Location Models for Airline Hubs Behaving as M/D/c Queues

by

Vladimir Marianov
Department of Electrical Engineering
The Catholic University of Chile, Santiago, CHILE
e-mail: *marianov@ing.puc.cl*

and

Daniel Serra
Corresponding author
Department of Economics, IET and GRES
Universitat Pompeu Fabra, Barcelona, SPAIN
email: *daniel.serra@econ.upf.es*

Abstract

Models are presented for the optimal location of hubs in airline networks, that take into consideration the congestion effects. Hubs, which are the most congested airports, are modeled as M/D/c queuing systems, that is, Poisson arrivals, deterministic service time, and c servers. A formula is derived for the probability of a number of customers in the system, which is later used to propose a probabilistic constraint. This constraint limits the probability of b airplanes in queue, to be lesser than a value α. Due to the computational complexity of the formulation, The model is solved using a meta-heuristic based on tabu search. Computational experience is presented.

Keywords: Hub location, Congestion, Tabu-search

Introduction

Networks involving hubs are important in transportation and telecommunications. In both cases, when there is traffic between several origins and several destinations, there are economical benefits if this traffic is concentrated on some arcs of the network. A hub is a point where traffic from several origins is concentrated and sent either to another hub, taking advantage of the economies of scale, or to its final destination. Since the eighties, models and heuristic algorithms have been presented for the location of a fixed number of hubs, or a number of hubs to be determined by the solution to the model. The goal in all cases has been to find the least expensive hub network (considering transportation and fixed costs), given the traffic volumes between each origin-destination node pair.

Most of the hub models and algorithms discussed in the literature have been oriented to the location of hubs in airline networks for transportation of both cargo and passengers (selection of certain airports as hubs). Although there is a profuse literature on the design of telecommunications networks, the models are generally not oriented to location of hubs alone, because in these networks there are some other difficulties, as for example, those related to arc capacities. Telecommunications models include typically the effects of transmission delay (on arcs) and congestion at nodes. The goal of these models is to keep average transmission delay and average waiting times at nodes, as low as possible. On the other hand, airline hub location models have never considered congestion; there is a parallel body of literature concerning congestion at airports.

As it is (particularly in the last few years) evident for all airline users, hub airports are more congested because they receive higher traffics than other airports in the

network. We formulate hub location models that consider congestion at hub airports. The models we propose are oriented to the design of hub networks, or to the analysis of existing hub networks with planning purposes, as for example, to determine in which existing hub it is worth adding an additional runway.

Congestion in airports is hard to deal with, because of several complicating issues. In the first place, arrival rate of the planes is highly variable throughout the day. Although flights follow a schedule, they are subject to delays at their origin airports and during the flight itself, which make their arrival non-deterministic. Secondly, although it can be considered constant during short periods of time, the service rate is also variable, mainly depending on weather conditions and the type of planes that are serviced. In general, service times are not independent, identically distributed, because there are passengers that transfer between flights. Thus, service time for a flight may depend on the arrival time of other flights. In the third place, at their arrival to the airport, airplanes have to go through three stages of service: landing at a landing runway, service at a gate and departure through a takeoff runway. Service at the gate includes waiting for passengers transferring from other flights. The probabilistic distributions of these service times are difficult to determine, although they can be approximated. Due to these issues, detailed models for congestion at airports are very complicated and it is extremely difficult to use them for planning purposes. Thus, some of the detail must be avoided through approximations, if models for planning purposes are to be developed.

We focus on the queue formed by airplanes waiting for landing. Under certain assumptions, the analysis is applicable to takeoff runways or combined landing-departure runways. We use a peak hour analysis, assuming that during the peak

hour the average arrival rate and the service rate are both constant. This allows us to model an airport as a M/D/c queuing system, i.e. Poisson arrivals, deterministic service time and several (c) servers. This election of probabilistic modeling is justified in the body of the paper. We state an analytic formula for the steady state probabilities of different numbers of customers in an M/D/c system, found using an approach similar to the one used by Gross and Harris (1974) for single server queues and Prabhu (1965) for waiting time distribution. Later, we use this formula for the development of a deterministic equivalent of a probabilistic constraint in an integer optimization formulation for the location of congested hubs.

In the next section, we review briefly the related literature. Next, we develop the probabilistic analysis and the probabilistic constraint. We present the full model and show some computational experience obtained through the use of exact and heuristic algorithms.

Literature Review

Hubs

The optimal location of hubs in a network is more difficult than other location problems because, as opposed to what happens in these, hubs interact with each other. This interaction results in non-linear models (O'Kelly, 1986a; O'Kelly, 1987 or Aykin, 1988), which can be linearized through the replacement of non-linear products in the objective by new variables (O'Kelly and Lao, 1991). This linearization squares the number of integer variables of the formulation. In the telecommunications case, simpler versions of the problem have been formulated, as the star-star concentrator location problem (Gavish, 1991; Pirkul, Nagarajan,

1992; Pirkul, Narasimhan, De, 1988; Marianov *et al*, 1995) in which the hubs are not connected to each other, but instead, they are connected to a central point. Thus, there is no interaction between them, and the problem becomes simpler.

