# An analytical theory of multi-echelon production/distribution systems

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science

Memorandum COSOR 89-10

# AN ANALYTICAL THEORY OF MULTI-ECHELON PRODUCTION/DISTRIBUTION SYSTEMS

L.J.G. Langenhoff and W.H.M. Zijm

Eindhoven University of Technology
Department of Mathematics and Computing Science
P.O. Box 513
5600 MB Eindhoven
The Netherlands

Eindhoven, April 1989
The Netherlands

# AN ANALYTICAL THEORY OF MULTI-ECHELON PRODUCTION/DISTRIBUTION SYSTEMS

L.J.G. Langenhoff* and W.H.M. Zijm#

Abstract

In this paper, we study inventory control problems arising in multi-echelon production/distribution chains. In these chains, material is delivered by outside suppliers, proceeds through a number of manufacturing stages, and is distributed finally among a number of local warehouses in order to meet market demand. Each stage requires a fixed leadtime; furthermore, we assume a stochastic, stationary end-item demand process.

The problem to balance inventory levels and service degrees can be modelled and analyzed by defining appropriate cost functions. Under an average cost criterion, we study the three most important structures arising in multi-echelon systems: assembly systems, serial systems and distribution systems. For all three systems, it is possible to prove exact decomposition results which reduce complex multi-dimensional control problems to simple one-dimensional problems. In addition, we establish the optimality of base-stock control policies.

* Eindhoven University of Technology, Department of Mathematics and Computing Science,

# Nederlandse Philips Bedrijven, B.V., Centre for Quantitative Methods
and
Eindhoven University of Technology, Department of Mathematics and Computing Science.

# 1. Introduction

This paper is concerned with the planning and control of the materials flow in an integrated production/distribution chain. At the beginning of the chain, components and raw materials are ordered from outside suppliers. Upon arrival, a certain amount may be used immediately in the subsequent manufacturing stage, whereas another part is temporarily stored in a component store, until it is required for manufacturing. The manufacturing stage may be subdivided in several phases, e.g. additional component manufacturing, subassembly and final assembly, separated by intermediate stockpoints. Finished goods are initially stored in a depot from which they are distributed to a number of local warehouses. Hence, this depot combines two important functions, it serves as a distribution center but at the same time it is used as a central warehouse. In particular, not all goods available at the depot <u>have to be</u> distributed among the local warehouses immediately; if not needed, a certain quantity is held at the depot or, in other words, the decision how to allocate this quantity is postponed, thus taking advantage of the latest available information about actual demand.

Distribution to local warehouses constitutes the final phase, i.e. we assume that the direct sales organisations are situated at these local warehouses from which finished goods find their way to the retail sector. When referring to the planning and control of the movement of materials and finished goods in such a chain, the phrase "logistics control" or "control of a logistic chain" is often used. Here logistics control is used in its broader industrial context, including both materials management and physical distribution management. An example of a logistic chain is pictured in fig. 1.

The control of the materials flow in such a chain generally tries to achieve a good balance between two important objectives. On the one hand we wish to attain a desired <u>customer service level</u>, which, in view of the uncertain demand, may for instance be expressed as the percentage of demand that can be satisfied immediately. Demand is defined here as orders from the retail sector which are placed at, and should be fulfilled from, the sales organizations at the local warehouses. Such a desired service degree naturally influences the amount of inventory, in particular the amount of <u>safety stock</u>, which should be available at any time at the local warehouses. Inventory is needed due to the presence of leadtimes, orders placed by the local warehouses at the depot generally require a certain distribution leadtime before they arrive at the local warehouses. If not available at the depot, this time may even increase significantly because the requested goods have to be produced and sent to the central depot, requiring again some time which is now referred to as the manufacturing leadtime. If the relevant components or raw materials are not immediately available, it may take an even unacceptable long time before the local warehouses are able to fulfill the requests of the retail sector; as a result of such a bad delivery performance a severe loss of market share may eventually occur. In other words, a desired customer service level at the market side induces the implicit definition of service levels at all inner stock points in the chain. The presence of leadtimes throughout the chain (often due to the fact that only a limited production capacity is available) requires the build up of inventories at least at the local warehouses and possibly at the depot and/or at the component stores.

The other side of the coin is that inventories represent a substantial amount of capital which cannot be used for other (investment) purposes, while in addition stockholding costs again consume a substantial amount of money. Hence, the second objective is to keep the total amount of inventory investment at the minimum possible level. To illustrate the importance of this, take a percentage of 20 % of the FAV (Factory Accounting Value) of finished goods as inventory holding costs (not uncommon in industrial companies); these costs result from insurance costs, materials handling and storage costs, interests and the like. Assuming a <u>turnover</u> (i.e. annual sales divided by the value of average inventory) of 4, and assuming also an annual net profit of 4 to 5 % of annual sales, it is easily found that in such a, not unrealistic, case the

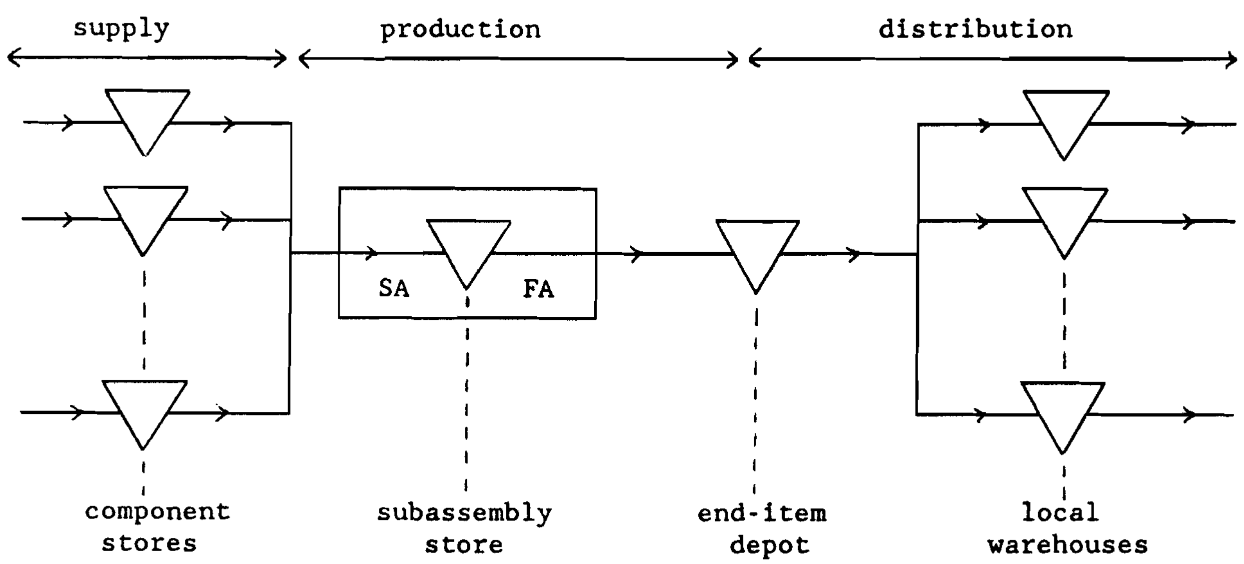


Fig. 1. Example of a logistic chain

amount of capital consumed by inventory holding costs equals the total net profit of an industrial company. Saving only ten percent of inventory implies a substantial amount of capital available for more profitable investments.

When discussing inventories, we refer to both the inventories held in intermediate stores and warehouses as well as work-in-process inventories, materials that are transported or transformed in some manufacturing process. High inventory levels at different stages are an indication of long leadtimes throughout the chain. For products with a relatively short commercial life cycle this may even result in final inventories that are obsolete, ultimately leading again to a loss of market share because the company is unable to serve the market right in time with the right products. Long leadtimes are an indication of the inflexibility of a manufacturing company to react adequately to rapidly changing market conditions. We will not investigate this matter further in this paper but it is the main reason for all efforts spent in cutting leadtimes as much as possible. Consequently, inventory levels have to decrease.

Let us briefly take a closer look at the structure of logistic chains as described above. Observe that rather than a simple serial structure one finds so-called assembly and distribution structures. Assembly structures typically arise when several types of components are needed to compose one final product type. Distribution structures indeed come up when a total amount of finished goods has to be distributed among a number of final locations, as in the logistic chain pictured in fig. 1, where the depot acts both as a central warehouse and as a distribution center. A large production quantity is ordered from the factory, upon arrival the goods are distributed among the local warehouses, or held in stock if not needed immediately. We emphasize this stockkeeping function of the depot as opposed to another situation, which again is modelled and analyzed as a distribution structure, but where the total arriving volume is always split up indeed.

This second interpretation of a distribution structure stems from a Hierarchical Production Planning context, facing an uncertain demand again. Hierarchical Production Planning systems reflect the fact that is does not make much sense generally to base production decisions on long term forecasts of specific end-item demand, whereas forecasts on an aggregate level (for a family of products) are much more reliable in general. Aggregate long term forecasts permit factories to agree with sales organisations on a certain production volume (on a family level basis), allowing them to order already some long leadtime components which can be used in a variety of end items within the family, while furthermore volume commitments are sufficient in general for a detailed capacity planning. Some periods later, the volume commitment has to be split up (disaggregated) into end item production quantities (the so-called mix replenishment); production of these end item quantities consumes an additional leadtime. The determination of volume commitments and at some later time the mix replenishment quantities constitutes a problem similar in structure as the pure distribution problem; a large total volume has to be split up, not to different destinations but to different end items (compare e.g. De Kok[1984]). However, contrary to the distribution case, here indeed the <u>total</u> volume is usually split up.

The literature on the planning and control of logistic chains is rapidly growing, although many contributions cover only certain aspects of the planning and control process. We will not try to give a complete review of the literature but only mention some important contributions which partly motivated our own work. Forrester's "Industrial Dynamics" is now recognized as a pathbreaking study on the cyclical variation of stocks in large production-/distribution chains (Forrester[1961]). Clark and Scarf[1960] proved the optimality of a Base Stock Control rule for serial inventory systems. They introduced the concept of echelon stock (all stock in a stockpoint <u>and downstream</u>, see the next section) and showed that an integrated control system should be based on this echelon stock rather than on local stock (and local forecasts) only. Their analysis is based on Dynamic Programming and assumes a Discounted Cost structure. Eppen and Schrage[1981] discuss control policies in a "one depot/multi-warehouse" situation where the depot acts as a pure distribution center (no stockkeeping function). They investigate, under an average cost criterion, whether such a (possibly only administrative) depot can play a role in decreasing safety stocks, under a balancing assumption, to be discussed below. Federgruen and Zipkin[1984] discuss approximate inventory allocation policies for these systems, see also Zipkin[1984]. In the same spirit as Clark and Scarf, assuming again a discounted cost structure, Schmidt and Nahmias[1985] analyze a two stage assembly system (with two components only) in detail.

The above studies have in common that they allow safety stocks to be present at all upstream stockpoints, and sometimes even explicity <u>advocate</u> to installation of stocks at upstream points rather than on downstream points, either because upstream stocks are cheaper (less added value) or because upstream stocks still can be allocated to a variety of final destinations, or a variety of different final products in case of commonality of components. Comparing this with widespread systems like Material Requirements Planning (MRP, see e.g. Orlicky[1975]) and its extension to Manufacturing Resources Planning (MRP II, see e.g. Wight[1981]), we see quite the opposite. To be fair, it should be noted that MRP systems are mainly intended to plan and control the factory production process, hence only a part of the logistic chain. But the main point is that MRP systems explicitely forbid safety stocks at any level upstream of the so-called Master Production Scheduling (MPS) level. A Master Production Schedule usually defines the production quantities of end-items (or subassemblies if final assembly is done on an assembly-to-order basis), safety stocks may be situated only at that level whereas all upstream production and procurement actions are driven (pushed) by this MPS, in a deterministic sense by using the product structure (the so-called Bill of Materials or BOM). However, in case an MPS is defined at end-item level, forecasts of end-item demands are needed for a horizon which covers the entire leadtime from ordering materials from outside suppliers up to final product manufacturing. The Eppen and Schrage study described above clearly showed that this may be a serious disadvantage. An important qualitative study in this respect is the paper by Whybark and Williams[1976]. A more complete treatment can be found in Vollmann, Berry and Whybark[1984], they discuss among others MRP in the context of a complete Manufacturing Planning and Control system.

