10.48047/jocaaa.2021.29.06.39

On a Priority Queueing Inventory System with Varying Demand Quantities

Shajeeb P U[1,2], Jaison Jacob[3], Abdul Rof V[4,*]

[1]Department of Mathematics, Cochin University of Science and Technology, Kochi, Kerala 682022, India.

[2]Department of Mathematics, Govt. Victoria College, Palakkad, Kerala 678001, India.

[3]Department of Mathematics, St. Aloysius College, Elthuruth, Thrissur, Kerala 680611, India.

[4]Department of Mathematics, Korambayil Ahamed Haji Memorial Unity Women's College, Manjeri, Kerala, India.

**abdulrof@unitywomenscollege.ac.in* (Corresponding author)

Abstract

A single server, two - priority queueing inventory (inventory with positive service time) is considered under the replenishment policy (s, S). Priority 1 customers (Type 1) have unlimited waiting space; they arrive according to a Poisson process. Each of these customers requires one unit of inventory. Priority 2 (Type 2) customers have a finite capacity waiting space. Their arrival is also according to a Poisson process, and service times have exponential distribution with parameter depending on the number of item they require, which can be 2, 3, . . . , a, where,
a < s. Priority discipline is non-preemptive. This system is analyzed under the condition for stability. The system state distribution is derived. Performance characteristics needed for an appropriate cost function are computed, and the cost function is numerically analyzed.

Keywords: Queueing inventory system, Type 1 and Type 2 customers, Continuous Time Markov Chain, Non-preemptive priority, Matrix Geometric Method

1. INTRODUCTION

In any service station where customer service requires a positive amount of time, queues are likely to form. In traditional queueing theory, the service process continues as long as there is at least one customer present and the server is available, without regard to resource availability. This contrasts with inventory systems, where service cannot proceed if inventory is unavailable, even if the customers and the server are ready. Classical inventory models assume that queues form only when inventory is depleted and that customers can wait during stockouts, based on the assumption of negligible service time. However, in reality, fulfilling an order requires a positive amount of time, which makes inventory models with positive service times more realistic. Such models are referred to as queueing inventory systems.

Over the past three decades, queueing inventory models have attracted considerable attention from researchers. The foundational contributions in this area are independently credited to Melikov and Molchanov [4] as well as Sigman and Simchi-Levi [5]. An extensive literature survey on queueing inventory systems can be found in [6] and [9]. A recent study on customer priority queueing inventory system is presented in [7], where the authors analyze a queueing inventory system with items that have phase-type (PH) distributed life times and class-based customer access priorities. In [8], the authors focus on a single-server production inventory model with server failure and customer impatience during downtime.

In [2], authors investigate a queueing inventory model that incorporates two distinct service channels: a single-server facility (Channel I) and a bulk service facility (Channel II) and two types of customers each receiving the same commodity: type I customers are served by Channel I, while type II customers are served by Channel II. In [3], the authors consider batch demands in the

10.48047/jocaaa.2021.29.06.39

context of two models, both of which assume that the demands occur according to a versatile Markovian point process.

The theme of this paper is the analysis of a single-server, single-commodity, two - priority queueing inventory (Q I) system operating under the (s, S) inventory control policy. Priority 1 customers (Type 1) have an unlimited waiting space whereas priority 2 customers (Type 2) are provided only a finite capacity waiting station. Each Type 1 customer demands exactly one unit of the inventoried item and holds non-preemptive priority over Type 2 customers. Each customer of later type requires a random number of inventory units , taking values $j = 2, 3, \ldots, a$ (where $a < s$, the level of safety stock reorder), with demand probability q_j for j units. Type 1 customers are served one unit of inventory, and their service times follow an exponential distribution with rate μ_1. In contrast, the service time parameter for Type 2 customers depends on the number of units the customer demands. However, if the number of units in the inventory is less than the requirement, then the customer is served with the available number of items.

The framework of the model is particularly relevant in scenarios where prioritizing single-unit demands over bulk orders is strategically advantageous. Periodically, manufacturers offer price reductions on their products based on the quantity purchased. However, retailers often find it more profitable to prioritize single-unit sales rather than bulk orders, especially when inventory holding costs are significantly lower than potential profit margins. Consequently, customers purchasing just one unit are given higher priority.

The model also incorporates scenarios involving emergency items with limited availability, which are distributed through a rationing system to ensure sparing usage. Therefore, priority 2 customers (Type 2) are only served when no single-unit demand exists, justifying the restricted waiting space capacity for these bulk-order customers.

