Simulation Tool for Evaluating Effectiveness of Freeway Incident Response Operations

TETI G. NATHANAIL AND KOSTAS G. ZOGRAFOS

A methodology for evaluating the effectiveness of freeway incident management (FIM) systems is presented. Evaluation of the effectiveness of a FIM system requires the establishment and estimation of the measures of effectiveness. For the estimation of the measures of effectiveness of a freeway incident response system, a simulation model that focuses on the operations of freeway emergency response units (FERUs) has been developed. Various aspects of the system are covered. First, an analysis of the arrival rates of mobile servers to the incidents to be serviced is done. These incidents occur randomly, follow discrete temporal and spatial distributions, and are of different natures and severities. Then, on the basis of the number of the mobile servers, the area in which these servers provide service to emergency calls is divided into smaller districts and the mobile servers are assigned to each one of the districts. The proposed simulation model generates freeway incidents according to their spatial, temporal, and severity characteristics. Alternative dispatching policies have been integrated in the model for the assignment of FERUs to incidents. The proposed model has the ability to estimate the total time that has elapsed between the occurrence of the incident and the completion of the service and to provide alternative emergency response policies. The value of alternative measures of effectiveness is calculated as a function of the total service time, and the performance of these policies is evaluated.

In many real-life events an emergency response is required. Examples of such events are accidents involving vehicles or pedestrians on the road network, medical emergencies, fire, failure of electric and other utility networks, and others. These events occur at a time and place that are not known in advance. These events are referred to as incidents. Depending on the nature of the incident, specialized authorities are called to provide assistance and service the incident.

An emergency response system is considered efficient when it provides fast and appropriate service to the incidents that need to be serviced. The main attributes of an emergency response system are the appropriate equipment, specialized personnel, and the time required for service to be provided to the incident. Incident restoration time is considered the time that elapses between the incident occurrence and the completion of the provided service. The restoration time can be divided into four components, as shown in Figure 1 (*1*):

- Time T_1 is the time from the incident occurrence until the notification of the response authority about the incident, or the detection time. Identification of the incident's characteristics, such as the type of incident and its location and severity, is included in the detection time.
- Time T_2 is the time that has elapsed between the notification of the responding authority about the incident and the allocation or assignment of the most appropriate server to the restoration of the incident, or the dispatch time.
- Time T_3 is the time that has elapsed between the assignment of the response unit and the arrival of the unit at the site of the incident, or the travel time to the incident scene.
- Time T_4 is the time required for all the necessary activities undertaken by the response unit to completely restore the incident, or the service time. Depending on the nature of the incident, service time can be divided into on-scene time (if special actions are required at the scene of the incident) and removal time (if transfer of the involved person or equipment to some other location is necessary).

The study described here presents a tool for evaluating alternative policies of dispatching and allocating emergency response units. More specifically, the proposed simulation tool has the capacity to examine alternative dispatching and assignment strategies of freeway emergency response units (FERUs) and to estimate the results of the incident restoration time. For this purpose the study focuses on the response operations as they are conducted in an area where incident occurrences generate a demand for service. A microscopic simulation model has been developed. The model assesses the roadway network characteristics and the response operations according to the selected policy to be applied. Thus, it generates the calls for service in time and space and identifies their priority for service. Furthermore, it selects the most appropriate server, according to a dispatching policy, and assigns the incident to the server, providing routing instruction to the FERUs. Finally, it estimates the time to the completion of the service and directs the server to the next incident location if there is another incident waiting to be serviced. The restoration time for each incident and the total restoration time are the measures of effectiveness applied in a case study to compare the different response policies tested.

EMERGENCY RESPONSE OPERATIONS

Emergency response systems can be divided into two major categories:

- Response systems in which the service is provided at a specific place, such as a hospital, and
- Response systems in which the servers travel to the point of the demand and either provide service at the site of the call or transfer the involved person or equipment to a servicing facility.

