QUANTITATIVE MODELS FOR POLICE PATROL DEPLOYMENT

by

KENNETH RICHARD CHELST

B.A., Yeshiva College
(1970)

M.S., New York University
(1972)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF
PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1975

Signature of Author..
Alfred P. Sloan School of Management,
August 22, 1975

Certified by..
Thesis Supervisor

Accepted by..
Chairman, Departmental Committee on Graduate Students

QUANTITATIVE MODELS FOR POLICE PATRCL DEPLOYMENT

Kenneth Richard Chelst

Submitted to the Alfred P. Sloan School of Management
on August 22, 1975
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

A primary task of the police administrator is to make the most efficient and effective use of his manpower. In addition because police provide a public service, an important issue in the deployment of police personnel is that each segment of the community be allocated an equitable share. Models are presented that focus on the spatial component of these deployment goals.

Interactive computer models are formulated to aid a police planner design patrol sectors which represent a good balance among often conflicting objectives (e.g. response times, workloads). The models guide the decision maker iteratively through a series of alternative sector designs while providing him with information about a spectrum of performance measures. An integral part of the system is a set of algorithms that can modify an initial sector design to greatly improve imbalances in either workload, preventive patrol coverage or response time. Computational experience is presented.

A second set of models is presented which focuses on the more effective deployment of randomly patrolling police units as measured by the probability of intercepting a crime in progress. The discussion begins with a presentation of a basic search theoretic model of police patrol which is used to calculate the probability of intercepting a crime. As part of the analysis of the model's input parameters, we discuss the critical need for a police patrol related data base and outline some of its salient features (e.g. duration and observability of various crime types). Then, using the model we explore the differences between overlapping and non-overlapping patrol sectors.

The development of methodologies for deploying patrol units proceeds in several stages. First, we analyze the impact on the classical search theory allocation problem of various characteristics of crimes (e.g. random arrival, short duration, multiple independent targets). A continuous time differential equation model of search and detection provides the vehicle for carrying out much of this analysis. Optimal solutions for a number of classical search problems are presented including simple closed form expressions for determining if a region should be excluded from the search. The main result of this analysis, however, is the generation of a number of important insights which simplify the development of algorithms for deploying police.

An algorithm for deploying a tactical patrol force (i.e. limited or no responsibility for calls for service) is presented. The measure of effectiveness that is used is the weighted (a user specified index that weights the various crime types) probability of intercepting a crime. An essential component of the algorithm is its ability to perform sensitivity analysis on the various input parameters.

Lastly we outline the development of an algorithm for effectively allocating the patrol time of standard patrol units. Once again the measure used is the weighted probability of intercepting a crime. The discussion closes with a description of what questions need to be answered before a total model of police patrol can be developed.

Thesis Supervisor: Richard C. Larson

Title: Associate Professor of Urban Studies and Electrical Engineering

ACKNOWLEDGEMENTS

To my advisor, Richard C. Larson, goes my appreciation for his guidance throughout the course of this work, for his continual emphasis on the real world side of applied research and for his patience during the evolution of my literary style. I would also like to express my gratitude to Philip M. Morse for his weekly consultations in which he volunteered to serve as a 'unique' sounding board whose 'reflected wave had much more energy than its incident wave'. Many thanks also to Arnold Barnett for his wise witticisms.

A special thank you to Keith Stevenson for his editorial assistance, to Rick Jarvis for sharing his dissertation writing experience as to all the "lent ears" at the Operations Research Center and IRPP. Fondly will I remember my jaunts with hilarious Harold who helped humble my moments of harassment.

To the Sabbath, a day of Rest, for its great redeeming value to an overworked youth and to my son, Dov, for his consideration for a sometimes overbearing father, I would like to express my deepest gratitude.

And above all, I owe a debt of gratitude to my wife, Tamy, that can never be repaid. Her boundless energies in typing all drafts of this opus all within the span of a few short weeks were nothing short of amazing; her editorial comments, her belief in my ability to write when others scoffed and her constant encouragement eased the way when there seemed to be an abyss between me and the next sentence. Without her willingness to sacrifice (vacations that never were, a promising career put in abeyance), this thesis would still be only a dream (or nightmare).

Most of all, I am grateful for her sharing my life, sharing the burdens which otherwise would have been unbearable, and for giving me so much of herself. I would therefore dedicate with love this thesis to her though, in fact, it is already at least as much hers as it is mine.

Support for this work was provided by the National Science Foundation under grant GI38004 and by the Harry J. Carmen Dissertation Fellowship of the Educational and Cultural Trust Fund of the Electrical Industry.

Table of Contents

List of Tables

List of Figures

Figure		Page

CHAPTER 1

INTRODUCTION

1.0 Background

1.0.1 Role Definition

"There's a holdup in the Bronx,
Brooklyn's breaking out in fights.
There's a traffic jam in Harlem
That's backed up to Jackson Heights.
There's a scout troup short a child;
Khruschev's due at Idlewild.
Car 54 where are you?"

The above excerpt from the lead song of a popular 1960's television series offers a sprinkling of the manifold activities of its heroes, two New York City patrolmen. Despite changes in the Kremlin and in New York's international airport during the past decade, the policeman of the seventies is still confronted daily with a kaleidoscope of responsibilities. Given the multiplicity of roles the police officer, and more specifically the patrolman, must assume, it is not hard to recognize the complex problem a decision maker faces in attempting to maximize the efficiency and equity with which the patrol force operates.

On the most basic level decisions need to be made as to which tasks fall within the proper purview of a policeman's role. Which tasks can be handled by civilians within the department and which responsibilities could be delegated entirely to other public agencies? Should police be the prime providers of emergency medical transport? Should a patrol officer be summoned if a cat is caught in a tree, an individual locked out of his home, or even a child lost? Should traffic control be the responsibility of police?

These same role identification questions can be asked both on a macroscopic level, police in general, and on a microscopic level, the patrol officer. Should time be taken from preventive patrol activities in order to respond to a two day late complaint of stolen property for the sole purpose of filing a perfunctory report on the incident? Should a city tie down more than half of its force two or three times a day to guard school crossings? There are no simple answers to these questions. And as the recent commission on Standards and Goals [8] suggested, the answers will and should be community dependent. The response should reflect an amalgamation of community expectations, needs and priorities, while cognizant of the limitations on police resources. An attempt at expanding the duties of a patrol officer will of necessity encroach on already existent responsibilities. All too often, the task most easily encroached upon is preventive patrol, a task which many police feel prevents crimes by posing a threat of apprehension. Since the amount of preventive patrol time is usually determined by how much time is 'left' over from other activities [4], this activity is very easily nibbled away by expanding the patrolman's role. More importantly, in the current situation, with our cities suffering from financial crises along with the rising crime rates, the questions posed here concerning role delineation have been transformed into pragmatic budgetary issues. Can the cities afford to have police provide all the services they presently do?

1.0.2 Performance Measures

Defining clearly the policeman's role in the community is, however, only one of a series of steps that should precede the process of allocating available resources. An obvious second issue to resolve is the specification of measures to be used in assessing the effectiveness of the police in performing their assigned functions. Ideally these should measure output. How have specific crime levels changed as a result of fielding more patrol units? Are the various segments of the community satisfies with the delivery of police services? Unfortunately it is often extremely difficult to measure directly police effectiveness. Instead a surrogate is often adopted; in some instances it will be a process measure which describes how the system performs (e.g. response time) rather than what its impact is. The measure of response time is purported to reflect citizen satisfaction and patrol coverage or arrests to reflect the efficiency with which police deal with crime. In other instances the surrogate may be an input measure describing, for example, the number of patrol units on duty at any one time. An alternative approach to measuring police performance is to measure the efficiency of the system. Can the same level of services be provided at a lower cost (e.g. introduce civilian dispatchers) or can more services be provided without increasing the cost?

Because police provide a public service, measuring of performance can not be limited just to issues of overall effectiveness and efficeincy. Each section of a community has a right to receive its equitable share of police services. In a recent court case [1], one segment of the Washington, D.C. community sued the city and required it to prove that police services had been distributed equitably. Equity, however, is not always easily defined or measured. What constitutes an equitable distribution of patrol coverage: the same number of patrol hours per street mile, per part I crime, per what?

Problems of performance measurement are not however limited to systemic issues. The performance of an individual police officer is, in many ways, even more difficult to assess [2, 7]. Does a particular officer use too much force when making an arrest? Is he efficient in his preventive patrol activities?

Even with the development of perceptive measures of performance, the multiplicity of goals and measures complicates the decision-making task. Individual goals and performance measures do not stand isolated. Decisions must be made as to what to emphasize and tradeoffs must be analyzed either because of limited resources or because of conflicting goals or both. Equity criteria may suggest one form of deployment (e.g. dispersal) while efficiency criteria warrant another (e.g. crime concentrated).

1.0.3 Relationship Between Input and Output

Having determined an appropriate multifaceted role for the police and patrol officer, and developed measures to evaluate the degree to which the goals are achieved, the decision maker can now begin the task of allocating his limited resources to achieve specified goals. He must, of course, have the capability of determining how many patrolmen are needed to achieve a specified response time and how that response time impacts on crime levels (in terms of crime prevention and crime solution after the fact). He also needs to determine the level of preventive patrol coverage needed to produce a 5, 10 or 20% reduction in the street crime level. Determining the link between various input and output measures would at least allow the decision maker to begin analyzing alternatives as to their relative effectiveness.

The difficulties involved in uncovering the relationship between input and output in the area of crime are in part reflections of the fact that the problems the police attack are not only police problems but, on a larger scope, societal problems. Just as the health of an individual does not necessarily depend on the type of hospital closest to him, the level of crime in an area is not a direct function of only the efficiency of local police patrol. The causes of crime are complex, the police being but one component of a criminal justic system designed to control and react to incidences of crime. For some crimes the police can pose almost no deterrent effect. Even for such crime categories as street crime where patrol has potential impact, the effectiveness of patrol is inseparable from the courts and corrections. Doubling the interception probability and solution rate of crimes can not take criminals off the street or pose a deterrent threat if criminals are not prosecuted and imprisoned. On the other hand, a relatively small increase in interception or solution could have a long term impact if all convicted criminals were sentenced to life imprisonment or successfully rehabilitated while in prison. (Obviously, we are not suggesting the former nor realistically expecting the latter).

1.0.4 Interactions

We have outlined above three separate steps: (1)definition of role; (2)specification of performance measures; (3)determination of relationship between input and output; that a rational decision maker should take in allocating resources. However, we also recognize the interaction between these steps. The specification of performance measures requires a realistic appraisal of how police can affect a specific situation. For example, in measuring the impact of various patrol strategies, the performance measures used should relate to only a delineated fraction of all part I crimes, those observable by a passing patrol car.

Another decision to expand the role of police to include, for example, responding to emergency medical situations, should be sensitive to the implications of this decision. What amount of expenditure in police resources will be required? How will taking on these added responsibilities detract from police performance in their other duties?

If the above discussion has conveyed a feeling for the tremendous array and spectrum of multifaceted problems facing the police decision maker, then it has achieved its goal. The reader should therefore not expect the following pages to contain a single all encompassing model of police patrol which can be used to deploy patrol units to optimize whatever goal is specified. The perceptive reader can not realistically expect a methodology for deploying patrol units to eliminate or even reduce significantly serious observable crimes. What will follow is a series of models that examine different aspects (performance measures) of patrol and which facilitate the exploration for more equitable and efficient deployment strategies.

1.0.5 The Cost of Police

Two and one half billion dollars a year was the estimate of the cost of police given by the President's Crime Commission in 1967 [10]. A lion's share of this total went to pay some 420,000 people working for approximately 40,000 separate agencies. With spiraling inflation since then, the cost is no doubt significantly higher. Since 1959 the cost of police in 12 of the nation's largest cities has more than tripled [9]. Presently the cost of police in those cities is running at approximately $65 per capita. Of the total departmental budget police patrol consumes by far the largest share. In many cities patrol costs represent 40 to 50 percent (sometimes even more) [4] of the budget. With the <u>average</u> starting salary for a policeman (in those 12 cities) now at over $10,100, the annual direct cost (excluding pension) of fielding round-the-clock one two-man patrol car is over $100,000.

In the light of the budget crunch faced by many cities in recent years, the rising cost of police is forcing city administrators into some very difficult decisions. In New York City more than one thousand patrolmen were recently laid off as part of a massive economy drive [3]. With the above discussion and figures as background, it is all too obvious that there is a critical need to deploy police in the most effiecient way possible. It is this particular issue (efficient allocation of patrol) that many of the models that will be presented here were designed to address.

1.1 Objectives

Despite the complexity of the police role, two rather simple assertions can be made: (1)The police provide a public service; (2)A primary goal of police patrol is to suppress crime. The first assertion suggests that equity is an important criterion for deploying police. Thus the first set of models that we will present focuses on developing equitable precinct level deployment strategies through the modification of sector boundaries. The second class of models, based on search theory, focuses on observable street crimes and is used to deploy police so as to maximize the weighted (by crime type) probability of intercepting a crime. One underlying theme common to the development of both models is that they have the flexibility to incorporate local considerations and priorities when applied to a particular situation.

The sector redesign models are part of an overall interactive system geared towards generating an equitable sector configuration. However, rather than using some one absolute measure of equity, the system allows the user to decide what, for his environment, is most equitable. He can measure equity in terms of only one performance measure (e.g. workload, preventive patrol coverage, or travel time) or he can explore tradeoffs between these particular measures to achieve an intermediate range of imbalances in each of the parameters.

A key component of the above system is the hypercube queuing model developed by Larson [5, 6]. It is a probabilistic model which is used to estimate the operating characteristics of a particular sector configuration, generating such statistics as each unit's workload, precinct wide average travel time, sector travel times, and frequency of intersector dispatches. An interactive system iteratively guides the user by offering several alternative modifications to the present configuration, which by hypercube estimates, will reduce present inequities in certain performance measures.

The development of the search theoretic models of police (interceptive) patrol starts with the introduction of several modifications of the earliest models of search and detection developed by Koopman. Using the descriptive search model as a basis, we carry out in two stages an extensive analysis to develop methodologies for efficiently allocating police patrol effort. The first stage determines characteristics of optimal search strategies for the general class of targets that arrive and depart randomly. Using the results of this stage, the range of alternatives that need be analyzed in order to find an efficient allocation is narrowed significantly. The major consequence of these insights is that the allocation, for example, of a tactical patrol force or the free time of an individual patrol unit can be reformulated so that straightforward heuristic techniques can be applied to develop a good allocation of patrol effort.

Although the search theory models consider only one measure, the weighted probability of interception, local conditions can still be incorporated. The weights can be used to capture the subjective importance placed on each crime type, which may vary from locale to locale.

1.2 Outline of the Chapters

Chapter II contains a review of the relevant literature and is divided into four sections. The first section reviews other models presently used by police in constructing sectors as well as presenting a brief description of the development of the hypercube model. Next the general search theory literature is reviewed as to its applicability to problems in which targets arrive and depart randomly. Then the literature applying search theory to police patrol is analyzed in detail. Lastly since an obvious implicit assumption of all of the search theory models is that patrol can impact on crime, relevant data from a number of experiments will be presented and discussed.

The third chapter describes an interactive system for sector redesign. Because the decision to make the system interactive greatly influenced many aspects of the system's development, the motivation for choosing that orientation is discussed. Following that, the superstructure of the system is described, outlining how the system functions in guiding the decision maker to an equitable configuration. The workhorse of the system, however, is a set of individual programs which can focus on a particular performance measure (e.g. workload, preventive patrol coverage or travel time). By altering iteratively the sector configuration, they attempt to improve on an imbalance in that measure. These programs are analyzed in detail both in terms of the algorithms used, as well as in terms of the underlying properties each program attempts to exploit. A number of applications are also included.

In the course of presenting this system one tangential issue that is discussed is the non-equivalence of balanced workloads and balanced preventive patrol coverage. Finally we introduce a method for attacking another related sector design problem. This involves taking an existing n sector configuration and assigning the n sectors to less than n patrol units in a way that minimizes the workload imbalance.

All of the succeeding chapters (four through eight) discuss the search theoretic models of police patrol. Chapter IV begins with a discussion of the basic model, including a comparison of overlapping and non-overlapping patrol sectors. Following the initial presentation we proceed to analyze in qualitative terms how randomness in the arrival and departures of the targets (i.e. crimes) impacts on optimal patrol strategies. This analysis is carried out in terms of classical search theory by focusing on the phenomenon of diminishing return. We show how randomness slows the process of diminishing return and when combined with the short duration of crimes tends to generate optimal strategies which limit patrol to only a small section of an entire region.

Chapter V continues the discussion of the impact of randomness in the arrival and departure of targets on optimal search strategies but with a quantitative focus rather than a qualitative one. Using a differential equation model of search, we analyze for a number of situations cyclical search strategies involving two regions. As part of our analysis we present simple analytic expressions that specify when to search only one of the two regions.

The first example that is analyzed involves a patrol dividing its search efforts equally between two identical regions (i.e. same crime rate and detection rate). The question that we resolve is "How frequently should the searcher change regions?". The finding was that the average expected number of crimes in progress increases as the length of stay in a region increases; the magnitude of this is also measured. The second example more realistically models police patrol as it assumes that time is lost (because of travel) from the search process when switching between regions. Again a method is presented for finding how long each region should be searched before traveling to the other region. We show that because crimes are of short duration even small travel times between regions make switching back and forth between regions an inefficient policy. The ramifications of this is that in allocating search effort among regions in a realistic patrol environment, solutions which require a patrol unit to travel between two even relatively high crime regions should be avoided. This conclusion played a key role in the development of the algorithms presented in Chapters

VI and VII.

The last problem analyzed with the differential equation model involves two regions of equal size but with differing crime rates. Two sets of conditions are determined under which it is preferable to search only one region. One condition describes when to search only one region if no minimum duration of a search of each each region has been specified. The second condition also specifies when to limit search to the high crime region but this time as a function of the minimum duration imposed on a visit to each region.

Chapters VI and VII build on the results of the two previous chapters to develop algorithms for (1)deploying a tactical patrol force and (2)finding the best region in which a standard patrol unit should concentrate its patrol. The two algorithms, besides addressing specific problems, are also meant to be representative of the potential application of search theory to a broad spectrum of crime related patrol allocation issues.

The tactical patrol force algorithm of Chapter VI allocates a specified number of patrol units among a set of high crime regions. In the process of finding the optimal strategy only solutions which allocate an integer number (or zero) of patrol units to each region are considered because, as was noted earlier, the time lost in travel between regions will, in general, more than outweigh any benefits that might be derived from switching regions.

The application of the algorithm to a number of examples is presented and is partially intended to point out some of the data that is critical to developing effective patrol strategies. For one thing, it is not sufficient to know just the total observable crime rate of a region but it is also necessary to know the breakdown by crime type. However, an equally important, but less recognized, data need is the observalbe duration of each crime type (either an average or preferably a probability distribution) and an estimate for the probability that a passing patrol car will detect the crime when it is potentially observalbe.

The tactical patrol force algorithm is flexible in that it is not limited by the specific form of the probability distribution function chosen to describe the duration of a crime. In addition the objective function which is optimized allows for different weights to be assigned to each crime type to reflect the importance associated with intercepting that type of crime. Lastly an integral part of the algorithm is its capability of performing sensitivity analysis on each of it input parameters.

Chapter VII is intended to be exploratory and offers one approach to finding the optimal patrol region for a standard patrol car (i.e. a patrol unit that has responsibility for answering calls for service). The objective function to be maximized is the same as in the previous algorithm, the weighted probability of intercepting a crime. In addition some of the issues relating to combining two patrol sectors into a single larger sector enabling overlapping patrol are also discussed.

The last chapter presents a summary of the most significant results of our search theoretic modeling. Chapter VIII also introduces some specific conclusions implied by our analysis about the potential for developing more effective patrol strategies. However the major focus of the chapter is to outline a number of important issues that need to be resolved in order to implement models such as ours. It has already been noted that there is a need for developing a data base more relevant to patrol. One other specific issue discussed is the need to design carefully experiments whose major purpose is to determine how saturation patrols disperse crime in terms of both the magnitude of the effect and how it varies over time. Experiments of this type, because of focus, are really distinct from those that attempt to measure the reduction in crime in the saturated region while monitoring some of the side effects (dispersion phenomena) which reduce the actual impact of the saturation patrol effort.

REFERENCES 1

1. Block, Peter B., Equality of Distribution of Police Services--A Case Study of Washington D.C., The Urban Institute, Washington, D.C., February 1974.

2. Cohen, Bernard and Jan M. Chaiken, Police Background Characteristics and Performance, New York City Rand Institute, D.C. Heath and Company, Lexington, MA, 1973.

3. "How New York City Lurched to the Brink", Time, Vol. 105: 16, June 16, 1975.

4. Larson, Richard C., Urban Police Patrol Analysis, MIT Press, Cambridge, MA, 1972.

5. Larson, Richard C., "A Hypercube Queuing Model for Facility Location and Redistricting in Urban Emergency Services", Computers and Operations Research, Vol. 1, No. 1, March 1974.

6. Larson, Richard C., Urban Emergency Service Systems: An Iterative Procedure for Approximating Performance Characteristics, The New York City Rand Institute, R-1493-HUD, Janurary 1974, to appear in Operations Research.

7. Marx, Gary, Alternative Measures of Police Performance. Working Paper WP-12-74. Massachusetts Institute of Technology Operations Research Center, Cambridge, MA, October 1974.

8. National Advisory Commission of Criminal Justice Standards and Goals, Police Report, U.S. Government Printing Office, Washington, D.C. 1973.

9. Odoni, Amadeo, Vol 1, Chapter 3, Innovative Resource Planning in Urban Public Safety Systems Final Report to National Science Foundation, to appear in 1976.

10. President's Commission on Law Enforcement and Administration of Justice, The Challenge of Crime in a Free Society, U.S. Government Printing Office, Washington, D.C., 1967.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

Before proceeding with the actual review of the relevant literature, it is important to understand first its intended purpose. The literature review does not purport to describe the state of the art in police patrol deployment. For a discussion of this scope, the reader is referred to Urban Police Patrol Analysis [25] by Larson and to a critical review of policy related research written by Gass and Dawson [16]. The focus here will be instead on models and methodologies which relate either to sector design or to the allocation of patrol effort to intercept crimes. However, even for this more limited set of models, this presentation will have as its only goal to place in perspective the research presented in later chapters with that of earlier work in the area. This issue will be of special importance when we discuss search theory since an initial question we had to resolve was "Is the existing general search theory literature directly applicable to police patrol issues?". Unfortunately, except for some of the more basic results, the answer was negative. It is a point we will explore initially here and will expand upon later in Chapters IV and V. With the above disclaimer and statement of purpose in mind we proceed with the review.

The literature to be discussed falls into four categories:

1. Sector Design--Proscriptive and Descriptive Models
2. Search Theory--General
3. Search Theory--Applied to Police Patrol
4. Patrol's Impact on Crime

2.1 Sector Design--Proscriptive and Descriptive Models

The redesign of a precinct's outdated sector configuration is a problem police administrators must tangle with periodically. The methods used range from an experienced patrol officer attempting to eyeball an acceptable configuration to the more sophisticated techniques involving computerized data analysis of calls for service and crime data. This data analysis is then used as input into a second set of programs to generate a sector design which optimizes one particular performance measure. Our interest is naturally in methods of the latter type. Gass [15], Heller et al [19] and Dean [9], to cite a few examples, have used techniques developed for political redistricting (set covering and transportation/heuristic) to produce workload balanced sector configurations. Incorporated in their approach, typically are constraints that force the sectors to be contiguous and compact rather than elongated. In Detroit [10] a different approach was taken to produce balanced workloads which in ways is analogous to some of the models described in Chapter III. For the existing configuration the individual workloads were estimated by summing up the work generated by each census block within a sector. Then the configuration was iteratively modified by switching groups of census blocks between over and underutilized units.

Common to these static models of workloads is the assumption that the workload of a patrol unit and the workload generated by its associated sector are equivalent. Larson [28], however, has shown this not to be true and in a recent study in New Haven [8], the internally generated workload underestimated a patrol unit's workload by as much as 60%.

Bammi [3] and a study in San Jose [1] both focus on response time as the performance measure of major concern in designing beats. Although each presents a different method for minimizing precinct wide response time, neither method is based on a solid theoretical foundation. In Bammi's analytic model a number of independence assumptions are made about the operation of each of the patrol units. It, however, has been shown [27] that in a queuing system (a group of patrol units deployed to answer calls for service can be viewed as a queuing system), the state (answering a call for service or on patrol) of a particular server is not independent of the rest of the system. The San Jose study, on the other hand, claimed that by minimizing the frequency of intersector dispatches the precinct wide response time is minimized. Although a justification for this assumption is given, no proof is offered nor is it clear that the method they used guarantees that even the frequency of intersector dispatches is minimized.

Besides the questions of internal validity that we have raised there is a fundamental difference in orientation of the above models and the system we describe in Chapter III. Each of the above models focuses essentially on only one parameter, with the user having at most limited control, through the use of constraints, over what the program will generate as its suggested configuration. Police patrol, though, is a multifaceted activity with multiple and often conflicting objectives. As such, an approach to sector design should be oriented more towards tradeoffs between performance measures than optimizing a single performance measure without regard to the impact on any of the others. The approach described in Chapter III contrasts sharply with this as the goal of the system is to <u>guide</u> the user towards what he judges to be an equitable configuration, while explicitly taking into account tradeoffs between balanced workloads, preventive patrol coverage, response time, etc.

An obvious prerequisite for the development of a system similar to that of Chapter III is the existence of a methodology (descriptive model) for calculating the various performance measures. Two methods presently available for evaluating a particular sector configuration are Larson's hypercube queuing model, which we use[1] [27], and a simulation, the prototype of which was also developed by Larson [25]. Although, conceptually the interactive approach described later could have been structured about a more flexible simulation, the speed of calculation was too critical a parameter since the system iterates through a series of alternative designs. Thus, the more rapid hypercube model was chosen, instead, as the foundation for the system's development.

The essential difference between using just the hypercube or simulation model and the structured system of Chapter III is that the former are purely descriptive models. They can, therefore, provide no guidance on how to improve on the present configuration other than by pointing out inequities. Instead the user must develop on his own a series of potential alternatives which he then evaluates individually with one of the models. With a structure superimposed on these models, it is possible for the computer to generate alternatives which modify the sector configuration in the direction the user has specified.

To summarize, the methodology to be presented here has the flexibility of focusing on any one of a series of performance measures, while encouraging the user to consider tradeoffs between measures. In addition the decision making process is left in the hands of the user. This is an advantage which should not be underestimated as it significantly enhances the system's implementability, a point which we elaborate on later.

2.2 Search Theory: General

As was stated in the introduction to this chapter, the literature reviewed will include only those papers directly related to the problem of maximizing the probability of intercepting randomly arriving and departing targets. Specifically we will review the foundations of search theory developed by Koopman [24] as well as the companion work of Charnes and Cooper [7]. Following that we will focus on the search literature that looks at periodic cyclical search strategies since the random arrival and departure of multiple independent targets often necessitates optimal strategies of this type. For a more comprehensive discussion of the literature the reader is referred instead to Morse [32] and Moore [30].

No discussion of search theory can be really complete without some introductory remarks about Koopman's pioneering work in the field. In a series of three articles [22, 23, 24] Koopman looked at several aspects of the search and detection problem and outlined, first, the basic negative exponential model with its fundamental characteristic of diminishing return.

> "As the available search effort increases the probability of interception increases less than linearly."

(It is this simple exponential model, coupled with a heuristic procedure to find a more efficient allocation of police patrol, that forms the backbone of the algorithms developed in Chapters VI and VII.) Building on this model, he then proceeded to develop a methodology for finding the optimal allocation of search effort over a region. In his example the target is assumed stationary and is located somewhere in a region A with a known probability density function continuous in the region. Charnes and Cooper [7] later developed an algorithm for the discrete analog to Koopman's problem.

One obvious difficulty in applying either Koopman's or Charnes and Cooper's methodology to police patrol is that their methodologies do not include travel time between points or regions and could, therefore, generate unimplementable solutions. Secondly their approaches do not lend themselves to the incorporation of classes of targets (crimes) each with a different mean duration. However, of greater significance is that the randomness in a crime's arrival and departure and the independent arrival of crimes limit the potential usefulness of their methods. Using the concept of diminishing return we explore in Chapter IV exactly why this is so. For the moment, though, it should be noted that randomness and multiple independent crimes require that the optimal solution specify not only how much effort to allocate to each point or region (which is all the above methods can determine) but also how should that effort be sequenced. This leads us to our next area of discussion, the literature involving sequential search strategies.

2.2.1 Sequencing Search Effort

Blachman and Proschan [5] analyzed the optimal sequencing of search effort among a series of regions into which targets arrive randomly atvarying rates. However, in contrast to our problem, once the target has arrived it never departs. The objective function to be maximized is a gain function which is a non-increasing function of the delay between a target's arrival and the beginning of the detecting look. Only cyclic searches are considered. Although a method for determining almost optimal strategies is presented, Barnett [4] points out that the theorem for determining which regions should not be searched is meaningless when targets are allowed also to depart. Thus this approach provides no insight in the police context into the characterization of when not to search a region. It is this specific question which the differential equation model of Chapter V successfully addresses, by generating very simple analytic expressions to determine when to exclude a region.

Gilbert [17] and later Kisi [21] analyzed a two-box search problem in which the objective is to minimize the expected length of time until detection. The first result Gilbert obtains is that if the target is equally likely to be in either box, the optimal policy is a limit strategy which involves switching instantaneously from one box to the other. We obtain, in Chapter V, an analogous result for randomly arriving and departing multiple targets using the differential equation model. Gilbert then proceeds to find an optimal strategy when non-zero switching times are included. However, because there is only a single target (and thus continuous unbounded diminishing return) no matter how much time is lost in switching regions, diminishing return eventually makes switching regions profitable. In Chapter IV we show that as a result of the arrival and departure of multiple targets diminishing return in the police context has a lower limit, and as a result it may pay to stay in only one region. Using the differential equation model this fact is confirmed and we are able to quantify the limited range of switching times for which cycling between regions is advantageous.

Two papers which more directly relate to the work presented here are by Moore [31] and Barnett [4]. Moore analyzed the impact on optimal search strategies of random visibility of a target both in terms of duration and initiation. An interesting aspect of his presentation is that it compares optimal strategies which take account of these factors with a blind application of Koopman's model disregarding all of these characteristics of the target visibility. In general for small amounts of search effort (crimes are relatively short) the two strategies often provided significantly different results. Although Moore suggests the possible analog between police patrol and some of his examples, the police patrol situation is really not comparable. One obvious reason is that his examples involve one target arriving somewhere in one of n regions. However, in the police context, each region is generating crimes independently of the other, which as we discuss in Chapter IV, tends to increase the likelihood that search should be limited to a single region. In addition we have noted that optimal strategies for targets arriving and departing in different regions require the specification not only of how much effort to allocate to each region but also of the sequence of that effort. Moore's paper does not discuss sequencing because the problem he analyzes involves only a single target.

There is, however, a more subtle, and in many ways, more significant limitation in applying Moore's results to police patrol. He considers the problem in which the target's visibility begins at a random instant (equivalent to a crime beginning) by introducing a probability distribution function for the target's appearance. However, this distribution function is not calculated relative to the start of the entire search process but relative rather to the search of each particular region. Thus his distribution function defines the probability of the target appearing in region i five minutes after the searcher has begun searching region i or appearing in region j three minutes after the searcher has begun searching region j, independent of how much time has elapsed before entering region i or region j. Thus conditioned on the target appearing in region i, Moore's example assumes that the target's time of arrival depends of the time at which search is begun in region i. However, the time at which a crime would occur in region i is more likely to be dependent on the searcher leaving region i and beginning search in region j than the time at which the searcher enters region i. We have discussed this point in somewhat greater detail because in stating the problem to be solved, Moore does not seem to make the above assumption; however, in the equations used to find the optimal strategy he, in fact, does.

Barnett [4] considers the situation in which targets are arriving in a Poisson manner with the rate varying by region. In addition a probability distribution is assigned to the duration of the target. One restrictive (in terms of general applicability) assumption of the problem he models is that a search of any region discovers, with probability one, all targets still present in that region. Barnett first proves that the optimal strategies are cyclical, a fact we use in applying the differential equation model. Then for the two region example he finds analytically the optimal sequence of search. With this analysis he then generates a sufficient condition for excluding regions from the search in the N region allocation problem. Although the model incorporates some of the characteristics of a search for crimes, it can not, at present, be directly applied to finding optimal deployment strategies in a realistic police environment.

The above discussion is by no means meant as an exhaustive survey of the literature that relates to the search problem we consider. We have chosen, for our discussion, papers that are representative of previous analyses of some of the different individual characteristics of the search process involved in police patrol. However, none of these models captures all of the characteristics

1. multiple independent targets
2. random arrival of targets
3. departure of targets
4. different target types (mean duration)
5. time lost in travel between non-contiguous regions

In our review we have attempted to show how in each model or methodology the elimination of one of the above characteristics makes it difficult to generalize its results to police patrol.

In Chapter IV we begin by first developing a qualitative understanding of how each of the above characteristics affects the optimal solution. Then Chapter V introduces a differential equation model which can incorporate all of them. However, even in our development of the differential equation model, the model is not offered as part of an algorithm for deploying police patrol because of the present computational difficulty in analyzing more than two regions. Instead it is used to provide a number of generalizable insights specific to the police environment (crimes of short duration, rapid rate of completion relative to rate of detection). These insights form the basis of a number of simplifications essential to the development of the patrol allocation algorithms of Chapters VI and VII. Thus, in effect, we have come full circle. We started with Koopman's most basic search model of random patrol, which calculates the probability of intercepting a target. Then we proceeded to show that a number of existing optimal allocation methodologies can not be applied to the police patrol. This motivated the development of the differential equation model, which then allowed us to develop patrol allocation algorithms using essentially the basic Koopman (not his allocation model) model for calculating the probability of interception when patrol is random.

2.3 Search Theory: Applied to Police Patrol

A number of earlier works have suggested and discussed the applicability of search theory to police patrol. A natural question is "How does the present work build on the earlier research and in what areas does it differ?". Before proceeding with the discussion there is an issue of definition that should be clarified. All search theoretic models of patrol, to date, are models of interception patrol and not of preventive patrol. They do not purport to predict the probability of preventing a crime only the probability of intercepting it. Whether or not optimal patrol interception strategies are also optimal preventive patrol strategies is subject to debate. One would hope that increases in the interception probability could serve as a crime deterrent; yet the perceived threat of interception may be a more significant factor. Strategies which optimize one measure need not optimize the other. Therefore, one of the major limitations of any of the existing search models of patrol is that they do not also capture the preventive aspect of patrol (assuming that crimes can be prevented) nor do they include the interception probabilities due to rapid response to a report of a crime in progress. In a study by the New York City-Rand Institute [18] it was found that of the criminals arrested at the scene of the crime by a patrol unit, approximately 50% were the result of a response to a call and 50% were the result of patrol initiated action. Thus, ideally a patrol model should incorporate at least both possibilities for intercepting a crime when attempting to determine the most effective patrol strategy.

The first set of works to be reviewed are those by Elliott [11, 12, 13]. Applying the Koopman exponential model for random patrol, Elliott calculates the probability of intercepting crimes of various duration as a function of the total time it takes to cover as many street miles as there are in the region patrolled. More importantly using data from Syracuse, he attempted to validate the model using as he admits relatively meager data. The result was that the estimate was within an order of magnitude of the observed frequency of interception. If in fact the estimate is within an order of magnitude of the actual frequency of interception it is probably coincidental. For one thing only reported type I crimes were considered which typically significantly underestimate the actual number of type I crimes, thereby decreasing the actual fraction of crimes intercepted. Secondly, his search theoretic estimates assume patrol was uniformly distributed across the city. However patrol is likely to be concentrated in higher crime areas which would also throw the estimates off significantly.

One last point about the model validation requires comment as there is a basic flaw in his application of Koopman's model to police patrol. To calculate the average speed of patrol Elliott determined the average number of miles patrolled during a tour and divided by eight yielding an average patrol speed of 4.4 miles per hour. An implicit assumption is therefore that there is no distinction between a patrol unit busy 60% of its time responding to calls for service but averaging 11 m.p.h. when on patrol and another patrol unit which does not respond to calls for service and averages 4.4 m.p.h. while on patrol. In Chapter IV we show the two are not equivalent . For the range of values Elliott considered the two alternatives do not yield very different estimates. However, if this distinction were not made when saturation patrols were considered, the result would be a serious overestimate of the probability of interception.

One last point to be discussed involves the estimation of the observable duration and conditional probability of detection of a crime. In Chapter VI we discuss the problem of obtaining estimates by interviewing patrol officers. Elliott has an interesting alternative suggestion. Once the Koopman model has been validated for police patrol, it would be possible to estimate the product of the duration and observability by allocating patrol uniformly in a region and finding the fraction of crimes of each type that are intercepted. In summary Elliott touched on many of the basic issues of applying the Koopman model to calculating the probability of interception; however no analysis is presented of, perhaps, the more important problem: how to improve on present deployment strategies involving randomly patrolling cars.

Olson [34], on the other hand, focuses more on the optimal allocation problem and develops a method for deploying a tactical (does not respond to calls for service) patrol force. His algorithm uses the Charnes and Cooper method to allocate a patrol force of n cars to a region which is subdivided into smaller groups of blocks. Since the resultant allocation to each region is not constrained to be integer, all fractional allocations are rounded off to produce only integer allocations of search effort to each group of blocks. Recognizing also that the specific solution may vary with the size of the subdivisions, the algorithm is repeated a number of times for different sized groupings of blocks and the results compared to obtain a more global optimal solution.

Chapter VI presents an analogous model which also generates integer allocations of manpower to each region. However the approach is not based on the Charnes and Cooper algorithm. The reason for not following this approach is, as Olson noted, because the resulting optimal allocation is not constrained to be integer. Consequently rather than round off to a solution that may or may not be optimal, it seemed more appropriate to start out by constraining the allocation to be integer by following a steepest ascent algorithm which serially allocated the patrol force unit by unit. Also with the latter approach multiple crime types (different durations) are no complication at all and sensitivity analysis is an immediate consequence of the optimality condition as described at the end of Chapter VI.

Olson and Wright [33] used a Markovian decision model to make an optimal allocation of effort (for a standard patrol car) in a manner that yields a random patrol schedule. The motivation for developing this model is that Koopman type allocation models may generate specified coverage levels that are infeasible. In developing their model, because of a lack of data on detection probabilities, they assume that the strategy which maximizes space time-coincidence maximizes also the probability of detection. This is not true, but the model's development is not really affected by this assumption. The model presents an alternative approach to the problem we discuss in Chapter VII. A discussion of the different orientation of the two methodologies will be presented later in this section.

Rosenshine [35] models an urban street grid as a flow network with the flows corresponding to the patrol coverage. He then develops an algorithm for determining the minimum total effort necessary to generate a set of flows that satisfy a given constraint on the minimum patrol effort allocated to each arc. The algorithm is designed to assure that patrol is 'as random as possible'. However as Gass and Dawson [15] point out, the degree of randomness required to thwart prediction of patrol routing is not likely to require the rather cumbersome and complicated approach outlined by Rosenshine. However, an equally important limitation

[36] on the algorithm is that it does not address the allocation problem, since the minimum constraint on the patrol coverage of each street (i.e. arc) is assumed to be given.

Blumstein and Larson [6] present a basic model for calculating the probability of interception which is a linear approximation of the model presented in Chapter IV. They use the model to obtain upper bounds on how frequently an individual patrolman is likely to come across a crime in progress. Using data for one large U.S. city and assuming, for example, that burglaries have a duration of twenty minutes, they estimate that an individual patrol officer can expect a maximum of four burglary-in-progress detection opportunities per year. Their discussion closes with a brief description of how the detection rate can be increased by changing the values of the model's input parameters. In section 4.1.3 we will elaborate on their discussion of changing the input parameters.

Larson in his book Urban Police Patrol Analysis [25] presents a detailed discussion of many of the issues surrounding the deployment of preventive patrol. Were we to plagiarize his six closing questions, entitled "Extensions and Further Work", we could still use the same heading. "To what extent is the crime distribution modified by patrol strategies?" is still an open and crucial question as is the question "What is the conditional probability of detecting a crime in a particular physical environment?". This last question will be the subject of extensive discussion in Chapters VI and VIII.

In discussing patrol models he introduces the Koopman allocation model as a possible conceptual framework for determining the optimal allocation of patrol effort. Upon concluding his presentation of the method, which generated for each point a patrol frequency, he raises the question of "To what extent is an optimal patrol coverage function realizable?". One obviously unrealizable function arises if a connecting street between two streets with non-zero patrol coverage is allocated no patrol. There is, however, a potentially more substantive barrier which undercuts the feasibility of this approach. In applying the Koopman model to a three minute crime the crime's duration plays a dual role. It not only represents the search effort but also constrains the sequencing of the search. As we elaborate in Chapter IV, an optimal patrol coverage function which assigns twice the patrol coverage to point A as to point B requires that the ratio be maintained not just over the eight hour tour but also over every possible three minute span and similarly for every pair of points.

Reviewing the approaches of Olson and Wright [33], Larson [25] and Rosenshine [35], there is a fundamentally different orientation to the problem of allocating a patrol unit's preventive patrol time compared to the orientation of the algorithms in Chapter VII. Their focus is on mapping out optimal routes through the city streets with the concurrent problems of realizability. The focus in Chapter VII is instead on the question of specifying the more gross allocation issues of which regions to patrol at all and which not. Along with this simplification comes the flexibility of incorporating into the model multiple crime types as well as assigning weights to the various crimes.

2.4 Patrol's Impact on Crime

The above section title could just as well havebeen written with a question mark "Patrol's Impact on Crime?". It is a question whose answer will have to await further research. The discussion here, therefore, will be brief, merely pointing out a number of cases which seem to shed some light on the issue.

The question of patrol's impact on crime has of late been split into two distinct questions. What is the impact on crime of highly visible easily avoidable patrol cars randomly (sometimes it seems aimlessly) patrolling the streets? What is the impact of other forms of patrol strategies (e.g. plain-clothes, directed patrol, saturation patrol, stakeouts, decoys, etc.)?

Two of the more well known contradictory studies that relate to the first question are the 20th Precinct study in [35] New York City and the now well known Kansas City experiment [20]. The former was not designed as an experiment and represents a post hoc analysis of some of the affects of an increase in manpower. This naturally raises some question of the validity of the claimed cause-effect relationship. Some of the effects claimed were a 36% decrease in total felonies visible from the street and a 49% decrease in visible (from the street) grand larceny. The Kansas City experiment, in which attempts were made to maintain controls, on the other hand, showed no discernible impact from doubling and tripling the number of units in a sector (which more than triples the potential number of preventive patrol hours). Should the Kansas City study prove the more generalizable (the results could conceivably be city dependent), it does not, however, negate the necessity for development of methodologies for determining optimal patrol strategies. As the authors of the report repeatedly point out, the experiment showed only that routine preventive patrol in marked police cars has little value in preventing crime or making citizens feel safe. As an alternative the experiment suggests that perhaps deployment strategies should be based instead on specific crime prevention and service goals.

Larson [29], however, in a detailed review of the Kansas City experiment raises a number of serious questions about whether or not the claimed experimental conditions were maintained. His conclusion is that it is not at all clear how generalizable the findings in Kansas City really are. Section 8.3.1 contains a description of some of the highlights of his analysis of the experiment.

In contrast to the above, two representative examples which claim to display the effectiveness of crime directed strategies involve a street crime unit (SCU) in New York City [14] and a burglary prevention program in Seaside, California [2]. In the first example the street crime unit averaged 8.2 man days per arrest against an average for all uniformed officers of 167 man days. In the Seaside experiment a special two man unit was assigned to reduce burglaries. The result was that over a two month period they made more arrests than the entire 54 man force did during the previous year. In addition there was a 25% reduction in burglaries and a 50% reduction in average loss.

Although both of these examples support the claimed effectiveness of crime directed police strategies, neither was carried out in a controlled experimental setting. And despite the often almost unanimous support among police for strategies of these types, their effectiveness remains to be proven.

FOOTNOTES 2

1 There are actually two versions of the hypercube model, an exact and an approximate one. We use the approximate version.

REFERENCES 2

1. Adams, R., S. Kolodney, R. Marx, and P. Wormeli, Operational Analysis of Police Field Force Command and Control in San Jose, prepared for San Jose Police Department by Sylvania Electronics Systems - Western Division, Mountain View, CA, August 1968.

2. Anderson, Dale Wl, Seaside Police Department Crime Prevention Program: Phase One/Burglary, Seaside Police Department, CA, 1973.

3. Bammi, D., Design of Patrol Beats to Minimize Response Time to Calls for Service, Ph.D. Dissertation, Illinois Institute of Technology, Chicago, IL, 1972.

4. Barnett, A., "On Searching for Events of Limited Duration," Working Paper WP-11-74. Massachusetts Institute of Technology Operations Research Center, Cambridge, MA, September 1974, submitted for publication.

5. Blachman, N. and F. Proschan, "Optimum Search for Objects Having Unknown Arrival Times", Operations Research, Vol. 7, No. 5, 1959.

6. Blumstein, A. and R.C. Larson, "A Systems Approach to the

Study of Crime and Criminal Justice", in Operations Research for Public Systems, P.M. Morse and L.S. Bacon (Eds.), The MIT Press, Cambridge, MA, 1967.

7. Charnes, A. and W.W. Cooper, "The Theory of Search: Optimum Distribution of Search Effort", Management Science, Vol. 5, October 1958.

8. Chelst, K.R., Implementing the Hypercube Queuing Model in the New Haven Department of Police Services: A Case Study in Technology Transfer, New York City- Rand Institute Report, R-1566/6-HUD, New York City, NY, July 1975.

9. Dean, B.U., A. Reisman, A. Daughety, R. Ehresman, V. Huckfeldt, M. Kiley and C. Pehmezcilar, A Preliminary Systems and Allocation Study of the Cleveland Police Department, Case Western Reserve University School of Management, Cleveland, OH, January 1970.

10. Detroit Police Department, Resource Allocation System, Volume II, Touche Ross & Company, Detroit, MI, June 1972.

11. Elliott, J.F., "Random Patrol", Law Enforcement Science

and Technology II, Proceedings of the Second National Symposium on Law Enforcement Science and Technology, Illinois Institute of Technology Research Institute, Chicago, IL, 1970.

12. Elliott, J.F. and T.J. Sardino, Crime Control Team: An Experiment in Municipal Police Department Management and Operations, Charles C. Thomas Publisher, Springfield, IL, 1971.

13. Elliott, J.F., Interception Patrol, Charles C. Thomas Publisher, Springfield, IL, 1973.

14. Exemplary Programs, Office of Technology Transfer, National Institute of Law Enforcement and Criminal Justice, Washington, D.C., 1974.

15. Gass, S.I., "On the Division of Police Districts into Patrol Beats", Association for Computing Machinery, Proceedings of the 23rd Conference, Las Vegas, NV, 1968.

16. Gass, S.I. and J.M. Dawson, An Evaluation of Policy Related Research: Reviews and Critical Discussion of Policy-Related Research in the Field of Police Protection, prepared as a final report to NSF-RANN, Mathematica, Inc., Bethesda, MD, October 1974.