The salient features of the present paper are:

- **It presents a single server, two - priority queueing inventory model that incorpo- rates both single and varying demand quantities from customers.**.
- **In this model, the customers are to join one of the two different queues based on their demand quantities and priority is given to customers with single unit requests.**

The remaining sections of this paper are organized as follows. Section 2 covers the mathematical formulations of the model, along with their stability conditions, steady-state probability vector, performance measures, and the development of the cost function. Section 3 presents the analysis of inventory cycle time. Numerical illustrations and an analysis of the cost function are provided in Sections 4 and 5. Finally, Section 5 concludes the article and outlines the scope of future research.

Notations and abbreviations used:

- (s, S) ordering Policy: An inventory control strategy where a new order is placed when the inventory level falls to or below s and the order replenishes stock up to a target level S.
- e = Column vector of appropriate order with all its entries as 1's.
- I_k = Identity matrix of order k.
- CTMC = Continuous Time Markov chain
- $f * g$ = Convolution of f and g
- $\lceil x \rceil$ = Ceiling integer value of x

2. ANALYSIS OF THE MODEL

We consider a single-server queueing inventory system with two priority classes and an (s, S) inventory policy. Customers arrive and request either one unit or multiple units (between 2 and a where $a < s$) of a single commodity. Customers are categorized into two types: Type 1 customers request a single unit, while Type 2 customers request between 2 and a units. The arrivals of Type 1 and Type 2 customers follow independent Poisson processes with rates λ_1 and λ_2, respectively. The service times for both types are

10.48047/jocaaa.2021.29.06.39

exponentially distributed, with rates μ_1 for Type 1 customers and μ_2 for Type 2 customers. Type 1 customers have an unlimited waiting space, while Type 2 customers are allocated a finite capacity waiting station. Type 1 customers are given priority over Type 2 customers, but the system operates under a non-preemptive service policy. When a Type 2 customer begins service, they specify the number of units they require. The probability that a Type 2 customer demands j units (where $= 2,3, \dots, a$) is denoted by q_j, with $\sum_{j=2}^{a} q_j = 1$. The service rate for a Type 2 customer demanding j units is $q_j\mu_2$, provided that sufficient inventory is available. If the inventory level is at least 2 but is insufficient to meet the full demand of a Type 2 customer, all units in stock are provided to the customer. The replenishment lead time is exponentially distributed with rate β.. The model also incorporates the following assumptions:

- Type 1 customers have infinite waiting capacity.
- The number of Type 2 customers in the system cannot exceed a fixed positive integer N.
- A Type 2 customer is served only if there are no Type 1 customers in the queue and the inventory level is at least two.
- Let i be the inventory level and j be the number of units demanded by a Type 2 customer. If $2 \leq i < j$, then all i units are provided to the customer at a service rate of $q_j\mu_2$.
- If, at the completion of a service, the inventory level drops to zero, the arrival of Type 2 customers is blocked.

2.1. Mathematical Formulation

We use the following notation:

- $N_1(t)$: The number of Type 1 customers in the system at time t
- $N_2(t)$: The number of Type 2 customers in the system at time t
- $I(t)$: The number of units of item in the inventory at time t

The process $\{X(t): t \geq 0\} = \{(N_1(t), I(t), N_2(t)); t \geq 0\}$ is a Continuous-Time Markov Chain (CTMC) with the state space $\Omega = \{(n_1, i, n_2); n_1 = 0,1,2 \dots; i = 0,1,2, \dots, S; n_2 = 0,1,2, \dots, N.\}$

For each $n = 0,1,2, \dots$, define $\mathcal{L}(n) = \{(n, i, m); i = 0,1,2, \dots S; m = 0,1,2 \dots N.\}. Thenwe$ may write the state space as :

$$\Omega = \bigcup_{n=0}^{\infty} \mathcal{L}(n)$$

10.48047/jocaaa.2021.29.06.39

and we refer to $\mathcal{L}(n)$ as the n^{th} level of the process. The main transitions involved are the following.

- Transitions due to arrivals:

1. Transition due to arrival of Type 1 customers:

$$(n_1, i, n_2) \to (n_1 + 1, i, n_2)$$

with rate λ_1 for $n_1 = 0,1,2, \cdots : n_2 = 0,1,2, \dots, N : i = 0,1, \dots S$

2. Transition due to arrival of Type 2 customers:

$$(n_1, i, n_2) \to (n_1, i, n_2 + 1)$$

with rate λ_2 for $n_1 = 0,1,2, \cdots : n_2 = 0,1,2, \dots, N - 1 : i = 1,2,3, \dots S$

- Transitions due to service completion:

