METHODS FOR ALLOCATING URBAN EMERGENCY UNITS

Jan M. Chaiken and Richard C. Larson
The New York City-Rand Institute is a nonprofit research institution founded, as its Articles of Incorporation state, "... primarily to conduct programs of scientific research and study, and provide reports and recommendations, relevant to the operations, planning or administration of the City of New York." The Institute was established in 1969 by the City of New York and The Rand Corporation as a center for the continuing application of scientific and analytic techniques to problems of urban life and local government. It is governed by a Board of Trustees appointed jointly by the City and Rand.
METHODS FOR ALLOCATING URBAN EMERGENCY UNITS

Jan M. Chaiken and Richard C. Larson

This study was sponsored in part by the U. S. Department of Housing and Urban Development and in part by a National Science Foundation grant to The New York City—Rand Institute and The State University of New York at Stony Brook. Its contents, however, do not purport to represent the official views or policy of its sponsors.

THE NEW YORK CITY RAND INSTITUTE
PREFACE

This Report is a survey of current research on allocation of municipal emergency service systems, with the emphasis on police patrol cars and fire engines and ladders. It was performed as part of a study at the New York City-Rand Institute on the deployment of emergency services.

We hope that personnel of emergency service agencies, especially in the planning and research sections, will find this Report useful as a guide to potentially promising policy changes, and that researchers concerned with allocation problems will become better informed about related work by others which is described here.

A related report by Edward Blum [4] places this research topic in the broader context of the objectives which cities have in supplying services to the public, and discusses the pressures on emergency service systems which have generated the need for research and policy change related to allocation of their units.
SUMMARY

Many urban emergency service systems, including police departments, fire departments, and ambulance services, share common operational problems related to the allocation and dispatch of vehicles which respond to calls for service from the public. Recent studies at the New York City-Rand Institute and elsewhere have been directed at the use of quantitative methods for solving such allocation problems. This Report reviews the major research topics and describes those results which have produced, or are likely to produce, substantial improvements in system performance when implemented.

The aspects of allocation policy discussed are the following:

- Determining the number of units (vehicles) to have on duty in each geographical area at different times of the day or week.
- Selecting the unit(s) to respond to a particular incident.
- Determining the locations or patrol areas for the units on duty, and designing patrol coverage patterns.
- Deciding when units should be redeployed to improve service in areas where a large number of units are temporarily busy.

A determination of the number of units to have on duty, by time and place, can provide guidance in formulating the operating budget of an urban service system. A more common problem for agency administrators is to improve service by appropriately deploying a given total number of units. A variety of models have been developed for both of these purposes. Their effective use requires more and different kinds of data than are usually collected to implement traditional "rule-of-thumb" allocations. For instance, detailed estimates are required of the travel speeds of units and the rates of calls for service in various geographical regions. By using such models, an administrator can usually find alternative operating policies which
-vi-

- Reduce the average delay between receipt of an emergency call for service and the arrival of a unit at the scene;
- Reduce the disparity in workload among the units; and
- Enhance the amount of preventive patrol where needed, in the case of police patrol cars.

Additional research is required to convert estimated improvements in such measures to more fundamental indicators of the quality of service provided, such as the number of lives saved at fires and the number of crimes prevented.

Even after the total number of units to have on duty has been selected, service can often be improved by modifying the method by which particular units are chosen to respond to each call for service. For instance, it is usually possible to redesign the primary response areas of units in a manner which reduces travel times, relieves workload imbalances, or reduces the number of out-of-area responses. One study has shown that dispatching the closest unit(s) to an incident is not always the best strategy for minimizing average response time(s). Moreover, the design for response areas which minimizes average response time often relieves the hardest-working units of some responses as well, thereby reducing workload imbalance. Another model has been used to predict the amount of out-of-area responses. When applied to police patrol areas, it shows that the fraction of responses which take units outside their assigned areas is often large enough to call into question the desirability of selecting nonoverlapping patrol areas for police cars.

At the current time, the problem of where to locate units and their facilities has not been analyzed as carefully as the number of units needed, or their response areas. If only one new facility is needed, many properties of system operation appear to be quite insensitive to its exact location, which can therefore be selected by the use of simple models. If the problem is to determine which facilities should be occupied, utilizing information about the availability of units at the moment the decision is to be made, the desired solution appears to be so difficult to find exactly that even high-speed computers cannot make the calculations with sufficient rapidity to be
practical. Instead, a variety of algorithms for finding approximate solutions have been developed.

The allocation of crime-preventive patrol activities by police cars has been the subject of several analytical studies. One model, based on "search theory," provides guidelines for the frequency with which cars should patrol individual streets, in accordance with the crime rates which are experienced there. However, even with improved allocation of patrol effort, the probability of a patrol car intercepting a crime in progress has been found to be so small that there is some question whether the threat of apprehension provided by preventive patrol is actually great enough to deter crime.

Given the insights provided by analytical models, one can use large-scale simulation models to test detailed changes in allocation procedures. These models imitate the actual operations anticipated in a particular geographical area more accurately than analytical models, which are necessarily more simplified. Simulation models can predict the effect of new procedures on a wide range of performance variables: travel times to particular types of incidents, workload of units, delays experienced before a unit can be dispatched, extent of preventive patrol, availability of reserve units for major emergencies, and many others. They provide a consistent framework from which to study the effects of increased manpower, new technology, or new operating procedures. An agency administrator who has such a model available to him can compare the benefits of a number of proposed changes, as well as their costs and other considerations, before selecting among them.
ACKNOWLEDGMENTS

This work was supported in part by the U.S. Department of Housing and Urban Development under Contract Number H-1056 and in part by the National Science Foundation under Grant Number GI-5 to the State University of New York at Stony Brook and the New York City-Rand Institute. We wish to thank our colleagues at the New York City-Rand Institute, especially those whose work is referenced, and members of the Boston Police Department and the New York City Police and Fire Departments for their assistance to us during our discussions of the topics in this Report.
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREFACE</td>
<td>iii</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ix</td>
</tr>
<tr>
<td><strong>Chapter</strong></td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. DESCRIPTION OF AN URBAN EMERGENCY SERVICE SYSTEM</td>
<td>2</td>
</tr>
<tr>
<td>III. FACTORS INFLUENCING THE SELECTION OF AN ALLOCATION POLICY</td>
<td>5</td>
</tr>
<tr>
<td>IV. DETERMINING THE NUMBER OF UNITS TO HAVE ON DUTY IN EACH AREA</td>
<td>12</td>
</tr>
<tr>
<td>A. Methods Based on Geography and Land Use</td>
<td>12</td>
</tr>
<tr>
<td>B. Another Traditional Approach: Workload or Hazard Formulas</td>
<td>13</td>
</tr>
<tr>
<td>C. Analytical Methods: Queuing Models</td>
<td>16</td>
</tr>
<tr>
<td>D. Analytical Methods: Response-Time Models</td>
<td>22</td>
</tr>
<tr>
<td>E. Analytical Methods: General</td>
<td>25</td>
</tr>
<tr>
<td>V. DESIGN OF RESPONSE AREAS</td>
<td>29</td>
</tr>
<tr>
<td>Single-Sector Models</td>
<td>29</td>
</tr>
<tr>
<td>Influence of Intersector Cooperation</td>
<td>31</td>
</tr>
<tr>
<td>VI. LOCATING UNITS AND FACILITIES</td>
<td>36</td>
</tr>
<tr>
<td>VII. CRIME-PREVENTIVE PATROL</td>
<td>42</td>
</tr>
<tr>
<td>VIII. METHODS FOR EVALUATING PROPOSED CHANGES IN ALLOCATION PROCEDURES</td>
<td>47</td>
</tr>
<tr>
<td>IX. CONCLUSION</td>
<td>50</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>51</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

Urban police and fire departments, emergency ambulance services, and similar urban emergency service systems comprise an important class of governmental service agencies that until recently has not benefited from systematic analyses of operational problems. These systems operate in a complicated environment that includes temporally and spatially varying demand patterns; both explicit and implicit administrative, legal, and political constraints; and often ill-defined mixtures of objectives.

Our purpose in this Report is to review the operational problems of these agencies which are related to the deployment of their vehicles and to report current progress on the quantitative solution to these problems. The discussion focuses on the methods which are available, the extent of improvement that can be expected as a result of quantitative study, and the types of solutions that can be obtained. References are given so that the interested reader can pursue details which are not given here. Several of the discussed problems (e.g., determining how many units to have on duty) are common to many urban service systems we are considering, while others (e.g., allocation of preventive patrol effort) are experienced by only one agency.

Although we will not focus on the mathematical details of any of the methods, we hope to be sufficiently precise to bring out the following points:

- The performance of emergency service systems is often affected in counterintuitive ways by changes in procedures or deployment.
- Popular operational "rules of thumb" usually yield levels of performance which can be substantially improved by other methods.
- Simple changes in administrative procedure can often produce more significant improvements than expensive hardware systems or increases in manpower.
- The current state of research is sufficiently advanced in several areas so that many cities' emergency service systems could experience significant improvement simply by applying what is currently known.
- Additional research is needed in a number of areas.
II. DESCRIPTION OF AN URBAN EMERGENCY SERVICE SYSTEM

The class of urban emergency service systems we are considering is characterized by the following properties:

- Incidents occur in the city, giving rise to calls for service (or alarms); the times and places at which these incidents occur cannot be specifically predicted in advance.
- In response to each call, one or more emergency service units (vehicles) are dispatched to the scene of the incident.
- The rapidity with which the units arrive at the scene has some bearing on the actual or perceived quality of the service.