T. G. Nathanail, TRD International S.A., 18, Navarinou Square, Thessaloniki, 54622, Greece. K. G. Zografos, Department of Operations Research and Management, University of Patras, Rion-Patra, 26500, Greece.

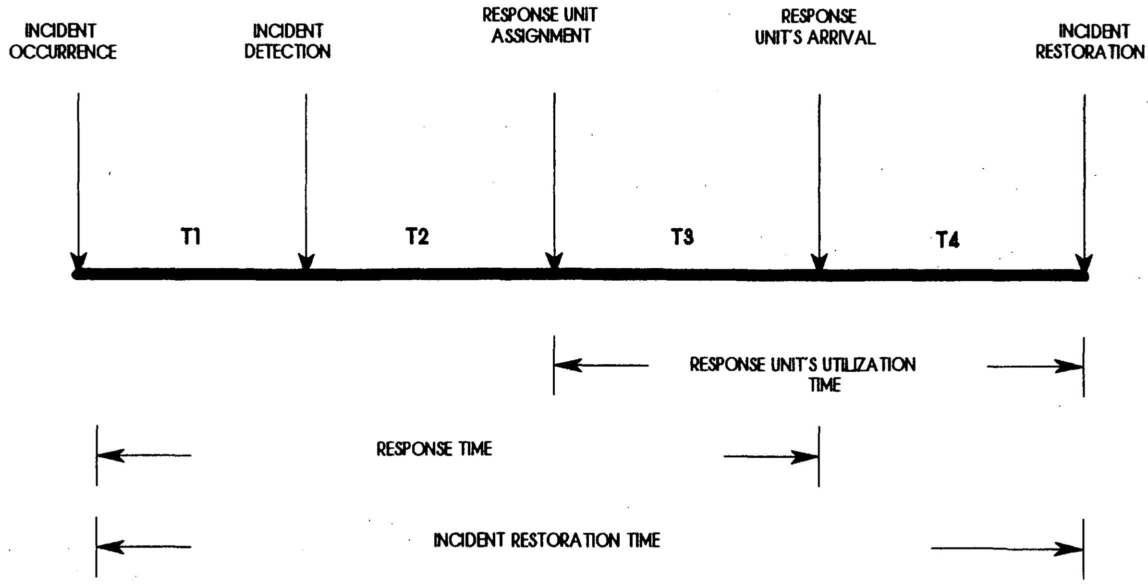

FIGURE 1 Time components of incident duration.

In the present study emphasis is given to the second category, in which equipped vehicles arrive at the scene of an incident to provide the necessary assistance. The sequence of the servers' operations for the incident clearance is the main attribute of a dispatching policy. The first in–first out (FIFO) and the nearest neighborhood (NN) policies are the ones tested in the present case study (*1,2*).

FIFO dispatching policy dispatches the server to the point where demand for service occurs first. Two disadvantages characterize the FIFO policy. The first is that under high workloads there is an excess delay to the servicing of other calls because of the travel time spent by the server. Also, sites with demand for service may be overpassed by the server because there is another call for service that arrived first.

The NN policy seems to result in less waiting time for service under high workloads because it dispatches the server to the closest location with a need for assistance, regardless of the time of occurrence of the calls.

Variations of these dispatching policies require that an area be divided into districts and that each district be assigned to one and only one server or require the accumulation of demand before a tour is being scheduled for a server (traveling salesman policy) (*2*).

Finally, a different priority for servicing the calls may be given on the basis of the type of call and the urgency and the severity of the incidents (*1,3*).

EMERGENCY RESPONSE METHODOLOGY: A GENERAL APPROACH

Simulation of incident restoration operations is a complicated process that includes the generation of incidents and their characteristics, the representation of response unit operations, and the estimation of the total incident duration and its components. A flow diagram of the model simulating the operations of the incident response units is shown in Figure 2.

The first module of the simulation model is the generation of incidents by their characteristics, such as time, location, and type of incident. The following assumptions have been considered for the incident generation module:

- Incidents follow a Poisson distribution in terms of time of occurrence, which is described by an interarrival time.
- The distribution of incidents in terms of location of occurrence is assumed to follow a uniform distribution. The locations of the area under study that have a higher probability of incident occurrences can also be considered.
- The incidents are generated by type and attributes. An indicator of their priority order is also given.