17. Gilbert, E.N., "Optimal Search Strategies", J. Soc. Indust. Appl. Math., Vol. 7, 1959.

18. Greenwood, P.W., An Analysis of the Apprehension Activities of the New York City Police Department, New York City-Rand Institute, Report R-529-NYC, September 1970.

19. Heller, N.B., R.E. Markland and J.A. Brockelmeyer, "Partitioning of Police Districts into Optimal Patrol Beats Using a Political Districting Algorithm: Model Design and Validation", paper presented at the Operations Research Society of America meeting, Anaheim, CA, 1971.

20. Kelling, G.L., T. Pate, D. Dieckman, and C.E. Brown, The Kansas City Preventive Patrol Experiment: A Technical Report, available from the Police Foundation, Washington, D.C., October 1974.

21. Kisi, T., "On an Optimal Searching Schedule", Journal of the Operations Research Society of Japan, Vol 8, No. 2, February 1966.

22. Koopman, B.O. "The Theory of Search: II. Target Detection", Operations Research, Vol. 4, 1956.

23. Koopman, B.O. "The Theory of Search: I. Kinematic Bases",

Operations Research, Vol. 4, 1956.

24. Koopman, B.O., "The Theory of Search: III. The Optimum Distribution of Searching Effort", Operations Research, Vol. 5, 1957.

25. Larson, R.C., Urban Police Patrol Analysis, MIT Press, Cambridge, MA, 1972.

26. Larson, R.C., "A Hypercube Queuing Model for Facility Location and Redistricting in Urban Emergency Services", Computers and Operations Research, Vol. 1, No. 1, March 1974.

27. Larson, R.C., Urban Emergency Service Systems: An Iterative Procedure for Approximating Performance Characteristics, The New York City-Rand Institute, R-1493-HUD, January 1974, to appear in Operations Research.

28. Larson, R.C., "Illustrative Police Sector Redesign in District 4 in Boston", Journal of Urban Analysis, Vol. 2, 1974.

29. Larson, R.C., "What Do We Know About Preventive Patrol: A Review of the Kansas City Preventive Patrol Experiment", published by Public Systems Evaluation, Inc., Cambridge, MA, July 1975, to appear in Journal of Criminal Justice.

30. Moore, L.M., A Review of Search and Reconnaissance Theory Literature, Department of Industrial Engineering, University of Michigan, Technical Report No. 70-4, Ann Arbor, MI, January 1970.

31. Moore, L.M., A Characterization of the Visibility Process and Its Effect on Search Policies, Systems Research Laboratory, Department of Industrial Engineering, University of Michigan, Report No. SRL 2147 TR 71-3, December 1971.

32. Morse, P.M. Search Theory, Working Paper WP-01-74, Massachusetts Institute of Technology Operations Research Center, Cambridge, MA, January 1974.

33. Olson, D.G. and G.P. Wright, Models for Allocating Police Preventive Patrol Effort. Krannert Graduate School of Industrial Administration, Purdue Univeristy, Paper No. 404, West Lafayette, IN, April 1973.

34. Olson, D.G. "A Preventive Patrol Model", paper presented at Operations Research Society of America meeting, Miami, FL, 1969.

35. Press, S.J., Some Effects of an Increase in Police Manpower in the 20th Precinct of New York City, New York

City-Rand Institute Report R-704-NYC, October 1971.

36. Rosenshine, N., "Contributions to a Theory of Patrol Scheduling", Operational Research Quaterly, Vol. 21, 1969.

CHAPTER 3

AN INTERACTIVE APPROACH TO POLICE SECTOR DESIGN

3.0 Introduction

A primary task of the police administrator is to make the most efficient and effective use of his manpower. To do so he must allocate his force in a way that reflects the temporal and geographical variations in crime level and other demands for service. This chapter concerns itself with a geographical allocation problem, specifically the design of police beats or sectors.

The computer models described in this chapter are formulated to aid a district (precinct) commander in designing police patrol sectors which represent, to him, a good balance of several somewhat conflicting goals. The models do not focus on optimizing any one performance measure (e.g. average response time) and are built on the belief that the judgement of the experienced police manager must be a major motivating force behind any sector design. Consequently, their purpose is to guide the decision-maker through a series of alternative sector designs and at each stage provide him with the information he considers necessary to choose between alternative designs.

3.1 An Interactive Design

The decision to develop computer models which require user interaction strongly influences all aspects of their design. A discussion of some of the reasons behind this decision is certainly warranted and may also provide some additional insight into the potential uses of the model. The first question that is raised in designing sectors is "On what basis do we judge good sector design?". One of the best suggestions was offered by Vollmer [15]: "construct beats so that every patrolman carries his share of the burdens and each section of the community receives its share of police protection".[1] These <u>equity</u> criteria are important and the major thrust of the models is to aid the manager in producing an equitable configuration. However, they are certainly not the only important criteria. Superbeat [14], a computerized sector design procedure designed at Illinois Institute of Technology, attempts to minimize district-wide average response time. Larson [9], in a more complete list of objectives than offered here, includes, for example, minimizing the number of cross sector dispatches.

If one sector configuration could be devised that would satisfy all the various criteria, user interaction might not be that critical. Unfortunately, these various criteria are often conflicting. For example, one may cite the two equity criteria--equal patrolman workloads and equal community protection (as measured by average response time, for example).

the successful implementation of innovations in both the public and private sectors. Little [11] offers as one reason for implementation failure of operations research models that 'managers don't understand the models and people tend to reject what they don't understand'. Colton [4] states that a source of problems in police computer use is the gap that often exists between technical and sworn personnel. An interactive system requiring the involvement of the police manager bridges the gap and provides the manager with a fuller understanding of the potential uses and limitations of their particular computer models.

This introduction is intended only as an initiation into the goals, problems and conflicts that exist in the design of sectors; it is in no way meant to be an exhaustive analysis of these aspects of sector design. For more detailed discussions of these issues, the reader is referred to Chapman [2] and Larson [9]. The following sections of this chapter describe and analyze one particular user-interactive model for sector design. The model's assumptions and data requirements are outlined and a description of the overall system is provided. Finally, there follows a discussion of five subsystems whose major purpose is to guide the police manager in producing a sector design which satisfies various equity criteria.

In regions where the density of crime is low if we are to maintain equal workloads, the size of the sector would have to be above average. The increased size of the sector would have a tendency to produce an average response time for a call for service in that sector above that found for a call for service in the high-crime small sector. The role of a police administrator who has intimate knowledge of his officers and community is crucial here in striking an acceptable balance between these and other such conflicting goals. He should be able to determine the effect of various workload imbalances on police morale and also, depending upon the political realities of his community, determine what represents an acceptable imbalance in average response time.

A second reason for desiring user interaction is that a number of objectives are not easily quantifiable. Larson [9] mentions as one objective in sector design the maintenance of neighborhood integrity so that, where possible, sector boundaries do not cut traditional neighborhoods in two. Another example of a non-quantifiable objective is evidenced in a recent proposal to redesign a sector in Boston. The district commander asked that the sector boundaries lie along the main thoroughfares in order to facilitate police patrol [13].

A final reason for developing an interactive system rather than a set of packaged programs is that this will facilitate implementation of the models. A number of recent articles have addressed the problem of what factors influence

3.2 Model Assumptions

A district or precinct is subdivided into a set of smaller regions called beats or sectors. Each sector consists of a set of reporting areas (subregions) or atoms. A reporting area is the smallest geographical unit for which police data are collected.

In combining a group of atoms to form a sector, only one major constraint (C1) is placed on the design of the sector.

(C1) A sector must consist of a set of contiguous atoms. There will, however, be one additional constraint that will be applied in a heuristic manner.

(C2) Every attempt will be made to maintain <u>compactness</u> in the sector designs.

The dispatch policy (i.e. the procedure by which units are assigned to answer calls for service) used in this model (MCM) is as follows. Each sector is assigned a patrol unit which has primary responsibility for that sector . Primary responsibility involves two things. 1) The patrol unit performs preventive patrol (time not spent answering calls for service) in that sector. <u>While on preventive patrol the time spent in each atom in the sector is assumed to be in direct proportion to the number of calls for service emanating from that atom</u>. 2) Any calls for service that arise in the sector are assigned to the patrol unit with primary responsibility for that sector if it is free regardless of its location. If it is not free, the <u>available</u> unit <u>estimated</u> to be closest is dispatched to the call.

The estimated distance of a unit from a call for service (for dispatch purposes) is calculated as follows. The unit is assumed to be at the weighted (on the basis of calls for service) center of gravity of its sector. The call for service is assumed to be at the center of mass of the atom from which it comes. The distance between the two centers is the sum of the (absolute) differences of the respective (X,Y) coordinates.

If when calls arrive no unit is available, the calls are queued in the order of their arrival and the first unit to become available is dispatched to the call at the head of the queue. This completes a description of the only dispatch policy presently allowed in the model. However, it is not difficult to change the model to allow for a wider range of dispatch policies (Larson [8,10]) and it is envisioned that future versions of this model will contain this added flexibility. The various statistics (e.g. workloads, travel times, etc.) which describe the functioning of the district's patrol units are calculated by using Larson's approximation procedure [10]. A description of this procedure is not offered here and the reader is referred to Larson [10]. However, one remark on this procedure is worth making. The operating statistics such as workloads and average travel times to a sector or atom include not only intrasector dispatches but also include intersector dispatches.

3.3 Data Requirements

The basic unit for which data must be supplied is the reporting area or atom. For each atom the following must be given

(D1) Location of the center of mass of the atom.

(D2) The call rate for the atom in terms of calls per year or any other unit of time as long as the user is consistent.[2]

(D3) An atom contiguity vector. For each atom a list of all the atoms contiguous to it must be provided.[3]

In addition, if the user is interested in tracking and possibly constraining the size of each sector either in terms of area or in terms of street miles, then additional data must be input for each atom.

(D4) Number of street miles in the atom.

(D5) Area of the atom.

Finally, the user must specify for the district as a whole.

(D6) Number of sectors.

(D7) Average workload for the units.

(D8) An initial sector design.

3.4 System Design

Once the data files have been input, the system is ready to operate. The system's initial interaction with the user will be to request that he choose a major concern from among five possibilities:

1. Equalize workloads.
2. Equalize preventive patrol coverage.
3. Equalize average travel times to a sector (region).
4. Equalize average travel times to an atom (subregion).
5. Equalize workloads of amalgamated sectors.

For each of the five options, a subsystem has been designed which will guide the user in an iterative manner towards improving a particular performance measure (workload, preventive patrol coverage, average travel time) imbalance. At the core of each subsystem is a program that operates in either one of two modes at the discretion of the user (See Figure 3.1). In one mode the program generates several alternative small modifications (transferring an atom from one sector to another) of the present configuration that improve upon the imbalance in the performance measure specified by the user. The user may then choose from among these alternatives, whereupon the modification is carried out and several new alternatives are generated. In the other mode, after the user has specified the imbalance to be corrected, the program carries out on its own a <u>series</u> of small modifications of

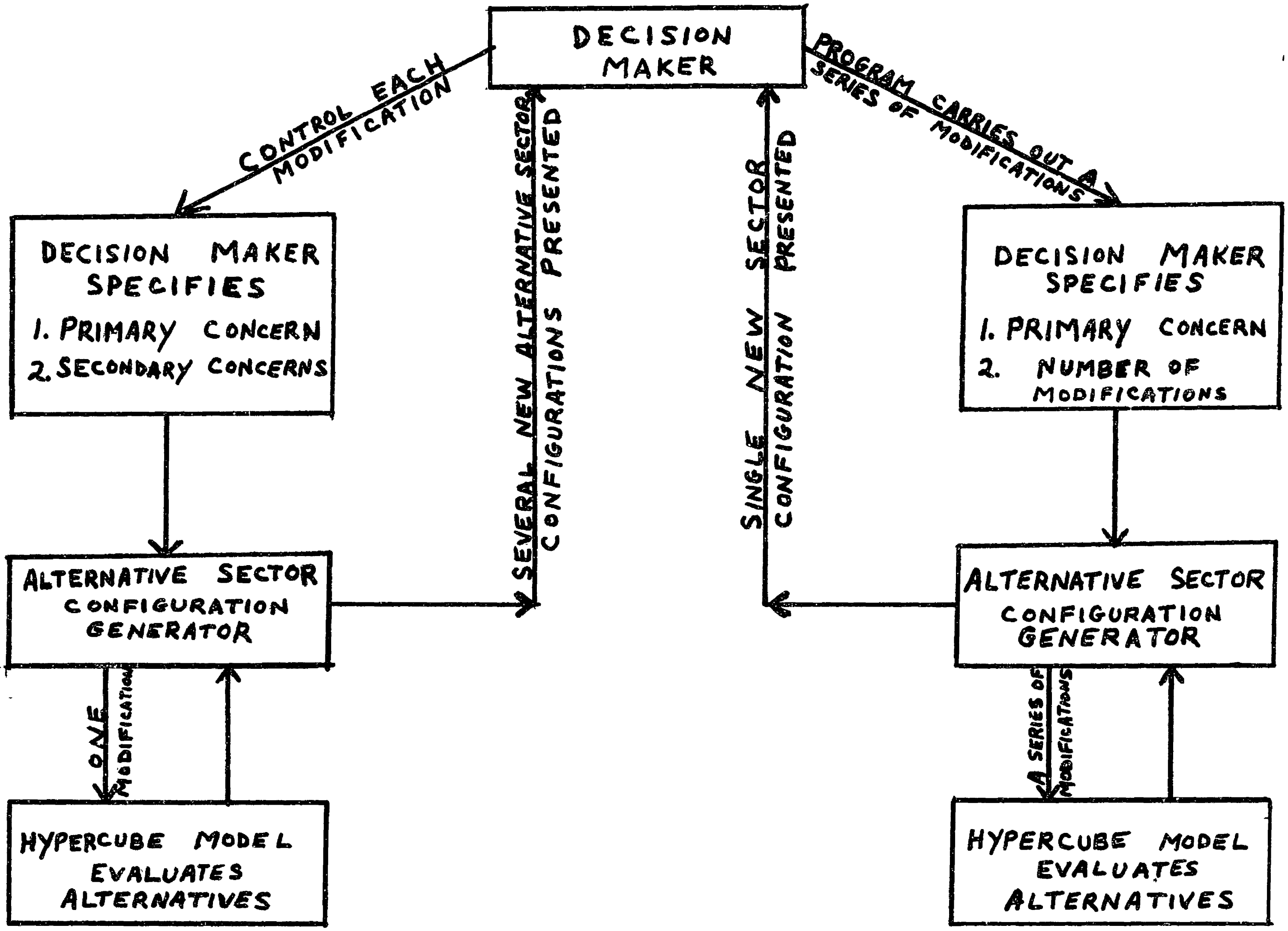


Figure 3.1:An Interactive System for Sector Design

the present configuration in order to improve on that imbalance.

Before proceeding with the description of the system, we will discuss briefly the distinction between the goal of subsystem 1 (equalize workloads) and subsystem 2 (equalize preventive patrol coverage). In addition we will outline the basic problem that the last subsystem (equalize workloads of amalgamated sectors) is designed to address.

3.4.1 Option 2: Equalizing Preventive Patrol Coverage

In calculating the average workload for a precinct and the individual workload for each patrol unit, we have included not only calls for service but also time spent for meals. Consequently, the fraction of time a patrol unit has available for preventive patrol is simply one minus its workload. Therefore, if the workloads are balanced, then obviously the number of patrol hours allocated to each sector are also balanced. However there are other definitions of an equitable allocation of patrol effort which do not coincide with requiring that the total number of patrol hours for each sector be the same. For example, a decision maker might question if it is equitable for two sectors, of different size or with different crime rates, to be allocated the same number of preventive patrol hours. A decision maker might feel that a better standard of equity is for the number of patrol hours per street mile to be the same in each sector or, alternatively, for the number of patrol hours for each Part I type street crime to be the same in each sector.

In response to these and other alternative definitions of equitable preventive patrol coverage the interactive system was designed (Option 2) with the flexibility to allow the decision maker to specify his own measure of equity. Equity can be defined as requiring that patrol hours be allocated in proportion to street mileage or street crimes or in proportion to any other parameter (e.g. the product of street mileage and crime rate) that is specified.

3.4.2 Option 5: Amalgamating

Past and present police literature have recognized that crimes and calls for service are not uniformly distributed over time and space. A number of manpower allocation methodologies (Walton [16], Ficklin [5], McClaren [12], Larson [7], to cite a few) have been developed in an attempt to mirror and take advantage of this nonuniformity. Once temporal and district-by-district allocation plans have been decided upon, the next step is to design the beat structure for each district. One option available is to design a different beat structure for each of the various police shifts (early morning, day, night, weekend). However, to avoid the necessity of drawing up a set of extensively different beat boundaries for each shift and the confusion it might produce, police often settle on a compromise. The basic beat design is formulated for the shift which has the maximum number of cars. For shifts with lower call rates and fewer men assigned, they simply combine some of the sectors (i.e. one man will patrol two or more sectors). The fifth subsystem will attempt to determine which sectors should be combined to produce a minimum workload imbalance.

To aid the manager in his decision-making he can request that additional information be provided at each step. He can receive data with regard to any or all of the factors listed below:

1. Car Workloads
2. Preventive patrol coverage
3. District wide average travel time
4. Sector travel times
5. Three (user may specify own number) worst atom travel times
6. Percentage of cross-sector dispatches
7. Number of street miles in each sector

Although the option exists to ask for all of the above factors, it is advisable that the user be selective and pinpoint the information he feels is critical in choosing between alternative designs. The natural tendency, of course, is to ask for whatever information is available; unfortunately, trying to work with too many variables at one time may only serve to confuse and slow the iterative process. Consequently, the police manager should first analyze what parameters he considers crucial before attempting to use the system. In this way he will avoid the problem of information becoming misinformation (Ackoff [1]). An alternative to a constant flow of information is to set up a series of constraints. For example:

(C3) A maximum and/or minimum limit is placed on the size of a sector. Whenever a potential sector design violates this constraint the user is notified.

or

(C4) If the user's major concern were sector travel times, he might set a constraint on the maximum workload for any particular sector car.

3.5 Subsystem Designs

The next section of this chapter will discuss the methodology used by each of the first four subsystems to guide in the construction of a satisfactory sector configuration. The fifth subsystem, combining sectors, will, however, be discussed separately as it entails a fundamentally different approach than the others. In each of the four subsystems, the user proceeds iteratively from an initial sector design until an acceptable design is reached. The general procedure (GP) for each of the four subsystems is as follows:

(GP1) An initial sector design is specified by the user.

(GP2) At each iteration the most recent design is modified by transferring one atom from one sector to another without violating the sector contiguity constraint. (C1)

(GP3) A transfer is specified by the atom transferred, the sector losing the atom and the sector receiving the atom. The criteria for determining these variables will vary with each subsystem, and even the order in which they are determined will vary.

(GP4) At each iteration the user will be offered a choice between two or three atom transfers which reduce the disparity in the parameter of major concern (either workload, sector travel time or atom travel time).

(GP5) Each of the candidate atoms offered the user is first screened to make sure that its transfer will not violate the contiguity constraint (C1). This involves checking the structure of two sectors.

a) The sector, R, receiving the atom, A, must presently be contiguous to that atom. This fact can be determined by checking if any of the atoms presently in sector R are contiguous to atom A.

b) The sector, L, that loses the atom, A, must not be left split in two. To determine if sector L is still contiguous, the sector is modelled as a network with nodes of the network representing atoms and the arcs between nodes representing the contiguity of atoms. The issue then becomes "Is it possible to travel from any one particular node in the network to every other node in the network?". To answer this question a node labeling algorithm is used that is analogous to the shortest route algorithm of Ford and Fulkerson [6]. (See Appendix A.)

(GP6) Information will be provided to the user about the effect of each transfer on the parameter of major concern and on any other parameters he has previously specified.

(GP7) The user can also offer another candidate for transfer, in which case, the effect of the transfer on the various parameters will be calculated by the system.

(GP8) The user selects the atom to be transferred and the transfer is then carried out. Steps GP2 thru GP8 are repeated until an acceptable sector design is reached.

There is one additional powerful option available to the user. Computer programs have been developed that will reduce the imbalance of any one particular parameter without user interaction at each iteration. The option is as follows:

(GP9) The user can allow any one of these programs to run for n (user defined) iterations or until the sector or atom with the worst imbalance is brought within p% (user defined) of the average.

(GP10) Upon reaching the user-defined stopping point, the new sector design is described to the user and the system returns to step GP2.

3.5.1 Typical Scenarios

These relatively non-interactive programs offered in step GP9 have the potential of significantly reducing the various imbalances unaided and all but eliminating any imbalances in car workloads and sector travel times (see computational experience sections). However, it is envisioned that the non-interactive programs will also be used in a more limited (carry out 5 iterations) fashion without realizing their full potential. The motive behind restrained usage of the programs is that full utilization, for example, of the program to reduce workload imbalances could result in totally unacceptable travel time imbalances.[5] These probrams (GP9) might be used instead to speed up one dimensional searches for an acceptable sector design. For example a typical scenario might be

1. The user decides that the initial design has too large a workload imbalance and requests the program that reduces workload imbalances to run on its own for five iterations.

2. Reviewing the resultant design, the user now decides to shift focus and concentrate on travel time. He requests that the program that balances sector travel times carry out three iterations.

3. Reviewing the newly generated beat configuration, he decides to improve once again the workload imbalance, however, with himself now in complete control of each iteration.

4. The user requests the generation of three alternative atom transfers that reduce the workload imbalance and that supplementary information be provided about precinct-wide average travel time and sector travel times.

5. The user proceeds to choose from among the three, the atom transfer which has the most beneficial impact on precinct-wide travel time and which, of course, improves also the workload imbalance.

6. The user then may request three more potential transfers and proceed again with the above process.

Having completed our discussion of the general or shared procedures (GP1-GP10) that are common to each of the first four subsystems, we will next focus on the unique characteristics of each of these subsystems. The discussion of each of these subsystems will be presented in three parts. The first part will analyze how an atom transfer affects the parameter of interest. The hope is that this analysis will provide additional insight into the development and procedure of each of the subsystems. Secondly, the manner of selecting each of the atoms suitable for transfer will be described including a description of the programs that run without direct user interaction. Finally, some computational results with the above-mentioned programs will be presented.

Before proceeding with the description of the subsystems, it should be noted and emphasized that the main focus of these subsystems is the equitable distribution of workloads and police services as measured by response times. As a result, no program to reduce precinct-wide average travel time is presented here. However, the effect of any atom transfer on district-wide average response time is monitored at each step and the user can take into account this effect when choosing between alternative designs. In addition, the interactive, iterative approach, presented in this paper will provide the user with a better understanding of the various tradeoffs that exist in sector design. District-wide response time is usually increased in order to reduce imbalances in the sector and atom travel times. Workload imbalances must often increase in order to reduce travel time imbalances and vice versa.

3.6 Workloads

3.6.1 Effect of Atom Transfer

Shifting an atom out of (into) a sector has two effects[6], a primary effect (E1) which reduces (increases) the sector car's workload and two complementary effects, (E2) and (E3), which counter-balance and reduce the magnitude of the primary effect.

(E1) Shifting an atom out of (into) a sector directly reduces (increases) the primary responsibility of the sector car, thereby, decreasing (increasing) his workload.

(E2) The first secondary effect is that by decreasing (increasing) the sector car's primary responsiblity, the car is now more (less) frequently available for intersector dispatches, consequently, increasing (decreasing) his workload.

(E3) The other secondary effect is that by transferring an atom out of (into) a sector, S, and into (out of) an adjoining sector, A, the adjoining sector will have an increased (decreased) workload. This will in turn necessitate more (less) intersector dispatches into sector A. The result is that the sector car for sector S will be called on more (less) frequently to respond to calls for service in sector A.

The magnitude of these second order effects (E2) and (E3) will vary according to the centrality of location of the sector. The more centrally located the sector, or in other words, the larger the region for which this sector car is the 2nd and 3rd most preferred car respectively, the greater the magnitude of this second order effect [10]. Consequently, an atom transfer in outlying sectors will in general produce a greater change in workload than an atom transfer in centrally located sectors.

3.6.2 Computer Program

The program, if reduced to its essence, can be described by outlining the steps by which it determines the previously mentioned (GP3) three variables involved in each atom transfer, (the atom transferred and the two sectors).

(WS1) The first step in equalizing workloads is to find the sector with the workload furthest from the mean, either above or below it.

Unfortunately, this program could not concentrate solely on reducing high workloads without simultaneously increasing the workloads of underworked cars. This is true because the average workload is a constant which depends only on the total call rate for the district and the number of sectors and not upon the particular sector configuration.

If the workload for sector car A is the one furthest from the average and it is above average, then procedures (WS2) and (WS3) are followed.

(WS2) A search is made of the sectors contiguous to sector A to determine which of the contiguous sectors has the sector car B with the lowest workload.

(WS3) The atom transferred from A to B is the atom in sector A, that is contiguous with B, and whose center of mass is closest to the center of gravity (weighted call rates) of sector B. This atom transfer must of course not violate any contiguity constraint (C1)[7] (See also GP5). If it does violate (C1), the second closest atom is transferred.

Although simple calculations could have been carried out to determine which atom transfer from A to B would have had the most dramatic effect on the workload imbalance, this simpler criterion (WS3) was used for one basic reason. This criterion proved to be a good heuristic method for maintaining a degree of compactness (C2) in the sector receiving the atom. To summarize (WS2) and (WS3), the procedure is to first determine the two sectors which were to be involved in a transfer and then to select an atom on the basis of compactness.

If the worst imbalance occurs in an under-utilized sector car, A, then procedures (WS4) and (WS5) are followed.

(WS4) A search is made for the closest (as defined in WS2) contiguous atom, C, to sector A rather than first searching for the sector B that is contiguous to A and whose sector car B has the highest workload.

(WS5) Once the atom, C, that is to be transferred into A has been found, the sector that will lose the atom is implicitly determined.

In short, the atom to be transferred was chosen without first selecting the 2nd sector to be involved in the transfer. The reason for not paralleling the procedure outlined in (WS2) and (WS3) is that this method proved to be more efficient at maintaining compactness in the sector receiving the atom.

In the mode (GP2-GP8) in which the user decides at each iteration which atom will be transferred, the procedures, (WS1) through (WS5), are repeated to produce the two best (compactness criteria) candidates to reduce the worst workload imbalance. A third atom transfer is offered which involves the sector car with the second worst workload imbalance.

In the non-interactive mode, (CP9), the computer program iteratively carries out the above steps, (WS1-WS5), until the user defined stopping point is reached.

3.6.3 Computational Experience

The non-interactive program was run on data from District 4 in Boston. Seventy reporting areas were to be combined into 6 sectors. If the number of reporting areas is too large (a number as yet undetermined), it may prove necessary to group the reporting areas into slightly larger regions before using the models. The average workload for the 6 cars was 50%. The measure of imbalance that was used in this program was the ratio of the highest workload to the lowest workload. Minimizing the imbalance is equivalent to bringing the ratio closer to one. In the initial design, the car with the lowest workload was in sector 5 and was busy 42% of the time. The highest workload was for sector car 2 with a workload of 56%. The ratio of highest to lowest was 1.33. In other words, sector car 2 was busy 33% more of the time than sector car 5 was. After 14 iterations, the sector car with the lowest workload was sector car 6 (49.5%) and the sector car with the highest workload was car 5 (50.5%). The ratio of highest to lowest workloads was 1.02, a significant improvement.

The program was run for 20 iterations; however, no improvement occurred after the 14th iteration. Table 3.1 contains the workloads of the cars for the 14 iterations. In general, once the ratio of highest to lowest workload was less than 1.05, it was difficult to find an atom transfer that could reduce further the imbalance.

The 20 iterations required 17 c.p.u.[8] seconds or an average .85 seconds per iteration. The core storage required was 200K. The limited number of iterations required to eliminate a significant workload imbalance and the rapidity of the iterations would serve to validate the feasibility of an iterative approach.

Table 3.1 Balancing Workloads:District 4,Boston

Iteration	CARS 1	2	3	4	5	6	Ratio=High/Low
Initial	.498	* .559	.502	.494	** .421	.526	1.33
1	.498	* .559	.500	.493	** .424	.526	1.32
2	.494	* .558	.499	.492	** .431	.525	1.30
3	.483	* .558	.499	.491	** .443	.527	1.26
4	.496	* .549	.498	.488	** .442	.527	1.24
5	.497	* .543	.496	.486	** .458	.520	1.18
6	.497	* .534	.497	.497	** .458	.518	1.17
7	.494	* .533	.496	.497	** .463	.517	1.15
8	.492	* .534	.492	.494	** .478	.510	1.12
9	.496	* .514	.510	.495	** .476	.509	1.08
10	.484	* .515	.510	.496	** .486	.509	1.06
11	.497	.507	* .510	.494	** .485	.506	1.05
12	.498	.507	* .507	.494	** .489	.505	1.04
13	.497	* .511	.502	** .492	.505	.493	1.04
14	.497	.505	.502	.496	* .505	** .495	1.02

* Highest Workload

** Lowest Workload

3.7 Preventive Patrol Coverage

Earlier in this chapter we discussed the distinction between balancing workloads and balancing preventive patrol coverage. However, despite the differences between the two, it was not necessary to develop an entirely separate subsystem in order to provide a capability for balancing preventive patrol coverage. Instead this capability was easily provided by making several minor modifications of the workload balancing subsystem. In essence the function of these changes was so that when the user specifies preventive patrol coverage as the major concern, the program uses, for example, one minus the workload divided by the sector's street mileage in determining which sector has a patrol allocation that is furthest from the mean instead of looking at the workloads. Once that sector has been located then if its patrol allocation is above average, the sector is enlarged by adding an atom and if its allocation is below average the sector is modified by removing an atom. Before proceeding with the description of the other changes, it is important to stop and realize one consequence of the above substitution for workloads. If the street mileage or crime rate is not approximately the same for each sector then not only are balancing workloads and preventive patrol coverage not equivalent but also the two goals of <u>necessity</u> conflict.

In addition to the above substitution, the only other changes involve the obvious need to input data either about the street mileage or crime rate of each atom and calculate and update (after each atom transfer) the total street mileage or crime rate of each sector. Except for the modifications described above, no additional effort was required in adapting the workload balancing subsystem to be used alternatively to balance preventive patrol coverage. Thus, for example, the selection of the atom to be transferred between sectors follows the same procedures as described in the previous section (steps WS2 through WS5).

3.8 Sector Travel Times

3.8.1 Effect of Atom Transfer

Transferring an atom, C, out of a sector affects the average travel time to a call in that sector in three ways.[9]

(E4) First, if the average travel time to that particular atom C, is above (below) the average for that sector, removing the atom will lower (raise) the average travel time for the sector.

(E5) By decreasing the workload of the sector car, it reduces the need for intersector dispatches into this sector. Since in general, <u>inter</u>sector dispatches have larger travel times than <u>intra</u>sector dispatches, it will reduce the average travel times to calls in that sector.[10]

(E6) Since the atom removed from the sector will usually be relatively far from the center of gravity of the sector, the average distance from the center of gravity to the atoms in the sector will decrease.

3.8.2 Computer Program

A discussion of the computer program that aides the user in reducing sector travel time imbalances will also center on how the three variables (atom transferred and the sectors between which the atom is transferred) are determined.

(SS1) First, the sector A, with the worst average travel time is located.

(SS2) The sectors contiguous to sector A are compared to determine which sector, B, has the lowest average travel time.

(SS3) Of the atoms in A contiguous to sector B, the atom farthest away from the center of gravity of A is transferred. (E6)

In steps (SS1-SS3), no explicit attention is paid to the first effect (E4) in choosing the transferable atom as it is assumed that there is a strong correlation between the distance from the center of a sector to an atom in that sector and the average travel time for the atom. The farther away from the center of gravity the atom is, the larger the average travel time to the atom. In addition, even if the travel time for the atom transferred were below the average for the sector, the second and third effects (E5) and (E6) could outweigh the first effect. In the sample runs, this, in fact, happened a couple of times. Also by disregarding (E4) and concentrating on the distance (E6) in choosing an atom, a greater degree of compactness is maintained in the sector losing the atom.

(SS4) If, however, it is found that the removal of a particular atom under step (SS3) increases the average travel time for the sector A, then step (SS3) is repeated to find the second farthest atom from the center of gravity.

At first it seemed reasonable to design a slightly simpler procedure instead of step (SS2) followed by (SS3).

(Alt) From the sector with the highest average travel time, transfer the atom farthest from its center of gravity and contiguous to some other sector.

However, this procedure often eventually led to the following scenario, a scenario which was repeated by all of the first three subsystems whenever the imbalance was reduced below 5%.

Atoms were transferred out of an outlying sector A, which had a high average travel time, into a neighboring

> sector B until the average travel times in both sectors were equally above average. At this point the situation was ripe for a "ping-pong" effect, e.g. transference of the same atom back and forth between the two sectors (A and B).

By using the criteria of lowest travel time to first choose among neighboring sectors (SS2), the above scenario was all but eliminated. Sector B would not transfer its atoms to sector A but rather to some more centrally located sector, C, which had an even lower travel time than either A or B.

The above procedures (SS1-SS4) are used to find two atoms in the sector with the worst travel time and one in the sector with the second worst travel time whose removal will reduce the average travel time for their respective sectors. The user must then choose among these three possible atom shifts. In the non-interactive mode (GP9), the above procedures (SS1-SS4) are repeated until a user defined stopping point is reached.

The procedure designed to equalize sector travel times differs in one fundamental way from the one designed to equalize workloads. With regard to sector travel times, the programs try to reduce high average travel times and not increase low travel times, although, of course, the two are not independent. The reason for this is that unlike the average workload, the average district-wide travel time is not a constant and will vary with the particular sector configuration. It is hoped that by concentrating solely on reducing high travel times, it may be possible simultaneously to reduce the district-wide average travel time (which did not happen in the test runs[11]) or at least hold any increases to a minimum.

3.8.3 Computation Experience: District 4, Boston

The non-interactive program was run on data from District 4 in Boston. In the initial sector design (6 sectors) the average travel distance for the district was .543 miles[12]. Sector 5 had the worst average travel distance .696 miles (28% above average) and sector 2 had the best average travel distance .483 miles (11% below average). The ratio of the worst to best travel distances was 1.44 or in other words, on the average it took 44% longer to respond to a call in sector 5 than it did to a call in sector 2. After 13 iterations, the sector with the highest average travel distances was sector 5 (.591) and the sector with the lowest average travel distance was sector 2 (.559). The ratio of highest to lowest was now 1.06. Unfortunately, some of this reduced imbalance was paid for by a 6% increase in the district-wide average travel distance which now was .577 miles.

No improvement was made after the 13th iteration and each iteration required on the average 1.5 c.p.u. seconds. The reader will note that iterations in this subsystem are significantly longer than for the first subsystem. One reason for this is that the calculation of the average travel times is much more time consuming than the calculation

Table 3.2 Balancing Sector Travel Distances:District 4,Boston

Iteration	SECTORS (TRAVEL DISTANCES) 1	2	3	4	5	6	Ratio=High/Low
Initial	.510	**.483	.521	.547	*.696	.513	1.44
1	.535	**.490	.525	.551	*.684	.514	1.40
2	.536	**.492	.527	.553	*.672	.529	1.37
3	.538	**.494	.529	.557	*.665	.541	1.35
4	.549	**.498	.531	.557	*.656	.542	1.32
5	.551	**.501	.534	.561	*.636	.560	1.27
6	.578	**.508	.538	.564	*.636	.563	1.25
7	.602	**.515	.546	.570	*.626	.564	1.22
8	.607	**.523	.553	.577	*.609	.589	1.16
9	*.612	**.527	.557	.583	.579	.605	1.16
10	.604	**.543	.565	.588	.573	*.605	1.11
11	*.605	**.553	.572	.590	.567	.597	1.09
12	.593	**.564	.577	.592	**.564	*.596	1.06
13	.586	**.559	.572	.586	*.591	.581	1.06

* Largest Sector Travel Distance (in Miles)

** Smallest Sector Travel Distance (in Miles) of the workloads. However, the results still validate the feasibility of an iterative procedure. Table 2 contains the sector travel times for the 13 iterations.

It is worth noting that in the final design sector car 5 had a workload of 35% and sector car 2 had a workload of 56%, or 60% higher than that of sector car 5. This result highlights the previously mentioned tradeoffs that exist in sector design and underscores the necessity for police manager involvement in making these tradeoffs.

3.8.4 Computational Experience: New Haven

In a recent technology transfer project involving the New Haven Department of Police Services [3], the non-interactive program was used to explore how one particular precinct's [13] configuration could be modified to improve on imbalances in travel distances. In the succeeding paragraphs we will describe how the program improved imbalances in sector travel distances as well as its impact on the other performance measures. The results display even more clearly the potential of this sector redesign program since the initial imbalance in sector travel distances as estimated by the hypercube model was a factor of 2 as compared to 1.44 in the previous example.

The iteration by iteration changes in the sector travel distance and precinct wide average travel distance are summarized in Table 3.3. Initially the minimum average sector distance was .46 miles, in sector 3, while the maximum was .92 miles, in sector 7. After fifteen iterations of the program (22 c.p.u. seconds using 250K of core storage), the imbalance was reduced to 1.29 with the maximum average sector travel distance now .74 miles, 20% less than it was before. However, once again, we find that this improvement was at the expense of the precinct wide average travel distance which increased by 6% from .63 miles to .67 miles. In addition there was a deterioration in the workload imbalance, which changed from 1.29 to 1.59 (see Table 3.4). Interestingly, during the first seven iterations, there was an improvement in imbalances in both travel distances (from 2 to 1.78) and workloads (1.29 to 1.19) which suggests that within an intermediate range of imbalances an improvement in one imbalance (e.g. sector travel distance) need not be made always at the expense of the other (e.g. workload).

Although this subsystem in general and its interactive program in particular do not focus on imbalances in atom travel distances, an attempt at reducing high sector travel distances will naturally also impact on atom travel distances. Table 3.5 describes how the distribution of response distances in the New Haven precinct changed as a result of modifications in the sector design. Prior to the redesign, five atoms had average travel distances greater than one mile, with a high of 1.4 miles. Four more had average response distances between one mile and .9 miles. In the new design

Table 3.3 Balancing Sector Travel Distances:New Haven

Iteration	SECTORS (TRAVEL DISTANCES) 1	2	3	4	5	6	7	Ratio=High/Low	Average Travel Distance
Initial	.559	.513	**.463	.689	.521	.475	*.923	1.993	.634
1	.559	.513	**.464	.688	.529	.510	*.916	1.976	.634
2	.561	.514	**.465	.684	.545	.566	*.887	1.906	.630
3	.561	.515	**.465	.684	.547	.572	*.886	1.903	.630
4	.561	.516	**.467	.682	.561	.627	*.869	1.863	.632
5	.561	.516	**.467	.680	.574	.669	*.848	1.817	.632
6	.561	.516	**.467	.679	.581	.685	*.842	1.803	.633
7	.565	.516	**.468	.691	.580	.685	*.835	1.783	.634
8	.564	.518	**.469	.683	.653	.787	*.799	1.702	.649
9	.574	.521	**.473	.708	.655	*.785	.760	1.659	.652
10	.576	.523	**.485	.709	.654	*.783	.759	1.615	.652
11	.597	**.547	.580	.711	.649	*.761	.754	1.392	.658
12	.616	**.565	.635	.719	.646	*.755	.749	1.336	.666
13	.622	**.577	.656	.728	.644	.740	*.755	1.309	.669
14	.623	**.577	.657	.731	.644	.740	*.751	1.302	.670
15	.627	**.577	.657	.735	.645	*.742	.737	1.286	.671

Table 3.4 The Impact On Workloads Of Balancing Travel Distances: New Haven

Iteration	CAR WORKLOADS 1	2	3	4	5	6	7	Ratio=High/Low	Ratio=Low/High
Initial	.43	.38	.40	* .45	.38	** .35	.42	1.286	.778
1	.43	.38	.40	* .44	.38	** .36	.41	1.222	.818
2	.43	.38	.40	* .44	.38	** .37	.40	1.189	.841
3	.43	.38	.40	* .44	.38	** .37	.40	1.189	.841
4	.43	.38	.40	* .44	.38	** .37	.39	1.189	.841
5	.43	.38	.40	* .44	.38	** .38	** .38	1.158	.864
6	.43	.38	.40	* .44	.39	.38	** .38	1.158	.864
7	.43	.38	.40	* .44	.38	.39	** .37	1.189	.841
8	.43	.38	.40	* .44	.40	.43	** .32	1.375	.727
9	.43	.38	.40	* .45	.40	.43	** .31	1.452	.689
10	.43	.38	.40	* .45	.40	.43	** .31	1.452	.689
11	.43	.39	.41	* .45	.40	.42	** .30	1.500	.667
12	.44	.39	.41	* .45	.40	.41	** .30	1.500	.667
13	.44	.40	.41	* .46	.40	.41	** .30	1.533	.652
14	.44	.40	.41	* .46	.40	.41	** .30	1.533	.652
15	.44	.40	.41	* .46	.40	.41	** .29	1.586	.630

* Highest Workload ** Lowest Workload

no response distance was greater than one mile and only two were greater than .9 miles. Much of this improvement, however, was obtained at the expense cf atoms with short response distances as the number with response distances of less than .6 miles decreased from 33 to 12.

Table 3.5: Distribution of Atom Travel Distances: New Haven

	Over 1 Mile	.9-1	.8-.9	.7-.8	.6-.7	.5-.6	.4-.5	.3-.4
Initial	5 atoms	4	14	9	13	20	12	1
Final	0	2	13	21	30	10	2	0

3.9 Atom Travel Time

3.9.1 Effect of Atom Transfer

In discussing the effect of an atom transfer on the various atom travel times, two classes of atoms must be considered: 1) the atom transferred and 2) the remaining atoms in the sector, A, from which the atom was removed.

(E7) If the atom transferred, T, is closer (further) to the center of gravity of its new receiving sector than it was to the center of gravity of its old sector, there will tend to be a decrease (increase) in the average travel time for that atom T.[14]

(E8) The remaining atoms will be affected in two ways. For one, since there will be a reduction in the sector car's workload, there will be a decrease in the number of intersector (relatively far) dispatches into the sector. This will tend to reduce the atom travel times of all the remaining atoms.

(E9) Secondly, removing an atom, T, shifts the center of gravity of the old sector in a direction away from that atom, T. Consequently, atoms in the direction opposite from atom T will now be closer to the center of gravity of the sector which tends to reduce their travel times.[14]

3.9.2 Computer Program

(AS1) The program procedes to identify the atom with the worst travel time, T.

(AS2) A check is made to determine whether or not atom T is closer (and contiguous) to the center of gravity of a sector other than the one to which it is presently assigned. If the closest center of gravity is a sector, C, other than its own, the preferred option offered the user is to transfer atom T to sector C. (E7)

(AS3) The one (or two) atom(s), F, farthest away from atom T that is in the same sector and also

> contiguous to another sector is offered as a possible atom transfer. If that atom, F, is contiguous to more than one sector, it is to be transferred to the sector whose center of gravity is closest (maintain compactness, C2). (E9)

In most cases, step (AS2) will prove fruitless and only (AS3) will yield any potential atom transfers. Unfortunately, the last step (AS3) utilizes effects (E8) and (E9) which are rather indirect as compared to (E7) and as compared to those effects available in subsystems 1 and 2.

3.9.3 Computational Experience

In the initial sector design the district-wide average travel distance was .543 miles. Atom 70 (worst) had an average response distance of .965 while atom 23 (best) had an average response distance of .446. It required, on the average, more than twice as long (2.16) to respond to a call in atom 70 as in atom 23.

The optimal design produced by the non-interactive program was reached in 9 iterations. At that point, the worst atom was atom 60 which had an average response distance of .76, a reduction of 21% in the maximum average response distance for the atoms. Atom 23 still had the lowest average travel distance .477. The imbalance ratio (worst/best) was now 1.59, still significant, but a major improvement over the previous imbalance. Once again, part of the decrease paid for by increasing the district-wide travel distance .584 (an 8% increase) and by increasing travel distances for atoms with a low average travel distance. Because of this increase, the more appropriate measure of the program's worth would seem to be the previously mentioned maximum average response distance for the atoms. By either measure, the program significantly reduced the districts' imbalance. However, there is still a great deal of room for improvement and various modifications of the present program are being explored to see if the imbalance can be reduced further.

Each iteration required, on the average, 1.55 c.p.u. seconds, about the same as for the sector travel time program. Table 3 contains a list of the initial and final atom travel times.

To summarize, the computational results of the three non-interactive programs indicate that an iterative procedure is certainly feasible (time-wise). Each iteration requires a maximum of 1.55 c.p.u. seconds and the number of iterations needed to significantly reduce imbalances is not large.

Table 3.6 Balancing Atom Travel Distances: District 4, Boston

ATOM NUMBER	AVERAGE TRAVEL DISTANCE INITIAL	FINAL	ATOM NUMBER	AVERAGE TRAVEL DISTANCE INITIAL	FINAL
1	.494	.542	36	.510	.500
2	.512	.583	37	.539	.581
3	.458	.522	38	.634	.678
4	.592	.690	39	.482	.522
5	.540	.644	40	.507	.543
6	.512	.624	41	.580	.622
7	.501	.615	42	.611	.650
8	.534	.613	43	.537	.576
9	.484	.570	44	.651	.688
10	.458	.549	45	.710	.751
11	.448	.540	46	.623	.665
12	.575	.613	47	.547	.583
13	.529	.577	48	.539	.581
14	.508	.562	49	.537	.605
15	.504	.560	50	.672	.733
16	.486	.567	51	.656	.675
17	.461	.539	52	.557	.606
18	.463	.549	53	.485	.551
19	.463	.521	54	.782	.584
20	.589	.629	55	.715	.695
21	.517	.563	56	.551	.625
22	.457	.488	57	.507	.568
23	**.446	**.477	58	.511	.556
24	.467	.509	59	.606	.672
25	**.446	.483	60	.640	*.760
26	.484	.515	61	.604	.744
27	.485	.519	62	.591	.682
28	.506	.483	63	.640	.663
29	.499	.515	64	.709	.706
30	.462	.519	65	.668	.720
31	.465	.545	66	.703	.533
32	.447	.538	67	.781	.542
33	.460	.496	68	.819	.549
34	.536	.500	69	.757	.527
35	.662	.581	70	*.965	.688

* Largest Travel Distance (in Miles)

** Smallest Travel Distance (in Miles)

3.10 Amalgamate Sectors

The fifth subsystem will attempt to address the problem of which sectors should be combined to form larger sectors with a minimum imbalance in workloads. Because of the relatively limited flexibility of design, allowing only for gross modifications, (adding one sector to another rather than just one reporting area), it may be impossible to come up with a design that does not have significant workload imbalances. In addition, as a result of this limited flexibility, it is not possible to design an iterative procedure that parallels the previous subsystems. In an iterative procedure that would modify an initial beat configuration by transferring areas the size of sectors, the changes in workloads would be highly unpredictable because of the gross effects resulting from each modification. In a sense it would be no better than a poor unordered method of carrying out a total enumeration of all possible combinations.