Examples of such emergency service units include fire engines and ladder trucks, police patrol cars and scooters, ambulances, emergency repair trucks for gas, electric and water services, and tow trucks.

Although all urban emergency service systems share the characteristics listed above, they may differ in three significant ways:

First, there are some emergency units that are ordinarily found at fixed locations at the time they are dispatched, while others are mobile. For example, police cars are typically patrolling a specified geographical area (patrol beat or sector) when they are dispatched to an incident.

This distinction is important for both administrative and analytical purposes. For instance, it is possible, in principle, to select the location, size, and shape of police patrol sectors at will, whereas the response areas of fire units must be designed in relation to the locations of the fire stations. Also, the dispatch strategy for mobile units can, in principle, be improved by a variety of techniques for estimating their locations; these are not needed for dispatching units with fixed locations. Among these techniques are the following: installing an automatic car-locator system, querying the cars as to their locations, and asking units that are near a stated location to identify themselves.

* These last two methods are often impractical using voice radio communications, since radio channels may be congested.
It is interesting to note that the distinction between mobile and fixed-location units begins to break down during periods of high demand. At such times the units may be dispatched directly from one incident to the next, or they may be dispatched while en route from a previous incident to their home location. This pattern is also common for emergency repair services, in which the driver may contact the dispatcher at the end of one service to determine where he should go next.

Second, emergency services differ in the urgency of the calls they receive and in their ability to discriminate among types of calls in advance. For example, false alarms of fire are not urgent at all, but it is difficult to know which particular alarms are false, even when records of previous false-alarm histories are maintained. On the other hand, a telephone call to the police reporting a past burglary can be identified as not requiring the immediate response of a patrol car. Some requests for service are generated by field personnel, as when a fire chief signals a second alarm \(^*\) at a fire, or a patrolman calls for assistance over his radio. Such calls can be immediately identified as reliable and of high priority.

The ability of an emergency service to distinguish the priority of its calls affects how it can react when many units are busy. If a call can be identified as not urgent, the dispatcher may decide not to send any units, or he may hold the call in queue to await the availability of a unit near the scene of the reported incident. He may even place a call in queue when some of his units are still available, so as to protect his ability to dispatch units to future high-priority incidents. However, if all alarms must be assumed to be urgent, none of them can be placed in queue for any substantial period of time. Then, if many units are unavailable, the dispatcher must either send fewer units to each alarm or dispatch units from greater distances.

Third, for some kinds of emergency services, the time the units spend between servicing calls is used for another important activity. For example, it is widely believed that routine patrol by police cars acts as a deterrent to certain types of crime \([33, 66]\). If police cars

\(^*\) A second or higher alarm is a request by the chief for the dispatch of additional units to the scene of the fire. The number and types of units dispatched to a second alarm are planned in advance.
spent nearly all of their time handling calls for service, the preventive patrol function would suffer. This kind of important secondary function is not present in all emergency service systems and should be distinguished from routine internal functions: rest, meals, and training for the men; maintenance of equipment; and the preparation of written reports.

For units which do have an important secondary function, questions involving the dispatch of units cannot be answered exclusively in terms of the expected effect on the service provided at emergencies. Thus, for example, it may be desirable to place some calls to the police in queue simply to preserve the deterrent patrol. A fire dispatcher would rarely have occasion to make such a decision, since the available fire units are not engaged in any activity which could be judged more important than responding to an alarm.
III. FACTORS INFLUENCING THE SELECTION OF AN ALLOCATION POLICY

The decisions made by dispatchers and field personnel that affect the response patterns of units are usually based on a collection of rules and procedures which constitute the allocation (or deployment) policy of the emergency service system. Essentially, the allocation policy determines the following properties of the system:

(1) The total number of units of each type on duty at any one time. (This may differ by time of day, day of the week, or season of the year.)
(2) The number of men assigned to each unit.
(3) The location or patrol area of each unit.
(4) The priority attached to different types of calls and the circumstances under which calls are queued.
(5) The number of units of each type dispatched to each reported incident.
(6) The particular units dispatched.
(7) The circumstances under which the assigned locations of units are changed. (This operation is variously called relocation, move-up, redeployment, repositioning or reinforcement.)
(8) When relocations are required: the number of units relocated, the particular units relocated, and their new locations.

This Report will focus on methods which can be used to select or improve the decisions made in relation to one or more of these components of the allocation policy. We should note, however, that many operational aspects of an emergency service system which are not part of the allocation policy may have a major influence on the quality of service provided. Among the most important are the assignments given to particular individuals; the procedures followed at the scene of an incident; and support functions -- maintenance, supply, training, and administration.
Since cities differ widely in their properties and in the demands they place on their emergency service systems, it is clearly impractical to try to specify a single "optimal" allocation policy which can apply to all locales. Nor does "optimal" have much meaning for systems having a large number of objectives, many of them ill-defined. But the theory underlying the allocation of emergency units is now sufficiently well developed that we can identify the quantitative factors which should be taken into account in selecting a policy and the methods which prove useful in analysis.

One critical influence on the selection of an allocation policy is the nature of the geographical area to be served. Gross measures of geographical properties, such as area or street-mileage, while interesting, are rarely adequate for allocation purposes. What is really required is an estimate of the time it will take for an emergency unit to travel from one point in the city to another or to patrol a specified region.

Thus, one needs to know the nature of the street network and the locations of impediments to travel, such as rivers, railroads, and parks. In addition, it is important to know the speed at which emergency vehicles travel. In locations where traffic congestion sometimes prevents rapid travel, it may be necessary to measure the travel speeds at different times of day, and on different days of the week. In view of the importance of travel time in selecting an allocation policy, it is noteworthy that the required data are rarely collected by emergency service agencies. The analyst ordinarily has to design an experiment to obtain this information.

The population density and land-use patterns of different parts of the city also affect the allocation policy. For fire departments, such information provides an estimate of the threat to life and property which would be associated with a fire in a particular location. It also indicates the kind of equipment which might be required to extinguish the fire. For police departments, such information would indicate the number of commercial establishments whose doors and windows have to be checked by patrol units, the locations where crowds may gather, etc.

The distribution of calls for service is a factor of great importance. Although it is not possible to predict the exact time and location of calls
for service, statistical analysis of past records can often provide an estimate of the numbers of calls of various types which can be expected from a given geographical region in a specified period of time (arrival rates) [48].

It is not unusual to find these arrival rates varying by as much as a factor of 20 at different times of day and by a factor of 100 from one location to another. Moreover, the fraction of the total number of calls which is of a particular type (e.g. outside crimes, false alarms) also varies substantially with time and place. The first two figures illustrate some of these phenomena for fire alarms in New York City. Figure 1 shows the average number of alarms which occurred in each hour of the day in 1968, together with the hourly alarm rate for a particularly busy and a particularly quiet day. Figure 2 shows the geographical distribution of these alarms for the Borough of Brooklyn in New York City, using an arbitrary collection of neighborhoods.

The times at which calls for service are received by an emergency system are found to have certain probabilistic properties which are common to all types of emergency services and to all cities. This makes it possible to develop methods of analyzing incidence data which can be applied with generality. After correcting the data for the obvious phenomenon that one major event can trigger many phone calls (or alarms from boxes), all referring to the same incident, the resulting arrival times are typically found to fall into a pattern which is known as a time-dependent Poisson process [50]. This fact is of considerable significance to the analyst, since it means that a single function of time is adequate to describe each arrival process, certain simplified models can be utilized, and the arrival process in a large region can be easily specified if one knows the arrival rates in smaller subregions.

Other important factors which affect the allocation policy are the number of units of each type actually required to handle each kind of incident, and the length of time each of them has to work at the incident.

*These times vary from incident to incident and must be treated as random variables which are called the service times.
Fig. 1 - Total fire alarms received in New York City by hour -- 1968 data
Fig. 2 - Total fire alarms received in 1968 in sub-areas of the Borough of Brooklyn.

NOTE: Total alarms includes false alarms and emergencies.
For example, certain kinds of fires can be extinguished in a minute or so by a single man with a portable fire extinguisher; others require an engine to pump water through a hose for about 5 minutes; and still others cannot be extinguished without a ladder truck as well as an engine and may require 30 minutes or more.

All the factors mentioned above will have some impact on the performance which can be expected from any particular allocation policy. But the most fundamental difficulty arises in trying to measure performance. How do we know when an allocation policy is good? Which ones are better than others? These questions typically have answers which are specific to the type of service system being considered, so that different methodologies have been developed for studying police patrol cars, ambulances, fire units, etc.

However, the various methods that have been developed can be categorized according to the allocation problems which they address, the extent to which they take into account the factors mentioned above, and the criteria they utilize for measuring performance. In the sections which follow, specific examples of each type of method will be discussed and analyzed.

There are basically two ways in which an agency administrator can use these methods. The first, and most common, application entails the selection of improved methods for assigning a predetermined total number of men. Even in cases where there is a possibility of hiring additional personnel, it is important to determine what level of performance can be achieved with existing manpower. The cost of even a single additional emergency unit is usually large enough to justify whatever analysis may be required to bring about the same performance level without the added unit.*

The second application, which has been recommended in planning and administration texts [33, 67], is to use quantitative methods for determining the overall number of men required to meet some prespecified objectives. Since personnel costs often constitute as much as 90 to 98 percent of the budget of an emergency service, this use of quantitative methods to derive

*Typically, a little more than 5 man-years are required annually to fill a single post around the clock. Thus, the manpower costs of operating a two-man patrol car in New York are approximately $120,000 per year, and the cost of operating a single fire engine with its complement of men now often exceeds $500,000 per year.
an allocation policy is virtually equivalent to determining the budget level of the service. However, this happens only on rare occasions. More typically, the administrator of an emergency service in a major city is faced with a total budget (or authorized strength) which he cannot change very much in one year.

