AN ALGORITHM FOR THE INITIAL DISPATCH OF FIRE COMPANIES

G. M. CARTER, E. IGNALL, K. L. RIDER

R-1997-NYC

NOVEMBER 1977

THE
NEW YORK CITY
RAND
INSTITUTE

Rand
SANTA MONICA, CA. 90406

This study was sponsored by the City of New York. Its contents, however, do not necessarily represent the official views or policies of the City of New York.

Library of Congress Cataloging in Publication Data

Carter, Grace M.
An algorithm for the initial dispatch of fire companies.

([Report] - Rand Corporation ; R-1997-NYC)
1. Fire-engines--Dispatching--Mathematical models. I. Ignall, Edward, joint author. II. Rider, Kenneth Lloyd, 1938- joint author. III. Title. IV. Series: Rand Corporation. Rand report ; R-1997-NYC.
AS36.R3 R-1997 [TH9310.5] 081s [illegible]
ISBN 0-8330-0000-4 7[illegible]-[illegible]

The New York City-Rand Institute was a joint venture of the City of New York and The Rand Corporation. During the period 1969-1975 it operated as 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 functioned as a center for the continuing application of scientific and analytic techniques to problems of urban life and local government. It was governed by a Board of Trustees appointed jointly by the City and Rand. Because of the City's budgeting constraints, the Institute began a formal process of dissolution in the fall of 1975. Costs of publishing the present report were borne by The Rand Corporation.

Published by The Rand Corporation

PREFACE

This report documents a procedure for deciding how many and which fire companies to dispatch to a newly reported incident. Designed for the computerized Management Information and Control System (MICS) now being installed by the Fire Department of New York City, the procedure was developed with the support of New York City under a contract with the New York City-Rand Institute.

SUMMARY

When an incident is reported to a local fire department, someone must decide how many and which companies (firefighters and their apparatus, usually either a pumping engine or a ladder truck) to dispatch to the scene. Traditional policies, based on the assumptions that the companies desired are always available and that the workload imposed on the firefighters is unimportant, do not work efficiently at the high alarm rates now characteristic of parts of New York and other cities. We report here a procedure that makes good initial dispatch decisions regardless of alarm rate.

Roughly speaking, a good procedure will get speedy initial response to those incidents that need it. Conceptually, we define serious incidents as those at which initial response matters; operationally, in terms of the number of engines and ladders working at the incident. We cross-classify this incident typology with the (incomplete) information available to the dispatcher when he must make the decision. We develop an objective function that includes the expected current "loss" associated with each possible dispatch decision and the future consequences of that decision.

The current loss considers the loss at the incident and a workload factor. The loss at the incident is approximately the chance that the incident is serious multiplied by a weighted average of the response times of various units to the incident (the latter depending on the dispatch decision). The future loss is approximated by the expected value of the loss at the first new incident, if any, occurring in the interval from now until a unit reports back the situation at the scene of the current incident. (This depends on the dispatch decision, which determines who is available to respond to that new incident.)

This research builds on earlier work by Swersey and by Carter, Chaiken, and Ignall. That work separated the questions of how many units to send and which ones to send and treated them in idealized situations. The resulting algorithm has several parameters that can be set by the user, and thus is a range of initial dispatch policies.

ACKNOWLEDGMENTS

The subject of initial dispatch was among the first considered at the New York City-Rand Institute in the late 1960s. So the list of references is only a partial acknowledgment of our intellectual and other debts.

In the development of the algorithm, the notion of a serious fire is central. Many hours were spent on this seemingly simple and innocent topic, and all members of the fire project were involved at one time or another. Chief Homer Bishop and Marc Leopold of the FDNY were very helpful in this effort.

The computer specification of the algorithm and of the files of alarm and company information it needed was done by Maggie Goetz, Joan Held, Toby Kosloff, and Carol Shanesy. Some initial work was done by Mei Ling. Clara Lai prepared the exogenous events tape for the simulation test.

The adaptive response policy described in the final section was apparently devised by the entire fire project staff. Intellectual title to it, then, belongs as much or more to Arthur Swersey, Peter Kolesar, Warren Walker, and Edward Blum as to the authors of this report. That policy was also the outgrowth of "Adaptive Response," a policy in which Chief Francis Ronan played a major role.

We thank David Jaquette and Jan Chaiken for helpful comments on a draft version of the report.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS vii

Section

I. INTRODUCTION 1

II. INITIAL DISPATCH OBJECTIVES 6
- Response Time and Losses Due to Inadequate Responses 6
- Unnecessary Responses 8
- Information Available at Time of Dispatch 10
- Objective Function for Initial Dispatch 11

III. MATHEMATICAL FORMULATION 15
- The "How Many" Decision 16
- The "Which Companies" Decision 21

IV. THE INITIAL DISPATCH ALGORITHM 25

V. AN EXAMPLE 29

VI. RESULTS OF SIMULATION TESTS IN NEW YORK CITY 35

VII. ADAPTIVE RESPONSE 39
- Defining an Adaptive Response Policy 40
- Example of a Simple Adaptive Policy 41
- Discussion and Extension 43

Appendix

A. DETAILS OF ALGORITHM DEVELOPMENT 45

B. ALGORITHM FLOWCHART 48

C. SIMULATION DETAILS 53

REFERENCES 55

I. INTRODUCTION

When an alarm is reported to a fire department, a decision has to be made: How many of the department's fire companies, and which ones, should be sent to the alarm?

The goal of an initial dispatch of fire companies is to get the necessary groups of men and equipment to an incident within an appropriate time. The resources needed and the appropriate response time differ from incident to incident. A small fire in a single family dwelling, for example, might call for a quick response of two engines and a ladder. A serious fire in a large commercial structure might require an initial response of several engines and ladders, with a followup (second alarm) of several more. On the other hand, a rubbish fire would probably need only a single engine company within a reasonable length of time. How can the dispatch decision be tailored to the requirements of the incident?

In the past--when alarm rates were low and false alarms were relatively few--the decision was fairly simple. When the dispatcher had some information about the nature of the incident (based on a telephone call or other verbal report), he would dispatch what appeared to be required. An alarm from a street box (in which case the dispatcher knows only that someone pulled the handle on that box) was presumed to signal a serious fire, and a preplanned response, judged to be what was needed initially for serious fires in that area, was used.

Sending the full preplanned response to box alarms made obvious sense. If the fire turned out to be minor, the extra units simply returned to their houses. Since the alarm rate was low, there was little danger that they would be needed while they were out. If the fire was large, the required units were on the scene quickly. In addition, under the traditional dispatching policy the initial dispatch always sent the closest companies--there was no reason to do otherwise.

But today we see rising alarm rates and tight municipal budgets. In New York City, many individual fire companies make over 5000 responses annually. Thus the pressure to match the initial dispatch to what is needed at the fireground is greater than it was years ago.

In New York City, alarm assignment cards (running cards) such as the one in Fig. 1 are used in initial dispatch and redeployment decisions. Companies are shown on the card in order of their distance from the location (in Fig. 1, from alarm box 2277). The companies on the first line constitute the full first-alarm assignment; those on the second line are the second-alarm assignment--they are called if an officer at the scene decides they are needed; and so forth.

Use of alarm assignment cards assumes that all companies listed will be available for dispatch when the alarm is received. Although in the past this was usually the case, with recent alarm rate increases one or more of the companies to be dispatched may be unavailable. In some instances the dispatcher will "replace" an unavailable company with another company listed on the same (or the next) line of the running card; in others, he will not, and fewer companies will be sent to the incident.[1]

In New York, the practice for street box alarms and some telephone alarms in the past few years has been to send engines[2] from the two houses closest to an incident if they are available. If one or more of these engines is unavailable, the (one) closest engine among those available at the time would be sent. The number of engines sent, therefore, would depend on the number available. We call this a *variable* response policy. In contrast, what we call a *fixed* response policy would specify a fixed number of engines (for example, two) to be sent to the incident, regardless of company availability. This fixed number need not be the same for different alarm boxes or for the same box at different times of day.

A variable response policy is a natural one for a busy city to use because it can reduce workload. While a fixed policy might modify response by time of day, this policy modifies a dispatch depending on the actual conditions at the time of an alarm. The more companies that

---

[1]Note that we are considering the question of replacing unavailable units only on the *initial* dispatch. We are not considering methods that adjust the initial response according to reports from units that reach the scene.

[2]The same policy is used for ladders.

ERS	BERGEN AND BROOK AVES (730, 540)	BRONX

ENGINES	LADDERS	SPECIAL UNITS	CHIEFS	Covering/Relocating Units Engines		Covering/Relocating Units Ladders	Covering Chiefs
60R	55 17		B26				B14
83	19	R3 S5	D 6 B55	42- 71 22- 60	96- 73	43- 17 59- 19	D5
82 94 69	42			90- 82 48- 92	67- 69	48- 42	
68 36 35	29			43- 36			
80 45 91	44			38- 45 88- 46	312- 91	56- 40	

ZONES	CIRCUIT	PRECINCT	ADMIN. CO.	NEAR BOXES	ORDER	DATE
	9	042	L055		056/75	03/27/75

Fig. 1--Assignment card (running card)

are busy fighting fires, the less chance that two companies of a given type will be dispatched to an alarm.

This policy, however, has serious disadvantages. For example, almost all engines might be available except for one of the two closest to a given location. A reasonable policy would, in this case, send two engines to a new alarm at that location. But a variable "two engine" policy would send only one. This could lead to a situation where a false alarm at a given location is sent two engines and two ladders, while a nearby serious structural fire reported over the phone a short time later would be sent one engine and one ladder. Thus, the decision on how many companies are sent to an alarm is not controlled by the department, in the sense that it is determined by the sequence of recent alarms.

In this report, we describe an algorithm that determines which and how many fire companies should be dispatched when the alarm is first received. It cannot help a fire department decide how many engines to send to an incident that is known to be a residential fire; it is intended to help dispatchers make the best use of the information available to them when the type of incident is uncertain.

Our goal was to develop a procedure for dispatching fire companies that would overcome the problems of the existing system, produce *good* initial dispatches, and still be simple enough to require negligible computation time. We defined *good* dispatches as ones that we, the public, and the Fire Department could all agree were appropriate; would not present increased danger of loss of life or property over current dispatch procedures; would improve workload distribution and response time to serious fires; and would keep available companies well distributed in the region, thus reducing the need for relocations to fill gaps in coverage.

The algorithm has been designed to be implemented in the computerized Management Information and Control System (MICS) of the FDNY, an on-line, real-time system that provides easy entry of up-to-date information about the status of each company. By using the computer's capability to store and process this information, the algorithm overcomes the deficiencies of existing methods of making initial dispatches.

In addition, some of the methods described can be used without a computer to improve manual initial dispatch methods.

While developed for New York City, the algorithm could be incorporated in computer-assisted dispatching systems in other large cities. Such a system, whether it implements a city's traditional dispatching policy or a procedure like the one described in this report, is a substantial undertaking. The additional tasks that adapting the algorithm would impose on a city include:

1. Analysis of historical alarm records.
2. Programming the algorithm and its interface with the rest of the computer-assisted dispatch system.
3. Simulation testing of the algorithm, to help choose suitable values for the weight to give serious incidents relative to nonserious ones, and other parameters that are described later in this report.

Tasks 1 and 3 would require several man-months of analytic work (by an operations researcher, for example). The flowchart in App. B would facilitate task 2 (see Fig. B.1).

Section II discusses the objectives of an initial dispatch policy. In Sec. III, we formulate algorithms for deciding how many and which companies to dispatch. These algorithms are used to develop the initial dispatch algorithm in Sec. IV. Section V gives an example of the use of the algorithm, and Sec. VI presents the simulation test results. For cities that do not need a dynamic algorithm, a procedure called adaptive response is discussed in Sec. VII.