Hierarchical Production Planning (HPP) systems have been proposed and analyzed by e.g. Hax and Meal[1975] and Bitran, Haas and Hax[1981], [1982] (the last paper dealing with two stage systems). A review of the literature on HPP can be found in Hax and Candea[1984], ch. 6. Again, these authors mainly address the production field, but they clearly recognize the need to make decisions on an aggregate level first, before getting into details on a mix replenishment level. Their approach is based exclusively on mathematical programming techniques, in particular the stochastic environment is not considered explicitely (e.g. safety stock levels have to be defined in advance). Also, commonality aspects are not dealt with. On the other hand, limited capacity (not present in most stochastic models) plays an explicit role in their analysis. An interesting attempt to combine an MRP approach with a HPP philosophy has been made by Meal, Wachter and Whybark[1987].

Finally, we mention other approaches which have drawn attention recently. Just In Time (JIT) systems are mainly suitable for controlling production in a relatively stable repetitive manufacturing environment (cf. Schonberger[1982].

An example of an implementation of such a (pull) JIT philosophy is the well-known Kanban system introduced by Toyota. JIT can be (and has been) extended to procurement of raw materials but is certainly not suitable for controlling an entire logistic chain. Optimized Production Technology (OPT) is a production control technique based on two simple rules: 1. assure an efficient use of bottleneck capacities (i.e. avoid idleness) by installing buffers in front of these capacities, and 2. do nòt dispatch work to a manufacturing system at a rate higher than can be handled by the bottleneck (see Goldratt[1988]). We mention OPT here because it is often viewed as an alternative to MRP or Base Stock Control, although in our opinion it is not more than an interesting shopfloor scheduling technique.

We conclude this section with an outline of the contents of this paper. We intend to present a unified analytical treatment of the control of an integrated production/distribution system as described above, in a stochastic environment (i.e. under stochastic demand). The analysis mainly draws on above mentioned work of Clark and Scarf[1960], Eppen and Schrage[1981] and Schmidt and Nahmias[1985], but considers an average cost criterion. This choice is based on several arguments. The three papers just mentioned together indeed cover the entire logistic chain, they present a thorough analysis of the control problem, optimality of echelon stock based inventory policies has been studied explicitely (although under restrictive conditions and for simple systems only), these policies have an intuitively appealing interpretation, and stochastic aspects are explicitely modelled. All other approaches consider only parts of the chain and generally lack a thorough mathematical analysis (except for Hierarchical Production Planning).

In section 2, we state our assumptions, formulate the model and briefly consider single echelon systems. Section 3 is devoted to an average cost analysis of serial line systems. A careful definition of echelon-based holding and penalty cost functions appears to be crucial in the derivation of decomposition results for echelon systems; the structure of these functions however is not immediately evident. Explicit proofs are given of the main results; these proofs are believed to be new. In section 4, we establish, under an average cost criterion, results for assembly structures similar to those obtained by Schmidt and Nahmias[1985]. The formulation in terms of average costs permits a relatively easy extension of the results to systems with more than two components, an extension which is not very easy to make under a discounted cost structure as was noted by Graves[1988]. A multi-echelon distribution structure is discussed in section 5. In particular, the notion of (un)balancedness of the stock positions in several final warehouses will get attention. In sections 4 and 5, we elaborate on structural aspects of the models but no proofs are presented; although somewhat more complicated, the central ideas of these proofs are similar to those given in section 3. In section 6 finally, we mention possible extensions and recommendations for future research.

## 2. Model and assumptions. The single installation system

Throughout this paper, we will assume that demand originates at the lowest installations (i.e. the installations at the downstream side of the logistic chain only). Although this assumption is not strictly necessary (indeed it can be shown that the forthcoming analysis remains valid if for instance outside demand for components is allowed), the notations and analysis complicate considerably. Demand in subsequent periods is assumed to be represented by i.i.d. (independently, identically distributed) random variables (hence demand is assumed to be stationary). The inventory holding cost, as well as the penalty cost at the demand side of the chain will be linear (with respect to both time and quantity). This is not really a restriction. As noted in the introduction one is generally interested in attaining a desired customer service level while at the same time minimizing inventory levels. By choosing appropriate linear holding and penalty cost functions, one easily arrives at the desired objectives (there is a one-to-one relationship between costs and service criteria, see e.g. Silver and Peterson[1985]). Also, it is well known for single installation models (Scarf[1960]) that convexity of the one-period penalty and holding cost function is generally sufficient to prove the optimality of base stock policies. Indeed, our analysis will be heavily based on convexity properties of appropriate functions.

A further assumption concerns the absence of fixed production and distribution costs. In a second paper on serial systems, Clark and Scarf[1962] mention that this assumption, made in their earlier paper, was seriously criticized, and they present approximations to deal with the more difficult set-up cost case. However, in many large industrial companies, decisions on replanning frequencies are made on a higher, strategic level. In general, sales organizations order once per month, or once per two weeks, and decisions on production planning (both on an aggregate level as with respect to product mix) are made with the same frequency (and then cover the full product range indeed). Therefore, fixed costs are already taken account of on a higher level, determining the planning and replanning frequencies on the lower, operational levels.

Finally, we assume that all excess demand is backlogged.

Before we continue let us define the terms echelon stock and echelon inventory position more precisely. Following Clark[1958], we use the term "echelon stock" for all stock at a given installation plus in transit to or on hand at any installation downstream minus the backlogs at the most downstream installations (which do not have a successor). The chain under consideration is called the echelon. Note that an echelon stock may be negative, indicating that the backlogs are larger than the total inventory in that echelon. Consider fig. 2. The echelon stock, associated with warehouse 1, is simply the stock or backlog in that warehouse. The echelon stock, associated with depot 3, is the stock in that depot plus any stock on hand at or in transit to either warehouse 1 or 2. The echelon stock, associated with component store 4, includes all components on hand in that component store, plus all components on hand at or in transit to any downstream installation, no matter whether these components have been used already in assembled products. The "echelon inventory position" finally denotes the echelon stock plus the materials already ordered but not yet available at the highest (most upstream) installations. When an echelon consists of one installation only, this definition coincides with the one given in Silver and Peterson[1985]. Echelons are numbered according to the highest installation in that echelon.

The analysis of the single installation (or warehouse) model will prove to be a basic building block in this paper. Therefore, we start with a brief

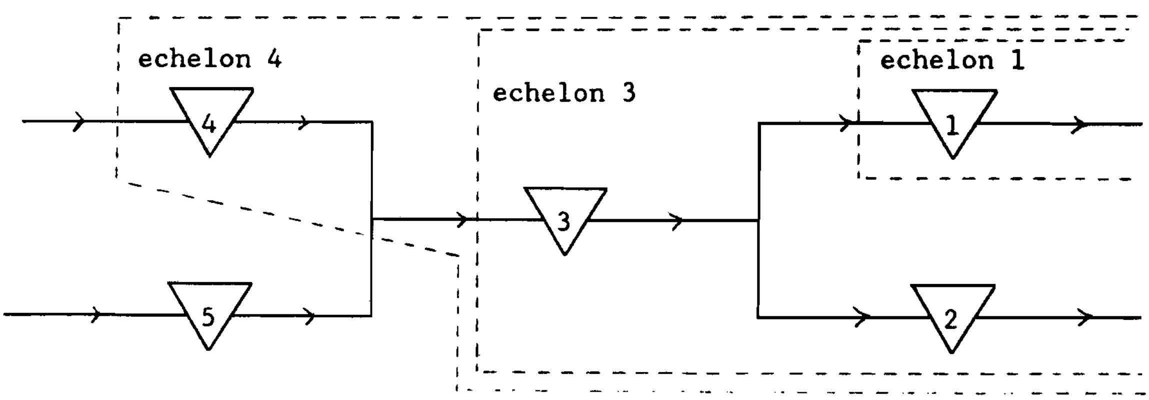


Fig. 2. Echelons associated with installations.

discussion of this model, in which demand for a single product has to be met immediately from stock. If demand cannot be met immediately, it is backlogged. Once per period, a stockholder may order goods from an outside supplier. We assume that the supplier is always able to satisfy any demand, and that it takes exactly $\ell$ periods (the delivery leadtime) until ordered goods arrive at the warehouse. Ordering costs for the stockholder are c per unit (these costs are assumed to be linear), no fixed ordering costs are assumed. Holding and shortage (penalty) costs are h and p respectively, per unit and per period (again we assume linearity of holding and penalty costs).

Let a denote the physical stock available at the warehouse at the beginning of a period, just after goods have arrived. If $a < 0$, a denotes a backlog. Let F denote the distribution function of the one period demand. The expected holding and penalty costs at the end of a period, L(a) say, can be expressed by the well-known Newsboy formulas

$$L(a) = h\int_0^a (a-u)dF(u) + p\int_a^\infty (u-a)dF(u) \qquad \text{for } a \geq 0,$$

$$L(a) = p\int_0^\infty (u-a)dF(u) \qquad \text{for } a < 0.$$

Note that L(a) is convex, and that $L(a) \to \infty$ for $|a| \to \infty$.

Let, at the <u>beginning</u> of period t, x denote the inventory position and suppose it is decided to raise this inventory position to y (where $y \geq x$). The costs, resulting from such a decision, can be expressed as

$$c(y-x) + \int_0^\infty L(y-u_\ell)dF_\ell(u_\ell)$$

where $F_\ell$ denotes the distribution function of the $\ell$-period cumulative demand and $\ell$ is the (fixed) delivery leadtime. Here the first term stems from the ordering costs while the second term expresses the expected holding and penalty costs at the <u>end</u> of period $t+\ell$. If now, at the beginning of <u>every</u> period, the inventory position is increased to y, the associated average costs of such a policy for the infinite horizon problem can be written as

$$c\mu + \int_0^\infty L(y-u_\ell)dF_\ell(u_\ell) \qquad (2.1)$$

where $\mu$ denotes the expected one period demand.

It has been shown (Scarf[1960], Iglehart[1963]) that the optimal average cost policy for the infinite horizon problem is indeed a policy with the structure described above, i.e. at the beginning of each period the inventory position is raised to the same level, S say. S is determined as the infimum of those values which minimize expression (2.1) (note that the ordering costs do not play any role in the minimization). Formally, if we define

$$D(y) = \int_0^\infty L(y-u_\ell)dF_\ell(u_\ell) \tag{2.2}$$

and S is the smallest value which minimizes D(y), then the optimal average cost policy for the infinite horizon problem can be stated as follows:

| | |
|---|---|
| order (S-x) | if $x \le S$ , |
| order nothing | if $x > S$ . |

Clearly all states $x > S$ are transient. The optimal average costs are equal to $c\mu + D(S)$. Note that D(y) is convex and that $D(y) \to \infty$ for $|y| \to \infty$ again. The fact that the ordering costs do not play any role in the minimization will be intuitively clear but it is worth to mention here that it depends heavily on the assumption of complete backlog (which implies that all demand has to be ordered eventually). The fact that this feature is no longer present in lost sales models make the latter essentially more difficult to analyze.

