=j}^{N} S_k \otimes I_m$

Table 17. Transition rate submatrices from level $l > 1$ to $l-1$.

From	To	Rate Matrix
$(l, p, 0, 1)$	$(l-1, p+q-j, 0, 1)$	$\theta S_j \otimes I_m$
$(l, p, n', 1)$	$(l-1, p+q-j, n'-1, 1)$	$S_j \otimes I_m$

Table 18. Transition rate submatrices from level l to itself.

From	To	Rate Matrix
$(l,p,0,1)$	$(l,p,0,1)$	$\theta S_0 \oplus D_0$
$(l,p,n',1)$	$(l,p,n',1)$	$S_0 \oplus D_0$
$(l,p,0,1)$	$(l,p+i,0,1)$	$I_{t_2} \otimes D_i$
$(l,p,n',1)$	$(l,p+i,n',1)$	$I_{t_2} \otimes D_i$
$(l,p,0,1)$	$(l,p-j,0,1)$	$\theta S_j \otimes I_m$
$(l,p,n',1)$	$(l,p-j,n'-1,1)$	$S_j \otimes I_m$

Table 19. Transition rate submatrices from level $l \geq 1$ to $l+1$.

From	To	Rate Matrix
$(l,p,0,1)$	$(l+1,p+i-q,0,1)$	$I_{t_2} \otimes D_i$
$(l,p,n',1)$	$(l+1,p+i-q,n',1)$	$I_{t_2} \otimes D_i$

About the transitions:

- $(0,p,n',0) \to (1,p+i-q,n',1)$ denotes the transition associated with the arrival of i customers to the system, which activates service.
- $(0,p,n',1) \to (0,p-j,n'-1,1)$ denotes the transition associated with service completion of j customers (with inventory at the service commencement epoch).
- $(l,p,0,1) \to (l,p-j,0,1)$ denotes the transition associated with service completion of j customers without inventory at the service commencement epoch.
- $(0,p,L-1,0) \to (0,p,L,0)$ denotes the transition associated with processing of a unit item while on vacation. The server remains idle as the maximum inventory level is reached.
- $(l,p,0,1) \to (l,p,0,1)$ denotes the transition associated with no service completion or arrival.

8. Steady-State Analysis for Model II

In this section, the condition for ergodicity is investigated for the queuing-inventory system in which successive arrivals as well as service batch sizes are independent (Model II). The steady-state probability vector of system states is derived for the same.

8.1. Ergodicity Condition

$$\mathcal{B}' = \begin{bmatrix} F'+C'_q+A'_q & C'_{q-1}+A'_1 & C'_{q-2}+A'_2 & \cdots & C'_1+A'_{q-1} \\ C'_1+A'_{q-1} & F'+C'_q+A'_q & C'_{q-1}+A'_1 & \cdots & C'_2+A'_{q-2} \\ \vdots & \vdots & \vdots & \vdots & \\ C'_{q-1}+A'_1 & C'_{q-2}+A'_2 & C'_{q-3}+A'_3 & \cdots & F'+C'_q+A'_q \end{bmatrix} \tag{55}$$

The steady state probability vector $\mathbf{z} = (\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_q)$ of the generator $\mathcal{B}' = B'_0 + B'_1 + B'_2$ satisfies

$$\mathbf{z}\mathcal{B}' = \mathbf{0}, \mathbf{z}\mathbf{e} = 1 \tag{56}$$

The matrix $\mathcal{B}'$ is block-circulant, and hence, the solution to the equations is given by

$$\mathbf{z} = \frac{1}{q}(\mathbf{e}'(q) \otimes \mathbf{w}) \tag{57}$$

where $\mathbf{w}$ is a solution to the equation

$$\mathbf{w}(F'+C'_q+A'_q+C'_1+A'_{q-1}+..+C'_{q-1}+A'_1) = \mathbf{0}. \tag{58}$$

$$\mathcal{F}' + C'_q + A'_q + C'_1 + A'_{q-1} + .. + C'_{q-1} + A'_1 = \begin{bmatrix} T_1 & 0 & 0 & 0 & \cdots & 0 \\ T_2 & T_3 & 0 & 0 & \cdots & 0 \\ 0 & T_2 & T_3 & & \cdots & 0 \\ & & \ddots & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & T_2 & T_3 \end{bmatrix} \tag{59}$$

