1. Introduction

Effective supply chain management (SCM) is currently recognized as a key determinant of competitiveness and success for most manufacturing and retailing organizations, because the implementation of supply chain management has significant impact on cost, service level, and quality. Numerous strategies for archiving these targets have been proposed and investigated in both practice and academic over the past decades. One such strategy, commonly practiced in multi-location supply chain systems facing stochastic demand, allows movement of stock between locations at the same echelon level or even across different levels. These stock movements are termed lateral transshipments, or simply, transshipments. As a demand occurs under the implementation of transshipment strategy, there will be three possible activities—the demand is met from the stock on-hand or it is met via transshipment from another location in the system or it is backordered. In another words, firstly, if a location’s on-hand inventory level is greater than the demand size, then the demand is met. Secondly, if the on-hand inventory level is positive but less than the demand size, then it is used to partially satisfy the demand and the remaining demand is met either via transshipment or is backordered. Thirdly, if the on-hand inventory level is zero, the demand is met via transshipment or is backordered under the assumption of no lost sale. In addition to the same echelon level transshipment, when neither one location’s same level partners in the same region nor its designated supplier/warehouse/or distribution center lack sufficient inventory to meet the demand, the unmet remaining demand can be fulfilled from the upper-level supplier which may not belong to the same geographical region. This practice is defined as cross-level transshipment. The illustration of transshipment is depicted in Figure 1. Therefore, transshipment policy can improve stock availability, i.e., customer service level, without increasing stock level which may induce higher inventory relevant cost. In another words, transshipments enable the sharing of stock among locations, they facilitate each location as a secondary, random supply source for the remainder. Thus, the locations’ replenishment can be coordinated and even combined in order to avoid excessive inventory costs.

Transshipment research is motivated by observations from various industries. It has gained increasingly attention in medicine, apparel, and fashion goods, particularly by those retailers with brick and click outlets, or critical repairable spare parts of equipment-intensive
industries such as airlines, nuclear power plants, and complex machines. They are also suitable for retailers that require long replenishment lead times from suppliers located closer to each other or spend significant funds on construction and operation of storage facilities to prevent costly shortage penalty.

Figure 1 Transshipment in a supply chain system

One of the prerequisite of successful implementation of transshipment is well-established information systems. At present many large modern companies connected by information systems can control the relationships of many branches, and thus they may be ready to reap cost reduction and service improvement associated with lateral transshipment.

In the past decades, a considerable amount of research has been dedicated to this field. The list of research papers dealing with transshipments to date is quite long and no attempt is made here to exhaust it. In this chapter we mainly focus on presenting a comprehensive description, classification, methodologies and solution procedures, and research directions for further study of transshipment in supply chain system. (Köchel, 1998) conducted a preliminary survey on transshipment, however, a considerable amount of research over the past decade have not been covered. The main aim of this paper is to present a systematic survey on the development of transshipment studies.

This chapter is organized as follows. In Section 2, we introduce the characteristics of the transshipment problem. We identify the scope, introduce the basic assumptions, explain the elements of the transshipment policy, introduce the common inventory control policies, and define the performance measurement. Section 3 classifies the transshipment problems based
2. Characteristics of transshipment

2.1 Scope
Effective supply chain management (SCM) has become an important management paradigm. A great amount of studies have shown that substantial benefits can be obtained from SCM. Basically, SCM is an effective and systematic approach of managing the entire flow of information, material and services in fulfilling a customer demand (Chase, 1998). In this chapter we are mainly focused on material flow management in the supply chain system. At present many quantitative models have been proposed to provide decision support for the management of materials in supply chains (see, Tayur et al., 1998). Moreover, since the network of entities that constitute the entire supply chain is typically too complex to analyze and optimize globally, it is often desirable to focus on smaller parts of the system so as to gain a in-depth understanding of its characteristics, performance and tradeoffs involved. One such part that is attracting growing attention is the local distribution network, consisting of multiple retail outlets (stocking locations), which are supplied by one or more sources.

The overall performance of the distribution network, whether evaluated in economic terms or in terms of customer service, can be substantially improved if the retailers collaborate in the occurrence of unexpectedly high demand, which may result in shortages in one or more retailing outlets. Collaboration usually takes the form of lateral inventory transshipment from a stock outlet with a surplus of on-hand inventory to another outlet that faces a stockout. Since the cost of transshipment in practice is generally lower than both the shortage cost and the cost of an emergency delivery from the designated warehouse and the transshipment time is shorter than the regular replenishment lead time, lateral transshipment simultaneously reduces the total system cost and increases the fill rates at the retailers. A group of stocking locations that share their inventory in this manner is to form a pooling group, since they effectively share their stock to reduce the risk of shortages and provide better service at lower cost.

There is a considerable amount of literature on SCM over the past decades. Some papers have provided literature survey for some specific topics. For example, (Ganeshan et al., 1998) provided a taxonomic review of the SCM research in three categories: competitive strategies, firm-focused tactics, and operational efficiency. (Tsay et al., 1998) reviewed the recent literature on supply chain contracts. (Tan, 2001) provided a review of the evolution of the SCM philosophy. (Sahin & Robinson, 2002) provided a review of the prior research on information and physical flow coordination. (Li & Wang, 2007) focused on coordination mechanisms that can align the objectives of individual supply chain members. This paper mainly concentrates on an analysis of the operation of transshipment in a group consisting of multiple retailing outlets and a single/or multiple upper sources in a two-echelon supply chain system.

Other researchers have also examined the effectiveness of lateral shipments for Repairable or recoverable items (e.g. Lee, 1987; Axsäter, 1990; Wong et al., 2005 & 2006), while still
others have focused on consumable products (see, e.g., Jonsson & Silver, 1987; Archibald et al., 1997; Cohen et al., 1986; Robinson, 1990).