1. Transition due to service completion of a Type 1 customer:

$$(n_1, i, n_2) \to (n_1 - 1, i - 1, n_2)$$

with rate μ_1 for $n_1 = 1,2,3, \cdots : n_2 = 0,1, \dots, N : i = 1,2,3, \dots S$

2. Transition due to service completion of a Type 2 customer:

$$(0, i, n_2) \to (0, i - j, n_2 - 1)$$

with rate $q_j \mu_2$ for $n_2 = 1,2,3, \dots, N; i = j, j + 1, \dots S; j = 2,3, \dots, a$

3.

$$(0, i, n_2) \to (0,0, n_2 - 1)$$

with rate $q_j \mu_2$ for $n_2 = 1,2, \dots, N; i = 2,3, \dots j - 1$

- Transition due to replenishment:

$$(n_1, i, n_2) \to (n_1, S, n_2)$$

with rate β for $n_1 = 0,1, \dots; n_2 = 0,1, \dots, N; i = 0,1, \dots s$

Using the lexicographical ordering of states, the infinitesimal generator matrix Q can be expressed in block form as

$$Q = \begin{pmatrix} A_{00} & A_0 & & & \\ A_2 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

All non-zero block matrices A_{00}, A_0, A_1 and A_2 are square matrices of order $(S + 1)(N + 1)$. We first define the following component matrices:

10.48047/jocaaa.2021.29.06.39

$$L_1 = \lambda_1 I_{N+1};\ M_1 = \mu_1 I_{N+1};\ B = \beta I_{N+1};$$

$$L_2 = \begin{pmatrix} 0 & \lambda_2 I_N \\ \vdots & \\ 0 & 0 \\ \cdots & 0 \end{pmatrix}; M_2 = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ & & & \vdots \\ & \mu_2 I_N & & \\ & & & 0 \end{pmatrix}; L_2^* = \begin{pmatrix} \lambda_2 I_N & & 0 \\ & & \vdots \\ \underline{0 \quad \cdots \quad 0} \end{pmatrix}$$

$$\mathrm{d}M_2^* = \begin{pmatrix} 0 & 0 & \cdots \\ 0 & & \mu_2 I_N \\ \vdots & & \end{pmatrix}$$

- The nonzero entries of A_{00} correspond to the transition rates within the level $\mathcal{L}(0)$ and are given by :

$$A_{00} = \begin{pmatrix}
B_0 & & & & & & & & & B \\
 & B_1 & & & & & & & & B \\
M_2 & & B_1 & & & & & & & B \\
\left(\sum_{j=3}^{a} q_j\right) M_2 & q_2 M_2 & & B_1 & & & & & & B \\
\vdots & & \ddots & & \ddots & & & & & \vdots \\
q_a M_2 & q_{a-1} M_2 & \cdots & q_2 M_2 & & B_1 & & & & B \\
\ddots & \ddots & \cdots & \ddots & & \ddots & & & & \vdots \\
 & q_a M_2 & q_{a-1} M_2 & \cdots & q_2 M_2 & & B_1 & & & B \\
 & & q_a M_2 & q_{a-1} M_2 & \cdots & q_2 M_2 & & B_2 & & 0 \\
 & & \ddots & \ddots & \cdots & \ddots & & \ddots & & \vdots \\
 & & & & q_a M_2 & q_{a-1} M_2 & \cdots & q_2 M_2 & B_2 & 0 \\
 & & & & & q_a M_2 & q_{a-1} M_2 & \cdots & q_2 M_2 & B_2
\end{pmatrix}$$

where $B_0 = -L_1 - B;\ B_1 = L_2 - L_2^* - L_1 - B$; and $B_2 = L_2 - L_2^* - L_1$.

- A_0 contains the transition rates from level $\mathcal{L}(i)$ to the level $\mathcal{L}(i+1)$ (for $i = 0,1,2, \ldots$) and has the form:

$$A_0 = \begin{pmatrix} L_1 & & & \\ & L_1 & & \\ & & \ddots & \\ & & & L_1 \end{pmatrix}.$$

- The matrix A_1 describes the internal transition rates for each level $\mathcal{L}(i)$ for $i = 1,2,3, \ldots$ and has the form:

$$A_1 = \begin{pmatrix} B_0 & & & & & & & B \\ & C_1 & & & & & & B \\ & & \ddots & & & & & \vdots \\ & & & C_1 & & & & B \\ & & & & C_2 & & & 0 \\ & & & & & \ddots & & \vdots \\ & & & & & & C_2 & 0 \\ & & & & & & & C_2 \end{pmatrix}.$$

