Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

SIMULATION OF ALLOCATION POLICIES FOR A SERIAL INVENTORY SYSTEM UNDER ADVANCE DEMAND INFORMATION

Bhoomica M. Nataraja
Zümbül Atan

Department of Industrial Engineering and Innovation Sciences
Eindhoven University of Technology
P. O. Box 513
5600 MB Eindhoven, THE NETHERLANDS

ABSTRACT

In this paper, we simulate allocation policies for a two-stage inventory system that receives perfect advance demand information (ADI) from customers belonging to different demand classes. Demands for each customer class are generated by independent Poisson processes while the processing times are deterministic. All customers in the same class have the same demand lead time (the difference between the due date and the requested date) and back-ordering costs. Each stage in the inventory system follows order-base-stock-policies where the replenishment order is issued upon arrival of a customer order. The problem requires a fast and reliable method that determines the system performance under different policies and ADI. Thus, we employ discrete event simulation to obtain output parameters such as inventory costs, fill rates, waiting time, and order allocation times. A numerical analysis is conducted to identify a reasonable policy to use in this type of system.

1 INTRODUCTION

Advances in information technology in recent years, such as enterprise resource planning (ERP) and electronic data exchange (EDI), have enabled supply chain partners to collaborate and improve operational efficiencies. One type of information that helps to meet the ever-increasing customer demand at reduced costs is the information related to customers' future requirements known as Advance Demand Information (ADI). Customers place their orders with a due-date. The difference between the order arrival date and the due-date is called the demand lead time. Additionally, studies have shown that the integration of ADI with base-stock-type control policies leads to significant improvement in efficiencies (Karaesmen et al. 2002).

Conventional systems without ADI use the standard allocation method, i.e., first-come-first-served (FCFS). In these systems, the customer expects the items to be delivered immediately upon placing an order. On the contrary, in the presence of ADI, the order placement and delivery dates do not coincide. While applying the FCFS principle in this case, it is essential to determine *when* the products must be *allocated* for a specific customer. Note that the terms "reservation" and "allocation" are used interchangeably. The allocation policy translates into the prioritization of customer orders and represents a time-based reservation. Time plays a key role in the allocation. An allocation made too early or too late means incorrect prioritization and additional back-ordering costs due to delays. In this study, FCFS is considered as the general allocation policy where the products are allocated as soon as the order is placed. We further analyze policies such as priority allocation, allocation deadline, level rationing and last-minute allocation policies.

The primary motivation behind this study is the increased information sharing between supply chain companies. Moreover, it is important for companies to satisfy the demand of all their customers with specified contractual agreements. This research can benefit many companies in the automotive, high-tech, and apparel industries. We can consider the case of the Faurecia Group, one of the world's leading automotive suppliers, as an example of a supplier that manages its production based on the ADI provided by its customers such as PSA, Renault, BMW, and Toyota. The major challenge, in this case, was to know the optimal date and the level of stock to begin the production. The research presented in this paper considers two key issues: i) Inventory systems with ADI and ii) Allocation policies. A brief review of the relevant studies in these areas is given in Sections 1.1 and 1.2.

1.1 Inventory Systems with ADI

The work of Hariharan and Zipkin (1995) is one of the first papers to incorporate ADI into inventory management. Chen (2001) considers a serial inventory system with delivery due dates and uses a last-minute allocation policy. Liberopoulos (2008) investigates the trade-off between finished-goods inventory and advance demand information for a model of a single-stage make-to-stock supplier. They model the supply process as $M/M/1$, $M/D/1$ and $M/D/\infty$ queues to find the optimal order-base-stock level.

Our study can benefit manufacturing companies using ADI. Gilbert and Ballou (1999) study a steel distribution facility that handles its inventory by determining a portion of its demand as ADI. They provide an estimation of the cost savings associated with various levels of ADI. Another application field for our problem/model are spare part inventory systems that serve multiple classes of demands. Koçağa and Şen (2007) model such a system with two classes of demands: one with demands that are due immediately and another with ADI.

Further, the literature on ADI can also be based on early fulfilment. Karaesmen et al. (2004) consider a single-stage make-to-stock system for one class of demand where early delivery is allowed. Sarkar (2007) allows early fulfilment for a system with multiple demand classes as an alternative strategy for ADI. Further, Wang and Toktay (2008), in their work, delay the fulfilment decision until the end of the period, thus proving that increasing demand lead time is more beneficial than reducing supply lead time when early fulfilment is possible. Available to Promise (ATP) is an alternative to ADI that addresses order fulfilment problems. Kilger and Meyr (2008) have proposed an ATP search method to allocate orders to customers. Meyr (2009) further proposes ATP allocation for different customer classes with the basic idea that ATP is held back in anticipation of later arriving, more profitable orders even if a less profitable order already requests the product stock.

1.2 Allocation Policies

Marklund (2006) introduces the idea of employing a reservation policy at the upper-level echelon. The policies analyzed were *complete reservation policy* where the reservations are made based on an FCFS basis at the time the order is placed (as early as possible) and *last-minute allocation policy*, in which reservations are made at the time when the order is to be shipped (as late as possible).

Chen and Samroengraja (2000) study two allocation policies: the *past priority allocation* (PPA) where back-ordered units are allocated on an FCFS strategy; and the *current priority allocation policy* (CPA), where the allocation of current, as well as backlogged orders, is made at the last minute to avoid earmarking of the inventories. According to their numerical study, the CPA policy performs better on average; however, it does not dominate the PPA policy.

Further, when considering conventional systems without ADI but with heterogeneous demand, some of the literature, worth mentioning, is as follows: Topkis (1968) shows that the optimal policy has a particular threshold structure that reserves items in stock for future (uncertain) demands of more valuable customers. Frank et al. (2003) analyze the rationing problem for two classes of customers where the demands of the first class must be completely satisfied, but the demands of the second class can be partially satisfied.

The remainder of this paper is organized as follows: In Section 2, we introduce the notations and analyze the model. We introduce the different allocation policies used in the model in Section 3. The simulation model is presented in Section 4. In Section 5, we summarize the results of the numerical analysis. Finally, we present our conclusions and suggest future research directions in Section 6.

