Kybernetika

Abdullah S. Karaman

Approximate evaluation of continuous review (R,Q) policies in two-echelon inventory systems with stochastic transportation times

Kybernetika, Vol. 53 (2017), No. 3, 461–479

Persistent URL: `http://dml.cz/dmlcz/146937`

KYBERNETIKA — VOLUME *53* (2017), NUMBER 3, PAGES 461–479

APPROXIMATE EVALUATION OF CONTINUOUS REVIEW (R, Q) POLICIES IN TWO-ECHELON INVENTORY SYSTEMS WITH STOCHASTIC TRANSPORTATION TIMES

Abdullah S. Karaman

This paper considers a distribution inventory system that consists of a single warehouse and several retailers. Customer demand arrives at the retailers according to a continuous-time renewal process. Material flow between echelons is driven by reorder point/order quantity inventory control policies. Our objective in this setting is to calculate the long-run inventory, backorder and customer service levels. The challenge in this system is to characterize the demand arrival process at the warehouse. We present a Markovian methodology to elucidate and approximate this process. We illustrate the use of this methodology in the distribution inventory system under stochastic transportation times with identical and non-identical retailers.

Keywords: inventory control, Markovian analysis, stochastic lead-times, distribution inventory systems

Classification: 60J27, 90B05

1. INTRODUCTION

A supply chain is defined as a system that moves goods, services, and information from points-of-origination to points-of-consumption. It includes a set of processes to efficiently link suppliers, manufacturers, distributors, and retailers in order to acquire raw materials, transform them into final products, and ship the final products to intermediate storage locations, retailers, and customers. Clearly, effective flow in the supply chain requires concerted activity across all the business functions. Replenishment, at any node in the supply chain, at the right quantity and at the right time is achieved by using proper inventory control policies.

In this paper, we consider a distribution inventory system including a single warehouse and several retailers. Demand arises in the retailers in the form of a stationary stochastic process. An inventory control policy is used to maintain inventories at the retailers above certain threshold levels. The central warehouse (distribution center) supplies the retailers, which in turn replenishes its inventory according to a policy from an outside supplier with unlimited capacity.

Initially, [33] considered a depot-base system for repairable items where demand for items follow compound Poisson processes at the bases. An analytical solution was

DOI: 10.14736/kyb-2017-3-0461

given to determine the optimal base-stock levels for each item subject to a limited system investment. Later, [28] investigated the identical system where the replenishment is made in batches. They provided a power approximation method to determine the optimal batch sizes and safety stocks. [17] and [32] studied the system where each facility follows a continuous review (R, Q) policy and the identical retailers face stationary Poisson demand. An approximate model was presented to calculate the system service levels in [17], and an optimization framework was developed to maximize the system fill-rate subject to a safety stock constraint in [32]. The distribution inventory system with one-for-one replenishment was explored in [3], and a periodic review control policy was used in [14]. [21], on the other hand, explored non-identical retailers under an (R,Q) policy. [16] also analyzed non-identical retailers operating under an (s, S) policy.

In the aforementioned studies regarding multi-echelon distribution networks, the main idea has been to decompose the system into smaller subsystems (i.e., decompose the system to a warehouse and retailers with their own procurement and demand arrival processes). Effective demand inter-arrival times at the warehouse and effective lead-times at the retailers were identified. Then, procedures for the single-location models were used to obtain desired performance measures.

[3, 4, 5, 15, 34], and [36] considered the multi-echelon distribution inventory system with divergence in their solution methodologies. [34], and [3, 4, 5] exploited solution methodologies based on the approach to match every supply unit with a demand unit. In other words, they kept track of each supply unit and its sojourn time in the system and calculated the holding and backorder costs accordingly. [15], on the other hand, disaggregated the backorders at the warehouse among the retailers and then computed the long-run inventory levels. [36] calculated the probability rules of the waiting times observed by retailers' replenishment orders in the warehouse.

A common assumption of the preceding studies related to the distribution inventory system was constant transportation times between the external supplier and the warehouse, as well as between the warehouse and the retailers. An exception to this was [35] where they assumed stochastic transit times under base-stock policies. Some reviews of the multi-echelon systems were [9, 18, 19], and [20].

In several settings, the arrival process is a superposition of different arrival streams. An example is a queue to which the arrival process is the superposition of separate arrival streams, each of whose inter-arrival times is of Erlang distribution. Practical applications include production line with input and output layers, that is, input to downstream machines is the output of upstream stations [37]; pooled production-inventory systems [10]; single server queues with Markovian arrival processes [27]; and multiservice network using ATM multiplexer [30], among others. An important characteristic of the superposed process is that although the individual streams are independent from each other, the inter-arrival times of the superposed process may no longer be independent. Additionally, exact analysis of the superposed process becomes computationally impractical as the number of the superposed streams increases. As a result, most of the work in this area delves into approximations. Typical methods approximate the superposed processes by renewal processes, which may be inadequate in capturing the temporal dependence (i.e., the autocorrelation).

[1] developed a hybrid approximation scheme that combines the stationary-interval

method and the asymptotic method of [38]. Both methods determine the approximating renewal process by identifying moments for the intervals between successive points and fitting a convenient distribution to the moments. [11] developed an approximation using Super-Erlang chains, which takes into account the local and long-term behavior of the second-order measures of the nonrenewal process being approximated. [12] analyzed a queue using the Super-Erlang chains in which the arrival process is the superposition of separate arrival streams, each of whose inter-arrival and service time distributions are of phase-type.

The above approximation methods were based on first order and second order statistics. However, [22] developed higher order approximations for the single server queue with general inter-arrival and service time distributions. Similarly, [8] used a three parameter renewal approximation in predicting the mean waiting time in a queue with deterministic service times. [37], on the other hand, proposed an approximation method based on state-space aggregation.

In this study, we develop a model based on decomposition approximations to study the distribution inventory system under stochastic transportation times. Our aim is to analyze system behavior using some key performance metrics such as the time averages of inventory and backorder levels, and the customer service levels. The literature for multi-echelon systems under stochastic transportation times is scarce and needs further attention, as opposed to the abundant literature under constant lead-times. In addition, we present a technique to characterize the demand arrival process at the warehouse as a superposition of inter-arrival times of Erlang distributions. We illustrate the practical use of this approach in the two-echelon distribution inventory system. The developed approximations are validated against the simulation, yielding good agreement of robust performance metrics.

The rest of the paper is organized as follows. In Section 2, we describe the multi-echelon distribution inventory system. In Section 3, we illustrate the modelling approach decomposing the system into smaller subsystems. Section 4 presents our methodology to analyze the demand arrival process at the warehouse. Section 5 includes the steady-state analysis of subsystems and Section 6 includes the numerical results. Finally, Section 7 concludes the paper.

2. MULTI-ECHELON DISTRIBUTION INVENTORY SYSTEMS

We consider a distribution inventory system comprising a single warehouse (W) and N retailers, as shown in Figure 1. The retailers face independent, stationary unit Poisson demand and have their own operating characteristics. They follow continuous review (R_i, Q_i) inventory control policies, (i.e., when the inventory position, inventory on hand plus outstanding orders minus backorders, at retailer i down-crosses its reorder point R_i, it orders a replenishment batch size of Q_i from the central warehouse). (R, Q) policies are appropriate under certain settings. First, they are used under continuous review (tightly controlled) and for slow moving items. Second, they are used when the ordering cost is high. The order arrives after a lead-time delay (including just the transportation time), if the warehouse has sufficient inventory on hand. Otherwise, it experiences additional delays due to stockouts at the warehouse. We assume that it is possible to have several outstanding orders from a retailer at any point in time. Any excess demand at a retailer is backlogged and filled as soon as the replenishment order arrives from the warehouse, in a first-in first-out manner.

