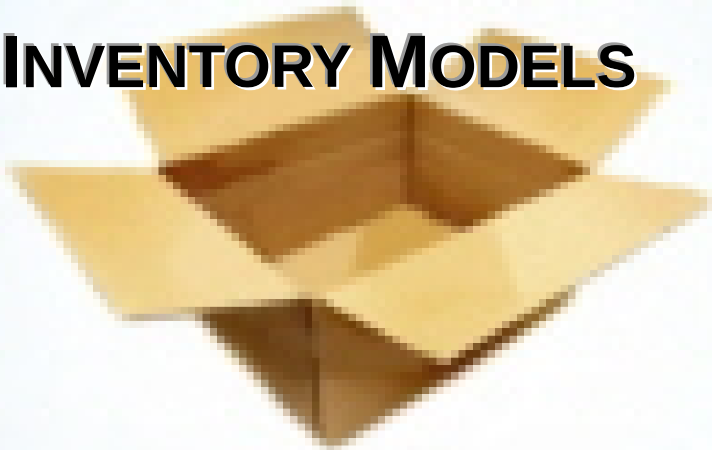
LATERAL TRANSSHIPMENTS

IN

INVENTORY MODELS



MIN CHEN

11.2008

BMI THESIS

Lateral Transshipments in Inventory Models

MIN CHEN

11.2008

vrije Universiteit amsterdam

PREFACE

The BMI Thesis is the final report to acquire the Master's degree in Business Mathematics and Informatics. Business Mathematics and Informatics (BMI) is a multi-disciplinary master program that is supplying the students with the integrated knowledge of Business Economics, Mathematics and Computer Science principles. It teaches students to solve the management and operational problems in the industry with a quantitative background.

This BMI thesis starts with the brief introduction of lateral transshipments. Different kinds of lateral transshipment policies are studied by the researchers. Based on the lateral transshipment models, we are interested in recognizing how the different models improve the service level and cut the total cost by applying the lateral transshipment policy.

After reading this thesis, one is able to know the details about lateral transshipment policies, and how these lateral transshipment policies maximize customer satisfaction and reduce management costs in inventory systems.

I would like to express my sincere gratitude to my supervisor Marco Bijvank for his guidance, encouragement, constructive critiques, enthusiasm and patience throughout this research. He has been a great mentor and guide to me. This thesis would not have been possible without him.

A few words of thanks seem inadequate to express my appreciation for those who have supported and helped me over the last two years during my study at VU University Amsterdam, The Netherlands.

Min Chen
Amstelveen, 2008

SUMMARY

This paper focuses on the literature study on lateral transshipment policies in inventory models. Emergency lateral transshipments combined with different types of replenishment policies can help to improve the service level of inventory systems. The problem of how to determine the optimal stocking levels under certain service level constraints is also presented in this paper. Furthermore, preventive lateral transshipments appear to be more effective in reducing the risk of a stock out before the demand is arriving. The new lateral transshipment policy (also referred to as service level adjustment) combines emergency and preventive lateral transshipments to efficiently respond to customer demands. The algorithm how a service level adjustment policy can be applied in an inventory system is illustrated and the relevant system cost is derived.

The objective of this paper is to give a complete overview of the different types of lateral transshipment policies in an inventory system, and how they reduce the cost and improve the service level.

We consider the use of a periodic review system for this study, but additional studies of lateral transshipments are needed to include other characteristics of the supply chain as well. For example, we have assumed that the shipment times from the depot to the bases are the same for all bases. Future work should relax this assumption, and incorporate more realistic and accurate factors.

CONTENTS

Chapter 1 Introduction

In an inventory system, spare parts supply chains often consist of multiple local warehouses and one or a few central warehouses. Local warehouses, located close to customers, are used to ensure prompt replenishments of spare parts to customers. The central warehouse(s) replenishes the local warehouse, but they can also perform an emergency delivery to the customers when a local warehouse runs out of stock. Since emergency deliveries from one central warehouse to a customer cost a lot of time and money, alternative solutions are common in practice. In this paper, we consider the alternative of ***lateral transshipments***. Lateral transshipment is defined as *a local warehouse which provides stocked items to another local warehouse which is out of stock or to prevent out-of-stock occurrences. In other words, these local warehouses exchange their inventory on the same echelon level*. Different local warehouses can operate individually and rely on regular or emergency replenishments from the central warehouse when they are out of stock. It may however be more useful to have a quicker backup from other local warehouses as well. Lateral transshipments can be seen a form of pooling. Physically, there are multiple stock points, but they have access to each other's inventory when needed. Cost savings can be expected because of this pooling. In case a local warehouse is out of stock at the moment of a customer's request, it first tries to obtain the required parts from a neighboring local warehouse. This costs less time than an emergency delivery from the central warehouse.

In many industries and service organizations, the reliance on two-echelon inventory systems for repairing and supplying recoverable items is becoming more and more prevalent. This paper focuses on discussing the two-echelon inventory system involving a central warehouse (or supplier) and multiple local warehouses (or retailers) with lateral transshipments as an option under the various inventory replenishment policies. The model is depicted in Figure 1.

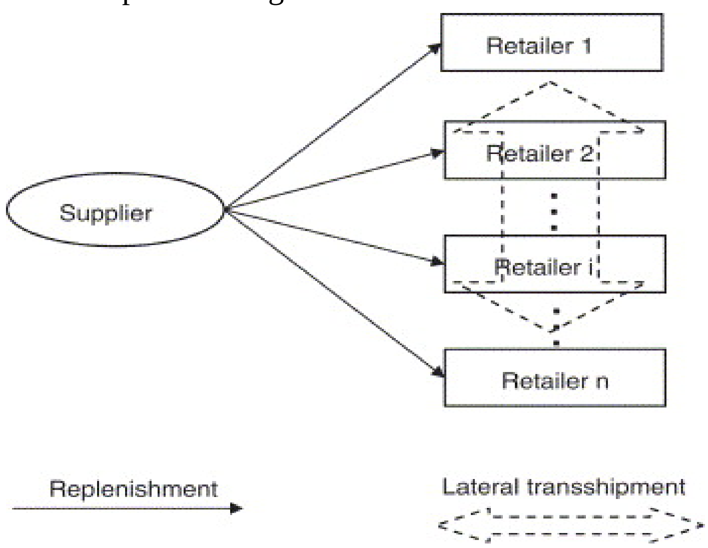


Figure 1: Lateral transshipment in inventory system.