10.48047/jocaaa.2021.29.06.39

where $C_1 = L_2 - L_2^* - L_1 - M_1 - B$; and $C_2 = L_2 - L_2^* - L_1 - M_1$

- A_2 represents the transition rates from level $\mathcal{L}(i)$ to level $\mathcal{L}(i-1)$ for $i = 1,2,3, \ldots$, expressed as:

$$A_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ M_1 & & & & 0 \\ & M_1 & & & 0 \\ & & \ddots & & 0 \\ & & & M_1 & 0 \end{pmatrix}.$$

2.2. Steady state analysis

This section examines the stability of the system. Let $\pi = (\pi_0, \pi_1, \pi_2, \ldots, \pi_S)$ denote the steady-state probability vector of the generator matrix $A = A_0 + A_1 + A_2$, where each $\pi_j = \left(\pi_j^0, \pi_j^1, \pi_j^2, \ldots, \pi_j^N\right)$. The block structured matrix A is given by:

$$A = \begin{pmatrix} D_0 & & & & & & & B \\ M_1 & D_1 & & & & & & B \\ & \ddots & \ddots & & & & & \vdots \\ & & M_1 & D_1 & & & & B \\ & & & M_1 & D_2 & & & \\ & & & & \ddots & \ddots & & \\ & & & & & M_1 & D_2 & \\ & & & & & & M_1 & D_2 \end{pmatrix}$$

where, $D_0 = L_1 + B_0$; $D_1 = L_1 + C_1$; and $D_2 = L_1 + C_2$

The steady state vector π satisfies the equations

$$\pi A = 0, \pi e = 1 \tag{1}$$

This yields the following system of equations:

$$\pi_0 D_0 + \pi_1 M_1 = 0 \tag{2}$$
$$\pi_j D_1 + \pi_{j+1} M_1 = 0;\ j = 1,2, \ldots, s \tag{3}$$
$$\pi_{s+j} D_2 + \pi_{s+j+1} M_1 = 0;\ j = 1,2, \ldots, S - s - 1 \tag{4}$$
$$(\pi_0 + \pi_1 + \cdots + \pi_s) B + \pi_S D_2 = 0 \tag{5}$$

Solving these equations recursively, we obtain :

$$\pi_j = \begin{cases} \dfrac{\beta}{\mu_1^j} \pi_0 (-D_1)^{j-1} & \text{for } j = 1,2, \ldots, s+1 \\ \dfrac{\beta}{\mu_1^j} \pi_0 (-D_1)^s (-D_2)^{j-s-1} & \text{for } j = s+2, s+3, \ldots, S \end{cases}$$

10.48047/jocaaa.2021.29.06.39

Using the normalizing condition from equation (1), π_0 is determined by :

$$\pi_0\left(I+\frac{\beta}{\mu_1}I+\frac{\beta}{\mu_1^2}(-D_1)+\cdots+\frac{\beta}{\mu_1^{s+1}}(-D_1)^s+\frac{\beta}{\mu_1^{s+2}}(-D_1)^s(-D_2)+\cdots\right.$$
$$\left.+\frac{\beta}{\mu_1^S}(-D_1)^s(-D_2)^{S-s-1}\right)e=1$$

Theorem 1. Let $\pi=(\pi_0,\pi_1,\pi_2,\ldots,\pi_S)$ be the steady-state probability vector of the generator matrix $A=A_0+A_1+A_2$, where $\pi_j=\left(\pi_j^{(0)},\pi_j^{(1)},\ldots,\pi_j^{(N)}\right)$. Then the system is stable if and only if

$$\lambda_1<(1-\pi_0\mathbf{e})\mu_1$$

Proof.

For the CTMC $\{X(t):t\geq 0\}$ to be stable, a necessary and sufficient condition is $\pi A_0\mathbf{e}<\pi A_2\mathbf{e}$, (see [1]). Simplifying this inequality yields the stability condition $\lambda_1<\left(1-\sum_{j=0}^{S}\pi_j^{(0)}\right)\mu_1$.

Remark 1. The stability condition derived in Theorem 1 can be refined by introducing the constraint that Type 1 customer arrivals are blocked when the inventory is depleted. Under this restriction, the stability condition simplifies to $\lambda_1<\mu_1$.