$\mathbf{w} = (w_0, w_1, w_2, ..., w_L)$ satisfy the set of equations:

$$\begin{aligned} w_L T_3 &= 0 \\ w_{L-1} T_3 + w_L T_2 &= 0 \\ w_{L-2} T_3 + w_{L-1} T_2 &= 0 \\ &\vdots \\ w_0 T_1 + w_1 T_2 &= 0 \end{aligned} \tag{60}$$

where the determinants of T_2 and T_3 are nonzero and that of T_1 is zero. Hence, the solution of is of the form $\mathbf{w} = (w_0, 0, 0, ..., 0)$ with w_0 satisfying

$$w_0 T_1 = 0, w_0 \mathbf{e} = 1. \tag{61}$$

The queuing system is stable if and only if the left drift rate exceeds that of the right drift [20]. That is, $\mathbf{z} B'_0 e < \mathbf{z} B'_2 e$.

Therefore, the given system is stable if and only if

$$\mathbf{w}. \sum_{i=1}^{q} (i A'_i . \mathbf{e}) < \mathbf{w}. \sum_{i=1}^{q} (i C'_i . \mathbf{e}). \tag{62}$$

8.2. Steady-State Probability Vector

Let $\mathbf{x}'$ be the steady state probability vector of $\mathcal{Q}_5$. We partition this vector as

$$\mathbf{x}' = (\mathbf{x'_0}, \mathbf{x'_1}, \mathbf{x'_2} \ldots),$$

where $\mathbf{x'_1}, \mathbf{x'_2}, \ldots$ are of dimension $t' = q \times (L+1) \times t_2 \times m$. Under the stability condition, we have

$$\mathbf{x'_i} = \mathbf{x'_1} R^{i-1}, i \geq 2 \tag{63}$$

where the matrix R is the minimal nonnegative solution to the matrix quadratic equation

$$R^2 B'_2 + R B'_1 + B'_0 = \mathbf{0}. \tag{64}$$

The vectors $\mathbf{x}'_0$ and $\mathbf{x}'_1$ are obtained by solving the equations

$$\begin{aligned} \mathbf{x}'_0 B'_{00} + \mathbf{x}'_1 B'_{10} &= \mathbf{0}, \\ \mathbf{x}'_0 B'_{01} + \mathbf{x}'_1 (B'_1 + R B'_2) &= \mathbf{0}, \end{aligned} \tag{65}$$

subject to the normalizing condition,

$$\mathbf{x}'_0 \mathbf{e} + \mathbf{x}'_1 (I - R)^{-1} \mathbf{e} = 1. \tag{66}$$

9. Performance Measures for Model II

In this section, formulas of key performance measures for Model II are presented to aid comparison with the main queuing-inventory system considered in this paper.

1. Expected number of customers in the system:

$$E_{NS} = \sum_{l=0}^{\infty} \sum_{p=0}^{q-1} (lq + p) x'_{(l,p)} \mathbf{e}, \tag{67}$$

where $n = lq + p$ is the number of customers in the system and $x'_{(l,p)}$ is the probability that the system is found in super-state n.

2. Expected number of inventory available:

$$E_I = \sum_{n'=1}^{L-1} n'[\sum_{p=0}^{q-1}\sum_{k_1=1}^{t_1}\sum_{k_2=1}^{t_2}\sum_{k_3=1}^{m} x'_{(0,p,n',0,k_1,k_2,k_3)} + \sum_{l\geq 0}\sum_{p=0}^{q-1}\sum_{k_2=1}^{t_2}\sum_{k_3=1}^{m} x'_{(l,p,n',1,k_2,k_3)}] + L[\sum_{p=0}^{q-1}\sum_{k_2=1}^{t_2}\sum_{k_3=1}^{m} x'_{(0,p,L,0,k_2,k_3)} + \sum_{l\geq 0}\sum_{p=0}^{q-1}\sum_{k_2=1}^{t_2}\sum_{k_3=1}^{m} x'_{(l,p,L,1,k_2,k_3)}], \tag{68}$$