A challenge in future years will be to implement both uses of these methods, with current systems performing effectively under a given budget level and required future budget levels determined from analyses of predicted demands and models of operation.
IV. DETERMINING THE NUMBER OF UNITS TO HAVE ON DUTY IN EACH AREA

A. METHODS BASED ON GEOGRAPHY AND LAND USE

The present locations of a city's fixed facilities, such as fire houses and police precinct stations, were determined for the most part by the manner in which the city grew from a small community to its present size. As each new area of the city became populated, it was necessary to add facilities. But when the geographical growth stopped, it may have appeared that there were enough facilities, and further construction would take place primarily for the purpose of replacing a deteriorating physical plant.

Usually the administrator of an emergency service has no opportunity to consider the possibility of relocating more than a few of these facilities, although in rare instances, such as the case in which Hogg [31] studied potential locations for 19 fire stations, planners may undertake to design or redesign an entire city. As long as the emergency vehicles are located at a specific set of facilities and are dispatched from there (whether this is absolutely necessary or not), it will not appear to an administrator that he has any flexibility in selecting the locations of his units. Thus, part of his allocation policy is determined by the geography of existing facility locations.

This reliance on geographical factors is often reinforced by the establishment of standards or regulations which apply to many cities. For example, the Standard Grading Schedule of the American Insurance Association [2] is used in most U.S. cities* to establish fire insurance rates. As a rule, cities will attempt to meet as many standards in the schedule as possible, so as not to have a lower rating than necessary. But for cities with population over 200,000, the only criteria provided by the schedule for the number of fire engines and ladders to be located in each part of the city are based exclusively on geography and land use.

*Not including New York City.
For certain parts of the city (called "high-value districts") the Standard Grading Schedule requires every point to be no further than 1 mile from an engine company and no further than 1.25 miles from a ladder company. Moreover, within 1.5 miles of any point there must be at least 3 engine companies, and within 2 miles at least 2 ladders. These standards are slightly more stringent for areas which may require substantial water flow for firefighting and are less stringent for residential areas; but for each type of area, the same kind of geographical standard applies.

Allocation methods which rely exclusively on such geographical standards must be regarded as inadequate, since each city has its own particular problems which cannot be taken into account in a standard which applies to all cities. The main deficiency of geographical standards is that they are meant to be substitutes for standards involving the time between receipt of a call for service and the arrival of the units. But this response time depends on many factors aside from geographical ones: the delays incurred in dispatching the units, the speed at which the units can travel, and the probability that particular units will be available. (It is little comfort to know that a fire house is within a mile of your home if the units located at that house would very likely be busy at the time you had a fire.)

Thus, as a general rule, it is not possible to determine whether an adequate number of units are located in each geographical area solely by inspecting a map of the city which shows the home location of each unit. Alternative methods for analyzing this problem will be considered in later sections.

B. ANOTHER TRADITIONAL APPROACH: WORKLOAD OR HAZARD FORMULAS

At the opposite extreme from methods which rely on a single factor such as geography are those which attempt to combine, in some intuitive or judgmental manner, virtually all factors which might be relevant for allocating personnel. These methods are commonly based on workload or hazard formulas, and they give the appearance of accuracy because of the large number of factors taken into account.
Perhaps the most well-known formula was developed for police
use by O. W. Wilson [65, 66, 67]. Wilson combined indicators of activity
(such as number of arrests, number of calls for service of particular
types, number of accidents, and number of doors and windows to be checked)
with other factors (such as number of street miles, number of licensed
premises, and number of crimes) to arrive at a "hazard score" for each
area. Suppose we denote by $f_{ij}$ the amount of factor $j$ associated with
area $i$ ($i = 1, 2, \ldots, I = \text{number of areas}; j = 1, 2, \ldots, J = \text{number}
of factors$). Then the city-wide amount of factor $j$ is

$$F_j = f_{1j} + f_{2j} + \ldots + f_{Ij}.$$  

For instance, $F_2$ may be the total number of street miles in the city and
$f_{42}$ may be the total street miles in police precinct 4. To arrive at
the hazard score for each area, an arbitrary importance weighting $w_j$
is assigned to factor $j$. For example, arrests might have a weight of 1,
FBI Part II crimes a weight of 2, and FBI Part I crimes a weight of 4.
This means that each reported Part I crime is judged to be four times
as important (in the sense of its hazard) as an arrest. The hazard
score for area $i$ is then calculated as

$$H_i = w_1 \frac{f_{1i}}{F_1} + w_2 \frac{f_{12}}{F_2} + \ldots + w_J \frac{f_{IJ}}{F_J}.$$  

The method then specifies that the total number of men (or patrol
cars) are to be distributed among the areas in direct proportion to
their hazard scores. Thus, if the hazard score for area 1 is 7 percent
of the total hazard score for the city, then 7 percent of the total
number of patrol units on duty should be assigned to area 1.

This procedure often produces unsatisfactory allocations that may
have to be "juggled" by hand computations to arrive at a reasonable al-
location. The inherently linear form of a hazard formula precludes
description of the highly nonlinear and complex interactions among
system components which are often observed in practice. Such a formula
also attempts a simple deterministic depiction of a system in which
many of the variables are probabilistic. The major difficulty, however,
arises in trying to determine how to improve the selection of the subjective weightings, given information about the performance of the system.

As an illustration of such a difficulty, consider the weighting given to the number of calls for police service that require the dispatch of a patrol unit. In practice this weighting may be as small as 5 percent, in which case the insensitivity of the resulting allocations to large variations in call-for-service rates results in large queues of call reports forming at predictable periods in certain areas of the city. These queues would be diminished somewhat by (1) giving the corresponding weighting a greater numerical value or (2) allocating units in accordance with separate calculations of the formula for short time periods (e.g. using hourly data in the formula rather than annual data). But such a change may affect in unknown, and possibly undesirable, ways other properties of the system, such as effectiveness of detecting crimes in progress.

Thus, the general problems associated with a hazard formula persist, not only because of its mathematical shortcomings, but in addition because such a formula does not relate meaningful measures of system effectiveness to operational policies. That is, a hazard formula, comprising a subjectively weighted sum of quantities, does not provide an administrator with such policy-relevant information as the response time of his units, the probability that a patrol car will intercept a crime in progress, or any of a number of other operational quantities. An administrator who is considering a possible alternative strategy or allocation level would like to know its probable effect on such measures of effectiveness.

In some instances, a hazard formula may have the perverse effect of indicating a need for additional personnel in areas which already are relatively overallocated. This can arise because factors such as number of arrests and reported crimes depend on the number of personnel currently allocated to an area. For example, crime suspects are more likely to be apprehended in an area which is sufficiently staffed than in one with saturated resources. The hazard score of the former area
would be artificially inflated by the high arrest rate of that area, thereby "substantiating" the need for additional personnel. The real need for personnel would be in the saturated area, but a hazard formula would not fully reveal this need.

Wilson should be credited with introducing quantitative methods into a policy area which previously had depended on "command judgment" alone, but recent developments in the theory of allocating emergency units have made reliance on a hazard formula unnecessary.

C. ANALYTICAL METHODS: QUEUING MODELS

A number of useful approaches to determining the number of units to have on duty have been based on queuing theory, which is a field of operations research concerned with the performance of service systems where customers, ships, telephone calls, etc., may have to wait "in line" until they receive the desired service. In applications of queuing theory to urban emergency service systems, the "customers" are the alarms, or calls requiring dispatch of a vehicle, and the consequences of having to place such a call in queue may be quite serious. Indeed, it is characteristic of emergency systems that a person's life or well-being may well depend on the immediate dispatch of a unit.

Thus, a primary objective of all urban emergency systems is to reduce to a low level the possibility that a call which requires the immediate dispatch of a vehicle will have to be placed in queue for more than a few seconds. Should a situation arise where nearly all calls must be placed in queue for 20 or 30 minutes, then the precise geographical distribution of the units (designed to keep travel time down to a few minutes) is largely irrelevant.

The nature of the arrival times and service times of calls is such that one can never guarantee that every call will be able to obtain the immediate dispatch of a unit (except if the number of emergency vehicles is infinite—or, as a practical approximation, much too large for the budgets of most cities). Thus, the objective of a queuing analysis of an emergency service must be to assure that the probability of an important call encountering a queue is below some specified threshold (such as 1 in 50)
or, what is more or less equivalent, that the average time to wait in queue is below some specified limit (such as 1 minute).

The difficulty of the analysis depends on how many different types of calls are received and on the circumstances under which they may be placed in queue. To take a simple hypothetical example, we might imagine a city in which each police patrol car is assigned a geographical response area ("beat") in such a way that no other car responds to calls in his area. Then, whenever a given patrol car is busy, all calls from his response area would have to be placed in queue. Such an arrangement constitutes a "single-server queuing system," and, given reasonable assumptions, standard textbooks provide formulas which give good estimates of the average time a call will wait in queue or the probability that it will experience a queue [49, 51]. One way of using these formulas to determine how many patrol units are needed would be as follows: A threshold would be selected for the maximum value of the expected waiting time (or probability of a queue) to be permitted in any beat, and beats would be selected small enough to assure that the threshold is not exceeded. The total number of beats designed in this way would then determine how many units are needed.

