Annals of Operations Research 136, 259–283, 2005

Filtering Policies in Loss Queuing Network Location Problems

ELENA FERNÁNDEZ e.fernandez@upc.es
Dpt. d' Estadística i Investigació Operativa, Universitat Politècnica de Catalunya, Barcelona, Spain

YOLANDA HINOJOSA yhinojos@us.es
Dpt. de Economía Aplicada I, Universidad de Sevilla, Sevilla, Spain

JUSTO PUERTO puerto@us.es
Dpt. de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Sevilla. 41012 Sevilla, Spain

Abstract. In this paper we deal with two stationary loss queuing network location models. We analyze the influence of filtering policies on the locational aspect of the problems. We assume that requests for service are placed at nodes of a transportation network and they arrive in time as independent homogeneous Poisson processes with different input rates. The considered policies only cover a given proportion of requests even if there are idle service units. This proportion is stationary and fixed in advance and only depends on the node where the request is originated. The objective is to find the location of the facilities together with the filtering policy to be applied that minimize the expected total cost per unit time with respect to a given cost structure. Properties and computational results are presented enabling the resolution of these problems efficiently and showing the good performance of filtering policies in terms of both the overall operating costs, and the demand that is served.

Keywords: facility location, multichannel queues

Most of the times, in the real world, requests of any kind of service cannot be known in advance since they are not deterministic. Patients requesting medical services, homes damaged asking for emergency repair services, fires whose extinction need fire brigades, incidents requiring police control... are typical examples of such situations. All these cases share two characteristics: (i) In order to provide service, some kind of displacement is required; and (ii) the actual instants of time when service is required are not known beforehand. Thus the decision-maker who manages the design of these systems should take into account, together with the optimization aspects, the stochastic behavior of the system. The first issue that has to be considered is the effectiveness measure to be used: maximizing the social satisfaction (minimizing average waiting time or maximizing the throughput of the system ...) and/or minimizing the operating cost. The first criterion looks for the equity and the second one for the efficiency of the system. "A priori" these two approaches are non-equivalent, although each one of them can be desirable depending on the perspective.

In this paper we consider the economic aspect of some location demand-responsive problems that fall into the above framework. We study the location of service centers with finite availability (finite number of service units or capacity) when the points where requests may occur and demand rates are given. Previous approaches to these kind of problems can be found in the literature (see e.g., Chiu and Larson, 1985; Frenk, Labbé, and Zhang, 1993). In these papers the authors study localization properties assuming that if a request for service arrives, and some service unit is idle, it is immediately dispatched to serve the demand. If all the service units are busy the request is lost. Our goal is to extend the analysis to more general policies that may reject a certain proportion of requests even when some service unit is idle. This proportion is set in advance and fixed for all the process. We prove results that show that when filtering (coverage/rejection) policies are applied the global efficiency of the system is improved, both in percentage of served demand and operating costs, while localization properties similar to those of Chiu and Larson (1985) and Frenk, Labbé, and Zhang (1993) still hold. It is clear that for systems without a backup unit, filtering policies may be sometimes unacceptable. However, there are situations where these policies could be applied. Consider, for instance, the case of a company for emergency repair services that receives requests of service coming from different areas. The company can decide not to cover an area or to cover only a proportion of the requests coming from this specific area if this would reduce the overall operating costs. This is also the case of planes used to extinguish fires in mountains or national parks. The high cost of dispatching a plane makes their use only acceptable when the magnitude of the fire is large enough thus justifying the cost involved, while for normal situations trucks are used. A different application to Computer Science can be found in Xu, Righter, and Shantikumar (1992). Despite their importance these anticipated policies have been hardly considered in the literature (some exceptions are Batta, 1988; Lippman, and Ross, 1971; Carrizosa, Conde, and Muñoz, 1998). The reason is the mathematical difficulty to handle these models.

We deal with loss queuing-location models where service centers with a finite number of service units have to be located. It is assumed that requests for service come from different populated areas and they arrive in time as independent homogeneous Poisson processes with different input rates. No queue is allowed and in addition, the system may reject a proportion of requests even when some service unit is idle. Our goal is to find the coverage of the requests together with the locations of the service centers so as to minimize the overall operating costs (both, travel costs and rejection costs). The proportion of requests covered by each service center must be set in advance and fixed for all the process, since the optimal locations of the service centers depend on them. Thus, the resulting filtering (coverage/rejection) policies are anticipated and stationary. We introduce decision variables to model the coverage proportions of the requests coming from the different areas. Once these decision variables are set in the model, the actual requests of service are independent homogeneous Poisson processes that are obtained by filtering the original ones with Bernouilli random variables whose probabilities of success are the coverage proportions (decision variables).

Since the classical acceptance policies in loss queuing models are particular cases of filtering policies the latter will outperform the former. By a similar argument dynamic policies will do even better. Unfortunately, we are not aware of any analytical tool to handle such policies, which justifies our approach. Other kinds of models could also be used to handle the above examples (e.g., coverage models). However, in our opinion, having different alternatives to handle a problem is more an advantage than a drawback. In addition, note that our models do not only take into account distances and on-scene service times, but also rejection costs. This supposes an important difference with "limiting coverage-radius strategies", that in our opinion are not appropriate when rejection costs occur, like in the case of our first example. On the other hand, it is clear that dynamic policies would be more appropriate to handle real life situations if they were available. In this regard our policies can be considered as an incremental step towards better understanding of stochastic location models and as building blocks for future developments in this area. Since no method can be readily applied to solve some of the potential applications that we mention, it is specifically in this context where these potential applications are useful to illustrate the capabilities and disadvantages of our models.

In our analysis the model with one service center is first considered. This gives insight to other models with more than one service center. The first model is solved exactly and filtering policies are compared with other standard policies that do not allow rejection. Second, we address a model with two service centers: a primary service center where filtering is permitted and a secondary service center that assumes the remaining proportion of requests not covered by the filtering in the primary center. The first model may leave part of the requests uncovered whereas the second one guarantees the whole coverage of the requests. Notice that since we do not allow queues any request that arrives to any of the two service centers when all their service units are busy is lost. We develop solution procedures for this second model based on a combination of enumerative and iterative methods.

Although our approach is mainly algebraic we can also use it to highlight the queuing insight into the problem. Note that loss queuing location models do not consider set-up costs since in these models the number of service units are not decision variables. However we will show how to perform a trade-off analysis to make strategic decisions on the number of service units once the optimal operating costs and their coverage levels are known for each number of units.

The results obtained in our computational experiments show that when filtering policies are used there is a considerable reduction in the overall operating costs. Additionally, the number of requests that are actually served by the system increases substantially with respect to the case when filtering is not applied. This is specially true for the model with two service centers.

The paper is organized as follows. In Section 1 the mathematical formulation of the model with one service center is presented. In Section 2 we derive localization results for this model and we develop an algorithm, which allows us to solve efficiently the problem. Section 3 includes an example, which shows the improvement of our model with respect to the model without filtering. Section 4 deals with an extension of the model, in which we consider two service centers rather than one. Section 5 derives localization results and develops a branch and bound procedure to solve the new model. In the last section we report computational results for some randomly generated problems. We first compare the performance of the two models and then we illustrate the capabilities of the models in the process of strategic decision making. The paper ends with some conclusions and an appendix where some technical details about the algorithms are described.

1. The model with one service center

We consider a demand-responsive service system in a region in which c mobile service units are garaged at the same home location (service center). Customers requesting on-scene service by a mobile unit come from different populated areas. Each area has its own arrival rate and service time, related to the travel distance from the service center to the area. Thus, the domain is modelled as a transportation network $N = (V, A)$, where the node set V ($|V| = n$) represents the populated areas and the edge set A the links between them. Requests occur at the nodes of the network, being the arrival process from the different nodes an independent homogeneous Poisson process. No queue is allowed. For ease of presentation, we assume in the formulation that the service units are dispatched to serve the requests. Nevertheless, there is no restriction in considering that the requests are sent to the service center.

The standard policy consists of not serving only those requests that arrive when all the service units are busy. In this paper we consider a filtering policy where a proportion of requests are discriminated according to their provenance so that they are excluded in advance even if the service center has some idle service units. The introduction of this new filtering policy is justified in order to enhance the global efficiency of the system.

For $i = 1, \dots, n$, let $\lambda_i \geq 0$ be the Poisson arrival rate of the requests that come from node i (say i-requests), being the arrival process from the different nodes independent; and let $w_i \geq 0$ be the mean time of on-scene service, which may depend on nodal location i. If the service units are garaged at point $z \in N$, $d(i, z)$ denotes the travel distance in N from z to node i. Finally, let ϑ be the travel velocity.

Using this notation, the expected service time of an i-request is $s_i(z) = w_i + \beta d(i, z)/\vartheta$, i.e. the mean time of on-scene service w_i plus the travel time which is proportional to the distance divided by the travel velocity $d(i, z)/\vartheta$. This proportion is determined by a travel factor $\beta \geq 1$ (for instance $\beta = 2$ is a round trip factor). Any served i-request induces a cost to the system proportional to the travel time given by $r_i(z) = \alpha\beta d(i, z)/\vartheta$ where $\alpha > 0$ represents the cost per unit travel time; and any non-served i-request induces a cost $\hat{r}_i \geq 0$. It is also assumed that we are considering a real loss system where not all the $\hat{r}_i = 0$. Otherwise, the optimal policy would be not to serve any request.