where $\sum_{p=0}^{q-1}\sum_{n_2=1}^{N}\sum_{k_1=1}^{t_1}\sum_{k_2=1}^{t_2}\sum_{k_3=1}^{m} x_{(0,p,n',0,k_1,k_2,k_3)}$ is the probability that the system is found in a state with n' items in stock while on vacation and processing inventory, $\sum_{p=0}^{q-1}\sum_{k_2=1}^{t_2}\sum_{k_3=1}^{m} x_{(0,p,L,0,k_2,k_3)}$ is the probability that the system is in a state with inventory L while the server is idle, and $\sum_{l\geq 0}\sum_{p=0}^{q-1}\sum_{k_2=1}^{t_2}\sum_{k_3=1}^{m} x_{(l,p,n',1,k_2,k_3)}$ is the probability that the system is found in a state with inventory n' while serving customers.

3. Probability that the server is on vacation and processing inventory:

$$P_{SVI} = \sum_{p=0}^{q-1}\sum_{n'=1}^{L-1}\sum_{k_1=1}^{t_1}\sum_{k_2=1}^{t_2}\sum_{k_3=1}^{m} x'_{(0,p,n',0,k_1,k_2,k_3)}, \tag{69}$$

where $x_{(0,p,n',0,k_1,k_2,k_3)}$ is the probability that the system is in a state with p customers and processing inventory (with inventory processing phase being k_1).

4. Probability that the server is idle:

$$P_I = \sum_{p=0}^{q-1}\sum_{k_2=1}^{t_2}\sum_{k_3=1}^{m} x'_{(0,p,L,0,k_2,k_3)}, \tag{70}$$

where $x_{(0,p,L,0,k_2,k_3)}$ is the probability that the system is in a state with p customers and L items in inventory.

5. Probability that the server is on vacation:

$$P_V = P_{SVI} + P_I. \tag{71}$$

6. Fraction of time that the server is busy serving a batch of customers without having inventory at the commencement of service:

$$T_{WI} = \sum_{l=0}^{\infty}\sum_{p=0}^{q-1}\sum_{k_2=1}^{t_2}\sum_{k_3=1}^{m} x'_{(l,p,0,1,k_2,k_3)}, \tag{72}$$

where $x_{(l,p,0,1,k_2,k_3)}$ is the probability that the system is in a state with $n = lq + p$ customers and n_2 customers are served without inventory while the service phase is k_2.

7. Fraction of time that the server is busy serving a batch of customers with inventory available at the beginning of service:

$$T_I = \sum_{l=0}^{\infty}\sum_{p=0}^{q-1}\sum_{n'=1}^{L}\sum_{k_2=1}^{t_2}\sum_{k_3=1}^{m} x'_{(l,p,n',1,k_2,k_3)}, \tag{73}$$

where $x_{(l,p,n',1,k_2,k_3)}$ is the probability that the system is in a state with $n = lq + p$ customers, serving customers(when n' items are in stock).

10. Comparitive Analysis of Model I with Model II

The effect of increasing parameter values L and θ on the behavior of a queuing-inventory system with and without Markov-dependent assumptions on successive arrival as well as service batch sizes are compared in this section.

Example 2. *This example studies the effect of L, the maximum number of additional items produced during a vacation on performance measures,*

- *expected number in the system, E_{NS};*
- *expected number of inventory available, E_I;*
- *probability that the server is on vacation, processing inventory, P_{SVI};*
- *probability that the server is idle, P_I;*
- *fraction of time server serves without inventory, T_{WI}; and*
- *fraction of time server serves with inventory, T_I,*

for two queuing-inventory systems of Models I and II (with more or less the same traffic intensity).