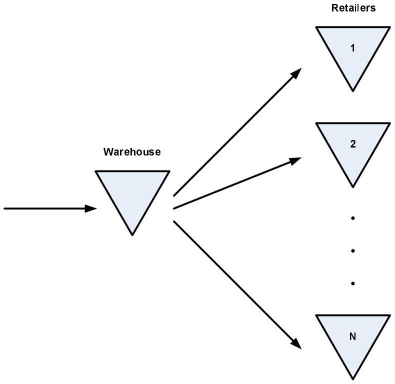


Fig. 1. A two-echelon distribution inventory system.

Demand at the warehouse includes replenishment orders from the retailers and is satisfied immediately provided that the warehouse has available stock on hand. The unsatisfied demand is backordered. The warehouse, in turn, orders from an outside supplier who is assumed to have sufficient inventory on hand at any point in time, based on an (R_W, Q_W) inventory control policy. Hence, the effective lead-time includes only the transportation time.

We assume that all replenishment batch quantities are multiples of a batch size q for convenience. Therefore, the inventory position at the warehouse is always a multiple of q and partial shipments are not allowed. In addition, we assume all transportation times between facilities are phase-type distributed because of their generality and versatility. Phase-type random variables include exponential random variables, their finite sums and mixtures, and certainly can approximate any distribution with a wide-range of variability at a desired accuracy level ([2, 29, 35]). We assume, in particular, all transportation times follow a k-phase Erlang (Erlang-k) distribution. Moreover, we assume the orders are processed sequentially in the transportation system. That is, no overtaking is possible and the orders are received in the same order they were placed. Assuming independent, identically distributed random variables, on the other hand, results in parallel processing of orders and lets orders to cross in time. [35] modelled the transportation times accordingly. See also [39] for the use of phase-type distributions in inventory-control models.

3. MODELLING APPROACH

It is plausible that the entire system can be modelled using a Markovian approach by keeping track of the inventory levels, demand arrival processes, and replenishment processes simultaneously at the warehouse and retailers (due to the Poisson demand arrivals and phase-type transportation times). However, as the system size increases, the use of the exact (Markovian) methods becomes computationally impractical due to the fast growing state-space of the underlying Markov chain. Indeed, we next present a decomposition procedure, which uses single-location models as building blocks to analyze the entire distribution inventory system.

Note that, the system with one-for-one (i.e., $Q = 1$) replenishment policies is a special case and easily analyzed, since the demand process at the warehouse is a superposition of N independent Poisson processes and still a Poisson process. On the other hand, the distribution system with (R_i, Q_i) inventory control policies at the retailers is quite difficult to analyze since the demand process at the warehouse is a superposition of N independent Erlang distributions.

Here, we propose an approximation approach that decomposes the system into two sets of subsystems. Each subsystem consists of an inventory holding buffer with its own stock keeping policy, replenishment and demand arrival processes. We treat each subsystem as a single-location model requiring modest computational effort. Finally, we link the subsystems to each other. The decomposition method adopted is based on [2, 23, 25]. Let us introduce the following notation:

λ_i : demand rate at retailer i, $i = 1, 2, \ldots, N$,
q largest common factor of $Q_W, Q_1, Q_2, \ldots, Q_N$,
TT_W : transportation time between supplier and warehouse,
TT_i : transportation time between warehouse and retailer i, $i = 1, 2, \ldots, N$,
$\Omega(W)$: subsystem involving warehouse,
$\Omega(i)$: subsystem involving retailer i, $i = 1, 2, \ldots, N$,
$M_j^{'}$: node modelling replenishment to facility j, $j = W, 1, 2, \ldots, N$,
$M_j^{''}$: node modelling demand arrival process to facility j, $j = W, 1, 2, \ldots, N$,
I_j : inventory level in $\Omega(j)$, $j = W, 1, 2, \ldots, N$.

The principles of decomposition are illustrated in Figure 2. The first subsystem, $\Omega(W)$, includes the warehouse, which uses an (R_W, Q_W) inventory control policy. Node $M_W^{'}$ models the effective replenishment process and node $M_W^{''}$ models the effective demand arrival process for the warehouse. Similarly, the subsystems, $\Omega(i)$, include retailer i, $i = 1, 2, \ldots, N$. An (R_i, Q_i) policy is used to control inventory level. Node $M_i^{'}$ represents the effective replenishment process and node $M_i^{''}$ represents the demand arrival process at retailer i, respectively.

3.1. Replenishment times

For subsystem $\Omega(W)$, the variable $U_W^{'}$ represents the effective replenishment time at the warehouse. Since the supplier has always sufficient raw material to replenish the warehouse, the effective replenishment time consists only of the transportation lead-time from supplier to the warehouse. That is,

$$U_W^{'} = TT_W.$$

For subsystems $\Omega(i)$, the variable $U_i^{'}$ represents the effective replenishment time at retailer i. The retailer order is filled as soon as it is received, if the warehouse has sufficient stock on hand. Otherwise, it is delayed until sufficient number of units arrive

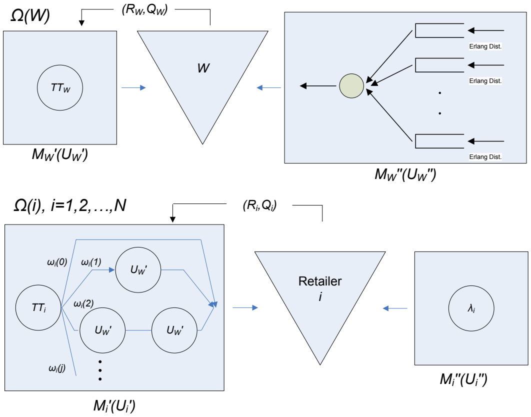


Fig. 2. Subsystems $\Omega(W)$ and $\Omega(i)$, $i = 1, 2, \ldots, N$.

at the warehouse. Let us assume $\omega_i(0)$ be the conditional probability that there are zero units missing in the warehouse (that is, the warehouse has sufficient stock on hand) when a replenishment request arrives from retailer i. Similarly, let us assume $\omega_i(j)$, $j = 1, 2, \ldots$ be the conditional probability that there are j batches missing, that is, for any j, $(j-1) * Q_W + 1, (j-1) * Q_W + 2, \ldots, j * Q_W$ units missing in the warehouse when a replenishment is requested from retailer i. Then, the effective lead time to retailer i is given by:

$$U_i^{'} = \begin{cases} TT_i & \text{w.p.} \quad \omega_i(0), \\ TT_i + j \times U_W^{'} & \text{w.p.} \quad \omega_i(j). \end{cases}$$

It is clear that, with probability $\omega_i(0)$, there is sufficient stock at the warehouse and retailer i's order experiences no additional delays. On the other hand, with probability $\omega_i(j)$, the warehouse misses j batches resulting in a delay in the replenishment process. This delay, however, is j replenishment lead times from the supplier to the warehouse.

3.2. Demand inter-arrival times

The retailers face customer demand according to a Poisson process with rate λ_i, $i = 1, 2, \ldots, N$. Equivalently, the effective demand inter-arrival times at retailer i are independent and follow an exponential distribution with rate λ_i, for $i = 1, 2, \ldots, N$.