To deal with this model we introduce two kinds of decision variables z and x_i, $i = 1, \dots, n$. Each x_i stands for the proportion of covered i-requests i.e., request that are not excluded in advance by the filtering policy. z is the location for the service center that we

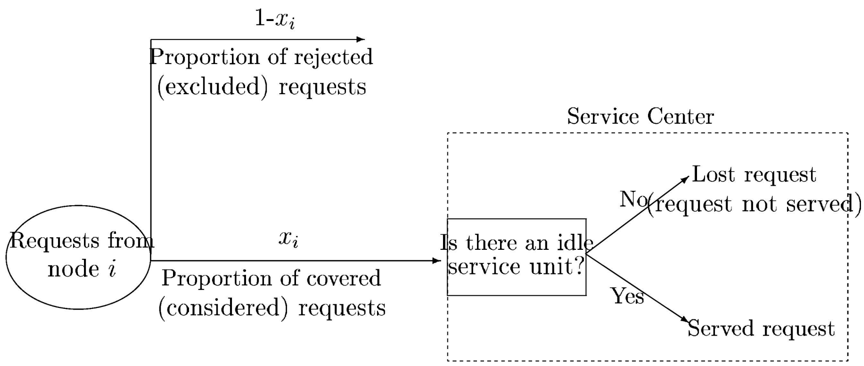

Figure 1. One service center with filtering policy.

are looking for. It is worth noting that covered requests need not be served, since queuing is not allowed. Thus, covered requests that arrive when the system is busy are lost (see figure 1). The goal is to find the location z of the service center and the proportions $\mathbf{x} = (x_1, \ldots, x_n)$ of requests to be covered that minimize the expected total cost per unit time.

Obviously, if $z \in N$ and $\mathbf{x} \in [0, 1]^n$ are fixed the system behaves as a $M/G/c/c$ queue with arrival rate $\lambda(\mathbf{x})$, mean service time $s(\mathbf{x}, z)$ and throughput $\rho(\mathbf{x}, z)$ given by

$$\lambda(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i x_i,$$

$$s(\mathbf{x}, z) = \sum_{i=1}^{n} \frac{\lambda_i x_i}{\lambda(\mathbf{x})} s_i(z),$$

$$\rho(\mathbf{x}, z) = \lambda(\mathbf{x}) s(\mathbf{x}, z) = \sum_{i=1}^{n} \lambda_i x_i s_i(z).$$

The key point that validates our model is the fact that the filtering policy is *anticipated* and *stationary*. This means that once the filtering policy is established the rate of arrivals that are going to be covered/rejected is known for each node i and, thus, the arrival process is never disrupted. Notice that we use the well-known fact that the result of filtering a Poisson process with input rate λ, by a Bernouilli random variable with success probability p, can be seen as a new Poisson process with input rate λp. This supposes a great difference with a model where the rejection decision were taken dynamically, which would be much more difficult to handle. Nevertheless, filtering policies can be considered as an approximation to dynamic ones. The approximation is obtained by (1) selecting a fixed number of time instants, (2) for each time instant estimating its input process, and then (3) identifying the optimal filtering policy associated with each time instant.

Let $\Psi_c(\mathbf{x}, z)$ denote the probability that not all the service units are busy. It is well-known (see e.g., Medhi, 1991) that in our model

$$\Psi_c(\mathbf{x}, z) = 1 - \frac{\rho(\mathbf{x}, z)^c / c!}{\sum_{k=0}^{c} \rho(\mathbf{x}, z)^k / k!}.$$

The expected total cost per unit time is given by the sum of the cost of the served requests plus the cost of the non-served requests. The expected number of i-requests per unit time that receive service is $\lambda_i x_i \Psi_c(\mathbf{x}, z)$ and the expected number of non-served i-requests per unit time is $\lambda_i(1 - x_i) + \lambda_i x_i (1 - \Psi_c(\mathbf{x}, z))$, i.e. the expected number of rejected (excluded) i-requests due to the filtering policy plus the expected number of "a priori" covered i-request that arrive when all the service units are busy (lost requests). Then, the expected total cost per unit time is:

$$\hat{F}(\mathbf{x}, z) = \Psi_c(\mathbf{x}, z) \sum_{i=1}^{n} \lambda_i x_i (r_i(z) - \hat{r}_i) + \sum_{i=1}^{n} \lambda_i \hat{r}_i. \quad (1)$$

Since the term $\sum_{i=1}^{n} \lambda_i \hat{r}_i$ does not depend on the decision variables $(\mathbf{x}, z)$ one can consider equivalently the following objective function for the model:

$$F(\mathbf{x}, z) = \Psi_c(\mathbf{x}, z)(r(\mathbf{x}, z) - \hat{r}(\mathbf{x})) \quad (2)$$

where, $r(\mathbf{x}, z) = \sum_{i=1}^{n} \lambda_i r_i(z) x_i$ and $\hat{r}(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i \hat{r}_i x_i$.

Therefore the model is formulated as:

$$\min_{\mathbf{x} \in [0,1]^n, z \in N} F(\mathbf{x}, z) = \Psi_c(\mathbf{x}, z)(r(\mathbf{x}, z) - \hat{r}(\mathbf{x})). \quad (3)$$

Hereafter and by similarity with the c-Server-Single-Facility-Loss-Median model (*c-SFLM*) of Chiu and Larson (1985) or Frenk, Labbé, and Zhang (1993), we will call this problem the Single-Facility-Loss-Median model with Filtering (*SFLMF*). We note in passing that none of these models consider set-up costs since the number of service units is fixed prior to the solution of the problem and thus is not a decision variable.

2. Localization results for the *SFLMF* problem and resolution approach

In this section we derive localization results for the *SFLMF* model. In order to prove these results we introduce the following auxiliary problem for each fixed $\bar{\mathbf{x}} \in [0, 1]^n$:

$$\min_{z \in N} t(\bar{\mathbf{x}}, z) := (\beta/\vartheta) \sum_{i=1}^{n} \frac{\lambda_i \bar{x}_i}{\lambda(\bar{\mathbf{x}})} d(i, z). \quad (4)$$

It is worth noting that the objective function $t(\bar{\mathbf{x}}, z)$ of Problem (4) is the expected travel time for the feasible solution $(\bar{\mathbf{x}}, z)$ of Problem (3). Therefore, Problem (4) is the well-known Weber or median problem on a network N with respect to the set of existing facilities placed at the nodes of N (see e.g., Hakimi, 1964). In addition, in order to simplify the notation we also introduce $w(\bar{\mathbf{x}})$ the expected on-scene service time,

$$w(\bar{\mathbf{x}}) = \sum_{i=1}^{n} \frac{\lambda_i \bar{x}_i}{\lambda(\bar{\mathbf{x}})} w_i .$$

The optimization Problem (3) we are interested in is different from (4). However, we prove in the next theorem that an optimal solution of (3) can be obtained among the minimizers of $t(\bar{\mathbf{x}}, z)$. This result extends the localization result for the *c-SFLM* model, (see Theorem 2 in Chiu and Larson, 1985 or Lemma 1.1 in Frenk, Labbé, and Zhang, 1993).

Theorem 2.1. Let $\bar{\mathbf{x}}$ be a fixed nonzero vector in $[0, 1]^n$. If z^* solves problem $\min_{z\in N} t(\bar{\mathbf{x}}, z)$, then $z*$ solves the problem $\min_{z\in N} F(\bar{\mathbf{x}}, z)$.

Proof. We only need to show that $\frac{dF(\bar{\mathbf{x}},z)}{dt(\bar{\mathbf{x}},z)} \geq 0, \forall z \in N$. First, observe that

$$\frac{dF(\bar{\mathbf{x}}, z)}{dt(\bar{\mathbf{x}}, z)} = \alpha \Psi_c(\bar{\mathbf{x}}, z)\lambda(\bar{\mathbf{x}}) + \frac{d\Psi_c(\bar{\mathbf{x}}, z)}{dt(\bar{\mathbf{x}}, z)}(r(\bar{\mathbf{x}}, z) - \hat{r}(\bar{\mathbf{x}})).$$

Recall that $\rho(\bar{\mathbf{x}}, z) = \lambda(\bar{\mathbf{x}})s(\bar{\mathbf{x}}, z) = \lambda(\bar{\mathbf{x}})(w(\bar{\mathbf{x}}) + t(\bar{\mathbf{x}}, z))$. Therefore, we can write the above expression as

$$\frac{dF(\bar{\mathbf{x}}, z)}{dt(\bar{\mathbf{x}}, z)} = \alpha \Psi_c(\bar{\mathbf{x}}, z)\lambda(\bar{\mathbf{x}}) + \frac{d\Psi_c(\bar{\mathbf{x}}, z)}{d\rho(\bar{\mathbf{x}}, z)}\lambda(\bar{\mathbf{x}}) \left(\alpha(\rho(\bar{\mathbf{x}}, z) - \lambda(\bar{\mathbf{x}})w(\bar{\mathbf{x}})) - \hat{r}(\bar{\mathbf{x}})\right).$$