2 PROBLEM CONTEXT

2.1 Inventory Model

We consider a two-stage serial inventory system with ADI serving a single product to two customer classes, as illustrated in Figure 1. Similar to the model considered by Hariharan and Zipkin (1995), we consider that both the stages use order base-stock policies to replenish their inventory. This policy works like a conventional one-for-one replenishment policy, except that replenishment orders are triggered by customer orders instead of actual demands. When a customer order occurs at stage 2, a replenishment order is placed for stage 1. Stage 1 replenishes its inventory from an uncapacitated supplier. The lead time for these replenishments, L_1, is constant. The transportation lead time from stage 1 to stage 2, L_2, is also constant. However, the replenishment lead time for stage 2 is stochastic due to possible stockouts at stage 1. All customers of class j have a demand lead time LD_j (time from a customer order until the associated customer demand is realised). $LD = \frac{1}{\lambda}\sum_{j=1}^{2}\lambda_j LD_j$, denotes the average demand lead-time.

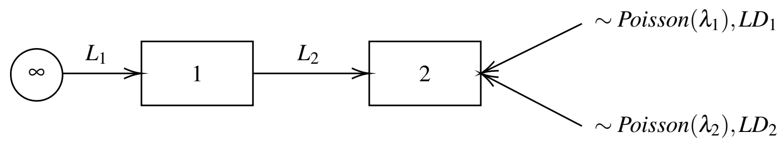


Figure 1: Two-stage serial system with two demand classes.

2.2 Model Assumptions

We consider the following assumptions. The order in each customer class follows an independent Poisson process, and the arrival rate for class j is λ_j arrivals per time unit. Let $\lambda = \lambda_1 + \lambda_2$, be the total arrival rate. We assume perfect ADI, i.e., each customer order includes a due date and, once placed, cannot be revised. It is assumed that the customers will not accept early deliveries. Every arriving demand is for one unit, and all unfulfilled demands are back-ordered. There are no economies of scale, so there is no incentive for batch orders. Customer's preferences for demand lead-times are outside the control of the inventory system. Further, the allocation decision is final, i.e., an allocated item cannot be used to fulfil another order.

2.3 Model Analysis

Figure 2 illustrates the sequence of events. Every order has an exact due date given by LD_j for customer class j, and there are three cases to consider:

1. When $LD_j \geq L_1 + L_2$, all demands can be met while holding no inventory at either stage, i.e., replenishment order for stage 1 is delayed until $L_1 + L_2$ before the due date (no early fulfilment).
2. Suppose $L_2 \leq LD_j < L_1 + L_2$, then, there is no need to hold stock at stage 2. Thus, this system can be considered as a single-stage system.
3. When $LD_j < L_2 < L_1 + L_2$, we have a conventional two-stage serial inventory system.

Based on this observation, we will consider cases with $\mathbf{LD_j < L_1 + L_2}$ throughout the study. Let h_i be the inventory holding cost rate at stage i and b be the back-order cost at stage 2. We assume $0 \leq h_1 \leq h_2$. Note that if $h_1 > h_2$, then the item should be kept at stage 2 to reduce costs. Further, if the customer classes have different back-ordering costs, the average back-order cost at stage 2 will be $b = \frac{1}{\lambda}\sum_{j=1}^{2}\lambda_j b_j$. The average delay from the receipt of an order until allocation is $LA_j \geq 0$, and a customer j placing an order at time t with demand lead-time LD_j will have an item allocated from inventory at $t + LA_j$.

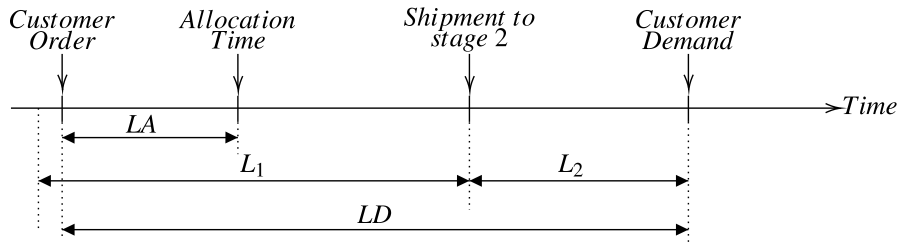


Figure 2: Sequence of events for $L_2 < LD < L_1 + L_2$.

Let s_i be the local base stock level at stage i. Assuming that each stage starts with inventory s_i and an empty supply system, then, each customer order at stage 2 triggers a replenishment order for stage 1. Stage 2 monitors its inventory position and places orders with stage 1, similar to a single location operating with a base stock policy. This situation is analogous for stage 1, which subsequently treats the incoming orders as its demand. Thus, every stage experiences its own demand process. For $t \geq 0$, we define:

- $I_i(t)$: local inventory at stage i
- $B_i(t)$: local back-orders at stage i
- $IN_i(t)$: local net inventory at stage i = $I_i(t) - B_i(t)$
- $IO_i(t)$: inventory on order at stage i
- $IP_i(t)$: local inventory order-position at stage i = $IN_i(t) + IO_i(t) = I_i(t) - B_i(t) + IO_i(t)$
- $IT_i(t)$: inventory in transit to stage i
- $D(t_1, t_2]$: cumulative customer demand from time t_1 through time t_2, $t_1 < t_2$
- $D^-(t_1, t_2]$: cumulative customer orders from time t_1 through time t_2, $t_1 < t_2$

From the above definition, we have $IO_i(t) - IT_i(t) = B_{i-1}(t)$. We assume that the outside source immediately responds to the orders from stage 1, thus, $IT_1(t) = IO_1(t) \implies B_0(t) = 0$. Further, $IP_i(t) = IN_i(t) + IO_i(t)$ is non-stationary over time and thus, we cannot define $IP_i(t) = s_i$, $\forall t > 0$. So, we define a modified inventory position $IP_i^-(t) = IP_i(t) - D(t, t+LD]$. This modified inventory depends on the demand during the length of the demand lead time. From this, it can be implied that $IP_i^-(t)$ has a stationary behavior over time and $IP_i^-(t) = s_i$. Based on this information, the inventory system can be evaluated following the conservation-of-flow law.