2.3. Steady state Probability vector

Assume the system is stable. We now derive the steady state probability vector **x** of the infinitesimal generator matrix Q. The vector **x** satisfies the following conditions:

$$\mathbf{x}Q=0,\mathbf{xe}=1 \tag{6}$$

Partition **x** according to the number of Type 1 customers in the system :

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

From equation (6) we have,

$$\mathbf{x}_0A_{00}+\mathbf{x}_1A_2=0 \tag{7}$$
$$\mathbf{x}_{n-1}A_0+\mathbf{x}_nA_1+\mathbf{x}_{n+1}A_2=0;n=1,2,3,\ldots \tag{8}$$

Under the stability condition, the solution takes the matrix geometric form (see Neuts):

$$\mathbf{x}_{n+1}=\mathbf{x}_0R^{n+1};n=0,1,2,\ldots$$

where R is the minimal non-negative solution of the matrix quadratic equation

$$R^2A_2+RA_1+A_0=0$$

The normalizing condition given in (6) yields $\mathbf{x}_0$ as:

$$\mathbf{x}_0(I-R)^{-1}\mathbf{e}=1 \tag{9}$$

2.4. Some important system characteristics

10.48047/jocaaa.2021.29.06.39

1. Expected number of Type 1 customers in the system : $E_1 = \sum_{n=1}^{\infty} \sum_{i=0}^{S} \sum_{m=0}^{N} nx_n(i,m)$.
2. Expected number of Type 2 customers in the system : $E_2 = \sum_{m=0}^{N} \sum_{n=0}^{\infty} \sum_{i=0}^{S} mx_n(i,m)$
3. Expected number of items in the inventory : $E_I = \sum_{i=0}^{S} \sum_{n=0}^{\infty} \sum_{m=0}^{N} ix_n(i,m)$
4. Expected reorder rate : $E_R = \mu_1 \sum_{n=1}^{\infty} \sum_{m=0}^{N} x_n(s+1,m) + \sum_{m=1}^{N} \sum_{j=2}^{a} q_j \mu_2 x_0(s+j,m)$
5. Expected loss rate of Type 2 customers : $E_L = \lambda_2 \sum_{n=0}^{\infty} \sum_{m=0}^{N} x_n(0,m)$
6. Expected shortage of inventory :

$$E_S = \sum_{j=1}^{a} \sum_{i=0}^{j-1} \sum_{r=0}^{a-(j-i)} \sum_{m=1}^{N} \sum_{n=0}^{\infty} (j-i) q_{j-i+r} x_n(r,m)$$

2.5. Cost function

Let

1. C_F = Fixed ordering cost of the inventory per unit item
2. C_I = Holding cost of the inventory per unit item.
3. C = Holding cost of customers per unit time.
4. C_L = Cost due to loss of customers per unit time
5. C_S = Penalty cost per unit shortage of inventory.

We define the cost function

$$K = C_F E_R + C_I E_I + C(E_1 + E_2) + C_L E_L + C_S E_S$$

3. Inventory Cycle Time

Let L denote the random variable representing the lead time in the (s, S) inventory policy under consideration. When the inventory level drops to s, an order is placed to raise the inventory level to S. We define T as the random variable representing the time between two consecutive instances when the inventory level hits s, which we refer to as the inventory cycle time. The distribution of T depends on the number of customers (both Type 1 and Type 2) in the system when the order is placed and the length of the lead time. To analyze the distribution of T, we introduce the following notation:

- t_0 : The initial time at which the cycle begins
- τ : The stopping time of the first replenishment ($\tau > t_0$).
- ν : The stopping time when inventory next reaches s again.

Then we have : $T = \nu - t_0, I(t_0) = s, I(\tau) = S$ and $I(\nu) = s$. Let $N_1(t_0) = n, N_2(t_0) = m$ and $I(\tau) = r$. Now, we consider various scenarios detailed as follows:

10.48047/jocaaa.2021.29.06.39

- Scenario 1: $n \geq S-r, 0 \leq r \leq s$ In this case, a cycle is completed with $S-r$ service completions of Type 1 customers as follows: Starting with s units, the inventory level drops to r due to $s-r$ service completions of Type 1 customers. This means the time before replenishment follows an Erlang distribution with parameter μ_1 and $s-r$ stages, denoted as $E_{\mu_1,s-r}$. Replenishment then occurs with a probability of $\left(\frac{\mu_1}{\beta+\mu_1}\right)^{s-r} \cdot \left(\frac{\beta}{\beta+\mu_1}\right)$. Following replenishment, the inventory level drops from S to s due to $S-s$ service completions of Type 1 customers, where the time is Erlang distributed with parameter μ_1 and $S-s$ stages, denoted as $E_{mu_1,S-s}$. Therefore the conditional distribution of the cycle time T is given by :

$$\sum_{r=0}^{s} \left(\frac{\mu_1}{\beta+\mu_1}\right)^{s-r} \cdot \left(\frac{\beta}{\beta+\mu_1}\right) E_{\mu_1,s-r} * E_{\mu_1,S-s}$$

where * denotes convolution.