Since $\frac{d(1-\Psi_c(\bar{\mathbf{x}},z))}{d\rho(\bar{\mathbf{x}},z)} \geq 0$ (see page 512 of Chiu and Larson, 1985), $\frac{d\Psi_c(\bar{\mathbf{x}},z)}{d\rho(\bar{\mathbf{x}},z)} \leq 0$. Besides, $\alpha \geq 0$ and $\hat{r}(\bar{\mathbf{x}}) \geq 0$ thus,

$$\frac{dF(\bar{\mathbf{x}}, z)}{dt(\bar{\mathbf{x}}, z)} \geq \alpha\lambda(\bar{\mathbf{x}}) \left(\Psi_c(\bar{\mathbf{x}}, z) + \frac{d\Psi_c(\bar{\mathbf{x}}, z)}{d\rho(\bar{\mathbf{x}}, z)}\rho(\bar{\mathbf{x}}, z) \right).$$

Furthermore, by page 513 of Chiu and Larson (1985)

$$\Psi_c(\bar{\mathbf{x}}, z) - \frac{d(1 - \Psi_c(\bar{\mathbf{x}}, z))}{d\rho(\bar{\mathbf{x}}, z)}\rho(\bar{\mathbf{x}}, z) \geq 0,$$

thus $\Psi_c(\bar{\mathbf{x}}, z) + \frac{d\Psi_c(\bar{\mathbf{x}},z)}{d\rho(\bar{\mathbf{x}},z)}\rho(\bar{\mathbf{x}}, z) \geq 0$. Hence, $\frac{dF(\bar{\mathbf{x}},z)}{dt(\bar{\mathbf{x}},z)} \geq 0$. □

Remark 2.1. There may exist $z^* \in \arg\min F(\bar{\mathbf{x}}, z)$ such that $z^* \notin \arg\min t(\bar{\mathbf{x}}, z)$, but there always exists $z_0 \in \arg\min t(\bar{\mathbf{x}}, z)$ such that $F(\bar{\mathbf{x}}, z_0) = F(\bar{\mathbf{x}}, z^*)$. Indeed, $F(\bar{\mathbf{x}}, z^*) \leq F(\bar{\mathbf{x}}, z)$ for all $z \in N$ and if $z^* \notin \arg\min t(\bar{\mathbf{x}}, z)$ there exists $z_0 \in N$ such that $t(\bar{\mathbf{x}}, z_0) < t(\bar{\mathbf{x}}, z^*)$. By the proof of Theorem 2.1, F is non-decreasing in $t(\bar{\mathbf{x}}, z)$. Then, $F(\bar{\mathbf{x}}, z_0) \leq F(\bar{\mathbf{x}}, z^*)$. This implies $F(\bar{\mathbf{x}}, z_0) = F(\bar{\mathbf{x}}, z^*)$.

The above theorem implies that if there exists a finite dominating set (FDS) D of candidates to be optimal solutions for all the median problems (4), which is independent of the filtering policy **x**, then D is also a FDS for Problem (3). Hence, the resolution of Problem (3) reduces to solve a finite number of pure queuing problems:

$$\min_{\mathbf{x}\in[0,1]^n} F(\mathbf{x}, \bar{z}) \quad \text{for each } \bar{z} \in D. \tag{5}$$

Each one of these problems determines the optimal filtering policy $\mathbf{x}^*(\bar{z})$ associated with a given location $\bar{z}$. The optimal solution for Problem (3) is given by the location z^* and the policy $\mathbf{x}^*(z^*)$ where the minimum of $F(\mathbf{x}, z)$ is achieved.

On the other hand, Hakimi's Theorem (see Hakimi, 1964) ensures that, for any median problem on a network, there exists a node which is an optimizer. Therefore, by applying Theorem 2.1 we can conclude that the set of nodes V is a FDS of candidates to be optimal solutions for Problem (3). We next present an algorithm to solve the *SFLMF* problem on a network $N = (V, A)$.

Algorithm 2.1

Step 1.
For each fixed $\bar{z} \in V$, do:
Solve the pure queuing problem (5) with $\bar{z}$ fixed.
Obtain the optimal filtering policy $\mathbf{x}^*(\bar{z})$ and the optimal value
$F(\mathbf{x}^*(\bar{z}), \bar{z}) := F^*(\bar{z})$.

Step 2.
Obtain $\min_{z\in V} F^*(z) := F^*$, $z^* \in \arg\min_{z\in V} F^*(z)$ and $\mathbf{x}^* = \mathbf{x}^*(z^*)$.

A important task in the above algorithm is how to solve the pure queuing problem (5) when the location z of the facility has been fixed. Carrizosa, Conde, and Muñoz (1998) give procedures for solving the resulting pure queuing problem, which are an extension of the well-known $c\mu$-rule. These procedures enable the actual resolution of Problem (5) by solving one unimodal one-dimensional optimization problem per each location $\bar{z}$ fixed. Moreover, as shown in Hansen, Poggi, and Ribeiro (1991), for the case $c = 1$ it is possible, using binary search techniques, to find an optimal solution for each $\bar{z}$ fixed with overall complexity $O(n)$.

3. Efficiency of the *SFLMF* model

As we asserted in the introduction, the *SFLMF* problem is an extension of the *c-SFLM* queuing-location problem, where both the facility location and the coverage rates of requests coming from each node are decision variables. This is a very interesting extension because even in very easy examples it improves the efficiency of the system (in both, expected number of served demand and operating costs) with respect to the model with

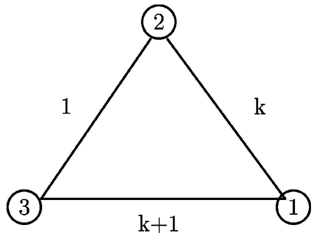

Figure 2. Network of Example 3.1.

the classical acceptance policy. An easy example illustrates this remark showing that the improvement of the *SFLMF* model with respect to the *c-SFLM* model can even be unbounded.

Example 3.1 Consider the *SFLMF* problem on the network $N = (V, A)$ where $V = \{1, 2, 3\}$ and $A = \{(1, 2), (1, 3), (2, 3)\})$ depicted in figure 2. The requests are located at the nodes of the network, and the distances between nodes appear on the edges. The parameters of the problem are $c = 1$; $\alpha = \beta = \vartheta = 1$; $w_i = 0$ for all $i = 1, \ldots, 3$; $\lambda_1 = 0.4$, $\lambda_2 = 0.25$, $\lambda_3 = 0.35$; and $\hat{r}_1 = k^2$, $\hat{r}_2 = k + 1$, $\hat{r}_3 = k + 2$.

In order to solve the *1-SFLM* problem, the results in Frenk, Labbé, and Zhang (1993) imply that an optimal location is obtained by finding the node z of V which minimizes the average service time $s((1, 1, 1), z)$. It is easily seen that the optimal location of the *1-SFLM* problem is node 2 for any $k \geq 0$. This is the solution when filtering policies are not allowed.

Now, to obtain an optimal solution of the *SFLMF* problem, we just have to solve three pure queuing problems (one for each node) using the approach earlier described in Section 2. After some algebra one finds the optimal solution of the *SFLMF* problem at the point $(\mathbf{x}^*, z^*) = ((1, 0, 0), 1)$ for all $k \geq 2$. Then, let us compare the costs associated with the optimal solutions of both problems. Since the expected total cost per unit time is $\hat{F}(\mathbf{x}, z) = \sum_{i=1}^{3} \lambda_i \hat{r}_i (1 - x_i \Psi_1(\mathbf{x}, z)) + \Psi_1(\mathbf{x}, z) \sum_{i=1}^{3} \lambda_i d(i, z) x_i$ we obtain:

$$\hat{F}((1, 1, 1), 2) = (0.4k^2 + 0.6k + 1.95)\frac{0.4k + 0.35}{0.4k + 1.35} \quad \text{(filtering not allowed)},$$
$$\hat{F}((1, 0, 0), 1) = 0.6k + 0.95 \quad \text{(filtering allowed)},$$

thus,

$$\lim_{k \to \infty} \frac{\hat{F}((1, 1, 1), 2)}{\hat{F}((1, 0, 0), 1)} = +\infty.$$

This shows that drastic reductions in costs can be obtained if filtering policies are allowed. In addition, although with the filtering policy the percentage of covered population is reduced (in this example only the requests from node 1 are considered, which represents a coverage of a 40% of the overall requests), the population that receives service increases. In fact, since the expected number of requests that receive service per unit time is $\sum_{i=1}^{3} \lambda_i x_i \Psi_1(\mathbf{x}, z)$, the expected proportion of the total demand to be served is:

$$\Psi_1((1,1,1),2) = \frac{1}{1.35 + 0.4k} \quad \text{(filtering not allowed)},$$
$$0.4\Psi_1((1,0,0),1) = 0.4 \quad \text{(filtering allowed)}$$

and it is straightforward to see that $\frac{1}{1.35+0.4k} < 0.4$ for all $k \geq 3$. This simple example shows how in order to improve the use of the resources it is sometimes convenient to perform a selection of the population to cover. In this particular case, the efficiency of the system is optimized considering only the requests in node 1.