Hub location models can be classified in two. Planar hub location problems, in which demand is concentrated at discrete points and hubs can be located anywhere on the plane, and network hub location problems, in which both demand and hubs are located at discrete points on a plane. Planar problems have been studied by O'Kelly, 1986a; Aykin, 1988; Aykin, 1995. In the case of airline networks, these models are not realistic, because they would result in airports being built anywhere. Consequently, we will refer to the second type, that is, network problems. These can be, in turn, classified in several types. The *p*-hub-median problem on a network (O'Kelly, 1986b; Klincewicz, 1991) locates a pre-specified number *p* of hubs on nodes of a network, and allocates the demand points to them. The uncapacitated hub location problem seeks the location of an unknown number of hubs at nodes of a network and the allocation of demand points to them, so that the sum of investment and transportation costs is minimized (O'Kelly, 1992). For reviews of these, see Campbell, 1994, who also proposes three new classes of models, analogous to the usual location problems: the *p*-hub center problem, the hub set covering problem and the hub maximal covering model. Later, Marianov, Serra and ReVelle (1998) formulated a maximum-capture hub location model, for competitive environments. Most of the models seek cost minimization. An exception are Campbell's, 1994, maximal covering models, which seek the location of *p* hubs, so as to maximize the coverage (by pairs of hubs) of weighted demand (represented by origin-destination pairs). No further definition of

coverage is offered. Most of these models can be written in their single-allocation or multiple-allocation versions (a demand node is allocated to one hub, or to several hubs, depending on the destination of the traffic).

Different solution methods have been proposed for the hub location models. Aykin (1994) proposes lagrangian relaxation. Klincewicz (1989 and 1991) shows the results of applying Greedy Random Adaptive Search Prcedure (GRASP) and Tabu Search to the problem. Skorin-Kapov and Skorin-Kapov (1994) use tabu search and compare the results with other algorithms.

None of the hub models consider congestion.

Congestion and airports

When dealing with location, most of the spatial queuing networks have been presented in the context of emergency systems. Many of them assume a single server in the region under study, because of the complexity of the multiple-server, non-Markovian queues. Berman, Larson and Chiu (1985), formulate the Stochastic Queue Median, a heuristic algorithm to locate optimally one server on a congested network, behaving as an M/G/1 queue. A model by Batta (1988) considers a situation with a selective rejection of calls. Batta, Larson and Odoni (1988) present a model and an algorithm for locating one server with calls of different priorities. Batta (1989) presents a model to study the effect of using expected service time dependent queuing disciplines on optimal location of a single server. Based on the one server Stochastic Queue Median, Berman, Larson and Parkan (1987) develop two heuristics for locating p servers on a congested network. Berman and Larson (1985) solve the problem of districting for a two-server network in the presence of

queuing, seeking minimum average response time. Later, Berman and Mandowsky (1986), use the Stochastic Queue Median, combined with this 2-server districting algorithm, to develop a general location - districting iterative algorithm for two units, and for *n*-nodes, *m*-server networks. All of these models are nonlinear.

In general, a fairly large computational effort is required if these models are to be used for location of servers. All of the models are nonlinear, heuristic, and their objective is to minimize expected response time of servers that travel to the site of an emergency. Also, all models use approximations in order to model the system.

Some authors have used embedded queuing systems in optimization models for location. Marianov and ReVelle (1994 and 1996) formulated models for emergency services in networks behaving as M/M/k queues, and Marianov and Serra (1997 and 1998) considered congested fixed servers.

A few researchers have studied congestion at airports, mainly for obtaining a model of its behavior. Peterson, Bertsimas and Odoni (1995) propose a model of the set of landing strips or runways (considered as a single server), with a deterministic, time-varying flight arrival rate and deterministic, time-varying service rate. This model is not intended for locational purposes but for schedule policy making. Newell (1979) presents a complete analysis of the airport operations. Both papers review existing results on this subject.

M/D/c queuing systems

The single server M/D/1 queue (Poisson arrivals, deterministic service times, 1 server) has been completely characterized. However, few results are available in the literature on the computation of the steady-state probabilities of different

numbers of customers in a multiple server M/D/c queue (Poisson arrivals, deterministic service times, c servers). Syski (1986), Gross and Harris (1974), Prabhu (1965) and Saaty (1961) present a method for finding the generating function of these probabilities, based on a seminal development by Crommelin (1932). Altinkemer *et al*, (1998) propose an approximation to the average waiting time in an M/D/c queue with non-preemptive priorities. Prabhu (1965) presents a formula for the limiting waiting time cumulative distribution. Knessl *et al* (1990) review the results for multiple server queues and present an integral approach for the M/G/2 queue, being a particular case the M/D/2 queue. Chaudhry *et al* (1992), provide an algorithm for numerically finding the limiting distribution of the number in the system for a Bulk arrival M/D/c queue. Many approximations and bounds for the average waiting time are available (see Altinkemer *et al*, 1998 for a survey).

Model for the Location of Congested Hubs (MoLoCH), when the number of runways at each airport is fixed *a priori*.

In a detailed analysis of congestion at airports, Peterson *et al* (1995) model the transient behavior of a queuing system at an airport using a time horizon (one day) divided in small intervals (15 minutes). Since flights follow schedules, each interval t is characterized by its own deterministic arrival rate, λ_t, which is highly variable from one interval to the next. They also assume bulk service at the end of the period, according to a service rate μ_t, also deterministic, which may last for several periods, depending on the runway configuration and weather conditions. Later, they remark that there is almost always a variation around scheduled arrival times, so a time varying (from period to period) Poisson process could to be a better

representation of the arrival rates. They confirm the validity of a Poisson-arrivals assumption, as well as the fit of their own model using real data. They also remark that their model is not exact in the sense that services take place over the whole interval, rather than concentrating at the end of it. Their model, taking into account the mentioned remarks, is probably the best approximation to the real behavior of the congestion at an airport. However, this model is not suitable for location purposes unless it is embedded in a simulation-type of location algorithm. Such a simulation algorithm should have as a starting point the location of the hubs, and a proposed schedule plan for all flights, *for each proposed hub location set*. This makes such a location algorithm practically impossible to use. Furthermore, they present recursive formulas for the waiting time and queue length, which require extensive computation time and the knowledge of large amounts of data. This data includes the arrival rates at every 15-minute-period of the day, the service rates at all times, and all the transition probabilities of going from one service rate to a different one. Thus, the use of their model imbedded in a location optimization algorithm can be discarded.