In the second module the incidents are assigned to servers depending on the jurisdictions of the possible servers. The jurisdictions of the servers are defined in advance by the control center or may vary depending on the type of emergency. Districting models for areas with incident occurrences to minimize the overall servicing time have been proposed by Toregas et al. (*4*), Marlin (*5*), Nathanail (*6*), and Zografos et al. (*1*). If the server is available the incident response process commences; otherwise, the incident is placed in the server's queue and waits until the server is available.

In the third module, the appropriate server assigned to incidents is dispatched to the incident site. The travel time of the server to the location of the incident is then computed as a function of the traffic volume at the time of the assignment and the geometric characteristics of the sections of the roadway network that constitute the fastest path from the location of the server to the incident site. The attributes of the roadway network (geometric characteristics and traffic volumes at certain times of the day to estimate the actual travel speed of the mobile server) exist in a data base of the network used for the incident response process.

After the arrival of the server at the location of the incident, the on-site and off-site response operations are described by the fourth module of the simulation model. Depending on the type of incident, the on-site response time is estimated, assuming an average response time for each type of incident. If the vehicles involved in the incident must be transferred to a servicing facility (e.g., hospi-

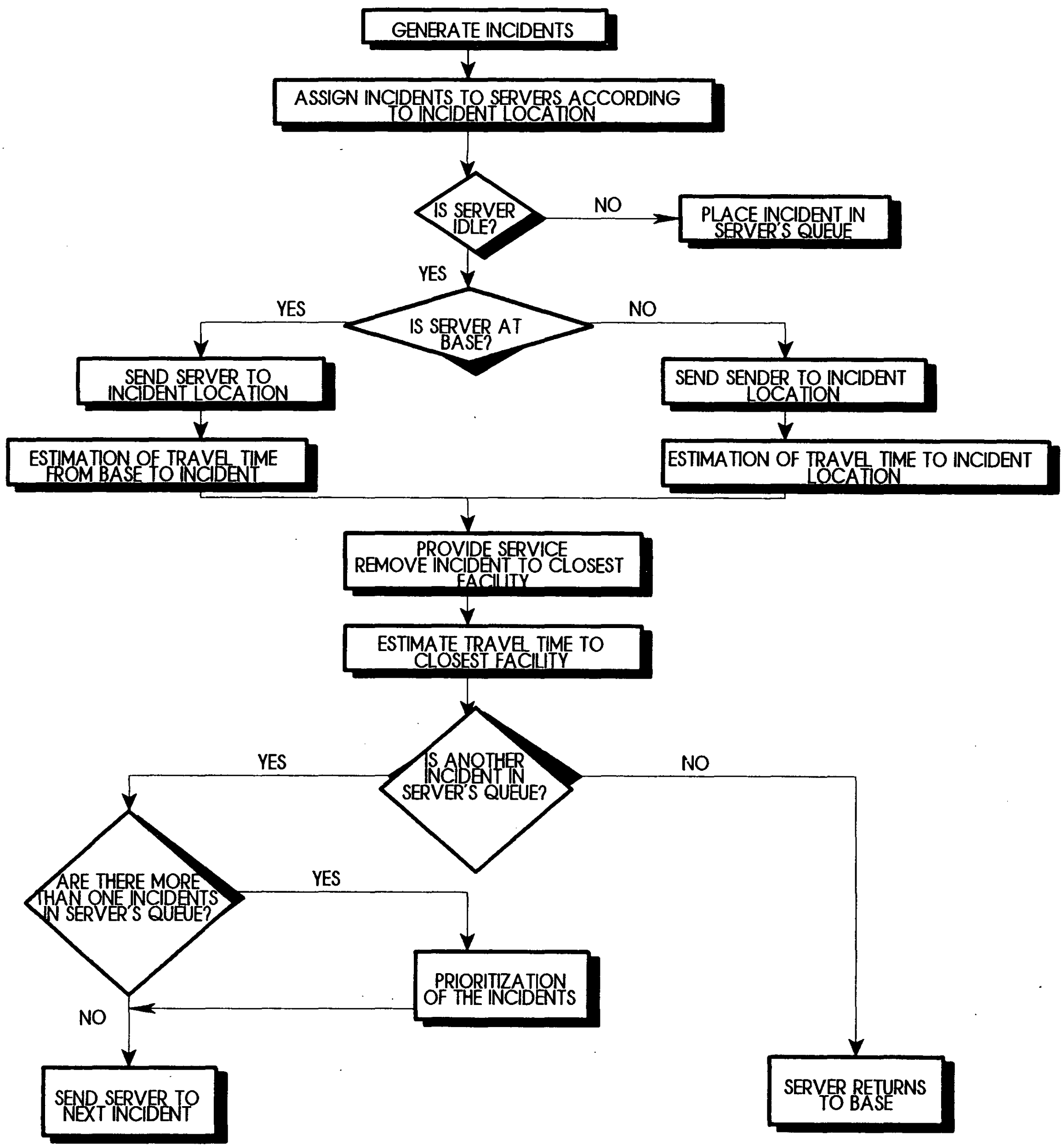

FIGURE 2 Simulation of incident response unit operations.

tal, garage, and accident investigation site), the facility is chosen among the available ones in the area and according to the dispatching policy on the basis of the criterion of the shortest path in terms of travel time from the incident site. The on-site response time and the travel time to the closest facility are then computed.

The server is considered available to service another incident after the completion of the required operations for the restoration of the previous incident. If an incident is in the server's queue, then the server is directed to the incident site and the server follows the same procedures described earlier. If more than one incident is in the server's queue, then these incidents are given a priority index according to their type and the dispatching policy, and the server is directed to the incident with the higher priority. If no incidents are waiting for service in the server's queue, then the server is directed to its initial position.