2.2 Assumptions

There are several basic assumptions that are commonly seen in the literature of transshipment such as the behaviors of demand occurrence, transshipment time, repair time, and transshipping priority rule etc. They are stated as follows.

The behaviors of demand occurrence are usually characterized by the time between demands and the distribution of demand size. The time between demands is commonly assumed to follow an Exponential or Gamma distribution. However, the distributions of demand size per each demand occurrence depend on the characteristics of the investigated industry. For example, it was taken as Weibull distribution for spare parts which have slow-moving, expensive and lumpy demand pattern (Kukreja & Schmit, 2005). (Needham & Evers, 1998) assume the normal distribution truncated at zero for military spare parts. A drawback of using the normal distribution is that it is less appropriate for low volume items (Silver & Peterson, 1998); however, it does not place any restriction on the values of the mean and variance (compared to, say, the Poisson distribution which requires that they must be equal). In addition, its properties are well known and it is typically the basis for examining continuous demand. Besides, (Wong et al., 2006) assumed the demand occur according to Poisson process with constant rate for reparable parts in equipment-intensive industries such as, airlines, nuclear power plants, and manufacturing plants using complex machines. These industries are confronted with the challenge of maintaining high system availability, while limiting the costly spare parts inventory simultaneously. Furthermore, in a large amount transshipment literature the behaviors of demand are alternatively characterized by assuming what distribution the average demand per time period follows (e.g., Needham & Evers, 1998; Tagaras, 1999; Herer & Rashit, 1999; Burton & Banerjee, 2005; Wong et al., 2006).

The majority literature assumes transshipment time to be negligible. (Kukreja & Schmidt, 2005) considered that a large utility company has all locations in five adjoining southeastern states. A part can be transshipped between any two locations within a working day. This transshipment time is acceptable to the management and is treated as negligible. Because many spare parts are for machines like turbines, pulverizers, etc., substantial time is needed to take off the part from the machine and prepare it to receive the new part. If the new part can be supplied via transshipment within one day, then the assumption of negligible transshipment times is acceptable. At present only some papers account for the non-negligible transshipment time. The transshipment time are assumed to be shorter than emergency supply. In other words, lateral transshipments are faster and cheaper than emergency supplies because all firms in the pooling group should be at close distance to each other. Otherwise it makes no sense to pool the item inventories. Therefore, a lateral transshipment is always preferred over an emergency supply from time and cost perspectives.

(Gong & Yucesan, 2006) formulated a multi-location transshipment problem with positive replenish lead time. They used simulation optimization by combining an LP/Network flow in corporate with infinitesimal perturbation analysis (IPA) to analyze the problem, and obtains the optimal base stock quantities through sample path optimization. (Wong et al., 2005 & 2006) addressed the analysis of a multi-item, continuous review model of a multi-
location inventory system of repairable parts with lateral transshipment and waiting time constraints, in which lateral and emergency shipments occur in response to stockouts. The objectives is to determine close-to-optimal stocking policies minimizing the total cost for inventory holding, lateral transshipments and emergency shipment subject to a target level for the average waiting times at all location. For the case of transshipment for spare parts, the repair time is usually assumed exponentially distributed. This assumption is probably not very realistic. However, (Axsäter, 1992) and (Alfredsson & Verrijdt, 1999) showed that the service performance of the system is insensitive to the choice of lead time distribution.

(Wong et al., 2005) showed that delayed lateral transshipments can improve the system performance. When a base having no backorders receives a repaired part and at the same time at least one base in the pooling group has backorders, it would be reasonable to send the repaired part to the base having backorders. Therefore, a delayed lateral transshipment occurs when the system has backorders and it is triggered by a repair completion.

One common transshipping priority rule for fulfilling the demands is assumed that a location receiving an order first satisfies its own backorder, if one exists, and then uses the remaining units to satisfy backorder(s) at other location(s) in a way that minimizes transshipping costs. The requested backorders are to be fulfilled according to first come first serve policy. However, if a transshipment request is indicated 1 day prior to the arrival time of the next cycle’s shipment from the upper echelon, no lateral shipments are made in the current cycle, due to the anticipated delivery of a relative large quantity the following day.

In addition to the first come first serve of demand handling rule, (Zhao et al., 2006) investigated a decentralized dealer network where each dealer is willing share his inventory. They considered the inventory sharing with multiple demand classes. Assume that each dealer faces his own customer demand with higher priority and transshipment requests from other dealers with lower priority.

A significant amount of literature in transshipment assumed that complete pooling policy is to be applied. This is part of the agreement between the cooperating companies. When the demand at a location cannot be met from on-hand inventory, it is met via transshipment(s) from other outlet(s) in a way that minimizes the transshipping cost. A unit demand is backordered if it cannot be satisfied via transshipment, in other words when there are no units in the system. In case companies do not want to share their last parts, one may introduce threshold parameters, aka partial pooling, and agree that a company does not supply a part by a lateral transshipment if the physical stock of the requested item is at or below the threshold level. A rule has to be added for how the values of the threshold parameters are chosen, or one may consider them as additional decision parameters. In principle, this extended model may be analyzed along the same lines as our current model. (Cantagalli, 1987) and (Needham & Evers, 1998) classified the transshipment policy as complete pooling and partial pooling for lateral transshipment. The partial pooling transshipment will be addressed further in the next sub-section.