Appendix A describes some details of the algorithm's development, App. B contains a flowchart of the algorithm, and App. C presents simulation details.

II. INITIAL DISPATCH OBJECTIVES

To evaluate an initial dispatch policy or to compare different policies, we need to look carefully at possible "errors" and their consequences. An initial dispatch can be "wrong" for a particular incident in two ways:

1. It can send fewer companies than are needed. The ensuing delay in getting the required companies to the scene can increase damage, injury, or deaths.
2. It can send more companies than are needed. The resulting unnecessary responses cost something in themselves, chiefly in stress on the firefighters. They also prevent the affected companies from responding to any alarm that is received before the companies are "released" from the current alarm.

Note that the appropriate initial dispatch for a serious incident need not be the total number of companies eventually needed there. For example, having the companies arrive more or less in groups (initial dispatch; followups) might help the officer in charge deploy his forces better than all-at-once arrival.

RESPONSE TIME AND LOSSES DUE TO INADEQUATE RESPONSES

The loss at an incident depends on the nature of the fire and how long it takes units to arrive (as well as other factors such as the time until discovery and the firefighting tactics used). Unfortunately, the precise relationship between response time and fire loss is not yet known.[1] Therefore, to formulate an initial dispatch policy, we will assume that a policy that reduces response times will, on the

---

[1] The most advanced work on this relationship has been done in Great Britain. See Hogg (1973) for an example, and Swersey et al. (1975) for a review of the situation.

average, reduce losses, and we will use response time as our measure of loss. For example, if a particular policy results in delayed responses to 50 serious fires, we can express that loss in terms of the additional response time to those fires (e.g., "the average additional response time for the second-arriving ladder company was 3 minutes"). A similar measure expresses loss in terms of the fraction of serious fires where there was a delay. For example, if there is a delayed response to 50 of 200 serious fires, we may say the initial response was adequate to 75 percent of the serious fires. We have found that *initial response adequacy* is more convenient to use than response time in some cases. When the alarm rate is low, we will see that these measures are nearly equivalent.

For both response time and response adequacy, different types of incidents should be separated. For example, short response times are important for structural fires but not for false alarms; adequate responses are more important for structural fires than for rubbish fires.

How can we separate incidents into types so that response times and response adequacy are reasonable substitutes for dollar damage, injuries, and fatalities? The key is to classify incidents as *serious* or *nonserious* according to the extent to which response times and adequacy affect life and property loss. The serious incidents are those structural fires (and other emergencies) at which slow or inadequate initial response will add to life and property loss. The nonserious incidents are false alarms, rubbish fires, fires in abandoned cars, and so forth where no exposure exists, so that the initial response matters little. Note that this notion of seriousness may not be the same as that used informally in a fire department. In fact, according to this definition, some fatal incidents are not classified as serious. For example, in the classic drinking-and-smoking fire, the victim is often dead and the fire smoldering in a chair when the incident is reported to the fire department. From an initial dispatch point of view, this kind of incident would not be considered serious.

How can past incidents be classified as serious or not? This is something that has to be worked out in conjunction with experienced fire officers. The following definition proved workable in New York City:

> A serious incident is any incident in a nonvacant structure that requires at least two ladders working; all others are not serious.[1]

The precise definition of seriousness is not crucial; the same incidents, more or less, are serious by any reasonable definition.[2]

UNNECESSARY RESPONSES

The consequences of an unnecessary response depend on the situation. What can be termed the direct cost--the gasoline and oil consumed, the cost of an accident if one occurs, and the wear and tear on the apparatus--is small. The indirect costs--resulting from needless stress on firefighters, or the chance that the company will be needed at another alarm--can be very large.

Consider the direct costs first. The cost of gasoline and oil for a two- or three-mile round trip would be less than $1. Wear and tear on the apparatus--tires, engine, etc.--would also be small. Typical accident rates are very small and would add less than $1 to the cost of an unnecessary response.[3]

What about indirect costs of firefighters responding to an incident where their services are not needed? The issue here is not one single such response, but the pattern of them. For the firefighter, there is a large difference between expecting to work on half the alarms to which he goes and expecting to work on one in twenty. And

---

[1] In New York, putting a second ladder to work usually indicates extensive search or overhaul, suggesting that an initial dispatch of two ladders instead of one would have some effect on the outcome.

[2] See Ignall, Rider, and Urbach (forthcoming).

[3] In New York City in 1973, there were 213 vehicle accidents involving units responding to and returning from 300,000 incidents. Most (but not all) involved engines or ladders. Figuring an average dispatch of four companies to each incident, the accident rate would be perhaps 200/1,200,000, or 1 for every 6000 responses. The loss at each accident would have to be $6000 to add $1 to the cost of a response.

there is a large difference between expecting to respond twenty times in a nine-hour tour and expecting to respond twice.

When the number of responses per hour is high, turnout may be slower at the end of the tour and firefighting may be less effective. Thus, response to several false alarms early in the tour can mean that a fire that occurs later will do more damage. In a similar way, when the chance that the response will not be necessary is large, turnout may be slower and morale may decline. In New York City in the late 1960s, this morale problem was one factor behind union pressures for more fire companies. Thus, a large enough increase in needless responses in a year in an area could mean that a new fire company will be created. There is no way to determine exactly how large this increase would have to be, but 5000 responses is the right order of magnitude. At an annual cost of over $500,000 for a company, the indirect cost of each needless response in this case would be over $100.

To capture the cost of unnecessary responses in units that are compatible with losses at incidents, we define a factor, W, in units of time. This factor, which is associated with a company response, is a surrogate for the cost of that response as discussed above. As we will see, the initial dispatch algorithm will reduce the number of unnecessary responses and increase response time if a larger value of W is chosen. Thus, the determination of this factor involves a subjective tradeoff between unnecessary responses and response time.

Another consequence of a needless response is the possibility that the unit in question will be needed to answer another alarm that occurs before the unit is released from the current one. If too many units are sent initially in response to an incoming alarm, the extra units will be unavailable to respond to other alarms that are received during the time it takes for the first units to arrive and determine that the extra units are not needed.[1] These later alarms may be

[1]Alternatively, all but the first-arriving unit can be dispatched on a conditional basis. That is, if another alarm is received, one or more of these units can be diverted to the new incident before the nature of the first one is known.

subject to the costs of an inadequate or delayed response. In areas with high alarm rates, this is an important consideration.

INFORMATION AVAILABLE AT TIME OF DISPATCH

The preceding discussion of loss considers incidents only in terms of an after-the-fact assessment of the situation at the scene of the incident. However, to implement an initial dispatch algorithm, we need to know or to estimate the requirements of an alarm that has just been received. Therefore, we need to be able to associate losses with incidents as they appear to the dispatcher when they are reported.

First, we have to systematically categorize what a dispatcher might know about an incoming alarm. We do this by defining various *alarm types* that reflect before-the-fact information. For example, we might define the following alarm types:

- Street box alarm.
- Telephone report--sounds like a structural fire.
- Telephone report--sounds like a rubbish fire or similar minor incident.

The alarm type, specifying the alarm as received, defines the information available to the dispatcher and relevant to the initial dispatch decision. We are interested, then, in defining alarm types that provide as much information as possible about the requirements of the incident.

In this report, we emphasize the separation of street box alarms into several alarm types. It was found in New York City that street box alarms can be divided reliably into *risk* classes. The chance that an alarm from a high risk box is a serious incident[1] is considerably higher than the chance that one from a low risk box is serious.[2] Thus knowing whether a street box alarm is from a high or low risk box is

---

[1]For these purposes, one that needs at least two engines and two ladders.

[2]See Carter and Rolph (1973).

useful information for the dispatcher. In a similar way, the time of day provides information. A box alarm received at 3 p.m. on a school day is less likely to be serious (because it is more likely to be false) than one received at 3 a.m. And an incident reported from more than one source--a telephone call and a pulled street box, or two (or more) calls--is much more likely to be serious than one that gets a single report. (In areas of New York City where an adaptive response policy[1] was in use, about half of all serious incidents in the late afternoon and evening that were first reported by box also had a follow-up telephone report before the first companies arrived at the scene.)

In any event, we will assume that an alarm classification scheme has been devised, and that the rates at which given types of alarms occur and the chance that a given type of alarm signals a serious fire have been estimated.

OBJECTIVE FUNCTION FOR INITIAL DISPATCH

Our objective is to reduce losses and workload at serious fires. As discussed previously, we will measure both of these in units of time. The initial dispatch algorithm is intended for situations where there is a significant chance of receiving another alarm (for another incident) before the companies finish handling the current one. Therefore, the calculation of expected losses cannot consider the current alarm only. In fact, our real problem is to find workable ways of combining the current and future consequences of a particular dispatch decision. That is, we want to first get expressions for the current loss, C, and future loss, F, of a possible dispatch decision, and then choose the decision that minimizes the sum, $Z = C + F$.

Before developing the initial dispatch algorithm, "losses at serious fires" must be more precisely defined. For the purposes of a dispatching policy, the most direct definition of a serious fire is "a fire that requires two ladders (engines) quickly," as discussed earlier in this section.

In general, the expected loss at a serious fire will be some (perhaps complicated) function of the time it takes all the needed

[1] See Sec. VII.

fire companies to arrive at the scene. As indicated previously, this function is not yet known. We will use a simple function of various response times that captures what appear to be the major determinants of loss. Following Keeney (1973), we give more weight to the response time of the first-arriving unit than to the second. We also give more weight to response times to serious fires than to other incidents. For simplicity, we chose a linear function.[1] To construct this function, we use the following concepts:

- Quantities to be determined by choice of how many and which companies to dispatch:

  R_1 = estimated time it takes the first-arriving company to respond.

  R_2 = estimated time it takes the second-arriving company to respond.

  $R(i)$ = estimated time it takes company i to respond.

- Information available at time of dispatch:

  $t_j(\underline{x})$ = estimated travel time from location of company j to location $\underline{x}$.

  $p(\underline{x}, m)$ = estimated probability that an alarm of type m, at location $\underline{x}$, is serious.

- Parameters chosen by user:

  β = importance of second-arriving company response time relative to first-arriving response time.

  r = importance of response time to a nonserious incident relative to response time to a serious incident.

Response time includes the time it takes to dispatch a company after receipt of an alarm, the turnout time u, the travel time, and the time for the company to set up for firefighting. Thus, if the second company is not immediately dispatched, the delay is included in the response time. For our purposes, we neglect the dispatch time and the set-up time.

The parameter β lies between 0 and 1; if $\beta = 1$, then the response time for the second-arriving unit is weighted just as heavily as the

---

[1]See Swersey (1972) and Ignall, Rider, and Urbach (forthcoming).

first-arriving unit's response time; if $\beta = 0$, the second-arriving unit's response time is not considered at all. A typical value could be $\beta = .5$. It might be tempting to set the parameter $r = 0$; however, our procedure would take this literally and send a company from many miles away to a nonserious incident. This is prevented by assigning r some small value, typically .01 or .02.

Calculation of the current loss is relatively straightforward. Consider an alarm of type m at location $\underline{x}$. If ladder i is the first-arriving ladder at the incident and ladder j is the second-arriving ladder, the expected loss measured in time units is

$$C = \{p(\underline{x}, m) + [1 - p(\underline{x}, m)]r\}R_1 + p(\underline{x}, m)\beta R_2 , \tag{1}$$

where $R_1 = R(i)$ and $R_2 = R(j)$. It is easier to work with C if we rewrite it as

$$C = q(\underline{x}, m)R_1 + p(\underline{x}, m)\beta R_2 ,$$

where $q(\underline{x}, m) = p(\underline{x}, m) + [1 - p(\underline{x}, m)]r$. That is, C is a weighted sum of the first- and second-arriving ladder response times. Typical weights on the response times are shown in Table 1.