Demand at the warehouse includes replenishment orders from the retailers. Since the retailers face independent, stationary Poisson demand and replenish their stock according to an (R_i, Q_i) policy, the inter-arrival times of the orders from the retailers follow Erlang distributions. As a result, the demand arrival process at the warehouse is a superposition of N independent Erlang distributions.

4. SUPERPOSITION OF ERLANG PROCESSES

An important characteristic of the superposed Erlang processes is that although the individual processes are independent from each other, the inter-arrival times of the superposed process may not be independent. Here, we present a methodology to characterize such arrival as a Markovian process (by keeping track of the individual arrival streams' phase structure and their transition behavior). Note that the state-space of the superposed arrival process increases considerably. Therefore, we also suggest a three-moment approximation scheme to efficiently use the methodology in practice. We illustrate the accuracy of the methodology on a number of test problems.

4.1. Preliminaries

A k-phase Erlang (Erlang-k) distribution is the sum of k exponential random variables. A phase diagram of the Erlang-k distribution with rate λ is shown in Figure 3. The Erlang-k distribution has also the following (α, T) representation:

$$\alpha^T = (1, 0, \ldots, 0), \quad T = \begin{bmatrix} -\lambda & \lambda & & & \\ & -\lambda & \lambda & & \\ & & -\lambda & \ddots & \\ & & & \ddots & \lambda \\ & & & & -\lambda \end{bmatrix}_{k \times k}.$$

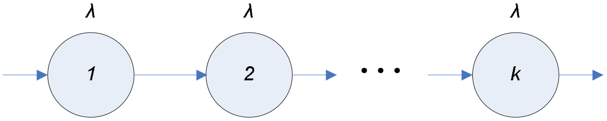


Fig. 3. Phase diagram of the Erlang-k distribution.

An important property of the Erlang-k distribution is that the residual (remaining) time has a mixture of generalized Erlang-k (*MGE-k*) distribution. This is due to the following arguments. At any point in time, the Erlang-k distribution, with probability $1/k$, is in any one of its exponential phases. Hence, the residual time has one exponential phase with probability $1/k$, the residual time has two exponential phases with probability $1/k$, and so on. The resulting *MGE-k* distribution has a graphical representation shown in Figure 4 with corresponding probabilities. The *MGE-k* distribution has also the following (α, T^*) representation [2]:

$$\alpha^T = (1, 0, \ldots, 0), \quad T^* = \begin{bmatrix} -\lambda & \frac{k-1}{k}\lambda & & & \\ & -\lambda & \frac{k-2}{k-1}\lambda & & \\ & & -\lambda & \ddots & \\ & & & \ddots & \frac{1}{2}\lambda \\ & & & & -\lambda \end{bmatrix}_{k \times k}.$$

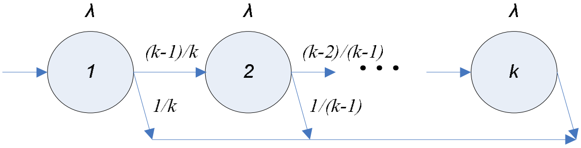


Fig. 4. Phase diagram of the remaining time of an Erlang-k distribution.

We borrow the following Definition 4.1 and Theorems (4.2, 4.3) from [29].

Definition 4.1. If L and M are rectangular matrices of dimensions $k_1 \times k_2$ and $k_1^{'} \times k_2^{'}$, their Kronecker product $L \otimes M$ is the matrix of dimensions $k_1 k_1^{'} \times k_2 k_2^{'}$, written in block-partitioned form as

$$\begin{bmatrix} L_{11}M & L_{12}M & \dots & L_{1k_2}M \\ \vdots & \vdots & & \vdots \\ L_{k_1 1}M & L_{k_1 2}M & \dots & L_{k_1 k_2}M \end{bmatrix}.$$

If X and Y are independent random variables with phase-type distributions $F(\cdot)$ and $G(\cdot)$, then the distribution $H(\cdot) = 1 - [1 - F(\cdot)][1 - G(\cdot)]$, corresponding to $\min(X, Y)$, is also phase-type.

Theorem 4.2. Let $F(\cdot)$ and $G(\cdot)$ have representations (α, T) and (β, S) of orders m and n respectively, then $H(\cdot)$, corresponding to $\min(X, Y)$, has the representation $[\alpha \otimes \beta, T \otimes I + I \otimes S]$.

Theorem 4.3. A finite mixture of phase-type distributions is a phase-type distribution. If $(p_1, \dots, p_k)$ is the mixing distribution and $F_j(\cdot)$ has representation $[\alpha(j), T(j)]$, $1 \leq j \leq k$, then the mixture has the representation $\alpha = [p_1\alpha(1), \dots, p_k\alpha(k)]$, and

$$T = \begin{bmatrix} T(1) & 0 & \dots & 0 \\ 0 & T(2) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & T(k) \end{bmatrix}.$$

4.2. Superposition of two erlang processes

We start from the simplest case; an arrival process that is a superposition of two independent Erlang processes. Let us denote by $F(\cdot)$ and $G(\cdot)$ two Erlang distributions with representations (α, T) and (β, S) of orders m and n, respectively.

Consider the superposed process at an arrival instance, (i.e., an instance at which an arrival just happened). Without loss of generality, let us assume that the arrival is from the first process. The amount of time until the next arrival from the first process follows an Erlang-m distribution. On the other hand, the amount of time until the next arrival from the second process follows an *MGE-n* distribution, since the remaining time of an Erlang-n distribution is an *MGE-n* distribution. In fact, the amount of time until the next arrival is distributed as the minimum of the Erlang-m and *MGE-n* distributions. From Theorem 4.2, this distribution has a representation $[\alpha \otimes \beta, T \otimes I + I \otimes S^*]$ where (β, S^*) is the corresponding representation of the *MGE-n* distribution. In a similar vein, let us assume that the arrival is from the second process. The amount of time until the next arrival is distributed as the minimum of *MGE-m* and Erlang-n distributions, and has a representation $[\alpha \otimes \beta, T^* \otimes I + I \otimes S]$ where (α, T^*) is the corresponding representation of *MGE-m* distribution.

Let us denote by $p(1)$ the probability that the arrival instance is from the first stream, and by $p(2)$ that it is from the second stream. Then, the superposed process is going to be a mixture of phase-type distributions. By Theorem 4.3, it is again a phase-type distribution with the corresponding representation, $\alpha = [p(1)(\alpha \otimes \beta), p(2)(\alpha \otimes \beta)]$, and

$$T = \left[\begin{array}{cc} T \otimes I + I \otimes S^* & \underline{0} \\ \underline{0} & T^* \otimes I + I \otimes S \end{array} \right].$$

4.3. Superposition of N Erlang Processes

Next, we generalize the methodology presented in the previous section to N independent Erlang processes. We first present a Corollary that follows from Theorem 4.2 to accommodate N phase-type distributions. If X_1, X_2, …, X_N are independent random variables with phase-type distributions $F_1(\cdot)$, $F_2(\cdot)$, …, $F_N(\cdot)$, then the distribution $H(\cdot) = 1 - [1 - F_1(\cdot)][1 - F_2(\cdot)] \ldots [1 - F_N(\cdot)]$, corresponding to $\min(X_1, X_2 \ldots X_N)$, is also phase-type.

