

AN EXACT AGGREGATION FOR INVENTORY DISTRIBUTION IN AN AUTOMATIC WAREHOUSING SYSTEM

Hideaki Yamashita
Tohoku University

Hiroshi Ohtani
NEC Corporation

Shigemichi Suzuki
Chiba Institute of Technology

(Received September 26, 1996; Revised September 24, 1997)

Abstract We consider a simple automatic warehousing system with a lot of storage spaces called slots in which only a single type of items are stored and retrieved. The purpose of this paper is to develop an efficient method for obtaining the marginal probability that each of the slots in the warehouse is full. The set of such probabilities for all the slots constitutes the spatial inventory distribution of the items in the system. This distribution, which we call the inventory distribution in short, enables us to evaluate key performance characteristics of the system such as the mean travel time of a crane for a single operation of storage or retrieval of items. Such characteristics cannot be obtained from the distribution of the total number of full slots alone.

We assume that inventories are controlled by an (s, S) reordering policy, and that received items are stored from the closest open slot to an I/O point and retrieved items are chosen randomly among currently full slots. Furthermore the time between retrievals and the time between placing of order and receipt of items are exponentially distributed. Under these assumptions the system can be modeled as a Markov chain. If we use joint inventory levels of slots, the number of states of the Markov chain amounts to 2^m , where m is the total number of slots. Here we devise exact aggregation methods of the states to reduce the size of the Markov chain. By exploiting the special structure the total computational complexity for obtaining the inventory distribution is reduced to $O(m^4)$.

1. Introduction

A typical automatic warehousing system has, among other things, racks with slots in which items are stored, a stacker crane which stores and retrieves items automatically, and a computer which controls the operations of the crane and the inventories as shown in Fig. 1. An order for items is placed according to a certain inventory policy. When items arrive at the warehousing system, they are assigned to pallets and then the pallets are put into the system from the I/O point. In the warehousing system items are controlled in the unit of pallet. The crane transports each pallet from the I/O point to one of open slots, and stores it in the slot. Retrieval of items will be done almost in the reverse order.

The computer chooses a slot for storing or retrieving according to a certain rule. This rule as well as the inventory control policy based on assumptions on request for retrieval of items and delivery lags has a great influence upon the performance of the warehousing system. As we will show in Section 2.3 the distribution of the total number of items in the warehouse can be obtained rather easily. However, this is not sufficient to evaluate such key performance characteristics as the mean travel time of the crane for an operation of storage or retrieval of items. We must know also how items are distributed in the system with respect to the I/O point. We will consider the marginal probability that each slot is full, then the set of such probabilities for all the slots in the warehouse will constitute the spatial inventory distribution of the items in the system. We call this spatial distribution as the inventory distribution in short. Once the inventory distribution is known, the distribution of

travel time can be determined from the geometrical configuration of slots in the warehouse and the physical performance of the crane such as speed and load/unload time of items. The distribution of traveling time is a key operating factor for evaluation of overall warehousing systems, though for that purpose such performance characteristics as queue lengths of pallets of items yet to be stored into slots and unfilled requests for retrieval of items must be also taken into account. The overall evaluation of warehousing system requires a comprehensive model-building process and will be a topic of for further research. In this paper we will devote ourselves to determination of the inventory distribution in the warehousing system operated by (s,S) policy in order to estimate the travel time distribution. Although the inventory distribution is very important in designing warehousing systems, all of the previous works on the related topics have not treated the inventory distribution adequately. In this paper we will show that the inventory distribution can be obtained with reasonable computational efforts by aggregation methods.

In some previous works analytical expressions are obtained for travel time of a crane in automatic warehousing systems. Based on the expected travel times of the crane, Hausman, Schwarz and Graves [4] compared three storage methods, i.e., randomized storage, turn-over based storage and class-based storage. Another paper by Graves, Hausman and Schwarz [3] investigated analytical and empirical results for various combinations of alternative storage assignment rules and interleaving rules. In [4] and [3], they have assumed that the I/O point is located at the corner of the rack and that the rack is square in time, i.e., the crane can get to the remotest point from the I/O point in the vertical direction and such point in the horizontal direction in the same length of time. Bozer and White [1] determined the expected travel time for alternative the I/O point locations and rack configurations. Hwang and Lee [2] and Chang *et al.* [5] further generalized the previous models by considering the acceleration and deceleration rates of an S/R machine. In these analyses an assumption is made that items are stored in the rack uniformly. This assumption makes the analysis easy. In real systems, however, this is not the case. There are open as well as full slots, and the distribution of full slots is not uniform. The assumption of uniformity gives rise to noticeable errors. Thus the purpose of this paper is to establish an efficient method for obtaining the inventory distribution to evaluate warehousing system more accurately.

Noguchi and Suzuki [7] have taken into consideration the inventory distribution in analyzing the automatic warehousing system. In their system, n different types of items were assumed to exist and inventories were controlled by an (s, S) reordering policy. They supposed that received items were stored from the closest open slot to the I/O point and retrieval was random among the same type of items. Under further assumptions of random arrival and random request for items this system can be formulated as a Markov chain with $(1 + n)^m$ states, where n and m are the numbers of item types and the slots respectively. It is almost impossible for practical sizes of m and n to solve the system of equilibrium equations for the steady state probabilities. Then they proposed an approximate solution method for this problem.

In this paper we employ the same model as in [7] but restrict ourselves to a single type of items. Then we show that aggregation leads us to a smaller Markov chain which provides us the inventory distribution with less computational efforts. More specifically we focus our attention on whether slot i ($i = 1, 2, \dots, m$) is full, the total number of full slots, the number of full slots among first i slots, and associated states, then we can get an aggregated Markov chain whose transition probabilities are all known. Although the model will have 2^m states if we use joint inventory levels of slots, the aggregation reduces the number of states to $O(m^2)$ for each slot. In addition, we propose $O(m^3)$ time algorithm for obtaining

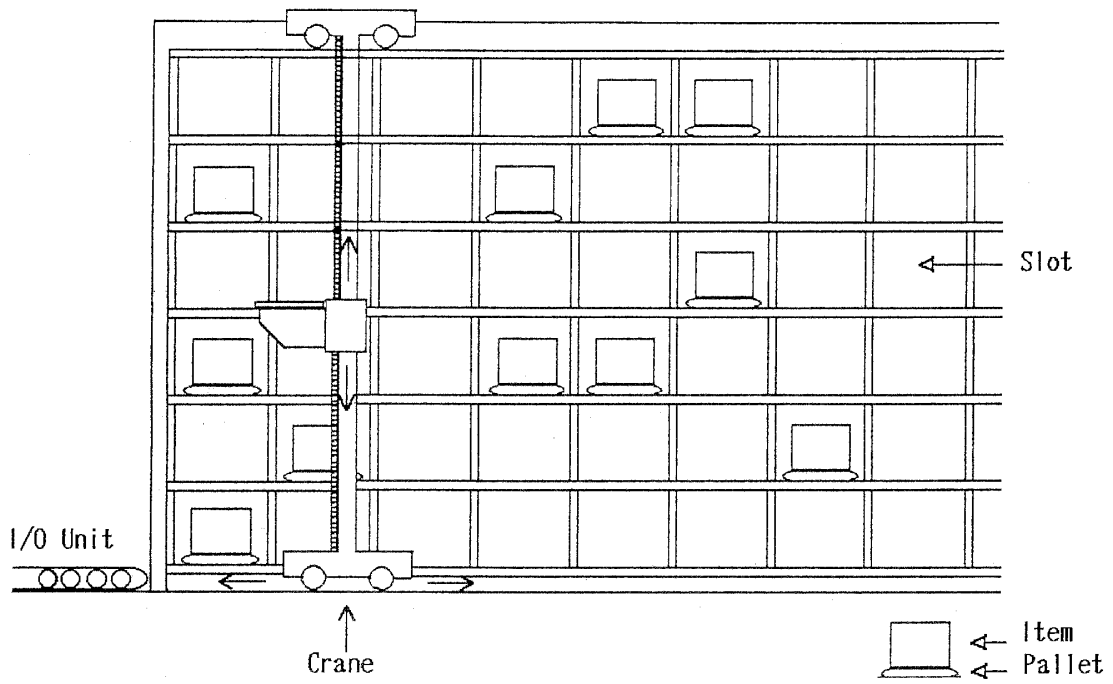


Figure 1: Automatic Warehousing System.

