RT-MAE-8101

A SIMULATION PROGRAM FOR EMERGENCY SERVICES - II

by

Carlos Alberto Barbosa Dantas

and

Eliete Colucci

Palavras Chave: Simulation, Queuing
(key words)

Classificação AMS: 68A55
(AMS Classification)

A SIMULATION PROGRAM FOR EMERGENCY SERVICES - II

Carlos A.B. Dantas and Eliete Colucci
Instituto de Matemática e Estatística - USP

Abstract - An extension of a simulation program for allocation of emergency units in an urban area or highway is given. Several measures of performance are proposed. A validation study is considered.

1. Introduction

The literature on problems of allocation of emergency units, such as ambulances, fire engines, repair vehicles, towing cars, etc. has grown considerably in the last fifteen years. Most of the research has concentrated on resource allocation problems in urban areas. As Larson [4] has formulated, the allocation problem of urban emergency services is composed of two related problems: the "districting" problem and the "location" problem. The districting problem can be formulated as follows: given a region with a certain spatial distribution of demands for service and given N response units spatially distributed in the region, how should it be partitioned into areas of primary responsability (districts) so as to best achieve some level or combination of levels of service?

The location problem can be formulated as follows: How should the N response units be located or positioned while not responding for calls for service?

In ambulance emergency services the district for a particular unit is the area in which calls for service are answered by that particular ambulance, provided it is free. Usually every hospital is located in a district but we may have several districts with no hospital.

Most of the analytic models developed suffer from some deficiencies: a) they have focused solely on intradistrict response of the units b) they have considered only one performance measure c) they have failed to incorporate the probabilistic nature of urban emergency services.

Larson [4] has proposed the hipercubic queuing model which overcomes the above limitations.

The simulation program we developed is used in the same manner as the hipercubic model, since we adopted some of the performance meansures of the hipercubic models and proposed some additional ones.

2. The Hipercubic Model

The hipercubic model is a queuing model where the servers are distributed spacially in a region and answer calls generated in that region. The region is divided in a certain number of minimal subregions which are called "atoms". For modeling purposes the atoms are the smallest subregions where the calls for service are registered. The union of a certain number of atoms will be called a district. In each district there is a unit of the emergency service. This unit while not answering a call may be stationed in a fixed location inside the district or may remain in movement inside it. In the last case, the location may be specified statistically by giving the relative amount of time it spends in the various atoms of the region.

The hipercubic model was developed based on the following assumptions: 1) The arrival process, ie, the calls for service in the region follow a Poisson Process with rate λ calls per unit of time. 2) The service times of all the units of the system have an exponential distribution with mean $1/\mu$ independent of the location of the call and of the unit sent. 3) Only one unit is sent to answer a call.We note that for a system with N units, if we do not take into account the identity of the units free or busy under hipothesis 1, 2 and 3 we have a queuing model of the type M/M/N.

A policy is defined when it is given a partition of the region in districts, the location of the units in them and a rule which assigns to each atom in decreasing order of priority the units which attend calls generated there. When all the units are busy there are two possibilities, either the call is rejected by the system or it enters a queue and is attended by the first unit to become free.

If the units are numbered from 1 to N and furthermore to each unit is assigned the number zero or 1 according as the unit is free or busy,the states of the systems in the case of 0-capacity are identified with sequences of 0's and 1's of lenght N, ie, the N-dimensional hipercube.

With this state space, under the hypothesis mentioned the process is a continous parameter Markov chain. Since the service times have a common exponential distribution with parameter μ it follows that the transition rate from a state with k to a state with k-1 units busy is μ, $1 \leq k \leq N$. The transition rates from a state with k to one with k+1 busy units are difficult to determine, because they depend on the state and on the priority rule with which the units are sent to answer the call. Larson in [4] constructed an algoritm to determine this rates.

Based on the transition rates one obtains the equations of detailed balancing whose solution gives the stationary distribution for the system with 2^N states. With the stationary distribution one obtains the workloads ρ_i which represent the amount of time unit i is busy servicing calls. Using the workloads one obtain the performance measures for the system as a whole and also for the units and the districts.

To obtain the stationary distribution a system of 2^N equations must be solved. Larson introduced an aproximate procedure which reduces the systems of equations to N. As it can be seen in [5] the approximation is quite good.