For stage 2, when $LD \geq L_2$, exactly $IO_2(t)$ demands will be realized in the time interval $(t, t+LD]$. The net inventory at stage 2 is

$$\begin{aligned} IN_2(t+LD) &= IN_2(t) + IT_2(t) - D(t,t+LD] = IP_2(t) - B_1(t) - D(t,t+LD] \\ &= IP_2^-(t) - B_1(t) = s_2 - B_1(t). \end{aligned}$$

Suppose, $LD < L_2$, in this case no ADI is provided to stage 1. Then, the net inventory at stage 2 is

$$\begin{aligned} IN_2(t+L_2) &= IN_2(t) + IT_2(t) - D(t,t+L_2] = IN_2(t) + IO_2(t) - B_1(t) - (D(t,t+LD] + D(t+LD, t+L_2]) \\ &= IP_2^-(t) - B_1(t) - D^-(t, t+L_2-LD] = s_2 - B_1(t) - D^-(t, t+L_2-LD]. \end{aligned}$$

Thus, the difference in the above two equations is the customer order during the time interval $(t, t+L_2-LD]$. So, we define, $D_2(t)$ = number of customer orders at stage 2 in the time interval $(t, t+(L_2-LD)^+]$ i.e., $D_2(t) \sim Poisson(\lambda(L_2-LD)^+))$. For $LD < L_1 + L_2$, we evaluate the inventory level of stage 1 as

$$IN_1(t+L_1+L_2) = IN_1(t) + IT_1(t) - D(t,t+L_1+L_2] = IP_1(t) - (D(t,t+LD] + D(t+LD, t+L_1+L_2]).$$

Now, if $LD \geq L_2$,

$$\begin{aligned} IN_1(t+L_1+L_2) &= IP_1(t) - D(t,t+LD] - D(t+LD, t+L_1+L_2] \\ &= IP_1^-(t) - D^-(t, t+L_1+L_2-LD] = s_1 - D^-(t, t+L_1+L_2-LD]. \end{aligned}$$

When $LD < L_2$,

$$\begin{aligned} IN_1(t+L_1+L_2) &= IP_1(t) - D(t,t+LD] - D(t+LD,t+L_2] - D(t+L_2,t+L_1+L_2] \\ &= s_1 - D^-(t,t+L_2-LD] - D^-(t+L_2-LD,t+L_1+L_2-LD]. \end{aligned}$$

Again, the inventory level at stage 1 depends on the customer orders during the time interval $(t,t+(L_2-LD)^+]$. Hence, we define $D_1(t)$ = number of customer orders at stage 1 in the time interval $(t,t+L_1+L_2-LD-(L_2-LD)^+]$ i.e., $D_1(t) \sim Poisson(\lambda(L_1-(LD-L_2)^+))$.

By omitting the time index to describe the model in equilibrium, we generalize the equations as below.

$$IN_i = s_i - B_{i-1} - D_i \tag{1}$$

$$B_i = [B_{i-1} + D_i - s_i]^+ = [-IN_i]^+ \tag{2}$$

Thus, the evaluation of this two-stage series system can be verified by an equivalent conventional system as proposed by Hariharan and Zipkin (1995) and Marklund (2006) using leadtimes shown in equation (3).

$$L'_2 = [L_2 - LD]^+, \quad L'_1 = [L_1 + L_2 - LD]^+ - L'_2 \tag{3}$$

2.4 Performance Measures

To evaluate the total average costs for FCFS, we consider the average values: b, (LD) and λ. A conventional two-stage serial system incurs the following total average cost (TC).

$$E[\sum_{i=1}^{2} h_i(I_i + IT_{i+1}) + bB_i] \tag{4}$$

where $E[IT_i] = \lambda L'_i$ is the number of units sent to stage $i's$ supply system per unit time which stay there for time L'_i. The difference between the two-stage series system with ADI and the conventional system is that the former has a longer stage-2 lead time, its average in-transit inventory is larger by the constant value λLD which incurs the stage-1 holding cost. Thus, the system defined in this paper incurs additional costs of $\mathbf{h_1 \lambda LD}$ (Hariharan and Zipkin 1995). We employ Two-Moment Approximation procedure (Graves 1985) to determine B_2, alternatively to numerical convolution of B_1 and D_2.

A key performance measure that plays an important role in the multi-customer problem is the fill rate. We define β_j as fill rate, the expected proportion of class-j demands met directly from stock. Thus, for $L_2 \leq LD < L_1 + L_2$, the fill rate is given by $\mathbb{P}(I_1 > 0) = \mathbb{P}(D_1 = S_1 - 1)$ as $I_1 = 0$ when $D_1 \geq s_1$. Here, we define L_{j_1} as the stage 1 lead time of the equivalent conventional system for customer class j. Since D_1 is Poisson-distributed with mean $\lambda L'_{j1} = \lambda(L_1 - (LD_j - L_2)^+)$ and the distribution of a Poisson random variable X with mean Λ is $P\{X = x\} = \frac{\Lambda^x}{x!}e^{-\Lambda}$, we obtain the fill rate for customer class j as

$$\beta_j = \mathbb{P}(D_1 < s_1) = e^{-\lambda L'_{j1}} \sum_{n=0}^{s_1-1} \frac{\lambda L'_{j1}}{n!}. \tag{5}$$

We consider another service measure: the expected waiting time until demand is fulfilled. This measure, *mean waiting time*, denoted by W_j for customer class j can be calculated using Little's law. For an average waiting time W and arrival rate λ, the average queue length is given by $L = \lambda W$. In our model, backorders form the queue length. Hence, the aggregate mean waiting time will be $W = E[B_2]/\lambda$.

3 ALLOCATION POLICIES

The main focus of this paper is to understand the performance of the model under different allocation strategies and to address the question *when* would it be ideal to allocate or reserve the item. Different policies (P) and allocation criteria are further explained in this section.

P = 0: First-Come-First-Served: Usually, this policy is considered the best when the customer classes have the same back-ordering costs and demand lead times, i.e., equal priorities. In this policy, both customer orders as well as back-orders are satisfied in the order of their arrival irrespective of the customer priorities.