2.3 Transshipment policies

As the literature and practice suggested, there are two classes of transshipment. (Lee et al., 2007) proposed that lateral transshipment can be divided into two categories: emergency lateral transshipment (ELT) and preventive lateral transshipment (PLT). ELT directs emergency redistribution from a retailer with ample stock to a retailer that has reached stockout. However, PLT reduces risk by redistributing stock between retailers that
anticipate stockout before the realization of customer demands. In short, ELT responds to stockout while PLT reduces the risk of future stockout. This concept of transshipment classification is similar to (Banerjee et al., 2003), wherein two kinds of policies were proposed: Lateral transshipment based on availability (TBA) and Lateral transshipment for inventory equalization (TIE). TBA transships stock to retailers with less than desirable levels until all stock is depleted. However, this policy is problematic when desired stock levels are determined incorrectly. TIE redistributes stock to match the target level of demand of each retailer whenever there are retailers with less than desirable stock levels. This policy does not respond to stockout after redistribution, because redistribution is performed once in every replenishment cycle. We will discuss these policies more in-depth and also address some recent developed transshipment rules in the following.

**No lateral shipments (NLS) policy:** In the literature NLS policy is usually used a baseline case for evaluating the effectiveness of the transshipment policies. Under this policy, no lateral transshipments are allowed. For example, at the end of each review cycle of 30 days, say, each retailer’s order size is determined from its individual order-up-to level and this quantity is received only from the supplier after the supply lead time elapses. Total backordering is allowed at each retail outlet. The sales will be lost, if the unmet sales can not be backlogged. For such case, shortage cost will be incurred.

**Lateral transshipments based on availability (TBA) policy:** This implies that either all current transshipment needs have been met, or the total available transshipment quantity among all the excess locations has been exhausted. TBA, or called ELT, allows the transshipment decisions to be made more than once during a review cycle, based on the transshipment order point signal. TBA mandates emergency redistribution from a retailer with ample stock to a retailer that has reached stockout (Lee, 1987). (Lee, 1987) presented a model that allows TBA between local warehouses that are part of a group. If a local warehouse cannot satisfy customer demands with its on-hand stock, TBA is used to fill the demands from a warehouse in the same group that has enough stock on hand. If TBA is impossible due to group-wide stockout, the unmet demand will be backordered. (Lee, 1987) derived expressions that approximate the fractions of demands that can be satisfied by stock on hand, TBA, and backordering, and in doing so, proved that applying lateral transshipment reduces total cost. (Axsätze, 1990) analyzed a system similar to that of (Lee, 1987), but with the modification of assuming that warehouses within each group are not identical. (Axsätze, 1990) derived steady-state probability by assuming exponentially distributed replenishment time. Analytical results were compared with simulation results to show that, in the case of non-identical warehouses, the proposed model gives better results. (Tagaras & Cohen, 1992) investigated a model with two locations and non-zero replenishment lead time. (Rudi et al., 2001) investigated the conflict between maximizing location and system profits in a two-location model. These models assumed a non-negligible lead time on the service of customer requests to allow the total demand to become apparent before transshipments are arranged. In highly competitive retail situations, such a delay would often lead to lost sales. However, if a transshipment is requested 1 day prior to the arrival time of the next cycle’s shipment from the upper-echelon supplier, no lateral shipments are made in the current cycle, due to it own nearly anticipated delivery of a relative large quantity the following day.

**Lateral transshipments for inventory equalization (TIE) policy:** Under this policy, the transshipment decisions are based on the concept of inventory balancing or equalization
through stock redistribution. One of the first such models is due to (Gross, 1963) who characterized an optimal policy for a two-location system in which replenishment and transshipment decisions are taken together at the beginning of each period. (Das, 1975) analyzed a variant of this model in which the transshipment decision is taken at a fixed point during each period. (Jonsson & Silver, 1987) examined a model in which the objective is to minimize backorders rather than cost. The transshipment decision is taken a fixed time before the replenishment decision and the model allows for non-zero transshipment lead time and an arbitrary number of locations. In this case, as opposed to the TBA policy described above, inventory redistribution occurs no more than once in every review cycle. There are three commonly used redistribution rules for TIE, or called PLT. Firstly, at the time of transshipment, inventories are redistributed among the retail locations through one or more lateral shipments, such that all locations will have an equal number of days’ supply (or, alternatively, equal runout times) just after the appropriate transshipment(s). Secondly, TIE redistributes stock to match the ratio of average demand of each retailer to that of the whole retailer whenever there are retailers with less than desirable stock levels. Thirdly, another redistribution policy proposed by (Bertrand & Bookbinder, 1998) adjusted stock to achieve equal marginal cost over all retailers just before the replenishment period. This method has the disadvantage of not being able to respond to stockout before or after redistribution because the redistribution policy is only performed at the end of the replenishment period. There are many other possible transshipment policies that can be devised based on the concept of TIE. For example, (Tagarus, 1999) compared two extreme policies, Random (RA) policy and Risk Balancing (RB) policy, which can be more or less easily implemented in practice. When one location faces a shortage, the decisions of the source location and the quantity of transshipment should take into account the risk of shortage in the following period.

The above mentioned policies have the disadvantage of not being able to respond to stockout before or after redistribution, and they cannot appropriately determine desired stock levels. (Lee et al., 2007) proposed a new lateral transshipment policy, service level adjustment (SLA), to effectively deal with retailer demand. The proposed policy can reduce risk by forecasting stockout in advance and efficiently responding actual stockout by combining TBA and TIE.