Calculation of the future loss is more difficult. Formally, we would sum the losses at future alarms given an expected alarm stream and dispatch policy. However, the future status of fire companies, and consequently of future decisions, will depend on the action taken for the incoming alarms. Therefore, a straightforward minimization of current plus future losses is almost intractable. In addition, an initial dispatch algorithm must be simple enough to allow the rapid calculation and recommendation of a dispatch. For these reasons, we will use an approximate calculation of the future losses of a dispatch decision.

Table 1

TYPICAL WEIGHTS ON FIRST- AND SECOND-ARRIVING LADDER RESPONSE TIME

$r = .02 \quad \beta = .5$

Alarm Type	p	Weight on First Unit Response Time (q)	Weight on Second Unit Response Time (βp)
Telephone (sounds like a structural fire)	.10	.118	.050
Box alarm	.05	.069	.025
Telephone (sounds like a car fire)	0	.020	0

III. MATHEMATICAL FORMULATION

Since alarms appear to follow a Poisson process,[1] a semi-Markov decision process would be a natural model for finding the best initial dispatch policy. The state of the system would be the status of the individual companies; transitions would occur when a company's status changed (e.g., when it became available after working at an incident) or an alarm was turned in.

Geography would be represented by identifying the locations of the alarms and companies, and calculating travel times accordingly. However, such a model would have a state space too large for effective computational procedures. Consequently, previous work (Swersey, 1972; Carter, Chaiken, and Ignall, 1972) eliminates important features in order to get insight on some aspects of the problem.

In this section, then, we present a mathematical formulation of the initial dispatch decision in which the problem is broken into two segments:

1. How many companies should be dispatched.
2. Which companies should be dispatched.

Each problem is given a separate mathematical formulation. In Sec. IV, the formulations will be combined to derive the initial dispatch algorithm.

Following Swersey (1972), our approach to future losses concentrates on the first future incident, if any, to occur before a company arrives at the scene of the current incident. To make the development easier to follow, we will drop from F those parts of the future losses that do not depend on the current decision.

Because the development of the algorithm uses cases that go beyond the assumptions of the models and makes use of various approximations, the algorithm is necessarily an approximation to the best policy.

---

[1] See Carter and Rolph (1975).

With this in mind, we provide a number of parameters that the user can vary. The initial dispatch algorithm therefore encompasses a set of policies.

THE "HOW MANY" DECISION

We build here on the work of Swersey (1972). For the purposes of illustration, he considers ladders, a fixed region that has eight ladders in it,[1] and only two possibilities for each incident: sending one ladder or sending two ladders. The relevant feature of an alarm is the chance that it signals a serious fire. The objective is the minimization of the sum of long-run average losses at the incidents and the "costs" of company response (as measured by the workload factor). These losses occur only at serious incidents and are determined by first- and second-ladder response times at those incidents.

Geography is not explicitly represented--travel times depend only on the number of companies currently available, not their identities (which are not distinguished in the model). Implicitly, then, the assumption is that the closest available ladder would always be dispatched, and if a second ladder were sent, it would be the second closest available unit.

Ladders that are busy are separated into those that are working at serious incidents and those working at nonserious incidents, with the latter becoming available more quickly. For our purposes, we will assume that service time is the same for both kinds of busy ladders, which allows us to lump them together and consider only the total number of ladders busy.

For this situation, the optimal dispatch policy is of the following form:

<u>Swersey Cutoff Theorem</u>: For each state of the system (i), where i is the number of ladders busy, there is a cuttoff probability, s*(i), such that two ladders should be sent

---

[1]The region should be large enough to include several ladders, but not so large that incidents in one part do not affect decisions in another.

if an alarm occurs with a probability serious greater than $s^*(i)$; otherwise, one ladder should be sent.

The value of s^* depends on the total number of ladders busy, i. Swersey found that s^* increases when i increases. He also found that the optimal cutoffs are larger when the alarm rate and the workload factor are larger.

We extend Swersey's algorithm by relaxing the assumption that the closest available ladder will always be dispatched and that the second ladder, if sent, would be the second closest available unit. In our situation, the state of the system is the status of the individual ladder companies, called $\underline{i}$. We will calculate a loss if one ladder is sent $Z_1(p, \underline{i})$, and a loss if two are sent $Z_2(p, \underline{i})$, and use the rule:

Send two if $Z_i(p, \underline{i}) > Z_2(p, \underline{i})$.
Send one otherwise.

Since both increase monotonically with increasing p, and Z_1 increases more rapidly than Z_2, this policy has the same form as the Swersey Cutoff Theorem.

Calculation of the current losses for either dispatch is straightforward. We use the loss function, Eq. (1).

Current Loss

Send One: Suppose ladder i alone is sent. Let ladder j be the closest of the other ladders.

$$R_1 = u + t_i(\underline{x}) ,$$
$$R_2 = [u + t_i(\underline{x})] + u + t_j(\underline{x}) .$$

The delay represented by the first term in R_2 can be smaller if the first engine arrives before ladder company i. When we combine the pieces here into the algorithm, we replace

$(u + t_i)$ by r_{min}, which is defined as u plus the smaller of t_i and the first engine time. The expected current loss with this dispatch is

$$C_1 = q(\underline{x}, m)[u + t_i(\underline{x})] + p(\underline{x}, m)\beta \times \{[u + t_i(\underline{x})] + u + t_j(\underline{x})\} . \tag{2}$$

Send Two: Suppose ladders i and j are sent, with i being closer than j. (The i and j here need have no relationship with the i and j in the previous paragraph.) Then,

$$R_1 = u + t_i(\underline{x}) ,$$

$$R_2 = u + t_j(\underline{x}) .$$

The expected current loss with this choice of responding companies is

$$C_2 = q(\underline{x}, m)[u + t_i(\underline{x})] + p(\underline{x}, m)\beta[u + t_j(\underline{x})] . \tag{3}$$

Future Losses

To consider future losses, we need the following definitions and concepts:

λ = total alarm rate in the region.

$\overline{p}$ = chance that an alarm in the region is serious.

$\overline{q} = \overline{p} + r(1 - \overline{p})$.

N = number of ladders stationed in the region.

n = number of those ladders that are now available (before any are sent to the incident).

Let $T_1(k)$ and $T_2(k)$ be the expected travel times of the first- and second-arriving ladders, respectively, when k ladders are available, and let $r_{min}(k)$ be the expected response time of the first-arriving unit of any kind. For each value of k, these three times can be estimated using the area of the region and models relating travel distance to area served (Kolesar and Blum, 1975) and travel time to travel distance (Kolesar and Walker, 1974).

It is not practical to calculate F exactly. Therefore, we compromise between ignoring the future and treating it exhaustively. As first suggested by Swersey (1972), a reasonable approximation is to look at the interval from now until the first unit arrives on the scene and the next alarm, if any, received in that interval. The future loss is taken to be the loss at that alarm times the chance that it occurs.

The significance of this interval is that once the first unit arrives at the scene, the number of companies actually needed will be determined by the firefighters. Extra units will be sent back, or additional units will be called for as needed. Thus, the major effect of the initial dispatch decision on company availability,[1] and thus on future losses, is felt in that interval.

The loss at the next alarm will depend on the expected seriousness of that alarm, and the expected initial dispatch and resulting response times to it. If there are k ladders available then and the chance that two are sent initially is d, then the expected loss of the next alarm, if it occurs, is

$$F = \bar{q}[T_1(k) + u] + \bar{p}\beta[T_2(k) + u + (1 - d)r_{min}(k)] \ . \qquad (4)$$

The quantity d depends on the algorithm, but it is sufficient to approximate it from the results of Ignall and Urbach (1975). (See App. A.) Equation (4) is an obvious extension of the formula for current losses.

The probability that an alarm comes during the time r_{min} when the alarm stream is Poisson is 1 - exp (λr_{min}). When λr_{min} is small, this may be approximated by λr_{min}, the expected number of new alarms in the interval.

[1] As opposed to the actual situation at the scene.

We now calculate expected future losses on the basis of the number of ladders sent to the current alarm:

Send One: The future loss is

$$F_1 = \left\{ \begin{array}{l} \bar{q}[T_1(n - 1) + u] \\ \\ + \bar{p}\beta[T_2(n - 1) + u + (1 - d)r_{min}\ (n - 1)] \end{array} \right\} \lambda r_{min} \ . \tag{5a}$$

The term inside the braces is the loss at any new alarm that occurs.

Send Two: The relationship is similar to (5a). We just replace $n - 1$ by $n - 2$. That is,

$$F_2 = \left\{ \begin{array}{l} \bar{q}[T_1(n - 2) + u] \\ \\ + \bar{p}\beta[T_2(n - 2) + u + (1 - d)r_{min}\ (n - 2)] \end{array} \right\} \lambda r_{min} \ . \tag{5b}$$

Combining Eqs. (2), (3), (5a), and (5b) so that

$$Z_1 = C_1 + F_1 \ , \quad \text{and} \quad Z_2 = C_2 + F_2 \tag{6a}$$

gives the loss in terms of response times to fire. We now consider workload. The workload "cost" is 2W if two ladders are sent and $W + p(x, m)W$ if one is sent (since the second ladder will eventually be sent if the alarm is serious). Thus, it is sufficient to add the difference, $W[1 - p(\underline{x}, m)]$ to Z_2, to obtain

$$Z_2 = C_2 + F_2 + [1 - p(\underline{x}, m)]W \ . \tag{6b}$$

To see the role of the workload factor W, consider the typical situation, in which r_{min} will be less than five minutes, so that in all regions of most cities, λr_{min} will be much smaller than one alarm at all times of the day. Suppose W were zero. Then in most situations, F_2 will be only slightly larger than F_1, so that unless $p(\underline{x}, m)$ is very nearly zero, Z_1 will be more than Z_2. (F_2 will be significantly larger than F_1 if n is small and λ is large.) Thus, choosing the smaller of Z_1 and Z_2 would send two ladders almost always. Adding a positive W (measured in units of response time) increases Z_2 and thus sends two ladders less often. W can depend on the region and the time of day and can be viewed as a parameter that permits us to adjust the average number of ladders dispatched.

Two important properties of this formulation are the following. The procedure will make an appropriate decision between any two candidate companies. There is no restriction to the closest and second-closest units. However, the "future" loss used by the algorithm does not depend on the candidate companies. The algorithm calculates the regional response time degradation, assuming that this degradation depends only on the number of companies dispatched, and not on which ones are dispatched. These observations suggest that an extension of of the formulation can readily be made to take into account specific differences between response time degradations if different companies are dispatched. The following section introduces some ideas that will be used to make this extension.

THE "WHICH COMPANIES" DECISION

We build here on the two-company model for deciding which companies to dispatch of Carter, Chaiken, and Ignall (1972). In it, one company is sent to each alarm, and the question is how to divide the region into "response areas." The geography is then explicitly identified, but the number dispatched is fixed and the region contains only two companies.

Straightforward extension of that model to situations with several types of alarms shows that, under certain conditions,[1] an optimal dispatch rule is of the following form. Company 1 rather than company 2 should be dispatched to an alarm of type m at location $\underline{x}$ if

[1] See, for example, Wrightson (1976) and Winston (1978).

$$p(\underline{x},\ m)t_1(\underline{x}) + \frac{\rho}{\rho + 1} T_1 \leq p(\underline{x},\ m)t_2(\underline{x}) + \frac{\rho}{\rho + 1} T_2 , \qquad (7)$$

where T_j is expected loss at an average alarm in the region if company j is not available (so that the other company serves the entire region), and $\rho = \lambda/\mu$, where λ is the alarm rate in the region and $1/\mu$ is the expected service time.