ROADWAY INCIDENTS AND THEIR CHARACTERISTICS

The main attributes of incidents are the time of occurrence, the location, and the severity. Incidents appear randomly in terms of their characteristics, which implies that incident response is based on online information.

For the purposes of applying the methodology described earlier, a case study in which an incident occurs on a roadway network is considered.

TABLE 1 Probability of Incident Occurrences on Four-Lane Highways with Adequate Shoulders

LOCATION	TYPE OF INCIDENT	NUMBER OF BLOCKED LANES	PROBABILITY OF OCCURRENCE
Lane	Accident	one	.72%
		two	.11%
		three	.02%
	Disablement	one	3.12%
		two	.03%
Shoulder	Accident	-	4.03%
	Disablement	-	91.97%

Incidents occurring on the roadway network can be divided into two categories in terms of severity, accidents, and other incidents. Accidents are incidents that involve personal injury or death. Incidents are minor events such as vehicle disablements (breakdowns) and spills. Furthermore, incidents are divided into categories that are related to the size of traffic blockage that they cause, which indicates an index of severity as well. This is expressed as the share of the roadway that they block, that is, the number of lanes. Incidents may occur on the shoulder, or they may occupy one, two, or more lanes of the highway. Finally, the time required for the clearance of the incident scene is another factor that differentiates incidents by severity. Incident clearance time is a direct derivative of the type of service required by the incident so that the situation is restored. Two types of service can be outlined. The first type of service is provided on site. The server is released after the completion of on-site restoration of the incident and is available to respond to another incident. The second type of service requires transfer of the vehicle involved in the incident to a service facility (i.e., accident investigation site, garage, etc.). Although the type of service required characterizes the incidents, it is mainly an attribute of the dispatching policy and depends on the structure of the response system and the equipment of the response unit.

In the present case study the incidents occurring on the roadway network were categorized by the number of lanes that they blocked. For a highway network consisting of four-lane highways with adequate shoulders, Lindley (*7*) has done an incident analysis in which the probability of incident occurrences by location and severity is estimated (Table 1).

