Extended Hypercube Models for Location Problems with Stochastic Demand

Burak Boyacı, Nikolas Geroliminis
Urban Transport Systems Laboratory (LUTS)
Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

March 29, 2013

Abstract

In spatial queues, servers travel to the customers and provide service on the scene. This property makes them applicable to emergency response (e.g. ambulances, police) and on-demand transportation systems (e.g. paratransit, taxis) location problems. However, in spatial queues, there exist a different service rate for each customer-server pairs which creates Markovian models with enormous number of states and makes these approaches difficult to apply on even medium sized problems. Because of demand uncertainty, the nearest servers to a customer might not be available to intervene and this can significantly increase the service times. In this paper, we propose two new aggregate models and an approximate solution method with a dynamic programming heuristic. Results are compared with existing location models on hypothetical and real cases.

1 Introduction

Emergency response systems and *on-demand transportation systems* are important for modern societies. The former protect public health, provide assistance and ensure safety whereas the latter provide services in areas with low passenger demand, where regular services are economically infeasible. Although both systems have low demand rates, the service availability is an important aspect for them. The service should be available when it is needed in an acceptable time. They should provide adequate coverage with rapid and reliable response times (Wisborg et al., 1994) while staying inside limited budgets. According to Emergency Medical Services (EMS) Systems Act, 95% of requests should be served within 30 minutes for rural areas and 10 minutes for urban areas (Ball and Lin, 1993).

Usually both emergency response and on-demand transportation systems have low demand to service ratio. The total demand is lower than the sum of available service capacity of the servers. However, these systems still experience congestions and even unavailabilities because of the stochastic structure of the demand and their performance decreases. In 2000, London Ambulance Service (LAS), which is one of the largest emergency response fleet with over 400 ambulances in Europe, was sued because of its high response times (Harpwood, 2000). In New York City, in 1999, 12779 taxis served about 240 million passengers (Schaller, 2006). Both emergency and on-demand transportation systems are highly variant and congested systems and need to be managed efficiently. For instance, deciding the borders of service regions and number of paratransit vehicles needed in each region for a required service rate is an interesting and important question for these systems.

2 Literature Survey

The early models in the literature dealing with the location of emergency response systems assume deterministic demand. *Median problems* locate the facilities on candidate locations or the Euclidean space to minimize average distance. The p-median problem (Hakimi, 1964), the Fermat-Weber problem (Weber, 1909) with its multifacility version (Cooper, 1963) are the initial location allocation based models in the literature used to deal with emergency response systems.

Coverage models are used to locate facilities on candidate locations to maximize total coverage and/or minimize number of facilities. Toregas et al. (1971) and Church and ReVelle (1974) take different aspects of this problem into consideration. Daskin (1983) and Batta et al. (1989) expands the two former deterministic models and add an explicit "availability" probability for the servers to model stochastic nature of the problem.

The first model that embeds a *queuing theory* to calculate implicit availability for the servers is the *hypercube queuing model* (HQM) (Larson, 1974). This model is suitable to analyze systems with a spatial queuing characteristic such as emergency, on-demand transportation, neighborhood services, door-to-door pickup and delivery systems

(Larson and Odoni, 1981). The HQM is a descriptive model and provides system performance measures for given demand and server locations Galvão and Morabito (2008).

The HQM (Larson, 1974) considers that time spend on the way to incident is negligible compared to service time on scene and assumes total service time is independent of the distance between the incident and server which is a fair assumption for fire brigades but not for ambulances or taxis. This simple model creates 2^n states for n servers. In order to cope with larger cases, Larson (1975) proposes a heuristic method. Atkinson et al. (2008) assume two different service rates (i.e. intra and interdistrict) for each server according to demand they served and propose a model with 3^n states. Recently Iannoni and Morabito (2007) and Iannoni et al. (2008) embed hypercube in a genetic algorithm framework with multiple dispatch (i.e. more than one server may be needed for an incident) option. Geroliminis et al. (2009) and Geroliminis et al. (2011) integrate location and distracting in the same framework and solve problem with steepest descent and genetic algorithms.

3 Hypercube Queuing Models

The HQM models include *hypercube* in their names because of the shape of the transition graph of the Markov chains used to represent these models. In Larson (1974)'s hypercube model each server is either available (0) or busy (1) and each state is a number in base 2. In Atkinson et al. (2008)'s model, number of states increases to three: available(0), busy with intradistrict (1) and busy with interdistrict (2). In both cases number of states is an exponential function of the number of servers (n), 2^n and 3^n respectively.

In order to deal with exponential number of states, we propose to model the problem in an aggregate way. In the two proposed *extended* hypercube queuing models (EHQM), a new structure called *bin* is introduced. Instead of taking each servers' state separately, the servers that are close to each other form structures called bins and state structure of the system is kept in bin level (e.g. the state variables contain number of available servers in each bin instead of each available server). With this modification, number of states becomes an exponential function of the number of bins and not the number of servers. The number of states for new models are $\prod_i (n_i + 1)$ and $\prod_i \binom{n_i + 2}{2}$ for 2^n and 3^n EHQM respectively where n_i is the number of servers assigned to bin i. The details of these extended models can be seen in Boyacı and Geroliminis (2011).