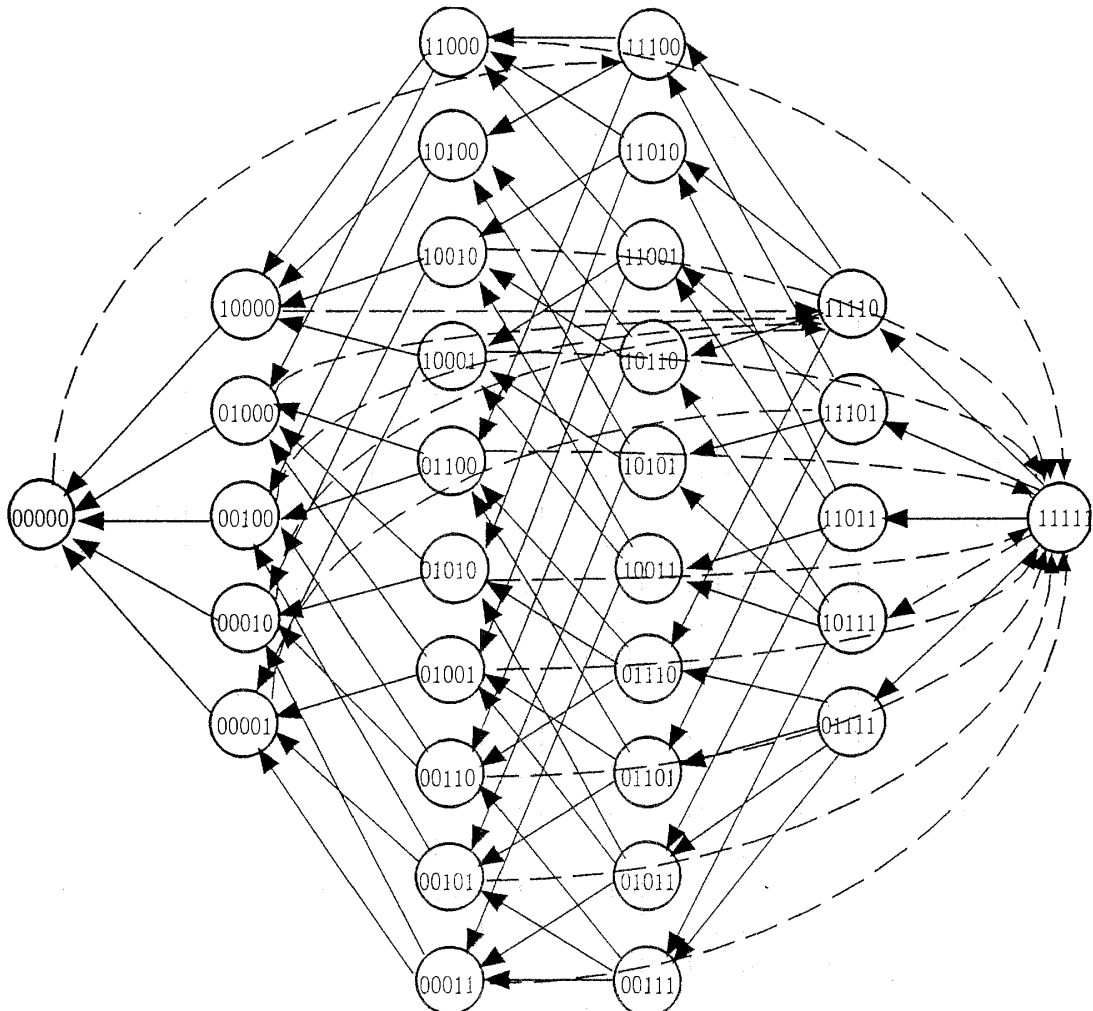


Figure 2: Global State Transition Diagram ($m = 5, q = 3, r = 2$).

the steady state probabilities of the aggregated states by utilizing the structure of the linear equilibrium equations for each slot. As a result, our algorithm makes it possible to obtain the inventory distribution for all the slots in $O(m^4)$ time. This gives a way to analyze the inventory distribution in an automatic warehousing system with a practical number of slots.

In Section 2 we will describe our model and assumptions and make preliminary analyses about the probability distribution of the total number of full slots and equilibrium equations for joint inventory levels of slots. There are two types of aggregation depending on the location of slots. In Sections 3 and 4 we will show those two aggregation methods and efficient methods for solving the aggregated systems of equilibrium equations respectively. Section 5 contains results of numerical experiments and Section 6 gives concluding remarks and comments on extensions.

2. Model and Preliminary Analysis

2.1. Assumptions and Notations

The assumptions and notations in our model will be described in the following:

- (a) Only a single type of items are treated in our model. The quantity of items is counted in the number of pallets.
- (b) There are m slots in the automatic warehousing system. To make the argument simple, the slots are assumed to be in a row and numbered from 1 (the closest slot to the I/O point) to m (the remotest slot from the I/O point).
- (c) Each slot can store one pallet of items.
- (d) Requests for retrieving one pallet come to this system according to a Poisson process with a constant rate μ . When a request for retrieval comes, the items on a pallet are taken out from one of the full slots which is chosen at random. If there are no items in the system at that time, the request is lost.
- (e) The inventory is controlled by an (s, S) reordering policy, i.e., if the inventory level decreases to r (reordering point), a reorder of q (reordering quantity) pallets of items will be placed, where r and q are prescribed constants. The lead time between placing of order and receipt of items is exponentially distributed with a constant parameter λ .
- (f) As soon as q pallets of ordered items arrive at the system, each pallet is stored in the closest open slot to the I/O point.
- (g) The number of slots is equal to the maximum inventory level, i.e., $m = r + q$. The reordering point is less than the reordering quantity, i.e., $r < q$.
- (h) The time required for storage or retrieval of pallets which consists of the travel time of the crane and the load/unload time of items at the I/O point and the slots is assumed to be zero, so we can neglect the waiting time for the crane.

Some of the assumptions will be explained to clarify our standing point: Storing of each pallet of arriving items in the closest open slot to I/O point (assumption(f)) is natural since it makes the travel time short. A pallet of requested items are assumed to be taken out at random from one of the full slots (assumption(c)). There are alternatives for this rule: One rule is of taking out a pallet from the slot to the closest to I/O point and another is of taking the oldest pallet in the warehouse. Those rules may either leave very old items in the warehouse or require history of inventories in the warehouse and time-consuming processing. A rule of assumption(c) is simple to operate and does not leave old items excessively. It may seem contradictory to neglect waiting time of the crane since we are interested in travel time of the crane. Remember that we are going to determine the inventory distribution in

the warehouse. If arrivals of ordered items and requests for retrievals of items are processed in FCFS (First Come First Served) basis, the inventory distribution of items at a moment of start of storage or retrieval of pallets does not depend on time required for storage or retrieval of pallets. This permits us to set time required for storage or retrieval arbitrarily. Hence as far as the inventory distribution is concerned we can neglect the time required for movement of the crane and processing of pallets. Of course the time associated with crane must be taken into account to evaluate the overall performance of the warehousing system.

