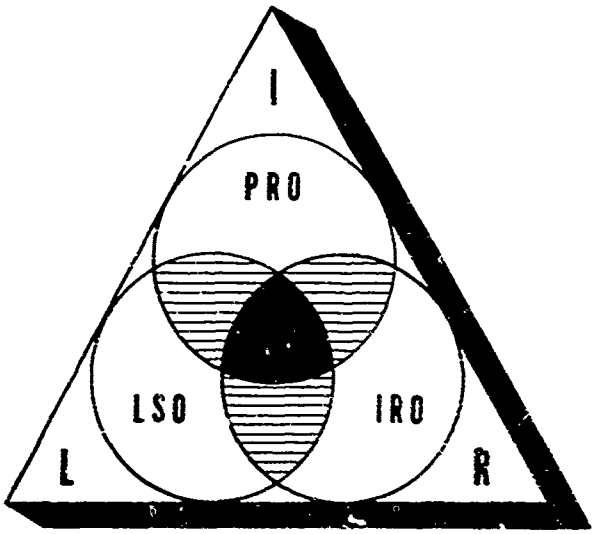
AD

TECHNICAL REPORT

AD 745941

A SUMMARY OF MULTI-ECHELON INVENTORY MODELS AND CONCEPTS

by D D C
RECEIVED
AUG 3 1972
B



W. Karl Kruse

June 1972

INSTITUTE OF LOGISTICS RESEARCH
US ARMY
LOGISTICS MANAGEMENT CENTER
FORT LEE VIRGINIA
23801

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
Springfield, Va. 22151

USAMC Inventory
Research Office
Frankford Arsenal

38

Approved for Public Release, Distribution Unlimited

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* AMC Inventory Research Office Institute of Logistics Research US Army Logistics Management Center	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

A Summary of Multi-Echelon Inventory Models and Concepts

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*

W. Karl Kruse

6. REPORT DATE June 1972	7a. TOTAL NO. OF PAGES 33	7b. NO. OF REFS 15
8a. CONTRACT OR GRANT NO. b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*	

10. DISTRIBUTION STATEMENT

Approved for Public Release: Distribution Unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY US Army Materiel Command

13. ABSTRACT

The AMC Inventory Research Office (IRO) has been involved in the development of multi-echelon inventory models for Army applications for the past several years. During research for a study for the Joint Logistics Review Board (JLRB) by IRO, several significant advancements were made in multi-echelon modelling. Subsequently, several more were noted by IRO in the literature and were put to use. This thesis summarizes the multi-echelon work done at IRO with particular emphasis on the efforts which originated with and followed the JLRB study. Philosophy of analysis is emphasized more than mathematical derivation. The techniques are exact where possible.

ia

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

UNCLASSIFIED

Security Classification

Unclassified
Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
inventory theory, levels, spare parts						

ib

Security Classification

A SUMMARY OF MULTI-ECHELON INVENTORY MODELS AND CONCEPTS

TECHNICAL REPORT

by

W. KARL KRUSE

JUNE 1972

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

AMC INVENTORY RESEARCH OFFICE
INSTITUTE OF LOGISTICS RESEARCH
US ARMY LOGISTICS MANAGEMENT CENTER
FT LEE, VIRGINIA

ic

Information and data contained in this document are based on input available at the time of preparation. Because the results may be subject to change, this document should not be construed to represent the official position of the U.S. Army Materiel Command unless so stated.

TABLE OF CONTENTS

	Page
TABLE OF CONTENTS	i
ABSTRACT	ii
1. INTRODUCTION	1
2. THE AMMIP METRIC MODEL	2
2.1 BASIC METHODOLOGY	2
2.2 OPTIMIZATION USING AMMIP-METRIC	5
3. REAL TIME MULTI-ECHELON MODELS	7
3.1 REAL TIME METRIC	7
3.2 IRO ALLOCATION MODEL	8
4. DELAY DUE TO STOCK-OUT AT A SUPPLIER WITH AN R,Q INVENTORY POLICY	8
4.1 EXPECTED STOCKOUT DELAY	8
4.2 PROBABILITY DISTRIBUTION OF CUSTOMER DELAY	10
4.3 PROBABILITY DISTRIBUTION OF CUSTOMER DELAY FOR ORDER SIZE ≥ 1	12
5. A TWO-ECHELON R,Q MODEL	13
5.1 A SINGLE ECHELON R,Q MODEL	13
5.2 A TWO-ECHELON, R,Q MODEL	16
6. AN EXACT TWO-ECHELON MODEL	22
7. SAFEGUARD PROVISIONING MODEL	24
7.1 THE SAFEGUARD SUPPLY AND MAINTENANCE SYSTEM.	25
7.2 A HEURISTIC SAFEGUARD STOCKAGE MODEL	25
REFERENCES	29
DISTRIBUTION	31

ABSTRACT

The AMC Inventory Research Office (IRO) has been involved in the development of multi-echelon inventory models for Army applications for the past several years. During research for a study for the Joint Logistics Review Board (JLRB) by IRO, several significant advancements were made in multi-echelon modelling. Subsequently, several more were noted by IRO in the literature and were put to use. This thesis summarizes the multi-echelon work done at IRO with particular emphasis on the efforts which originated with and followed the JLRB study. Philosophy of analysis is emphasized more than mathematical derivation. The techniques are exact where possible.