Summarizing, this section shows how the use of filtering policies improves the throughput of this queuing-location demand-responsive system. Nevertheless, this fact cannot be taken as a recommendation for the implementation of these policies in all the cases due to the percentage of uncovered population, even when filtering policies imply a reduction on the overall costs and an increase in the served demand. These policies can only be implemented, when they are used for emergency services, provided that backup systems are available. On the other hand, if a filtering policy is found to be optimal in a situation and a backup system does not exist it would be advisable to perform a trade off analysis. The goal of this analysis would be to determine whether or not to implement the backup system in order to globally save resources when combined with the filtering policy. Assume that $(\mathbf{x}^*, z^*)$ is the optimal solution of Problem (3) and $(\mathbf{1}, \hat{z})$ is the optimal solution for the *c-SFLM* model. The saving S achieved using $(\mathbf{x}^*, z^*)$ instead of $(\mathbf{1}, \hat{z})$ during a planning horizon of T time units would be $S = T(\hat{F}(\mathbf{1}, \hat{z}) - \hat{F}(\mathbf{x}^*, z^*))$. Therefore, any amount of money smaller than S could be invested either to implement or to improve the backup service and still the whole system would save money.

Finally, we would like to notice the technical advantages of filtering policies: by using Bernouilli filters the fundamental behavior of the system does not change which allows to use standard queuing techniques. Indeed, dynamical policies would be preferable if appropriate techniques to handle them were available. As we mentioned before the strategy that consists of discretizing time and repeatedly solving the filtering model can be taken as an approximation to more general dynamic policies.

4. The model with two service centers

Our experimental results show that the filtering policy used by the one service center model is efficient in economic terms and in percentage of served demand. However, its social impact may be unacceptable due to the proportion of uncovered demand. Table 1 shows that this proportion ranges between 16.8% and 26.1% in the experiments that we have performed. This justifies the interest on new models that allow requests selection while ensuring full coverage as well. The simplest model for this is a two service centers system in which filtering policies are used. These filtering policies

Table 1
Comparison between models without and with filtering policy.

		c-SFLM		*SFLMF*			*TFLMF*	
n	c^1, c^2	*o. f.* value	% served demand	% cost reduction	% population coverage	% served demand	% cost reduction	% served demand
10	$1, \ldots, 5$	3497.94	65.17	8.54	83.13	67.68	72.86	85.41
20	$4, \ldots, 8$	7462.81	65.29	12.92	81.78	68.75	77.11	86.14
30	$8, \ldots, 12$	11256.81	66.03	14.55	80.12	70.37	78.69	89.69
50	$14, \ldots, 18$	18144.32	68.37	16.40	82.04	73.45	79.44	90.51
100	$31, \ldots, 35$	38546.37	66.27	18.31	79.13	72.79	78.33	92.88
200	$91, \ldots, 95$	95445.34	61.36	18.19	73.88	69.56	74.33	94.53

establish the proportions of i-requests to be covered by the main service center. The secondary service center assumes the remaining proportions of requests rejected by the main service center due to the filtering policy. Notice that in this model requests belonging to covered proportions that find the main service center busy are not delivered to the secondary service center, they are simply lost by the whole system. Therefore, this analysis separates each input stream into two different ones according to a "a priori" Bernouilli random variable. One of them goes to the main service center and the other to the secondary service center.

In this section, we consider the location of two service centers with c^j $j = 1, 2$ service units respectively, embedded in a transportation network $N = (V, A)$, where requests come from the nodes of N ($|N| = n$). In what follows, we refer to the main service center as indexed by $j = 1$ and to the secondary service center as indexed by $j = 2$. We denote by $w_i^j \geq 0$ the mean time of on-scene service of i-requests, served in service center j with $j = 1, 2$. If the service center j is located at z^j, the expected service time of an i-request served in service center j is $s_i^j(z^j) = w_i^j + \beta d(i, z^j)/\vartheta$. Any i-request that is served in service center j induces a cost to the system proportional to the travel time given by $r_i^j(z^j) = \alpha^j \beta d(i, z^j)/\vartheta$. Any request that arrives to the main or to the secondary service center when all the service units are busy is lost. Any lost i-request in the main or in the secondary service center induces a cost $\hat{r}_i \geq 0$.

The decision variables are z^j and x_i, with $j = 1, 2$ and $i = 1, \ldots, n$. z^j is the location for service center j and each x_i stands for the proportion of i-requests that are covered by the main service center. Therefore, $1 - x_i$ is the proportion of i-requests that are covered by the secondary service center. The goal is to find the locations z^j for $j = 1, 2$ of the two facilities and the proportions $\mathbf{x} = (x_1, \ldots, x_n)$ that minimize the expected total cost per unit time.

4.1. Analysis of the model

Since the filtering variables $\mathbf{x} = (x_1, \ldots, x_n)$ are anticipated and stationary, if $z^j \in N$ and $\mathbf{x} \in [0, 1]^n$ are fixed the queuing system of each service center j behaves as a

$M/G/c^j/c^j$ queue with arrival rate $\lambda^j(\mathbf{x})$, mean service time $s^j(\mathbf{x}, z^j)$ and throughput $\rho^j(\mathbf{x}, z^j) = \lambda^j(\mathbf{x})s^j(\mathbf{x}, z^j)$ given by

$$\lambda^1(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i x_i, \quad \lambda^2(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i (1 - x_i),$$

$$s^1(\mathbf{x}, z^1) = \sum_{i=1}^{n} \frac{\lambda_i x_i}{\lambda^1(\mathbf{x})} s_i^1(z^1), \quad s^2(\mathbf{x}, z^2) = \sum_{i=1}^{n} \frac{\lambda_i (1 - x_i)}{\lambda^2(\mathbf{x})} s_i^2(z^2) \quad \text{and}$$

$$\rho^1(\mathbf{x}, z^1) = \sum_{i=1}^{n} \lambda_i x_i s_i^1(z^1) \quad \rho^2(\mathbf{x}, z^2) = \sum_{i=1}^{n} \lambda_i (1 - x_i) s_i^2(z^2).$$

Again, the validity of this stochastic model is based on a well-known property: the result of filtering a Poisson process by a Bernouilli random variable results in two independent Poisson processes.

Let $\Psi_{c^j}^j(\mathbf{x}, z^j)$ denote the probability that not all the service units in service center j are busy. It is well-known (see e.g., Medhi, 1991) that in our model

$$\Psi_{c^j}^j(\mathbf{x}, z^j) = 1 - \frac{\rho^j(\mathbf{x}, z^j)^{c^j}/c^j!}{\sum_{k=0}^{c^j} \rho^j(\mathbf{x}, z^j)^k/k!} \qquad \text{for } j = 1, 2.$$

The expected total cost per unit time is the sum of the operating cost of the served plus the lost requests. The expected number of i-requests that are served per unit time by the main service center is $\lambda_i x_i \Psi_{c^1}^1(\mathbf{x}, z^1)$ and the expected number of i-requests that are served per unit time by the secondary service center is $\lambda_i(1 - x_i)\Psi_{c^2}^2(\mathbf{x}, z^2)$. The expected number of lost i-requests per unit time in both, the main and secondary service centers, is $\lambda_i x_i(1 - \Psi_{c^1}^1(\mathbf{x}, z^1)) + \lambda_i(1 - x_i)(1 - \Psi_{c^2}^2(\mathbf{x}, z^2))$; therefore the expected cost per unit time is:

$$\Psi_{c^1}^1(\mathbf{x}, z^1)\sum_{i=1}^{n} \lambda_i x_i (r_i^1(z^1) - \hat{r}_i) + \Psi_{c^2}^2(\mathbf{x}, z^2)\sum_{i=1}^{n} \lambda_i (1 - x_i)(r_i^2(z^2) - \hat{r}_i) + \sum_{i=1}^{n} \lambda_i \hat{r}_i. \quad (6)$$

Since the term $\sum_{i=1}^{n} \lambda_i \hat{r}_i$ does not depend on the decision variables $(\mathbf{x}, \mathbf{z})$ one can consider equivalently the following objective function for the model:

$$F(\mathbf{x}, \mathbf{z}) = \sum_{j=1}^{2} \Psi_{c^j}^j(\mathbf{x}, z^j)(r^j(\mathbf{x}, z^j) - \hat{r}^j(x)) \quad (7)$$

where, $r^1(\mathbf{x}, z^1) = \sum_{i=1}^{n} \lambda_i r_i^1(z^1) x_i$, $r^2(\mathbf{x}, z^2) = \sum_{i=1}^{n} \lambda_i r_i^2(z^2)(1 - x_i)$, $\hat{r}^1(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i \hat{r}_i x_i$, and $\hat{r}^2(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i \hat{r}_i (1 - x_i)$.

Therefore the model is formulated as:

$$\min_{\mathbf{x} \in [0,1]^n, z^j \in N} F(\mathbf{x}, \mathbf{z}) = \sum_{j=1}^{2} F^j(\mathbf{x}, z^j) \quad (8)$$

being, $F^j(\mathbf{x}, z^j) = \Psi_{c^j}^j(\mathbf{x}, z^j)(r^j(\mathbf{x}, z^j) - \hat{r}^j(\mathbf{x}))$ for $j = 1, 2$.