3. The Simulation Model

In the hipercubic model the data such as arrival rate, mean service time, probability of a call ocurring in an atom and speed of the vehicles must be estimated. Estimates are obtained in general for a configuration of the system with a given number of units. If we want to operate with a different number of units the mean total service time must be obtained.

Larson in his model cannot consider policies with a queue for certain units while others are unoccuppied. For large distances this possibility might be convinient.

Besides the fact that usually total service times do not have an exponential distribution the facts mentioned above were the main reasons for the development of the simulation model in [1].

Recently we have extended the model by considering new meausures of performance and compared the results of the simulation with those of the aproximated hipercubic model [3].

Hypothesis: As in the hypercubic model the region in which the service operates is divided in atoms whose unions form the districts.

To construct the model the following hypothesis were assumed:

a) The arrival process in the region is a Poisson process with rate λ, namely the interarrival times are exponentially distributed with mean 1/λ. It follows that if f_k represents the probability that a call occurs in atom k, k=1,2,...,nº of atoms then the number of calls in atom k is a Poisson Process with rate λf_k.

b) Each unit can perform various types of service the type of service to be performed is selected from a discrete distribution.

c) The elapsed time of service at the scene (place of ocurrence) has an exponential distribution.

d) The velocity of the units when moving to the scene or while performing a task are constant.

e) The units while not attending a call are stationed in a district. There may be more than one unit in a district.

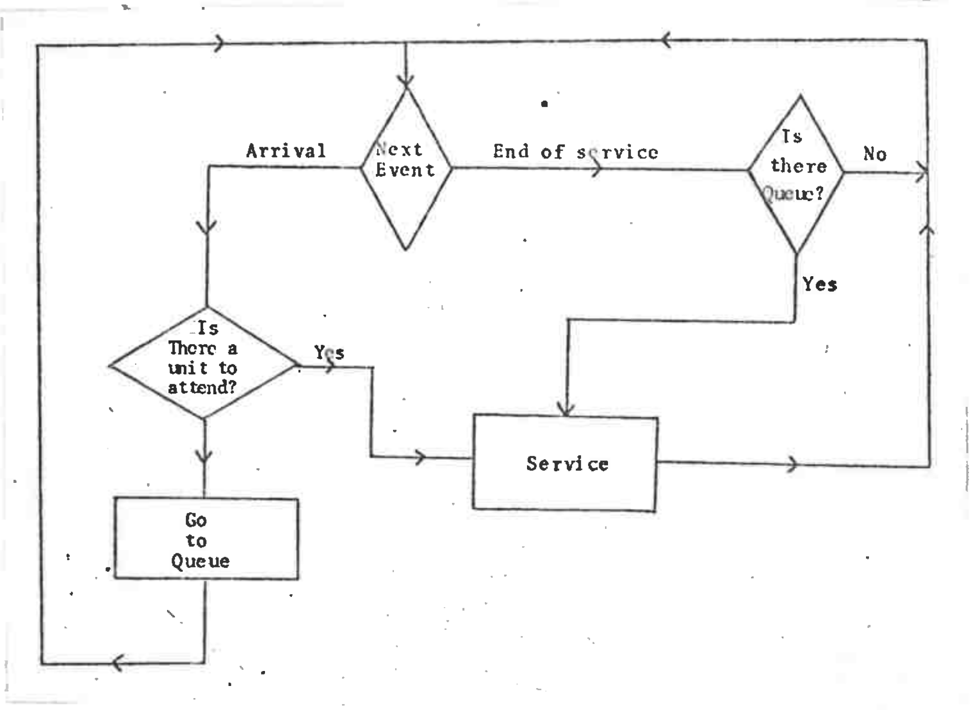
Remark

If the service times at the scene are not exponential it will be sufficient to replace the routine that generate the exponential variables by one that generates the other distribution.

The Program

The control of the program is done through the "event scheduling approach". An event will be an arrival or the end of a service.

A simplified version of the flowchart of the program is the following:



Next Event: In this block the arrival times are generated and it is decided which is the next event an arrival or the end of a service. The one that occurs first is the next scheduled event. Then simulated time is advanced to this scheduled time and all the variables (descriptors) associated with the simulation are adjusted.

Is there a unit to attend? This block determines the atom which originates the call and searches if there is a unit to attend it.