Although this example is instructive, no urban emergency service system actually operates in this manner, because other arrangements lead to fewer delays with the same number of units. The simplest generalization of this model which has been usefully applied to real emergency services is the following: There are a fixed number N of vehicles, perhaps all located at one place, such as a hospital, and each call requires the dispatch of one vehicle. A call is placed in queue only when all N vehicles are busy servicing prior calls. All calls are assumed to be identical in terms of their importance and service time.* With certain additional assumptions, the formula for the probability of a queue is known in this case [51].

*The distribution of the time required to service a call is the same for all calls.
If the calls are assumed to arrive according to a Poisson process at an average rate of \( \lambda \) calls per hour, say, and the service-time distribution is exponential with mean \( \frac{1}{\mu} \) hours (meaning that the probability of service lasting longer than \( t \) hours is \( \exp(-\mu t) \)), the probability of a call arriving when all \( N \) units are busy eventually approaches

\[
p(N) = p_0 \frac{r^N}{N!},
\]

where

\[
p_0 = \frac{1}{\sum_{n=0}^{N-1} \frac{r^n}{n!} + \frac{r^N}{N!(1 - \frac{r}{N})}},
\]

and

\[r = \frac{\lambda}{\mu}\] (assumed less than \( N \)).

Although the assumption of exponential service times rarely proves to be valid, these formulas may nonetheless provide useful approximations.

Such a model is a good description of the operation of ambulance services in many cities, as discussed by Stevenson [59]. Stevenson has applied this model to determining how large the number \( N \) of vehicles must be (depending on the arrival rate of calls) to assure that only 1 in 100 (or 1 in 20, or 1 in 10) callers must wait for the dispatch of an ambulance. Given an estimate of the arrival rate of calls for service during each time period, the administrator can select a desired threshold probability and determine how many ambulances to have on duty at various times.

The results of Stevenson's calculations have a property which is common to nearly all realistic queuing models: The number of units needed increases with the alarm rate, but not in direct proportion to the alarm rate. Thus, a doubling of the alarm rate would produce a requirement for fewer than twice as many units. This observation constitutes an additional argument against using call-for-service rates in a linear manner in workload formulas.

The identical model has been used in St. Louis for the allocation of police patrol cars [48]. The city is divided into nine patrol districts,
and a call is assumed to enter a queue whenever all the cars in its district are busy. For each 4-hour time period the Police Department estimates, using the N-server queuing model, how many cars will be needed so that at most 15 percent of each district's calls will experience a queuing delay.

The results of this queuing calculation are not the sole basis on which cars are assigned to districts, since St. Louis also has a "preventive patrol" force which does not respond to calls unless they have a very high priority. However, the use of queuing theory is an essential component of the allocation policy of the St. Louis Police Department.

The next step in complexity of queuing models involves the recognition that there are different types of calls. Each type may be characterized by its priority in relationship to other calls and by the length of time required to service the calls of that type. An example of a model which assumes that high priority calls are served first, but retains the assumptions that one unit responds to each call and that all service times have the same exponential distribution, was developed by Cobham [20]. Although in most cities calls are not explicitly assigned priorities according to specified rules, Larson has found this model useful as an approximation to current performance of police dispatchers and as a tool for analysis of the potential benefits of more precise priority schemes [43]. It has the advantage that it places emphasis on reducing the delays which are associated with important calls.

Although even this model is a rather crude description of most police response systems, the improvements in accuracy which can be achieved by time-consuming refinement of the model can only be justified if they will lead to improved allocation decisions. For example, greater realism would be introduced by such refinements as (1) permitting each priority level of calls to consist of mixtures of types, each requiring a different length of time for service, or (2) allowing the service time to vary with the number of units busy. But the effort required to design such models should not be made unless existing models are shown to be inaccurate in a way that affects decisions and unless a comparable effort will be devoted to collecting and analyzing service-time data.
The dependence of service times on the number of busy units is characteristic of most urban emergency systems, but it is difficult to measure quantitatively. One cause of the variation, which can be estimated at least roughly [44], is the increase in average travel time which occurs when distant units must be dispatched to calls. More important, however, is the fact that an incident may escalate when units do not arrive promptly. A small fire may become much larger and require a longer time to extinguish, or a reported marital dispute may turn into an assault before a patrol car arrives. Available data are rarely adequate to model these phenomena [34].

One refinement of the N-server queuing model has been found practical, and indeed necessary, for describing the number of units busy at the operations of a fire department. Fire departments typically dispatch several units to each alarm, and these units originally have distinct duties to perform, while the previous models assume that one unit is sent to each incident. In addition, fire units do not all complete service at the same time. Instead, some units may leave the fire scene when the fire is under control, while others will remain until extinguishment, and still others will continue to work after extinguishment on some duties known as "overhaul."

Chaiken [17] has developed a queuing model which allows for these features of fire operations. In particular, this model has the following characteristics:

- Different types of alarms may require different numbers of units of various kinds.
- The units may arrive singly, or in groups, and they may depart in similar fashion.
- The length of time the units are busy at the incident depends on the type of incident.

This model assumes that queues are never permitted to develop. Instead, whenever units are required in one region of the city, it is assumed they will be dispatched from there or from another region, if necessary. In applying this model, one does not try to assure that the probability
of encountering a queue is small. Instead, one requires that the probability of the necessity to dispatch units from a distant region is small.

Under quite general assumptions (which allow, in particular, nonexponential service time distributions), Chaiken [17] has shown that the probability of finding *n* fire units busy in a region approaches a limit \( P(n) \) which can be calculated as follows:

1. From data giving the time of dispatch and time of service completion for the various units at each incident, determine the average time \( t_j \) that exactly \( j \) units spend working together at an incident, for \( j = 1, 2, \ldots \). (This is an average over all incidents; for example, if half the incidents do not use 3 units at all, and the other half use 3 units together for 1 hour, then \( t_3 = 1/2 \) hour.)

2. Let \( \rho_1(j) = \lambda t_j, \ j \geq 1, \)

where \( \lambda \) is the arrival rate for all incidents, and calculate iteratively the following convolutions:

\[
\rho_2(j) = \sum_{m=1}^{j-1} \rho_1(j-m)\rho_1(m), \quad j \geq 2,
\]

\[
\rho_3(j) = \sum_{m=2}^{j-1} \rho_1(j-m)\rho_2(m), \quad j \geq 3,
\]

\[\vdots\]

\[
\rho_k(j) = \sum_{m=k-1}^{j-1} \rho_1(j-m)\rho_{k-1}(m), \quad j \geq k.
\]

3. Then the probability that no units are busy is

\[
P(0) = \exp(-\sum_{j=1}^{\infty} \rho_1(j)),
\]
and the probability that \( n \) units are busy (\( n \geq 1 \)) is

\[
P(n) = P(0) \sum_{k=1}^{n} \frac{\rho_k(n)}{k!}.
\]

Given estimates of the arrival rates and average service times, these calculations are readily carried out on a computer.

The details of an application of this model to the allocation of fire units in New York City is given in a report by Chaiken [16]. A general description of the results is that at low alarm rates, such as occur in the early hours of the morning, the numbers of units needed to meet the requirements of the queuing model are well below the numbers needed to meet geographical requirements such as described in Section A of this chapter; therefore, the geographical requirements predominate. However, in some parts of the city, at times of high alarm rate, the queuing model implies a need for more units than would be suggested by geography alone. The same model could be utilized for analyzing allocations in police departments which dispatch two or more cars to certain types of incidents.

**D. ANALYTICAL METHODS: RESPONSE-TIME MODELS**

The simple analytical models described in Section C cannot, by themselves, determine the number of units needed, because of the following problems:

1. Only at high alarm rates are queuing considerations as important as issues like geographical requirements, so that models based on queuing theory do not provide guidance for allocations when and where alarm rates are low.
2. For services which engage in important functions other than response to calls, these functions have not been taken into account.
3. The specific criteria chosen (probability of encountering a queue, average waiting time in queue) do not have a clear relationship to the true performance of the system.
The first two difficulties can only be handled by utilizing additional performance criteria.* To resolve the third difficulty, one would like to know the actual benefits which accrue by virtue of decreasing the delay until arrival of a unit. Such benefits might be stolen goods recovered, lives saved, property damage averted, etc. Although preliminary research along these lines has been performed [11, 27, 34], the currently available empirical information is not an adequate foundation for an administrator's use in selecting allocation policy.

Thus, at present, one is forced to use available performance measures such as response times. A careful and realistic use of such measures can provide reliable proxies for more fundamental measures, as has been discussed by Carter and Ignall [12].

The total response time of an emergency unit is generally defined to be the duration of the period between the initiation of a telephone call or an alarm and the arrival of the unit at the scene of the incident. Thus, the response time includes travel time as well as a number of queuing, communications and processing delays. (In some cases it may even be appropriate to include in the response time the period between the beginning of an incident and the transmission of an alarm, or initiation of a call.)

In general, one is interested in the average response time to all important incidents, or the probability that the response time will exceed a certain threshold, such as 5 minutes. Moreover, the response times of several different units may be of concern. Thus, formulas are required relating specified measures of response time to the number of units on duty, the nature of the geographical region to be served, arrival rates of alarms, and service times at incidents. The models used to calculate these formulas may have to take into account the probabilities of particular combinations of units being available, the dispatching rules, and a variety of other details of system operation.