Since our goal is to construct a model for strategic planning purposes (hub location), as opposed to a model for traffic control policy-making, as Peterson's *et al*, (1995), we do not focus on a transient or microscopic analysis of the delays at the airports. Rather, we use an *approximation* of the behavior of the airport from the point of view of congestion, which allows us to develop a tractable hub location model.

In order to construct the location optimization model, we use a peak-hour analysis. This is a worst case analysis, and the hope is, that improving the performance at

peak hours will reduce also the queuing effects during off-peak time. We focus on the landing runways. Takeoff runways could be assumed to behave in the same way as landing runways, only delayed by the service time at the gate. For the sake of clarity, we assume that there are different runways for landing and takeoff, so the demand for service is composed only by the landing flights. In the case of airports where the same runways are used for both landing and departure, the assumption can be made that the service time at the gates (between landing and departure) is approximately exponential. Then, the departure process is Poisson. Thus, two added Poisson processes, with different average arrival rates, compose the arrival process to the runways. If such assumption can be validated in practice, our analysis and model are valid also for this case. We use also some of the assumptions in Peterson *et al* (1995) model. In particular, as they suggest, during peak hours we assume that arrivals of flights to a hub located at node k follow a Poisson distribution. We also assume a deterministic service time at the same node. As opposed to what other authors do (Gerla and Kleinrock, 1977, Peterson *et al* 1995), we do not approximate the set of undistinguishable servers (landing runways) as one server. Instead, we use a more exact M/D/c model for the system. The first hub-location model we propose is based on the plant location model (Balinski, 1965). O'Kelly (1992) and Campbell (1994) presented different uncapacitated versions of the hub location model. Since we need to maintain congestion within acceptable limits, we choose to add an extra probabilistic "capacity" constraint, as follows:

HLRA1:

$$\text{Min} \sum_i \sum_j \sum_k \sum_m c_{ijkm} x_{ijkm} + \sum_k f_k y_k \qquad (1)$$

s.t.

$$\sum_k \sum_m x_{ijkm} = 1 \qquad \forall i, j \qquad (2)$$

$$x_{ijkm} \leq y_k \qquad \forall i, j, k, m \qquad (3)$$

$$x_{ijkm} \leq y_m \qquad \forall i, j, k, m \qquad (4)$$

$$P[\text{queue length at node } k \geq b] \leq a_q \qquad \forall k \qquad (5)$$

$$P[\text{waiting time at node } k \geq t] \leq a_t \qquad \forall k \qquad (6)$$

$$y_k \in \{0,1\}, \ 0 \leq x_{ijkm} \leq 1 \qquad \forall i, j, k, m \qquad (7)$$

where:

$$x_{ijkm} = \begin{cases} 1 \ \text{if traffic from node } i \text{ to node } j \text{ goes through hubs located a nodes } k \text{ and } m \\ 0 \ \text{otherwise} \end{cases}$$

$$y_k = \begin{cases} 1 \ \text{if there is a hob located at node } k \\ 0 \ \text{otherwise} \end{cases}$$

c_{ijkm} = transportation cost from node i to node j going through hubs located at nodes k and m.

f_k = fixed cost of locating a hub (with a predetermined number of runways) at node k.

α_τ, α_q = desired upper bounds for the probabilities of inadequate waiting times or queue lengths at hubs.

In this model, the objective minimizes transportation cost, as well as fixed costs of locating the hubs. The first constraint forces the traffic to go through one or two hubs (note that if both hub subscripts are the same, the traffic goes only through one hub). The next two equations force the traffic to go through hubs that are effectively located. The new probabilistic constraint (5) forces the probability of more than b airplanes waiting on queue to be less than α_q. The alternative

equation (6) is written in terms of waiting time: the probability of an aircraft having to wait for more than a time τ is constrained to be less than α_τ.

Probabilistic equations (5) and (6) must be rewritten in an analytic, hopefully linear form, for the model to be possible to solve. In order to write a deterministic, linear equivalent to equation (5), let p_s be the steady state probability of s customers being in the system with c servers. Then, the probability of more than b airplanes waiting on queue is

$$\sum_{s=b+c+1}^{\infty} p_s, \quad \text{or} \quad 1 - \sum_{s=0}^{b+c} p_s \tag{8}$$

that is, more than b airplanes on queue and c being served. The second form just uses the fact that the sum of all probabilities is one.

We want to restrict this probability to be less than or equal to α_q. In order to derive a useful formula, we need the expression for the probabilities of different numbers of customers in the system, p_s. It has not been possible to find the expression in the reviewed literature, so we find it. Given an arrival rate of $\boldsymbol{l}$ and a service time of $T = 1/\boldsymbol{m}$, where $\boldsymbol{m}$ is the service rate, we use the generating function for these probabilities (Prabhu, 1965):

$$P(z) = \frac{(1-z)\sum_{i=0}^{c-1} v_i z^i}{1 - z^c e^{lT(1-z)}}, \tag{9}$$

where c is the number of servers, and v_i is calculated using the equations:

$$\sum_{i=0}^{c-1} v_i = c - \boldsymbol{l}T \tag{10}$$

and

$$\sum_{i=0}^{c-1} v_i z_j^i = 0 \qquad j = 1,2,\ldots,c-1 \qquad (11)$$

In equation (11), z_j is the jth root of equation

$$1 - z^c e^{IT(1-z)} = 0 \qquad (12)$$

Note that the limiting distribution for p_s exists only for $c > IT$. Otherwise, the queue length tends to infinity. In this case, equation (12) has exactly c distinct roots z_j (one of which is 1), within and on the unit circle $|z| \le 1$. (Prabhu, 1965). Once these roots are known, the set of c equations composed by equations (10) and (11) have a determinant that does not vanish, and the c unknowns v_0, v_1, ... v_{c-1} are uniquely determined. Thus, for a complete knowledge of the generating function, equation (12) is solved first for the roots z_j and next, equations (10) and (11) are solved for the parameters v_i. We remark that, because of the how the generating function is constructed, and the form in which the parameters v_s are defined, for $s \le c - 1$,