P = 1: Priority Allocation: We evaluate this policy with two priority options: (1) customer class with the highest back-ordering cost; (2) customer with the shortest demand lead time. When the on-hand inventory at stage 1 is $IOH_1 > 0$, the demands are satisfied according to the FCFS strategy regardless of the customer class. In the presence of back-orders, the items arriving from the supplier to stage 1 are first allocated to the priority class and then to the non-priority class.

P = 2: Allocation Deadline Policy: Allocation deadline for any customer class is the difference between demand lead time and actual lead time. Customers are sorted in ascending order of their allocation deadlines. Both the arriving orders and the back-orders are allocated on arrival in the order of their deadlines.

P = 3: Inventory Rationing Policy: Assuming class 1 is the priority class and z is the threshold stock level, arriving demand of class 2 is satisfied if the $IOH_1 \geq z$; otherwise, only the demand of class 1 is satisfied. In the presence of back-orders, the item is allocated to class 2 if $IOH_1 \geq z$ and class 2 has the highest back-ordering cost; otherwise, the item is reserved for class 1. Thus, orders from both the classes are satisfied when $IOH_1 \geq z$; otherwise, only class 1 orders are satisfied.

P = 4: Last-minute Allocation: Allocation is delayed until L_2 time units before the arrival of the actual demand. As illustrated in Figure 3, $P4$ delay is the difference between the actual lead time and the time spent in the system so far i.e., $P4 = (LD_j - L_2 - \text{Time spent in the system})$. However, under this policy, the allocation occurs according to FCFS, but after $P4$ time units.

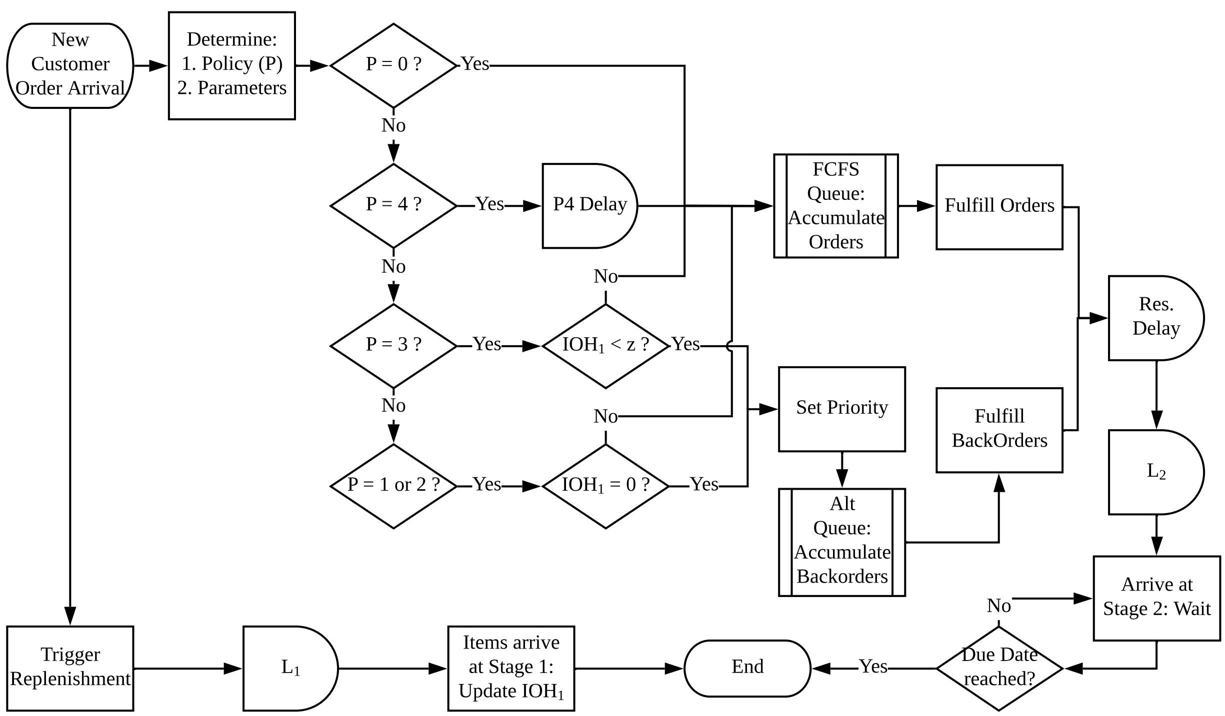


Figure 3: Flow of process.

4 SIMULATION MODEL

It is time-consuming to evaluate analytically different policies with various parameters. Thus, we employ discrete event simulation, which enables a dynamic change in the demand pattern and the model parameters. We have implemented the simulation model and its parameters in AnyLogic Simulation Modeling Software. To represent a realistic situation such as Assemble-to-Order, Buy-to-Order or Make-to-Order with a due date, we consider stage 1 as the Customer Order Decoupling Point which starts assembling or producing for L_2 time units with ample capacity and keeps the finished goods at stage 2. The stock at stage 1 buffers against the customer orders while we keep zero stock at stage 2.

4.1 Input Parameters

The input parameters (λ_j, b_j, LD_j, L_i , s_i $\forall$ stage i and customer class j) defined in Section 2 are used in the simulation model. In addition, we introduce a new notation pw_j to represent the customer order priority. The higher the weight (pw_j), the higher the priority.

4.2 Process Flow

The process flow is illustrated in Figure 3. The events that occur in the simulation are: (1) At the beginning, the inventory level of stage 1 is set to s_1 while we stock nothing at stage 2 ($s_2 = 0$); (2) Statistics are initialized to record the output parameters; (3) Upon generation of a customer order, a replenishment order for stage 1 is triggered which takes L_1 time to reach stage 1 and updates on-hand inventory at stage 1 (IOH_1). Simultaneously, the customer order is queued up in an FCFS queue or the alternative queue based on the policy; (4) The order seizes an item from the inventory and moves to stage 2 with a delay of L_2. There is a reservation delay which delays the order at stage 1 such that the items reach stage 2 only on the due date. However, this delay is set to zero for the FCFS base model; (5) Once the items arrive at stage 2, they are released out of the system only on the due date, and until then they wait at stage 2; (6) Various performance measures are recorded during the simulation.