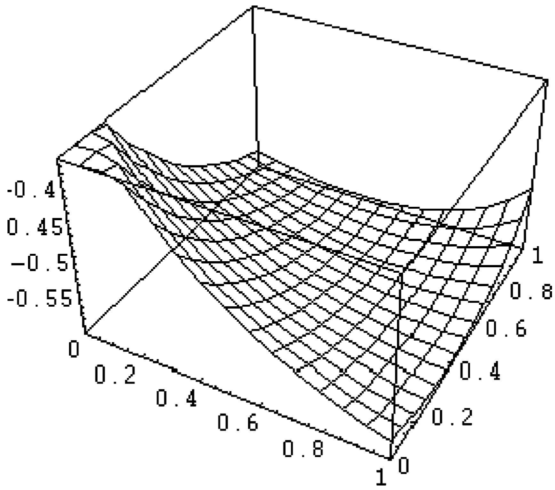

Figure 3. Objective function of Problem (13) for $n = 2$ and $c^1 = c^2 = 1$.

Hereafter we will call this problem the Two-Facilities-Loss-Median model with Filtering (*TFLMF*). The function F is sum of two fractional functions where none of them has good structural properties: (1) they are quotient of non-linear functions; (2) in general they are neither convex nor monotone (figure 3 represents the objective function F for z^1 and z^2 fixed) and; (3) they are not differentiable with respect to the location variables.

These reasons do not permit the use of standard optimization methods and lead us to exploit some properties of the problem on its locational aspect. We will prove that there exists a finite set of candidates for optimal solutions in the location variables. Thus, this reduces the problem to solve the allocation subproblem for each candidate location and to choose the best solution found.

5. Localization results and solution procedure for the *TFLMF* problem

In this section we derive localization results for the *TFLMF* model. For each fixed $\bar{\mathbf{x}} \in [0, 1]^n$, we observe that the *TFLMF* problem can be separated into two subproblems:

$$\min_{z^1 \in N} F^1(\bar{\mathbf{x}}, z^1) = \Psi^1_{c^1}(\bar{\mathbf{x}}, z^1)(r^1(\bar{\mathbf{x}}, z^1) - \hat{r}^1(\bar{\mathbf{x}})). \tag{9}$$

and,

$$\min_{z^2 \in N} F^2(\bar{\mathbf{x}}, z^2) = \Psi^2_{c^2}(\bar{\mathbf{x}}, z^2)(r^2(\bar{\mathbf{x}}, z^2) - \hat{r}^2(\bar{\mathbf{x}})). \tag{10}$$

Each one of the these subproblems is a *SFLMF* problem for a fixed $\bar{\mathbf{x}} \in [0, 1]^n$. Therefore, they can be solved independently. Indeed, as we did in Section 2 we can introduce the two following auxiliary problems for each fixed $\bar{\mathbf{x}} \in [0, 1]^n$:

$$\min_{z^1 \in N} t^1(\bar{\mathbf{x}}, z^1) := (\beta/\vartheta) \sum_{i=1}^{n} \frac{\lambda_i \bar{x}_i}{\lambda^1(\bar{\mathbf{x}})} d(i, z^1), \tag{11}$$

$$\min_{z^2 \in N} t^2(\bar{\mathbf{x}}, z^2) := (\beta/\vartheta) \sum_{i=1}^{n} \frac{\lambda_i (1 - \bar{x}_i)}{\lambda^2(\bar{\mathbf{x}})} d(i, z^2), \tag{12}$$

where the objective functions $t^1(\bar{\mathbf{x}}, z^1)$ and $t^2(\bar{\mathbf{x}}, z^2)$ are, respectively, the expected travel time for the feasible solution $(\bar{\mathbf{x}}, z^1)$ of Problem (9) and for the feasible solution $(\bar{\mathbf{x}}, z^2)$ of Problem (10). Problems (11) and (12) are Weber or median problems on a network $N = (V, A)$ with respect to the set of existing facilities placed at the nodes of N. Similarly to Section 2, we prove in the next theorem that an optimal solution of Problem (8) can be obtained by combining the minimizers of $t^1(\bar{\mathbf{x}}, z^1)$ with the minimizers of $t^2(\bar{\mathbf{x}}, z^2)$.

Theorem 5.1. Let $\bar{\mathbf{x}}$ be a nonzero vector in $[0, 1]^n$. If z^{1*} solves Problem (11) and z^{2*} solves Problem (12) then, (z^{1*}, z^{2*}) solves the problem $\min_{z^1 \in N, z^2 \in N} F(\bar{\mathbf{x}}, z^1, z^2)$.

Proof. Once $F(\bar{\mathbf{x}}, z^1, z^2)$ can be separated into $F^1(\bar{\mathbf{x}}, z^1)$ and $F^2(\bar{\mathbf{x}}, z^2)$ the proof runs parallel to the one of Theorem 2.1. □

Hakimi's Theorem (see Hakimi, 1964) together with Theorem 5.1 ensure that $V \times V$ is a FDS for Problem (8). We will use this result to give a procedure for solving Problem (8).

Next, we consider the following problem for each fixed $\bar{z}^1$ and $\bar{z}^2$:

$$\min_{\mathbf{x} \in [0,1]^n} F(\mathbf{x}, \bar{z}^1, \bar{z}^2) = F^1(\mathbf{x}, \bar{z}^1) + F^2(\mathbf{x}, \bar{z}^2) \tag{13}$$

Problem (13) is a nonlinear programming problem with linear constrains and differentiable in $(0, 1)^n$ but with non convex objective function (see figure 3). The objective function $F(\mathbf{x}, \bar{z}^1, \bar{z}^2)$ is sum of two fractional non-linear functions. These functions are well-known to be extremely hard to optimize. There exist several methods adapted to solve problems of this kind when the two terms of the sum are the ratio of linear functions (see e.g., Schaible, 1995). However, as far as we know no methods have been proposed in the literature for the general case. This fact leads us to develop "ad hoc" optimization approaches for the variable $\mathbf{x}$ minimization phase. These approaches, which are described in the Appendix, are heuristic procedures that provide approximated solutions for Problem (13). Thus, we reduce the resolution of the *TFLMF* problem to an enumeration scheme over the elements in the FDS $V \times V$.

Remark 5.1. Let $\mathbf{x}^*$ be the optimal solution of Problem (13).Then, $\mathbf{x}^*$ verifies the necessary optimality condition:

$$\nabla F(\mathbf{x}^*, \bar{z}^1, \bar{z}^2)'d \geq 0, \quad \forall\, d \text{ feasible direction at } x^*.$$

Let us assume that $0 < x_k^* < 1$, that is the kth component is fractional. Then, e^k and $-e^k$ are feasible directions at x^*. Therefore, $\nabla_k F(\mathbf{x}^*, \bar{z}_1, \bar{z}_2) = 0$. The reader can check that this conditions is equivalent to:

$$(\hat{\lambda}^{\mathbf{1}} \cdot \mathbf{x}^*)\hat{\Psi}'_{c^1}(\hat{\rho}^{\mathbf{1}} \cdot \mathbf{x}^*)\hat{\rho}_k^1 - (\hat{\lambda}^{\mathbf{2}} \cdot (\mathbf{1} - \mathbf{x}^*))\hat{\Psi}'_{c^2}(\hat{\rho}^{\mathbf{2}} \cdot (\mathbf{1} - \mathbf{x}^*))\hat{\rho}_k^2$$
$$= \hat{\Psi}_{c^2}(\hat{\rho}^{\mathbf{2}} \cdot (\mathbf{1} - \mathbf{x}^*))\hat{\lambda}_k^2 - \hat{\Psi}_{c^1}(\hat{\rho}^{\mathbf{1}} \cdot \mathbf{x}^*)\hat{\lambda}_k^1,$$

where for $j = 1, 2$ we denote $\hat{\rho}^{\mathbf{j}} = (\hat{\rho}_i^j)_{1\leq i\leq n}$, $\hat{\rho}_i^j = \lambda_i s_i^j(\bar{z}^j)$; $\hat{\lambda}^{\mathbf{j}} = (\hat{\lambda}_i^j)_{1\leq i\leq n}$, $\hat{\lambda}_i^j = \lambda_i(r_i^j(\bar{z}^j) - \hat{r}_i)$; and $\hat{\Psi}_{c^j}(t) = 1 - \frac{t^{c^j}/c^j!}{\sum_{i=0}^{c^j} t^i/i!}$.
Consider the following linear equation in the variables $y_1,\ y_2,\ y_3,\ y_4$:

$$(\hat{\lambda}^{\mathbf{1}} \cdot \mathbf{x}^*)\hat{\Psi}'_{c^1}(\hat{\rho}^{\mathbf{1}} \cdot \mathbf{x}^*)y_1 - (\hat{\lambda}^{\mathbf{2}} \cdot (\mathbf{1} - \mathbf{x}^*))\hat{\Psi}'_{c^2}(\hat{\rho}^{\mathbf{2}} \cdot (\mathbf{1} - \mathbf{x}^*))y_2$$
$$= \hat{\Psi}_{c^2}(\hat{\rho}^{\mathbf{2}} \cdot (\mathbf{1} - \mathbf{x}^*))y_3 - \hat{\Psi}_{c^1}(\hat{\rho}^{\mathbf{1}} \cdot \mathbf{x}^*)y_4. \qquad (14)$$

Since the coefficients of (14) do not depend on k the vector $(\hat{\rho}_l^1, \hat{\lambda}_l^1, \hat{\rho}_l^2, \hat{\lambda}_l^2)$ associated with any other fractional component x_l^* of $\mathbf{x}^*$ must fulfill the same linear equation (14). Thus, $(\hat{\rho}_k^1, \hat{\lambda}_k^1, \hat{\rho}_k^2, \hat{\lambda}_k^2)$ and $(\hat{\rho}_l^1, \hat{\lambda}_l^1, \hat{\rho}_l^2, \hat{\lambda}_l^2)$ must belong to the same linear variety. Notice that this arguments shows that having two fractional components in an optimal filtering policy x^*, although possible, is rather rare. This coincides with our computational experiments where we never found more than one fractional component in any optimal solution.