$$p_s = v_s - v_{s-1} \text{ or } v_s = \sum_{i=0}^{s} p_s .$$

Note that the roots z_j are complex numbers. However, it can be shown that the values of the parameters v_i, as well as the values of the probabilities p_s are real (see Appendix A.1). Once the generating function is known, the probabilities p_s are the coefficients of z^s in the generating function, when this function is written as

$$P(z) = \sum_{s=0}^{\infty} p_s z^s . \qquad (13)$$

In order to write equation (8) in the form of equation (13), we expand its denominator in series, in the region $\left|z^c\right| < \left|e^{-IT(1-z)}\right|$, which coincides with the region of interest $|z| \leq 1$.

$$\frac{1}{1 - z^c e^{IT(1-z)}} = \sum_{k=0}^{\infty} \left[z^c e^{IT(1-z)}\right]^k \qquad (14)$$

$$= \sum_{k=0}^{\infty} z^{kc} e^{klT} e^{-klTz}$$

$$= \sum_{k=0}^{\infty} z^{kc} e^{klT} \sum_{m=0}^{\infty} \frac{(-klTz)^m}{m!}$$

$$= \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} e^{klT} (-1)^m \frac{(klT)^m}{m!} z^{kc+m}$$

defining a new index $n = kc + m$,

$$\frac{1}{1 - z^c e^{IT(1-z)}} = \sum_{k=0}^{\infty} \sum_{n=kc}^{\infty} e^{klT} (-1)^{n-kc} \frac{(klT)^{n-kc}}{(n-kc)!} z^n \qquad (15)$$

The interested reader can verify (graphing the region of summation on the $n - k$ plane) that equation (15) is equivalent to

$$\frac{1}{1 - z^c e^{IT(1-z)}} = \sum_{n=0}^{\infty} \sum_{k=0}^{r} e^{klT} (-1)^{n-kc} \frac{(klT)^{n-kc}}{(n-kc)!} z^n \qquad (16)$$

where $r = \lfloor n/c \rfloor$, the floor function of n/c, that is, the largest integer smaller than or equal to n/c. Thus, using equations (9) and (16),

$$P(z) = (1-z) \sum_{j=0}^{c-1} v_i \sum_{n=0}^{\infty} \sum_{k=0}^{r} e^{klT} (-1)^{n-kc} \frac{(klT)^{n-kc}}{(n-kc)!} z^{n+j} . \qquad (17)$$

We define again a new summation index $s = n + j$, rewrite $r = \lfloor (s-j)/c \rfloor$, and

perform a reordering of terms analogous to that done to equation (15). Then, the

generating function can be rewritten as

$$P(z) = (1-z)\sum_{s=0}^{c-1}\sum_{j=0}^{s}\sum_{k=0}^{r} v_i e^{klT}(-1)^{s-j-kc}\frac{(klT)^{s-j-kc}}{(s-j-kc)!}z^s +$$
$$+ (1-z)\sum_{s=c}^{\infty}\sum_{j=0}^{c-1}\sum_{k=0}^{r} v_i e^{klT}(-1)^{s-j-kc}\frac{(klT)^{s-j-kc}}{(s-j-kc)!}z^s$$

After some short algebraic operations of the same nature as those already done,

we can find the coefficients p_s of z^s in the generating function:

$$p_0 = v_0$$

and, for $s \geq 1$,

$$p_s = \sum_{j=0}^{p}\sum_{k=0}^{r} v_i e^{klT}(-1)^{s-j-kc}\frac{(klT)^{s-j-kc}}{(s-j-kc)!} - \sum_{j=0}^{q}\sum_{k=0}^{u} v_i e^{klT}(-1)^{s-j-kc-1}\frac{(klT)^{s-j-kc-1}}{(s-j-kc-1)!} \quad (18)$$

where $p = \min(s, c\text{-}1)$, $r = \lfloor (s-j)/c \rfloor$, $q = \min(s\text{-}1, c\text{-}1)$ $u = \lfloor (s-j-1)/c \rfloor$.

Now, we use the fact that, for $s \leq c-1$, $p_s = v_s - v_{s-1}$, to simplify the formula to

$$p_s = v_s - v_{s-1} \qquad \text{for } s \leq c\text{-}1$$

and, for $s \geq c$,

$$p_s = \sum_{j=0}^{c-1}\left[\sum_{k=0}^{r} v_i e^{klT}(-1)^{s-j-kc}\frac{(klT)^{s-j-kc}}{(s-j-kc)!} - \sum_{k=0}^{u} v_i e^{klT}(-1)^{s-j-kc-1}\frac{(klT)^{s-j-kc-1}}{(s-j-kc-1)!}\right] \quad (19)$$