- Scenario 2: $s-r \leq n < S-r$ and $N_1(\tau) \geq S-s$ In this scenario, a cycle time is realized with $S-r$ service completions of Type 1 customers as follows: Starting with an inventory level of s units, it decreases to r due to $s-r$ service completions of Type 1 customers. Consequently, the time before replenishment is Erlang distributed with parameter μ_1 and order $s-r$, denoted as $E_{\mu_1,s-r}$. Replenishment then occurs with a probability of $\left(\frac{\mu_1}{\beta+\mu_1}\right)^{s-r} \cdot \left(\frac{\beta}{\beta+\mu_1}\right)$. Given that $N_1(\tau) \geq S-s$, at least $S-n-r$ Type 1 arrivals occur during the time interval (t_0, τ) with a probability of $\left(1-\sum_{k=0}^{S-n-r} \frac{(\lambda_1(\tau-t_0))^k \cdot \exp(-\lambda_1(\tau-t_0))}{k!}\right)$. After replenishment, the inventory level drops from S to s due to $S-s$ service completions of Type 1 customers, where the time is distributed as $E_{\mu_1,S-s}$. Hence, the conditional distribution of the cycle time is given by :

$$\sum_{r=0}^{s} \left(\frac{\mu_1}{\beta+\mu_1}\right)^{s-r} \cdot \left(\frac{\beta}{\beta+\mu_1}\right) E_{\mu_1,s-r}$$

$$* \sum_{n=s-r}^{S-r-1} \left(1-\sum_{k=0}^{S-n-2r} \frac{\left(\lambda_1(\tau-t_0)\right)^k \cdot \exp\left(-\lambda_1(\tau-t_0)\right)}{k!}\right) E_{\mu_1,S-s}$$

where * denotes convolution.

- Scenario 3: $s-r \leq n < S-r$ and $N_1(v-t_0) = 0$ In this scenario, we derive a lower bound for the cycle length distribution. The inventory level initially drops from s to $s-r$ due to $s-r$ service completions of Type 1 customers, which takes an Erlang distributed time of (

10.48047/jocaaa.2021.29.06.39

$E_{\mu_1,s-r}$. Afterward, a replenishment occurs with a probability $\left(\frac{\mu_1}{\beta+\mu_1}\right)^{s-r} \cdot \left(\frac{\beta}{\beta+\mu_1}\right)$. Following the replenishment, the inventory level decreases from S to $S-n+s-r$ due to $n-s+r$ additional service completions of Type 1 customers. At this point, the Type 1 customer queue empties, and service for Type 2 customers begins. Since $N_1(v-t_0)=0$, there are no arrivals of Type 1 customers until the inventory level returns to s. Consequently, the inventory level reaches the level s through the service of at least $\left\lceil\frac{S-n-r}{a}\right\rceil$ Type 2 customers. Thus, we obtain a lower bound for the cycle length, expressed as:

$$\sum_{r=0}^{s}\left(\frac{\mu_1}{\beta+\mu_1}\right)^{s-r}\cdot\left(\frac{\beta}{\beta+\mu_1}\right)E_{\mu_1,s-r} * E_{\mu_1,S-n+s-r} * \exp\left(-\lambda_1(v-t_0)\right)\cdot E_{\mu_2,\left\lceil\frac{S-n-r}{a}\right\rceil}$$

where $\lceil x\rceil$ denotes the ceiling function and $*$ denotes the convolution.

- Scenario 4: $n=0, m\geq\left\lceil\frac{S-r}{a}\right\rceil$ and $N_1(v-t_0)=0$ In this case also, we obtain a lower bound for the cycle length distribution. Since $n=0$ and $N_1(v-t_0)=0$ the cycle length is realized through the service completions of Type 2 customers alone. Starting with s, the inventory level reaches r by service completions of at least $\left\lceil\frac{s-r}{a}\right\rceil$ Type 2 customers which then follows a replenishment and consequently the inventory level decreases from S to s through the service completions of a minimum of $\left\lceil\frac{S-s}{a}\right\rceil$ Type 2 customers. This gives a lower bound for the cycle length as follows:

$$\sum_{r=0}^{s}\left(\frac{\mu_2}{\mu_2+\beta}\right)^{\left\lceil\frac{S-s}{a}\right\rceil}\left(\frac{\beta}{\mu_2+\beta}\right)\cdot E_{\mu_2,\left\lceil\frac{S-s}{a}\right\rceil} * E_{\mu_2,\left\lceil\frac{S-s}{a}\right\rceil}$$

4. Numerical Illustrations

In this section we provide an illustration to see how system characteristics vary with different parameters.