Corollary 4.4. Let $F_1(\cdot)$, $F_2(\cdot)$, …, $F_N(\cdot)$ have representations (α_1, T_1), (α_2, T_2), …, (α_N, T_N) of orders n_1, n_2, …, n_N, respectively. Then, $H(\cdot)$ has the representation $[\alpha_1 \otimes \alpha_2 \otimes \ldots \otimes \alpha_N, T_1 \otimes I_2 \otimes \ldots \otimes I_N + I_1 \otimes T_2 \otimes I_3 \otimes \ldots \otimes I_N + \ldots + I_1 \otimes I_2 \otimes \ldots \otimes I_{N-1} \otimes T_N]$.

Now, consider the superposed process at an arrival instance. Let us assume that the arrival is from the first process. The amount of time until the next arrival from the first process follows an Erlang-n_1 distribution. On the other hand, the amount of time until the next arrival from the second process follows an *MGE-*n_2 distribution, the amount of time until the next arrival from the third process follows an *MGE-*n_3 distribution, and so on. In fact, the amount of time until the next arrival is distributed as the minimum of Erlang-n_1, *MGE-*n_2, …, *MGE-*n_N distributions. The superposed distribution is defined by Corollary 4.4 and has the representation $[\alpha(1), T(1)] = [\alpha_1 \otimes \alpha_2 \otimes \ldots \otimes \alpha_N, T_1 \otimes I_2 \otimes \ldots \otimes I_N + I_1 \otimes T_2^* \otimes I_3 \otimes \ldots \otimes I_N + \ldots + I_1 \otimes I_2 \otimes \ldots \otimes I_{N-1} \otimes T_N^*]$ where (α_i, T_i^*) is the corresponding representation of the *MGE-*n_i distribution.

Similarly, if we assume that the arrival is from the second process, the amount of time until the next arrival is distributed as the minimum of *MGE-*n_1, Erlang-n_2, …, *MGE-*n_N distributions. The distribution is defined by Corollary 4.4 and has the representation $[\alpha(2), T(2)] = [\alpha_1 \otimes \alpha_2 \otimes \ldots \otimes \alpha_N, T_1^* \otimes I_2 \otimes \ldots \otimes I_N + I_1 \otimes T_2 \otimes I_3 \otimes \ldots \otimes I_N + \ldots + I_1 \otimes I_2 \otimes \ldots \otimes I_{N-1} \otimes T_N^*]$.

At this point, it is clear that the superposed process is going to be a mixture of phase-type distributions. By Theorem 4.3, it is again a phase-type distribution with the corresponding $[\alpha, T]$ representation, $\alpha = [p(1)\alpha(1) \otimes p(2)\alpha(2) \otimes \ldots \otimes p(N)\alpha(N)]$, and

$$T = \begin{bmatrix} T(1) & 0 & \ldots & 0 \\ 0 & T(2) & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & T(N) \end{bmatrix}.$$

Although the above methodology exactly characterizes the superposed process, it has limited practical use because of the fast growing state-space.

4.4. Approximating the superposition process

The idea of the approximation procedure requires superposing individual arrival steams iteratively, avoiding the state-space getting larger. Initially, we superpose two individual arrival streams and approximate the resulting stream by using a three-moment approximation. Then, we superpose the resulting arrival stream with the next arrival stream, and again use the three-moment scheme to approximate the resulting process. We continue in a similar manner until all the arrival streams are exhausted. Thus, we prevent the state-space getting larger at the expense of losing limited degree of accuracy. Here, we facilitate the three-moment approximation due to [31].

The three-moment approximation in [31] utilizes Erlang-Coxian (EC) distributions and its variants as shown in Figure 5. The EC distribution is simply an *MGE*-2 distribution appended to a generalized Erlang distribution. It also allows positive probability for mass at point zero. EC distribution has six parameters being estimated. A closed-form solution is derived in [31]. Empirical studies suggest that using three-moment approximation captures the skewness of the distribution and potentially brings an adequate degree of accuracy.

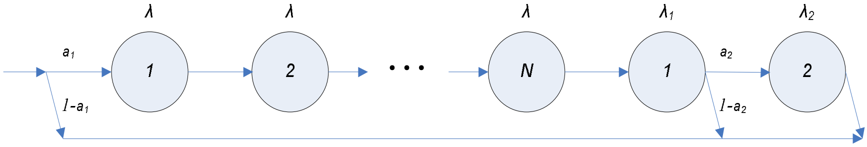


Fig. 5. Phase diagram of the Erlang-Coxian distribution.

Example 4.5. Consider a distribution inventory system comprising a single warehouse and three identical retailers. The retailers face Poisson demand with rate one and follow continuous review $(R, Q) = (5, 10)$ inventory control policies. Hence, the demand arrival process at the warehouse is a superposition of three Erlang-10 distributions with rate one. The final process using the superposition technique along with the three-moment approximation results in a first moment, $E[X] = 3.3334$, a second moment,

$E[X^2] = 16.8486$, a third moment, $E[X^3] = 102.9705$, and a squared coefficient of variation, $Cv^2 = 0.5163$.

If we directly employ the methodology for the superposition of the three arrival streams (without approximation), we get a first moment of $E[X] = 3.3333$, a second moment of $E[X^2] = 16.8363$, a third moment of $E[X^3] = 103.8908$, and a squared coefficient of variation, $Cv^2 = 0.5153$. A comparison of the approximation technique to the exact methodology is given in Table 1. As expected using the three-moment approximation results in an acceptable level of accuracy.

	Three-moment Approximation	Exact Methodology
$E[X]$	3.3334	3.3333
$E[X^2]$	16.8486	16.8363
$E[X^3]$	102.9705	103.8908
Cv^2	0.5163	0.5153

Tab. 1. Accuracy of the three-moment approximation method.

5. STEADY-STATE ANALYSIS OF THE SUBSYSTEMS

Each of the subsystems, $\Omega(j)$, $j = W, 1, 2, \ldots, N$ is a two-node subsystem with its own stock keeping policy, and replenishment and demand inter-arrival times of phase-type. Let us consider $\Omega(W)$. The triple $\{I_t, J_t, N_t, t \geq 0\}$ is a Markov chain where I_t represents the phase of $U^{'}_W$, J_t represents the phase of $U^{''}_W$, and N_t denotes the number of inventories in the warehouse. The essence of the phase-type random variables gives rise to a Markovian analysis and matrix-recursive procedures based on [13, 29] are used to obtain the steady-state probabilities. We assume all transportation times to follow a 2-phase Erlang distribution (Erlang-2). We avoid using higher phase Erlang distributions since they approach to a deterministic variable in the number of phases. We include the detailed analysis of subsystems in [24, 26].

5.1. An aggregation algorithm

The nature of the decomposition algorithm requires that subsystems supply data to each other. The required data includes the warehouse demand representation and $\omega_i(j)$, $j = 0, 1, 2, \ldots$ for $i = 1, 2, \ldots, N$. In the aggregation algorithm, the warehouse demand representation is used in the analysis of $\Omega(W)$. Similarly, $\omega_i(j)$'s, $j = 0, 1, 2, \ldots$ are used in the analysis of $\Omega(i)$ for $i = 1, 2, \ldots, N$.