If the values of c, l and T were known, it would be possible to compute the values

of p_s for all s, using equation (19), and the probabilistic constraint (5) could be

written as

$$1 - \sum_{s=0}^{b+c} p_s \leq a_q$$

or

$$\sum_{s=0}^{b+c} p_s \geq 1 - a_q \qquad (20)$$

for all nodes were the hubs are located. In this case, l is the airplane arrival rate to the hub; c is the number of runways at the same hub, and T the service time at each runway. Thus, this constraint would be

$$v_{c-1} +$$
$$\sum_{s=c}^{c+b} \left\{ \sum_{j=0}^{c-1} \left[\sum_{k=0}^{r} v_i e^{klT} (-1)^{s-j-kc} \frac{(klT)^{s-j-kc}}{(s-j-kc)!} - \sum_{k=0}^{u} v_i e^{klT} (-1)^{s-j-kc-1} \frac{(klT)^{s-j-kc-1}}{(s-j-kc-1)!} \right] \right\} \geq 1 - a_q$$

$$(21)$$

However, neither the location of hubs and runways nor the arrival rates to hubs are known before solving the model. The locations of the hubs are given by the values of the variables y_k, and the arrival rate to a hub located at node k is given by:

$$l_k = \sum_i \sum_j \sum_m a_{ij} x_{ijkm} + \sum_i \sum_j \sum_m a_{ij} x_{ijmk}, \qquad (22)$$

where a_{ij} is the known average rate of airplanes going from node i to node j, (through hub k) computed for the peak hour. Note that the airplanes coming directly from an origin node i (first term of the expression), plus the airplanes coming from another hub, m (second term of the expression) compose the total arrival rate to a hub k

If the expression for l_k given in equation (22) could be plugged in equation (21), we would obtain a nonlinear, deterministic equivalent to constraint (5). Unfortunately, this is not possible, because the values of the parameters v_s are computed numerically, starting from a known value of the arrival rate. However, for $c > lT$, i.e. if the equilibrium condition is satisfied, the left-hand side of equation (21) <u>must be</u>

decreasing with increasing values of λ. This is so, because physically, when the arrival rate increases, the queue length increases and, consequently, the probabilities p_s must increase for higher values of s and decrease for lower values of s. As an example, suppose there is one server. If the arrival rate is zero, evidently, $p_0 = 1$, and equation (20) is satisfied. As the arrival rate increases, p_0 decreases, and the probabilities p_s, for $s \geq 1$ increase. If the arrival rate keeps increasing, probabilities p_s with small values of s decrease more and more, until equation (20) is no longer satisfied. What this means is, since the left-hand side of equation (21) is equivalent to the left-hand side of equation (20), it decreases with increasing values l, and consequently, there must exist a continuous range of values of l, defined by $l \leq l_{max}$, that make the equation hold. Furthermore, l_{max} is the value of l that makes the equation (21) hold as equality. Thus, we can numerically solve equation (21) for the variable l and find the value l_{max}. Once this value is found, any smaller value of l will satisfy equation (21). What this means is that equation (21) is equivalent to equation

$$l \leq l_{max}$$

or, using equation (22), if a value of l_{max} is computed for each node k, (assuming that there are differences between nodes in terms of service time or number of servers),

$$\sum_i \sum_j \sum_m a_{ij} x_{ijkm} + \sum_i \sum_j \sum_m a_{ij} x_{ijmk} \leq l_{max,k} \qquad (23)$$

Equation (23) is the deterministic, linear equivalent to equation (5).

In synthesis, in order to write the model, the following steps must be taken:

1. Find the candidate nodes. For each candidate node, estimate the service time, T and the number c of runways that can be built.

2. For each candidate node with particular values of T and c, find λ_{max}. This is done iteratively, by giving the arrival rate λ a starting value, solving equation (12) for roots z_j (Appendix A.2), then the set of equations (10) and (11) for parameters v_i (Appendix A.1), and finally, checking if equation (21) holds as an equality. If it does, stop. If it does not, increase or decrease the value of the arrival rate, depending if the left-hand side of equation (21) is greater than or less than the right-hand side, respectively, and solve the equations again.

3. Once the values of λ_{max} have been found, use the model

$$\text{Min } \sum_i \sum_j \sum_k \sum_m c_{ijkm} x_{ijkm} + \sum_k f_k y_k \qquad (1)$$

s.t.

$$\sum_k \sum_m x_{ijkm} = 1 \qquad \forall i,j \qquad (2)$$

$$x_{ijkm} \leq y_k \qquad \forall i,j,k,m \qquad (3)$$

$$x_{ijkm} \leq y_m \qquad \forall i,j,k,m \qquad (4)$$

$$\sum_i \sum_j \sum_m a_{ij} x_{ijkm} + \sum_i \sum_j \sum_m a_{ij} x_{ijmk} \leq \lambda_{max,k} \qquad (23)$$

$$y_k \in \{0,1\},\ 0 \leq x_{ijkm} \leq 1 \qquad \forall i,j,k,m \qquad (7)$$

Model for allocation of runways

If the number of runways at each airport must be optimized, new variables and parameters must be defined. The model is:

$$\text{Min} \sum_i \sum_j \sum_k \sum_m c_{ijkm} x_{ijkm} + \sum_{k,c} f_k^c y_k^c \qquad (24)$$

s.t.

$$\sum_k \sum_m x_{ijkm} = 1 \qquad \forall i,j \qquad (2)$$

$$x_{ijkm} \le \sum_c y_k^c \qquad \forall i,j,k,m \qquad (25)$$

$$x_{ijkm} \le \sum_c y_m^c \qquad \forall i,j,k,m \qquad (26)$$

$$\sum_c y_k^c \le 1 \qquad \forall k \qquad (27)$$

$$\sum_i \sum_j \sum_m a_{ij} x_{ijkm} + \sum_i \sum_j \sum_m a_{ij} x_{ijmk} \le \sum_c I_{max,k}^c y_k^c \quad \forall k \qquad (28)$$

$$\sum_c y_k^c \ge 1 \quad \forall k \text{ where an airport must be located} \qquad (29)$$

$$y_k^c \in \{0,1\}, \ 0 \le x_{ijkm} \le 1 \qquad \forall i,j,k,m \qquad (30)$$

In this model, the new variable y_k^c is one if c runways are located at hub k. Constraints (25) and (26) have the same meaning as (3) and (4). Constraint (27) states that, for node k, only one variable y_k^c is one, thus an exact number of runways is located. The parameter I_{max}^c in constraint (28) has to be computed for each c. Constraint (29) is used when some locations k are preset.