4.3 Output Parameters

The output parameters are tabulated in Table 1. Similar to back-ordering costs and demand lead time, the average of waiting times of different customer classes is calculated as $WT = \frac{1}{\lambda}\sum_{j=1}^{2}\lambda_j WT_j$.

Table 1: Output parameters of the simulation model.

<table>
<tr><th colspan="2">Output Parameters</th><th>Collected at</th><th>Description</th><th>Notation</th></tr>
<tr><td colspan="2">Avg Back-ordering Costs</td><td>End of period</td><td>Total number of orders present in the system that have surpassed the due date * b_i</td><td>BC</td></tr>
<tr><td colspan="2">Avg Holding Costs</td><td>End of period</td><td>($IOH_1 \times h_1$) + (items between stage 1 and 2 $\times$ h_1) + (items in stage 2 $\times$ h_2)</td><td>HC</td></tr>
<tr><td colspan="2">Total Costs</td><td>End of period</td><td>$BC+HC$</td><td>TC</td></tr>
<tr><td rowspan="4">per Customer class j</td><td>Avg Fill Rate</td><td>Order exits stage 2</td><td>$\frac{\text{number of orders fulfilled within due date}}{\text{Total number of orders}}$</td><td>$FR_j$</td></tr>
<tr><td>Avg Allocation Lead Time</td><td>Order exits stage 1</td><td>Allocation Time - Order Generated Time</td><td>AL_j</td></tr>
<tr><td>Avg Waiting Time</td><td>Order exits stage 2</td><td>Exit Time at stage 2 - Order Generated Time - LD_i</td><td>WT_j</td></tr>
<tr><td>Avg Reservation Delay</td><td>Order exits stage 1</td><td>Shipment start time to stage 2 - Allocation Time</td><td>RT_j</td></tr>
</table>

4.4 Model Validation

The model run parameters for all the experiments are: a) Each period is equal to one day; b) The run length is set at 2 years, to represent a typical business operating cycle; c) The number of replications is set to 40 with random seed values and the variance obtained during the pilot runs was considered acceptable.

Table 2: Simulation and analytical results under FCFS policy, ± is based on 95% C.I.

#	(λ_1,λ_2)	(b_1,b_2)	(LD_1,LD_2)	(L_1,L_2)	(h_1,h_2)	(s_1,s_2)	TC		FR_1		FR_2		WT	
							Analy.	Sim.	Analy.	Sim.	Analy.	Sim.	Analy.	Sim.
1	(1,1)	(10,10)	(5,5)	(5,3)	(2,2.5)	(5,0)	36.155	37.454±0.707	0.285	0.27±0.013	0.285	0.28±0.014	0.757	0.751±0.023
2	(0.4,0.6)	(10,10)	(5,5)	(5,3)	(2,2.5)	(5,0)	15.615	15.965±0.191	0.815	0.794±0.01	0.815	0.79±0.01	0.135	0.129±0.009
3	(1,1)	(10,10)	(4,6)	(5,3)	(2,2.5)	(5,0)	36.155	39.984±0.511	0.100	0.111±0.005	0.629	0.602±0.008	0.757	0.868±0.014
4	(0.4,0.6)	(10,10)	(4,6)	(5,3)	(2,2.5)	(5,0)	15.810	16.9±0.217	0.629	0.623±0.012	0.947	0.907±0.005	0.121	0.172±0.009
5	(0.4,0.6)	(5,10)	(5,5)	(5,3)	(2,2.5)	(5,0)	15.346	15.138±0.164	0.815	0.789±0.011	0.815	0.782±0.009	0.135	0.138±0.008
6	(0.4,0.6)	(5,10)	(3,3)	(5,3)	(2,2.5)	(5,0)	14.763	17.683±0.386	0.440	0.43±0.013	0.440	0.438±0.012	0.876	0.867±0.031
7	(0.4,0.6)	(5,10)	(12,10)	(5,3)	(2,2.5)	(5,0)	27.549	28.055±0.248	1.000	0.958±0.004	1.000	0.956±0.003	0.000	0.000
8	(0.4,0.6)	(5,10)	(4,6)	(5,3)	(2,2.5)	(3,0)	16.413	17.776±0.222	0.238	0.243±0.01	0.677	0.646±0.009	0.561	0.661±0.016
9	(0.4,0.6)	(5,10)	(7,8)	(5,6)	(2,2.5)	(5,0)	20.625	20.797±0.254	0.629	0.613±0.016	0.815	0.787±0.009	0.222	0.236±0.012
10	(0.4,0.6)	(5,10)	(7,8)	(10,3)	(2,2.5)	(5,0)	25.527	23.409±0.413	0.285	0.27±0.012	0.440	0.402±0.013	1.113	1.131±0.037

To validate the simulation model, we vary the input parameters for the FCFS policy ($pw_1 = pw_2$). Furthermore, we calculate 95% confidence intervals for the means of the selected performance measures. Table 2 shows a close agreement between the simulated and analytical results for the cases considered. We make the following observations. We set different cases to obtain the conformity of the model. Cases 3 and 4 are special cases of 1 and 2 with different demand lead times. Case 2 is modified in terms of penalty cost to get case 5. We also check for three cases of demand lead times: Cases 1 through 5 for $L_1 + L_2 > LD > L_2$, case 6 for $LD = L_2$ and case 7 for $LD > L_1 + L_2$. Stock level s_1 in case 4 is decreased to 3 and verified in case 8. Finally, from cases 9 and 10, it can be inferred that an increase in L_1 has a more significant influence on the model than that of L_2.

5 NUMERICAL ANALYSIS

This section investigates the performances of allocation policies for the model in Section 4. Further, we seek answers to the following questions: Which system parameters affect the average cost the most? How does early fulfilment play a role in the system? How do the policies affect customer service levels?

5.1 Performance of the Allocation Policies

Our first set of results demonstrates the performance of the model under different allocation policies using the problem data in Table 3. Here, early fulfilment is not permitted, i.e., the items wait at stage 2, if the customer demand for the corresponding order has not arrived.