Here, the warehouse demand representation is obtained using the superposition technique described in Section 4, and $\omega_i(j)$'s, $j = 0, 1, 2, \ldots$ for $i = 1, 2, \ldots, N$ are evaluated as

$$\begin{aligned} \omega_i(0) &= Pr(I_W \geq Q_i | I_i = R_i), \\ \omega_i(j) &= \sum_{k=(j-1)Q_W+1}^{jQ_W} Pr(I_W = Q_i - k | I_i = R_i), \; j = 1, 2, \ldots . \end{aligned}$$

where I_W and I_i represent the inventory level at the warehouse and the retailers for $i = 1, 2, \ldots, N$, respectively. Note that the $\omega_i(j)$'s are arrival-point probabilities. In this setting, we use arbitrary time probabilities as a surrogate for the arrival-rate probabilities.

A summary of the algorithm is given in Table 2.

1. Initialize: Characterize the warehouse demand.
2. Analyze $\Omega(W)$, obtain its steady-state probabilities, compute $\omega_i(j),\ j = 0, 1, 2, \ldots,$ for $i = 1, 2, \ldots, N$.
3. Analyze $\Omega(i)$, obtain its steady-state probabilities, for $i = 1, 2, \ldots, N$.
4. Obtain average inventory and backorder levels, customer service level at retailer i, for $i = 1, 2, \ldots, N$.

Tab. 2. The aggregation algorithm for multi-echelon distribution inventory system.

6. NUMERICAL RESULTS

We test the accuracy of our aggregation algorithm by comparing its results against the simulation in a number of examples. The approximation procedure described above and the discrete-event simulation model runs are implemented on a Core i7 PC operating at 2.20 GHz. The simulation model is developed using the Arena[1] simulation software. Each simulation run includes 50,000,000 departures to provide point estimates and 95% confidence intervals for key performance metrics.

In this study, we focus on the average inventory levels (Inv.), average backorder levels (BO), and customer service levels (C.S.L.). Here, we define the C.S.L. as the probability of fully satisfying the demand of an arriving customer. We also compute the relative error (Rel. Error) as the difference between the aggregation algorithm and simulation results divided by simulation results.

Illustrative approximation and the simulation results are given in Tables 3 − 6 for different settings. We have three major experimental settings: a serial system in Table 3; a system including a single warehouse and three retailers in Table 4; and a system including a single warehouse and five retailers in Tables 5 and 6. We further assume identical retailers in Table 4, and non-identical retailers in Tables 5 and 6. In most of the settings, demand rate, λ, is varied while keeping other parameters constant.

In the serial setting, in Table 3, the retailer follows a continuous review $(R_1, Q_1) = (5, 10)$ inventory control policy and the warehouse follows a continuous review $(R_W, Q_W) = (10, 20)$ inventory control policy. The transportation times from the supplier to the warehouse and from the warehouse to the retailer follow 2-phase Erlang distributions (Erlang-2) with rate 1. The initial demand rate is $\lambda = 0.5$, which is incremented by 0.5 in the following experiments.

The results demonstrate that the relative error for the performance estimates varies from -4 % to 0% for the average inventory levels (Inv.), 0 % to 10 % for the backorder levels (BO), and -5 % to 0 % for the customer service levels (C.S.L.). In this particular

[1] Arena is a trademark of Rockwell Automation.

	W	Ret. 1
	$R_W = 10$	$R_1 = 5$
Parameters:	$Q_W = 20$	$Q_1 = 10$
	$\beta_S = 1$	$\beta_W = 1$

		λ=0.5			λ=1.0		
		Inv.	BO	C.S.L.	Inv.	BO	C.S.L.
W	Analytic	24.0	0.0	100%	23.0	0.0	100%
	Sim.	24.0	0.0	100%	23.0	0.0	100%
	Rel. Error	0%	N/A	0%	0%	N/A	0%
Ret. 1	Analytic	9.5	0.0	100%	8.5	0.0	99%
	Sim.	9.5	0.0	100%	8.5	0.0	99%
	Rel. Error	0%	N/A	0%	0%	N/A	0%

		λ=1.50			λ=2.0		
		Inv.	BO	C.S.L.	Inv.	BO	C.S.L.
W	Analytic	22.0	0.0	100%	21.0	0.0	100%
	Sim.	22.0	0.0	100%	21.0	0.0	100%
	Rel. Error	0%	N/A	0%	0%	N/A	0%
Ret. 1	Analytic	7.5	0.1	95%	6.6	0.3	89%
	Sim.	7.5	0.1	95%	6.6	0.3	89%
	Rel. Error	0%	0%	0%	0%	0%	0%

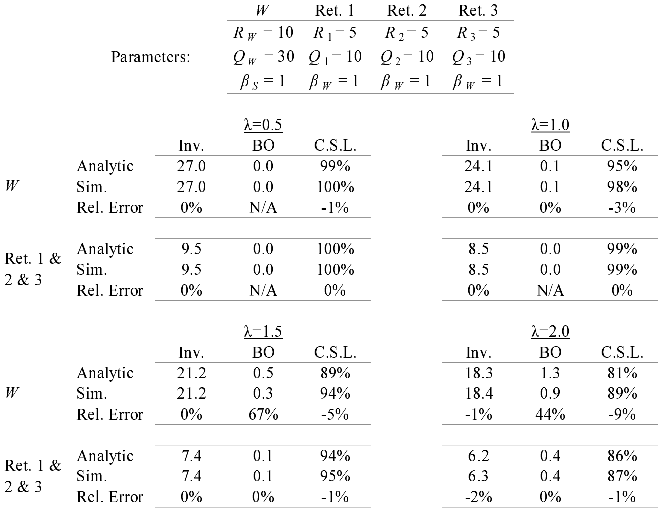
		λ=2.50			λ=3.0		
		Inv.	BO	C.S.L.	Inv.	BO	C.S.L.
W	Analytic	20.0	0.0	100%	19.0	0.0	99%
	Sim.	20.0	0.0	100%	19.0	0.0	99%
	Rel. Error	0%	N/A	0%	0%	N/A	0%
Ret. 1	Analytic	5.6	0.7	81%	4.5	1.6	70%
	Sim.	5.6	0.7	81%	4.5	1.6	70%
	Rel. Error	0%	0%	0%	0%	0%	0%

		λ=3.50			λ=4.00		
		Inv.	BO	C.S.L.	Inv.	BO	C.S.L.
W	Analytic	17.9	0.1	98%	16.9	0.2	96%
	Sim.	17.9	0.1	98%	16.9	0.2	96%
	Rel. Error	0%	0%	0%	0%	0%	0%
Ret. 1	Analytic	3.4	3.5	55%	2.2	8.7	38%
	Sim.	3.4	3.5	56%	2.3	7.9	40%
	Rel. Error	0%	0%	-2%	-4%	10%	-5%

Tab. 3. Accuracy of the approximation algorithm for the case of 1 Warehouse and 1 Retailer.

setting, the proposed approximation algorithm yields robust results as compared to simulation requiring modest computational effort. In addition, it is seen from the results that the relative error gradually increases with the demand rate (system load), which is also expected. In the particular instance where $\lambda = 4.0$, the C.S.L. is around 40% which is not common in practice. Even in this setting, the metrics Inv., BO and C.S.L. are being approximated to a reasonable error margin. The error margin for BO Level, though, is higher due to the insufficient capturing of low tail probabilities.

In the system with one warehouse and three identical retailers, in Table 4, the relative error for the system estimates varies from -2 % to 0 % for the average inventory levels, 44 % to 67 % for the backorder levels, and -3% to 0% for the customer service levels. Again, the percentage deviation gradually increases with the demand rate (system load).



