LOCATION AND INVENTORY OPTIMIZATION FOR SERVICE CENTERS

by

Angela Lan Chiu

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1993

Author ....
Department of Electrical Engineering and Computer Science
May 19, 1993

Certified by ....
Dimitris Bertsimas
Associate Professor
Thesis Supervisor

Certified by ....
John Tsitsiklis
Professor
Thesis Supervisor

Accepted by ....
Campbell L. Searle
Chair, Departmental Committee on Graduate Students

LOCATION AND INVENTORY OPTIMIZATION FOR SERVICE CENTERS

by

Angela Lan Chiu

Submitted to the Department of Electrical Engineering and Computer Science
on May 19, 1993, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering

Abstract

In this thesis, we model the location and inventory optimization problem for service centers in a real world environment. Large number of repair parts are stored and barely moved, referred to as slow-moving parts. Our objective is to find the optimal inventory level and locations for all the repair parts as well as the optimal locations for service centers, in order to minimize the total cost, which includes the inventory cost for all the parts in stock and the setup cost for service centers. Meantime, certain performance criterion (average customer satisfaction rate) has to be satisfied. Various models and algorithms are proposed to provide both insights of the problem and reasonably good solutions. Numerical results have shown impact of possible cost savings for service centers.

Thesis Supervisor: Dimitris Bertsimas
Title: Associate Professor

Thesis Supervisor: John Tsitsiklis
Title: Professor

Acknowledgment

I would like to express my deepest appreciation to both my research advisors Prof. Dimitris Bertsimas and Prof. John Tsitsiklis for their continual guidance, ideas, and encouragement. Working and discussing with them is a enjoyful educational experience for me.

I would also like to thank my graduate counsellor Prof. Alvin Drake for his advice and encouragement.

Special thanks to Bruce Arntzen, Gary Finckel, and Peter Metz at DEC who we worked with throughout the year. Their advice and information, especially the data (in Appendix A) Gary provided made this research more valuable to service centers in a real world environment.

I would like to thank Barry Kostiner and Sarah Stock for their help on GAMS.

Lots of thanks to all the members of LIDS for their help and for the nice environment they created for me to work in.

On a personal note, I would like to thank my friends: Gina, Ioannis, Leyuan, and Mei for being my "family" here, both inside and outside of school.

Last, but definitely not least, I wish to express my love to my parents Yao-Cheng and Pik-Siu. Their sacrifice made the journey to this country possible, where I have receive such excellent education. I owe everything to them and nothing can describe my feelings. My love to "my little sister" Wei for being the closest to me and sharing our life together throughout all these years. My love to "my husband-to-be" Zhuang-Bo for always being there for me throughout the good time and bad time. His love and support has made these two years at MIT as memorable as it could be.

This research is supported by the Leaders for Manufacturing Program (LFM) at MIT.

Contents

1 Introduction **8**

1.1 Motivation 8

1.2 Problem Definition 9

1.3 Review of Existing Research 12

1.4 Contribution and Organization of the Thesis 13

2 An Exact Analysis for a Class of Simple Models **15**

2.1 One-Part, One-Stockroom Model 16

2.1.1 Problem Definition 16

2.1.2 Analysis of the Model 16

2.1.3 Numerical Results 20

2.2 One-Part, Two-Stockroom Model 22

2.2.1 Problem Definition 22

2.2.2 Analysis of the Model 24

2.2.3 Assignment Policy 27

2.2.4 Numerical Results 28

3 An Optimization Approach to the One-Part, Multi-Stockroom Problem **32**

3.1 Inventory Location Problem 33

3.1.1 Problem Definition 33

3.1.2 Algorithm 33

3.1.3 A Covering Algorithm for the LOCATION Problem 34

3.1.4 A Greedy Algorithm for the LOCATION Problem 35
3.1.5 Adaptive Partitioned Random Search (APRS) Algorithm to the LOCATION Problem 36
3.1.6 Numerical Results 42
3.2 Assignment Policy Problem 47
3.2.1 Problem Formulation 47
3.2.2 Numerical Results 49

4 An Optimization Approach to the Multi-part, Multi-stockroom Problem 50
4.1 With Fixed Stockrooms 51
4.2 With Potential Stockrooms 51
4.2.1 Algorithm 52
4.2.2 Numerical Results 53

5 Conclusions 56

A DEC Virginia Service Region Data 58

B Reconstruction of DEC Virginia Service Region Data 61
B.1 Construction of Customers' Geographic Location 61
B.2 Estimation on Customers' Demands for the Given Parts of VAX 6000 62

C Approximate Closed-Form Formula for Average Customer Satisfaction Rate 64

List of Figures

2-1 α vs N with t_r = 24 hours, V = 1, 2, p = 0.00057 20

2-2 α vs N with t_r = 24 hours, V = 1, 2, p = 0.004 21

2-3 α vs N with t_r = 48 hours, V = 1, 2, p = 0.00057 22

2-4 α vs N with t_r = 24 hours, V = 1, 2, p = 0.004 22

2-5 Two stockrooms with intersected service regions 23

2-6 α vs N_1 with t_r = 24 hours, $N_{12} = N_2 = 50$ 28

2-7 α vs N_{12} with t_r = 24 hours, $N_1 = 100$, $N_2 = 50$ 29

2-8 α vs N_1 with t_r = 48 hours, $N_{12} = N_2 = 25$ 29

2-9 α vs N_{12} with t_r = 48 hours, $N_1 = 50$, $N_2 = 25$ 30

3-1 An example: DEC Virginia service district 42

3-2 A scaled map of DEC Virginia service district (1:10miles) 43

List of Tables

2.1 The effect of assignment policy to satisfaction rate with $t_r = 48$ hours, $N_1 = 100$, $N_{12} = N_2 = 50$ 30

3.1 Numerical results from enumeration with five current stockrooms and their 2-hour service regions 44

3.2 Numerical results from the covering algorithm with 140 potential stockrooms and their 2-hour service regions 44

3.3 Numerical results from the greedy algorithm with 140 potential stockrooms and their 2-hour service regions 45

3.4 Numerical results from the APRS algorithm with 140 potential stockrooms and their 2-hour service regions 45

3.5 Numerical results from enumeration with five current stockrooms and their 4-hour service regions 46

3.6 Numerical results from the covering algorithm with 140 potential stockrooms and their 4-hour service regions 46

3.7 Numerical results from the greedy algorithm with 140 potential stockrooms and their 4-hour service regions 46

3.8 Numerical results from the APRS algorithm with 140 potential stockrooms and their 4-hour service regions 46

3.9 The effect of assignment policy to satisfaction rate for multi-stockroom 49

Chapter 1

Introduction

1.1 Motivation

In today's world, new high-tech products are invented and stay in the limelight for a few years, and then are replaced by new models. Their repair parts, however, need to be around for decades. Furthermore, most such products are covered by service contracts which guarantee that the manufacturer will provide service within a few hours (typically 2 or 4 hours) after receiving a service call (although the frequency of calls is usually very low).

The results of these observations are that

- the number of repair parts (compared to active products) is very large; for example, a DEC workstation can have thousands of different parts in it,
- local service centers are required to stock a large number of parts in order to meet their service commitments.

These may lead to high inventories and low turns of usage, which again may cause a poor return on assets.

If the repair parts are very expensive, and have very low frequency of failure referred as slow-moving parts, it is evident that not every service center in a given district should have every part in stock. This raises the main question of this research, which is how many parts should be stocked and at which locations such that certain performance criterion can still be satisfied. These questions were originally raised from the logistic department of Digital Equipment Cooperation (DEC) in the U.S., and the research is supported by DEC.

The main objective of this research is to address the following issues:

1. Globally, how many inventories the whole district needs for each part, and where to locate the service centers if we have the freedom to choose from a set of potential locations?

2. At each service center, how many inventories to store for each repair part?

3. Administratively, how to assign each customer with a primary/secondary service center?

Our goal is to minimize the total cost which includes the setup cost for service centers, and the inventory cost for all the parts in stock. Meanwhile, we have to guarantee certain level of service performance, characterized by the average customer satisfaction rate.

A formal definition of the problem is stated in the next section.

1.2 Problem Definition

Consider a service district of interest. There are N customers who have signed service contracts with local service centers, where S different repair parts are stored. Customers' geographical locations are given and their demands for parts are all unit-sized, i.e., each demand requests for only one item of particular part. We make the following assumptions in order to model this problem.

1. The availability of a service person is ignored in our models. With slow-moving parts, the daily demands are usually very low. Therefore, the main issue here is the availability of the parts in stock rather than the availability of the service person. As a result, when we say that a customer's request is satisfied, it means that when the customer calls for service, there is a corresponding part in

stock and the part can be delivered to the customer within the service timing constraint t_s (2 or 4 hours). Therefore, we can define the service region of a service center as the area that can be reached from the center within certain t_s. In other words, the service timing constraint t_s determines the service region of each service center.

2. Each customer independently has a probability p_i of requesting for a part i ($i = 1, 2, ..., S$) at each day. In the case where several different parts are grouped together as a "kit", we can view each kit as one "part", so that our assumption is still valid.

Additional data provided in the model are the following,

1. Q_R: the set of all R stockroom locations that are given, $Q_R = \{1, 2, ..., R\}$; it can be either the set of existing stockrooms that are fixed, or the set of potential locations that we have the freedom to choose from,
2. C_{SR}: setup cost of a stockroom,
3. C_i: unit inventory cost of part i ($i = 1, 2, ..., S$),
4. α^*: the minimum satisfaction rate that has to be guaranteed for all the parts $i = 1, 2, ..., S$,
5. t_r: the replenishment time of stockrooms, t_r = 24 hours, or 48 hours, which means that the requests of parts at any given day will be replenished at the beginning of the next day or two days after.

The decision variables are defined as follows:

1. $Q_{SR} \subseteq Q_R$: the set of stockrooms selected from the R potential locations, $N_{SR} = |Q_{SR}|$; in the case where the R stockrooms are fixed, $Q_{SR} = Q_R$,
2. V_i: inventory level (number of items) of part i,

3. x_{kj}^i: 1 if the kth ($k = 1, ..., V_i$) inventory of part i is at the jth ($j \in Q_{SR}$) stockroom, 0 otherwise; $\underline{x}^i$ is the inventory location vector of part i with components x_{kj}^i,

4. π: an assignment policy that assigns each customer with a primary/secondary stockroom (i.e., a customer will call his/her primary stockroom first if a service request occurs); Δ is the set of all possible assignment policies.

The service performance criterion is characterized by the average satisfaction rate of part i ($i = 1, 2, ..., S$) denoted as α_i^π with an assignment policy π, which is averaged over all the customers $j = 1, 2, ..., N$. Let $\alpha_i^\pi(j)$ be the corresponding satisfaction rate of customer j for part i with policy π, i.e., it is defined as the probability that when customer j requests part i, the request is satisfied. Therefore, we have

$$\alpha_i^\pi = \frac{\sum_{j=1}^N \alpha_i^\pi(j)}{N}.$$

Since α_i^π is a function of the inventory locations $\underline{x}^i$ of part i, we also write it as $\alpha_i^\pi(\underline{x}^i)$.

With all the variables defined above, the problem can be formulated as

$$\min_{Q_{SR} \subseteq Q_R,\, V_i,\, \underline{x}^i,\, \pi \in \Delta} \quad N_{SR} C_{SR} + \sum_{i=1}^S V_i C_i$$

$$\text{s.t.} \quad \sum_{j \in Q_{SR}} x_{kj}^i = 1, \ \text{for all } k = 1, ..., V_i, \ i = 1, 2, ..., S$$

$$\alpha_i^\pi(\underline{x}^i) \geq \alpha^* \text{ for all } i = 1, 2, ..., S$$

The first group of constraints implies that each inventory has to be at exactly one of the selected stockrooms.

The fact that $\alpha_i^\pi(\underline{x}^i)$ is a function of both π and $\underline{x}^i$ makes the problem very difficult to solve. Therefore, in this research the problem is decomposed into two subproblems. First, we solve the problem above with a fixed π. Then with the decision variables found in the first subproblem, an optimal assignment policy is obtained which maximizes the corresponding satisfaction rate α_i^π.

1.3 Review of Existing Research

In the study of slow-moving parts, existing models are usually dealing with one stockroom at a time. The main issues considered are the inventory level, replenishment strategy (when to order and how many), and demand forecasting.

For example, Croston [Cros] discussed whether or not to stock a part (as well as the optimal inventory level); this is a major problem with slow-moving parts. The model Croston considered is a stockroom which is replenished at fixed review intervals to maintain the inventory of a part at some level R. Due to the slow-moving property, the average interarrival interval between demands is assumed to be much larger than the fixed replenishment interval. The goal of the system is to find the optimal inventory level R (possibly 0) so that the total cost is minimized. There are several costs considered (part unit cost, storage cost, replenishment cost, extra special order cost and customer dissatisfaction cost if the inventory available is insufficient to meet customers' demands). The analysis was done for the case of normally distributed demand quantities occurring with a fixed probability in each review interval, and zero replenishment lead times.

For the issue of replenishment strategy, Schultz's paper [Schu] introduces the concept of delayed replenishment orders for expensive parts subject to intermittent, unit-sized demands. Two commonly used, continuous review, inventory control systems for single parts under probabilistic demand are the order-point, order-quantity (s, Q) system and the order-point, order-up-to (s,S) system. Traditionally, replenishment orders are released as soon as the inventory position drops to s or lower. In an environment where lead times are small relative to the average time between demands, and holding costs are large relative to other inventory costs, it may be more economical to delay ordering to achieve holding cost saving with little additional risk or cost of shortage. The (0,1) policy with a replenishment delay d is used for slow-moving parts since the decision is usually whether or not to stock a part. Note that not stocking an part at all is equivalent to an infinite replenishment delay d. The inventory system modeled is one with complete backlogging, constant lead times, arbitrary inter-demand distributions, linear carrying costs, and emergency handling of shortages at a premium cost. Results are given for the optimal delay for inter-demand distributions with increasing failure rates.