Although the new two EHQMs have less states than simpler HQM models, the number of states is still too high to be applicable to real life problems. Specifically 3^n EHQM is more representative since it assumes different service rates for different bin-demand pairs however has more states. Therefore, we propose to use these models in some approximation algorithms to deal with real life instances. However, before going into details of this, we want to show the performance of our model and compare them with the maximal expected covering location problem (MEXCLP) Daskin (1983) formulation. MEXCLP is selected in this comparison since it (i) is simple to implement, (ii) takes probabilistic nature of the problem into consideration and (iii) has no parameters to fine-tune.

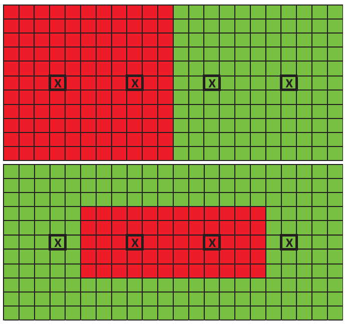
4 Performance of Extended Hypercube Queuing Model

In order to measure performance of the proposed 3^n EHQM, we have generated scenarios with different parameters (i.e. demand distribution, demand ratio, coverage distance, on scene service time). In each scenario, the model distributes 12 identical servers to four candidate locations to minimize total loss rate. Each scenario is solved both with MEXCLP and 3^n EHQM. In order to convert EHQM to an optimization model, an iterative procedure is implemented that solves a 3^n EHQM for each possible configuration of servers and reports the solution with the minimum lost rate.

In order to have an unbiased comparison, the two approaches are compared with a discrete event simulation which simulates a spatial queuing system of type $M/D/12/12/\infty$ with slight differences. The interarrival time of incidents are exponentially distributed whereas the service time is fixed and has a different value for each server-region pair, i.e., a system with stochastic demand rate but fixed service time. A single service time equals to sum of travel and on-scene service times. The former is a function of Manhattan distance between the region served and the server dispatched, and latter is a fixed predefined value. The queue size is zero and a demand that cannot be served by an available server is lost. The parameters and the two demand distribution used in experiments can be seen in Figure 1. Note that, each unit box in the figure is a subregion and creates a demand with some predefined rate. The total demand in the system has three different values (20,40,60) and distributed between these subregions accordingly. The area shown with the color red has higher demand than the area shown with color green and two different demand ratios (0.5 and 0.75) are used in the experiments. The subregions with a cross are candidate locations that servers can placed. The comparison of lost rate ratio to total demand and average service rate per server of the two approaches for these 24 scenarios can be seen in Figure 2.

demand distribution of type 1

demand distribution of type 2



demand distribution:	type 1 and 2
total demand rate (# of incidents / hour):	20, 40 and 60
demand ratio between regions:	0.5 and 0.75
on scene service time (minute):	5 and 10
number of servers:	12
server speed (unit distance/minute):	1
server coverage distance (unit distance):	5

Figure 1: Parameters used to compare the performance of 3^n EHQM and MEXCLP

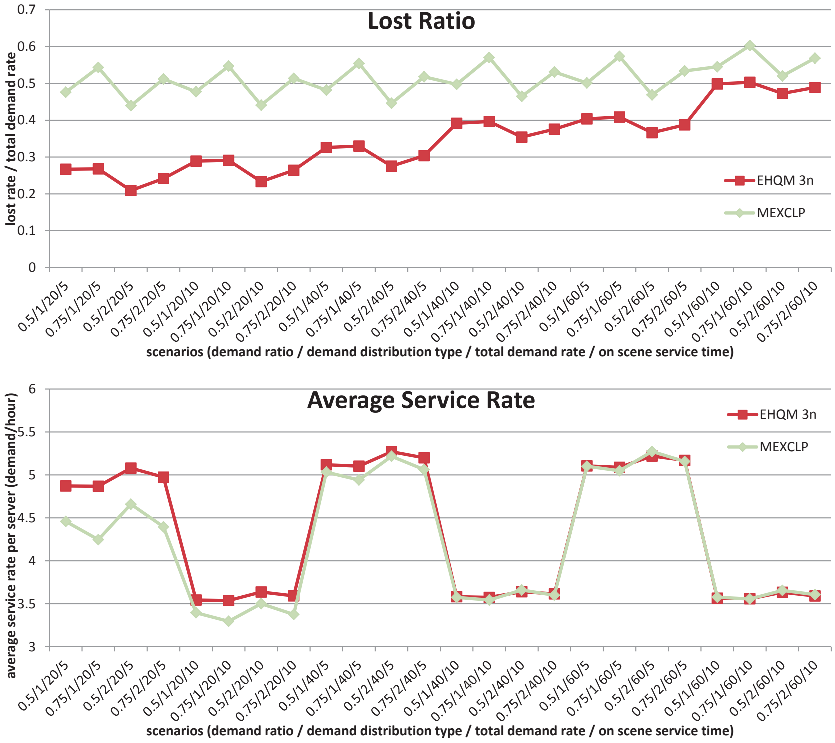


Figure 2: The comparison of lost rates and average service rates (per server) of 3^n EHQM and MEXCLP

