REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	9 Dec 96	

4. TITLE AND SUBTITLE
An Open Queueing Network Representation of the Reparable Item Problem.

5. FUNDING NUMBERS

6. AUTHOR(S)
Bradley David Silver

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Houston

8. PERFORMING ORGANIZATION REPORT NUMBER
96-034D

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DEPARTMENT OF THE AIR FORCE
AFIT/CI
2950 P STEET, BLDG 125
WRIGHT-PATTERSON AFB OH 45433-7765

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS

15. NUMBER OF PAGES
142

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

DTIC QUALITY INSPECTED 1

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.
DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA - Leave blank.
NTIS - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

AN OPEN QUEUEING NETWORK REPRESENTATION OF THE REPARABLE ITEM PROBLEM

A Dissertation

Presented to

the Faculty of the College of Business Administration

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

Bradley David Silver

August, 1996

AN OPEN QUEUEING NETWORK REPRESENTATION OF THE REPARABLE ITEM PROBLEM

APPROVED:

Everette S. Gardner, Jr.
Professor of Decision and Information Sciences
Chairman of Committee

Edward P. Kao
Professor of Decision and Information Sciences

Basheer M. Khumawala
Marvin Hurley Professor
Professor of Decision and Information Sciences

Elizabeth A. Anderson
Assistant Professor of Decision and Information Sciences

Sara M. Freedman, Dean
College of Business Administration

ACKNOWLEDGEMENTS

First and foremost, I wish to acknowledge God's enabling hand throughout this dissertation effort. This achievement would not have been possible without the strength and wisdom He supplied. I am particularly grateful to my wife, Joan, who not only patiently and faithfully endured the arduous dissertation process and the added burdens it placed upon her, but somehow found the energy and grace to give me an extra dose of support and encouragement at the moments I needed her most. Thanks to Daniel, Matthew, and Hannah who time after time had to make sacrifices while I worked on this effort. I also want to thank my chairman, Dr. Everette S. Gardner, Jr., for steering me toward the reparable item problem and providing encouragement throughout this research. Special thanks to Dr. Edward P. Kao for challenging me academically and introducing me to the power of queueing theory. Finally, I want to express my gratitude to my other committee members Dr. Basheer M. Khumawala and Elizabeth A. Anderson for their support and encouragement.

AN OPEN QUEUEING NETWORK

REPRESENTATION OF

THE REPARABLE ITEM PROBLEM

An Abstract of a Dissertation

Presented to

the Faculty of the College of Business Administration

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

Bradley David Silver

August, 1996

ABSTRACT

This research examined various approaches to the reparable item problem and demonstrated significant shortcomings in those approaches which hamper their effectiveness. The METRIC-based approaches had problems handling the variability of empirical data and state dependent behavior. Queueing approaches ran into significant state space problems when they attempted to solve realistically sized problems or address complex issues.

We develop a new paradigm for the reparable item problem which abandons stock levels as decision variables in favor of depot allocations of repair funding. It also assumes that the depot repair process is not constrained by the availability of unserviceable assets or by workshop capacity.

We use this new paradigm to propose an open queueing network representation of the repair process. This generates an item availability probability distribution function, thus opening up a broad range of different objective functions. We demonstrate a set of specific techniques for creating these representations from empirical data with a U. S. Air Force data set. In a comparison of the fitting and forecasting performance of our model out performed a METRIC based model 38 out of 40 comparisons.

Based upon the queueing network representations generated in this demonstration, we developed a global marginal allocation model to determine the best allocation of the depot's repair funding between competing bases and competing items at each of the bases.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

List of Tables

List of Figures

Figure | Page

CHAPTER 1
INTRODUCTION

THE REPARABLE ITEM PROBLEM

The reparable item class of inventory problems is a very specialized area of inventory policy with very narrowly defined, but very critical real-world applications. As the name implies, its defining characteristic is the fact that items are repaired and returned to the inventory. Most literature on inventory models tends to focus on "consumables." During use, consumables lose all their original utility. This happens in some cases, such as wooden pencils, light bulbs, and bombs, because the item is destroyed in the process of being used. In other instances, such as recapping a worn tire, diagnosing and repairing a faulty circuit card, and replacing the battery on a cheap watch, the item may not actually be destroyed during use, but repair is so costly relative to the replacement cost that repair is not economically viable. Because they are consumed during use, once consumables are used to satisfy a demand, they leave the inventory system and are no longer considered in the problem.

In addition to items which are literally consumable, consumable inventory models also apply to a large number of items which are "functionally consumable" in that they leave the inventory system under consideration once they satisfy an exogenous demand. For example, an automobile manufacturer assembles components, like engines, transmissions, and body parts, which are anything but "consumable." However, because the individual parts no longer have any bearing on the automobile manufacturer's inventory system once they are installed on an automobile, the manufacturer's inventory models can safely treat them as consumables.

In some inventory applications, however, it is economically feasible to repair the item, and the inventory system being modeled encompasses both the demand and repair segments of the process. For such items which are neither consumable nor functionally consumable, the consumable models fail in certain key respects. The key feature of the reparable item problem which consumable models do not address is the return flow of parts from the repair functions. Because of this cyclical flow of parts, the reparable item problem environment is often referred to as a "repair cycle."

There have been some efforts to incorporate this repair cycle aspect into the classic reorder point methods. Allen and D'Esopo (1968) developed a total cost function for reparable items predicated upon Poisson demand and being able to decompose the demand into its reparable and condemned components. The condemned portion of the decomposed demand stream represents the requirements for the item. Simon and D'Esopo (1971) further refined the expected backorder portion of this total cost function. However, even using this type of total cost function, the conventional inventory models do not completely account for the cyclical nature of the repair cycle.

A closely related type of problems is the machine repair problem. As long as the reparable items are at a single location, the machine repair type problem is appropriate. Another defining characteristic of the reparable item problem, however, is that it is multi-echelon. Although the machine repair problem focuses on the repair cycle aspect of the reparable item problem, unless the methodology also considers the multi-echelon facet, it too is an inadequate representation. Muckstadt and Thomas (1980) specifically addressed the question of whether or not a multi-echelon technique was required. Using an actual "industrial inventory system" they used the multi-echelon model to achieve a target level of "expected emergency backorders." They then used the single-echelon models to achieve the same level of support and compared the inventory costs of the two alternatives. Based upon the cost differentials, they concluded that for the specific situation they evaluated, the multi-echelon model was clearly superior. Hausman and Eskip (1994) extended Muckstadt and Thomas' research by improving upon the single-echelon model and looking at a larger range of fill rate levels. Their single-echelon model compared more favorably with the multi-echelon model in the higher fill rate range. Even so, they were only within 3-5 percent of the multi-echelon solutions.

Some additional features which are typical of many reparable item applications and which motivate the development of repair cycle models are long lead-times for the component parts and high cost of down time for the end item. The implication of the long lead-time characteristic is that in the short term, the number of components is fixed, with the only source of supply being the internal repair capability. The high cost of down time is not a characteristic which by itself disqualifies the consumable models. However, when coupled with the fact that the consumable models do not adequately handle the other characteristics of the reparable item problem, it provides significant incentive for developing specialized reparable item models to address the reparable item problem.

These traits of an environment in which a reparable inventory model is necessary typically restrict their applicability to large organizations where inventory is maintained for internal use and repair capability is intrinsic to the organization. The military environment, specifically the Air Force, is the most commonly cited application for the reparable item inventory model, but some other types of organizations which could potentially use reparable item models include utility companies, airlines, and railroads.

The Air Force Repair Cycle

As is the case with much of the earlier research on the reparable item problem, this effort will focus on the U. S. Air Force's repair cycle. The following sections describe the environmental context of the reparable item problem with particular emphasis on the Air Force's repair cycle environment and the system it uses for making inventory-related management decisions regarding these items. Because of the wide variation in actual practice from one aircraft to another, from component to component, and from base to base, this outline attempts to focus on the most likely repair cycle scenarios, noting prominent exceptions where applicable. Its purpose is not to provide a precise system description, but rather to provide a real world context for the repair cycle problem. Where appropriate specific comments about modeling a particular facet of the repair cycle are also included. For the interested reader, Silver, et al (1993) contains a more detailed process description of the Air Force's repair cycle.

Indenture Relationships

Just as the MRP literature deals with bills of material and the way in which component parts are built up into end items, a hierarchy of component parts features prominently in many versions of the reparable item problems. However, unlike the MRP system which assembles and forgets, the reparable item problem continues to cycle these component parts through the repair cycle where they have a direct bearing on the performance measures of the system. In the Air Force context, this hierarchy of parts is referred to as "indenture relationships." At the top of the hierarchy is the end item--the aircraft. Each aircraft, however is made up of multiple levels of component parts which make up the individual systems and sub-systems. At the highest level of indenture, an aircraft is viewed in terms of its component "systems." Examples of the various systems would be communications, navigation, and weapons control. These systems are, however, composed of "line-replaceable units" (LRUs). LRUs are self contained units which the crew chiefs, who are tasked with direct maintenance of the aircraft, can remove and replace. The term "line-replaceable" refers to the flight line where the maintenance action takes place. Some Air Force examples of LRUs include a radio transmitter unit, a cockpit airspeed indicator, and an inertial navigation unit. On newer aircraft, which were designed for rapid maintenance action, many of the LRUs are literally "black boxes" which have built in handles, slide in and out of easily accessible slots, and have a single plug-in connection with the aircraft.

At the next level of indenture are the component parts which make up the LRUs. In the case of the "black box" LRUs, they typically consist of a casing with a series of slots and circuit cards or other electrical components which go into those slots, much like the system unit of a personal computer. Repairing one of these LRUs often consists of a technician removing and replacing one of the component parts. Because these repairs take place in a specialized repair shop, instead of on the flight line, those component parts of the LRU which are themselves reparable are called "shop replaceable units" (SRUs). In some instances, the failure of the LRU is caused by a consumable part. Component parts of the LRU which are not reparable are referred to as "bits and pieces." Because they are managed with an economic order quantity (EOQ) inventory system, these bits and pieces are not given individualized consideration in reparable item problems.

Failure Process

The Air Force's repair cycle environment for any given part is multi-echelon, consisting of three levels, the flight line, the base's repair shop, and the depot. In a typical sequence of events at the base level, the pilot returns from a mission and reports a problem with one of the systems on his/her aircraft. The crew chiefs and other flight line technicians diagnose the failure and isolate a specific LRU they suspect of causing the problem. They then order a replacement LRU from base supply's stock, remove and replace the failed LRU, and send it to the specialized shop for diagnosis.

This process of LRU failure at the bases, and its underlying failure rate, is a key component of all reparable item models, and one of the most troublesome. Some models use a deterministic average number of failures per flying hour. Others model the failure with constant rate stochastic distributions (exponential, compound exponential, and negative binomial are the most common distributions). Still others incorporate state-dependent rates conditioned upon the number of operating aircraft or use time-dependent rates corresponding to the various stages of a likely conflict scenario. Later sections will describe in greater detail the ways in which the various reparable item models address the failure process. In spite of the vast resources dedicated to the problem, this failure process defies definition.

Repair Routing

Depending upon a variety of factors, the failed LRU is routed for repair at either the depot or the base. The rate at which LRUs are repaired at the base level or the "percent base repair" (PBR) is a key factor in multi-echelon reparable item models. Reparable item models simplify the routing process, usually choosing a fixed PBR which may vary from base to base. In practice, the PBR is dependent upon such factors as cause of the failure, availability of diagnostic and testing equipment and technical data, level of expertise in the base's repair shop, availability of SRUs, and repair shop workload for that LRU and other LRUs which require the same equipment.

Base Repair Process

The base process of actually repairing the LRU is, like the earlier stages, a complex and highly variable process. The factors cited in the previous paragraph which influence PBR also impact the time it takes to repair the LRU. The time an LRU spends in the base repair shop can be divided into three major categories: actual repair, awaiting maintenance (AWM), and awaiting parts (AWP). The actual repair time is the most consistent time in the process. Since LRUs are repaired on automated test equipment using pre-programmed test routines, there is very little opportunity for variability other than in the time required to actually remove and replace faulty components.

The times in the AWM and AWP portions, however, depend upon numerous factors which introduce large amounts of variation. Models which incorporate an indenture feature can reflect the AWP time, but the AWM time, which is largely dependent upon local conditions and management decisions, is difficult to model. The composite time that results from the base repair process is highly variable and thus proves difficult to accurately model. The Air Force term for this segment is "base repair cycle time" (BRCT). Repaired LRUs are returned to base supply, where they are stocked to meet future demand by the crew chiefs on the flight line.

Transportation of Unserviceables

Those LRUs which are routed to the depot for repair go through a transportation process which includes base processing time to prepare the LRU for shipment, transportation time, and depot in-processing time upon receipt. For many LRUs, the significance of this stage of the process is mitigated by the fact that the depot already has numerous unserviceable LRUs sitting on the shelf awaiting repair. Except for critical items, induction into depot repair depends primarily upon repair funding, not the availability of the unserviceable LRUs.

Depot Repair Process

Much like the base repair process, the depot's repair cycle times are highly variable, depending upon numerous factors. The automated test equipment and technicians at the depots are very flexible resources, able to repair a large number of different types of LRUs with similar functions. Because of this flexibility, the largest determinant of the rate at which a depot produces a given LRU is the competition from other similar LRUs for those scarce repair resources. The Air Force terminology for this segment of the repair cycle is "depot repair cycle time" (DRCT). LRUs which are repaired in the depot's maintenance function are either stocked at the depot to meet future base demand or used to fill open backorders from the bases.

Serviceable Shipments

When a serviceable LRU is sent to a base, either in response to a requisition or to fill an open backorder, it encounters processing time at the depot and transit time between the depot and the base. The time from when a base places an order for an LRU until the base receives that item out of depot stock, "order and shipping time" (O&ST), is a key component in the Air Force's current requirements system. As was the case in the other repair cycle segments, the O&ST segment is highly variable. Although most of the variation in this segment is introduced by the differences in shipping times between bases, but the different priorities assigned the requisitions also introduce some of the variability.

The SRU Repair Cycle

The repair cycle process for SRUs is virtually the same as for LRUs except for the fact that demand for SRUs is generated from the repair shops, either at the base or at the depot, instead of from the flight line maintenance organizations. In practice, there is less attention focused on the SRUs, so there is less management intervention in the SRU repair process. As such, the SRU repair cycle times are often less variable than the LRU repair cycle times.

COMPUTING REQUIREMENTS

The Air Force expends considerable resources in its ongoing efforts to gather and process data in order to solve the reparable item problem of how many of each item to repair or buy and where to stock these items. At the core of this process is the Recoverable Consumption Item Requirements System (D041). Feeding this system are a variety of systems at the individual bases and depots which engage in data reporting, data consolidation, and data processing. Silver, et al (1991) described these individual systems, their inputs to the requirements computation process, and the impact this data has on the Air Force's reparable item budget using an extensive collection of empirical data bases from the various data systems.

Every Air Force base reports any transactions involving reparable items to the depot systems on a daily basis as they occur. Systems at the depots convert these reports into the base-level rates, balances, and pipeline times. The D041 consolidates the inputs from all the depots to solve its reparable item problem.

The D041 begins solving its reparable item problem by forecasting demand for the individual items. This forecast is simply the historical failure rate multiplied by the projected usage. Usage for most items is measured in flying hours, but a small number of items use other measures such as number of missions or number of shots fired. By factoring in desired and actual stock levels at the bases as well as the historical PBR at the individual bases, the D041 computes the demand that the depot will see from the bases.

In addition to the base-level LRU demand, the depots have some LRU demand from aircraft overhaul functions located at each of the depots. For SRUs, the depots see additional demand as a result of failed SRUs from the depot repair process. The D041 uses the base and depot-level demand rates coupled with historical depot repair times to compute quarter by quarter repair and procurement requirements for each item. For the interest reader, Silver, et al (1992) contains a more detailed explanation of the D041's requirements computations.

DISSERTATION OVERVIEW

In Chapter 2, we review the relevant literature on the reparable item problem as well as the closely related machine repair problem and multi-echelon inventory problem literature. Within the reparable item literature, our review breaks out the METRIC-based literature and the queueing literature. The focus is on the assumptions, benefits, and limitations associated with the various approaches.

Chapter 3 highlights the problems with both the METRIC-based and queueing approaches to the reparable item problem. While the METRIC-based approaches are very sophisticated and have enjoyed widespread application, they are limited in their ability to model the extreme variability inherent in the reparable item processes. The queueing models are readily able to handle the process variability, but are hampered by severe state space problems when they try to address any problem even approaching a realistic size. This chapter also introduces a new paradigm which reduces the reparable item problem to a decision by the depot on the allocation of repair funding and develops the assumptions necessary to support this new paradigm.

In Chapter 4, we develop the basic concept of using an open queueing network to represent the reparable item problem. We describe the theories behind open queueing networks and how they can apply to the availability distributions of the individual items. Using the availability distributions for the individual items, we then show how we can derive end item availability.

In Chapter 5, we demonstrate the process of fitting an empirical availability distribution with an open queueing network representation. We compare the fitting and forecasting ability of the open queuing network representations with those predicted by METRIC-based models and show the clear superiority of the our open queueing network approach.

Chapter 6 demonstrates how the open queueing network representations derived in Chapter 5 can be incorporated into a marginal allocation routine to determine how the depot can best allocate its limited repair funding to maximize a given objective function.

Chapter 7 summarizes the results of this research and proposes opportunities for further research into using open queueing networks to represent reparable item systems.

CHAPTER 2
LITERATURE REVIEW

INTRODUCTION

The reparable item problem is a narrowly defined area of research which simultaneously addresses some of the key issues from both the machine repair problem as well as the multi-echelon inventory problem. The following figure illustrates the domain of the reparable item problem.

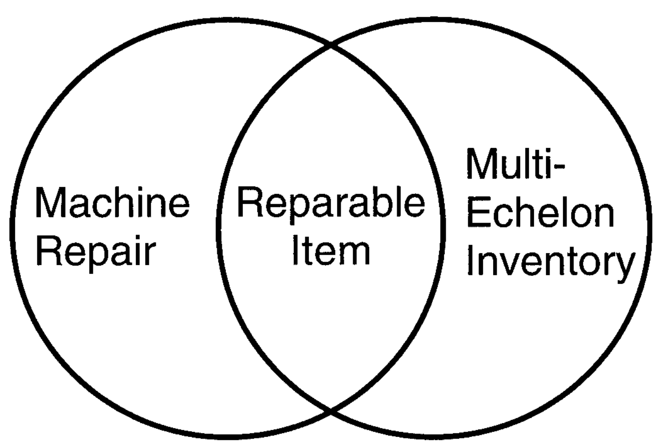


Figure 2-1: Reparable Item Problem Domain

Because the reparable item problem research is so closely related to both the machine repair problem research and the multi-echelon inventory problem research, this literature review will begin by surveying some of the key relevant research in these areas to lay a foundation for the reparable item problem research. This related research is often valuable in that it addresses specific aspects or key characteristics of the reparable item problem and suggests possible approaches to those issues. These efforts, however, fall short of simultaneously addressing the key issues of the reparable item problem. The reparable item problem research is a specific subset of both problem types in that it addresses aspects of both the machine repair and multi-echelon inventory problems in a unified fashion.

This chapter begins by reviewing the some of the key research in the machine repair problem and the multi-echelon inventory problem areas which have applicability to the reparable item problem. We will then look specifically at the reparable item problem research. Our review of the reparable item problem begins with the Multi-Echelon Technique for Recoverable Item Control (METRIC) model and its related extensions. We then briefly cover some simulation approaches to the reparable item problem. Finally, we examine queueing representations of the reparable item problem. The figure below outlines the various approaches to the reparable item problem and their evolution.