	W	Ret. 1	Ret. 2	Ret. 3
	$R_W = 10$	$R_1 = 5$	$R_2 = 5$	$R_3 = 5$
Parameters:	$Q_W = 30$	$Q_1 = 10$	$Q_2 = 10$	$Q_3 = 10$
	$\beta_S = 1$	$\beta_W = 1$	$\beta_W = 1$	$\beta_W = 1$

		λ=0.5			λ=1.0		
		Inv.	BO	C.S.L.	Inv.	BO	C.S.L.
W	Analytic	27.0	0.0	99%	24.1	0.1	95%
	Sim.	27.0	0.0	100%	24.1	0.1	98%
	Rel. Error	0%	N/A	-1%	0%	0%	-3%
Ret. 1 & 2 & 3	Analytic	9.5	0.0	100%	8.5	0.0	99%
	Sim.	9.5	0.0	100%	8.5	0.0	99%
	Rel. Error	0%	N/A	0%	0%	N/A	0%

		λ=1.5			λ=2.0		
		Inv.	BO	C.S.L.	Inv.	BO	C.S.L.
W	Analytic	21.2	0.5	89%	18.3	1.3	81%
	Sim.	21.2	0.3	94%	18.4	0.9	89%
	Rel. Error	0%	67%	-5%	-1%	44%	-9%
Ret. 1 & 2 & 3	Analytic	7.4	0.1	94%	6.2	0.4	86%
	Sim.	7.4	0.1	95%	6.3	0.4	87%
	Rel. Error	0%	0%	-1%	-2%	0%	-1%

Tab. 4. Accuracy of the approximation algorithm for the case of 1 Warehouse and 3 identical Retailers.

Here, the accuracy in the backorder levels in the warehouse is somehow surprising. This is because backorder levels are low and approximating small probabilities (low tail) does not seem to be quite successful. Backorder levels in retailers, though, are highly accurate. Other tables can be interpreted accordingly.

In order to delve into the less accurate backorder levels at the warehouse, we investigate the autocorrelation structure of the superposed demand arrival process at the warehouse. The magnitude of autocorrelation of the demand arrival process at the warehouse decreases as there are more channels to send replenishment orders. In fact, we observe the highest negative autocorrelation at the superposed process with two identical retailers. The negative lag-1 autocorrelation decreases as the number of superposed processes increases. In fact, as there are more channels to send orders, the superposed process will converge to a renewal process. To give an idea of the magnitude of the lag-1 autocorrelation, we consider a system with a single warehouse and two identical retailers, and a system with a single warehouse and three identical retailers, both with Erlang-10 distributions of rate one. The lag-1 autocorrelation is -0.5901 in the two retailers system, while it is -0.3972 in the three retailers system. The lag-1 autocorrelation is expected to converge to zero as the number of superposed processes increases. In a similar vein, a system with non-identical retailers, in general, has lower negative lag-1 autocorrelation when compared to a system with identical retailers. This, in addition, explains the less accurate backorder levels at the warehouse. Since we replace a process with significant level of lag-1 autocorrelation with a renewal process, it results in less accuracy in low tail probabilities.

As can be seen from the results, using arbitrary time probabilities as a surrogate of

		λ=1.0				λ=1.0			
		Parameters	Inv.	BO	C.S.L.	Parameters	Inv.	BO	C.S.L.
	Analytic	$R_W = 10$	18.0	1.1	83%	$R_W = 10$	18.0	1.3	79%
W	Sim.	$Q_W = 30$	18.0	0.9	87%	$Q_W = 30$	18.1	1.1	83%
	Rel. Error	$\beta_S = 1$	0%	22%	-5%	$\beta_S = 1$	-1%	18%	-5%
	Analytic	$R_1 = 5$	8.3	0.0	98%	$R_1 = 5$	8.3	0.0	98%
Ret. 1	Sim.	$Q_1 = 10$	8.3	0.0	98%	$Q_1 = 10$	8.3	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	N/A	0%	$\beta_W = 1$	0%	N/A	0%
	Analytic	$R_2 = 5$	5.7	0.1	95%	$R_2 = 5$	5.7	0.1	95%
Ret. 2	Sim.	$Q_2 = 5$	5.7	0.1	95%	$Q_2 = 5$	5.7	0.1	95%
	Rel. Error	$\beta_W = 1$	0%	0%	0%	$\beta_W = 1$	0%	0%	0%
	Analytic	$R_3 = 5$	10.6	0.0	98%	$R_3 = 5$	10.6	0.0	98%
Ret. 3	Sim.	$Q_3 = 15$	10.6	0.0	98%	$Q_3 = 15$	10.6	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	N/A	0%	$\beta_W = 1$	0%	N/A	0%
	Analytic	$R_4 = 5$	5.7	0.1	95%	$R_4 = 5$	8.3	0.0	98%
Ret. 4	Sim.	$Q_4 = 5$	5.7	0.1	95%	$Q_4 = 10$	8.3	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	0%	0%	$\beta_W = 1$	0%	N/A	0%
	Analytic	$R_5 = 5$	5.7	0.1	95%	$R_5 = 5$	8.3	0.0	98%
Ret. 5	Sim.	$Q_5 = 5$	5.7	0.1	95%	$Q_5 = 10$	8.3	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	0%	0%	$\beta_W = 1$	0%	N/A	0%

		λ=1.0				λ=1.0			
		Parameters	Inv.	BO	C.S.L.	Parameters	Inv.	BO	C.S.L.
	Analytic	$R_W = 10$	18.1	1.5	75%	$R_W = 10$	18.0	1.2	81%
W	Sim.	$Q_W = 30$	18.1	1.3	80%	$Q_W = 30$	18.0	1.0	85%
	Rel. Error	$\beta_S = 1$	0%	15%	-6%	$\beta_S = 1$	0%	20%	-5%
	Analytic	$R_1 = 5$	8.2	0.0	98%	$R_1 = 5$	8.3	0.0	98%
Ret. 1	Sim.	$Q_1 = 10$	8.2	0.0	98%	$Q_1 = 10$	8.3	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	N/A	0%	$\beta_W = 1$	0%	N/A	0%
	Analytic	$R_2 = 5$	5.6	0.1	95%	$R_2 = 5$	5.7	0.1	95%
Ret. 2	Sim.	$Q_2 = 5$	5.6	0.1	95%	$Q_2 = 5$	5.7	0.1	95%
	Rel. Error	$\beta_W = 1$	0%	0%	0%	$\beta_W = 1$	0%	0%	0%
	Analytic	$R_3 = 5$	10.6	0.0	98%	$R_3 = 5$	10.6	0.0	98%
Ret. 3	Sim.	$Q_3 = 15$	10.6	0.0	98%	$Q_3 = 15$	10.6	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	N/A	0%	$\beta_W = 1$	0%	N/A	0%
	Analytic	$R_4 = 5$	10.6	0.0	98%	$R_4 = 5$	5.7	0.1	95%
Ret. 4	Sim.	$Q_4 = 15$	10.6	0.0	98%	$Q_4 = 5$	5.7	0.1	95%
	Rel. Error	$\beta_W = 1$	0%	N/A	0%	$\beta_W = 1$	0%	0%	0%
	Analytic	$R_5 = 5$	10.6	0.0	98%	$R_5 = 5$	8.3	0.0	98%
Ret. 5	Sim.	$Q_5 = 15$	10.6	0.0	98%	$Q_5 = 10$	8.3	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	N/A	0%	$\beta_W = 1$	0%	N/A	0%

Tab. 5. Accuracy of the approximation algorithm for the case of 1 Warehouse and 5 non-identical Retailers with fixed demand rate.