As it can be seen in the graphs given in Figure 2, 3^n EHQM outperforms MEXCLP in both performance measures. The difference is more significant in the lost ratio graph since it is taken as the parameter to be optimized in the iterative procedure of 3^n EHQM. We can also observe that, the difference between methods are more in low demand rates.

5 Conclusions and On-Going Work

In this paper, two new hypercube queuing models that extend the existing two models (Larson, 1974; Atkinson et al., 2008) are proposed. We also compare the performance of one of the two models, namely 3^n EHQM, with a method from the literature, MEXCLP Daskin (1983). The results show 3^n EHQM significantly outperforms MEXCLP.

On our on-going work, we have already modeled an approximation algorithm that firstly partitions the whole problem area into manageable subareas. Then this approach merges each subarea iteratively by taking servers that can serve other subareas into consideration. We have also implemented a dynamic programming heuristic that locates predefined number of servers on candidate locations. We have started to test the model and heuristic, and the preliminary results are promising. The partitioning merging and dynamic programming will be presented in the full paper.

References

J.B. Atkinson, I.N. Kovalenko, N. Kuznetsov, and K.V. Mykhalevych. A hypercube queueing loss model with customer-dependent service rates. *European Journal of Operational Research*, 191(1):223 – 239, 2008.

Michael O Ball and Feng L Lin. A reliability model applied to emergency service vehicle location. *Operations Research*, 41(1):18–36, 1993.

Rajan Batta, June M Dolan, and Nirup N Krishnamurthy. The maximal expected covering location problem: Revisited. *Transportation Science*, 23(4):277–287, 1989.

B. Boyacı and N. Geroliminis. Extended hypercube models for large scale spatial queueing systems. In *11th Swiss Transport Research Conference*, 2011.

R. Church and C. ReVelle. The maximal covering location problem. *Papers in Regional Science*, 32:101–118, 1974.

Leon Cooper. Location-allocation problems. *Operations Research*, 11(3):331–343, 1963.

M.S. Daskin. A maximum expected covering location model: Formulation, properties and heuristic solution. *Transportation Science*, 17(1):48–70, 1983.

R.D. Galvão and R. Morabito. Emergency service systems: The use of the hypercube queueing model in the solution of probabilistic location problems. *International Transactions in Operational Research*, 15(5):525–549, 2008.

N. Geroliminis, M.G. Karlaftis, and A. Skabardonis. A spatial queuing model for the emergency vehicle districting and location problem. *Transportation Research Part B: Methodological*, 43(7):798 – 811, 2009.

N. Geroliminis, K. Kepaptsoglou, and M.G. Karlaftis. A hybrid hypercube - genetic algorithm approach for deploying many emergency response mobile units in an urban network. *European Journal of Operational Research*, 210(2): 287–300, 2011.

S.L. Hakimi. Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research*, 12(3):450–459, 1964.

Vivienne Harpwood. *Principles of Tort Law 3/E*. Routledge, 2000.

A. Iannoni, R. Morabito, and C. Saydam. A hypercube queueing model embedded into a genetic algorithm for ambulance deployment on highways. *Annals of Operations Research*, 157:207–224, 2008. 10.1007/s10479-007-0195-z.

A. Paula Iannoni and R. Morabito. A multiple dispatch and partial backup hypercube queuing model to analyze emergency medical systems on highways. *Transportation Research Part E: Logistics and Transportation Review*, 43 (6):755 – 771, 2007. Challenges of Emergency Logistics Management.

R.C. Larson. A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers & Operations Research*, 1(1):67 – 95, 1974.

R.C. Larson. Approximating the performance of urban emergency service systems. *Operations Research*, 23(5):845–868, September-October 1975.

R.C. Larson and A.R. Odoni. *Urban Operations Research*. Prentice-Hall, Englewood Cliffs, N.J., 1981.

Bruce Schaller. The new york city taxicab fact book. *Schaller Consulting, mars*, 2006.

C.R. Toregas, R. Swain, C.S. ReVelle, and L. Bergman. The location of emergency service facilities. *Operations Research*, 19(6):1363–1373, 1971.

Alfred Weber. *Über den standort der industrien*. JCB Mohr, 1909.

T. Wisborg, A.B. Guttormsen, M.B. Sørensen, and H.K. Flaatten. The potential of an anaesthesiologist–manned ambulance service in a rural/urban district. *Acta anaesthesiologica scandinavica*, 38(7):657–661, 1994.