2.2. Probability Distribution of Total Number of Full Slots

Let $X(t)$ be a random variable designating the total number of full slots in the warehouse at time t . From the assumptions on Poisson arrival of ordered items and random request for retrieval of items described in Section 2.1 the stochastic process $X(t)$ is ergodic and Markovian. Let $P(i)$ be the steady state probability that $X(t) = i$ ($i = 0, 1, \dots, m$). The system of equilibrium equations for $P(i)$ can be easily obtained as follows:

$$\begin{cases} \lambda P(0) = \mu P(1) & (i = 0), \\ (\lambda + \mu)P(i) = \mu P(i + 1) & (1 \leq i \leq r), \\ \mu P(i) = \mu P(i + 1) & (r < i < q), \\ \mu P(i) = \lambda P(i - q) + \mu P(i + 1) & (q \leq i < m), \\ \mu P(m) = \lambda P(r) & (i = m). \end{cases}$$

After a little algebra, we can express $P(i)$'s only with $P(q)$ as follows:

$$\begin{cases} P(0) = \frac{\mu}{\lambda} \left(\frac{\mu}{\lambda + \mu}\right)^r P(q) & (i = 0), \\ P(i) = \left(\frac{\mu}{\lambda + \mu}\right)^{r-i+1} P(q) & (1 \leq i \leq r), \\ P(i) = P(q) & (r < i < q), \\ P(i) = \left[1 - \left(\frac{\mu}{\lambda + \mu}\right)^{m-i+1}\right] P(q) & (q < i \leq m), \end{cases} \tag{2.1}$$

by using these equations and the condition

$$\sum_{i=0}^m P(i) = 1,$$

we can get $P(q)$ at first, and other $P(i)$'s can be obtained by (2.1) easily. Those probabilities along with other more detailed information will be used in Section 3.2 to compute the inventory distribution, i.e., the probability that each of the slot is full.

2.3. System of Equilibrium Equations for Joint Inventory Levels for Slot 1 to m

Let $X_i(t)$ be a zero-one random variable denoting whether slot i is full (1) or empty (0) at time t and x_i be a value which $X_i(t)$ assumes ($i = 1, 2, \dots, m$). It is easy to see from the assumptions described in Section 2.1 that the stochastic process $\{(X_1(t), X_2(t), \dots, X_m(t))\}$ is ergodic and Markovian. Let $P(x_1, x_2, \dots, x_m)$ be the stationary joint probability for inventory levels for slot 1 through slot m , namely,

$$P(x_1, x_2, \dots, x_m) = \lim_{t \rightarrow \infty} Pr(X_1(t) = x_1, X_2(t) = x_2, \dots, X_m(t) = x_m).$$

We assume that $P(x_1, x_2, \dots, x_m) = 0$ for impossible states, i.e., the states for which some x_i is neither 0 nor 1. The system of equilibrium equations for the warehousing system can be written down referring to, for example, Fig. 2 in case of $m=5, q=3,$ and $r=2$. The equations

are of different forms depending on the total inventory level as follows:

$$\left\{ \begin{array}{ll} \mu \sum_{j=1}^m P(x_1, \dots, x_j + 1, \dots, x_m) - \lambda P(x_1, \dots, x_m) = 0 & (f_m = 0), \\ d(f_m) \sum_{j=1}^m P(x_1, \dots, x_j + 1, \dots, x_m) - (\mu + \lambda) P(x_1, \dots, x_m) = 0 & (1 \leq f_m \leq r), \\ d(f_m) \sum_{j=1}^m P(x_1, \dots, x_j + 1, \dots, x_m) - \mu P(x_1, \dots, x_m) = 0 & (r < f_m < q), \\ \lambda \sum_{e_1+\dots+e_m=q} P(x_1 - e_1, \dots, x_i - e_i, x_{i+1} - e_{i+1}, \dots, x_m - e_m) \\ \quad + d(f_m) \sum_{j=1}^m P(x_1, \dots, x_j + 1, \dots, x_m) - \mu P(x_1, \dots, x_m) = 0 & (q \leq f_m < m), \\ \lambda \sum_{e_1+\dots+e_m=q} P(x_1 - e_1, \dots, x_i - e_i, x_{i+1} - e_{i+1}, \dots, x_m - e_m) \\ \quad - \mu P(x_1, \dots, x_m) = 0 & (f_m = m), \\ \sum P(x_1, \dots, x_m) = 1, & \end{array} \right. \tag{2.2}$$

where

$$f_i = x_1 + x_2 + \dots + x_i \quad (i = 1, 2, \dots, m),$$

$$d(f_m) = \mu / (f_m + 1),$$

and e_i ($i = 1, 2, \dots, m$) is either 0 or 1.

We will briefly explain the equilibrium equations (2.2). For $1 \leq f_m \leq r$, both an arrival of ordered items and a request for retrieval cause transition out of the state (x_1, x_2, \dots, x_m) with the rates λp and μp respectively, where p is the stationary probability of designated state. In addition, transitions from $(m - f_m)$ states with the inventory level $f_m + 1$ into the state (x_1, x_2, \dots, x_m) will occur due to a request for retrieval with the rates $\mu / (f_m + 1)$ since the pallet is retrieved at random among $(f_m + 1)$ full slots. For $r < f_m < q$, no transitions occur associated with arrival of ordered items. The transition rate out of the state (x_1, x_2, \dots, x_m) is μp and transitions from the states with the inventory level $f_m + 1$ into the state (x_1, x_2, \dots, x_m) are the same as the case $1 \leq f_m \leq r$. For $q \leq f_m < m$ the situation is the same as in the case $r < f_m < q$ except that there exist transitions from the states with the inventory level $(f_m - q)$ into the state (x_1, x_2, \dots, x_m) due to arrival of ordered items with the rate λp . For $q \leq f_m < m$, we have expressed the situation concretely that q arriving pallets have been put into empty slots up to slot m . The case $f_m = m$ is the same as the case $q \leq f_m < m$ except that there is no transition into the state (x_1, x_2, \dots, x_m) due to retrieval of items.

The system of equilibrium equations has 2^m unknown variables. If we can solve it, we, of course, obtain the marginal probabilities by

$$Pr(X_i(t) = 1) = \sum_{x_k, k \neq i} P(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_m) \quad (i = 1, 2, \dots, m).$$

But it is practically impossible, because real systems have more than thousand slots in some cases. Therefore we need some devices to overcome this difficulty.

3. Exact Aggregation Method I

3.1. Aggregation for Slots 1 to q , and m

In this section and next, we will propose an exact aggregation method for the inventory distribution, i.e., the probability that each of the m slots in the warehouse is full. This method does not give the probabilities for all the slots at the same time, but gives the probability for each slot individually. We will reduce the computation time drastically by aggregating associated states of the Markov process so as to keep the necessary information for obtaining the probability that the slot is full. We use two kinds of aggregation depending on which slots we are working on: The one is for obtaining the probability that each of slots

1 $\sim q$ and m is full and the other is for the rest of the slots. In this section, we will explain the former.

Our purpose is to establish an efficient method for computing the marginal probability $Pr(X_i(t) = 1)$ ($i = 1, \dots, m$).

Let $F_i(t)$ be a random variable designating the total number of full slots in slot 1 through slot i , namely

$$F_i(t) = X_1(t) + X_2(t) + \dots + X_i(t) \quad (i = 1, \dots, m),$$

and let f_i be a value which $F_i(t)$ assumes. We will introduce a stochastic process $\{(X_i(t), F_m(t))\}$ with $2m$ states for each i to reduce the number of states of the original Markov process. Note that there do not exist states $(1, 0)$ and $(0, m)$ for each i . Associated with each of $2m$ states (x_i, f_m) define a set $S_i(x_i, f_m)$ as ,

$$S_i(x_i, f_m) = \{(x_1, x_2, \dots, x_m) \mid \sum_{k=1, k \neq i}^m x_k = f_m - x_i\},$$

then

$$S = \left(\bigcup_{k=0}^{m-1} S_i(0, k) \right) \bigcup \left(\bigcup_{k=1}^m S_i(1, k) \right)$$

will be a partition of the set S of all the 2^m states of the original Markov process. There exists a one-to-one correspondence between a state (x_i, f_m) of the stochastic process $\{(X_i(t), F_m(t))\}$ and a set $S_i(x_i, f_m)$ of states of the original Markov process.

The stochastic process $\{(X_i(t), F_m(t))\}$ is also a Markov process for $1 \leq i \leq q$ and $i = m$. In order to show this, first define the transition rate from a state $x = (x_1, x_2, \dots, x_m) \in S_i(x_i, f_m)$ to a set of states $S_i(x'_i, f'_m)$ based on the transition rate $p(x, x')$ from a state x to a state $x' = (x'_1, x'_2, \dots, x'_m)$ of the original process as

$$p(x, S_i(x'_i, f'_m)) = \sum_{x' \in S_i(x'_i, f'_m)} p(x, x'). \quad (3.1)$$

Essentially due to the lumpability theorem of a Markov chain of Kemeny and Snell [4, p.124], the stochastic process $\{(X_i(t), F_m(t))\}$ is a Markov process if and only if for every pair of sets $S_i(x_i, f_m)$ and $S_i(x'_i, f'_m)$, $p(x, S_i(x'_i, f'_m))$ has the same value for every $x \in S_i(x_i, f_m)$. This common value will be the transition rate from a set of aggregated states $S_i(x_i, f_m)$ to another set $S_i(x'_i, f'_m)$ or equivalently from a state (x_i, f_m) to a state (x'_i, f'_m) of the stochastic process $\{(X_i(t), F_m(t))\}$.