1.1 Research objective

Lateral transshipments can be divided into **emergency lateral transshipments (ELTs)**, **preventive lateral transshipments (PLTs)**, and **service level adjustments (SLAs)**. ELT mandates emergency redistribution from a retailer with ample stock to a retailer that has reached a stock out. PLT reduces this risk by redistributing stock to prevent a stock out at a retailer before demand exceeds the inventory level. In other words, ELT responds to actual stock outs while PLT reduces the risk of possible future stock outs. The SLA policy combines the ELT and PLT policies together to reduce the risk of stock outs in advance and efficiently respond to actual stock outs. In this paper, we give a complete overview on how PLT, ELT and SLA policies play an important role in improving service levels and reducing inventory cost. The following research objectives are studied:

- To improve service levels and reduce inventory cost with emergency lateral transshipments in a system with one-for-one $(S-1, S)$ and (r, q) replenishment policies.

- To achieve equal marginal cost over all retailers just before the replenishment period with preventive lateral transshipments.

- To efficiently respond to customer demands with a service level adjustment policy by integrating emergency and preventive lateral transshipments.

The remainder of this paper is organized as follows. In Chapter 2, emergency lateral transshipments with two different replenishment policies are discussed. The mathematical formulations to calculate the expected number of backorders and the total system cost are developed for each replenishment policy. Next, the optimal stocking levels are determined based on service level constraints. In Chapter 3, we give an overview and examination of the effectiveness for preventive lateral transshipments. We propose service level adjustments using the service level in the remaining period (SLRP) as the service level in Chapter 4. Finally, we compare the three lateral transshipment policies in Chapter 5, and give our conclusions and the directions for further research.

Chapter 2 Emergency Lateral Transshipment (ELT)

2.1 Introduction

Multi-echelon inventory systems are usually used to provide service support for products where customers are distributed over an extensive geographical region. These systems can be characterized by lower echelons (bases) that serve as a first level of product support, and a higher echelon, consisting of a depot that serves as a second level of support in case the bases are not able to meet the customers' demand. The depot can also serve as a distribution center for the replenishment of stock to the bases. When stock is shipped from one base to another when it faces an out of stock occurrence it is called an emergency lateral transshipment.

There is a lot of literature that describes the emergency lateral transshipment policy in inventory models. See for example, Lee [1], Axsäter [2], Sherbroke [3], Alfrodsson and Verrijdt [4], Grahovac and Chakravarty [5], Kukreja et al. [6] and Wong et al. [7]. These papers focus on a one-for-one $(S-1, S)$ replenishment policy. Such policies are defined as policies, in which a replenishment order is placed as soon as the customer withdraws an item. S is called the base stock level. It is usually applied in supply chain systems in which the manufacturer supplies expensive, low-demand items to vertically integrated or autonomous retailers via a central depot.

Another type of replenishment policy is the (r, q) replenishment policy which is considered as a batch ordering policy, where r is the reorder level and q the replenishment quantity. It means that an order of a fixed size q is placed whenever the inventory position reaches the reorder point r. This policy is a very common and relatively straightforward policy for an inventory system. For continuous review (r, q) policy, Needham and Evers [8], Evers [9], Xu, et al .[10], Axsäter [11] and Huo [12] study how to make transshipment decisions when replenishment parameters (such as the order point and order size) are given. Table 1 presents a brief review of the previous studies on multi-echelon inventory systems with an emergency lateral transshipment policy.

The objective of this chapter is to determine the optimal stocking levels at each echelon based on service performance measures (e.g. expected shortages) for a multi-echelon system with emergency lateral transshipments.

Paper	# Echelons	# Items	Review	Policy
Lee (1987)	Two	Single	Continuous	$(S-1,S)$
Tagaras (1989)	Single	Single	Periodic	Order-up-to
Axsäter (1990)	Two	Single	Continuous	$(S-1,S)$
Tagaras and Cohen (1992)	Single	Single	Periodic	Order-up-to
Yanagi and Sasaki (1992)	Single	Single	Continuous	$(S-1,S)$
Sherbrooke (1992)	Two	Single	Continuous	$(S-1,S)$
Archibald et al. (1997)	Single	Multi	Periodic	Order-up-to
Needham and Evers (1998)	Two	Single	Continuous	(r,q)
Tagaras (1999)	Single	Single	Periodic	Order-up-to
Alfredsson and Verrijdt (1999)	Two	Single	Continuous	$(S-1,S)$
Grahovac and Chakravarty (2001)	Two	Single	Continuous	$(S-1,S)$
Kukreja et al. (2001)	Single	Single	Continuous	$(S-1,S)$
Evers (2001)	Single	Single	Continuous	(r,q)
Herer et al.(2002)	Single	Single	Periodic	Order-up-to
Xu et al. (2003)	Single	Single	Continuous	(r,q)
Axsäter (2003)	Single	Single	Continuous	(r,q)
Wong et al. (2005a)	Single	Single	Continuous	$(S-1,S)$

Table 1: A brief review of the literature on the multi-echelon inventory systems

2.2 Emergency lateral transshipment with $(S-1,S)$ replenishment policy

Emergency lateral transshipments from the same echelon level are one of the popular tools to handle stock outs at local warehouses. Since the cost of a lateral transshipment is generally lower than the shortage cost and the cost of an emergency delivery from the depot. Therefore emergency lateral transshipments reduce the total system cost. Simultaneously, the transshipment time is shorter than the regular replenishment lead time and therefore they increase the fill rate at the local warehouses. It is used by lots of companies to form a pooling group between retailers to reduce the risk of shortages and provide better service at lower cost. For repairable items with low demand volume and high value, a one-for-one $(S-1,S)$ replenishment policy is commonly used in modeling the multi-echelon inventory system.

2.2.1 Assumptions and notations

Consider a multi-echelon inventory system for repairable items with one depot (or central warehouse) and M bases (or local warehouses) which are grouped into n disjoint groups. Each of these groups is identical and shares the same pooling stocks. Let m_i be the number of identical bases in pooling group i. A one-for-one $(S-1, S)$ replenishment policy is used and emergency lateral transshipments between identical bases are allowed.