Note that Eq. (7) says that there may be incidents where the farther company should be dispatched (and the nearer one kept in its house). For example, suppose that company 1 is "better" than company 2, in the sense that its absence would give higher expected loss at the next fire (that is, $T_1 > T_2$). Then even if $t_1(\underline{x}) < t_2(\underline{x})$, it may be better to send company 2 to a type m incident at $\underline{x}$. This action would be preferable as long as

$$[t_2(\underline{x}) - t_1(\underline{x})] \leq \frac{\left(\frac{\rho}{\rho + 1}\right)(T_1 - T_2)}{p(\underline{x},\ m)} .$$

Thus the size of the time penalty one should be willing to pay on the current alarm $(t_1 - t_2)$ is larger (1) when the alarm rate (and hence ρ) is larger; (2) the more company 1 is better than company 2 (the larger $T_1 - T_2$); and (3) when the chance that the current alarm is serious $[p(\underline{x},\ m)]$ is small.

Equation (7) can be partially explained as follows. If company 1 is sent to the incoming alarm and does not return before the next alarm, then the next alarm will be served by company 2, and the loss will be T_1. If the alarms are Poisson and service is exponential, then $\rho/(\rho + 1)$ is the probability that the next alarm occurs before company 1 returns. Therefore, $\rho/(\rho + 1)T_1$ on the left-hand side of Eq. (7) is the expected loss at the next alarm if company 1 is sent in response to the incoming alarm. (We do not consider the loss at the next alarm if it occurs *after* company 1 returns because both companies will be available and the system starts over from scratch.)

Thus, the left-hand side of Eq. (7) can be seen as the sum of a current loss pt_1 and a future loss $[T_1\rho/(\rho + 1)]$ if company 1 is sent to the incoming alarm. The right-hand side is similarly a current loss plus a future loss, if company 2 is sent. It is this structure that we preserve in developing an algorithm that will handle a region with many companies and the possibility that more than one company will be sent to each incident. That is, we will calculate pt_i + $T_i\ \rho/(\rho + 1)$ for each company i in the region, and send the one with the smallest value of that sum.

Define $\theta_j = T_j\rho/(\rho + 1)$.[1] The larger the region considered in this computation, the smaller the role of any particular company in it and so the smaller the values of the $\{\theta j\}$. Thus the choice of region determines how the algorithm weighs current and future losses. To allow a range of tradeoffs, the computation of θ_j includes a region size factor, G. It turns out that θ_j is nearly proportional to G.

Using the above, it is easy to design a more general procedure that could choose a company to send for a dispatch of one unit. Current send-one loss would be given by Eq. (2); the future loss given by θ_j. So the algorithm would specify: Choose the company j* with the smallest value of

$$Z_j = C_1(j) + \theta_j \ . \tag{8}$$

The future loss with this procedure may be contrasted with that of the "how many" procedure of the previous subsection. With this method, future loss is company-specific and static in the sense that it does not depend on the current availability of other companies. With the "how many" algorithm, future loss is dynamic in its dependence on current availability, but does not depend on which company is chosen for dispatch.

θ_j turns out to be roughly proportional to the fraction of alarms within the region that are closest to company j. It thus serves as a measure of the relative amount of work that the company does in the region to which it belongs. We do not lose much by making this term

[1]The details of the computation of θ_j are given in App. A.

static. If a region's firefighting resources were depleted, the choice of a company dispatched to an incoming alarm would be more critical than if all the region's units were available. In this case, however, it is unlikely that we would make a decision to send other than the closest available company since the next closest company would probably be far away.

Finally, θ_j is in a sense a long-term cost since it considers the effects of having a company unavailable for some time, while the "how many" terms are operative only during the period r_{min}.

IV. THE INITIAL DISPATCH ALGORITHM

The future loss in the "how many" procedure formulated in Sec. III makes no distinction among the companies in a region. We will develop the initial dispatch algorithm by reformulating the "how many" choice, using notions of the "which companies" decision, so that future loss as well as current loss depends on the particular candidates for dispatch.

In choosing dispatches, it may be desirable to forgo strict optimality in terms of minimum average response times over a period of time by discounting the future losses. The reason for this is what we call the *principle of accountability*, which states that a public official will be held more accountable for what *has* happened than what *may* happen. If it turns out that a fire with a fatality was not sent the closest fire company, it will do the responsible official little good to argue that the closest company was being held in reserve for a potentially more serious fire.

Therefore, the algorithm incorporates a number of parameters and choices that can be adjusted so that individual dispatches will be defensible while average performance is improved. Section V shows the effects of adjusting these parameters on dispatch decisions. The algorithm performs the following steps separately for engines and for ladders:

- A unit that will definitely be sent to the alarm is selected. Because of the principle of accountability, parameters will be set so that this is almost certainly the closest unit for occupied structural fires.
- A candidate second unit is selected. This unit may not be the second closest available unit.
- A decision is made whether or not to send the candidate (second) company.

In mathematical terms we have the series of steps described below. (Note that times shown with a small t are retrieved from a data file; those shown with a capital T are calculated estimates.)

Step 1: Calculate

$$Z_1'(k) = q(\underline{x}, m)t_k(\underline{x}) + \theta_k , \tag{9}$$

and choose the company k* that minimizes Z_1'.

Equation (9) has the same form as Eq. (8); however, the current loss component is simpler than Eq. (2): u is eliminated as a constant term and β is set to equal 0. Setting $\beta = 0$ in this step completely discounts a second dispatch. This slight suboptimization will ensure, by the principle of accountability, that if exactly one company is actually sent, it is the best regardless of whether the value of β was correctly chosen. For the same reason, it may be desirable to delete θ_k (or give it less weight) in (9).

Step 2: After an engine and a ladder are selected in step 1, the response time of the first of them to arrive can be found by adding u to the shorter travel time. Call this r_{min}; it will be the response time of the first unit on the scene, except in a small number of cases where two units are sent and the closest unit is not chosen in step 1. Ignoring this possibility, as we do, will have a minimal effect on dispatching decisions.

Step 3: Calculate the send-one loss

$$Z_1(k^*) = q(\underline{x}, m)t_{k^*}(\underline{x}) + \theta_{k^*} + p(\underline{x}, m)\beta[t_i(\underline{x}) + r_{min}] , \tag{10}$$

where company i is the closest ladder other than k*. This step puts in the term that was left out in Eq. (9). Again, a constant term containing u is dropped.

Step 4: Given that we are sending company k*, we attempt to select the best company j to send as the second company if we send two companies. We calculate for all j ≠ k*:

$$Z_2'(k^*, j) = q(\underline{x}, m) \min [t_{k^*}(\underline{x}), t_j(x)]$$
$$+ \beta p(\underline{x}, m) \max [t_{k^*}(\underline{x}), t_j(\underline{x})]$$
$$+ \theta_j + [F_2(j) - F_1(j)]$$
$$+ [1 - p(\underline{x}, m)]W . \qquad (11)$$

The minimum and maximum terms are in Z_2 because if j is closer to $\underline{x}$ than k*, j will be arriving first if both units are sent. Future losses F_2 and F_1 are obtained from Eqs. (5a) and (5b) by considering an N-company region centered about company j. Thus the appropriate value of n depends on j, which is the reason for our notation $F_2(j)$ and $F_1(j)$. Including θ_j allows a company specific choice of the appropriate second ladder if one is sent. The workload factor W adds the workload "cost."

Select the company j* that minimizes $Z_2'(k^*, j)$. Note that W plays no role in this choice.

Step 5: In this step we calculate $Z_2(K^*, j^*)$, the actual send-two loss with candidate j*. This is the same as $Z_2'(k^*, j^*)$ in Eq. (11) except that θ_{j^*} is replaced by θ_{k^*} if company k* is closer than company j*. If $Z_2(k^*, j^*) \le Z_1(k^*)$, dispatch both k* and j*. If not, send only k*.

As we have mentioned, the algorithm has several parameters--β, r, W, and G--that enable the user to "tune" its behavior. From this viewpoint, the algorithm says (roughly):

a. How Many:

Send two if and only if

$$r_{min} \beta p(\underline{x}, m) > (F_2 - F_1) + [1 - p(\underline{x}, m)]W .$$

For a given incident (and its corresponding p and r_{min}), $F_2 - F_1$ is usually very small so that the ratio β/W usually determines how many companies are sent. The larger this ratio, the larger the average number of ladders dispatched.

b. Which Companies:

(i) If two companies are sent, the candidate second-arriving companies are ranked on the value of

$$\beta p(\underline{x}, m)t_k(\underline{x}) + \theta_k .$$

Since θ is almost proportional to G, the larger the ratio G/β, the more willing we are to choose, on the basis of future loss, a second ladder that is not the closest available.

(ii) If one company is sent, the candidate companies are ranked on the value of

$$\{p(\underline{x}, m) + r[1 - p(\underline{x}, m)]\} \; t_k(\underline{x}) + \theta_k .$$

Roughly, then, the larger the ratio G/r, the more willing we are to choose, on the basis of future loss, a first ladder that is not the closest one available.

As the following example shows, the algorithm will choose the more distant pair of companies only when their response times are nearly equal.

V. AN EXAMPLE

We now illustrate the algorithm with data for a hypothetical city. Table 2 shows how the ten ladder companies in a 36-square-mile city might share alarms and areas of coverage. We assume that the average proportion of alarms requiring two ladders is $\bar{p} = .013$ and that the

Table 2

DISTRIBUTION OF ALARMS AND VALUES OF θ IN A HYPOTHETICAL CITY

Company	Fraction of City's Alarms to Which Company Is Closest	Area of City in Which It Is the Closest Company (sq mi)	Values of $\theta(\times 10^{-3})$ for Two Times of Day	
			Busy (3 alarms/hr)	Not Busy (.86 alarm/hr)
1	.0625	4.0	1.05	.53
2	.0625	6.0	1.29	.65
3	.125	2.0	1.48	.74
4	.0625	3.6	1.00	.50
5	.0625	3.6	1.00	.50
6	.125	2.8	1.76	.88
7	.125	2.0	1.48	.74
8	.1875	2.8	2.64	1.32
9	.125	3.2	1.88	.94
10	.0625	6.0	1.29	.65

service time is one-half hour per incident. Using Eq. (A.1) for θ, described in App. A,

$$\theta_j = \frac{\lambda_j/\lambda}{1 + \mu/\lambda} \bar{q}(R_2 - R_1) = \frac{\lambda_j/\lambda}{1 + \lambda/\mu} \bar{q}(1/v)(c_2 - c_1) \sqrt{A_j} .$$

Substituting the above data and r = .01, (1/v) = 1.53 min/mile, $c_2 = 1.00$, and $c_1 = .60$, we get $\bar{q} = .0229$ so that

$$\theta_j = \frac{\lambda_j/\lambda}{1 + \mu/\lambda} .014 \sqrt{A_j} .$$

Table 2 gives the values of θ_j for a busy time of day with $\lambda = 3$ alarms/hr, and a quiet time of day with $\lambda = .86$ alarm/hr. Figure 2 shows a possible configuration of the companies.