In addition, we can obtain lower and upper bounds for Problem (13). Indeed, if $\mathbf{x}^{1*}$ is the optimal solution to the pure queuing problem, $\min_{\mathbf{x}\in[0,1]^n} F^1(\mathbf{x}, \bar{z}^1)$ and $\mathbf{x}^{2*}$ is the optimal solution to the pure queuing problem, $\min_{\mathbf{x}\in[0,1]^n} F^2(\mathbf{x}, \bar{z}^2)$ we obtain the following inequalities:

$$F^1(\mathbf{x}^{1*}, \bar{z}^1) + F^2(\mathbf{x}^{2*}, \bar{z}^2) \leq \min_{\mathbf{x}\in[0,1]^n} F(\mathbf{x}, \bar{z}^1, \bar{z}^2) \leq \min\{F(\mathbf{x}^{1*}, \bar{z}^1, \bar{z}^2), F(\mathbf{x}^{2*}, \bar{z}^1, \bar{z}^2)\}$$

Then, $\underline{F}(\bar{z}^1, \bar{z}^2) := F^1(\mathbf{x}^{1*}, \bar{z}^1) + F^2(\mathbf{x}^{2*}, \bar{z}^2)$ is a lower bound for Problem (13) and $\bar{F}(\bar{z}^1, \bar{z}^2) := \min\{F(\mathbf{x}^{1*}, \bar{z}^1, \bar{z}^2), F(\mathbf{x}^{2*}, \bar{z}^1, \bar{z}^2)\}$ is an upper bound for Problem (13). Thus, these bounds are used to reduce the search space in the following branch and bound procedure:

Procedure

Initialization.
  Incumbent=$+\infty$
  $(z^{1*}, z^{2*}) = (0, 0)$
  For $z \in V$ obtain $F^1(\mathbf{x}^{1*}, z)$ and $F^2(\mathbf{x}^{2*}, z)$

```
Steps.
 For (z^1, z^2) ∈ V × V do :
   Evaluate F_(z^1, z^2) and F̄(z^1, z^2)
   If F̄(z^1, z^2) < Incumbent {
           Incumbent := F̄(z^1, z^2)
           (z^{1*}, z^{2*}) := (z^1, z^2)
   }
   If F_(z^1, z^2) < Incumbent {
           Solve Problem (13) for (z^1, z^2) fixed.
           Let F*(z^1, z^2) be the objective value
           If F*(z^1, z^2) <Incumbent {
                   Incumbent:= F*(z^1, z^2)
                   (z^{1*}, z^{2*}) := (z^1, z^2)
           }
   }
 EndFor
```

6. Computational experience

We have run a series of computational experiments to evaluate the performance of the proposed models and to show how to use them as a tool for strategic decision making. Since no benchmark instances for these problems can be found in the literature, we have generated a battery of test problems. Problems correspond to planar graphs randomly generated with the LEDA library (see Mehlhorn and Näher, 1999). We assume that demand points are located at the nodes of the network. Thus, for each instance the number of admissible locations is n for the model with one service center and n^2 for the model with two service centers. Problems are divided into six groups of ten problems each one according to the number of nodes n. The considered values for n are 10, 20, 30, 50, 100, 200. For each problem we have considered 5 different values of service units, both for the primary and the secondary service centers. The range of values is the same for all problems of the same size, and for the two service centers. The ranges are depicted under the column "c^1, c^2" in Table 1. They have been set so that the expected proportion of overall demand served by the system in the *c-SFLM* model is approximately 65%. Thus, for each generated problem 25 instances have been considered for the different combinations of the number of service units in each service center. This gives a number of 250 instances for each size n so that in total we have solved 1500 instances. The other parameters of the problems are the following:

1. λ_i for $i = 1, \dots, n$ randomly generated in (0, 1).
2. $w_i^j = 0$ for $i = 1, \dots, n$; $j = 1, 2$.
3. The distances between nodes are randomly generated in [100, 5000]

4. $\alpha^1 = 3, \alpha^2 = 1; \beta = 2; \vartheta = 50$.
5. $\hat{r}_i$ for $i = 1, \ldots, n$ randomly generated in [1000, 2000].

Programs have been coded in FORTRAN and executed in a DELL Inspiron 8100UT PC with a Mobile Pentium III processor at 1 GHz and RAM 128 Mb of RAM.

6.1. Comparison of the two models

First of all, note that there is no real trade-off between filtering policies and location variables. Recall that there is a finite number of candidates to optimal locations and, for each potential location, we know how to obtain the optimal filtering policy. Thus, there are no conflicting objectives in our problem which means that, in the considered models, neither the location nor the filtering decision variables play a leading role.

In addition, in these models we can draw no conclusion on where filtering occurs more. Indeed, this would depend on the cost structure, which in our case depends on various elements and not only on the distance of clients from optimal locations. It is worth noting that rejection costs are an important component of the cost function in our models. Imposing adequately high rejection cost on different types of clients one could force filtering to spread in many different ways among the nodes of the network.

For these reasons our computational experiments are devoted to compare objective function value and percentage of served demand in the two models. The first series of experiments refer to the one service center problem and is oriented to compare the models without and with the filtering policy. The average results for each group of problems are depicted in Table 1 in the columns under the heading "*c-SFLM*" and "*SFLMF*", respectively.

Column "*o.f. value*" gives the objective function value for *c-SFLM*. Column "*% cost reduction*" under the heading "*SFLMF*" gives the percentage of reduction in the objective function value when the filtering policy is applied. As was expected the use of a filtering policy results in a reduction of the operating costs with respect to *c-SFLM*. However, in the "*c-SFLM*" the covered demand is 100% and thus, the use of a filtering policy also results in a reduction of the overall covered demand. This loss in the equity of the system is a collateral effect that needs to be considered. To this end, Column "*% population coverage*" gives the values of the overall covered demand. Nevertheless, notice that to cover more population does not imply to serve more demand. Recall that covered demand is the percentage of demand that is not excluded by the filtering policy while served demand is the demand actually served by the system. The latter can be lower because of busy periods of the service units. This fact can be seen in Columns "*% served demand*" under the headings "*c-SFLM*" and "*SFLMF*", which represent the percentage of served demand in each model. As can be seen the percentage of served demand in *SFLMF* is always higher than in *c-SFLM*. The explanation for this is given by the value of the function $\Psi_c(\cdot, \cdot)$, which represents the probability of finding free service units. In the experiments we have performed this probability ranges between 0.61 and 0.68

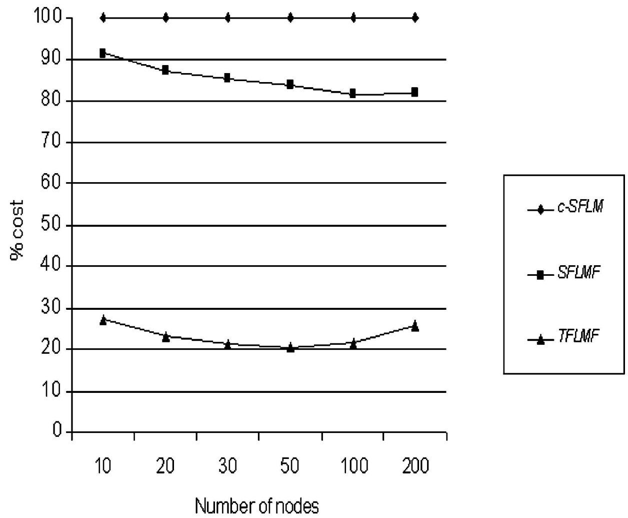

Figure 4. Cost reduction with respect to the *c-SFLM* model.

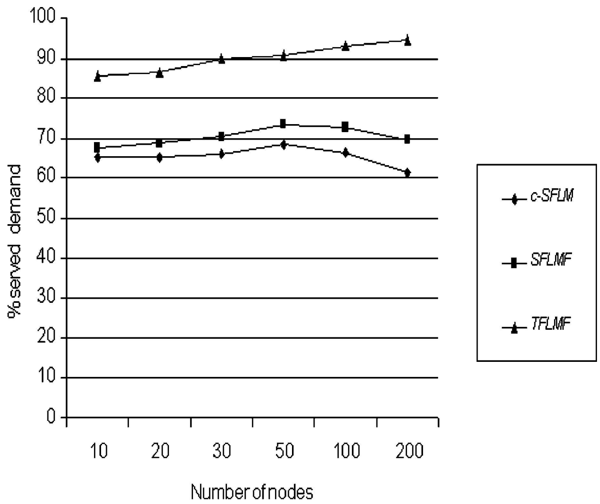

Figure 5. Percentage of served demand.

in *c-SFLM*, whereas in *SFLMF* it ranges between 0.8 and 0.94. Then, we can conclude that *SFLMF* performs a selection of the population in order to improve the efficiency of the system, resulting in a reduction in the operating costs together with an increase in the number of served demand (see figures 4 and 5). However, the use of a filtering policy might be unacceptable in terms of the proportion of uncovered demand. The two service centers model solves this problem since it recovers full coverage. In addition, the percentage of served demand with *TFLMF* (Column *% served demand* under the

Table 2
Comparison between methods for the *TFLMF* model.