*For a general discussion of criteria which are appropriate for police patrol, see the report of Kakalik and Wildhorn [36].
Several models show that average travel time is inversely proportional to the square root of the number of available units, with the proportionality constant dependent on the travel speeds and locations of the units. If \( N \) units are available and are randomly patrolling a region of area \( A \), a simplified model [43, p. 323] shows that the average travel time for a single unit dispatched to an incident is approximately \( 0.63\sqrt{A/N}/v \), where \( v \) is the response speed of the unit. If the units are not patrolling, but instead are located in such a way as to minimize average travel time [46], the constant 0.63 is replaced by 0.47. Similar formulas can be found for the average travel time of the second arriving unit when two are dispatched, etc. Some of the models used to calculate these formulas also provide estimates of the probability that travel time will exceed any specified threshold [40, 43].

These results cannot be applied directly to the problem of determining how many units to have on duty, since the number of units actually available to respond to a call will ordinarily be smaller than the number on duty. However, using the probability that \( n \) units will be busy, which is calculated according to the methods described in Section C, one can estimate the expected travel time, given that \( N \) units are on duty. Larson [43, p. 328] provides such an estimate for the case in which each available unit patrols the entire region randomly. It is similarly possible to calculate the expected travel time under the assumption that the available units are moved, if necessary, so as to occupy the locations which give minimum travel time.

Kolesar [40] has studied the numerical output from these models, as well as data derived from experiments and from more complex models, and he has concluded that a reasonable approximation in cases where unavailabilities are not too severe is to assume that the average travel time is inversely proportional to the square root of the average number of available units. This result leads to a method with the following steps for determining how many units to have on duty in a specific region:

1. Determine the constant of proportionality between average travel time and the inverse of the square root of the
average number of units on duty, using data collected in the region.

(2) Estimate the arrival rates for calls in the region.

(3) For each number M which is a candidate for the number of units to have on duty, use queuing models to calculate the probability of a queue* and the average number of units expected to be available.

(4) Estimate the average travel time, using the results of steps (1) and (3).

(5) Select the number of units on duty to be at least large enough to assure that the probability of a queue does not exceed a specified threshold, and add enough units to assure that the average travel time does not exceed another specified threshold.

This method provides a good approximation to the desired allocation for each region. These suggested allocations can then be tested with more accurate models (see Chapter VIII).

E. ANALYTICAL METHODS: GENERAL

A number of analytical models have been designed to take into account two or more performance criteria at once. Larson [43] has developed a dynamic programming model for allocation of patrol cars to precincts which incorporates a queuing model, a somewhat complex travel-time model, and a model of the frequency with which cars pass by an arbitrary point while on preventive patrol. In addition, it is possible to include restrictions on the smallest number of units which can be assigned to any one command and a variety of other criteria which may be supplied by police administrators.

With this allocation procedure, most of the individual policy objectives are stated in terms of constraints. For instance, for a particular command it may be decided that the average travel time should not exceed

* Or the probability that distant units will be needed.
4 minutes. Then, the algorithm supplies the command with enough patrol units so that this objective and all other policy objectives (constraints) are satisfied. Once the constraints are met, the queuing delay is treated as a variable to be minimized using whatever additional patrol units are available.

With limited police resources, it is possible that a specific set of policy objectives is unobtainable. If so, the algorithm indicates the additional number of patrol units required to meet the stated objectives. To allocate the number of units currently available, the algorithm then requires a more modest set of objectives.

Compared to the allocations derived from a hazard formula, the algorithm-derived allocations appear to reflect more fully the operating characteristics of the system. For instance, in one large city the results suggest that during periods of relative congestion (e.g., Friday and Saturday evenings), average dispatcher queuing delay can be decreased significantly by diminishing resources in residential commands with relatively light demands and increasing resources in "core-area" commands which are heavily loaded. Such a redeployment of resources does not noticeably degrade performance in residential commands, since a sufficient number of patrol units are retained to satisfy all policy constraints. Yet, average queuing delays in core areas can be reduced often from 30 minutes to less than 2 or 3 minutes.

Although such a finding may not be surprising, the calculation of the optimal reallocation would be extremely tedious without the assistance of a computer algorithm which can compute the effects of each alternative and quickly discard "bad" ones. And, without models of patrol activity, it would not be possible to predict whether each alternative allocation satisfies the policy constraints and reduces the delay at the dispatcher's position in the best possible way.

It is natural to assume that when units are added to a command, the number of responses made by each of its units will decrease, and this may be one of the secondary benefits of adding units which is of particular interest to an agency administrator. However, in instances where two or more units are ordinarily dispatched to each incident, but fewer are sent when
some of them are unavailable, this benefit is not nearly as large as one might expect, and, in fact, it may be almost nonexistent.

Carter and Ignall [13] developed a somewhat complex N-server queuing model for determining the extent to which an added fire unit actually provides relief to overworked units in its area. They showed that the predominant effect of an added unit in a high-alarm area is to "fill out" the response to nearby alarms. This means, for example, that if two ladder trucks are supposed to be dispatched to each incident, the new unit will increase the chances that two will actually be available for dispatch, rather than just one. Thus, if it is desired to reduce the workloads of units in addition to improving the response time, a greater number of units may be needed than is suggested by the simpler models.

In general, the quality of the allocations which an administrator can expect from any of the analytical models described here depends on how much effort he is willing to have his staff devote to the application. An analyst who is not a member of the concerned agency cannot make an appropriate determination of what constitutes an "excessively long" delay before the arrival of a unit, how much preventive patrol will be considered adequate, or what level of workload is "too great."

In the case of fire departments, where the various units dispatched to a single incident may arrive at different times, the analyst is not even in a position to know which arrival patterns are "better" than others. However, the field chiefs, who are completely familiar with their department's operating procedures at a fire, can provide valuable information. Through asking a series of questions such as, "Would you prefer two fire engines arriving 1.5 minutes after an alarm, or one arriving at the 1-minute mark and one at the 2-minute mark?", it is possible to derive a chief's utility function for arrival times. These, of course, will vary according to the chief chosen to respond to the questions and the type of incident he has in mind. Keeney [37] has determined the algebraic form that the utility function takes, and, through extensive interviews with one chief, has calculated the function which expresses his preferences for the response times of engines and ladders. This process differs substantially from showing the chief the form of the equation and asking him to supply the parameters (as one might do
with a workload formula), because the operational consequences of his answers are understood by the chief. Future research along these lines should make it possible to select the allocation of units to maximize the utility of the resulting patterns of arrival times, as interpreted by the chief.

One final comment: In regard to any of these methods for determining how many units to have on duty, it should be noted that there may be some difficulty in assigning individuals to shifts or tours of duty which best "fit" the desired assignments. Legal and administrative constraints can make this problem quite difficult. A heuristic approach is discussed by Edie [24]. A more general approach using a computer algorithm has recently been reported by Heller [29].
V. DESIGN OF RESPONSE AREAS

A problem commonly shared by all spatially distributed urban systems is the design of areas (districts or sectors or beats) that indicate where a particular patrol unit, fire engine, or ambulance is to have primary responsibility. In designing these administrative areas, agency administrators have stated several diverse (often conflicting) objectives:

- Minimization of response time.
- Equalization of workload.
- Demographic homogeneity of each area.
- Administrative convenience.

It is doubtful if any one mathematical technique for design of districts will fully take into account all the relevant factors. Yet, even some of the more simple, recently developed models have provided more insight into the problem than was previously available with ad hoc "rules of thumb" and, in fact, have shown several such rules to be invalid in most cases.

SINGLE-SECTOR MODELS

Traditionally, police planners have been instructed to design patrol sectors as squares, circles, or as straight lines along particular streets. The idea behind square or circular sectors is to keep at a minimum the time required for the patrol unit to travel to the scene of a reported incident in its sector. For instance, O. W. Wilson states that "... a square beat permits a maximum quadrilateral area with a minimum distance between any two possible points within it" [67, p. 228].

One factor not considered in this statement is that, for mobile patrol units, travel speeds may depend on direction of travel; then it

* More precisely: Among all quadrilaterals of a given area, the one with the smallest maximum distance between two points is a square.
will be desirable to design the sector so that the longer sector dimension corresponds to the direction with higher travel speeds. Using quantitative techniques, it is possible to predict the travel time characteristics of any proposed sector design and thereby determine which designs actually do minimize some indicator of travel time.

As an example, consider an urban region in which the streets form a mutually perpendicular grid (e.g., as occurs in central Manhattan and certain other cities), running, say, east-west and north-south. Then, a shortest route of travel for the assigned patrol unit requires the unit to traverse the total east-west distance, plus the total north-south distance, between the unit's initial position and the position of the incident.* Under the simplifying assumption that each patrol unit responds only within its own sector, Larson [43] has shown that the sector should be designed so that the average time required to travel east-west equals the average north-south travel time. Since it is not unusual to find regions (such as in Manhattan) where the north-south speed is about four times as great as the east-west speed, this implies that the sectors should also be four times as long in their north-south direction. In this case, the sector design can be expected to reduce average intrasector travel time by approximately 20 percent over that obtained by the rule-of-thumb design—square or circular sectors.

Some of these ideas have already been applied by Bottoms, Nilsson, and Olson in the city of Chicago [7]. They have constructed a new sector plan of the city using rectangular sectors designed so that the average intrasector travel time never exceeds approximately 3 minutes.