To see how the first step of the algorithm works, suppose that an alarm comes in at location $\underline{x}$, somewhere between companies 2 and 4. The difference between the loss from sending company 2 (alone) and the loss from sending company 4 (alone) is given, using Eq. (9), by

$$Z'(4) - Z'(2) = \overline{q}(\underline{x}, m)[t_4(\underline{x}) - t_2(\underline{x})] + \theta_4 - \theta_2 .$$

So the algorithm will send company 4 if

$$q(\underline{x}, m)[t_4(\underline{x}) - t_2(\underline{x})] < \theta_2 - \theta_4 .$$

If $q(\underline{x}, m) = .01$ (the minimum if $r = .01$), then during a busy period company 4 would be sent if its travel time were no more than .029 min (about 2 sec) greater than that of company 2. During a nonbusy period, this difference could be no greater than .015 min. If the workloads of the companies were substantially different, or if the total alarm rate were greater, then this allowable difference would be larger; but it would never be substantial, and it would vanish for any incident with even a slight chance of being serious. For example, if $\underline{x}$ were between company 7 and company 8, in a busy time the former would be sent to an alarm with $q = .01$ if its travel time were no more than .116 min (7 sec) more than that of company 8.

Now assume that the incident had occurred at location $\underline{y}$, and company 1 is definitely chosen as a unit to send. The decision now is whether to choose company 2 or company 4 as the candidate second company. Equation (11) shows the calculation for each company. Since in this particular case we are considering t_{k*} as fixed, and since the city consists of only one region so that $F_2 - F_1$ is the same for all companies, using Eq. (11) as the algorithm says:

<u>Send company 4 if</u>

$$\beta p(\underline{y}, m)[t_4(\underline{y}) - t_2(\underline{y})] + \theta_4 - \theta_2 > 0 .$$

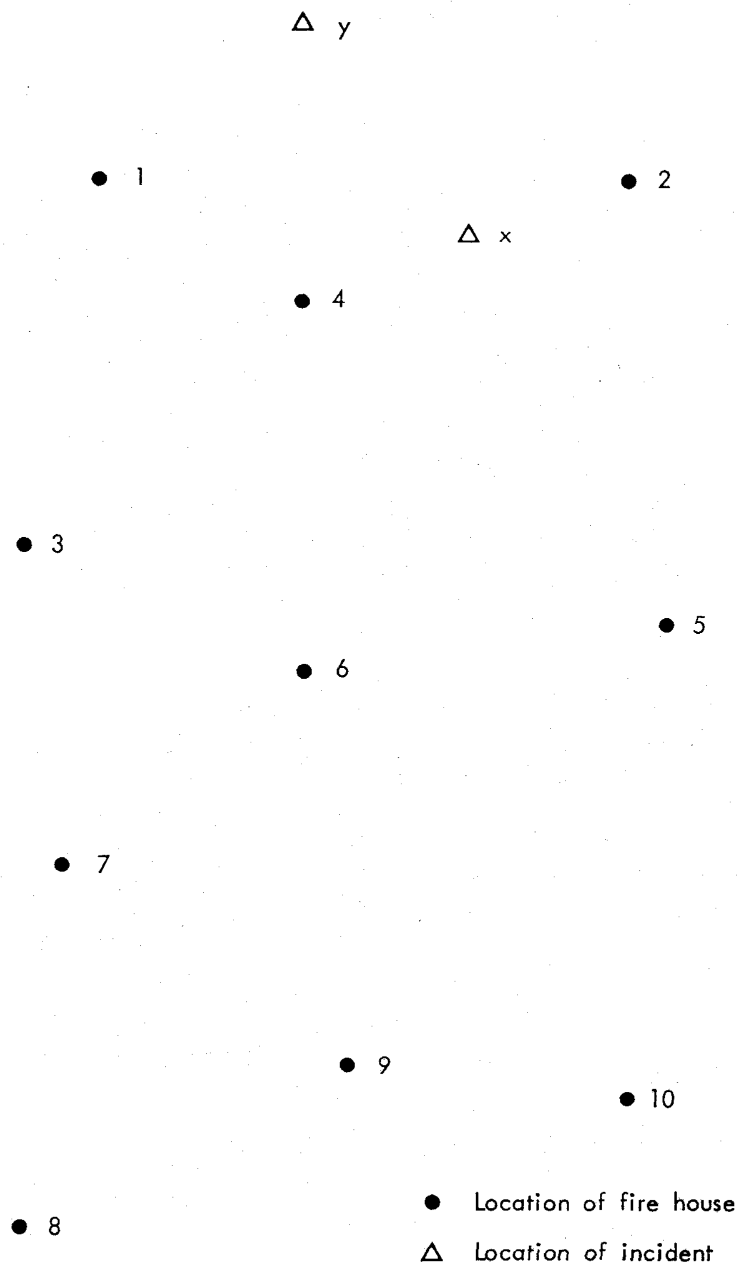

Fig. 2--Ladder company locations in a hypothetical city

As an example, assume that $\beta = .3$. If the probability of a serious fire is average (.013), company 4 will be chosen during a busy period if its travel time is no more than .075 min (4.5 sec) greater than that of company 2. Table 3 shows how the allowance travel-time difference varies by time of day and by the probability that the incident is serious. (If $p = 0$, then according to Eq. (11), company 4 would always be sent. But if $p = 0$, the algorithm would find $Z_2 > Z_1$ and therefore would not dispatch a second ladder.)

Table 3

ALLOWABLE TRAVEL-TIME DIFFERENCES
FOR SECOND-ARRIVING COMPANY

	Maximum Amount Travel Time of Company 4 Can Exceed That of Company 2[a] (min)	
Probability of a Serious Fire	Busy Time of Day	Nonbusy Time of Day
.01	.097	.048
.02	.048	.024
.03	.032	.018
.04	.024	.012
.05	.019	.010

[a]With company 4 still being chosen over company 2 to respond as second-arriving.

How does the algorithm decide whether or not to send the second company? Step 5 compares the send-two loss, Z_2, to the send-one loss, Z_1. Since the "definite" company 1 is closer than either company 2 or company 4, the comparison reduces to the following:

<u>Send company 2 if</u>

$$\beta p(\underline{y}, m) > \frac{[F_2(j) - F_1(j)]}{r_{min}} + \frac{[1 - p(\underline{y}, m)]W}{r_{min}} .$$

Using Eqs. (5a) and (5b), as well as the relevant approximations[1] for T_1 and T_2, we get:

$$\frac{F_2 - F_1}{r_{min}} = \begin{pmatrix}\text{Alarms}\\ \text{per}\\ \text{Minute}\end{pmatrix}\left(\sqrt{\frac{1}{n-2}} - \sqrt{\frac{1}{n-1}}\right)\left[\{\bar{q}C_1 + \bar{p}\beta[C_2 + (1-d)C_1]\}\ (1/v)\sqrt{\begin{matrix}\text{Area}\\ \text{of City}\end{matrix}}\right].$$

Under the assumptions of this section (and assuming $d = 1$ for convenience), the term in brackets is .162. Values of $(F_2 - F_1)/r_{min}$ are given in Table 4 as a function of n, the number of companies available before the dispatch is made,[2] for two values of the alarm rate.

First, suppose W is set equal to zero. From Table 4, if an alarm with probability serious = .013 (the average) is received (so that βp is .00390), the algorithm would send two ladders in all circumstances. Only if p were very low would less than two be sent. For example, if an alarm with probability serious = .003 is received, βp = .00090. The algorithm would send two ladders unless four or fewer ladders were available at the busy time of day.

If W is made positive, then two ladders will be sent less often. For example, suppose r_{min} = 2 min for this alarm at $\underline{y}$, and again suppose it has p = .013. Then, at the busy time of day, two ladders would be sent in all circumstances if W is less than .0031. So, if W = .0040, one ladder would be sent if three ladders were available; if W = .0060, one would be sent if three or four ladders were available; and so forth.

---

[1] See Sec. III and Apps. A and B.

[2] Since company 1 is being sent, and 2 and 4 are both being considered as second-arriving, n is at least three.

Table 4

EXAMPLE: INCREASE IN EXPECTED FUTURE LOSS IF TWO LADDERS RATHER THAN ONE ARE SENT INITIALLY

Number of Ladders Available Before Dispatch	Increase in Expected Future Loss	
	Busy Time of Day (3 alarms/hr)	Nonbusy Time of Day (.86 alarm/hr)
10	.00016	.00005
9	.00020	.00006
8	.00024	.00007
7	.00032	.00009
6	.00043	.00012
5	.00063	.00018
4	.00105	.00030
3	.00237	.00068

VI. RESULTS OF SIMULATION TESTS IN NEW YORK CITY

The summary response-time and workload statistics in Table 5 allow us to compare the performance of the traditional initial dispatch policy in New York City with the dispatch algorithm as various parameters of the algorithm are varied. These statistics come from simulation of department operations in response to actual alarms in the Bronx for the four weeks (672 hr) beginning July 5, 1972.

The simulation model used is the one first described in Carter and Ignall (1970) and documented in Carter (1974). The traditional policy was a variable two-engine and two-ladder response to street box alarms and telephone or verbal alarms that sounded structural.[1] Other telephone and verbal alarms were sent one engine and/or one ladder, as needed. The initial dispatch algorithm handled these other alarms in the same way as the traditional policy. For street box and telephone/verbal "sounds structural" alarms, the procedure given in Sec. IV was used. The values of $p(\underline{x}, m)$ for box alarms were estimated from 1967-1970 data using the empirical Bayes procedures described in Carter and Rolph (1973). For telephone/verbal alarms, boroughwide estimates of p were assumed to apply to all locations. In both cases, the values of p were adjusted to the time of day. Travel distances were measured using the coordinates of the fire station and the box in question, as described in Carter (1974). Travel times were determined from the square root linear time-distance relationship described in Kolesar and Walker (1974).

Alarm assignment cards were synthesized from these travel times. Actual alarm assignment cards are constructed using distances found by tracing a path through the streets on a map. (These distances are not recorded.) Since the distances had to be synthesized from (x, y) coordinates for the simulation, the use of actual cards would have handicapped the traditional policy, and it would sometimes appear *not* to be sending the closest companies.

---

[1] All telephone and verbal alarms that were structural, as well as those that were sent more than one engine and one ladder, were classified as "sounding structural."

Table 5

COMPARISON OF INITIAL DISPATCH POLICIES USING SIMULATION

		Under Traditional Policy	Under Policies with Approx. Same Total Workload as Traditional Policy: Always Send Closest Company			Sometimes Send a Farther Company	Under Policies that Vary the Total Workload (all with G = 1): Make More Responses		Save Responses
			G = 0	G = 1/2	G = 1	G = 2	W = 132	W = 265	W = 397
Response Time									
Average response times to fires in occupied structures (in minutes)	First-arriving engine	3.95	3.96	3.96	3.97	3.98	3.97	3.97	3.96
	Second-arriving engine (when needed)	5.57	4.87	4.75	4.81	4.87	4.55	4.85	5.57
	First-arriving ladder	4.01	4.02	4.03	4.04	4.04	4.03	4.04	4.03
	Second-arriving ladder (when needed)	5.49	5.04	5.02	5.03	5.07	4.59	5.01	5.47
Response Adequacy									
Number of fires in occupied structures that were dispatched	Two engines initially	457	589	596	598	599	653	588	513
	Two ladders initially	474	598	598	601	602	654	588	514
Workload									
Average number of responses by Bronx companies to all incidents in the 28-day period	Engines	[illegible]	318	321	323	326	357	314	283
	Ladders	[illegible]	369	373	375	377	413	361	325
Number of responses made by company that made the most responses	Engines	712 (E82)[a]	653 (E82)	623 (E85)	634 (E85)	638 (E85)	733 (E85)	637 (E85)	553 (E85)
	Ladders	714 (L31)	673 (L31)	655 (L58)	681 (L58)	697 (L58)	813 (L58)	674 (L58)	595 (L58)
Total operating hours of company that had the most responses	Engines	202.5 (E82)	195.4 (E82)	187.7 (E82)	183.8 (E82)	179.8 (E73)	183.2 (E82)	185.1 (E82)	179.9 (E82)
	Ladders	189.5 (L31)	188.0 (L31)	180.1 (L31)	177.1 (L42)	177.2 (L42)	180.3 (L42)	177.9 (L42)	174.5 (L42)
Relocations									
Total number of relocations made	Engines	372	294	283	284	288	293	287	288
	Ladders	226	43	39	38	41	22	43	66
Total time spent relocated (in hours)	Engines	335	374	368	361	373	359	363	350
	Ladders	280	67	63	60	66	42	68	97

[a] Engine Company 82. Other engines and ladders (L) are labeled similarly.