	ORD				FEAS			
n	v_1	N_1	g_1	t_1	v_2	N_2	g_2	t_2
10	962,58	65,02	19,89	0,0021	949,39	64,70	19,51	0,0588
20	1790,34	250,61	21,84	0,1353	1708,32	246,25	21,03	0,5736
30	2525,78	562,69	25,10	0,1766	2399,35	553,68	24,65	1,6636
50	3838,34	1487,34	29,28	0,3693	3729,85	1474,97	29,63	7,3105
100	9057,40	3907,53	28,23	2,2546	8354,92	3722,32	27,50	39,5486
200	26187,99	4840,35	24,10	10,4279	24503,83	5081,30	26,21	137,3010

heading "*TFLMF*") increases by a factor that ranges between 1.3 for $n = 10$ and 1.6 for $n = 200$, with respect to *c-SFLM*. On the other hand, Column "*% cost reduction*" under the heading "*TFLMF*" gives the percent reduction in the objective function value with respect to *c-SFLM*. As can be seen this reduction ranges between 72.8% for $n = 10$ and 79.4% for $n = 50$. In conclusion, we can assert that *TFLMF* improves the efficiency of the system with respect to the one service center models (in both, operating costs and percentage of served demand as we can see in figures 4 and 5) while covering the overall demand. Thus, the decision-maker should perform a trade-off analysis in order to decide whether or not implement the two service centers model with the filtering policy taking into account on the one hand, the installation cost of the second service center; and on the second hand, the reduction in operating costs and the increase of the percentage of served demand.

Table 2 shows the average results obtained in the optimization of the *TFLMF* problem. Columns are divided into two blocks. The first block refers to the ordering based method (ORD) and the second one corresponds to the feasible direction method (FEAS) (see Appendix). For each method $l = 1, 2$ ($l = 1$ represents ORD, $l = 2$ represents FEAS), four values are reported: v_l is the objective function value, N_l is the number of subproblems solved (i.e. the number of subproblems that could not be fathomed by bounding), g_l the gap with respect to the lower bound in percentage, and t_l the CPU time in seconds. Note that the gaps range between 19.5% and 29.6%. The reason for these high values is that the gaps have been computed with respect to the lower bound, which is very rudimentary. Note also that the best solutions provided by the two different methods need not correspond with the same location for the service centers. This explains that better values for the objective function do not always correspond with smaller gaps, since the gap depends not only on the objective function value, but also on the value of the lower bound on the node (which, in turn, depends on the location of the service centers.)

As can be seen, in terms of the objective function value the feasible direction method outperforms ORD for problems of all sizes, although the values obtained with ORD are nearly as good as those of FEAS. In terms of the required times, the times of ORD are smaller than those of the FEAS. Nevertheless, although the times needed by FEAS are higher than those of ORD they are still small for the sizes of the considered instances. Therefore, from an overall point of view we can not decide which is the method that has obtained the best results. On the other hand, although we cannot ensure optimality, our experiences with some small-size problems solved exactly suggest that our local optima obtained by FEAS are also global.

Finally, to get further insight on the performance of each of the considered models we have arbitrarily selected one 30 nodes instance and observed the obtained results as the total number of service units varies from 1 to 30. For a fixed value for the total number of service units, say c, the results for *TFLMF* correspond to averages over all

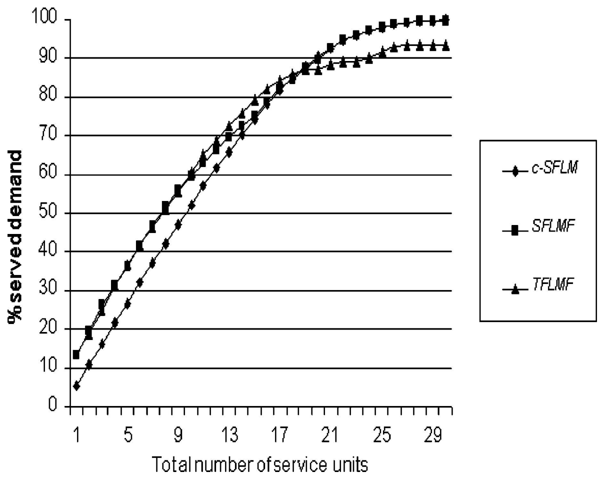

Figure 6. Percentage of served demand for $n = 30$.

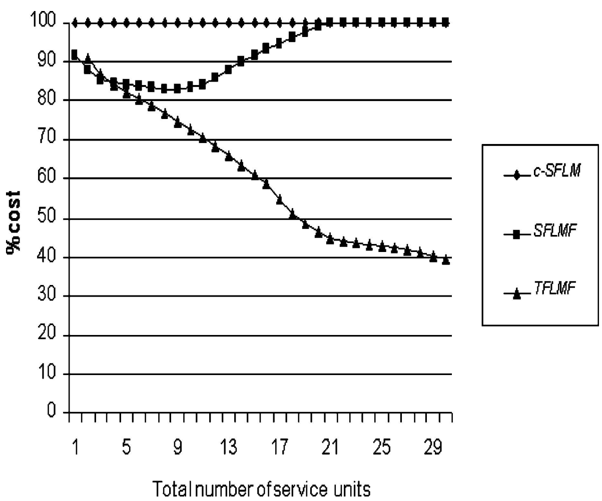

Figure 7. Cost reduction with respect to the *c-SFLM* model for $n = 30$.

combinations of service units in the primary and secondary service centers that add up to c. Figure 6 shows the percentage of the overall demand served with each model, whereas figure 7 shows the costs of *SFLMF* and *TFLMF*, evaluated in percentage relative to the cost of *c-SFLM*. First, observe that with the considered numbers of service units, the percentage of the overall demand served by *c-SFLM* nearly gives the full range, since it varies from 5.43 for $c = 1$ to 99.82 for $c = 30$. As was expected, from figures 6 and 7 we can see that as the number of service units increases the two models with one service center tend to be behave similarly. Indeed, as the number of service units increases the optimal filtering policy is to cover all requests ($x = (1, \ldots, 1)$) so that *c-SFLM* and *SFLMF* are the same model. Figure 6 also shows that for $c < 20$ the models with filtering policies are preferable to *c-SFLM*, in terms of the percentage of the overall demand served. Note that for $c \geq 20$ *c-SFLM* serves at least 90.48% of the demand so,

Table 3
Trade-off between savings and % of served demand.

c	% served demand	Savings
1	13.45	0
2	19.61	2007.24
3	26.27	3639.57
4	31.14	4809.76
5	36.35	5855.54
6	41.54	6891.23
7	46.68	7907.45
8	51.58	8886.70
9	55.78	9782.34
10	59.56	10630.33
11	62.87	11358.77
12	66.06	11982.87
13	69.43	12568.72
14	72.62	13134.98
15	75.10	13743.66
16	78.83	14369.56
17	81.90	14978.84
18	84.43	15504.00
19	87.43	15999.10
20	89.55	16441.05
21	92.39	16852.35
22	94.51	17247.60
23	96.00	17571.27
24	97.17	17825.09
25	98.05	18017.25
26	98.70	18157.47
27	99.15	18255.97
28	99.45	18322.57
29	99.65	18365.90
30	99.78	18393.06

from a practical point of view, there is really no need to implement any policy to improve the performance of the system in terms of the served demand. For $c < 10$, we can see that *SFLMF* and *TFLMF* serve approximately the same percentage of demand, but from $10 \leq c \leq 20$ the percentage of demand served by *TFLMF* improves that of *SFLMF*. It is worth mentioning that, in any case, for $c < 10$ the percentage of demand served with any of the models does not reach 60%, so the number of service units is not enough as to give an "acceptable" service with any of the models. When we analyze the results in terms of the cost function, again the two models with filtering policy behave better than *c-SFLM*. This is specially true for *TFLMF* where for all the considered numbers of service units, the value of the cost function is considerably smaller than that of *c-SFLM* (and also smaller than that of *SFLMF* excepting for $c = 2, 3$). If we just consider the values of $c < 20$, the percent reduction of the cost function value with respect to *c-SFLM* ranges from 9.13% for $c = 2$ to 53.62% for $c = 20$. If we consider higher values of c this percent reduction further increases up to 60.5% for $c = 30$, although in this case a trade off analysis would be desirable given that, as it has been pointed out before, the percent of the served demand is slightly higher for *c-SFLM* than for *TFLMF*. Therefore, from figures 6 and 7 we can conclude that when *c-SFLM* is not capable of serving an acceptable percentage of the total demand, filtering policies really pay off since they improve the percentage of the served demand and result in a considerable reduction of the cost function.