For ***Model 1****, (i.e., with Markov dependence (With MD)) consider two queuing-inventory systems where the arrival process is BMAP with the mean arrival rate of batches of customers 1 and with (D_0, D_1):*

$$D_0 = \begin{bmatrix} -2.2444 & 0.0673 \\ 0.0374 & -0.4489 \end{bmatrix}, D_1 = \begin{bmatrix} 2.0948 & 0.0823 \\ 0.0374 & 0.3741 \end{bmatrix}$$

Similarly, the service process BMSP is with the mean service rate of a batch of customers 3 and (S_0, S_1):

$$S_0 = \begin{bmatrix} -6.5214 & 0.0216 \\ 0.0216 & -1.3041 \end{bmatrix}, S_1 = \begin{bmatrix} 6.4347 & 0.0651 \\ 0.0216 & 1.2609 \end{bmatrix}$$

The correlation coefficient between two consecutive inter-arrival times is 0.2245, and the correlation coefficient between two consecutive inter-batch service times is 0.2792.

The queuing-inventory systems differ in underlying Markov chains for successive arrival as well as service batch sizes. Based on the underlying Markov chains for arrival and service batch sizes, they are classified into two classes:

- ***With MD 1*** *The probability of choosing the next arrival or service batch size is uniform, i.e., the successive arrival batch sizes form a Markov chain with tpm, P and the successive service batch sizes form a Markov chain with tpm, Q:*

$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, Q = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

  For this process, the traffic intensity $\rho = 0.3551$ and is indicated in black.
- ***With MD 2*** *The successive arrival batch sizes form a Markov chain with tpm, P and the successive service batch sizes form a Markov chain with tpm, Q:*

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.8 & 0.2 \end{bmatrix}, Q = \begin{bmatrix} 0.1 & 0.2 & 0.7 \\ 0.3 & 0.1 & 0.6 \\ 0.4 & 0.1 & 0.5 \end{bmatrix}$$

  For this process, traffic intensity $\rho = 0.2649$ and is indicated in blue.

  Fix $\theta = 0.7$.
  The inventory processing time follows $PH(\alpha, T), \alpha = [0.6, 0.4]$,

$$T = \begin{bmatrix} -2 & 2 \\ 0 & -2 \end{bmatrix}$$

The mean inventory processing time is 0.8.

For ***Model II*** *(i.e., without Markov dependence (Without MD)), consider the following queuing-inventory systems: Here, the mean arrival rate of batches of customers is normalized to 1 and the mean service rate of batches is 3. The correlation coefficient for consecutive inter-arrival times is 0.2245, and the coefficient for consecutive service times is 0.2792. The transition rates for different arrival and service batch sizes are constant for the first queuing-inventory system, while this is not true for the second.*

- ***Without MD 1***
  BMAP with representation (D_0, D_1, D_2)*:*

$$D_0 = \begin{bmatrix} -2.2444 & 0.0673 \\ 0.0374 & -0.4489 \end{bmatrix}, D_1 = D_2 = 1/2 \times \begin{bmatrix} 2.0948 & 0.0823 \\ 0.0374 & 0.3741 \end{bmatrix}$$

  BMSP with representation (S_0, S_1, S_2, S_3)*:*

$$S_0 = \begin{bmatrix} -6.5214 & 0.0216 \\ 0.0216 & -1.3041 \end{bmatrix}, S_1 = S_2 = S_3 = 1/3 \times \begin{bmatrix} 6.4347 & 0.0651 \\ 0.0216 & 1.2609 \end{bmatrix}$$

  For this process, traffic intensity $\rho = 0.3551$ *and is indicated in red.*
- ***Without MD 2***
  BMAP with representation (D_0, D_1, D_2)*:*

$$D_0 = \begin{bmatrix} -2.2444 & 0.0673 \\ 0.0374 & -0.4489 \end{bmatrix}, D_1 = D_2 = 1/2 \times \begin{bmatrix} 2.0948 & 0.0823 \\ 0.0374 & 0.3741 \end{bmatrix}$$

  BMSP with representation (S_0, S_1, S_2, S_3)*:*

$$S_0 = \begin{bmatrix} -6.5214 & 0.0216 \\ 0.0216 & -1.3041 \end{bmatrix} S_1 = S_2 = 1/3 \times \begin{bmatrix} 6.4347 & 0.0651 \\ 0.0216 & 1.2609 \end{bmatrix} S_3 = 7/3 \times \begin{bmatrix} 6.4347 & 0.0651 \\ 0.0216 & 1.2609 \end{bmatrix}$$