The relocation policy of Kolesar and Walker (1972) had already been seen as superior to the traditional alarm assignment card procedure, and so was likely to be adopted by the Fire Department when the MICS was installed. Thus we compared the performance of the traditional initial dispatch policy to that under the algorithm, when both used the Kolesar-Walker relocation policy.[1]

Table 5 shows that the algorithm improves second-arriving engine and second-arriving ladder response times to occupied structure fires that need them, while keeping workload about the same as under the traditional policy. These improvements result from the algorithm's attempt to consider the chance that the alarm is a fire in an occupied structure. Thus, although in the $G = 1/2$ case, for example, fewer responses are made with the algorithm, and so fewer incidents receive a two-engine and two-ladder initial response, more occupied structure fires receive two and two initially. In this case, two engines were initially dispatched to 457 of the 702 occupied structure fires under the traditional policy. With the algorithm, this figure increased 30 percent to 596 fires.

The slight degradation (about 1 to 2 sec) in first-engine and first-ladder times to occupied structure fires with the algorithm is a result of the closest company being in house slightly more often under the traditional policy. For ladders, one reason that the closest company was in-house more often is the more frequent relocation under the traditional policy. This policy resulted in at least five times as many ladder relocations as the algorithm did, and the total time spent relocated was at least four times as large.

The different values of the θ_j's induced by the differences in G led to very little variation in average response times. The smallest values of first-engine and ladder response time occur when $G = 0$, which makes all of the θ_j zero. The smallest values of second-engine and ladder response times occur when $G = 1/2$.

The different values of G have minor effects on average workload, as measured by the average number of responses made by Bronx engines

---

[1] For further details of the experimental conditions, see App. C.

and ladders. Among themselves and relative to the traditional policy, they do have a significant effect on the arrangement of that workload. Comparing the four values of G to the traditional policy we see that

1. Even when G = 0, the busiest engine and ladder companies make fewer responses and spend fewer hours traveling and at incidents.
2. As G is increased, the number of responses and operating hours of the busiest engine and ladder companies first decreases and then increases. Responses of the busiest engine and ladder companies are smallest for G = 1/2; operating hours of the busiest engine are smallest at G = 1/2; operating hours of the busiest ladder are smallest at G = 1. Note that when G = 1, the algorithm gave many responses to Engine 85, making it run more than Engine 82, but 82 still spends more time operating.

We have already mentioned that the algorithm substantially reduced ladder relocations. For engines, it reduced the number of relocations but not the time spent by relocatees filling in for other companies. Relocation implies a trip to another firehouse and back, and, while there, possible response to incidents in a less familiar area. Thus, all else being equal, less relocation is an advantage, and in this regard the algorithm then offers a significant advantage for ladders.

VII. ADAPTIVE RESPONSE

The simulation test results in Sec. VI show that the major savings in workload for the busiest company are obtained by using the algorithm with $G = 0$ (always sending the closest company). If alarm rates are moderate and relocations are not a problem, a static algorithm that does not depend on the number of companies currently available and that always sends the closest available companies may be satisfactory.

If we set $G = 0$ and let $F_2(j) = F_1(j)$, the initial dispatch algorithm reduces to:

1. Always send the closest available company.
2. Also send the second closest available company if

$$p(\underline{x}, m)\beta r_{min} > [1 - p(\underline{x}, m)]W ,$$

or, rearranging the terms, if

$$p(\underline{x}, m) > \frac{W}{\beta r_{min} + W} .$$

This motivates a simple static algorithm for initial dispatch--adaptive response.

An adaptive response policy is a specification of a particular fixed policy for each alarm. It is static in that the local situation--how many companies are actually available nearby right now--is not considered.

We evaluate a particular adaptive response policy in terms of

1. The fraction of serious incidents to which one ladder is initially dispatched, the fraction to which two ladders are dispatched, and so forth.
2. The number of responses made by companies in the region.

Our first measure, then, is response adequacy.

DEFINING AN ADAPTIVE RESPONSE POLICY

As an illustration, consider the policy to be used to dispatch ladder companies in response to *alarms reported by street box in a particular region at a specified time of day*. Assume that at least one ladder will always be dispatched, and that the problem is when to send a second ladder initially. An adaptive response policy will specify which alarm boxes will receive an initial dispatch of two ladders and which will receive only one ladder.

The procedure is the following. We label the boxes in the region (1, 2, 3, ..., N) in increasing order of the probability that an alarm from the box signals a serious fire. Let p_i be the chance that an alarm from box i signals a serious fire and λ_i be the hourly rate at which box alarms occur there. Our labeling implies that $p_1 \leq p_2 \leq p_3 \leq \ldots, \leq p_N$. The p_i and λ_i will be estimated values. Carter and Rolph (1973) give methods for getting them. Let

$$\lambda = \lambda_1 + \ldots + \lambda_N$$

be the total rate at which box alarms occur in the region. Then a cutoff probability is specified: this determines the boxes that will receive an initial dispatch of one ladder. If an alarm is received from box i and p_i is less than the cutoff value (s^*), one ladder is sent; if p_i is greater than or equal to s^*, two ladders are sent. The larger the value of s^*, the fewer the number of boxes to which two ladders are dispatched. By varying s^*, policies can be developed that dispatch an average number of ladders per box alarm anywhere in the range between 1.0 and 2.0. The ordered lists of alarm boxes make it easy to develop and evaluate such policies.

For example, suppose that $p_{m-1} < s^*$ and $p_m \geq s^*$. Then the average dispatch to a box alarm is given by

$$d(s^*) = 1 + \frac{\lambda_m + \ldots + \lambda_N}{\lambda} . \qquad (12)$$

The fraction of all the serious box incidents to which two ladders are dispatched is given by

$$a(s^*) = \frac{\lambda_m p_m + \ldots + \lambda_N p_N}{\lambda_1 p_1 + \ldots + \lambda_N p_N} . \quad (13)$$

So a(s*) measures response adequacy.

EXAMPLE OF A SIMPLE ADAPTIVE POLICY

The following example is based on data from New York City. Table 6 gives the information on the boxes. Empirical Bayes methods (Carter and Rolph, 1973) were used with 1967-1969 data to estimate the value

Table 6

1970 ALARMS FROM BOXES ASSIGNED TO "SUPER" BOX CLASSES[a]

Class	Box Alarms	Serious Occupied Structurals	Proportion Serious
1	3028	18	.0059
2	3017	33	.0109
3	3071	29	.0094
4	2905	43	.0148
5	3051	52	.0170
6	2954	61	.0206
7	3000	67	.0233

[a]Alarms from 3 p.m. to midnight for part of the Bronx. Classification based on 1967-1969 data.

of p_i for every box in the region. The boxes with the smallest estimated values of p_i were put in class 1 until that class of boxes had at least 3000 alarms between 3 p.m. and midnight in 1970; the boxes with the next smallest values were put in class 2 until that class had at least 3000 alarms; and so forth. These classes then define N = 7 "super" boxes, each with about 3000 box alarms per year in the 3 p.m. to midnight period (or .913 alarms/hr). So, for the super boxes,

$$p_1 = .0059, \ p_2 = .0109, \ \ldots, \ p_7 = .0233 ,$$

and

$$\lambda_1 = \lambda_2 = \ldots = \lambda_7 = .913 \ .$$

To calculate a, it is helpful to have the sum

$$\lambda_1 p_1 + \ldots + \lambda_N p_N = .0930 \ .$$

If s* = .005 is chosen, then two ladders are dispatched to all alarms and

$$d(s^*) = d(.005) = 2, \ a(.005) = 1 \ .$$

If s* = .008, then one ladder responds to alarms from box 1 and all others get two. As a result,

$$d(.008) = 1.857 \ (= 1 + 6/7) \ .$$

From Eq. (13),

$$a(.008) = 1 - \frac{.913 \times .0059}{.0930} = .942 \ .$$

That is, sending a second ladder 85.7 percent of the time gets that second ladder to 94.2 percent of the serious fires. Using a procedure that did not take box classes into account would require sending a second ladder to 94.2 percent of the incidents, significantly more than the 85.7 percent required by the adaptive policy.

Figure 3 shows how the adaptive policy compares to variable response policies over a range of average dispatch sizes. Fixed policies are represented by points A and B. Suppose that average ladder availability is .75. Then, according to Ignall and Urbach (1975), the probability that two units are dispatched is approximately the square root of the average availability. Thus, two ladders would be sent 57 percent of the time under a variable policy, so that its performance would be represented by point C. Note that moving to an adaptive

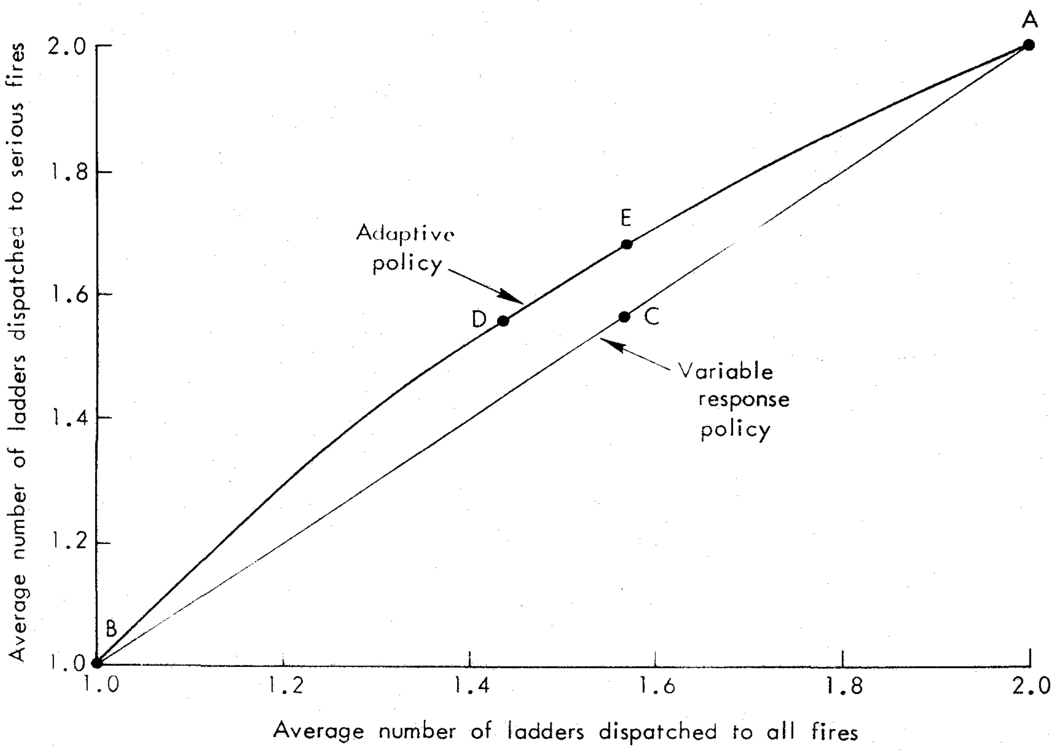

Fig. 3--Comparison of adaptive and variable response policies

response policy at point D will decrease initial ladder responses while keeping the same response adequacy as the variable policy. Also note that by moving to an adaptive response policy at point E, response adequacy is improved while initial ladder responses are unchanged.