3.10.1 A Slowest Ascent Algorithm

At present, no model for this subsystem has been developed but a number of directions are being explored. One approach is analogous to a steepest descent algorithm. For example, given a police district divided into b beats where there are only m cars to man, at least $b - m + 1$ and at most $2(b - m)$ sectors would be amalgamated into larger sectors.[15] The algorithm looks at all possible ways of combining only two sectors and then chooses the one with the smallest imbalance. The resulting design has only b - 1 sectors producing a sector design with only b - 2 sectors. The algorithm stops when there are only m sectors left.

The upper bound for the number of combinations that will have to be compared in the first iteration is $C_2^b = [b(b - 1)]/2$. For b = 6 this is 15. However, it is important to remember that because of a contiguity constraint, we combine only contiguous sectors so that this is only an upper bound. In the initial design for District 4, of the 15 possible combinations of the 6 sectors, only 8 satisfied this constraint. Although the number of combinations increases as b increases, it is also likely that the fraction of permissible combinations will decrease because there will, in general, be only a very limited number of sectors to which each sector will be contiguous. An example of this are the 48 continental states. Of the 1128 combinations of two, only 105 combinations, less than 10% would satisfy a contiguity constraint.

The slowest ascent algorithm, as described above, is complete as is; however, it is envisioned that it will be combined with a heuristic to reduce the number of combinations considered each time. The heuristic would be a simple procedure for characterizing which sectors are unlikely candidates for doubling up.

As with the previous systems, this methodology would also allow for user interaction. After considering the various alternatives for reducing the number of sectors from b to b - 1 to b - 2, etc., the user would then be provided at each step with information concerning the n (user defined) best options for reducing the number of sectors by one.

3.11 Present Stage of Development

For each of the first four subsystems, programs which independently equalize either workloads, preventive patrol coverage, sector travel times, or atom travel times have been written, debugged and run for data of District 4 in Boston. Although the nature of user interaction for each of these subsystems has been outlined, the structure of the user interaction has not yet been translated into computer programs. There are some questions that still have to be resolved with regard to the form of user interactive programs so as to make interaction as simple as possible. The ease of interaction is critical since the computer models are intended for the use of police managers with, at best, limited computer experience. Before a final version of the model is developed, it is hoped that feedback from various police managers will provide ideas for improving the interactive design but the basic format wil follow that of the interactive version of the hypercube model that was developed by Weisberg [17]. The last subsystem, doubling up on sectors, is still in the development stage and no computer programs have been written yet.

3.12 Summary

The models described in this paper represent an iterative approach to sector design. At each iteration a choice is made between alternative sector designs and the choice will depend upon the variables of primary and secondary concern to the police managers. Experience with the individual non-interactive programs has shown this approach to be rapid enough to be feasible. The usefulness of these models, however, will depend strongly on the individual police manager's ability to integrate quantitative and non-quantitative variables in choosing between sector designs.

FOOTNOTES 3

1 Determining what is a community's fair share of police protection is a difficult task and is a problem not really addressed in this chapter. It is hoped, however, that the use of quantitative measures will at least facilitate the evaluation of the level of police protection afforded each segment of the community.

2 The call rates are determined by using data aggregated for for the time period (patrol shift) for which this sector design is intended.

3 If two atoms are contiguous but there is a barrier between them (e.g. a river), they are treated as if they were non-contiguous.

4 Computational experience has shown that p should not be set less than 5.

5 This problem can be avoided, to some extent , by constraints similar to C4.

6 To simplify the analysis, each effect is treated as if it were independent of every other effect though in fact the effects are obviously related.

7 From now on it is to be understood that all atom transfers have been first checked by procedure (GP5) to see if they do not violate the contiguity constraint (C1).

8 All computational results were obtained on an IBM 370/165 computer at the M.I.T. Information Processing Center.

9 Each effect is treated separately even though the effects are interrelated.

10 However, one ripple effect is that by transfering the atom to an adjoining sector and increasing its workload, the average distance travelled on an intersector sector dispatch will tend to increase.

11 Computational results obtained by various modifications of the initial sector design used here seem to indicate that the district-wide average response time for this design was close to the minimum possible.

12 Because travel speeds were set to be 1, this number represents the average distance travelled in miles.

13 For consistency we continue our use of the words precinct to describe a collection of sectors and sector to describe a single patrol unit's area of responsibility even though in New Haven the respective terms are sector and beat.

14 The relationship between the distance to the center of gravity of a sector and the atom travel times is not as direct as it would seem. The reason is that the travel time for sector car A' to atom C is not calculated by using the distance from the center of gravity of sector A to the center of gravity of atom C. Instead it is calculated by finding the expected distance sector car A' must travel in order to reach the center of gravity of atom C. The two are not equivalent. Larson [ ].

15 It is not always 2(b - m) because a larger amalgamated sector car can be formed out of more than two smaller sectors.

REFERENCES 3

1. Ackoff, R.L., "Management Misinformation Systems", Management Science, December 1967.

2. Chapman, S.G., Police Patrol Readings(Second Edition). Chapter 4, "Patrol Force Distribution", Charles C. Thomas, publisher, Springfield, IL, 1972.

3. Chelst, K.R., Implementing the Hypercube Queuing Model in the New Haven Department of Police Services: A Case Study of Technology Transfer, New York City-Rand Institute Report, R-1566/6-HUD, New York City, NY, July 1975.

4. Colton, K., "Use of Computers by Police: Patterns of Success and Failure", Urban Data Service, Vol. 4, No. 4, April 1972.

5. Ficklin, L.R., "Police Manpower Formula?", published in Police Patrol Readings. Samuel G. Chapman, ed., Charles C. Thomas, publisher, Springfield, IL, 1972.

6. Ford Jr., L.R. and D.R. Fulkerson, Flows in Networks, Princeton University Press, Princeton, NJ, 1962.

7. Larson, R.C., Urban Police Patrol Analysis, M.I.T. Press, Cambridge, MA, 1972.

8. Larson, R.C., " A Hypercube Queuing Model for Facility Location and Redistricting in Urban Emergency Services", Computers and Operations Research, Vol. 1, No. 1, March 1974.

9. Larson, R.C., "Illustrative Police Sector Redesign in District 4 in Boston", Journal of Urban Analysis, Vol. 2, 1974.

10. Larson, R.C., Urban Emergency Service Systems: An Iterative Procedure for Approximating Performance Characteristics, The New York City-Rnad Institute, R-1493-HUD, January 1974, to appear in Operations Research.

11. Little, J.D.C., "Models and Managers: The Concept of a Decision Calculus", Management Science, Vol. 18, No. 9, May 1970.

12. McClaren, R.C., "The IACP Allocation and Distribution Method"(1967), published in Police Patrol Readings (Second Edition), Samuel G. Chapman, ed., Charles C. Thomas, publisher, Springfield, IL, 1972.

13. McKnew, M. personal communication.

14. Smith, S.B., et al., Superbeat: A System for the Effective Distribution of Police Patrol Units and Superbeat Program Manual, Illinois Institute of Technology, 1973.

15. Vollmer, A., "The Police Beat"(1933), published in Police Patrol Readings (Second Edition), Samuel G. Chapman, ed., Charles C. Thomas, publisher, Springfield, IL, 1972.

16. Walton, F.E., "Selective Distribution of Police Force", published in Police Patrol Readings (Second Edition), Samuel G. Chapman, ed., Charles C. Thomas, publisher, Springfield, IL, 1972.

17. Weisberg, R.W., Using the Interactive Hypercube Model, Massachusetts Institute of Technology Operations Research Center Technical Report TR-17-75, June 1975.

CHAPTER 4

THE BASICS OF SEARCH THEORY APPLIED TO POLICE PATROL

4.0 Introduction

Classical search theory addresses two classes of problems. The first class centers about calculating the probability of intercepting a target which exhibits certain characteristics (e.g. stationary, non-stationary, evasive, etc.) under a specific search pattern (e.g. random, parallel sweeps, cross over barrier patrol, etc.) [18]. Included in this class is the analysis of the operating characteristics of the detection instrument, be it radar, sonar or simply the human eye. The second class of problems involves the determination of an optimal search strategy.

In this chapter, we will begin our analysis of the application of search theory to problems of police patrol. The first sections of this chapter will focus on a basic search theoretic model of police patrol which is used to calculate the probability of intercepting a crime. The presentation will include an analysis of each of the model's input parameters as well as a discussion of the model's implicit assumptions. However, the discussion of how to obtain estimates of the various parameters and the inherent difficulties in the task will be postponed until our presentation of algorithms and examples in Chapter 6. Next, using the basic model, we will explore the differences between overlapping and non-overlapping patrol sectors.

The second half of this chapter will focus on issues relating to the optimal allocation of patrol effort. The major part of this section will involve a fundamental analysis of the characteristics of optimal search strategies when targets (i.e. crimes) arrive randomly and independent of each other and then depart after a rather limited duration. This analysis will use Koopman's pioneering work [13] in search theory as a background and show the limited applicability of his results to the situation in which the targets exhibit the above characteristics. The goal of this discussion will be to develop a qualitative understanding of the search process while pinpointing some misconceptions that have arisen in the application of earlier search theory results to police patrol. These qualitative concepts will be quantified later in Chapter 5 with the use of a differential equation model of search and detection. Finally Chapters 6 and 7 build on the insights developed here to create algorithms for deploying both tactical and standard patrol units.

4.1 A Search Theoretic Model of Police Patrol

A number of recent papers [7, 8, 9, 14, 19, 20] have applied the work of Koopman [12] to the problem of calculating the probability of a randomly patrolling police car intercepting a crime. The general model they follow, with perhaps one or two minor modifications, is as follows:

t = the observable duration of a crime,

d = probability of detecting the crime conditioned upon passing it in progress,

m = the total number of street miles in the area patrolled,

s = the speed of patrol,

p_i = probability of interception

$$\text{where } p_i = 1 - \exp(-s \cdot t \cdot d/m) \qquad (4.1)$$

The exponential form of this equation stems from the use of the exponential distribution as an approximation to the binomial distribution. The actual underlying model of patrol (the justification for equation (4.1)) treats the patrol path of s . t miles as a series of very small independent paths of length le. On each path of length le, the probability of finding the target is $(1 - d \cdot le/m)$. Consequently a path of length s · t miles has a probability of $(1 - d \cdot le/m)^{st/le}$ of not finding the target and $[1 - (1 - d \cdot le/m)^{st/le}]$ of finding it. Applying the exponential approximation to this binomial distribution yields equation (4.1).

In introducing this first basic model of patrol a number of assumptions were made. Some were made so as not to complicate our initial discussion of the model and will be relaxed later with only minor modifications of the model. While some others can not be eliminated without necessitating an almost entirely different approach to modeling police patrol. In the first category falls the assumption that at the time the crime is committed the patrol unit is on patrol and not responding to a call for service. Consequently this initial model is more relevant to a tactical patrol force or to the patrol units in a split patrol force that do not generally respond to calls for service. Later in section 4.1.2, we modify the model to make it applicable to standard patrol units which spend a good percentage (often more than 50%) of their time on tasks other than patrol.

Secondly equation (4.1) calculates the probability of intercepting a crime of a fixed observable duration of t minutes. However of at least equal importance is the ability to carry out similar calculations for crimes of a specific class (e.g. street robberies) which may have a common mean but whose observable duration is a random variable. In Chapter 6 we discuss calculating the probability of interception when a probability distribution is given for the observable duration.

One assumption that can not be modified without immediately getting enmeshed in complex game theoretic modeling is that crimes occur independent of the location of the patrolling vehicle. One justification for making this assumption is that we are considering only randomized non-predictable patrol strategies. This unpredictability limits, but by no means eliminates, the potential usefulness to a criminal of knowing the location of the patrol unit when committing a crime. For example, a criminal might wait to commit a crime until he knows that the patrol unit is at the other end of the sector. This would provide him with a minimum time span in which he need not be concerned about being intercepted by a passing patrol unit. With a standard patrol unit there is an additional potential violation of the independence assumption. Criminals may monitor the police radio in order to initiate their crimes when the local sector car is busy responding to a call for service [14]. Should it be found that criminals frequently use information about the location and status of patrol units when committing crimes, then the potential applicability of equation (4.1) and variations on it would be limited. One would expect though that at least for unmarked patrol units (e.g. taxicabs) with police officers in civilian clothes the validity of the independence assumption would not be a significant problem. In addition Larson [14] suggests a number of strategies (e.g. overlapping patrol sectors) which can be used to reduce the gain to a criminal of monitoring an individual patrol unit's activity. These strategies would thereby restore the independence between the time and location of a crime and the status of a patrol unit.

Although the above equation will be discussed throughout this and succeeding chapters, for now we would like to comment briefly on just one of its components. The observable duration of a crime, t, is perhaps, the most critical parameter of the equation. It represents the amount of time (in search theory terminology: the total search effort) available to a single patrol unit to detect a random crime even though that patrol unit may have four or eight hours to spend on preventive patrol. Its small magnitude, often two minutes or less, is a major reason why even very few potentially observable crimes are ever intercepted by patrolling police units. The only way to increase the amount of search effort available is by adding patrol units. Doubling the number of patrol units doubles the available search effort, but unfortunately does not double p_i, the probability of interception, because of the exponential nature of equation (4.1). When additional units are added, the model becomes

n = the number of patrol units,

$$p_i = 1 - \exp(-n \cdot s \cdot d \cdot t/m) \qquad (4.2)$$

This equation applies if each unit patrols its own separate sector of m/n miles. Also, by making a number of minor assumptions (to be discussed in the next section), it can be shown to apply as well to n units patrolling randomly the entire m miles.

4.1.1 Manpower Requirements

Equation (4.2) does not represent a manpower allocation formula. However, by looking at the inverse (i.e. solving for n), it is possible to determine the number of patrol units needed to obtain a specified level of interception, p_i (see Elliott [9]).

$$n = (-m/s \cdot t \cdot d) \ln(1 - p_i) \tag{4.3}$$

In using this equation, it is once again important to understand the impact of t, the observable duration of a crime. To obtain the same specified level of interception, significantly different manpower levels will be required for the crimes of purse snatching and commerical burglary because their durations vary so tremendously. Consequently careful consideration of the duration of various crime types should be given before developing crime specific desployment strategies. This awareness, in fact, may be the critical factor in choosing between totally different strategies--whether to increase patrol strength or set up decoys. In pursuing this approach, an obvious prerequisite is the estimation of the duration of the various crimes. (This issue will be discussed later in Chapter 6.)

4.1.2 Standard Patrol Car

The previous discussion and equations more accurately model a patrol unit whose sole responsibility is to search for crimes (e.g. a tactical patrol unit) rather than the typical patrol car which spends a significant fraction of its time responding to calls for service or handling matters totally unrelated to crime. To model the latter case the equations must be modified to allow for the fact that at the time the crime is being committed the patrol car may be responding to a call for service elsewhere or may be otherwise occupied.

Let b = the average fraction of time a patrol unit is busy and therefore not on patrol.

If t is small in the sense that during the entire duration of the crime a unit is likely to be either busy or free, then equation (4.1) becomes

$$p_i = [1 - b] \cdot [1 - \exp(-s \cdot t \cdot d/m)] \quad (4.4)$$

If, however, t is large such that during any time period a patrol car will be busy about $(b \cdot t)$ minutes, then a good approximation could be

$$p_i = 1 - \exp[-(1 - b) \cdot s \cdot t \cdot d/m)] \quad (4.5)$$

For an intermediate range of t the equation becomes more complicated.

f(a) = the probability density function of the fraction of time available for patrol during the period of t minutes.

$$p_i = \int_0^1 f(a) \cdot [1 - \exp (a \cdot s \cdot t \cdot d/m)]da \quad (4.6)$$

In this last expression the average available fraction of free time, (1 - b), is replaced with a probability distribution, f(a). This distribution, though, is not easy to determine because each unit's time available for patrol is a complex function of its interactions with the other patrol units in its precinct. (It will often respond to calls in other sectors as well as vice versa.) Its calculation requires a not at all obvious extension of the hypercube model. This is because the points in time at which an individual patrol unit becomes available for patrol do not constitute a renewal process [11].

Naturally, of the above three equations, the last is the most accurate. However, given the complexity of determining f(a) and the relatively short observable duration of crimes as compared to the time spent on a call for service, we suggest using expression (4.4) as a good approximation, albeit overestimation, for the probability of interception. In Chapter 7 we use this approximation in developing algorithms for finding the portion of a sector in which a standard patrol should concentrate its efforts.

4.1.3 Parameter Control

At this point it may prove useful to review the various parameters to see what control can be exercised to increase the probability of interception [4]. Obviously n, the number of patrol cars, can be increased. Also, b, the fraction of time busy, can be decreased by changing the dispatch policy to eliminate service calls that could be handled, just as well, by someone other than a patrol officer. This will result in an increased probability of intercepting a random crime. For example decreasing b from .6 to .4 by definition increases the patrol units availability, (1 - b), from .4 to .6. The result is a 50 percent increase in the probability of interception as calculated using the approximation (equation 4.3). However, this increased interception <u>rate</u> will not necessarily generate an equivalent increase in the actual number of <u>arrests</u> since criminals may react to the increased patrol by shifting their activities elsewhere. This issue of the impact on crime patterns of increased or concentrated patrol will be addressed in greater detail in Chapters 6 and 8.

The speed at which officers patrol, s, can also be increased; however, there is an inverse relationship between s and d since increasing the speed reduces the probability of observing a crime in progress [8]. On the other hand, there are alternatives for increasing d without reducing the speed of patrol. Better street lighting would increase the observability of street crimes. Instituting a concerted effort to have homeowners leaving for extended vacations notify the police would also increase the probability of a passing patrol car noticing a burglary in progress. Any activity around the vacated house would immediately be suspicious. Similarly a blinking light attached to the front of a store that a storeowner triggers to signal a robbery in progress would have a similar effect. To some extent m, the number of street miles, can also be controlled by concentrating patrol in the highest crime areas. Lastly, one pessimistic note of interest can be sounded with regard to t, observable duration of a crime. Increasing the frequency of patrols may cause criminals to work faster thereby reducing t.

4.2 Comparison of Overlapping and Non-Overlapping Patrol

Earlier we had stated that for an n man tactical patrol force, totally overlapping and non-overlapping patrols have the same probablity of intercepting a random crime assuming that crimes arise in a geographically homogeneous manner. In this section we will outline a proof of this as well as specify the underlying assumptions. In addition we will show that for a standard patrol force (often busy responding to calls for service) the two policies are not equivalent and that for a specific set of assumptions overlapping patrols have a higher probability of interception. Lastly at the close of this section we will digress from the search theoretic models in order to apply the hypercube queuing model [15] to measure how overlapping and non-overlapping patrols compare with regard to travel time.

4.2.1 Probability of Interception

The first problem to be addressed is the calculation of the probability of interception for non-overlapping and overlapping patrol when carried out by a tactical patrol force. If the region (m street miles) is divided into n separate sectors then each patrol unit is patrolling m/n street miles. In addition, since crimes are assumed to arise uniformly over the entire region, the probability of a random crime occurring in any particular sector is 1/n. Thus the probability of intercepting a random crime with non-overlapping sectors is just the sum (over all sectors) of the probability of a crime occurring in sector j and patrol unit j discovering it.

$$p_i = \sum_{j=1}^{n} \frac{1}{n} [1 - \exp(-s \cdot t \cdot d \cdot n/m)]$$

$$= 1 - \exp(-s \cdot t \cdot d \cdot n/m) \qquad (4.7)$$

If on the other hand all n patrol units independently patrol the m street miles then the probability of a particular patrol unit, j, not detecting the crime is $\exp(-s \cdot t \cdot d/m)$. (This is an approximation in that we have not included a second order effect, which involves unit j not intercepting the crime because another unit has already intercepted it.) Since each unit patrols independently of all the other units, the probability of none of the n units intercepting the crime is

$$[\exp(-s \cdot t \cdot d/m)]^n = \exp(-s \cdot t \cdot d \cdot n/m)$$

and

$$p_i = 1 - \exp(-s \cdot t \cdot d \cdot n/m)$$

the same as before.

The above analysis could have been performed similarly for geographically non-homogeneous crime rates and the result would have been the same. The only change in the analysis would have been that the 1/n term in equation (4.7) would have to be replaced by p_{jc}, the probability of a crime occurring in sector j given that it occurs somewhere in the entire region. However, since the sectors are still of equal size and the p_{jc} sum to one, p_i, the probability of interception, does not change.

Although we have shown that for equal sized sectors the search theoretic model produces the same probability of interception for totally overlapping and non-overlapping patrol, it would be incorrect to infer that the two strategies are totally equivalent. As we have noted earlier, the search model assumes that the criminal selects his victim independent of the location of the patrol unit. This, in fact, may not always be the case. Larson [14] points out that one advantage of overlapping patrol is that a criminal would have to keep track of all n patrol units in order to be sure that no patrol unit were nearby; while for non-overlapping patrols, he would only have to keep track of the local sector car. Thus from this perspective overlapping patrol might have a greater chance of catching(or perhaps deterring) a criminal. Conversely, non-overlapphing sectors allow a patrol unit to become better acquainted with its sector , therefore enabling it to notice more easily things which are out of the ordinary. The above two examples represent but a few of the issues that need to be considered when comparing overlapping and non-overlapping patrol but which are not incorporated into the search model. Throughout this section, though, we will be focusing on either one or two performance measures with our models. In the process we will be making a number of assumptions as well as leaving out numerous other issues (in general those that are not easily quantifiable) which should be kept in mind when translating our results to a specific application.

4.2.2 Standard Patrol Unit

In contrast to the previous discussion, the probability of interception by a standard patrol unit will be different for overlapping and non-overlapping patrols. Using equation (4.4) which assumes crimes of relatively short duration, we will compare non-overlapping and overlapping patrol strategies for standard patrol units. Three characteristics of the specific problem to be discussed are

1. Crimes arise uniformly in space.
2. The fraction of time each unit is busy is b and is independent of the particular policy.
3. Each sector is the same size.

For two patrol cars each patrolling its own sector and for small t, the equation for the probability of interception is

$$p_{in} = \sum_{j=1}^{2} \frac{1}{2} \cdot [1 - b] \cdot [1 - \exp(-2 \cdot s \cdot t \cdot d/m]$$

$$= [1 - b] \cdot [1 - \exp(-2 \cdot s \cdot t \cdot d/m)] \qquad (4.8)$$

In this expression the 1/2 is the probability of the committed crime occurring in sector j and (1 - b) is the probability that the patrol unit in sector j is free and on patrol. With a totally overlapping patrol policy the probability of interception, p_{io}, is

$$p_{io} = [(1 - b)/ (1 + b)] \cdot [1 - \exp(-2 \cdot s \cdot t \cdot d/m)] + [2b \cdot (1 - b)/(1 + b)] \cdot [1 - \exp(-s \cdot t \cdot d/m)] \quad (4.9)$$

The terms $[(1 - b)/(1 + b)]$ and $[2b \cdot (1 - b)/(1 + b)]$ represent the probability that two or one server respectively is not busy. (See Appendix B.) Thus the first term in expression (4.9) is the probability of (intercepting a crime and both units are on patrol) and the second term the probabilty of (intercepting a crime and only one unit is on patrol).

In Appendix C it is proven that for the above example the probability of interception with an overlapping patrol, p_{io}, is higher than that for a non-overlapping patrol, p_{in}. The magnitude of this improvement is

$$p_{io} - p_{in} = [b \cdot (1 - b)/ (1 + b)] \cdot [2(1 - \exp(-s \cdot t \cdot d/m)) - (1 - \exp(-2 \cdot s \cdot t \cdot d/m))] \quad (4.10)$$

An analysis of this gain shows that it is directly proportional to the probability of only one server being busy. In addition it monotonically increases as s or t increases. However the difference will be very small since the expression $[2(1 - \exp(-s \cdot t \cdot d/m)) - (1 - \exp(-2 \cdot s \cdot t \cdot d/m))]$ is close to zero. If the product, $s \cdot t \cdot d/m$ were .01 (approximately one percent chance of intercepting a crime) then the expression is equal to approximately .0001. In general these results seem to be analogous to the improvement produced when two one-server queuing systems merge into a one-queue, two-server system.

The above small improvement should not, however, be used as the real measure of the potential improvement that can be generated from overlapping patrol. Up to now we have considered only uniformly distributed crimes. Although it is easy for a police decision maker to allocate his resources when crimes are distributed uniformly, non-uniform crime rates allow for the concentration of patrol in the areas of high crime which would result in an overall higher probability of intercepting a random crime. In our concluding remarks on interception probabilities for both patrol strategies, we will outline some of the differences between the two strategies when calls are not distributed uniformly and the workloads are not the same.

Assume sector A has high crime and call for service rates and sector B, low crime and call rates. Consequently, if each sector is assigned a single patrol unit, we would expect for the unit in sector A to have the higher workload (even though we are allowing intersector dispatches). The result is that the area with the higher crime rate has fewer hours of preventive patrol and consequently a smaller probability of intercepting a crime there given the sectors are the same size. If instead, the two units jointly patrol the two sectors and call assignments are alternated, then, first of all, both units will have equal workloads which is an advantage in itself. Secondly, the higher crime rate sector will receive at least an equal share of patrol coverage. Perhaps more importantly the overlapping patrol policy provides the added flexibility to allocate an even larger proportion of the patrol effort in the higher crime area which is not possible when each patrol unit is assigned to its own sector. (It should be noted that the same type of flexible patrol allocation can be accomplished also with a split patrol force in which the responding cars are assigned separate sectors and the remainder of the patrol force is assigned to the highest crime areas.)

The potential payoff from being able to allocate a greater proportion of the patrol effort to the area where it is needed most will be significantly greater than the queuing type improvement discussed earlier in this section. The magnitude of the improvement will be directly related to the degree of nonuniformity in the crime rate and the amount of imbalance in workloads. For example assume sector

A generates 60% of the crimes and that its patrol unit's workload is .60 while sector B generates only 40% of the crimes and its unit's workload is .40. Secondly assume that the product, s · t · d/m, is .01. Then by changing from non-overlapping to overlapping sectors and even allocating patrol effort in proportion to the crime rate (which is not the optimal strategy) there would be approximately an 8% increase in the probability of interception. With an optimal allocation it would improve by 10% or more.

However, before rushing off to recommend a change from non-overlapping to overlapping sectors, the impact of this type of change on other performance measures must be evaluated. In the following section the focus will be on average travel time (which is also related to catching a criminal at the scene of the crime [21]) which in contrast to interception probability generally improves under non-overlapping patrol.

4.2.3 Overlapping Sectors: Impact on Travel Times

Larson in his book, Urban Police Patrol Analysis [14], discusses overlapping sectors as a potentially more flexible alternative to the more widely used non-overlapping sectors. The major advantages of overlapping sectors are that it decreases the probability of patrol coverage being reduced to zero and, in general, increases the difficulty for a criminal to monitor the activity of patrol units and plan his crimes accordingly. One issue Larson addressed in depth, as part of an evaluation of car locator systems, was a comparison of travel times in a system with overlapping sectors and perfect car location information and travel times in current systems with non-overlapping sectors and no explicit car location information.

The general conclusion was that:

"If sectors were eliminated and each car were to patrol uniformly one large area, independently of other cars, and if perfect resolution car location information were used to dispatch the closest available car, then the travel time characteristics of this overlapping sector system are nearly identical to those of SCM (strict center of mass dispatching; see Glossary for definition) system with nonoverlapping sectors."

Since his discussion of overlapping sectors was presented in the context of evaluating an automatic car locator system (which makes it feasible for even all of a precinct's units to patrol the entire precinct), the focus was on comparing overlapping patrol with perfect car location information to non-overlapping patrol under SCM or MCM (modified center of mass; see Glossary for definition). In this section, we will present a much more limited discussion than that of Larson of the impact on travel times of overlapping patrol but with no car location information. The tool to be used in this analysis is the hypercube model which has the capability of modeling various levels of overlapping patrol [15]. Our focus will be on combining only two sectors. It is a policy that can be implemented without an automatic car locator system and is presently used in varying degrees by many police departments in the United States. We will assume that a patrol unit's available time is divided among the sector's atoms in proportion to each atom's call rate. The dispatch policy will be MCM with one modification needed to describe the dispatching of patrol units within overlapping sectors. In our examples one of the pair of patrol units in the overlapping sectors was designated the primary responding car. When a call for service arises in the overlapping sectors, the primary car is always sent if it is available and only when it is not available is the other patrol unit dispatched. In general, though, the impact of overlapping patrol on travel times would be essentially the same if, instead, calls were shared equally by the two patrol units.

For a range of average workloads (from a low of .1 to a high of .9) we ran the hypercube model twice, using data from District 4 in Boston [16]. In one set of runs, sectors 5 and 6 were separate; in another they were combined to form one larger overlapping sector. (Calls for service were allowed to be queued.) For low utilization there was a significant degradation of travel times (See Figure 4.1). When the average workload was .1, combining the two sectors increased the average travel time to the combined region by 53%; for an average workload of .30 the increase, although smaller, was still 30%. As the average workload increased further to .50, the difference in travel times was reduced to 15%. Thus as the average workload increased the deleterious impact on travel times continues to decrease and as the system nears saturation (average workload .9) the differences between overlapping and non-overlapping are almost eliminated (1% difference).

The explanation of this phenomenon is as follows. With low average workloads and non-overlapping sectors, almost all calls for service in sector 5 will be answered by the patrol unit located in the same sector; the same is true for sector 6. However, when the two sectors are combined, the car that is dispatched to a call in sector 5 will often be (approximately 50% of the time) on patrol in sector 6 and have to travel much farther to the call. (Remember we do not use or have car location information.) However, when the average workload increases, even with non-overlapping sectors the patrol car in sector 6 will be sent frequently to calls in sector 5 because the local sector car will often be busy answering another call. In addition as the average workload rises significantly, larger and larger proportion of calls in sector 5 will be answered by sector cars that are even further away than car 6. As the proportion of these overly long travel times increases, they will tend to dominate a statistic such as the average travel time thereby reducing further the impact of overlapping pairs of sectors.

An interesting footnote to the comparision of the two policies is that the imbalance in response times between sectors 5 and 6 was not very different for the two policies.

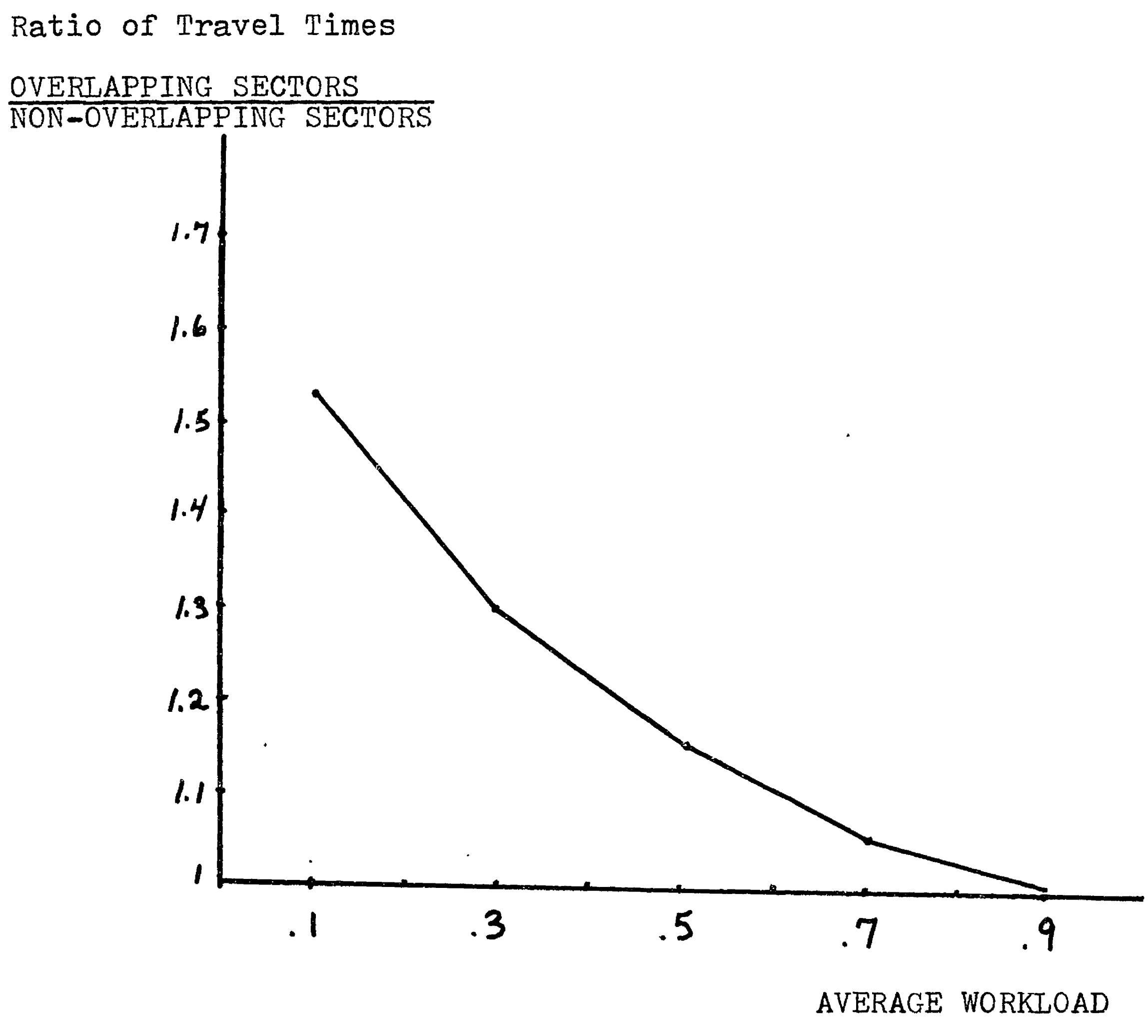


Figure 4.1: A Comparison of Travel Times for Overlapping and Non-overlapping Sectors for a Range of Workloads

For a .30 average workload, sector 5 had a 23% higher travel time than sector 6 with overlapping sectors and a 30% higher travel time with non-overlapping sectors. With a .50 average workload, sector 5's travel times were 34% and 36% higher with overlapping and non-overlapping patrol respectively.

4.2.4 Conclusions: Overlapping vs. Non-Overlapping Sectors

The only clear cut conclusion about the preferred strategy that can be stated with any certainty is that it will depend upon the particular circumstances. With this in mind we will present our conclusions for a number of situations; at times, however, our conclusions will be merely a clearer statement of the tradeoffs that the decision maker should consider. Our presentation will first focus on situations in which a car locator system is not available and later discuss how the recommendations would change if a car locator system were available.

Two nonquantifiable issues that will be relevant to each of the situations discussed are 1)Overlapping patrols reduce the criminal's abilbity to choose his place and time in such a way that he is assured of not being spotted by a passing patrol car; 2)Non-overlapping patrol enables the patrol officer to obtain a better knowledge of his particular sector. Each decision maker must assess the relative importance of each of these issues in his locality. Do criminals frequently plan their activities by monitoring patrol movements? How critical is it for the patrol officer to develop an extensive knowledge of the community he patrols?

With these issues as background, we proceed with a presentation of our conclusions for five different situations:

1) No Car Locator Information: Relatively Homogeneous Call Rates and Crime Rates

If the crime rate in a region were relatively homogeneous and no serious workload imbalances existed, then essentially the only advantage to overlapping patrol is the above mentioned difficulty of tracking. However, in the range of workloads within which most police departments operate overlapping sectors would significantly increase travel times, anywhere from 15 to 50 percent. Thus unless there were hard evidence to show that many crimes were planned and committed so as to coincide with the sector car being either unavailable or far away, non-overlapping sectors are preferable.

2) No Car Information: Low Utilization and Non-Homogeneous Call Rates and Crime Rates

Given the tremendous increase in travel time (on the order of 50%)[1] that would result from changing to overlapping sectors it would require extremely large disparities in crime rates to justify the change. It is important to remember that even for large disparities in crime rates the tradeoff under consideration is not just between higher police initiated interceptions and higher citizen initiated (police responding) on site apprehensions. The average travel time has wider significance since police also respond to a large number of non-criminal emergencies. In addition citizen perception of the quality of police services may be directly related to the rapidity with which police respond to a call for service. In short it would be necessary in this situation to estimate accurately both the impact on travel time and the potential increase in interception probabilities resulting from a better matching of patrol to crime. However we would expect the significant increase in travel time to outweigh the other benefits.

3) No Car Location Information: High Utilization and Non-Homogeneous Call Rates and Crime Rates

Once the utilization approaches .50, the increase in travel times resulting from overlapping sectors would seem to be in an acceptable range (15%) given other benefits that may accrue. With the variations in call rates, we would expect imbalances in workloads and an even more serious misallocation of patrol effort. With the potential for increasing interception probabilities by more than 10%, overlapping sectors are an attractive option to seriously consider. Under these circumstances it may be that the other issues we have mentioned will decide the issue in favor of one policy or the other.

4) Car Location Information: Homogeneous Call Rates and Crime Rates

With the introduction of a car locator system, travel time is no longer a serious consideration. However because call rates are geographically homogeneous, there is no real prospect of directly improving interception probabilities through the better allocation of patrol effort. The choice thus evolves into an analysis by the decision maker of the two issues with which we began this discussion, monitoring difficulty vs. neighborhood familiarity.

5) Car Location Information: Non-Homogeneous Call Rates and Crime Rates

The preference for overlapping sectors seems to be the most clear cut in this situation. Overlapping sectors would allow a better allocation of patrol without significantly degrading travel times. The only drawback would therefore be that with the larger regions to patrol, patrol officers may lack the intimate knowledge of their area that they might otherwise have.

4.3 Problems in Applying Search Theory to Patrol Allocation

The second question that has been approached from a search theoretic viewpoint is the allocation of police patrol effort in a region where the spatial distribution of crimes is non-uniform. On this problem Larson [14] has brought to bear graphical techniques of Koopman [13] while Olson [19] has made use of the equivalent but more easily computerized analytical techniques developed by Charnes and Cooper [5]. In applying these techniques to the allocation of patrol both authors have constrained the patrol car to patrol in a random[2] manner with a resultant reduction in the theoretical efficiency [14]. The motivation behind this type of patrol, as was mentioned earlier, is that it reduces the possibility of evasive action (a difficult thing to model) on the part of the criminal. Elliott [9], on the other hand, in allocating police resources has used a linear approximation to the exponential distribution to calculate the probability of interception. He then proceeded to develop a method for finding a totally deterministic patrol route which would optimize the probability of interception of a crime.

4.3.1 Classical Search Theory: The Allocation Problem

Before proceeding to analyze the allocation problem in the police context, it is necessary to provide first a brief introduction to the classical search theory problem. For simplicity, assume there are three regions, A, B, and C, of equal size but with different probabilities (which sum to 1) of a target being located in each and a total available search effort of 1 hour. Assume further that A has the highest probabilitiy of the target's being in it, B the second highest, and C the lowest. The optimal search strategy might be to search A for 40 minutes, B for 20 minutes and C not at all. This optimal solution would satisfy the following equilibrium characteristic.

A differential increase in the search allocation to A, dt_a, and a differential increase in the search of B, dt_b, will increase equally the probability of interception, p_i or $\frac{dp_i}{dt_a} = \frac{dp_i}{dt_b}$. In addition, both differential increases, dt_a and dt_b, have higher payoffs than an initial allocation of search to region C. However, as more than 1 hour becomes available, because there is diminishing return ($\frac{dp_i}{dt_a}$ is decreasing as t_a increases), the region C will eventually be allocated some search effort. Conversely, if less than one hour of search were available, for example only 5 minutes, it is conceivable that the optimal solution would direct all search efforts to region A. In short, the less effort available, the smaller the region to be allocated search.

4.3.2 Applying Classical Search Theory

In applying the Koopman and Charnes and Cooper techniques to police, the time t, the observable duration of a crime, once again plays an important and unfortunately dual role. For one thing t, or a multiple of t for several patrol units, represents the total search effort available for allocation and not the eight hours of patrol. Because t is typically very small, often less than five minutes, the likelihood that the optimal solution will be to patrol only a very small high crime area is great. Secondly, an optimal allocation (using classical search theory) for a 5 minute crime, of three minutes in region A and two minutes in B, does not directly address the problem of how to allocate the 4 hours or more available for patrol. An initial response [14] might be to spend three fifths of the time (2.4 hours) patrolling A and two fifths of the time (1.6 hours) patrolling region B. However, there is a fundamental distinction between patrolling A for 2.4 hours and then patrolling B for 1.6 hours and patrolling B for 2 minutes, A for 3 minutes, etc. Under the former strategy after the first three minutes in A and later after the first two minutes in B, the search is incurring a diminishing return that was not accounted for in the original solution (i.e. In the original solution no search effort and consequently no diminishing return occurred beyond the basic 3 minutes in A and 2 minutes in B.) This second constraint seems to place infeasible limitations on implementing the optimal strategy, especially when one uses numbers like 45 seconds for a street robbery [19]. This fact alone raises serious questions about the straightforward applicability of search theory to finding optimal strategies for looking for crimes of a short duration.

4.3.3 Implicit Assumptions

An even more fundamental objection must be raised, however, to the earlier applications of search theory to this problem. The equations used for calculating the probability of interception were similar to equation (4.1), which is of the form amenable to the classical search theory approach. This equation, however, as previously mentioned calculates only the probability of interception <u>conditional</u> on a crime occurring. In allocating search effort, though, at the time the search begins, a crime may or may not be in progress and, even if it is in progress, its observable duration over the time period of the search may no longer be t. In essence no adjustment was made for the fact that crimes are starting and ending at random points relative to the start of the search effort. In effect a straightforward application of search theory makes the following unacceptable implicit assumptions:

1. Crimes occur every t minutes.
2. Crimes last t minutes.
3. Patrol has started at the beginning of a cycle.

In the following sections we will do two things. Firstly, equations will be presented for calculating the probability of intercepting a crime during a search of fixed duration. These equations will not be conditioned on the presence of a target and will allow for the random arrival of a target. They represent the more appropriate equations to be used in allocating patrol effort if a classical search theory approach is to be used. Following that, we will discuss in qualitative terms, using basic search theory concepts (e.g. diminishing return) how random independent arrival and short finite duration of targets (e.g. crimes) impact on optimal search strategies. The results of this discussion will be to lead us in the direction of developing simple but potentially robust heuristic patrol allocation algorithms (Chapters 6 and 7) which do not make use of the allocation methodologies of Koopman [13] and Charnes and Cooper [5]. Instead they will be designed to take advantage of some of the insights we will develop here.

4.3.4 A Search of Fixed Duration

In order to avoid making the above three implicit assumptions in attacking the allocation problem, the first problem that must be resolved is the calculation of the probability of intercepting a crime of duration t_0 minutes during a search of t_1 minutes. However to simplify the analysis we will consider only the situation in which, at most, only one crime is present during the search. In addition crimes are assumed to arrive in a Poisson process.

Let $A(t)$= the probability that only one crime will arrive during a time period of t minutes.

1) The simplest case to solve is for $t_1=t_0$,

$$p_i = A(t_1 + t_0) \cdot \int_0^{t_0} \frac{2}{t_0 + t_1} \cdot [1 - \exp(-s \cdot t \cdot d/m]dt$$

$$= A(2t_0) \cdot \int_0^{t_0} \frac{1}{t_0} \cdot [1 - \exp(-s \cdot t \cdot d/m]dt \qquad (4.11)$$

The first term, $A(t_1 + t_0)$, is the probability of one crime occurring that could be detected during the search. It includes the possibility of a crime beginning before the search begins but which is still present (because of its finite duration) when the search gets underway as well as a crime which begins after the search has begun but before the search has ended. Because of the Poisson assumption, if an observable crime occurs, the probability distribution for the point at which the crime <u>begins</u> is uniformly distributed over the $t_1 + t_0$ minutes. This fact explains the presence of the second term $\frac{1}{t_1+t_0}$. The factor, 2, is introduced beacuse there are two equally likely possibilities for a crime to be present for t minutes during the search. Either it arrived after the search began and there are only t minutes left in the search effort or it began before the search commenced (to be exact t_0 - t minutes before) and will end t minutes into the search. The last term is just the probability of intercepting a crime that is detectable for only a span of t minutes.

2) For the case in which $t_1 < t_0$ (i.e. the search is shorter than the duration of the crime),

$$p_i = A(t_0 + t_1) \cdot \left[\int_0^{t_1} \frac{2}{t_0+t_1} \left[1 - \exp(-s \cdot t \cdot d/m)\right] dt + \frac{t_0-t_1}{t_0+t_1} \cdot \left[1 - \exp(-s \cdot t_1 \cdot d/m)\right] \right] \qquad (4.12)$$

The explanation of the first half of this expression (the integral) is analogous to that of the previous equation. The crime may be observable for only <u>part</u> of the search in one of two ways. Either it began during the search or it began at a point more than $t_0 - t_1$ minutes before the search did and therefore will end before the search itself is completed. The additional term is necessary to include all crimes that began within $t_0 - t_1$ minutes prior to the start of the search (with conditional probability of $\frac{t_0-t_1}{t_0+t_1}$) and were therefore detectable throughout the <u>entire</u> search.

3) The last possibility is $t_1 > t_0$ (i.e. the search is longer than the duration of the crime)

$$p_i = A(t_0 + t_1) \left[\int_0^{t_0} \frac{2}{t_0+t_1} \cdot \left[1 - \exp(-s \cdot t \cdot d/m)\right] + \frac{t_0-t_1}{t_0+t_1} \cdot \left[1 - \exp(-s . t_0 \cdot d/m)\right] \right] \qquad (4.13)$$

The first component looks at the search for a crime which is present less than t_0 minutes. The second component accounts for crimes for which the search effort of t_1 minutes totally overlaps the duration of the crime.

None of the above equations represent exponential distributions and consequently existing search theory is not easily applied to the optimal allocation problem. In addition the three equations only consider one crime type of duration t_0 minutes. However, it is relatively easy to expand the equations to allow for crime types of varying durations t_1, t_2, etc. This would be accomplished by introducing a summation of the various crime types weighted by their relative frequency. Unfortunately, it will not be as easy to translate optimal search strategies for single type crimes to several types of crimes.

4.4 How Certain Characteristics of Crimes Impact on the Optimal Allocation Problem

The introduction of randomness in the arrival and/or departure of a target is not a new problem. Blachman [2] and Blachman and Proschan [3] have developed search strategies that minimize the expected delay between the appearance and detection of targets with unknown arrival times; however, once the target arrives, it never departs. Barnett [1] in a more recent article developed some very interesting theorems about optimum search strategies for targets with unknown arrival times and finite departure times. Gilbert [10] carried out a detailed analysis of the impact of random intervals of target visibility on optimal search strategies and compared the results to a blind application (making no correction for random visibility) of Koopman's allocation methodology. Although Gilbert's results provide some important insights into the crime detection problem, the problem we are analyzing is sufficiently distinct from his (a point discussed earlier in the literature review) to require a separate analysis. Hopefully this analysis will complement the earlier work and add yet another perspective to understanding the impact of randomness and finite duration. This presentation will, on the whole, be on a more elementary and fundamental level than the earlier works. In the course of the analysis, we will construct abstract examples in order to analyze individually how the various characteristics of crimes (e.g.

unknown arrival times) impact on optimal patrol strategies. These examples are not, however, meant to be models of how police patrol is carried out. The goal of this analysis will be to develop a qualitative understanding of the impact of various factors within the context of the classical search theory results of Koopman rather than to generate optimal search strategies.