1. Introduction

In the past several years the AMC Inventory Research Office (IRO) has been involved in the development and application of multi-echelon inventory models. During that time, several reports were printed describing some of the multi-echelon work, but until now much of the recent efforts had not been published. This thesis is a summary and reference for all of IRO's relevant multi-echelon work.

Emphasis will be placed on the analytical philosophy behind the development of the models with the intention of stimulating interest in the multi-echelon area. Those models documented elsewhere are given only cursory treatment, while undocumented models are developed thoroughly.

Models of continuous review inventory systems will be the only ones considered, since periodic models have little application within the Army. An appropriate beginning is with the coincident but independent development of two identical models - the AMMIP model at IRO and the METRIC model at RAND. A chronological history from this point is developed in this report which leads to a description of all of IRO's current multi-echelon models.

Sections 2 and 3 cover the initial multi-echelon work which was done in the mid to late sixties. Sections 4 and 5, which are more detailed than the others, cover some of the more recent work which originated from IRO research during a

study for the Joint Logistics Review Board. A model developed at RAND in which an error was noted and corrected at IRO is briefly covered in section 6, since the analysis itself is a significant development. In section 7, a discussion of a heuristic multi-echelon model for the SAFEGUARD ABM is given.

2. The AMMIP-METRIC Model

The term "AMMIP-METRIC model" will be used to signify that the two are essentially identical models and can be discussed as one. The METRIC model has been formalized at RAND into a marketable computer program package, while IRO has kept the associated computational techniques in-house. Nevertheless, apart from computational techniques, the two models have remained identical.

The best references are [10] and [11]. From a theoretical viewpoint [11] is superior, and would be more valuable to one wanting to learn about the model.

2.1 Basic Methodology

A queuing theorem due to Palm is the basis of AMMIP-METRIC. Palm derived the distribution of the number of customers in an $M/G/\infty$ queue. His theorem states that if T is the average service time, and λ the customer arrival rate, then the number of customers in the queue is Poisson distributed with parameter λT, independent of the form of the service time distribution.

A clear and concise proof of Palm's theorem is given in [7] as a sidelight to another theorem.

Often, inventory systems are studied by analogy with queuing systems. The infinite channel system is frequently used. Demands are analogous to queue customers, and lead times are analogous to queue service times.

In the multi-echelon context, a lower echelon stockage location requisitions from the echelon above (queue customer arrives) and receives his stock a lead time later (queue service time). Now the lead time may be thought of as two segments - a normal time to respond and a delay incurred if no stock is available. The lead time is the sum of these two. In general, the delay due to stockout will depend upon the demand pattern and the stockage policy at the above echelon. But, using the queuing analogy, Palm's theorem says that provided demand on the above echelon is Poisson distributed, the number of requisitions from the lower echelon unit which have not been filled (in the queuing system) is Poisson and depends only on the average lead time, including stockout delays. Thus, if T_o is the average of the normal time to satisfy a requisition when there is no stockout delay and W is the stockout delay, the average lead time is $T = T_o + E(W)$.

At this point, yet another queuing analogy is used. In simple terms, for sake of discussion, when the lower echelon

requisition arrives at the above echelon, it is processed and sent to the stock room for shipment. At the stock room, it is serviced and shipped without delay if stock is available, but waits at the stock room as a backorder to be serviced in a FIFO priority if stock is not available. Here the well known $L = \lambda W$ queuing relation applies. Thus, $E(W)$ = expected backorders (L) divided by the demand rate (λ).

Now suppose that all stockage locations follow S-1,S policies and that all exogenous demands are Poisson distributed. The S-1,S policy has two ramifications. First, S-1,S policies merely pass on demands to the above echelons with no modification to their distribution. Thus, S-1,S policies insure Poisson demand at all locations in the multi-echelon system. Secondly, with S-1,S policies, the net stock at a stockage point is S minus the number on order. Knowledge of the number of units on order is, therefore, equivalent to knowledge of net stock, which is useful in forming cost or performance expressions.