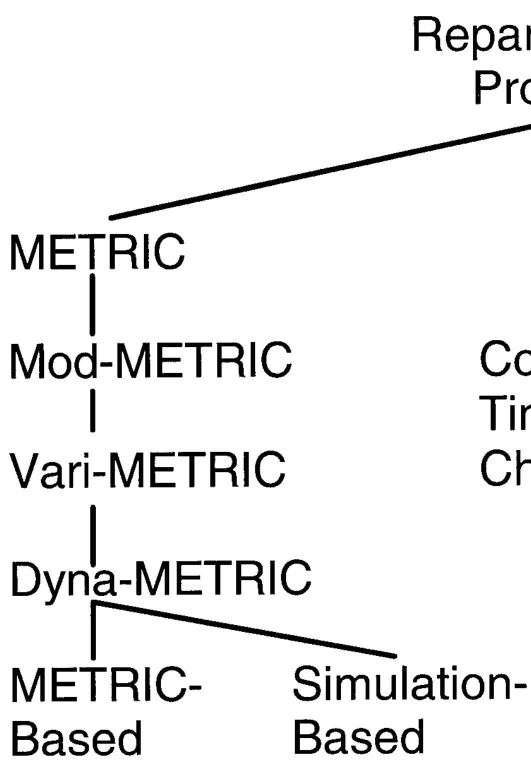


Figure 2-2: Reparable Item Problem Research

MACHINE REPAIR PROBLEMS

The machine repair problem concerns itself with the issues surrounding reparable items. Some typical issues include the optimal number of repairmen or repair channels. From an inventory standpoint, the machine repair problem can address the best stock levels for each of the component parts in a multi-indenture environment. The problem becomes a reparable item problem when these issues are addressed for a multi-echelon system.

Gross, Miller, and Soland (1983) used a closed queueing network and implicit enumeration techniques to determine the optimal allocation of both spares and repair channels in a steady state, single-echelon, reparable environment. Balana, Gross, and Soland (1989) extended this work to a transient environment. The closed queueing network consists of an operating node with working machines and spares and a repair node with a given number of repair channels. By constructing the infinitesimal generator for the network and using standard Markov chain procedures to solve for the transient probabilities, they were then able to use these probabilities to evaluate possible solutions using performance measures such as availability and expected backorders. These solutions formed the basis of their implicit enumeration optimizing routine which sought to minimize the cost of spares and repair channels while meeting minimum performance criteria. Their implicit enumeration technique used upper and lower bounds in conjunction with a binary search routine to significantly reduce the computational requirements of their approach over the explicit technique.

Ebeling (1991) examined the trade-off between stock levels and repair channels in a single-echelon multiple indenture environment. Since the number of repair channels is not an issue when using the METRIC-based assumption of uncapacitated repair, he used a queueing representation as the core of his analysis. Based upon the steady state probability formulations for an M/M/K queue with a fixed number of units derived in Gross and Harris (1974), he was able to compute the ready rate for individual components given a stock level and number of servers. Using this ready rate, he first computed the maximum availability at each feasible budget level for the individual components, then allocated the available budget to stock levels and repair channels for the various components with a dynamic programming formulation.

Dhakar, Schmidt, and Miller (1994) described an application of the machine repair problem for determining repair policy for steel and rubber rolls in a paper mill. These parts are characterized by their high cost, low demand, and their criticality to the operation. Using costs for ordering, holding, downtime, and expedited or emergency repair, they determined the optimal inventory thresholds for expediting existing repair orders or for submitting an emergency repair order. They used a detailed simulation of the paper mill's operations in conjunction with a direct search routine.

MULTI-ECHELON INVENTORY PROBLEMS

The multi-echelon inventory research revolves around the hierarchical relationship between demand generating "branches" and the "trunk" which supplies them. Clark and Scarf (1960) produced one of the earliest multi-echelon inventory efforts. Their research developed multi-echelon solutions from the existing single-level solution procedures and demonstrated the effectiveness of their techniques for a variety of supplier-installation hierarchical structures. Clark (1972) performed a comprehensive review of the multi-echelon literature through 1971. The framework he used for his review categorized the research by six "dichotomies" based upon the problem characteristic: deterministic vs. stochastic, single-product vs. multi-product, stationary vs. nonstationary, continuous review vs. periodic review, consumable product vs reparable product, and backlog vs. no backlog. Although the reparable item problem is specifically addressed in his framework definition, he covers it only in a general overview and in one reference to METRIC. One area in which METRIC-based models have been criticized is their inability to adequately model the variability of the demand process. Because modeling this variability is the focus of our research, this review pays particular attention to the multi-echelon inventory problems which specifically address the variability issue.

Svornos (1986) uses a decomposition approach to the multi-echelon inventory problem in which he addresses the specific distribution of stochastic lead-times. He uses a decomposition approach to derive the steady state distributions using single-echelon techniques for each of the individual locations which make up the multi-echelon structure. Performance at the lower echelon locations is, however, dependent upon the performance at the upper echelon locations. His specific focus was the distribution of the lead-time delay between echelons.

Svornos and Zipkin (1988) also take a decomposition approach to the multi-echelon inventory problem focusing on lead-time delay. They, however, further refine a technique for computing the lead-time which models both the mean and the variance of the demand during lead-time at both the depot and the bases.

Zipkin (1988) describes a technique for modeling the demand and lead-time processes in inventory systems as phase-type distributions. The most important conclusion from this research is that the distribution of lead-time demand is also phase-type, and that it has the same number of phases as the lead-time distribution. This phase-type parallel of Palm's theorem has the same potential for computational efficient solutions as Palm's theorem, but with the significantly enhanced flexibility of the phase-type distribution. Svornos and Zipkin (1991) apply the phase-type models to a multi-echelon inventory system and demonstrate the benefits of the more flexible phase-type distributions in capturing the variability of the transportation process.

METRIC-BASED MODELS

Introduction

The literature on reparable item inventory problems has been dominated by the Multi-Echelon Technique for Recoverable Item Control (METRIC) class of models. This is due, in large part, to the fact that the Air Force has actually applied the METRIC models to manage its reparable inventory. A number of authors associated with either the Air Force or contractors such as the RAND corporation and the Logistics Management Institute have been able to chronicle the various improvements on the original model. Demmy and Presutti (1981) is a thorough survey of the METRIC-based models within their Air Force repair cycle context with a particular emphasis on the theoretical development of the METRIC-based methodology. Their review is also noteworthy for its explanation of the ways in which METRIC-based models are actually applied in the Air Force's planning, budgeting, initial provisioning, replenishment, and distribution processes. The following sections describe the early roots of the METRIC models, review the enhancements to METRIC over the years, and introduce some of the queueing representations of the repair cycle.

METRIC's (S-1,S) Roots

The most common approach to the reparable item problem has been the METRIC- based models. These models have their roots in the (S-1,S) class of stockage policy models. Because of their typically high value and low demand, most inventory systems try to avoid stocking large quantities of reparable items. Sherbrooke (1992) observes that this result follows from the Wilson EOQ formula. A small value for demand in the numerator and a large value for cost in the denominator drives the optimal ordering quantity toward one. When Q=1, the resulting inventory policy is (S-1,S). This policy is a key assumption in reparable item problems.

One of the (S-1,S) roots of the METRIC models was Scarf's extension in Arrow, Karlin and Scarf (1958) of Palm's theorem to the (S-1,S) inventory system. Scarf equated the expected number of busy servers from Palm's theorem with the number of demands during lead-time in the (S-1,S) stockage policy setting. To get the expected number of backorders, he used Formula 2-1.

$$\sum_{s}^{\infty}(u-s)P(u|\lambda,t) \qquad (2\text{-}1)$$

Where u is the index of the summation (starting with s, the number of units stocked)

The probability term represents the Poisson probability of observing u demands given the arrival rate (λ) and lead-time (t) associated with the stockage situation.

Feeney and Sherbrooke (1966) extended Scarf's work by showing that the Poisson arrival assumption required by Palm's theorem could be relaxed using a compound Poisson arrival process instead. This gave a more realistic arrival stream than did the simple Poisson process. The only thing this changed in Scarf's formula for expected backorders was the probability term which changed from the simple Poisson probability to the compound Poisson probability.

These early roots of the METRIC models in Palm's theorem highlight two of their important assumptions/limitations. (1) Since Palm's theorem requires an M/G/∞ queue, the Marcovian arrival rate is a constant, not dependent upon the number of parts currently in repair. Sherbrooke (1992) suggests that in the context of the Air Force application, this assumption is not a significant limitation. When aircraft break, the flying load is usually distributed to the remaining serviceable aircraft. (2) Palm's theorem also assumes an infinite number of servers, and thus no waiting for repair. Gross (1982) performed queueing analysis on both the M/M/∞ queue assumed by METRIC-based models and the M/M/c queue which models constrained repair. For the M/M/c queues he had to use numerical solution techniques. Based upon his comparison of system performance measures such as fill rate and expected backorder level, he concluded, "The general direction of the errors may not be surprising, but the actual magnitude might be. Certainly, one should give careful thought prior to employing the ample service assumption." Proponents of the METRIC models, such as Sherbrooke (1992), however, minimize this limitation by suggesting that the priority system used by the depots is able, in practice, to expedite critical items, and that METRIC models using the standard repair rate actually produce conservative aircraft availability results.

METRIC

Sherbrooke (1968) reported RAND's efforts to incorporate this computation of expected backorders for a single-level (S-1,S) inventory system into a multi-echelon model. The key to this multi-echelon approach was using the stockage policy decisions at the next higher echelon (the depot in the Air Force application METRIC was designed to model) to compute the lead-time that METRIC uses at the lower level (the individual bases in the Air Force system).

Formula 2-2 below illustrates how the depot policy is used to determine base-level backorders:

$$\text{Lead-time} = (\text{PBR} \times \text{BRCT}) + ((1\text{-PBR}) \times (\text{OST} + \text{Depot Delay})) \qquad (2\text{-}2)$$

PBR=Percent Base Repair

BRCT=Base Repair Cycle Time

OST=Order and Shipping Time (from the depot)

Depot delay was computed using Little's (1961) Law shown below in Formula 2-3 below.

$$L=\lambda W \qquad (2\text{-}3)$$

Where: L=length of the queue
W=wait time in the queue).

The L for use in this formula was computed using Scarf's expected number of backorders formula. The λ was computed by superimposing the base demand rates on the depots using Formula 2-4:

$$\sum_{\forall bases} (1 - PBR) x\lambda \qquad (2\text{-}4)$$

This multi-echelon formulation enabled METRIC to model the tradeoff between the value of having stock at the base level vs the benefit (via reduced lead-time) of holding stock at the depot level. METRIC embedded this computation of expected backorders for individual items, into an optimizing routine in which the stock levels for the individual items at the depot and the bases were the decision variables and the total expected number of backorders was the objective function.

Mod-METRIC

One concern the Air Force had with METRIC was the fact that its objective function of minimizing total backorders gave equal weight to all reparable items. In actuality, the availability of the larger components (LRUs) was more important than the smaller component parts (SRUs) that go into them. One complicating factor was the fact that the availability of the LRUs is dependent upon the availability of the SRUs. RAND extended the basic METRIC model by modifying the base repair cycle time (BRCT) term in the lead-time equation to include a separate lead-time term for component parts.

When the SRUs required to repair the LRUs are available, BRCT is simply the time it takes the shop to repair the LRU (which is no change from the original METRIC model). The component SRUs have their own lead-times from the depot. When an SRU's demand during this lead-time exceeds its base stock, the time it takes the shop to return a serviceable LRU to the inventory must also include component lead-time. MOD-METRIC was further able to model the fact that the depot's SRU stock levels determine the lead-time for

SRUs at the base level. MOD-METRIC combined the base's SRU stock levels, the demand rate for the SRUs, and the SRU lead-times (which are in turn a function of depot stock levels for the SRU) to determine the likelihood of delays in the LRU repair process and, by extension, the BRCT.

By explicitly addressing the indenture relationships between end items and their components, MOD-METRIC's objective function more accurately reflected the Air Force's goal of minimizing end item backorders as opposed to minimizing all backorders. However, when the indenture relationships were added, the effect of changes in the individual decision variables became inseparable. This occurred because the value of a given SRU, as it relates to reducing end item backorders, is related to the stock of other SRUs. Silver (1972) addressed this by marginal allocation starting with no LRUs and adding them incrementally, using the remainder of the funding for SRUs. Muckstadt (1973) proposed a similar marginal allocation solution beginning with all the funding dedicated to LRUs and incrementally adding SRUs. Muckstadt (1978) later proposed using Fox and Landi's (1970) Lagrangian multiplier approach to achieve separability.

Vari-METRIC

METRIC and MOD METRIC both used Feeney and Sherbrooke's compound Poisson distribution for demand. However, the variance to mean ratio (VTMR) presupposed by the compound Poisson distribution did not sufficiently replicate the variability found in the empirical data. In an empirical study of over 23,000 Air Force reparable items, Stevens and Hill (1973) showed that the VTMR increases as the mean increases, but at a decreasing rate. They derived Formula 2-5 below in their research.

$$\text{VTMR} = 1.13 \times (\text{mean})^{.34} \qquad (2\text{-}5)$$

Stevens and Hill's VTMR formula is still used by the Air Force's requirements computation system. Sherbrooke (1984) extended Stevens and Hill's research on VTMR by showing that computing a VTMR based upon the mean is also inadequate. He found that demand was not independent from quarter to quarter and accordingly proposed a VTMR computation which took into account the number of quarters for which the VTMR is being computed.

Graves (1985) suggested that this understates the variability of demand, thus underestimating the number of backorders. He proposed using a negative binomial distribution to achieve more realistic variability. Sherbrooke (1986) supported this proposal and pointed out that the assumption of a compound Poisson had resulted in the allocation of too many LRUs.

Dyna-METRIC

The distinguishing features of the Dyna-METRIC models is their ability to handle non-stationary demand and repair processes and to compute transitory probabilities as opposed to steady state probabilities. In the Air Force setting, one of the most important applications of Dyna-METRIC is for war time scenarios in which the day-to-day demands as well as the system's repair capabilities are constantly changing. Because of the short time frames involved in wartime scenarios, the system never has a chance to reach equilibrium, and transient behavior is the key to evaluating a given logistics support system's performance.

Dyna-METRIC's time-dependent demand rates are an important part of modeling the war time scenario since varying sortie rates and mission types during the progression of the conflict produce highly variable failure rates for the various weapon systems involved. For example, the long-range attacks against the enemy's command and control structure initiated during the opening stage of a conflict would produce one set of failure rates for the weapon systems involved in that mission. The transition to close air support of ground forces during their advance, however, would produce a dramatically different set of failure rates for the weapon systems used in that phase of the conflict. The time-dependent repair rates also add valuable realism to the model. This feature can accurately reflect the fact that the military's industrial complex and its associated private sector contractors take some time to gear up to the expanded requirements.

Dyna-METRIC's ability to model the transient behavior of the logistics support system allows decision makers to see the evolving status of the logistics support system, particularly in the first days of any conflict scenario. Since the system would not reach steady state for a long time, the steady state behavior of the system is meaningless because the critical stage would be long past by the time the system ever reached steady state.

Hillestad and Carrillo (1980) and Hillestad (1982) outline the mathematical underpinnings of Dyna-METRIC. The basic building blocks of Dyna-METRIC are time-dependent distributions for demand rate ($d(\tau)$) and repair time ($F(\tau,t)$) instead of the stationary distributions originally used by METRIC. Dyna-METRIC's modeling hinges on a dynamic version of Palm's theorem which computes the number of units in repair at any given time ($m(t)$) using Formula 2-6.

$$m(t) = \int_0^t d(\tau)F(\tau,t)d\tau \qquad \text{(2-6)}$$

Carrillo (1989) details this adaptation of Palm's theorem to the dynamic environment. Dyna-METRIC applies variations of this basic formula to compute demands on base and depot repair capabilities. As in the earlier METRIC models, base stock levels determine backorders while depot stock levels as well as the base and depot repair rates influence the duration of those backorders. Pyles (1984) and Isaacson, Boren, Tsai, and Pyles (1988) describe Dyna-METRIC's implementation for the Air Force. In particular, they detail the user interfaces with the model. For the interested reader, Carrillo (1991) is a comprehensive review of the efforts to extend Palm's theorem to handle dynamic demand and repair rates.

Lateral Transshipment

The basic METRIC-based reparable item models use a strict arborescent resupply structure where a base can only be resupplied from levels above it in the hierarchical structure. In actual Air Force operations, however, bases will often use other bases at the same level in the hierarchical structure as a source of supply, especially when the backorder is causing a not fully mission capable condition. While this reduces backorder time, it results in additional transportation and ordering costs and reduces capability at the supplying base.

Cohen, Kleindorfer, and Lee (1986) developed a lateral resupply approach for use in the multi-echelon inventory problem. This approach made use of "pooling groups" or specifically defined groups of users which share assets laterally and are characterized by "sharing rates" which describe the degree to which the individual users share with each other. Using expected costs and expected response times as their objective, they constructed a branch and bound routine which aggregated demand into a single pseudo-location to get a lower bound, then used a linear transportation problem to solve for an actual upper bound. Cohen, Kamesam, Kleindorfer, Lee, and Tekerian (1990) report how they were able to apply these lateral resupply techniques in IBM's service logistics system.

Lee (1987) proposed a model specifically aimed at the reparable item environment based upon lateral transshipments between the bases in the "pooling groups" from the earlier research. He used a two-step optimization process which found the optimal levels for base stock for each level of depot stock, then used a dynamic optimization to determine the best level for depot stock.

Axsater (1990) observed that Lee's model did not account for bases with dissimilar demand patterns and that it also did not perform very well when the percentage of demands satisfied through lateral support was very high. His extension of Lee's research focused on modeling a state-dependent demand rate which is a function of on-hand balance in order to more accurately model the demand process. His technique revolves around iteratively computing the probabilities a demand is met from: base stock, lateral resupply, and backordered at the depot. Using upper and lower bounds on these percentages, Axsater shows that his iterative routine converges very rapidly, usually within five iterations.

Dada (1992) approaches the lateral resupply issue from a Markov chain perspective. He acknowledges that any exact Markov chain representation is "computationally intractable" for realistically sized problems because of the immense state space. However, by aggregating the individual locations into a single pseudo-warehouse and assuming independent processes for demand and delay, a finite state Markov chain model is computationally feasible. Using the steady state characteristics of the resulting Markov chain and the constraints of the underlying system, he disaggregates from the approximation using a minimization problem to compute the lower bound on performance measures such as fill rate and a maximization problem to get the upper bound. He demonstrated his technique by comparing it's results with the exact solutions of a variety of small scale problems for which the exact solutions were computationally feasible.

Simulation-Based Extensions

In spite of the progress METRIC-based models have made in achieving more and more realistic failure representations, RAND has begun development of Dyna-METRIC 5, a multi-echelon inventory model for reparable items in which, Monte Carlo sampling has replaced the analytic computation of probabilities based on extensions of Palm's theorem. Isaacson and Boren (1988) indicate that one of the primary reason for abandoning the earlier versions is that they, "did not accurately represent the uncertainty in demand and repair, especially the queuing caused by repair constraints." Buyukkurt and Parlar (1993) also used the simulation approach to model the repair cycle in their research to evaluate alternative distribution policies. They cited the fact that METRIC's infinite source and unlimited repair capability assumptions do not produce state-dependent failure and repair rates as their rationale for resorting to simulation.

QUEUEING APPROACHES

METRIC-based models have been widely criticized for their simplifying assumptions of an infinite source and infinite repair. The assumption of an infinite source ignores the possibility of the failure rate decreasing as the number of units which are already unserviceable increases. The infinite repair assumption ignores any possible queueing for limited repair resources. Using queueing terminology, the model which results from these two assumptions is an $M/G/\infty$ queue. By their very nature, queuing models are not hampered by the assumptions of a failure rate which is independent of the state of the system and infinite servers required by the METRIC-based models. As a result, queueing models have greater flexibility in modeling the failure and repair processes. They also offer greater latitude in being able to model other decision variables like the number of servers at various stages in the repair cycle. For the reader who is not familiar with queueing theory and techniques, Appendix A explains some of the key conceptual foundations for the queueing approaches we will be reviewing.

Continuous Time Markov Chain Representation

The most basic queueing approach to the reparable item problem is an exact representation of the system using a continuous time Markov chain (CTMC). The key to a CTMC approach is developing an appropriate state definition paradigm then completely and accurately defining the transitions between those states. Once this is accomplished, the infinitesimal generator of the CTMC, can be used to compute both the non-stationary and steady state probabilities of each of the states. Because it explicitly models the transition rates between each pair of feasible states and computes individual probability distributions for each of the states, the CTMC approach to queueing is very precise and very flexible.

In their research, Gross, Kioussis, and Miller (1987) apply the CTMC approach to the reparable item problem. They begin by showing how the CTMC approach can be used to explicitly model the transitions between all the feasible states in an entire reparable item system. Although this exact method can very faithfully model the real world, they point out that it quickly runs into size problems with the required state space necessary to represent all the feasible states.

These state space size problems are the motivation for their "network decomposition approach." By assuming independence between bases, they model the multi-echelon reparable item inventory system problem by "decomposing" the larger network into separate "local" models for each base and for the depot. The appeal of using these local models is that each one has a considerably reduced state space requirement compared with the CTMC representation of the entire system.

Their decomposition approach begins by solving the local depot model for its steady state probability distribution. In order to link the depot-level and base-level local models, their approach uses a set of performance indicators which define the support the base-level receives from the depot as a function of the state of the base-level system. They compute these state-dependent performance indicators using the steady state solution to the depot-level problem. They then use the depot's performance indicators for each base to build that base's CTMC infinitesimal generator, from which they can compute the base-level steady state probability distribution. Using the base's steady state probability distribution, they compute a different set of state-dependent performance indicators which defines the base's demands upon the depot. Using these performance indicators from all of the bases, they are able to build a new CTMC infinitesimal generator for the depot. This iterative process continues until convergence is achieved.

The significant reduction in the state space they achieved with the smaller local models enabled them to more efficiently solve the larger problem. Their model produced separate steady state probability vectors for each of the bases and the depot. From the steady state probability vectors for the individual bases, they were able to derive key local performance indicators. By assuming independence between the bases, they were also able to derive aggregated performance indicators for the larger system.

Network of Queues Representation

Although the continuous time Markov chain representation, or a decomposition of the complete chain, offers the most precise representation of the reparable item problem, the network of queues approach, because of the product form solutions associated with it, has a distinct advantage in computational efficiency while, in many cases, being able to retain a large degree of flexibility in representation.

One of the drawbacks of the network of queues approach is that it cannot model state-dependent routing. In a multi-echelon reparable item system with multiple bases, the routing of repaired units from the depot depends upon the number of backorders at the individual bases. The network of queues approach cannot accommodate this state-dependent routing because the product form solutions described by Jackson (1957) for open networks and Gordon and Newell (1967) for closed networks require fixed routing probabilities between nodes.

Another reality of reparable item problems which is not readily accommodated by the network of queues approach is the fact that each base has its own requisitioning objective which limits the number of units in the base subsystem. This "blocking" in the individual queues at the base level is not only dependent upon the base's requisitioning objective, but also the number of units in the other nodes at that base. While approximation methods which allow the network of queues approach to deal with blocking at a single node have been proposed, blocking of a subsystem, such as the network of queues representation of a base, which consists of multiple nodes is significantly more complex. The ramifications of these drawbacks need to be evaluated in the context of the intended application to determine if the computational efficiency of the network of queues approach is acceptable for the specific application.

Mirasol (1964) was one of the earliest attempts to represent the reparable inventory system using the network of queues methodology. Using a closed queueing network representation, he researched the system performance tradeoffs associated with allocating resources to the various stages in the repair cycle process. His model incorporated the effect on system performance of the multiple indenture relationships between an end item and its component parts all the way down the product structure to the consumable components.

Mirasol's representation is based upon an end item with multiple indentured components within the context of a single echelon repair and inventory system. Mirasol formulated his model of this reparable inventory system as a closed queuing network with five nodes. These five nodes represent stages which incorporate the various activities in the repair process. In the first stage, serviceable assemblies (or LRUs using the Air Force convention) fail while installed on end items during their operation. The second stage represents transit time to the repair facility.

The third stage is a series of parallel paths corresponding to each of the components (SRUs in the Air Force convention) which make up the failed LRU.

Each of these paths model the waiting time for their respective component. If the failed SRU is available in stock, then this stage is bypassed entirely. Mirasol assumes independence between the probability an individual component caused the end item's failure and the probability of that component's being unavailable. This allows him to use the product of these two probabilities as the probability of entering a given path.

This third stage is the key to Mirasol's model since sequential solutions of the sub-models for each of the component parts determine the probability of unavailability, and by extension the probability of being able to bypass the third stage as well as the waiting time in the individual paths. In these sub-models, the first stage represents serviceable stock of the SRU, thus the probability of an SRU's unavailability is the probability of an empty first stage in that particular sub-model. By extension, the complementary probability is the probability that this stage will be bypassed for that particular SRU. The duration of a component's awaiting parts path is the inverse of the conditional output rate from the final stage of the component's sub-model given that its first stage is empty.

The fourth stage models the repair time which consists of time awaiting maintenance, diagnosis, removal of the failed component, and replacement of the serviceable component. The fifth stage corresponds to the transit time of the serviceable assembly back to the inventory location. The states in Mirasol's model are defined by the number of items in each of the stages. Since Mirasol models these stages as a closed Jackson network, he is able to use balancing equations and the resulting product form solutions to derive the steady state probabilities for the individual states.

Ahmed, Gross, and Miller (1992) use a network of queues representation in conjunction with METRIC to capitalize on the computational efficiency of the METRIC-based models and the precision of the queueing approach. They point out that infinite source, infinite repair channel models (which they designate as ∞/∞) are inherently more efficient than the more precise finite source, finite channel models (designated as *f/f*). The METRIC-based models are classic examples of ∞/∞ models. The computational efficiency of the ∞/∞ models make them a natural choice of the decision maker as long as their simplifying assumptions do not significantly detract from their accuracy.

The authors propose using a single comparison between the *f/f* and ∞/∞ models to establish how closely the ∞/∞ model follows the *f/f* model, then use this difference to adjust the results of the more computationally efficient ∞/∞ model to estimate the performance of the *f/f* model. Once calibrated, the ∞/∞ model can be used repeatedly to model the *f/f* system based upon the original comparison. Although Ahmed, Gross, and Miller do not address any differences in performance between the ∞/∞ and *f/f* models based upon actual data, their research is important in that it explores one possible way to work around the simplifying assumptions of the METRIC-based models.

CONCLUSIONS

From this review of the reparable item problem literature, it appears that the state of the art is at a crossroads. On one hand are the METRIC-based models with their long history of application in the Air Force reparable item system. This history of application is, however, tempered by the limitations imposed by the simplifying assumptions inherent in the entire approach. On the other hand are the queueing models. Although they are a relatively new approach to the reparable item problem, they show great promise in being able to more accurately and flexibly model reparable item systems. This potential is, however, qualified by the technical problems with state space requirements. In the next chapter, we will expand upon the difficulties with both of these approaches to the problem and introduce a new queueing paradigm.

CHAPTER 3
CHANGING REPARABLE ITEM PARADIGMS

PROBLEMS WITH TRADITIONAL APPROACHES

Decision Variables and Objective Functions

These traditional representations of the reparable item problem approach the reparable item problem from a perspective which misinterprets the essence of the problem's real world applications. The foundational shortcomings of these models are most obvious in their definition of the problem, specifically in the decision variables and objective functions these models use.

In the case of the METRIC-based models, the number of units of each item that are to be stocked at each of the bases and at the depot are the decision variables. The objective of these METRIC-based models then is to minimize the expected number of backorders as a function of these decision variables subject to a limited budget for buying those parts.

In the queueing literature, Mirasol's (1964) decision variables are the stock levels for the various items and the number of repair servers. He introduces an objective function called "strategic unavailability" which is the product of the unavailability rate and the mean duration of that unavailability. His model attempts to minimize strategic unavailability by allocating a limited budget between additional stock levels for each of the individual items and additional repair servers for those items. Mirasol suggests that the decision variables can be extended to include other system characteristics such as transportation capability, thus modeling a more complete set of tradeoffs in the repair cycle process.

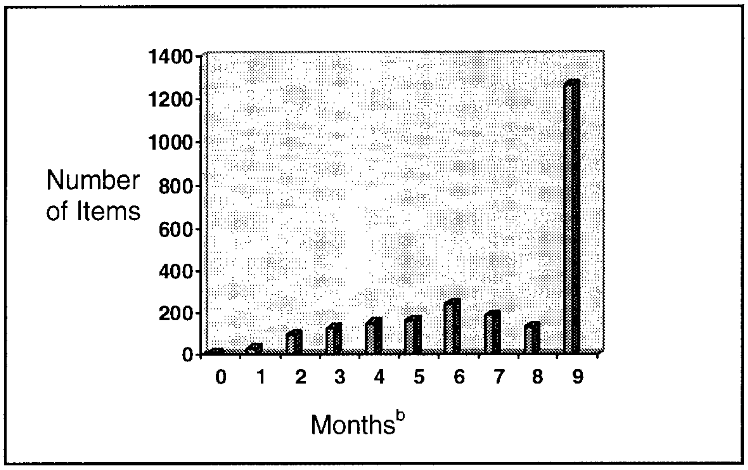
Gross, Kioussis, and Miller (1987) define the cases they use to demonstrate their decomposition approach with a variety of system characteristics which, in an optimizing routine, could be used as decision variables. Because of the flexibility of the queueing approach, stock levels, the number of servers, failure rates, and repair rates at the individual bases and the depot could all be considered as decision variables. Since the focus of their research was demonstrating the accuracy of their decomposition approach, Gross, Kioussis, and Miller did not specifically address objective functions. They did, however, point out that the values for a wide variety of performance measures can be derived from their approach's steady state probability distributions:

> . . . we can compute various means and other moments and also probabilities such as the probability that all bases have the desired number of operational units (system availability) or individual bases have the desired number of operational units (base availability). (Gross, Kioussis, Miller, 1987)

From the standpoint of decision variables and objective functions the perspectives behind these earlier approaches are somewhat detached from the realities of actual reparable item problems.

All the earlier approaches use the stock levels at individual locations as decision variables. The METRIC-based models rely exclusively upon these levels as their decision variables. Although the queueing approaches include other decision variables, stock levels are featured prominently.

In reality, stock levels are meaningless unless there are assets available to fill those stock levels. However, one common feature of real world reparable item applications is long procurement lead times. Since the number of assets in the system is fixed over any planning horizon shorter than the procurement lead-time, this means that in practice, stock level decision variables are inappropriate for any application with a planning horizon shorter than the procurement lead-time. This effectively eliminates many, if not most, potential reparable item problem applications. For example, in the Air Force context, the lead-time for reparable items consists of both administrative and procurement lead-time segments. Administrative lead-time is the time it takes the Air Force to award a contract to meet a given requirement. Procurement lead-time is the time it takes the source of supply to provide the contracted items. In the case of large orders with multiple deliveries, procurement lead-time is computed using the first significant delivery quantity. This practice keeps procurement lead-time and the resulting computed requirements from being inflated. Figure 3-1 below shows the reported historical values (in months) for administrative lead-time at one of the Air Force's depots.



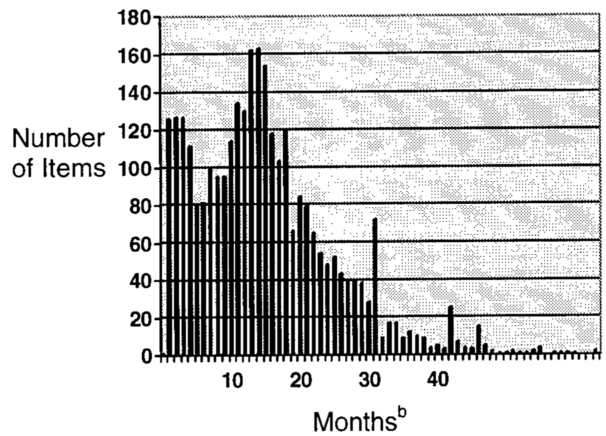
[a] Administrative Leadtime is the time it takes to award a procurement contract from the time a requirement is identified

[b] The Administrative Leadtime used in the Air Force's requirements computation process is rounded to the nearest month

Figure 3-1: Administrative Leadtime[a]

Out of 2,399 different items at this depot which used historical data to compute the administrative lead-times, 1,265 (52%) had lead-times of nine months or longer, and that's just the administrative component of lead-time, before procurement actually begins.

Figure 3-2 on the next page shows the corresponding values for procurement lead-time at the same depot.



[a] Procuremnet leadtime is the time it takes the source of supply to provide the items after the contract is awarded

[b] The procurement leadtime used by the Air Force's requirements computation process is rounded to the nearest month. It is computed using the first significant delivery quantity.

Figure 3-2: Procurement Lead-time[a]

There were 3,022 items at this depot which used historical data to compute the procurement lead-times. Of these, 1,702 (56%) had lead-times greater than a year. There was even a group of 487 (16%) whose lead-times exceeded two years. Given the large number of items with long lead-times, the use of stock levels as a decision variable in the Air Force reparable item context appears to be of limited relevance.

From an objective function perspective, the METRIC-based objective of minimizing expected backorders and other such average values for performance indicators have dominated the literature. Queueing models, however, have significant potential in their ability to generate probability distributions instead of a single value for the average of some performance indicator. It should be intuitively obvious that a probability distribution contains more information than a single average value. In actual reparable item environments, decision makers are more interested in the probability of an event than in some average value.

For example, in the context of the reparable item problem, a decision maker is more interested in the probability of having the desired number of end items serviceable than in the average availability. The following charts illustrate four cases all with a mean of five. If the decision maker is interested in having at least seven end items available, the probability of this happening ranges from 0 (in Figure 3-3) to 40 percent (in Figure 3-4) depending upon the selected distribution.

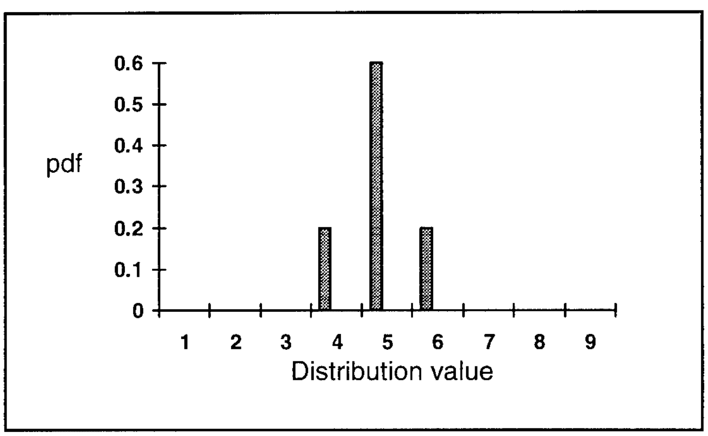


Figure 3-3: Sample unimodal distribution

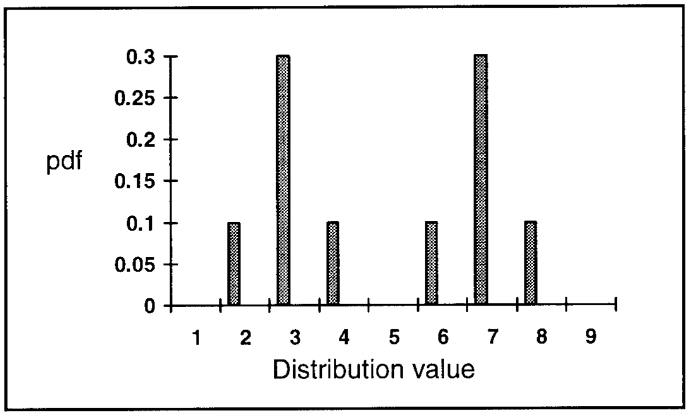


Figure 3-4: Sample bimodal distribution

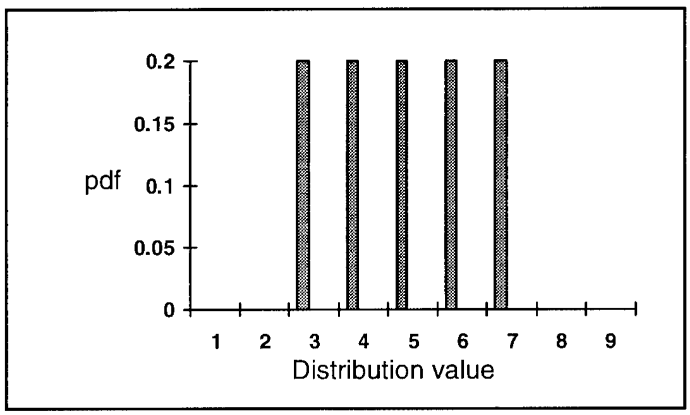


Figure 3-5: Sample uniform distribution

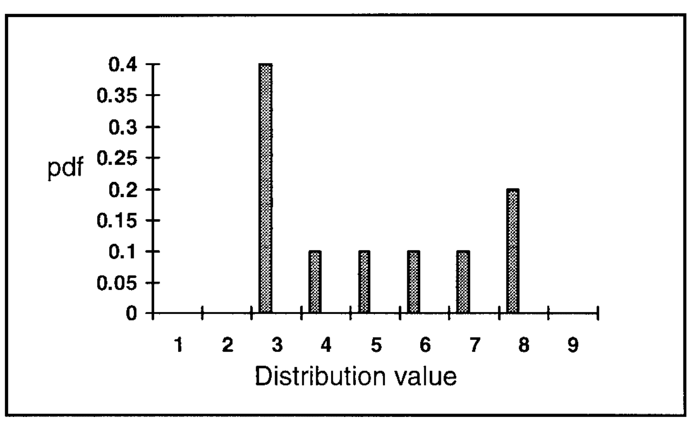


Figure 3-6: Sample concave distribution

Although average values, such as the METRIC-based expected number of backorders, enjoy widespread real world usage, using probability distributions instead of average values would open the door for objective functions with much greater practical relevance to reparable item decision makers.

Fitting Real World Failure Distributions

Another problem with these traditional approaches to the reparable item problem is the way in which they represent the failure processes of the individual components. The queueing models proposed by Mirasol (1964) and Gross, Kioussis, and Miller (1987) both use exponential failure rates. The METRIC-based models have represented the failure process with more complex failure distributions such as compound Poisson (METRIC and MOD-METRIC), negative binomial (Vari-METRIC), and even time-dependent rates (Dyna-METRIC).

In spite of the progress METRIC-based models have made in achieving more and more realistic failure representations, RAND has begun development of Dyna-METRIC 5, a multi-echelon inventory model for reparable items in which, Monte Carlo sampling has replaced the analytic computation of probabilities based on extensions of Palm's theorem. Isaacson and Boren (1988) indicate that one of the primary reason for abandoning the earlier versions is that they, "did not accurately represent the uncertainty in demand and repair, especially the queuing caused by repair constraints."

In his analysis of the variability of demand, Sherbrooke (1984) evaluated the worldwide demands for a sample of 1,030 Air Force reparable items over a period of four years and another sample of 810 items over a period of two and a half years. Based upon the demand patterns he observed in this empirical data, he concluded,

> . . . neither the gamma-Poisson models or compound Poisson models are in agreement with the data. While negative binomial state probabilities are still a useful representation of demand during a period of time, a better model of how demands occur in time is needed.

In a RAND study which analyzed actual component failure data for the F-15, F-16, and C-5, Crawford (1988) also found that the variability of demand was a problem. He concluded, ". . . logistics models that assume constant means and Poisson arrival processes do a poor job of modeling real-world demands for real-world airplane parts." For a selected group of "problem items" for the F-15, he showed that only 25 percent of the part-base combinations had variance to mean ratios (VTMRs) of less than two. By contrast, 20 percent had VTMRs in excess of eight. He illustrated the impact of VTMR using a Dyna-METRIC run for a squadron of 24 aircraft in a typical wartime scenario. With a VTMR of 1, the expected number of not fully mission capable (NFMC) aircraft by day 30 would be 5. However, increasing the VTMR to 2 resulted in a corresponding increase in the NFMC rate to approximately 9. When the VTMR is 5, the expected number of NFMC aircraft increases to 16 out of a squadron of 24.

State Space Size Problems

One of the key obstacles to applying the CTMC approach to the reparable item process is the size of the states space. The CTMC approach can make reparable item models much more representative of reality by adding such features as state-dependent failure rates and multiple capacitated repair servers instead of the infinite calling population and infinite repair server features which are commonly cited as limitations of the METRIC-based models. This reality, however, brings with it large state spaces.

When the individual states in a CTMC are composites of a number of indicators throughout the system being modeled, the number of states can grow rapidly. For example, one way to define the states in a multi-echelon reparable item system with a single depot is the number of serviceable units, the number of units in repair, and the number of backorders with the depot at each of the bases and the number of units which are serviceable and in repair at the depot .

The number of feasible states in this representation is constrained by the fact that Air Force reparable item stockage policy limits the number of units (serviceable or in repair) and backorders at the base level to an established "requisitioning objective." The number of feasible states is further restricted by the limited number of units in the entire system. Because of these restrictions, one of the base-level components of the state definition is redundant as is one of the depot-level components. In light of these constraints, the number of states required for this representation is the product of the number of feasible states at each of the individual bases plus the requisitioning objective at the depot.

Given the standard reparable item problem limitation on the number of units at a given base, computing the number of feasible states at a given base is analogous to an indistinguishable balls in urns (non-negative solutions) problem where the number of state definition components at the base-level corresponds to the number of urns (even though one is redundant, all three are used for this computation), and the requisitioning objective corresponds to the number of balls. The number of feasible states is given by Formula 3-1.

$$\frac{(n+r-1)!}{(r-1)!n!} \qquad (3\text{-}1)$$

Where: n is the base's requisitioning objective
r is the number of variables in the state description
(3 in this case)

Given the state space requirement for this type of queueing representation, it becomes apparent that the number of states would be prohibitive for any realistically sized problem. A numerical example will illustrate the state space difficulties with real world problems. A relatively small base may have 12 assigned aircraft of a given type (larger bases have 48 or even 72 assigned aircraft of a single type). If we assume a requisitioning objective of just 5 units more than the number of assigned aircraft, the number of feasible states at that single base is a very manageable 171. Taking into account more than one base and the depot's stock level, the number of required states is given by Formula 3-2.

$$\text{Number of states} = RO_D + \prod_{\forall Bases} (RO_n + 1) \qquad \text{(3-2)}$$

The number of states increases dramatically when more bases are added to the model. Assuming a requisitioning objective of 5 at the depot, the following table illustrates the state space problem as identical bases are added to the problem:

Table 3-1: Queueing Model State Space Requirements

Number of Bases[a]	**Number of Required States**[b]
1	176
2	29,246
3	5,000,216
4	855,036,086

[a] Assumes identical bases where n=17 (12 assigned aircraft + authorized stock of 5)

[b] Number of base states computed from Formula 3-1, total required states computed from Formula 3-2

While the growth in the number of states is rapid, the growth in the size of the infinitesimal generator is even more rapid because of its matrix format. With just four such bases, the CTMC infinitesimal generator representing the problem would have more than 7 x 10^{17} cells.

These prohibitive state space requirements associated with using a single CTMC infinitesimal generator to model the entire system provided the motivation for the Gross, Kioussis, and Miller (1987) decomposition approach discussed earlier. Since the individual bases are treated separately in their model, the number of required states for each of the base models is given by the number of non-negative solutions to the ball and urn problem described earlier. Table 3-2 below outlines some numerical examples of the required state size for the decomposed local base models.

Table 3-2: Decomposed Base-Level State Space Requirements

Number of Assigned Aircraft[a]	**Number of Required States[b]**
12	171
24	465
48	1485
72	3081

[a] The value of n used in Formula 3-1 is the number of assigned aircraft + an assumed authorized stock of 5

[b] Computed using Formula 3-1 with n based upon the number of assigned aircraft and the value of r fixed at 3

Even using the decomposition approach proposed by Gross, Kioussis, and Miller, the local base models would require a large number of states and entail computations and manipulations with very large matrices. For example, the case shown in Table 3-2 in which 3081 feasible states are needed to represent a base with 72 assigned aircraft would require a 3081 x 3081 matrix, or 9,492,561 individual cells.

The local model of the depot employed by Gross, Kioussis, and Miller is also hampered by its state space requirements and the corresponding size of the matrix which must be manipulated. The state definition for their local depot model separately reflects the number of backorders owed to each base as well as the number of serviceable units at the depot, and the number of units in depot repair. Because it is assumed that the depot will release any serviceable stock before allowing base backorders and because one of the components of the depot's state description is redundant, the number of required states is given by Formula 3-3.

$$RO_D + \prod_{\forall Bases} (RO_n + 1) \qquad (3\text{-}3)$$

Where RO is the requisitioning objective for the depot or a given base depending upon the subscript.

As the number of bases and the number of units allowed at each base increases, the state space grows rapidly. The numerical examples in the Table 3-3 show the state space requirements for the local depot model given various combinations of requisitioning objectives and the number of bases in the problem. All cases assume an authorized base stock level of 5 and a depot requisitioning objective of 5.

Table 3-3: Decomposed Depot-Level State Space Requirements[a]

Number of Bases	12 planes	24 planes	48 planes	72 planes
1	23	35	59	84
2	329	905	2,921	6,246
3	5,837	27,005	157,469	493,044
4	104,981	810,005	8,503,061	38,950,086
5	1,889,573	24,300,005	459,165,029	3,077,056,404
.				
.				
125	$> 10^{156}$	$> 10^{184}$	$> 10^{216}$	$> 10^{236}$

[a] Computed using Formula 3-3, where RO_n = number of assigned aircraft + authorized stock of 5 and $RO_D = 5$

These numbers are just the state space requirements. Recall that the CTMC approach uses n x n matrices, where n is the number of required states, so the matrix size would be even more prohibitive.

Gross, Gu, and Soland (1993) explored a series of alternative methods for solving the large CTMCs generated by the reparable item problem. Their research found that a two-phase iterative procedure was the best of their alternative methods. However, even using this technique, solving two and three-base problems with one million states took 4 and 5 hours. Based upon these running times, they acknowledge that approximation or simulation techniques might be more appropriate for realistically sized problem.

A NEW REPARABLE ITEM PARADIGM

Introduction

When examined in the context of realistic applications, the traditional approaches to the reparable item problem exhibit some fundamental shortcomings. Although METRIC-based models are actually in use in the Air Force's reparable item program, the focus implied by their decision variables limits their relevance, and their specialized structure limits their flexibility to change this focus. The METRIC-based models also suffer from their limited capabilities in modeling the highly variable component failure process.

The queueing models have not developed the same degree of sophistication as the METRIC-based models, but the queueing techniques have demonstrated the flexibility to handle a wider variety of decision variables and model more complex failure processes than the METRIC-based approaches. The current queueing approaches to the reparable item problem, however, have state space problems when they are faced with realistically sized problems.

This research addresses the reparable item problem from an entirely different perspective than the existing literature. The approach predicated upon this new perspective is able to avoid these shortcomings of the earlier approaches. This new perspective stems from two basic questions, "What types of answers do the decision makers in actual reparable item systems really need?" and "What are the real world constraints these decision makers deal with on a daily basis?"

The remainder of this chapter will explain the alternative perspective of the reparable item problem upon which this research is predicated, contrasting it with the traditional perspectives and their shortcomings discussed earlier. In the process, it develops and justifies a series of assumptions which form the foundation of this new perspective.

Repair Allocation as the Decision Variables

One of the key shortcomings of the traditional models of the reparable item problem demonstrated earlier is their use of the stock levels at the individual locations as key decision variables, or in the case of the METRIC-based models, the only decision variables. The new perspective taken in this research hinges on a new set of decision variables to replace the traditional stock level decision variables.

Our selection of decision variables stems from the basic question, "What decisions do real world reparable item managers make on a recurring basis?" Because of the typically low condemnation rates for reparable items, their significant ordering costs, and their relatively small repair costs, real world decision makers don't make stock level procurement decisions on a regular basis. Even when decision makers take procurement action, the long lead-time delays described earlier postpone the impact of the new stock levels into the long-term planning horizon. Using the Air Force repair cycle as an example, only one requirements computation a year includes procurement decisions, and even then, in FY 96, out of 18,099 "active" items, there was procurement action on only 4,424 (24%) of these items.

Unlike procurement action, repair action is a recurring task for repair cycle decision makers. Since reparable items are, by definition, constantly being repaired, decision makers are repeatedly faced with the decision of how many of which item to repair for which location. In contrast to the single procurement computation each year, the Air Force performs four requirements computations a year to determine repair requirements. Because of the contrast in relevance between procurement and repair decisions, the new perspective focuses on repair allocation to each of the different part-location combinations as a decision variable alternative to the stock level decision variables used in earlier research.

ASSUMPTIONS

In order to evaluate the effect of these repair allocations, we must be able to translate them into end item availability. We will develop the methodology for doing this in the subsequent chapters. For the time being, however, it is sufficient to view the translation as a complex process dependent upon the interaction of a wide variety of factors throughout the repair cycle. There are, however, two key simplifying assumptions which form the basis for our technique for converting the repair allocation variables into end item availability at the individual locations without the intervening complexities of the global repair cycle entering into the equation.

Ample Supply of Unserviceable Parts

The first assumption is that an ample supply of unserviceable items exists at the centralized repair facility. This allows us to truncate the repair cycle at the point where the individual locations return a failed item to the central repair facility for repair and ignore the complexities of retrograde traffic and "starved" repair servers. Since this research specifically targets the Air Force's reparable item problem, we use the characteristics and practices associated with the Air Force's reparable item program to justify this simplifying assumption.

Because of fleet downsizing and reduced flying hour funding, many line items in the Air Force reparable inventory have enough unserviceables available on the depot shelf to guarantee that routine base retrograde will be sufficient to ensure that the depot repair process is never starved. In these cases, this assumption is clearly justified. Since Air Force reparable policy requires the bases to return an unserviceable when they establish a backorder, it follows that the flow of unserviceables back to the depot should always be comparable to the flow of backorders, and by extension should be comparable to the repair requirement. As a result, most Air Force line items should fit into this category.

There will, however, always be some outlier cases in which the unserviceables are not readily available. In these cases, the Air Force intensively manages the retrograde process to keep depot repair from being constrained by carcass shortages. Two of the specific tools they use are the reparable item movement control system (RIMCS) and the "express table."

RIMCS codes are assigned to each stock number by the depot repair function to tell bases what priority the retrograde shipments should have. Those items for which unserviceables are in short supply at the depot receive priority processing and premium transportation from the bases. Certain unserviceables will even be shipped Federal Express to ensure rapid retrograde to the repair facility. On the depot end of the retrograde shipment, depot processing can be expedited using the "express table" concept under which a specialized processing line bypasses ordinary in-checking procedures and moves critical unserviceables directly from the freight receiving function to the repair technician.

The end result of all this specialized management of unserviceables is the fact that even when an ample supply of unserviceables does not exist, the depot repair process is rarely without a sufficient stock of unserviceables to support the required output rate, and the ample unserviceables assumption holds true.

Table 3-4, adapted from data supplied by the Logistics Management Institute, illustrates this assumption.

Table 3-4: Air Force Supply of Unserviceables

Months of Supply[a]	Number of NSNs	Percentage of NSNs[b]
0	475	10.0
1	60	1.3
2	97	2.0
3	75	1.6
4	95	2.0
5	66	1.4
6	96	2.0
7	46	1.0
8	83	1.7
9	60	1.3
10	58	1.2
11	40	0.8
12	206	4.3
13 - 24	583	12.3
25 - 36	380	8.0
> 36	2336	49.1

[a] Months of supply was computed by dividing the unserviceable depot inventory by the monthly demand rate for the item. Months of supply was rounded to the nearest number.

[b] This is the percentage of all 4,756 USAF reparable items with base-level retrograde shipments to the depot falling into a given category.

These 4,756 items represent all Air Force reparable items with base-level retrograde shipments to the depot. Months of supply was computed by dividing the unserviceable depot inventory by the monthly demand rate for the item.

To put this into perspective, Silver, et al (1991) analyzed a complete year of unserviceable shipments, consisting of over 500,000 transactions, and found that the average shipment time was only 14.4 days. Given the RIMCS procedures for rapid retrograde handling, improvements in the retrograde process since that time, and the fact that the data was only available in one month segments preventing us from identifying those units in the one month category which actually beat the shipping time, it would be a very conservative estimate to say that the supply of unserviceables was only insufficient to cover depot repair requirements for unserviceables during the retrograde shipment time in 11.3 percent of the cases.

Simplified Repair Capacity Constraints

The first assumption tells us that centralized repair is not constrained by the availability of unserviceable assets. The second assumption is that a component's repair capacity at the centralized location is not constrained by personnel, equipment, or other such limitations, rather, it is constrained only by total available repair funding. This further simplifies the problem's capacity constraints from a joint function of funding and physical capacity to a simple total repair budget constraint.

The assumption that capacity is only constrained by available funding is similar to the infinite server assumption behind the METRIC-based models. Sherbrooke (1992) and other proponents of the METRIC-based models defend this assumption based upon management intervention in the repair process which ensures that needed items are expedited. In the specific context of the Air Force's depot repair process, this assumption is also supported by the fact that there are numerous civilian contractors and other Department of Defense depots with the capability to repair these items. For many reparable items, civilian contractors, other Air Force depots, and even Army, Navy or Marine Corps depots already actively bid against the primary Air Force depot for the workload.

It should be noted that the repair cost function may not be linear because of the increased marginal cost of buying repair capacity from the next lowest bidder when the lowest bidder runs out of repair capacity. The non-linear nature of the cost function does not, however, detract from the fact that, in the Air Force reparable item context, repair funding is the only true constraint on the repair rate for a given item.

The New Reparable Item Problem

Based upon these assumptions, the depot can be envisioned as providing parts for the individual bases at some rate which is constrained only by the available repair budget. In keeping with this paradigm, the performance indicators for the individual bases are a function of the quantity of the different types of items the depot-level chooses to repair for each of the bases. Accordingly, the reparable item problem reduces to determining the optimum rates at which the depot should repair the various parts for each of the bases while staying within the budget limitation.

CHAPTER 4
MODELING THE BASE-LEVEL REPAIR CYCLE WITH OPEN QUEUEING NETWORKS

INTRODUCTION

In the previous chapter, we developed the basic outline of our new paradigm for modeling the reparable item problem. At the core of this new perspective is the conclusion that base-level performance is a direct function of the repair allocation decision variables and that the reparable item problem can be reduced to deciding the repair rate for each LRU-base combination. Given this redefinition of the reparable item problem, our objective is to allocate depot repair in such a way as to get the best global performance possible from a limited depot repair budget.

Before we attempt to optimize global performance, however, we must develop the basic building blocks that make up the global model. The reparable item problem has a well defined hierarchical structure. The "best global performance possible" we're trying to achieve from our given repair budget is actually a composite of the performance levels at the individual bases. This base-level performance, which in the Air Force context is usually measured in terms of aircraft availability, is in turn a function of the availability of the individual subsystems. Continuing down the hierarchical structure, subsystem availability is dependent upon the availability of its component LRUs, the SRUs which make up the LRUs, and the consumable bits and pieces which go into the SRUs. For the sake of simplicity, in this research, we will address availability measures beginning at the LRU level in the hierarchy. Thus, the basic building blocks of our model are the representations of each of the LRUs at the individual bases.

The remainder of this chapter develops the specifics of the open queueing network models we use in this research to represent the individual LRUs. We begin by outlining the basic open queueing network representation of an LRU at the base level, then describe how more complex LRU repair cycle representations give this approach great flexibility in modeling the base-level repair cycle process and the resulting LRU availability distributions. In addition to the flexibility inherent in this approach, we also point out its computational efficiency. We then provide a numerical example to demonstrate the general modeling methodology. Finally, we explain our aggregation technique for deriving the end item availability distribution for an individual base given the component LRU availability distributions and illustrate using a simplified numerical example.

NETWORK REPRESENTATION OF LRU AVAILABILITY

An open queueing network is essentially a set of nodes representing individual queues with constant Poisson arrival rates from outside the network and fixed routing percentages between the nodes within the network. In order to represent the base-level repair cycle for an individual LRU with an open queueing network, our basic task is defining a network which accurately reflects the availability distribution of the LRU as a function of the depot repair rate. The following sections describe the open queueing representation in general terms. The next chapter will explore the techniques for actually fitting empirical availability data.

Basic LRU Model

The following illustration depicts the most basic open queueing network representation of the base repair cycle.

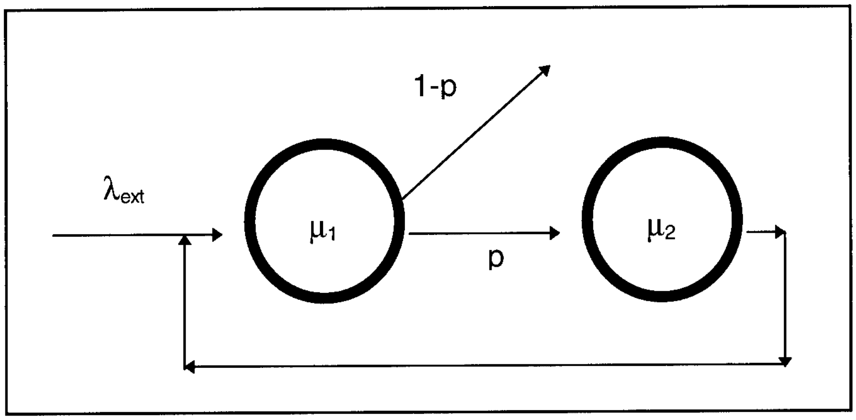


Figure 4-1: Base Repair Cycle Network Representation

It should be noted that this network representation is not an attempt to directly model the base-level process. Although there is no literal queueing of LRUs at the in-use node, we introduce pseudo servers and pseudo queues as an artificial device in order to replicate the base-level availability distribution for each LRU. In this pseudo network, the depot repair production rate is reflected as the external arrival rate (λ_{ext}) to the base's in-use node. This node represents those LRUs which are either installed on aircraft or are in stock awaiting installation. The "service rate" at the in-use node (μ_1) is the base's failure rate for the LRU being modeled. This rate represents a composite failure rate for all of that particular type of LRU at the base.

Leaving the in-use node are two paths. With probability p, corresponding to the percent base repair (PBR), the path branches to a base repair node. The rate along this path is the arrival rate at the repair node. This rate can be computed by recursively using the traffic equations for the network. With probability (1-p), the path leaves the open network, representing an unserviceable return to the depot. Because of the assumption of ample unserviceables at the depot, this path is of no further concern in our problem. The service rate at the base repair node (μ_2) is the base's repair rate for that particular LRU. As with the in-use node, this service rate represents a composite rate for the combined efforts of all "servers" who repair that particular LRU. The base repair node has 100% routing back to the in-use node of the model.

Jackson (1957) demonstrated that for an open queueing network, the steady state distributions for the queue lengths at the individual nodes behave as if the nodes were independent, in spite of the fact that their queue lengths are obviously not independent. This characteristic of the open queueing network leads to the expression in Formula 4-1 for the steady state queue length distribution at any given node:

$$\pi(n) = (1-\rho)\,\rho^n \qquad (4\text{-}1)$$

Where a node's traffic intensity (ρ) = the node's arrival rate (λ) / the node's service rate (μ).

Because the limiting queue length distributions for the individual nodes follow this computationally simple equation and behave independently, key performance indicators for each of the nodes can be efficiently computed, independent of what is happening at the other nodes in the network. From the reparable item problem perspective, the distribution of particular interest is the steady state queue length for the in-use node. In the repair cycle context, this is the distribution for the number of serviceable LRUs. These serviceable LRUs translate into aircraft availability in the depot-level marginal allocation routine. Since this representation produces a probability distribution, not just an average number, the process is not only computationally efficient, but also gives this approach great flexibility in the types of objective functions it can support in the depot-level marginal allocation routine.

Modeling a Complex Failure Process

One of the key hindrances to using queueing representations of the reparable item process is the state space size problem discussed earlier. However, because the individual nodes behave independently, there is no need to enumerate the possible states of the entire network, or even "decomposed" segments of the network. This characteristic of the open Jackson network means that the approach taken by this research is not affected by the large state space requirements which render earlier queueing approaches computationally intractable for realistically sized problems. A byproduct of this freedom from state space limitations is added flexibility in using more complex models of the base repair cycle process which can more accurately replicate the complex failure rates observed in practice.

By representing the in-use and serviceable LRUs as an single node, the simple example of the base repair cycle in the previous section lumps all these LRUs together and implies that the failure rate for the composite group behaves as if it were an M/M/1 queue. As discussed in the previous chapter, this is an obvious over-simplification of reality. To lay the foundation for modeling more complex, and thus more realistic, failure processes, we will start with the end product the base-level repair cycle model produces for use in the depot-level marginal allocation routine and work our way back to the base-level repair cycle representation.

This end product is the steady state probability distribution for the number of serviceable LRUs. For example, in the case of the earlier simple exponential failure rate illustration, this LRU availability distribution is a geometric distribution defined by Formula 4-2.

$$\pi(n) = (1-\rho)\,\rho^{n} \qquad (4\text{-}2)$$

Our objective in creating a network to model more complex failure patterns is to make this LRU availability distribution replicate, within acceptable limits, the real world availability which results from varying levels of depot repair. Since the depot-level model only uses the end product of the base-level network--the LRU availability distribution--it is not necessary to replicate the actual failure process, it is only necessary to model the failure process in such a way as to make this distribution fit the empirical availability data.

The only criteria Jackson (1957) required in order for a network to have the characteristics described earlier were: Poisson arrival rates from outside the network to any nodes with external arrivals, exponential service rates at each of the nodes, and fixed probabilistic routing between the network's various nodes and departure from the network. These criteria give us great flexibility in creating a base-level network which produces an acceptable LRU availability distribution.

By increasing the number of in-use nodes, arranging the nodes in parallel or in series, and adjusting the routing percentages and service rates, it is possible to manipulate the resulting LRU availability distribution (which for the multiple in-use node network is simply the convolution of all the individual in-use nodes) to fit the required distribution. The next chapter discusses the specific techniques for using these network features to fit an actual availability distribution. The following figures illustrate some of the different types of configurations which could be used.

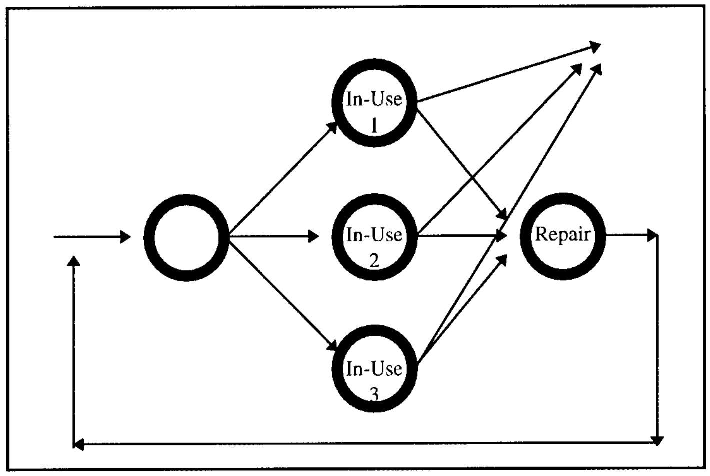


Figure 4-2: Three In-Use Nodes in Parallel

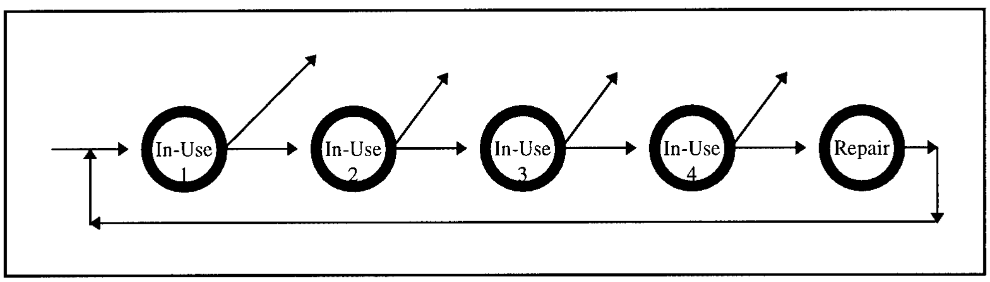


Figure 4-3: Series of Four In-Use Nodes

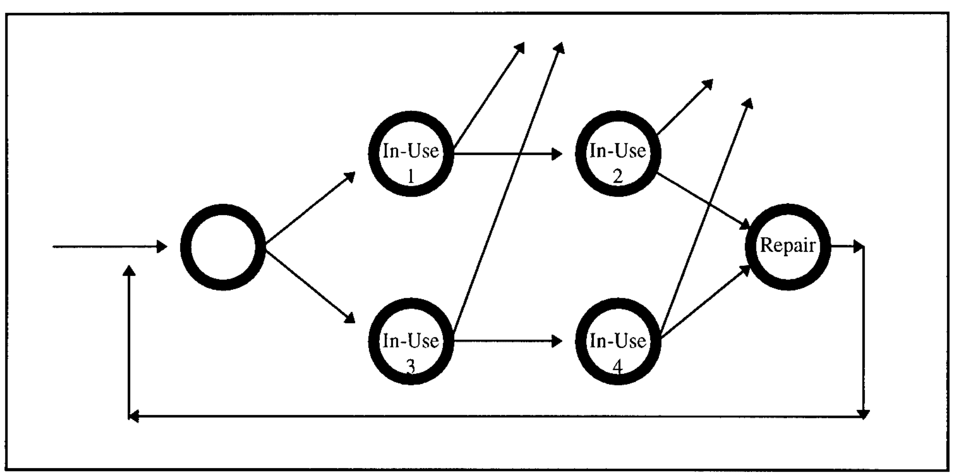


Figure 4-4: Two Parallel Series of In-Use Nodes

Solving for the Availability Distribution

We will use the following example with two in-use nodes in parallel to demonstrate the methodology for deriving the LRU availability distribution from any given queueing network representation of the base-level repair cycle process.

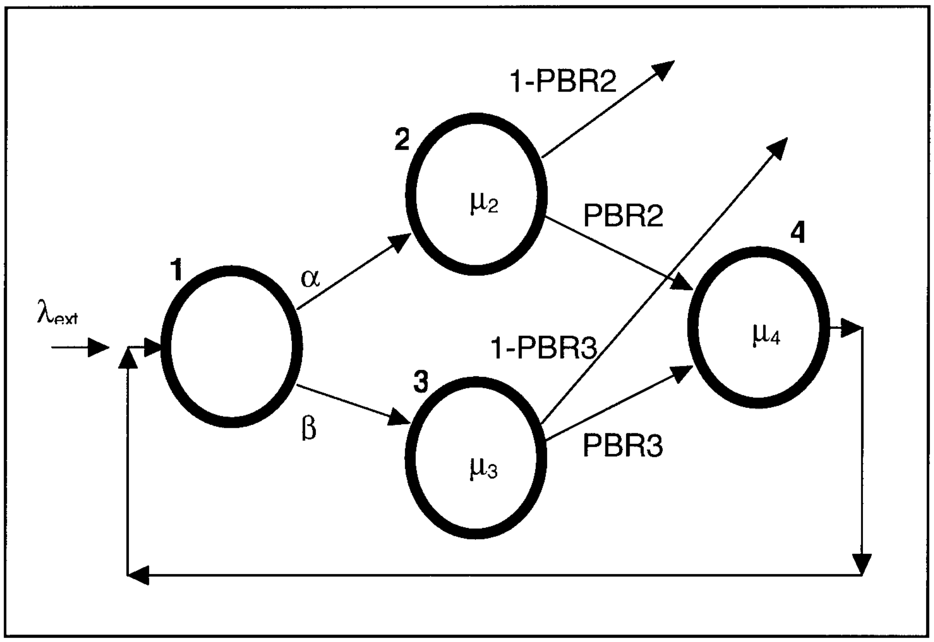


Figure 4-5: Base Repair Cycle Open Queueing Network

The first step in the process is to derive the traffic equations for each of the nodes based upon the specific structure of the network. The general form of the traffic equation for any node i is given by Formula 4-3.

$$\lambda_i = \lambda_{External} + \sum_{\forall nodes(j)} \lambda_j P_{ji} \qquad (4\text{-}3)$$

Where: λ_i is the composite arrival rate at node I
$\lambda_{External}$ is the arrival rate at node i from outside the network
P_{ji} is the fixed routing probability of going from node j to node i

The traffic equations for the sample network are:

$$\lambda_1 = \lambda_{ext} + \lambda_4 \qquad (4\text{-}4)$$

$$\lambda_2 = \alpha \times \lambda_1 \qquad (4\text{-}5)$$

$$\lambda_3 = \beta \times \lambda_1 \qquad (4\text{-}6)$$

$$\lambda_4 = (PBR2 \times \lambda_2) + (PBR3 \times \lambda_3) \qquad (4\text{-}7)$$

The next step is to solve for the composite arrival rates at each of the nodes. For example, in our sample network, the traffic equations express λ_2 and λ_3 in terms of λ_1. We can, in turn use these equations to express λ_4 in terms of λ_1.

$$\lambda_4 = (PBR2 \times \alpha \times \lambda_1) + (PBR3 \times \beta \times \lambda_1) \qquad (4\text{-}8)$$

Substituting this expression for λ_4 into the first traffic equation we can solve for λ_1 as a function of λ_{ext}, α, β, PBR2, PBR3 which are defined in another part of the problem:

$$\lambda_1 = \frac{\lambda_{ext}}{\left(1-\left(\left(PBR2*\alpha\right)+\left(PBR3*\beta\right)\right)\right)} \qquad (4\text{-}9)$$

Once actual values of λ_{ext}, α, β, PBR2, PBR3, μ_2, and μ_3 are supplied, we can solve for the remaining λ_i's. Using the μ_i's in the equation $\rho=\lambda/\mu$, we can also compute all the corresponding ρ_i's.

In the next step, we use these ρ_i's in the equation for steady state queue length probabilities:

$$P\{x_i=n\} = (1-\rho_i)\rho_i^n \qquad (4\text{-}10)$$

to derive the availability distributions for each of the nodes. Using Formula 4-10, even very large values of n have a probability of occurring. However, for the reparable item problem, we are only interested in availability probabilities for those values of n less than or equal to the number of available end items, since any available units above and beyond those in service on end items simply go into stock and therefore have no effect on end item availability. Accordingly, in the Air Force context, the LRU availability distributions are truncated at the base's number of primary assigned aircraft (PAA), and the probabilities associated with the remaining values of n are lumped into that final entry.

As a numerical example, let $\lambda_{ext} = 4$, $\alpha = 0.7$, $\beta = 0.3$, PBR2 = 0.6, PBR3 = 0.5, $\mu_2 = 8$, $\mu_3 = 6$, $\mu_4 = 7$, and PAA = 4. Using these values in the equations derived earlier, the computed values for ρ_2 and ρ_3 are .8140 and .4651 respectively. The resulting probability distributions for each of these nodes (truncated at the PAA of 4) are:

$$\pi_2 = [\ .1860\ \ .1514\ \ .1233\ \ .1003\ \ .4389\ ]$$

$$\pi_3 = [\ .5349\ \ .2488\ \ .1157\ \ .0538\ \ .0468\ ]$$

Since LRU availability is the convolution of the probability distributions at all the in-use nodes, the steady state availability distribution for the LRU represented by these nodes (also truncated at the PAA of 4) is:

$$\pi_{Total} = [\ .0995\ \ .1273\ \ .1251\ \ .1119\ \ .5362\ ]$$

This open queueing network modeling technique for individual LRUs implements our new paradigm for the reparable item problem in that it directly converts the depot repair rate into LRU availability.

DERIVING END ITEM AVAILABILITY

In order to derive end item availability from the availability distributions for the individual LRUs, we must model the associated indenture relationships. Much as the bill of materials (BOM) defines the hierarchical relationships between the component parts and the final product in the MRP problem, there are specifically defined hierarchical relationships between the component LRUs and the end item. However, unlike the MRP problem, in which requirements for the component parts are derived from demand for the final product, our reparable item problem works in the opposite direction, deriving end item availability through the hierarchical structure from the availability of the component LRUs. In order to take full advantage of this hierarchical relationship, we will introduce an assumption of "complete cannibalization."

Complete Cannibalization Assumption

Our basic approach for calculating the end item availability distribution from the component LRU availability distributions relies upon an assumption of "complete cannibalization." Complete cannibalization refers to the maintenance practice of pooling serviceable components from all end items under repair in order to produce the maximum number of serviceable end items possible. For example, if a maintenance organization had five end items, each of which was unserviceable because of a different component part, practicing complete cannibalization means they would take serviceable components from one of the unserviceable end items to repair the other four unserviceable end items.

The degree of cannibalization actually observed in the Air Force repair cycle process is not 100%, but it is very high. Some items with seals or specialized bonding to the end item are not cannibalized. However, since most reparable items on newer weapon systems have been specifically designed for easy removal and repair, the practice of cannibalization is almost 100%. Although complete cannibalization is an obvious simplification, in most cases it is a reasonable one. Without the complete cannibalization assumption, the model would have to associate individual failures with specific aircraft. Although this could be done by stipulating that each failure of an LRU which cannot be cannibalized would reduce the upper bound on the number of available aircraft by one, this complication will not be addressed any further in our research.

Weakest Link Availability

Assuming complete cannibalization, the number of serviceable end items is constrained by the smallest number of serviceable units of any of its component LRUs. As discussed earlier, if the number of serviceable units for each of the component LRUs exceeds the number of assigned aircraft, availability cannot exceed 100%. The following simplified examples based upon an aircraft consisting of just five LRUs illustrates the calculation of the number of available aircraft:

Table 4-1: Aircraft Availability Examples

LRU #1	LRU #2	LRU #3	LRU #4	LRU #5	PAA	Number Available
2	2	1	4	2	3	1[a]
3	3	4	3	4	2	2[b]

[a] In this case, the number available is constrained by LRU #3
[b] In this case, the number available is constrained by the PAA

We can extend this "weakest link" approach to aircraft availability to the process of computing a base's end item availability distribution from the availability distributions for its component LRUs. In order to do this, it is necessary to deal with cumulative distribution functions (c.d.f.) as opposed to simple probability distribution functions. We will denote the c.d.f. with F(x), which is defined as $P\{X \leq x\}$. Because we are specifically interested in the likelihood of availability exceeding a given level, we will use the compliment of the c.d.f., or 1-F(x).

In computing this compliment of the c.d.f. for an end item at a given base, we must systematically determine the "weakest link" among the end item's component LRUs. We start the process with a 1 x (PAA+1) vector for each LRU. The entries in these vectors represent the probability of observing n serviceable LRUs for n = 0, 1, ... PAA. We must first convert each of these probability distributions into their corresponding compliment of the c.d.f. The individual entries in this complimentary distribution vector represent P{x>n} for n = 0, 1,

...(PAA-1). Note that the entry for PAA is redundant since its probability is 0. In the applied context, this vector displays the probabilities that more than n end items will be available. Table 4-2 illustrates the steps in this process starting with a probability distribution f(x) for an LRU at a base with a PAA of 4:

Table 4-2: Availability Distribution Examples

n	0	1	2	3	4	
f(x)	.1	.2	.3	.2	.2	$P\{x = n\}$
F(x)	.1	.3	.6	.8	1.0	$P\{x \leq n\}$
1-F(x)	.9	.7	.4	.2	0	$P\{x > n\}$

In order to compute the base's end item compliment of the c.d.f., visualize arranging all the individual LRU vectors in a (# of LRUs) by (PAA) matrix. Each entry in the system availability vector is simply the minimum value from its corresponding column in this composite matrix. Table 4-3 demonstrates this process using the compliment of the c.d.f. for five LRUs at a base with four assigned aircraft.

Table 4-3: Base Availability Example

n	0	1	2	3
LRU #1	.9	.7	.4	.2
LRU #2	.8	.6	.4	.2
LRU #3	.9	.8	.7	.6
LRU #4	.7	.6	.5	.4
LRU #5	.8	.7	.5	.1
Base Avail	**.7**	**.6**	**.4**	**.1**

Summary

We have shown how starting with an open network representation of each of the individual LRUs at a given base we can, using the hierarchical relationship between the end item and its component parts, derive the base end item availability distribution. The base availability distributions for each of the individual bases will eventually be embedded in the global model where they will be used to compute the value of the objective function, and by extension give the global model a basis for determining the best allocation of limited repair funding.

CHAPTER 5
FITTING NETWORK PARAMETERS USING EMPIRICAL DATA

INTRODUCTION

In the previous chapter, we developed the open queueing network paradigm from a conceptual standpoint. In this chapter, we will demonstrate how this base-level modeling process might be implemented using empirical data. The starting point of the process is the construction of the empirical availability distributions. For the purpose of this demonstration, we will be using Air Force supply data. We will begin by describing the Air Force data we used to construct our availability distributions. Since the Air Force does not currently track LRU availability, we had to build a simulation to convert the data which is currently collected into the availability distributions required for our research. We will describe how our simulation model derives an availability distribution from the existing data. We will then describe how we generated network parameters to represent the base-level repair cycle in order to fit the empirical availability distributions from the simulation model. Finally, we will demonstrate our technique with an actual Air Force data set and compare the results with the theoretical availability distribution METRIC would produce using Palm's theorem.

CONSTRUCTING EMPIRICAL AVAILABILITY DISTRIBUTIONS

Before we can begin the process of constructing an appropriate representation of the base repair cycle, we need to have some information on the system we're trying to model. Since current approaches focus on stock levels and demand distributions, the existing data is expressed in these terms. The Air Force also maintains detailed transaction histories at the individual bases. In order to perform our availability based analysis, we had to translate these data records and supply transactions into an LRU availability distribution. We did this in a two-step process. In the first step, we isolate those transactions which affect the serviceable balance for the item from a much larger set of all supply transactions. The second step is to process these relevant transaction through a availability simulation which models the daily availability level. The collection of these daily availability levels over an 18-month period yields the LRU empirical availability distributions we will use to fit and test our network representation.

Data Collection

The Air Force Supply Data Bank collects various types of data on the base-level repair cycle process. Twice a year, the bases capture a snap shot of the state of the process at a specific point in time. Three types of these semi-annual records are of particular interest to this research. An LRU's "item" record captures indicative data on the LRU as well as the number of LRUs in stock. "Detail" records come in a variety of formats. Readiness Spares Package (RSP) detail records reflect the number of units authorized in support of war readiness as well as the number of serviceable units currently stocked in the package. Backorder detail records show any "holes" in aircraft for the LRU. The "Repair Cycle" record collects historical performance measures for the repair cycle such as the percentage of LRUs which can be repaired at the base and the length of time it takes the base to repair the LRU. The "Routing Identifier" record contains data elements such as the order and shipping time for a given source of supply or depot.

In addition to the semi-annual snap shot, each base also records every supply action involving reparable items. Each of the transactions in this data base have a similar structure. Some of the key data elements contained in this structure are described below. The "transaction identifier code" (TRIC) is a three-position field which identifies the type of transaction which is reflected by the particular record. Some of the key transactions we will use in this research are the issue (ISU or MSI), turn-in (TIN), due-out release (DOR), shipment

(SHP), and receipt (REC). The TRIC on a transaction record is modified by a variety of other data elements such as "type transaction phrase code" (TTPC) which more specifically identifies what is happening with that transaction, "supply action taken code" and "maintenance action taken code" which describe what action the supply or maintenance activity performed on the unit and whether or not they were able to repair the unit, "supply condition code" which tells whether the asset is serviceable or unserviceable, and "activity code" which defines the type of organization originating the action.

The program we used to collect the relevant transactions selected and excluded transactions based upon these data elements. It used TRIC to select only those transactions which could possibly affect the serviceable balance. Even within a given TRIC, only certain transactions were applicable to our analysis. For example, the ISU TRIC is used to reflect supply issues to maintenance, which we want to include in our analysis, as well as internal supply transfers, which do not affect availability. Our data collection program used the activity code data element to distinguish between these internal transfers and actual issues to maintenance. For TIN transactions, however, the program had to use the supply action taken code to distinguish between these internal transfers and actual maintenance actions.

Even isolating the appropriate supply action is not enough since a single supply action often produces multiple transactions. In the case of an issue from supply to maintenance, the computer produces two ISU transactions one to change the computer records to reflect the new serviceable balance and one to create a memo that maintenance owes supply a part. For the purpose of our data collection, we are only interested in the ISU transaction changing the serviceable balance. Our program used the TTPC data element to isolate the appropriate ISU transaction.

For some transaction types, the program had to differentiate based upon the supply condition code. For instance, there are SHP transactions for both serviceable and unserviceable assets. Our program eliminated the unserviceable transactions since the simulation decreases the serviceable balance when the asset is issued from supply, thus including unserviceable shipments would double count that decrease in serviceable balance. For the same reason, the data collection program only captured serviceable turn-ins.

Availability Simulation

The simulation begins with the records contained in the semi-annual snap shot and constructs the number of LRUs available at that specific point in time by adding the number of LRUs in stock from the item record, the number of RSP assets from the RSP detail records, and the number installed on aircraft from the base's PAA less any "holes" reflected on the backorder detail records. From this starting point, the simulation processes the transactions gathered in the previous step chronologically, incrementing or decrementing the physical inventory balance as appropriate for each individual transaction. At the end of a day's transactions, the LRU availability distribution is adjusted to reflect an occurrence of that day's ending balance. The simulation proceeds through the transaction data set, recording the resulting end-of-day inventory balances for each day. The FORTRAN code for this simulation is contained in Appendix B.

For the purpose of this analysis, we divided the 18 months of transactions into two 9-month segments. We used the availability distribution from the first segment to fit a distribution and the second distribution to test that fit both against the actual second period observations and the availability predictions based upon METRIC. Appendix C contains the availability distributions for the 20 individual items selected for this analysis as produced by our availability simulation. Subsequent sections will discuss the comparisons we performed using these distributions in greater detail.

FITTING A NETWORK REPRESENTATION

Given this empirical availability distribution, we developed a program to fit an appropriate network representation. This fitting process was greatly simplified by two key characteristics of our network representation. Firstly, since we are only interested in each node's contribution to the total availability distribution, it does not matter how a particular node fits into the in-use portion of the network. Secondly, since we are using the queueing network as an artificial construct to generate a desired probability distribution, the service rates at the individual nodes are not constrained in any way. Given this flexibility, there are any number of equivalent ways of building a network of queues to generate a given number of nodes with specific traffic densities. For example, by arranging the nodes in series and not allowing any departures from the system until after the last node, we can achieve the required density at a given node simply by setting its service rate according to Formula 5-1 below:

$$\mu_{\text{Node } i} = (\lambda_{\text{External}} / \text{PBR}) / \text{Desired Traffic Density at Node I} \qquad (5\text{-}1)$$

The task of fitting a distribution thus reduces to selecting a collection of traffic densities which appropriately fits the empirical data. From this set of densities, we can build any number of equivalent networks, each of which produces the same densities at the individual nodes, and as a result, the same total availability distribution.

In order to simplify the process, we limited the number of possible nodes to four and the possible traffic densities to increments of 0.1 ranging from 0.1 to 0.9. This enabled us to use explicit enumeration to select our network parameters. We compared the alternative sets of parameters using the Kolmogorov-Smirnov (K-S) statistic. The K-S test statistic compares two cumulative distribution functions (CDFs), and is defined as the largest difference between the two distributions at any point along the x axis. For the interested reader, Pfaffenberger and Patterson (1977) and Netter, Wasserman, and

Whitmore (1993) contain more detailed descriptions of the K-S goodness-of-fit test.

TESTING THE FIT

We used the first 9 months of our data set as the fit period to determine the parameters of our network representation. The remaining 9 months served as our holdout sample for testing purposes. We compared the availability distribution forecasted by our network representation with the actual availability distribution from this 9 month holdout sample. In addition to comparing the network representation availability distribution with the actual availability distribution, we also compared it with a theoretical METRIC-based availability distribution.

In their application of Palm's theorem to the reparable item problem, the METRIC-based models assume that the number of demands during lead-time follows a Poisson distribution with a mean of λt, where λ is the demand rate and t is the repair cycle lead-time. Based upon these assumptions of the METRIC-based models, we derived a theoretical LRU availability distribution to compare with the actual availability distribution from the 9-month fit period data as well as the holdout data set.

In order to generate the METRIC-based availability distribution, we needed two parameters, the demand rate (λ) and the lead-time (t). For our

METRIC model's demand rate, we used the demand rate from the item record captured at the beginning of the period. In order to compute the lead-time, we used METRIC's definition of lead-time shown below in Formula 5-2.

$$\text{Lead-time} = (\text{PBR} \times \text{BRCT}) + (1-\text{PBR}) \times (\text{O\&ST} + \text{depot delay}) \qquad (5\text{-}2)$$

We captured the various components of Formula 5-2 for each stock number from the following sources. The base repair cycle record (one of the records contained in the semi-annual snap shot) contains PBR as well as the number of serviceable turn-ins and the cumulative number of repair days from which we can compute the Base Repair Cycle Time (BRCT). We extracted Order and Shipping Time (O&ST) from the base's routing identifier records (another type of record contained in the semi-annual snap shot). For depot delay time, we used the 16.6 day standard developed in Silver, et al (1991). Using Palm's theorem, we computed the METRIC-based availability distribution by working backward from the stock level, computing the probability from a Poisson distribution with mean of λt (which we defined earlier) of observing the number of demands required to achieve the specified availability level. The parameters we used to generate the METRIC-based availability distributions are shown in Appendix D.

Results

We performed the K-S goodness-of-fit test for both the network model and METRIC-based model on the 9-month fitting sample as well as the 9-month holdout sample. The following table summarizes the results of these comparisons between the availability distributions from the actual data, our network representation, and the METRIC-based model.

Stock Number	Network Fit	METRIC Fit	Network Forecast	METRIC Forecast
1560-01-152-6387JH	.0625	.2301	.0545	.5410
1630-00-758-3758	.1691	.6553	.1080	.4389
1650-00-757-3862	.1891	.7940	.2521	.4579
1650-00-930-3160	.1002	.5737	.3770	.6132
1660-00-573-6482	.1045	.5205	.2230	.5691
1680-00-852-0803	.1024	.7538	.2170	.6663
2840-00-066-9925RV	.2851	.4160	.3050	.7245
2925-00-939-1473RV	.0518	.3903	.1992	.3903
2995-00-759-9072	.0967	.4845	.4813	.5958
4510-00-740-1074JH	.1564	.3675	.2362	.6741
4810-00-573-6461TP	.1503	.2888	.1720	.4685
5998-00-064-8059NT	.0877	.2107	.1224	.3592
5999-00-067-3622NT	.1542	.7004	.2955	.9304
5999-00-121-4721JH	.2183	.6565	.5043	.8321
6105-00-960-9879	.0162	.1859	.4859	.5182
6130-00-967-2610NT	.0816	.0758	.1675	.1994
6605-00-955-3029JH	.0302	.2091	.2219	.2958
6610-00-051-0989	.1293	.6292	.2584	.6366
6610-00-056-7150	.4082	.4509	.5517	.4435
6615-01-018-1635	.1429	.8257	.2822	.8445

Table 5-1: Network vs. METRIC vs. Empirical

Although there was a wide range of performance levels for the network representations in terms of the K-S test statistic, the network representations consistently out-performed the METRIC-based model. In only 2 of the 40 comparisons between the two models was the METRIC-based model able to out perform the network representations.

CHAPTER 6
ALLOCATING REPAIR FUNDING

INTRODUCTION

In Chapter 4, we developed the concept of using an open queueing network representation to model an item's availability distribution as a function of the level of depot repair allocation for that part. Then in Chapter 5, we used an empirical data set to demonstrate the actual technique of fitting a network representation to an empirical availability distribution. In this chapter, we will expand upon this item by item representation by incorporating the individual representations into a marginal allocation routine which allocates available repair funding between the possible individual LRU-base combinations. We will begin by describing the marginal allocation technique and applying it to our queueing network approach to the reparable item problem. We will specifically discuss the marginal allocation technique's objective function, constraints, and actual operation in the context of our reparable item problem. Finally, we will demonstrate the marginal allocation approach using the open queueing network representations of empirical data derived in Chapter 5.

THE MARGINAL ALLOCATION TECHNIQUE

The basic concept behind marginal allocation is to incrementally allocate resources where they will generate the largest marginal benefit, or greatest improvement in the objective function per unit of resource expended. Marginal allocation is a widely used technique in the reparable item literature, Silver (1972) and Muckstadt (1973) use marginal allocation to achieve seperability between LRUs and SRUs in their versions of Mod-METRIC in determining stock levels. Miller (1968, 1974) uses it to allocate newly repaired items at the depot in the context of his Transportation Time Look Ahead technique. This research extends the marginal allocation technique to the question of which items the depot should repair for which bases in order to get the "most bang for the buck" out of a constrained repair budget.

The key to adapting the marginal allocation technique to our new paradigm of the reparable item problem is the way in which we have reduced the problem to determining the rate at which the depot provides parts for the individual bases as constrained by the available repair budget. In the context of a marginal allocation routine, the allocation being determined is the number of each item the depot should repair for each of the bases, the objective function is derived from the availability distributions for the individual items, and the key constraint is the depot's repair budget. Figure 6-1 shows, at a macro level, how we have adapted the marginal allocation technique to make this repair rate determination.

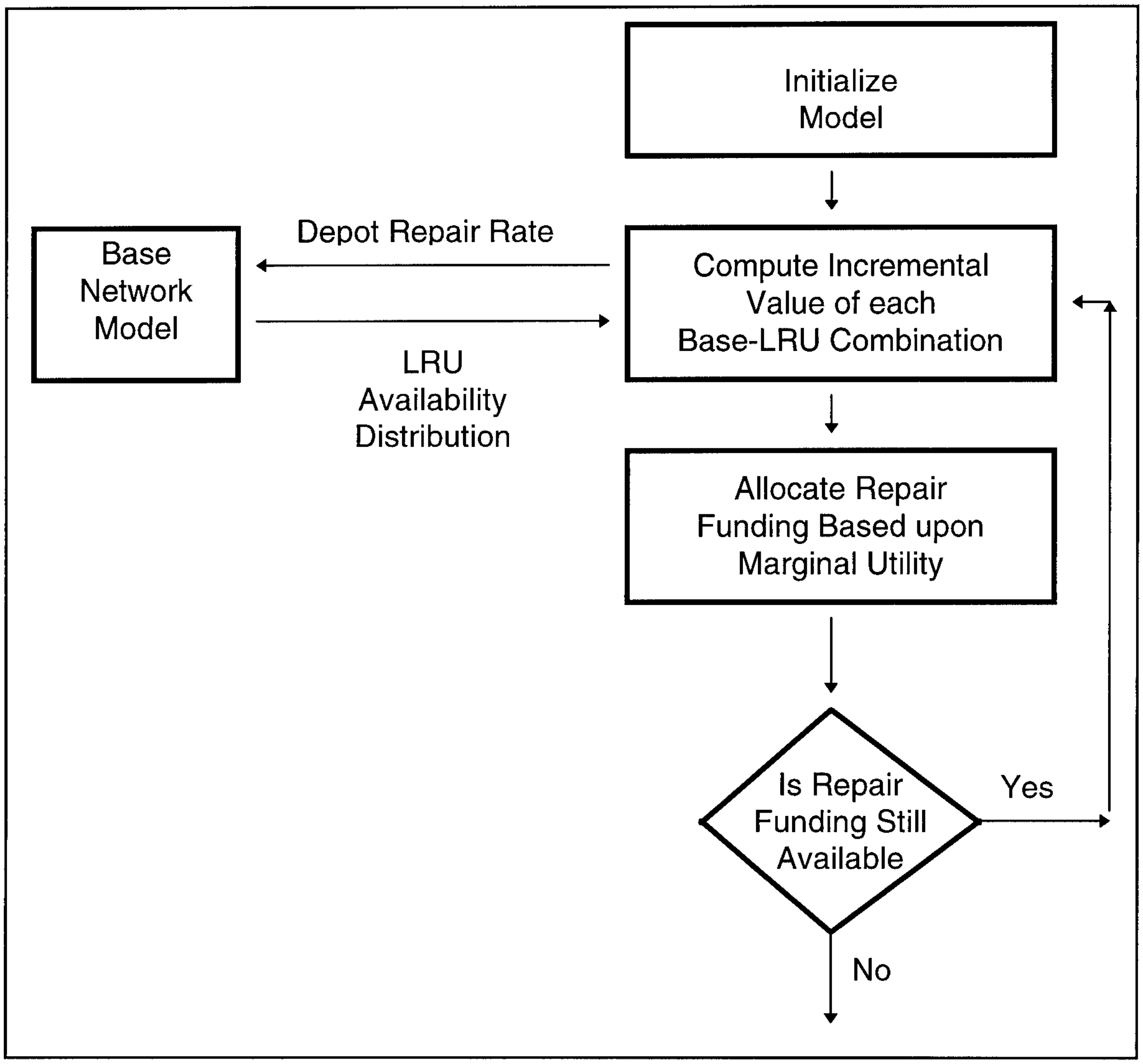


FIGURE 6-1: Reparable Item Problem Flow Chart

Objective Function

At the core of marginal allocation is the objective function. An objective function enables marginal allocation routines to compare the value of alternative allocations. Using the objective function, a marginal allocation routine compares the current value of the objective function with its value when each of the possible allocations is incrementally increased while the other possible allocations are held constant. This difference between the before and after objective function values is the incremental value of allocating resources to that particular alternative. Since a single allocation to the different alternatives often require varying amounts of a resource, marginal allocation routines express the incremental value as a function of a single unit of the resource. For example, in the repair context, the repair cost varies from item to item, but the incremental value of each potential allocation can still be expressed as the improvement in the objective function per $1,000 in repair spending. The marginal allocation routine makes the best allocation based upon this incremental value, then, because the incremental value of a given allocation usually changes after an allocation has been made, recomputes the incremental value of each alternative allocation and repeats the allocation process.

As discussed in the previous chapter, an objective function based upon a probability distribution is much more meaningful than one based upon a simple average. In the reparable item problem literature expected end item availability is a widely used performance measure. Using the queueing network approach, we can produce a probability distribution version of end item availability which measures the probability that end item availability will exceed a given level. We will use this type of objective function in some of the marginal allocation demonstrations at the end of this chapter.

Since our model produces probability distributions all the way down to the individual item level, there is much more information that would be available in formulating objective functions uniquely tailored for specific users. For example, in actual Air Force applications, this basic objective function should probably be modified to account for the fact that, in the global mission perspective, all bases are not created equal. The critical nature of some bases' missions makes their availability rates more important than other bases with the same aircraft. The objective function could, however, accommodate this reality by weighting the probability of a given level of availability to reflect a base's importance to the mission. For example, a base with a front line readiness mission might receive a weight of 1.5, while a base with a reserve mission might have a weight of 1.0, and a training base may only be weighted at .5. Thus, a one percent increase in the probability of achieving a given level of readiness at a front line base would have three times the value in the objective function as that same one percent increase at a training base. Another possible objective function which we will demonstrate at the end of this chapter is based upon the probability that an item's stock of wartime spares, or Readiness Spares Package (RSP), meets or exceeds a given percentage of the required level.

Given an objective function and the queueing network representations of the individual item, we can develop a module for our marginal allocation routine which computes the value of the objective function from the repair rates for the individual items. To compute the incremental value of any proposed allocation, we simply subtract the value returned by this module using the repair rates before the allocation from the value returned by this module based upon the repair rates after the allocation.

Constraints

Another key to the marginal allocation technique are the constraints on the problem. Implicit in the need for a marginal allocation program is the fact that resources are limited. Thus, the most obvious constraint on possible solutions is that the allocations of resources not exceed this resource limitation. This constraint is typically used as the stopping point for the marginal allocation routine. In the context of our proposed reparable item problem paradigm, the limiting resource is repair funding. The repair funding constraint can be as simple as subtracting a fixed repair cost from the total repair budget whenever the model makes a repair allocation, stopping the routine when there is not enough repair funding to make another allocation. It can, however, take a more complex form such as allowing a step function in the repair cost at certain levels of production.

In addition to the basic resource constraint, marginal allocation routines can incorporate a variety of other constraints at any number of points in the process. Some other types of possible constraints include a minimum acceptable objective function value at all locations or maintaining objective function value parity between locations within a given range. It should be noted that the objective functions referenced in these additional constraints need not be the same as the objective function used to determine the allocation of resources. They can be any function of the availability distributions for the individual items, or for that matter, a function of any available data elements. For example, in the demonstrations at the end of this chapter, we use a minimum acceptable probability of .75 that each item has no backorders to initialize the marginal allocation routine even though we use other objective functions to make the additional allocations. As an example of constraints within the allocation routine itself, some of our demonstrations incorporate a parity check before each allocation which will preempt an allocation if the value of some objective function for the receiving base exceeds that of the other base by some threshold. This type of flexibility in constructing constraints makes the marginal allocation technique a powerful management tool.

In addition to the obvious benefits to applicability, using a constraint or constraints to set the initial conditions for the marginal allocation routine can reduce computational requirements and in the case of queueing networks, ensure the marginal allocation routine is operating in the representation's feasible range. The closer the starting conditions are to the final solution, the fewer iterations the marginal allocation routine must go through. For this reason, it is obviously beneficial to initialize the beginning repair rates in a reasonable range. With networks of queues, there is always a danger of unrealistic conditions causing abnormal results. In the case of our reparable item representations, the queueing parameters are derived from the system's behavior within a specific range of traffic intensities. If the initial depot repair rates produce uncharacteristically low traffic intensities, the low intensity behavior of the queues may differ dramatically from their behavior under realistic intensities.

MARGINAL ALLOCATION DEMONSTRATION

Using the 20 items from Chapter 5, we constructed a two-location marginal allocation model with 10 items at each location. In order to demonstrate the impact of varying objective functions and constraints, we ran four variations of this routine. We initiated two of the models with a requirement that all items have a minimum probability of .75 of having no backorders. We initiated the other two models by requiring that all items have a minimum probability of .50 of having 50 percent or more of their RSP requirement on hand. In two of the variations, the objective function was the minimum probability across all items of having three or more items in stock. In the other two variations, the objective function was the minimum probability across all items of having 50 percent or more of the RSP requirement on hand. For each of the two objective functions, we constructed a model which had a constraint which preempted allocation if the value of the objective function at one base exceeded the objective function value at the other by more than a 0.1 threshold.

The MATLAB code for the marginal allocation routine with the fixed level objective function and the preemptive parity check is contained in Appendix E. Table 6-1 lists the data elements used by the marginal allocation routines and the sources of the data. The data set itself is provided in Appendix F.

Table 6-1: Marginal Allocation Routine Data Elements

Data Element	Source
Traffic intensity at the four network nodes	Generated by the empirical data fitting routine performed in Chapter 5
Calibration Repair Rate	Collected from the transaction data analyzed in Chapter 5
Number of leading calibration zeros	The number of leading zeros in the empirical distribution between the first non-zero entry for the specific item and the smallest non-zero availability level observed across all items
Repair cost	A data element on the semi-annual item record
RSP Requirement	A data element on the semi-annual RSP detail records
Repair Budget	Computed by multiplying the actual receipts for each item during the second nine months of transactions by its respective repair cost

Demonstration Results

The final allocations produced by each of the models are shown in Table 6-2. The actual allocation sequence for each of the models is given in Appendix G.

Table 6-2: Allocation Results by Model

Item # (Base)	Level Objective w/o Parity	Level Objective w/ Parity	RSP Objective w/o Parity	RSP Objective w/ Parity
1 (1)	12	14	10	10
2 (1)	6	7	4	4
3 (1)	1	1	8	8
4 (1)	10	12	20	21
5 (1)	9	10	9	9
6 (1)	11	12	11	12
7 (1)	8	9	5	5
8 (1)	11	12	11	11
9 (1)	16	18	16	16
10 (1)	1	1	1	1
11 (2)	10	9	6	6
12 (2)	13	13	10	10
13 (2)	13	12	11	11
14 (2)	11	7	10	11
15 (2)	8	8	7	7
16 (2)	14	12	8	8
17 (2)	12	11	9	9
18 (2)	12	11	9	9
19 (2)	20	20	20	20
20 (2)	20	18	27	19

It was interesting to note that the effect of the parity constraint was much more pronounced for the allocations produced using the level-based objective function than it was for those using the objective function based upon RSP requirements. The step-by-step allocation results illustrated the impact upon computational requirements of the constraint used to initialize model conditions. The level-based constraint did not allocate as many assets in the initial pass. As a result, the level-based models required almost three times as many iterations as did those based upon RSP requirements (59 iterations without parity and 58 iterations with parity compared with 19 iterations without parity and 20 with parity).

CHAPTER 7
CONCLUSIONS / RECOMMENDATIONS

CONCLUSIONS

This research examined the various approaches to the reparable item problem reported in the literature and demonstrated shortcomings in those approaches which hamper their effectiveness in real world applications. We saw that the METRIC-based approaches, in spite of their sophistication and years of actual large scale applications, are ill-equipped to handle the variability of empirical data and cannot address state-dependent behavior in a system. The newer queueing approaches successfully model state-dependent behavior, and by extension the variability of the empirical data, but run into state space problems when they attempt to solve realistically sized problems or address complex issues such as indenture relationships.

In a departure from these existing approaches, we developed and justified a new paradigm for the reparable item problem. Because the actual number of on-hand assets typically bears little resemblance to the authorized stock levels, which the traditional methods use as their decision variables, we abandoned stock levels as decision variables in favor of depot allocations of repair funding. In another departure from convention, we justified two key assumptions about the depot repair process--it is not constrained by the availability of unserviceable assets and it is not constrained by workshop capacity.

Based upon our new paradigm, we proposed and developed a new approach to the reparable item problem which uses open queueing networks to model the repair process. Taking advantage of the computational efficiency of the open queueing network representation and the detailed data available from our representations, we replaced the simplistic expected number of backorders performance measure used by the METRIC-based models with the more information-laden item availability probability distribution function. We showed how this improved performance measure opens up a broad range of different objective functions which give real world managers much greater flexibility in tailoring the reparable item problem to their actual requirements.

Empirical Results

We demonstrated our new paradigm and its resulting open queueing network approach to the reparable item problem, using a U. S. Air Force data set. From this data set, we extracted 18 months of transactions for a sample of 20 reparable items and the corresponding indicative data for those items. We first had to produce a simulation routine to convert the available data elements into the availability distribution format upon which our paradigm is predicated. We then developed the specific procedures and programs to determine the best queueing representations for the empirical availability distributions. In order to compare the performance of our model with METRIC, we also had to develop a program to convert the indicative data for an item into a METRIC-based theoretical availability distribution.

When we compared the performance of our queueing network models, both in fitting and forecasting the empirical data, with the performance of a METRIC-based model, our models out-performed the METRIC-based models in 38 out of 40 comparisons. As would be expected from METRIC's implicit assumption that the source of supply will attempt to stock up to the authorized stock levels, the METRIC-based models generally performed best when the availability distributions were skewed toward full stock and performed worse when the availability distributions deviated away from full stock. The queueing models, on the other hand, performed more consistently across the spectrum of availability distributions.

Based upon the queueing network representations generated for the sample items in our data fitting demonstration, we developed a global marginal allocation model to illustrate the marginal allocation of the depot's repair funding between competing bases and competing items at each of the bases. We developed and coded a basic marginal allocation routine for these items. We then produced two versions of the model based on different objective functions to demonstrated the flexibility inherent in the marginal allocation technique. We then modified these two models to include initialization constraints. These initializing conditions reduced the computational requirements from 218 iterations to 59 iterations for one model (a 73 percent reduction) and from 212 iterations to 20 iterations for the other model (a 91 percent reduction). We also modified the two basic models to demonstrate the use of constraint functions within the body of the marginal allocation process.

These results illustrate the promise of our new paradigm and the associated open queueing approach to the reparable item problem. They show the practical application potential associated with this perspective of the reparable item problem.

RECOMMENDATIONS

This research demonstrated at a basic level the principles of our new paradigm and illustrated some of its potential. However, in order to exploit this potential, there is much that needs to be done in terms of future research.

Future Research on Fitting Empirical Data

In our demonstration of fitting empirical data, we used an inefficient explicit enumeration search procedure to determine the best fit for a given empirical data set. Because of the computational limitations imposed by our search routine, we artificially limited the range of network structures we and traffic intensities we considered. One avenue of future research would be to develop a more efficient search routine which could reduce the computational requirements for fitting a given empirical data set.

There are a variety of extensions which could flow from this more efficient search routine. One such extension would be to expand the range of network structures and traffic intensities considered in the fitting process. By examining networks with more than four nodes and traffic intensities in increments smaller than 0.1, future researchers would be able to achieve better fitting performance from their network representations. A more efficient fitting routine would also enable future researchers to more fully explore the relationship between the improvement in the fit performance of the network representations and their predictive value. In our analysis, we found that increasing the number of nodes improved the fit performance, but that it did not necessarily improve the predictive value of the network representation in all cases.

Future research along this avenue could examine the correlation between improvement in forecasting performance and the characteristics of the empirical distribution being fitted or even the characteristics of the items which the empirical distribution represent. Researchers could classify empirical distributions by their shape or other such characteristics, then evaluate within each class the benefit, in terms of predictive value, of having a greater range of possible network representations available to the fitting routine. This type of analysis would be able to produce more accurate network representations from the forecasting perspective, which is the most important perspective in the application context.

Future research could also examine the multi-period prediction performance of the queueing network and ways to adapt the network representation to accommodate available data from period to period. We limited our demonstration to a single fit period and a single prediction period. Multi-period research, however, could analyze the benefits over time of a broader range of alternative network representations in the fitting stage. This extension to our research would also be able to explore fitting procedures using more than one fitting period. While our demonstration fit an empirical distribution by simply minimizing the value of the K-S test statistic, research into a multi-period fitting routine could develop and test a variety of other objective functions which it would attempt to minimize. Using time series forecasting as a model, some possible candidates would be the MAD or MSE measures using the K-S test statistic as the error value. Another component of this multi-period research extension would be exploring ways in which to make the network definition process adaptive over time. Once again relying on time series forecasting as a model, future researchers could compare the forecasting performance of network representation versions of the naive, simple moving average, and exponential smoothing models. A naive network representation model might simply use a fixed number of nodes and select the service rates for those nodes which best fits the last period's empirical data to forecast for the next period. A network representation version of exponential smoothing might also use a fixed number of nodes, but adjust the service rates at each of the network representation's in-use nodes by the dampened differences between the current service rates and the best fit service rates. Part of the multi-period fitting routine might be a comparison between different versions of these models with different numbers of nodes.

Yet another extension in the multi-period area would be proactive forecasting using additional information. For example, the Air Force's D041 requirements computation process relies upon forecasted flying hours to determine requirements. Since flying hours per period have an impact upon failure rates per period, future multi-period research could attempt to incorporate forecasted flying hours into the construction of the queueing network representations. This future research would have to be application specific, focusing on how the objective function for the specific application is affected by a variety of factors other than the time series of availability distributions. The research would also have to develop and test specific techniques for incorporating these forward looking indicators into the construction of the network representations.

Future Research on Marginal Allocation Techniques

Our approach's use of availability distributions as opposed to the traditional expected number of backorders opens up a wide variety of alternatives for the real world manager who would actually apply our techniques. Our demonstration only gave a glimpse of the flexibility our approach offers.

Since objective functions and constraints are, by their very nature, application specific, future research into the marginal allocation process would have an applied focus. Some possible avenues include developing appropriate objective functions for a specific application. Future researchers could also explore, within a specific application context, the impact of initializing constraints upon computational requirements. They could further customize a specific application with internal constraints.

Promise of the New Paradigm

This research represents a revolutionary departure from the established traditions of the reparable item problem. By abandoning the conventions of the existing literature and replacing them with a new paradigm which promises to deliver greater flexibility and applicability, this research opens a new frontier on the reparable item problem. Future researchers have much to gain by exploring our new paradigm.

APPENDIX A
FOUNDATIONS OF QUEUEING

INTRODUCTION

In its broadest sense, queueing applications revolve around the phenomenon of lines or queues. Whether it's the line at the grocery store (for which it is axiomatic that I will always end up in the longest one), a backlog of parts at a machine in a job shop, data packets awaiting transmittal in a communications network, or any of a vast number of other types of queues, there are some basic principles which can be used to analyze the performance of these systems. These fundamentals of queueing theory form the foundation for our analysis of the reparable item problem. The following sections describe some the key principles we will use in our research.

Notation

Before delving into the actual principles of queueing theory, the basic notation used to describe a queue is an instructive introduction of some of the queueing issues we will be exploring. Three of the key features of a queue are the distributions of its interarrival and service times and the number of service channels. These three features make up the basic notation for describing a queue illustrated below:

Arrival Process / Service Process / Number of Service Channels

For example, "M/M/1" describes a queue with exponential interarrival and service times and a single service queue. Queueing literature also refers to exponential processes as "Markovian," hence the "M" in the notation.

Two other process distributions which commonly show up in the literature are the general (G) distribution and the phase-type (PH). A general distribution in the queue description is typically used to indicate that the conclusions are can be generalized to the point where they depend only upon the distribution's mean and not the other specific characteristics of the distribution. The use of an $M/G/\infty$ queue in the derivation of Palm's theorem is a classic example of using the general distribution to emphasize the broad applicability of the results. The phase-type distribution denotes a queue with a specialized structure, the properties of which can be exploited in a variety of ways. We will describe the phase-type distribution in greater detail in a later section. In addition to this basic queue description notation, terms can be included which define the number of items in the total system, describe the queueing discipline, and other system characteristics. For the purpose of this research, however, we will not make use of the more complex notations.

MARKOV CHAINS

The behavior of queues can be addressed from a purely mathematical, formula-based standpoint as Gross and Harris (1974) have done in their queueing text. Another approach, however, is to approach queueing theory using the Markov chain embedded in the queue to analyze its behavior as Neuts (1981, 1989) and Kao (1997) have done. The embedded Markov chain for a queue with n possible states represents that queue as an n x n matrix. In the case of discrete time Markov chains, this matrix reflects the transition probabilities between state pairs with the row index defining the "from" state and the column index defining the "to" state. For continuous time Markov chains (CTMC), the cells contain transition rates instead of one step transition probabilities. In the context of the reparable item, which typically assumes a continuous review inventory policy, the CTMC is more appropriate than the discrete time version, so we will concentrate on the CTMC.

Because it can explicitly model the transition rates between each pair of feasible states, the CTMC is a very flexible modeling tool. By appropriately defining the states, a CTMC approach can be used to model many different types of queues or systems of queues. For example, to model a simple reparable item problem we could use a composite state definition which includes the number of items installed or in serviceable stock at the base, the number of items in repair at the base, and the number of items backordered with the depot. This example also illustrates one of the key limitations of the CTMC approach--state space size. Since the CTMC not only explicitly enumerates all possible states, but creates an n x n matrix, the state space size is computationally prohibitive for large problems and complex state definitions. When computationally feasible, however, the CTMC's versatility makes it a valuable approach to modeling stochastic processes.

The matrix containing the transition rates is called an infinitesimal generator of the CTMC and typically designated as "Q." Once generated, it can be used to compute the steady state probabilities of each of the states as well as their transient probabilities at any given point in time. Kao (1997) shows that by simply replacing the first column of Q with ones and inverting the resulting matrix, the first row of the inverted matrix yields the steady state probabilities of each of the possible states. For the time dependent case, he demonstrates that the probability vector at time t given a specific beginning state is simply the corresponding row of the matrix e^{Qt}. The fact that the steady state and transient probabilities, and by extension a wide variety of performance indicators, are so readily available via simple matrix operations, makes the CTMC approach a powerful modeling tool.

NETWORK OF QUEUES

A queueing network is, in the most general case, any series of queues which feed one another. A specific subset of queueing networks which has been the focus of much research are those networks in which the individual queues exhibit Markovian behavior (Poisson arrival rates and exponential service times) and departure routing is probabilistic and state independent. This specific subset of queueing networks has some special characteristics which can be exploited to overcome the state space problem with CTMCs discussed earlier. We will expand upon these characteristics in the next sections.

A key way of categorizing networks of queues is "open" vs. "closed." In an open queueing network arrivals from outside the network and departures from the network are permitted, while in a closed network, the only arrivals are the departures from other queues in the network, and no departures from the network are allowed. A key result of these definitions is the fact that the number of units in a closed network is fixed, while the number in an open network is unconstrained.

Open Queueing Networks

In his seminal research into the behavior of the individual queues imbedded in an open network of Markovian queues, Jackson (1957, 1963) observed that these individual queues appear to behave independently allowing their marginal distributions to have "product forms." Product form means that the limiting probability for a given vector **n**, which denotes the number of units in each of the queues in the network, is given by Formula A-1.

$$P(\mathbf{n}) = \Pi_{\forall i}\ (1-\rho)\rho_i^{n^i} \qquad \text{(A-1)}$$

Hence we apply the term product form. The ρ_i 's in the formula are the traffic intensities at each of the nodes in the network, and are computed by dividing a node's arrival rate (λ_i) by its service rate (μ_i). For any given node, λ_i is the sum of the arrival rates from the other nodes and any external arrivals. By capitalizing on the Markovian characteristics of the individual queues, it is possible to solve for the individual λ_i 's given the external arrival rate(s) and the departure routing probabilities for each of the nodes.

Closed Queueing Networks

Gordon and Newell (1967) extended Jackson's work by showing that closed queueing networks also could have product form solutions. Computing the product form solution of the closed queueing network is complicated by two key differences between open and closed networks. The first complication is the absence of outside input to provide a beginning point for computing the arrival rates at the individual nodes. Without any external arrivals, the set of balance equations which define the closed system is the summation of internal transitions at each of the nodes in the system. Formula A-2 below applies to each of the nodes in the system:

$$\lambda_j = \sum_{\forall i} \lambda_i P_{ij} \qquad \text{(A-2)}$$

Where λ_j is the arrival rate into node j and P_{ij} is the probability of a departure from node i being routed to node j.

There is not, however, a unique solution to this set of balance equations. Because of the constraint on the total number of units in the system, this set of balance equations contains a redundant equation and there is no unique solution to the set of balance equations. This redundant equation allows any given λ_j to be set to an arbitrary number > 0. As a result of such a substitution, there is a unique solution for the remaining λ's.

The marginal probabilities at each of the nodes in the open network are predicated upon the fact that there is no constraint on the number of units in the system. For the closed network, this condition is not valid, so the methodology for determining the product form solution for the closed queueing network must account for the fact that some states which would be possible in an open queueing network are no longer feasible. The product form solution for the closed network adjusts for this fact by "normalizing" the results from an open network product form solution. The closed network methodology begins by solving for the joint probability associated with each feasible state using the open network product form, denoted by $\pi(n)$. Each of these joint probabilities is then normalized into its closed network probability, denoted by $\pi_N(n)$ where N is the limit on the number of units in the system, by dividing it by the sum of the joint probabilities for all the feasible states as shown in Formula A-3.

$$\pi_N(n) = \pi(n) \,/\, \Sigma_{n \in \text{feasible states}}\ \pi(n) \qquad \text{(A-3)}$$

Conclusion

Whether representing a Markov chain or an open or closed network, these steady state probability vectors represent a valuable collection of information for the analyst. This is one of the reasons the queueing approaches to inventory management have seen a growing degree of attention.

APPENDIX B
AVAILABILITY DISTRIBUTION SIMULATION

This simulation transforms the transaction history for a given item into an availability distribution.

```
C   This program builds the serviceable balance distribution
C   for an individual NSN from that NSN's transaction file

C   Define Variables

        CHARACTER NSN*15, TRIC*3, TTPC*2, TYEAR*3
        INTEGER QTY, TDAY, TDATE
        INTEGER DISTRIB(100)

C   Open files

        open(1,file='/home/bsilver/files/nsn7757.tran')
        open(2,file='/home/bsilver/files/nsn7757.dist')


C   Initialize serviceable balance current date, backfill to
beginning date

        servbal=50
        READ(1,10,end=99)NSN,TRIC,TTPC,QTY,TYEAR,TDATE
        if(tyear.eq."093")tday=tdate-284
        if(tyear.eq."094")tday=tdate+82
        if(tyear.eq."095")tday=tdate+447

        distrib(50)=tday-1
  5     currdate=tday
```

```
      do 15, j=1,3000
        if(tday.eq.currdate)then
           if(tric.eq.'ISU')servbal=servbal-qty
           if(tric.eq.'MSI')servbal=servbal-qty
           if(tric.eq.'DOR')servbal=servbal-qty
           if(tric.eq.'SHP')servbal=servbal-qty
           if(tric.eq.'TIN')then
             oldbal=servbal
             servbal=servbal+qty
             print*,'increase TIN',oldbal,servbal
           end if
                      if(tric.eq.'REC')then
             oldbal=servbal
             servbal=servbal+qty
             print*,'increase REC',oldbal,servbal
           end if
           READ(1,10,end=99)NSN,TRIC,TTPC,QTY,TYEAR,TDATE
           if(tyear.eq."093")tday=tdate-284
           if(tyear.eq."094")tday=tdate+82
           if(tyear.eq."095")tday=tdate+447
         else
           distrib(servbal)=distrib(servbal)+1
           currdate=currdate+1
        end if

   15 continue

   99 do 35, j=1,100
        WRITE(2,20)distrib(j)
   35 continue

   10 FORMAT(A15,A3,A2,I6,A3,I3)
   20 FORMAT(I6)

      end
```

APPENDIX C
AVAILABILITY DISTRIBUTIONS

The following empirical availability distributions were derived from transaction histories at a single Air Force base over an 18-month period using an availability simulation described in Chapter 5. The top series represents the first 9 months and the bottom series the second 9 months. Leading zeros have been truncated except where necessary to match the starting points in both distributions. The final entry in each series represents full stock.

1560-01-152-6387JH	2	44	66	39	116			
	2	5	22	39	199			
1630-00-758-3758	30	26	147	39	25			
	51	71	7	63	74			
1650-00-757-3862	0	0	0	4	29	0	53	36
	22	61	21	5	7	22	7	
	2	18	30	5	0	19	12	20
	14	2	10	34	25	45	30	
1650-00-930-3160	0	0	3	13	10	12	34	47
	6	2	4	4	35	7	13	24
	0	11	12	11	11	8		
	7	41	43	4	34	3	4	14
	12	3	2	2	0	7	16	11
	29	17	4	0	6	5		
1660-00-573-6482	0	65	34	35	10	7	3	113
	3	10	101	34	9	5	75	29

1680-00-852-0803	30	41	76	0	14	46	0
	17	43					
	78	51	7	42	0	4	7
	66	11					
2840-00-066-9925RV	29	17	207	14			
	21	219	26	0			
2925-00-939-1473RV	17	14	33	35	0	14	154
	21	75	10	0	2	5	154
2995-00-759-9072	0	62	40	84	26	16	39
	1	30	3	12	6	1	214
4510-00-740-1074JH	54	10	22	14	5	2	10
	96	21	33				
	0	8	29	26	25	3	17
	26	116	0				
4810-00-573-6461TP	22	36	134	18	57		
	17	91	132	20	6		
5998-00-064-8059NT	30	4	89	14	130		
	17	35	17	0	198		
5999-00-067-3622NT	3	27	6	12	25	43	33
	58	12	43	5			
	3	119	7	48	29	34	13
	12	1	0	0			
5999-00-121-4721JH	17	27	0	6	7	48	91
	11	60	0				
	18	58	32	49	40	30	7
	14	12	6				
6105-00-960-9879	6	20	10	231			
	2	59	163	42			
6130-00-967-2610NT	20	21	89	137			
	3	80	14	170			
6605-00-955-3029JH	47	39	42	139			
	27	7	4	229			
6610-00-051-0989	36	44	11	0	51	12	0
	53	44	16				
	1	82	32	19	0	15	29
	31	7	51				

6610-00-056-7150	0 7	41 3	7 11	16	20	53	0	0
	10 52	17 36	8 22	12	61	27	13	8
6615-01-018-1635	3 6	12 6	0 33	12 28	22 29	23 5	51 37	0
	11 82	25 0	68 14	0 5	2 5	2 10	5 4	11

APPENDIX D
METRIC-BASED MODEL PARAMETERS

Stock Number	Dmd	PBR	BRCT	O&ST
1560-01-152-6387JH	35	.50	13.4	7
1630-00-758-3758	7	0	0	7
1650-00-757-3862	30	.42	7.6	7
1650-00-930-3160	107	.41	7	7
1660-00-573-6482	11	.23	1	7
1680-00-852-0803	8	0	0	10
2840-00-066-9925RV	12	0	0	7
2925-00-939-1473RV	28	.07	1	7
2995-00-759-9072	28	.13	1.3	7
4510-00-740-1074JH	56	.98	10.5	7
4810-00-573-6461TR	35	.88	14.8	7
5998-00-064-8059NT	15	0	0	7
5999-00-067-3622NT	15	0	0	7
5999-00-121-4721JH	16	.05	1	7
6105-00-960-9879	7	.30	3	11
6130-00-967-2610NT	14	.06	1	7
6605-00-955-3029JH	9	0	0	7
6610-00-051-0989	11	.07	1	7
6610-00-056-7150	48	.34	1.2	7
6615-01-018-1635	66	.97	2.3	7

APPENDIX E
MARGINAL ALLOCATION ROUTINE

This marginal allocation routine initializes its allocation levels using the constraint that each NSN must have a probability of no backorders greater than or equal to .75. It uses the probability of two or more items in stock as its objective function and features a preemptive parity check with a threshold of 0.1. The variations on this routine are described in greater detail in Chapter 6.

```
function [out]=margin

load nsntbl;

% Structure of nsntbl:
%(1-4): original traffic intensity at nodes 1-4
%  (5): original repair rate
%  (6): number of leading calibration zeros
%  (7): repair cost
%  (8): RSP authorization
%  (9): current allocation (start w/ all 0's)
% (10): current availability value
% (11): marginal utility to objective function

% Initialize allocation levels to meet minimum objective of
% P(availability >= 1) >= .75 for all NSNs
```

```
for i=1:20
  while nsntbl(i,10) < .50
    nsntbl(i,9)=nsntbl(i,9)+1;
   distrib=[zeros(1,nsntbl(i,6)),build( ...
            [nsntbl(i,1:5),nsntbl(i,9)])];
    nsntbl(i,10)=1-(sum(distrib(1:4)));
  end
end

budget=556000-(nsntbl(:,9)'*nsntbl(:,7));

% begin allocation process

allocate=[];

% Base A Selection Process

objfun=min(nsntbl(1:10,10));
maxutil=0;

for i=1:10
  distrib=[zeros(1,nsntbl(i,6)),build( ...
          [nsntbl(i,1:5),nsntbl(i,9)+1])];
  newprob=1-(sum(distrib(1:4)));
  newavail=nsntbl(1:10,10);
  newavail(i)=newprob;
  nsntbl(i,11)=(min(newavail)-objfun)/nsntbl(i,7);
  if nsntbl(i,11) > maxutil
    maxutil=nsntbl(i,11);
    bestnsn=i;
  end
end
basea=[maxutil,bestnsn];
paritya=min(newavail);

% Base B Selection Process

objfun=min(nsntbl(11:20,10));
maxutil=0;

for i=11:20
  distrib=[zeros(1,nsntbl(i,6)),build( ...
          [nsntbl(i,1:5),nsntbl(i,9)+1])];
  newprob=1-(sum(distrib(1:4)));
  newavail=nsntbl(11:20,10);
  newavail(i-10)=newprob;
  nsntbl(i,11)=(min(newavail)-objfun)/nsntbl(i,7);
  if nsntbl(i,11) > maxutil
    maxutil=nsntbl(i,11);
    bestnsn=i
  end
end
baseb=[maxutil,bestnsn];
parityb=min(newavail);
```

```
% Continuing Allocation Routine

while budget >= max(nsntbl(:,7))
  pardiff=paritya-parityb;
  parity=0;
  if (paritya-parityb) > .1
    parity=1;
  end
  if (parityb-paritya) > .1
    parity=2;
  end

  if (basea(1) >= baseb(1) & parity==0) | parity==2
    allocate=[allocate;basea(2),basea(1)*1000, ...
              baseb(1)*1000,pardiff,nsntbl(basea(2),9)];

% Update # allocated and availability

    nsntbl(basea(2),9)=nsntbl(basea(2),9)+1;
    distrib=[zeros(1,nsntbl(basea(2),6)),build( ...
            [nsntbl(basea(2),1:5),nsntbl(basea(2),9)])];
    nsntbl(basea(2),10)=1-sum(distrib(1:4));
    budget=budget-nsntbl(basea(2),7);

% Select next Base A candidate

    objfun=min(nsntbl(1:10,10));
    maxutil=0;

    for i=1:10
      distrib=[zeros(1,nsntbl(i,6)),build( ...
              [nsntbl(i,1:5),nsntbl(i,9)+1])];
      newprob=1-(sum(distrib(1:4)));
      newavail=nsntbl(1:10,10);
      newavail(i)=newprob;
      nsntbl(i,11)=(min(newavail)-objfun)/nsntbl(i,7);
      if nsntbl(i,11) > maxutil
        maxutil=nsntbl(i,11);
        bestnsn=i;
      end
    end
    basea=[maxutil,bestnsn];
    paritya=min(newavail);
  end
```

```
  if (baseb(1) > basea(1) & parity==0) | parity==1
    allocate=[allocate;baseb(2),basea(1)*1000, ...
              baseb(1)*1000,pardiff,nsntbl(baseb(2),9)];
    nsntbl(baseb(2),9)=nsntbl(baseb(2),9)+1;
    distrib=[zeros(1,nsntbl(baseb(2),6)),build( ...
            [nsntbl(baseb(2),1:5),nsntbl(baseb(2),9)])];
    nsntbl(baseb(2),10)=1-sum(distrib(1:4));
    budget=budget-nsntbl(baseb(2),7);

% Select next Base B candidate

    objfun=min(nsntbl(11:20,10));
    maxutil=0;

    for i=11:20
      distrib=[zeros(1,nsntbl(i,6)),build( ...
              [nsntbl(i,1:5),nsntbl(i,9)+1])];
      newprob=1-(sum(distrib(1:4)));
      newavail=nsntbl(11:20,10);
      newavail(i-10)=newprob;
      nsntbl(i,11)=min(newavail)-objfun;
      if nsntbl(i,11) > maxutil
        maxutil=nsntbl(i,11);
        bestnsn=i;
      end
    end
    baseb=[maxutil,bestnsn];
    parityb=min(newavail);
  end
end

% Terminal allocation of remaining budget

while budget > 0
  best=0;
  for i=1:20
    if nsntbl(i,11) > best
      if nsntbl(i,7) <= budget
        best=nsntbl(i,11);
        bestnsn=i;
      end
    end
  end
  nsntbl(bestnsn,9)=nsntbl(bestnsn,9)+1;
  budget=budget-nsntbl(bestnsn,7);
end

save nsntbl
save allocate

out=nsntbl(:,9);
```

APPENDIX F
MARGINAL ALLOCATION ITEM INDICATIVE DATA

The following table shows the indicative data on the item used in the marginal allocation routine. Chapter 6 describes the source and usage.

Item #	Node 1	Node 2	Node 3	Node 4	Initial Rate	Lead Zeros	Repair Cost	RSP Level
1	3	5	5	6	10	0	4896	2
2	3	4	4	4	4	1	1336	2
3	3	7	7	8	14	4	2277	8
4	4	8	7	8	22	3	6750	12
5	1	1	6	8	8	0	864	5
6	1	1	5	7	9	1	1371	5
7	2	3	3	5	5	1	3945	2
8	1	1	1	9	11	1	931	4
9	4	4	5	5	12	0	1922	3
10	1	1	7	8	7	4	4897	6
11	3	3	4	5	5	1	7667	2
12	4	6	2	6	8	1	3394	3
13	4	6	6	7	9	0	4982	6
14	5	5	6	7	8	3	4961	10
15	2	3	5	9	7	0	1450	2
16	3	4	5	6	8	1	677	2
17	1	1	7	4	8	1	642	2
18	3	3	3	8	9	2	2621	4
19	1	1	1	8	16	0	416	5
20	4	6	7	8	16	0	854	14

APPENDIX G
STEP-BY-STEP ALLOCATION

The following series show the step-by-step repair allocations for each of the four marginal allocation routines demonstrated in Chapter 6.

Level-Based Objective Function without Parity Constraint

Item #	New Balance
15	7
13	9
19	18
20	14
11	7
16	9
18	8
14	4
12	9
17	10
20	15
19	19
13	10
16	10
18	9
11	8
20	16
14	5
12	10
15	8
17	11
16	11
18	10
20	17
13	11
19	20
14	6

Level-Based Objective Function without Parity Constraint (Continued)

Item #	New Balance
12	11
11	9
20	18
16	12
14	7
18	11
13	12
12	12
17	12
20	19
14	8
11	10
16	12
14	9
13	13
12	13
20	20
18	12
14	10
16	14
14	11
6	10
4	7
9	15
8	10
1	12
5	9
2	6
4	8
7	8
9	16
6	11

Level-Based Objective Function with Parity Constraint

Item #	New Balance
15	7
13	9
19	18
20	14
11	7
16	9
18	8
14	4
12	9
17	10
6	10
4	7
9	15
20	15
19	18
8	11
1	12
13	10
5	9
2	6
16	10
18	9
4	8
11	8
7	8
20	16
14	5
12	10
9	16
6	11
15	8
4	9
17	11
1	13
16	11
18	10
20	17
13	11
9	17
4	10
19	20
14	6
12	11
8	12
11	9
2	7
4	11
7	9

Level-Based Objective Function with Parity Constraint (Continued)

Item #	New Balance
20	18
16	12
6	12
9	18
14	7
1	14
18	11
5	10
4	12
13	12

RSP-Based Objective Function without Parity Constraint

Item #	New Balance
9	16
3	8
18	7
11	5
20	19
12	9
15	7
17	8
14	9
16	7
13	11
18	8
19	20
11	6
17	9
12	10
16	8
18	9
4	21
6	12

RSP-Based Objective Function with Parity Constraint

Item #	New Balance
9	16
3	8
18	7
11	5
20	19
12	9
15	7
17	8
14	9
16	7
13	11
18	8
19	20
11	6
17	9
12	10
16	8
18	9
14	10

BIBLIOGRAPHY

Ahmed, Mohamed A., Donald Gross, Douglas R. Miller, "Control Variate Models for Estimating Transient Performance Measures in Repairable Item Systems," *Management Science*, **38**, 388-399, (1992).

Allen, Stephen G., Donato A. D'Esopo, "An Ordering Policy for Repairable Stock Items," *Operations Research*, **16**, 669-675, (1968).

Arrow, K.J., S. Karlin, H. Scarf, *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford, CA, 304-307, (1958).

Axsater, Sven, "Modelling Emergency Lateral Transshipments in Inventory Systems," *Management Science*, **36:11**, 1329-1338, (1990).

Balana, Arturo R., Donald Gross, Richard M. Soland, "Optimal Provisioning for Single-Echelon Repairable Item Inventory Control in a Time-Varying Environment," *IIE Transactions*, **21**, 202-212, (1989).

Buyukkurt, Demirbag M. and M. Parlar, "A Comparison of Allocation Policies in a Two-Echelon Repairable-Item Inventory Model," *International Journal of Production Economics*, **29:3**, 291-302, (1993).

Carrillo, Manuel J. "Extensions of Palm's Theorem: A Review," *Management Science*, **37**, 739-744, (1991).

Carrillo, Manuel J. "Generalizations of Palm's Theorem and Dyna-METRIC's Demand and Pipeline Variability," R-3698-AF, The Rand Corp., Santa Monica, CA (1989).

Clark, Andrew J. "An Informal Survey of Multi-Echelon Inventory Theory," Naval Research Logistics Quarterly, **19:4**, 621-649, (1972).

Clark, Andrew J., Herbert Scarf, "Optimal Policies for a Multi-Echelon Inventory Problem," *Management Science*, **6**, 475-490 (1960).

Cohen, Morris A., Paul Kleindorfer, Hau Lee, "Optimal Stocking Policies for Low Usage Items in Multi-Echelon Inventory Systems," *Naval Research Logistics Quarterly*, **33**, 17-38, (1986).

Cohen, Morris A., P.V. Kamesam, P. Kleindorfer, H. Lee, A. Tekerian, "Optimizer: IBM's Multi-Echelon Inventory System for Managing Service Logistics," *Interfaces* **20:1**, 65-82, (1990).

Crawford, Gordon B., "Variability in the Demands for Aircraft Spare Parts: Its Magnitude and Implications," R-3318-AF, The Rand Corp., Santa Monica, CA (1988).

Dada, Maqbool, "ATwo-Echelon Inventory System with Priority Shipments," *Management Science*, **38:8**, 1140-1153, (1992).

Demmy, W. Steven, Victor J. Presutti, "Multi-Echelon Inventory Theory in the Air Force Logistics Command," *TIMS Studies in the Management Sciences*, **16**, 279-297, (1981).

Dhakar, T.S., C.P. Schmidt, D.M. Miller, "Base Stock Level Determination for High Cost Low Demand Critical Repairable Spares," *Computers in Operations Research*, **21:4**, 411-420, (1994).

Ebeling, Charles E., "Optimal Stock Levels and Service Channel Allocations in a Multi-Item Repairable Asset Inventory System," *IIE Transactions*, **23:2**, 115-120, (1991).

Feeney, G.J., C.C. Sherbrooke, "The (S-1,S) Inventory Policy Under Compound Poisson Demand," *Management Science*, **12:5**, 391-411, (1966).

Fox, B.L., D.M. Landi, "Searching for the Multiplier in One Constraint Optimization Problems," *Operations Research*, **18:2**, 253-262, (1970).

Gordon, W. J., G. F. Newell, "Closed Queueing Systems with Exponential Servers," *Operations Research*, **15**, 254-265, (1967).

Graves, Stephen C., "A Multi-Echelon Inventory Model for a Repairable Item with One-for-One Replenishment," *Management Science*, **31:10**, 1247-1256, (1985).

Gross, Donald, Carl M. Harris, *Fundamentals of Queueing Theory*, John Wiley and Sons, New York, (1974).

Gross, Donald, "On the Ample Service Assumption of Palm's Theorem in Inventory Modeling," *Management Science*, **28:9**, 1065-1079, (1982).

Gross, Donald, Douglas R. Miller, R. M. Soland, “A Closed Queueing Network Model for Multi-Echelon Repairable Item Provisioning,” *IIE Transactions*, **21**, 202-212, (1989).

Gross, Donald, Leonidas C. Kioussis, Douglas R. Miller, “A Network Decomposition Approach for Approximating the Steady-State Behavior of Markovian Multi-Echelon Reparable Item Inventory Systems,” *Management Science*, **33**, 1453-1468, (1987).

Gross, Donald, B. Gu, R.M. Soland, “Iterative Solution Methods for Obtaining the Steady-State Probability Distributions of Markovian Multi-Echelon Repairable Item Inventory Systems,” *Computers in Operations Research*, **20:8**, 817-828, (1993).

Hausman, W.H., N.K. Erkip, “Multi-Echelon vs. Single-Echelon Inventory Control Policies for Low-Demand Items,” *Management Science*, **40:5**, 597-602, (1994).

Hillestad, R.J., M.J. Carrillo, “Models and Techniques for Recoverable Item Stockage When Demand and the Repair Process are Nonstationary--Part I: Performance Measurement, The RAND Corporation, Santa Monica, CA, (1980).

Hillestad, R.J., “Dyna-METRIC: Dynamic Multi-Echelon Technique for Recoverable Item Control,” R-2785-AF, The RAND Corporation, Santa Monica, CA, (1982).

Isaacson, Karen E., P. Boren, C.L. Tsai, R. Pyles, “Dyna-METRIC Version 4: Modeling Worldwide Logistics Support of Aircraft Components,” R-3389-AF, The RAND Corporation, Santa Monica, CA, (1988).

Isaacson, Karen E., Patricia Boren, “Dyna-METRIC Version 5: A Capability Assessment Model Including Constrained Repair and Management Adaptations,” R-3612-AF, The RAND Corporation, Santa Monica, CA, (1988).

Jackson, J. R., “Networks of Waiting Lines,” *Operations Research*, **5**, 518-521, (1957).

Jackson, J.R., “Jobshop-Like Queueing Systems,” *Management Science*, **10:1**, 131-142, (1963).

Kao, E. P. C., *An Introduction to Stochastic Processes*, Duxbury Press, Belmont, CA, (1997).

Lee, Hau L., “A Multi-Echelon Inventory Model for Repairable Items With Emergency Lateral Transshipments,” *Management Science*, **33:10**, 1302-1316, (1987).

Little, J.D.C., "A Proof for the Queueing Formula: $L=\lambda W$," *Operations Research*, **9:3**, 383-387, (1961).

Miller, Bruce L., "A Real Time Metric for the Distribution of Serviceable Assets," RM-5687-PR, The RAND Corporation, Santa Monica, CA, (1968).

Miller, Bruce, L., "Dispatching from Depot Repair in a Recoverable Item Inventory System: On the Optimality of a Heuristic Rule," *Management Science*, **21:3**, 316-325, (1974).

Mirasol, Noel M., "A Queuing Approach to Logistics Systems," *Operations Research*, **12:5**, 707-724, (1964).

Muckstadt, John A., "A Model for a Multi-Item, Multi-Echelon, Multi-Indenture Inventory System," *Management Science*, **20:4**, 472-481, (1973).

Muckstadt, John A., "Some Approximations in Multi-Item, Multi-Echelon Inventory Systems for Recoverable Items," *Naval Research Logistics Quarterly*, **25**, 377-394, (1978).

Muckstadt, John A., L.J. Thomas, "Are Multi-Echelon Inventory Methods Worth Implementing in Systems with Low-Demand-Rate Items?," *Management Science*, **26:5**, 483-494, (1980).

Nahmias, Steven, "Managing Reparable Item Inventory Systems: A Review," *TIMS Studies in the Management Sciences*, **16**, 253-277, (1981).

Netter, John, William Wasserman, G.A. Whitmore, *Applied Statistics*, Allyn and Bacon, Boston, (1993).

Neuts, Marcel F., *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Johns Hopkins University Press, Baltimore, (1981).

Neuts, Marcel F., *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, Marcel Dekker, Inc., New York, (1989).

Pfaffenberger, Roger C., James H. Patterson, *Statistical Methods for Business and Economics*, Richard D. Irwin, Inc., Homewood, IL, (1977).

Pyles, R.A., "The Dyna-METRIC Readiness Assessment Model: Motivation, Capabilities and Use," R-2286-AF, The RAND Corporation, Santa Monica, CA, (1984).

Sherbrooke, Craig C. "METRIC: A Multi-Echelon Technique for Recoverable Item Control," *Operations Research*, **16**, 122-141 (1968).

Sherbrooke, Craig C., "Estimation of the Variance-to-Mean Ratio for AFLC Recoverable Items," AFLC/MMMA Final Report, Sherbrooke & Associates, Potomac, MD, (1984).

Sherbrooke, Craig C., "Vari-Metric: Improved Approximations for Multi-Indenture, Multi-Echelon Availability Models," *Operations Research*, **34:2**, 311-319 (1986).

Sherbrooke, Craig C., *Optimal Inventory Modeling of Systems, Multi-Echelon Techniques*, John Wiley & Sons, New York, (1992).

Silver, B. D., W. B. Faulkner, C. E. Miller, G. F. Jarnagin, P. McKenna, "Reduction of Recoverable Pipeline," AFLMC Final Report LS902029, Gunter AFB, AL, (1991).

Silver, B. D., B. J. Allen, J. Dietz, W. B. Faulkner, T. Gage, G. F. Jarnagin, P. McKenna, M. Dubois, A. Tallent, "Analysis of Retail-Wholesale Data Interfaces," AFLMA Report LS912083, Gunter AFB, AL, (1992).

Silver, B. D., W. B. Faulkner, T. H. Ondracek, W. P. Quinones, E. J. Scarpa, L. T. Whitsel, "Analysis of the Depot Repair Process," AFLMA Report LS922128, Gunter AFB, AL, (1993).

Silver, Edward A., "Inventory Allocation Among an Assembly and its Repairable Subassemblies," *Naval Research Logistics Quarterly*, **19:2**, 261-280, (1972).

Simon, R. M. and D. A. D'Esopo, "Comments on a Paper by S.G. Allen and D.A. D'Esopo: 'An Ordering Policy for Reparable Stock Items' ", *Operations Research*, **19:4**, 986-989, (1971).

Stevens, R.J., J.M. Hill, "Estimating the Variance-to-Mean Ratio for Recoverable Items in the ALS Marginal Analaysis Algorithms," Working Paper Number 49, Office of DCS/Comptroller, Wright-Patterson AFB, OH, (1973).

Svoronos, Antony, "A General Framework for Multi-Echelon Inventory and Production Control Problems," Doctoral Dissertation, Graduate School of Business, Columbia University, New York, (1986).

Svoronos, Antony and Paul Zipkin, "Estimating the Performance of Multi-Level Inventory Systems," *Operations Research*, **36:1**, 57-72, (1988).

Svoronos, Antony and Paul Zipkin, "Evaluation of One-for-One Replenishment Policies for Multiechelon Inventory Systems," *Management Science*, **37:1**, 68-83, (1991).

Zipkin, Paul, "The Use of Phase-Type Distributions in Inventory-Control Models," *Naval Research Logistics*, **35**, 247-257, (1988).