The analysis will address the following issues:

1) The nature of diminishing return from two perspectives:

   a. The _a priori_ diminishing return that accompanies the allocation of additional search effort.

   b. _Sequential_ (or _marginal_) diminishing return occurs when conditional on not finding the target, the probability of locating the target during the next period of search is less than in the previous period.

2) The impact on diminishing return of

   a. unknown arrival times

   b. finite duration

   c. targets arriving independently of one another

3) In classical search theory when effort is allocated between two regions A and B, the order in which the areas are searched is irrelevant.

   a. Analyze the impact on this of random arrivals, finite duration and independent targets.

4.4.1 Diminishing Return

a) Assume a target is located somewhere in a region A and the observer has a probability p of locating the target with a single glance. If an observer makes two random glances in the region he has a probability, p_a

$$p_a = 1 - (1 - p)^2 = 2p - p^2 \tag{4.14}$$

of detecting the target. The increase, Δp_a, in the total probability of detection as a result of the second glance is not the same as for the first glance (which was p) but rather

$$\Delta p_a = p - p^2$$

This example portrays the most fundamental form of diminishing return which relates to the a priori allocation of search effort. It should be noted, however, that from a sequential point of view, the probability of finding the target on the second glance given that it was not found on the first glance is still p.

b) Suppose a target is located with certainty somewhere in either region A or region B and each region is broken into four equal areas (Figure 4.2). In addition, the target is twice as likely to be in any one of the A boxes as in any one of the B boxes. Assume also that a single scan can survey an entire box and determine with certainty if there is a target in that box. Consequently the probability of finding the target with a single glance in region A is

$$p_a = 2/12 = .167$$

and in region B is

$$p_b = 1/12 = .083$$

If two independent glances were cast in region A, the *a priori* probability of locating the target is

$$^3p_{aa} = (3/4 \cdot 4/12) + (1/4 \cdot 2/12) = .292$$

and the increase in the probability of interception is

$$\Delta p_{aa} = p_{aa} - p_a = 7/24 - 4/24 = .125$$

Again there is a diminishing return but now with one distinction, that even when the search is viewed sequentially there is diminishing return. Given that the target was not found on the first glance, the probability of finding it on the second glance is only

$$^4p_{a\underline{a}} = 3/4 \cdot 2/10 = .15$$

as compared to .167 for the first glance. This second form of diminishing return results from the target not being constrained to be in region A. Also if we look at the effect on B of the first unfruitful glance in A, we find that the probability of finding the target in B on the second glance has increased to

$p_{ab} = 1/10 = .1$

as compared to .083 previously. In effect the conditional probabilities of finding the target in A or B in the succeeding glances have begun to change in opposite directions. From this simple example it begins to become apparent why as the available search effort increases there is a greater likelihood that the optimal search policy will include searching region B. The two opposing effects that generate the above property are very closely related but, as will be seen later, they can be separated.

As the search of region A continues beyond the second glance to a third and fourth glance, diminishing return continues. The result is that by the fourth glance the optimal strategy would allocate that look to region B rather than region A.

4.4.2 Unknown Arrival Times

To introduce randomness and yet analyze an almost equivalent problem, the following case is constructed.

A target is always present in either region A or region B with the relative likelihoods as displayed in Figure 4.2. The one modification is that the target can remain in its place for only the time span of two glances at which point another target replaces it. The second target selects its position independently of the first target's position but with the same relative likelihoods as before. At the time the search begins the target is equally likely to have just arrived as to have been present for the time period of one glance. A first glance in A has a payoff of

$$p_a = 2/12 = .167$$

the same as before. However, during the second glance we may no longer be looking for the same target since the first one may have disappeared. The probability of finding the target on the second glance is now

$$p_{a\underline{a}} = 1/2\,[3/4 \cdot 2/10 + 2/12] = .1583$$

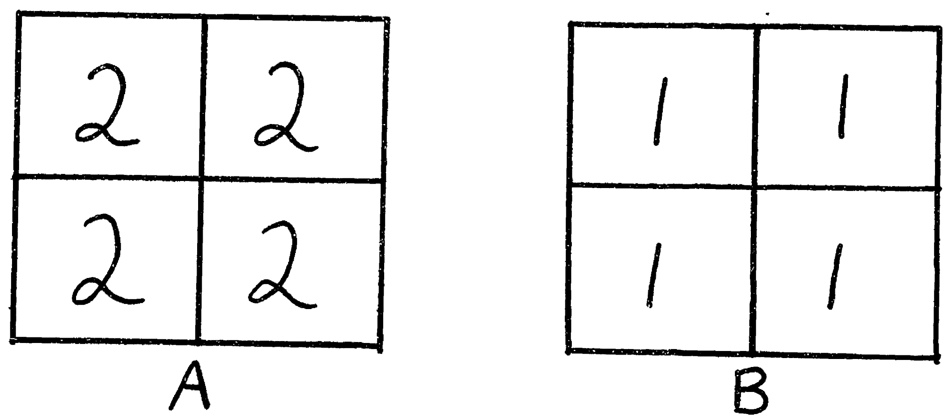
which is greater than the .15 obtained by the nonrandom case. If the process is viewed _a priori_, then the probability of finding the target with two glances in A is

$$p_{aa} = 1/6 + 5/6 \cdot (.1583) = .299$$

and

$$\Delta p_{aa} = .299 - .167 = .132$$

which is greater than the .125 for the nonrandom case. From either vantage point, it is apparent that the introduction of randomness in the target's arrival and departure slows the process of diminishing return because probabilistically we may not be duplicating our earlier effort. The implication of this is that under an optimal search strategy more effort would have to be available in the random case than in the non-



PROBABILITY OF INTERCEPTING A TARGET

	Nonrandom Arrivals	Random Arrivals
1st Look in A	.167	.167
2nd Look in A	.150	.1583
2 Looks in A	.291	.299
3rd Look in A	.1323	.1584
1st Look in B	.088	.088
2nd Look in B after 1st Look was in A	.100	.092

Figure 4.2: The Impact of Random Arrival and Departure of Targets on the Probability of Interception and on Diminishing Return

random case before the equilibrium criterion suggests searching B.

The above example can be easily extended to include continuous search. It is also apparent that as the lifecycle of the target increases beyond two glances the rate of diminishing return increases and in the limit approaches the non-random arrival case.

4.4.3 Finite Duration

The next component of the random case to be evaluated is the impact of the finite duration of a target. To do so we will allow the search to continue beyond the second glance. The third glance will again have diminishing return, increasing the *a priori* probability of interception less than the second. However, if we look at it from a sequential perspective, the probability of finding the target on the third glance in A conditional on not finding it during the first two is .1584. Thus diminishing return continued only as long as the life (in this example two glances) of an individual target. As a result, if it is not advisable to search B during the first two glances, it is never advisable to do so. The same will be true for the generalized continuous problem. For targets (e.g. crimes) that last for a finite time period, t, the conditional probability of finding the target in an incremental time period will approach and attain a steady state value when the search is carried on for longer than t minutes.

(The reason why in the above example there was an increase from .1583 to .1584 is explained by the following. As a result of not finding the target in the first two looks in A, there is now a slightly greater than .50 probability (in fact .505) that the target has just departed and that a new target is arriving at the beginning of the third glance. However, in the police context, with assumed Poisson arrival of crimes, this phenomenon will not occur and a steady state value will be reached.)

A somewhat intuitive explanation of this bound on sequential diminishing return is as follows. Since a target survives only t minutes, the fact that the target was not found t + 1 minutes ago in a particular region can yield no information about whether the target is now present in that region. Thus as a search is carried on longer than t minutes, the searcher is continually gaining new information about the likelihood of a target being present now but at the same time the information that was obtained more than t minutes ago is becoming outdated at the same rate. The system, in a sense, is in a state of dynamic equilibrium as long as search is continued.

4.4.4 Independently Arriving Targets

There is one final complication that has to be analyzed before closing the door on diminishing return. For this discussion there will no longer be a constraint on the number of targets present and in addition the two regions will generate targets independent of each other although in the same proportion as before (A generates twice as many as B). In this case not finding a target in region A no longer yields any information about the possible presence of a target in B. So that when diminishing return occurs in region A there is no accompanying increase in the probability of finding the target in B.

The above description of the impact of randomness on diminishing return can be summarized as follows:

1. Randomness slows diminishing return.
2. Sequential diminishing return lasts for a period equal to the duration of a crime.[5]
3. Searching region A does not increase the potential return from search in region B because of the independence of the two regions.

These three results compounded by the short duration of crimes unite to increase the likelihood that the optimal allocation of patrol (search) effort over several nonuniform crime regions would be to concentrate on only the highest crime region within the patrol sector.

4.4.5 Sequencing Search Effort

It was mentioned earlier that even with the straightforward application of search theory to the allocation problem there are serious constraints on implementing the optimal solution. For example, an optimal solution (for crimes of 5 minute duration) of searching region A for 3 minutes and B for 2 minutes can not be interpreted as equivalent to allocating 60% of your effort to region A and 40% to B without simultaneously constraining the search to be cyclical with frequent shifting between regions. However, we will soon show that randomness may require that the optimal strategy include even more frequent shifts between A and B than are implied by the finite observable duration of the target.

In allocation problems of the type that Koopman analyzed, the optimal solutions are one dimensional in that they specify only the amount of effort to allocate to each region. The manner in which this effort is carried out is irrelevant. The probability of intercepting a target is the same whether you search A for 3 minutes and B for 2 minutes or you search A for 1.5 minutes, B for 1 minute, A for 1.5 minutes and, again B for 1 minute. In essence, for nonrandomly arriving targets the only important aspect of the search is how much territory was covered in each region. This fact is more a consequence of the lack of time dimension or dependence in the search rather than a result of the exponential form of the equation (4.1) [6].

This concept of time independence is illustrated in Figure 4.3 which represents the sequential rate of return of a search in one of two regions. As the search is carried out sequential diminishing return occurs. However, if the search effort is interrupted, put in abeyance, and resumed several minutes later, the rates immediately prior to the interruption and immediately after the resumption of search are the same, and during the interruption the potential rate of return is constant. In contrast to this, randomness introduces a time dependence into the rate of return, Figure 4.4. When the search is interrupted and later resumed, the rate of return after the interruption is higher than before the interruption because of new arrivals and throughout the interruption the potential rate of return is increasing. Under these circumstances, it is fairly obvious that there is now a difference between searching A for 3 minutes, B for 2 minutes and searching A for 1.5 minutes, B for 1 minute, A for 1.5 minutes and B for 1 minute, with the latter strategy having a higher payoff (but also being more difficult to implement by a patrol unit). As a result of this phenomenon of time dependence, the optimal search strategy must include both the quantity of search allocated to each region and the sequencing of search. In determining this optimal strategy, it is also necessary to realize that the two components are not independent.

It would now be useful to present a basic but interesting case that displays the impact of time dependence on even the <u>quantity</u> of search allocated to each region. The easiest type of optimal solution to implement is, of course, one that involves a search in only one region, A. A necessary condition for this type of solution is that the marginal rate of return

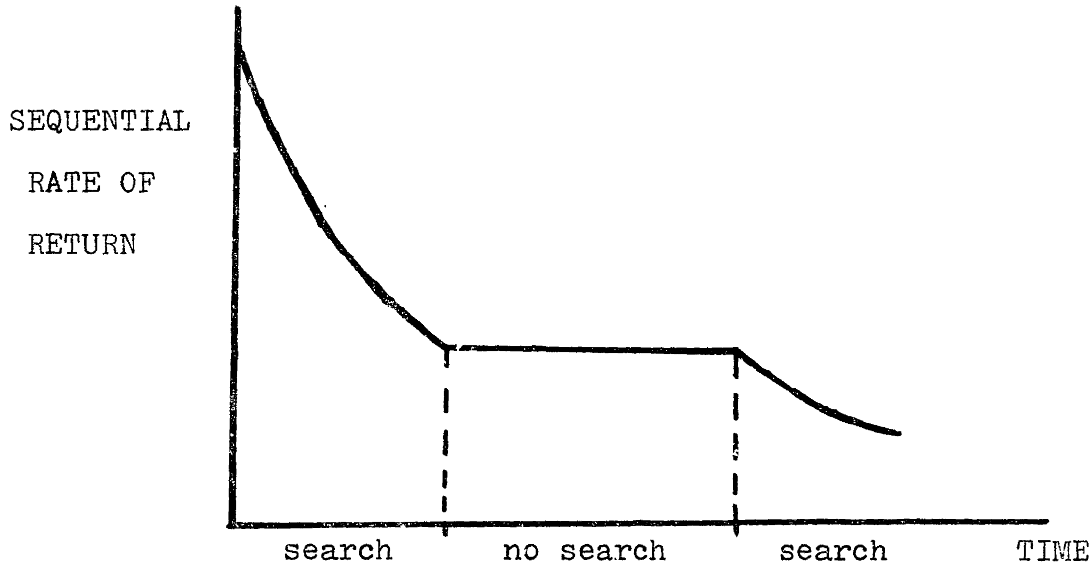


Figure 4.3:The Sequential Rate of Return as a Function of Time for Nonrandom Targets:Time Independent Search

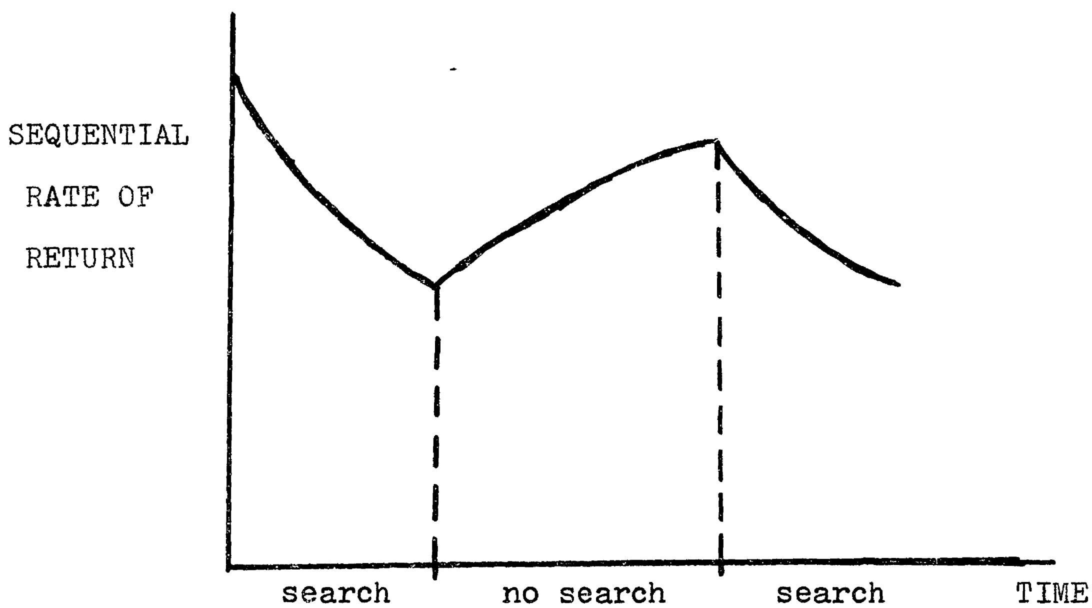


Figure 4.4:The Sequential Rate of Return as a Function of Time for Random Targets:Time Dependent Search

at the end of t (the total available search effort) minutes in A be greater than the initial return from search in B; however, because of the time dependence factor, this is no longer a sufficient condition.[6] The above condition means that 't' minutes in A is preferred to 't - dt' minutes in A followed by 'dt' minutes in B; it does not imply that 't' minutes in A is preferred to 't/2 - dt' minutes in A, followed by 'dt' minutes in B followed by 't/2' in A. In the latter case we are balancing search effort in A against search effort in B and an increased rate of return from the second half of the search effort in A. An obvious question to be resolved is what are necessary and sufficient conditions for allocating all your effort to only one region?

A second interesting consequence of time dependence of the search is as follows. Assume a target is present for 30 minutes. Then for nonrandom arrivals there is no distinction between one unit searching for 30 minutes and two units searching simultaneously for 15 minutes. However, with random arrivals there will be a distinction between sequential and simultaneous search with sequential search preferred. Similarly increasing the number of hours of patrol in a sector by reducing the length or number of busy periods of the sector car has a higher payoff than an equivalent increase in the number of patrol hours produced by adding manpower. In the latter case, part of the additional patrol will be carried out simultaneously with the already allocated patrol effort, while in the former case the additional patrol will always be performed at points in time when there would otherwise not have been any patrol in that sector.

4.4.6 Summary and A Different Approach to Patrol Allocation

In the preceding sections a number of general properties of a search for randomly arriving and departing targets were discussed. Three factors were described: 1)random arrival, 2)finite short duration, and 3)independent targets, which all tend to increase the likelihood (for nonhomogeneous regions) that the optimal strategy will be to search only the high crime area. A fourth factor, which has a similar impact and which will be duscussed in detail in Chapter 5, is the time lost in travel between regions that are allocated patrol effort. In addition we have described the difficulties (because of time dependence) of implementing solutions that would result from applying classical search theory to the problem. In order to exploit the realization that optimal strategies will require the concentration of patrol to very limited portions of a sector as well as to avoid generating unimplementable theoretical optimums, we suggest approaching the optimal allocation problem form a different perspective.

In classical search theory the allocation problem is posed as follows:

> There exist several predetermined regions with varying target probabilities. How should the available search effort be allocated (and sequenced) among these regions?

An alternative to the above formulation which reverses the process is:

> Assume that one and only one contiguous area will be searched. What is the optimal size and location of this area?

Chapter 7 presents an algorithm for constructing from a collection of atoms (each several square blocks in size), the single region in which to concentrate the sector car's patrol. However, any plan that would implement a strategy of this type must weigh carefully the implications of leaving a portion of a sector without regular police patrol. It may be necessary to moderate the 'theoretical' optimum by making random widely (timewise) dispersed, highly visible, high speed passes through the low crime areas with the purpose being to create an impression of presence. It is possible, though that this impression of presence [17] may be created without any special effort since the sector car will still be responding to calls for service arising <u>throughout</u> the sector. An equally important problem that is not necessarily easy to solve is the reaction of local community groups to their sections of the sector receiving no patrol. Lastly, a patrol plan of this nature would have to be continually reevaluated as the spatial distribution of crimes may vary in response to police presence. This type of reevaluation would have to be on a frequent basis, probably at least every week, in order to keep on top of changing patterns of crime. The obvious critical question in this regard is "How quickly and in what way do criminals respond to changes in patrol strategies?" Unfortunately, it is a question about which we know little especially in a quantitative sense. In Chapter 8 we will describe an experiment which focuses on obtaining at least a partial answer to this question.

The approach suggested above has one additional benefit. Thus far all discussion of patrol allocation has been limited to a search for crimes of a single class (same observable duration). However, the standard patrol does not usually focus on one single class of crimes. This complication does not present any significant problems in calculating the probability of interception or space-time coincidence. All that would be required to account for the multiplicity of crime types is to take the basic equation (4.1) and sum or integrate it over the various crime types weighted by their relative frequencies. The introduction of multiple crime types would, however, seem to seriously complicate the classical allocation approach as even the amount of search effort available to allocate will depend on the crime type. In the reversed allocation methodology the impact of multiple crime types would be basically the same as on calculating the probability of interception since equation (4.1) forms the basis of the procedure.

Before proceeding with the development of algorithms that build on the work of this chapter, we will attempt to quantify a number of concepts we have discussed here qualitatively. In the next chapter a differential equation model will be used to determine under what conditions should a region be excluded from search. In addition we will quantify the impact of travel time on optimal allocations of search.

FOOTNOTES 4

1 A decision maker, however may be more willing to accept a 50% increase if the change in travel time is from 1 minute to 1.5 minutes than he would be if the change were from 4 to 6 minutes.

2 Throughout the paper the term random when applied to patrol will describe the unpredictable non-patterned search effort. When applied to the target, it is describing its time of arrival and departure.

3 The first subscript will refer to where the first glance was, the second subscript to the second glance and so on.

4 A line under the last subscript will mean that we are focusing sequentially on that glance. For example, $p_{a\underline{b}}$, means the probability of finding the target on the second glance in b given the first glance was unfruitful in a.

5 The duration of a crime is assumed fixed. If a probability distribution for the duration of the crime is inserted (with a finite mean), sequential diminishing return will approach a non-zero bound. See also Barnett [1].

6 See Barnett [1] for a discussion of a similar necessary but not sufficient condition.

REFERENCES 4

1. Barnett, A., On Searching for Events of Limited Duration, (Working paper WP-11-74), Massachusetts Institute of Technology Operations Research Center, Cambridge, MA, September 1974, submitted for publication.

2. Blachman, N., "Prolegomena to Optimum Discrete Search Procedures", Naval Research Logistics Quarterly, Vol. 6, No. 4, 1959.

3. Blachman, N. and F. Proschan, "Optimum Search for Objects Having Unknown Arrival Times", Operations Research, Vol. 7, No. 5, 1959.

4. Blumstein, A. and R.C. Larson, "A Systems Approach to the Study of Crime and Criminal Justice", in Operations Research for Public Systems, P.M. Morse and L.S. Bacon (eds.), MIT Press, Cambridge, MA, 1967.

5. Charnes, A. and W.W. Cooper, "The Theory of Search: Optimum Distribution of Search Effort", Management Science, Vol. 5, October 1958.

6. Dobbie, J.M. "Search Theory: A Sequential Approach", Naval Research Logistics Quarterly, Vol. 10, 1963.

7. Elliott, J.F., "Random Patrol", Law Enforcement Science and Technology II, Proceedings of the Second National Symposium on Law Enforcement Science and Technology, Illinois Institute of Technology Research Institute, Chicago IL, 1970.

8. Elliott, J.F. and T.J. Sardino, Crime Control Team: An Experiment in Municipal Police Department Management and Operations, Charles C. Thomas Publisher, Springfield, IL, 1971.

9. Elliott, J.F., Interception Patrol, Charles C. Thomas, Publisher, Springfield, IL, 1971.

10. Gilbert,E.N., " Optimal Search Strategies", SIAM Journal, Vol. 7, 1959.

11. Jarvis, J.P., Optimization in Stochastic Service Systems with Distinguishable Servers, Massachusetts Institute of Technology Operations Research Center, Technical Report, TR-19-75, June 1975.

12. Koopman, B.O., "The Theory of Search I and II", Operations Research, Vol. 4, 324-346, 503-531, 1956.

13. Koopman, B.O., "The Theory of Search: III. The Optimum

Distribution of Searching Effort", Operations Research, Vol. 5, 1957.

14. Larson, R.C., Urban Police Patrol Analysis, MIT Press, Cambridge, MA, 1972.

15. Larson, R.C., "A Hypercube Queuing Model for Facility Location and Redistricting in Urban Emergency Services", Computers and Operations Research, Vol. 1, No. 1, March 1974.

16. Larson, R.C., "Illustrative Police Sector Redesign in District 4 in Boston", Journal of Urban Analysis, Vol. 2, 1974.

17. Larson, R.C., "What Do We Know About Preventive Patrol? A Review of the Kansas City Preventive Patrol Experiment", published by Public System Evaluation, Inc., Cambridge, MA, July 1975, to appear in Journal of Criminal Justice.

18. Morse, P.M., Search Theory, (Working paper WP-01-74), Massachusetts Institute of Technology Operations Research Center, Cambridge, MA, January 1974.

19. Olson, D.G., A Preventive Patrol Model, paper presented at Operations Research Society of America meeting, Miami, FL, 1969.

20. Olson, D.G. and G.P. Wright, Models for Allocating Police Preventive Patrol Effort, Krannert Graduate School of Industrial Administration, Purdue University, Paper No. 404, West Lafayette, IN, April 1973.

21. President's Commission on Law Enforcement and Administration of Justice, The Challenge of Crime in a Free Society, U.S. Government Printing Office, Washington, D.C., 1967.

CHAPTER 5

A DIFFERENTIAL EQUATION MODEL OF SEARCH AND DETECTION

5.0 Introduction

Thus far the description and analysis of the dynamics of a search for randomly arriving and departing targets has emphasized the development of a qualitative understanding of this process. A recurring theme of this analysis has been that in allocating search effort among competing regions the optimal solution is likely to concentrate all the search in one region. In this chapter we will introduce a model that will be used to quantify the conditions under which the optimal strategy is to search only a single regions. The model (which was suggested by Philip M. Morse) is a set of differential equations which characterize the continuously changing state of the system in terms of the number of crimes in progress. The model will then be applied in three cases to determine the optimal sequence of search between two regions for crimes of a single type. The three cases are

1. Two regions with equal crime rates and no time lost in switching between regions;

2. Two regions with equal crime rates and time lost in switching between regions;

3. Two regions with differing crime rates and no time lost in switching between regions.

This chapter will close with a discussion of extensions of the above examples to allow for multiple crime types and more than two regions.

5.1 The Differential Equation Model

The model consists of two differential equations for each region, one to describe the system (city, precinct, sector, or reporting area) when no search is in progress in that region, the other to describe the system when a search is in progress.

Let S= the state of the system (expected value of the number of crimes in progress).

A= the rate at which crimes arrive (assumed independent of the state of the system).

Crimes depart from the system in either one of two ways. Either crimes leave because they have finished or because they have been interrupted during a period of search.

F= the constant of proportionality for the rate at which crimes finish. (This will depend upon the duration of a crime.)

I= the constant of proportionality for the crime interception rate. (This will depend on the street mileage in the region and the observability of the crime.)

Both of these rates are proportional to the number of crimes in progress.

Using the above parameters (S, A, F, I), a set of differential equations to describe the changing system can be written as follows. During a period of no search the system is described by

$$\frac{ds}{dt} = S' = A - F \cdot S$$

The equation states simply that the system is changing because targets are arriving at a rate A and are departing, as a result of finishing at a rate, $F \cdot S$, proportional to the state of the system. The solution of the differential equation is

$$S = C_1 \exp(-F \cdot t) + A/F \qquad (5.2)$$

C_1 is a constant that depends on the boundary conditions and A/F represents the steady state number of crimes in progress if no search were ever carried out. During a period of search the equation becomes

$$\frac{ds}{dt} = S' = A - (F + I) \cdot S \qquad (5.3)$$

whose solution is

$$S = C_2 \exp(-(F + I) \cdot t) + A/(F + I) \qquad (5.4)$$

C_2 is also a constant that depends on boundary conditions with $A/(F + I)$ the steady state number of crimes in progress during an unending period of search.

5.1.1 Model Assumptions

These equations assume that the arrival process of crimes is Poisson (i.e.random) and that the duration of a crime is exponentially distributed. S is therefore the expected value of the number of crimes. This can be shown by setting up queuing type equations for the probabilities P_n, where n is the number of crimes in progress at any one time. The general equation would be of the form

$$\frac{dP_n}{dt} = A \cdot P_{n-1} + F \cdot (n+1) \cdot P_{n+1} - (A+n \cdot F) \cdot P_n \tag{5.5}$$

with the equation for P_0 just

$$\frac{dP_0}{dt} = F \cdot P_1 - A \cdot P_0 \tag{5.6}$$

By setting $S = \sum nP_n$, we arrive at equation (5.1), which describes the system during a period of no search.

5.2 Two Regions--Equal Crime Rates - No Time Lost in Transfer

The first problem considered with the differential equation model is finding the optimal strategy for sequencing the search effort of a single patrol unit between two regions, R1 and R2, of equal size. The objective is to minimize the steady state average expected number of crimes in progress. In our analysis two assumptions will be made about the optimal solutions.

1. The optimal allocation of search effort divides the available effort equally between the two regions.

2. The optimal strategy will be cyclic of the form X minutes spent in R1, followed by X minutes in R2, and then back to R1 for X minutes and so on.

The first assumption is intuitively appealing because of symmetry arguments and will be proven later in this chapter. The second assumption is motivated by a number of earlier search theory papers [2, 3, 5] in which the optimal search strategies were cyclical. In the situation most analogous to ours, Barnett [1] proved that for n regions in which targets arrive and depart randomly, the strategy which maximizes the probability of intercepting a random crime must be cyclic. Given the above assumptions, the problem that remains to be resolved is "What is the optimal cycle length or equivalently the optimal value of X?".

Although the objective is to minimize the total (in both regions) average expected number of crimes in progress, it is more convenient to analyze first the search process from the separate perspectives of the two regions. Each region will be viewed independently as experiencing alternating periods of search and no search. This split can be made because the regions are generating arrivals and departures independent of each other; consequently searching R2 can yield no information about the likelihood of a crime being in progress in R1. This split also makes it clear that the solution to the first two region problem will in addition answer the following question.

> Given that only 50% of a patrol car's time is available for patrol, what is the distinction between short numerous intervals of search and a few long intervals of search?

Before proceeding with the analysis, it might clarify the discussion to first outline three steps common to each of the examples presented in the chapter:

1. Two differential equations are defined for each region. One to describe the dynamics of the region while it is being searched, the other to describe the region while no search is going on.

2. For each region two boundary conditions of the following form are established:

> Under steady state conditions, the level of crime in each region at the end of a period of search must equal the level of crime at the beginning of a period of no search and vice versa.

These conditions are simply a continuity constraint on S, the number of crimes in progress, not allowing discrete shifts in the values of S as a result of the patrol unit entering or departing a region. Once the boundary conditions are defined, each of the constants in the differential equations is determined.

3. The average level of crime, $\bar{S}$, in the total area is calculated by integrating each of the four equations over its respective period of search or no search, summing the four values and dividing by the cycle length. The resultant expression for $\bar{S}$ is a function of the parameters of interest in each of the particular examples and is subsequently analyzed to determine an optimal strategy.

Turning back to the original problem, we will focus the analysis on only one of the regions, R1, with all discussions applying similarly for R2 because of the symmetry of the search. In R1 the two equations describing the periods of search and no search are just the previously defined equations (5.1) and (5.3) and whose solution equations, (5.2) and (5.4), are rewritten here:

$$S = C_1 \exp(-F \cdot t) + A/F$$
$$S = C_2 \exp(-(F + I) \cdot T) + A/(F + I)$$

The continuity constraint on S is used to generate the following equations:

$$C_2 \exp(-(F + I) \cdot X) + A/(F + I) = C_1 + A/F \qquad (5.7)$$

The left hand side represents the end of a search period of duration X minutes and the right hand side the beginning of a no search period. Conversely,

$$C_1 \exp(-F \cdot X) + A/F = C_2 + A/(F + I) \qquad (5.8)$$

Solving for C_1 and C_2 as functions of X yields:

$$C_1 = [-A \cdot I/(F \cdot (F + I))] \cdot [1 - \exp(-(F + I) \cdot X)] / [1 - \exp(-(2F + I) \cdot X)] \qquad (5.9)$$

$$C_2 = [+A \cdot I/(F \cdot (F + I))] \cdot [1 - \exp(-F \cdot X)] / [1 - \exp(-(2F + I) \cdot X)] \qquad (5.10)$$

The expected level of crime in each of the regions at any instant in time can now be written as a function of only one parameter, X, the half cycle length. To compare various cycle lengths, equations (5.2) and (5.4) are combined into a single equation representing the average expected level of crime in R1 or R2 during a cycle.

$$\text{AVE. } S1 = \overline{S1} = \frac{1}{2X}\int_0^X C_1 \exp(-F \cdot t) + A/F\, dt + \frac{1}{2X}\int_0^X C_2 \exp(-(F + I) \cdot t) + A/(F + I)\, dt \qquad (5.11)$$

with C_1 and C_2 as defined in equations (5.9) and (5.10) respectively. Integrating equation (5.11) yields

$$\overline{S1} = -(C_1/(2F \cdot X)) \cdot (1 - \exp(-F \cdot X)) + A/2F - (C_2/(2(F + I) \cdot X)) \cdot (1 - \exp(-(F + I) \cdot X) + A/2(F + I) \qquad (5.12)$$

The average for the entire region, $\overline{S}$, is merely twice that of any individual region.

$$\overline{S} = \overline{S1} + \overline{S2} = 2 \cdot \overline{S1} \qquad (5.13)$$

In Appendix D it is then shown that $\overline{S}$ (or equivalently $\overline{S1}$) monotonically increases as X increases. In other words, the average expected level of crime in each region decreases as the frequency of transfers between the two regions increases. To use the terminology of Chapter 4, by spacing out the available search effort as much as possible the amount of diminishing return is reduced to its absolute minimum. This result is directly comparable to that of Gilbert [3], who showed (using a probabilistic search model) that it is optimal to switch from one region to another whenever the region being searched has received a longer time of search than the other region. He also defined a limit strategy which approaches the theoretical optimum strategy as switching becomes instantaneous.

Earlier we had noted an equivalence between sequencing search between regions and scheduling the patrol time of a single unit that spends 50% of its time answering calls for service. Thus from one perspective if somehow preventive patrol activities were scheduled in long blocks of time (i.e. increase X), the model predicts a decrease in the probability of interception.[1] However, this decrease will be very small since even the two extremes, infinitely long cycles and infinitesimally small cycles, typically differ by less than 2%. (See Table 5.1 for a ratio greater than 30) Therefore other aspects of the problem, not presently captured by the model, are likely to be more crucial. From a psychological perspective, scheduling patrol could modify the attitude of patrolling officers towards this activity. At present, patrol is often viewed as time left over from their main activity of responding to calls for service. Scheduling patrol in long blocks of time might then increase the value of patrol in the eyes of the officer. This change would manifest itself quantitatively by increasing I, the rate of detection. If it were possible to actually specify I as a function of the cycle length then also this issued could be easily included into the model. The optimal cycle length then would not be infinitesimally small.

5.2.1 Magnitude of Impact of Short Cycles

Having shown that shorter cycles increase the probability of interception, the obvious next step is to determine the magnitude of the improvement produced by the shorter cycles. Because $\overline{S}$, $\overline{S1}$ and $\overline{S2}$ are monotonically increasing functions of X, their highest values occur as the cycle length approaches infinity. For example the limit of $\overline{S1}$ as the cycle length increases is

$$\lim_{2X \to \infty} \overline{S1} = (1/2)\cdot(A/F) + (1/2)\cdot(A/F + I)) = \frac{A \cdot (2F + I)}{2F \cdot (F + I)} \qquad (5.14)$$

The first term, A/F, is the steady state level of crime for the region not being searched and A/(F + I) is the steady state level of crime for a region being searched.

For short cycles the average level of crime in R1 is calculated by applying L'hôpital's Rule twice to equation (5.12) (see Appendix E) to yield

$$\lim_{2X \to 0} \overline{S1} = 2A/(2F + I) \qquad (5.15)$$

The ratio then of the average crime level, $\overline{S1}$, (similarly for $\overline{S}$) for short cycles as compared to long cycles is

$$\text{Ratio} = \text{short/long} = \lfloor 2A/(2F + I) \rfloor / \lceil A(2F + I)/2F \cdot (F + I) \rceil \qquad (5.16)$$

which reduces to

$$1 - (I/(2F + I))^2 \qquad (5.17)$$

which not surprisingly is independent of the arrival rate, A, of crimes. Thus if crimes were discovered at a rate equal to the rate at which they are completed then

$$F = I$$

and

$$\text{Ratio} = 1 - I^2/9I^2 = 1 - 1/9$$

Under these circumstances optimally sequencing the same total amount of search effort in each region reduces the average level of crime by 1/9. If, however, the two policies are compared with regard to their relative impacts on the crime level, then the reduction resulting from shorter cycles seems more significant. For example, consider the reduction in the average crime level produced by long cycles. If no search were carried out the average number of crimes in progress would be A/F. The ratio of the average number of crimes in progress in long cycles over the average number when no search is carried out is

$$\text{Ratio} = \frac{\text{Long cycles}}{\text{No search}} = \left[\frac{A \cdot (2F + I)}{2F \cdot (F + I)}\right] \Big/ [A/F] \qquad (5.18)$$

$$= 1 - I/(2 \cdot (F + I)) \qquad (5.19)$$

which for I=F is 1 - 1/4. For short cycles the ratio is

$$\text{Ratio} = \frac{\text{Short cycles}}{\text{No search}} = \left[\frac{2A}{2F + I}\right] \Big/ [A/F] \qquad (5.20)$$

$$= 1 - I/(2F + I) \qquad (5.21)$$

which for I=F is 1 - 1/3. As a result long cycles reduce the level of crime by 25% while short cycles by 33 1/3%. In relative terms this means that short cycles reduced the crime level one-third more ((1/3)/(1/4) = 1 1/3) than long cycles.

The measure of effectiveness that was just introduced, namely the ratio of a policy's crime level to the crime level when no search is carried out has a deeper significance than may be apparent at first. As we shall show, there is a simple relationship between this measure and the probability of intercepting a crime.

Under the two limiting policies (i.e. infinite and infinitesimal cycles), it is possible to calculate directly the probability of intercepting a random crime. (We are assuming, still, that crimes arrive in a Poisson process (rate A), have an exponential lifetime (mean 1/F) and are discovered at an exponential rate (mean 1/I).) For long cycles the fraction of crimes that are intercepted during a period of search is simply I/ (F + I). Since half of all the crimes occur during the period of search, the fraction of all crimes that will be intercepted is

$$\frac{I}{2(F + I)} \tag{5.22}$$

Calculating the probability of interception for infinitesimally short cycles is also relatively straightforward and is based on the following argument. No matter how long a crime lasts, during half the lifetime of the crime, the patrol unit will be searching the same region as the crime is in because the cycles are infinitesimally small. Consequently the probability of intercpeting a random crime is

$$\int_0^\infty F \exp(-F \cdot t) \cdot (1 - \exp(-I \cdot t/2)dt \tag{5.23}$$

where $F \exp(-F \cdot t)dt$ is the probability that a crime has a lifetime of 'near' t minutes and $(1 - \exp(-I \cdot t/2)$ is the probability of intercepting a crime of that duration. Equation (5.23) reduces to

$$= 1 - (F/(F + I/2)) \cdot \int_0^\infty (F + I/2) \cdot (1 - \exp(-(F + I/2) \cdot t))dt \tag{5.24}$$

$$= 1 - \frac{2F}{2F+I} = \frac{I}{2F+I} \tag{5.25}$$

Not surprisingly the probability of intercepting a crime under each of the two policies is exactly the same as the reduction in the average crime level (Equations (5.21) and (5.22)) as calculated with the differential equation model.

Table 5.1 contains a comparison of the two limiting policies for a range of F and I. When crimes are discovered at a rate equal to their competion rate, short cycles as compared to long cycles reduce the average number of crimes in progress by 11% while they increase the probability of intercepting a crime by 33%. As the rate at which crimes are completed increases relative to the discovery rate to a factor of 5, the improvement produced by short cycles is less than 1% for the average number of crimes, but 9% for the probability of interception. Increasing the relative discovery rate still further to a factor of 50, which brings it closer to the level

RATIO=F/I	AVERAGE EXPECTED NUMBER of CRIMES	PROBABILITY of INTERCEPTION		
	Short/Long	Short	Long	(Short/Long)-1
1	.111	.333	.250	.333
2	.040	.200	.167	.20
3	.020	.143	.125	.143
4	.0123	.111	.100	.111
5	.0083	.091	.083	.091
10	.0023	.0476	.0455	.0476
20	.00059	.0244	.0238	.0244
30	.00027	.0164	.0161	.0164
40	.00015	.0123	.0122	.0123
50	.000098	.0099	.0098	.0099

Table 5.1: A Comparison of Short and Long Cycles with Regard to Interception Probabilities and the Average Expected Number of Crimes in Progress

at which police typically function, reduces the impact of short cycles to a .01% improvement in crime levels and a 1% improvement in the probability of interception.

It should be obvious already from our presentation that only the ratio of F to I is needed in comparing the relative impact of the two extreme policies. This can be seen clearly by rewriting equation (5.17) as

$$\text{Ratio} = \text{short/long} = 1 - \frac{1}{(2 \cdot (F/I) + 1)^2} \qquad (5.26)$$

Perhaps more significantly, the probability of intercepting a crime under either of the two policies is also dependent upon only the ratio F/I. Reformulating equations (5.22) and(5.25) for long and short cycles respectively yields

$$\text{Prob. of Interception (Long Cycles)} = \frac{1}{2(F/I + 1)} \qquad (5.27)$$

$$\text{Prob. of Interception (Short Cycles)} = \frac{1}{2F/I + 1} \qquad (5.28)$$

A similar reduction in the number of significant input parameters also arises later in the chapter when we consider a broader range of search strategies.

5.3 Two Regions - Equal Crime Rates - Time Lost in Transfer

In Chapter Four's analysis of the search process in classical search theoretic terms, it was noted that because of the short duration of crimes and their randomness that an optimal solution is likely to be to search only the highest crime region. In that analysis, though, one component of the actual search process was not included, namely, the time to travel between two regions. This time is either lost totally from the search or is, at best, a period in which the detection rate is significantly reduced. This added motivation for limiting search to one region will be analyzed through a modification of the previous example. A parameter L will be introduced that will represent the time lost every time a transfer between the two regions occurs. Then, using the differential equation model, two questions will be addressed.

1. What is the optimal value of X, the time spent in R1 before switching to R2? The optimal value is obviously no longer infinitesimal.

2. Is there a simple analytic expression which specifies for which values of L it does not even pay to switch regions?

Although papers by Gilbert [3] and Kisi [4] have addressed problems in which time losses for switching were included, those problems, however, did not allow for arriving and departing targets. Since the optimal strategy for static targets requires that eventually each region be searched, no method was, therefore, presented which can be used in problems similar to ours to determine when a region ought to be excluded from the search. This will be a principal question addressed here. In tackling the problem with the differential equation model, we will again make the assumptions that the optimal solution is cyclic (both Kisi and Gilbert have cyclic strategies) and that the search effort will be divided equally between the two regions. Optimal solutions which limit the search to only one region will appear in the analysis as solutions in which the optimal value of X is infinite.

5.3.1 Applying the Model

In the application of this model to this second problem, which ascribes a penalty for switching regions, the basic equations which describe the periods of search and no search do not vary from those of the previous example. The solutions to these equations for one of the regions are rewritten here for convenience.

$$S = C_1 \exp(-F \cdot t) + A/F \qquad (5.29)$$

describes a period of no search and

$$S = C_2 \exp(-(F + I) \cdot t) + A/(F + I) \qquad (5.30)$$

describes a period of search. However, the introduction of the switching time, L, does affect the continuity boundary conditions since a no search period in each region has a duration of 'X + 2L' minutes. (The L is multiplied by 2 because two switches must occur, one leaving region R1 and

and a second upon returning to R1, during any single cycle.) The result is that although equation (5.7), which states that the level of crime at the end of a period of search is equal to that at the beginning of a period of no search, remains the same,

$$C_2 \exp(-(F + I) \cdot X) + A/(I + F) = C_1 + A/F \tag{5.31}$$

equation (5.8) is modified to be

$$C_1 \exp(-B \cdot (X + 2L)) + A/F = C_2 + A/(F + I) \tag{5.32}$$

Solving the equations for C_1 and C_2 yields

$$C_1 = -[A \cdot I/(F \cdot (F + I))] \cdot [\exp(-(F + I) \cdot X) - 1] / [\exp(-((2F + I) \cdot X) + (2F \cdot L)) - 1] \tag{5.33}$$

$$C_2 = +[A \cdot I/(F \cdot (F + I))] [\exp(-(F \cdot X) + (2F \cdot L)) - 1] / [\exp(-((2F + I) \cdot X) + (2F \cdot L)) - 1] \tag{5.34}$$

We now have equations for the expected number of crimes in a region at any point in time as well as solutions for the constants C_1 and C_2. The next step is to define a single equation that specifies the average expected number of crimes, $\overline{S1}$, in a region for a complete cycle of length $2X + 2L$. That equation is generated by integrating equation (5.29) (a period of search) over a time period of X minutes and equation (5.30) (a period of no search) over a time period of $X + 2L$ minutes. The resultant equation divided by the length of a cycle is just

$$\overline{SI} = \frac{1}{2(X+L)} \int_0^X C_2 \exp(-(I + F) \cdot t) + A/(I + F)dt + \frac{1}{2(X+L)} \int_0^{X+2L} C_1 \exp(-F \cdot t) + A/F\, dt \qquad (5.35)$$

Integrating out, substitutung for C_1 and C_2 and combining the terms produces the following equation for $\overline{SI}$, as a function of the variable X and the parameters L, I and F.

$$\overline{SI} = A/F - \left[(A \cdot I \cdot X)/(2(X + L) + F \cdot (F + I))\right] + \left(\frac{A}{2(X+L)}\right) \cdot \left(\frac{I}{F \cdot (F+I))}\right)^2 \cdot \left[1 + \frac{2 - \exp(-(I + F) \cdot X) - \exp(-(X + 2L) \cdot F)}{\exp(-2F \cdot X - 2F \cdot L - I \cdot X) - 1}\right] \qquad (5.36)$$

Although equation (5.36) is a relatively complicated expression for the average expected number of crimes in progress, once values have been assigned to F, I and L, it is a function of only one variable, X. Therefore, it is a straightforward task to carry out a one dimensional search for the optimum value of X for the given parameters. In Figure 5.1 the optimum value of X is shown for a range of L, F and R, where R is the ratio of F, the finishing rate, and I, the interception rate. There are graphs for two values of F, one for F equals 20, in which case crimes last an average of three minutes, and one for F equals 10, in which case crimes average 6 minutes. For each F a range of ratios is presented with R as high as 20 (Crimes are completed twenty times as fast as they can be detected.) and as unrealistically low as .01 which means that crimes are intercepted at a rate one hundred times as fast as they finish on their own. The lower values for R were included less for realism than to display how the curves behave as R approaches the limit of zero.