In the same study (*7*) mean detection and clearance times were estimated by type of incident for incidents occurring on a highway network (Table 2). Because no data on the clearance times required by the shoulder accidents and disablements were available, for the purposes of the case study these times can be assumed to be 10 min for shoulder accidents and 5 min for shoulder disablements. Note that the detection and clearance times were considered to be given, whereas the dispatch and travel times were estimated by the simulation model, assuming a certain dispatching policy.

The incident arrival rate was estimated to follow a Poisson distribution, and a generation rate of 20 incidents per 1 million vehicle miles was assumed. This number is only an indication and may be estimated in each case on the basis of historical incident and traffic data collection.

CASE STUDY AND RESULTS

The methodology for evaluating alternative policies for dispatching emergency response units to incidents occurring on a roadway was tested on a hypothetical roadway network for the FIFO and NN dispatching policies and for the allocation of one, two, three, and four incident response units. The network consisted of a highway connected to an arterial by entrance and exit ramps. The time frame of the simulation was a 3-hr period during the evening peak. The generation of the incidents in time and space and their characteristics followed the distributions given earlier. Setting the starting time of simulation at 0:00 hr and the beginning of the corridor on which incidents occur at 0.00 m, the incident generation module resulted in the incident characteristics described in Table 3.

Because of the complexity of the incident restoration problem on a highway network, the applied policy should also consider the prevailing situation on the network, such as traffic volumes and the time of incident occurrence (peak versus off-peak hour) to provide adequate service to the persons involved in the incident and the rest of the users of the network by minimizing travel delays. However, consideration of the minimization of motorist delay because of lane blockage caused by incidents is not described here and can be found in other reports by the authors (*1,6*).

The average incident duration and the time components for the two applied dispatching policies and for different numbers of response units are given in Figures 3 and 4. In all cases it was

TABLE 2 Detection and Clearance Times by Type of Incident on Highways

TYPE OF INCIDENT	DETECTION TIME (min)	CLEARANCE TIME (min)
one-lane accident	10	10
two-lane accident	10	15
three-lane accident	10	20
one-lane disablement	10	5
two-lane disablement	10	10
shoulder accident	10	N/A
shoulder disablement	10	N/A

TABLE 3 Result of Simulation Module for Incident Generation Simulation

Incident identification number	Time of incident appearance (hour:min:sec)	Distance of incident from beginning of corridor	Type of incident
1	0:00:00	31485.84	shoulder disablement
2	0:26:41	4852.416	shoulder disablement
3	0:39:17	34686.24	shoulder disablement
4	0:56:00	34015.68	shoulder disablement
5	0:58:23	32125.92	shoulder disablement
6	1:05:30	34046.16	shoulder disablement
7	1:10:13	6769.61	shoulder accident
8	1:12:40	42885.36	two-lane accident
9	1:35:26	32796.48	one-lane disablement
10	2:00:24	8436.86	one-lane disablement
11	2:01:12	7336.54	shoulder disablement
12	2:02:00	33863.28	shoulder disablement
13	2:02:18	32583.12	shoulder disablement
14	2:14:24	4346.45	shoulder disablement
15	2:20:24	20141.18	shoulder disablement
16	2:26:54	3154.68	shoulder disablement
17	2:36:18	22792.94	shoulder disablement
18	2:43:48	3602.74	shoulder disablement
19	2:46:48	2847.14	shoulder disablement
20	2:57:24	32461.20	shoulder accident
21	2:59:42	35783.52	shoulder disablement

assumed that the detection and clearance times are not affected by the dispatching policy, and their values were obtained from the literature, (Table 2). The general trend shown in Figures 3 and 4 indicates that the total incident duration decreases as the number of response units increases. Application of the FIFO policy resulted in a 73 percent reduction in incident duration when the number of response units was increased from one to two, a 22.1 percent reduction when the number of units was increased from two to three, and an 18 percent reduction when the number of units was increased from three to four. Similarly, when the NN policy was applied, there was a 66 percent reduction in incident duration when the number of units was increased from one to two, a 17.2 percent reduction when the number of units was increased from two to three, and a 17.8 percent reduction when the number of units was increased from three to four. Furthermore, use of FIFO policy resulted in a 32.6 percent longer incident duration than use of the NN policy, resulting in a high utilization rate of the response unit by FIFO policy, as indicated in the study of Bertsimas (*2*).