  For this process, traffic intensity $\rho = 0.2663$ *and is indicated in green.*

From Figure 4 on the next page, the following observations could be made (The numbers of the subfigures from which observations are made are given in brackets.):

Effect of L **on performance measures**: For both models, the increase in L increases not only the availability of inventory (Figure 4b) but also the time spent in inventory processing (Figure 4c). The expected queue length decreases as the service rate is higher with inventory (Figure 4a) and the server goes on vacation more often, as expected. As L increases, the fraction of time that the server serves customers with inventory increases (Figure 4e,f) and the idle time is reduced (Figure 4d).

Results of Comparison of Model I with Model II

- In Model I (the graphs of which are indicated in black and blue), the server goes on vacation once there are not enough customers to initiate the next batch service as specified by the Markov chain rule for service batch sizes. This increases the number of inventory processed as well as the idle time compared to Model II (for almost the same values of traffic intensity ρ). The server serves with inventory for a higher fraction of time, which results in a lower number of customers in the system as service is provided at a faster rate.
- As can be seen from the description, transition rates in Model I (with MD 1) and Model II (without MD 1) are the same and, hence, they have the same value for ρ though both models differ as specified in the first bulleted item. The values of the key performance measures such as the expected number in the system, the expected inventory available, and the probability that the server is on vacation processing inventory nearly coincide while the probabilities of the server being idle and of the server serving with or without inventory differs by a fraction of 5×10^{-2} (for the same value of ρ), as can be seen from the graphs. (The graphs in black and red have

ρ = 0.35, while the graphs in blue and green have ρ = 0.26). The values of performance measures for both models are more or less the same when the mean number of arrivals during an average service time, ρ, coincide.

- For both models, a slight increase in ρ (from 0.26 to 0.35) results in a noticeable increase in the expected number of customers in the system and reduces the inventory available and idle time. For all the observations made in the previous example (Example 1, the effect of increasing ρ on performance measures for Model I remains valid for Model II also.
- In addition, it is to be noted that, for Model I, it is the transition probability matrices P and Q that determine the behavior of the queuing-inventory system. However, for Model II, it is the stochastic matrices associated with the arrival and service processes that determines the behavior of the system.

Example 3. *Consider two queuing-inventory systems with inventory processing as in Example 2. To study the effects of θ, the factor that slows the service process (without inventory at the service commencement epoch) on various performance measures for both models, consider queuing-inventory systems with arrivals and service processes for Model I (with MD 2) and Model II (without MD 2), as in the previous example. We fix $L = 3$.*

As can be seen from Tables 20 and 21 for both models, the increase in θ considerably decreases the expected number of customers in the system, i.e., a higher service rate without inventory at the service commencement epoch, and reduces the queue length. In addition, a higher service rate and a small queue mean that the server goes on vacation frequently, resulting in an increased duration of time for inventory processing and hence the available inventory. This leads to an increase in the fraction of time that the server serves with inventory, which leads to a further reduction in queue length.

Table 20. Effect of θ on performance measures for Model I (with MD 2).

	E_{NS}	E_I	P_I	P_{SVI}	P_V	T_{WI}	T_I
$\theta = 0.2$	441.5593	0.2933	0.0723	0.0425	0.1148	0.8712	0.0107
$\theta = 0.3$	27.3600	1.2671	0.3082	0.1899	0.4980	0.4494	0.0506
$\theta = 0.4$	8.9726	1.6326	0.3931	0.2475	0.6407	0.2866	0.0703
$\theta = 0.5$	4.7130	1.8172	0.4344	0.2769	0.7113	0.2034	0.0827
$\theta = 0.6$	3.2188	1.9233	0.4573	0.2940	0.7514	0.1548	0.0910
$\theta = 0.7$	2.5500	1.9899	0.4713	0.3052	0.7765	0.1238	0.0969
$\theta = 0.8$	2.1955	2.0346	0.4804	0.3131	0.7935	0.1025	0.1082

Table 21. Effect of θ on performance measures for Model II (Without MD 2).