Looking at the series of curves in Figure 5.1 a number of points stand out, but perhaps the most striking characteristic is the asymptotic nature of each curve. For each value of R as L approaches some limiting value (different for each R), the optimal value of X goes off to infinity. However, an infinite value for X is the equivalent of searching only one region. This means that for given values of F and R, as L increases above some point the optimal solution is always to search one region. For example for crimes of a three minute duration (F=20) when R equals 10, as L increases above (approximately) .3 minutes it no longer pays to switch regions; the same is true for crimes of 6 minutes, R equal to five and as L increases above 1 minute. Another interesting general observation is that the asymptote for each curve is less than the mean duration of the crime and in the limit as R approaches zero the asymptote approaches the mean duration of a crime. In other words, if the travel time between regions is greater than the mean duration of a crime, then, no matter what the detection capabilities, switching regions is counter-

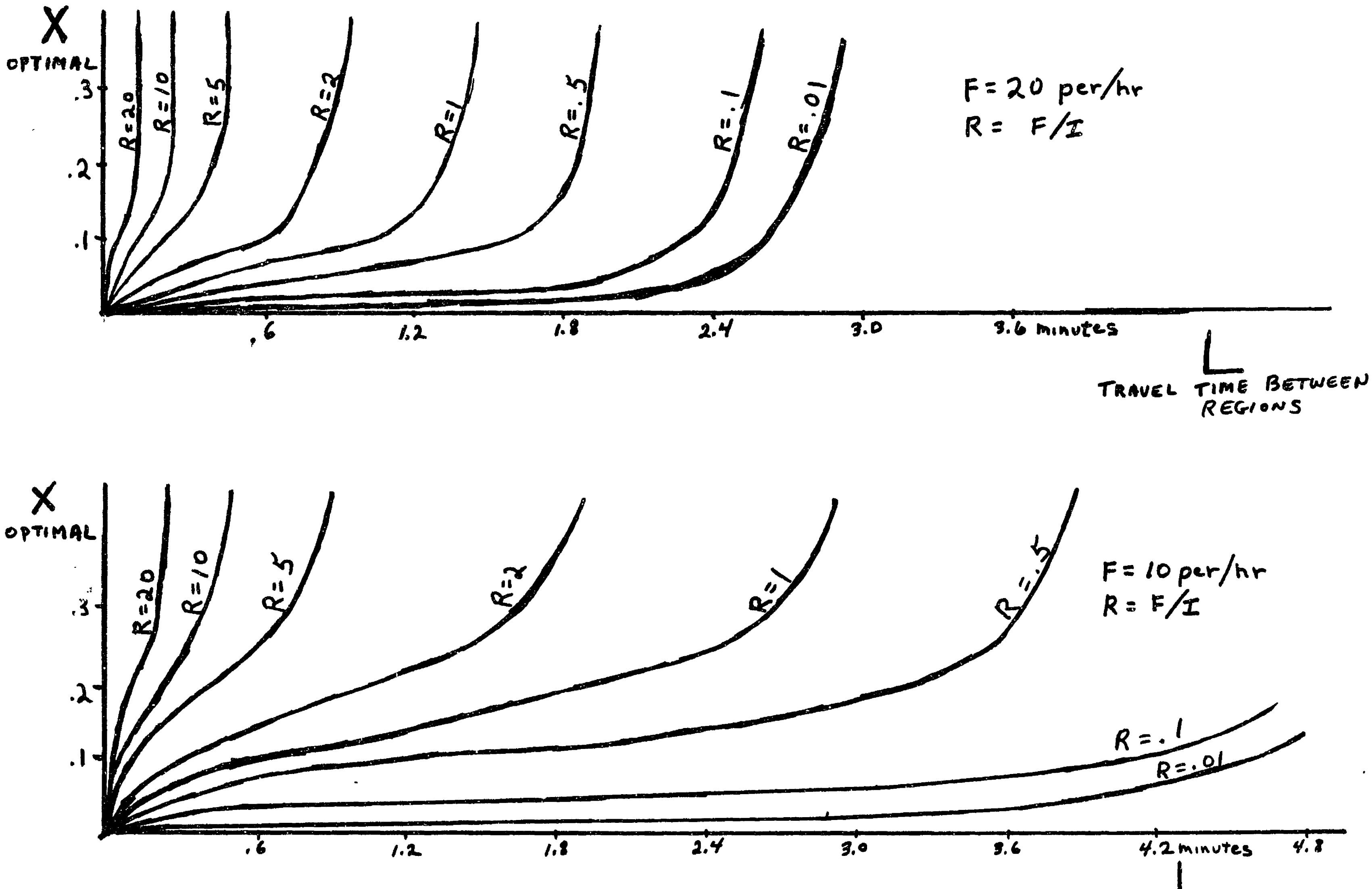


Figure 5.1: Optimal Search Strategies as a Function of the Time to Travel Between Regions

productive. The higher payoff that would accrue initially from switching to the other region is outweighed by the time lost in travelling there.

Two additional remarks about the curves, which are corollaries of the above points, are

1) For a given value of L as the ratio, R, decreases, the optimum X also decreases, and
2) As F decreases (from 20 to 10) the value of the asymptote for a given R increases.

All of the above issues will be addressed more formally in the succeeding section in which an analytic expression for the asymptote as a function of F and R is derived.

5.3.2 When to Search Only One Region

The development of an analytic expression to search only one of the regions builds on the realization that the curve for the average expected level of crime ($\overline{SI}$) as a function of X (i.e. the length of a visit to a region) can take on only one of two forms. (See curves B and C in Figure 5.2) For both forms as X approaches zero, the average number of crimes in progress increases and approaches A/F, the steady state value when no search is carried out. As X increases, however, the curves behave differently. In the first curve (B), $\overline{SI}$ monotonically decreases asymptotically to a limiting value. Consequently the optimal value for X is infinite (i.e. search only one region). In the second curve (C) as X increases, the

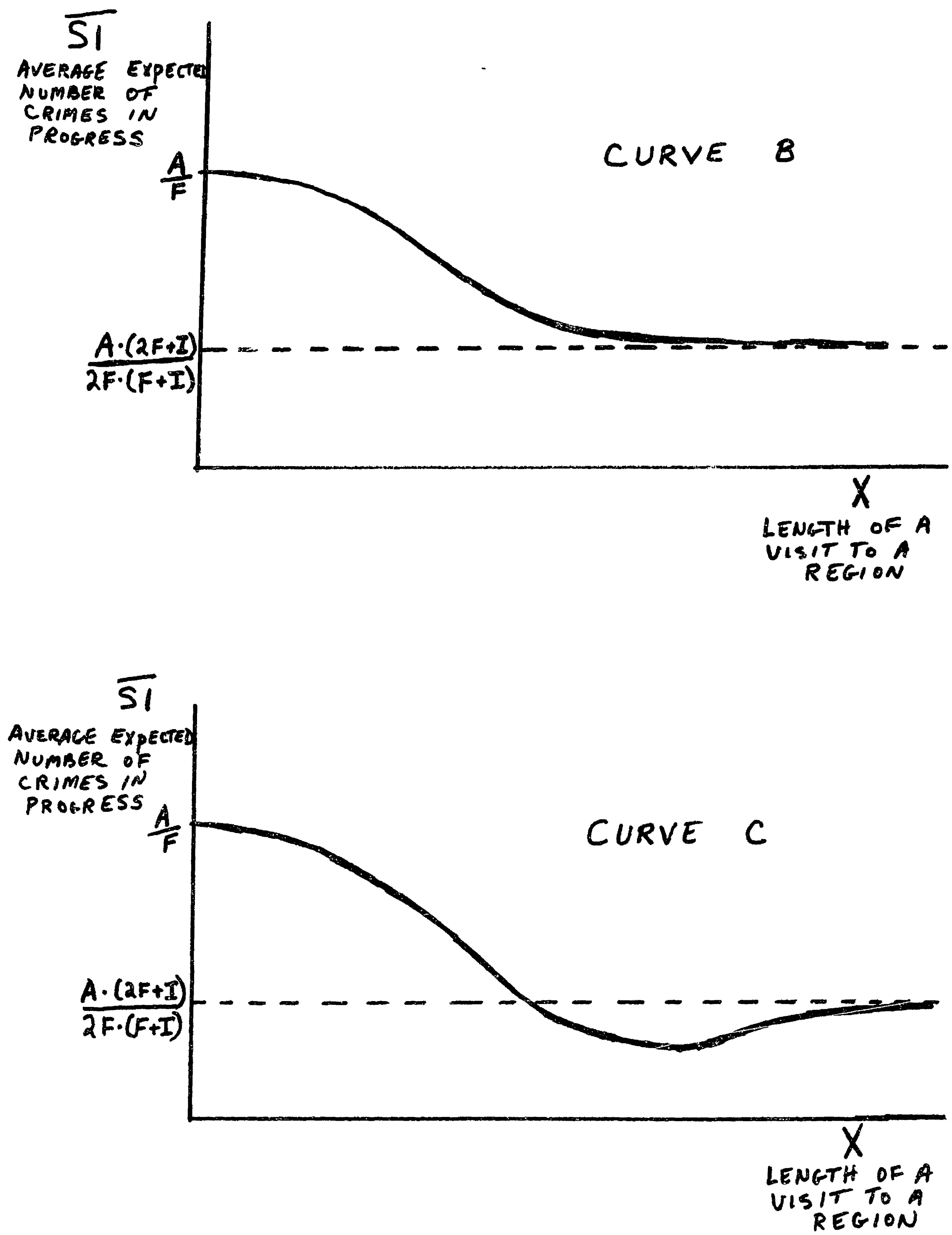


Figure 5.2: Two Types of Curves for the Average Expected Number of Crimes in Progress as a Function of the Length of a Visit to a Region

level of crime again decreases until a minimum value is obtained for a finite value of X. Then as X continues to increase, $\overline{SI}$ increases asymptotically to a limiting value. In both instances the limiting value is $A \cdot (2F + I)/(2F\cdot(F+I))$ (See equation (5.14).)

The key, then, to our problem lies in analyzing the derivative of $\overline{SI}$, $d\overline{SI}/dX$, (or equivalently $d\overline{S}/dX$) as X approaches infinity to determine not whether the derivative approaches zero, since it always will, but rather the direction from which it approaches zero. If the derivative approaches zero from the positive direction then some finite X is optimal, otherwise infinite X is optimal.

The derivative of $\overline{SI}$ is 'simply'

$$\begin{aligned}\frac{d\overline{SI}}{dX} = {} & \frac{-A \cdot I}{2F\cdot(I+F)\cdot(X+L)^2} \cdot \left[L + \frac{I}{F\cdot(I+F)}\right] \\ & + \frac{A \cdot I^2}{2(X+L)\cdot F^2\cdot(I+F)^2} \left[\frac{-1}{X+L} \cdot \frac{2 - \exp(-(I+F)\cdot X) - \exp(-F\cdot(2L+X))}{\exp(-2F\cdot(X+L) - I\cdot X) - 1}\right. \\ & + \frac{(\exp[-2F\cdot(X+L) - I\cdot X] - 1) \cdot ((I+F)\cdot\exp(-(I+F)\cdot X + F\exp(-F\cdot(X+2L))}{(\exp(-2F\cdot(X+L) - I\cdot X) - 1)^2} \\ & \left. + \frac{[2 - \exp(-(I+F)\cdot X) - \exp(-F\cdot(X+2L))]\cdot[(2F+I)\cdot\exp(-2F\cdot(X+L) - I\cdot X)]}{(\exp(-2F\cdot(X+L) - I\cdot X) - 1)^2}\right]\end{aligned} \tag{5.37}$$

However, with regard to the limiting behavior of this equation some of the **components** can be immediately disregarded since some approach zero faster than others. The slower components, of course, dominate in the limit, which allows us to eliminate all expressions that approach zero faster than $1/(X + L)^2$. Therefore the problem reduces to analyzing

$$\lim_{X\to\infty} \frac{d\overline{SI}}{dX} = \frac{-A\cdot I}{(X+L)^2\cdot 2F\cdot(I+F)} \left[L + \frac{I}{F\cdot(I+F)} + \frac{I}{F\cdot(I+F)} \cdot \left(\frac{2-\exp(-(I+F)\cdot X) - \exp(-F\cdot(2L+X))}{\exp(-2F\cdot(X+L) - I\cdot X) - 1} \right) \right] \tag{5.38}$$

(In Appendix F we prove that all other expressions in equation (5.37) approach zero faster than $1/(X+L)^2$.) The focus though is on the direction from which the derivative approaches zero which reduces the expression of interest still further to

$$\lim_{X\to\infty} -\left[L + \frac{I}{F\cdot(F+I)} + \frac{I}{F\cdot(F+I)} \cdot \left(\frac{2-\exp(-(I+F)\cdot X) - \exp(-F\cdot(2L+X))}{\exp(-2F\cdot(X+L) - I\cdot X) - 1} \right) \right] \tag{5.39}$$

The limit of this expression is

$$-\left[L + \frac{I}{F\cdot(I+F)} + \frac{I}{F\cdot(I+F)} \cdot \frac{2}{-1} \right]$$

or equivalently

$$\frac{I}{F\cdot(F+I)} - L = \frac{1}{F(F/I + 1)} - L \tag{5.40}$$

This will be negative and the optimal value of X is infinite whenever

$$L > \frac{1}{F\cdot(F/I + 1)} \qquad (5.41)$$

This inequality states that if the travel time is greater than the mean lifetime of a crime (I/F) divided by one plus the ratio of the competion rate to the interception rate (F/I), then is is not worthwhile to switch regions. This inequality consequently confirms the graphical analysis (Figure 5.1) that was presented earlier. For example, since $1/(F/I + 1)$ is always less than one if L, the travel time is greater than $1/F$ (the mean duration of a crime), it must also be greater than the previous expression.

$$L > \frac{1}{F} > \frac{1}{F}\cdot\frac{1}{(F/I)+1}$$

This means that if it takes longer to travel between regions than the average duration of a crime, then no matter what the detection rate is, the optimal strategy will be to search a single region. Perhaps more significantly if the completion rate of crimes were, for example, ten times the interception rate (by no means an unreasonable figure) and the travel time were longer than one-eleventh the duration of a crime (e.g. 11 seconds for two minute crimes; 16 seconds for three minute ones, etc.), it does not pay to switch regions.

In light of the above inequality, the short duration of crimes, and high rate of completion relative to interception, the propensity will be to search one region even when the two regions generate crimes at the same rate. Once the crime rates vary this tendency is further compounded. Unequal call rates alone, even without penalties for transferring between regions, may also produce optimal strategies which limit the search to a single region. It is this last phenomenon that is explored in the next section. However, in order to isolate the effects of just unequal crime rates we will not include in the example time lost in travel between regions.

5.4 Two Regions - Differing Crime Rates - No Time Lost

5.4.1 Introduction

The final example to be analyzed involves two regions with differing crime rates. RL, the low crime region, has crimes arriving at a rate of A, while RH, the high crime region, has crimes arriving at a rate of M • A where M is greater than one. In both regions all crimes have the same completion and interception rates, F and I respectively. Once again our analysis will revolve about cyclic policies, this time of the form , X minutes in RL followed by K • X minutes in RH.

Although one problem of interest is, of course, the finding of the (K,X) pair which minimizes the average expected number of crimes in progress, the discussion will not be limited to that, since in all instances the optimum is approached as X tends towards zero, an unimplementable optimum. Therefore the development will also address the issue of the optimum value of K for a given value of X. Setting X to be a specified value is interpretable as establishing a feasibility constraint on the search process. The constraint states that whenever a patrol unit enters the low crime region, RL, it must patrol there for <u>at least</u> X minutes. Completing the discussion of this section and more in line with the thrust of the rest of the chapter is the description of conditions under which patrolling only the high crime region is optimal. In the analysis solutions of that form will appear as the optimal value of K being infinite.

The discussion that will follow, then, can be categorized briefly as:

1. A single expression is developed to be used in searching for the optimal value of K for a given X.
2. An analytic expression is found for K optimal as X approaches zero.
3. Evolving directly from 2 is an expression, in terms of F and I, for determining, for which values of M are the optimal K infinite for all values of X.
4. Lastly an expression is found which specifies the values of X, as a function of M, F, and I, for which the optimal K is again infinite.

5.4.2 Applying the Model

Unlike the previous two examples in which it was possible to focus on only one region because of symmetry, it is now necessary to define a different set of differential equations for each region. In the low crime region the equations remain as before with the solutions just

$$SL = C_1 \exp(-F \cdot t) + A/F \tag{5.42}$$

during a period of no search and

$$SL = C_2 \exp(-(F + I) \cdot t) + A/(F + I) \tag{5.43}$$

during a period of search. However, the continuity boundary conditions have changed since a cycle as viewed from the perspective of the low crime region is now X minutes of search followed by K·X minutes of no search. The simultaneous equations that are generated by the constraint to be used to solve for C_1 and C_2 are

$$C_2 \exp(-(F+I)\cdot X) + A/(F+I) = C_1 + A/F \tag{5.44}$$

which equates the level of crime at the end of a period of search to the level at the beginning of no search, and

$$C_1 \exp(-F\cdot K\cdot X) + A/F = C_2 + A/(F+I) \tag{5.45}$$

which equates the levels at the end of a period of no search and the beginning of a period of search. Solving for C_1 and C_2 yields

$$C_1 = \frac{-A\cdot I}{F\cdot(I+F)} \cdot \left[\frac{1-\exp(-F+I)\cdot X}{1-\exp(-(F\cdot K + F + I)\cdot X}\right] \tag{5.46}$$

and

$$C_2 = \frac{A\cdot I}{F\cdot(I+F)} \cdot \left[\frac{1-\exp(-F\cdot K\cdot X)}{1-\exp(-(F\cdot K + F + I)\cdot X}\right] \tag{5.47}$$

In the high crime region, RH, the differential equations are slightly different from before because crimes are arriving at a rate of MA. The equations therefore are

$$\frac{dSH}{dt} = SH' = M\cdot A - F\cdot SH \tag{5.48}$$

during no search and

$$SH' = M \cdot A - (F+I) \cdot SH \qquad (5.49)$$

during search. The solutions are similar to equations (5.42) and (5.43) only with $M \cdot A$ instead of just A.

$$SH = C_3 \exp(-F \cdot t) + M \cdot A/F \qquad (5.50)$$

$$SH = C_4 \exp(-(F+I) \cdot t) + M \cdot A/(I+F) \qquad (5.51)$$

To solve for C_3 and C_4, we use these simultaneous equations, which are almost the mirror image of equations (5.41) and (5.45), and produce the following solutions:

$$C_3 = \left(\frac{-M \cdot A \cdot I}{F \cdot (F+I)}\right) \cdot \left(\frac{1-\exp(-(F+I) \cdot K \cdot X)}{1-\exp(-(F+I) \cdot K \cdot X - F \cdot X)}\right) \qquad (5.52)$$

$$C_4 = \left(\frac{M \cdot A \cdot I}{F \cdot (F+I)}\right) \cdot \left(\frac{1-\exp(-F \cdot X)}{1-\exp(-(F+I) \cdot K \cdot X - F \cdot X)}\right) \qquad (5.53)$$

Now that the preliminary groundwork has been laid out, the next step is to develop a single expression for the average expected number of crimes in progress, $\overline{S}$, in the two regions combined. This expression is generated by integrating each of the four equations ((5.42). (5.43), (5.50) and(5.51)) over their associated portions of the cycle and dividing their sum by the length of a cycle, $(K + 1) \cdot X$. Carrying out the above, results in

$$\overline{S} = \frac{1}{(K+1)\cdot X} \int_0^{K\cdot X} C_1 \exp(-F\cdot t) + A/F \, dt$$

$$+ \frac{1}{(K+1)\cdot X} \int_0^{X} C_2 \exp(-(F+I)\cdot t) + A/(I+F) \, dt$$

$$+ \frac{1}{(K+1)\cdot X} \int_0^{X} C_3 \exp(-F\cdot t) + M\cdot A/F \, dt$$

$$+ \frac{1}{(K+1)\cdot X} \int_0^{K\cdot X} C_4 \exp(-(F+I)\cdot t) + M\cdot A/(I+F) \, dt \qquad (5.54)$$

The first two components represent the average number of crimes in progress in the low crime region, $\overline{SL}$, and the last two, the average in the high crime region $\overline{SH}$. After integrating out, substituting where necessary, and combining terms wherever possible, the result is (and by no means simply)

$$\overline{S} = \frac{M\cdot A}{F+I} + \frac{A}{F} + \frac{A\cdot I\cdot(M-1)}{(K+1)\cdot F\cdot(F+I)} - \left(\frac{1}{(K+1)\cdot X}\right) \cdot \left[\frac{A\cdot I^2}{F^2\cdot(F+I)^2}\right] \cdot$$

$$\cdot \left[\frac{M\cdot(1-\exp(-F\cdot X))\cdot(1-\exp(-(F+I)\cdot K\cdot X))}{1-\exp(-(F\cdot(K+1) + I\cdot K)\cdot X)}\right.$$

$$\left. + \frac{(1-\exp(-F\cdot K\cdot X)\cdot(1-\exp(-(F+I)\cdot X))}{1-\exp(-(F\cdot(K+1)+I)\cdot X)}\right] . \qquad (5.55)$$

As complicated as this expression looks , finding the optimal value of K for a given set of M, F, I, and X still requires only a one dimensional search that is easily carried out by computer. (A patient individual could find the optimum with a calculator.) Before displaying graphically the optimal value of K for a range of M, F, I and X, we will note one facet of the equation that should be stored for later reference. The first two components, M·A/(F+I) and A/F, do not vary with either K or X. The third <u>decreases</u> as <u>K</u> increases, is positive (M is greater than 1) and is also <u>independent</u> of X. The last component is unique both in that it is negative and also in that it is a function of <u>both</u> K and X. This information will later be used to determine for which values of X is the optimal value of K infinite (Section 5.4.5).

Figures 5.3 through 5.6 graph the optimal value of K for different values of X. In all four figures crimes are of a five minute duration (F=12), but both I and M are allowed to vary. A common characteristic of all curves, except for the M equal to one curve, is that as X (the minimum duration of a visit to the low crime region) grows larger, the optimal K eventually becomes infinite. (In the next section an analytic expression will be presented for determining that value of X.) This cutoff value or asymptote decreases both as M increases and as R (the ratio of F to I) increases. The latter can be seen, for example, by comparing the curves for which M is 1.2 in each of the four figures. In the first figure the cutoff value of X is at approximately .43 hours and in the succeeding figures the cutoff decreases to .33, .23 and .18 hours respectively. Another aspect of these curves that stands out is that as R increases the initial value (at X equal to zero) of K optimal increases. Thus it is not surprising that for large enough R it does not pay to visit

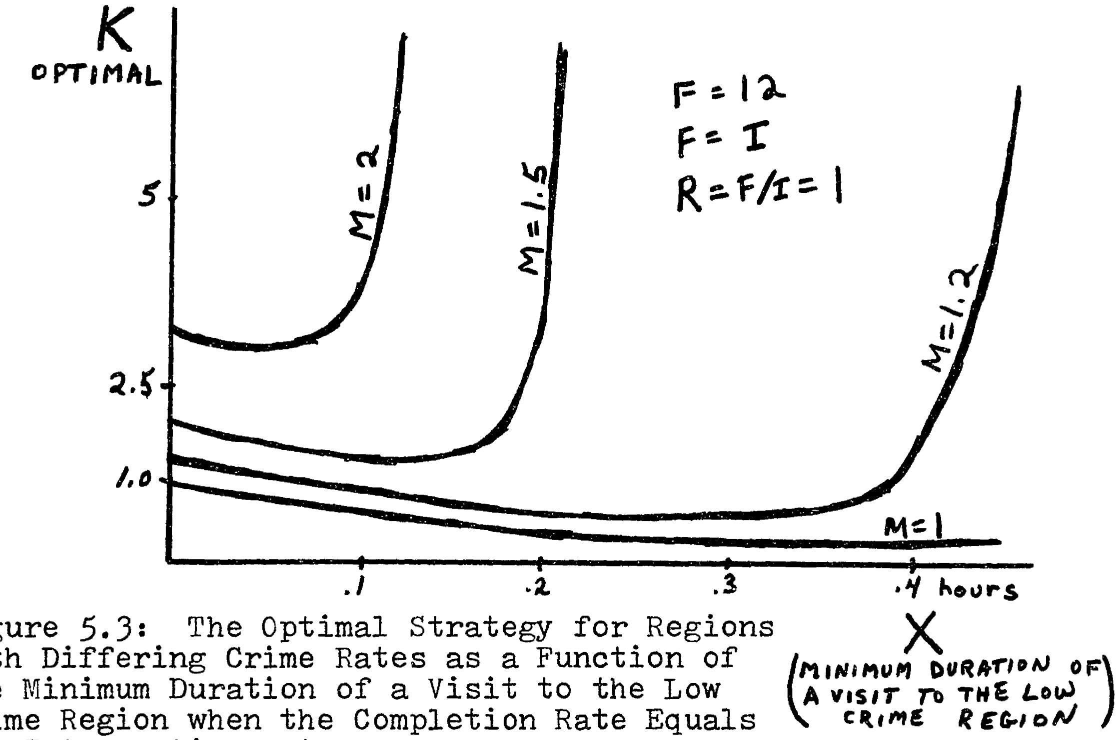


Figure 5.3: The Optimal Strategy for Regions with Differing Crime Rates as a Function of the Minimum Duration of a Visit to the Low Crime Region when the Completion Rate Equals the Interception Rate.

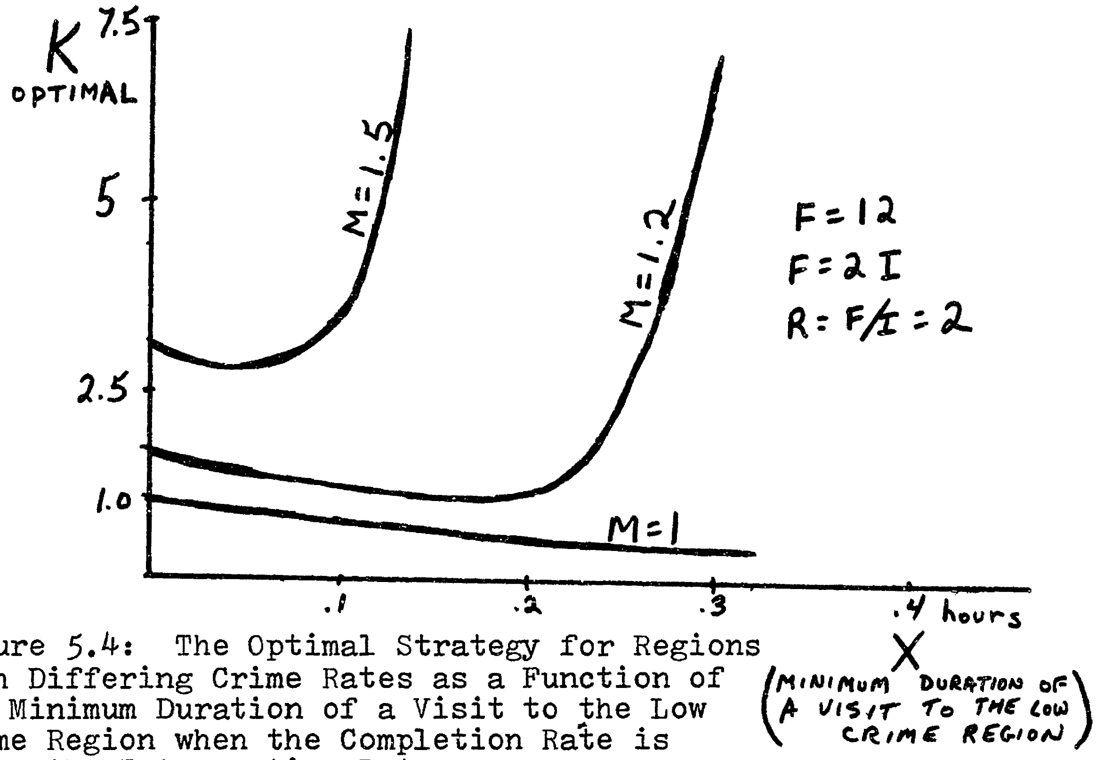


Figure 5.4: The Optimal Strategy for Regions with Differing Crime Rates as a Function of the Minimum Duration of a Visit to the Low Crime Region when the Completion Rate is Twice the Interception Rate

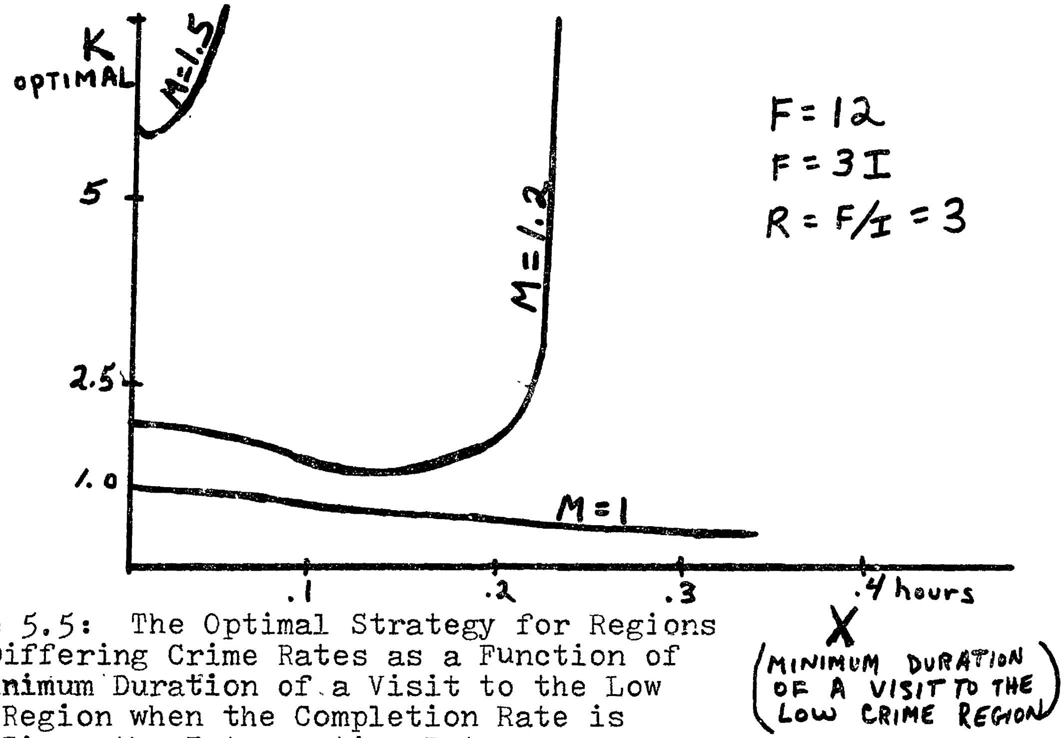


Figure 5.5: The Optimal Strategy for Regions with Differing Crime Rates as a Function of the Minimum Duration of a Visit to the Low Crime Region when the Completion Rate is Three Times the Interception Rate

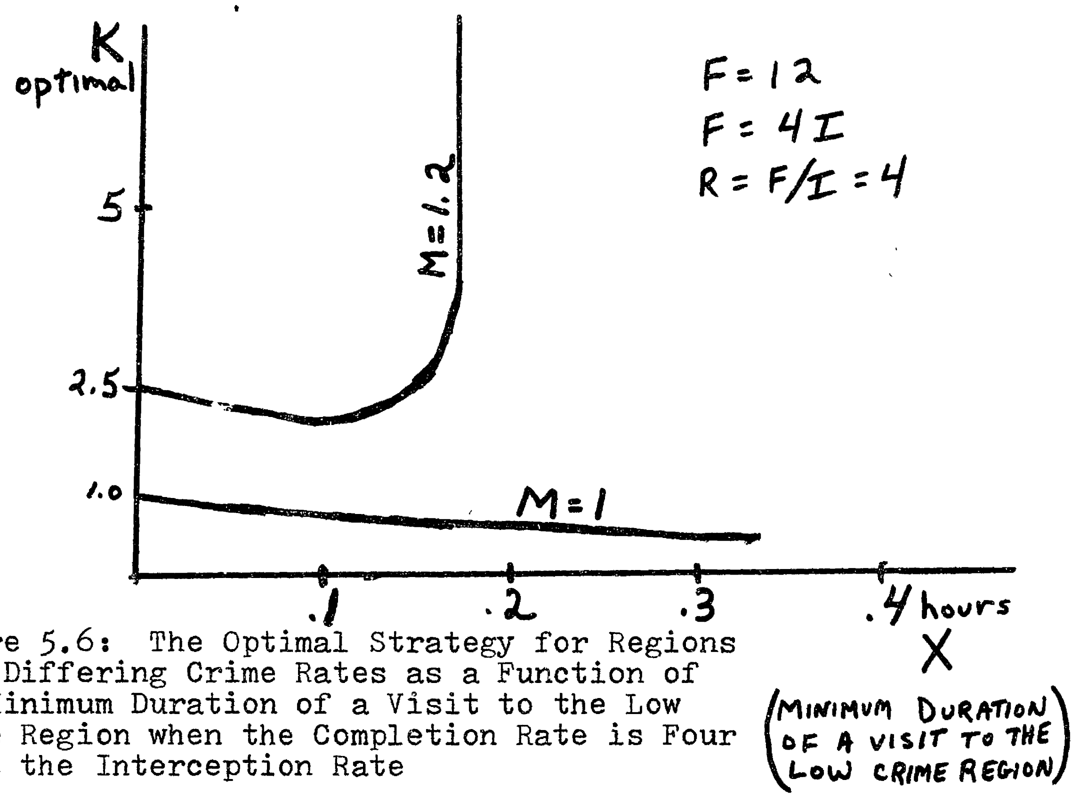


Figure 5.6: The Optimal Strategy for Regions with Differing Crime Rates as a Function of the Minimum Duration of a Visit to the Low Crime Region when the Completion Rate is Four Times the Interception Rate

the low crime region (K optimal is infinite) even though the minimum duration of a visit to that region is infinitesimally short (X near zero).

In completing the discussion of these figures, the focus will be on one behavioral characteristic of the curves that seems counterintuitive. In each of the curves as X initially increases from zero, the optimal value of K decreases. This means that by constraining the search in the low crime area to be of a minimum continuous duration, the optimal solution may reduce the total fraction of time (K/K+1) spent in the high crime area, even to the point where it is allocated less than half the search. This is precisely what occurred for the curve marked M=1.2 in Figure 5.3.

This phenomenon occurs consistently for the M=1 curves. Even though the two regions generate crimes at the same rate, if a constraint is placed on the minimum continuous duration of a search in one of the regions the optimal solution may allocate less than half the search to the other region. The explanation that eventually became apparent was that there were two conflicting forces at work. The more intuitive one was that as X increases the searcher is forced to incur increasing diminishing return in any visit to the low crime region and eventually the cost for visiting the region RL becomes so prohibitive that it no longer pays to search the region. This force dominates in the long run. However there is an advantage in general to having short cycles and as X increases there is only one way of restraining the increase in cycle length and that is by decreasing K. This propensity for shorter cycles seems to dominate the behavior of the optimal value of K for small values of X turning K optimal initially into a decreasing function of X.

Because the differential equation model allows for an explicit constraint on the continuous duration of a search in only one of the regions, and because the optimal value of K can be less than one, any desired constraint on the other region must be made post hoc. For example, suppose that both the high and low crime regions were to have the same lower limit on the duration of a visit. The differential equation model might find the optimal K to be less than one, which violates the high crime region constraint. Under those circumstances, the optimal feasible solution is to set K equal to one.

Although in the above discussion we have noted that the optimal K may decrease with X, this should in no way be confused with the impact that increasing X has on the crime level. Figure 5.7 shows that the level of crime increases as X increases even though the searcher consistently uses the optimal strategy for that X and that the optimal value of K decreases initially as X increases. (See the corresponding curve in Figure 5.6).

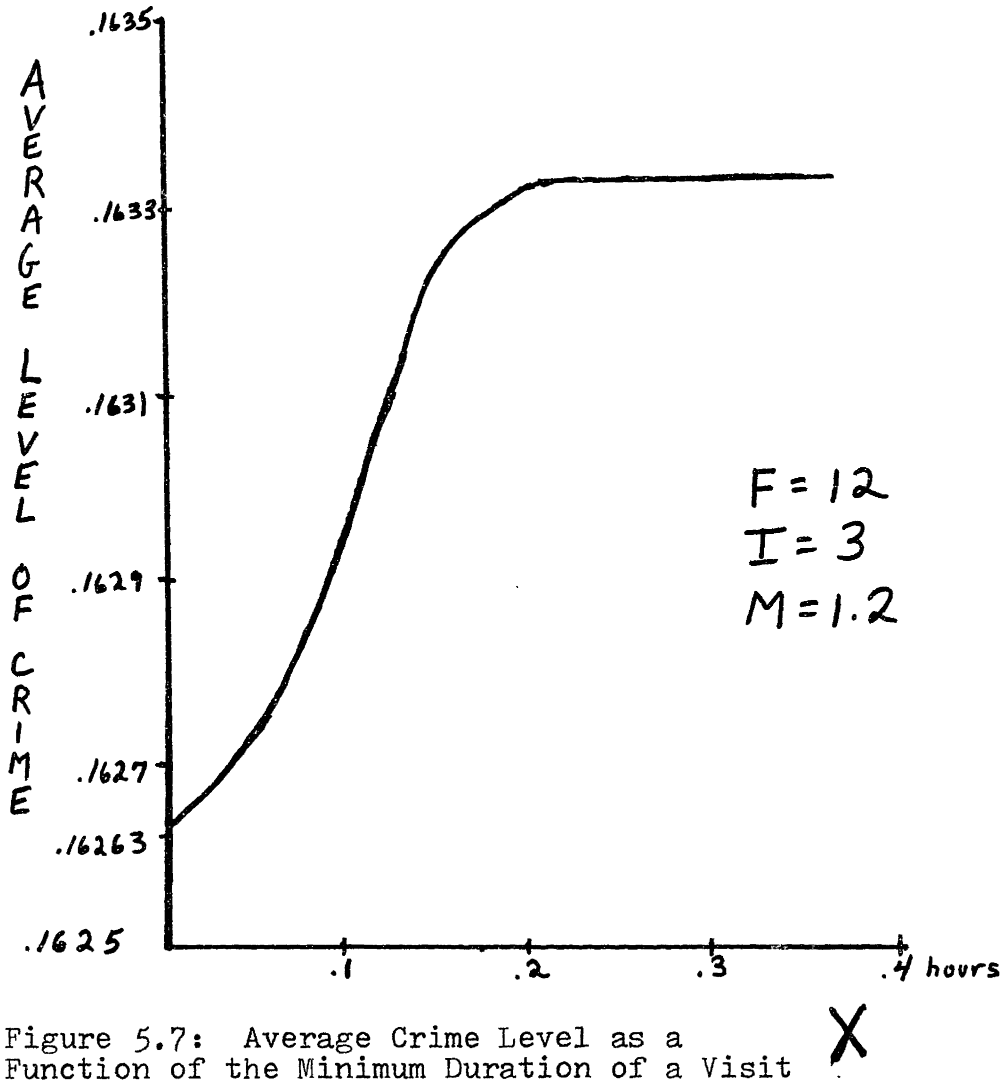


Figure 5.7: Average Crime Level as a Function of the Minimum Duration of a Visit to the Low Crime Region:Optimal Strategies are Consistently Used

5.4.3 No Constraint on the Minimum Duration of a Visit

In the previous section, an expression for $\overline{S}$, the average number of crimes in progress, was presented and was used to determine the optimal values of K over a range of values for X. In this section we will narrow our perspective to telescope in on the limiting behavior of $\overline{S}$ as X approaches zero (i.e. no constraint on the minimum duration of a visit to a region). Although as X approaches zero the optimal solutions are no longer implementable, the analysis of the limit properties will prove significant, mostly because the absolute minimum of $\overline{S}$ is also approached as X goes to zero. (See Figure 5.3 in the previous section for an illustrative example.) In addition this analysis will be facilitated tremendously by the relative tractability of the expression for the limit as compared to the expression for $\overline{S}$ itself.

The obvious first step, then, is to determine the limit of $\overline{S}$ as X goes to zero. Referring back to the general equation for $\overline{S}$ (equation (5.55)) the only component that is affected by changing X is

$$-\frac{1}{K+1}\cdot\left(\frac{A\cdot I^2}{F^2\cdot(F+I)^2}\right)\cdot\left[\frac{1}{X}\left(\frac{M\cdot(1-\exp(-F\cdot X))\cdot(1-\exp(-(F+I)\cdot K\cdot X))}{1-\exp(-(F\cdot(K+1)-I\cdot K)\cdot X)}\right.\right.$$
$$\left.\left.+\frac{1-\exp(-F\cdot K\cdot X))\cdot(1-\exp(-(F+I)\cdot X)}{1-\exp(-(F\cdot(K+1)+I)\cdot X)}\right)\right] \qquad (5.56)$$

In order to find the limit of that expression, it is not possible to substitute in X equal to zero because the result is zero divided by zero. Instead L'hôpital's rule is applied twice to the expression and the resultant value for its limit is

$$\frac{-A\cdot I^2}{F\cdot(F+I)\cdot(K+1)} \cdot \left[\frac{M}{F\cdot(K+1) + I\cdot K} + \frac{1}{F\cdot(K+1) + I}\right] \qquad (5.57)$$

Combining this with the components of $\overline{S}$ which are not functions of X, yields the following for the limit of $\overline{S}$

$$\lim_{x\to 0} \overline{S} = \frac{M\cdot A}{I+F} + \frac{A}{F} \div \frac{A\cdot I\cdot(M-1)}{(K+1)\cdot F\cdot(F+I)} - \frac{A\cdot I^2}{F\cdot(F+I)\cdot(K+1)} \cdot \left[\frac{M}{F\cdot(K+1) + I\cdot K} + \frac{1}{F\cdot(K+1)+I}\right] \qquad (5.58)$$

Simpler than the general expression for $\overline{S}$, this expression can be analyzed to find the optimal K by the standard procedure of setting the derivative equal to zero and solving for K. However, prior to doing just that, we will again (See section 5.1) display the equivalence between the two statistics, the probability of intercepting a crime and the reduction in crime as a result of the search process.

Since we are considering cycles that are infinitesimal in length, it is possible to calculate directly the probability of intercepting a random crime for a cycle of K·X minutes in RH and X minutes in RL. For any crime arising in RH and lasting for T minutes the patrol unit will be in the same region on patrol for (K/K+1)·T minutes during the commission of the crime. Therefore the probability of intercepting the crime is

$$\int_0^\infty F\exp(-F\cdot t)\cdot(1-\exp(-I\cdot(K/K+1)\cdot t)\ dt \qquad (5.59)$$

which integrated yields

$$\frac{I\cdot(K/K+1)}{F + I\cdot(K/K+1)} \qquad (5.60)$$

The equation can be interpreted as follows. Imagine two streams of targets leaving a system in two separate Poisson processes, one at a rate F (completion), the other at a rate I (interception). In addition targets that wish to leave in the second stream (rate I) must also face a lottery before they can leave. This lottery allows them to leave in that stream with a probability of K/K+1 (i.e. the fraction of time the searcher is in the high crime region). Of all the targets that leave the system, the fraction that depart in the second stream is just the above expression. Similarly for the low crime region, RL, the probability of intercepting a random crime is analogously

$$1 - F/(F + I\cdot(1/(K+1)) = 1 - [F\cdot(K+1)/((K+1)\cdot F + I] \qquad (5.61)$$

Since the proportion of all crimes that occur in region RH is (M/(M+1)) and in region RL, 1/(M+1), the probability of intercepting a random crime is

$$\frac{M}{M+1} \cdot \left(1 - \frac{F\cdot(K+1)}{F\cdot(K+1) + I\cdot K}\right) + \left(\frac{1}{M+1}\right)\cdot \left(1 - \frac{F\cdot(K+1)}{F\cdot(K+1)+I}\right)$$

$$= 1 - \frac{F\cdot(K+1)}{M+1} \cdot \left[\frac{M}{F\cdot(K+1) + I\cdot K} + \frac{1}{F\cdot(K+1)+I}\right] \qquad (5.62)$$

With regard to the differential equation model the measure that will be compared to the interception probability, is again simply the proportional reduction in the average number of crimes in progress as a result of the search process. When no search is carried out in both regions the average number of crimes in progress is

$$M\cdot A/F + A/F = A\cdot \frac{(M+1)}{F} \qquad (5.63)$$

The reduction in the crime level is then just one minus the ratio of, the number of crimes in progress for the particular search strategy (equation (5.58)) over the average when no search is carried out. This is abbreviated as

$$1 - \lim_{x\to 0} \frac{\overline{S}}{(A\cdot(M+1)/F)} \qquad (5.64)$$

Substituting in the relevant expression for the lim of $\overline{S}$ and combining the appropriate terms the expression reduces to not surprisingly

$$1 - \frac{F\cdot(K+1)}{M+1} \cdot \left[\frac{M}{F\cdot(K+1) + I\cdot K} + \frac{1}{F\cdot(K+1)+I}\right] \qquad (5.65)$$

the same as the probability of interception, proving the equivalence between the two statistics even in the more general problem (two regions with unequal call rates).