In the model presented here it is not hard to understand why a control limit policy of the type described above is optimal. The fact that the one period costs, apart from the ordering costs which do not play an essential role, are completely determined by the order-up-to level, and not by the initial inventory level, together with the convexity properties of D(y), are in fact sufficient. Rigorous proofs can be given by means of Dynamic Programming arguments (Scarf[1960]) which remain valid in the case of fixed ordering costs K (independent of the order size). The N-period cost functions, in which the action for the first period still can be chosen whereas the remaining N-1 periods are controlled optimally, then satisfy a so-called K-convexity property. This in turn is sufficient to prove that the optimal policies are still of a control limit type. The optimal policy is characterized by two critical numbers, s and S say, such that, if the inventory position before ordering x is smaller that s, an amount S-x is ordered whereas nothing is done in the other case.

It is also important to note that the linear holding and penalty costs are not essential in the above analysis; sufficient is the fact that D(y) is convex. This property will be basic in the forthcoming analysis where we have to analyze more general cost functions, arising from attributing extra penalties to higher installations if they are not able to respond properly to requests of lower installations.

Finally, we remark that the above analysis can be straightforwardly extended to the case in which the delivery leadtimes $\ell$ are stochastic, under the condition that an order placed in period t cannot arrive later than an order placed in period t+1. Stochastic leadtimes offer a possibility to model the unability of a supplier to deliver requested goods within $\ell$ periods, due to e.g. the unavailability of materials in case of a manufacturing process. However, in a multi-echelon situation, such a model ignores the explicit correlation between out of stock positions at higher (upstream) stockpoints and (too) late deliveries and therefore backlogs occurring at lower echelons. These models therefore do not sufficiently explain the cyclic movements of stocks observed by Forrester[1960]). Models which incorporate the above indicated correlation can be expected to provide a better framework for analyzing multi-echelon situations.

## 3. <u>Average cost analysis of serial systems</u>

In this section, we present a detailed analysis of the control problem for serial production/inventory systems, under the assumptions described above, for the average cost criterion.

Consider first the two stage system, shown in fig. 3. Let $\ell_1$ denote the delivery leadtime for goods ordered by installation 1, <u>if these goods are available at installation 2</u>, and let $\ell_2$ denote the delivery leadtime for goods ordered by installation 2 (which always can be supplied). As before, let $F_\ell$

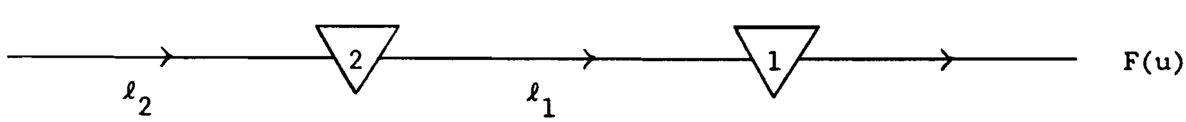


Fig. 3. A two echelon serial system.

denote the distribution of the $\ell$-period cumulative demand, for all $\ell$. If $\ell=1$, we suppress the index. Demand originates in the system at the lowest installation only. By $h_2$ we denote the (linear) holding charge associated with all stock in installation 2 and in transit to installation 1, and by $h_1+h_2$ the holding charge in installation 1. An additional charge $h_1$ is natural since these charges are generally calculated as a percentage of the value of the product (compare the remarks in the introduction), added value therefore implies an additional holding charge. In case of a stockout at installation 1, a (linear) penalty cost p is incurred. Finally, we do not assume any variable purchasing (or ordering) costs, since, due to the complete backlog assumption, these costs do not play an essential role in minimizing average costs for the infinite horizon problem (compare section 2).

Before we set up equations for the cost functions, let us see how inventory costs can be attributed to echelons instead of to installations. This will provide some intuition for choosing appropriate echelon cost functions which subsequently allow an exact decomposition. Clark and Scarf[1960] did not relate echelon cost functions to more traditional holding and penalty costs for single installations but it is well known that, in each part of the chain, holding costs have to be attributed only to the value added in that part. However, when backlogs occur, it is less intuitively obvious how to transform installation cost functions to echelon cost functions. Nevertheless, this translation will prove to be the key to the derivation of decomposition results as we well see in the sequel.

Let $x_1$ and $x_2$ denote the echelon stock associated with installation 1 and installation 2 respectively (hence $x_2$ is all stock at installation 2 and anywhere downstream, minus eventually existing backlogs). Clearly $x_1 \le x_2$. We may distinguish the following cases.

1. $0 \le x_1 \le x_2$. No backlogs exist. Note that the stock at installation 2 plus the stock in transit between the two installations equals $x_2-x_1$. Inventory holding costs are therefore equal to $(h_1+h_2)x_1 + h_2(x_2-x_1) = h_1x_1 + h_2x_2$.
2. $x_1 \le 0 \le x_2$. A backlog exists at installation 1. The physical stock at installation 2 plus the stock in transit between the two installations equals $x_2-x_1$. Holding costs are hence equal to $h_2(x_2-x_1) = -h_2x_1 + h_2x_2$.
3. $x_1 \le x_2 < 0$. Again a backlog exists. Although $x_2 < 0$ there still may be an amount of stock equal to $x_2-x_1$ at installation 2 plus in transit to installation 1. Holding costs are again equal to $h_2(x_2-x_1) = -h_2x_1 + h_2x_2$.

Summarizing, the above analysis suggests that an inventory holding cost $h_2x_2$ may be attributed to echelon 2, independent of the sign of $x_2$. With respect to echelon 1 we find that an inventory holding cost $(h_1+h_2)x_1-h_2x_1$ is attributed if $x_1 \geq 0$, and a cost $-h_2x_1$ if $x_1 < 0$.

Next, we define the one-period holding and penalty costs for both echelons. Let

$$L_1(x_1) = (h_1+h_2)\int_0^{x_1}(x_1-u)dF(u) + p\int_{x_1}^{\infty}(u-x_1)dF(u) - h_2x_1 \qquad \text{if } x_1 \geq 0,$$

$$L_1(x_1) = p\int_0^{\infty}(u-x_1)dF(u) - h_2x_1 \qquad \text{if } x_1 < 0,$$

and

$$L_2(x_2) = h_2x_2 \qquad \text{for all } x_2.$$

The cost function $\hat{L}_1(x_1)$, associated with the real costs incurred at echelon 1 (i.e. installation 1), can be written as

$$\hat{L}_1(x_1) = (h_1+h_2)\int_0^{x_1}(x_1-u)dF(u) + p\int_{x_1}^{\infty}(u-x_1)dF(u) \qquad \text{if } x_1 \geq 0,$$

$$\hat{L}_1(x_1) = p\int_0^{\infty}(u-x_1)dF(u) \qquad \text{if } x_1 < 0,$$

from which we immediately deduce that $L_1(x_1) = \hat{L}_1(x_1) - h_2x_1$, for all $x_1$.

Following the analysis in the preceding section, let us at the beginning of period t decide to return the echelon inventory position $x_2$ of echelon 2 to a level $y_2$ (we suppose $x_2 \leq y_2$). At the beginning of period $t+\ell_2$ the echelon stock of echelon 2 will then be equal to $y_2-u_{\ell_2}$ (where $u_{\ell_2}$ denotes the demand in periods t, t+1, ... , $t+\ell_2-1$). At that moment, let us decide to raise the echelon inventory position $x_1$ of echelon 1 (installation 1) to $y_1$. Note that the quantity $y_1-x_1$ can only be shipped if $y_2-u_{\ell_2} \geq y_1$ (the echelon stock of echelon 2 must be at least the order-up-to level of echelon 1). As usual, if $y_2-u_{\ell_2} < y_1$ we will ship as much as possible to installation 1, while the remaining amount is backlogged. From the definition of echelon inventory positions it follows immediately that we must have $y_2 \geq y_1$.

As a result of these two decisions, we will, at the end of period $t+\ell_2+\ell_1$, be confronted with an expected holding and penalty cost function (where the holding costs cover all system inventory) which we denote by $D^{(2)}(y_1,y_2)$ (defined only on $y_1 \leq y_2$). The question of interest is the calculation of the optimal values of $y_1$ and $y_2$. First we need the following theorem:

Theorem 3.1. The cost function $D^{(2)}(y_1,y_2)$ can be written as

$$D^{(2)}(y_1,y_2) = C_1(y_1) + C_2(y_1,y_2)$$

where

$$C_1(y_1) = \int_0^\infty L_1(y_1-u_{\ell_1})dF_{\ell_1}(u_{\ell_1})$$

$$C_2(y_1,y_2) = \int_0^\infty L_2(y_2-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) + \int_{y_2-y_1}^\infty [C_1(y_2-u_{\ell_2})-C_1(y_1)]dF_{\ell_2}(u_{\ell_2})$$

Proof. Recall that $D^{(2)}(y_1,y_2)$ is defined only for $y_1 \leq y_2$. Consider the two cases $y_2-u_{\ell_2} \geq y_1$ and $y_2-u_{\ell_2} < y_1$. In the first case, the stock of echelon 2 is sufficient to satisfy the demand of echelon 1. A quantity $y_2-u_{\ell_2}-y_1$ is left at installation 2 and will not be at installation 1 before the start of period $t+\ell_2+\ell_1+1$ (and hence will be charged at the end of period $t+\ell_2+\ell_1$ at a rate $h_2$). In the second case, the amount of stock available at or in transit to installation 1 (not the inventory position) will be raised to $y_2-u_{\ell_2}$ and nothing is left at installation 2.

Furthermore, note that in periods $t+\ell_2+1, \ldots, t+\ell_2+\ell_1$ goods have arrived in the system (but not in installation 1 yet) which have been ordered in periods $t+1, \ldots, t+\ell_1$ respectively. Since all demanded products have to be ordered eventually, we may attribute a total charge $h_2\ell_1\mu$ to these goods. Now, define

$$\hat{D}^{(1)}(y) = \int_0^\infty \hat{L}_1(y-u_{\ell_1})dF_{\ell_1}(u_{\ell_1})$$

then we may write

$$D^{(2)}(y_1,y_2) = \int\limits_{y_2-u_{\ell_2}\geq y_1} [\hat{D}^{(1)}(y_1) + h_2(y_2-u_{\ell_2}-y_1)]dF_{\ell_2}(u_{\ell_2}) +$$

$$+ \int\limits_{y_2-u_{\ell_2}<y_1} [\hat{D}^{(1)}(y_2-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) + h_2\ell_1\mu =$$

$$= \int_0^\infty [\hat{D}^{(1)}(y_1) + h_2(y_2-u_{\ell_2}-y_1)]dF_{\ell_2}(u_{\ell_2}) + h_2\ell_1\mu +$$

$$+ \int\limits_{y_2-u_{\ell_2}<y_1} [\hat{D}^{(1)}(y_2-u_{\ell_2}) - \hat{D}^{(1)}(y_1) - h_2(y_2-u_{\ell_2}-y_1)]dF_{\ell_2}(u_{\ell_2}) =$$

$$= \int_0^\infty \hat{L}_1(y_1-u_{\ell_1})dF_{\ell_1}(u_{\ell_1}) - h_2(y_1-\ell_1\mu) + \int_0^\infty h_2(y_2-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) +$$

$$+ \int\limits_{y_2-u_{\ell_2}<y_1} \int_0^\infty [\hat{L}_1(y_2-u_{\ell_2}-u_{\ell_1}) - \hat{L}_1(y_1-u_{\ell_1}) - h_2(y_2-u_{\ell_2}-y_1)]$$

$$dF_{\ell_1}(u_{\ell_1})dF_{\ell_2}(u_{\ell_2}) =$$

$$= \int_0^\infty [\hat{L}_1(y_1-u_{\ell_1}) - h_2(y_1-u_{\ell_1})]dF_{\ell_1}(u_{\ell_1}) + \int_0^\infty h_2(y_2-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) +$$

$$+ \int\limits_{y_2-u_{\ell_2}<y_1} \int_0^\infty [\{\hat{L}_1(y_2-u_{\ell_2}-u_{\ell_1}) - h_2(y_2-u_{\ell_2}-u_{\ell_1})\}$$

$$- \{\hat{L}_1(y_1-u_{\ell_1}) - h_2(y_1-u_{\ell_1})\}]dF_{\ell_1}(u_{\ell_1})dF_{\ell_2}(u_{\ell_2}) =$$

$$= \int_0^\infty L_1(y_1-u_{\ell_1})dF_{\ell_1}(u_{\ell_1}) + \int_0^\infty L_2(y_2-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) +$$

$$+ \int\limits_{y_2-u_{\ell_2}<y_1} \int_0^\infty [L_1(y_2-u_{\ell_2}-u_{\ell_1}) - L_1(y_1-u_{\ell_1})]dF_{\ell_1}(u_{\ell_1})dF_{\ell_2}(u_{\ell_2}) .$$

By substituting in the last equations the functions $C_1$ and $C_2$ the theorem is proved.