	E_{NS}	E_I	P_I	P_{SVI}	P_V	T_{WI}	T_I
$\theta = 0.2$	551.3419	0.2232	0.0545	0.0353	0.0898	0.8920	0.0077
$\theta = 0.3$	35.3123	1.0959	0.2635	0.1869	0.4444	0.5082	0.0412
$\theta = 0.4$	11.4672	1.4583	0.3464	0.2448	0.5912	0.3399	0.0590
$\theta = 0.5$	5.7789	1.6511	0.3887	0.2792	0.6680	0.2487	0.0705
$\theta = 0.6$	3.7503	1.7656	0.4130	0.3002	0.7131	0.1936	0.0782
$\theta = 0.7$	2.8338	1.8389	0.4280	0.3141	0.7421	0.1575	0.0836
$\theta = 0.8$	2.3500	1.8887	0.4379	0.3241	0.7620	0.1323	0.0876

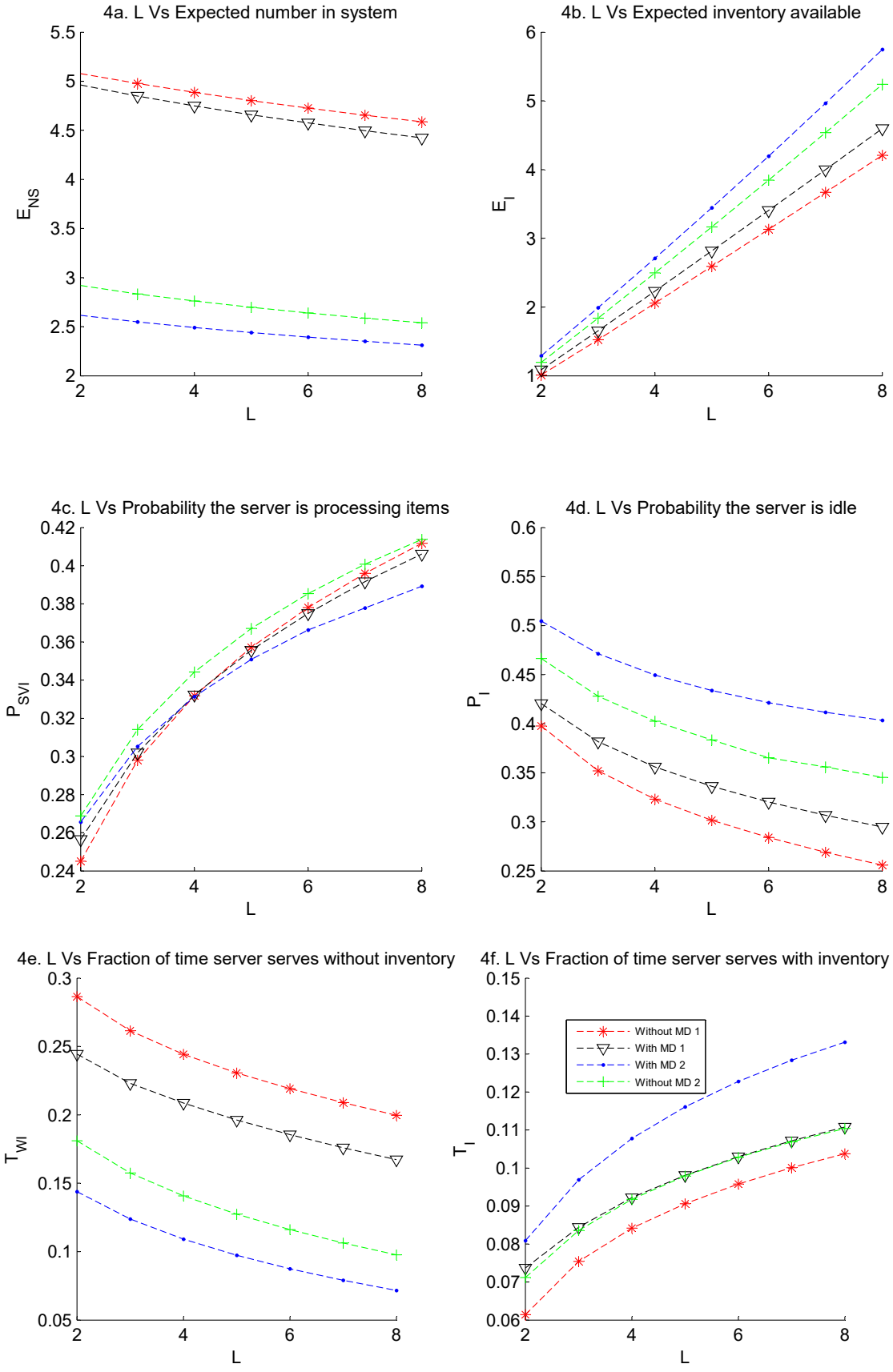


Figure 4. Effect of L, maximum inventory to be processed on performance measures.

11. Conclusions

Here the dynamics of a queuing-inventory system that has immense practical applications—especially for large organizations where both arrival and demand for service are in batches—are analyzed. The assumption that a single item (inventory) is required for service of a batch, although seemingly restrictive, is realistic. For example, if during vacation, the server is engaged in booking activity, before the start of service itself, the server has prior information regarding the type of service to be offered, and this reduces the service time. For the queuing system considered in [8], idle time (the period of time that the server stops service due to a lack of customers, as specified by the service batch size MC rule) could not be used optimally, but in this model, it can be utilized to reduce service time further.

The queuing-inventory system was studied as a multi-dimensional Markov chain. The structure of the generator matrix was not QBD; thus, it was brought to that form by suitably redefining the level. The steady-state probability vector was computed, and distributions relating to a vacation cycle were analyzed. Some important performance measures were computed. The impact of correlation in successive inter-arrival as well as service times on the behavior of the queuing system was studied in Example 1. The present model was then compared with a model without the Markov-dependent assumption. Conclusions based on the results of computations were presented. A cost function was constructed for the queuing system under study, and an optimal value of L that minimizes the cost was found for a specific example.