A Heuristic Procedure to Solve the Model

The models presented in the previous section have the common characteristic of having thousands of variables and constraints for relatively small networks. Therefore, the use of traditional optimal solution methods such as linear

programming plus branch and bound (LP+BB) can become very burdensome in terms of computing times and, for relatively large networks, these methods cannot be applied. On the other hand, the deterministic constraints create an additional problem in finding integer solutions since the specified parameters are not equal to 0 or 1. This implies that the number of branches is likely to increase dramatically (see ReVelle, 1993, on Integer Friendly Programming).

Therefore, it is necessary to develop some alternative solution procedures to solve these problems. In this section we offer a heuristic for the Location of Congested Hubs (MoLoCH) model, when the number of runways at each airport is fixed *a priori.* The procedure has two phases: a construction phase and an improvement phase.

In the first phase (construction phase), a greedy adding procedure with random substitution (GRASP) is used to find the initial location of the facilities, where in each iteration the vertex with the best objective value is added to the set of locations.

In the second phase, a one opt exchange heuristic, based on the well-known Teitz and Bart (1968) procedure is used. For each one-opt exchange of facilities, the objective is computed and, if its value is better than before the trade, the new set of locations is stored. Otherwise, the old solution is restored. The procedure is executed for all facilities and potential locations, until no improvement is obtained.

Since the one-opt algorithm only considers vertices that improve the objective, the heuristic may end in a local optimum. In order to avoid being trapped in a local optimum, a tabu search procedure is developed, similar to the one presented by Benati and Laporte (1994). In essence, this tabu search explores a part of the

solution space by repeatedly examining all neighbors of the current solution, and moving to the best neighbor even if this causes the objective function to deteriorate. To avoid cycling, recently examined solutions are inserted in a constantly updated tabu list. At each iteration, a facility is selected, the m vertices that are closest to it are considered candidate nodes for it. For each of the candidates the objective is computed and the one that is not declared tabu with the highest objective is chosen. If the value of the new solution improves the objective, the new solution is stored as the best one, and the vertices where the facility has moved to is declared tabu for t iterations. Otherwise, the new solution is still implemented but it is not considered as the best solution so far. If all neighbor vertices are declared tabu, then the one with the lowest tabu tag is chosen as the new solution. The number of one opt trades needs to be fixed *a priori.*

Once the number of one-opt trades is reached, the tabu procedure is re-started using as initial solution the p_l nodes that were least visited in the previous tabu phase. This is known as the diversification step.

Observe that in each one opt trade a new solution is found. This solution may not be feasible due the nature of the capacity-like constraint (the deterministic constraints). If this happens, a penalty is added to the objective. The penalty value is proportional to the extent of the total violation of the capacity –like constraints and is added to the final objective.

Observe that we have assumed that the number of airports (hubs) that need to be located is known, but in fact, when solving the MoLoCH, we do not know how many airports we are going to locate. Therefore, the heuristic has to be modified in the following way: The procedure starts with 2 hubs and the algorithm is applied

and the best solution (if any) is stored. Then, a new hub is added and the new solution is stored if it is better than the best one so far. We keep adding hubs until a feasible solution is found. Once the feasible solution is found, we keep on adding hubs until the objective value is greater than a given percentage of the best solution so far. The rational for this procedure is the following: as we increase the number of hubs, the fixed cost component of the objective function increases while the transportation cost component decreases. If both functions are convex, the total cost function is also convex and there is only one minimum that is global. But if the fixed cost function decreases as the number hubs increases, but is not convex, and/or if the transportation cost function increases as the number of hubs increases but is not convex, the total cost function to minimize may present some local minima. Therefore, it is necessary to explore the neighborhoud of the best solution to see if we are in a local minimum.

Computational experience

In order to test the heuristic, we randomly generated 900 instances on a 30-vertex square grid, with uniform flow distributions between vertices in [0,5]. The right hand side of the capacity-like constraint was set to 1200, 1400 and 1600 to see how the tightness of the constraint affects the solution of the algorithm. The transportation costs between vertices were obtained by computing their euclidean distance. Savings in hub-to-hub transportation were set to 50%. Fixed costs were set to 10.000, 25.000 and 50.000.

The number of one opt trades in the tabu phase of the heuristic was set to 200*(30-p), where p is the number of hubs to locate. The diversification phase was set to 3. The size of the neiborhood was set to between 4 and 8.

For each generated network, the model was also solved to optimality using complete enumeration. That is, all possible combinations of hub locations were found for p=2,3,4…until no improvement of the objective. Therefore, we were able to examine the performance of the heuristic in terms of accuracy. Results are presented in Table 1.

In this table, the first column indicates the fixed cost parameter for each hub used in the objective. That is, the fixed cost was the same for all candidate nodes. In the second column, the value of the right hand side (RHS) of the capacity-like constraint used. Again, the same capacity was considered for all potential hub locations. In the third column the average left hand side of the capacity-like constraint is shown for each fixed cost and RHS. The average value of the total transportation cost is indicated in the fourth column. The fifth one shows the average number of hubs that are finally located. Finally, the last three columns indicate the performance of the heuristic in terms of optimality. The sixth column shows the number of times that the one-opt heuristic found the optimal solution, while in the next column indicates the number of times that the tabu phase improved the solution of the one-opt phase and obtained the optimal solution. Finally, the eighth column shows the number of times that the algorithm did not find the solution and the ninth one the percentage of the average deviation of the objective value when its optimal value was not found.

As expected, as the fixed cost increases, the average cost also increases, since less hubs are located. For each RHS, the behaviour of all instances in terms of transport costs, average number of hubs and the average left hand side of the capacity-like constraint was very similar.