Consider the two echelon situations previously discussed. A stockage location in the above echelon sees some average response to its requisitions and by Palm's theorem the number on order by the stockage location is Poisson. On the average it backorders (net stock less than zero)

$$\bar{B} = \sum_{n=1}^{-\infty} n \cdot p(n)$$

where

n = net stock = on hand - backorders

$p(n)$ = probability net stock equals n

= probability on order equals $S-n$

Then $E(W) = \bar{B}/\lambda$ where λ is the demand rate on the above echelon location. This is then added to the normal response time provided to the lower echelon and Palm's theorem is used again to get the distribution of net stock at the lower echelon location.

Note that the use of Palm's theorem is only an approximation. Delay at the upper echelon is conditional on the demands occurring on the lower echelon. The model does not recognize this dependence and assumes that delay is independent of lower echelon demands. Also, it does not recognize that backordered requisitions are not likely to cross over in the real world as the use of Palm's theorem implies, i.e., lead times are assumed independent. Despite these faults, the model has been found to provide good approximations for the low demand items for which it was designed.

2.2 Optimization Using AMMIP-METRIC

There are at least three types of objectives for which an optimization procedure is required. One might wish to minimize a total cost expression, or minimize the investment required to achieve a performance target, or achieve the best performance subject to an investment constraint.

Optimization using the AMMIP-METRIC model is made difficult, however, because, in general, the objective functions are not convex as are those of most single echelon inventory models. Often there are small bumps in the objective function surface which prevent common optimization techniques that rely on convexity from proceeding to proper termination. Moreover, these bumps can arise with small changes in parameters with the result that dramatically different allocations occur even though the parameters differ by only a little. This was observed to occur on a heuristic algorithm developed by IRO. This being a very undesirable property, IRO developed two algorithms, both of which are considered satisfactory.

The first algorithm is described in [3]. It was designed to minimize total cost equal to the sum of inventory and backorder costs in a two echelon system. The algorithm produces exact optimal solutions, but its lack of applicability to other than minimum cost objectives led IRO to develop a flexible heuristic algorithm which does not hang up because of non-convexity, and which can solve either of the three objectives listed above.

The algorithm operates by adding one unit of inventory at a time to the location where the greatest improvement in total backorders occurs. In the terminology of search procedure, it is a steepest ascent method. Termination occurs

when either total inventory, stock availability, or total backorders meet or exceed their targets. Establishment of an availability goal is equivalent to minimizing cost (inventory plus backorder cost) since a necessary condition for minimum cost is that availability equal $1-C_H/C_B$ where

C_H = holding cost per unit per unit time

C_B = backorder cost per unit per unit time

In most cases, this procedure leads to the optimal solution. Where it did not in the cases examined, the solution differed from optimum by no more than one unit at any single stockage location. This degree of error is acceptable.

3. Real Time Multi-Echelon Models

"Real Time" as used here denotes a stockage decision process which relies on real time system information to allocate available assets as opposed to allocation by a predetermined decision rule such as a reorder point, reorder quantity rule which does not depend on any other system conditions. There are two models of interest which are of this type. Both will be treated casually.

3.1 Real Time METRIC

Reference [6] provides a complete description of Real Time METRIC. This model provides a decision rule for shipping depot stock to the bases which is geared to the occurrence

of an event in the system. Stated in its simplest form the rule affirms a shipment to base j if base j's "need" is greater than the depot's "reluctance" to ship. "Need" is defined in terms of base backorders, but "reluctance" is an abstract concept defined by a parametric equation, whose parameters have been set to yield the best results. While philosophically appealing, Real Time METRIC is somewhat lacking in rigor.

3.2 IRO Allocation Model

In contrast to real time METRIC, this model, which is described in [2], operates only when a stock imbalance or scarcity occurs. A short term horizon is defined and a decision for allocation of available assets is produced which minimizes delay to customers. The model is mathematically sound, but is limited to a three area depot, one wholesale depot system.

4. Delay Due to Stock-out at a Supplier With an R,Q Inventory Policy

The importance of delay due to stockout in multi-echelon models can be appreciated from the discussion of the AMMIP-METRIC model in section 2. This section describes two approaches to stockout delay for the more general continuous review R,Q model.

4.1 Expected Stockout Delay

The reference for this section is [13] in which Simon

developed an expression for average delay using these assumptions:

a. The demand process on the supplier is Poisson with parameter λ.

b. The supplier's lead time is either deterministic or exponential. For each lead time case, the delay expressions were found from the basic relationship

$$E(\tau) = \sum_{b=0}^{\infty} E(\tau|b) \cdot Pr[B=b] \qquad 4.1$$

where

τ = customer delay due to stockout

$E(\tau|b)$ = expected customer delay given b backorders at his arrival.

$Pr[B=b]$ = probability backorders equal b at the customers arrival (for Poisson customer arrivals this is the same as the probability that backorders equal b at random point in time).

Using complex reasoning involving order statistics, Simon was able to find $E(\tau|b)$. Since $Pr[B=b]$ has been derived in other works (see [1], Chapter 4) he was able to find $E(\tau)$.