1. Effect of μ_1 on performance measures and total cost

   Fix $S=12, s=5, a=4, N=3, \beta=17, \lambda_1=6; \lambda_2=7, \mu_2=10$

   Table 1: Variation of Performance measures and total cost with respect to μ_1

μ_1	E_1	E_2	E_l	E_R	E_L	E_S	K
8	3.006464	1.875919	5.568557	0.919903	0.852625	0.463808	321.5448
10	1.506109	1.499352	4.823672	0.86787	0.805346	0.357657	258.0546

10.48047/jocaaa.2021.29.06.39

13	0.863315	1.149202	4.103205	0.820618	0.770603	0.265034	213.4086
15	0.672922	0.993098	3.772726	0.799885	0.758016	0.225684	195.205
18	0.50638	0.823941	3.407348	0.777589	0.746369	0.184437	176.0922
20	0.435028	0.739453	3.22168	0.766478	0.741301	0.164386	166.6988

Remark 2. The analysis of Table 1 reveals that all performance measures ($E_1, E_2, E_l, E_R, E_L, E_S$) decrease monotonically as the service rate μ_1 increases from 8 to 20 . The most significant improvements occurring at lower μ_1 values (8-13). The total cost K exhibits strong sensitivity to μ_1. Additionally, the table shows a non-linear pattern in the performance measures as μ_1 increases, suggesting an optimal operational zone between $\mu_l = 13$ and 18, where the system achieves balanced cost-performance efficiency.

2. Effect of μ_2 on performance measures and total cost

Fix $S = 12, s = 5, a = 4, N = 3, \beta = 17, \lambda_1 = 6; \lambda_2 = 7, \mu_1 = 10$

Table 2: Variation of Performance measures and total cost with respect to μ_2

μ_2	E_1	E_2	E_I	E_R	E_L	E_S	K
9	1.505076	1.499341	4.821342	0.860681	0.818568	0.360565	257.7512
11	1.50722	1.499383	4.826463	0.87545	0.792778	0.354897	258.4007
13	1.509648	1.499519	4.833066	0.891788	0.769557	0.349807	259.2093
15	1.512306	1.49976	4.84061	0.90967	0.748722	0.345252	260.1566
18	1.516629	1.500319	4.852961	0.939226	0.721396	0.339304	261.7995
20	1.519676	1.50081	4.86158	0.960609	0.705428	0.335846	263.0211

10.48047/jocaaa.2021.29.06.39

Remark 3. Table 2 demonstrates the impact of varying the service rate μ_2 (from 9 to 20) on system performance while holding other parameters constant. Notably, increasing μ_2 leads to slight but consistent growth in all expected measures (E_1, E_2, E_I) as well as the total cost K. The measures (E_R, E_L, E_S) exhibit mixed trends: E_R increases, while E_L and E_S decrease.

3. Effect of λ_1 on performance measures and total cost

Fix $S = 12, s = 5, a = 4, N = 3, \beta = 24, \lambda_2 = 7;, \mu_1 = 14, \mu_2 = 11$

Table 3: Variation of Performance measures and total cost with respect to λ_1

λ_1	E_1	E_2	E_I	E_R	E_L	E_S	K
5	0.557467	0.903127	3.532533	0.66395	0.735024	0.198221	176.7267
6	0.752386	1.080453	3.924719	0.817093	0.77189	0.246891	204.2063
7	1.002925	1.256238	4.303895	0.972684	0.801572	0.295812	232.4291
8	1.336907	1.430519	4.673440	1.130517	0.826235	0.34473	262.0636
9	1.804447	1.603353	5.035748	1.290385	0.847285	0.393502	294.188
10	2.505865	1.774804	5.392553	1.452085	0.865667	0.442045	330.8611

Remark 4. Table 3 demonstrates the impact of increasing the arrival rate λ_1 (from 5 to 10) on system performance. All metrics exhibit monotonic growth, with E_1 showing the most dramatic increase indicating heightened congestion for priority customers. The total cost K rises substantially from 176.73 to 330.86 .

4. Effect of β on performance measures and total cost Fix $S = 12, s = 5, a = 4, N = 3, \lambda_2 = 7; \lambda_1 = 6 \mu_1 = 10, \mu_2 = 11$

Table 4: Variation of Performance measures with respect to β

β	E_1	E_2	E_I	E_R	E_L	E_S	K
14	1.513415	1.486742	4.835222	0.877562	0.778045	0.349949	258.4681
17	1.50722	1.499383	4.826463	0.87545	0.792778	0.354897	258.4007

10.48047/jocaaa.2021.29.06.39

20	1.504259	1.50827	4.819332	0.874381	0.803441	0.358557	258.3768
23	1.502688	1.514786	4.813221	0.873935	0.811488	0.361357	258.3645
26	1.501788	1.519727	4.807857	0.87387	0.81776	0.363559	258.3549
29	1.501239	1.523579	4.803088	0.874039	0.822776	0.365334	258.3454

Remark 5. Table 4 shows that as β increases from 14 to 29, E_1 decreases slightly while E_2 shows a moderate increase, indicating a redistribution of queue lengths between priority classes. Notably, the total cost K remains nearly constant as β increases. E_R shows negligible change across the tested β range. These results suggest that while β influences the priority-based queue distribution, it has impact on overall system efficiency and cost.