Recently, (Minner & Silver, 2005) observed that choosing the better of two extreme policies leads to performance that is nearly as good as a more complex analysis that takes account of the future impact of a transshipment on the cost at the location sending the shipment. These extreme policies under investigated are (i) never transship and (ii) always transship when there is a shortage at one location and stock available at another. They developed an analytical approach for estimating the expected cost. Thus, it provided a mechanism for selecting the better policy between these two extreme policies. (Burton & Benerjee, 2005) examined the cost effects of two lateral (intra-echelon) transshipment approaches in a two-echelon supply chain network, with a single supply source at the higher echelon and multiple retail locations at the lower. Through a series of simulation experiments under different operating conditions, they found an ad hoc emergency transshipment approach appears to be significantly more effective in terms of several important criteria, as compared to a more systematic transshipment technique based on stock level equalization.
In view of the transshipment, most research has focused on determining when to fill transshipment requests from other dealers, ignoring the decision of determining when to send transshipment requests to other dealers. With an exception, (Zhao et al., 2006) developed optimal inventory transshipment policies that incorporate both types of decisions in a decentralized system. They devised threshold rationing and requesting levels for determining the optimal inventory and transshipment decisions for each individual dealer. They also considered a decentralized two-dealer network and use a game theoretic approach to characterize the equilibrium inventory strategies of the individual dealers. The research is classified based on transshipment policies are summarized in Table 1.

<table>
<thead>
<tr>
<th>TBA(ELT)</th>
<th>TIE(PLT)</th>
<th>Combined or Newly developed</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Axsäter, 1990)</td>
<td>(Das, 1975)</td>
<td>(Lee et al., 2007)</td>
</tr>
<tr>
<td>(Rudi et al., 2001)</td>
<td>(Bertrand &amp; Bookbinder, 1998)</td>
<td>(Burton &amp; Benerjee, 2005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Zhao et al., 2006)</td>
</tr>
</tbody>
</table>

Table 1. Transshipment policies

2.4 Inventory control policies in transshipment

Transshipment policies are incorporated with traditional inventory control policies which are classified based on two fundamental questions: when to replenish and how much to order. Commonly used inventory control policies such as (S-1,S), (Q,R), (R,S), and (s,S) will be discussed as follows.

Inventory control policy (S-1,S)
Continuous one-for-one stock replenishments (S-1,S) is a commonly used inventory control policy for a system in cooperation with transshipment. It means whenever any stock is withdrawn, a replenishment order is released. This control policy is especially suitable for slow-moving and expensive items. The first to deal with continuous one-for-one inventory policies in multi-echelon systems with transshipment were (Dada, 1984) and (Lee, 1987). One can refer to the following research for more in-depth description, (Lee, 1987), (Axsäter, 1990), (Sherbrooke, 1992), (Yanagi & Sasaki, 1992), (Alfredsson & Verrijdt, 1999), (Grahovac & Chakravarty, 2001), (Kukreja et al., 2001), and (Wong et al., 2002 & 2005).

(Lee, 1987) developed a method of determining the minimum cost inventory position for a system that allows transshipments between identical locations and finds approximations to measures of system performance including the expected number of backorders and transshipments. Expressions derived approximate the fractions of demands that can be satisfied by stock on hand, TBA, and backordering, and in doing so, proved that applying lateral transshipment reduces total cost.

Both (Axsäter, 1990) and (Sherbrooke, 1992) proposed similar approximations for systems that allow transshipments between non-identical locations. (Axsäter, 1990) analyzed a system similar to that of (Lee, 1987), but with the modification of assuming that warehouses within each group are not identical. Steady-state probability is derived by assuming exponentially distributed replenishment time. Analytical results were compared with simulation results to show that, in the case of non-identical warehouses, this proposed model gives better results.
Recently, (Grahovac & Chakravarty, 2001) formulated and solved the proposed model based on (S-1,S) policy. They reached some counter-intuitive conclusion that is saving is not always accompanied by a reduction in overall reduction inventory in the supply chain. These opposing trends suggest that new extra incentives are needed to enforce the transshipment arrangement. In addition, (Kukreja et al., 2001) developed a heuristic to determine replenishment and transshipment policies for a system with non-identical locations under the objective of minimizing cost.

(Wong et al., 2005) extended the single item model of (Wong et al., 2005) to a model of multiple items. They analyze a two-location, multi-item, continuous-review system for repairable items with one-for-one stock replenishments and determine policies for all items that minimize the total cost subject to a target level for average waiting time. However, these models are only appropriate for slow moving, expensive and/or repairable items.

**Inventory control policy (Q,R)**

The continuous (Q,R) policy consider a reorder quantity, reorder point (Q,R) system instead, where an order of fixed size (Q) is placed whenever the reorder point (R) is reached. The (Q,R) system is a very common and relatively straightforward system whose primary drawback is associated with demands of appreciable magnitude (Silver & Peterson, 1985). There are considerable amount of research on (Q,R) inventory control policy for system with transshipment. Readers can refer to the following research for more in-depth description, (Needham & Evers, 1998), (Evers, 2001), and recently (Xu et al., 2003), (Minner et al, 2003) and (Axsäter, 2003 a,b).

(Needham & Evers, 1998) examines the interaction of relevant costs and transshipment policies and presents a method for determining the point at which the benefits of transshipments outweigh their costs. Simulation and sensitivity analysis identify the relevant costs drivers and are used to construct a decision making tool for managers contemplating the implementation of transshipments. Simulation results indicate that the cost of a stockout is the primary determinant in the transshipment decision, with higher stockout cost levels generally increasing the likelihood that transshipment usage will lead to lower overall cost. A meta-model is proposed as a practical means of providing insight into when emergency transshipments should be employed.