DISCUSSION AND EXTENSION

We have looked at how the choice of s* induces a specific policy and the resulting value of d(s*), the average initial ladder dispatch to an incident, and a(s*), the average fraction of serious box incidents to which two ladders are initially dispatched. It is possible to choose either d or a, and let it determine s*. How to do so should be clear from the example.

So far we have ignored secondary responses to alarms where more than two ladders work and to alarms where two are needed and one is sent. We now show how to determine total ladder responses--including those to higher alarms and those where two are needed but one is sent.

Let π_j be the proportion of all box incidents at which j ladders work, and r_i be the proportion of box alarms at box i at which two or more ladders work. Then, the expected total hourly number of responses is given by

$$M(s^*) = \lambda d(s^*) + (\pi_3 + 2\pi_4 + 3\pi_5 + \dots)\lambda$$
$$+ (\lambda_1 r_1 + \dots + \lambda_{m-1} r_{m-1}) .$$

Note that the π terms do not depend on s*. Also, M(s*) can be rewritten as

$$M(s^*) = \lambda(\pi_1 + 2\pi_2 + 3\pi_3 + \dots)$$
$$+ [\lambda_m(1 - r_m) + \dots + \lambda_N(1 - r_N)] ,$$

where the last term gives the needless second ladder responses. (We are assuming that at least one ladder is necessary; that is, $\pi_0 = 0$.)

Recall that this discussion has been for ladder company responses to box alarms in a given period and region. We can combine different time periods and develop a single ranked list of possible alarms[1] and use a single cutoff. Or we can have separate cutoffs for different times of the day. A similar choice is possible for different regions.

Combination into a single list allows maximum benefit in the sense of highest response adequacy for a fixed number of company responses. But we may not wish to create such a list. Serious fires late at night can be different from serious daytime fires. Workload is not the problem late at night that it may be during the day. And, as described in Rider (1976), combining regions can lead to equity problems when companies are allocated to regions.

[1]It might look like "box 1, evening; box 2, evening; box 1, midday; box 3, evening; ...; box N, late night."

Appendix A

DETAILS OF ALGORITHM DEVELOPMENT

EVALUATION OF PROBABILITY THAT TWO LADDERS ARE SENT TO THE NEXT INCIDENT IF IT IS SERIOUS

The quantity called d is used in Eq. (4) to determine the expected response time to any alarm occurring during the interval r_{min} and depends on (1) the number of companies available during the interval, and (2) the number that would be sent initially to a new alarm.

We assume that the number of companies available will remain at $n - 1$ if one company is sent to the current incident and at $n - 2$ if two are sent.[1] We assume that between one and two units will be sent to any new alarms. The precise chance that it will be two depends on the algorithm itself. An approximation that will serve is to assume that the chance here is the chance that would apply if the variable response policy were used. From Ignall and Urbach (1975), the chance that two are sent in the future would be $[(n - 1)/N]^2$ if we send one now, and $[(n - 2)/N]^2$ is we send two now. Because the calculation is not sensitive to this factor, we use $[(n - 1.5)/N]^2$ in both cases.

CALCULATION OF EXPECTED LOSS AT THE NEXT INCIDENT IN THE REGION IF COMPANY j IS UNAVAILABLE

This quantity, called T_j, is used to determine which company (or companies) to send to an incident. The size of the region is determined by the user. We permit this by defining a factor G that can transform a particular region[2] into a smaller or larger region. G is then chosen by the user.

Expected future loss (in Eq. (7)) is $\theta_j = T_j\rho/(\rho + 1)$. We will find, in Eq. (A.2) below, that θ_j is approximately proportional to G.

---

[1] This ignores the possible return of companies working at other incidents during the interval. The calculation is not sensitive to this approximation.

[2] In New York City, we chose the area served by 8 ladder companies.

Consequently, choosing $G = 0$ will ignore the future entirely (in the "which" decision, but not the "how many" decision). Thus if Eq. (7) is used, the closest companies will always be sent.

We calculate T_j for company j by first considering the expected loss if company j is unavailable (and all others are available), and the closest remaining unit is sent to the next alarm in the region. We can subtract from this the loss at the next alarm if all companies are available (including j) since this will not depend on the choice of j. The result is the *increase* in loss to the next alarm if company j is not available. We call this T_j. (A detailed treatment of this loss would consider the possibility that company j was the second-closest company to a new alarm, third-closest, etc. For the purpose of making initial dispatch decisions, however, such refinements would probably make little difference.)

Following Eq. (1) for the first-arriving companies only, we take the increase in loss to the next alarm to be proportional to the increase in the response time to that alarm multiplied by the probability it is serious:

$$T_j = \bar{q}(R_2 - R_1) \quad \text{if the next alarm is in the area closest to company j;}$$

$$T_j = 0 \quad \text{otherwise,}$$

where R_1, R_2, and $\bar{q}$ are as defined previously. If company j is the closest ladder in a region of area A_j, the square-root law for distances applies and the distances are long, then $R_2 - R_1$ can be approximated by $(1/v)(c_2 - c_1)\sqrt{A_j}$, where v is average speed, and c_2 and c_1 are square-root law constants.[1]

The probability of the next alarm occurring in area j is given by λ_j/λ, where λ_j is the alarm rate in area j and λ is the alarm rate in the entire region. Therefore,

$$T_j = \frac{\lambda_j}{\lambda}\,\bar{q}(R_2 - R_1) \ .$$

[1]See Kolesar and Blum (1973).

Defining $\theta_j = \rho/(\rho + 1)T_j$ as the analog of future loss in Eq. (7), and substituting λ/μ for ρ, we have

$$\theta_j = \frac{\lambda_j}{\lambda + \mu} \bar{q}(R_2 - R_1) \ . \qquad \text{(A.1)}$$

Another question is what size to make the region *for the purpose of this calculation.* We fix a nominal region "size" of 8 companies and use a divisor G to allow regions of different sizes to be considered. That is, the number of companies in the region is 8/G, so that when G is 4, we are effectively considering a 2-company region; when G is .8, we have a 10-company region, etc. Letting λ be the alarm rate in the (nominal) 8-company region, the alarm rate in the effective area is roughly λ/G and, using

$$\frac{\lambda_j}{\lambda/G + \mu} = G \frac{\lambda_j}{\lambda + G\mu} ,$$

we get

$$\theta_j = G \frac{\lambda_j}{\lambda + G\mu} \bar{q}(R_2 - R_1) \ . \qquad \text{(A.2)}$$

When λ is much larger than μ, then θ_j is approximately proportional to the product of the alarm rate in "its" area, and the increase in first-arriving ladder response time there if it is unavailable. Also, a small value of G (a large region) implies that the T_j's are all small and thus very close to one another--the larger the region, the less any particular company matters to it. In the limit, when G = 0, the T_j are all zero.

Appendix B

ALGORITHM FLOWCHART

The routine described in the flowchart shown in Fig. B.1 implements the initial dispatch decision algorithm described in this report. For simplicity we assume that other subroutines will be used when the dispatcher has already decided on the number of units to dispatch, as would always be the case for follow-up dispatches. We assume further that the calling routine has verified the accuracy of the input data.

The routine assumes that there are two kinds of units--engines and ladders--and the index J is used to process each in turn. It also assumes the choice is whether to send one or two units.[1] The input argument BOX gives the location of the incident and is used as an index into the array EQ (I, J, BOX) which contains the index of the I*th* closest company of type J to the Box, I = 1, 2, ..., NDIM. The data on each company are stored in the arrays defined in Table B.1. The probability that the current alarm is serious is found in the variable P. The program uses a function: TRVLT (K, BOX) which calculates the travel time for company K to reach the location Box. In some cities it might be convenient to use an array that contains these distances when the companies are located at a fire station. The remaining data used by the program are the constants:

β = importance of second-arriving company response time relative to first-arriving response time.

r = importance of response time to a nonserious incident relative to response time to a serious incident.

U = turnout time.

W = workload factor.

N = number of companies in each N-company region.

[1] If the choice is whether to send H + 1 or H + 2 units, then choose the closest available H units and use the routine to decide whether to send 1 or 2 more, as well as which ones.

Table B.1

DATA REQUIRED FOR EACH COMPANY

AVAIL (K)	=	1 if company K is available for dispatch, 0 if otherwise.
AREA (K)	=	Area of the N-company region centered on company K.
LAMDA(K)[a]	=	Alarm rate in same region.
RING(K, M)	=	Identity of *mth* company in same region; m = 1, 2, ..., N.
PSER (K)[a]	=	Probability that an alarm in the ring will be for a serious fire.
THETA(K)[a]	=	Expected future loss if company K is not available (calculated from Eq. (A.2)).

[a]Should either be updated periodically to change with time of day or also indexed by time of day.

C_1 = square-root law constant such that the average first due response time in an area of size A when N companies are available is given by $C_1 \cdot \sqrt{A/N}$. Note that C_1 includes velocity.

C_2 = square root constant defined similarly to C_1 for second due response time.

C_M = square root constant defined similarly to C_1 for the average minimum response time ($C_M \le C_1$).

The square root approximation provides a simple means of calculating the expected increase in future loss if two companies are sent to the current alarm. We have, from Eqs. (5a) and (5b),

$$F_2 - F_1 = \lambda r_{min} \Big[\bar{q}[T_1(n-2) - T_1(n-1)] + \beta\bar{p}\{T_2(n-2) - T_2(n-1) + (1-d)[r_{min}(n-2) - r_{min}(n-1)]\} \Big] ,$$

where λ, $\bar{q}$, $\bar{p}$, $T_1(n)$, $T_2(n)$ and $r_{min}(n)$ refer to an N-company region surrounding the company that is a candidate for being sent as second due to this alarm. Using the square-root law to calculate $T_1(n)$ and $T_2(n)$, and assuming $r_{min}(n)$ is proportional to $T_1(n)$, this reduces to

$$F_2 - F_1 = \lambda r_{min} \left\{ \bar{q}C_1 + \beta\bar{p}[C_2 + (1 - d)C_M] \right\} \sqrt{A} \left(\frac{1}{\sqrt{n - 2}} - \frac{1}{\sqrt{n - 1}} \right),$$

where A is the area of the same N company region.