An extension of the present work to the case in which the server provides items to each customer served in a batch is proposed to be taken up. This will be highly challenging. Therefore, we propose to relax the service process to a phase-type distributed one. This assumption could be used in the arrival process as well. Another possibility is to extend the work to retrial queues with a finite buffer and an infinite capacity orbit.

Author Contributions: Conceptualization, methodology, investigation, writing—original draft preparation, supervision, and funding acquisition, A.K.; validation, data curation, writing—review and editing, and visualization, A.N.J.; formal analysis and writing—review and editing, D.K. All authors have read and agreed to the published version of the manuscript.

Funding: Research was supported by the Indo-Russian project: INT/RUS/RFBR/377, funded by DST. This paper was supported by the RUDN University Strategic Academic Leadership Program (recipient D.K., formal analysis) and funded by RFBR according to the research project number 19-29-06043 (recipient D.K., writing—review and editing).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors express their gratitude to the referees for valuable suggestions that improved the quality of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following notations and abbreviations are used in this manuscript:

MAP	Markovian arrival process
$BMAP$	batch Markovian arrival process
$BMSP$	batch Markovian service process
$\mathbf{e}$	the column vector of 1s of appropriate order
$\mathbf{e}(j)$	the column vector of 1s of order j
$\mathbf{e}'$	the transpose of $\mathbf{e}$
$\mathbf{e}_i(j)$	the column vector of order j with 1 in the ith position and 0 elsewhere
$\mathbf{0}$	zero-matrix of appropriate order

I	identity matrix of appropriate order
I_r	identity matrix of order r
$CTMC$	continuous-time Markov chain
MC	Markov chain
$LIQBD$	level-independent quasi birth and death process
tpm	transition probability matrix
LST	Laplace–Stieltjes transform
TC	tagged customer