The performance of the heuristic in computing terms was quite satisfactory, as shown in Table 1. The worst perfomance was obtained when the fixed cost was set to 10.000 and the RHS to 1200, were 6% of the runs (6 out of 100) did not obtain the optimal solution. In this instance, the average deviation from optimality was equal to 6%. In general, as the number of hubs decreases, the performance of the heuristic tends to improve, as it is usual when one-opt heuristics are used in location models.

Computing time was very similar across the different instances. On average, the heuristic took 25 seconds on a Pentium II 366 PC with 256 Mb of RAM.

Conclusions

In this paper a new hub location model has been formulated. This model locates hubs so as to minimize total costs (Fixed and transportation costs) while taking into account congestion. The model also considers the number of runways to open in each hub. Two versions of the model are offered: in the first one, the number of runways is fixed *a priori,* while in the second one the number of runways to open in each hub is determined by the model itself. The key feature of the model is the transformation of the probabilistic constraint that states that the amount of congestion in a hub cannot exceed a given threshold with a probability alpha into a

deterministic linear constraint. Note that the hubs are modeled as M/D/c queuing systems, and a novel procedure is developed for solving exactly such systems.

A one-opt meta-heuristic has been used to obtain model solutions. This heuristic has two phases. The first one finds an initial solution to the problem and the second one tries to improve it, by using first the well known Teitz and Bart (1968) one-opt heuristic, and then executing a tabu search. The heuristic was tested in 900 different 30 nodes-networks with success.

Further research on congested hub systems could include more exact models of the congestion, although any improvement on these could lead to intractable formulations.

Appendixes

A.1. The roots of equation

$$1 - z^c e^{lT(1-z)} = 0 \qquad\qquad (12)$$

are, in general, complex numbers. Some of them may have real values. If z_j is a complex a root of the equation, so is z^*_j, the conjugate complex. In effect, if

$$z_j = a + ib = me^{iq}$$

is a root of equation (12), then

$$1 - m^c e^{icq} e^{lT(1-a-ib)} = 0$$

$$1 - m^c e^{i(cq - blT)} e^{lT(1-a)} = 0$$

$$1 - m^c [\cos(cq - blT) + i\,\mathrm{sen}(cq - blT)] e^{lT(1-a)} = 0$$

$$1 - m^c e^{lT(1-a)} \cos(cq - blT) + ie^{lT(1-a)} \mathrm{sen}(cq - blT) = 0 \qquad (A1)$$

both the real and the imaginary part of the left-hand side of the equation must vanish, for the equation to hold. Since both vanish, the same happens for z^*_j, the conjugate of z_j, since for this conjugate, the left-hand side of equation (12) is just the conjugate of equation (A1).

In order to compute the parameters and v_i, we use the equations (10) and (11). Since z_j and z^*_j, are roots of equation (12), the following two equations are part of the set (11):

$$v_0 + v_1 z_j + v_2 z_j^2 + \ldots + v_{c-1} z_j^{c-1} = 0$$

and

$$v_0 + v_1 z^*_j + v_2 z^{*2}_j + \ldots + v_{c-1} z^{*c-1}_j = 0 .$$

Note that z^{*s}_j is the conjugate of z^s_j. We can replace these equations for the sum of them and the difference of them. The sum of both equations is an equation whose coefficients are real (the imaginary part vanishes). The difference is an equation whose coefficients are purely imaginary. If we divide the equation by i, the root of (-1), the equation is entirely real. Thus, all the equations are real. Since the right-hand side of these equations, as well as equation (10) are real, so are the values of the parameters v_i and, consequently, the values of the probabilities p_s.

A.2. In order to solve equation (12), write $z = \beta/\alpha$, with $\alpha = \lambda T/c$, and plug in the equation (12), which can be then put in the following form (Syski, 1986):

$$\boldsymbol{b}e^b = \sqrt[c]{1}ae^a.$$

The cth root of 1 is:

c	roots
1	1
2	1, -1
3	1, e^{i120^o}, e^{-i120^o}
4	1, -1, e^{i90^o}, e^{-i90^o}

and so on.

References

Altinkemer K., Bose I., Pal R., 1988: "Average Waiting Time of Customers in an M/D/k queue With Nonpreemptive Priorities", *Computers Ops. Res.* **25**, 4, 317-328.

Aykin, T, 1988: "On the location of hub facilities", *Transportation Science*, **22**, 155 - 157.

Aykin, T., 1994, "Lagrangian Relaxation based approaches to capacitated hub-and-spoke network design problems", *European Journal of Operational Research*, **79**, 501 - 523.

Aykin, T., 1995: "The hub location and routing problem", *European Journal of Operational Research*, **83**, 200 - 219.

Balinski, M. L., 1965: "Integer Programming: Methods, Uses, Computation", *Management Science*, **12**, 253 – 313.

Batta R, 1988: "Single Server Queuing - Location Models with Rejection", *Transportation Science*, **22**, 209 - 216.

Batta R, 1989: "A Queuing - Location Model with Expected Service Time dependent Queuing Disciplines", *European Journal of Operational Research*, **39**, 192 – 205.

Batta R, Larson R, Odoni A, 1988: "A Single - Server Priority Queuing - Location Model", *Networks*, **8**, 87 – 103.

Berman O, Larson R, 1985: "Optimal 2 - Facility Network Districting in the Presence of Queuing", *Transportation Science*, **19**, 261 – 277.

Berman O, Larson R, Chiu S,1985: "Optimal Server Location on a Network Operating as a M/G/1 Queue", *Operations Research*, **12(4)**, 746 – 771.

Berman O, Larson R, Parkan C, 1987: "The Stochastic Queue p - Median Location Problem", *Transportation Science*, **21**, 207 – 216.

Berman O, Mandowsky, R, 1986: "Location - Allocation on Congested Networks", *European Journal of Operational Research,* **26**, 238 – 250.