We return now to our interrupted search for that analytic expression at the end of the rainbow which will be used to calculate the optimal value of K for limiting strategies. As was mentioned earlier, the procedure to be followed requires first the differentiation with respect to K of the expression for the limit of $\overline{S}$

$$\frac{d \lim\limits_{x\to 0} \overline{S}}{dK} = \frac{A\cdot I\cdot(1-M)}{F\cdot(F+I)\cdot(K+1)^2} - \frac{A\cdot I^2}{F\cdot(F+I)\cdot(K+1)^2} \cdot$$

$$\cdot \left[\frac{M}{F\cdot(K+1) + I\cdot K} + \frac{1}{F\cdot(K+1)+I^2}\right]$$

$$+ \frac{A\cdot I^2\cdot K}{F\cdot(F+I)\cdot(K+1)^2} \cdot \left[\frac{M\cdot(F+I)}{(F\cdot(K+1) + I\cdot K)^2} + \frac{F}{(F\cdot(K+1)+I)^2}\right] \qquad (5.66)$$

Although at first glance the derivative seems too complicated to be easily solvable for K, when the derivative is set to zero, it can be rewritten in a form more amenable to solution. By combining all the terms into a single fraction, the expression can be reformulated as a multiple of a quadratic expression of the type

$$f(K)\cdot(AK^2+BK+C) \qquad (5.67)$$

In particular the expression to be set equal to zero is

$$0= \frac{A\cdot I}{(F\cdot(K+1)+I)^2\cdot(F\cdot(K+1)\ +\ I\cdot K)^2}\cdot\Big[K^2\cdot\big[(1-M)\cdot F^2+I\cdot(2F+I)\big]$$

$$+\ K\cdot(2(1-M)\cdot F\cdot(F+I)\ +\ F^2-M\cdot(F+I)^2\Big] \qquad (5.68)$$

Thus any value of K which makes the quadratic component zero makes the entire expression zero. Therefore by using the formula for solving a quadratic equation we find K to be

$$K(\text{optimal})= \frac{-2F\cdot(F+I)\cdot(1-M)\pm 2I\cdot(2F+1)\cdot\sqrt{M}}{2(F+I)^2\ -\ 2M\cdot F^2} \qquad (5.69)$$

or equivalently

$$= \frac{-2R\cdot(R+1)\cdot(1-M)\pm 2(2R+1)\cdot\sqrt{M}}{2\cdot(R+1)^2\ -\ 2M\cdot R^2} \qquad (5.70)$$

where $R=F/I$. With the above substitution, it is clear that the optimal value for K does not depend upon the independent values of F and I but only on their ratio, thereby reducing, once again, the number of critical parameters. Table 5.2 displays the optimal K for a range of M and R. One particular value of M is of special interest. For M equal to one, which means that the two regions generate crimes at the same rate, the optimal value of K does not depend upon R but is always one (i.e. equal search in both regions), as was assumed earlier in the chapter. The obvious reason for this is that for M equal to one, equation (5.70) becomes

M \ R	.1	.5	1	2	5	10	20	30	40
	The Optimal Value of K								
1	1	1	1	1	1	1	1	1	1
1.1	1.06	1.10	1.15	1.27	1.71	3.00	85.0	∞	∞
1.2	1.12	1.20	1.32	1.59	3.01	45.0	∞	∞	∞
1.3	1.17	1.30	1.49	1.97	6.16	∞	∞	∞	∞
1.4	1.22	1.40	1.67	2.45	25.0	∞	∞	∞	∞
1.5	1.28	1.51	1.87	3.04	∞	∞	∞	∞	∞
2	1.52	2.05	3.12	13.1	∞	∞	∞	∞	∞
2.5	1.74	2.64	5.16	∞	∞	∞	∞	∞	∞
3	1.95	3.31	9.20	∞	∞	∞	∞	∞	∞
3.5	2.15	4.09	21.2	∞	∞	∞	∞	∞	∞
4	2.33	5.00	∞	∞	∞	∞	∞	∞	∞

Table 5.2: Optimal Cyclic Strategies for Regions with Differing Crime Rates and No Constraint on the Minimum Duration of a Visit to a Region

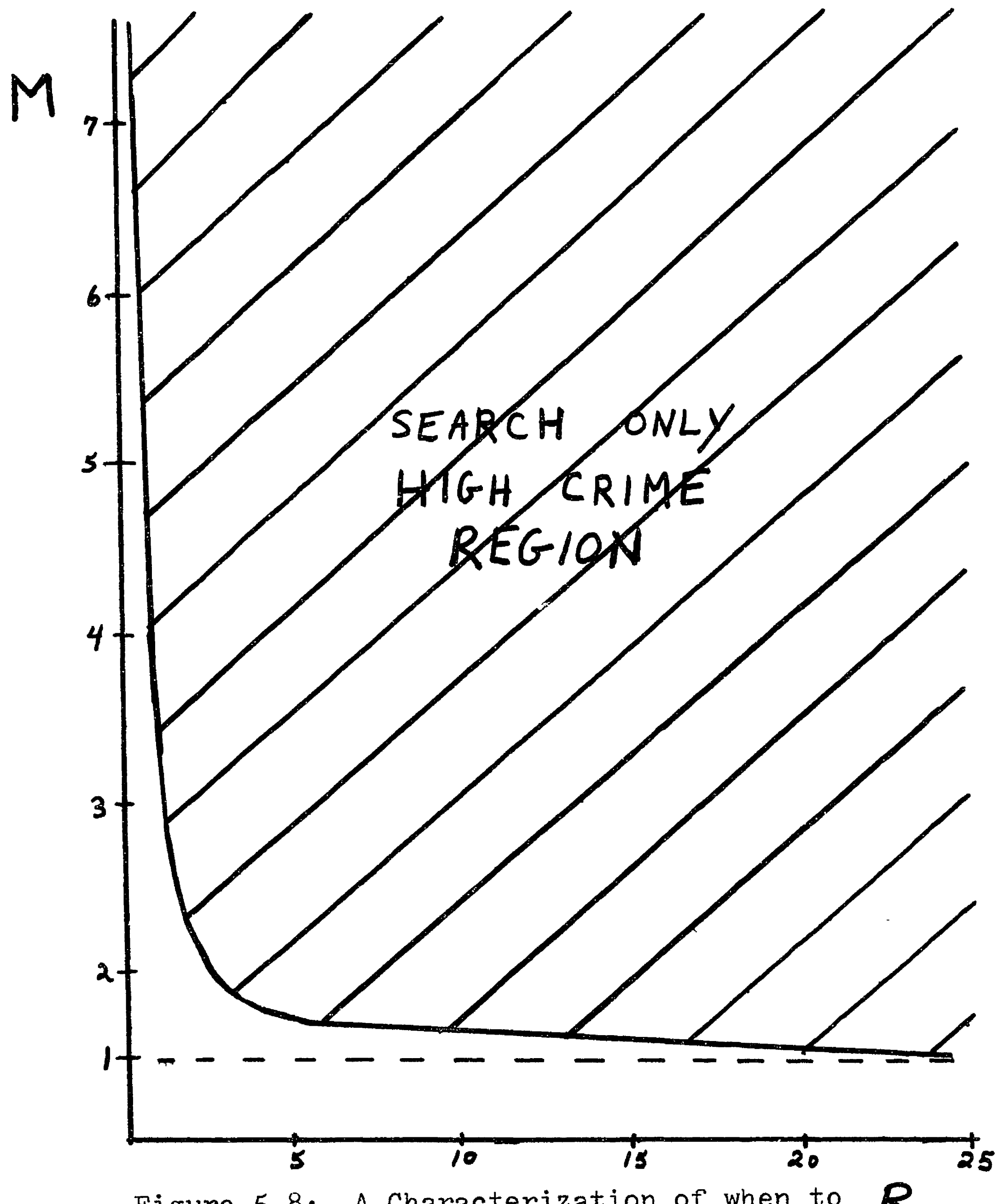


Figure 5.8: A Characterization of when to Search only One Region as a Function of Two Parameters: M, the Ratio of the Two Crime Rates, and R, the Ratio of the Completion and Interception Rates of Crimes

$$K(\text{optimal}) = \frac{2(2R+1)}{2(R+1)^2-2R^2} = \frac{2(2R+1)}{2(2R+1)} = 1 \qquad (5.71)$$

One aspect of the quadratic equation that needs clarification is that since there are two solutions it is possible for both solutions to be positive. As it turns out that cannot happen. In Appendix G it is proven that for values of M for which both values of the numerator in expression (5.69) are positive, the corresponding denominator will be negative. This last result provides, in addition, an almost too obvious bonus in that it contains the key to finding under what conditions the optimal solution is infinite.

5.4.4 When to Search Only One Region

Whenever both solutions to the aforementioned quadratic equation are negative, it seems intuitively clear that for M greater than one the optimal solution will be at the upper bound of the feasible region, namely K optimal will be infinite (i.e. search only one region). A direct corollary, therefore, of the last result is that the optimal solution is infinite when the denominator is negative or zero. Consequently if M (the ratio of the crime rates in the high and low crime areas) is larger than $(R+1)^2/R^2$,

$$M > \frac{(R+1)^2}{R^2} \qquad (5.72)$$

the search should be limited to the high crime area. To be rigorous, the above proof only showed that in the limit as X approaches zero that the inequality specifies whether or not the low crime region should be searched. It, however, is clear that if it does not pay to ever search RL for an infinitesimally short duration there will be no incentive to ever search RL if the minimum duration of a visit is instead some number larger than zero.

Now would seem the appropriate time to step back for a second to attempt to obtain a more intuitive feeling for the above inequality. The method to be used, and which in some ways could be considered this dissertation's other theme song, is to analyze the limits. As R increases, the right side of the inequality approaches one. So that when crimes are finishing at a fast pace relative to their interception rate, even small differences in the two regions force the concentration of search into one region. One explanation is that the relatively rapid completion rate does not allow for much diminishing return in the high crime region; nor, unfortunately, can even a total concentration of search effort in that region impact enough on the crime level to reduce it below that of the low crime region. Conversely as R goes to zero, the right hand side approaches infinity. This means that when the interception rate is relatively high even for large disparities in the crime rates there is reason to visit the low crime region every once in a while. Because of the relatively high interception rate there will be a significant amount of diminishing return from focusing search in RH. In addition because of the same high interception rate even infrequent searches in RL can have a high payoff in terms of reducing that region's crime leve. The two reasons together therefore conspire to include RL in the search.

Figure 5.8 displays the convex region in which no search would be carried out. For example if R (the ratio of completion to interception) were only 10 then if RH generated 21% more crimes than RL, only RH would be searched. <u>If more realistically R were 20, 30, or 40, then if RH generated respectively 10%, 7% or 5% more crimes than RL, again only the high crime region would be searched</u>. The result is a second (the first was based on travel times between regions) strong limitation on the likelihood of searching more than one region. If there are even small variations in the crime rate and the two areas are relatively large (i.e. the interception rate, which is a function of the size of the area, is large compared to the completion rate), the model suggests limiting search to the high crime area.

The last parameter to be microscopically examined for symptoms that may contribute to the tendency to limit the search to one region is the parameter X, the specified minimum duration of a visit to the low crime region. Tucked away earlier in the chapter and stored for later use was the key to determining the critical X above which it is no longer profitable to search region RL. Remember that in the expression for $\bar{S}$ (equation (5.55)) only two components, one positive and one negative, varied with K, with only the negative one a function also of X. The two components are

$$\frac{1}{(K+1)} \cdot \frac{A \cdot I \cdot (M-1)}{F \cdot (F+I)}$$

and

$$-\frac{1}{(K+1) \cdot X} \cdot \frac{A \cdot I^2}{F^2 \cdot (F+I)^2} \cdot \left[\frac{M \cdot (1-\exp(-F \cdot X)) \cdot (1-\exp(-(F+I) \cdot K \cdot X))}{1-\exp(-(F \cdot (K+1) + I \cdot K) \cdot X)} \right.$$

$$\left. + \frac{(1-\exp(-F \cdot K \cdot X) \cdot (1-\exp(-(F+I) \cdot X))}{1-\exp(-(F \cdot (K+1)+I) \cdot X)} \right]$$

If for some X the absolute value of the first component were greater than that of the second component for all values of K, the sum of the two would, of course, always be positive. However, since by driving K to infinity both expressions can be forced to zero, the obvious way to minimize the average number of crimes when the sum is positive, is to let K be infinite. The problem therefore reduces to finding under what conditions the following inequality holds.

$$\frac{A \cdot I \cdot (M-1)}{F \cdot (F+I)} > \frac{1}{X} \cdot \frac{A \cdot I^2}{F^2 \cdot (F+I)^2} \cdot (M \cdot f_1(X)+f_2(X)) \qquad (5.73)$$

The problem is all but solved by the realization that both $f_1(X)$ and $f_2(X)$ are bounded from above by one. This is easily seem once it is recognized that both $f_1(X)$ and $f_2(X)$

are of the form

$$\frac{(1-\exp(-B))\cdot(1-\exp(-C))}{1-\exp(-(B+C))} \qquad (5.74)$$

which is always less than one. As a result the inequality of interest becomes

$$\frac{A\cdot I\cdot(M-1)}{F\cdot(F+1)} > \frac{1}{X} \cdot \frac{A\cdot I^2}{F^2\cdot(F+I)^2} \cdot (M+1) \qquad (5.75)$$

Solving for X yields

$$X > \frac{(M+1)\cdot I}{(M-1)\cdot F\cdot(F+I)} = \frac{(M+1)}{(M-1)} \cdot \frac{1}{(F\cdot(F/I + 1)} \qquad (5.76)$$

which depends, not only on the ratio of F to I, but also on their individual values.

One easily identifiable characteristic of this inequality is that as M approaches one, the right hand side tends towards infinity. Interestingly and perhaps not too surprisingly this expression, except for the multiple (M+1/M-1), is identical to the inequality obtained earlier for L, the time lost in switching, equation (5.41)

$$L > \frac{1}{F\cdot(F/I + 1)} \quad .$$

As would be expected, the bound on X is strictly greater than the corresponding bound on L (i.e. (M+1/M-1) is greater than one) since in the latter case the time is totally lost while in the former, it merely has a reduced payoff because of the lower crime rate. In fact, this last expression can be considered a special case of equation (5.71). If we treat the travel time as if an equivalent time were spent in a second region in which no crimes occur then M (the ratio of crime rates in the high and low crime areas) would be infinite. In the limit as M approaches infinity, the expression M+1/M-1 approaches one. This reduces equation (5.76) to the inequality found previously for L.

5.5 Summary

In this chapter a differential equation model of a search process was developed which has the potential for capturing the dynamics of a sequential cyclic search strategy. In displaying the application of the model to a number of examples, we have derived a number of independent quantifiable constraints which limit the number of regions to be searched and which are summarized in Table 5.3. With regard to the two time parameters, L and X, we have found that the key limits result first from the short mean duration of a crime, (1/F), which is then further reduced (divided by R+1) by the rate of crime completion relative to the rate of interception. The last parameter based constraint, which involved M, unlike the other two, was influenced solely by the ratio of the completion rate to the interception rate.

The development of these constraints strongly affected and simplified the construction of algorithms for allocating police patrol. In Chapter 6 we consider the allocation of a tactical patrol force among various high crime areas of a city or precinct. The algorithm we present was made possible by the travel time constraint. It allowed us to consider only solutions in which patrol units do not divide their time (and hence do not have to travel) between two or more of the regions. Chapter 7 presents an algorithm for finding the optimal region for a standard patrol unit to patrol. The algorithm constructs a single contiguous (again the travel constraint) region from

Table 5.3: A List of Three Criteria for When Search Should Be Limited to One Region

$$L > \frac{1}{F \cdot (F/I + 1)}$$

$$X > \frac{M+1}{M-1} \cdot \frac{1}{F \cdot (F/I + 1)}$$

$$M > \left(\frac{(R+1)}{R} \right)^2$$

a series of small regions. For the algorithm to be effective, the regions need to be small so that neither the X, the minimum duration of a visit to the area, nor the F/I is too large. (Remember that I depends also on the size of the area.) If, however, the regions were large in the above sense, there would be little flexibility in constructing the region to patrol since the algorithm would be reduced to finding the atom with the highest payoff. Before concluding this chapter, we will briefly discuss extensions of the model to multiple crime types and more than two regions.

5.6 Extensions

5.6.1 Multiple Crime Types

Of the two extensions to be considered, multiple crime types will be analyzed first because conceptually the extension is simple. In order to include the multiple crime types in the model, it is necessary to add for each crime type (different I and F) a pair of differential equations for each region. Then to calculate the state of the system, S, under a particular search strategy, each of the crimes is initially treated separately. The total average number of crimes in progress is then just the sum of all the individual averages. The resultant expression is, however, still a function of only X or K and X, and can be analyzed in the same manner as before. For example, if N crime types were introduced into the last example, the inequality involving X could easily be generalized from equation (5.76) to be

$$X > \frac{\displaystyle\sum_{i=1}^{N} \frac{A(j)\cdot I(j)^2\cdot(M(j)+1)}{F(j)\cdot(F(j)+I(j))}}{\displaystyle\sum_{i=1}^{N} \frac{A(j)\cdot I(j)\cdot(M(j)-1)}{F(j)^2\cdot(F(j)+I(j))^2}}$$

Lastly in the event that the different crime types are to be weighted differently, this too can be easily incorporated in the search for optimal strategies. This is done by introducing the appropriate weighting factors when adding together the average number of crimes in progress to generate the overall average.

5.6.2 More Than Two Regions

Introducing more than two regions into the problem, significantly increases the complexity of analysis. Although any of the earlier three inequalities can, without difficulty be applied pairwise to all the regions to exclude as many as possible, the complications enter when more than two regions are still left. The heart of the problem lies in the difficulty of defining one single cycle type to be analyzed. In the two region case, in addition to assuming a cyclical search pattern, we made a second reasonable simplification in considering only simple cycles of the form, X minutes in one region followed by X or K•X minutes in the second region and back to the first region again for X minutes. We did not, however, allow a single cycle to contain more than a single visit to each region. Thus cycles of the form, X minutes in RL followed by K•X minutes in RM followed by Y minutes in RL and back to RH for J•Y minutes and then finally back to RL for X minutes, were not considered. With three or more regions, though it is not as easy to justify considering only cycles of the form, visit R1, then R2, then R3 and back to R1 and disregard cycles of the form R1, R2, R1, R3, R2 or some variation on it. The result is that in order to analyze more than two regions, the optimal solutions must be determined separately for each 'reasonable' alternative cycle. In addition as the number of regions increase the number of alternative cycles increases faster than linearly, which only further compounds this complication. Therefore an important primal step in expanding the model in this direction is the development of a methodology for selecting 'reasonable' cycles.

FOOTNOTES 5

1 Although Appendix D proves only that when allocating 50% of your search effort to a region the optimal cycle length is infinitesimally small, an analogous proof can be derived to show that for any fraction of time spent in a region the optimal dispersion of that search is also in infinitesimal cycles.

REFERENCES 5

1. Barnett, A., On Searching for Events of Limited Duration, (Working Paper WP-11-74), Massachusetts Institute of Technology Operations Research Center, Cambridge, MA, September 1974, submitted for publication.

2. Blachman, N. and F. Proschan, "Optimum Search for Objects Having Unknown Arrival Times", Operations Research, Vol. 7, No. 5, 1959.

3. Gilbert, E.N., "Optimal Search Strategies", SIAM Journal, Vol. 7, 1959.

4. Kisi, T., "On an Optimal Searching Schedule", Journal of the Operations Research Society of Japan, Vol. 8, No. 2, February 1966.

5. Matula, D., "A Periodic Optimal Search", American Mathematical Monthly, Vol. 71, 1964.

CHAPTER 6

AN ALGORITHM FOR DEPLOYING A TACTICAL PATROL FORCE

6.0 Introduction

In the preceding chapters we have developed and expanded upon a theoretical foundation of search theory applied to police patrol. Building on that foundation, we will present algorithms for both deploying a tactical patrol force (Chapter 6) and for finding the optimal patrol region for an individual patrol unit (Chapter 7). In general the impact of the earlier work was in the form of simplifying assumptions that were used to determine the structure of the optimal solution. This structure naturally facilitates the search for optimality by reducing the class of solutions that need to be surveyed. Later, in the discussion of each model's assumptions, we will describe these simplifying assumptions, their justification and their specific relationship to the algorithm.

The first algorithm to be presented provides a methodology for deploying a tactical patrol force. Throughout the chapter we will use the term 'tactical patrol force', in a broad sense, to describe a patrol force with the following characteristics: 1)Its major focus is on crime with, at most, very limited responsibility for responding to calls for service; 2)The deployment of the force can be rather flexible (not limited to individual patrol units patrolling separate sectors). Thus, the term is meant to include, for example, the portion of a split patrol force that has been relieved of most of its responsibility for calls for service.

The optimal allocation of a tactical patrol force is formulated here as the following problem:

> Allocate M patrol units among R high crime regions to maximize the weighted probability of intercepting a random crime.

The mathematical programming formulation of this problem is

$$\text{MAX} \sum_{i=1}^{C} W(i) \cdot \sum_{i=1}^{R} F(i,j) \cdot P(i,j,\ N(j))$$

$$\text{s.t.} \qquad \sum_{j=1}^{R} N(j) = M$$

$$N(j) \geq 0 \text{ and integer} \qquad (6.1)$$

The W(i) are the (subjective) weights assigned to each of C crime types which reflect the relative importance of intercepting different types of crimes. F(i,j) is the (relative or absolute) frequency of each crime type i in each region j. The N(j) are the control variables which represent the number of patrol units allocated to region j. Lastly, P(i,j, N(j)) is the probability of intercepting crime type i in region j when N(j) units are patrolling the region. The functional form of P(i,j, N(j)) will be discussed in later sections. For now we limit ourselves to noting that it is, in addition, a function of these input parameters: the street mileage of the region, the speed of the patrol car, and the duration and observability of the individual crime types. In our later discussion of these parameters, we will pay special attention to the issues and problems that arise in obtaining estimates of the last two parameters, duration and observability. One point that will be emphasized is the critical need for research concerning these aspects of crimes. Good estimates are essential both in the development of effective patrol strategies and in obtaining a more accurate picture of the potential impact of police patrol on street crime.

The presentation of the optimal allocation algorithm will proceed in the following sequence. First its underlying assumptions and data requirements will be spelled out. Next the individual steps of the algorithm will be described in conjunction with a discussion of calculating P(i,j, N(j)). Finally, closing out the discussion of the basic algorithm will be an illustrative application of the model to a problem involving the allocation of ten patrol units to five prototype high crime areas. The second half of the chapter will focus on algorithms for carrying out sensitivity analysis on each of the input parameters. It will be clear from our discussions that we consider this sensitivity analysis capability not an added frill but an essential part of the methodology for deploying the patrol force. It is not hard to see why we have developed such an attitude, given the subjectivity of some of the input parameters (e.g. weights), changeability of others (e.g. F(i,j), the frequency of different crimes in each region), and the difficulty of 'accurately' estimating the crime type related parameters.

6.1 Model's Assumptions

A detailed step by step description of the algorithm will be presented in section 6.3. However, in order to place each of the underlying assumptions in perspective, it will be necessary at least to sketch the various components of the algorithm. The most fundamental component of the algorithm is the expression for calculating $P(i,j, N(j))$, the probability of intercepting a crime of type i in region j if $N(j)$ patrol units are assigned to that region. The expression used here is essentially the same as the one introduced earlier in Chapter 4 (equation (4.2)) for calculating the probability of intercepting a crime when n units are on patrol. Consequently the discussion in Chapter 4 of that equation's assumptions is equally applicable here. The most crucial assumption was the independence between the location of a crime within a region and the location of a patrol unit in the same region. This assumption is most justifiable if the tactical patrol force consists of plainclothes policemen traveling in unmarked cars. For a more visible patrol force (e.g. part of a split patrol force), this assumption may not be valid. However until information is available about the magnitude of the dependence, it is impossible to assess how inaccurate the model's estimates of interception probabilities are. Of course any implementation of the model should keep this issue in mind, recognizing that the model may be overestimating the interception probabilities.

A second, related assumption is that the frequency of crimes in the different regions is not <u>immediately</u> affected by the deployment of the patrol force. The key word is 'immediately'. We sould expect criminals to <u>eventually</u> shift their activities away from regions with a high risk of interception. However, if this process is not too rapid, the algorithm can be rerun as the changing crime patterns begin to emerge. As soon as one or two criminals are caught in a particular region, it may be necessary to rerun the algorithm (or use sensitivity analysis) to find a new optimal allocation. The new input data might even anticipate the disappearance of crimes of a particular class from the region in which the criminals had just been caught. As we shall see the algorithm (and its sensitivity analysis) is simple and fast enough to allow for even on line usage to update the patrol allocation. Largely, the speed and frequency with which the patrol force can be redeployed to react quickly to, or even anticipate, crime pattern changes is limited more by command and control issues than by any difficulties in rerunning the algorithm. If however criminals react so quickly that the present optimal deployment plan is outdated even before any criminals are caught, the algorithm to be described may be of little value. The only alternative under those circumstances is to take a game theory approach to the problem.

There is one assumption of the original expression (4.2) that will be relaxed. The algorithm does allow for the specification of a probability distribution for the observable duration of a crime. On the other hand, one potentially important aspect of police patrol that is not captured by the algorithm nor by any of the search theory models is the possible crime preventive component of patrol.

The second major component of the algorithm maps out the route followed while approaching the optimal allocation of all n patrol units. It starts out by determining the best region in which to place a single patrol unit. Having optimally allocated the first unit, it then determines the best location for the second unit. It proceeds iteratively until all n patrol units are deployed. The key simplification that lies at the heart of this procedure is that at each step we allocate the patrol time of each additional patrol unit to one and only one region. We do not consider dividing that time between one or more regions which would require the unit to travel back and forth between the regions. The motivation for considering only integer allocations to each region derives from our analysis (Chapter 5) of the two region problem in which time was lost (from search) while travelling between the regions. Our analysis there showed that even for identical regions the time lost travelling would usually outweigh the benefits that might accrue from the transfer. This assumption immediately reduces the number of possible solutions from a non-countable infinite set to a relatively small finite set. The algorithm described later in detail, in essence, is just a procedure for searching the finite set for an optimal solution without enumerating all possible solutions.

There is one important assumption that we have not made. We have not assumed that the best way to deploy a tactical patrol force is for the units to patrol randomly. It may well be that stakeouts are more effective against burglaries and decoys more effective against muggings. The goal of this chapter is to develop an algorithm to deploy optimally a randomly patrolling tactical patrol force. Once that is achieved, it is then possible to compare the relative effectiveness of two completely different strategies (e.g. random patrol versus stakeouts). Without this capability, however, a comparison of the two strategies would be like comparing for sweetness a ripe Mc Intosh apple with an unripe Delicous apple.

6.2 Input Data

An obvious prerequisite for using a model in a real environment is that it be possible to generate the model's input parameters from existing data. However, there are instances where the development of a model antecedes the existence of usable data. Under those circumstances the creation of a model may prove a catalyst in the development of the needed data by focusing on which data is needed and why that data is critical. It is hoped that the search theoretic models presented here provide such a stimulus to the development of an accurate street crime data base. As we shall see, the data needs of this model are not really model specific in that any attempt at increasing the efficiency with which police intercept street crimes requires that the same data be collected or generated. With this in mind we proceed with the discussion of the data.

6.2.1 Crime Weights

Early on in the development of the model, it was recognized that different values may be attached to intercepting different crimes. This facet of the problem was incorporated into the model by allowing for different weights to be associated with intercepting each of the various types of crimes. These weights may be city or precinct specific reflecting the subjective assessment of only the local decision maker. In that case the decision maker would, of course, be required to go through the process of quantifying his subjective feelings. This process would be aided by techniques in the field of statistical decision theory, for assessing utility functions. The source of these weights can also be based on a broader theoretical and empirical foundation. Maltz, in a recent article [2], suggests a number of measures of effectiveness for crime reduction programs which could be used to determine the weights. One existing measure is the Crime Seriousness Index (CSI) developed by Sellin and Wolfgang [5]. Questionnaires were administered to various groups affiliated with the criminal justice system in order to determine how the groups assessed the seriousness of each crime relative to a standard crime. The weights were found to be fairly consistent from group to group. Maltz suggests other measures which are multidimensional extensions of CSI. The use of these vector measures in the search thoretic framework presented here will require first the application of multiattribute utility theory to produce a single objective function.

An alternative to the somewhat subjective measures offered by Sellin and Wolfgang and by Maltz would be measures based on the existing criminal justice code. The criminal code has, in a sense, already quantified society's attitudes towards the various crimes by assessing different penalties (prison terms) for the commission of different crimes. A variation on the same theme would be to weight crimes according to the actual prison terms handed down by judges as a measure of the relative seriousness of each crime type.

Another less subjective measure, that can be used alone or in conjunction with other measures similar to CSI, attempts to capture the impact on the crime rate of intercepting a particular criminal. This measure is the relative rate at which criminals involved in a particular class of crimes commit crimes of that type. For example if a burglar carries out 5 burglaries per month and an auto thief steals 10 cars per month, auto theft would be assigned a weight twice that of burglary. As was noted earlier this measure need not, however, be used alone but can also be multiplied by the associated value of the CSI for that class of crimes. With this type of weights, the objective of the allocation model would be to maximize the longer range weighted impact on crime levels. The concept of weighted impact suggested here has broader applications. This concept can be applied to measure any crime reduction program that focuses on the criminal rather than on the crime. Thus recently suggested mandatory sentences for certain specific types of crimes can easily be evaluated using the above measure. Another program already being implemented in different parts of the country is an emphasis in the courts on bringing to speedy trial criminals charged with certain types of crimes. This measure can be used to optimally allocate resources among competing crime types in order to maximize the program's weighted impact on crime.

6.2.2 Crime Frequency

The second parameter to be discussed is the relative frequency of each crime type in each region. In our example we used data from the published FBI Uniform Crime Reports. One oft mentioned problem with crime data from this source or police records is that it seriously underestimates the actual number of crimes. From our perspective, however, there is an even more serious problem with available police data especially since the introduction of victimization studies have begun to determine levels of unreported crime. The problem is that the typical data often does not distinguish between a personal robbery committed on the street or in the hallway of the tenth story of a high rise apartment house. No distinction is made between a rape in which the initial encounter was on the street and one which occurs through a rapist gaining entrance to the victim's residence. Similarly, the data may not distinguish between a burglary which occurs in a location where a patrol car has access and a burglary where a patrolman would need telescopic, x-ray vision in order to notice anything unusual. In short if the decision maker is choosing between competing patrol strategies and allocations, he needs complete information about what fraction of crimes in each category can possibly be affected (intercepted) through street patrols.

The changing nature ofthis parameter also requires that the crime data be continuously tracked and updated. The algorithms for sensitivity analysis can help this tracking process. They point out how sensitive the optimal solution is to changes in the frequency of the various crime types. This allows the major part of the ongoing data analysis to be concentrated in a limited number of critical (with regard to the optimal solution) crime categories.

6.2.3 Duration and Observability of Crimes

The last set of data to discuss and the most difficult to obtain is the crime descriptive data. Police personnel are certainly aware that the chances of a patrol car passing an auto theft in progress are very slim because the crime is so short. Similarly they recognize that the likelihood of a passing patrol officer spotting a burglary in progress is very small. Yet ask them how long an auto theft takes or how observable a burglary is and you realize that these questions have never been asked even though their knowledge and information gained through experience is sufficient to at least partially answer these questions. Even departments that have begun to review their crime data to determine what fraction of crimes are observalbe have not asked to what degree and for how long can the observable crimes be seen. The succeeding paragraphs will discuss, in a somewhat anecdotal style, how we obtained the crime descriptive data used in the examples. However in the application of the model presented later, we will suppress all names of crime types. The reason for doing this is that the numbers we used do not have a very solid foundation and are the result of a very limited number of discussions with police personnel. Our goal in these discussions was not to generate a good data base (which is a massive undertaking) but rather to make sure that the hypothetical examples we constructed do not use <u>absurd</u> values for observability and duration. Thus to avoid any unjustified inferences being drawn from our later examples, the names have been changed to protect the innocent. Instead of the more eye catching terms such as rape, manslaughter and commerical burglary , we have generally substituted the innocuous terms Type I, II and III crimes. However, our discussion here will use some standard names in order to illustrate one way to approach the data collection task. In addition we will attempt to generalize from our limited experience about the possible problems to be faced in generating the data.

The initial stumbling block, and by no means unique to police, were the patrolmen's uneasiness with attempting to define a single number for the average duration. Instead they often referred to those rare instances when crimes took an unusually long time (e.g. the rape in which the woman was held captive for ten hours). This problem of outliers was easy to overcome mainly because our primary concern was the <u>observable</u> duration. Therefore, even in the extremely long crimes, typically, only during a small fraction of the time could a passing patrol officer have possibly noticed something suspicious. As the discussions continued it seemed that patrolmen were little more comfortable with assigning a small range to the average observable duration. One estimate for a commercial burglary was between 5 and 10 minutes. For our examples an estimate of 7.5 minutes was used. However, in future discussions it would be worthwhile to explore the range in more detail asking if 5 minute crimes were more frequent than 10 minute ones and if so, how much more, etc., in order to derive a functional form for the probability distribution of the observable duration. An important tool used in assessing the observable duration and even more so the observability of a crime, was the constant comparisons between crime categories in order to be assured of consistent estimates.

The determination of the observability of a crime proved a much more difficult task. The first barrier was the need for the officers to comprehend and internalize the concept of conditional probability. The reason for this is that observability parameters describe the probability of noticing (intercepting) a crime conditioned on passing it while it is in progress. This problem was compounded by the fact that there is not much in a policeman's past experience that he can call on to estimate the average observability. How frequently does a patrolman pass a crime in progress? The result was that some of the estimates may be low since the conditional probability of seeing a crime seemed to be, at times, confused with the probability of passing a crime in progress and seeing it, a probability which is very small. The direction that, however, seemed most useful in assessing the observability of a crime was to try to build a profile of a typical crime from each category and then analyze each component of the crime separately. For example a profile of a potentially interceptible rape might be described as follows:

> In the initial encounter between the man and woman there might be a struggle lasting for approximately a minute which a passing patrol car would have a better than 50-50 chance of spotting (set at .6). After that the woman would then be forced to a more secluded location nearby in which the actual rape occurred. During that time, about 5 minutes, there is almost no chance for a passing patrol unit to spot anything. This probability was assessed at .02. After the completion of the rape the rapist might run (20 seconds) which would attract attention with a probability of .5. Thus the average observability during the total crime was about .14.

The same was done for other crimes. Another example is personal burglary:

> During the commission of the crime, which includes the breaking and entering, there is almost no chance of spotting the burglar (assessed at .01). However, when leaving the premises with stolen property (last half minute of a ten minute crime) there is a 50-50 chance of a passing patrol officer being suspicious.

These numbers, as the officers pointed out, will often depend on when the crime was committed. At 2:00 A.M., in the morning, a man lugging a stereo or television is likely to attract more attention than at 2:00 P.M., in the afternoon. For our examples we disregarded all such variations.

In determining the observability of a crime the statistic that was calculated was the average over the duration of a crime. The average observability is a sufficient statistic because each region is allocated its own group of patrol units and those patrol units are assumed to be on patrol in their respective regions throughout the course of the crime. If, however, patrol units were leaving and entering the region in which the crime occurs, then it would be important to know during exactly which stages of the crime it is most observable. Moore [3] addresses the problem of detecting targets whose observability changes over time and as can be seen from his analyses it is a very complicated problem to analyze.

In the above discussions we have attempted to sketch some of the issues and problems that arose in generating our sample data. The problems addressed here are certainly crucial in effectively deploying a patrol force to attack street crimes; yet no significant research to date has been done in this area. Although we recognize the inherent problems in determining these parameters, we also realize that these are the numbers that are needed.

6.3 An Algorithm to Allocate a Tactical Patrol Force

The algorithm to allocate a tactical patrol force is unabashedly simple and involves but two stages. During the initial stage each region is analyzed separately to determine the payoff (weighted probability of intercepting a random crime) resulting from introducing a single patrol unit into that region. The regions are then ranked in order of their potential payoff from a single patrol unit, with the region ranked highest allocated the first patrol unit.

Once the first unit has been allocated, the next stage, requiring even fewer calculations than the first, iteratively allocates the additional (N-1) units. Because it is assumed that adding a patrol unit to the region, RH, does not affect the probability of intercepting a crime in any of the other regions, the payoff from adding a unit to any of the other regions has not changed. Therefore only in region RH does the incremental payoff of adding a patrol unit change. As a result it is necessary to evaluate for that region the following expression

$$\sum_{i=1}^{C} W(i) \cdot F(i,j) \cdot (P(i,j,2) - P(i,j,1)) \quad \text{for } j=RH \qquad (6.2)$$

Then using the just updated incremental payoff for RH, its rank relative to the other regions is also updated. The region now ranked first with regard to the marginal payoff of an additional patrol unit is allocated the second patrol unit. Afterwards the updating process is repeated for that region. Thus we cycle again through the entire procedure described above until all N patrol units are allocated. The general form of expression (6.2) is

$$\sum_{N=1}^{C} W(i)\cdot F(i,j)[P(i,j,N(j)+1) - P(i,j,N(j))] \quad \text{for } j=RH \qquad (6.3)$$

with N(j) the number of patrol units allocated at present to the region RH. The second stage is a marginal allocation procedure which can be briefly summarized as

1. Allocate the k^{th} patrol unit to the region ranked first on the incremental payoff list.
2. Update that region's incremental payoff.
3. Update its rank on the incremental payoff list.
4. Return to step 1.

An important feature of the steepest ascent algorithm just presented is that in the process of determining the optimal allocation of N patrol units, it also finds the optimal allocation of any number of patrol units less than N. In all instances the solution is a global optimum (See section 6.4 for a discussion.).

In the above algorithm, an obviously central calculation is the determination of P(i,j,N(j)), the probability of intercepting a crime of type i in region j with N(j) patrol units in that region. The basic formula to be used in the calculation was introduced previously in Chapter 4. It assumes that the units are randomly patrolling their assigned region and is written

$$1-\exp(-N(j)\cdot S\cdot T(i)\cdot OB(i)/M(j)) \qquad (6.4)$$

where

N(j)= number of patrol units in region j

S = patrol speed

T(i)= observable duration of crime type i

OB(i)= average (over the observable duration of the crime) observability of crime type i

M(j)= number of street miles in region j.

The above equation, though, calculates only the probability of intercepting a crime of fixed duration, T(i). However, in the developement of the algorithm it was recognized that all crimes of a single type are obviously not clocked by a special timer with a 24 second buzzer signalling the crime to end. Consequently, the programmed version of the algorithm allows the user to specify any one of four distributions, (1)deterministic, (2)exponential, (3)k^{th} order Erlang, and (4)uniform, for the duration of any one or all of the crime types. Even the above four options do not really reflect the full flexibility of the algorithm. It is easy to include any distribution for which it is possible to evaluate either

$$1 - \int_0^{\infty} f(t) \cdot \exp(-N \cdot S \cdot t \cdot OB/M) dt \qquad (6.5)$$

for a probability distribution function, f(t) or

$$1 - \sum_{t=0}^{\infty} p(t) \cdot \exp(-N \cdot S \cdot t \cdot OB/M) \qquad (6.6)$$

for a probability mass function, p(t). It is even envisioned that in order to describe accurately the observable duration of a crime, it may be necessary to divide the total duration into two or more components, each with its own probability distribution. For example, from an operational perspective, the commission of a crime can often be divided into three stages:

(1) The initial encounter between criminal and victim (target) including perhaps a struggle (breaking in);

(2) Time during which the criminal actually obtains his desired goal;

(3) The criminal's hasty departure from the scene of the crime with the goods.

The degree of variation in the duration of each of these stages may not be the same, requiring therefore three distributions (assumed independent) to describe the crime. Expression (6.5) could then be modified to be

$$1 - [\int_0^{\infty} f(t_1) \cdot (\exp(-N \cdot S \cdot t_1 \cdot OB(1)/M))\, dt_1$$
$$\cdot \int_0^{\infty} g(t_2) \cdot (\exp(-N \cdot S \cdot t_2 \cdot OB(2)/M))\, dt_2$$
$$\cdot \int_0^{\infty} h(t_3) \cdot (\exp(-N \cdot S \cdot t_3 \cdot OB(3)/M))\, dt_3] \qquad (6.7)$$

Notice that the average observability is also allowed to vary with each stage. The three integrals in the expression are multiplied together because each calculates separately the probability of not intercepting the crime during their respective stages. The probability of not intercepting the crime at all is therefore just the product of the three. It is assumed for now that the duration of each stage is independent of the other two stages.

Prior to programming the present algorithm, with its flexibility of also handling three non-deterministic distributions, expression (6,5) was first evaluated for each of the distributions. For the exponential distribution with mean (1/A) it is

$$1 - \int_0^\infty A \cdot \exp(-A \cdot t) \cdot \exp(-N \cdot S \cdot t \cdot OB/M)\, dt$$

$$1 - \frac{A}{A + (N \cdot S \cdot OB/M)} = \frac{N \cdot S \cdot OB}{M \cdot A + N \cdot S \cdot OB} \qquad (6.8)$$

for an Erlang of order k it is

$$1 - \int_0^\infty \frac{A}{(k)} (A \cdot t)^{k-1} \cdot \exp(-A \cdot t) \cdot \exp(-S \cdot t \cdot N \cdot OB/M)\, dt$$

$$= 1 - \left(\frac{A}{A + (S \cdot N \cdot OB/M)}\right)^k \qquad (6.9)$$

and lastly for the uniform distribution of [A,C] it is

$$1 - \int_A^C \frac{1}{C \cdot A} \exp(-S \cdot t \cdot N \cdot OB/M)\, dt$$

$$= 1 - \frac{M}{(C-A) \cdot S \cdot N \cdot OB} \left[\exp(-S \cdot N \cdot OB \cdot A/M) - \exp(-S \cdot N \cdot OB \cdot C/M) \right] \quad (6.10)$$

The above discussion of alternative distribution functions for each crime type leads to an obvious question, "How much difference does it make whether we use a constant for the duration time or an exponential distribution (with the same mean) for it?". In the next section we will present two examples of applying the algorithm. The first example assumes constant duration; the other, exponential.

6.4 Allocating Ten Patrol Units to Five Regions

6.4.1 Equal Total Crime Rates: Different Distribution of Crime Types

The algorithm requires two sets of data, one to describe each different crime type, the other to describe each region. It is necessary to specify for street mileage and the frequency of each crime type for each region. To simplify the analyses in our example all regions will contain 15 miles of streets. In the first example, in order to focus on how even the distribution of crime types, alone, can affect the optimal solution, the total crime rates were set so as not to differ from region to region. However, the distributions of crimes by type within each region do differ. In order to generate realistic distributions, we used National Crime Panel Surveys data [4] for the five largest cities (Chicago, Detroit, Los Angeles, New York and Philadelphia). (The appropriateness of this source of data for the model was discussed earlier in the chapter.) Focusing on six potentially detectable crime types (e.g. personal robbery, commercial robbery, auto theft, personal burglary, etc.), we determined for each city the fraction of crimes that fell within each of the six classes. Thus, in the summary of the region input data for the first example that is contained in Table 6.1, the distribution of crimes by type in each region is the same as that found in one of the five cities.

For the crime type descriptive data, we first assigned a weight of one to each crime so that in effect we were maximizing the (non-weighted) probability of intercepting a crime. The data describing the average duration and average observability of each crime type is summarized in Table 6.2. The generation of this data was described in the previous section. As for the probability distribution of the observable duration of a crime, two separate examples are presented in order to display how changing the distribution can affect the optimal solution. One example assumes a deterministic distribution (i.e. constant duration); the other, an exponential distribution with the same mean. Our choice of distributions is not, however, meant to imply that either distribution duplicates the real world although the exponential is likely to be closer to reality.

6.4.2 Optimal Solution: Crime Durations Assumed Constant

The most striking aspect of the optimal allocation (described in Tables 6.3 and 6.4) is the wide disparity in the allocation of units, Region A was allocated six and C none, even though the total crime rates in both regions were the same. Even more striking, perhaps, was that the first five units and six of the first seven units were assigned to A (Table 6.3) with Region B allocated one of the seven. In order to comprehend why A received the largest number of units, better described as a whale's share, we will first look at Table 6.4, which displays the probability of intercepting each of the crimes in each region.

REGION DATA

Regions	Miles	Frequency of Various Crime Types I	II	III	IV	V	VI
A	15	0.022	0.202	0.142	0.465	0.136	0.033
B	15	0.015	0.161	0.132	0.468	0.174	0.050
C	15	0.016	0.104	0.153	0.530	0.171	0.026
D	15	0.009	0.184	0.105	0.276	0.324	0.102
E	15	0.010	0.194	0.150	0.388	0.199	0.059

Table 6.1: Region Descriptive Data:Frequency of Various Crime Types in each Region

CRIME DATA

Crime Type	Weight	Distribution	Mean	Average Observability
I	1.0	Deterministic	0.100*	0.140
II	1.0	Deterministic	0.066	0.100
III	1.0	Deterministic	0.037	0.100
IV	1.0	Deterministic	0.167	0.030
V	1.0	Deterministic	0.125	0.040
VI	1.0	Deterministic	0.066	0.060

Table 6.2: Crime Descriptive Data:Weight, Mean, Distribution, and Average Observability of each Crime Type

* The mean is in hours not minutes.

The first point that stands out is that crime type I has the highest probability of interception in the five regions as a whole, .034, and similarly in each individual region. The region with the highest interception probability, .081, is region A, as would be expected. A distant second, but of greater significance because of its higher frequency, was crime type II. It had an overall interception probability of .015 far below that of Type I. This was still more than one-third higher than its next nearest competitor, crime type IV which had an interception probability of .011. The cause of the higher interception probabilities for type I and type II crimes obviously lies in the original data describing the duration and observability of the crimes (Table 6.2).

In the data we used, type I crimes had an observable duration of six minutes and a patrol car passing during that time had, on the average, a one in seven chance of noticing something suspicious. This average observability is significantly higher than for any of the other crimes, especially those crimes of longer observable duration. In addition its own observable duration is the third highest. Both factors combined generate an interception probability over twice that for type II crimes.

Comparing types II and IV, we see that although the latter has a duration two and a half times the former, its average observability is less than one-third that of type II.

This explains why the probability of intercepting type II crimes is 33% higher than that for type IV. Combining the above with the fact that Region C, which receives no units, generates a larger fraction of the type IV crimes than it does type II crimes, explains why type II crimes have an overall 46% higher (.0148 versus .0102) probability of being intercepted.

Turning back to the region data, Table 6.1, we can now understand why patrol is concentrated in region A. Both type I and II crimes, which have the highest interception rates, are a larger proportion of the crimes in that region than in any other region. While in region C (no patrol units) these crime types, together, make up only 12% of its crimes, the lowest for any region.

The above discussion explains why Region A is assigned the most patrol units and C the least. It does not, however, fully justify why so many are allocated to the former and none to the latter and so few to all the other regions. In order to understand this phenomenon, we must first realize that the algorithm is not using the following logic: Assign to Region A proportionately more patrol units as it has a higher proportion of more interceptible crimes, and consequently a random crime occurring there has a higher chance of being intercepted by a patrol unit. Using the concept of diminishing return presented in Chapter 4, it is possible, though, to explain the lopsided optimal solution.

SEQUENTIAL ALLOCATION

Patrol Unit Number	1	2	3	4	5	6	7	8	9	10
Allocated to Region	A	A	A	A	A	B	A	E	B	D

Table 6.3: Sequential Allocation:The Order in Which Patrol Units areAllocated to the Various Regions Under the Optimal Policy: Equal Total Crime Rates-Duration of Crimes Assumed Constant

RESULTS OF OPTIMAL ALLOCATION

Regions	Cars Alloc.	Probability of Interception of Crimes I	II	III	IV	V	VI	Average
A	6	.081	.039	.022	.030	.030	.024	.0314
B	2	.028	.013	.007	.010	.010	.008	.0102
C	0	.000	.000	.000	.000	.000	.000	.0000
D	1	.014	.007	.004	.005	.005	.004	.0051
E	1	.014	.007	.004	.005	.005	.004	.0051
Average		.034	.015	.007	.010	.008	.007	.0103

Table 6.4: Probability of Interception of each Crime Type in each Region under the Optimal Allocation Strategy: Equal Total Crime Rates-Duration of Crimes Assumed Constant

The initial payoff for introducing a single patrol unit into A was only about 3% higher than that for B. However, as more units are allocated, none are assigned to region B until the incremental payoff from adding another unit to region A has been reduced, by diminishing returns, below that of an initial payoff in the other region (B). Because the diminishing return is occurring so slowly, cutting into the incremental payoff by only slightly more than one-half of one percent for each additional unit assigned to A, it is not until the sixth unit is to be allocated that diminishing return has erased what was initially only a 3% advantage. The same analysis, of course, can be used to explain why region C is not allocated patrol units even though the initial difference between it and A was only 4%.

6.4.3 Equal Total Crime Rates--Duration Assumed to be Exponential

In the next example the input data (region and crime type) to the model is the same as before except that the observable duration of each crime type is assumed to follow an exponential probability distribution with the mean the same as before (Table 6.2). When the algorithm was run with this slightly modified data, the optimal solution (Table 6.6) significantly decreased region A's allocation, reducing it from six to four. The two cars were reallocated, one each to regions D and E, which were now allocated two cars apiece. An analysis of the rate of diminishing return shows clearly how exponentiality affected the optimal solution.

In both examples the initial difference in payoff between a unit assigned to A and one assigned to B was only 2.5%. For the constant duration, it was noted earlier that adding a patrol unit decreases the incremental payoff by only slightly more than one-half of one percent. However, introducing exponential distributions increases the rate of diminishing return to almost 1.2%. Thus by the time the fourth unit is allocated here (see Table 6.5), the first in line to receive that patrol unit is no longer region A; instead it is region B. In the first example, B did not receive any patrol units until the sixth one was allocated (Table 6.3). This increased rate of diminishing return also affects region C. Sensitivity analysis carried out on the optimal solution showed that if an additional unit (the eleventh) became available, it would be allocated to region C if the crime duration distribution were exponential but to region A if the distribution were deterministic.