Under (Q,R) continuous review policy, (Evers, 2001) and (Minner et al., 2003) developed heuristics to determine whether or not to make a transshipment in a multilocation inventory system facing a stockout. (Axsäter, 2003a) developed an approximate method of determining the replenishment policy for a continuous review multilocation inventory system in which a location facing a stockout sources items from locations with lower shortage costs whenever possible. In this paper, transshipments are only allowed in one direction, i.e., the flow of transshipment is in only one direction. Such policies can be of interest if the warehouses have very different shortage costs. Another interpretation is substitution in an inventory system. That is when a demand for a low quality item cannot be met directly, the item can be replaced by another high quality item. He provide a simple and efficient approximate technique for policy evaluation in such systems. Under the assumption that no further transshipment will take place, (Axsäter, 2003b) extends (Axsäter’s, 2003a) modeled and proposed the decision rule and develop a heuristic to determine whether or not to make a transshipment in response to a stockout.
Inventory control policy (s, S)
Other studies of transshipment assume periodic review policies and they usually assume no order setup cost, so that an order-up-to or base-stock policy is appropriate. In a (s, S) inventory policy, an order is placed every time the inventory position drops to or below s, and the order size is the difference between S, the order up to level, and the inventory position at the time of placing the order. The outlets operate independently and follow a continuous review (s, S) inventory policy. In fact, the interactions in terms of transshipments make the system complex, and analytical results for such a system seem to be intractable.
This type of inventory system with multiple locations interacting in terms of complete pooling of stock, and each location following a (s, S) type inventory control policy has not been dealt with in the past in the literature.
Some studies devised (s,S) inventory policy in transshipment such as (Krishnan & Rao, 1965), (Cantagalli, 1987), (Tagaras, 1989 & 1999), (Robinson, 1990), (Tagaras & Cohen, 1992), (Archibald et al., 1997), (Rudi et al., 2001), and (Herer et al., 2002). The scopes and contributions of these studies are discussed below.
The model of (Krishnan & Rao, 1965) minimized cost in a multilocation system with zero replenishment and transshipment lead times. (Cantagalli, 1987) evaluated the impact of four different emergency transshipment policies using the (s,S) inventory system. In the paper variant transshipment policies, complete pooling and partial pooling rules are examined. Though (s,S) systems are in use, they tend to be difficult to work with in terms of establishing the control parameters (Silver & Peterson, 1985). (Robinson, 1990) characterized the form of close-to-optimal policies for similar systems. (Tagaras & Cohen, 1992) examined a model with two locations and non-zero replenishment lead time. (Rudi et al., 2001) investigated the conflict between maximizing location and system profits in a two-location model. These models require a non-negligible lead time on the service of customer requests to allow the total demand to become apparent before transshipments are arranged. In highly competitive retail situations, such a delay would often lead to lost sales.
Moreover, (Herer et al., 2002) determined how much to replenish and how much to transship each period; thus this work can be viewed as a synthesis of transshipment problems in a static stochastic setting and multi-location dynamic deterministic lot sizing problems. They provide interesting structural properties of optimal policies which enhance our understanding of the important issues which motivate transshipments and allow us to develop an efficient polynomial time algorithm for obtaining the optimal strategy.
Recently, (Hu et al., 2005) adopted the major assumptions in (Krishnan & Rao, 1965) for an N-location inventory system but extended their one-period, base-stock inventory model to a multi-period, general (s, S)-type model. While (Robinson, 1990) considers the simultaneous determination of base stock inventory policies at each store as well as the transshipment decisions, they focus on the development of an appropriate (s,S)-type policy for a multi-location inventory system with centralized ordering. The focus of this research is to investigate the effect of transshipment costs on the optimal(s,S) ordering policy that minimizes inventory and transshipment costs. Then this (s,S) ordering policy is then compared with a simplified policy that assumes free and instantaneous transshipments. In general, the results indicated that using transshipments seem to be a very cost effective way of reducing inventories for situations with a large number of stores where transshipment costs are small relative to the stock-out plus holding costs.
Other studies assume periodic review policies and they usually assume no order setup cost, so that an order-up-to or base-stock policy is appropriate. Examples are (Gross, 1963), (Krishnan & Rao, 1965), (Das, 1975), (Hoadley & Heyman, 1977), (Cohen et al., 1986), (Tagaras, 1989 & 1999), (Robinson, 1990), (Tagaras & Cohen, 1992), (Archibald et al., 1997), (Rudi et al., 2001), and (Herer et al., 2002). Different from those examples, (Herer & Rashit, 1999) introduce the existence of non-negligible fixed and joint replenishment costs.

In summary, lateral transshipments have been analyzed from the perspective of both continuous time (e.g. Lee, 1987; Axsäter, 1990; Sherbrooke, 1992), as well as discrete periods (see, e.g. Showers, 1979; Archibald et al., 1997; Kochel, 1998). The relevant research is classified based on transshipment policies are summarized in Table 2.

<table>
<thead>
<tr>
<th>(S-1,S)</th>
<th>(Q,R)</th>
<th>(R,S)</th>
<th>(s,S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Dada, 1984), (Lee, 1987), (Axsäter, 1990), (Sherbrooke, 1992), (Yanagi &amp; Sasaki, 1992), (Alfredsson &amp; Verrijdt, 1999), (Grahovac &amp; Chakravarty, 2001), (Kukreja et al., 2001), (Wong et al., 2002 &amp; 2005), (Silver &amp; Peterson, 1985), (Needham &amp; Evers, 1998), (Evers, 2001), (Xu et al., 2003), (Minner et al, 2003), (Axsäter, 2003 a,b).</td>
<td>(Gross, 1963), (Krishnan &amp; Rao, 1965), (Das, 1975), (Hoadley &amp; Heyman, 1977), (Cohen et al., 1986), (Tagaras, 1989 &amp; 1999), (Robinson, 1990), (Tagaras &amp; Cohen, 1992), (Archibald et al., 1997), (Rudi et al., 2001), (Herer et al., 2002).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Inventory control policies