Benat Si and G. Laporte 1964: "Tabu Search Algorithms for the ($r|X_p$)-Medianoid and ($r|p$)-Centroid Problems". *Location Science* **2**(4), 193-204.

Campbell, J. F., 1994: "Integer programming formulations of discrete hub location problems", *European Journal of Operational Research*, **72**, 387 - 405.

Crommelin C.D., 1932: "Delay Probability Formulae When the Holding Times are Constant", *P. O. Elec. Engrs. J.*, **25**, 41-50.

Gavish B., 1991: "Topological Design of Telecommunications networks - Local access design methods", *Annals of Operations Research*, **33**, 17 - 71.

Gerla M., Kleinrock L., 1977: "On the Topological Design of Distributed Computer Networks", *IEEE Transactions on Communications*, **COM-25**, 1, January 1977, 48 – 60.

Gross D., Harris C.M., 1974: *Fundamentals of Queuing Theory*, New York, John Wiley and Sons, 1974.

Kleinrock L., 1975: *Queuing Systems: Volume I, Theory*, New York, Wiley Interscience, 1975.

Klincewicz, J. G., 1989: "Avoiding local optima in the *p*-hub location problem using tabu search and grasp", AT&T Bell Laboratories, Holmdel, N.J.

Klincewicz, J. G., 1991: "Heuristics for the *p*-hub location problem", *European Journal of Operational Research*, **53**, 25 - 37.

Knessl C., Matkowsky B.J., Schuss Z. and Tier C., 1990: "An Integral Equation Approach to the M/G/2 Queue", *Operations Research*, **38**, 3, 506-518.

Marianov V, ReVelle C, 1994: "The Queuing Probabilistic Location Set Covering Problem and Some Extensions", *Socio-Economic Planning Sciences*, **28(3)**, 167 – 178.

Marianov V., Revelle C., 1996: "The Queueing Maximal Availability Location Problem: a Model for the Siting of Emergency Vehicles", *European Journal of Operations Research*, **93**, 110 - 120.

Marianov V., Serra D., 1998: "Probabilistic Maximal Covering Location-allocation Models for Congested Systems", *Journal of Regional Science*, **38**(3), 401-424.

Marianov V., Serra D., 1998: "Location Set Covering Location-Allocation Models with Constrained Waiting Time or Queue Length for Congested Systems", *UPF working Paper* 177. Submitted for publication.

Marianov V., Pérez M., Ríos M., 1995: "Location of Nodes on Communications Networks Considering a User Pre-Specified Minimum Reliability Under Failures", Third International Conference on Telecommunications Systems: Modeling and Analysis, Nashville, Tennessee, March 16-19.

Marianov V., Serra D., ReVelle C., 1999: "The location of hubs in a competitive environment" *European Journal of Operational Research*, **114**, 363-371.

Newell, G. F., 1979: "Airport Capacity and Delays", *Transportation Science*, **13**, 201 - 241.

O'Kelly, M. E., 1986a, "The location of interacting hub facilities", *Transportation Science*, **20**, 92 - 106.

O'Kelly, M. E., 1986b "Activity levels at hub facilities in interacting networks", *Geographical Analysis*, **18**, 343 - 356.

O'Kelly, M. E., 1987, "A quadratic integer program for the location of interacting hub facilities", *European Journal of Operational Research*, **32**, 393 - 404.

O'Kelly, M. E., 1992: "Hub facility location with fixed costs", *Papers in Regional Science: The Journal of the RSAI*, **71**, 293 - 306.

O'Kelly, M. E., Lao, Y., 1991, "Mode Choice in a Hub-and-Spoke Network: A Zero-One Linear Programming Approach", *Geographical Analysis*, **23**, Nr 4, 283-297.

Peterson M., Bertsimas D., Odoni A., 1995: "Models and Algorithms for Transient Queuing Congestion at Airports", *Management Science,* **41** (8), 1279 – 1295.

Pirkul H., Nagarajan V., 1992: "Locating concentrators in centralized computer networks", *Annals of Operations Research*, **36**, 247 - 262.

Pirkul H., Narasimhan S. and De P., 1988: "Locating Concentrators for Primary and Secondary Coverage in a Computer Communications Network", *IEEE Transactions on Communications*, **36** N° 4, 450 - 458.

Prabhu N.U., 1965: *Queues and Inventories*, New York, John Wiley and Sons.

ReVelle C. , 1993: "Facility Siting and Integer Friendly Programming*", European Journal of Operational Research* **65**, 147-158

Saaty T.L., 1961: *Elements of Queuing Theory*, New York, Mc. Graw-Hill, 1961.

Skorin-Kapov, D, Skorin-Kapov, J., 1994, "On Tabu Search for the Location of Interacting Hub Facilities", *European Journal of Operational Research*, **73**, 502 - 509.

Syski, R. 1986: *Introduction to Congestion Theory in Telephone Systems*, Amsterdam, North Holland.

Teitz, M, and P. Bart 1968: "Heuristic Methods for Estimating the Generalized Vertex Median on a Weighted Graph" Operations Research **16**, 955-965.

Table 1: Results 30-node network

Fixed Cost	RHS	Avge. LHS	Avge. transport cost	Avge # of hubs	% of optimal solutions		% of non-opt solutions	% deviation
					T&B	Tabu		
10000	1200	1002	88376	3,8	85%	9%	6%	6%
	1400	1325	100245	3,16	90%	6%	4%	3%
	1600	1443	102224	2,72	94%	3%	3%	5%
25000	1200	1061	145945	3,08	89%	5%	6%	1%
	1400	1395	156128	2,52	92%	5%	3%	2%
	1600	1480	165607	2,16	95%	5%	0%	0%
50000	1200		152833	2,88	91%	7%	2%	4%
	1400		172375	2,12	93%	7%	0%	0%
	1600		182278	2,00	98%	2%	0%	0%