□

Theorem 3.1 indicates already the possibility to decompose the system. It is easily shown that

$$\frac{\partial}{\partial y_1} D^{(2)}(y_1,y_2) = C_1'(y_1)F_{\ell_2}(y_2-y_1) \tag{3.1}$$

$$\frac{\partial}{\partial y_2} D^{(2)}(y_1,y_2) = \frac{\partial}{\partial y_2} C_2(y_1,y_2) = h_2 + \int_{y_2-y_1}^{\infty} C_1'(y_2-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) \tag{3.2}$$

$$\frac{\partial}{\partial y} D^{(2)}(y,y) = h_2 + \int_0^{\infty} C_1'(y-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) \tag{3.3}$$

Furthermore, $C_1(y_1)$ is convex (since $L_1(y_1)$ is convex). Let $S_1$ be the infimum of the values which minimize $L_1(y_1)$. Then, by taking second derivatives, one easily verifies that also $C_2(S_1,y_2)$ is a convex function in $y_2$. Finally, $D^{(2)}(y,y)$ is convex in y. These observations will prove to be helpful in determining the global minimum of $D^{(2)}(y_1,y_2)$. We have

<u>Theorem</u> 3.2. Let $C_1(y_1)$ be minimized in $S_1$, $C_2(S_1,y_2)$ in $S_2$ and $D^{(2)}(y,y)$ in $S_{1,2}$. Then the global minimum of $D^{(2)}(y_1,y_2)$ on $\{(y_1,y_2) \mid y_1 \le y_2\}$ is reached in $(S_1,S_2)$ if $S_1 \le S_2$ and in $(S_{1,2},S_{1,2})$ if $S_1 > S_2$.

<u>Proof</u>. In order to determine all possible minima of the function $D^{(2)}(y_1,y_2)$ we set its partial derivatives to zero. For convenience, we assume $F(u) = 0$ for $u = 0$ and $F(u) > 0$ for $u > 0$. Other cases can be handled analogously but will not be discussed here. From equations (3.1), (3.2) and (3.3) and the definitions of $S_1$, $S_2$ and $S_{1,2}$ we find

$$\frac{\partial}{\partial y_1} D^{(2)}(y_1,y_2) = 0 \quad \Rightarrow \quad y_1 = S_1 \text{ or } y_1 = y_2,$$

$$\frac{\partial}{\partial y_2} D^{(2)}(S_1,y_2) = 0 \quad \Rightarrow \quad y_2 = S_2,$$

$$\frac{\partial}{\partial y_2} D^{(2)}(y_2,y_2) = 0 \quad \Rightarrow \quad y_2 = S_{1,2}.$$

If $S_1 > S_2$, the solution $(S_1,S_2)$ is not feasible hence in this case the minimum of $D^{(2)}(y_1,y_2)$ on its defining domain is reached in $(S_{1,2},S_{1,2})$. The same holds if $S_1$ does not exist (i.e. $S_1 = \infty$, for instance when $h_1 = 0$). The more interesting case arises when $S_1 < \infty$ and $S_1 \le S_2$. In this case, two local minima exist, in $(S_1,S_2)$ and in $(S_{1,2},S_{1,2})$ respectively, and we have to prove that the global minimum is reached in the first point.

To this end, we first show that $S_{1,2} \ge S_1$ if $S_1 \le S_2$ and $S_1 < \infty$. Since $C_2(S_1,y_2)$ is convex in $y_2$ it follows that

$$h_2 + \int_{y_2-S_1}^{\infty} C_1'(y_2-u_{\ell_2})dF_{\ell_2}(u_{\ell_2})$$

is monotone non-decreasing in $y_2$. Since $S_1 \le S_2$ we have

$$0 = h_2 + \int_{S_2-S_1}^{\infty} C_1'(S_2-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) \ge h_2 + \int_{0}^{\infty} C_1'(S_1-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) .$$

But the latter expression is equal to the derivative of $D^{(2)}(y_2,y_2)$ in $S_1$ (compare equation (3.3)). Since this derivative is monotone non-decreasing it follows immediately that $S_{1,2} \ge S_1$.

Next, we show that the global minimum of $D^{(2)}(y_1,y_2)$ is attained in $(S_1,S_2)$. Note that, since $C_1(y_1)$ is convex and since a distribution function is always nonnegative, equation (3.1) implies that, for fixed $y_2 \ge S_1$,

$$\frac{\partial}{\partial y_1} D^{(2)}(y_1,y_2) \ge 0 \qquad \text{for } S_1 \le y_1 \le y_2.$$

Since this holds for any $y_2 \ge S_1$, we have in particular

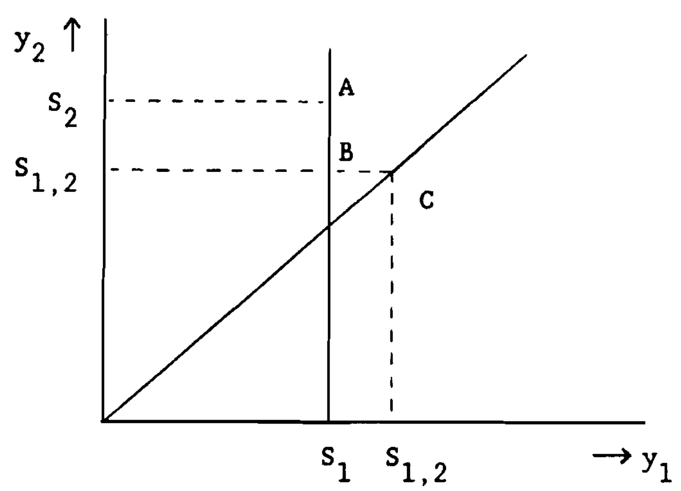


Fig. 4. $A = (S_1,S_2)$, $B = (S_1,S_{1,2})$, $C = (S_{1,2},S_{1,2})$.

$$D^{(2)}(S_1,y_2) \le D^{(2)}(y_2,y_2) \qquad \text{for all } y_2 \ge S_1. \qquad (3.4)$$

Furthermore, since $C_2(S_1,y_2)$ is convex, the same holds for $D^{(2)}(S_1,y_2)$. Therefore

$$D^{(2)}(S_1,S_2) \le D^{(2)}(S_1,y_2) \qquad \text{for all } y_2 \ge S_1. \qquad (3.5)$$

Since $S_{1,2} \ge S_1$, a combination of (3.5) and (3.4) (see also fig. 4) yields

$$D^{(2)}(S_1,S_2) \le D^{(2)}(S_1,S_{1,2}) \le D^{(2)}(S_{1,2},S_{1,2})$$

which was the result to be proved.

□

In fig. 4 it was implicitely assumed that $S_{1,2} \le S_2$. Indeed we have

<u>Lemma</u> 3.3. If $S_1 < \infty$ and $S_1 \le S_2$ then $S_1 \le S_{1,2} \le S_2$.

<u>Proof</u>. Since $C_1'(y) \ge 0$ for $y \ge S_1$, we have

$$h_2 + \int_{y-S_1}^{\infty} C_1'(y-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) \le h_2 + \int_0^{\infty} C_1'(y-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) .$$

Substituting $y = S_{1,2}$ and recalling the definition of $S_{1,2}$ yields

$$0 = h_2 + \int_0^{\infty} C_1'(S_{1,2}-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) \ge h_2 + \int_{S_{1,2}-S_1}^{\infty} C_1'(S_{1,2}-u_{\ell_2})dF_{\ell_2}(u_{\ell_2}) .$$

The latter expression is equal to the value of $\frac{\partial}{\partial y_2} D^{(2)}(S_1,y_2)$ in $S_{1,2}$. Since $\frac{\partial}{\partial y_2} D^{(2)}(S_1,y_2)$ is monotone non-decreasing it follows that $S_2 \ge S_{1,2}$.

□

If we define

$$\Lambda_2(y) = [C_1(y) - C_2(S_1)]*I_{\{y<S_1\}}$$

with I an indicator function, then the function $D^{(2)}(S_1,S_2)$ may be written as

$$D^{(2)}(S_1,S_2) = \int_0^{\infty} L_1(S_1-u_{\ell_1})dF_{\ell_1}(u_{\ell_1}) + \int_0^{\infty} (L_2(S_2-u_{\ell_2})+\Lambda_2(S_2-u_{\ell_2}))dF_{\ell_2}(u_{\ell_2})$$

Here $\Lambda_2(y)$ can be interpreted as the expected increase in costs at installation 1, due to insufficient stock in echelon 2. Adding these costs then to echelon 2 yields an artificial penalty function for stockouts in installation 2. Note that $L_2(y) + \Lambda_2(y)$ is again convex.

Theorem 3.2 shows that $D^{(2)}(y_1,y_2)$ can be minimized by subsequently solving one-dimensional problems: first we minimize $C_1(y_1)$, yielding $S_1$, and next $C_2(S_1,y_2)$, yielding $S_2$. If $S_1 \le S_2$ (the "normal" case) we are finished, if not we minimize the one-dimensional function $D^{(2)}(y,y)$, yielding $S_{1,2}$.

Up to now, we only solved a one-period problem; we have minimized the expected holding and penalty costs in period $t+\ell_1+\ell_2$. However, $y_1$ and $y_2$ have been chosen arbitrarily in advance (in particular $y_1$ might depend on $x_2$). Subsequently, $D^{(2)}(y_1,y_2)$ was minimized. Since no ordering costs are present (and since there are no capacity restrictions) the resulting minimum costs clearly do not depend on $x_1$ and $x_2$, the echelon inventory positions at the beginning of period t just prior to ordering. Hence, we may repeat the decisions, associated with the minimum value of $D^{(2)}(y_1,y_2)$ in every period, resulting in an average cost optimal policy for the infinite horizon problem.

The extension of theorem 3.1 to an N-echelon serial system is straightforward. Define for these systems

$$L_1(x_1) = \sum_{n=1}^{N}(h_n)\int_0^{x_1}(x_1-u)dF(u) + p\int_{x_1}^{\infty}(u-x_1)dF(u) - \sum_{n=2}^{N} h_n x_1 \quad \text{if } x_1 \ge 0,$$

$$L_1(x_1) = p\int_0^{\infty}(u-x_1)dF(u) - \sum_{n=2}^{N} h_n x_1 \quad \text{if } x_1 < 0,$$

$$L_n(x_n) = h_n x_n \quad \text{for all } x_n \quad (n = 2,\dots,N).$$

Then the following result can be proved along the same lines as theorem 3.1.