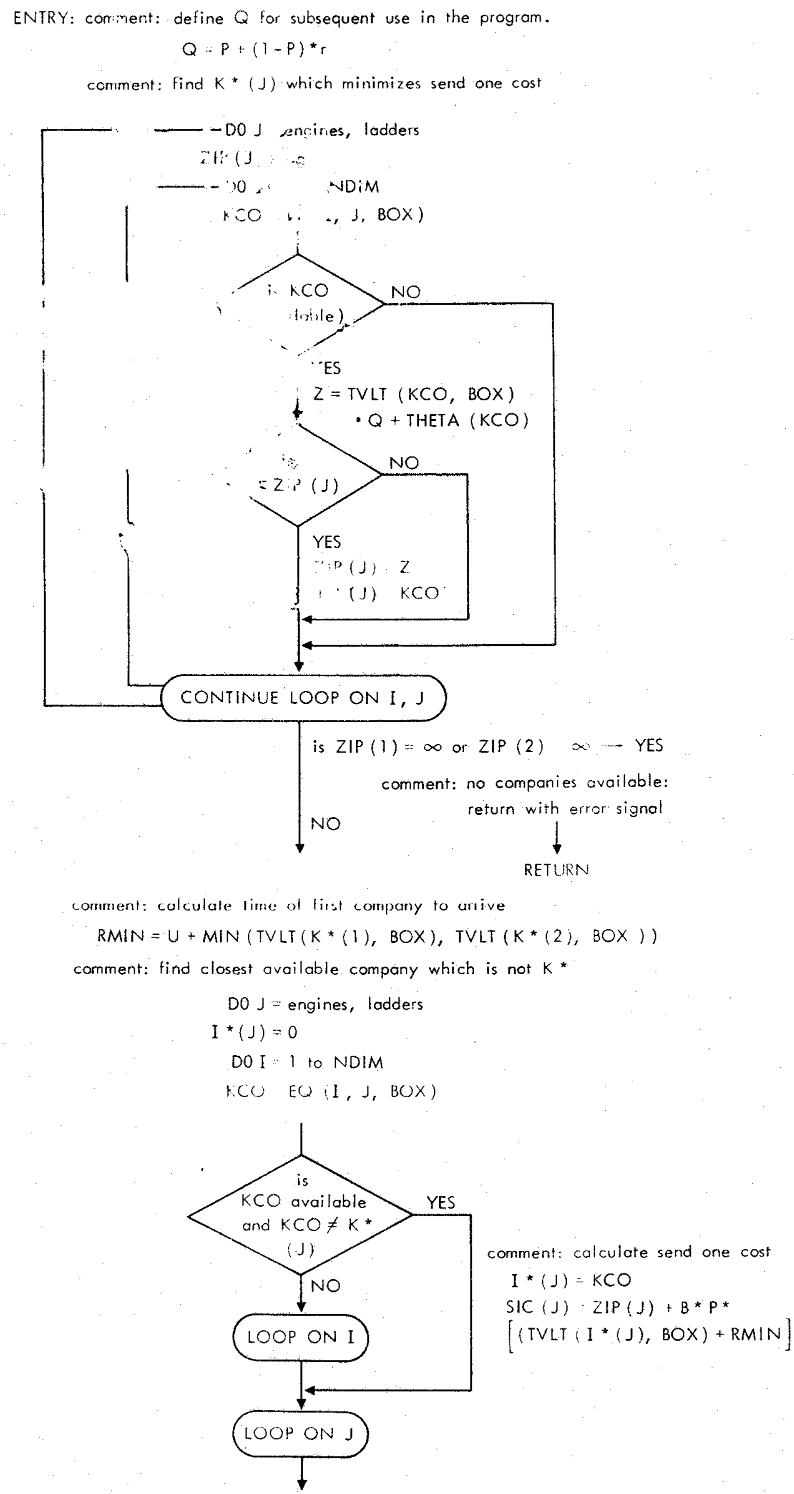

Fig. B.1--Algorithm flowchart

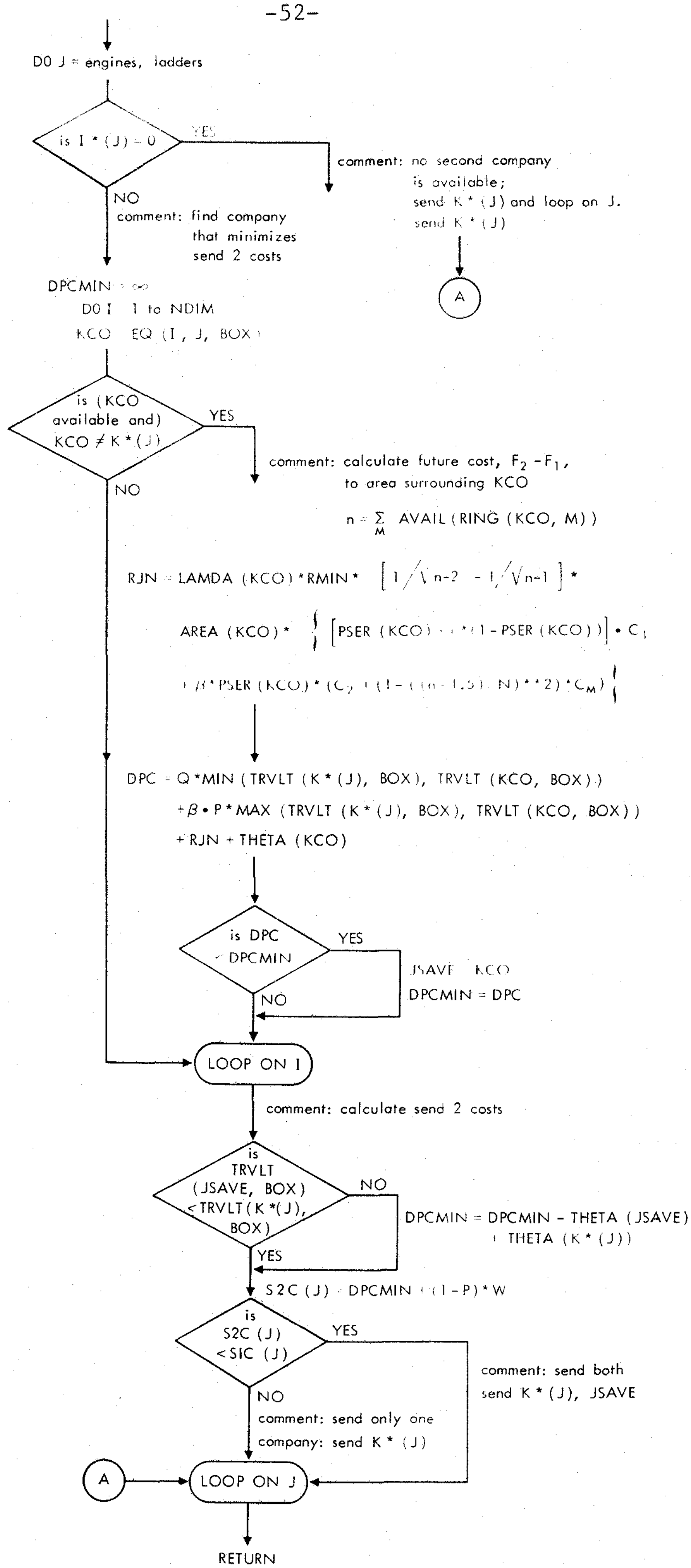

Fig. B.1--Continued

Appendix C

SIMULATION DETAILS

Section VI gave some information about the simulation experiment used to test the initial dispatch algorithm. This appendix supplements that information. We assume that the reader is familiar with the simulation program described in Carter (1974).

Parameter values of $\beta = .4$, $r = .01$, and $N = 8$ companies were chosen.

The exogenous events tape was prepared from a file on actual incidents in the Bronx and Manhattan. All Bronx incidents in the 28-day period (7-5-72 to 8-1-72) were included. If the box number reported for an incident was not legitimate, a box was chosen at random from the list of legitimate boxes and the incident assigned there. Eighty-four of the 6905 incidents in the Bronx were so assigned. Manhattan incidents at boxes numbered higher than 1500 (locations north of 127th Street, approximately) triggered DOWN events (working one engine and one ladder from Manhattan, chosen at random, for a time appropriate to the incident type). Some 50 actual Manhattan incidents were lost because the records were not readable.

The number of engines and ladders working at actual incidents were adjusted as Table C.1 specifies. Note that the next to the last "condition" is a second alarm, and that all larger incidents are treated as if they were third alarms. (As it happens, there were no incidents more severe than third alarms.) Table C.1 also gives work times and other data used in the simulation.

Engines and ladders in upper Manhattan were available to respond to Bronx alarms or relocate to the Bronx unless busy at DOWN events. The 32 engines and 26 ladders simulated in the Bronx were those companies that existed in early 1975.

Table C.1

SIMULATION DATA

x = number of engines working on tape
y = number of ladders working on tape

Condition	No. of Higher Alarms	Number to Work in Simulation		Release Mode	Arrival to Call (min)	Time Spent Working (min)	
		Engines	Ladders			Engines	Ladders
False alarm	0	1	1	1	1	4	4
If $x + y \leq 2$	0	x	y	1	1	Use 18 min for whichever companies work	
$x = 1 \quad y = 2$	0	1	2	2	0	60	60, 40
Other $x + y = 3$	0	2	1	1	1	30, 20	30
$x + y = 4$	0	2	2	2	0	75, 45	75, 45
$x + y = 5$ or 6	0	3	2	2	0	150, 105, 60	150, 90
$7 \leq x + y \leq 10$	1	7	3	2	0	240, 180, 120, 90, 90, 60 60	180, 135, 105
$x + y \geq 11$	2	11	4	2	0	360, 300, 270, 240, 240, 180, 150, 120, 120, 90, 90	330, 270, 180, 135

REFERENCES

Carter, G., *Simulation Model of Fire Department Operations: Program Description,* The Rand Corporation, R-1188/2-HUD/NYC, December 1974.

Carter, G., J. Chaiken, and E. Ignall, "Response Areas for Two Emergency Units," *Operations Research,* Vol. 20, No. 3, May-June 1972, pp. 571-594.

Carter, G., and E. Ignall, *A Simulation Model of Fire Department Operations: Design and Preliminary Results,* The Rand Corporation, R-632-NYC, December 1970. Published in *IEEE Transactions on Systems Science and Cybernetics,* Vol. SSC-6, No. 4, October 1970, pp. 282-293.

Carter, G., and J. Rolph, *New York City Fire Alarm Prediction Models: I. Box Reported Serious Fires,* The Rand Corporation, R-1214-NYC, May 1973.

-----, *New York City Fire Alarm Prediction Models: II, Alarm Rates,* The Rand Corporation, R-1215-NYC, January 1975.

Hogg, J., *Losses in Relation to the Fire Brigade's Attendance Time,* Fire Research Report 5/73, Home Office, London, 1973.

Ignall, E., P. Kolesar, A. Swersey, W. Walker, E. Blum, G. Carter, and H. Bishop, "Improving the Deployment of New York City Fire Companies," *Interfaces,* Vol. 5, 1975, pp. 48-61.

Ignall, E., K. Rider, and R. Urbach, *Fire Severity and Response Distance: Initial Findings,* The Rand Corporation, R-2013-NYC (forthcoming).

Ignall, E., and R. Urbach, *The Relationship Between Firefighting Unit Availability and the Number of Units Dispatched,* The Rand Corporation, P-5420, May 1975.

Keeney, R., "A Utility Function for the Response Times of Engines and Ladders to Fires," *Journal of Urban Analysis,* Vol. 1, No. 2, 1973, pp. 209-222.

Kolesar, P., and E. Blum, "Square Root Laws for Fire Engine Response Distances," *Management Science,* Vol. 19, No. 12, September 1973, pp. 1368-1378. Revised and updated version, *Square-Root Laws for Fire Company Travel Distances,* The Rand Corporation, R-895-NYC, June 1975.

Kolesar, P., and W. Walker, *An Algorithm for the Dynamic Relocation of Fire Companies,* The Rand Corporation, R-1023-NYC, September 1972. Published in *Operations Research,* Vol. 22, No. 2, 1974, pp. 249-274.

-----, *Measuring the Travel Characteristics of New York City's Fire Companies,* The Rand Corporation, R-1449-NYC, April 1974.

Rider, K., "A Parametric Model for the Allocation of Fire Companies in New York City," *Management Science,* Vol. 23, No. 2, 1976, pp. 146-158.

Swersey, A., "A Markovian Decision Model for Deciding How Many Units to Dispatch," in "Models for Reducing Fire Engine Response Times," Ph.D. Thesis, Columbia University, New York, 1972.

-----, *Reducing Fire Engine Dispatching Delays,* The Rand Corporation, R-1458-NYC, December 1973.

Swersey, A., E. Ignall, H. Corman, P. Armstrong, and J. Weindling, *Fire Protection and Local Government: An Evaluation of Policy-Related Research,* The Rand Corporation, R-1813-NSF, September 1975.

Winston, W., "Optimal Assignment of Customers in a Two-Server Congestion System with No Waiting Room," to appear in *Management Science,* 1978.

Wrightson, C. W., *Stochastic Models for Emergency Service Systems,* Operations Research Center, ORC 76-18, University of California at Berkeley, 1976.