The same principles must be kept in mind when drawing boundaries between the response areas of units which have fixed locations. If all

*Several interesting applications of this "right-angle distance metric" and other metrics have been discussed by Fairthorne [26] and Smed and Jeffcoate [57]. Certain complications involving one-way streets or obstacles such as railroad tracks increase the average travel time from that which is estimated using this metric. Larson [45] has computed the effects of these complications, finding that the increase in the average is quite small, but that certain responses involving one-way streets may require 3 additional minutes for the unit to reach the scene.
the calls in a district are to be served by units from the fixed facility in that district, then the dividing lines must consist of points which are an equal travel **time** from two facilities, rather than an equal travel **distance** [46]. Modifications of these principles which result from the use of more realistic models will be discussed below.

**INFLUENCE OF INTERSECTOR COOPERATION**

When an incident is reported from a response area whose units are busy, most emergency service systems will dispatch a nearby available unit from another response area. Such an arrangement is nearly mandated by queuing considerations, but it introduces subtle complications into the design of response districts. In the case of mobile units, it even raises questions about the need for restricting response areas of the units to be nonoverlapping. These consequences of intersector cooperation will be discussed for two examples.

**Police Patrol: "Flying"**

Police administrators are often heard to argue in favor of assigning patrol units to nonoverlapping sectors in order to establish a "sector identity" on the part of the patrol officer. This identity, which derives from patrolling and from citizen contacts made while responding to calls for service, is supposed to cause the officer to feel responsible for public order in his sector. However, given nonoverlapping sectors, one can show by a simple probabilistic argument [44] that the fraction of dispatches which cause a unit to travel outside its own sector is usually equal to or greater than the fraction of time that units in that area are unavailable for dispatch. This result does not appear obvious at first glance, and it has been quite difficult to convince police administrators of the following type of statement: "If your patrol units are 'busy' 40 percent of the time (a typical value), then at least 40 percent of all dispatch assignments cause the assigned patrol unit to leave its 'own' patrol sector. In turn, at least 40 percent of all citizen contact occurring while responding to calls for service takes place in sectors other than the patrol unit's 'own' sector."
The predicted amount of intersector dispatches (called "flying" in some police departments) has been substantially verified both by our own work [44] and by the reports of others. For instance, in a report by McCormack and Moen [47] (both law enforcement officers supported as Office of Law Enforcement Assistance Fellows), based on analysis of a one-week sample of radio dispatches from the Mission Police District (San Francisco), it was found that in almost 50 percent of the cases, a unit from another area answered the call. The 50-percent figure checks roughly with that predicted by analysis, since the Mission District cars were "busy" during the sample week a portion of time which can be estimated as slightly greater than 50 percent. In a more precise test of the model in New York City [44], the data showed that the extent of intersector dispatching is never significantly less than the percentage of time unavailable, and it may be significantly more. Intersector dispatches ranged from 37 to 57 percent of the total.

The extent of flying brings into question not only the philosophy behind nonoverlapping sectors, but also a widely popular statistical procedure for computing workloads of police patrol cars. Usually a sector is associated with a workload which is proportional to the number of calls for service generated from within the sector. Thus, for instance, sector A would have a workload three times as great as sector B if sector A generated three times as many calls for service as sector B. And, it would usually be assumed that a patrol unit assigned to sector A would work three times as hard as a unit assigned to sector B. Using our knowledge of the flying phenomenon, we know that the latter assumption is false and, in fact, the car assigned to sector B may be dispatched to calls in sector A sufficiently often so that both may work about equally hard. Thus, to keep track of the workload of a patrol unit, one must record the dispatch assignments of each unit and not the rate of calls for service from individual sectors.

There is one additional property of nonoverlapping sector systems that we should mention. It involves the "randomness" of preventive patrol. With nonoverlapping sectors, preventive patrol coverage in a sector is reduced to zero whenever the corresponding patrol unit is busy.
Anyone, including potential criminals, can monitor a patrol unit's activity in some manner (e.g., visual observation, listening to the police radio) and determine when a particular car is not patrolling. Then, since units are assigned to nonoverlapping sectors, a crime can be committed with near-zero probability that the patrol unit will pass during the commission of the crime.

Given the undesirable features of a nonoverlapping sector system, how can an administrator revise and improve operations? First, if the sector concept is to be retained, the large percentage of calls which are low priority (i.e., they do not require rapid response)* can be "stacked" and handled by the car assigned to the sector of the call when that car becomes available. This procedure reduces the amount of flying and enhances "sector identity."

Second, the sector concept can be modified so that patrol units are assigned to overlapping areas (or sectors). This procedure enlarges the area with which each patrol officer should develop an identity. In addition, it increases the randomness of patrol, a desirable outcome which is not achieved simply by stacking on nonoverlapping sectors.

Clearly, the number of possible combinations of alternatives is very large. Fortunately, quantitative methods using mathematical models of operations can structure one's thinking about these alternatives and, in fact, can predict the extent of improvement to be gained by any particular combination [43].

Response Areas for Fire Units

A fire unit's primary response area consists of all points to which it would be dispatched if an alarm were generated there and all units were available. In the event of unavailabilities, the unit may also respond to alarms elsewhere. Fire departments have traditionally designed response areas so that the dispatched units are the ones closest to the fire. This means that all points on the dividing line between two districts are equally close to some pair of companies.

*For instance, even for those calls which are related to crimes, typically 75 percent are "nonemergency" and thus do not require rapid response [62, p. 91].
With the modification of interpreting "closest" in the sense of shortest travel time, one would expect this procedure to minimize overall response time. However intuitively reasonable this rule of thumb may appear, a recent analytical study by Carter, Chaiken, and Ignall [14] has shown that equal-travel-time dividing lines are usually not optimal and that overall average travel time is minimized by following a policy that often requires a unit other than the closest unit to be assigned to a particular fire.

The philosophy underlying this result is one that often appears in systems with unpredictable demands in the near future. It may be preferable to incur an immediate cost (e.g., travel time) that is slightly greater than the minimum possible immediate cost so that the system (e.g., the collection of all fire apparatus) is left in a state which best anticipates future demands. That is, assigning, say, the second closest unit to the most recent fire may result in favorable positioning of units for the next reported fire, thus minimizing overall travel time. Assignment of the closest unit to the first fire might have required an unusually large amount of time to respond to the next reported fire.

One can understand this phenomenon intuitively by considering the region illustrated in Fig. 3, in which one of the two units shown is to be dispatched to each incident, and the alarm rate is higher on the right than on the left. If an incident occurs at point a (which is closer to unit 2 than to unit 1), it appears natural to dispatch unit 2. However, there may be a good chance of another incident occurring while unit 2 is busy at point a. In such a case, it is likely to be located on the right side, say at point b. Unit 1 would then travel a long distance to point b, and it is clear that the average travel distance for these two dispatches would have been shorter if unit 1 had been dispatched to a, and unit 2 to b. The cited report [14] provides a formula for the location of the dividing line which minimizes average response time. For this example, it would be a vertical line somewhat to the right of the y-axis.

Carter, Chaiken, and Ignall have also shown that the boundaries which minimize overall average travel times will, in many cases, also reduce workload imbalance (where workload imbalance is defined to be the difference in the fraction of time worked by the busiest unit and
by the least busy unit). Thus, implementation of their derived procedures results in two types of gains—travel time reduction and workload imbalance reduction.

Their boundary-structuring procedures have been worked out in detail for the case of two cooperating units; current research is being directed at extending the results to systems with many cooperating units. The qualitative features of the results have already been used to arrive at preferable dividing lines in New York City Fire Department operations—and these results are currently being implemented.

![Diagram](attachment:image.png)

**Fig. 3 - Response areas for two units**
VI. LOCATING UNITS AND FACILITIES

Typical location problems of urban emergency service systems are:

- Which site to select for an additional facility whose approximate location has been decided.
- Where to place a new facility when consolidating two or more existing facilities into one.
- Where to locate units at the start of a tour (prepositioning).
- How, and under what circumstances, to change the locations of units during a tour to correct for unavailabilities as they develop (repositioning).

In rare instances, an agency administrator may have to select the sites for a large number of facilities.

Although there is an extensive literature on the subject of "facility location," most of it is presented in economic terms and ignores the probabilistic aspects of the operations of emergency service systems. ReVelle, Marks, and Liebman [53] have recently reviewed a variety of location theory models related to public sector problems, but none of the models appears to have been applied to an emergency service system. The work of Larson and Stevenson [46] is the beginning of a theory of facility location specifically designed for emergency services. Although further research is needed to eliminate some of their simplifying assumptions, this work tends to show that in a realistic environment the optimal location of a new facility is rather insensitive to the precise location of existing facilities.

A considerably larger body of analytical work has been completed, or is underway, concerning the repositioning of units during the course of a tour. Techniques for repositioning which are arising out of this work can provide large cities the opportunity to obtain substantial improvements in performance from their existing units. Two examples are discussed below.

*"Tour" refers to the period of time during which a specific group of men will be working.
Local Repositioning: Police Patrol Cars

Consider once again the situation illustrated in Fig. 3, except that now we will assume that each unit patrols its area randomly, and the demands are uniformly distributed over the entire region. Each unit responds in its own area, unless it is unavailable, in which case a second unit responds. The question of interest is, "At the moment when one of the units becomes busy, is there any advantage to repositioning the remaining available unit? If so, how should this be accomplished?"

Whenever one unit becomes unavailable, consider the following three alternatives for the free unit:

Alternative 1: Free unit does nothing (i.e., no repositioning).
Alternative 2: Free unit is assigned to patrol both areas uniformly (i.e., "uniform repositioning").
Alternative 3: Free unit is stationed on the boundary line between the two areas at the north-south halfway point (i.e., "fixed-point repositioning").