Now we will show that $p(x, S_i(x'_i, f'_m))$ has the same value for every $x \in S_i(x_i, f_m)$. Consider a state $x \in S_i(x_i, f_m)$ and the set of states S_x to which transitions are possible in one step caused by retrieval or receipt of items from x . If $f_m \leq r$ receipt is possible and S_x has a single element x' and the rate of transition is λ since received items are stored from the closest open slot to the I/O point (see assumption (f)). In case of $i \leq q$, $x' \in S_i(1, f_m + q)$ is the only state after receipt of items for any $x \in S_i(x_i, f_m)$, but for $m > i > q$ there exist y' and z' such that

$$y' \in S_i(0, f_m + q) \quad \text{and} \quad z' \in S_i(1, f_m + q)$$

are the states right after receipt of items. This means that the transition rate $p(x, S_i(0, f_m + q))$ does not necessarily have the same value for every $x \in S_i(0, f_m)$. This implies that the stochastic process $\{(X_i(t), F_m(t))\}$ is not Markovian if $m > i > q$, but it will be Markovian if $i \leq q$ and the transition rate defined in (3.1) has a common value for transitions by retrieval of items.

Now we will consider transitions caused by retrieval of items. If $f_m \neq 0$, retrieval is possible and items in one of full slots will be taken out at random. There are f_m states in S_x in this case. We will consider two sub-cases, i.e., the cases $x_i = 0$ and $x_i = 1$. In the former case all the possible transitions are into one of the states in $S_i(0, f_m - 1)$ and

$$p(x, S_i(0, f_m - 1)) = f_m(\mu/f_m) = \mu \tag{3.2}$$

for all $x \in S_i(0, f_m)$ because of Assumption (d). In the latter case two kinds of transitions are possible: One is a transition into one of the states in $S_i(0, f_m - 1)$ and the other is a transition into one of the states in $S_i(1, f_m - 1)$. Because of Assumption (d) again, we have

$$p(x, S_i(0, f_m - 1)) = \mu/f_m \text{ and } p(x, S_i(1, f_m - 1)) = (f_m - 1)\mu/f_m$$

for all $x \in S_i(1, f_m)$. We have shown that the transition rate (3.1) has a common value for every $x \in S_i(x_i, f_m)$ for transitions by retrieval of items. The discussions above now have established that the stochastic process $\{(X_i(t), F_m(t))\}$ is a Markov process if $i \leq q$. We denote by $P_i(x_i, f_m)$ the steady state joint probability distribution for the inventory level of slot i and the inventory level of the system. Then we can derive the system of equilibrium equations for $P_i(x_i, f_m)$. There is, of course, one equation for each state (x_i, f_m) . The equilibrium equations for $P_1(x_1, f_m)$ are derived as a typical example in the following by classifying them with the value of f_m . The system of equations are written down in the following in the descending order of f_m since those equations are referred to in this order later in this paper:

$$\left\{ \begin{array}{ll} \lambda P_1(0, r) + \lambda P_1(1, r) - \mu P_1(1, m) = 0 & (f_m = m), \\ \lambda P_1(0, f_m - q) + \lambda P_1(1, f_m - q) \\ \quad + a(f_m)\mu P_1(1, f_m + 1) - \mu P_1(1, f_m) = 0 & (q \leq f_m < m), \\ a(f_m)\mu P_1(1, f_m + 1) - \mu P_1(1, f_m) = 0 & (r < f_m < q), \\ a(f_m)\mu P_1(1, f_m + 1) - (\mu + \lambda)P_1(1, f_m) = 0 & (1 \leq f_m \leq r), \\ b(f_m)\mu P_1(1, f_m + 1) + \mu P_1(0, f_m + 1) \\ \quad - \mu P_1(0, f_m) = 0 & (r < f_m < m), \\ b(f_m)\mu P_1(1, f_m + 1) + \mu P_1(0, f_m + 1) \\ \quad - (\mu + \lambda)P_1(0, f_m) = 0 & (1 \leq f_m \leq r), \\ \mu P_1(0, 1) + \mu P_1(1, 1) - \lambda P_1(0, 0) = 0 & (f_m = 0), \\ \sum_{f_m=0}^{m-1} P_1(0, f_m) + \sum_{f_m=1}^m P_1(1, f_m) - 1 = 0, & \end{array} \right. \tag{3.3}$$

where

$$a(f_m) = f_m/(f_m + 1), \quad b(f_m) = 1/(f_m + 1).$$

Since the state transitions in the aggregated Markov process are similar to those in the original process explained in detail in Section 2.3, we will explain only the case $q \leq f_m < m$ here. Transitions into state $(1, f_m)$ by receipt of items are from states $(0, f_m - q)$ and $(1, f_m - q)$ respectively both with transition rate λ which correspond to the first and the second terms. A transition into state $(1, f_m)$ by retrieval is from state $(1, f_m + 1)$ with the rate $p((1, f_m + 1), (1, f_m)) = f_m\mu/(f_m + 1) = a(f_m)\mu$ which corresponds to the third term. A transition out of this state $(1, f_m)$ is possible only by retrieval of items. The transition rate is μ from (3.2) which corresponds to the last term. We will show an example of the aggregation for the model with $m = 5, q = 3$, and $r = 2$. The state transition diagram of the system before aggregation is shown in Fig. 2 which corresponds to the equilibrium equations (2.2). After aggregation the state transition diagram for the states (x_1, f_m) will

be as shown in Fig. 3 which corresponds to the first seven equations of (3.3). In general the state transition diagram for the states (x_1, f_m) will be as shown in Fig. 4. (3.3) is a system of linear equilibrium equations with $2m$ unknown variables. The number of variables is much less than that in (2.2).

Among m variables $X_1(t), \dots, X_m(t)$, q variables $X_1(t), \dots, X_q(t)$ are the random variables which have the same characteristic as $X_1(t)$. We can apply the same aggregation method as above to each of slots $2 \sim q$ and exactly the same system of equilibrium equations as (3.3) can be obtained for $P_i(x_i, f_m)$ for each i . This means that the probability that slot i is full is the same for $i = 1, 2, \dots, q$. We can see it intuitively because of the rules for storage and retrieval in Section 2.1. The states of slot m can be described similarly to those of slots 2 through q and the same type of aggregation can be applied. The system of equilibrium equations for $P_m(x_m, f_m)$ is as follows:

$$\left\{ \begin{array}{ll} \lambda P_m(0, r) + \lambda P_m(1, r) - \mu P_m(1, m) = 0 & (f_m = m), \\ \lambda P_m(1, f_m - q) + a(f_m)\mu P_m(1, f_m + 1) \\ \quad - \mu P_m(1, f_m) = 0 & (q + 1 \leq f_m < m), \\ a(f_m)\mu P_m(1, f_m + 1) - \mu P_m(1, f_m) = 0 & (r < f_m < q + 1), \\ a(f_m)\mu P_m(1, f_m + 1) - (\mu + \lambda)P_m(1, f_m) = 0 & (1 \leq f_m \leq r), \\ \lambda P_m(0, f_m - q) + b(f_m)\mu P_m(1, f_m + 1) \\ \quad + \mu P_m(0, f_m + 1) - \mu P_m(0, f_m) = 0 & (q \leq f_m < m), \\ b(f_m)\mu P_m(1, f_m + 1) + \mu P_m(0, f_m + 1) \\ \quad - \mu P_m(0, f_m) = 0 & (r < f_m < q), \\ b(f_m)\mu P_m(1, f_m + 1) + \mu P_m(0, f_m + 1) = 0 \\ \quad - (\mu + \lambda)P_m(0, f_m) & (1 \leq f_m \leq r), \\ \mu P_m(0, 1) + \mu P_m(1, 1) - \lambda P_m(0, 0) = 0 & (f_m = 0), \\ \sum_{f_m=0}^{m-1} P_m(0, f_m) + \sum_{f_m=1}^m P_m(1, f_m) - 1 = 0, & \end{array} \right. \quad (3.4)$$

which is also a system of linear equilibrium equations with $2m$ unknown variables.