Since application of the different dispatch policies mainly affects dispatch and travel times, an analysis of the share of these two time components was conducted. Results of such an analysis are given in Figures 5 and 6, which indicate the percentages of incidents for which the travel and dispatch times were within a certain range of the total incident duration. The general trends of the analysis show that with both dispatching policies, dispatch and travel times constituted an important portion of the total incident duration. In more than 50 percent of the observations, the dispatch and travel times constituted 60 percent of the total incident duration when one unit was assigned to respond to the incident. For more than one response unit the dispatch and travel times constituted an average of 20 percent of the total incident duration in most observations.

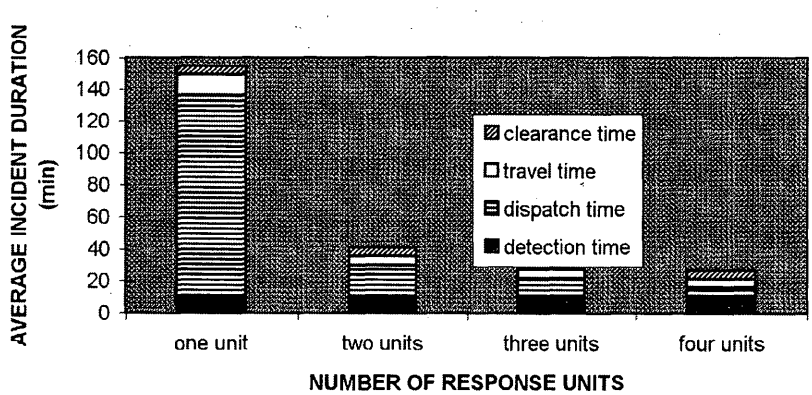

FIGURE 3 Average incident duration for FIFO dispatching policy.

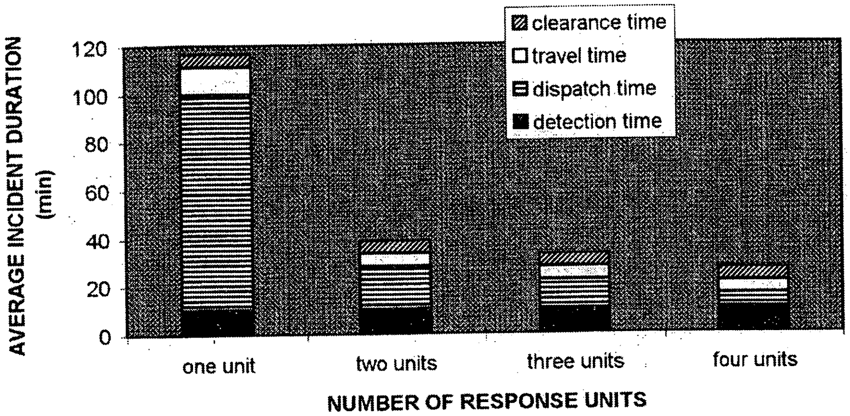

FIGURE 4 Average incident duration for NN dispatching policy.