There is one last ancillary effect that should be noted. Because each region's crime, when broken down into the six categories, is different, shifting units between regions affects the overall probability of intercepting crimes of a particular type. Region A has a higher proportion of type I crimes while in D and E criminals prefer crime types V and VI. As a result, shifting the units out of A decreases the

SEQUENTIAL ALLOCATION

Patrol Unit Number	1	2	3	4	5	6	7	8	9	10
Allocated to Region	A	A	A	B	E	D	A	B	E	D

Table 6.5: Sequential Allocation:The Order in Which Patrol Units are Allocated to the Various Regions Under the Optimal Policy: Equal Total Crime Rates-Duration of Crimes Assumed Exponential

RESULTS OF OPTIMAL ALLOCATION

Regions	Cars Alloc.	Probability of Interception of Crimes I	II	III	IV	V	VI	Average
A	4	.053	.026	.015	.020	.020	.016	.0208
B	2	.027	.013	.007	.010	.010	.008	.0102
C	0	.000	.000	.000	.000	.000	.000	.0000
D	2	.027	.013	.007	.010	.010	.008	.0102
E	2	.027	.013	.007	.010	.010	.008	.0102
Average		.029	.015	.007	.010	.010	.008	.0103

Table 6.6: Probability of Interception of each Crime Type in each Region under the Optimal Allocation Strategy: Equal Total Crime Rates-Duration of Crimes Assumed Exponential

probability of intercepting a type I crime by 17% (from .034 to .029) and increases the probability of intercepting a type V crime by 13% (from .0083 to .0096) and a type VI crime by 16% (from .0068 to .008). Results analogous to the above will appear in the next example when we consider regions with differing total crime rates.

6.4.4 Differing Total Crime Rate: Duration Assumed Exponential

Previously we analyzed the impact of the differing crime categories and the functional form of the observable duration. In this next example we explore how differing overall crime rates compound the above effects and, more importantly, address the question of what gains are associated with the optimal allocation. An obvious prerequisite for measuring the gains generated by an optimal deployment is the existence of a standard with which to compare. The straw man to be used is the linear model which allocates patrol units to each region in direct proportion to its total crime rate.

The crime type descriptive data used here is the same as in the previous example including the assumption that the observable duration is exponential (which seems to be more realistic than a deterministic distribution). However, the region data was modified in the following manner. The frequency of each crime type in regions A and C was tripled so that both regions had total rates of three crimes per unit time. Region B's total crime rate was doubled in the same manner while the crime rates in regions D and E remained unchanged. A linear allocation model would allocate three patrol units each to A and C, two to B and one each to D and E. With this allocation a random crime would be intercepted with a probability of .0123 or approximately one chance in 80. The interception probabilities of the different crime types range from a low of .008 for type VI crimes to a high of .035 for type I crimes and are summarized in Table 6.8.

The optimal allocation algorithm was run on this problem and it suggested that seven patrol units be assigned to region A, three to B and none to the other three regions. The imbalance between A and C which have the same overall crime rate, was due to region A generating a higher frequency of crimes with higher interception rates (crime types I and II). However, with the optimal allocation, the probability of intercepting a random crime increased 22% from .0123 to .0152. The increases in the individual crime categories, though, varied widely. For crime types V and VI, the increases were less than 10% and for types I and II, they were greater than 30%. The determining factor in the size of the increase in a particular crime category was the distribution of that crime type among the regions. Regions B, D and E (allocated no patrol units) generate a high proportion of type V and type VI crimes.

To add another dimension to the comparison, we have introduced another statistic which is the inverse of the

SEQUENTIAL ALLOCATION

Patrol Unit Number	1	2	3	4	5	6	7	8	9	10
Allocated to Region	A	A	A	A	C	A	C	A	C	A

Table 6.7: Sequential Allocation: The Order in Which Patrol Units are Allocated to the Various Regions Under the Optimal Policy: Different Total Crime Rates-Duration of Crimes Assumed Exponential

RESULTS OF OPTIMAL ALLOCATION

Regions	Cars Alloc.	Probability of Interception of Crimes I	II	III	IV	V	VI	Average
A	7	.089	.045	.026	.034	.034	.027	.0355
B	0	.000	.000	.000	.000	.000	.000	.0000
C	3	.040	.020	.011	.015	.015	.012	.0151
D	0	.000	.000	.000	.000	.000	.000	.0000
E	0	.000	.000	.000	.000	.000	.000	.0000
Average		.048	.021	.011	.015	.012	.008	.0152
Linear Model		.035	.015	.009	.012	.011	.008	.0123

Table 6.8: Probability of Interception of each Crime Type in each Region under the Optimal Allocation Strategy: Compared to Linear Allocation Model: Different Total Crime Rates-Duration Assumed Exponential

probability of interception. This statistic is the expected number of crimes committed until one is intercepted. It can be interpreted in terms of interdicting a career path of a criminal. Thus with the linear allocation, a non-discriminating criminal (commits all types of crimes) would, on the average, commit 81 crimes (1/.0123) before being intercepted by a passing patrol unit. Under the optimal allocation this is reduced to 61 crimes (1/.0152). Similarly a criminal who specializes in type II crimes would have his expected career length reduced 36%. These improvements suggest that for regions with disparate crime rates, the algorithm can significantly (percentage-wise) increase the probability of interception. The increases generated in this last example are perhaps more impressive when compared to the goals of a one million dollar crime reduction proposal submitted to LEAA [1]. In it Atlanta's police department set a goal of increasing on-site apprehensions by 5%, with the term on-site apprehensions including all criminals captured within an hour of the crime. A better deployment of a tactical patrol force might by itself generate that level of improvement.

At this point, however, we would like to reiterate our earlier remarks. Even with the improvements (over proportional allocation) that the model generates, the resultant allocation of randomly patrolling units may not be the best strategy. The probabilities of interception are still only on the order of .015. Other totally different strategies, such as stakeouts, may be able to generate higher probabilities especially when the focus is on only one crime type (e.g. burglary).

6.5 Sensitivity Analysis

The following sections present algorithms which perform sensitivity analysis on each of the allocation model's input parameters. The development of each algorithm revolves around understanding how changes in a particular parameter affect the following necessary and sufficient condition of optimality:

"If there exists no region such that adding a patrol unit to that region increases the weighted probability of intercepting a crime, more than removing a patrol unit from some region reduces the same objective function, the solution is optimal.

This optimality condition can be formalized:

$$\max_{j} \sum_{i=1}^{C} W(i)\cdot F(i,j)\cdot[P(i,j,N(j)+1) - P(i,j,N(j))]$$

$$\leq \min_{k} \sum_{i=1}^{C} W(i)\cdot F(i,k)\cdot[P(i,k,N(k)) - P(i,k,N(k)-1)] \qquad (6.11)$$

The left hand side of the inequality represents the increase produced by adding a patrol unit to region, j, and the right hand side describes the decrease resulting from removing a patrol unit from region, k. If the above inequality does not hold, there must exist a pair of regions, j and k, such that adding a patrol unit to j has a higher incremental payoff than the decrease produced by removing a unit from region k. In which case, the present solution can be improved upon by switching one unit from region k to region j. This proves that the above condition is necessary for optimality. The sufficiency of this condition is a direct consequence of diminishing return. If transferring one unit does not improve upon the present solution, transferring more than one certainly can not. Because of diminishing return the second patrol unit that is added to a region increases the interception probabilities there less than the first; while conversely, the second unit removed from a region decreases the probability of interception even more than the first.

6.5.1 Frequency of the Different Crime Types in Each Region

The first algorithm analyzes the optimal allocation with respect to the frequency of each crime type in each region. The need for sensitivity analysis on this parameter is a natural consequence of its changeability. Crime patterns change over a period of time in regard to both the absolute and relative frequency of each crime type. In addition since the allocation model suggests concentrating the patrol force in a limited number of regions, criminal reaction to police deployment may speed up the process of change significantly. The goal of sensitivity analysis on this set of parameters will be twofold: (1)To determine how much the frequency of a particular crime type in each region can vary before affecting the optimality of the present deployment; (2)To pinpoint the crime type-region pairs to which the optimal solution is most sensitive. The decision maker would use this information in determining which crime rates to monitor most closely for changes.

Compared to the other parameters, sensitivity analysis on the crime type frequencies is easy to carry out because increasing the frequency of a crime in a particular region, j, affects the incremental payoff in only that region. The determination of when the optimal solution changes, therefore, requires making only one comparison. That comparison will be between region j and the region (other than j) with the minimum decrement. As F(i1,j), the frequency of crime type i1 in region j, increases, the incremental payoff of region j increases linearly. (The slope of the line W(i1)·[P(i1,j,N(j)+1 - P(i1,j,N(j))].) Thus to find the upper limit, UF(i1,j), to which F(i1,j) can rise without altering the optimal solution, it is necessary to determine the intersection point of two straight lines. One line is the incremental payoff equation (a function of F(i1,j)) of the region j; the other is a constant representing the minimum decrement. The intersection point is

$$UF(i1,j)= \frac{\left[\left(\min_{k,k\neq j} \sum_{i=1}^{C} W(i)\cdot F(i,k)\cdot(P(i,k,N(k))-P(i,k,N(k)-1)\right) - \sum_{i=1,\ i\neq i1}^{C} W(i)\cdot F(i,j)\cdot(P(i,j,N(j)+1)-P(i,j,N(j)))\right]}{[W(i1)\cdot(P(i1,j,N(j)+1)-P(i,j,N(j)))]} \qquad (6.12)$$

If the F(i1,j) increases above this limit, UF(i1,j), the the present optimal solution can be improved by transferring a unit into region j from the region with the present minimum decrement.

The calculation of the lower limit on F(i1,j) is directly analogous to the above. Instead of Equation (6.7) the relevant expression for determining the two line intersection point is

$$LF(i1,j) = \text{Max}\left[0,\ \frac{\left[\underset{k,k\neq j}{\text{Max}} \sum_{i=1}^{C} W(i)\cdot F(i,k)\cdot\left(P(i,k,N(k)+1) - P(i,k,N(k))\right) - \sum_{i=1,i\neq i1}^{C} W(i)\cdot F(i,j)\cdot(P(i,j,N(j))-P(i,j,N(j)-1))\right]}{\left[W(i1)\cdot(P(i1,j,N(j)) - P(i1,j,N(j)-1)\right]}\right] \quad (6.13)$$

If the frequency of F(i1,j) decreases below its lower limit then transferring a patrol unit from region j into the region with the maximum incremental payoff, improves on the present deployment. There are, however, two distinctions between searching for the upper and lower limits on F(i1, j). Firstly the lower limit LF(i1,j) can not be less than zero. Secondly, for regions which presently are not allocated any patrol units, decreasing the frequency of any crime in that region can not alter the optimal solution.

As an illustration, sensitivity analysis was carried out.

on the last example in which in the optimal solution, region A, was allocated seven patrol units and region C, three. Table 6.9 summarized the results of this analysis. Notice, first of all, that there would have to be significant increases (at least double) in any of the individual crime rates in the regions B, D or E before the optimal solution would allocate any patrol units to these regions. However, the optimal solution is far more sensitive to changing crime patterns in regions A and C. In the extreme it is highly sensitive to any increases or decreases in the rate of crime type III in either of the two regions. An increase (decrease) of only 3% in the type III crime rate in region A will warrant transferring a unit from (to) region C to (from) A. The same is true, in reverse, for changes in region C's rate. There are some crime types, even in these regions, to which the optimal solution is not overly sensitive. Changes of 30% or more in the type VI crime rate of either of the two regions do not alter the optimal solution. Also, for example, a 30% increase or 14% decrease in the rate of crime type I in region A would not affect the optimal solution.

The methodology for sensitivity analysis described here has been of very limited scope. It analyzes the impact of changing only one of the F(i,j) while all others are assumed constant. Naturally crime rates are not constrained to change in this manner. There may be an across the board increase in the total crime rate of one region, or one

Table 6.9: Sensitivity Analysis on the Frequency of the Different Crime Types in the Different Regions

REGION A

Crime	Frequency	U Bound	Region Loses	L Bound	Region Gains
I	.065	.084	C	.056	C
II	.606	.640	C	.587	C
III	.426	.484	C	.394	C
IV	1.395	1.439	C	1.370	C
V	.408	.452	C	.383	C
VI	.099	.153	C	.069	C

REGION B

Crime	Frequency	U Bound	Region Loses	L Bound	Region Loses
I	.030	.358	A	-	-
II	.322	1.003	A	-	-
III	.264	1.477	A	-	-
IV	.936	1.845	A	-	-
V	.348	1.259	A	-	-
VI	.100	1.232	A	-	-

REGION C

Crime	Frequency	U Bound	Region Loses	L Bound	Region Gains
I	.048	.056	A	.032	A
II	.312	.329	A	.280	A
III	.459	.490	A	.402	A
IV	1.590	1.613	A	1.547	A
V	.513	.536	A	.470	A
VI	.078	.107	A	.025	A

REGION D

Crime	Frequency	U Bound	Region Loses	L Bound	Region Gains
I	.009	.714	A	-	-
II	.184	1.646	A	-	-
III	.105	2.711	A	-	-
IV	.276	2.229	A	-	-
V	.324	2.281	A	-	-
VI	.102	2.533	A	-	-

REGION E

Crime	Frequency	U Bound	Region Loses	L Bound	Region Gains
I	.010	.714	A	-	-
II	.194	1.655	A	-	-
III	.150	2.752	A	-	-
IV	.388	2.338	A	-	-
V	.199	2.153	A	-	-
VI	.059	2.487	A	-	-

particular crime may increase or decrease in all of the regions, or some combination of both may occur.

The sensitivity of the optimal solution to the former type of change (all crimes in one region) can be analyzed through a natrual extension of the methodology already presented. In order to find an upper bound, UB, on the increase in the total crime rate in region j1 (assuming the relative frequency of the crime types remains the same), it is necessary to modify equation (6.11) to be

$$UB(ji) = \min_{j,\, j \neq j1} \frac{\sum_{i=1}^{C} W(i) \cdot F(i,j) \cdot (P(i,j,N(j)) - P(i,j,N(j)-1))}{\sum_{i=1}^{C} W(i) \cdot F(j1) \cdot (P(i,j1,N(j1)+1) - P(i,j1,N(j1)))} \qquad (6.14)$$

In addition it is also possible to relax the assumption that as the total crime rate in the region increases the relative frequency of each crime type in that region remains fixed. This added complication is accounted for by introducing a vector, V(i), of values into the denominator, which reflects the expected relative rate of change in each crime category.

Our discussion of analyzing global changes in a single crime category will be postponed until we present the algorithm for sensitivity analysis on the weights, W(i), associated with each crime type. As we shall see, the two problems are equivalent.

6.5.2 Crime Weights

In contrast to the crime rates, the development of this sensitivity analysis capability was not motivated by the changeability of the crime weights. Instead the motivation lies in that the crime weights will reflect, at least in part, a subjective assessment of the relative seriousness of each crime type. Because of the difficulty in accurately translating subjective attitudes into quantifiable measures, it is important to determine the range over which each weight can vary without affecting the present optimal solution.

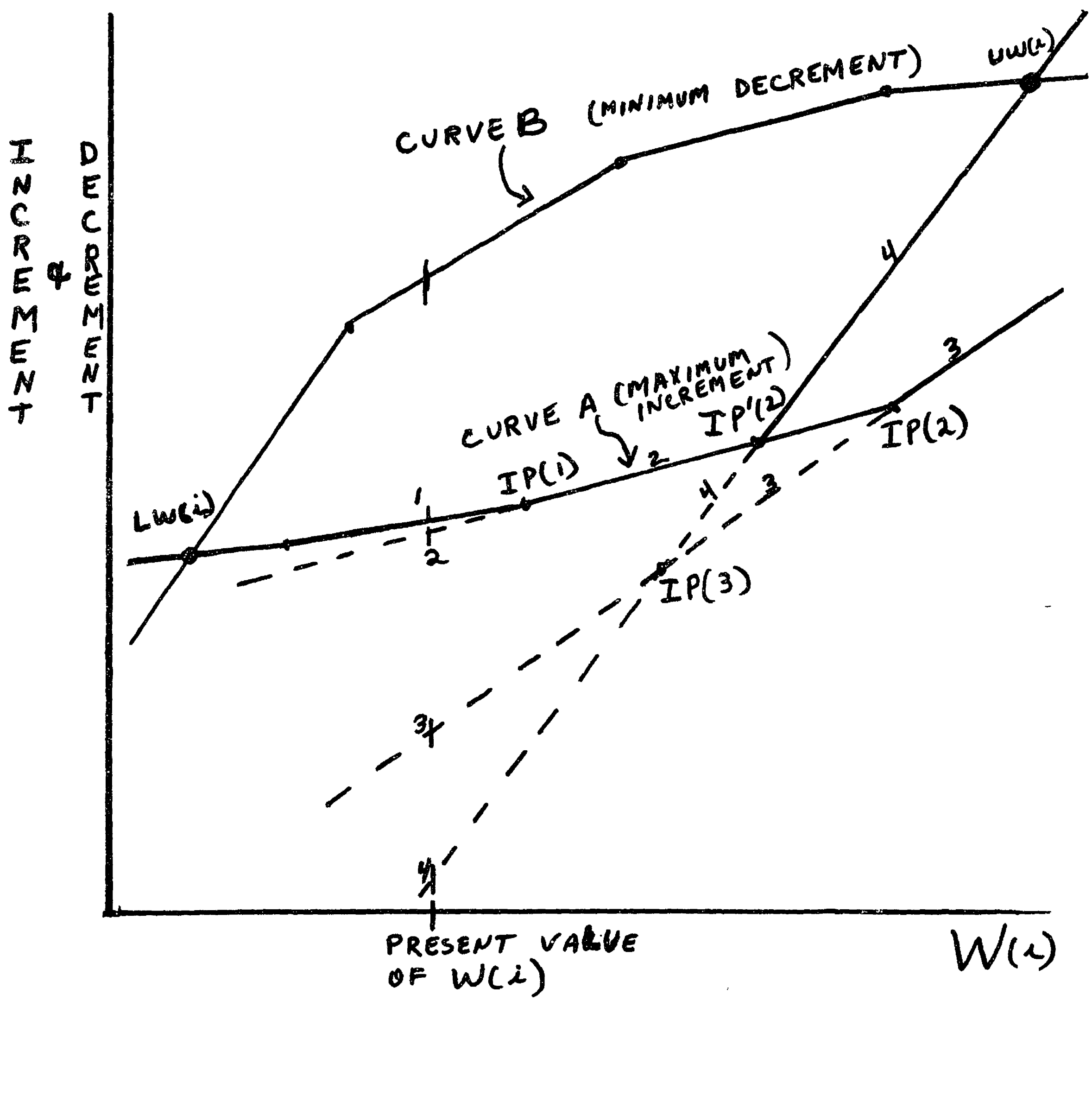
As before, the essence of the sensitivity analysis algorithm consists of determining for what values of W(i) does the optimality condition, maximum increment less than the minimum decrement, no longer hold. This task is complicated by the compound effect produced by changing the weight of a specific crime category. As the weight of a single crime type increases not only does the increment associated with adding a unit to a region change but the ranking of the regions according to their potential increment also changes. Thus for one value of W(i), region j might have the highest increment, while for another value, region k may have the highest increment. Consequently, even though each individual region's incremental payoff is a linear function of the crime weight, W(i), the maximum increment is a piecewise linear convex function of W(i) (See Figure 6.1). Analogously, the minimum decrement produced by removing a patrol unit from a region is a piecewise linear concave function of W(i).

The sensitivity analysis algorithm we constructed consists of two components. The first component maps the two piecewise linear functions, for the maximum increment and minimum decrement over the entire range of positive values of W(i) (in Figure 6.1, curves A and B respectively). The second component determines the intersection point of the two curves thereby determining the upper, UW(i), and lower, LW(i), limits within which W(i) can vary without affecting the present optimal solution. As W(i), however, increases above the upper limit, UW(i), the present allocation can be improved upon by tranferring a patrol unit. Since each line segment represents a different region, the two regions corresponding to the intersecting lines at UW(i) are the ones between which the patrol unit should be transferred.

The process of mapping a piecewise linear function requires the specification of only the starting point and slope of each line segment. The algorithm presented here uses the list which ranks the regions (for the present value of W(i)) according to their incremental payoffs as a starting point in determining these numbers. Analyzing increases and decreases in W(i) separately, it performs a series of pair wise comparisons and calculations with consecutively ranked regions on that list to generate the maximum increment curve. For example, in order to determine the effect of increasing W(i) the following computations are performed.

As was mentioned earlier, each region's incremental payoff is a linear function of W(i) with a slope of F(i,j)·(P(i,j,N(j)+1) - P(i.j.N(j))). This expression is used to compare the slopes of the regions ranked R and R+1. If the region ranked R+1 (lower payoff) has a smaller slope, then for increases in W(i) its incremental payoff will always be dominated by that of region R. Therefore, for increases in W(i), it can never represent the maximum increment and can be eliminated from further consideration. If, on the other hand, its slope is greater than R, a second claculation is performed. The intersection point of the two lines, IP(R), is determined. Thus for W(i) greater than IP(R) the region ranked R+1 dominates and for W(i) less than IP(R) the region ranked R dominates. Next IP(R) is compared to the previous intersection point IP(R-1). If the point IP(R) occurs prior to IP(R-1), relative to increases in W(i), the region ranked R can be eliminated from consideration. Using Figure 6.1 as an example the justification is as follows. The lines marked 1 through 4 represent the regions whose incremental payoffs are respectively ranked 1 through 4 for the present value of W(i). Although the line marked 3 dominates the line marked 4 up to the point IP(3), it itself is dominated by the line marked 2. Between the points IP(3) and IP(2), line 3 is now dominated by both lines 2 and 4 and above IP(2) it is also dominated by line 4. Thus over the entire range of increasing W(i) line 3

Figure 6.1: A Graphic Description of the Method for Finding the Maximum Increment and the Minimum Decrement



can never represent the maximum increment. Once the line ranked R (in our example, R is 3) is eliminated, the intersection point, IP'(R-1), of the lines ranked R-1, (2), and R+1, (4), is determined and compared to the previous intersection point, IP(R-2). The final product of all of the above comparisons of slopes and intersection points is a list which is a subset of the original incremental payoff list. This new list is simply an ordering of the line segments in the seqeunce that they appear in the piecewise linear curve as W(i) increases. When the last comparison has been completed the maximum increment has been completely specified since the starting point of each line segment is the already determined intersection point of consecutively ranked line segments.

The above series of comparisons are then reversed to find the maximum increment curve for decreasing values of W(i). For example if the incremental payoff for the region ranked R+1 has a higher slopethan that of the region raked R, it can be eliminated from the list. Lastly an analogous set of comparisons are performed in order to graph the minimum decrement curve. The upper and lower limits on W(i) are then found by locating the intersection points of the two curves.

Table 6.10 summarizes the results of sensitivity analysis carried out on the weights. The optimal solution is least sensitive to a <u>change</u> in the weight on type VI crimes, with the solution totally insensitive to a <u>decrease</u> in the weight.

This insensitivity is the result of a combination of two factors: (1)Type VI crimes make up less than 5% of all the crimes in these six categories, (2)they have the second lowest probability of interception with only the interception rate of type III crimes lower. At the other end of the spectrum, even minor changes in the weights on either type IV or V crimes affects the optimal solution. If the weight on type IV crimes were greater than 1.1, instead of the present value of 1, then shifting a patrol unit from region A to region C improves upon the present deployment. Conversely if the weight were less than .840, a patrol unit should be transferred from region C to region A.

In the above example, and throughout the discussion of sensitivy analysis, the focus has been on determining over what range of values does the present solution remain optimal. Changes in the input data which are just above or below the limits will require the reallocation of only one unit to achieve the new optimal solution. As a contrast is might be interesting to illustrate how drastically the optimal solution can change when far more significant changes occur, for example, in the weights. Consider a situation in which a decision is made to place an emphasis on catching criminals who specialize in type IV and type V crimes. To reflect this emphasis the weights on these crime types are doubled. Using the same data as in the last example, we find that the optimal solution is exactly the reverse of before. Region C is now

CRIME TYPE	WEIGHT	UPPER BOUND	REGIONS LOSES	REGIONS GAINS	LOWER BOUND	REGIONS LOSES	REGIONS GAINS
I	1.0	2.6	C	A	.31	A	C
II	1.0	1.12	C	A	.93	A	C
III	1.0	1.74	A	C	-	-	-
IV	1.0	1.10	A	C	.84	C	A
V	1.0	1.2	A	C	.66	C	A
VI	1.0	4.1	C	A	-	-	-

Table 6.10: Sensitivity Analysis on the Crime Weights

CRIME TYPE	PARAMETER	MEAN	UPPER BOUND	REGION LOSES	REGION GAINS	LOWER BOUND	REGION LOSES	REGION GAINS
I	OBSERVABILITY	.14	.65	A	C	.03	A	C
	DURATION	.10hrs	.46	A	C	.02	A	C
	PRODUCT	.014	.065	A	C	.003	A	C
II	OBSERVABILITY	.10	.12	C	A	.09	A	C
	DURATION	.066	.08	C	A	.06	A	C
	PRODUCT	.0066	.008	C	A	.006	A	C
III	OBSERVABILITY	.10	.16	A	C	-	-	-
	DURATION	.037	.06	A	C	-	-	-
	PRODUCT	.0037	.006	A	C	-	-	-
IV	OBSERVABILITY	.03	.033	A	C	.025	C	A
	DURATION	.167	.184	A	C	.142	C	A
	PRODUCT	.005	.0055	A	C	.0043	C	A
V	OBSERVABILITY	.04	.048	A	C	.026	C	A
	DURATION	.125	.150	A	C	.081	C	A
	PRODUCT	.005	.006	A	C	.0032	C	A
VI	OBSERVABILITY	.06	-	-	-	-	-	-
	DURATION	.066	-	-	-	-	-	-
	PRODUCT	.004	-	-	-	-	-	-

Tabel 6.11: Sensitivity Analysis: Crime Descriptive Data

allocated seven patrol units as compared to three before, with the remaining three assigned to region A, which had seven before. Although changes of the above magnitude can be explored through sensitivity analysis algorithms, the more appropriate response it to rerun the original algorithm for the changed data base.

One final point to be discussed with reference to the weights is their relationshipto the crime frequencies, F(i,j). In expression (6.11), which describes the conditions for optimality, the two parameters always appear as the product, W(i)·F(i,j). Consequently, doubling the weight on a particular crime type is equivalent to doubling the frequency, in each region j, of crimes of type i. Therefore, the above sensitivity analysis on the weights can also be applied to changes in the total frequency of the different crime types (assuming the distribution among regions remains the same). Therefore by referring back to Table 6.10, we can determine that type II crimes would have to increase 74% across the five region area before the optimal deployment strategy would change; while a decrease in that crime's total rate would not affect the optimal solution. On the other hand, a 10% increase in the five region type IV crime rate would necessitate a reassignment of one patrol unit from A to C and a 12% increase in type II crimes would have the reverse impact.

6.5.3 The Observability and Duration of a Crime

The last set of parameters to be analyzed are the two crime descriptive ones. The need for a sensitivity analysis capability on these parameters is unfortunately all too obvious. No data exists at present with regard to these parameters nor is there research being carried out to determine these numbers. Consequently early applications of search theoretic models will have, at best, only very rough estimates to work with and even with extensive research in the future, it is not clear how good a set of estimates can be obtained.

Sensitivity analysis will be performed simultaneously on the observability and duration of each crime type. This can be done because the critical input parameter in calculating the probability of interception (see equations(6.6) through (6.10)) is the product of the two numbers, 1/A(i) (mean duration) and OB(i) (observability), and not their individual values. Of course once the limits on the product have been calculated, the limits on the individual parameters follow directly. Aside from this simplification, sensitivity analysis of these parameters is far more complex than for the weights or crime frequencies.

The key problem is that not only is the incremental payoff for a region not a linear function of 1/A(i) or OB(i), it is not even a monotonically increasing function of these parameters. Assume for example that the duration of a crime is exponential. The payoff of adding an additional patrol unit to region j is:

$$\sum_{i=1}^{C} F(i,j)\cdot W(i)\cdot \left[\frac{(N(j)+1)\cdot S\cdot OB(i)/A(i)}{((N(j)+1)\cdot S\cdot OB(i)/A(i)+M(j)} - \frac{N(j)\cdot S\cdot OB(i)/A(i)}{(N(j)\cdot S\cdot OB(i)/A(i)) + M(j)}\right] \quad (6.15)$$

Notice that as all the OB(i) or 1/A(i) approach either <u>zero</u> or <u>infinity</u>, the incremental payoff approaches zero (if N(j) is not zero). This phenomenon can be explained as follows. For large OB(i) there is a high probability of intercepting the crime even with only one patrol unit; thus additional patrol units can not have much of an impact. At the other extreme for small OB(i), since there is only a small probability of intercepting a crime, adding one more patrol unit can not significantly increase (in absolute terms) the probability of intercepting a crime. For intermediate values, though, the incremental payoff is certainly not zero.

The nonlinearity of the incremental payoff expression complicates the task of sensitivity analysis but the non-monotonicity has an even more profound effect. It necessitates a different understanding of the upper and lower limits. For example, as OB(i) increases above its present value it may reach a point, UOB(i), above which the present solution is not optimal because the maximum increment is greater than the minimum decrement. However as OB(i) increases further the maximum decrement will begin to decrease which may result in its becoming once again less than the minimum decrement leaving the present solution still optimal.

Although cognizant of the above issues, the algorithm to be presented will focus only on determining the initial bounds within which the product, OB(i)·1/A(i), can vary without changing the optimal solution. However, the algorithm will not, for example, analyze the behavior of the optimal solution as the product increases significantly above the bound. Even for this limited problem, sensitivity analysis will be carried out using a direct brute force approach because of the non-linearity of the incremental payoff.

Starting from the present value of the product, the product is increased by a user specified percentage. The incremental payoff and decrement are compared. If the maximum is greater than the minimum then the present optimal solution is not optimal for the new value of the product. If, however, the maximum increment is still less than the minimum decrement, then the product is increased again by the same amount and the comparison is repeated. This procedure is repeated until we reach a value of product for which the present solution is not optimal or reach a user specified reasonable upper bound on the product. The same methodology is used to determine a lower bound on the product with an obvious implicit bound of zero on the product. The above approach yields only approximations of the upper and lower limits; however, these approximations can be made as accurate as the user desires by specifying the magnitude of each step. In the example presented below, the product of OB(i) and 1/A(i) was first increased by five percent of its initial value at each step and later the process was reversed, and the initial value was repeatedly decreased by five percent.

Table 6.11 summarizes the results of the analysis. Not surprisingly the results parallel those for the sensitivity analysis of the crime weights, Table 6.10. Again the optimal solution is least sensitive to changes in type VI crime data and relatively insensitive to changes in type I and type III crimes. If the estimates were found to be low by one minute (.017 hours) or more, then the optimal solution would reallocate one patrol unit from region A to region C. The sensitivity analysis of type I crimes shows an interesting phenomenon occurring. Whether the actual mean was above .46 hours or below .02 hours, to achieve optimality, the present solution would have to be modified in <u>both</u> cases by reallocating a unit from A to C. This phenomenon can not occur in sensitivity analysis of the weights and crime frequencies but is possible here because of the aforementioned non-monotonicity of the marginal return.

6.6 Summary

In the preceding sections of this chapter we presented a flexible algorithm for optimally deploying a tactical patrol force among several competing regions. The examples presented in the chapter were chosen with two purposes in mind. The first purpose was to display the importance of knowing not only the total crime rate of each region but also the distribution of crimes by type for each region. The second was to show the basic nonlinearity of the optimal allocation. Instead of allocating patrol in direct proportion to the crime level the optimal solution tends to focus on just the highest crime regions. In the example presented this allocation produced a 22% higher probability of interception than a proportional (to total crime rates) allocation of patrol. However of perhaps even greater significance than the above insights is that through the perspective of the allocation model we can see what the critical variables are in developing optimal patrol strategies. Besides the frequency of observable crimes of each type, the important parameters are the observability and duration of each crime type. The need for data of this type seems almost intuitively obvious and yet no extensive research has yet been done in this area.

Included in the algorithm are components for performing sensitivty analysis on each of the input parameters. The allocation model, we noted, pointed out general data needs for developing effective patrol strategies. The sensitivity analysis algorithms analogously pinpoint, on a more microscopic level, the critical parameters in the particular regions under consideration (i.e. slight variations can affect the optimal solution). It will be these parameters that require the most accurate estimates since small inaccuracies might change the optimal solution.

REFERENCES 6

1. Kupersmith, G., High Impact Anti-Crime Program: Sample Impact Evaluation Components, U.S. Department of Justice, LEAA, NILE & CJ, U.S. Government Printing Office, Washington, D.C., July 1974.

2. Maltz, M.D., "Measures of Effectiveness for Crime Reduction Programs", Operations Research, Vol. 23, No. 3, 1975.

3. Moore, L.M., A Characterization of the Visibility Process and Its Effect on Search Policies, Systems Research Laboratory, Department of Industrial Engineering, The University of Michigan, Report No. SRL 2147 TR 71-3, Ann Arbor, MI, December 1971.

4. National Crime Panel Surveys of Chicago, Detroit, Los Angeles, New York and Philadelphia, Crime in the Nation's Five Largest Cities: Advance Report, U.S. Department of Justice, LEAA, National Criminal Justice and Statistics Service, Washington, D.C., April 1974.

5. Sellin, T. and M.E. Wolfgang, The Measurement of Delinquency, Wiley, NY, 1964.

CHAPTER 7

ALLOCATING THE PATROL TIME OF SINGLE AND MULTIPLE PATROL UNITS

7.0 Introduction

The tactical patrol force allocation model that was presented in Chapter VI is oriented towards deployment issues on a precinct or city wide level. In this next chapter we will present issues more relevant to patrol strategies on a sector or two sector level, involving a standard patrol car with responsibility for calls for service. In our search for and discussion of efficient patrol strategies, we will limit the class of solutions surveyed to those consisting of patrol cars concentrating their patrol efforts in a single contiguous region. This constraint is based on the results of earlier chapters. In Chapter V we showed that in allocating search effort between noncontiguous regions, the optimal strategy will usually be to search only one region rather than incur the time lost (from search) in travel between the regions. Thus, the general problem we analyze here involves constructing from small building blocks, called atoms (2 square blocks in size), the best contiguous region in which to concentrate patrol. In applying an algorithm which concentrates patrol in a limited area to the police environment, it may be necessary to set aside a minimal amount of patrol time to be allocated over the entire sector. (This point will be elaborated on later in this chapter.) On the whole, though, the discussion will be exploratory, suggesting one approach to the problem and presenting some insights as to the eventual size of the region.

7.1 Methodology

The objective function to be maximized is the same as before, the weighted probability of intercepting a crime. As a result the data requirements do not change except for the addition of two sets of data. Since the algorithm constructs a contiguous region out of a set of individual atoms, information is needed to describe which atoms are contiguous to one another. Secondly, since this analysis involves standard patrol cars, some estimate must be given for the average workload of the unit. The best available method for calculating this statistic is the hypercube queuing model [1].

The basic steps in the algorithm are listed below:

1. Calculate for each atom the payoff (i.e. weighted probability of intercepting a crime) resulting from a single patrol unit spending all of its 'free' time patrolling just that one atom.

2. Rank the atoms in order of their payoff.

3. Incorporate the highest ranked atom into the region to be patrolled.

4. List all atoms (not yet in the region) that are contiguous to an atom that is already included in the heavily patrolled region and rank them according to their <u>initial</u> payoff (step 1).

5. Determine if adding the highest ranked atom on the contiguity list increases the total weighted probability of intercepting a crime. If it increases the payoff include it into the patrolled region. If not, test the next highest ranked atom, etc. If no atom can increase the payoff go on to step 6 otherwise return to step 4.

Once the point is reached that no further single atom expansions of the patrol area increase the payoff, we proceed to check another alternative.

6. Determine if removing an atom can increase the payoff. Remove the atom which produces the greatest increase and return to step 4. (Only those atoms which can be removed without leaving the region split in two are considered in this step.) If no improvement can be generated, the algorithm stops.

Consequently, the stopping point of the algorithm is a local optimum at which adding or removing a single atom does not increase the weighted probability of intercepting a crime. There are, of course, numerous approaches that can be followed in trying to determine if the present local optimum can be improved upon. One alternative is to see if adding or removing pairs of atoms from the present patrol region can improve on the present solution.[1] In our programmed version of the above algorithm the alternative we chose was to rerun the algorithm a second time. However, this time we forced the construction of the patrol region to begin with the inclusion of an atom that was not in the original solution. In all our trials the region constructed in the second pass through the algorithm was the same as the first.[2]

Before proceeding with a discussion of how the weighted probability of intercepting a crime is calculated, we would like to comment briefly on steps 4 and 5 in the algorithm. In step 5 it is not sufficient to test only the highest ranked atom for inclusion. The rankings used here were not generated by calculating how much the addition of each contiguous atom would increase the payoff. Instead we used the <u>initial</u> rankings, which compared each atom's potential payoff from a single patrol unit patrolling only that atom. The reason for taking the second approach (which in general is a good surrogate) is that otherwise it would have been necessary to recalculate at each iteration how much every contiguous atom would change the present payoff. However, even though in step 4 we use the ordering in the original unchanging list to determine which atom to add (as if to say the marginal and initial rankings are the same), in step 5, we do not assume that if the contiguous atom ranked first (in the initial list) can not increase the payoff that no contiguous atom can.

7.1.1 The Weighted Probability of Interception

The calculation of the weighted probability of intercepting a crime in this algorithm takes on the same general form as in Chapter 6 with one modification. An additional parameter is included to reflect the possibility that during the commission of the crime the patrol unit may be busy responding to a call and therefore not on patrol. Using the following parameters:

Let A= set of all atoms contained in the patrol area

B= average fraction of time the patrol unit is busy

F(i,j)= the frequency of crime type i in atom j

$F(i) = \sum_{j, j \in A} F(i,j)$, the total frequency of crime type i in the patrol area

M(j)= the street mileage in atom j

$M = \sum_{j \in A} M(j)$, the total street mileage in the patrol area

P(i,M)= probability of intercepting a crime of type i in the patrolled region (M street miles)

The expression for the weighted probability of intercepting a random crime is

$$\text{Payoff} = (1-B) \sum_{j,\, j \in A} \sum_{i=1}^{c} F(i,j) \cdot W(i) \cdot P(i,M(i))$$

$$= (1-B) \sum_{i=1}^{c} F(i) \cdot W(i) \cdot P(i,M) \qquad (7.1)$$

There are two underlying assumptions that produce a number of simplifications used in generating the above expression. The first one, which was discussed earlier in Chapter 4 (section 4.1), assumes that because the duration of a crime is small relative to the average duration of a call for service, that a patrol unit will either be busy or free (on patrol) throughout the entire duration of a crime. The result is that although the average workload affects the total payoff, its impact is independent of which atoms are included in the heavily patrolled area. It therefore can be ignored when deciding which atom to include in the patrol area. However, when the two patrol car example is presented, it will be apparent that the value of the average workload will affect the construction of the optimal region to patrol.

In addition the patrol effort is assumed to be distributed uniformly over the patrolled arèa and is not distributed in proportion to the crime rates of the individual included atoms. As a consequence only the total street mileage and total frequency of each crime type in the patrolled region is used. If, however, each atom were allocated a different proportion of patrol, then each atom would need be treated separately. For each atom it would be necessary to calculate the following:

> "Conditioned on a crime of duration t_0 minutes occuring in atom i, what is the probability distribution for the total time t, t less than t_0, during the commission of the crime, that the patrol unit is searching the same atom i."

As was discussed in Chapter IV this is an extremely difficult thing to calculate. This assumption (i.e. patrol uniformly distributed) is not as unreasonable as it first might seem because it is unlikely that the crime rates of the atoms in the region patrolled will be widely disparate. The second example will focus on this point.

In terms of equation (7.1) the process of searching for atoms to be included into the patrolled area, involves balancing two opposite effects. Adding more atoms increases each F(i) thereby increasing the total payoff. At the same time adding to the size of the region increases the street mileage, thereby decreasing the probability of intercepting a crime in the area patrolled.

7.2 Single Patrol Car

For the first example we constructed a sector of 15 atoms (approximately 8 square city blocks) each of which contained .4 street miles (see Figure 7.1). The distribution of crimes within each atom was generated from the earlier five region data. Atoms one through five have the same distribution as the five regions of the earlier examples. For the other ten atoms the distributions were generated by averaging the distribution of pairs of the first five atoms. As a result the range of distributions is limited and the total crime rate in each atom is the same. The patrol unit was assumed busy 50% of the time. The crime descriptive data is the same as in Chapter 6 (See Table 6.2).

The optimal region to patrol included almost the entire sector (excluding only atom 3) because of the uniformity of the crime rates. Overall there was a .0064 chance of intercepting a random crime. In the atoms that were patrolled the probability of intercepting a crime ranged from a low of .00677 in atom 13 to a high of .00703, only a difference of 4%, reflecting the small variations in the distributions. Table 7.1 summarizes these results.

The above example represents one extreme in terms of the size of the optimal patrol region. However, if the crime rate were less uniform, then the size of the region in which patrol is to be concentrated would decrease. In the next example the total crime rate in atom 1 was changed; it was

ATOM NUMBER	PROBABILITY OF INTERCEPTING A CRIME THERE
1	.00703
2	.00686
3	.00000
4	.00681
5	.00683
6	.00694
7	.00689
8	.00692
9	.00693
10	.00679
11	.00684
12	.00685
13	.00677
14	.00678
15	.00683
Average (random crime)	.00640

CRIME TYPE	PROBABILITY OF INTERCEPTION
I	.01673
II	.00845
III	.00460
IV	.00607
V	.00623
VI	.00515

Table 7.1: Probability of intercepting a crime in the heavily patrolled region of the sector

increased to 1.2 times that of any of the other atoms. The resultant solution limited patrol to just atom 1. To understand why this happened, we will analyze the changes that occur when an atom is added to the patrol region.

When the algorithm is applied in this example, the first atom included in the patrol area is naturally atom 1. As the algorithm considers expanding the size of the patrol area, the underlying issue is whether or not half of the patrol time now presently allocated to atom 1 should be allocated instead to some other atom. The question then is, what is the payoff from this half of the patrol time when it is assigned to atom 1 and what is its payoff when it is allocated to some other atom?

The first point to realize is that for a typical crime described in Table 6.2, the product, t·CB (observable duration and average observability), is on the order of magnitude of .005. In addition we are assuming that the patrol speed is 15 miles per hour, the patrol unit is busy 50% of the time and that the observable duration of a crime has an exponential distribution. Thus if half the patrol effort is allocated to atom 1 (.4 street miles), the probability of intercepting a crime is approximately

$$(1-B) \cdot \left(\frac{s \cdot t/2 \cdot OB}{s \cdot t/2 \cdot OB + M} \right) = .5 \cdot \left(\frac{15 \cdot .005/2}{15 \cdot .005/2 + .4} \right) = .043$$

If however all the effort were allocated to the single atom, the probability of interception increases to .079. Due to

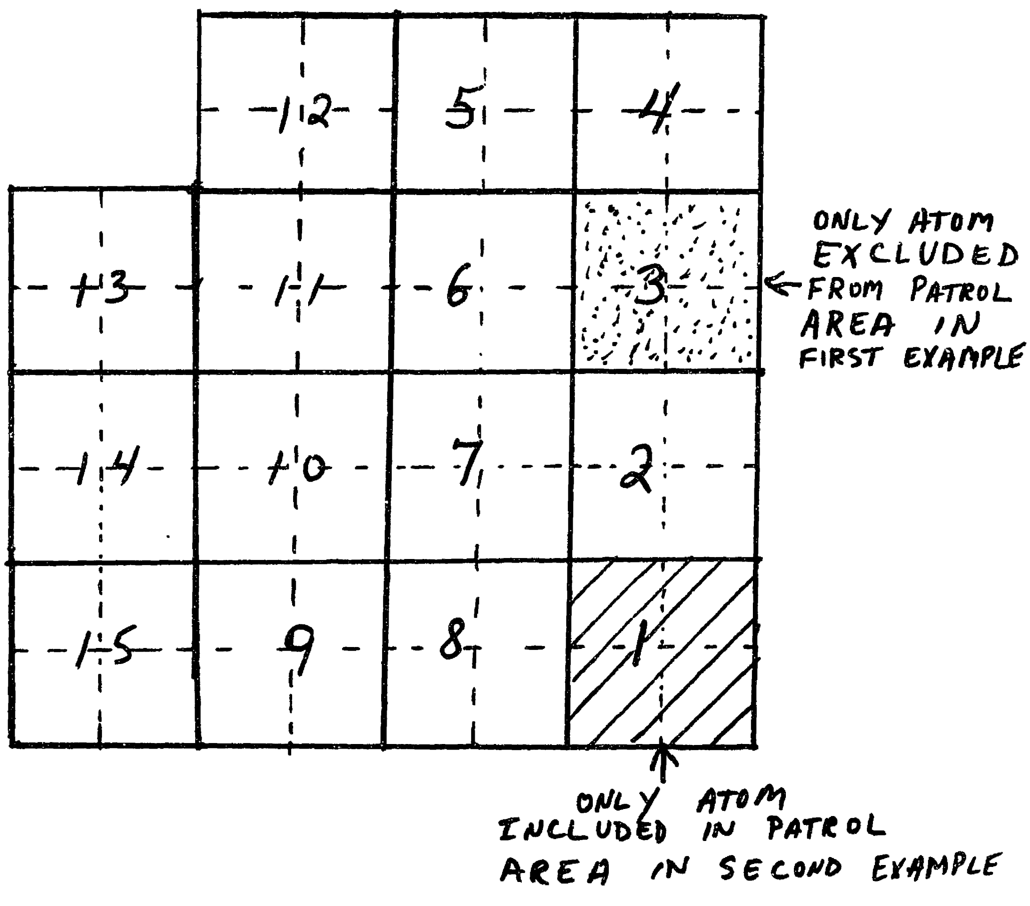


Figure 7.1: The Optimal Region in Which to Concentrate Patrol

diminishing return the yield from the additional search effort is only about 84% as great as the initial effort. If instead this additional effort were allocated to another atom, the interception probablity there would also be .043; however, the crime rate there is only 83% (1/1.2) of that in atom 1. Consequently concentrating all the search in atom 1 has an overall higher payoff than splitting the search betwen atom 1 and a second atom. Extrapolating from the above, the atoms to be included in the patrolled area are not likely to have crime rates that vary, from highest to lowest, by more than about 17% assuming the observability and duration are on the order of magnitude of the data used in our example.

The algorithm presented in this chapter is intended only as an outline of one approach to finding an optimal area to patrol and is not offered as a finished product. For one thing optimal solutions similar to the last example, which limit patrol to a two square block area, will likely prove counter-productive as criminals move to less well patrolled, more productive areas. Obviously reactions similar to the above must be taken into account in developing realistic more effective patrol strategies. Thus algorithms of the type described here could for instance be incorporated in a game theoretic approach that attempts to anticipate possible criminal reaction to patrols. Alternatively, a proportion of the patrol effort might be set aside for making relatively high speed patrols in the lower crime areas. This last alternative could even be handled within the general framework of an algorithm similar to the one presented in this chapter. However, the possible repercussions of concentrating patrol in only a small section of the sector (several atoms) may not be as large as one might expect. Larson [2] in his review of the Kansas City experiment pointed out that a high level of visibility was maintained in beats devoid of regular patrol as a result of cars responding (often with sirens going) to calls for service in the unpatrolled areas.