Go to Queue: If there is no unit to attend a call this block sends the unit to a queue. The queues are formed for the units in the district where the call was generated.

Is there a queue? The unit that finishes a service searches if there are queues in the districts which it attends in decreasing order of priority.

Service: This block generates the type of service, the elapsed time of service at the scene; for services executed in two stages it generates the location where the second stage is to be performed. It computes all the travel times using the velocity and the matrix of the distances among the atoms. The total service time is computed adding the travel times with the service time at the scene.

The policies: To specify a policy for the simulation it is necessary to add to the definition given for the hypercubic model (see 2) a table. This table assigns to every unit the districts which it attends in decreasing order of priority at the moment it finishes a service.

4. Example

Before studying the validation of the simulation model, we will give an example of its outputs. We will consider a towing system in a highway. The highway was divided in 21 atoms and five districts were formed with one towing car in each one. Calls may require towing or not. In the first case the vehicles are towed to given locations along the highway. This information is summarized by a vector $x=(x_i, i=1,...,21)$ where x_i denotes the destination of a vehicle towed from atom i.

The total service time (TST) is decomposed in either of the following ways:

$$(1) \quad TST = TT + STC_1 + RT$$

$$(2) \quad TST = TT + STC_2 + T_0T + RT_0T$$

whether the call does not or does require towing.

Here: TT = travel time to the atom of the call

STC_1= service time at the place of the call without towing

STC_2= service time at the place of the call with towing

T_0T = time to tow

RT = return time to location of the unit

RT_0T= return time to location of the unit from towing place.

The policy adopted is specified by tables 1, 2 and 3.

TABLE 1

District	Location (atom)	Atoms					
1	3	1	2	3	4	5	
2	8	6	7	8	13		
3	9	9	10	11	12	20	21
4	15	14	15	16	17	18	
5	19	19					

TABLE 2

District	Units		
1	1	4	2
2	2	3	1
3	3	5	
4	4	1	2
5	5	3	

TABLE 3

Unit	Districts			
1	1	4	2	
2	2	3	1	4
3	3	5	2	
4	4	1		
5	5	3		

Tables 1 gives the location of the units in the districts and the atom that form them. Table 2 is associated with a call: it gives the units which attend every district in decreasing order of priority. Table 3 is associated with an end of service. It lists in decreasing order of priority the district a unit shall attend whenever there are calls waiting.

Four hundred periods of 18 hours of operation were simulated with the following data: call rate: $\lambda=3.24$ calls per hour; speed of the units = 60Km/h; probability a call requires towing = 0,33; STC_1 and STC_2 exponentially distributed with means 30' and 15' respectively.

The tables that gives the probability a call occurs in each atom and the one giving the places the vehicles are towed to are omited.

Output

Regionwide fraction of dispaches which are interdistrict = 0,37

Regionwide workload imbalance = 0,105

Performance measures associated with units

Unit	W	Disp out of district	TST	TT	STC	T_0T	RT	L_q
1	.5041	.51	48.72	14.17	17.57	12.00	14.89	.0378
2	.4528	.49	46.15	13.26	17.77	12.11	12.97	.0273
3	.5578	.37	43.53	11.02	17.65	8.46	13.35	.1088
4	.5347	.18	49.23	11.64	17.12	16.19	17.64	.0770
5	.5388	.33	52.29	9.02	17.63	7.97	24.30	.1059
Mean	-	-	47.81	11.77	17.55	11.26	16.52	-

Performance measures associated with districts

District	Fr. of Inter Disp.	............MEAN............			
		TTD	W'	WTQ	WTQQ
1	.4134	12.22	22.03	4.36	19.61
2	.3530	9.54	18.88	3.56	19.07
3	.3487	10.06	37.39	8.72	22.99
4	.4260	15.07	20.03	4.98	23.31
5	.3019	10.11	37.39	10.72	28.89

The regionwide workload imbalance is the maximum difference among the workloads of the units; W = workload is the fraction of time that a unit is busy; Fr. Disp out of District is the fraction of dispaches performed by an unit out of its district;L_q is the expected lenght of the queue for each unit; Fr of inter disp - is the fraction of dispaches in a particular district attended by units of other districts; TTD is the travel time to a district; W' is the fraction of time all units which attend a dis trict are busy, ie, the fraction of time an arriving call will enter a queue; WTQ = waiting time in queue; WTQQ = waiting time in queue given there is a queue.