2.5 Performance measures

As above mentioned, the implementation of supply chain management has significant impact on cost, service level, and quality. Emergency transshipments represent one way in which logistics managers can reduce inventories while simultaneously maintaining customer service levels. Therefore, the commonly used performance measures to evaluate the effectiveness of transshipment are the costs and service level. The relevant costs considered in the transshipment model are similar to those of inventory research. They are stockout cost (aka, shortage cost), holding cost, transportation cost and ordering cost. Stockout costs were used to assign a penalty when a customer request could not be filled. Two holding costs were classified as in-storage and in-transit, the in-transit holding cost is usually lower than the in-storage cost. Two transportation costs considered are routine and rush transportation costs. The routine transportation costs are taken from full-truckload (FTL) rates. Nevertheless, rush transportation costs are taken from less-than-truckload (LTL) rates. Ordering costs in the simulation model were accumulated each time an order was placed from either a retail center or a distribution center.

Some other papers addressed the transshipment problem with variant objectives. (Jönsson & Silver, 1987) examined a model in which the objective is to minimize backorders rather than cost. (Lee, 1987) developed a method of determining the minimum cost inventory position...
for a system that allows transshipments between identical locations and finds approximations to measures of system performance including the expected number of backorders and transshipments. (Bertrand & Bookbinder, 1998) considered this model with the objective of minimizing cost for the case of zero transshipment lead time. (Rudi et al., 2001) investigated the conflict between maximizing location and system profits in a two-location model. (Lee et al., 2007) considered future demands, current stock quantity, and the degree of stockout, the service level proposed in their study can be used as criteria to evaluate the performance of lateral transshipment. It is called service level for the remaining period (SLRP) which is based on the concept of safety stock. Refer to (Lee et al., 2007), a summary of service levels from the previous research are shown in Table 3.

<table>
<thead>
<tr>
<th>Service level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle-service level (Lee &amp; Larry, 2002)</td>
<td>the desirable probability of not occurring stockout in any cycle</td>
</tr>
<tr>
<td>Customer service level (Yan et al., 2003)</td>
<td>Portion of demand met</td>
</tr>
<tr>
<td>Customer service level (Biswas &amp; Narahari, 2004)</td>
<td>Order fill rate: the fraction of demand are met from on-hand stock</td>
</tr>
<tr>
<td></td>
<td>Probability of on-time delivery: the fraction of demand are met fulfilled</td>
</tr>
<tr>
<td>Service level (Surrie &amp; Wagner, 2002)</td>
<td>(i) $\alpha$-service level: the fraction of incoming order are fulfilled</td>
</tr>
<tr>
<td></td>
<td>from on-hand stock</td>
</tr>
<tr>
<td></td>
<td>(ii) $\beta$-service level: the proportion of incoming order quantity</td>
</tr>
<tr>
<td></td>
<td>are fulfilled from on-hand stock</td>
</tr>
<tr>
<td></td>
<td>(iii) $\gamma$-service level:1-mean demand not fulfilled /mean demand per</td>
</tr>
<tr>
<td>Service level (Lee et al., 2007)</td>
<td>the probability of not occurring stockout for retailer during its</td>
</tr>
<tr>
<td></td>
<td>remaining period RP</td>
</tr>
</tbody>
</table>

Table 3. Definitions of service level

3. Classifications of transshipment problems

One can think of the following important features that should be taken into account when trying to present existing research systematically: (1) the number of item(s) in inventory system, (2) the number of locations in the pooling group, (3) the number of warehouses/supplier(s) (4) the replenishment lead time from the warehouse(s), (5) the demand process, (6) the timing of transshipment (preventive or emergency), (7) the measure of performance (cost or service level), (8) the storage space or waiting time constraint, (9) the direction of transshipment, and (10) the reparability of stocked items. Some have been addressed in the previous sections. Here just focus on the topics that have not been mentioned.

3.1 Number of item(s)

Most of the transshipment related research deals with single-item problems in which only one item at a time is considered. Such problems are typical when we use an item approach. Under an item approach, inventory levels for each individual item are set independently.
An alternative approach, denoted as the system approach by (Sherbrooke, 2004), considered all items in the system when making inventory-level decisions, and may lead to large reductions in inventory costs in comparison to an item approach. At present only a limited number of papers addressing lateral transshipments in the context of multi-item problems. (Archibald, 1997; Wong et al., 2005 & 2006).

(Archibald et al., 1997) considered a two-location, multi-item, multi-period, periodic review inventory system subject to a storage space limitation for all items. The demand is assumed to follow Poisson distribution and unlimited transshipments during a period in response to stockouts. (Wong et al., 2006) investigated a two-location, multi-item, continuous-review system for repairable items with one-for-one replenishments. The optimization problem is to determine stocking policies for all items that minimize the total system cost subject to a target level for the average waiting time for an arbitrary request for a ready-for-use part at each of the two locations. In their model, the decisions with respect to different items are coupled because of the multi-item service measure that is used. However, the solution procedure has a limitation since it requires a long computation time to solve rather large problems.

To overcome that limitation, (Wong et al., 2005) developed a simple and efficient solution procedure to obtain close-to-optimal solutions for the multi-item problem with lateral transshipments. The model is further extended to the case with multiple (and not limited to two) locations. Further, they also analyze the magnitude of the savings obtained by using the multi-item approach and lateral transshipments.