It is straightforward to show [43, p. 351] that alternatives 1 and 2 have the same average travel distance and the distance for alternative 3 is 75 percent as large. Thus, in the sense of average travel distance, uniform patrol repositioning (alternative 2) offers no advantage over no repositioning (alternative 1). On the other hand, fixed-point repositioning offers a 25-percent reduction in average travel distance when compared to alternatives 1 and 2. Similar results hold [43, p. 353] for regions of four cooperating sectors and for more complicated examples.

The results suggest that any local repositioning (among nearby sectors) is advantageous in terms of travel distance, only if patrol is concentrated near the boundaries of the appropriate sectors. In practice, strict fixed-point repositioning may not be advisable because of lost preventive patrol coverage; still, if the free unit must remain patrolling, a large part of the travel-distance reduction can be retained provided the patrol occurs near the appropriate sector boundaries. In fact, we have heard patrolmen remark that on an informal basis two units will occasionally agree to "cover" both sectors when the other unit is unavailable; this

* Similar results hold if travel time rather than travel distance is used.
covering usually takes the form of concentrated patrol near the common sector boundary. To gain travel-distance reductions when such covering occurs, it is necessary that the dispatcher be aware of the identity of the cooperating units so that he can assign the covering unit to a call in the busy unit's sector.

**Global Repositioning: Relocation of Fire Units**

By "global repositioning," we mean the reassignment of one or more available units to areas which may be at some distance from the areas to which they are currently assigned. For many years, urban fire departments have relocated available units when a number of units in one area become busy fighting a large fire. Indeed, these relocations are preplanned, so that when a second alarm (or higher) is sounded, certain specified units respond to the fire while other specified units simultaneously move into some of the fire houses which have been vacated by units at the fire. Such large-scale repositioning is not as widely used in other urban service systems, although situations continually arise (e.g., congestion at the police-precinct level) in which repositioning of forces would reduce dispatch delays or provide some preventive patrol.

The absence of global repositioning as a standard technique in police patrol operations may be explained by the fact that an accumulation of small incidents, rather than a single large incident, is most often the cause of whatever unavailabilities exist. Even fire departments are not likely to provide relocation guidelines for dispatchers to use in cases when several small fires produce as many vacant fire houses as a large fire might.

In part, the analysis of relocations is identical to the following question which arises in considering facility locations: "Given N potential locations (i.e., fire houses or patrol sectors) for units in a region, and n (which is less than N) available units, which sites should be occupied and which unoccupied?" However, the relocation problem is more complex than this, because the change in locations will only be temporary. For example, if having a unit at site C provides a 2-percent improvement in average response time over having a unit at site D, one might choose to build a facility at C rather than D; however, one would
probably not choose to move a unit from D to C for a short period of
time when site C is empty. Thus, the following additional factors come
into consideration in determining whether to make a relocation:

- How long is the expected duration of the existing unavailability?
- How many units will have to relocate to accomplish the desired
final locations?
- How long will it take for the units to travel to their new loca-
tions?
- Is the magnitude of the expected improvement in performance
large enough to warrant the effort required to move units?
- Is there a good reason to believe that the units to be moved will
be needed at their present locations in the near future?

Work still in progress at the New York City-Rand Institute is de-
signed to produce an algorithm which will operate in a computerized dis-
patch system and will calculate preferred relocations both for large
fires and for unavailabilities which occur through an accumulation of
minor incidents. With this system, no relocations will be preplanned, since
the calculations can be based on the actual status of the units at each moment.
Thus, the dispatchers will be provided with suggested relocations even under
circumstances where preplanned relocations do not currently exist.

Several approaches have been taken toward developing a relocation
algorithm, and the final version will probably be a combination of them.
Swersey [61] developed an integer programming model to determine which
fire houses should be empty and which full. His objective was basically
to minimize the expected travel time to incidents, taking into account
the expected time that busy units would remain busy, but, in addition,
his objective function provided a penalty for each unit relocated. Once
a solution to this model has been found, a standard assignment problem
can be solved to determine which units should move to which empty houses.

Unfortunately, it was not possible to solve Swersey's model rapidly
enough, using either branch-and-bound or approximate heuristic techniques
[63], to make it appropriate for giving immediate assistance to dispatchers.
The model can, however, be used to solve the simpler problem of determining where to preposition $n$ units (fewer than the number of houses) in order to minimize expected response time when all $n$ units are available.

The relocation method which is now planned for actual implementation has been developed by Kolesar and Walker [39] based on a suggestion of Chaiken. It is applied separately for the two types of fire units, fire engines and ladder trucks; for specificity, we will discuss the algorithm for engines. The notion of relocating units to achieve minimum expected travel time is abandoned in this model, since numerous arrangements of unavailabilities can occur which have an expected travel time close to the minimum. Instead, every location at which a fire might occur is guaranteed a minimal level of coverage. Conceptually, minimum coverage is defined to be the presence of at least one available engine which, if dispatched, could arrive at the location within a specified period of time, such as 5 minutes. A location which fails to meet this minimum standard is said to be "uncovered."

Most locations would be considered uncovered only if the closest three engines were unavailable, although some locations are uncovered when the closest two engines are busy. For convenience, a "response neighborhood" is defined as the collection of all points which would be uncovered if a particular two or three engines were unavailable. If no response neighborhoods are uncovered, no relocations of engines are needed, since an engine could reach the scene of any fire in a suitably short period of time. Even if one or more response neighborhoods are uncovered, it may not be sensible to schedule a relocation if the unavailable units will complete their service and return to their fire houses before a replacement could arrive from elsewhere. For these reasons, the relocation algorithm takes into account only those neighborhoods which are uncovered and will remain uncovered for at least, say, 20 minutes.

If such neighborhoods exist then the algorithm mandates at least one relocation, and it determines the minimum number of empty fire houses which must be newly filled so that none of the neighborhoods in
question remains uncovered. Since many different ways of filling the houses may eliminate the uncovered neighborhoods, the algorithm chooses among them on the basis of average travel time to incidents or a similar measure of performance.

Associated with each house to be filled is a previously prepared list of units which are suitable for relocating there. The algorithm must simply choose one engine from each list, but this has to be done carefully to avoid creating new uncovered neighborhoods or assigning one unit to relocate into two houses. Unless an extremely large number of engines are unavailable at once, it is not difficult to find a suitable collection of relocations, which may then be displayed for the dispatcher.

*There is no efficient general method for solving problems of this type (known as set-covering problems), so an approximate method is used.
VII. CRIME-PREVENTIVE PATROL

Although other urban service agencies have certain patrolling functions (e.g., fire departments "patrol" areas looking for fire hazards), the patrolling function is most important in urban police operations. A patrol unit is said to be performing "crime-preventive patrol" when passing through an assigned area, with the officers checking for crime hazards (e.g., open doors and windows) and attempting to intercept any crimes in progress. By removing opportunities for crime, preventive patrol activity is supposed to prevent crime. By posing the threat of apprehension, preventive patrol is supposed to deter individuals from committing crimes.

Mathematical models of the preventive patrol activity can be useful not only for resource allocation purposes, but also for the structuring of controlled experiments that would attempt to determine the relationships between preventive patrol effort and crime rates. Much debate in police circles has centered about the crime-preventive effectiveness of patrol. Most experiments which have been performed to measure this effectiveness have been marred by poor experimental design and by the irregular way in which crimes are reported [52]. We would expect the quantification of some of the features of preventive patrol to assist future studies in defining critical parameters and suggesting relationships among them, thereby creating the formal structure needed for controlled experimentation and analyses.

Most mathematical studies of police crime-preventive patrol have occurred in the past several years, although some earlier work in "search theory" is also relevant to the problem. The term "random patrol" was introduced into police literature in 1960 by Smith [58] who stressed the need for unpredictable patrol patterns. Blumstein and Larson [5] developed a simple analytical model for spatially homogeneous random patrol in order to estimate the probability that police would pass a crime in progress. Elliott [25] quoted one of Koopman's [41] search theory results and attempted also to compute probabilities of space-time coincidence of crime and patrol. Bottoms, Nilsson, and Olson [7] have applied some of these ideas to operational problems in the Chicago Police Department.
The present studies of preventive patrol are uncovering important physical parameters which have not previously been defined and are suggesting means for better allocating patrol effort. To give some idea of the types of models being formulated, we will outline the development of a very simple model that can be used to provide an estimate of the probability that a patrol unit will intercept a crime in progress. An important feature of the model is a new physical parameter which plays a key role in determining the crime-intercept probability.

To develop this new parameter, imagine that we take each street segment in a patrol sector and assign it a coverage number between zero and one. The number one is assigned only to those segments which are to receive maximum possible patrol coverage. The coverage number assigned to each other segment is in direct proportion to the relative amount of patrol coverage to be given the segment, compared to a maximum coverage segment. For instance, a segment assigned the coverage number 0.25 would be patrolled, on the average, one-fourth as frequently as a maximum coverage segment. Denote the coverage number assigned to segment \( i \) by \( e_i \) and the length (street-mileage) of segment \( i \) by \( l_i \) (\( i = 1, 2, \ldots, I = \) total number of street segments). Then the parameter

\[
L = e_1 l_1 + e_2 l_2 + \ldots + e_I l_I
\]

is called the "average effective length of a sector sweep." By this terminology we mean that \( L \) can be shown to be the average distance traveled between passings of any particular point contained in a maximum coverage segment. In models of preventive patrol activity, the parameter \( L \) appears to play a much more important role than the total street-mileage

\[
D = l_1 + l_2 + \ldots + l_I,
\]

which is usually used by police departments in resource allocations.

To illustrate the interpretation of \( L \), imagine that the total street-mileage \( D \) in a sector equals 10 miles. Assume that 1 mile of the sector should receive maximum possible patrol coverage. This region could correspond to 1 mile of store fronts, for instance. Assume that points in
the remainder of the sector (say, a residential district) should receive one-ninth the coverage of points in the highly patrolled region. Then, 
\[ L = 1 \cdot 1 + 9 \cdot 1/9 = 2 \text{ miles.} \]  
This means that, on the average, after passing one of the stores in the highly patrolled region, the patrolling unit will travel 2 miles before again passing the store. If the speed of patrol is 10 mph, the average time between passings of the store would be one-fifth hour, or 12 minutes. Similarly, one can show that the average time between passings of a house in the residential district is 108 minutes.

For a crime of short duration \( T \) which occurs on segment \( i \), one can argue that a reasonable upper-bound estimate of the probability of space-time coincidence of crime and patrol is

\[
P_c = s_e \frac{T}{L},
\]

where \( s = \) speed of patrolling vehicle, and

\[
P_p = \text{fraction of total time spent patrolling (the remainder of the time being spent answering calls and performing other duties).}
\]

Even this simple formula provides insight into certain features of system operation. It illustrates that the crime-intercept probability increases directly with the fraction of time spent patrolling, the duration of the crime, the patrol speed, and the assigned coverage number, while it decreases in inverse proportion to the average effective length of a sector sweep. The relationship between response and patrol activities is also made apparent by the formula. During periods when numerous calls for service occur, the fraction \( P_p \) of time spent patrolling is small, and therefore the crime-intercept probability is low. A police administrator can, of course, choose to screen calls and/or place them in queue so that \( P_p \) is increased, but then the system's average response time to incidents will increase. By using this formula with queuing models, it is possible to provide quantitative estimates of the trade-off between response time and the interception of crimes.

In the example given above, suppose the patrol unit actually spends 50 percent of its time patrolling (i.e., \( P_p = 0.5 \)) * and that we are

\[ * \text{This is a typical value.} \]
interested in estimating the probability that a crime of duration one minute (possibly a street robbery) will be intercepted by the patrolling unit. If the street robbery occurs along the mile of store fronts, $P_c = 1/24 = 0.042$. If the street robbery occurs in the residential district, $P_c = 1/216 = 0.0046$. These results could be interpreted to mean that on the average (for this example) no more than 1 out of 24 street robberies occurring along the store fronts and no more than 1 out of 216 street robberies occurring in the residential district would be intercepted by a patrolling vehicle. Even increasing the fraction of time spent patrolling ($P_p$) to its maximum possible value (1.0) would still leave these intercept probabilities remarkably small. Similarly low detection probabilities have been computed for presently operating systems. They bring to question whether the threat of apprehension provided by preventive patrol is actually great enough to deter crime.

Since crimes are most likely to occur in certain regions during certain time intervals, any effective allocation of preventive patrol effort must take into account the relative frequencies of occurrence of crimes at various places and times. Given these frequencies, the objective is to maximize detection probabilities by assigning more patrol effort to the crime-prone regions at the appropriate times. In terms of our model, the problem is to find the coverage numbers ($e_i$) that maximize detection probability. Linear hazard formulas have traditionally been used to perform this task; they imply that the preventive patrol coverage should be directly proportional to frequency of crime occurrence. However, by structuring a model of preventive patrol operation one finds that the allocation of preventive patrol effort is mathematically similar to an allocation of search effort problem studied by Koopman [41]. His results, when applied to the problem of preventive patrol, yield the following (very nonlinear) properties:

1. On street segments with sufficiently low crime rates, no preventive patrol effort should be allocated.
2. On street segments which should receive preventive patrol effort, the effort should grow slower than linearly with crime rate.
Although much more refinement of Koopman's techniques is required before they can be implemented by police, we would expect the qualitative features of his solution to hold (i.e., an allocation of effort which grows slower than linearly with crime rate and street segments where no patrol effort should be expended).

Recently, M. Rosenshine [54] analyzed the problem which arises from the fact that the topology of streets may be such that the set of desired coverage numbers \( e_i \) may not be feasible. For instance, a block of storefronts may require very great coverage and all surrounding blocks might require relatively little coverage. In order to get to the high coverage block, a patrol unit may have to cover (or patrol) the neighboring blocks more often than is suggested by analysis. Rosenshine shows how to devise a feasible set of coverage numbers which most closely resembles the desired set. He also considers various aspects of the "randomness" of patrol.
VIII. METHODS FOR EVALUATING PROPOSED CHANGES IN ALLOCATION PROCEDURES

An agency administrator is typically faced with a number of proposed changes in his allocation policy at one time. For example, the results of the models described in previous chapters may suggest that he should add units at certain times of day, select new locations for some units, change response or patrol areas, and modify the procedures for relocating units. In addition, certain technological innovations such as automatic car-locator systems may have been proposed to accomplish some of the same objectives. Before making a choice among the alternatives, the administrator will want to have a realistic comparison of the benefits which can be expected from each approach.

For a thorough evaluation of such a comprehensive change in allocation procedures, one generally has to turn to much more complex and detailed models than the ones already discussed. Large-scale simulation models are typically used for this purpose. They can provide information about the effect of a proposed policy change on a wide range of variables: response times to particular types of calls, workload of units, queuing delays, availability of units, etc. Such simulation studies have been undertaken by Savas [56], to investigate the reduction of travel times which could be achieved by spatially repositioning ambulances; by Swersey [60], to analyze the operations of the dispatch centers of the New York City Fire Department; by Carter and Ignall [12], to compare a wide range of combinations of fire department allocation policies; by Larson [43], in a study of the allocation of police patrol and the potential benefits of utilizing a car locator system; and by Adams and Barnard [1], to study the value of an automated dispatch system for the San Jose Police Department. Recent work on efficient computer coding of geographical data [35, 8] has been of some assistance in designing such simulation models of urban emergency service systems.

A common feature of these studies has been the finding that rather simple and inexpensive administrative innovations can often make a contribution to system performance which is equivalent to that of much more expensive hardware or increases in manpower. Swersey's study [60] provides such an example. In this case, the fire dispatching office in
Brooklyn was experiencing an increase in alarm rates and consequent delays prior to dispatch of units which were beginning to be of some concern to the Department. The simulation showed that computerized methods for recording, storing, or retrieving location information about alarms would not solve the essential difficulty, which had to do with the fact that a single man had final responsibility for every dispatch decision. Swersey's suggestions for dividing this responsibility, a basically administrative change which has been implemented, provided substantially decreased delays during peak-alarm hours.

Similarly, Larson's simulation [43] has demonstrated that the absence of an explicit priority structure for calls to police departments produces unnecessary delays for urgent and moderately important calls. Most departments have been reluctant to implement such a structure, stating that their policy is to provide rapid service to all citizens. However, some departments [21, 22] have begun to implement priority structures based on quantitative information derived from such models.

In addition, the Larson simulation was used to study the best use of automatic vehicle-locator systems in police departments. The technology of such systems is well developed [38, 55], and recently field tests and operational installations have been reported [10, 19, 23, 30, 32, 68]. Each system provides a central dispatcher with estimates of the positions of all service units (e.g., buses, patrol cars, taxicabs) and other status information (e.g., current speed and direction, current type of activity).

In the Larson study [43, p. 289] analysis showed that superimposing an automatic vehicle-locator system on a patrol force assigned to non-overlapping sectors causes an average travel-time reduction in the order of 10 to 20 percent, the exact value depending on the fraction of time each car is busy, number of sectors in a command, spatial distribution of calls, etc. Analysis also showed that a system with fully overlapping sectors utilizing car position information has approximately the same travel-time characteristics as current nonoverlapping sector systems without car position information. Thus, if there are reasons to want overlapping sectors, even to the extent of not assigning sectors to cars, there would be little or no degradation in travel time characteristics of the overlapping sector system, compared to current systems,
provided high resolution car position information is available. Apparently, the prepositioning advantages gained by assigning cars to mutually exclusive sectors are recovered by knowing exact car positions in a system with no deliberate spatial prepositioning. *(As mentioned in Chapter V, arguments based on regional identity and randomness of patrol seem to favor some type of overlapping sector plan.)*

This analysis is an example in which applying technology to a system "operating as usual" may not fully utilize the new technological capabilities.

*These statements are subject to the assumptions of the models used, the most critical of which is the assumption that the cars patrol independently of each other.*
IX. CONCLUSION

Although in many instances we do not yet know how to make the link between ultimate measures of performance of emergency systems and the quantities which can be studied using analytical models, it is now apparent that many models are sufficiently developed to be of great assistance to agency administrators when carefully used. Many of the research goals for allocation of police patrol forces proposed in 1968 in a study for the Department of Justice [6, p. 168] are now being approached, if not achieved. Wide interest in quantitative methods is evidenced by reports of applications of such techniques in police departments in Boston [42], New York [64], St. Louis [48], Chicago [18], Cleveland [28], Tucson, Arizona [3], Phoenix, Arizona [15], and Great Britain [9]. The whole subject of the allocation of fire units has been developed in the past two or three years and has given an entirely new complexion to fire research. In the next few years we expect that the models will improve in their sophistication and utility and agency administrators will make increasing use of quantitative models as their virtues become apparent.
REFERENCES


44. Larson, R., Measuring the Response Patterns of New York City Police Patrol Cars, New York City-Rand Institute, R-673-NYC/HUD, April 1971.