5. Model Validation

From the hypothesis in section 3 we can easily conclude that for a system with one unit we have an M/G/1 queue if service is the total service time (TST). In fact the arrival process is Poisson.

The total service times are i.i.d. because given the instant a call occurs its location is obtained selecting an atom from a discrete distribution independent of the instant of arrival, the service times at the scene are i.i.d. exponentially distributed random variables independent of the location of the call, and the distances travelled depend on the atom selected.

For an M/G/1 queue with arrival rate λ, average service time $1/\mu$ the expected queue lenght L_q and the average waiting time in queue W_q are given by:

$$L_q = \frac{\rho^2 + \lambda^2\sigma_S^2}{2(1-\rho)} \qquad W_q = \frac{L_q}{\lambda}$$

where $\rho = \frac{\lambda}{\mu}$ and σ_S^2 is the variance of the total service time.

The expected value and variance of TST are given by:

$$(3)\quad E(TST) = E(TT + STC_1 + RT|X=1)\,P(X=1) + \\ + E(TT + STC_2 + T_0T + RT_0T|X=2)\,P(X=2)$$

$$(4)\quad Var(TST) = E[Var(TST|X)] + Var[E(TST|X)]$$

where X=2 if the call require towing and X=1 if it does not.

We ran the simulation of four hundred periods of operation of the system with one unit centrally located for three different arrival rates: λ= .25, .50 and .75 calls per hour and with the same remaining data as in the example considered.

The table below gives the results obtained for L_q and W_q in the simulation and a comparison with the values obtained for them in the M/G/1 model. The times are given in minutes.

	λ = .25		λ = .5		λ = .75	
	SIM	M/G/1	SIM	M/G/1	SIM	M/G/1
TST	57.99	-	56.43	-	56.52	-
ρ	.2462	.2416	.4759	.4619	.7011	.7064
L_q	.0479	.0463	.237	.242	.894	.85
W_q	11,3'	11,1'	28,15'	29'	72'	68'

In our simulation program we are developing a routine to obtain confidence intervals for the above parameters. The results will appear in a forthcoming paper.

REFERENCES

1 Dantas, C.A.B. and Ferrari, P.A. - "Simulação de Serviços de Emergência" - Atas do 3º SINAPE - 1978 - p. 24-29.

[2] Dantas, C.A.B. - "Uma Aplicação de um Serviço de Emergência" Atas do 3º SINAPE - 1978 - p. 142-148.

[3] Dantas, C.A.B. and Colucci, E. - "Serviços de Emergência:Comparação dos Resultados de Simulação e do Modelo Hipercúbico" - 4º SINAPE.

[4] Larson, R.C. - "A Hipercubic Queuing Model for facility location and Redistricting in urban emergency services"- Computers and Operations Research - 1 - 67-95 - 1974.

[5] Larson, R.C. - "Approximating the Perfomance of Urban Emergency Service Systems", Operations Research, 23 - nº 5 - 845-868.

RELATÓRIOS TÉCNICOS
DO
DEPARTAMENTO DE ESTATISTICA
TITULOS PUBLICADOS

7901 - BORGES, W. de S. On the limiting distributions of the failure time of composite material. São Paulo, IME-USP, 1979. 22p.

7902 - GALVES, A.; LEITE, J.G.; ROUSSIGNOL, M. The invariance principle for the one-dimensional symmetric simple exclusion process. São Paulo, IME-USP, 1979. 9p.

8001 - MENTZ, R.P. et al. Exploratory fitting of autoregressive and moving models to Well-behaved time series data. São Paulo, IME-USP, 1980. 16p.

8002 - MORETTIN, P.A. Walsh spectral analysis. São Paulo, IME-USP, 1980. 27p.

8003 - RODRIGUES, J. Robust estimation and finite population. São Paulo, IME-USP, 1980. 13p.

8004 - BORGES, W. de S. & RODRIGUES, F.W. On the axiomatic theory of multistate coherent structures. São Paulo, IME-USP, 1980. 10p.

8005 - MORETTIN, P.A. A central limit theorem for stationary processes São Paulo, IME-USP, 1980. 5p.