6.2. *Analysis of results for strategic decision making*

We next illustrate how to use optimal filtering policies to make decisions on the parameters of the system for *SFLMF*. A similar analysis applies to *TFLMF*. We consider again

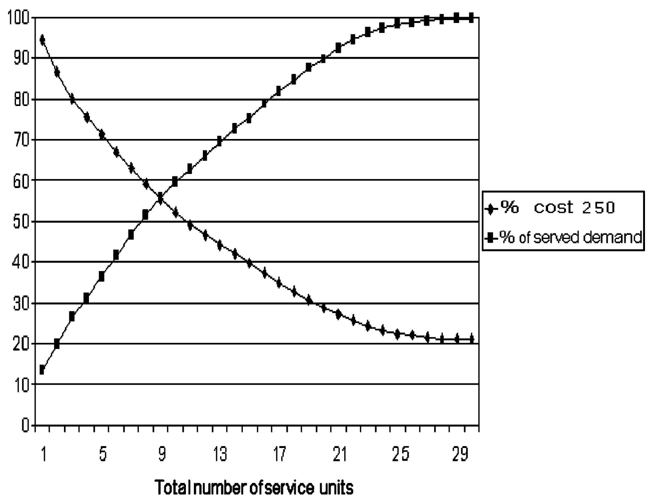

Figure 8. Operating costs and % of served demand.

the 30 nodes instance of figures 6 and 7. Table 3 shows the price that could be invested per unit time to have additional service units, in order to increase the percentage of served demand. Row i of column "*Savings*" gives the difference of the optimal operating cost between the systems with one and i service units. Note that it is not surprising that the saving increases with the number of service units since loss queuing models do not consider facility set-up costs. Thus, the costs of the system decrease when the number of service units augment since the percentage of served demand increases. Finally, figure 8 depicts the optimal operating cost and their associated percent of served demand for different values of unit service units. Thus, it can be used as a managerial tool to decide on the number of service units in order to ensure a given percentage of served demand.

7. Conclusions

This paper studies the influence of filtering policies on loss-network-queuing location problems. When one single service center is considered and filtering policies are allowed we prove results that show that the efficiency of the system can be improved while maintaining localization properties similar to those of the model without filtering policy. However, the social impact of applying filtering policies can be unacceptable due to the proportion of uncovered demand. This fact justifies the study of a new model that allows the use of filtering policies but guarantees full coverage of demand by adding a secondary service center that assumes the requests not covered by the primary service center. The resolution of these problems uses: (1) the existence of a FDS for an associated median problem on the network; and (2) the optimal admission policy of a two service centers loss-queuing model with filtering. We have tested two different methods to solve the above mentioned queuing problem. Combining these two facts, we give a branch and bound scheme that allows solving the location problem in the above settings by enumeration. The results of the computational experiments on a set of randomly generated problems confirm that the new model, in addition to recovering full coverage of demand, leads to an improvement of the efficiency of the system in both, reduction of the operating costs and increase of the percentage of served demand. This performance is not surprising since classical acceptance policies are particular cases of filtering policies. Within this framework our work is a step towards handling more realistic queuing location problems.

Although we have limited our study to the case where customers are placed at the nodes of a network, the results presented here can be extended to different spatial distributions of customers. Another possible extension of the *TFLMF* model is to consider a system with $J > 2$ facilities where the filtering would decide the proportion of requests to be served by each of the facilities. Finally, it would be interesting to study the extension of the model to include the determination of the required number of service units which will imply to introduce additional decision variables.

Appendix

We have tested two different procedures for solving the pure queuing Problem (13). The first one is a simple method based on the ordering of the variables. The second one is the iterative procedure that results from applying the reduced gradient method to Problem (13).

We denote by x_i^j the proportion of i-requests assigned to service center j.

Ordering based method

The optimal solution of the pure queuing problem with one service center is obtained by considering the ordering $\phi_i(z) = s_i(z)/(\hat{r}_i - r_i(z))$ (see Carrizosa, Conde, and Muñoz, 1998). Thus, ϕ gives a measure of the adequacy of requests to the service center at z. When the problem is extended to the case of two service centers we have two measures, one for each center. These measures provide us with two different orders (one over each service center) that may be used to allocate requests on each center. By doing this we use the same rationale behind Johnson's algorithm for the flow-shop problem with two machines (see Johnson, 1954). Let $\phi_i(z^j)$ be the function that evaluates the adequacy of assigning request i to service center at z^j. We order decreasingly the values $\phi_i(z^j)$, with $i = 1, \ldots, n$ and $j = 1, 2$ in a single list L. At each step, the procedure selects the first value $\phi_i(z^j)$ from L and assigns completely request i to service center z^j. Then, the list is updated by deleting from it the two pairs corresponding to request i. The procedure continues until the list is empty. In practice, this procedure can be implemented as follows:

Initialization.
  $x_i^j = 0, \; i = 1, \ldots, n, \quad j = 1, 2.$

For $i = 1, \ldots, n$ do
  Evaluate $\phi_i(z^1)$ and $\phi_i(z^2)$
  If $\phi_i(z^1) \leq \phi_i(z^2)$ then $(x_i^1 := 1)$
  else $(x_i^2 := 1)$

EndFor

Feasible direction method

It is the iterative procedure that results from applying Wolfe's reduced gradient method (see Bazaraa, Sherali, and Shetty, 1993) to Problem (13). Let $\mathbf{x}^c$ denote the current point and $S = [0, 1]^n$ be the feasible domain. The maximum descent feasible direction at $\mathbf{x}^c$ can be found by solving:

$$\min_{\mathbf{x} \in S} \nabla f(\mathbf{x}^c)(\mathbf{x} - \mathbf{x}^c) = -\nabla f(\mathbf{x}^c)\mathbf{x}^c + \min_{\mathbf{x} \in S} \nabla f(\mathbf{x}^c)\mathbf{x}.$$

The solution to this problem is given by:

$$x_i = \begin{cases} 1 & \text{if } \nabla_i f(\mathbf{x}^c) \leq 0 \\ 0 & \text{otherwise} \end{cases}.$$

Hence, the descent direction d at $\mathbf{x}^c$ is:

$$d_i = \begin{cases} 1 - x_i^c & \text{if } \nabla_i f(\mathbf{x}^c) \leq 0 \\ -x_i^c & \text{otherwise} \end{cases}.$$

The step size is obtained by applying line search along the above descent direction. We take as initial point $\mathbf{x}^*$ such that $F(\mathbf{x}^*, \bar{z}^1, \bar{z}^2) = \min\{F(\mathbf{x}^{1*}, \bar{z}^1, \bar{z}^2), F(\mathbf{x}^{2*}, \bar{z}^1, \bar{z}^2)\}$, where $\mathbf{x}^{1*}$ and $\mathbf{x}^{2*}$ are the optimal solutions of the pure queuing problem $\min_{\mathbf{x}\in[0,1]^n} F^1(\mathbf{x}, \bar{z}^1)$ and $\min_{\mathbf{x}\in[0,1]^n} F^2(\mathbf{x}, \bar{z}^2)$, respectively.

Acknowledgments

The authors want to thank Spanish Ministry of Science and Technology trough grant numbers BFM01-2378, MTM2004-0909, TIC2003-05982-005-04 and TIC2000-1750-c06-03. The authors are grateful to two anonymous referees for their careful reading and insightful comments on the paper.

References

Batta, R. (1988). "Single-Server Queueing-Location Models with Rejection." *Transportation Sci.* 22, 209–216.

Bazaraa, M.S., H.D. Sherali, and C.M. Shetty. (1993). *Nonlinear Programming: Theory and Algorithms*. New York: John Wiley Interscience.

Carrizosa, E., E. Conde, and M. Mñoz. (1998). "Admission Policies in Loss Queueing Models with Heterogeneous Arrivals." *Management Sci.* 44(3), 311–320.

Chiu, S.S. and R.C. Larson. (1985). "Locating an *n*-Server Facility in a Stochastic Environment." *Comp. Oper. Res.* 12, 509–516.

Frenk, J.B.G., M. Labbé, and S. Zhang. (1993). "A Note on a Stochastic Location Problem." *Oper. Res. Letters* 13, 213–214.

Hakimi, S.L. (1964). "Optimum Distribution Centers and Medians of a Graph." *Oper. Res.* 12, 450–459.

Hansen, P., M.V. Poggi de Aragao, and C.C. Ribeiro. (1991). "Hyperbolic 0-1 Programming and Query Aptimization in Information Retrieval." *Math. Programming* 52, 255–263.

Johnson, S.M. (1954). "Optimal Two- and Three-Stage Production Schedules with Set-Up Times Included." *Naval Res. Logistic Quarterly* 1, 61–68.

Lippman, S.A. and S.H. Ross. (1971). "The Streetwalker's Dilemma: A Job Shop Model." *SIAM J. Applied Mathematics* 20, 336–342.

Medhi, J. (1991). *Stochastic Models in Queueing Theory*. Academic Press.

Mehlhorn, K. and S. Näher. (1999). *LEDA Library, Version 4.0.* Sasbrücken: Max-Planck-Institut für Informatik.

Schaible, S. (1995). "Fractional Programming." In Horst R. and Pardalos (eds.), *Handbook of Global Optimization.* Kluver Academic, pp. 495–608.

Xu, S.H., R. Righter, and G. Shantikumar. (1992). "Optimal Dynamic Assignment of Customers to Heterogeneous Servers in Parallel." *Oper. Res.* 40, 1126–1138.