7.3 Extensions to Two Patrol Units

In attempting to develop more flexible effective patrol strategies, one alternative we discussed earlier (Section 4.2) involves combining an area normally divided between two patrol units into a single overlapping sector. Calls for service would be shared equally (automatically balancing workloads) with each unit's patrol effort concentrated in only high crime areas within the enlarged joint sector. In searching for the region or regions in which to concentrate patrol, an algorithm should consider and compare these alternatives:

1. The two patrol units patrol two disjoined sections of the larger sector, which are selected to maximize the overall weighted probability of intercepting a crime.
2. The two patrol units concentrate their efforts in one single optimal region.
3. The units patrol partially overlapping regions with the overlap centered on the highest crime areas.

The construction of two separate patrol areas (alternative 1) can be accomplished by using iteratively an algorithm analogous to that described in the previous section. However to locate the optimal region to be patrolled jointly by the two patrol units, equation (7.1) for calculating the probability of interception must be modified. The new equation becomes

$$\sum_{i=1}^{c} p(1)\cdot F(i)\cdot P(i,M,1) + \sum_{i=1}^{c} p(2)\cdot F(i)\cdot P(i,M,2) \qquad (7.2)$$

where p(1) and p(2) are the respective probabilities that one and two patrol units are on patrol. P(i,M,1) and P(i,M,2) are the probability of intercepting a crime of type i in a region of size M when one and two patrol units respectively are searching for crimes. If the duration of a crime is assumed exponential with mean t then

$$P(i,M,2) = \frac{2\cdot S\cdot OB\cdot t}{2\cdot S\cdot OB\cdot t + M}$$

If the two patrol units respond to calls for service only within their sector then

$$p(1) = 2b\cdot(1-b)/(1+b)$$

and

$$p(2) = (1-b)/(1+b)$$

with b representing the average workload of each of the patrol units. In general, though, the two units will be part of a larger precinct and will respond to calls for service outside of their sector. Under those circumstances, it would be necessary to use the hypercube model to calculate these probabilities as they are intertwined with the operation of the rest of the precinct. We discussed earlier (Section 4.2.3) that as a result of overlapping sectors travel time would increase. Consequently, the hypercube could be used at the same time to measure also the magnitude of this increase.

An algorithm which would also consider partially overlapping patrol (alternative 3) would need to use both equations. Equation (7.1) would be used to calculate the probability of intercepting a crime in the non-overlapping parts of the patrol area and equation (7.2) would be used for the overlapping regions.

7.4 Summary

The above discussion is an introduction to some of the issues and approaches to developing a methodology for better utilizing the patrol time of a standard patrol car. To the insightful reader, though, this discussion may have raised more questions than it answered. How does one incorporate response time, a commonly used surrogate for police effectiveness, along with the search theoretic model into a single model of patrol? How quickly do crime patterns change in response to concentrated patrol efforts? These are just some of the unanswered questions that immediately come to mind. Without answers, though, it will be impossible to develop a total composite picture of the multidimensional interaction between police patrol and street crime.

FOOTNOTES 7

1 This alternative is not likely to help under the following circumstances: At one end of the sector is an atom i with the highest crime rate. However, the crime rate drops sharply in the surrounding atoms. At the other end of the sector there is a larger high crime area but with no one atom having a crime rate as high as atom i. Thus the algorithm would concentrate patrol in i even though it would be better to patrol the other high crime area. The second alternative for improving on the local optimum would avoid this problem.

2 The fact that the algorithm began with a particular atom does not necessarily mean that that atom will appear in the final patrol area. Remember the algorithm allows for the removal of atoms if their removal will increase the total payoff.

REFERENCES 7

1. Larson, R.C., "A Hypercube Queuing Model for Facility Location and Redistricting in Urban Emergency Services", Computers and Operations Research, Vol. 1, No. 1, March 1974.

2. Larson, R.C., "What Do We Know About Preventive Patrol? A Review of the Kansas City Preventive Patrol Experiment", published by Public Systems Evaluation, Inc., Cambridge, MA, July 1975, to appear in Journal of Criminal Justice.

CHAPTER 8

SUMMARY AND CONCLUSIONS

8.0 Introduction

The models that were presented earlier focus on two distinct but related issues in police patrol deployment. The first set of models address deployment issues through the perspective of sector design, with the emphasis on problems of equity. These models take a relatively broad view of police in that the crime directed activities of police are not the sole focus (e.g. travel time is an important consideration for all emergent calls whether or not they are crime related.). The second set of models, however, focus only on crime with their goal being to generate increased probabilities of intercepting crimes in progress. One of the models works within the framework of a sector configuration. Its function is to determine where in each sector the local patrol unit should concentrate its patrol time. The second model is, however, not constrained by sector boundaries as it allocates a tactical patrol force to the high crime areas in a precinct or city. When all of the above models are used together, it is possible for a police decision maker to address a range of deployment questions from several perspectives.

In the succeeding sections we will review in greater detail each of these models. In addition we will present a number of general conclusions that have resulted from our continuing work in model development. Finally, this chapter will close with a discussion of a number of directions for future research.

8.1 Summary

8.1.1 Sector Design

The interactive system for designing sectors that was presented offers a different approach to the problem of sector design than has, in general, been followed until now. With the development of the hypercube queuing model, it has become possible to focus on multiple criteria. This system is just a logical superstructure that can be applied to the hypercube. It attempts to guide the user towards his preferred goal while utilizing, to the fullest extent possible, the rich array of performance measures the hypercube model estimates.

The system, in general, focuses on various definitions of equity. Through a series of examples we have attempted to explore some of the conflicts and tradeoffs between, for example, balancing workloads and balancing travel times. Similarly we have analyzed the distinction between balanced workloads and balanced preventive patrol coverage. If crime rates do not mirror general calls for service, the two objectives will almost certainly conflict. One potentially important aspect of the system is that the very nature of the system tends to direct the potential user towards making explicit decisions as to his preferences for tradeoffs between important performance measures.

Although we have also touched on some of the conflicts between equity (e.g. balanced travel times) and efficiency (e.g. minimized precinct-wide travel times), there is much that requires exploration. One example that comes to mind is a consequence of our work in search theory. A strategy that maximizes the probability of interception will tend to allocate to the higher crime areas a proportion of the patrol effort that is <u>greater</u> than the fraction of crimes that they generate. On the other hand, balanced travel times are achieved by allocating to the regions with the highest call rate a proportion of the patrol force that is <u>less</u> than the fraction of calls they generate. How does one choose between the two alternatives? As we have emphasized before, only the local decision maker can decide.

The system as described in Chapter III is complete conceptually. However, as in the development of the interactive interface with the hypercube queuing model [4], much will be learned from user feedback and incorporated into the final structure of the system. In addition, although the system was designed with the police in mind, it has potentially broader applications. For example, the issues of balancing travel times and workloads are also relevant issues for the deployment of emergency medical vehicles. Perhaps the most generalizable aspect of the system, though, is simply its different approach to redistricting. Instead of focusing on one performance measure, the system guides the decision maker through a process which involves repeated evaluation of his relative preferences in order to reach a final acceptable solution.

8.1.2 Search Theory Applied to Police Patrol

In this section we will discuss the second set of models which are based on search theory. In Chapters VI and VII we presented two sets of algorithms which can be applied to different facets of police patrol deployment. The algorithms are conceptually simple and therein lies much of their strength and flexibility. By being able to use only the basic formula for the probability of interception, it became possible to build a great deal of flexibility into the algorithms. Consequently, they are not limited by the form of the probability distribution for the observable duration of a crime. In addition a value structure (either the decision maker's or the community's) is easily incorporated in the models in order to reflect the relative seriousness of the various crimes. Similarly, if the relationship between interception probabilities and crime levels can be discovered (in terms of both deterrence and removing criminals from the street), the algorithms can be modified so as to focus instead on the longer range goals of crime reduction.

However, in reviewing the contributions of this work in applying search theory to the deployment of police, it is not really the specifics of the algorithms that are of the most significance. Instead, in many ways, the real significance lies in the foundation that was laid in Chapters IV and V. There we discussed the basic character of the search for crime through the analysis of diminishing return and the introduction of a differential equation model. By analyzing travel time between regions and the various charcteristics of crimes (e.g. short duration, limited observability, random arrival), we were able to justify the restructuring of the deployment problems discussed in the later chapters. Thus in deploying a tactical patrol force we could ignore solutions that would require patrol units to travel between regions. Similarly in developing an algorithm for deploying a standard patrol car, we could focus on constructing a single contiguous region to patrol. And, as was noted before, it is in this restructuring and simplification of the problem that it becomes possible to build into applications of search theory to police patrol, performance measures other than just the probability of intercepting a random crime.

It may be necessary, however, to modify the solutions generated by these algorithms. Their tendency is to concentrate patrol forces in some areas to the exclusion of others and this may violate political and equity constraints. In addition if the disparities in patrol regions become too obvious to criminals, the best theoretical strategies may be self defeating. This last issue can be addressed either by expanding the model to include game theoretic concepts to anticipate criminal reactions or perhaps by simply applying some ad hoc modification of the solution that makes direct use of police familiarity with the local crime problem.

8.2 Conclusions

A major consequence of the relatively short observable duration of crimes (i.e. small probability of interception) is that patrol should be heavily concentrated in the highest crime areas assuming that there are significant geographic variations (a range of 20% or more) in crime levels. Concentrating the patrol efforts of a single patrol unit in the highest crime area of its sector could itself produce a 5% increase in the probability of interception. However, some highly visible patrol should still be allocated to other parts of the sector for two reasons. One is because of political constraints which might make it infeasible to eliminate regular patrols of the entire sector. (Later in section 8.3, we discuss some aspects of the Kansas City experiment which involved leaving entire beats without patrol.) Secondly, by maintaining visible presence in other parts of the sector, the rate at which crime patterns shift away from the present intensely patrolled area to other parts of the sector should be slowed. This would increase the chances of intercepting a crime before the present area of focus is no longer the optimal patrol region. This brings us to a necessary component of any implementation of concentrated patrols. Crime patterns must be continuously monitored. Concurrently, there must be a readiness to shift, if necessary, a patrol unit's area of concentrated patrol every week (or even more frequently) in order to keep pace with a shifting crime rate.

In Chapter VI we disussed an important aspect of monitoring crime patterns that should be kept in mind. We demonstrated that in allocating patrol effort it is not sufficient to look at just the total crime rates of the various regions. The crime rate of each region should be analyzed in order to breakdown the crime rate into more and less interceptible crimes, using duration and observability as the criteria.

On a larger scale, such as a precinct, a better allocation of patrol would have a higher payoff, increasing the probability of interception by more than 10%. The major problem is that often if there are differences in workloads the patrol unit with the least time to patrol is likely to be responsible for the highest crime area. This, of course, is a poor matching of patrol to crime levels. One alternative is to redesign the sectors to balance workloads or patrol coverage. However there are two alternatives which have a higher potential for allowing for concentrated allocations of patrol in the highest (precinct-wide) crime areas. They are overlapping sectors and a split patrol force. Their key asset is that they both have larger blocks of patrol effort that can be flexibly allocated to where it is most needed. However, both have the disadvantage of increasing the average travel time to a call for service. This increased travel time could be reduced with, for example, flexible dispatching algorithms (closest car dispatched) feasible with an automatic car locator system [3]. (See section 4.2.3.)

One last point we would like to make relates also to the issue of analyzing crimes with regard to their observability and observable duration. When considering different tactics for attacking a specific category of crimes (e.g. burglary), one component in the analysis should be an attempt at estimating the observable duration of the crime and the degree of observability. The quantification of these characteristics may affect the ultimate choice of tactics and will likely provide, at least, a good idea of how effective any of the strategies is likely to be.

8.3 Evaluating a Police Patrol Experiment

8.3.1 Using Models to Monitor Experimental Conditions

Larson in his review [5] of the Kansas City preventive patrol experiment discussed the need for models to be used in the design and in the monitoring of the proposed experimental conditions. He showed that in the reactive beats of the experiment (i.e. no routine preventive patrol) a significant visible presence was still provided by units responding to calls for service in those beats.

This result has important ramifications for the patrol strategies described in Chapter VII. The algorithm presented there tended to concentrate patrol in only a small section of a sector whenever crime rates were not uniformly distributed over the sector. One concern of ours in applying the model was how removing routine patrol from the rest of the sector would be perceived. The Kansas City experiment seems to indicate that even in areas not receiving routine patrol, it is possible for patrol units just responding to calls for service in those areas to maintain the earlier levels of visibility (e.g. by increasing the use of sirens).

Larson also demonstrated (using a simple model) that under the particular experimental design utlized in Kansas City one should not expect a marked increase in travel distance in the reactive beats. Of perhaps greater relevance, though, to this work, were Larson's findings on the levels of patrol in the various beats.

> "As shown by a simple mathematical model, even doubling or tripling of patrol effort--as was done in the proactive beats--does not adequately reflect routine levels of patrol experienced in other cities."

Consequently, the changes (or lack of change) in crime levels that occurred in the proactive areas are not necessarily representative of what would occur if preventive patrol were concentrated in the highest crime areas at levels suggested in Chapter VII.

8.3.2 Crime Statistics

The rest of this section will focus on a different aspect of evaluating a patrol experiment, the collection of appropriate crime data. The obvious guiding principle is to determine the relationship between the patrol force and each piece (or sample) of data. This principle should be applied to both crime and arrest statistics. It suggests that one appropriate question to ask a victim is whether or not the crime could have been observed by a passing patrol unit. Classifying the occurrence of crimes as inside or outside a structure (as was done in the Kansas City experiment) may be a good surrogate but does not directly answer the question. A drugstore robbery may be spotted by a passing patrol car (albeit with only a small probability) especially if the robber or robbers have to flee from the scene of the crime. On the other hand, a robbery in the middle of a housing project may occur in the open and still not be observable by a patrol car cruising the city streets. The standard, however, needs to be applied to all crime types and not just robbery. Was it a burglary of a fourth floor apartment or of a one family house? Did the commercial burglary occur on the 50th story of a skyscraper or did the criminals break into a warehouse, drive a ten ton truck up to the door, load it to capacity and drive away.

A direct corollary of the above suggestion is that the categorization of crimes go beyond just a yes-no description of observability. As our examples in Chapters VI and VII pointed out, it is important to be able to rate the level of observability. Would a passing car have had a clear or obstructed view of the crime? How long did the crime last? Was it at night or in the daytime? Crime patterns might be shifting in response to the patrol experiment from more to less observable crimes but that fact would not be apparent from the usual crime data that is collected.

An analogous breakdown of arrest data is also necessary. Was the arrest made by a passing patrol car which spotted the crime in progress? Or was it the result of a rapid response to a triggered burglar alarm? If the arrest were made not at the scene of the crime, was it still in anyway related to the speed with which the police responded to the original report of a crime?

Many of the issues we have raised here are not related solely to the problem of evaluating a patrol experiment. In essence the data we have described (summarized in Table 8.1) should be routinely gathered in order to assess accurately what the potential impact of patrol is. We do not, however, underestimate the difficulty in obtaining this data. In many crimes against property there may be no way of determining even the exact time of the crime, much less whether or not it was observable. However, even with all the inherent difficulties (and cost) we feel the potential payoff certainly justifies carrying out the collection and analysis of the above described data.

There is perhaps an even more significant problem in obtaining 'good' patrol related data than the difficulty and cost of gathering the potentially available data. Often the small sample size (*e.g.* not many crimes of a particular type) makes it extremely difficult to obtain good estimates and establish whether or not any significant changes have occurred as a result of the experiment. Our suggestion to break down further the crime categories will, however, tend to magnify this problem of small sample size.

8.4 Future Directions of Research

8.4.1 Introduction

Although in this last section we will be discussing extensions of the present work, many of the issues that will be raised are extremely fundamental and in some ways their resolution may be a prerequisite for applying our models. Many of the questions were originally asked by the Crime Commission in 1967 [6]. In our restatement of some of their points, we will focus on specific needs and how the resolution of a particular issue fits into the entire framework of developing effective police strategies. Perhaps the ability to be more specific and structure the problems is in part a measure of the progress that has been made since 1967.

8.4.2 Data Requirements

There is no doubt that significant work needs to be done in developing a crime data base that is oriented specifically toward the patrol question. Many departments have begun to analyze their crime data in order to determine the actual level of street crime that is observable and therefore potentially affected by patrol strategies. Thus a city like Atlanta, in analyzing its crime patterns, has found that 55% of its robberies occur on the street, while in contrast, 60% of its rapes occur in a dwelling [2]. However, as we have pointed out, a binary system of labeling crimes as observable or not is insufficient. Crucial questions that need to be answered are "For how long was the crime observable? During that time period how detectable was it?". These questions are not specific to a search theoretic model of patrol but rather strike at the heart of the question of what kind of an impact we reasonably expect a patrol force to have on a particular class of crimes.

Finding answers to these questions, however, will not be easy. Typically, no one will be standing by during the commission of a crime with a stopwatch and light meter to check on visibility. One source of data to answer these questions is the police themselves. The accumulated years of patrol experience are a starting point for obtaining estimates for these parameters which could be supplemented by interviewing victims and criminals. (Remember though there is a bias since the criminals that will be interviewed are the ones that were caught.) In addition, despite our opening remark, police departments sometimes do have motion pictures of specific types of crimes. Laboratory type experiments also can be performed to replicate different types of crimes in order to estimate the above parameters as was done in one experiment in Syracuse by Elliott [1]. No one method is necessarily going to yield very accurate estimates; however, by using all of the above approaches, it should be possible to obtain estimates that are sufficiently accurate to help answer deployment questions.

One interesting issue, besides the probability of interception by a passing patrol car, that the above data can be used to address, involves determining the relationship between response time and apprehending the criminal in the vicinity of the crime. A model of this phenomenon would seem to require the following:

1. What is the probability of an individual spotting the crime and then summoning the police?
2. How long between the observation and the telephone call is made?
3. At what point during the crime was it observed?
4. How long after the crime was observed will the crime last and/or the criminal be in the immediate vicinity?

Much of the data that we have suggested gathering can also be applied to answering these questions. This coupled with the more easily obtainable measures of the police response system can be used to model a major component of the relationship between apprehension and rapid response.

8.4.3 Expanded Patrol Model

The search theoretic models of patrol we have discussed have captured only one aspect of patrol, the patrol initiated probability of intercepting a crime. However, criminals can also be apprehended by a patrol force rapidly responding to information about a crime in progress. Thus, a first step in developing an expanded model of patrol effectiveness is to develop a model of patrol which incorporates both manners of intercepting a crime. Earlier in our discussion of overlapping patrol we had seen the need for such a two faceted model. Overlapping patrol increases the probability of police initiated interceptions but also increases the travel time. A model which could relate travel time to interception probabilities (citizen initiated) would make possible a detailed analysis of tradeoffs that arise in choosing between overlapping and non-overlapping patrol.

In the second stage this reactive model should be expanded to include a preemptive model which focuses on deterrence (assuming that some types of patrol can deter criminals). The model would, of course, require information about the relationship between both visible and plainclothes patrol and deterrence. Two other factors, however, that need to be included are how the probability of interception and police response time may also deter crimes. Thus the expanded model would include:

A. Probability of Interception
   1. Patrol initiated action
   2. Rapid response

B. Deterrence (as a function of)
   1. Visibility of the patrol force
   2. Probability of interception
   3. Rapid response

With this model (by no means easily developed) it then is possible to evaluate alternative strategies and develop mathematical techniques for finding strategies which are more effective overall.

8.4.4 Saturation and Displacement

One effect of major concern in deploying police is the displacement of crimes. Even if a strategy is effective in reducing one class of crimes in one particular region, the crimes may simply be displaced geographically (as in the New York City 20th precinct study [7]) or criminals may switch to different types of crimes. Consequently, experiments involving new police tactics sometimes contain controls for assessing as part of the evaluation whether or not there was a displacement effect. However, what we are suggesting here is that rather than attempt to measure this phenomenon by tacking on controls to other experiments this issue should be addressed directly. The reason for this different emphasis will become clear as we describe the goals of a possible experiment.

The experiment would have as its basic goal not just the determination of when displacement occurs or doesn't occur but also a detailed mapping out of a number of relationships.

1. How much displacement occurs and of what type if a patrol car passes a random point on the average every hour, half hour, fifteen minutes, etc.?
2. What is the time lag (and time decay) between implementation of a <u>specific</u> level of saturation patrol and displacement?
3. How many criminals are caught before and during the process of displacement until the system stabilizes?

In order to answer these questions an experiment would obviously have to be repeated a number of times for differing levels of saturation. This variability will, in general, be difficult to carry out if this is only part of a larger experiment whose major goal is to determine if a particular strategy is effective. Our discussion here should be interpreted as only an outline of the purpose and direction of the experiment. There are obviously a number of political considerations that will have to be dealt with in designing the actual experiments. Successful implementation of an experiment of this type would require that personnel at all levels (including patrolmen) be cognizant of the goals of the experiment and of their importance. Although this is good advice in any police experiment, it is especially crucial here. The experiment would have to involve committing a not insignificant number of patrol units to an experiment whose ultimate payoff is by no means immediate (or obvious) as its purpose is to develop a data base necessary for designing effective patrol strategies. This data base, for example, would be crucial to applying game theory to the deployment problem.

8.4.5 Concluding Remarks

In our discussion of some issues that need clarification, our focus has been on only a limited number of specific recommendations and almost exclusively on patrol. Our emphasis on patrol does not mean to suggest that other police tactics are ineffectual. However, before it is possible to make a valid comparison between patrol and stakeouts, decoys or investigative services, it is necessary to determine how effective patrol can be and under what circumstances it is most effective. Lastly, we have not even touched on the interaction between police and other components of the criminal justice system. This omission was made even though we realize that the nature and magnitude of the interaction between patrol and courts (or corrections) will sometimes depend on the type of patrol strategy. Patrol strategies which maximize deterrence and those which maximize apprehension will interact differently with courts and their relative impact on crime levels may well be a function of what the courts do. However until more is known about how court actions alone and in conjunction with police actions impact on crime levels, a more comprehensive approach to police deployment is not possible. Thus, for now, we have had to limit the discussion to an isolated analysis of police patrol. Implementation of models such as ours should, however, at least involve a qualitative analysis of the impact of the courts and, where possible, attempt to coordinate the activities of the courts with the actions of the police.

FOOTNOTES 8

1 The fact that the algorithm began with a particular atom does not necessarily mean that that atom will appear in in the final patrol area. Remember the algorithm allows for the removal of atoms if their removal will increase the total payoff.

REFERENCES 8

1. Elliott, J.F. and T.J. Sardino, Crime Control Team: An Experiment in Municipal Police Department Management and Operations, Charles C. Thomas, Publisher, Springfield, IL, 1971.

2. Kupersmith, G., High Impact Anti-Crime Program: Sample Impact Project Evaluation Components, U.S. Department of Justice, LEAA, NILE & CJ, U.S. Government Printing Office, Washington, D.C., July 1974.

3. Larson, R.C., Urban Police Patrol Analysis, MIT Press, Cambridge, MA, 1972.

4. Larson, R.C., "A Hypercube Queuing Model for Facility Location and Redistricting in Urban Emergency Services", Computers and Operations Research, Vol. 1, No. 1, March 1974.

5. Larson, R.C., "What Do We Know About Preventive Patrol? A Review of the Kansas City Preventive Patrol Experiment", published by Public Systems Evaluation, Inc., Cambridge, MA, July 1975, to appear in Journal of Criminal Justice.

6. President's Commission on Law Enforcement and Administration

of Justice, The Challenge of Crime in a Free Society, U.S. Government Printing Office, Washington, D.C., 1967.

7. Press, S.J., Some Effects of an Increase in Police Manpower in the 20th Precinct of New York City, New York City-Rand Institute Report R-704-NYC, October 1971.

GLOSSARY

Atom
A region or area within the city that is sufficiently small so that all spatial distributions over the region can be approximated to be uniform.

Beat
An area or region in which one patrol unit has preventive patrol responsibility. Same as sector.

Call for service
A communication to police originating from a citizen, an alarm system, a police officer, or other detector, reporting the need for on-scene police assistance.

Car locator system
A mthod or device which provides the dispatcher with improved estimates of the positions of available patrol units. This is to be distinguished from usual manual position estimation methods which usually entail guessing an available unit's position, using a center-of-mass criterion.

Center-of-mass
The point in a sector or an atom, respectively, which is the statstically average positon of the patrolling unit or the reported incidents, respectively.

Dispatch assignment
A directive by the dispatcher to a patrol unit assigning the unit to respond to the scene of a reported incident, or call for service.

Dispatch policy
A set of rules regarding the immediate assignment of patrol units to reported incidents. It specifies the conditions under which a reported incident of a particular priority from a particular location is entered into a queue of waiting incident reports or is handled immediately by an assigned patrol unit.

Dispatcher
An individual who has responsibility for assigning radio-dispatchable patrol units to reported incidents.

Dispatching strategy
Usually the component of the dispatch policy pertaining to distance estimation techniques.

Home sector
The sector in which a patrol unit is assigned to perform preventive patrol.

Intercept probability
The likelihood that a patrolling unit will intercept a crime while in progress

Intersector assignment
A dispatch assignment to a sector other than the unit's home sector.

MCM (Modified Center of Mass Dispatching Strategy)
If a call for service arises in atom j and the home sector car is unavailable, the dispatcher assigns the available unit with the minimum estimate travel time to atom j.

Overlapping sectors
Sectors that at least partially share common regions or areas.

Patrol allocation
The entire process of determining the total required number of patrol units, their spatial and temporal assignments, and rules governing their operation. Usually used here to describe just the spatial assignment.

Patrol deployment strategy
A set of rules specifying the spatial distribution of available patrol units, including sector and command design, patrol coverages, and repositioning.

Patrol frequency
The number of times per hour that a patrolling unit passes a particular point.

Patrol status
The condition of a patrol unit, particularly pertaining to dispatch availability. In some police departments the dispatch status of a patrol unit is restricted to one of two possibilities: available or unavailable; in others, finer distinctions are made, including such possibilities as meal break, auto maintenance, patrol initiated action, station-house, or type of incident currently being serviced.

Patrol unit
A footpatrolman; or an assigned pair of footpatrolment; or a patrol car, scooter, or wagon and its assigned police officer(s). Occasionally the term patrol car is used as a substitute for this more general term.

Precinct
An area or region comprising several sectors that is administratively distinct, usually having a station-house used as a base of operations. A patrol officer is usually assigned

to one precinct for a period of time. Dispatch assignments are nearly always intra-precinct assignments.

Preventive patrol
An activity undertaken by a patrol unit, in which the unit tours an area, with the officer(s) checking for crime hazards (for example, open doors and windows) and attempting to intercept any crimes while in progress.

Probability density function
A nonnegative function for which the probability that the corresponding random variable lies between x and x+ x (x small) is approximately equal to the function evaluated at x multiplied by x.

Queue
A waiting line, as of customers before a checkout counter or incident reports before a dispatcher.

Random patrol
A preventive and interceptive patrol in which the patrolling unit selects unpredictable patrol paths.

Right-angle distance
The sum of the total east-west and north-south distances between two points, given that the directions of travel are oriented east-west and north-south.

Search effort
The amount of time (man-hours) available to search for a target or targets.

Search theory
A body of literature that analyzes the problem of searching for targets. It includes models for calculating the probability of intercepting a target under various patrol strategies and for a range of target behaviors (e.g. stationary, moving). Part of the literature discusses the optimal allocation of search effort.

Sector
Same as beat.

Sector identity
A term applied to an officer's personal commitment to maintain public order and provide effective police service within his home sector.

Simulation
A method of replicating the operations of a system with a computer model that incorporates the same statistical behaviors as found in the actual system.

spatial distribution
The relative allotment of some quantity (for example, reported incidents) to each region of the city.

SCM (Strict Center of Mass Dispatching Strategy Strategy)
If a call for service arises in sector i and the home sector car is unavailable, the dispatcher assigns the available unit with the minimum estimated travel time to the center of mass of sector i.

Temporal distribution
The relative allotment of some quantity (for example, patrol unit) to each time of day.

Travel time
The time required for the dispatched patrol unit to travel to the scene of the reported incident.

Utilization factor
The fraction of time a patrol unit is unavailable to respond to dispatch requests. Sometimes it is assumed that a unit can only be unavailable because of call-servicing duties. Sometimes called utilization rate.

Workload
Some measure of the time spent by a patrol unit on a number of prescribed duties, particularly calls for service.

Most of this glossary appears in a book by Richard C. Larson (Urban Police Patrol Analysis, MIT Press, Cambridge, MA, 1972) and is reprinted with the author's permission.

APPENDIX A

AN ALGORITHM TO CHECK IF A SECTOR IS CONTIGUOUS

An Algorithm To Check If A Sector Is Contiguous

```
CONTIG: PROCEDURE OPTIONS(MAIN);
     DECLARE  DIMEN FLOAT BINARY;
     /* DIMEN: THE NUMBER OF ATOMS */
     GET LIST(DIMEN);
     BEGIN;
       DECLARE  A_A_TIG(DIMEN,DIMEN) FIXED BINARY (1,0),
               (LABEL(DIMEN),  SCAN(DIMEN)) FIXED BINARY(4,0),
                LIST1  FIXED DECIMAL;
       /* A_A_TIG: ATOM CONTIGUITY MATRIX */
       /* SCAN(K)=1: NODE K IS SCANNED    */
       /* LABEL(J)=1: NODE J IS LABELED   */
       /* LIST1: NUMBER OF SCANNED NODES PLUS ONE */
       GET LIST(A_A_TIG);
       SCAN=0;
       SCAN(1)=1;   LABEL(1)=1;   LIST1=2;
       DO K=1 TO DIMEN;
          IF SCAN(K)=0 THEN
            DO;
            PUT LIST('SECTOR IS NOT CONTIGUOUS');
            GO TO FINISH;
            END;
          I=SCAN(K);
          DO J=1 TO DIMEN;
             IF A_A_TIG(I,J)=1 & LABEL(J)¬=1 THEN
               DO;
               SCAN(LIST1)=J;
               LIST1=LIST1+1;
               IF LIST1 > DIMEN+.5 THEN GO TO FINE;
               LABEL(J)=1;
               END;
             END;  /* LOOP ON J */
           END;  /* LOOP ON K */
     FINE:PUT LIST ('SECTOR IS CONTIGUOUS');
     END;  /* BEGIN BLOCK */
FINISH: END CONTIG;
```

APPENDIX B

A TWO SERVER QUEUING SYSTEM

A TWO SERVER QUEUING SYSTEM

The problem addressed here is to calculate the probability of zero, (P_0), or one, (P_1), server being busy given that each server is busy an average of 100·b percent of the time.

Let a = arrival rate of calls for service

u = average service time for a call

P_i = the probability that there are i calls presently in the system, either being serviced or in queue

The following equations can then be written that describe the system in steady state.

$$u \cdot P_1 = a \cdot P_0$$

$$2u \cdot P_2 = a \cdot P_1$$

$$\vdots \qquad \vdots$$

$$2u \cdot P_i = a \cdot P_{i-1}$$

$$\vdots \qquad \vdots$$

Equivalently

$$P_1 = (a/u) \cdot P_0$$

$$P_2 = (a/2u) \cdot P_1 = (a/u) \cdot (a/2u) \cdot P_0$$

$$\vdots \qquad \vdots$$

$$P_i = (a/u) \cdot (a/2u)^{i-1} \cdot P_0$$

But $\sum_{i=0}^{\infty} P_i = 1$

$$1=P_0+(a/u)\cdot P_0+(a/u)\cdot(a.2u)\cdot P_0+\cdots+(a/u)\cdot(a/2u)^{i-1}\cdot P_0+\cdots$$

$$1=P_0 + (a/u)\cdot P_0/(1-a/2u)$$

$$1=(P_0 +(a/2u)\cdot P_0)/(1-a/2u)$$

However, the average utilization of each server wil be a/2u so that $b= a/2u$.

Consequently $P_0= (1-b)/(1+b)$

and $P_1=(a/u)\cdot P_0 = 2b\cdot P_0 = 2b\cdot(1-b)/(1+b)$.

APPENDIX C

OVERLAPPING SECTORS vs. NON-OVERLAPPING SECTORS

Overlapping Sectors vs. Non-overlapping Sectors

Overlapping Sectors

$$P_{io} = \frac{(1-b)}{(1+b)} \cdot [1-\exp(-2S \cdot T \cdot OB/M)] + \frac{2b(1-b)}{1+b} \cdot [1-\exp(S \cdot T \cdot OB/M)]$$

Non-overlapping Sectors

$$P_{in} = (1-b) \cdot [1-\exp(-2S \cdot T \cdot OB/M)]$$

$$P_{io}-P_{in} = \frac{1-b}{1+b} \cdot [1-\exp(-2S \cdot T \cdot OB/M)] + \frac{2b(1-b)}{1+b} \cdot [1-\exp(-S \cdot T \cdot OB/M)]$$

$$-(1-b) \cdot [1-\exp(-2S \cdot T \cdot OB/M)]$$

$$= \frac{b(1-b)}{1+b} \cdot [1+\exp(-2S \cdot T \cdot OB/M)] - 2\exp(-S \cdot T \cdot OB/M)$$

In order to prove that the above equation is always greater than zero (i.e. P_{io} is greater than P_{in}), it is only necessary to show that

$$1 + \exp(-2S \cdot T \cdot OB/M) - 2\exp(-S \cdot T \cdot OB/M) \geq 0$$

since b and (1-b) are always greater than zero.

By substituting 'X' for 'S·T·OB/M' in the inequality, we obtain

$$f(X) = 1 + \exp(-2X) - 2\exp(-X) \geq 0$$

To show that f(X) is always greater than or equal to zero, it is sufficient to prove that when X=0, f(X)=0 and that when X is greater than zero f(X) is monotonically increasing.

For X = 0

$$1-\exp(-2X)-2\exp(-X) = 1 + 1 - 2 = 0$$

In addition the derivative of f(X) is

$$f'(X) = -2\exp(-2X) + 2\exp(-X)$$

However, for X > 0 the following is always true:

$$\exp(-X) > \exp(-2X)$$

which means that the function monotonically increases. Consequently, the function, f(x), is always greater than or equal to zero. Q.E.D.

APPENDIX D

Average Level of Crime Increases as the Cycle Length Increases

Average Level of Crime Increases as the Cycle Length Increases

$$\bar{S} = (-C_1/2F\cdot X)\cdot(1-\exp(-F\cdot X)) + A/2F$$
$$- (C_2/2(F+I)\cdot X)\cdot(1-\exp(-(F+I)\cdot X)) + A/2(F+I)$$

with

$$C_1 = \frac{(-A\cdot I/F\cdot(F+I))\cdot(1-\exp(-(F+I)\cdot X))}{1-\exp(-(2F+I)\cdot X}$$

$$C_2 = \frac{(A\cdot I/F\cdot(F+I))\cdot(1-\exp(-F\cdot X))}{1-\exp(-(2F+I)\cdot X)}$$

Replacing C_1 and C_2 in the original equation yields

$$\bar{S} = (A\cdot I/2F(F+I)) \cdot \frac{(1-\exp(-F\cdot X))\cdot(1-\exp(-(F+I)\cdot X))}{X \cdot (1-\exp(-(2F+I)\cdot X}$$
$$\cdot \left[\frac{1}{F+I} - \frac{1}{F}\right] + A/2(F+I)) + A/2F$$

Since the expression $\left[\frac{1}{F+I} - \frac{1}{F}\right]$ is negative, and the expressions $A/2F\cdot(F+I)$, $A/2(F+I)$, and $A/2F$ are constants, then to prove that $\bar{S}$ is a monotonically increasing function of X, it is necessary and sufficient to show that H(X),

$$H(X) = \frac{(1-\exp(-F\cdot X))\cdot(1-\exp(-(F+I)\cdot X))}{X\cdot(1-\exp(-(2F+I)\cdot X)}$$

is a monotonically decreasing function of X.

Let $f(X) = \exp(-F \cdot X)$

$g(X) = \exp(-(F+I) \cdot X)$

Then H(X) can be written as

$$H(X) = \frac{(1-f)\cdot(1-g)}{X\cdot(1-fg)}$$

$$H'(X) = \frac{X(1-fg)(-f-f+fg'+gf') - (1-f)(1-g)(-Xfg'-Xgf'+(1-fg))}{(X(1-fg))^2}$$

Since the denominator is always positive, the problem reduces to proving that the numerator is negative

Let N(X) equal the numberator of H'(X)

$$N(X) = X(-f'g'+fg'+gf'+fgf'+fgg'-f^2gg'-fg^2f'+fg'+gf'-f^2g'-gff'$$
$$-gfg'-g^2f'+f^2gg'+fg^2f') - (1-f)(1-g)(1-fg)$$

$$N(X) = X(-f'-g'+2fg'+2gf'-f^2g'-g^2f') - (1-f)(1-g)(1-fg)$$

But $1-fg = (1-f) + (1-g) - (1-f)(1-g)$

Hence $(1-f)(1-g)(1-fg) = (1-f)^2(1-g) + (1-g)^2(1-f) - (1-f)^2(1-g)^2$

so

$$N(X) = (-Xf'-(1-f))(1-g)^2 + (-Xg'-(1-g))(1-f)^2 + (1-f)^2(1-g)^2$$

Let $(1-f)^2(1-g)^2 = \frac{(1-f)^2(1-g)^2}{2} + \frac{(1-f)^2(1-g)^2}{2}$

$$N(X) = \left(-Xf'-(1-f) + \frac{(1-f)^2}{2}\right)(1-g)^2 + \left(-Xg'-(1-g) + \frac{(1+g)^2}{2}\right)(1-g)^2$$

If each of the two terms is shown to be negative then the proof is complete.

$$\text{Let } u(X) = -Xf' - (1-f) + \frac{(1-f)^2}{2}$$

$$\text{Then } u'(X) = -Xf'' - f' + f' - f'(1-f) = -Xf' - f' + ff''$$

Replacing f with the original exp(-F·X) and combining terms yields

$$u'(X) = [F\exp(-F\cdot X)]\cdot[-F\cdot X + 1 - \exp(-F\cdot X)]$$

Now let $r(X) = (-F\cdot X + 1 - \exp(-F\cdot X))$

$$r'(X) = -F + F\exp(-F\cdot X) \leq 0$$

But $r(0) = 0 + 1 - 1 = 0$

which imples that $r(X) \leq 0$

Consequently, $u'(X)$ is a product of a positive, $F\exp(-F\cdot X)$, and a negative function, $r(X)$, which implies that

$$u'(X) \leq 0$$

But $u(0) = -0F'(0) - (1-F(0)) + \frac{(1-F(0))^2}{2} = 0$

As a result $u(X)$, which is also a decreasing function of X, must be less than or equal to zero.

Similarly $v(X) = -X\cdot g' - (1-g) + \frac{(1-g)^2}{2} \leq 0$

so $H'(X) = u(X)\,(1-g)^2 + v(X)\,(1-F)^2 \leq 0$

and $H(X)$ is a decreasing function of X. Q.E.D.

APPENDIX E

Average Level of Crime for Shortest Cycles

Average Level of Crime for Shortest Cycles

The problem is to find the limit of $\overline{S}$ as 2X approaches zero.

$$\lim_{2X\to 0} \overline{S} = -[C_1/2F\cdot X]\cdot[1-\exp(-F\cdot X)] + A/2F$$
$$- [C_2/2(F+I)\cdot X]\cdot[1-\exp(-(F+I)\cdot X)]$$
$$+ A/2(F+I)$$

with C_1 and C_2 as defined in equations (5.9) and(5.10). Replacing C_1 and C_2 into the original equation yields:

$$\overline{S} = [A\cdot I/2F\cdot(F+I)]\cdot \left[\frac{(1-\exp(-F\cdot X))\cdot(1-\exp(-(F+I)\cdot X))}{X\cdot(1-\exp(-(2F+I)\cdot X)}\right]$$
$$\cdot \left[\frac{1}{F+I} - \frac{1}{F}\right] + A/2(F+I) + A/2F$$

The limits of the numerator and denominator for the first term in the previous expression both approach 0 as '2X' approaches 0. It turns out that in order to calculate the limit of the first term, L'Hôpital's rule will have to be applied twice in succession.

$$\text{Let } N = [1-\exp(-F\cdot X)] \cdot [1-\exp(-(F+I)\cdot X)]$$
$$D = X\cdot[1-\exp(-(2F+I)\cdot X)]$$

The derivatives of the numerator, N, and denominator, D, are

$$N' = [1-\exp(-F\cdot X)]\cdot[F+I]\cdot[\exp(-(F+I)\cdot X)]$$
$$+ [1-\exp(-(F+I)\cdot X)]\cdot[F\exp(-F\cdot X)]$$

$$D' = X\cdot[2F+I]\cdot[\exp(-(2F+I)\cdot X)] + [1 - \exp(-(2F+I)\cdot X)]$$

However, once again the limits of n' and D' are zero. Taking the second derivative of each term generates the following two expressions.

$$N'' = -(F+I)^2\cdot\exp(-(F+I)\cdot X) + (2F+I)^2\cdot\exp(-(2F+I)\cdot X) - F^2\exp(-F\cdot X)$$

$$\lim_{2X\to 0} N'' = -(F+I)^2 + (2F+I) - F^2 = 2F(F+I)$$

$$D'' = (2F+I)\cdot[\exp(-(2F+I)\cdot X)]\cdot[-X\cdot(2F+I)+1] + (2F+I)\cdot\exp(-(2F+I)\cdot X)$$

$$\lim_{2X\to 0} D'' = 2(2F+I)$$

Therefore $\lim_{2X\to 0} N/D = F(F+I)/(2F+I)$

It is now possible to calculate the $\lim_{2X\to 0} \overline{S}$

$$\lim_{2X\to 0} \overline{S} = [A\cdot I/(2F\cdot(F+I))]\cdot[F(F+I)/(2F+I)]\cdot[-I/F(F+I)] + A/2(F+I) + A/2F$$

$$= 2A/(2F+I)$$

APPENDIX F

APPENDIX F

This appendix is intended to show that the expressions that were left out in going from equation (5.37) to (5.38) all approach zero faster than $1/(X+L)^2$ when X approaches infinity. First we write down the eliminated expressions. They are

$$R_1 = \frac{A\cdot I^2}{2(X+L)\cdot F^2\cdot(I+F)^2}\cdot \frac{[\exp(-2F\cdot(X+L) - I\cdot X)-1]\cdot[(2F+I)\cdot\exp(-2F\cdot(X+L) - I\cdot X)]}{(\exp(-2F\cdot(X+L) - I\cdot X) - 1)^2}$$

$$R_2 = \frac{A\cdot I^2}{2(X+L)\cdot F^2\cdot(I+F)^2}\cdot \frac{[2 - \exp(-(I+F)\cdot X) - \exp(-F\cdot(X+2l)]\cdot[(2F+I)\cdot(X+L) - I\cdot X)]}{(\exp(-2F\cdot(X+L) - I\cdot X) - 1)^2}$$

We will prove that R_1 approaches zero faster than $1/(X+L)^2$ and analogous proof applies to R_2. To prove this it is necessary to show that

$$\lim_{X\to\infty} R_1/(1/(X+L)^2)=0$$

which equals

$$\lim_{X\to\infty} \frac{K\cdot(X+L)^2}{(X+L)}\cdot\frac{[\exp(-2F\cdot(X+L)-I\cdot X)-1]\cdot[(I+F)\exp(-(I+F)X)+F\exp(-F(X+2L))]}{(\exp(-2F\cdot(X+L))-1)^2}$$

(K is a constant.) The denominator approaches 1 as X goes to infinity and the expression $[\exp(-2F\cdot(X+L)-I\cdot X)-1]$ approaches minus 1 and thus both components can be ignored. However as X goes to infinity the expression (X+L) also goes to infinity and the expression $[(I+F)\cdot\exp(-(I+F)\cdot X)+F\exp(-F(X+2L))]$ goes to zero. Thus the proof that R_1 goes to zero faster than $1/(X+L)^2$ goes to zero reduces to proving that

$$\lim_{X\to\infty} (X+L)\cdot[(I+F)\cdot\exp((-(I+F)\cdot X)+F\exp(-F(X+2L))] = 0$$

However, if it can be shown that in general

$$\lim_{X\to\infty} (X+L)\cdot K_1\exp(-K_2\cdot X-K_3) = 0$$

then the above also is true. This can be rewritten as

$$\lim_{X\to\infty} \frac{K_1(X+L)}{\exp(K_2\cdot X+K_3)}$$

The application of l'Hopital's Rule yields

$$\lim_{X\to\infty} \frac{K_1}{K_2\cdot\exp(K_2\cdot X+K_3)}$$

which obviously approaches zero as X goes to infinity.

Q.E.D.

APPENDIX G

A Proof that at most One of the Two Solutions is Positive

A Proof that at most One of the Two Solutions is Positive

In order to show that at most one of the two solutions is positive, we will sketch a proof that

$$\frac{(M-1)2R(R+1) - 2(2R+1)\sqrt{M}}{2(R+1)^2-2MR^2}$$

never generates a positive solution. The denominator will be negative when

$$M > \frac{(R+1)^2}{R^2}$$

and equal to zero when M equals that expression. However, what happens to the numerator at the turning point. Substituting that expression for M into the numerator yields

$$\left[\left(\frac{R+1}{R}\right)^2 - 1\right] \cdot 2R(R+1) - 2(2R+1)\sqrt{\left(\frac{R+1}{R}\right)^2}$$

$$= \frac{[(R+1)^2 - R^2] \cdot 2R(R+1)}{R^2} - \frac{2(2R+1)\cdot(R+1)}{R}$$

$$= \frac{2(R+1)\cdot(2R+1)}{R} - \frac{2(2R+1)\cdot(R+1)}{R} = 0$$

Thus at the point the denominator is zero the numerator is also zero. However as M increases above $\left(\frac{R+1}{R}\right)^2$ (i.e. the denominator is negative), it is clear the numerator has become positive. The first term is increasing faster (almost proportional to M) than the second term (which increases with the square root of M). Consequently when the numerator is positive the denominator is negative and vice versa.

BIOGRAPHY

Kenneth Chelst was born in Bronx, New York on April 4, 1948. He attended parochial school in the Bronx and was graduated from Yeshiva University High School, Manhattan, in June 1965.

He received his B.A. degree cum laude in Physics and Mathematics from Yeshiva University in 1970. As an undergraduate he spent a year of study abroad in Israel engaged in Talmudic research. He was later ordained as a rabbi from Rabbi Isaac Elchanan Theological Seminary, Yeshiva University in June 1972.

Along with these studies, he attended the New York University School of Engineering and Sciences where he was awarded an M.S. degree in Operations Research in June 1972.

As a doctoral student at the Massachusetts Institute of Technology, he served as a research assistant on a port planning project and thereafter on the Innovative Resource Planning Project. It is from the latter that his dissertation area has arisen. He has also served as a consultant to the New York City-Rand Institute where he was in charge of a technology transfer project with the New Haven Police Department. Mr. Chelst alsohas served as a part-time lecturer in mathematics at Babson College, Wellesley, Massachusetts.

In his last year at M.I.T., he was a recipient of the Harry J. Carmen Dissertation Fellowship of the Electrical Industry.

He is a member of the Operations Research Society of America and has been scheduled to present two papers at its fall 1975 convention.