		λ=1.0				λ=1.0			
		Parameters	Inv.	BO	C.S.L.	Parameters	Inv.	BO	C.S.L.
	Analytic	$R_W = 10$	18.0	1.1	83%	$R_W = 10$	18.0	1.3	79%
W	Sim.	$Q_W = 30$	18.0	0.9	87%	$Q_W = 30$	18.1	1.1	83%
	Rel. Error	$\beta_S = 1$	0%	22%	-5%	$\beta_S = 1$	-1%	18%	-5%
	Analytic	$R_1 = 5$	8.3	0.0	98%	$R_1 = 5$	8.3	0.0	98%
Ret. 1	Sim.	$Q_1 = 10$	8.3	0.0	98%	$Q_1 = 10$	8.3	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	N/A	0%	$\beta_W = 1$	0%	N/A	0%
	Analytic	$R_2 = 5$	5.7	0.1	95%	$R_2 = 5$	5.7	0.1	95%
Ret. 2	Sim.	$Q_2 = 5$	5.7	0.1	95%	$Q_2 = 5$	5.7	0.1	95%
	Rel. Error	$\beta_W = 1$	0%	0%	0%	$\beta_W = 1$	0%	0%	0%
	Analytic	$R_3 = 5$	10.6	0.0	98%	$R_3 = 5$	10.6	0.0	98%
Ret. 3	Sim.	$Q_3 = 15$	10.6	0.0	98%	$Q_3 = 15$	10.6	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	N/A	0%	$\beta_W = 1$	0%	N/A	0%
	Analytic	$R_4 = 5$	5.7	0.1	95%	$R_4 = 5$	8.3	0.0	98%
Ret. 4	Sim.	$Q_4 = 5$	5.7	0.1	95%	$Q_4 = 10$	8.3	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	0%	0%	$\beta_W = 1$	0%	N/A	0%
	Analytic	$R_5 = 5$	5.7	0.1	95%	$R_5 = 5$	8.3	0.0	98%
Ret. 5	Sim.	$Q_5 = 5$	5.7	0.1	95%	$Q_5 = 10$	8.3	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	0%	0%	$\beta_W = 1$	0%	N/A	0%

		λ=1.0				λ=1.0			
		Parameters	Inv.	BO	C.S.L.	Parameters	Inv.	BO	C.S.L.
	Analytic	$R_W = 10$	18.1	1.5	75%	$R_W = 10$	18.0	1.2	81%
W	Sim.	$Q_W = 30$	18.1	1.3	80%	$Q_W = 30$	18.0	1.0	85%
	Rel. Error	$\beta_S = 1$	0%	15%	-6%	$\beta_S = 1$	0%	20%	-5%
	Analytic	$R_1 = 5$	8.2	0.0	98%	$R_1 = 5$	8.3	0.0	98%
Ret. 1	Sim.	$Q_1 = 10$	8.2	0.0	98%	$Q_1 = 10$	8.3	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	N/A	0%	$\beta_W = 1$	0%	N/A	0%
	Analytic	$R_2 = 5$	5.6	0.1	95%	$R_2 = 5$	5.7	0.1	95%
Ret. 2	Sim.	$Q_2 = 5$	5.6	0.1	95%	$Q_2 = 5$	5.7	0.1	95%
	Rel. Error	$\beta_W = 1$	0%	0%	0%	$\beta_W = 1$	0%	0%	0%
	Analytic	$R_3 = 5$	10.6	0.0	98%	$R_3 = 5$	10.6	0.0	98%
Ret. 3	Sim.	$Q_3 = 15$	10.6	0.0	98%	$Q_3 = 15$	10.6	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	N/A	0%	$\beta_W = 1$	0%	N/A	0%
	Analytic	$R_4 = 5$	10.6	0.0	98%	$R_4 = 5$	5.7	0.1	95%
Ret. 4	Sim.	$Q_4 = 15$	10.6	0.0	98%	$Q_4 = 5$	5.7	0.1	95%
	Rel. Error	$\beta_W = 1$	0%	N/A	0%	$\beta_W = 1$	0%	0%	0%
	Analytic	$R_5 = 5$	10.6	0.0	98%	$R_5 = 5$	8.3	0.0	98%
Ret. 5	Sim.	$Q_5 = 15$	10.6	0.0	98%	$Q_5 = 10$	8.3	0.0	98%
	Rel. Error	$\beta_W = 1$	0%	N/A	0%	$\beta_W = 1$	0%	N/A	0%

Tab. 6. Accuracy of the approximation algorithm for the case of 1 Warehouse and 5 non-identical Retailers with varying demand rate.

the arrival-point probabilities, however, does not impact the accuracy in the measures related to retailers (this information has been supplied by the warehouse to retailers). The relative errors for retailers, in general, are highly accurate.

Planning and operation of the studied distribution inventory system requires an optimization framework. Optimal configuration of the batch ordering policies specifies the amount of inventory to hold and move across the two-echelon system. A possible approach to handle the optimization scheme is to use a minimum-cost objective function. Such functions consider long-run averages of inventories and backorders, and assign cost penalties for both. Typical optimization frameworks are detailed in [6, 7].

7. CONCLUSION

In this study, we have considered a distribution inventory system consisting of a single warehouse and several retailers. We have developed a decomposition model that segregates the warehouse from the retailers. The challenge in this system has been elucidating the demand arrival process at the warehouse. We have proposed a procedure to analyze the demand arrival process at the warehouse as a superposition of independent Erlang processes. We have built in a moment matching method into the analysis of the superposed arrival process and showed its applicability in the distribution inventory setting. As a result, this has saved a great deal of computational effort and given rise to a computationally efficient way to solve the decomposed subsystems. The results have been highly accurate and acceptable, in view of the computational savings. Higher errors have been observed in small backorder probabilities, yet these errors are acceptable since they occur in low backorder levels.

(Received June 16, 2016)

REFERENCES

[1] S. L. Albin: Approximating a point process by a renewal process, II: Superposition arrival processes to queues. Oper. Res. *32* (1984), 5, 1133–1162. DOI:10.1287/opre.32.5.1133

[2] T. Altiok: Performance Analysis of Manufacturing Systems. Springer Series in Operations Research and Financial Engineering. Springer, New York 1997. DOI:10.1007/978-1-4612-1924-8

[3] S. Axsäter: Simple solution procedures for a class of two-echelon inventory problems. Oper. Res. *38* (1990), 1, 64–69. DOI:10.1287/opre.38.1.64

[4] S. Axsäter: Exact and approximate evaluation of batch-ordering policies for two-level inventory systems. Oper. Res. *41* (1993), 4, 777–785. DOI:10.1287/opre.41.4.777

[5] S. Axsäter: Exact analysis of continuous review (R,Q) policies in two-echelon inventory systems with compound Poisson demand. Oper. Res. *48* (2000), 5, 686–696. DOI:10.1287/opre.48.5.686.12403

[6] S. Axsäter: Approximate optimization of a two-level distribution inventory system. Int. J. Product. Econom. *81* (2003), 545–553. DOI:10.1016/s0925-5273(02)00270-0

[7] S. Axsäter and J. Marklund: Optimal position-based warehouse ordering in divergent two-echelon inventory systems. Oper. Res. *56* (2008), 4, 976–991. DOI:10.1287/opre.1080.0560