His final expressions unfortunately provided little

insight into the delay process. However, for deterministic lead times, it was shown in [8] that his result simplifies to $E[B]/\lambda$ which, of course, then agrees with the intuitively appealling $L = \lambda W$. However, Simon indicates that this relationship is not correct for the exponential lead times. He cites computational experience in which as much as 53% difference was observed from $E[B]/\lambda$. This is somewhat surprising in view of the general applicability of $L = \lambda W$.

4.2 Probability Distribution of Customer Delay

A requirement for understanding this section is an understanding of chapter 4 of Hadley and Whitin [1], in which the probabilistic properties of an R,Q inventory policy are given. A thorough description of the contents of this section is in [4].

If the demand process on the supplier is Poisson and if his lead time is deterministic, then the probability that a customer arriving at time 0 waits longer than τ is

$$H(\tau) = \frac{1}{Q} \sum_{j=1}^{Q} \Pr[d(\tau-T,0) \geq R+j] \quad 0 \leq \tau \leq T \qquad 4.2$$

where $d(\tau-T,0)$ = demand in the interval $[\tau-T,0)$

T = suppliers deterministic lead time

This follows simply from the fact that the suppliers assets at $\tau-T$, all of which will be available for issue no later than τ, and the demand in the interval $[\tau-T,0)$ determines

if the customer waits longer than τ. If $d(\tau-T,0)$ is greater than assets at $\tau-T$, the customer will wait longer than τ.

While equation (4.1) is strictly valid only for a Poisson demand process, intuition indicates that it should be a good approximation for active items provided $T-\tau$ is large enough so that the effect of residence time of assets at $T-$ is negligible.

The finding of expected values requires intergration over τ from 0 to T. For τ close to T, $T-\tau$ will be small, and thus only a Poisson demand process can be used.

To find $E(\tau)$ we use

$$E(\tau) = \int_0^T H(x)dx \quad .$$

For a Poisson demand process

$$E(\tau) = \frac{1}{\lambda Q} \sum_{j=1}^{Q} \sum_{k=1}^{\infty} P(R+j+k,\lambda T) \tag{4.3}$$

where

$$P(x,\lambda T) = \sum_{m=x}^{\infty} \exp(-\lambda T)\frac{(\lambda T)^m}{m!} \quad .$$

Since $\frac{1}{Q} \sum_{j=1}^{Q} P(R+j+k,\lambda T)$ is the probability that backorders are greater than k, we see that $E(\tau)$ reduces to $E(\tau) = E[B]/\lambda$ which agrees with the results in section 4.1.

In a similar manner

$$E(\tau^2) = 2\int_0^T xH(x)dx$$

$$= \frac{2}{\lambda^2 Q} \sum_{j=1}^{Q} \sum_{k=R+j+1}^{\infty} (k-R-j-1)P(k,\lambda T) \quad , \tag{4.4}$$

Then

$$\mathrm{Var}(\tau) = E(\tau^2) - [E(\tau)]^2 \quad . \tag{4.5}$$

4.3 Probability Distribution of Customer Delay for Order Size ≥ 1

Again a complete reference for this section is [4]. When customer demand is always for one unit, there is no problem in defining customer delay. But if a demand can be ≥ 1 unit, then it is possible that all, part, or none of the demand will be delayed due to stockout. As such, several possible definitions of delay can be made.

In order to overcome this definition problem, the delay distribution is derived for all individual units of the demand. If a demand is for U units, each is identified by an index j. Using arguments similar to those in section 4.2, the probability that the jth unit waits less than τ is

$$G_j(\tau) = 0 \ , \ j > R + Q \ , \ 0 \leq \tau < T$$

$$G_j(\tau) = \sum_{a=\max(j,R+1)}^{R+Q} \Pr[A(\tau-T)=a]\Pr[d(\tau-T,0) \geq a-j] \tag{4.6}$$

$$1 \leq j \leq R+Q, \ 0 \leq \tau < T$$

$$G_j(T) = 1 \quad \text{for all } j \ .$$

Here again, as with equation 4.2, $G_j(\tau)$ is strictly valid only for a compound Poisson demand process.

Derivation of the pdf by individual units provides

flexibility in developing measures of delay. For example, to get the expected value of the average wait of the demand taken over all units use

$$E_u(\tau) = \sum_{u=1}^{\infty} \frac{f(u)}{u} \sum_{j=1}^{u} E(\tau_j) .$$

where

$E(\tau_j)$ = expected delay of the jth unit

$f(u)$ = probability the order size is u.

An alternative measure might be

$$E\ (\tau) = \sum_{u=1}^{\infty} E(\tau_u) f(u) .$$

i.e. the expected delay until satisfaction of the entire demand.

5. A Two-Echelon R,Q Model

Again for this section, familiarity with Hadley and Whitin [1], Chapter 4 is required. As yet, there are no other references for this section.

Based on the results of the previous section, several approximations are used to model a two-echelon inventory system. The policies at both echelons are of the general continuous review R,Q type. The items must be either completely consumable or completely reparable.

5.1 A Single Echelon R,Q Model

If a supply point uses an R,Q inventory policy and is

replenished in a deterministic lead time, L, the probability distribution of its net stock position can be derived. Defining net assets at time t, A(t), as on hand + on order - due out, and noting that all on orders at t-L will have been received into on hand stock by time t we have

$$N(t|L) = A(t-L) - d(t-L,t) \tag{5.1}$$

where

$N(t|L)$ = net stock at time t with lead time L

$d(t-L,t)$ = demand in the interval $[t-L,t]$.

Exact solutions for the steady state pdf of net stock have been found only for compound Poisson demand processes, but successful approximations have been used. For example, approximating the pdf of assets as uniform equal to 1/Q which is true only when all demands are of unit order size, or using a normal distribution for lead time demand.

Except for exponetially distributed lead times, no exact method exists for finding the pdf of N for random lead times. Hadley and Whitin, however, suggest as an approximation

$$Pr[N(t)=n] = \int_{L=0}^{\infty} Pr[N(t|L)=n]\, g(L)dL \tag{5.2}$$

where g(L) is the pdf of the leadtime. This will be a good approximation, provided the chances of more than one order outstanding are negligible. Equivalent to (5.2) is the use

of the marginal lead time demand distribution directly in (5.1).

If the marginal distribution of demand cannot be found in tractable form, it might be necessary to hypothesize a reasonable form for the distribution and set its parameters appropriately. In this light a useful result is found in Parzen [10] where it is shown that

$$\mathrm{Var}[X] = E[\mathrm{Var}(X|Y)] + \mathrm{Var}[E(X|Y)]. \qquad (5.3)$$

Using lead time demand, d_L, in place of X, and L in place of Y in (5.3) yields

$$\mathrm{Var}(d_L) = E[\mathrm{Var}(d_L|L)] + \mathrm{Var}[E(d_L|L)] \qquad (5.4)$$

For compound Poisson demand distributions where $\mathrm{Var}[d_L|L] = \lambda L\ \mathrm{VMR}$ and as an approximation for others, (5.4) can be changed to

$$\mathrm{Var}[d_L] = \mathrm{VMR}\ \lambda\ S\ E(L) + (\lambda S)^2\ \mathrm{Var}(L) \qquad (5.5)$$

where

λ = demand rate

S = average order size

VMR = variance to mean ratio of lead time demand quantity

Along with the expected lead time demand, $\lambda\ S\ E(L)$, $\mathrm{Var}(d_L)$ can be used to set the parameters of the hypothesized distribution (provided of course it is a two parameter distribution).

The military services plan to determine their lead time

demand in this manner using a convenient approximation to the normal distribution which gives closed form expressions for the optimal parameters.

5.2 A Two-Echelon, R,Q Model

The logic behind this model is much the same as the AMMIP-METRIC model, although a few more approximations must be made. Recall that AMMIP-METRIC developed measures for the top echelon independent of the lower echelons, and then used these to determine the effect of top echelon stock on the lower echelon. In this sense, the two-echelon R,Q model is like AMMIP-METRIC.

In this case the mean and variance of customer delay due to stockout at the top echelon are determined as a function of its stockage policy. Then these are related to the marginal lead time demand distribution to determine the pdf of net stock at the lower echelon locations.

As previously mentioned, the use of equation (5.2) on (5.1) is identical to using

$$N(t) = A(t-L) - d_L \tag{5.6}$$

where d_L has the probability function

$$Pr[d_L=X] = \int_{L=0}^{\infty} Pr[d(t-L,t)=X]\ g(L)dL$$

Consider a two echelon supply system with several stockage locations in the bottom echelon and only one in the

top echelon from whom the bottom echelon points order. As with the AMMIP-METRIC model, the bottom echelon lead time is thought of in terms of a normal response plus a stockout delay at the top echelon. In the simplest case, the normal response is deterministic, with all randomness coming from the stockout delay. Clearly, if g(L) can be found in terms of stockout delay, the basis for a two-echelon model is created.

If the lower echelon lead time, L, is equal to

C+W

where

C = deterministic normal response

W = delay due to stockout

then g(L), $L \geq C$ is equal to the stockout delay density function h(•) at the point L-C, i.e. g(L) = h(L-C). However, even in the simplest situation we were unable to obtain a closed form expression for the marginal distribution of lead time demand. This was tried by using a Poisson distribution to represent both demand on the top echelon (it cannot be if the bottom echelon locations order quantities greater than 1), and demand on the bottom echelon.

We decided, therefore, to assume a form for the marginal distribution and set its parameters as discussed in Section 5.1. The negative binomial distribution was selected since

it is gaining acceptance within the military services to represent demand likelihood. Of course, any other distribution can be used in place of the negative binomial.

With $L = C+W$ we have

$$E(L) = C+E(W)$$

$$Var(L) = Var(W)$$

Using equation 5.5

$$Var[d_L] = VMR\ AYD\ (C+E(W)) + (AYD)^2\ Var(W)$$

$$E[d_L] = AYD\ (C+E(W)) \tag{5.7}$$

where AYD is annual yearly demand and C and W are expressed in years.

Demand on the top echelon will depend on the demands on the lower echelon stockage points and their reorder quantities as well. In section 4 it was indicated that, at best, measures of the mean and variance of stockout delay could be obtained exactly only for compound Poisson demand distributions. Moreover, we have been unsuccessful in finding tractable expressions for anything but a pure Poisson demand process. (Equations 4.3 and 4.5) By tractable expression is meant one which can be quickly evaluated by computer. We decided, therefore, on the following intuitive approach.

In order to limit the number of computations, the reorder quantities at the bottom echelon were assumed to be Wilson Q's. This eliminates searching for the optimum Q's

Since the optimum will be larger than the Wilson, and since the tendancy within the services is to keep order quantities small, this is not felt to be a serious limitation.

Establishing the Q's establishes the demand pattern on the top echelon. We will assume that the demand process on the bottom echelon is Poisson. Then demands on the top echelon from a particular lower echelon unit occur with Gamma distributed inter arrival times.

That is if

λ_i = demand rate on location i

Q_i = reorder quantity of location i

ξ_i = time between placement of orders on top echelon by location i

then ξ_i is distributed as

$$f(\xi_i) = \lambda_i \frac{e^{-\lambda_i \xi_i} (\lambda_i \xi_i)^{Q_i - 1}}{\Gamma(Q_i)} . \tag{5.8}$$

Pelczynski [9] has derived relationships for the mean and variance of the number of order placements in a random time interval, $n_i(t)$, for Gamma distributed time between order placements. In terms of the parameters of (5.8) then

$$E[n_i(t)] = \frac{\lambda_i}{Q_i} t .$$

The expression for $Var[n_i(t)]$ is not as simple, but without much difficulty it can be evaluated on a computer. However, the limiting form as $t \to \infty$ is simply

$$\lim_{t \to \infty} \mathrm{Var}[n_L(t)] = \frac{1}{Q_i^2} [\lambda_i t + \frac{(Q_i-1)(Q_i+1)}{6}]$$

which can be used for some two echelon systems. In general, the appropriateness of this form depends on the magnitude of $\exp(-\lambda_i t)$; the smaller, the better. In particular, if the top echelon is the wholesale level, then t would be its procurement lead time. For any but the most inactive items, the approximation will be good. Assuming that demands from the lower echelon units are independent of one another, then the mean of the quantity demanded on the top echelon in a random period t is

$$E[d(t)] = \sum_{i=1}^{N} \frac{Q_i \lambda_i}{Q_i} = \sum_{i=1}^{N} \lambda_i t \tag{5.9}$$

and the corresponding variance is

$$\mathrm{Var}[d(t)] = \sum_{i=1}^{N} Q_i^2 \, \mathrm{Var}[n_i(t)] \tag{5.10}$$

where N is the number of stockage locations in the lower echelon.

Here a critical assumption is used. While neat forms for expectation and variance of stockout delay were obtained for the Poisson only, we assume the form of the expression is valid for any probability distribution. Thus, wherever a Poisson probability function appears in the expression, the corresponding function for another distribution is used. This we assume provides a good approximation.

This is analogous to Hadley and Whitin [1] replacing the Poisson by the normal in the net stock probability equation even though the equations were exact only for the Poisson. Moreover, it is reassuring to note that the Poisson cumulative distribution appears explicity in the expressions, and replacing the Poisson by another distribution does not destroy the interpretation of the expressions.

Thus, expected delay will still be $E[B]/\lambda$ as equation (4.3) was interpreted. While no intuitive interpretation was made for the expression for variance of delay, any interpretation of equation (4.4) will not change with a substituted distribution.

All the ideas having been covered, the computational aspects of the model will be summarized in instruction form.

1. Compute the Q_i for each lower echelon location using the Wilson formula and select reorder points, R_i.

2. Use equations (5.9) and (5.10) to determine mean and variance of quantity demanded on the top echelon during its lead time.

3. Assume demand on the top echelon in the lead time is distributed as a negative binomial random variable and compute its parameters using the results of step 2.

4. Assume a top echelon R and Wilson Q and use equations (4.3) and (4.4) with the negative binomial of step 3 to get expectation and variance of stockout delay.

5. Use equation (5.4) to obtain mean and variance of lower echelon lead time demand for each lower echelon unit.

6. Use expression (5.6) to determine probability functions for net stock at each lower echelon stockage location, and also for the top echelon.

7. Form an appropriate objective function and find the optimum reorder points.

6. An Exact Two-Echelon Model

In [14] a two-echelon model was developed and was claimed to be exact. However, there was an error in the development. Nevertheless, the basic methodological approach was valid. A corrected methodology was developed in [5].

6.1 Methodology

The model is developed from these assumptions:

a. The top echelon, or depot, uses an R,Q continuous review replenishment policy and a repair as received repair policy.

b. The lower echelon locations use S-1,S replenishment policies and repair as received repair policy.

c. The demand process on the lower echelon is Poisson.

d. All repair and replenishment times are deterministic.

e. There are probabilities that a failed item can be repaired locally, or if not locally repaired then at depot.

It is beyond the scope of this report to reproduce the

mathematical expressions, but the ideas behind these are interesting and will be presented.

When a stockage location uses an S-1,S inventory policy, knowledge of the number of units on order plus in repair is equivalent to knowledge of the net stock, since net stock plus on order plus in repair is always equal to S. By assumptions c, d and e the number in repair at time t is Poisson distributed. The number on order at time t consists of those orders which have not had sufficient time to be filled, plus those which have had sufficient time but are unfilled because of stockout delay at the depot. If t_j is the order and ship time for lower echelon location j, then any demands placed on the depot in the interval $(t-t_j,t)$ cannot be satisfied by t and are, therefore, in the on order quantity at time t. Again, these demands are Poisson distributed by assumptions c, d, and e.

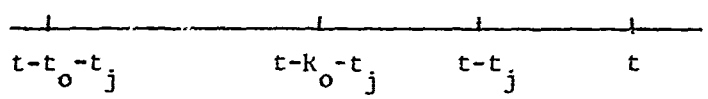


FIGURE 6.1

Figure 6.1 will be helpful for the remainder of the discussion. R_o is the depot repair time, and t_o is the depot lead time, where $R_o \leq t_o$ (a similar development is required for $R_o > t_o$).

Demands by location j on the depot prior to $t-t_j$ but later

than $t-t_j-t_o$ are potential candidates for on order at time t, assuming that the depot's reorder point is ≥ -1. Now depot assets at $t-t_o-t_j$ plus any failed items returned to the depot in the interval $(t-t_o-t_j, t-R_o-t_j)$ can possibly be made available to location j by time t. If, for example, these total to A, a demand from location j in $(t-t_o-t_j, t-t_j)$ will be satisfied by t if it is one of the first A demands after $t-t_o-t_j$. Thus, if total demands on the depot in $(t-t_o-t_j, t-t_j)$ are $d > A$, and those from location j are $d_j \leq d$, the probability x of the d_j are on order at time t is the probability that d_j-x of the first A demands on the depot are from location j.

These are the concepts behind the model development. Once they are understood, it is merely the use of the probability calculus to produce the mathematical expressions for the probability function of net stock at location j. The final expressions are lengthy and involve quadruple summations; however, they are simplified when the item is either all consumable or always able to be repaired.

7. SAFEGUARD Provisioning Model

This model was developed at IRO for the SAFEGUARD Logistics Command's use in provisioning decisions. It is included because it is a good example of how the techniques and ideas described previously can be modified and coupled with other models.

7.1 The SAFEGUARD Supply and Maintenance System

The SAFEGUARD Supply and Maintenance System will be a two echelon structure, the sites at the bottom, and a depot on top. Removal and replacement of failed modules (called ORU's for On-Line Replaceable Units) is the basic maintenance concept. The failed ORU's, most of which are reparable, are either repaired on site with probability, p, or evacuated to depot on a direct exchange basis. Of those returning to depot, there is a probability q the item cannot be repaired at all.

The SAFEGUARD ICP will control the depot operations of maintenance and supply. Depot stocks are replenished by procurement of an amount Q when attrition reduces depot assets to the reorder point R. Failed items returned to depot enter the repair facilities immediately with no batching. Likewise repair at the site is immediate also.

7.2 A Heuristic SAFEGUARD Stockage Model

A most important requirement of the SAFEGUARD ABM is that a target availability be achieved. This objective precludes the direct use of a single item inventory model in which performance is measured in terms of the item alone. At least two system availability stockage models have been developed to aid in provisioning ORU's to achieve the target. But these are single echelon models and can only answer the question of how many ORU's are required on site to

achieve the target at minimum site spares cost. While it is possible to conceive of these models being expanded to handle the depot spares decision as well, it could not be done without considerable effort and perhaps loss of computational feasibility.

The output from one of these system availability models is a list of spare ORU requirements such that the system availability is achieved with the least investment in ORU spares at site. In producing this list, of course, the model had to measure the per dollar impact of a spare on system availability. Spare ORU's affect system availability through the replacement time which is composed of a normal segment that does not depend on spares plus a delay segment that does depend on spares. The output of the availability model can, therefore, be interpreted as a list of the most economical tolerable delays due to ORU stockout on site. Most important is that as long as these average delays due to stockout are achieved, the system will meet its goal.

Now average stockout delay is a supply measure which has been discussed throughout this report. In fact, all the inventory models discussed are capable of approximating average delay due to stockout at the site as a function of site and depot stockage policies. All of this suggests a heuristic multi-echelon optimization procedure which uses a suitable multi-echelon inventory model to achieve at least

cost the most economical tolerable delays produced by the system availability model.

Recalling the description of the AMMIP model in section 2, Palm's theorem was used to approximate the number of items in the pipeline where demand for the item was Poisson distributed. But at the SAFEGUARD site, for a given ORU, the number installed may be small. Consequently, failures cannot be approximated by a Poisson since they are state dependent. However, there is a theorem analogous to Palm's which gives the probability distribution of the number of items in the pipeline when both failures and lead times are state dependent. A derivation appears in [12]. If the state of the system is denoted by m, demands are Poisson with rate λ_m, and lead times have an arbitrary distribution, then the state probabilities depend on only the λ's and the average lead time.

As with the AMMIP-METRIC model, average lead time is computed from a normal lead time plus an average delay. Then the site pipeline distribution is computed by the above theorem. Thus if

T = the average time to return an unserviceable ORU to serviceable condition on site = average pipeline time.

f(x) = probability the number in the pipeline is x

then

$$f(x) = C\,\lambda_1\lambda_2\cdots\lambda_{x-1}/T^x \tag{7.1}$$

where C is a normalizing constant. Note that $T = p\,t_r + (1-p)(t_s+W)$ where t_r = average site repair time

t_s = average replenishment time from depot

W = average delay due to stockout at depot.

To determine depot delay, we are forced to return to assuming depot demands are Poisson. A single echelon model described in [15] is used to determine average depot backorders for a given depot R and Q. Depot delay is then computed as average backorders divided by the depot demand rate. This is an exact expression and has been derived in the same manner as was equation (4.3).

Using equation (7.1) the average number of site backorders can be found for any site spares level, S_i, and any depot policy, R,Q. Optimization over S_i and R with a Wilson Q is accomplished by search.

REFERENCES

1. Hadley, G. and Whitin, T.M., Analysis of Inventory Systems, Prentice-Hall, Inc., Englewood Cliffs, N.J. 1963.

2. Kaplan, A., and Deemer, R., Stock Allocation in a Multi-Echelon System, AMC Inventory Research Office, April 1971.

3. Kruse, W.K., "An Algorithm for Computing Optimum Stock Levels in a Two Level Maintenance System," AMC Inventory Research Office, January 1970.

4. _____, "Distribution of Customer Wait at a Supply Point Whose Inventory Policy is R,Q," AMC Inventory Research Office February 1972.

5. _____, and Kaplan, A., "Comments on Simon's Two Echelon Inventory Model" AMC Inventory Research Office, January 1972.

6. Miller, B.L., A Real Time METRIC for the Distribution of Serviceable Assets, The RAND Corporation, RM-5187-PR, October 1968.

7. Mirasol, N., "The Output of an M/G/∞ Queuing System is Poisson," Operations Research, 11, 282-284 (1963).

8. Parzen, E., Stochastic Processes, Holden-Day, Inc., San Francisco, 1965.

9. Pelczynski, S., The Distribution of Demands Associated with Erlangian Inter-Arrival Times, Australian Defense Scientific Service Aeronautical Research Laboratories, Note 313, March 1969.

10. Rosenman, B., and Hoekstra, D., A Management System for High Value Army Aviation Components, AMC Inventory Research Office, Report No. TR 64-1, October 1964.

11. Sherbrooke, C.C., METRIC: A Multi-Echelon Technique for Recoverable Item Control, The RAND Corporation, RM-5078-PR, November 1966. Also published in Operations Research, 16, 122-141, 1968.

12. _____, MINE: Multi-Indenture NORS Evaluator, The RAND Corporation, RM-5826-PR, December 1968.

13. Simon, R.M., Waiting Time Till Service for a Simple Inventory Model, The RAND Corporation, RM-5927-PR, March 1969.

14. _____, Stationary Properties of a Two Echelon Inventory Model, The RAND Corporation, RM-5928-PR, May 1969.

15. _____, and D'Esopo, D.A., Comments on a Paper by S.G.Allen and D.A. D'Esopo: "An Ordering Policy for Reparable Stock Items", Operations Research, 19, 986-988, 1971.