5. Cost optimization

Fix $S = 11, s = 5, a = 4, N = 3, \beta = 17, \lambda_2 = 7; \lambda_1 = 6, \mu_1 = 14, \mu_2 = 11$

Table 5: Variation of Cost with respect to s and S

s \ S	11	13	15	17	19	21
5	255.3403	290.1662	294.640451	316.5495	330	356
6	…	264.2802	320.373248	305.6478	350.2696	340.6037
7	…	…	345.124406	324.9167	372.2885	360.2104
8	…	…	…	308.5787	331.5715	354.3984
9	…	…	…	⋯	340.7366	368.0285
10	…	…	…	…	…	350.4209

Remark 6. From Table 5, we see that for fixed s, the total cost K generally increases with S, as seen when $s = 5$ where costs rise from $255.34(S = 11)$ to $356(S = 21)$. However, this growth is non-monotonic in some cases (e.g., $s = 6$ shows a dip at $S = 17$ before rising again). Second, for fixed S, costs tend to decrease then increase as s grows - for $S = 15$, costs fall from $294.64(s = 5)$ to $264.28(s = 6)$ before climbing to $345.12(s = 7)$, suggesting an optimal s value exists between 5-7. The most economical configuration appears at $(s, S) = (6,13)$ with $K = 264.28$, while the most expensive occurs at $(5,21)(K =$

10.48047/jocaaa.2021.29.06.39

356). These patterns demonstrate the critical trade-off between holding costs (increasing with) and shortage risks (decreasing with s) in inventory management.

5. Conclusion and future work

In this paper, we analyzed a single-server two-priority queuing inventory system with positive lead time, where customers request varying quantities of a commodity. The system is analyzed under stability. The inventory cycle time for the model was examined and a cost function was developed for the system. The cost dynamics were illustrated numerically across various parameters. In future work, we plan to investigate a system that categorizes customers according to their demand units and provides separate queues for each demand units ranging from 1 to a.

References

[1] Neuts M.F.(1994). Matrix-Geometric Solutions in Stochastic Models - An Agorithmic Approach. Dover Publication Inc., New York, 2nd ed.

[2] Krishnamoorthy, A., Joshua, A. N., and Kozyrev, D. (2021). Analysis of a batch arrival, batch service queuing-inventory system with processing of inventory while on vacation. Mathematics, 9(4):419. Available online: https://www.mdpi.com/2227-7390/9/4/419

[3] Chakravarthy, S. R., and Rumyantsev, A. (2020). Analytical and simulation studies of queueinginventory models with MAP demands in batches and positive phase-type services. Simulation Modelling Practice and Theory, 103, 102092. Available online: https://doi.org/10.1016/j simpat.2020.102092.

[4] Melikov, A. Z., and Molchanov, A. A. (1992). Stock optimization in transportation/storage systems. Cybernetics and Systems Analysis, 28(3), 484-487. Springer.

[5] Sigman, K., and Simchi-Levi, D. (1992). Light traffic heuristic for an M/G/1 queue with limited inventory. Annals of Operations Research, 40, 371-380.

[6] Krishnamoorthy, A., Shajin, D., and Narayanan, W. (2021). Inventory with positive service time: A survey. Queueing Theory, 2, 201-237.

[7] MJ, J., Rumyanstev, A., Krishnamoorthy, A., et al. (2024). On a queueing-inventory model with age-based selling of items to distinct priority groups. Operations Research Forum, 5(4), 1 – 25.

[8] Mathew, N., Joshua, V. C., Krishnamoorthy, A., Melikov, A., and Mathew, A. P. (2023). A production inventory model with server breakdown and customer impatience. Annals

10.48047/jocaaa.2021.29.06.39
of Operations Research, 331(2), 1269-1304.
[9] Rangaswamy, M., et al. (2023). Queueing-inventory systems: A survey. arXiv preprint arXiv:2308.06518.
[10] Latouche, G., and Ramaswami, V. (1999). Introduction to matrix analytic methods in stochastic modeling. SIAM.