Theorem 3.3. Consider a policy which, at the beginning of every period, increases the echelon inventory position of echelon n to $y_n$ $(n = 1,2,\dots,N)$. Let $D^{(N)}(y_1,y_2,\dots,y_n)$ be the associated average costs (which is defined only on $\{(y_1,y_2,\dots,y_n) \mid y_1 \le y_2 \le \dots \le y_N\}$). We have

$$D^{(N)}(y_1,y_2,\dots,y_n) = C_1(y_1) + \dots + C_N(y_1,\dots,y_N)$$

where

$$C_1(y_1) \qquad = \int_0^\infty L_1(y_1-u_{\ell_1})dF_{\ell_1}(u_{\ell_1})$$

$$C_n(y_1,\ldots,y_n) = \int_0^\infty L_n(y_n-u_{\ell_n})dF_{\ell_n}(u_{\ell_n}) +$$

$$+ \int_{y_n-y_{n-1}}^\infty [C_{n-1}(y_1,\ldots,y_{n-2},y_n-u_{\ell_n})-C_{n-1}(y_1,\ldots,y_{n-2},y_{n-1})]dF_{\ell_n}(u_{\ell_n})$$

for $n = 2,\ldots,N$.

□

Finally, we present an algorithm to minimize $D^{(N)}(y_1,\ldots,y_N)$. Define for convenience

$$D^{(n)}(y_1,\ldots,y_n) = C_1(y_1) + \ldots. + C_n(y_1,\ldots,y_n) \qquad n = 1,2,\ldots,N.$$

Then the following procedure yields the global minimum of $D^{(N)}(y_1,\ldots,y_N)$ in the area $\{(y_1,y_2,\ldots,y_n) \mid y_1 \le y_2 \le \ldots \le y_N\}$.

<u>Step 1</u> (Initialization). $n := 1$. Minimize $D^{(1)}(y_1)$. Let $\hat{S}_1$ denote the value that minimizes $D^{(1)}(y_1)$.

<u>Step 2</u>. $n := n+1$. If $n > N$ <u>stop</u>.
Let $(\hat{S}_1,\ldots,\hat{S}_{n-1})$ minimize $D^{(n-1)}(y_1,\ldots,y_{n-1})$. Minimize next $D^{(n)}(\hat{S}_1,\ldots,\hat{S}_{n-1},y_n)$ and let $\hat{S}_n$ denote the corresponding minimizing value. If $\hat{S}_n \ge S_{n-1}$ then goto 2.

<u>Step 3</u>. Let k be the smallest index such that $\hat{S}_k > \hat{S}_n$. If $k = 1$ then goto 4. Minimize $D^{(n)}(\hat{S}_1,\ldots,\hat{S}_{k-1},y,\ldots y)$. Let $S_{k,n}$ denote the corresponding minimizing value. Set $\hat{S}_m := S_{k,n}$ for $m = k,\ldots,n$. If $\hat{S}_k \ge \hat{S}_{k-1}$ then goto 2, else goto 3.

<u>Step 4</u>. Minimize $D^{(n)}(y,\ldots,y)$. Let $S_{1,n}$ denote the corresponding minimizing value. Set $\hat{S}_m := S_{1,n}$ for $m = 1,\ldots,n$. Goto 2.

Again, we find that the N-dimensional problem can be solved completely by successively solving a sequence of one-dimensional problems. This finally leads to

Theorem 4.4. The procedure outlined above yields the global minimum of $D^{(N)}(y_1,\ldots,y_N)$ in a finite number of steps. The associated policy (which in every period increases the echelon inventory position of echelon n to $\hat{S}_n$) is average cost optimal for the infinite horizon problem.

□

## 4. Average cost analysis of assembly systems

Consider a production system, in which several components are assembled into a single end item. Without loss of generality we may assume that only one component of each type is needed to assemble one product (define appropriate units). Components have to be delivered by outside suppliers. Let $\ell_i$ denote the leadtime for an order of component i to arrive at the component store. Final assembly of an end item is obviously possible only if at least one copy of each component type is available; the components are subject to what is known as dependent demand. In particular, it is possible to arrive at a situation in which there is plenty of stock of several components which nevertheless may be useless since one critical component is missing. Indeed, MRP systems are explicitely addressing these dependent demand structure by deriving production and needed component quantities from a given Master Production Schedule (MPS). This MPS separates the stochastic demand for end items from the production system which is deterministic; the only safety stocks allowed are safety stocks of end items. As a result of such an approach, the height of these end-item safety stocks has to reflect the uncertainty in demand during the entire leadtime from procurement of components up to and including final assembly. If this total leadtime is long, we therefore need high safety stock levels.

Below, we describe an alternative procedure, recognizing still the dependent demand structure but nevertheless allowing for component safety stocks. The key observation is that orders for different components have to be coordinated (in a way to be described below); coordinated component safety stocks then permit a considerable reduction of end-item inventory levels. In particular, end-item safety stocks now only have to reflect the uncertainty in demand during the final assembly leadtime. The resulting policy can be shown to be average cost optimal, within the cost framework used throughout this paper.

The results presented in this section can be viewed as an average cost equivalent of the discounted cost analysis of Schmidt and Nahmias[1985] and are in the same spirit as the results of the preceding section. We analyze systems with an arbitrary number of component types (Schmidt and Nahmias only treat the two component case and, as Graves[1988] remarks, it is not clear how their results can be extended to more general structures).

Consider the system pictured in fig. 5. As before, only final products are subject to outside demand. Components in the system (in stock at the component store or as part of work-in-process in the assembly phase) are subject to a holding cost $h_i$ (for component type i), final products are stored at a holding cost $h_0 + \sum_{n=1}^{N} h_n$, while a penalty p is incurred if demand cannot be met immediately and has to be backlogged. All costs are calculated at the end of a period.

At this point, the reader should recall the definition of an echelon stock and of an echelon inventory position, in particular for assembly systems (cf. section 2). Note that the echelon stock of a component includes components already assembled in end items stored in the final product warehouse. As before, echelons are numbered according to the highest installation in that echelon. Let $x_i$ denote the echelon stock of echelon i, for $i = 0,1,\ldots,N$ (where a negative value denotes a backlog again). Define

$$L_1(x_0) = (h_0+\textstyle\sum_{n=1}^{N} h_n)\displaystyle\int_0^{x_0} (x_0-u)dF(u) + p\int_{x_0}^{\infty} (u-x_0)dF(u) - \textstyle\sum_{n=1}^{N} h_n x_0 \quad \text{if } x_0 \geq 0,$$

$$L_1(x_0) = p\int_0^{\infty} (u-x_0)dF(u) - \textstyle\sum_{n=1}^{N} h_n x_0 \quad \text{if } x_0 < 0,$$

$$L_n(x_n) = h_n x_n \quad \text{for all } x_n \quad (n = 1,\ldots,N).$$

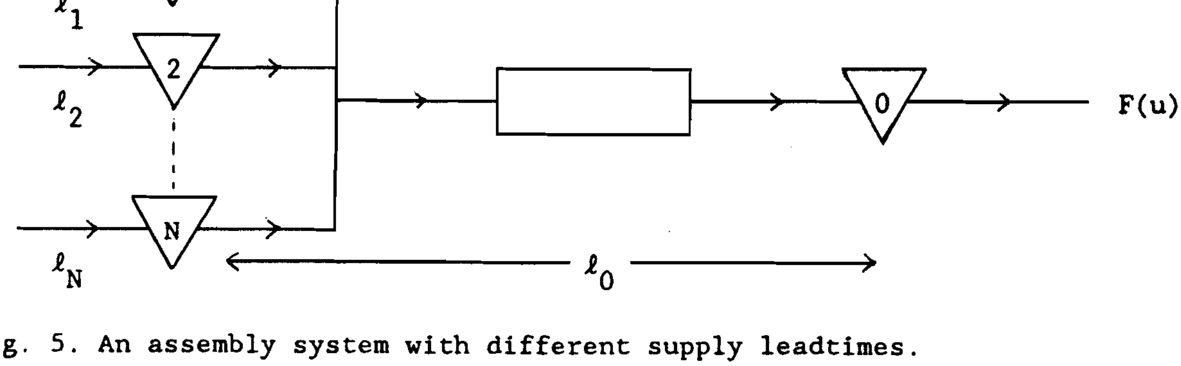


Fig. 5. An assembly system with different supply leadtimes.

As before, note that $L_0(x_0) + \sum_{n=1}^{N} h_n x_0$ represents the real costs incurred at the end of a period in installation 0, if at the beginning of this period the available stock is equal to $x_0$.

Since there is no uncertainty in supply of components (each order for component type i is delivered after exactly $\ell_i$ periods) it is easily verified that components with equal order leadtimes can be handled as one "aggregate" component (i.e. the same order-up-to level is used for these components). In particular, if all components would have equal order leadtimes the system could be analyzed as a two stage serial system. Therefore, it is no restriction to assume in the sequel that $\ell_1 < \ell_2 < \ldots < \ell_N$.

In the preceding section we evaluated a decision made at the beginning of period t and a decision made at the beginning of period $t+\ell_2$ on their joint impact on the costs at the end of period $t+\ell_2+\ell_1$. In the same spirit we will now consider decisions made at the beginning of period t (with respect to echelon N), at the beginning of period $t+\ell_N-\ell_{N-1}$ (with respect to echelon N-1) and so on, until, at the beginning of period $t+\ell_N$, we have to decide on the assembly of the final product. Finally, we evaluate these decisions on their joint impact on the costs at the end of period $t+\ell_N+\ell_0$.

Suppose, at the beginning of period t, we decide to raise the echelon inventory position of component N to a level $y_N$, say. As a result, the echelon stock of component N at the beginning of period $t+\ell_N$ will be equal to $y_N-u_{\ell_N}$ ($u_\ell$ denotes the cumulative $\ell$-period demand again). Clearly, an echelon stock of component N-1 (or any other component) higher than $y_N-u_{\ell_N}$ at time $t+\ell_N$ does not make any sense (and only leads to higher inventory costs) since the assembly of any end-item requires a copy of <u>each</u> component type. Of course, the same holds vice versa (the echelon stock of component type N should not be higher than that of component type N-1) but the important observation is that <u>the echelon stock of component N-1</u> at time $t+\ell_N$ <u>still can be influenced after time t</u> (up to time $t+\ell_N-\ell_{N-1}$). Furthermore, note that between time $t+\ell_N-\ell_{N-1}$ and time $t+\ell_N$ both echelon stocks are subject to <u>the same stochastic demand</u> $u_{\ell_{N-1}}$. It therefore seems reasonable to apply the following ordering rule for component N-1:

<u>Ordering rule for echelon N-1</u>: Choose a <u>potential</u> order-up-to level $y_{N-1}$. At time $t+\ell_N-\ell_{N-1}$, increase the echelon inventory position of component N-1 to $y_{N-1}$ if $y_N-u_{\ell_N-\ell_{N-1}} \geq y_{N-1}$, and to $y_N-u_{\ell_N-\ell_{N-1}}$ if $y_N-u_{\ell_N-\ell_{N-1}} < y_{N-1}$. Here $u_{\ell_N-\ell_{N-1}}$ denotes the outside demand (translated in terms of components of type N) between t and $t+\ell_N-\ell_{N-1}$.

More generally, let $y_N$, $y_{N-1}$, ... , $y_{k+1}$ be chosen potential order-up-to levels for component types N, N-1, ... , k+1. Next, choose a potential order-up-to level $y_k$ for component type k. At time $t+\ell_N-\ell_k$, we increase the echelon inventory position of component k to $\min(y_k, y_{k+1}-u_{\ell_{k+1}-\ell_k}, \dots, y_N-u_{\ell_N-\ell_k})$. Here $u_{\ell_m-\ell_k}$ denotes the outside demand (translated in terms of components of type m) between t and $t+\ell_m-\ell_k$, for $m = k+1,\dots,N$.

The above intuitive logic can be made rigorous. Indeed, it can be shown, by exploiting convexity properties of appropriate cost functions again, that the class of policies indicated above is dominant in the set of all possible policies. More precisely stated: for each policy there exists a corresponding policy satisfying the ordering rules given above, which has lower average costs. A detailed analysis is presented in Langenhoff and Zijm[1989].

At time $t+\ell_N$ finally, we decide to increase the echelon inventory position of final products to a level $y_0$, say. However, if the echelon stock of any component is smaller than $y_0$, i.e. if

$$\min(y_N-u_{\ell_N}, y_{N-1}-u_{\ell_{N-1}}, \dots, y_1-u_{\ell_1}) < y_0 \tag{4.1}$$

we simply assemble as much as possible while furthermore a backlog occurs at some of the component stores.

Let $D^{(N)}(y_0,y_1,\dots,y_N)$ denote the resulting costs of these N+1 decisions (at time points t, $t+\ell_N-\ell_{N-1}$, ... , $t+\ell_N-\ell_1$ and $t+\ell_N$); these costs arise in period $t+\ell_N+\ell_0$. It is not hard to see that we must have $y_0 \le y_1 \le \dots \le y_N$. By applying methods based on convexity properties of appropriate functions, similar to those discussed in the previous section, the following theorem can be proved.

Theorem 4.1. Let $\ell_1 < \ell_2 < \dots < \ell_N$. Then

$$D^{(N)}(y_0,y_1,\dots,y_n) = C_0(y_0) + C_1(y_0,y_1) + \dots + C_N(y_0,y_1,\dots,y_N)$$

$$C_0(y_0) = \int_0^\infty L_0(y_0-u_{\ell_0})dF_{\ell_0}(u_{\ell_0}),$$

$$C_1(y_0,y_1) = \int_0^\infty L_1(y_1-u_{\ell_1})dF_{\ell_1}(u_{\ell_1}) + \int_{y_1-y_0}^\infty [C_0(y_1-u_{\ell_1})-C_0(y_0)]dF_{\ell_1}(u_{\ell_1})$$

$$C_n(y_0,y_1,\ldots,y_n) = \int_0^\infty L_n(y_n-u_{\ell_n})dF_{\ell_n}(u_{\ell_n}) +$$

$$+ \int_{y_n-y_{n-1}}^\infty [C_{n-1}(y_0,y_1,\ldots,y_{n-2},y_n-u_{\ell_n-\ell_{n-1}})-$$

$$-C_{n-1}(y_0,y_1,\ldots,y_{n-2},y_{n-1})]dF_{\ell_n-\ell_{n-1}}(u_{\ell_n-\ell_{n-1}}) \qquad (n=2,\ldots,N).$$

□

The reader may note the striking resemblance between the result of theorem 4.1 and the structure of an (N+1)-echelon serial system (compare theorem 3.3), with leadtimes $\ell_n-\ell_{n-1}$ $(n = 2,\ldots,N)$, $\ell_1$ and $\ell_0$. The results would even correspond completely if the first term of $C_n(y_0,y_1,\ldots,y_n)$ would be

$$\int_0^\infty L_n(y_n-u_{\ell_n-\ell_{n-1}})dF_{\ell_n-\ell_{n-1}}(u_{\ell_n-\ell_{n-1}})$$

Indeed, our decision structure strongly resembles the decision structure in an (N+1)-echelon serial system. At time $t+\ell_N-\ell_{N-1}$ we order up to $y_{N-1}$ only if $y_N-u_{\ell_N-\ell_{N-1}} \geq y_{N-1}$, otherwise we limit our order to $y_N-u_{\ell_N-\ell_{N-1}}$. But <u>no backlog occurs</u> for components of type N-1, we could have ordered more but it would have resulted in temporarily useless stocks only. If $y_N-u_{\ell_N-\ell_{N-1}} > y_1$, there is <u>not</u> a part of the order of components of type N that is delayed, since at time $t+\ell_N-\ell_{N-1}$ there is simply no stocking point for an order of type N, released at time t. The complete order for components of type N <u>keeps moving towards its component store</u>. A similar remark holds for the other components. This explains why the first term of $C_n(y_0,y_1,\ldots,y_n)$ has to correspond indeed with the full leadtime $\ell_n$ for component n, instead of with the leadtime difference $\ell_n-\ell_{n-1}$.

Despite the differences mentioned above, theorem 4.1 and the results of the preceding section suggest that an optimal policy for the control of assembly systems can be found in the same way as for the control of serial systems. Indeed we may formulate the following theorem.

<u>Theorem</u> 4.2. The function $D^{(N)}(y_0,y_1,\ldots,y_n)$, defined on $\{y_0,\ldots,y_N|y_0\leq\ldots\leq y_N\}$ can be minimized by subsequently minimizing a series of one-dimensional convex functions. The procedure to be used is completely similar to the one described

at the end of section 3. Let the global minimum of $D^{(N)}(y_0, y_1, \ldots, y_n)$ be reached in $(\hat{S}_0, \hat{S}_1, \ldots, \hat{S}_N)$. Then an average cost optimal policy for the infinite horizon problem can be formulated as follows:

at each decision moment, increase the economic inventory position

- of echelon 0 to $\hat{S}_0$,
- of echelon N to $\hat{S}_N$,
- of echelon k to $\min(\hat{S}_k, \hat{S}_{k+1} - u_{\ell_{k+1}-\ell_k}, \ldots, \hat{S}_N - u_{\ell_N - \ell_k})$, for $k = 1, \ldots, N-1$. Here $u_{\ell_m - \ell_k}$ denotes the outside demand (translated in terms of components of type m) in the last $\ell_m - \ell_k$ periods, for $k+1 \le m \le N$, $k = 2, \ldots, N$.

□

Concluding, for convergent assembly systems it is again possible to decompose the system into a series of single echelon problems, even though several component types run parallel through the system. Unfortunately, a divergent distribution system yields more difficulties. These systems will be treated in the next section.

## 5. Average cost analysis of distribution systems

Consider a system, consisting of one depot and N local warehouses. The depot acts both as a central warehouse and as a distribution center. Orders are placed at the depot by the local warehouses. The depot may order goods from an outside supplier (a factory, say); this supplier is able to deliver any order within $\ell_2$ periods. The order leadtimes for the local warehouses are all equal to $\ell_1$. Local warehouses are numbered from 1 to N, the depot has index N+1 (cf. fig. 6).

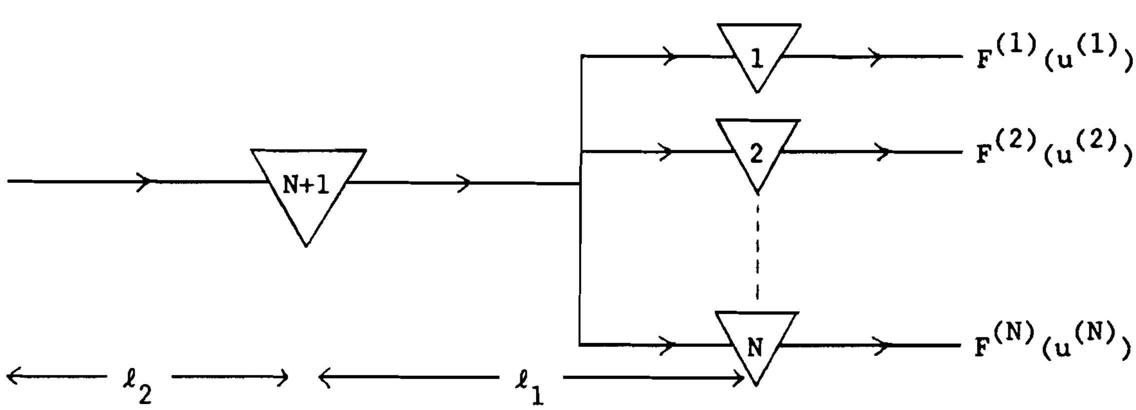


Fig. 6. A distribution system with equal distribution leadtimes.

Outside demand is experienced at each local warehouse, but not at the depot. We assume that demand at different local warehouses can be represented by independent random variables $u^{(n)}$ with distribution function $F^{(n)}(u^{(n)})$. $F(u)$ is the distribution of the cumulative one-period demand u at all local warehouses. No transfer of goods between local warehouses is permitted.

A cost structure similar to the one used in the preceding sections is assumed. At the depot and in transit to the local warehouses, a holding cost of $h_2$ per unit per period is incurred, while at the local warehouses we charge a holding cost $h_1+h_2$ for each unit of stock, per period (the inventory holding costs attributed to added value are the same for all local warehouses). Furthermore, a penalty cost of p per unit per period is incurred at each local warehouse in case of a stockout (again equal for all local warehouses).

The structure of distribution systems gives rise to specific problems not encountered in the models discussed previously in this paper. If there is not sufficient stock at the depot to fulfill requests of all local warehouses, the question arises how to distribute the available products, or in other words, how to allocate the echelon stock of the depot among these local warehouses.

A reasonable allocation rule seems to be based on the minimization of the expected holding and penalty costs at the local warehouses. Hence, suppose that at the beginning of a period the echelon stock of the depot is equal to $x_{N+1}$, while the inventory positions of the local warehouses equal $x_n$. If local warehouse n wishes to increase its inventory position to $y_n$ (n=1,...,N) and if

$$x_{N+1} < \sum_{n=1}^{N} y_n$$

then an allocation decision has to be made. Define, as before,

$$\hat{L}_n(x_n) = (h_1+h_2)\int_0^{x_n}(x_n-u^{(n)})dF^{(n)}(u^{(n)}) + p\int_{x_n}^{\infty}(u^{(n)}-x_n)dF^{(n)}(u^{(n)}) \quad \text{if } x_n \geq 0,$$

$$\hat{L}_n(x_n) = p\int_0^{\infty}(u^{(n)}-x_n)dF^{(n)}(u^{(n)}) \quad \text{if } x_n < 0,$$

and

$$\hat{C}_n(y_n) = \int_0^{\infty}\hat{L}_n(y_n-u_{\ell_1}^{(n)})dF_{\ell_1}^{(n)}(u_{\ell_1}^{(n)})$$

then the allocation which minimizes the expected holding and penalty costs appears as the solution $(z_1,\dots,z_N)$ of the following problem

$$\text{(P)} \qquad \min \sum_{n=1}^{N} \hat{C}_n(z_n)$$

$$\text{under } \sum_{n=1}^{N} z_n = x_{N+1}$$

$$z_n \geq b_n \tag{5.1}$$

where $b_n$ denotes the amount of stock available at or in transit to local warehouse n, just prior to the allocation decision to be made. Problem (P) can be solved by sequentially solving a series of relaxed problems with the standard Lagrange multiplier technique (see Appendix), and by exploiting convexity properties of $\hat{C}_n(z_n)$. The difficulty, however, is in the form of the answer. The solution _may_ depend on the values $b_n$ $(n = 1,2,\dots,N)$, not only on $x_{N+1}$. This generally rules out the possibility to derive a decomposition result similar to those presented earlier. The dependence of the values $b_n$ did not appear in the situations discussed in the previous sections since there an installation always had to supply goods to at most one downstream installation and since trivially $x_{N+1} \geq b_n$ for all n.

In order to be able to decompose the system again (which is the only way to keep things tractable from a numerical point of view) we have to make an additional assumption. We will come up with a Balance Assumption similar to the one made by Eppen and Schrage[1981], but first we provide some intuition. The assumption to be made should rule out the set of inequalities (5.1) as serious constraints in the optimization problem (P), i.e. the optimum of problem (P) should also be the optimum of the relaxed problem (P'), defined by

$$\text{(P')} \qquad \min \sum_{n=1}^{N} \hat{C}_n(z_n)$$

$$\text{under } \sum_{n=1}^{N} z_n = x_{N+1}$$

By using a Lagrange multiplier technique and next eliminating this multiplier, we arrive at

$$\frac{\partial}{\partial z_1}\hat{C}_1(z_1) = \frac{\partial}{\partial z_2}\hat{C}_2(z_2) = \dots = \frac{\partial}{\partial z_N}\hat{C}_N(z_N) \; ,$$

$$\sum_{n=1}^{N} z_n = x_{N+1}$$

It is not hard to show that

$$\frac{\partial}{\partial z_n}\hat{C}_n(z_n) = (h_1+h_2)F^{(n)}_{\ell_1+1}(z_n) - p(1-F^{(n)}_{\ell_1+1}(z_n)),$$

from which we finally obtain

$$\begin{aligned} &F^{(1)}_{\ell_1+1}(z_1) = F^{(2)}_{\ell_1+1}(z_2) = \dots = F^{(N)}_{\ell_1+1}(z_N) \\ &\textstyle\sum_{n=1}^{N} z_n = x_{N+1} \end{aligned} \tag{5.2}$$

In the sequel, we will denote a solution of the set of equations (5.2), corresponding to an optimal solution of problem (P'), with $z_n[x_{N+1}]$, for $n=1,2,\dots,N$, to emphasize the dependence on $x_{N+1}$. The optimal solution of problem (P') corresponds with a distribution of products among the local warehouses such that the stockout probabilities after $\ell_1+1$ additional periods are equal at these local warehouses. This situation was called an "equal fractile position" by Eppen and Schrage[1981]. Vice versa, <u>if</u> it is always possible to reach an equal fractile postition for the inventory positions at the local warehouses, <u>then</u> the minimum values of problem (P) and (P') are identical, and hence in particular an allocation <u>depends solely on</u> $x_{N+1}$. Quoting Clark and Scarf[1960], this means that it is assumed that the values $b_n$ prior to the allocation decision should not be "seriously out of balance".

We now formally state our balance assumption (a slightly adapted version of the one given by Eppen and Schrage[1981] who considered only the stockless depot situation).

<u>Balance Assumption</u>: If, at the beginning of a period, the echelon stock of the depot does not allow a distribution of products among the local warehouses such that all requests can be fulfilled, then it is still possible to allocate the totally available echelon stock of the depot in such a way that an equal fractile position is reached for the local warehouses.

Simulation studies (cf. Van Donselaar and Wijngaard[1987]) as well as approximation techniques (cf. Eppen and Schrage[1981], Federgruen and Zipkin[1984]) indicate that the Balance Assumption is not a serious restriction, i.e. it rarely happens that inventory positions are seriously out of balance. For completeness we also state the following Allocation Rule.

<u>Allocation Rule</u>: If, at the beginning of a period, the echelon stock of the depot does not allow a distribution of products among the local warehouses such that all requests can be fulfilled, then we allocate the <u>total</u> echelon stock of the depot in such a way that an equal fractile position is reached for the local warehouses.

Next, define

$$L_n(x_n) = \hat{L}_n(x_n) - h_2 x_n \qquad \text{for all } x_n \ (n = 1,2,\dots,N)$$

$$L_{N+1}(x_{N+1}) = h_2 x_{N+1} \qquad \text{for all } x_{N+1},$$

and let $D^{(N+1)}(y_1, \dots, y_N, y_{N+1})$ denote the average cost in period $t+\ell_2+\ell_1$ if in period t the echelon inventory position of the depot is raised to $y_{N+1}$ and in period $t+\ell_2$ the local warehouse inventory positions are increased to $y_n$ $(n=1,\dots,N)$. Note that this definition of $D^{(N+1)}(y_1, \dots, y_N, y_{N+1})$ only makes sense in the area $\{(y_1, \dots, y_N, y_{N+1}) \mid \sum_{n=1}^N y_n \le y_{N+1}\}$. If the echelon stock of the depot at time $t+\ell_2$ is not sufficient we apply the Allocation Rule. Then we have

<u>Theorem</u> 5.1.

$$D^{(N+1)}(y_1,\dots,y_N,y_{N+1}) = \sum_{n=1}^N C_n(x_n) + C_{N+1}(y_1,\dots,y_N,y_{N+1})$$

where

$$C_n(y_n) = \int_0^\infty L_n(y_n-u_{\ell_n})\,dF_{\ell_n}(u_{\ell_n}), \qquad \text{for } n = 1,\dots,N,$$

and

$$C_{N+1}(y_1,\dots,y_N,y_{N+1}) = \int_0^\infty L_{N+1}(y_{N+1}-u_{\ell_{N+1}})\,dF_{\ell_{N+1}}(u_{\ell_{N+1}}) +$$

$$\int_{y_{N+1}-y_1}^\infty \sum_{n=1}^N \{C_n(z_n[y_{N+1}-u_{\ell_{N+1}}]) - C_n(y_n)\}\,dF_{\ell_{N+1}}(u_{\ell_{N+1}}) \ .$$

□

Let $C_n(y_n)$ attain its maximum in $S_n$, for $n = 1,\dots,N$. Although the analysis becomes considerably more complex, it still can be shown that the functions $C_{N+1}(S_1,\dots,S_N,y_{N+1})$ and $D_{N+1}(z_1[y],\dots,z_N[y],y)$ are convex again. Using this, we finally may prove

Theorem 5.2. Let $C_n(y_n)$ be minimized in $S_n$ $(n = 1,2,\dots,N)$, let furthermore $C_{N+1}(S_1,\dots,S_N,y_{N+1})$ attain its minimum in $S_{N+1}$ while $D_{N+1}(z_1[y],\dots,z_N[y],y)$ is minimized in $\hat{S}$. Then the global minimum of $D^{(N+1)}(y_1,\dots,y_N,y_{N+1})$ on the area $\{(y_1,\ \dots,\ y_N,\ y_{N+1})\ |\ \sum_{n=1}^{N} y_n \le y_{N+1}\}$ is reached in $(S_1,\dots,S_N,S_{N+1})$ if $\sum_{n=1}^{N} S_n \le S_{N+1}$ and in $(z_1[\hat{S}],\dots,z_N[\hat{S}],\hat{S})$ if $\sum_{n=1}^{N} S_n > S_{N+1}$.

Moreover, the stationary policy associated with this global minimum is average cost optimal for the infinite horizon problem.

□

With theorem 5.2 again a decomposition result has been found which enables us to solve a complex (N+1)-dimensional control problem by successively minimizing a series of convex functions of one variable. However, note that the minimization of $C_{N+1}(S_1,\dots,S_N,y_{N+1})$ and $D_{N+1}(z_1[y],\dots,z_N[y],y)$ is not an easy task since it requires in particular the calculation of $z_n[y_{N+1}-u_{\ell_{N+1}}]$, the solution of (5.2), for all values $u_{\ell_{N+1}} \ge y_{N+1} - \sum_{n=1}^{N} y_n$. Explicit expressions for $z_n[y_{N+1}-u_{\ell_{N+1}}]$ can only be given if further assumptions on the demand distribution are made.

Eppen and Schrage[1981] assume that demand at each local warehouse is normally distributed with parameters $(\mu^{(n)},\sigma^{(n)})$. However, necessary is only the normality of the $(\ell_1+1)$-period demand at each local warehouse (with parameters $\mu^{(n)}_{\ell_1+1}$ and $\sigma^{(n)}_{\ell_1+1}$, say), an assumption which is more often at least approximately valid if $\ell_1$ is not too small. Then (5.2) leads to

$$\frac{z_1-\mu^{(1)}_{\ell_1+1}}{\sigma^{(1)}_{\ell_1+1}} = \frac{z_2-\mu^{(2)}_{\ell_1+1}}{\sigma^{(2)}_{\ell_1+1}} = \dots = \frac{z_N-\mu^{(N)}_{\ell_1+1}}{\sigma^{(N)}_{\ell_1+1}}$$

$$\sum_{n=1}^{N} z_n = x_{N+1}.$$

A simple calculation then shows that

$$z_n[x_{N+1}] = \mu^{(n)}_{\ell_1+1} + \frac{x_{N+1}-\sum_{n=1}^{N}\mu^{(n)}_{\ell_1+1}}{\sum_{n=1}^{N}\sigma^{(n)}_{\ell_1+1}}\,\sigma^{(n)}_{\ell_1+1} \tag{5.3}$$

Federgruen and Zipkin[1984] remark that the assumption of normality can be relaxed further. Only needed is the assumption of a common distribution function $\hat{F}$ such that

$$F^{(n)}\left(\frac{y - \mu^{(n)}_{\ell_1+1}}{\sigma^{(n)}_{\ell_1+1}}\right) = \hat{F}(y) \qquad \text{for } n = 1,2,\ldots,N, \qquad (5.4)$$

which clearly holds for the normal distribution but also for exponential distributions and for Gamma, Weibull and Pareto distributions with the same scale parameter.

For situations in which (5.4) is not fulfilled, De Kok[1988] developes an approximative method to allocate stocks. He addresses the problem of splitting a volume commitment into appropriate mix replenishment quantities, which corresponds with the stockless depot case (recall the discussion in the introduction). His approach is based on the observation (also apparent from (5.3)) that the determination of the $z_n[x_{N+1}]$ should actually attempt to equalize "normalized safety stocks", i.e. to equalize the quantities

$$\frac{z_n[x_{N+1}] - \mu^{(n)}_{\ell_1+1}}{\sigma^{(n)}_{\ell_1+1}} .$$

Finally, we briefly comment on the stockless depot model analyzed by Eppen and Schrage[1981]. They considered a situation with a fixed holding costs h throughout the system (in our notations, this means that $h_1=0$). Furthermore, they assumed a balance condition similar to ours and in addition they worked with the same allocation rule. As mentioned already, they explicitely addressed the stockless depot case only. However, if we take $h_1=0$ in our model (in which the depot <u>may</u> hold stock), we find $S_n = \infty$ for $n \le N$. Hence, the only function to be minimized is the function $D_{N+1}(z_1[y],\ldots,z_N[y],y)$, leading to the solution $(z_1[\hat{S}],\ldots,z_N[\hat{S}],\hat{S})$, which corresponds with the situation in which stock is never held at the depot. In other words: under the inventory holding cost assumptions made by Eppen and Schrage[1981] and assuming the same balance condition, their policy remains optimal in the stockholding depot case as well.

## 6. Conclusions and suggestions for further research

In this paper, we have presented a unified framework to model and analyze an integrated production/distribution system under an average cost criterion. Under the assumptions stated in section 2, it can be proved that a policy based on echelon inventory positions, is optimal. By carefully defining the one period expected holding and penalty cost functions $L_n(x_n)$, it is possible to decompose the systems such that the calculation of the optimal order-up-to levels reduces to the minimization of a series of one-dimensional functions (or, equivalently, to analyze a series of single installation models with a convex cost function). These decomposition result holds for serial systems as well as for assembly systems (with an even surprisingly similar structure) and, under an additional balance assumption, also for distribution systems.

It will be clear that the framework developed so far enables us now to analyze complicated combined assembly/distribution structures as the one shown in fig. 1 without any difficulty. First, we analyze the single depot - multi warehouse distribution structure which yields a convex cost function for the echelon associated with the depot (in most cases this will be the function $D_{N+1}(S_1,\ldots,S_N,y_{N+1})$). This latter echelon next serves as the last installation (with the convex cost function just derived) in a serial or an assembly structure, etc.

It has been noted by Clark and Scarf already that we do not have to restrict ourselves to situations where demand occurs at final installations only. Demand at intermediate levels (for instance, the situation where a central depot directly supplies finished products to some key customers) can be incorporated without more than notational difficulties; in particular, the decomposition results and the optimality proofs remain valid. On the other hand, the framework presented in this paper suffers from a number of limitations which we will point out and discuss in some more detail below.

The first limitation concerns the absence of fixed ordering costs. As indicated already in section 2, we do not consider this as a serious restriction as far as the impact on reordering frequencies for an item are concerned. Most industrial companies indeed decide to order almost each item once per month, or once per two months, in combination with orders for other items. In other words, the general policy is to order every period each group of items (each family), i.e. the reordering frequency is decided upon on a higher level. As a result, fixed ordering costs do not play a role any more as far as these reordering frequencies are concerned, they occur in each period.

However, the assumption of no fixed ordering costs has also been criticized since it precludes lotsizing. In our opinion, such a statement does not properly reflect the level on which logistic chains have to be controlled. Lot sizes generally arise due to capacity limitations (not present in our model so far), changeover times of machines and the like, hence on a lower level. The echelon-based policies discussed in this paper should in practice operate on a high level, for instance the depot order-up-to level may serve as the basis for a factory production plan (e.g. a Master Production Schedule when an MRP system is used to control the factory). In other words: it is specified what quantities have to be produced in a specific period and it is left to the factory controller how to execute the plan, including the determination of lotsizes and, on a shopfloor control level, the determination of production schedules. At these levels, lotsizing plays an important role, not at the level of control of an entire logistic chain.

In passing by, we have indicated that it is possible indeed to combine a Base Stock control system on a higher level with other systems for controlling parts of the chain in more detail, parts which are recognized by the Base Stock system as one black box. For instance, it it possible to derive a Master Production Schedule from the Base Stock system, control the factory by means of MRP, whereas component availability is assured again by the Base Stock system.

We have also touched upon the absence of any capacity restriction. This indeed is a serious limitation but we feel that it can be overcome. For single installation models it has been shown by Federgruen and Zipkin[1986] that, in the precence of finite capacity, an adapted order-up-to policy is still optimal under weak assumptions (which are completely fulfilled in our model). The optimal policy is the one to be expected intuitively: order up to a fixed level S or as much as possible, i.e. if the single period production capacity equals c and our inventory position prior to decision equals x then we order min(S-x,c). However, this optimality property cannot be extended to the case with fixed ordering costs K, i.e. in this case the optimal policy is generally not of the expected (s,S) type (see e.g. Wijngaard[1972]).

We believe that capacity limitations can be built in in our multi-echelon model as well, such that the decomposition results remain valid. In fact, the phenomenon is quite similar to the restricted availability of materials in an upstream installation. In both situations we order up to a predetermined level or as much as possible. The same holds for a multi-product environment in which we first have to agree on a volume production quantity and later decide on mix replenishment quantities. In the case where the requested volume cannot be made due to a lack of capacity, we may apply the allocation rule, formulated in section 5, to determine the mix replenishment quantities.

An important extension is related to the inclusion of nonstationary demand patterns. In particular, we often observe seasonal patterns in market demand, as well as fluctuations in available capacity. In order to cope with these fluctuations it might be appropriate to build up some seasonal stock for periods in which a peak demand is expected. The most common approach is to use Linear Programming (see e.g. Bitran, Haas and Hax[1981]) on a very high aggregation level to determine when to produce for which period (or better, when to reserve capacity already). We believe that the Base Stock Control approach can be extended to these types of nonstationary demand, at least when these seasonal patterns can be forecasted with some accuracy, to cope with this seasonality as well as with the "normal" stochastic fluctuations around a (time-dependent) expected value. A simple idea is the following: let $S_T$ be the order-up-to level if we have to produce for T periods (instead of one as we did up to now). By comparing $S-\sum_{k=1}^{T-1}(c_k)$ for different values of T (where $c_k$ denotes the available capacity in period k) it is easy to determine how much one has to produce in advance for future periods in order to prevent running into trouble because of a lack of capacity in the peak season. Such an approach can be viewed as an adaptation of Land's algorithm (see e.g. Silver and Peterson[1985]) for a stochastic environment and will be part of future research.

References

Bitran, G.R., E.A. Haas and A.C. Hax[1981], Hierarchical Production Planning: A Single Stage System, Operations Research, 29, pp. 717-743,

Bitran, G.R., E.A. Haas and A.C. Hax[1982], Hierarchical Production Planning: A Two Stage System, Operations Research, 30, pp. 232-251,

Clark, A.J.[1958], A dynamic, single-item, multi-echelon inventory model, RM-2297, The Rand Corporation, Santa Monica,

Clark, A.J. and H. Scarf[1960], Optimal policies for a multi-echelon inventory problem, Management Science 6, pp. 475-490,

Clark, A.J. and H. Scarf[1962], Approximate solutions to a simple multi-echelon inventory problem, in K.J. Arrow et. al. (eds.), Studies in applied probability and management science, Stanford University Press, Stanford,

De Kok, A.G. [1988], Hierarchical production planning for consumer goods, CQM-Note nr. 71, Centre for Quantitative Methods, Philips, Eindhoven,

Eppen, G. and L. Schrage[1981], Centralized ordering policies in a multi-warehouse system with lead times and random demand, in L.B. Schwarz (ed.), Multi-level production/inventory control systems: theory and practice (Studies in the Management Sciences, Vol. 16), North-Holland, Amsterdam,

Federgruen, A. and P. Zipkin[1984], Approximations of dynamic, multilocation production and inventory problems, Management Science 30, pp. 69-84,

Federgruen, A. and P. Zipkin[1986], An inventory model with limited production capacity and uncertain demands, I: The average cost criterion, Math. Oper. Res. 11, pp. 193-207,

Forrester, J.W.[1961], Industrial Dynamics, The M.I.T. Press, Cambridge, Massachusetts,

Goldratt, E.M.[1988], Computerized shop floor scheduling, Int. J. Prod. Res. 26, pp. 443-455,

Graves, S.C.[1988], Safety stocks in manufacturing systems, J. Manuf. Oper. Mgt. 1, pp. 67-101,

Hax, A.C. and D. Candea[1984], Production and Inventory Management, Prentice-Hall, Inc., Englewood Cliffs, NJ,

Hax, A.C. and H.C. Meal[1975], Hierarchical integration of production planning and scheduling, in M.A. Geisler (ed.), Logistics (Studies in the Management Sciences, Vol. 1), Elsevier, North-Holland, Iglehart, D.L.[1963], Optimality of (s,S) policies in the infinite horizon dynamic inventory problem, Management Science 9, pp. 259-267,

Langenhoff, .L.J.G. and W.H.M. Zijm[1989], Analysis of stochastic assembly models, in preparation,

Meal, H.C., M.H. Wachter and D.C. Whybark[1987], Material Requirements Planning in Hierarchical Production Planning systems, Int. J. Prod. Res. 25, pp. 947-956,

Orlicky, J.A.[1975], Materials Requirements Planning, McGraw-Hill, New York,

Scarf, H.[1960], The optimality of (s,S) policies in the dynamic inventory problem, in K. Arrow, S. Karlin and P. Suppes (eds.), Mathematical Methods in the Social Sciences, Stanford University Press, Stanford,

Schmidt, C.P. and S. Nahmias[1985], Optimal policy for a two-stage assembly system under random demand, Operations Research 33, pp. 1130-1145,

Schonberger, R.J.[1982], Japanese Manufacturing Techniques, The Free Press (MacMillan), New York,

Silver, E.A. and R. Peterson[1985], Decision Systems for Inventory Management and Production Planning, 2nd. ed., Wiley, New York,

Van Donselaar, K. and J. Wijngaard[1987], Commonality and safety stocks, Engineering Costs and Production Economics 12, pp. 197-204,

Vollmann, Th. E., W.L. Berry and D.C. Whybark[1984], Manufacturing planning and control systems, Dow Jones-Irwin, Homewood, Illinois,

Whybark, D.C. and J.G. Williams[1976], Materials Requirements Planning under uncertainty, Decision Sciences 7, pp. 595-606,

Wight, O.[1981], MRP II - Unlocking America's Productivity Potential, Oliver Wight Limited Publications, Essex Junction,

Wijngaard, J.[1972], An inventory model with constrained order capacity, TH-report 72-WSK-63, Eindhoven University of Technology, Eindhoven,

Zipkin, P.[1984], On the imbalance of inventories in multi-echelon systems, Math. Oper. Res. 9, pp. 402-423.

## Appendix

Let V be a finite index set. Consider the following optimization problem

(P) $\min \sum_{n \in V} \hat{C}_n(z_n)$

under $\sum_{n \in V} z_n = b$

$z_n \geq b_n$ for all $n \in V$.

in which the $\hat{C}_n(z_n)$ are nonlinear, convex differentiable functions. Let V' be as subset of V. We will need the following relaxation of (P):

(Q[V']) $\min \sum_{n \in V'} \hat{C}_n(z_n)$

under $\sum_{n \in V'} z_n = b$

Next, we formulate a solution procedure for problem (P).

Step 1 (Initialization). $k := 0$. Set $V(0) := V$. Solve problem $Q(V[0])$ by means of the Lagrange multiplier technique and denote its solution by $\{z_n(0);\ n \in V(0)\}$.

Step 2. Set $\hat{z}_n := z_n(k)$ for $n \in V(k)$, $\hat{z}_n := b_n$ for $n \in V \backslash V(k)$. If $\hat{z}_n \geq b_n$ for all $n \in V$ then stop. The solution $\{\hat{z}_n\ ;\ n \in V\}$ is optimal for problem (P).

Step 3. $k := k+1$. Set $V(k) := V(k-1) \backslash \{n \mid z_n < b_n\}$. Solve problem $(Q[V(k)])$ by means of the Lagrange multiplier technique and denote its solution by $\{z_n(k);\ n \in V(k)\}$. Goto step 2.

This procedure ends in a finite number of steps since V is finite. It is easily shown, by exploiting the convexity of the functions $\hat{C}_n(z_n)$, that this algorithm solves problem (P) to optimality.

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science
**PROBABILITY THEORY, STATISTICS, OPERATIONS RESEARCH AND SYSTEMS THEORY**
P.O. Box 513
5600 MB Eindhoven - The Netherlands
Secretariate: Dommelbuilding 0.03
Telephone: 040 - 47 3130

---

List of COSOR-memoranda - 1989

| Number | Month | Author | Title |
|---|---|---|---|
| M 89-01 | January | D.A. Overdijk | Conjugate profiles on mating gear teeth |
| M 89-02 | January | A.H.W. Geerts | A priori results in linear-quadratic optimal control theory |
| M 89-03 | February | A.A. Stoorvogel<br>H.L. Trentelman | The quadratic matrix inequality in singular $H_\infty$ control with state feedback |
| M 89-04 | February | E. Willekens<br>N. Veraverbeke | Estimation of convolution tail behaviour |
| M 89-05 | March | H.L. Trentelman | The totally singular linear quadratic problem with indefinite cost |
| M 89-06 | April | B.G. Hansen | Self-decomposable distributions and branching processes |
| M 89-07 | April | B.G. Hansen | Note on Urbanik's class $L_n$ |
| M 89-08 | April | B.G. Hansen | Reversed self-decomposability |
| M 89-09 | April | A.A. Stoorvogel | The singular zero-sum differential game with stability using $H_\infty$ control theory |
| M 89-10 | April | L.J.G. Langenhoff<br>W.H.M. Zijm | An analytical theory of multi-echelon production/distribution systems |