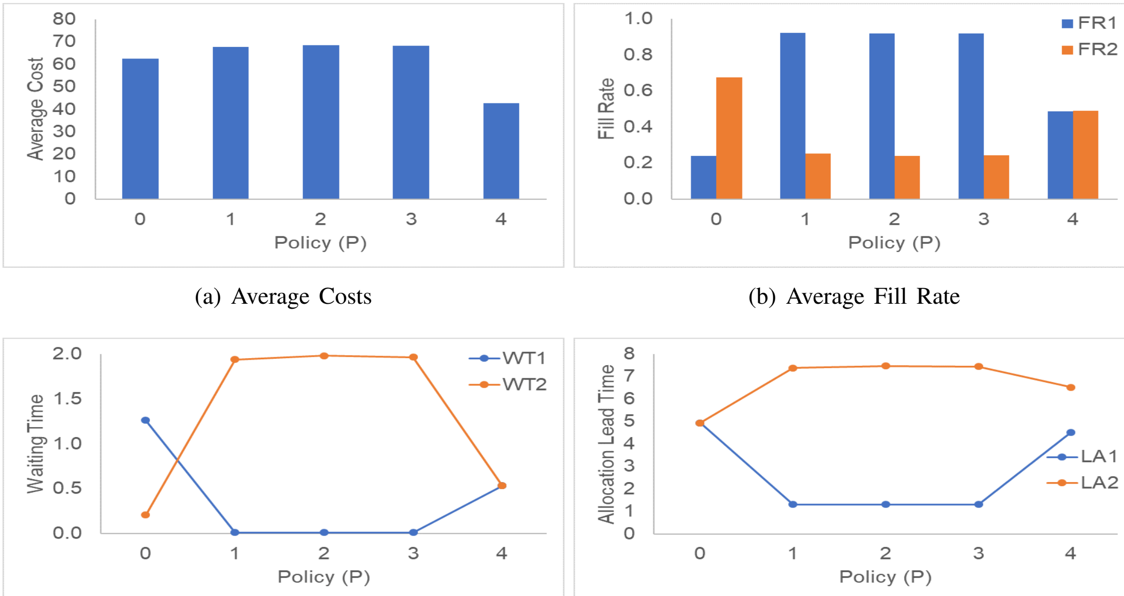
Table 3: Input data for allocation policy analysis.

Parameter	(λ_1,λ_2)	(b_1,b_2)	(LD_1,LD_2)	(L_1,L_2)	(h_1,h_2)	(s_1,s_2)	(pw_1,pw_2)
Value	(0.8, 1.2)	(20, 10)	(7, 9)	(10, 3)	(3, 5)	(10, 0)	(10, 1)

From the plots in Figure 4, we make several observations. First, the average costs incurred under the policy 4 is the least since the items wait at stage 1 and $h_1 < h_2$. Second, we observe from Figure 4(b) that the fill rates are higher for prioritized customers under $P = 1,3$ and customers with the shortest LD under $P = 2$. Third, policy 4 dominates all the policies in terms of waiting time. Finally, the allocation lead time for policy 0 is, on average, 5 time units after order arrival. Another interesting observation is that the allocation lead times under policy 4, $LA_1 = 4.532$ and $LA_2 = 6.529$, give the same difference as $(LD_1, LD_2) = (7,9)$. As expected, $LA_1 < LA_2$ under policies 1, 2 and 3 due to customer class 1 prioritization.

The assumption that early fulfilment is forbidden is not realistic in all scenarios. Nowadays, early fulfilment is appreciated, which would lead to a reduction in holding costs. For the test case in Table 3, when the early fulfilment is permitted, we observe a significant decrease in costs as the item does not incur the holding costs at stage 2. Last-minute allocation policy would not perform differently since it always delays allocation by $DueDate - L_2$ time at stage 1. The costs are reported in Table 4.

In Section 4.3, we introduced reservation time delay. After allocation of an item, the item is reserved/delayed at stage 1 until L_2 time units to the due date. So far, RT was zero. With early fulfilment



(a) Average Costs (b) Average Fill Rate

(c) Average Waiting Time (d) Average Allocation Lead Time

Figure 4: Simulation results under different allocation policies (early fulfilment forbidden).

forbidden and reservation delay initiated, we can decrease costs. If early fulfilment is forbidden, then the introduction of a reservation delay would prove to be a beneficial option.

To obtain a near-optimal stock level for stage 1, we perform an enumeration by varying s_1 from 0 to 30 and record the costs at the end of each iteration. The simulation run for each policy stops when the minimum cost value is reached and records the corresponding s_1 value in that iteration. From Table 4, we observe that the near-optimal stock value increases when early fulfilment is permitted. This is to be expected since s_1 is primarily a contributor to the costs.

Table 4: Near-optimal values of s_1 obtained for test case in Table 3.

Early Fulfillment	P = 0		P = 1		P = 2		P = 3		P = 4	
	s_1	Cost	s_1	Cost	s_1	Cost	s_1	Cost	s_1	Cost
No	12	57.884	13	60.091	14	60.312	14	60.297	13	33.996
Yes	19	29.806	19	29.667	19	29.679	19	29.729	13	33.765

5.2 Sensitivity Analysis

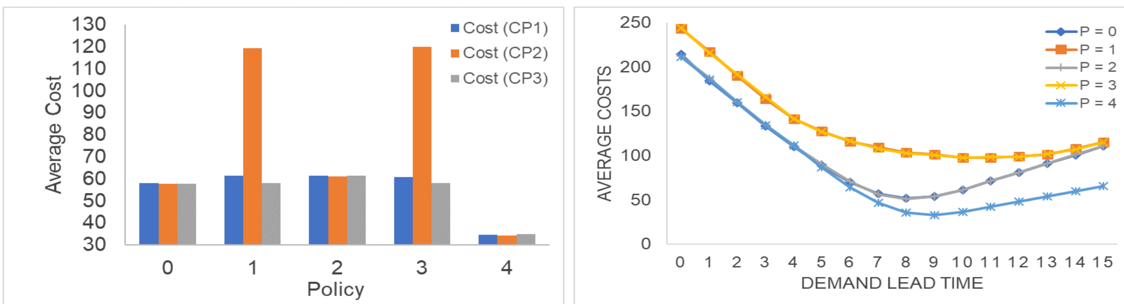
The results in Section 5.1 suggest that there is a trade-off between system costs and fill rates. In this section, we perform a sensitivity analysis to understand the impact of system parameters on performance measures. We use the data in Table 3 with $(s_1, s_2) = (12, 0)$ and vary one parameter at a time to observe its effect.