A frequent problem with slow-moving parts is that there is little or no historical data for demand forecasting. For an excellent overall literature review and analysis regarding this issue, see Williams [Will].

The main difficulty of our problem which is also the main difference from other existing research models is the constraint on average customer satisfaction rate α^i for a particular part i. In most of the existing models, there is no such constraint on α^i, instead a dissatisfied demand results in a dissatisfaction cost. Moreover, in order to evaluate α^i in our problem, we have to take all the stockrooms with their corresponding service regions into account. In contract, conventional inventory control models study only one stockroom at a time, independent of others.

1.4 Contribution and Organization of the Thesis

The main contribution of this research is that we model the location and inventory optimization problems for service centers in a real world environment, especially for multi-stockroom case with service timing constraints. Various approaches and algorithms are developed, which provide good solutions with possible cost saving for service centers. We believe that our models and algorithms can be used in a real world environment.

The rest of the thesis is organized as follows. As a preparation for more complex models, we first study in Chapter 2 the problem with both the one-part, one-stockroom model, and the one-part, two-stockroom model. These are the cases where satisfaction rate α can be evaluated exactly and various effects of the main parameters to the average customer satisfaction rate can be examined. Then, in Chapter 3 we study a more complicated model with one-part, R-stockroom, where the simplest one-part, one-stockroom model in Chapter 2 can be used to obtain a lower bound of the minimum inventory level and various algorithm are developed to proceed after that. Moreover, the optimal assignment policy can be formulated as a combinatorial optimization problem, which can be solved using a software package called GAMS/MINOS. Finally, in Chapter 4 we study the core problem of the thesis which is the S-part, R-stockroom model, where the one-part, R-stockroom model in Chapter 3 can be used to find the lower bounds of the minimum inventory levels for every part individually. Beginning with the lower bounds, an algorithm is developed to find the minimum inventory levels and their locations for all the parts. Conclusions of the thesis are summarized in Chapter 5.

Chapter 2

An Exact Analysis for a Class of Simple Models

In this chapter, we study a class of simple models where customer satisfaction rate can be written in a closed-form. The goal of the chapter is to examine the effects to the average satisfaction rate of various problem parameters: inventory level, number of customers, probability of customer's request for service and replenishment time. The locations of stockrooms are assumed to be fixed (not subject to optimization). Therefore, the problem can be decomposed into a one-part problem for each repair part.

Given the location of a stockroom, the service region of the stockroom can be determined according to the timing constraint of service. For example, if the timing constraint of service is 4 hours, the service region of a given stockroom would be the area that can be reached from the stockroom within 4 hours. In the case when the locations of stockrooms are far enough so that their service regions do not overlap with each other, each stockroom can be studied independently which forms the one-part, one-stockroom model. On the other hand, when the service regions of two stockrooms intersect with each other, customers in different subregions are related to each other in terms of their satisfaction rates. Therefore, we need to consider both stockrooms at the same time in order to evaluate the average satisfaction rate over all the customers. This becomes the one-part, two-stockroom model. However, as the number of stockrooms whose service regions overlap with each other is larger than two, it becomes very difficult to write down the exact closed-form formula for the customer satisfaction rate. In this case, approximate formulae and simulation will be used in the next chapter.

The rest of this chapter studies two simple models where exact analysis can be obtained. The first one is the simplest *one-part, one-stockroom model.* The second one is the *one-part, two-stockroom model* where the service regions of the two stockrooms intersect with each other. Since only one part is considered in the models, the index i for part i is dropped in our analysis.

2.1 One-Part, One-Stockroom Model

In this section, we study the simplest case where only one repair part and one stockroom are considered. The effects of the main parameters of the model are examined as shown below.

2.1.1 Problem Definition

Given a single repair part and the location of a stockroom and its service region determined by service timing constraint (typically 2 or 4 hours), the parameters of the model are:

1. N: number of customers within the service region,
2. p: probability of a customer requesting a part at any given day,
3. t_r: replenishment time of the stockroom (typically 24 or 48 hours),
4. V: inventory level of the part (typically 1 or 2 because of the low p).

2.1.2 Analysis of the Model

The satisfaction rate of customer i (i = 1, 2, ..., N), denoted as $\alpha(i)$, is defined as the probability that when customer i requests for a part at any given day the request is satisfied. Here, there is only one possible assignment policy, which is assigning all the customers with the given stockroom as their primary ones. Thus, the variable π in the satisfaction rate is dropped for this model. Since all the N customers are identical in terms of both demand pattern and service pattern (they are all served by the same stockroom), $\alpha(i)$ is the same for all customers $i = 1, 2, ..., N$, thus the same as the average customer satisfaction rate α. Therefore, we can obtain α by finding the satisfaction rate for a particular customer, e.g., customer 1.

In order to obtain the closed-form formula for the satisfaction rate α, we need to define some variables and functions in the model first.

At any given day, customer 1 has a service request, call it request A. During the same day that request A comes, the other $N-1$ customers' requests for services are i.i.d. Bernoulli random variables with parameter p independent of A. Let X be the number of requests from customers 2, ..., N during the same day that request A comes, X is a binomial random variable with parameter $N-1$ and p. Similarly, let X' be the number of requests from all the N customers during the day before request A comes, X' is also a binomial random variable with parameter N and p.

Now, we can define some of the functions used in the analysis.

- $b(n,k)$: probability that there are k requests from n potential customers, and

$$b(n,k) = Binomial(n,k) = p^k(1-p)^{n-k}\frac{n!}{(n-k)!k!},$$

- $f_X(k)$: probability mass function of random variable X, and

$$f_X(k) = b(N-1,k),$$

- $f_{X'}(k)$: probability mass function of random variable X', and

$$f_{X'}(k) = b(N,k),$$

- $a_1(n)$: probability that request A is not the first one among the requests from other n potential customers, note that request A is not from one of the n customers, then

$$a_1(n) = \sum_{k=1}^{n} b(n,k)\frac{k}{k+1},$$

- $a_2(n)$: probability that request A is the first one among the requests from other n potential customers, then

$$a_2(n) = \sum_{k=0}^{n} b(n,k)\frac{1}{k+1},$$

- $a_3(n)$: probability that request A is not one of the first two requests among all the requests from other n potential customers, then

$$a_3(n) = \sum_{k=2}^{n} b(n,k)\frac{k-1}{k+1},$$

- $a_4(n)$: probability that request A is the second one among the requests from other n potential customers, then

$$a_4(n) = \sum_{k=1}^{n} b(n,k)\frac{1}{k+1}.$$

where

$$a_1(n) = 1 - a_2(n),$$

$$a_3(n) = a_1(n) - a_2(n) + b(n,0) = 1 - 2a_2(n) + b(n,0),$$

$$a_4(n) = a_2(n) - b(n,0).$$

Next, customer satisfaction rate α can be written as a function of the number of customers (N) and the probability of request (p), given the replenishment time (t_r) and certain inventory level (V). Since $\alpha = 1 - q$, where q is the probability where a particular request A comes it is not satisfied, it is equivalent to writing q as a function of N and p given t_r and V, denoted by $q(t_r, V)$ as shown below.

t_r = 24 hours, V = 1

With replenishment time being 24 hours and one inventory, q is the probability that request A is not the first one among the requests from other $N-1$ potential customers. Therefore

$$q(24,1) = a_1(N-1).$$

t_r = 24 hours, V = 2

With replenishment time being 24 hours and two inventories, q is the probability that request A is not one of the two requests among all the requests from other $N-1$ potential customers. Therefore

$$q(24,2) = a_3(N-1).$$

t_r = 48 hours, V = 1

With replenishment time being 48 hours and one inventory, q depends on not only the other requests during the same day that request A comes, but also the requests during the day before. Therefore,

$$\begin{aligned} q(48,1) &= \text{prob}(X' \geq 1) + \text{prob}(X' = 0)q(24,1) \\ &= 1 - b(N,0) + b(N-1,0)q(24,1). \end{aligned}$$

t_r = 48 hours, V = 2

With replenishment time being 48 hours and two inventories, similarly q depends on not only the other requests during the same day that request A comes, but also the requests during the day before. Therefore,

$$\begin{aligned} q(48,2) &= \text{prob}(X' \geq 2) + \text{prob}(X' \geq 1)q(24,1) + \text{prob}(X' = 0)q(24,2) \\ &= 1 - b(N,1) - b(N,0) + b(N-1,1)q(24,1) + b(N-1,0)q(24,2). \end{aligned}$$

2.1.3 Numerical Results

Based on the closed-form formulae, we can obtain the following numerical results under various conditions.

t_r = 24 hours, V = 1, 2

With 24 hours replenishment time, satisfaction rate α can be plotted as a function of number of customers N for both $V = 1$ and 2 on the same graph given customer's probability of request p. The following two plots are obtained using two different values of p. Meantime, we can use the plot to find out the minimum inventory level in order to serve certain number of customers given the value of satisfaction rate α^* that has to be guaranteed.

$p = 0.00057$ (=1/250/7), average one request every 7 years The corresponding part we consider here is T201500 of VAX 6000, which is the one with relatively high demands among VAX 6000 group. Here, we assume there are 250 working days per year.

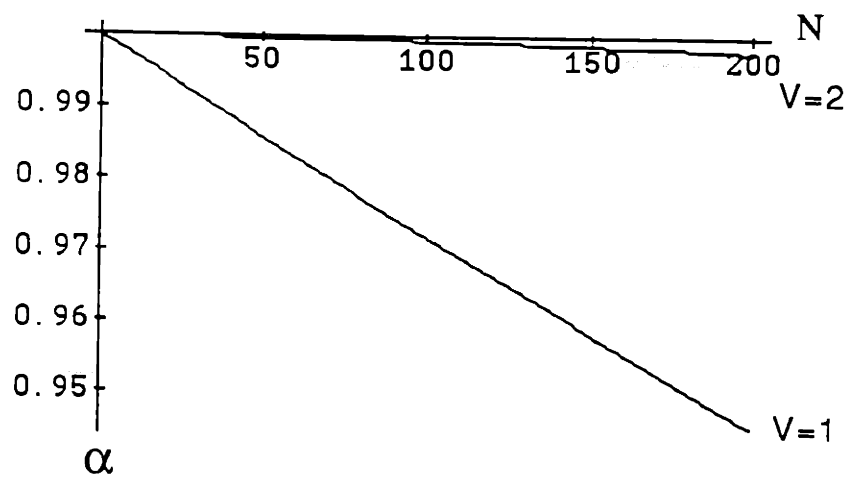

Figure 2-1: α vs N with t_r = 24 hours, V = 1, 2, p = 0.00057

Therefore, if the required $\alpha^* = 95\%$ (which is typical value for 4-hour service), we can see from Figure 2-1 only one inventory is needed to serve 150 customers.

p = 0.004 (=1/250), average one request per year Here, we just take an arbitrary value of p which is higher than the previous one.

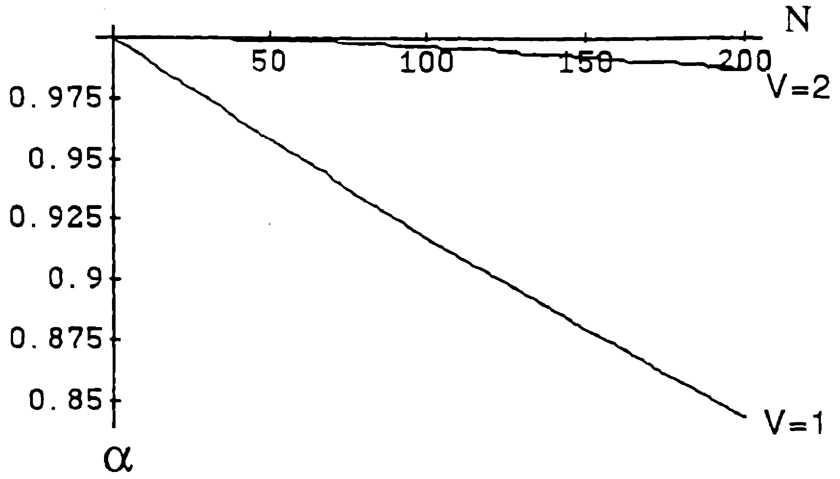

Figure 2-2: α vs N with t_r = 24 hours, V = 1, 2, p = 0.004

With the given value of p, if the required α^* is also 95%, then two inventories are needed to serve 150 customers based on Figure 2-2.

t_r = 48 hours, V = 1, 2

With 48 hours replenishment time, the value of satisfaction rate would be lower than the one with 24 hours replenishment time under the same conditions, as shown below.

p = 0.00057 (=1/250/7), average one request every 7 years Part T201500 of VAX 6000 is considered here.

From Figure 2-3, if the required α^* is also 95%, then two inventories are needed to serve 150 customers, which is higher than the one with 24 hours replenishment time.

p = 0.004 (=1/250), average one request per year If the required α^* is again 95%, then two inventories can only serve 50 customers at most based on Figure 2-4.

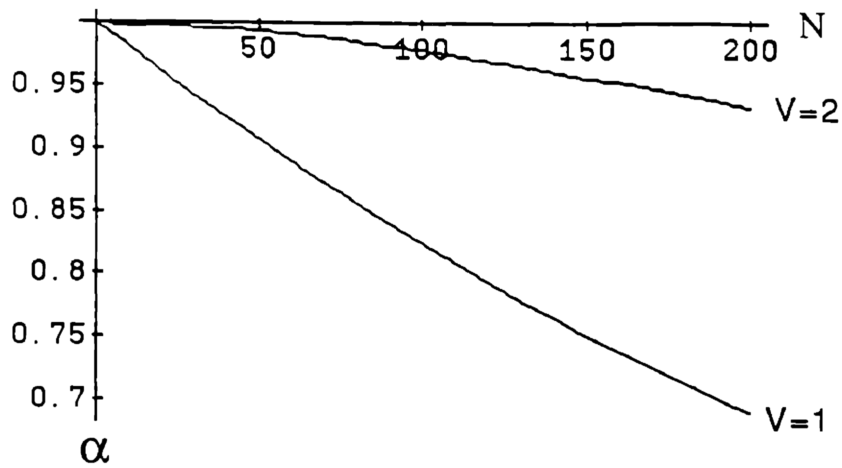

Figure 2-3: α vs N with t_r = 48 hours, V = 1, 2, p = 0.00057

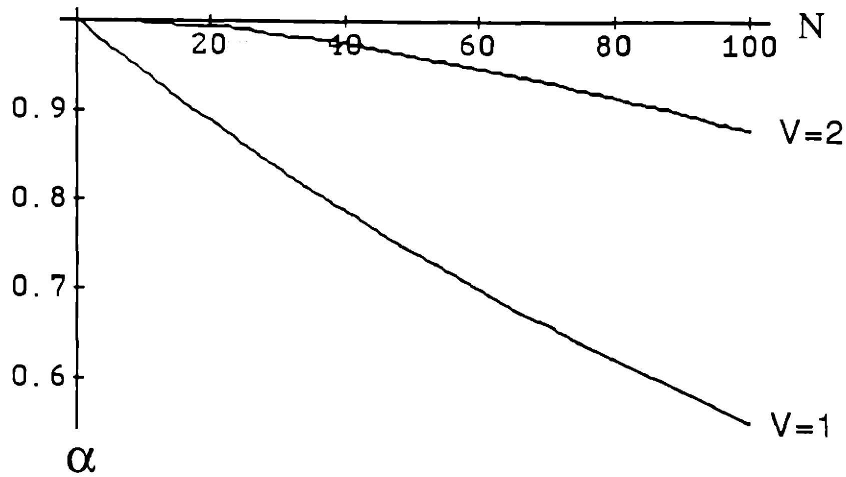

Figure 2-4: α vs N with t_r = 24 hours, V = 1, 2, p = 0.004

2.2 One-Part, Two-Stockroom Model

In this section, we study the case where the service regions of two stockrooms intersect with each other, as shown in Figure 2-5.

2.2.1 Problem Definition

Given the locations of two stockrooms and their service regions intersecting with each other, here are the parameters should be provided in the model,

1. N_1, N_2, N_{12}: number of customers only served from stockroom one, stockroom two, respectively, and number of customers served from both stockrooms,

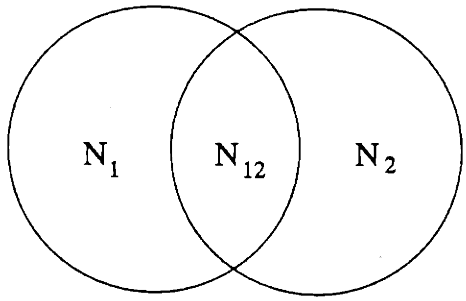

Figure 2-5: Two stockrooms with intersected service regions

2. p: probability of a customer requesting a part per day,
3. t_r: replenishment time of both stockrooms (typically 24 or 48 hours),
4. V_1, V_2: inventory level at stockroom 1 and 2, respectively, and $V_1 = V_2 = 1$ (typical value for a slow-moving part with low p), and denote by INV_1, INV_2 the inventory at stockroom 1 and 2, respectively.

Moreover, there is an assignment policy problem we need to consider: if there is a service request from one of the N_{12} customers, which of the two stockrooms should he/she call for service first? This problem can be solved by assigning each customer with a primary stockroom and a secondary stockroom, i.e., an N_{12} customer would ask his/her primary stockroom first for a service, and ask the secondary one if the primary one is out of stock. We assume that cN_{12} customers $(0 \leq c \leq 1)$ are assigned with stockroom 1 as the primary and stockroom 2 as the secondary, and $(1-c)N_{12}$ customersare assigned with stockroom 2 as the primary and stockroom 1 as the secondary. Let us denote the first group of people as N_{12}^1 $(= cN_{12})$, and the second group as N_{12}^2 $(= (1-c)N_{12})$.

With all these data and an assignment policy with constant c (the parameter π is dropped for simplicity), we can compute the satisfaction rate of the customers in each of three regions as $\alpha_1, \alpha_2, \alpha_{12}$, where α_i is the probability that when a customer from N_i requests for a part it is satisfied, and it is the same for all the N_i customers, i = 1, 2, 12. Therefore, the average customer satisfaction rate α is calculated by

$$\alpha = (\alpha_1 * N_1 + \alpha_2 * N_2 + \alpha_{12} * N_{12})/N$$

where $N = N_1 + N_2 + N_{12}$.

Next, we will write α as a function of the parameters listed above. By varying each parameter, the effect of each main parameter can be examined. Also, from these plots, the minimum inventory level V_{min} can be found given that certain satisfaction rate α^* has to be guaranteed.

2.2.2 Analysis of the Model

Again, in order to obtain closed-form formula for the satisfaction rates $\alpha_1, \alpha_2, \alpha_{12}$, we need to define some variables and functions in the model first.

At any given day, we consider a particular service request A_i from one of the group N_i customers, $i = 1, 2, 12$. Define X_j, $j = 1, 2, 12$, as the total number of requests from group N_j customers (besides A_i if $i = j$) during the same day as A_i, and X'_j as the total number of requests from group N_j customers during the day before A_i comes.

We introduce the following notations:

- $f_{X_j}(k)$: probability mass function of random variable X_j, and

$$f_{X_j}(k) = b(N_j, k) \text{ if } j \neq i, \text{ and } f_{X_i}(k) = b(N_i - 1, k)$$

- $f_{X'_j}(k)$: probability mass function of random variable X'_j, and

$$f_{X'_j}(k) = b(N_j, k), \text{ for all } j = 1, 2, 12.$$

Next, customer satisfaction rate $\alpha_1, \alpha_2, \alpha_{12}$ can be written as a function of the number of customers (N_1, N_2, N_{12}) and the probability of request (p), given the replenishment time (t_r) and certain inventory level (V_1, V_2). Since $\alpha_i = 1 - q_i$, where

q_i is the probability that where a particular request A_i comes it is not satisfied, it is equivalent to writing q_i as a function of N_i and p given t_r and V_i, denoted by $q_i(t_r, V_1, V_2)$ as shown below.

t_r = **24 hours,** $V_1 = V_2 = 1$

With replenishment time being 24 hours and one inventory at both stockrooms, q_1 is the probability that request A_1 is not the first one among X_1 other requests from N_1, plus the probability where even if A_1 is the first one among X_1, INV_1 has already been taken away by the first request from N_{12}, call it A_{12}. Therefore

$$\begin{aligned} q_1(24,1,1) &= \sum_{k=1}^{N_1-1}[f_{X_1}(k) * \text{prob}(A_1 \text{ is not the first})] + \sum_{k=0}^{N_1-1}[f_{X_1}(k) * \text{prob}(A_1 \text{ is the first})] * \\ &\quad \{\sum_{k=2}^{N_{12}}(f_{X_{12}}(k) * [\text{prob}(A_1 \text{ is not the first two}) + \\ &\quad \text{prob}(A_1 \text{ is the second}) * \text{prob}(A_{12} \text{ takes } INV_1)] + \\ &\quad f_{X_{12}}(1) * \text{prob}(A_1 \text{ is the second}) * \text{prob}(A_{12} \text{ takes } INV_1)\} \\ &= a_1(N_1-1) + a_2(N_1-1)\{\sum_{k=2}^{N_{12}}[b(N_{12},k) * (\frac{k-1}{k+1} + \frac{1}{k+1}s_1)] + b(N_{12},1)\frac{1}{2}s_1\}, \end{aligned}$$

$$\begin{aligned} \text{where } s_1 &= \text{prob}(A_{12} \text{ takes } INV_1) \\ &= \text{prob}(A_{12} \text{ is from } N_{12}^1) + \\ &\quad \text{prob}(A_{12} \text{ is from } N_{12}^2) * \text{prob}(INV_2 \text{ has been taken away by one of } X_2 \text{ before } A_{12}) \\ &= c + (1-c)a_1(N_2). \end{aligned}$$

Similarly, by symmetry we can write

$$q_2(24,1,1) = a_1(N_2-1) + a_2(N_2-1)\{\sum_{k=2}^{N_{12}}[b(N_{12},k) * (\frac{k-1}{k+1} + \frac{1}{k+1}s_2)] + b(N_{12},1)\frac{1}{2}s_2\},$$

where $s_2 = \text{prob}(A_{12} \text{ takes } INV_2) = 1 - c + ca_1(N_1)$.

$$
\begin{aligned}
q_{12}(24,1,1) &= \text{prob}(A_{12} \text{ is not one of the first two among } X_{12}) + \\
&\quad \text{prob}(A_{12} \text{ is the second one among } X_{12}) * \\
&\quad \text{prob}(\text{the first of } X_{12} \text{ takes one of } INV_1, INV_2) * \\
&\quad \text{prob}(A_{12} \text{ cannot take the other one}) + \\
&\quad \text{prob}(A_{12} \text{ is the first one among } X_{12}) * \text{prob}(A_{12} \text{ cannot take either one}) \\
&= a_3(N_{12}-1) + a_4(N_{12}-1)[a_2(N_2)a_1(N_1) + a_2(N_1)a_1(N_2)] + \\
&\quad a_2(N_{12}-1)a_1(N_1)a_1(N_2).
\end{aligned}
$$

t_r = 48 hours, $V_1 = V_2 = 1$

With replenishment time being 48 hours and one inventory at both stockrooms, similar analysis can be used to get the q_i's. As a result, the following equations are obtained,

$$
\begin{aligned}
q_1(48,1,1) &= 1 - b(N_1,0) + \\
&\quad b(N_1,0) * \{1 - b(N_{12},0) - b(N_{12},1) + \\
&\quad b(N_{12},1)[c + (1-c)(a_1(N_2) + a_2(N_2)a_1(N_1+N_{12}-2))] + \\
&\quad b(N_{12},0)[(1-b(N_2,0))a_1(N_1+N_{12}-1) + b(N_2,0)q_1(24,1,1)]\},
\end{aligned}
$$

$$
\begin{aligned}
q_2(48,1,1) &= 1 - b(N_2,0) + \\
&\quad b(N_2,0) * \{1 - b(N_{12},0) - b(N_{12},1) + \\
&\quad b(N_{12},1)[1 - c + c(a_1(N_1) + a_2(N_1)a_1(N_2+N_{12}-2))] + \\
&\quad b(N_{12},0)[(1-b(N_1,0))a_1(N_2+N_{12}-1) + b(N_1,0)q_2(24,1,1)]\},
\end{aligned}
$$

$$
\begin{aligned}
q_{12}(48,1,1) &= 1 - b(N_{12},0) - b(N_{12},1) + b(N_{12},1)[cr_1 + (1-c)r_2] + \\
&\quad b(N_{12},0) * \{[1 - b(N_1,0) - b(N_1,1)][1 - b(N_2,0) - b(N_2,1)] +
\end{aligned}
$$

$$[1 - b(N_1,0) - b(N_1,1)]b(N_2,0)a_1(N_2 + N_{12} - 1) +$$
$$[1 - b(N_2,0) - b(N_2,1)]b(N_1,0)a_1(N_1 + N_{12} - 1) +$$
$$b(N_1,0)b(N_2,0)q_{12}(24,1,1)\},$$

$$\text{where} \quad r_1 = 1 - \text{prob}(A_{12} \text{ is satisfied when } X'_{12} = 1 \text{ from } N^1_{12})$$
$$= 1 - a_2(N_1)b(N_2,0)a_2(N_2 + N_{12} - 1),$$

$$\text{where} \quad r_2 = 1 - \text{prob}(A_{12} \text{ is satisfied when } X'_{12} = 1 \text{ from } N^2_{12})$$
$$= 1 - a_2(N_2)b(N_1,0)a_2(N_1 + N_{12} - 1).$$

2.2.3 Assignment Policy

Optimal Assignment Policy

The optimal assignment policy is the one which maximizes the average satisfaction rate. Since each assignment policy π is associated with a constant c, the satisfaction rate can be written as a fuction of c, denoted as $\alpha(c)$. Then we can obtain the optimal assignment policy by finding the $c^* = \arg\max_c(\alpha(c))$, s.t. $0 \leq c \leq 1$.

As we know,

$$\alpha = (\alpha_1 * N_1 + \alpha_2 * N_2 + \alpha_{12} * N_{12})/(N_1 + N_2 + N_{12})$$
$$= [(1 - q_1)N_1 + (1 - q_2)N_2 + (1 - q_{12})N_{12}]/(N_1 + N_2 + N_{12}),$$

where q_1, q_2, and q_{12} are all linear functions of c. Therefore α is also a linear function of c. As a result, c^* must be one of the end points 0 or 1.

Let us assume $N_1 \geq N_2$. It is obvious that the assignment policy having all the N_{12} customers call stockroom 2 first ($c = 1$) is better than the assignment policy having all the N_{12} customers call stockroom 1 first ($c = 0$). Therefore the optimal c^* equals 1. In general, the optimal assignment policy says that the primary stockroom of every N_{12} customer is the one with less customers only covered by its service region.

Heuristic Assignment Policy

In practice, assignment policy for a customer in the overlapping area is usually determined by the distances between the customer and the two stockrooms, i.e., having the closer one being the primary, and the further one being the secondary.

2.2.4 Numerical Results

With closed-form formula of the satisfaction rate and the optimal assignment policy given above, we can obtain some numerical results as shown below. The part we consider here is T201500 of VAX 6000 with $p = 0.00057$ (=1/250/7, average one request every 7 years). The inventory levels at both stockrooms are assumed to be one. First, we consider the case where replenishment time is 24 hours.

t_r = 24 hours

With fixed value of p, the satisfaction rates of customers in the three subregions α_1, α_2, and α_{12} are all functions of numbers of customers in the three subregions N_1, N_2, and N_{12}, so is the average customer satisfaction rate α. Therefore, we can obtain plots of α versus each parameter separately given the other two. Due to the symmetry of N_1 and N_2, we only need to plot α versus N_1, and omit the plot α versus N_2.

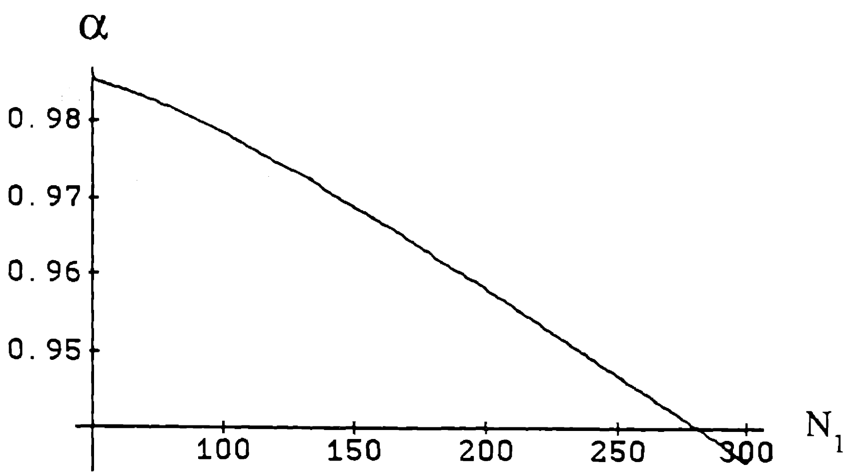

Figure 2-6: α vs N_1 with t_r = 24 hours, $N_{12} = N_2 = 50$

From Figure 2-6 we can see, if the required $\alpha^* = 95\%$, with one inventory at both stockrooms, we can have at most 225 N_1 customers given $N_{12} = N_2 = 50$.

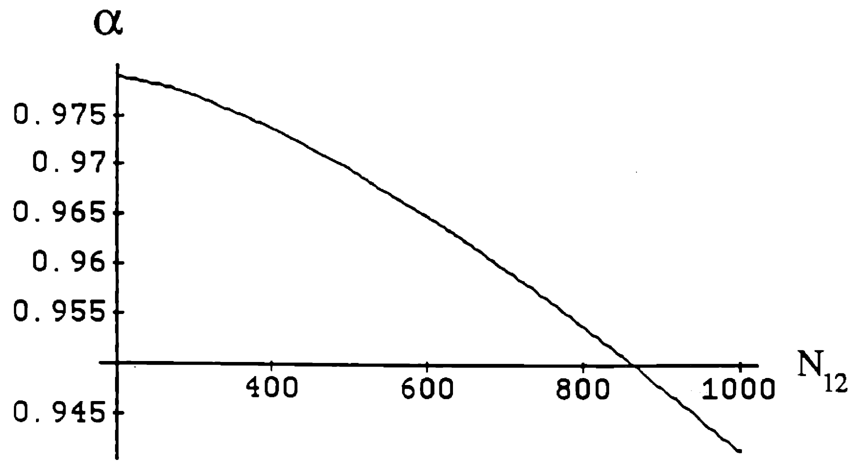

Figure 2-7: α vs N_{12} with $t_r = 24$ hours, $N_1 = 100$, $N_2 = 50$

Similarly based on Figure 2-7, if the required $\alpha^* = 95\%$, with one inventory at both stockrooms, we can have at most 850 N_{12} customers given $N_1 = 100$, $N_2 = 50$.

t_r = 48 hours

Similar to the case with 24 hours relpenishment time, we can obtain plots of α versus N_1 and N_{12} as shown below.

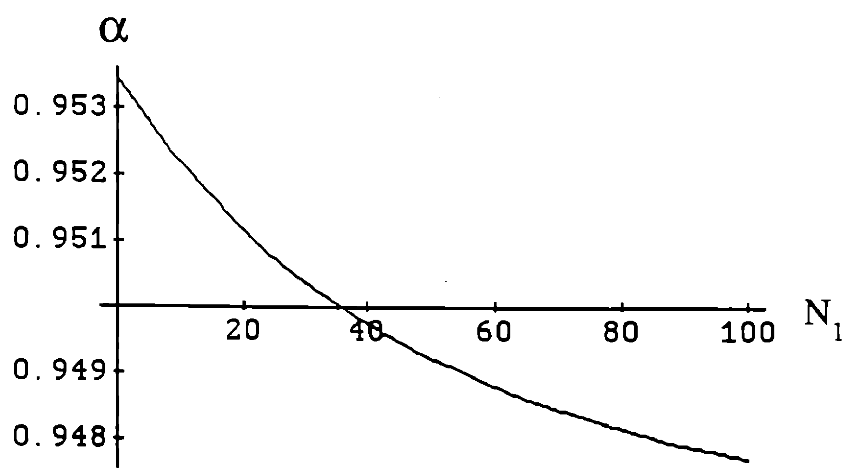

Figure 2-8: α vs N_1 with $t_r = 48$ hours, $N_{12} = N_2 = 25$

We can see form Figure 2-8, if the required $\alpha^* = 95\%$, with one inventory at both stockrooms, we can have at most 35 N_1 customers given $N_{12} = 50$, $N_2 = 25$.

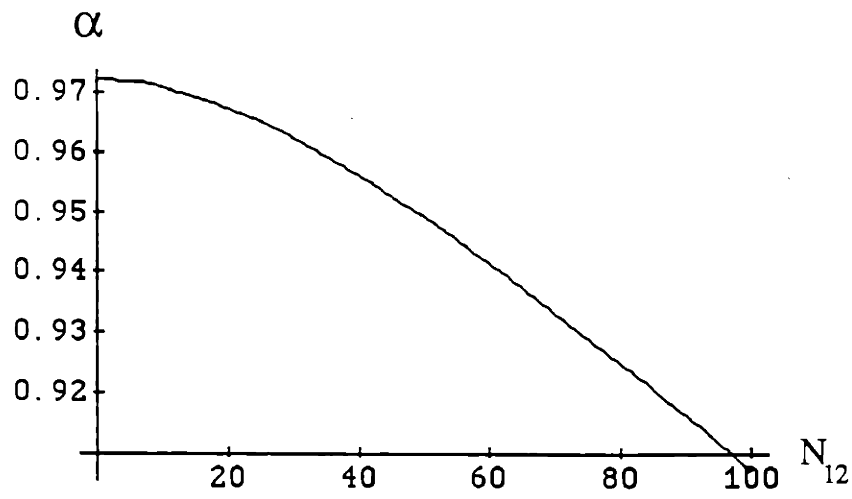

Figure 2-9: α vs N_{12} with t_r = 48 hours, N_1 = 50, N_2 = 25

Similarly based on Figure 2-9, if the required α^* = 95%, with one inventory at both stockrooms, we can have at most 45 N_{12} customers given N_1 = 50, N_2 = 25.

As it has been proved in the analysis, the optimal assignment policy is the one that assign all the N_{12} customers with stockroom two as their primary one, and stockroom one as their secondary one assuming $N_1 \geq N_2$. Let us denote the corresponding satisfaction rate as α_{opt}. Denote by α_{opt^c} the satisfaction rate which corresponds to the assignment policy that is opposite to the optimal one, i.e., the primary stockroom of N_{12} customers is stockroom one in stead of stockroom two. The following table shows that the effect to satisfaction rate α the assignment policy makes with t_r = 48 hours, N_1 = 100, N_{12} = N_2 = 50, and various values of p.

probability of request p	optimal satisfaction rate α_{opt}	satisfaction rate α_{opt^c}
0.00057	93.02%	92.66%
0.001	87.96%	87.32%
0.002	57.16%	55.07%

Table 2.1: The effect of assignment policy to satisfaction rate with t_r = 48 hours, N_1 = 100, N_{12} = N_2 = 50

As we can see from Table 2.1, the difference between α_{opt} and α_{opt^c} is very small for low value of p of interest. Therefore, for the purpose of slow-moving parts, the assignment policy is not the main issue to consider. The heuristic policy (assigning customers with primary/secondary stockroom by distances between them) is reasonably acceptable.

Chapter 3

An Optimization Approach to the One-Part, Multi-Stockroom Problem

In this chapter, we address the problem of finding the level and locations of inventories for a single part in order to guarantee a certain average customer satisfaction rate. At the same time, we will try to obtain an optimal assignment policy given the inventory level and locations.

As we mentioned in Chapter 2, when the total number of stockrooms is larger than 2, it becomes very difficult to derive a closed-form formula for the average customer satisfaction rate. Therefore, approximate formulae and simulations are going to be used as alternatives in this chapter.

In Chapter 2, it has been shown that different assignment policies do not affect the value of average customer satisfaction rate very much. Therefore, it is not wise to solve an assignment policy problem with every set of inventory locations; instead, we can decompose the problem in two steps. First, we find the optimal inventory level and their locations with a heuristic assignment policy. Then with the optimal inventory locations, we find the corresponding optimal assignment policy.

3.1 Inventory Location Problem

3.1.1 Problem Definition

In this section, we consider the inventory location problem. Let R be the total number of stockrooms which can be either the existing ones, or the potential ones. The following heuristic assignment policy is used, for simplicity: customers are assigned to stockrooms according to the distances between customers and stockrooms, i.e., the closest stockroom is the primary stockroom, and the second closest one is the secondary stockroom. Given the minimum average customer satisfaction rate α^* that has to be guaranteed, our goal is to find the minimum amount of inventory (V^*) needed as well as its location. The main idea of the approach to this problem is as follows.

3.1.2 Algorithm

First, we view all the R stockrooms as one stockroom (with infinitely large service region) which serves all the N customers. Then we use the simple one-part, one-stockroom model discussed in Chapter 2 to find the minimum number of inventory corresponding to α^*. This serves as a lower bound for V^*, denoted by V_{LB}. The binary variable x_{kj} is defined in Chapter 1, i.e., $x_{kj} = 1$ if the kth inventory ($k = 1, ..., V$) is at the jth ($j = 1, ..., R$) stockroom, 0 otherwise; $\underline{x}$ is the inventory location vector with components x_{kj}. Therefore, the average customer satisfaction rate can be written as a function of $\underline{x}$, denoted as $\alpha(\underline{x})$

The algorithm starts with $V = V_{LB}$, and V is incremented by 1 after each iteration until the corresponding satisfaction rate is at least as high as α^*. The algorithm is implemented as follows:

1. $V = V_{LB}$

2. Find α_{max} by solving the (LOCATION) problem as stated below,

$$\text{(LOCATION)} \quad \alpha_{max} = \max_{\underline{x}} \alpha(\underline{x})$$

s.t. $\sum_{j=1}^{R} x_{ij} = 1$ for all $i = 1, 2, ..., V$

3. if $\alpha_{max} \geq \alpha^*$, stop, and $V^* = V$, $\underline{x}^* = \underline{x}$; otherwise, let $V = V + 1$, go back to step 2.

If the value of R is relatively small (e.g. 5), problem (LOCATION) can be solved by enumerating all the possible combinations of locations (there is a total of R^V different sets of locations) and select the one which maximizes the satisfaction rate. However, when R and V are large, it may become impossible to enumerate all the combinations. In this case, the problem can be viewed as a global optimization problem where the search region is $\{1, ..., R\}^V$. We can attack this problem using one of the following three algorithms. First, a heuristic algorithm is used to locate inventories at the V stockrooms that cover as many customers as possible. In other words, the objective function of (LOCATION) which is the satisfaction rate has been changed to the percentage of customers covered by the selected stockrooms. Let us call this algorithm the covering algorithm. Next, the greedy algorithm is used as a second heuristic algorithm to again maximize the satisfaction rate. It performs well in the case where a stockroom service region is relatively small compared to the whole district. Finally, a recently developed algorithm called Adaptive Partitioned Random Search (APRS) [Tang] is used to find a reasonably good solution under all conditions. In the next three subsections, we introduce each of the three algorithms in details.

3.1.3 A Covering Algorithm for the LOCATION Problem

Due to the low probability of demand for the slow-moving parts under consideration, once a customer is covered by one of the stockrooms, his/her satisfaction rate is usually very close to 1. Here, we assume that the number of customers sharing the inventory at the same stockroom is not very large, which is usually the case in practice. Take the numerical results in section 2.1.3 as an example. With 48 hours replenishment time, a stockroom with unit inventory of part T201500 of VAX 6000 (p = 0.00057) can support 50 customers with satisfaction rate at least as high as 95%.

Therefore, the solution of covering algorithm which locates the given V inventories at the stockrooms which cover the maximum percentage of customers before everyone is covered provides a suboptimal solution to the original (LOCATION) problem (maximizing average customer satisfaction rate). Although the whole sample space still has a total of R^V different sets of locations, each function evaluation time is much less than the million days used in the simulation of the satisfaction rate. However, this algorithm does not work well when the number of inventories increases to the point where selected stockrooms can cover all the customers. Many combinations of stockrooms which can cover all the customers may result in different satisfaction rates and will be shown in the numerical results in Section 3.1.6.

3.1.4 A Greedy Algorithm for the LOCATION Problem

Let us first reformulate the (LOCATION) problem in order to apply a greedy algorithm. As in Chapter 1, Q_R is defined as the set of all the stockrooms locations 1, 2, ... R. Denote by S_V the set of locations for the given V inventories, and consider only sets S_V of cardinality V. (That is, no location will have more than one unit of inventory, which is usual the case for slow-moving parts.) For a given inventory level V, the average satisfaction rate can be written as a function of S_V, denoted as $\alpha(S_V)$. The problem becomes,

$$\text{(LOCATION)} \quad \alpha_{max} = \max_{S_V \subseteq Q_R} \alpha(S_V)$$

The main idea of the greedy algorithm is to optimize the objective function with respect to one inventory location at a time. The algorithm for the (LOCATION) problem is as follows:

1. **(Initialization)** $S_0 \leftarrow \emptyset, k \leftarrow 1.$
2. **(Main Loop)** For $k = 1, 2, ..., V$

$$j_k \leftarrow \arg \max_{j \in Q_R \setminus S_{k-1}} \alpha(S_{k-1} \cup \{j\})$$

$$S_k \leftarrow S_{k-1} \cup \{j_k\}$$

3. **(Output)** S_V, $\alpha(S_V)$.

Given the amount of inventories V and the set of potential locations $Q_R = \{1, 2, ..., R\}$, the algorithm outputs a set S_V of V locations with function value $\alpha(S_V)$.

Clearly, the greedy algorithm returns a optimal solution for $V = 1$. For general V, let us change the objective function to be the number of customers covered by the V selected stockrooms, as in the covering algorithm. The problem then becomes a special case of the delta coverage problem (P_1) in [BeBL]. [BeBL] has proved that the objective function is nondecreasing and submodular, and therefore the greedy algorithm guarantees a solution within 37% from the optimal solution value [NeWF]. It has also been stated in [BeBL] that the result of the greedy algorithm for (P_1) can be very "close" to the optimal value in practice. Therefore, we can expect a similar performance for our original problem, especially when the service region is relatively small compared with the whole district so that the satisfaction rate can be approximately decomposed by the average of the sum of the satisfaction rate of customers covered by each of the V stockrooms. However, for some other cases the results from greedy algorithm may not be good enough if we need to find a high value (very close to 1) of α. In the next section, we introduce a recently developed APRS algorithm to obtain reasonably good results for all cases.

3.1.5 Adaptive Partitioned Random Search (APRS) Algorithm to the LOCATION Problem

The APRS algorithm is a recently developed algorithm for global optimization. In [Tang], it has been demonstrated that APRS is simple in terms of its easy implementation, robust in terms of its applicability to high dimensional and discrete search regions, and efficient in terms of its computation time. And most importantly, it is generic in that it makes no assumptions on the objective function. A brief introduction to APRS is given below.

Let $f : A \rightarrow R$ be a real valued objective function defined in a region $A \subset R^d$. The global optimization problem often considered in literature is aimed to finding

$x^* \in A$ such that

$$f(x^*) = \max\{f(x) : x \in A\}$$

where x^* and $f(x^*)$ are referred to as the global maximizer and the global maximum, respectively, which we assume to exist. Numerically, the global optimization problem defined above is approximated by finding a point in the set

$$B(\varepsilon) = \{x \in A : f(x^*) - f(x) \leq \varepsilon\}.$$

The complexity of the problem is determined mainly by properties of the feasible region A and the objective function f. For example, for a generic continuous function defined in A, the Lebesgue measure of the set $B(\varepsilon)$ can be arbitrarily small with respect to the Lebesgue measure of the entire feasible region A. Therefore, given ε and for an arbitrary continuous function, there is no algorithm which can produce a point $x \in B(\varepsilon)$ with probability one in finite time.

APRS takes more realistic point of view and considers maximizing the expectation of the largest function value that is obtained for a given number of function evaluations. More specifically, the problem can be formulated as follows. Partition the search region A into $K \geq 2$ subregions A_i, $1 \leq i \leq K$, such that $A_i \neq \emptyset$, $\cup_{i=1}^{K} A_i = A$ and $A_i \cap A_j = \emptyset$ for $i, j = 1, ..., K$ and $i \neq j$. An independent identical distributed (i.i.d.) random sampling scheme is employed for each of K subregions. The observed function value $f(x)$ in subregion i ($i = 1, ...K$) can be considered as a random variable. Let $F_i(y)$ denote the induced distribution function of the observed function value $f(x)$ in subregion i. That is, $F_i(y) = \Pr(f(x) \leq y)$ given that function evaluations are taken from subregion i.

Denote by N the total number of function evaluations allowed. A sequential sampling policy π corresponds to a rule that determines a sequence of subregions where function evaluations are taken at each $t = 1, ..., N$. (The decision at time step t may depends on the outcomes of previous time step evaluations.) Let Δ be the set of all admissible sampling policies. Then, the partitioned random search (PRS) problem is defined as to find an optimal policy π^* such that

$$J_N(\pi^*) = \max_{\pi \in \Delta} J_N(\pi) = \max_{\pi \in \Delta} E[max\{f(x(1)), ..., f(x(N))\}],$$

where $f(x(t))$ is the function value sampled at time t. The *promising index* $I_i(M)$ of distribution function $F_i(y)$ at M (for all t) is defined as the expected gain to be added to M if the next evaluation is taken from subregion i. That is

$$I_i(M) = E[(f(x) - M)^+ | x \in A_i] = \int_M^\infty [1 - F_i(y)]dy.$$

Note that $I_i(-\infty) = E[f(x)|x \in A_i]$. Define $M(t)$ as the largest function value ever obtained up to time t, $1 \leq t \leq N - 1$. That is,

$$M(t) = \max\{f(x(1)), ..., f(x(t))\}.$$

The following theorem is proved in [Tang].

Optimal Sampling Theorem Given that $F_i(y)$ is piecewise continuous and $|I_i(-\infty)| < \infty$ for $1 \leq i \leq K$, sampling the subregion with the maximum promising index at $M(t)$ for $1 \leq t \leq N$ is the optimal sampling policy.

In [Tang], it has been shown that if promising indices are available whenever needed, the larger the number of partitioned subregions K, the larger the performance value $J_N(\pi^*)$ for a given N. However, in practice the probability distribution $F_i(y)$ ($1 \leq i \leq K$) is unknown. In [Tang], a procedure for estimating the promising indices is proposed and requires at least three function evaluations for each subregion. In other words, if the search region A is partitioned into too many small subregions, many function evaluations will be used to estimate the promising indices before the optimal sampling policy can be started. This leads to the adaptive version of the algorithm as APRS in which only the most promising subregion (the one with largest promising index) will be further partitioned into a certain number of small subregions.

The Main Idea of APRS APRS requires that at any time, the entire search region A is partitioned into K subregions. That is, if one of K subregions is chosen to be further partitioned, it is partitioned into $(K-1)$ smaller subregions and all the other $(K-1)$ remaining subregions are aggregated into a single subregion. Therefore, no part of A is ever excluded from consideration for possible sampling in the future. This operation is called as *repartition.* The subregion being chosen to be partitioned is referred to as the *working region.* For example, at the very beginning, the working region is the entire search region A.

It has been observed in [Tang] that if a subregion has both the largest promising index and the largest function value being sampled, this subregion is far more promising than any of the other subregions and hence is chosen as the working region, otherwise the optimal sampling policy is applied.

In this way, the "size" of the working region can decrease exponentially as the search procedure progresses. Since the estimates of the promising indices are subject to errors, it is desirable to be able to control the reduction speed of the "size" of the working region. Therefore, after a fixed number of repartitions, the working region is redefined. This operation is referred to as *adjustment.* The fixed number of repartitions between two consecutive adjustments is called the *adjustment delay* l. More specifically, the adjustment is conducted such that

$$\mu(A_w(k)) = r^k \mu(A)$$

where $\mu(A_w)$ and $\mu(A)$ are the measure of the sizes of the current working region A_w and the entire search region A, respectively, and k is the total number of adjustments being conducted. The reduction speed of the size of the working region is then controlled by taking the value of r between 0 and 1, where r is referred as the *shrinking factor.* Furthermore, the adjusted working region should have the point where the largest function value has been sampled at its "center" as much as possible. The search procedure is terminated if the "size" of the current working region is less than some prespecified number TOL.

In summary, the APRS described above is characterized by four parameters: the number of partitions K, the adjustment delay l, the shrinking factor r, and the stopping condition parameter TOL.

Implementation of APRS with a Hypercube Search Region A

If the search region A of the objective function $f(x)$ is a d-dimensional hypercube as defined by

$$A = \{x \in R^d | a_1 \leq x_1 \leq b_1, ..., a_d \leq x_d \leq b_d\},$$

APRS algorithm is implemented as follows:

1. Initially, the entire region A is taken as the working region. It is partitioned in the manner such that each of the first d_p dimensions is partitioned into k_0 (e.g., 2) subintervals with equal length. d_p and k_0 are referred to as the *partition depth* and the *one dimension partition degree*, respectively. Thus, the number of subregions $K = k_0^{d_p}$.

2. A fixed number (e.g., 4) of function evaluations is performed in each subregion and promising indices are estimated (see [Tang] for index estimation). The optimal sampling policy (taking the next sample in the subregion with the largest promising index) is applied to the K subregions until the largest function value obtained is from the subregion whose promising index is also the largest. In addition, the number of applications of the optimal sampling policy is upper bounded by some prespecified number (e.g., 5). This is to prevent the search procedure from being trapped in there.

3. This subregion with the largest promising index is taken as the working region and is further partitioned into $k_0^{d_p}$ smaller hypercubes in a similar fashion as in step 1 (each of the d_p dimensions with relatively longer length is partitioned into k_0 subintervals with equal length). The remaining unpartitioned subregions are aggregated into a single subregion with irregular shape. Thus, the total number of subregions becomes $K = k_0^{d_p} + 1$.

4. If the subregion with irregular shape should be chosen as the working region, the entire search region A is taken as the working region to make the repartition simple and go to step 1. This operation is referred as the *restoration.*

5. Let the adjustment delay l be d/d_p if $\mathrm{mod}(d/d_p) = 0$ or $\lfloor d/d_p \rfloor + 1$ if $\mathrm{mod}(d/d_p) \neq 0$. In other words, after l consecutive repartitions have been conducted since the last adjustment or restoration, a new adjustment is called. The length of the longest dimension of the working region is used as the measure of its "size".

6. The search procedure is terminated if the length of the longest dimension of the working region is less than or equal to TOL.

Thus, if A is a hypercube, the APRS is characterized by four parameters: the partition depth d_p, the one dimension partition degree k_0, the shrinking factor r, and the stopping condition parameter TOL. The adjustment delay is embedded in the algorithm.

When the Search Region A is Discrete: In the case when the d-dimensional search region A is discrete (for our purpose, $d = V$), and is defined by

$$A = \{x \in Z^d | x_j \in \{1, 2, ..., N_j\}, j = 1, ..., d\},$$

where N_j is some positive integer ($N_j = R$ for the LOCATION problem), the APRS designed for a hypercube search region can still be applied here with care.

More specifically, a hypercube A' is defined by

$$A' = \{x \in R^d | r_j < x_j \leq N_j, j = 1, ..., d\},$$

where r_j is some positive number slightly smaller than 1 (e.g., $r_j = 0.95$). It is clear that $A \subset A'$. Partitions are performed on A', and the subregions A'_i, $1 \leq i \leq K$, take forms of

$$A'_i = \{x \in R^d | a_j < x_j \leq b_j, j = 1, ..., d\}.$$

When a subregion A_i' is chosen to be sampled, only those feasible discrete points in A_i' will be sampled equally likely. Thus, there is no need to make the stopping parameter TOL smaller than 1.

The numerical results in next subsection are obtained by applying each of the three algorithms.

3.1.6 Numerical Results

According to the data obtained from DEC Virginia service district, we construct the following geographic map. Since the customer locations are not given in the data, we assign locations to every customer so that the results are consistent with the data we have (see Appendix B).

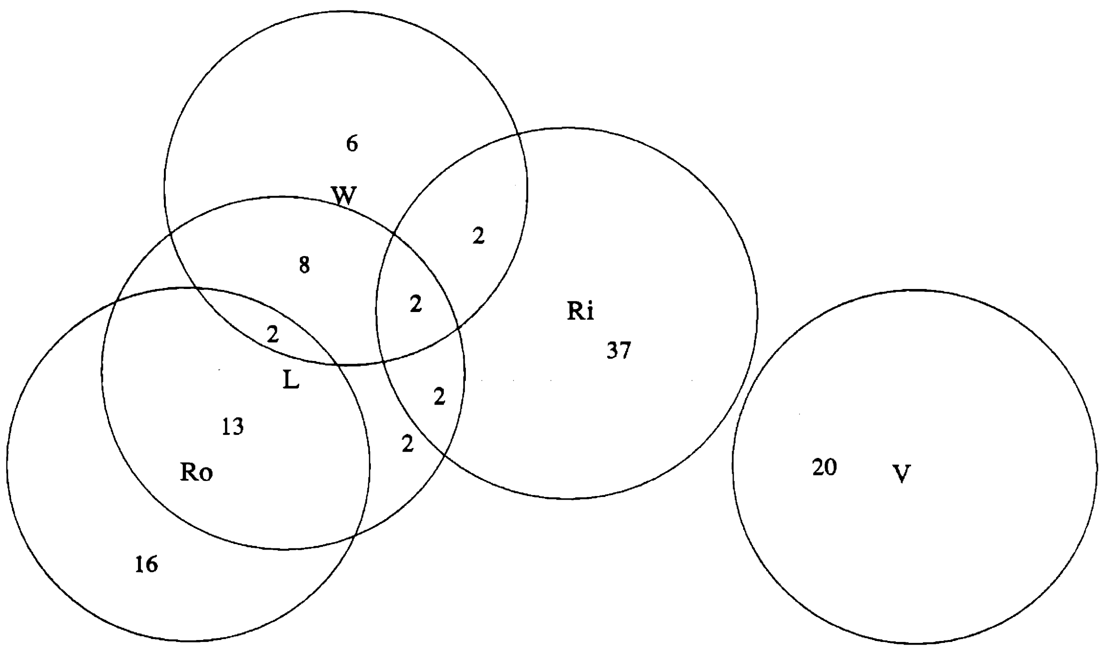

Figure 3-1: An example: DEC Virginia service district

In Figure 3-1, Ro, L, W, Ri, and V are five current stockrooms. More specifically, (1) Ro stands for the stockroom at Roanoke, (2) L at Lynchburg, (3) W at Waynesboro, (4) Ri at Richmond, and (5) V at Virginia Beach. The number indicated at each disjoint subregion is the number of the customers in the area. As a whole, there are total 110 customers in the Virginia district.

Here, we assume that the heuristic assignment policy is used, i.e., customers are assigned to stockrooms according to the distances between them. The part we consider here is part T2014BA of VAX6000. Customers all have a probability 0.0004 of requesting a service at any given day (average one service call per 10 years). The replenishment time for all the stockrooms is 48 hours.

Given the inventory level V, we can find the best two combinations of locations $(\underline{x})$ in terms of the average customer satisfaction rate by enumerating all the 5^V possible choices. The corresponding average traveling distances between stockrooms and customers are also obtained according to the scale in Figure 3-2, and all the results are shown in Table 3.1.

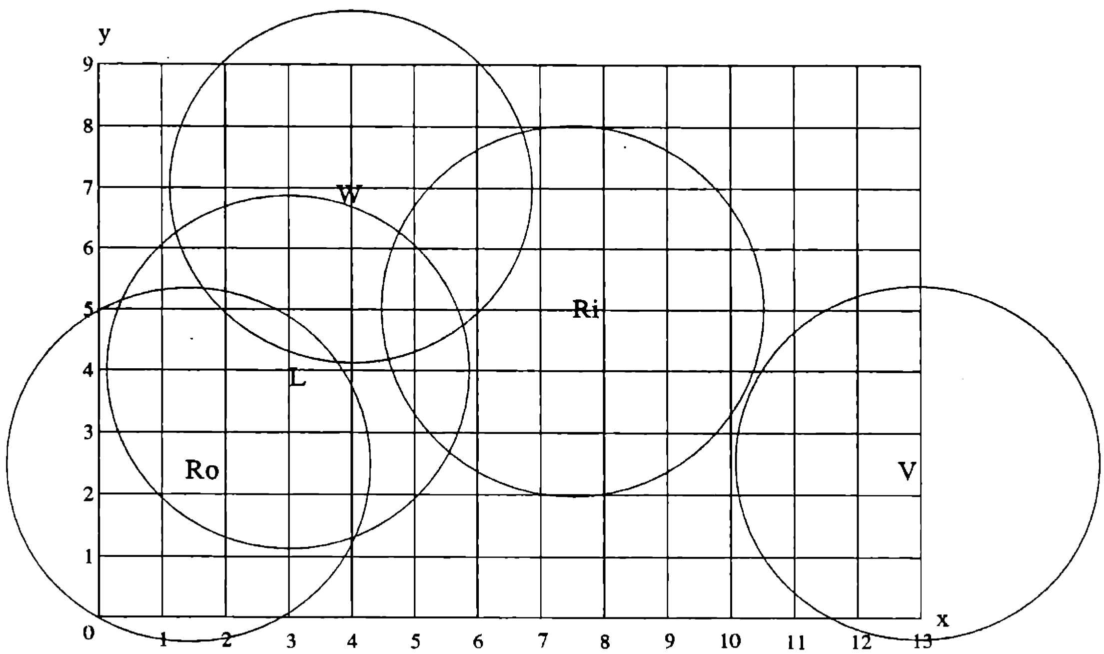

Figure 3-2: A scaled map of DEC Virginia service district (1:10miles)

Next, assume we have the freedom to locate stockrooms at any one of the potential 140 locations, whose positions are represented by the coordinates (x, y), where $x = 0, 1, ..., 13$, and $y = 0, 1, ..., 9$ in Figure 3-2. Now, since the value of R (=140) is large, we use the proposed three algorithms to find the the best locations for the different levels of inventory (V). The corresponding results are shown in Table 3.2, 3.3 and 3.4.

From the numerical results, we can see that all three algorithms perform almost

inventory level (V)	inventory locations ($\underline{x}$)	satisfaction rate (α)	average traveling distance
1	Ri	40.83%	1.604
	L	33.19%	2.153
2	Ro, Ri	67.78%	1.633
	L, Ri	66.68%	1.766
3	Ro, Ri, V	86.01%	1.619
	L, Ri, V	84.92%	1.725
4	Ro, W, Ri, V	97.74%	1.571
	L, W, Ri, V	90.21%	1.623

Table 3.1: Numerical results from enumeration with five current stockrooms and their 2-hour service regions

inventory level (V)	inventory locations ($\underline{x}$)	percentage of customers covered	satisfaction rate (α)
1	(7,5)	43.64%	43.57%
2	(3,4), (9,4)	72.73%	72.20%
3	(2,4), (7,4), (11,2))	88.18%	87.78%
4	(1,2), (3,6), (8,4), (11,2))	100.00%	99.42%

Table 3.2: Numerical results from the covering algorithm with 140 potential stockrooms and their 2-hour service regions

equally well and give better results than the current stockrooms, as expected. The covering algorithm works extremely well for small V, in which case not all customers are covered. Comparing the results of the greedy algorithm and APRS, the number of points APRS uses for a given V is very close to the one the greedy algorithm uses (i.e. $140V$), which is a linear function of R (=140). And it is much less than the total number of solutions 140^V. Moreover, the locations found by both algorithms are exactly the same for $V = 1, 2$ and very close for larger values of V. This is because the service region of one stockroom is relatively small to the whole district so that inventories can be located one by one optimally. However, since we do not have a closed-form formula for the satisfaction rate, we cannot determine whether the results obtained are globaly optimal.

inventory level (V)	inventory locations ($\underline{x}$)	satisfaction rate (α)	sample points
1	(7,5)	43.57%	140
2	(2,3), (7,5)	71.67%	280
3	(2,3), (7,5), (12,3)	89.36%	420
4	(2,3), (4,8), (7,5), (12,3)	97.42%	560

Table 3.3: Numerical results from the greedy algorithm with 140 potential stockrooms and their 2-hour service regions

inventory level (V)	inventory locations ($\underline{x}$)	satisfaction rate (α)	average travel distance	sample points
1	(7,5)	43.57%	1.751	56
2	(2,3), (7,5)	71.67%	1.760	233
3	(2,3), (6,6), (11,3)	89.54%	1.906	441
4	(2,2), (4,6), (7,4), (11,3)	99.63%	1.627	617

Table 3.4: Numerical results from the APRS algorithm with 140 potential stockrooms and their 2-hour service regions

When the service timing constraints increases to 4 hours, the radius of the circles in the map is assumed to be doubled. In a similar fashion, we obtain the results shown by Table 3.5, 3.6, 3.7 and 3.8.

Again, all the three algorithms generate the same result for the case with $V = 1$. As V increases to 2, it is easy to locate inventories at two stockrooms that can cover all the customers, and the result of the covering algorithm is not as good as the one of APRS algorithm (both sets of locations cover 100% of customers). On the other hand, since the service region of one stockroom is relatively larger than before, the given two inventories should be located at each part (left and right) of the whole region so that all the customers are covered and the area (around Ri) where customers are most concentrated is in the overlapped region of the two selected stockrooms. With the greedy algorithm, since the first inventory is fixed at the center area around Ri, the final result is not as good as the one from APRS algorithm. Therefore, as a whole, the APRS algorithm performs well under all the conditions, and the covering

inventory level (V)	inventory locations ($\underline{x}$)	satisfaction rate (α)	average traveling distance
1	Ri	75.92%	3.073
	L	72.38%	3.436
2	L, V	98.85%	3.123
	L, Ri	95.01%	2.517
	Ro, Ri	94.67%	2.400
3	Ro, Ri, V	99.72%	1.928
	L, Ri, V	98.81%	1.928

Table 3.5: Numerical results from enumeration with five current stockrooms and their 4-hour service regions

inventory level (V)	inventory locations ($\underline{x}$)	percentage of customers covered	satisfaction rate (α)
1	(6,3)	84.55%	82.99%
2	(0,3), (8,4)	100.00%	98.96%

Table 3.6: Numerical results from the covering algorithm with 140 potential stockrooms and their 4-hour service regions

inventory level (V)	inventory locations ($\underline{x}$)	satisfaction rate (α)	sample points
1	(6,3)	82.99%	140
2	(6,3), (12,4)	93.71%	280
3	(2,4), (6,3), (12,4)	99.87%	420

Table 3.7: Numerical results from the greedy algorithm with 140 potential stockrooms and their 4-hour service regions

inventory level (V)	inventory locations ($\underline{x}$)	satisfaction rate (α)	average travel distance	sample points
1	(6,3)	82.99%	3.556	66
2	(5,3), (8,4)	99.56%	2.654	244
3	(2,5), (5,3), (13,4)	99.89%	2.429	499

Table 3.8: Numerical results from the APRS algorithm with 140 potential stockrooms and their 4-hour service regions

algorithm and greedy algorithm can be used according to the specific conditions one will have.

From the results, we can also find the minimum inventory level and their locations with a given satisfaction rate that has to be guaranteed. Let us take the current five stockrooms as an example. If the minimum 2-hour satisfaction rate is 65% and 4-hour satisfaction rate is 95%, from Table 3.1 and 3.5 we can see that the best choice of locations is (L, Ri) with $V = 2$. If the minimum 2-hour satisfaction rate goes up to 85%, then the best one would be (Ro, Ri, V). When traveling costs are taken into consideration, Table 3.5 shows that the set of locations with higher satisfaction rate does not always associate with lower average traveling distance. For example, the table shows that the best locations in terms of satisfaction rate is (L,V) with $V = 2$. Howerver, both set of locations (L,V) and (L,Ri) are feasible if the required satisfaction rate $\alpha^* = 95\%$. Therefore, the district manager may choose the set (L,Ri) with less average traveling distance, which also has the minimum inventory level $V = 2$ and still keep at least 95% satisfaction rate. Moreover, the numerical results have shown that with the typical values of α^* (used at DEC service districts), at most three inventories of each part are needed for the whole district, where the current five stockrooms store 4 or even 5 inventories for some parts (see Appendix A). Thus, our solution can provide impact of possible cost saving to service centers.

With the best solution of inventory level and locations found above, we try to find the corresponding optimal assignment policy as shown in the next section.

3.2 Assignment Policy Problem

3.2.1 Problem Formulation

The assignment policy problem can be stated as follows. Given the locations of all the stockrooms and all the customers in the district of interest, we are trying to find the optimal assignment policy in order to maximize customer satisfaction rate by solving an integer programming problem as shown below.

With the inventory level V^* and locations obtained from last section, we assume unit inventory at each stockroom, which is usually the case for slow-moving parts, and renumber those stockrooms with unit inventory as $j = 1, 2, ..., V^*$. We assume that each customer can get services from at most the closest two stockrooms. According to the geographic map of the local service district, the following data can be obtained.

- M_{jk} = number of customers served by both stockroom j and stockroom k if $j \neq k$; note $M_{jk} = M_{kj}$,

- M_{jj} = number of customers only served by stockroom j.

The decision variable N_{jk} we want to find is defined as,

- N_{jk} = number of customers assigned with stockroom j as their primary one and with stockroom k as their secondary one if $j \neq k$; note that N_{jk} is not the same as N_{kj},

- N_{jj} = number of customers only served by stockroom j as their primary, obviously $N_{jj} = M_{jj}$.

Therefore, the only variables we need to find are N_{jk}'s for $j \neq k$. Denote $\underline{N}$ as the decision vector with components N_{jk}'s, and the average satisfaction rate can be written as an approximate nonlinear function of $\underline{N}$ as $\alpha(\underline{N})$ (see Appendix C). Then the problem can be formulated as a combinatorial optimization problem with an nonlinear objective function $\alpha(\underline{N})$ and linear constraints as below.

$$\text{(ASSIGNMENT)} \quad \max_{\underline{N}} \alpha(\underline{N})$$

$$\begin{aligned} \text{s.t.} \qquad N_{jk} + N_{kj} &= M_{jk} \quad \text{for all } j, k, \text{ s.t. } j < k \\ N_{jk} &\geq 0 \quad \text{for all } j, k \\ N_{jk} &\in Z \quad \text{for all } j, k \end{aligned}$$

This integer programming problem can be solved using a package called GAMS/MINOS. The numerical results are shown in the next subsection.

3.2.2 Numerical Results

Given the data of M_{ij}'s, we can obtain the optimal assignment policy using GAMS/MINOS. Let us denote the corresponding satisfaction rate as α_{opt}. Denote by α_{opt^c} the satisfaction rate which corresponds to the assignment policy that is opposite to the optimal one, i.e., N_{ij} group of customers (primary i, secondary j) will become N_{ji} group of customers (primary j, secondary i). The following table shows the effect of the assignment policy to the average satisfaction rate with t_r = 48 hours, $M_{11} = M_{12} = M_{23}$ = 100, M_{22} = 50, M_{33} = 0, and various values of p. Here, the optimal assignment policy is $M_{12} = N_{21}$ and $M_{23} = N_{32}$, i.e., having all M_{12} customers call stockroom 2 first, and all M_{23} customers call stockroom 3 first.

probability of request p	optimal satisfaction rate α_{opt}	satisfaction rate α_{opt^c}
0.0004	96.73%	95.16%
0.001	91.11%	87.68%
0.002	81.04%	75.88%

Table 3.9: The effect of assignment policy to satisfaction rate for multi-stockroom

As we can see from Table 3.9, the difference between α_{opt} and α_{opt^c} is very small for low value of p of interest. Therefore, for the purpose of slow-moving part, the assignment policy is not the main issue to consider. The heuristic policy (assigning customers by distances between customers and stockrooms) is reasonably acceptable, and also easy to use in practice since customers' assignment policies are independent of each other unlike the optimal one.

Chapter 4

An Optimization Approach to the Multi-part, Multi-stockroom Problem

In this chapter, we consider the core problem of this whole research, as defined in Chapter 1. We have S different types of parts. Each part i ($i = 1, 2, ..., S$) has a probability p_i that a customer will request one part i at any given day. Let R be the total number of stockrooms which can be either the existing ones, or the potential ones (if we have the freedom to locate stockrooms). The heuristic assignment policy is chosen to be used for simplicity, where customers are assigned to stockrooms with inventory according to the distances between them, i.e., the closest one to be the primary stockroom, and the second closest one to be the secondary stockroom. Given the minimum customer satisfaction rate α^* that has to be guaranteed, our goal is to find the optimal locations of stockrooms and the minimum number of inventory (V_i^*) needed as well as their locations ($\underline{x}^{*i}$) for all parts $i = 1, 2, ..., S$. The problem can be divided into two subproblems,

1. When the locations of all the stockrooms are fixed, this becomes a location problem for the inventories only;

2. When the stockroom locations are the potential ones, the problem becomes a location problem both for the inventories and for the stockrooms.

Now, let us consider the two subproblems separately.

4.1 With Fixed Stockrooms

In the case when the stockrooms are fixed, we do not need to consider the setup cost for the stockrooms. The only cost we need to consider is the inventory cost for all the S parts. Let C_i be the unit inventory cost for part i, and V_i be the inventory level of part i, then the problem becomes,

$$\min \sum_{i=1}^{S} C_i V_i \qquad \text{s.t. } \alpha_i \geq \alpha^* \text{ for all } i = 1, 2, ..., S$$

which is equivalent to

$$\min V_i \qquad \text{s.t. } \alpha_i \geq \alpha^* \text{ for all } i = 1, 2, ..., S$$

Now, all the parts can be solved independently using the one-part, multi-stockroom model in Chapter 3 to find the minimum inventory level required in order to keep certain satisfaction rate level.

4.2 With Potential Stockrooms

When the stockroom locations are potential ones, the setup cost for stockrooms needs to be considered as well as the inventory cost for parts in stock. Let us assume that the setup cost is much larger than the inventory cost, in which case, we want to keep the number of stockrooms as low as possible. This may result in storing several different parts in the same stockroom.

Although, in this case we cannot treat each part separately as before, the one-part, multi-stockroom model in Chapter 3 can still be used to find a lower bound for the inventory level needed for each part and restrict the number of stockrooms to be the same as the largest lower bound. Then we can check to see if these set of inventory levels are feasible in terms of satisfaction rate. If not, we increase the inventory levels such that the increase of the inventory cost is minimum at each iteration, and go back to check the feasibility again. Repeat this procedure until a feasible set of inventories and their optimal locations are found. The formal algorithm is given as follows.

4.2.1 Algorithm

First, we view each of the S parts individually. By applying the one-part, multi-stockoom model in Chapter 3 to each part i ($i = 1, 2, ..., S$), we obtain the minimum number of inventory corresponding to α^*, which serves as a lower bound for V_i^*, denoted by V_i^{LB}. Fix the number of stockrooms denoted as N_{SR} to be the number of different inventory locations of part i^*, where $i^* = arg \max_i V_i^{LB}$, and $N_{SR} = V_{i^*}^{LB}$ if every stockroom has unit inventory of part i^*. Obviously, part i^* is also the one with $\max_i(p_i)$.

With the lower bound vector $\underline{V}^{LB}$ (with components V^{LB}'s) and N_{SR}, the algorithm is implemented as follows:

1. $V_i = V_i^{LB}$ for all $i = 1, 2, ..., S$

2. check if $\underline{V}$ is feasible by solving the (FEASIBILITY) problem as stated below,

$$\text{(FEASIBILITY)} \quad d_{min} = \min_{\underline{x}} \sum_{i=1}^{S} (\alpha^* - \alpha_i(\underline{x}^i))^+$$

   where $\underline{x}^i$'s are the inventory locations of V_i. And d_{min} measures how far away α_i's are from the required α^*.

3. if $d_{min} = 0$, stop, and let $V_i^* = V_i$, $\underline{x}_i^* = \underline{x}_i$ for all $i = 1, 2, ..., S$; otherwise, let $V_i = V_i^{LB} + \Delta V_i$, $\Delta V_i = 0, 1, ...,$ $i = 1, 2, ..., S$, where ΔV_i's are the ones which result in the minimum increase of total inventory cost from the previous iteration, then go back to step 2.

Here, the total inventory cost is defined as $\sum_{i=1}^{S} C_i V_i$, where C_i is the unit inventory cost for part i.

Now, let us see why the algorithm proposed will terminate within finite iterations with fixed number of stockrooms N_{SR}. As defined above, N_{SR} is the number of different locations of $V_{i^*}^{LB}$ inventories, where part i^* is the one with $\max_i(p_i)$. Let all the parts i ($i = 1, ..., S$) have the same inventory level as part i^* at the same locations. Since the inventory level $V_{i^*}^{LB}$ and the corresponding inventory locations are obtained using the one-part, multi-stockroom model with respect to the required α^*, the corresponding α_{i^*} must be as large as α^*. So are the other α_i's with lower p_i's than p_{i^*}. Therefore, we have a feasible solution with the fixed N_{SR}. Since the value of $V_{i^*}^{LB}$ is finite, the algorithm must terminate within finite steps.

The method to solve the (FEASIBILITY) problem is similar to the one we use to solve the (LOCATION) problem in Chapter 3. Again we can enumerate all the choices of locations $\underline{x}^i$'s for small values of R and N_{SR}. For large values of R and N_{SR}, we can use APRS algorithm to obtain solutions that are close to the optimal one. Since the value of N_{SR} is usually very small for slow-moving parts, with each sample set of locations of N_{SR} stockrooms in APRS, the locations of V_i's are chosen to be the best one within the N_{SR} locations by enumeration for all the part $i = 1, 2, ..., S$.

Besides the inventory problem, we still have an assignment policy problem in the multi-stockroom case. Since we have shown in Chapter 3 that various assignment policies do not affect the satisfaction rate very much, we can simply take the part i^* with maximum probability of request p_i and solve the assignment policy problem in Chapter 3 with respect to this part. Again, the heuristic assignment policy is very acceptable here.

The numerical results shown in the next subsection are obtained using APRS algorithm and heuristic assignment policy.

4.2.2 Numerical Results

Again, with the data provided by the DEC Virginia service district, customer locations are fixed. The 140 potential locations of stockrooms are chosen to be used here with 48 hours replenishment time. The minimum customer satisfaction rate α^* that needs to be guaranteed is assumed to be 95% with 4-hour service timing constraint. The two parts (part 1 and part 2) are chosen to have arbitrary $p_1 = 0.004$ and $p_2 = 0.003$. First, using the one-part, multi-stockroom model in Chapter 3 for each part, we obtain the following results:

- part 1: $V_1^{LB} = 3$, at (2,5), (5,3) and (13,4), with $\alpha_1 = 98.17\%$,
- part 2: $V_2^{LB} = 2$, at (5,3) and (8,4), with $\alpha_2 = 96.00\%$,

Next, by applying APRS to the (FEASIBILITY) problem , the following results are obtained with 443 samples.

- part 1: $V_1 = 3$, at (4,3), (6,3) and (8,2), with $\alpha_1 = 98.01\%$,
- part 2: $V_2 = 2$, at (4,3) and (8,2), with $\alpha_2 = 95.54\%$,

Since $d_{min} = 0$, it means that the solution is feasible, and we can stop.

In the example, after obtaining the lower bounds V_1^{LB}, V_2^{LB}, and the corresponding locations, we can also use the following two heuristic algorithms to check the feasibility of the inventory levels.

The first heuristic algorithm locates stockrooms at the set of inventory locations of the part with largest inventory level. Inventories of all the other parts are located within these selected stockrooms optimally through enumeration. As in the example above, since part 1 has a larger lower bound for its inventory level, 3 stockrooms are located at (2,5), (5,3) and (13,4). The two inventories of part 2 are chosen to be located at (2,5) and (13,4) which are the best two locations out of the three in terms of the average satisfaction rate. However, the corresponding average satisfaction rate of part 2 is 92.03% which is lower than the required 95%. Therefore, with this heuristic algorithm, the set of lower bounds is not a feasible solution. But the result from APRS does show that it is feasible. In this case, this heuristic algorithm does not perform well in terms of solving the feasibility problem. However, if the required satisfaction rate α^* is 92% instead of 95%, then the heuristic algorithm found a set of locations that makes the inventory levels feasible. The biggest advantage of the algorithm is that it uses very few points (for our example, only 3 different sets of locations are evaluated for part 2) to have a quick check of the feasibility. One can use this as a start point.

Since the parts we consider here are mainly slow-moving part, once a customer is covered by a stockroom service region, his/her satisfaction rate is close to 1. Therefore, the average customer satisfaction rate is more sensitive to low values of inventory level (where not every customer is covered) than to high values of inventory level (where almost all the customers are covered). Therefore, we can propose a second heuristic policy. The algorithm first selects the part with lowest inventory level, denoted as V_{i_*} inventories, and locates the same number of stockrooms at the corresponding locations of V_{i_*}. Then the rest of $N_{SR} - V_{i_*}$ stockrooms are located one by one optimally by greedy algorithm with respect to the parts with $V_{i_*} + 1$, ... V_{i^*} inventories. For example, for the two parts we use above, two stockrooms are first being fixed at (5,3) and (8,4), the inventory locations of the two inventories of part 2. Then, one more inventory is added optimally by greedy algorithm, which results at (6,4) with average satisfaction rate being 97.87% for part 1. Thus, the algorithm shows that the set of inventory levels is a feasible solution. This second heuristic has its simplicity and better performance than the first one, but with more points (R points for each extra inventory).

As a whole, the numerical results show that the algorithm does provide a reasonably good solution to our problem. Two heuristic algorithm for the (FEASIBILITY) proposed above serve as a simple and quick check for the feasibility problem.

Chapter 5

Conclusions

In this thesis, we modeled the location and inventory optimization problem for service centers in a real world environment. Large numbers of repair parts are stored and barely moved, referred to as the slow-moving parts. Given a service district of interest, the problem is to optimize the inventory levels and locations for all the parts as well as the locations of selected stockrooms (if we have the freedom to choose them). Then with the set of selected stockroom locations, the problem tries to find an optimal assignment policy in terms of the average satisfaction rate. In Chapter 1, we formulated the problem as a combinatorial optimization problem with nonlinear constraints (the closed-form formula for average customer satisfaction rate in the constraints is difficult to obtained).

In order to find approaches to the problem, we first studied a class of simple models (e.g. the simplest one-part, one-stockroom model) in Chapter 2, where customer satisfaction rate can be written as a closed-form function of the parameters in the models. Then, we examined the effects of several main parameters to the satisfaction rate α. Numerical results have shown that higher inventory level, low probability of demand, less number of covered customers, and shorter replenishment time result in higher satisfaction rate. Also, the optimal inventory level can be obtained from the plots of satisfaction rate versus various inventory levels. Moreover, in the model where service regions of two stockrooms intersect, the optimal assignment policy for the customers served by both stockrooms is proved to be the one having those customers first call the stockroom which has less customers only in its service region.

Then in Chapter 3 we studied a more complicated model with one-part, multi-stockroom, where the simplest one-part, one-stockroom model in Chapter 2 can be used to find a lower bound of the minimum inventory level. An algorithm was developed to find the optimal inventory level and the corresponding locations. The (LOCATION) problem in the algorithm can be attacked using three proposed algorithms: covering algorithm, greedy algorithm and APRS algorithm. Numerical results showed that all three algorithms provide reasonably good solutions to the problem, and the APRS algorithm is the one with the most robust solutions under all conditions. Not surprisingly, the results also showed unit-sized inventory or no inventory at all for each part at each stockroom, which is the usual case for slow-moving parts. On the other hand, inventories are located so that there are more customers who can get service from those locations, especially for low inventory levels. Moreover, the optimal assignment policy was formulated as a combinatorial optimization problem, and can be solved using package GAMS/MINOS. Numerical results have shown that different assignment policies do not make much difference to the average satisfaction rate. Therefore, a simple heuristic policy (assigning customers with stockrooms according to the distances between them) is very acceptable.

Finally, in Chapter 4 we studied the core problem of the thesis which is the multi-part, multi-stockroom model, where the one-part, multi-stockroom model in Chapter 3 can be used to provide a lower bound for the minimum inventory level of each part individually. Based on the lower bounds, an algorithm was developed to find the minimum inventory levels and locations for all the parts, assuming that stockroom setup cost is much larger than inventory cost. Numerical results using APRS algorithm showed that the algorithm did find a reasonably good solution to the problem.

As a whole, the main contribution of the thesis is that we modeled the inventory optimization problem raised by the service centers in a real world environment, and developed various algorithms of solving the problem. Numerical results have shown impact of possible cost saving to service centers.

Appendix A

DEC Virginia Service Region Data

From: Gary L. Finckel @CSO, Logistics Planning

Attached is data for the Virginia Parts Location Problem

The following table shows the number of VAX 6000 customer contracts within the time periods specified. So, for instance, with the Waynesboro site, there are 18 customer contracts within 2 hours of Waynesboro, 72 contracts between 2 and 4 hours from the site, and 20 contracts greater than 4 hours from the site. There are 110 contracts in total.

Stockroom Location	$<$ 2hours	$>$ 2hours and $<$ 4hours	$>$ 4hours
Waynesboro	18	72	20
Richmond	43	61	6
Virginia Beach	20	43	47
Lynchburg	24	66	20
Roanoke	30	62	24

VAX 6000 Part Data

I am not able to get part failure rates for the VAX 6000 parts on a timely basis. However, the data below should suffice. This data shows the activity history for these parts.

Part	Southern Consump.s	12 months Loans	Inventory Value	Virginia Consump.s	19 months Loans	Virginia Current Stocking
T201500	157	247	$3,034	12	22	1-1-1-1-0
T2014BA	111	169	$2,805	13	14	1-0-1-1-0
H7214A	96	172	$309	5	7	1-0-1-1-0
T2011YA	88	126	$1,499	0	0	1-1-0-1-0
H7215A	77	148	$305	4	6	1-1-1-1-0
T104300	79	102	$438	5	7	2-1-1-1-0
H7206A	67	135	$426	4	6	0-0-1-1-0
T104500	45	117	$721	5	18	1-0-1-0-1
202917601	37	67	$84	3	3	1-1-1-1-0
T104600	34	110	$585	7	20	1-0-1-0-1
T201200	32	54	$779	3	9	1-1-1-1-0
122502411	30	38	$25	2	4	0-0-0-0-0
122784801	28	27	$140	0	0	0-0-0-0-0
T201100	28	49	$1,440	2	3	1-1-1-1-0
T103500	19	50	$589	2	3	0-1-0-0-0
541817201	14	18	$274	2	4	0-0-1-1-0
T104400	12	19	$747	0	0	0-0-0-0-0

The above table shows information about each part. For instance, with part T201500:

Within the entire southern area, for a 12 month period, there were 247 loans that occurred, on which there were 157 consumptions. Recall that a loan is when a part is taken out of the stockroom, while a consumption is a loan on which the part gets used to replace a bad part at the customer site. The cases when an part is loaned but not consumed (247-157=90) represent pure loans. In these cases the part might have been loaned for testing or "just in case".

The inventory value is $3,034.

The loans that occurred in Virginia, for a 19 month period, was 22. There were 12 consumptions that occurred on these loans.

The last set of numbers represent the current stocking strategy. For the T201500 it is 1-1-1-1-0, or:

Stockroom Location	Units in Stock
Richmond	1
Waynesboro	1
Virginia Bch	1
Roanoke	1
Lynchburg	0

Appendix B

Reconstruction of DEC Virginia Service Region Data

B.1 Construction of Customers' Geographic Location

The data of DEC Virginia service district shown in Appendix A is imcomplete in terms of the geographic location of every customer. The information it provides is the number of customers in each of the 2-hour and 4-hour service regions of every stockroom. Therefore we need to assign locations to all the customers so that the result is as close as possible to the given information. More specifically, the customers' locations are assigned as follows.

1. According to the geographic map of Virginia, we first construct a scaled map (1:10miles) as shown in Figure 3-2 on page 43.

2. Then we assign location coordinates (x, y) to each of the five current stockrooms according to their positions in the map as follows:

    - Roanoke (Ro) at (1.5, 2.5),
    - Lynchburg (L) at (3, 4),
    - Waynesboro (W) at (4, 7),

- Richmond (Ri) at (7.5, 5),
- Virginia Beach (V) at (13, 2.5).

3. For simpicity, we assume that a 2-hour service region of a stockroom corresponds to the area within a circle with a 30-mile radius (3 in the scaled map), and a 4-hour service region of a stockroom corresponds to the area within a circle with a 60-mile radius. By drawing the 2-hour service regions of all the five stockrooms, we obtain a map as shown in Figure 3-1 on page 42. Then, according to the given data, we assign the number of customers in each of the disjoint area as shown in Firue 3-1.

4. Finally, we assign location coordinates (x, y) to each customer according to the numbers in Figure 3-1.

In practice, the service region of a stockroom might be some irregular shape of area according to the road condition.

B.2 Estimation on Customers' Demands for the Given Parts of VAX 6000

Recall that the probability of demand p_i is defined as the probability that a customer would reqire for a part i at any given day. For the given parts of VAX 6000, the information of p_i is not given in Appendix A. However, we can use the average daily demand for part i over the given 19 month period as the estimates of the p_i.

To be more specific, we use the number of loans over the 19 months as the total number of demand from the 110 customers, denoted by L_i for part i. Assuming that there are 250 working days per year, the probability p_i can be computed as the average daily demand for part i per customer. Then, we have

$$p_i = \frac{12L_i}{19 * 250 * 110}.$$

Take part T201500 as an example. There are 22 loans in 19 months. According to the formular above, we have $p = 0.00051$. To be conservative, we round the number to 0.00057 which corresponds average one demand every 7 years.

Appendix C

Approximate Closed-Form Formula for Average Customer Satisfaction Rate

We consider a single part with unit-sized inventory at each stockroom j ($j = 1, 2, ..., V$). Again, we assume each customer can get services from at most the closest two stockrooms. According to customers' geographic locations, the following data can be obtained.

- M_{jk} = number of customers served by both stockroom j and stockroom k if $j \neq k$; note $M_{jk} = M_{kj}$,
- M_{jj} = number of customers only served by stockroom j.

The decision variable N_{jk} is defined as,

- N_{jk} = number of customers assigned with stockroom j as their primary one and with stockroom k as their secondary one if $j \neq k$; note that N_{jk} is not the same as N_{kj},
- N_{jj} = number of customers only served by stockroom j as their primary, obviously $N_{jj} = M_{jj}$.

Denote F_j as the total number of customers with j as their primary stockrooms, then

$$F_j = \sum_{k=1}^{V} N_{jk}.$$

Now the average satisfaction rate can be written as an approximate nonlinear function of $\underline{N}$ (the vector with components N_{jk}'s), denoted by $\alpha(\underline{N})$.

Since all the N_{jk} customers are identical in terms of both demand pattern and service pattern (they are all served by the same primary and secondary stockrooms), they have the same satisfaction rate, denoted by α_{jk} $(j, k = 1, 2, ...V)$. Therefore the average satisfaction rate

$$\alpha(\underline{N}) = \frac{\sum_{j=1}^{V} \sum_{k=1}^{V} N_{jk}\alpha_{jk}}{N}$$

where N is the total number of customers. Similar to the analysis in Chapter 2, define q_{jk} as the probability where a particular request A_{jk} comes from one of the N_{jk} customers, it is not satisfied. Then $\alpha_{jk} = 1 - q_{jk}$. Therefore, it is equivalent to writing out a closed-form formula for the q_{jk}'s.

Let us define some additional variables and functions used in the formulae below. At any given day, a particular request A_{jk} is called from one of the group N_{jk} customers $(j, k = 1, ..., V)$. Define for all $i, h = 1, ..., V$,

- X_{ih}: the total number of requests from group N_{ih} customers (besides A_{jk} if $i = j$ and $h = k$) during the same day as A_{jk},
- $f_{X_{ih}}(l)$: probability mass function of random variable X_{ih}, then

$$f_{X_{ih}}(l) = b(N_{ih}, l) \text{ if } i \neq j, \text{ and } h \neq k$$

$$f_{X_{jk}}(l) = b(N_{jk} - 1, l)$$

- Y_i: the total number of requests from group F_i customers (besides A_{jk} if $i = j$) during the same day as A_{jk},

- $f_{Y_i}(l)$: probability mass function of random variable Y_i, then

$$f_{Y_i}(l) = b(F_i, l) \text{ if } i \neq j,$$

$$f_{Y_j}(l) = b(F_j - 1, l)$$

- X'_{ih}: the total number of requests from group N_{ih} customers before the day A_{jk} comes,

- $f_{X'_{ih}}(l)$: probability mass function of random variable X'_{ih}, then

$$f_{X'_{ih}}(l) = b(N_{ih}, l)$$

- Y'_i: the total number of requests from group F_i customers before the day A_{jk} comes,

- $f_{Y'_i}(l)$: probability mass function of random variable Y'_i, then

$$f_{Y'_i}(l) = b(F_i, l)$$

- INV_i: the inventory at stockroom i.

Now let us first look at the simple case with 24 hours replenishment time.

Replenishment time t_r = 24 hours

For $j = 1, ..., V$, by ignoring the effect of customers whose secondary stockroom is j, q_{jj} is the probability that request A is not the first one among the requests from all the other $F_j - 1$ potential customers. Therefore

$$q_{jj}(24) \approx a_1(F_j - 1).$$

Similarly, for all $j, k = 1, ..., V$, $j \neq k$,

$$q_{jk}(24) \approx \text{prob}(A_{jk} \text{ is not one of the first two among all other } X_{jk} \text{ requests}) +$$

prob(A_{jk} is the second one among all other X_{jk} requests) $*$

[prob(the first request from N_{jk} customers takes INV_j) $*$

prob(INV_k is taken by one of the Y_k requests before A_{jk}) $+$

prob(the first request from N_{jk} customers cannot take INV_j)] $+$

prob(A_{jk} is the first one among all other X_{12} requests) $*$

prob(INV_j is taken by one of the requests from $(F_j - N_{jk})$ customers) $*$

prob(INV_k is taken by one of the Y_k requests)

$$\approx a_3(N_{jk}-1) + a_4(N_{jk}-1)\,[a_2(F_j - N_{jk})a_1(F_k) + a_1(F_j - N_{jk})] + a_2(N_{jk}-1)a_1(F_j - N_{jk})a_1(F_k).$$

Replenishment time t_r = 48 hours

For $j = 1, ..., V$, by ignoring the effect of customers whose secondary stockroom is j,

$$q_{jj}(48) \approx 1 - b(F_j - 1, 0) + b(F_j - 1, 0)a_1(F_i - 1).$$

And for all $j, k = 1, ..., V$, $j \neq k$,

$$\begin{aligned} q_{jk}(48) \approx\ & 1 - b(M_{jk}, 0) - b(M_{jk}, 1) + \\ & b(N_{jk}, 1)b(Nkj, 0) * [1 - b(F_k - N_{kj}, 0) + b(F_k - N_{kj}, 0)a_1(N_{jk} + F_k - 1)] + \\ & b(N_{jk}, 0)b(Nkj, 1) * [1 - b(F_j - N_{jk}, 0) + b(F_j - N_{jk}, 0)a_1(N_{kj} + F_j - 1)] + \\ & b(M_{jk}, 0) * \{ [1 - b(F_j - N_{jk}, 0)] * [1 - b(F_k - N_{kj}, 0)] + \\ & [1 - b(F_j - N_{jk}, 0)]b(F_k - N_{kj}, 0)a_1(N_{jk} + F_k - 1) + \\ & [1 - b(F_k - N_{kj}, 0)]b(F_j - N_{jk}, 0)a_1(F_k - 1) + \\ & (F_j - N_{jk}, 0)b(F_k - N_{kj}, 0)q_{jk}(24)\}. \end{aligned}$$

Bibliography

[BeBL] O. Berman, D. Bertsimas and R. C. Larson, (1992), "Locating Discretionary Service Facilities II: Maximizing Market Size, Minimizing Inconvenience", Operation Research Center OR271-92, MIT.

[Cros] J. D. Croston, (1974), "Stock Levels for Slow-Moving Items", *Oper. Res. Quart.*, 25, 123-130.

[NeWF] G. L. Nemhauser, L. A. Wolsey and M.L. Fisher, (1978), "An Analysis of Approximations for Maximizing Submodular Set Functions", *Mathematical Programming*, 14, 265-294.

[Schu] C. R. Schultz, (1989), "Replenishment Delays for Expensive Slow-Moving Items", *Management Science*, Vol. 35, No. 12, 1453-1462.

[Tang] Z. B. Tang, (1992), "Adaptive Partitioned Random Search to Global Optimization", Preprint.

[Will] T. M. Williams, (1984), "Stock Control with Sporadic and Slow-Moving Demand", *J. Oper. Res. Soc.*, 35, 939-948.