References

1. Chaudhry, M.L.; Templeton, J.G.C. *A First Course in Bulk Queues*; John Wiley and Sons: New York, NY, USA, 1983.
2. Lucantoni, D.M. New Results on the Single Server Queue with a Batch Markovian Arrival Process. *Commun. Stat. Stoch. Model.* **1991**, *7*, 1–46. [CrossRef]
3. Neuts, M.F. Versatile Markovian Point Process. *J. Appl. Probab.* **1979**, *16*, 764–779. [CrossRef]
4. Ramaswami, V. The N/G/1 Queue and Its Detailed Analysis. *Adv. Appl. Probab.* **1980**, *12*, 222–261. [CrossRef]
5. Artalejo, J.R.; Gómez-Corral, A. Markovian arrivals in stochastic modelling: A survey and some new results. *SORT–Stat. Oper. Res.* **2010**, *34*, 101–144.
6. Chakravarthy, S.R. The batch Markovian arrival process: A review and future work. In *Advances in Probability and Stochastic Processes*; Krishnamoorthy, A., Raju, N., Ramaswami, V., Eds.; Notable Publications: Long Island, NJ, USA 2001; pp. 21–49.
7. Vishnevskii, V.M.; Dudin, A.N. Queueing Systems with Correlated Arrival Flows and Their Applications to Modeling Telecommunication Networks. *Autom. Remote Control* **2017**, *78*, 1361–1403. [CrossRef]
8. Krishnamoorthy, A.; Joshua, A.N. A $BMAP/BMSP/1$ queue with Markov dependent arrival and Markov dependent service batches. *J. Ind. Manag. Optim.* **2020**. [CrossRef]
9. Divya, V.; Krishnamoorthy, A.; Vishnevsky, V.M.; Kozyrev, D.V. On a Queueing System with Processing of Service Items under Vacation and N-policy with Impatient Customers. *Queueing Model. Serv. Manag.* **2020**, *3*, 167–201.
10. Kazirmsky, A.V. Analysis of BMAP/G/1 Queue with Reservation of Service. *Stoch. Anal. Appl.* **2006**, *24*, 703–718. [CrossRef]
11. Hanukov, G.; Avinadav, T.; Chernonog, T.; Spiegel, U.; Yechiali, U. A queueing system with decomposed service and inventoried preliminary services. *Appl. Math. Model.* **2017**, *47*, 276–293. [CrossRef]
12. Divya, V.; Krishnamoorthy, A.; Vishnevsky, V.M. On a queueing system with processing of service items under vacation and N-policy. In *Distributed Computer and Communication Networks. DCCN 2018*; Vishnevskiy, V., Kozyrev, D., Eds.; Communications in Computer and Information Science; Springer: Cham, Switzerland, 2018; Volume 919.
13. Baek, J.; Dudina, O.; Kim, C. A Queueing System with Heterogeneous Impatient Customers and Consumable Additional Items. *Int. J. Math. Comput. Sci.* **2017**, *27*, 367–384. [CrossRef]
14. Chakravarthy, S.R.; Maity, A.; Gupta, U.C. An '(s, S)' inventory in a queueing system with batch service facility. *Ann. Oper. Res.* **2017**, *258*, 263–283. [CrossRef]
15. Shajin, D.; Dudin, A.N.; Dudina, O.; Krishnamoorthy, A. A two-priority single server retrial queue with additional items. *J. Ind. Manag. Optim.* **2020**, *16*, 2891–2912. [CrossRef]
16. Levi, Y.; Yechiali, U. Utilization of idle time in an M/G/1 queueing system. *Manag. Sci.* **1975**, *22*, 202–211. [CrossRef]
17. Doshi, B.T. Queueing systems with vacations—A survey. *Queueing Syst.* **1986**, *1*, 29–66. [CrossRef]
18. Takagi, H. *Queueing Analysis, Volume I: Vacation and Priority Systems, Part 1*; North Holland: Amsterdam, The Netherlands, 1991; p. 488.
19. Tian, N.; Zhang, Z.G. *Vacation Queueing Models. Theory and Applications. International Series in Operations Research and Management Science*; Springer: New York, NY, USA, 2006; Volume 93, p. 386. [CrossRef]
20. Neuts, M.F. *Matrix-Geometric Solutions in Stochastic Models—An Algorithmic Approach*; The John Hopkins University Press: Baltimore, MD, USA, 1981; p. 352.