5.2.1 Effect of Arrival Rate

We vary arrival rates for both the customer classes from 0.25 to 2. We noted that under the FCFS policy $(P = 0)$, $WT_1 > WT_2$ even when $\lambda_1 < \lambda_2$ possibly because $LD_2 > LD_1$, giving the system more time to process orders from customer class 2. As expected, under policies 1, 2, and 3, priority class had shorter waiting times. When the total arrival rate doubles, policies 0 through 3 show a drastic decrease in the fill rates. However, under policy 4, both customer classes have equal waiting times and a gradual decrease in fill rates.

5.2.2 Effect of Customer Priority

Policies 1, 2, and 3, thus far, provided similar results since $pw_1 > pw_2$. Here, we evaluate scenarios: CP1: Priority to Class 1; CP2: Priority to Class 2; CP3: Equal priority to both classes. The CP2 scenario indicates the worst case, while the scenarios with CP1 indicate the best case concerning the costs. Average costs are shown in Figure 5(a). The significant increase in cost for CP2 is seen for policies 1 and 3 due to the non-prioritization of Class 1, which has a higher back-ordering cost and prioritization of Class 1 which has a higher arrival rate but a longer *LD*. Little difference is found between CP1 and CP3. It is obvious to say that prioritization is suitable for classes with higher back ordering costs than that with higher arrival rate.

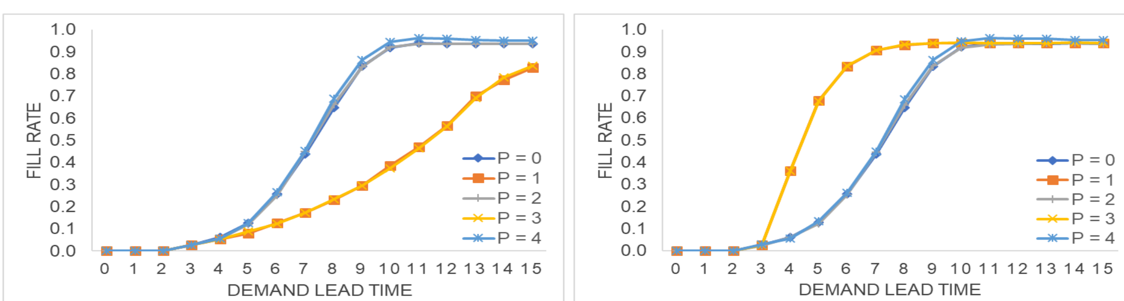


(a) Average costs vs. customer class priority. (b) Average costs vs demand lead time.

Figure 5: Effect of customer class priority and demand lead time on average costs.

5.2.3 Effect of Demand Lead time

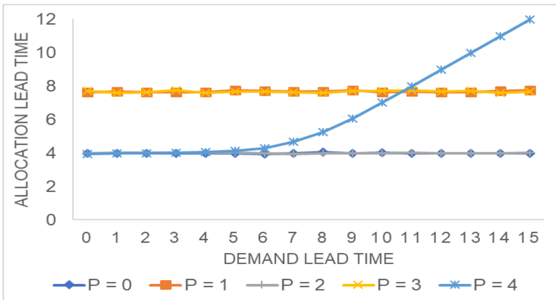
The demand lead time is incremented from 0 to 15 and $LD = LD_j$, $\forall j = 1,2$ with $pw_2 > pw_1$. First, we analyze the costs of the different policies for varying lead times. Since the demand lead time is the same for both the customer classes, the allocation deadlines would be equal. Thus, policy 2 acts as FCFS. Similarly, policy 1 and policy 3 give the same results, as shown in Figure 5(b). Policy 4 dominates the other policies when $LD > L_1 + L_2$ as the orders incur holding costs of stage 1 while orders under other policies wait at stage 2 incurring higher holding costs. In general, all the policies incur higher costs when $LD < L_2$, i.e., when no ADI is available. Costs decrease as the demand lead time increases. However, at some point for the policies, the costs start increasing as the items are held at either stage 1 or stage 2 until the due date.



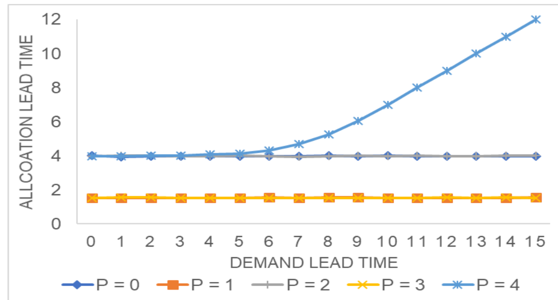
(a) FR_1 *vs. LD* (b) FR_2 *vs. LD*

Figure 6: Effect of demand lead time on fill rates.

Both the customer classes have the same fill rates under policies 0, 2, and 4. In Figures 6(a) and 6(b), we observe a left shift in FR_2 as class 2 customers are prioritized. Thus, considering fill rates, policies 1 and 3 dominate the other policies. It is evident that as the demand lead time increases, fill rates increase for both the customer classes, thus proving investment in ADI beneficial. Mean waiting times are vertical reflection of fill rate plots, i.e. they reach zero when $LD = max(L_1, L_2)$ under policies 0, 2, and 4.



(a) AL_1 *vs.* LD



(b) AL_2 *vs.* LD

Figure 7: Effect of demand lead time on allocation lead time.