If one of the bases can not satisfy a customer demand with its stock on hand, the emergency lateral transshipment policy will be applied to fill the demand from another base in the same group which has enough stock on hand. If the emergency lateral transshipment is impossible due to group-wide stock out, the demand is backordered. For the situation of more than one base with stock on hand, there are three prioritized sourcing rules [16] to determine the supplier for the emergency lateral transshipments. The three prioritized rules are random choice, choice based on maximum stock on hand and choice based on the smallest number of outstanding orders. In this model, we assume the random choice as the main prioritized rule for emergency lateral transshipments.

The following notations and assumptions are used in this chapter:

- the demand rates at the bases in the same pooling group are identical,
- all failed items can be repaired in the depot which is assumed to have an infinite number of servers. It is according to an exponential distribution with mean repair time $(1/\mu)$ for a failed item,
- orders are coming as first-in first-out (FIFO) at both depot and bases,
- the transportation time for a lateral transshipment to pooling group i is deterministic and denoted as T_i for a pooling group i,
- demand at each base of pooling group i is according to a Poisson distribution with rate λ_i. So, $\lambda = \sum_{i=1}^{n} m_i \lambda_i$ represents the total arrival rate to all bases and the depot,
- the emergency lateral transshipment times are substantially lower than the normal resupply time. Otherwise, the emergency transshipment will not be chosen,
- a_i is the unit cost for a lateral transshipment in pooling group i,

- N_i is the average number of emergency lateral transshipments per unit time at pooling group i.

The model can be seen in Figure 2.

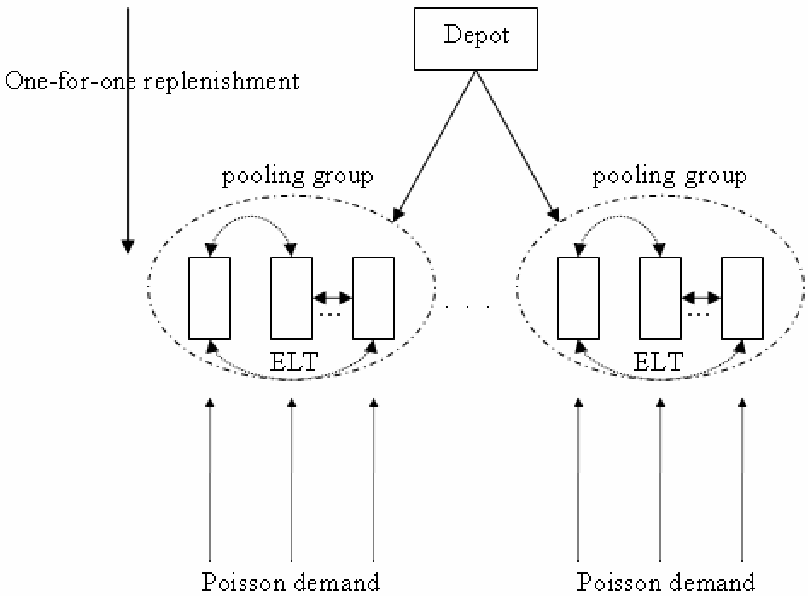


Figure 2: Emergency lateral transshipment with one-for-one replenishment policy

2.2.2 Model formulation and specification

In this model, we are interested in calculating the expected number of backorders and the quantity of emergency lateral transshipments in order to determine the optimal inventory stocking levels S in the multi-echelon inventory systems. The measurement of the models performance will be according to the expected total number of backorders, the total system cost and the improvement of the customer service level.

According to Lee's theory [1], the expected number of emergency lateral transshipments N_i per unit time at pooling group i is equal to the proportion (α_i) of arriving demands met by emergency lateral transshipments multiplied by the total arrival rate $(m_i\lambda_i)$ at pooling group i:

$$N_i = m_i\lambda_i\alpha_i. \quad (2.1)$$

The total number of backorders in the system consists of backorders that are to be met by emergency lateral transshipments at pooling group i and to be met by the depot.

Following the well-known result of Palm (1938), the number of items on order at the depot is Poisson distributed with the parameter (λ/μ) as long as successive resupply times are independent. Hence, the expected number of backorders at the depot can be determined as (according to the theorem provided by Graves [17] and Axsäter [2]):

$$E(B_0) = E(Q_i - S_0)^+$$

$$= \sum_{k=S_0}^{\infty} (k - S_0)\exp(-\frac{\lambda}{\mu})\frac{(\lambda/\mu)^k}{k!} \quad , \tag{2.2}$$

where Q_i is the number of orders outstanding at pooling group i at time t, S_0 is the base stock level at the depot, and $[x]^+$ denotes $\max[0,x]$.

Combining Equation (2.2) with Little's formula, which provides the expected number of items in transit for emergency lateral transshipments as $\sum_i N_i T_i$, the total expected number of backorders $E(B_i)$ at pooling group i is derived by Lee [1] as:

$$E(B_i) = E(Q_i - m_i S_i)^+ + N_i T_i$$

$$= \sum_{k=m_i S_i}^{\infty} (k - m_i S_i)\exp(-\frac{m_i \lambda_i}{\mu})\frac{(m_i \lambda_i/\mu)^k}{k!} + N_i T_i \quad , \tag{2.3}$$

where S_i is the base-stock level at a base of pooling group $i,(i=1,...,n)$. The first part of Equation (2.3) expresses the expected number of backorders at pooling group i that are to be met by the depot. The second part is the expected number of items in transit for emergency lateral transshipments.

So, by adding Equation (2.2) for all pooling groups and Equation (2.3), we can get the total expected number of backorders involved in this system.

In this section, we evaluate the performance of inventory system in terms of the overall transportation, inventory holding, and backorder costs. Equation (2.4) expresses the total system cost per unit time (derived by Lee [1]):

$$C_{total} = h(S_0 + \sum_{i=1}^{n} m_i S_i) + v\sum_{i=1}^{n} E(B_i) + \sum_{i=1}^{n} a_i N_i \ , \tag{2.4}$$

where h is holding cost per unit time, v is the backorder cost per unit backordered.

The first summation represents all inventory holding costs. The second summation represents all waiting costs or backorder costs that customers transfer to the supply chain. The third summation is the total additional cost coming from the emergency shipments and transshipments in the inventory system.

2.2.3 Determination of optimal stocking levels

The key component in the design of multi-echelon inventory systems for recoverable items is the determination of the base-stock levels at each echelon. There is a two phase method to determine these optimal base-stock levels at the bases in each pooling group i and the depot (Lee [1]). The first phase is to find the value of S_i at each pooling group i to minimize the cost C_{total} in Equation (2.4) for each given S_0. The second phase is to optimize the value of S_0 at the depot. For a given S_0, the optimal base-stock level S_i at each pooling group i, is the one that minimizes Equation (2.5).

$$\begin{aligned} C_i(S_i \mid S_0) &= hm_iS_i + vE(B_i) + \alpha_i N_i \\ &= hm_iS_i + vE(Q_i - m_iS_i)^+ + (vT_i + a_i)m_i\lambda_i\alpha_i . \end{aligned} \quad (2.5)$$

The optimal cost C_i for pooling group i is denoted as $C_i^*(S_0)$ for a given value of S_0, where

$$C_i^*(S_0) = \min_{S_i} C_i(S_i \mid S_0) . \quad (2.6)$$

By solving Equation (2.6), we always find the best value of S_i which result in the optimal cost.

The second phase in the method is to look for the optimal value of S_0 at the depot, with the following proposition as stopping criterion (from Lee [1]):

Suppose $\overline{S_0}$ *is such that*

$$h \geq \sum_{i=1}^{n} C_i^*(\overline{S_0}) - \sum_{i=1}^{n} C_i^*(S_0 \to \infty) . \quad (2.7)$$

Then all $S_0 > \overline{S_0}$ *can not be optimal.*

2.3 Emergency lateral transshipment with (r,q) replenishment policy

The (r,q) inventory replenishment policy is a generalization of the one-for-one or base-stock policy, and represents a more realistic model of a common manufacturing problem faced by management. Comparing to the one-for-one stock policy, the (r,q) replenishment policy is optimal when there are costs involved for every orders. The one-for-one policy is also costly when there is a shipping cost associated with each replenishment shipment to the inventory. Emergency lateral transshipments with an (r,q) replenishment policy can result in much lower costs and high service levels. In this section, we focus on a multi-echelon inventory system without pooling groups.

2.3.1 Assumptions and notations

Consider a multi-echelon model with a depot (central warehouse) and M bases (local warehouses). Spare parts are stored in different locations in the system. The demand at each base is according to a Poisson distribution with demand rate λ_i at each base i. The (r,q) replenishment policy is used. Emergency lateral transshipments between each base at a lower echelon level are allowed. When there is more than one base available for the emergency lateral transshipments, the base performing the transshipment is assumed to be chosen randomly. In other words, if there are no parts available at one base, the part is ordered with an emergency lateral transshipment from another randomly chosen base which has stock on hand.

The following notations and assumptions will be used:

- the average lead time for a replenishment from the depot to base i is denoted as L_i,
- $D(\lambda)$ denotes as lead time demand, which is a random variable following as a Poisson distribution with mean λL_i,
- the order quantity is denoted by q_i for base i,
- the reorder point at base i is represented by r_i for the (r,q) replenishment policy,

- the number of bases is m,
- the required high service level is ξ,
- all the inventory bases are the same.

The model can be seen in Figure 3.

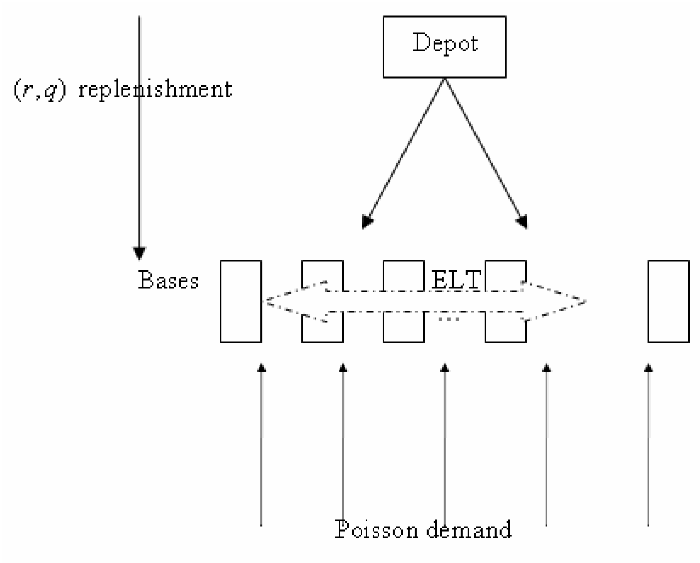


Figure 3: Emergency lateral transshipment with (r,q) replenishment policy.

2.3.2 The analytical model

The decision rule in this model is to find the reorder points at all locations that minimize the sum of inventory holding and transshipment costs.

With emergency lateral transshipments, the demand at base i can be satisfied under three conditions. The demand at base i is met by an emergency lateral transshipment with probability α_i, demand at base i is met immediately by stock on hand with probability β_i, or the demand is met by a backorder with probability θ_i. Evidently,

$$\alpha_i + \beta_i + \theta_i = 1. \tag{2.8}$$

Each base serves the customers of its designated territory and controls its inventory level independently and continuously with a reorder-point, fixed order-quantity (r,q) policy. The base places a replenishment order of size q at the depot or at another base which has ample capacity to fulfill all replenishment requests within a deterministic replenishment lead time.

As defined, base i has a positive on–hand inventory level during the proportion β_i of the time, and no inventory on hand during the remaining proportion $1-\beta_i$ of the time. In other words, when the inventory is positive, the base is facing the Poisson demand λ_i plus a demand from other bases due to emergency transshipments with an average rate $\alpha_i \lambda_i / \beta_i$. When the base has no inventory on hand, the demand that has to be backordered comes in with rate $\alpha_i \lambda_i / (1-\beta_i)$ which is the rate λ_i minus the demand satisfied by the other bases due to emergency transshipments.

Thus, the demand rate at base i when it has positive inventory on hand is denoted as:

$$g_i = \lambda_i (1+\alpha_i / \beta_i) = \lambda_i (1-\theta_i) / \beta_i . \tag{2.9}$$

Similarly, the demand rate at base i when it has negative inventory on hand is denoted as:

$$\eta_i = \lambda_i (1-\alpha_i / (1-\beta_i)) = \lambda_i \theta_i / (1-\beta_i) , \tag{2.10}$$

So, the expected number of backorders $E(B_i)$ at a base i is expressed as Equation (2.11) from [14] when there is no positive on-hand inventory level:

$$E(B_i) = \frac{1}{q_i} \sum_{x=r_i+1}^{r_i+q_i} \left[\left(\sum_{k=x}^{\infty} \frac{(k-x)(\eta_i L_i)^k}{k!} \right) \Big/ \left(\sum_{k=x}^{\infty} \frac{(\eta_i L_i)^k}{k!} \right) \right] . \tag{2.11}$$

According to Axsäter’s theory [13] and Huo’s theory [14], the reasonable approximation for the expected number of backorders in the whole system consisting of the expected number of backorders in depot $E(B_0)$ and in base i, $E(B_i)$, is determined by Equation (2.12):

$$E(B) = E(B_0) + E(B_i) = (M-1)E(B_i) . \tag{2.12}$$

Axsäter’s and Huo’s methods to approximate the number of backorders in the whole system $E(B)$ indicates a situation in which the system is just about to leave or has just reached a state where no individual base has positive stock on hand. This means that one of the bases has no stock on hand and no backorders. Some of the other $M-1$ bases may have backorders, though, and a reasonable approximation for the total expected number of backorders in the whole system is according to Equation (2.12).

2.3.3 Optimizing reorder point

The objective of this section is to find the optimal reorder points for all locations such that the total system cost is minimized subject to a certain service level constraint ξ. The total system cost consists of inventory holding, and backorder costs. The average on-hand inventory level at base i is denoted as $G_i(x)$ when the current inventory position is x. It is determined according to Equation (2.13) from [14]

$$G_i(x) = E\left[(D_i(\lambda) - x)^- \right]$$

$$= \sum_{k=0}^{x-1} (x-k)\exp(-\lambda L_i)\frac{(\lambda L_i)^k}{k!} \quad , \tag{2.13}$$

where $D_i(\lambda)$ denotes the lead time demand, which is a random variable that follows a Poisson distribution with mean λL_i at base i.

Equation (2.14) expresses the total system cost per unit of time (from [14]).

$$C_{total} = Mh\frac{1}{q_i}\sum_{x=r_i+1}^{r_i+q_i} G_i(x) + M\alpha_i\lambda_i a \, , \tag{2.15}$$

where h is the holding cost, a is the lateral transshipments cost per unit. According to Axsäter’s theory [13], we assume the order quantity q_i is determined in advance for each base i in order to solve Equation (2.15). After that, we can always find the best reordering points r_i that minimize the total system cost subject to a certain service level constraint ξ in this system.

Chapter 3 Preventive Lateral Transshipment (PLT)

3.1 Introduction

In this chapter we consider a pooling group consisting of multiple stocking locations at the same echelon which places regular orders at a depot according to periodic reviews with a base stock replenishment policy. The local warehouses of the pooling group may resort to lateral transshipments when there is a stock out, or when there appears to be significant risk of an imminent stock out in the near future. The decision whether to move inventories between these stock points in a pooling group to anticipate to a possible stock out is considered to be a preventive lateral transshipment (PLT).

There are two major redistribution policies towards making a preventive lateral transshipment decision at the same echelon level: transshipment based on availability (called a TBA policy) and transshipment based on inventory equalization (called a TIE policy). In this section, we will give an overview on how these two policies can be applied in inventory systems including their relevant system cost.

3.2 Assumptions and notations

Consider a two-echelon model with a depot (or supplier) at the higher echelon and M bases (or retailers) at the lower echelon. We assume that the supplier has unlimited inventory at the depot, and no other supply sources exist for the retailers at the bases. Although, several products are stocked at each base, we focus on a single item. Demand for the product at each base is stochastic and stationary. The delivery lead time from the depot to any bases is the same and deterministic. Lateral transshipments between any of the retail locations are possible and lead times are negligibly small.

The following assumptions and notations are used:

- expected daily demand during a review period at base i is λ_i,
- the length of the review period is R,
- supply lead time from depot to any base is L,
- the available inventory at base i at time t is $I_i(t)$,
- the unit holding cost per period at any base is h,
- shortage (or backorder) cost per unit and period is v,
- unit lateral transshipment cost between two bases is a,

- the lateral transshipments lead times are negligible.

At the beginning, we define the reorder point for a lateral transshipment to base i $(i = 1, 2, ..., M)$ in this system as:

$$S_i = \lambda_i - 1, \quad i = 1, 2, ..., M \ . \tag{3.1}$$

Since demand at each base is stationary, a transshipment order point can be easily implemented at each retail location. When the available inventory at one or more bases drops down to order level S_i, it indicates the likelihood of an impending shortage. So, the lateral transshipments can be applied for this situation to reduce the risk of a shortage. At a time t, we define a set of bases J, which require a lateral transshipment:

$$J = \{i \mid I_i(t) \le S_i\}, \quad i = 1, 2, ..., M \ . \tag{3.2}$$

The shortage at base j for $j \in J$ equals the expected needed transshipment quantity:

$$SH_j(t) = \lambda_j - I_j(t), \quad j \in J \ . \tag{3.3}$$

When a base is not included in J, it does not mean that items can be transshipped to those bases. Therefore, the safety stock level has to be determined for each base, which should be sufficient to fulfill the demand until the receipt of the next delivery. This is given by Equation (3.4) from [15]:

$$E[I_i(t)] = \lambda_i(\tau_{t,R} - t + R), \ i = 1, 2, ..., M \ . \tag{3.4}$$

where $\tau_{t,R}$ is the deterministically known scheduled time of receipt of the next shipment from the depot at higher echelon.

Thus, the bases which have excess inventory on hand can be used for making lateral transshipments to the bases which have a shortage. Therefore, we define a set of bases K, which have stock levels above the average stock level to fulfill demand:

$$K = \{i \mid I_i(t) > E[I_i(t)]\}, \quad i = 1, 2, ..., M \ . \tag{3.5}$$

The excess inventory for the base k for $k \in K$ can be used for lateral transshipments at time t:

$$A_k(t) = I_k(t) - E[I_k(t)], \quad k \in K \ . \tag{3.6}$$

Thus, the transshipments can be performed from any bases $k \in K$ to any bases $j \in J$.

3.3 Lateral transshipments based on availability

The lateral transshipment based on availability (TBA) policy is one type of preventive lateral transshipment in order to reduce the risk of a stock out. The objective of this policy is to redistribute stock to retailers with ample stock levels to fulfill customer demands. This policy can also be considered as a reactive policy. The difficulty is to determine the desired stock levels correctly such that future demand is satisfied.

We denote the time instance at which the available stock falls down to the transshipment order point at a base for the first time as t. At this time instance t the shortage $SH_j(t)$ for base j is calculated, as well as the excess stock $A_k(t)$ for the other bases.

The following procedure is used to perform the lateral transshipment:

1) Determine the set J and K of the bases with shortages and excess stock, respectively. Next, set the excess quantities $A_k(t)$ for base $k \in K$, and shortage quantities $SH_j(t)$ for base $j \in J$.
2) Rank the bases with excess stock according to the available quantities $A_k(t)$ in descending order, and rank the bases with shortages in descending order. Consequently, $A_{[1]}(t) \geq A_{[2]}(t) \geq ... \geq A_{[|K|]}(t)$, $SH_{[1]}(t) \geq SH_{[2]}(t) \geq ... \geq SH_{[|J|]}(t)$.
3) Next, the total shipment quantity at time t from the base with the largest excess stock to the base which needs the most replenishments is determined by Equation (3.7) from [15]:

$$Q_{[1][1]} = min\{A_{[1]}(t), SH_{[1]}(t)\}. \qquad (3.7)$$

4) The quantity of Equation (3.5) is reallocated according to step (3). Repeat step (1) to step (3) until $Q_{[1][1]} = 0$. As a result, either all transshipments are performed or the total amount of stock available for the transshipments is exhausted.

3.4 Lateral transshipments based on inventory equalization

The lateral transshipment based on inventory equalization (TIE) policy redistributes stocks to match the ratio of average demand of each retailer whenever there are retailers with less than desirable stock levels. This policy can be also considered as the proactive policy where transshipment decisions are based on the concept of inventory balancing or equalization through stock redistribution. The rationale behind this policy is to correctly determine the transshipment quantities in every base. The transshipped quantities must be those that equalize the probability of a stock out for every base during one review period.

During one review period, inventories are redistributed among the bases through lateral transshipments, such that all bases have an equal number of days' supply. According to the TIE policy, the expected equalized inventory level $E[I_i(t)]$ in Equation (3.4) at base i at time t can be further expressed as Equation (3.8) from [15].

$$E[I_i(t)] = \frac{\lambda_i}{\sum_{i=1}^{M} \lambda_i} \left[\sum_{i=1}^{M} I_i(t) \right], \tag{3.8}$$

where $I_i(t)$ is the available inventory at location i at time t.

For achieving inventory equalization at base i for $i \in M$; if $I_j(t) - E[I_j(t)] < 0$, the shortage at base j for $j \in J$ equals the expected needed transshipment quantity is calculated from Equation (3.3); if $I_k(t) - E[I_k(t)] > 0$, the excess inventory for the bases k for $k \in K$ can be used for lateral transshipments at time t as determined by Equation (3.6).

3.5 System performance analysis

The expected total system cost per period is adopted here as the criterion for evaluating the performance of the system. In general, the total cost consists of the transportation cost from the central depot, inventory holding cost h, backorder cost v and lateral transshipments cost a. As the transportation cost for regular supply from the central depot to all bases is the same, the expected total transportation cost is constant, independent of the base stocks and the transshipment policy, and can be disregarded. Thus, assume the system has reached steady state, the expected cost per period can be written as Equation (3.9) from [16].

$$C_{total} = \sum_{i=1}^{M} \left\{ h \cdot OH_i + v \cdot BO_i + \sum_{j=1}^{M} a \cdot X_{i,j} \right\}, \tag{3.9}$$

where OH_i , BO_i denote as the expected on-hand inventory and expected backorder level at base i, $X_{i,j}$ is the expected quantity transshipped from base i to j at the end of a period after redistribution, respectively.

The economically optimal solution to the problem of operating the inventory system under consideration is the combination of ordering and transshipment policies that minimizes the total expected cost per period C_{total}. The decision variable that completely determines the ordering policy is the order-up-to levels S_i.

At the beginning, in Equation (3.1), we assume the base stock level S_i equals to expected daily demand rate minus 1. But, this is not optimal base stock level in the inventory system. Suppose the base i faces random demand with mean λ_i and standard deviation σ_i per period. The demand is stationary, normally distributed and independent at other bases.

The cost C_{total} of an independent retailer i following an order-up-to S_i periodic review policy consists only of the holding, backorder and lateral transshipments costs can be expressed as Equation (3.9). Based on the assumption of normally distributed demand, the on-hand inventory OH_i and backorder level BO_i can be determined by Equation (3.10) and (3.11) from [16] respectively.

$$OH_i = \int_{y=0}^{S_i} (S_i - y) f(y) dy, \tag{3.10}$$

$$BO_i = \int_{y=S_i}^{\infty} (y - S_i) f(y) dy, \tag{3.11}$$

where y and $f(y)$ are the demand during $L_i + 1$ periods and its normal density function, respectively.

The expected quantity $X_{i,j}$ transshipped from base i to j at the end of a period after redistribution is expressed as Equation (3.12) from [15].

$$X_{i,j} = min\{max(A_i(t)), max(SH_j(t))\}. \tag{3.12}$$

In Equation (3.12), the quantity $X_{i,j}$ at time t is transshipped from the base i with the largest excess stock to the base j which needs the most replenishment.

Equation (3.13) from [16] with the effective safety factors ω_i implies the minimum expected cost per period C_{total}.

$$S_i = \lambda_i (L_i + 1) + \omega_i \sigma_i \sqrt{L_i + 1} \,. \tag{3.13}$$

The optimal S_i value is computed for all bases per period, and the minimum C_{total} is determined. The total fill rate ϑ is defined as Equation (3.14) from [16].

$$\vartheta = 1 - (\sum_{i=1}^{M} BO_i \Big/ \sum_{i=1}^{M} \lambda_i) \,. \tag{3.14}$$

So, the redistribution of inventory between bases before the realization of demand is to reduce the risk of future stock outs. From Equation (3.13), we can determine the optimal order-up-to levels S_i to reduce the total system cost by Equation (3.9) and increase the fill rate and better service at the bases by Equation (3.14).

Chapter 4 Service Level Adjustment Policy (SLA)

4.1 Introduction

A new lateral transshipment policy is called the service level adjustment or service level agreement (SLA), and combines emergency lateral transshipments with preventive lateral transshipments. This policy can reduce the risk by forecasting stock outs in advance and efficiently respond to actual stock outs. In this section, we give a description on how this policy can be applied in an inventory system and what its relevant total system costs are.

4.2 Assumptions and notations

Consider a single item two-echelon inventory system with one depot and M bases. In this system, the item has high stock out cost. Demand for the item at base i is normally distributed and independent with mean λ_i and standard deviation σ_i. The average lead time L required to replenish each base from the central depot is assumed to be equal and deterministic. We assume that the lead time of emergency lateral transshipments between each base is extremely shorter than the lead time of replenishments from the depot to the bases. In this chapter, we focus on the service level adjustment policy operates based on several replenishment scenarios during one review replenishment period.

The following notations and assumptions will be used in this chapter:

- expected normal demand rate during a review period at base i is λ_i, standard deviation is σ_i,
- the length of the review period is R,
- t $(t \le R)$ is the index for time within a review period R,
- supply lead time from depot to any base is L,
- the reorder point for a lateral transshipment to base i $(i = 1, 2, \ldots, M)$ is S_i,
- the current stock or inventory level at base i at period t is $I_i(t)$,
- on-hand stock of base i at period t is $OH_i(t)$,
- quantity for backorder of base i at period t is $BO_i(t)$,

- $X_{k,j}$ is the quantity of lateral transshipments from base k to base j,
- the unit holding cost per period at any base is h,
- shortage (or backorder) cost per unit and period is v,
- unit lateral transshipment cost between two bases is a,
- ordering cost per order is c,
- $\text{int}(.)$ is a function to return the integer value,
- $\Phi(.)$ is the cumulative distribution function of standard normal distribution with mean 0 and variance 1,
- z_α is z-value with proportion α of the area under the normal curve, where $\alpha = \Phi((z_\alpha - \lambda)/\sigma)$,
- β is defined as the target level of $SLRP$,
- γ is defined as the lower level of $SLRP$,
- the lateral transshipments lead times are negligible.

In this paper, we use another definition of service level which is different from previous studies. We define the new service level to coincide with the service level adjustment policy to determine the desired stock level. For base i during one replenishment period R, we define $RP(t)$ as the remaining time until next replenishment period starts. It can be seen from Figure 4 and is expressed as Equation (4.1).

$$RP(t) = R - t\ . \tag{4.1}$$

If a base does not have enough stock to react to future demands before receiving ordered products, they satisfy emergency demands by having products transshipped from other bases with ample stock. According to Lee's theory [18], the new service level is defined as $SLRP$ which is the abbreviation of service level remaining period. The $SLRP$ indicates the probability that no stock outs occur at base i during the remaining period $RP(t)$. It can be seen from Figure 4. The mathematical formulation for the $SLRP$ at base i during the remaining period $RP(t)$ is according to Equation (4.2) from [18].

$$\begin{aligned} SLRP_i[RP(t)] &= P(D_{i,RP(t)} < I_i(t)) \\ &= \Phi\left[\frac{I_i(t) - \lambda_i(RP(t))}{\sigma_i(RP(t))}\right], \end{aligned} \tag{4.2}$$

where $D_{i,RP(t)}$ is the demand of base i during the period $RP(t)$.

This equation expresses the probability that the demand during the remaining period $RP(t)$ at base i is less than the current stock level. In other words, it corresponds to the probability that base i has not reached a stock out during the remaining period $RP(t)$.

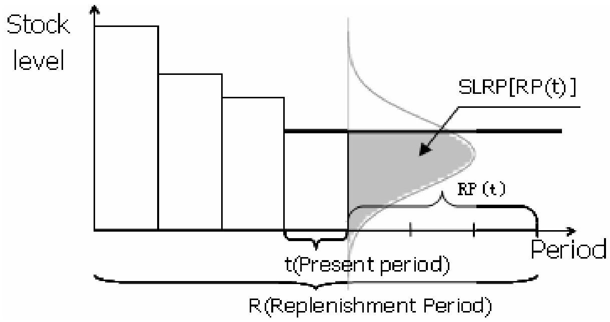


Figure 4: SLRP (figure from Lee [18])

4.3 Service level adjustment analysis

The service level adjustment policy is performed in the multi-echelon inventory system based on several replenishment scenarios. Based on the assumptions and notations above, the service level adjustment policy determines the quantities for lateral transshipments at a reviewing period according to the following steps.

In case, there is replenishment from the central depot in the current period t:

1. if $OH_i(t) > \lambda_i(t)$, the base i can immediately meet the customer demand with on-hand stock.
2. if $OH_i(t) < \lambda_i(t)$, the base i can meet the customer demand by replenishment from the central depot and also by lateral transshipments from other bases with on-hand stock.

In case, there is no replenishment in the current period t:

First, we use the upper and lower levels of the $SLRP$ to search for base k ($k \in K$) with a high stock level that satisfies Equation (4.3) and for base j ($j \in J$) with a low stock level that satisfies Equation (4.4) (Lee [18]).

1. $\text{int}\left\{\lambda_i(RP(t)) + z_\alpha \sigma_i(RP(t)) + 0.5\right\} < OH_i(t) - BO_i(t)$. (4.3)

2. $\text{int}\left\{\lambda_i(RP(t)) + z_\gamma \sigma_i(RP(t)) + 0.5\right\} \geq OH_i(t) - BO_i(t)$. (4.4)

Secondly, calculate the quantity for lateral transshipments between bases according to Equation (4.5) and (4.6) from Lee [18]:

1. The excess inventory for the bases $k \in K$ can be used for lateral transshipments at period t. The possible amounts to transship are

$$A_k(t) = OH_k(t) - \max\left\{\text{int}\left\{\lambda_k(RP(t)) + z_\alpha \sigma_k(RP(t)) + 0.5\right\}, 0\right\}, \quad k \in K. \quad (4.5)$$

2. The shortage at base $j \in J$ equals the expected needed transshipment quantity at period t. The amounts shortage are

$$SH_j(t) = \max\left\{\text{int}\left\{\lambda_j(RP(t)) + z_\beta \sigma_j(RP(t)) + 0.5\right\}, 0\right\} - (OH_j(t) - BO_j(t)), \quad j \in J. \quad (4.6)$$

Thirdly, we adjust the quantity for stock level at each base when the available stock is smaller than the quantity of required stock or stock outs. It is followed Equation (4.7) to adjust required stock quantity.

$$SH_j(t) = \text{int}(BO_j(t) \cdot (\sum_k A_k(t)) / \sum_j BO_j(t)). \quad (4.7)$$

Fourthly, redistribute the highest available stock from base k to the highest required stock at base j. The total quantity of lateral transshipments $X_{k,j}(t)$ from base k to base j at the current period t is expressed in Equation (4.8) (from Lee [18]).

$$X_{k,j}(t) = \min[\max(SH_j(t)), \max(A_k(t))]. \quad (4.8)$$

Fifthly, update the quantity of required stock and available stock according to Expressions (4.9) and (4.10).

$$SH_j(t) - X_{k,j}(t) \to SH_j(t). \quad (4.9)$$

$$A_k(t) - X_{k,j}(t) \to A_k(t). \quad (4.10)$$

Finally, if the total quantity of shortage stock at base j satisfies Equation (4.11), the service level adjustment policy in the current period is completed.

$$\sum_j SH_j(t) = 0. \quad (4.11)$$

4.4 System performance analysis

The fundamental rationale for the lateral transshipments is the potential for achieving higher customer service levels. In this model, the *SLRP* service level is based on the concept of the safety stock. So the holding cost should be considered inside the model. Furthermore, the backorder, transportation and ordering costs should also be considered in the system. The total cost during whole period can be written as Equation (4.12) (from Lee [18]):

$$C_{total} = \sum_{t} \left\{ \sum_{i} (h \cdot OH_i(t) + v \cdot BO_i(t)) + \sum_{k} \sum_{j} (a \cdot X_{k,j}(t) + c \cdot e_{k,j}(t)) \right\}, \tag{4.12}$$

where $e_{k,j}(t)$ is a constant value of 1 when lateral transshipment is performed between base k and base j at period t.

Chapter 5 Conclusions and Future Work

We have presented models for two-echelon inventory systems in which lateral transshipments are allowed. A very common problem in inventory theory is to determine the optimal reoder point and order quantity in order to minimize total costs. In this paper, we illustrate three types of lateral transshipment policies: emergency lateral transshipments, preventive lateral transshipments and service level adjustments. For the model which applies emergency lateral transshipments together with a $(S-1, S)$ or (r, q) replenishment policy, we present the calculations to characterize the service performance (e.g. expected shortages). For the model which uses preventive lateral transshipments, we have examined the relative effectiveness of two rather simple lateral transshipment methodologies: lateral transshipments based on availability and lateral transshipments based on inventory equalization. Finally, we propose a new lateral transshipment policy, called service level adjustment, which reduces the risk of stock outs in advance and efficiently responds to customer demands. These three lateral transshipments policies can maximize customer satisfaction and reduce management costs in certain inventory systems.

We considered the usage of a periodic review system for this study, but additional studies of lateral transshipments are needed to include the characteristics that incorporate inventory policy systems according to the characteristics of products in the supply chain. Furthermore, our findings are obviously valid only under the specific operating assumptions of the model studied in this chapter. There are many variations of that model which present practical interests and constitute potential topics of future research. Some of the most extensions are the following: (a) non-negligible transshipment times; (b) different costs at each base; (c) dependent demand; (d) more than one depot in the inventory system; (e) alternative transshipment policies; (f) lost sales. Better understanding of the properties and performance of more complex systems with some of these characteristics may have significant implications for the design and operation of supply chains.

References

1. H Lee, A multi-echelon inventory model for repairable items with emergency lateral transshipments, *Management Science* **33** (1987), pp. 1302–1316.

2. S Axsäter, Modeling emergency lateral transshipments in inventory systems, *Management Science* **36** (1990), pp. 1329–1338.

3. C Sherbrooke, Multi-echelon inventory systems with lateral supply, *Naval Research Logistics* **39** (1992), pp. 29–40.

4. P Alfredsson and J Verrijdt, Modeling emergency supply flexibility in a two-echelon inventory system, *Management Science* **45** (1999), pp. 1416–1431.

5. J Grahovac and A Chakravarty, Sharing and lateral transshipment of inventory in a supply chain with expensive lowdemand items, *Management Science* **47** (2001), pp. 579–594.

6. A Kukreja, CP Schmidt and DM Miller, Stocking decisions for low-usage items in a multilocation inventory system, *Management Science* **47** (2001), pp. 1371–1383.

7. H Wong, GJ Van Houtum, D Cattrysse and D Van Oudheusden, Multi-item spare parts systems with lateral transshipments and waiting time constraints, *European Journal of Operational Research* **171** (2006), pp. 1071–1093.

8. PM Needham and PT Evers, The influence of individual cost factors on the use of emergency transshipments, *Transportation Research* **E34** (1998), pp. 149–160.

9. PT Evers, Heuristics for assessing emergency transshipments, *European Journal of Operational Research* **129** (2001), pp. 311–316.

10. K Xu, PT Evers and MC Fu, Estimating customer service in a two-location continuous review inventory model with emergency transshipments, *European Journal of Operational Research* **145** (2003), pp. 569–584.

11. S Axsäter, Evaluation of unidirectional lateral transshipments and substitutions in inventory systems, *European Journal of Operational Research* **149** (2003), pp. 438–447.

12. PH Zipkin, *Foundations of inventory management*, McGraw-Hill, Boston (2002).

13. S Axsäter, Evaluation of installation stock based (R, Q) polieies for two-level inventory systems with Poisson demand, *Operations Research* **46** (1998), pp. 135–145.

14. J Huo, Batch ordering policy of multi-location spare parts inventory system with emergency lateral transshipments, *Chinese Language Journal*, **27** (2007), pp.62-67 .

15. A Banerjee, J Butrton, S Banerjee, A simulation sudy of lateral transshipment in single supplier, *International Journal of Production Economics*, **93-94** (2005), pp. 169-178.

16. G Tagaras, Pooling in multi-location periodic inventory distribution systems, *The International Journal of management Science*, **27** (1999), pp. 39-59.

17. S Graves, A multi-echelon inventory model for a repairable item with one-for-one replenishment, *Management Science*, **31** (1985).

18. H Lee, J W Jung, Y S Jeon, An effective lateral transshipment policy to improve service level in the supply chain, *International Journal of Production Economics*, **106** (2007), pp.115-126.