[8] B. Balcıoğlu, D. L. Jagerman, and T. Altiok: Approximate mean waiting time in a $GI/D/1$ queue with autocorrelated times to failures. IIE Trans. *39* (2007), 10, 985–996. DOI:10.1080/07408170701275343

[9] B. M. Beamon: Supply chain design and analysis: Models and methods. Int. J. Product. Econom. *55* (1998), 3, 281–294. DOI:10.1016/s0925-5273(98)00079-6

[10] S. Benjaafar, W. L. Cooper, and J.-S. Kim: On the benefits of pooling in production-inventory systems. Management Sci. *54* (2005), 548–565. DOI:10.1287/mnsc.1040.0303

[11] G. R. Bitran and S. Dasu: Approximating nonrenewal processes by Markov chains: Use of Super-Erlang (SE) chains. Oper. Res. *41* (1993), 5, 903–923. DOI:10.1287/opre.41.5.903

[12] G. R. Bitran and S. Dasu: Analysis of the $\sum Ph_i/Ph/1$ queue. Oper. Res. *42* (1994), 1, 158–174. DOI:10.1287/opre.42.1.158

[13] J. A. Buzacott and D. Kostelski: Matrix-geometric and recursive algorithm solution of a two-stage unreliable how line. IIE Trans. *19* (1987), 4, 429–438. DOI:10.1080/07408178708975416

[14] G. P. Cachon: Exact evaluation of batch-ordering inventory policies in two-echelon supply chains with periodic review. Oper. Res. *49* (2001), 1, 79–98. DOI:10.1287/opre.49.1.79.11188

[15] F. Chen and Y.-S. Zheng: One-warehouse multiretailer systems with centralized stock information. Oper. Res. *45* (1997), 2, 275–287.

[16] E. P. Chew and L. C. Tang: Warehouse-retailer system with stochastic demands – Non-identical retailer case. Europ. J. Oper. Res. *82* (1995), 1, 98–110. DOI:10.1016/0377-2217(93)e0279-7

[17] B. L Deuermeyer and L. B. Schwarz: A model for the analysis of system service level in warehouse-retailer distribution systems: The identical retailer case. Institute for Research in the Behavioral, Economic, and Management Sciences, Krannert Graduate School of Management, Purdue University, 1979.

[18] E. B. Diks, A. G. de Kok, and A. G. Lagodimos: Multi-echelon systems: A service measure perspective. Europ. J. Oper. Res. *95* (1996), 2, 241–263. DOI:10.1016/s0377-2217(96)00120-8

[19] S. S. Erenguc, N. C. Simpson, and A. J. Vakharia: Integrated production/distribution planning in supply chains: An invited review. Europ. J. Oper. Res. *115* (1999), 2, 219–236. DOI:10.1016/s0377-2217(98)90299-5

[20] A. S. Eruguz, E. Sahin, Z. Jemai, and Y. Dallery: A comprehensive survey of guaranteed-service models for multi-echelon inventory optimization. Int. J. Product. Econom. *172* (2016), 110–125. DOI:10.1016/j.ijpe.2015.11.017

[21] R. Forsberg: Exact evaluation of (R,Q)-policies for two-level inventory systems with Poisson demand. Europ. J. Oper. Res. *96* (1997), 1, 130–138. DOI:10.1016/s0377-2217(96)00137-3

[22] M. K. Girish and J.-Q. Hu: Higher order approximations for the single server queue with splitting, merging and feedback. Europ. J. Oper. Res. *124* (2000), 3, 447–467. DOI:10.1016/s0377-2217(99)00174-5

[23] C. Z. Gurgur and T. Altiok: Approximate analysis of decentralized, multi-stage, pull-type production/inventory systems. Ann. Oper. Res. *125* (2004), 1–4, 95–116. DOI:10.1023/b:anor.0000011187.52502.37

[24] A. Karaman and T. Altiok: Approximate Analysis of Batch Ordering Policies in Distribution Inventory Systems. Technical Report TR-2007-050, Rutgers University, Department of Industrial and Systems Engineering, 2007.

[25] A. Karaman and T. Altiok: Approximate analysis and optimization of batch ordering policies in capacitated supply chains. Europ. J. Oper. Res. *193* (2009), 1, 222–237. DOI:10.1016/j.ejor.2007.10.018

[26] A. S. Karaman: Performance Analysis and Design of Batch Ordering Policies in Supply Chains. PhD Thesis, Rutgers, The State University of New Jersey, 2007.

[27] D. M. Lucantoni: New results on the single server queue with a batch Markovian arrival process. Commun. Statist. Stoch. Models *7* (1991), 1, 1–46. DOI:10.1080/15326349108807174

[28] K. Moinzadeh and H. L. Lee: Batch size and stocking levels in multi-echelon repairable systems. Management Sci. *32* (1986), 12, 1567–1581. DOI:10.1287/mnsc.32.12.1567

[29] M. F. Neuts: Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. Dover Publications, 1994.

[30] I. Norros, J. W. Roberts, A. Simonian, and J. T. Virtamo: The superposition of variable bit rate sources in an ATM multiplexer. IEEE J. Selected Areas Commun. *9* (1991), 3, 378–387. DOI:10.1109/49.76636

[31] T. Osogami and M. Harchol-Balter: A closed-form solution for mapping general distributions to minimal PH distributions. In: Computer Performance Evaluation. Modelling Techniques and Tools (P. Kemper and W. H. Sanders, eds.), Lecture Notes in Computer Science *2794*, pp. 200–217. Springer Berlin Heidelberg, 2003. DOI:10.1007/978-3-540-45232-4_13

[32] L. B. Schwarz, B. L. Deuermeyer, and R. D. Badinelli: Fill-rate optimization in a one-warehouse N-identical retailer distribution system. Management Sci. *31* (1985), 5, 488–498. DOI:10.1287/mnsc.31.4.488

[33] C. C. Sherbrooke: Metric: A multi-echelon technique for recoverable item control. Operr. Res. *16* (1968), 1, 122–141. DOI:10.1287/opre.16.1.122

[34] A. Svoronos and P. Zipkin: Estimating the performance of multi-level inventory systems. Oper. Res. *36*(1988), 57–72. DOI:10.1287/opre.36.1.57

[35] A. Svoronos and P. Zipkin: Evaluation of one-for-one replenishment policies for multiechelon inventory systems. Management Sci. *37* (1991), 1, 68–83. DOI:10.1287/mnsc.37.1.68

[36] H. Tempelmeier: A multi-level inventory system with a make-to-order supplier. Int. J. Product. Res. *51* (2013), 23–24, 6880–6890. DOI:10.1080/00207543.2013.776190

[37] M. van Vuuren and I. J. B. F. Adan: Approximating multiple arrival streams by using aggregation. Stoch. Models *22* (2006), 3, 423–440. DOI:10.1080/15326340600820398

[38] W. Whitt: Approximating a point process by a renewal process, I: Two basic methods. Oper. Res. *30* (1982), 1, 125–147. DOI:10.1287/opre.30.1.125

[39] P. Zipkin: The use of phase-type distributions in inventory-control models. Naval Res. Logistics *35* (1988), 2, 247–257. DOI:10.1002/1520-6750(198804)35:2¡247::aid-nav3220350209¿3.0.co;2-l

Abdullah S. Karaman, Department of Industrial Engineering, American University of the Middle East, P. O.Box 220 Dasman, 15453. Kuwait.
e-mail: abdullahkaraman@yahoo.com