3.2. Solution Method for $P(X_i(t) = 1)$ ($i = 1 \sim q$ and m)

Systems of equations (3.3) and (3.4) both have $2m$ unknowns. These equations can be solved with any of standard methods for a system of linear equations. But we will develop a more efficient method by making use of the special structure of the equations. Once $P_i(1, f_m)$'s are known the marginal probability that each of slots $1 \sim q$ and m is full can be obtained by

$$Pr(X_i(t) = 1) = \sum_{f_m=1}^m P_i(1, f_m) \quad (i = 1 \sim q \text{ and } m). \quad (3.5)$$

Now we will present a method for solving Equations (3.3) and (3.4). This method will use the distribution $P(i)$ of the total number of full slots described in Section 2.2. For this purpose we will transform the first four equations of (3.3) as follows observing that $P(i) = P_1(0, i) + P_1(1, i)$:

$$\left\{ \begin{array}{ll} P_1(1, m) = P(m), \\ P_1(1, f_m) = \frac{\lambda}{\mu} P(f_m - q) + a(f_m)P_1(1, f_m + 1) & (q \leq f_m < m), \\ P_1(1, f_m) = a(f_m)P_1(1, f_m + 1) & (r < f_m < q), \\ P_1(1, f_m) = a(f_m) \frac{\mu}{\lambda + \mu} P_1(1, f_m + 1) & (1 \leq f_m \leq r). \end{array} \right.$$

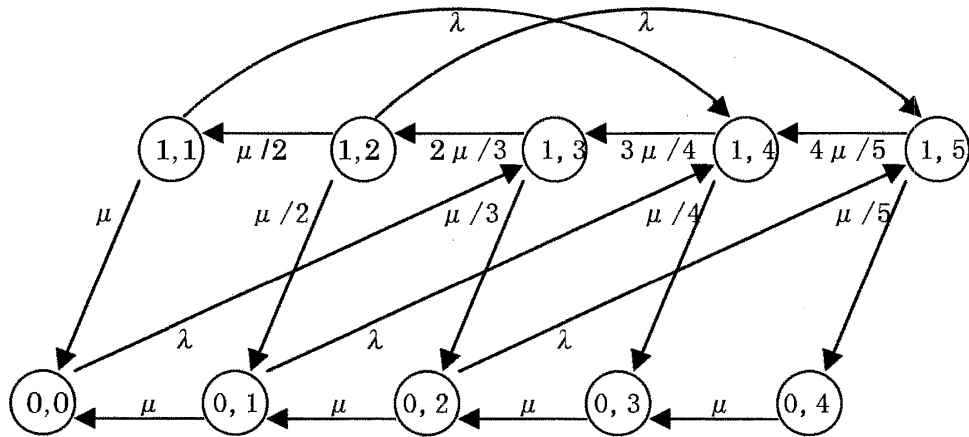


Figure 3: State Transition Diagram for States (x_i, f_m) ($m = 5, q = 3, r = 2$).

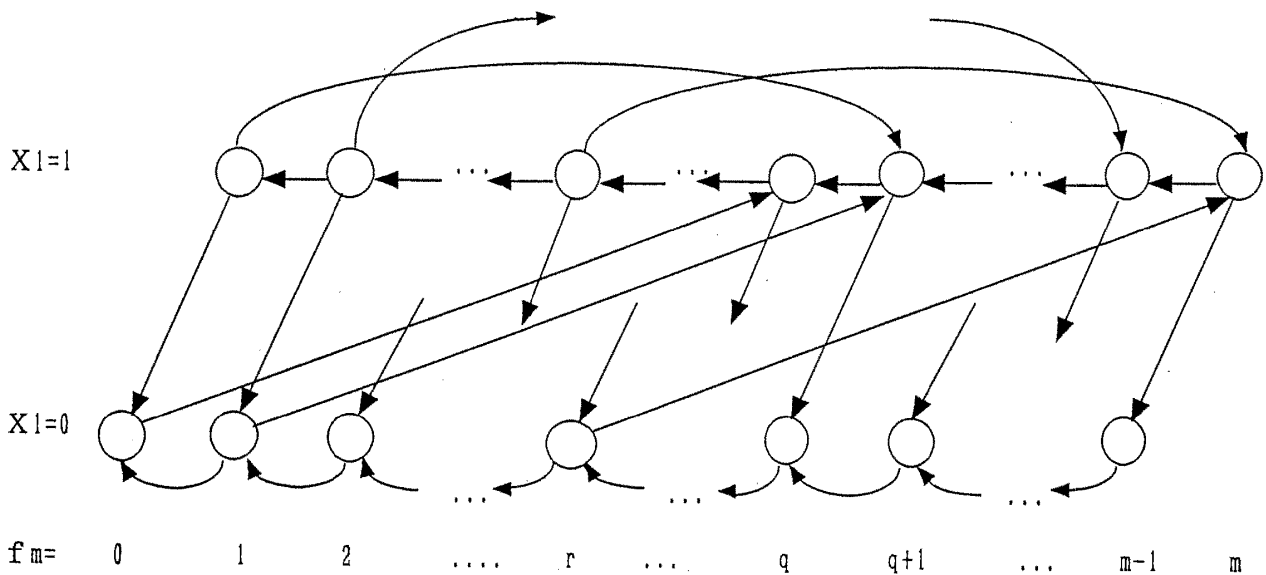


Figure 4: State Transition Diagram for States (x_i, f_m) .

Now $P_1(1, f_m), f_m = m, m - 1, \dots, 1$, are obtained recursively. Similarly, we can transform the first four equations of (3.4) to

$$\left\{ \begin{array}{l} P_m(1, m) = P(m), \\ P_m(1, q + 1) = [(q + 1)/(1 - \frac{\lambda}{\mu} \sum_{j=1}^{r-1} \frac{j}{q + j} (\frac{\mu}{\lambda + \mu})^{r-j+1})/m]P(m), \\ P_m(1, f_m) = \frac{f_m}{q + 1} (\frac{\mu}{\lambda + \mu})^{r-f_m+1} P_m(1, q + 1) \\ \\ P_m(1, f_m) = \frac{f_m}{q + 1} P_m(1, q + 1) \\ P_m(1, f_m) = \frac{f_m}{q + 1} [1 - \frac{\lambda}{\mu} \sum_{j=1}^{f_m-q-1} \frac{j}{q + j} (\frac{\mu}{\lambda + \mu})^{r-j+1}] P_m(1, q + 1) \end{array} \right. \quad \begin{array}{l} (1 \leq f_m \leq r), \\ (r < f_m \leq q + 1), \\ (q + 2 \leq f_m < m). \end{array} \quad (3.6)$$

The third and fourth equations of (3.6) are obtained from the fourth and third equations of (3.4), respectively. The rest of the equations of (3.6) are derived by substituting those equations into the second equation of (3.4). We can obtain $P_m(1, f_m)$ ($f_m = 1, \dots, m$) from (3.6).

Now we can obtain $P_i(1, f_m)$ ($f_m = 1, \dots, m$) for $i = 1, \dots, q$ and m with much less effort than solving (3.3) and (3.4) by standard numerical methods such as Gaussian elimination method. The computational complexities for solving them are dominated by those for obtaining $P_i(x_i, f_m)$ for $i = q + 1 \sim m - 1$ as we will see in Section 4.2.

4. Exact Aggregation Method II

4.1. Aggregation for Slots $q + 1$ to $m - 1$

Exact aggregation of 2^m states S of the original Markov process into $2m$ states $S_i(x_i, f_m)$ was successful to obtain $Pr(X_i(t) = 1)$ for $i = 1, 2, \dots, q$, and m . The Markov process thus obtained was $\{(X_i(t), F_m(t))\}$. This kind of aggregation will fail for $i = q + 1, q + 2, \dots, m - 1$. The reason for this was briefly explained in Section 3.1. We will examine the situation again and propose another method of exact aggregation which enables us to compute $Pr(X_i(t) = 1)$ for $i = q + 1, q + 2, \dots, m - 1$.