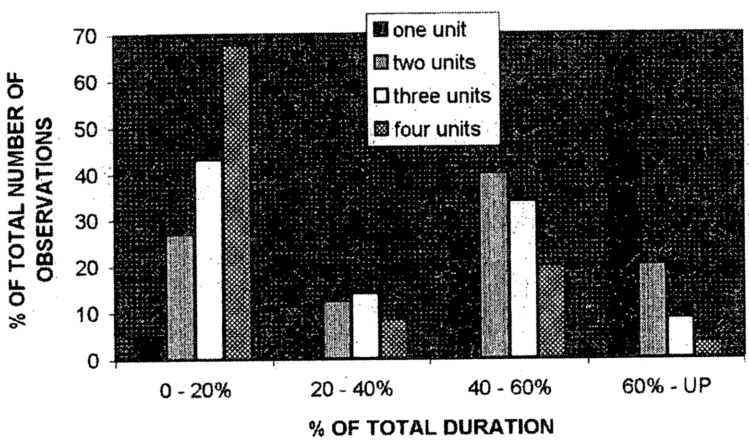

FIGURE 5 Average dispatch and travel time shares for FIFO policy.

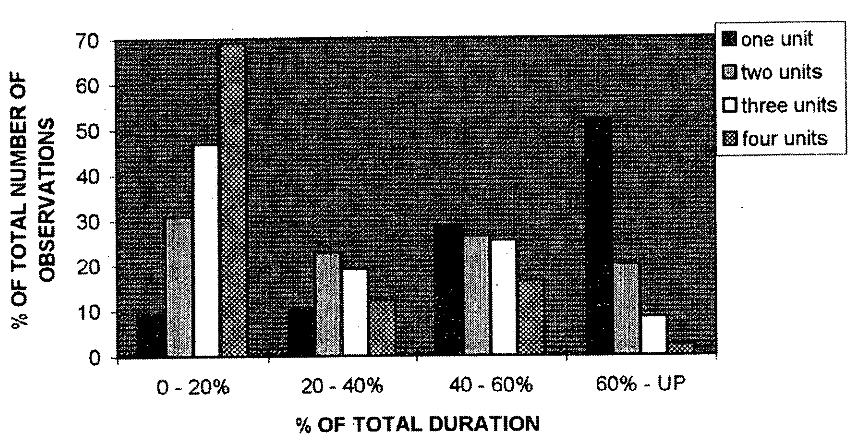

FIGURE 6 Average dispatch and travel time shares for NN policy.

CONCLUSIONS

A methodology for evaluating alternative policies of dispatching FERUs (and emergency vehicles in general) has been described. A simulation model that generates the operations of the FERU has been developed on the basis of a selected dispatching policy. The simulation model estimates the total incident duration as a measure of effectiveness of the dispatching policy and provides decision makers with a useful tool for selecting the optimum policy. Work in progress involves the development of a cost–benefit model that will be able to quantify the total cost of the freeway emergency response system for alternative levels of emergency response resources and to estimate the corresponding benefits in user costs, that is, fuel consumption, travel, and social impacts.

REFERENCES

1. Zografos, K. G., T. Nathanail, and P. Michalopoulos. Analytical Framework for Minimizing Freeway-Incident Response Time. *Journal of Transportation Engineering,* ASCE, 1993.
2. Bertsimas, V. R. Stochastic and Dynamic Vehicle Routing in the Euklidean Plane: The Multiple-Server Capacitated Vehicle Case. Working paper. Operation Research Center. Massachussetts Institute of Technology, Cambridge, 1990.
3. Jarvis, S. A Simple Procedure for Ambulance Allocation in Areas of Low Demand. In *Urban Public Safety Systems*, Vol. 5. Massachussetts Institute of Technology, Cambridge 1977.
4. Toregas, C., R. Swain, C. ReVelle, L. Bergman. The Location of Emergency Service Facilities. *Operations Research*, Vol. 19, 1970.
5. Marlin, P. G. Application of the Transportation Model to a Large-Scale "Districting" Problem. *Operations Research*, Vol. 8, 1987.
6. Nathanail, T. *Freeway Incident Delay Minimization Through the Optimum Deployment of Traffic Restoration Units.* M.S. thesis. University of Miami, Miami, Fla., 1991.
7. Lindley, J. A. *Transportation Research Circular 344: The Urban Freeway Congestion Problem.* TRB, National Research Council Washington, D.C., 1989.

Publication of this paper sponsored by Committee on Transportation Safety Management.