3. 2 Number of levels and locations in the system
Most of the previous study is focused on dealing with the transshipment problem in a two-echelon supply chain network, where it includes a single source supplier/warehouse at the higher level and multiple (two or more than two) retailers at the lower level. The assumptions for simple problem structure are necessary for the reason of computational tractability in the process of finding the optimal solution. Especially, the earlier study addressed relatively simple model with two stock outlets and/or one single period, thus limiting their practical application. To alleviate the loss of realism, the recent researchers have attempted heuristic approximation and/or simulation approaches in their analyses for the supply chain system with increased members.(for example, Robbinson, 1990; Dis & de kok, 1996; Needham & Evers, 1998; Tagaras, 1999; Chiou et al., 2007)

When lateral transshipment occurs only among the same level retailers, it is called intra-echelon. In contrast, transshipment can be conducted across different levels. For example, in case there are two or more suppliers/warehouses at the upper level, the retailers seek transshipment from other supplier/warehouse when its designated supplier/warehouse can not fulfill its emergency delivery request. Therefore, the stock shipping operation for each retailer can be regular replenishment from its designated supplier, intra-echelon lateral transshipment, or inter-echelon transshipment from the supplier of the other region. Both (Needham & Evers, 1998) and (Chiou et al., 2007) considered allowing inter-transshipment across two levels.

3.3 Constraints: space, capacity, and time
Space, capacity, and time constraints are three factors that can affect significantly the system performance, either costs or service level. Not many works have been done in the areas of
transshipment problem accounting for these factors. (Wong et al., 2006) investigated multi-item spare parts system, minimizing the total costs for inventory holding, lateral transshipments and emergency shipments subject to a target level for the average waiting time per demanded part at each of the two locations. In their model, the waiting time consideration is taken into account.

(Van Houtum & Zijm, 2000) classified inventory systems as two categories: service model and cost model. In a service model, the objective is to minimize the total system costs subject to a set of service level constraints, such as space, capacity, and time constraints. In a cost model, however, the service constraints are replaced with shortage penalty costs. Although in general the cost models are analytically more tractable, they have a serious limitation in that the penalty costs are generally hard to estimate. In this case, the service level constraints are constraints on the maximum expected waiting time. Hence they considered a service model rather than a cost model.

(Archibald et al., 1997) analyzed a multi-period, periodic-review model of a two-location inventory system in which lateral transshipments can occur at any time during the period. They formulated the two-location, single-item inventory problem as a Markov decision process and then extend the results to a two-location, multi-item inventory problem with limited storage space. In fact, this kind of optimization problem with space, capacity, and time constraints is appropriate to be analyzed by Lagrange relaxation (Porteus, 2002 and Wong et al., 2005). However, this problem is only a two-location problem. It can be extended to a problem with multiple locations.

3.4 Transshipment direction

Transshipment direction is associated to the concepts of shortage cost differentiation and the usage of substitution item. Most previous papers focus on transshipments that are not limited to one direction. However, in some cases the decision makers also have to consider unidirectional transshipments (see e.g., Tagaras & Cohen, 1992; Axsäter, 2003; Liu & Lee, 2007). Especially, (Axsäter, 2003) presented a simple technique for evaluating policies with unidirectional transshipment. That is, such transshipment policy is only allowed in one direction. Such policy can be of interest if the warehouses have very different shortage costs. It may be irrational to transship items from a warehouse with higher shortage cost to the warehouse with lower shortage cost. Another interpretation is substitution in an inventory system. When a demand for a low quality item cannot be met directly, the item can be replaced by another high quality item. The simulation study of their performance gives a good picture of how the considered lateral transshipments or substitutions affect the inventory system.

In contrast, (Liu & Lee, 2007) proposed Markovian models for multi-item base-stock inventory policies where uni-directional substitutions are allowed among part types. They identified two substitution cases: substitution of incoming demand and substitution of backlogged demand for spare part management. As the number of part types increases, computational effort required to solve the Markovian models increases rapidly. In order to reduce computation burden, an approximation approach based on the decomposition of multi-dimensional state transition is developed for systems with two or more part types. In addition, there are also a number of papers that consider substitution in inventory systems. (Bassok et al., 1999) provided exact results for a single-period Newsvendor type
model. Various substitution models are also analyzed in e.g., (Pasternack & Drezner, 1991), (Bitran & Dasu, 1992), (Gerchak et al., 1996), (Hsu & Bassok, 1999).

4. Methodology

The methodologies adopted for investigating transshipment models can be divided into two classes: analytical and simulation.

4.1 Analytical approach

In (Wong et al., 2005a), considering the transshipment problem with waiting time constraints, the system behavior with respect to an item i is independent of all other items and may be described by a two-dimensional Markov process. This problem is appropriate to be analyzed by Lagrange relaxation which was applied to general constrained optimization problems.

(Kukreja & Schmidt, 2005) analyzed a model for lumpy demand parts in a multi-location inventory system with transshipment by using analytical and simulation techniques. They derived analytical results for the mean and variance of the lead-time demand at various locations and then use simulation methodology to determine inventory control policies for such a system. In particular, when the demand cannot be met fully from the location’s on-hand stock, a dynamic programming recursion was used to and the lowest transshipment cost solution for satisfying demand at the location.

In (Wong et al., 2006), an integer-programming problem with a nonlinear objective function and non-linear constraints was structured for multi-item multi-location spare parts systems with lateral transshipment and waiting time constraints. Four different heuristics were developed and evaluated in terms of their total costs and computation times. The results showed that the greedy-type heuristic has the best performance.

In (Archibald, 2006), for a given replenishment decision, the problem of minimizing the long run average cost per period was modeled as a Markov decision process. The state of the system is the stock level in each of the locations at a review epoch. The decision is the number of items to order for each location. Due to the storage limit at the locations, the number of states and decisions are finite. Therefore, the problem is an infinite horizon, average cost Markov decision process with finite state and action spaces (see e.g., Puterman, 1994).