Suppose that $q < i < m, f_m \leq r$, and $X_i(t) = 0$, then receipt of items is possible in the state $(0, f_m)$ in the Markov process $\{(X_i(t), F_m(t))\}$. On receipt of items a transition occurs and the new state will be $(1, f_m + q)$ if $f_i \geq i - q$ and $(0, f_m + q)$ if $f_i < i - q$. The lumpability condition fails to hold for such a case since the state does not have information on $F_i(t)$. We need more information to get $Pr(X_i(t) = 1)$ for $i = q + 1, q + 2, \dots, m - 1$. To overcome the difficulty we will introduce a stochastic process $\{(F_i(t), F_m(t))\}$ or equivalently $\{(F_i(t), G_i(t))\}$ where $G_i(t)$ is a random variable defined by

$$G_i(t) = X_{i+1}(t) + X_{i+2}(t) + \dots + X_m(t).$$

The number of states of the stochastic process is less than $(m^2 + 4m + 4)/4$ (since $F_i(t)$ and $G_i(t)$ have $(i + 1)$ and $(m - i + 1)$ states respectively) which is still much smaller than that of the original Markov process.

We will show that the stochastic process $\{(F_i(t), G_i(t))\}$ is Markovian almost in the same way as we proved that $\{(X_i(t), F_m(t))\}$ was so. First we will define the sets of states by

$$T_i(f_i, g_i) = \{(x_1, x_2, \dots, x_m) \mid F_i(t) = f_i, G_i(t) = g_i\} \quad (q < i < m),$$

where g_i is a value which $G_i(t)$ assumes. All the sets $T_i(f_i, g_i)$ for each i clearly constitutes a partition of S . Define the transition rate from a state $x \in T_i(f_i, g_i)$ to a set of states $T_i(f'_i, g'_i)$ by

$$p(x, T_i(f'_i, g'_i)) = \sum_{x' \in T_i(f'_i, g'_i)} p(x, x'). \tag{4.1}$$

If we can show that for every pair of sets $T_i(f_i, g_i)$ and $T_i(f'_i, g'_i)$, the transition rate has the same value for every $x \in T_i(f_i, g_i)$, then an aggregated Markov process can be constructed. This common value forms the transition rate from an aggregated state $T_i(f_i, g_i)$ to another $T_i(f'_i, g'_i)$ of the aggregated Markov process or equivalently from a state (f_i, g_i) to another state (f'_i, g'_i) of the Markov process $\{(F_i(t), G_i(t))\}$.

We will show that the transition rate defined by (4.1) actually has a common value. First let $x \in T_i(f_i, g_i)$. If $f_i + g_i (= f_m) \leq r$, then a transition from x due to receipt of items is possible with the rate λ . Every possible state of the original Markov process after this transition belongs to the state $T_i(f'_i, g'_i)$, where

$$f'_i = \min(f_i + q, i) = \begin{cases} f_i + q & (q \leq i - f_i), \\ i & (q \geq i - f_i), \end{cases} \tag{4.2}$$

$$g'_i = g_i + \max(q - (i - f_i), 0) = \begin{cases} g_i & (q \leq i - f_i), \\ f_i + g_i + q - i & (q \geq i - f_i), \end{cases} \tag{4.3}$$

and the transition rate in (4.1) has the same value λ for every $x \in T_i(f_i, g_i)$.

If $f_i + g_i (= f_m) \neq 0$, then transitions from a state $x \in T_i(f_i, g_i)$ due to retrieval of items are possible into states in $T_i(f_i - 1, g_i)$ and $T_i(f_i, g_i - 1)$ whose transition rates

$$p(x, T_i(f_i - 1, g_i)) = f_i \mu / f_m \text{ and } p(x, T_i(f_i, g_i - 1)) = g_i \mu / f_m \tag{4.4}$$

are the same for every $x \in T_i(f_i, g_i)$ and the sum is equal to μ .

From the above discussion the stochastic process $\{(F_i(t), G_i(t))\}$ has been proved to be Markovian. We will denote by $P'_i(f_i, g_i)$ the steady state joint probabilities for the inventory level of slots $1 \sim i$ and that of slots $i + 1 \sim m$. Then we can derive the system of equilibrium equations for $P'_i(f_i, g_i)$ using the transition rates obtained above. It will be as follows:

$$t_{i,m} \mu P'_i(f_i, g_i + 1) + \sum_{j=0}^{g_i} \lambda P'_i(i - q + g_i - j, j) - \mu P'_i(f_i, g_i) = 0 \quad (f_i = i), \tag{4.5.1}$$

$$s_{i,m} \mu P'_i(f_i + 1, g_i) + t_{i,m} \mu P'_i(f_i, g_i + 1) + \lambda P'_i(f_i - q, g_i) - \mu P'_i(f_i, g_i) = 0 \quad (q \leq f_i < i), \tag{4.5.2}$$

$$s_{i,m} \mu P'_i(f_i + 1, g_i) + t_{i,m} \mu P'_i(f_i, g_i + 1) - \mu P'_i(f_i, g_i) = 0 \quad (f_m > r, f_i < q), \tag{4.5.3}$$

$$s_{i,m} \mu P'_i(f_i + 1, g_i) + t_{i,m} \mu P'_i(f_i, g_i + 1) - (\mu + \lambda) P'_i(f_i, g_i) = 0 \quad (1 \leq f_m \leq r), \tag{4.5.4}$$

$$\mu P'_i(1, 0) + \mu P'_i(0, 1) - \lambda P'_i(0, 0) = 0 \quad (f_m = 0), \tag{4.5.5}$$

$$\sum_{f_i} \sum_{g_i} P'_i(f_i, g_i) = 1, \tag{4.5.6}$$

where

$$s_{i,m} = (f_i + 1) / (f_m + 1), \quad t_{i,m} = (g_i + 1) / (f_m + 1),$$

and $P'_i(*, *)$'s are assumed to be zero for undefined states. The state transition diagram for the states (f_i, g_i) is shown in Fig. 5. Since each of the equations of (4.5.1) ~ (4.5.6) can be derived in almost the same way, we will explain only the first equation which is for $f_i = i$. We will consider the transitions into state (i, g_i) as well as the transitions from the state. The transitions into state (i, g_i) are possible by both retrieval and receipt of items. Since the

first i slots are full after transition, retrieval must be from one of the slots $i + 1, i + 2, \dots, m$. This means that a transition is from state $(f_i, g_i + 1)$ of the aggregated Markov process whose transition rate is $t_{i,m} = (g_i + 1)/(f_m + 1)$ from the second equation of (4.4) which correspond to the first term. The transitions into state (i, g_i) due to receipt of items are possible from a set of aggregated states, i.e., $(i - q + g_i, 0), (i - q + g_i + 1, 1), \dots, (i - q + 1, g_i - 1)$, and $(i - q, g_i)$ according to equations (4.2) and (4.3). Then we will get the second term of (4.5.1) because each transition rate is λ . Since $i > q$, transitions from the state are possible only by retrieval of items and the sum of these transition rates is μ as was observed above which corresponds to the last term.

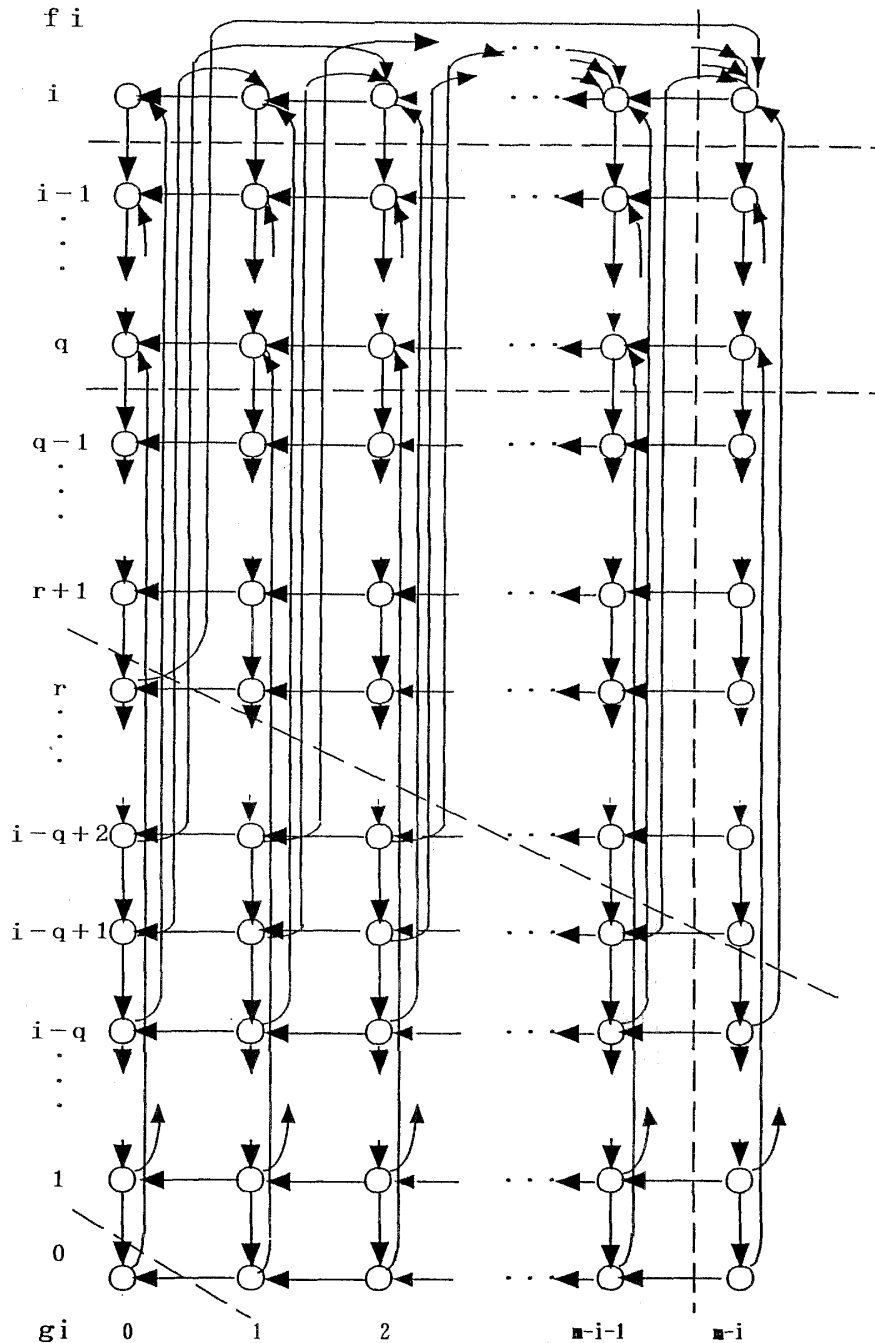


Figure 5: State Transition Diagram for States (f_i, g_i) .

4.2. Solution Method for $Pr(X_i(t) = 1)$ ($i = q + 1, \dots, m - 1$)

If the solution of the system of equilibrium equations (4.5.1) ~ (4.5.6) is obtained, we can get the expectation of F_i by

$$E[F_i] = \sum_{f_i} \sum_{g_i} f_i P'_i(f_i, g_i) \quad (i = q + 1, q + 2, \dots, m - 1). \tag{4.6}$$

Then we can obtain the marginal probability that slot i is full by

$$Pr(X_i(t) = 1) = E[X_i] = E[F_i] - E[F_{i-1}] \quad (i = q + 1, \dots, m - 1), \tag{4.7}$$

where $E[F_q]$ can be obtained by

$$E[F_q] = E[X_1 + X_2 + \dots + X_q] = E[X_1]q = Pr(X_1(t) = 1)q.$$

$Pr(X_1(t) = 1)$ is, of course, obtained by (3.5) after solving Equations (3.3) using the method explained in Section 3.3.

(4.5.1) ~ (4.5.6) is a system of linear equilibrium equations with $O(m^2)$ unknown variables: More precisely the number is less than $(m^2 + 4m + 4)/4$. Now, let us show an algorithm for solving it in $O(m^3)$ time by utilizing its structure. First we will reduce the number of variables to $m - i + 2$ by expressing each of $P'_i(f_i, g_i)$ ($f_i = 0, \dots, i, g_i = 0, \dots, m - i$) as a linear combination of $P'_i(q, k)$ ($k = 0, \dots, m - i$), i.e.,

$$P'_i(f_i, g_i) = \sum_k C_{f_i, g_i, k} P'_i(q, k), \tag{4.8}$$

where obviously

$$C_{q, g_i, k} = \begin{cases} 1 & (g_i = k), \\ 0 & (\text{otherwise}), \end{cases} \quad (k = 0, 1, \dots, m - i) \tag{4.9}$$

holds and the rest of the coefficients are obtained recursively to be explained below. To simplify the notation, we use $C_{f, g, k}$ instead of $C_{f_i, g_i, k}$, s and t instead of $s_{i, m}$ and $t_{i, m}$. The coefficient $C_{f, g, k}$ can be obtained from (4.5.1) ~ (4.5.5) as follows: First suppose that $P'_i(q, k)$ ($k = 0, \dots, m - i$) are known for the time being, then $P'_i(q - 1, g_i)$ can be expressed in terms of $P'_i(q, g_i)$ using (4.5.3) for $g_i = m - i, m - i - 1, \dots, 0$ and the coefficients $C_{f, g, k}$ will be obtained as in (4.10.1). The same relation holds for $r < f + g$ and $f < q$. The coefficients are obtained in the descending order of f and k until $f = 0$ yielding (4.10.2) and (4.10.3). Note that the coefficients on the right hand sides of (4.10.1) through (4.10.5) are all known by the time evaluation of the coefficients on left hand sides to be carried out. Next the coefficients $C_{f, g, k}$ are obtained for $f = i$ and $q < f < i$ yielding (4.10.4) and (4.10.5) which correspond to (4.5.1) and (4.5.2).

$$C_{f, g, k} = sC_{f+1, g, k} + tC_{f, g+1, k} \quad (r < f + g, f < q), \tag{4.10.1}$$

$$C_{f, g, k} = \frac{\mu}{\lambda + \mu} (sC_{f+1, g, k} + tC_{f, g+1, k}) \quad (1 \leq f + g \leq r), \tag{4.10.2}$$

$$C_{0, 0, k} = \frac{\mu}{\lambda} (C_{1, 0, k} + C_{0, 1, k}) \quad (f + g = 0), \tag{4.10.3}$$

$$C_{f, g, k} = tC_{f, g+1, k} + \frac{\lambda}{\mu} \sum_{j=0} C_{f-q+g-j, j, k} \quad (f = i), \tag{4.10.4}$$

$$C_{f, g, k} = sC_{f+1, g, k} + tC_{f, g+1, k} + \frac{\lambda}{\mu} C_{f-q, g, k} \quad (q \leq f < i). \tag{4.10.5}$$

As (4.10.1) through (4.10.5) are the recursive formulae with the initial terms $C_{q,g,k}$ given by (4.9), we can obtain all $C_{f,g,k}$. We then construct a system of linear equilibrium equations for $P'_i(q, g_i)$ by equating $P'_i(q, g_i)$ to the corresponding expression obtained from (4.10.5) with $f_i = q$ for $g_i = 0, 1, \dots, m - i$ as follows:

$$\begin{cases} P'_i(q, g_i) = \sum_k (v_i C_{q+1,g,k} + w_i C_{q,g+1,k} + \frac{\lambda}{\mu} C_{0,g,k}) P'_i(q, k) \\ \sum_f \sum_g \sum_k C_{f,g,k} P'_i(q, k) = 1, \end{cases} \quad (g_i = 0, \dots, m - i), \quad (4.11)$$

where

$$v_i = (q + 1)/(q + g_i + 1), \quad w_i = (g_i + 1)/(q + g_i + 1),$$

and the second equation of (4.11) is obtained from (4.5.6).

(4.11) is a system of linear equations with $O(m)$ unknown variables. If we solve it by a standard numerical method and substitute the solution to (4.8), we can get $P'_i(f_i, g_i)$ ($f_i = 0, \dots, i, g_i = 0, \dots, m - i$). Thus we can calculate $E[F_i]$ by substituting them to (4.6). Finally the marginal probabilities $Pr(X_i(t) = 1)$ for $i = q + 1, q + 2, \dots, m$ are given by (4.7).

Now we consider the time complexity. For each slot i ($i = q + 1, \dots, m - 1$) the coefficients $C_{f,g,k}$ are obtained in $O(m^3)$ time by (4.10.1)~(4.10.5). A standard numerical method requires $O(m^3)$ time for solving (4.11). Then we can get $P'_i(f_i, g_i)$ in $O(m^3)$ time by (4.8), $E[F_i]$ in $O(m^2)$ time by (4.6), and $Pr(X_i(t) = 1)$ in constant time by (4.7). Thus we can obtain the marginal probability that each slot is full in $O(m^3)$ time. Since there are m slots in all our algorithm enables us to obtain the inventory distribution in $O(m^4)$ time.

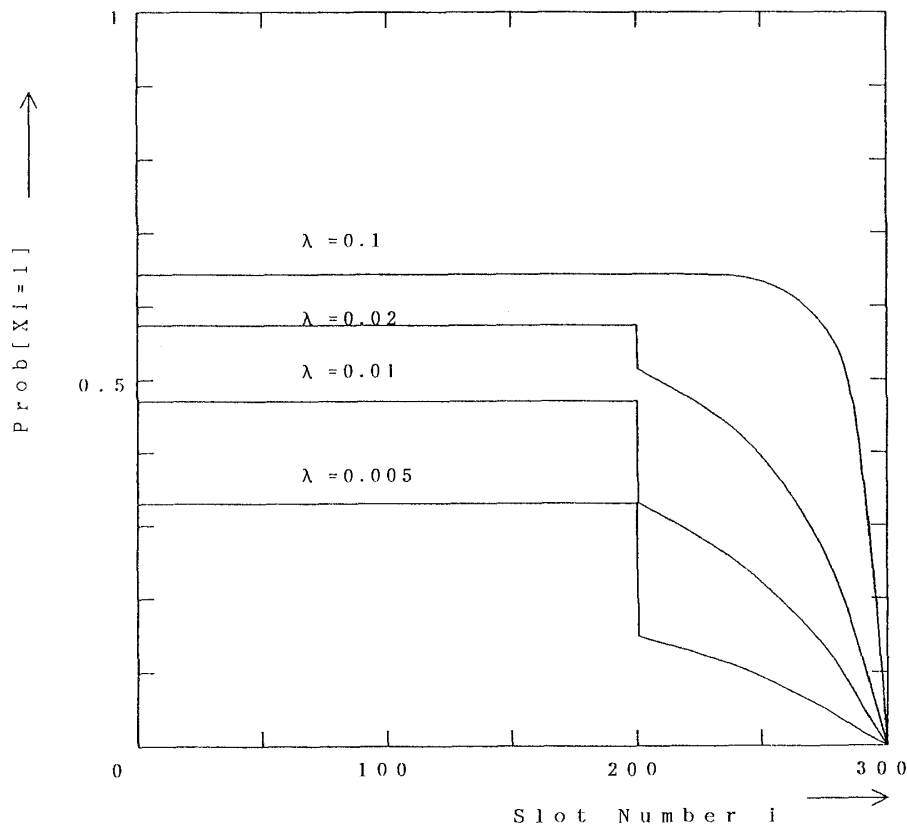


Figure 6: Inventory Distribution ($m = 300, r = 100, q = 200, \mu = 1.0$).

5. Numerical Experiments

We have computed the inventory distribution for some examples. In our examples the number of slots m is 300, and the reordering point r is 100, so that the reordering quantity q is 200. In this case the number of variables of the equilibrium equations (2.2) is more than 2.0×10^{90} . On the other hand, the equilibrium equations for $P'_i(f_i, g_i)$ have at most 20200 unknown variables, and (4.11) which we solve indeed has only at most 100 unknown variables. We have fixed μ at 1.0, and varied λ from 0.1 to 0.005. When $\lambda = 0.1$ for example, the mean time between placing an order and receipt of the items is ten times as long as the mean time between retrievals, i.e., when the reordered items arrive at the system, f_m is 90 on the average.

The results are shown in Fig. 6. As is expected, the following relations hold in the numerical results.

$$Pr(X_i(t) = 1) = Pr(X_{i+1}(t) = 1) \quad (i = 1, \dots, q - 1),$$

$$Pr(X_i(t) = 1) > Pr(X_{i+1}(t) = 1) \quad (i = q, \dots, m - 1),$$

and the probability that each slot is full increases as the arrival rate of items λ increases. We can also observe the great difference between $Pr[x_{200}]$ and $Pr[x_{201}]$ at $\lambda = 0.005$ in particular. It means that the system often becomes almost empty by the time reordered items arrive at the system.

6. Conclusions

We consider a simple automatic warehousing system in which only one type of items are stored and retrieved, and focus our efforts on deriving the inventory distribution. This is one of the key factors in designing an automatic warehousing system. We have assumed that inventories were controlled by an (s, S) reordering policy, and that received items are stored from the closest open slot to the I/O point and a retrieved pallet with items is chosen randomly among currently full slots. Under further assumptions on arrival of items and request for retrieval of items the system can be modeled as a Markov chain. If we use joint inventory levels of items, the number of the states of the Markov chain amounts to 2^m , where m is the total number of slots. m easily becomes several hundreds to a thousand, so that trying to solve the equilibrium equations of the Markov chain directly is not practical for real systems. So we devised exact aggregation methods for the states. The number of variables in the system of the equilibrium equations of the Markov chain has been reduced to $O(m^2)$ or less by the aggregation for each slot. By solving m such systems we could obtain the inventory distribution. We also developed an efficient solution method which required $O(m^3)$ time for each system of equilibrium equations by making use of the special structure of the system. Thus the time complexity for obtaining the inventory distribution has been reduced to $O(m^4)$. This enables us to get the exact inventory distribution for real systems even though our model has simple assumptions such as only one type of items can be treated.

In order to make our model more realistic we would like to extend or apply our analysis to the following cases:

1. various types of items are treated in the system,
2. different storage methods such as class-based storage are employed,
3. waiting times of items before storage and retrieval are taken into account.

We believe that the method presented in this paper is a good starting point to the extensions mentioned above.

Acknowledgments

The authors are grateful to the anonymous referees for their helpful comments.

References

- [1] Y.A.Bozer and J.A.White: Travel-time models for automated storage/ retrieval systems. *IIE Transactions*, **16** (1984) 329-338.
- [2] D.-T.Chang, U.-P.Wen, and J.T.Lin: The impact of acceleration/deceleration on travel-time models for automated storage/retrieval systems. *IIE Transactions*, **27** (1995) 108-111.
- [3] S.C.Graves, W.H.Hausman, and L.B.Schwarz: Storage-retrieval interleaving in automatic warehousing systems. *Management Science*, **23** (1977) 935-945.
- [4] W.H.Hausman, L.B.Schwarz, and S.C.Graves: Optimal storage assignment in automatic warehousing systems. *Management Science*, **22** (1976) 629-638.
- [5] H.Hwang and S.B.Lee: Travel-time models considering the operating characteristics of the storage and retrieval machine. *International Journal of Production Research*, **28** (1990) 1779-1789.
- [6] J.G.Kemeny and J.L.Snell: *Finite Markov Chains* (Springer-Verlag, New York, 1976).
- [7] A.Noguchi and S.Suzuki: An approximate solution method for obtaining the travel time distribution of a crane in an automatic warehouse. *Transactions of the Japan Society of Mechanical Engineers*, **54-C** (1988) 798-804, in Japanese.

Hideaki Yamashita
Faculty of Economics
Tohoku University
Aoba-ku, Sendai, 980-8576, Japan.
E-mail: hideaki@econ.tohoku.ac.jp