The allocation lead time (AL) also shows an interesting output with different demand lead times in Figure 7. Policies 0 and 2 allocate an item after 4 time units of order arrival irrespective of the demand lead time. Under policies 1 and 3, $AL_2 < AL_1$ as customer class 2 has a higher priority. As expected, under policy 4, AL increases with LD when $LD > L_2$. Finally, we analyze the effect of demand lead time on the optimal stock level (s_1^*). The graph in Figure 8 indicates that for all the policies, s_1^* decreases as LD increases. We expected that stock level goes to zero when $LD = L_1 + L_2$ and this occurs under policy 0, 2, and 4. However, the priority-based policies (1 and 3) show a higher base-stock level $s_1 > 0$ even when $LD > L_1 + L_2$.

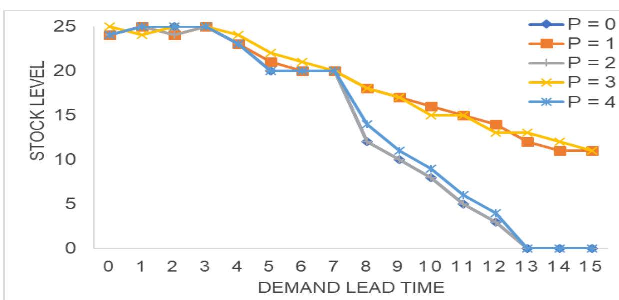


Figure 8: Near-optimal base stock level s_1^* for different demand lead times.

6 CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

In this paper, we used simulation experiments to investigate the performance of a two-stage serial system with ADI and two customer classes. We verified the model analytically and studied the impact of several input parameters on the performance measures. Based on numerical analyses, we infer that the last-minute allocation policy (P=4) is cost-effective in all the situations. However, this is suitable for companies that operate with lower service levels. When it comes to prioritization, customers with highest back-orders should be prioritized rather than the ones with shorter demand lead time. This consideration holds for both priority allocation (P=1) and level-rationing (P=3). If tardiness is an important performance criterion, then the allocation deadline policy (P=2) performs best while maintaining the desired fill rates. The simulation model developed offers a reliable and efficient method to deal with the dynamic environment faced by a firm and remodel the operations in terms of customer service and operating costs.

It might be interesting to see how the policy parameters change when different types of ADIs (Thonemann 2002) are used. Further, the model can be scaled to more than two demand classes. The challenge is to calculate the policy parameters for this divergent problem analytically; but, the simulation model can be used to solve the problem. Furthermore, the model can be extended to capacitated production-inventory problems or periodic-review systems. Time of allocation can be considered as a decision variable to study how the system performs under different allocation lead times. These improvements will provide more insights for a company to fulfil its business objectives.

REFERENCES

Chen, F. 2001. "Market Segmentation, Advanced Demand Information, and Supply Chain Performance". *Manufacturing & Service Operations Management* 3(1):53–67.

Chen, F. and R. Samroengraja. 2000. "A Staggered Ordering Policy for One-Warehouse, Multiretailer Systems". *Operations Research* 48(2):281–293.

Frank, K. C., R. Q. Zhang, and I. Duenyas. 2003. "Optimal Policies for Inventory Systems with Priority Demand Classes". *Operations Research* 51(6):993–1002.

Gilbert, S. M. and R. H. Ballou. 1999. "Supply Chain Benefits from Advanced Customer Commitments". *Journal of Operations Management* 18(1):61–73.

Graves, S. C. 1985. "A Multi-Echelon Inventory Model for a Repairable Item with One-for-One Replenishment". *Management Science* 31(10):1247–1256.

Hariharan, R. and P. Zipkin. 1995. "Customer-Order Information, Leadtimes, and Inventories". *Management Science* 41(10):1599–1607.

Karaesmen, F., J. A. Buzacott, and Y. Dallery. 2002. "Integrating Advance Order Information in Make-to-Stock Production Systems". *IIE Transactions* 34(8):649–662.

Karaesmen, F., G. Liberopoulos, and Y. Dallery. 2004, Feb. "The Value of Advance Demand Information in Production/Inventory Systems". *Annals of Operations Research* 126(1):135–157.

Kilger, C. and H. Meyr. 2008. "Demand Fulfilment and ATP". In *Supply Chain Management and Advanced Planning: Concepts, Models, Software, and Case Studies*, edited by H. Stadtler and C. Kilger, 181–198. Berlin, Heidelberg: Springer.

Koçağa, Y. L., and A. Şen. 2007. "Spare parts inventory management with demand lead times and rationing". *IIE Transactions* 39(9):879–898.

Liberopoulos, G. 2008. "On the Tradeoff Between Optimal Order-Base-Stock Levels and Demand Lead-Times". *European Journal of Operational Research* 190(1):136 – 155.

Marklund, J. 2006. "Controlling Inventories in Divergent Supply Chains with Advance-Order Information". *Operations Research* 54(5):988–1010.

Meyr, H. 2009. "Customer Segmentation, Allocation Planning and Order Promising in Make-to-Stock Production". *OR Spectrum* 31(1):229–256.

Sarkar, S. 2007. *The Effect of Advance Demand Information on a Pull Production System with Two Customer Classes*. Ph.D. thesis, Virginia Tech.

Thonemann, U. 2002. "Improving Supply-Chain Performance by Sharing Advance Demand Information". *European Journal of Operational Research* 142(1):81–107.

Topkis, D. M. 1968. "Optimal Ordering and Rationing Policies in a Nonstationary Dynamic Inventory Model with n Demand Classes". *Management Science* 15(3):160–176.

Wang, T. and B. L. Toktay. 2008. "Inventory Management with Advance Demand Information and Flexible Delivery". *Management Science* 54(4):716–732.

AUTHOR BIOGRAPHIES

BHOOMICA M NATARAJA is a Ph.D. candidate in the Department of Industrial Engineering at the Eindhoven University of Technology. Her research interests include simulation modelling and stochastic optimization techniques, with a focus on multi-echelon inventory systems and warehouse storage policies. Her email address is b.m.nataraja@tue.nl.

ZÜMBÜL ATAN is an Associate Professor in Department of Industrial Engineering at Eindhoven University of Technology. She does research on optimization of multi-echelon supply chains subject to demand and supply uncertainty, revenue management and retail operations in collaboration with multiple international universities and companies. Her email address is z.atan@tue.nl.