New approaches such as the game theory approach for solving the transshipment problem have drawn attention from researchers. For example, (Reyes, 2005) solved the transshipment problem for maintaining stable conditions in the logistics network by using the well-known Shapley value concept from cooperative game theory.

4.2 Simulation approach

Due to the complexities involved in the analytical modeling and solution of multi-echelon supply chain problems, some researchers have attempted heuristic approximations and/or simulation approaches, in efforts to maintain at least some degree of realism in their analyses.

(Needham & Evers, 1998) investigated the interaction of relevant costs and transshipment policies via simulation study and presented a method for determining a threshold value at which the benefits of transshipments outweigh their costs. They found that the cost of a
stockout is the primary determinant in the transshipment decision, with lower stockout cost levels generally decreasing the likelihood that transshipment usage. A meta-model was also proposed as a practical means of providing insight into when emergency transshipments should be employed. (Ozdemir et al., 2006) analyzed a capacitated transshipment problem. They modeled it as network flow problem embedded in a stochastic optimization problem. They tackled the problem by proposing a solution procedures based on infinitesimal perturbation analysis (IPA). IPA is an efficient simulation-based optimization technique (Ho et al., 1979). IPA-based methods have also been introduced to analyze supply chain problems (Glasserman & Tayur, 1995; Here et al., 2006). With IPA, the idea is to use the expected value of the sample path derivative obtained via simulation, instead of using the derivative of the expected cost, in a gradient search algorithm to update the order-up-to level for each stock location.

(Gong & Yucesan, 2006) utilized simulation optimization by combining an LP/Network flow in corporate with infinitesimal perturbation analysis (IPA) to analyze the problem, and obtained the optimal base stock quantities through sample path optimization. (Zhao & Sen, 2006) conducted a comparison of sample-path based simulation and stochastic decomposition for multi-location transshipment problems proposed by (Herer et al., 2006), considering one supplier, and N non-identical retailers who face uncertain customer demands. Each retailer reviews its own inventory periodically, and replenishes its stock by placing orders with the supplier. They investigated the performance of two methods: infinitesimal perturbation analysis with Stochastic Quasi Gradient (IPA/SQG) and Random Stochastic Decomposition (RSD). The computational results showed that while IPA/SQG and RSD methods provide solutions of similar quality, the amount of computational time required by RSD is significantly lower because it takes advantage of the special structure of the two-stage stochastic linear program.

5. Conclusion and directions for further research

The list of research papers dealing with transshipments to date is quite long and no attempt is made here to exhaust it. In this chapter we mainly focus on presenting a comprehensive description, classification, methodologies and solution procedures, and research directions for further study of transshipment in supply chain system. In view of the transshipment problems in a supply chain system, they can be characterized by four considerations: basic assumptions, transshipment policies, inventory control policies, and performance measurement. We further discuss the transshipment problem based on the features such as: the number of item(s) in inventory system, the number of locations in the pooling group; the storage space or waiting time constraint, and the direction of transshipment. Next, some literatures are discussed and classified into two methodologies of solution procedures, analytical and simulation. In this paper, we attempt to increase the understanding of the properties, characteristics, and methodologies of transshipment problem. Although numerous researches have contributed in this area, the investigated structure is much simpler than the practical. There still exits rich research opportunities for considering more complex systems with more echelons, items and locations. Some extensions are pointed out as follows.

The transshipment directions in two-location groups are much easier to specify than those of larger groups with more locations. Some alternative transshipment policies and priority rules are taken into account when there are multiple potential senders or receivers for each
transshipment request. In addition, the source of the transshipment system may be more than one in practice. There are multiple sources supplying to multiple warehouses, and each warehouse supplies a group of retailers. While the transshipment can be made across different echelon, not only at the same level, the transshipment sequence options and complexity of the network system increases significantly.

The effectiveness of a wider range of cost parameters and other methods of inventory control besides continuous review policies (S-1, S), (Q, R) and periodic review policies (s, S) need still further investigation.

Non-negligible trans-shipment times and the timing of transshipment incorporation with different inventory control policies are interesting topics for further research.

The space limit and the capacitated sources those considerations reflect more practical situation are also needed further study. Therefore, there still exists rich opportunity for further research in this topic.

The transshipment problem can be in corporate with vertical emergency shipment from two or more sources. Such policies combine inter-echelon emergency shipment and intra-echelon transshipment. In summary, there exist rich research opportunities in the areas of transshipment for supply chain systems.

6. Acknowledgements

This work was partially supported by the National Science Council of Taiwan, ROC under Grant No. NSC NSC-95-2213-E-212-048.

7. Reference


Chiou, C.-C. “A Simulation Study on the Effectiveness of Transshipment rules and Inventory Control Policies,” The 37th International Conference on Computers and Industrial Engineering, Alexandria, Egypt, October 20-23 2007


www.intechopen.com


Traditionally supply chain management has meant factories, assembly lines, warehouses, transportation vehicles, and time sheets. Modern supply chain management is a highly complex, multidimensional problem set with virtually endless number of variables for optimization. An Internet enabled supply chain may have just-in-time delivery, precise inventory visibility, and up-to-the-minute distribution-tracking capabilities. Technology advances have enabled supply chains to become strategic weapons that can help avoid disasters, lower costs, and make money. From internal enterprise processes to external business transactions with suppliers, transporters, channels and end-users marks the wide range of challenges researchers have to handle. The aim of this book is at revealing and illustrating this diversity in terms of scientific and theoretical fundamentals, prevailing concepts as well as current practical applications.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: