Scientific
Research

Modeling Customer Reactions to Congestion in Competitive Service Facilities

Mohammad Saidi-Mehrabad, Ebrahim Teimory, Ali Pahlavani*

Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran.
Email: pahlavani@iust.ac.ir

ABSTRACT

This paper reviews classic approaches for modeling customers' choice behavior in competitive facility planning problems. They are either deterministic or probabilistic and work by a utility function based on some factors whether customer-independent or dependent. This paper focuses especially on congestion, the most important factor in customer to service or fixed-server systems. Various behaviors which customers may divulge when they face with a congested facility are extensively studied. We also define a new congestion-sensitivity reaction which has not been considered in the literature. Relevant modeling approaches are proposed to formulate customers-sensitivity to congestion. An illustrative example is also given to analyze and compare the proposed approaches.

1. Introduction

A large part of planning problems which ask for firms' location and pricing decisions occurs in a situation in which there exist other facilities providing the same or homogonous service or product. In the relevant models, planner aims to devise the better alternatives for a firm competing for customers' purchasing power with other firms. For example, the problem of locating shop centers, banks, ATMs, super-markets and restaurants could be modeled using this paradigm [1]. This problem known as competitive facility location model maximizes market share, revenue or profit.

Prior to coming to a decision using the model, it is required to determine how customers behave or how they decide to choose a facility and furthermore what is their expected expenditure.

There are two main categories on retail facilities choice models: descriptive-determinist approach and explicative-stochastic approach.

Descriptive approaches are based on observation. They rely on unreal assumptions such as customers choose the closest facility. Most classic location problems such as p-median [2] and MCLP [3] are often formulated based on this assumption. Hotelling [4] was the first on studying a competitive location planning model using a descriptive approach. MAXCAP [5] is a well known competitive location problem based on this approach. Customers' purchasing power is distributed among different

facilities according to a deterministic or zero-one approach which is called also full capture [6]. In this case, the whole demand of a customer is captured by a facility which is the best for him/her according to a utility function. Conventionally, the utility function is defined based on only distance or travelling time. This is true when differences between facilities are negligible, or in areas where shopping opportunities are few and transportation is difficult [7]. In many cases however, facilities are multiform, *i.e.*, they do differ in other aspects than the mere site where they are located, and customers will take these differences into account in the way they feel attracted to them [8].

In the explicative approach for formulating customers' behavior, historical information is implemented to comprehend dynamics of retail selling competition and how customers choose purchasing opportunities. Spatial interaction model as the most important branch of the explicative approach is first developed by Huff [9]. Spatial interaction is the process whereby entities at different points in physical space make contacts, demand/supply decisions or locational choices [10].

Spatial interaction models postulate that customers compare alternatives based on their evaluation of the total utility of the facility and not merely on its location. Huff argued that when customers have several alternatives, they may consider visiting different facilities rather than restricting their patronage to only one facility. Based

on this claim, Huff coined his idea that assumes the customers' behavior to be probabilistic rather than deterministic. He defined a utility function as $A_j.d_{ij}^{-\beta_2}$ where A is the facility's attraction measure and d is the distance to the facility and β_2 is the sensitivity of customers to distance. In his model, the probability of patronizing facility j by customer i (x_{ij}) is determined as

$$x_{ij} = \frac{u_{ij}}{\sum\limits_{k \in E} u_{ik}} = \frac{A_j.d_{ij}^{-\beta_2}}{\sum\limits_{k \in E} A_k.d_{ik}^{-\beta_2}}, \forall i \in N, j \in E \qquad (1)$$

where the denominator of Equation (1) sums up the utilities of customer i from all facilities (E).

As a result, if there are D_i customers resided at demand point i, the expected number of customers visiting facility j will be

$$E_{ij} = D_i.x_{ij}, \forall i \in N, j \in E \qquad (2)$$

Later, the inclusion of other characteristics in Huff's model originated other models. According to [11], characteristics of a retail facility could be categorized into two groups. The characteristics included in the first group are independent of customer's origin (e.g. product quality, price, facility's convenience level and its size). The other group includes characteristics that are dependent on the customer's origin such as distance or travelling time.

By a Multinomial Logit model [12], the above probability is given as the following.

$$x_{ij} = \frac{\exp(V_{ij})}{\sum\limits_{k \in E} \exp(V_{ik})}, \forall i \in N, j \in E \qquad (3)$$

where V_{ij} is the utility perceived by customer i from facility j. Conventionally, this utility is expressed as a linear additive function of facilities' characteristics.

As pointed before, in addition to distance there are other criteria affecting customer's choice behavior. Authors in [13] developed a competitive location and design model in which customers decide based on distance and some other design variables. These may be quality, congestion level or offered price. Among them, congestion is very important especially in a competitive service market. In the service sector, customers' impatience to being served has been considered as a main issue of competitive advantage. Convenience in terms of service speed is usually accounted for premier on price. In this atmosphere, a competitor will succeed if it responses fairly to this requirement.

Customers divulge their impatience by reacting to the level of congestion at facilities. For service-to-customer systems, the congestion is reflected to customers by waiting time or response time and for customer-to-service systems it is measured by waiting time or system occu-

pancy level.

For a competing firm that plans to maximize its market share, the congestion should be taken into account as a main customer's choice criterion. A considerable part of the literature is devoted to an approach by which customers consider the congestion of facilities at their origins. In this approach, it is assumed that customers know facilities' congestion level at the beginning and decide a facility or a set of facilities based on a measure such as mean waiting time or mean occupancy level and/or total admissions. It has been extensively studied by various researchers.

For instance, Lee and Cohen [14] studied the existence and uniqueness of equilibrium demand for service facilities serving congestion-sensitive customers. Congestion is considered by customers in their initial decisions. In [15] MAXCAP model is improved to include waiting time as a customers' choice criterion along with travelling time. However it utilizes a deterministic choice approach and assumes all facilities to be single server. A multi-server facility location problem was developed [16] in which customers' demand is distributed according to a Multinomial Logit model based on travelling and waiting time. There have been also some simultaneous optimization models. A simultaneous location and capacity optimization model is presented for a competitor in a market with customers considering the mean waiting time in their initial choices [17]. Aboolian *et al.* [18] presented a competitive web server location and design problem in which customers make choice based on the difference in expected response times between new and old facilities. Demand elasticity to congestion has been also studied in facilities planning issues [14,19]. In a different approach [20], the author formulated a model for locating multiple-server, congestible facilities. He defined demand to be elastic to travelling time and also system occupancy level. However the customers' allocation is deterministic by a central planner.

This paper criticizes the common approach for modeling congestion-sensitivity of customers involved in a competitive service market. We study the obvious reactions of customers that face by congested facilities. We also develop five different approaches for formulating such reactions in competitive planning models. Except one approach that presents a learning process on congestion level, the other approaches follow a two steps framework. In the first step customers decide probabilistically based on a utility function depending on distance and offered price. In the second step they take congestion into account and determine whether to patronize a facility or not. The manner how they react to the congestion defines the behavior.

The rest of the paper is organized as follows. Section 2 describes our proposed frameworks for modeling customers' reactions to congestion. Section 3 gives some

experimental results on the models and finally Section 4 concludes the paper and proposes future research issues.

2. Our Modeling Frameworks

Suppose that the market is a network that includes some nodes ($N = \{1, 2, ..., n\}$) as demand origins and also as potential facility sites. Let $E \subset N$ ($|E| = q$) be the set of our firm's facilities and $E' \subset N$ ($|E'| = q'$) be the set of other competitors' facilities. There are also some edges (G) each of them indicates the availability of a direct path between two nodes. The network is in a metric space equipped with distance d being the shortest path distance.

Without loss of generality, it is assumed that customers arrive from multiple infinite sources according to a Poisson process with mean demand generating rate $\gamma_i, \forall i \in N$. They are served with FIFO discipline in facilities which utilize m servers all with exponentially distributed service time with mean $1/\mu$. Buffer volume of each facility is also limited to K. We have the following performance indicators:

(Utilization factor) $\rho = \lambda / \mu$ (4)

(Mean queue length) $L = \sum_{n=m}^{K} (n-m) \Pr_n$ (5)

(Mean waiting time) $w = L / \overline{\lambda}$ (6)

where λ is the arrival rate of a facility and defined as the sum of demand generating rates of demand nodes patronizing the facility, $\overline{\lambda}$ is the effective arrival rate and $\Pr_n$ is the probability that there are n customers at a facility. This probability is a function of the arrival rate, λ and is defined according to the structure of queuing system.

With the conventional approach for formulating congestion-sensitivity as explained in Section 1, customers consider all criteria simultaneously and they have to make a definite decision on destination facilities when they are at their origins. The planner assumes that they cannot deviate from their initial decisions. Obviously this is not a real adaptation from human decision making. Since customers usually follow a changing mood and moreover don't know all criteria simultaneously, they follow a sequential decision making process.

Whether they employ a simultaneous or sequential decision making, the probability distribution of their demand should be determined.

We define this probability according to Multinomial Logit model [12] as

$$x_{ij} = \frac{e^{-vc_{ij}}}{\sum_{k \in E \cup E'} e^{-vc_{ik}}}, \forall i \in N, j \in E \cup E' \qquad (7)$$

where v is a parameter defined as $v = \pi / \sigma\sqrt{6}$, and σ is the standard deviation in taste of the customers [16]. The dispersion in facility choice increases with smaller values

for v resulted from higher values of σ.

The main indicator of the probability is the cost incurred by customers to being served, c_{ij}. The determining factors of this cost may differ for different customers.

The manner through which the congestion is included in customers' choices is the main issue considered by this paper. We describe different reactions of customers to congested competitive facilities and present appropriate approaches to determine the effective arrival rates of facilities and the firms' market share.

2.1 Customers are Insensitive to Congestion

For the case which the arriving customers are not congestion-sensitive, c_{ij} in Equation (7) is defined as the sum of offered price and cost of travelling time to the facility.

$$c_{ij} = p_{ij} + f.t_{ij}, \forall i \in N, j \in E \cup E' \qquad (8)$$

where t_{ij} is the travelling time between nodes i and j, parameter f is the cost of unit time and p_{ij} is the service price offered by the facility located at j to customer i.

We have the following term for the effective arrival rate of facility j.

$$\overline{\lambda}_j = \lambda_j = \sum_{i=1}^{n} \gamma_i.x_{ij}, \forall j \in E \cup E' \qquad (9)$$

2.2 Customers Revise their Decisions According to their New Observations on Waiting Times

In this case the congestion level is stated by mean waiting time. It is assumed that customers initially don't know anything about waiting time levels in a new established facility and they cannot foreknow congestion level. Therefore, at their first trip they choose facilities based on factors other than congestion. Their experienced waiting times are included in their second trip. This process will be continued until an equilibrium demand distribution is found.

Therefore c_{ij} is defined as the sum of offered price, cost of travelling time to the facility and cost of waiting time at the facility.

$$c_{ij} = p_{ij} + f.(t_{ij} + w_j), \forall i \in N, j \in E \cup E' \qquad (10)$$

where w_j is the mean waiting time of facility j according to Equation (6). The effective arrival rate in this case will be as the following

$$\overline{\lambda}_j = \lambda_j(1 - \Pr_K), \forall j \in E \cup E' \qquad (11)$$

where K is the maximum capacity of the system and probability $\Pr_K$ denotes the probability that there are K customers at the facility. The state probabilities of the considered queuing system are computed according to [21] as

$$Pr_k = \begin{cases} Pr_0 \dfrac{\rho^k}{k!} & for \quad k \le m \\ Pr_0 \dfrac{\rho^k}{m!m^{k-m}} & for \quad m \le k \le K \\ 0 & for \quad k > K \end{cases} \quad (12)$$

$$Pr_0 = \left[1 + \sum_{n=1}^{m} \frac{\rho^n}{n!} + \frac{\rho^m}{m!} \sum_{n=m+1}^{K} \left(\frac{\rho}{m} \right)^{n-m} \right]^{-1} \quad (13)$$

By this approach, it is assumed that at the first usage of network after a new facility's establishment or network redesign, the mean waiting times of facilities are not known to the customers and their renewed knowledge about waiting time levels affects their next choices. To formulate this framework, we present a procedure with the following steps:

1) Set $t = 0$ and $w_j^{(t)} = 0, \forall j \in E \cup E'$,

2) Compute $c_{ij}^t = p_{ij} + f.(t_{ij} + w_j^{(t)}), \forall i \in I, j \in E \cup E'$,

3) Compute $x_{ij}^t, \forall i \in I, \forall j \in E \cup E'$ using Equation (7),

4) Compute the arrival rates as $\lambda_j^{(t+1)} = \sum_{i=1}^{n} \gamma_i.x_{ij}^t$, $\forall j \in E \cup E'$ and the effective arrival rates using Equation (11),

5) Compute $w_j^{(t+1)}, \forall j \in E \cup E'$ using Equation (6),

6) Check convergence condition. If it holds, stop with the current arrival rates else set $t = t + 1$ and go to step 2. Convergence is reached when the value of two successive results for $\lambda_j, \forall j \in E \cup E'$ become close together, i.e., $\sum_{j \in E \cup E'} |\lambda_j^{(t+1)} - \lambda_j^{(t)}| < \varepsilon$, where ε is a nonnegative small real number.

Since the approach assumes that decisions made by competing firms change the congestion level and the customers need to learn how to apply it in their choices, it better suits with decision making situations which highly affect the congestion level. Therefore it is efficient for competitive location and design planning models rather than pricing models.

2.3 Customers Balks from Entering the Facility When they Arrive

Similar to Subsection 2.2, it is assumed that customers initially don't know anything about the congestion level of facilities. However they never consider congestion at their origins but behave in a sequential manner. At the first step they decide based on the sum of offered price and cost of travelling time i.e.

$$c_{ij} = p_{ij} + f.t_{ij}, \forall i \in N, j \in E \cup E' \quad (14)$$

At the second step, they react to the congestion when they arrive at a facility. As stated in [20], in the case of non-essential services, some of the arriving customers will choose not to wait if they see a long queue, i.e., they balk from waiting in the queue. We define a parameter $\beta \in [0,1]$ which accounts for the decrease of the demand with respect to the system's occupancy level faced by the customer.

Parameter β_k is the percentage of the customers willing to wait in the queue given that k other customers are attending in the facility. It is defined as

$$\beta_k = \begin{cases} 1 - \dfrac{\max(0; k - m)}{K - m} & if\ k <= K \\ 0 & otherwise \end{cases} \quad (15)$$

where m is the number of servers and K is the maximum possible capacity of the facility. A typical instance for balking function for $m = 2$ and $K = 10$ is given in **Figure 1**.

With x_{ij} defined by Equations (7) and (14), the percentage of customers i that patronize facility j and joins the queue, given that there are k other customers in the facility, is

$$\bar{\lambda}_{ijk} = \beta_k.\gamma_i.x_{ij}, \forall i \in N, j \in E \cup E', k = 1, 2, ..., K \quad (16)$$

As a result, the effective arrival rate of facility j would be as the following:

$$\begin{aligned} \bar{\lambda}_j &= \sum_{i=1}^{n} \bar{\lambda}_{ij} = \sum_{i=1}^{n} \sum_{k=0}^{K} \bar{\lambda}_{ijk}.Pr_k(\lambda_j) \\ &= \sum_{i=1}^{n} \sum_{k=0}^{K} \beta_k.Pr_k(\lambda_j).\gamma_i.x_{ij}, \forall j \in E \cup E' \end{aligned} \quad (17)$$

where λ_j is defined as

$$\lambda_j = \sum_{i=1}^{n} \gamma_i.x_{ij}, \forall j \in E \cup E' \quad (18)$$

The state probabilities (Pr_k) of the system are derived according to death and birth flow diagram [21] as the following,
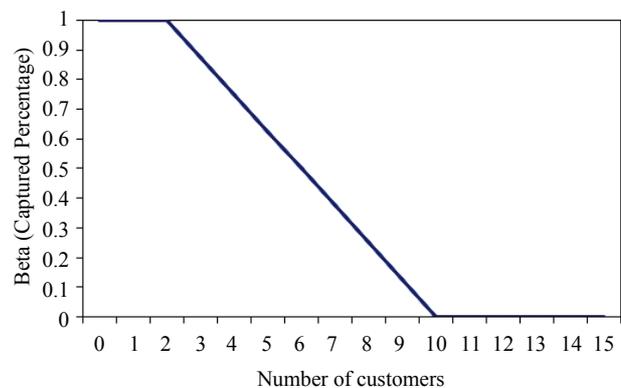

Figure 1. A typical balking function

$$\Pr_n = \begin{cases} \Pr_0 \dfrac{\rho^n}{n!}, & 0 \le n \le m \\[2mm] \Pr_0 \dfrac{\rho^n}{m!} \left[\dfrac{\beta_{n-1}\beta_{n-2}\cdots\beta_m}{m^{n-m}} \right], & m < n \le K \\[2mm] 0 & n > K \end{cases} \qquad (19)$$

$$\Pr_0 = \left[\sum_{n=0}^{m} \frac{\rho^n}{n!} + \frac{1}{m!} \sum_{n=m+1}^{K} \rho^n \left[\frac{\beta_{n-1}\beta_{n-2}\cdots\beta_m}{m^{n-m}} \right] \right]^{-1} \qquad (20)$$

2.4 Joined Customers Renege from Waiting

In the previous case, it is assumed that all the customers that join the queue stay until served by a server. However, it is also possible for an impatient customer to depart the queue, *i.e.*, he/she may renege from waiting. In this case, after joining the queue each customer will wait a certain length of time to being served. If the service has not begun by then, he/she departs. This time is a random variable whose density function is $r(t) = \alpha e^{-\alpha t}$. Consequently, the effective arrival rate will be as the following:

$$\bar{\lambda}_j = \sum_{i=1}^{n} \sum_{k=0}^{K} R_k . \beta_k . \Pr_k(\lambda_j) . \gamma_i . x_{ij} \qquad (21)$$

where λ_j, β_k, γ_i and x_{ij} are the same as Subsection 2.3 and R_k is the probability that a new arrived customer will survive to be serviced given that there are k customers in the facility on arrival and given that it joins. In [22] it is proved that

$$R_k = \frac{\delta}{\delta + k - m + 1} \qquad (22)$$

where $\delta = \dfrac{m.\mu}{\alpha}$.

The state probabilities are derived according to death and birth flow as the following,

$$\Pr_n = \begin{cases} \Pr_0 \dfrac{\rho^n}{n!}, & 0 \le n \le m \\[2mm] \Pr_0 \dfrac{\rho^m}{m!} \eta^{n-m} \left[\dfrac{\beta_{n-1}\beta_{n-2}\cdots\beta_m}{(\delta+1)_{n-m}} \right], & m < n \le K \\[2mm] 0 & n > K \end{cases} \qquad (23)$$

$$\Pr_0 = \left[\sum_{n=0}^{m} \frac{\rho^n}{n!} + \frac{\rho^m}{m!} \sum_{n=m+1}^{K} \eta^{n-m} \left[\frac{\beta_{n-1}\beta_{n-2}\cdots\beta_m}{(\delta+1)_{n-m}} \right] \right]^{-1} \qquad (24)$$

where $\eta = \lambda/\alpha$.

2.5 Balked or Reneged Customers may Veer from their Initial Destinations

In the case of essential services or a fierce competitive market, the balked or reneged customers may go directly to another facility rather than coming back to their origins, *i.e.*, they veer and deviate from their initial decisions. In this case the second facility would indirectly capture their demand. We assume that customers do such upturns only for one cycle due to travelling time and cost issues.

Figure 2 illustrates the two situations in a simple network where there is a single demand node with four customers and three facilities are serving them. Initially three customers choose to patronize facility F1 and one of them chooses facility F2 (**Figure 2(a)**). Because of unbearable congestion at facility F1, the customers balk or renege from waiting and return to their origins (**Figure 2(b)**). But this is not the case for all situations. They may go directly to another facility close to facility F1 with the aim of being served in a less congested facility (**Figure 2(c)**).

We need to determine the percentage of customers which may divulge such behavior. Assume that $\Pr_l^B = \sum_{k=0}^{K} (1-\beta_k).\Pr_k(\lambda_l), \forall l \in E \cup E'$ is the balking probability for the customers interested initially in facility l and $\Pr_l^R = \sum_{k=0}^{K} \beta_k.(1-R_k).\Pr_k(\lambda_l), \forall l \in E \cup E'$ is the reneging probability for them. To reflect the veering behavior in the model, we also define a variable z_{lj} which stands for the probability that a customer balked or reneged from facility l, will choose facility j. This proba-
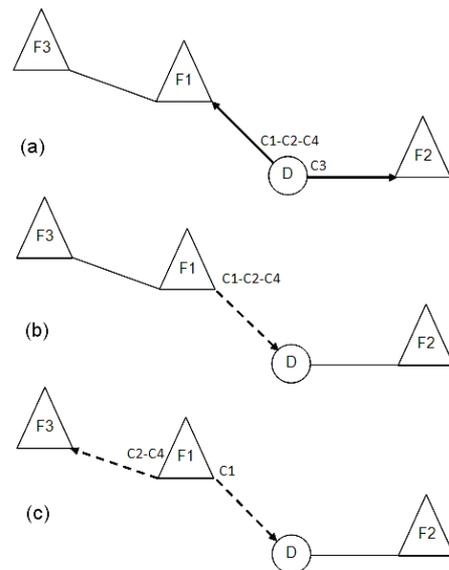

Figure 2. An example for comparison of two congestion-sensitivity manners: (a) Customers dispatching from the demand node; (b) customers balk or renege and return to their origins; (c) Some of balked or reneged customers go to another facility (veer)

bility depends on the difference of prices offered by two facilities and travelling time between them as follows:

$$z_{lj} = (1 - Q_l / d_m) \cdot \frac{e^{-c_{lj}}}{\sum_{k \in E \cup E'} e^{-c_{lk}}}, l \in E \cup E', k, j \neq l$$

(25)

$$c_{lj} = p_{ij} - p_{il} + f.t_{lj}, j, l \in E \cup E', l \neq j \quad (26)$$

where d_m is the maximum distance between two nodes in the network and Q_l is the centrality index for facility l. It is defined as

$$Q_l = \frac{\sum_{j \neq l} d_{lj}}{q + q' - 1}, \forall l \in E \cup E' \quad (27)$$

We implement the centrality index to determine how far the facility is from other facilities, *i.e.*, the density of network's areas. The density measure is utilized to define what part of balked or reneged customers from a facility will go to other facilities. For a facility established in a dense area, the probability that a balked or reneged customer will go directly to another facility will be higher than that for sparse areas. **Figure 3** illustrates the effect of centrality index for the example depicted by **Figure 2**.

2.5.1 Only Balked Customers Choose to Go to Another Facility

In this case, the arrival rate could be partitioned into two parts, one part for directly captured demand and another one for indirectly captured demand of balked customers. Therefore, Equation (18) becomes

$$\lambda_j = \sum_{i=1}^{n} \gamma_i.x_{ij} + \sum_{\substack{l \in E \cup E' \\ l \neq j}} \sum_{i \in N} \lambda_{il}.\mathrm{Pr}_l^B.z_{lj}, \forall j \in E \cup E' \quad (28)$$
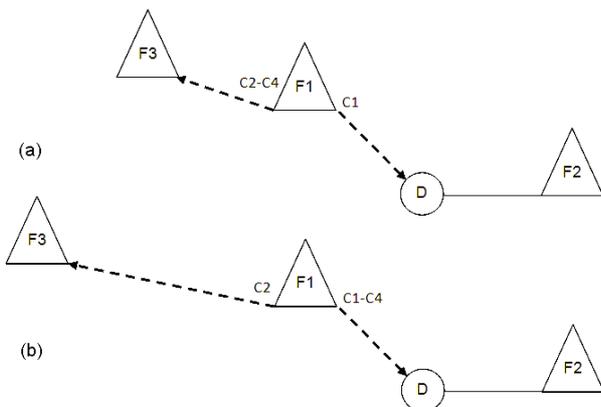

Figure 3. An example illustrating the effect of centrality index: (a) Facility F3 is closer to F1 and larger part of balked or reneged customers decide to go to F3; (b) Facility F3 is farther to F1 and smaller part of them decide to go to F3

The effective arrival rate in this case is the same as Equation (17) where probabilities Pr_k are computed using Equations (19)-(20).

2.5.2 Both Balked and Reneged Customers Choose to Go to Another Facility

In this case, the arrival rate could be partitioned into three parts, the first part for directly captured demand, the second part for indirectly captured demand of balked customers and the third part for indirectly captured demand of reneged customers.

Therefore, Equation (18) becomes

$$\lambda_j = \sum_{i=1}^{n} \gamma_i.x_{ij} + \sum_{\substack{l \in E \cup E' \\ l \neq j}} \sum_{i \in N} \lambda_{il}.\mathrm{Pr}_l^B.z_{lj}$$

$$+ \sum_{\substack{l \in E \cup E' \\ l \neq j}} \sum_{i \in N} \lambda_{il}.\mathrm{Pr}_l^R.z_{lj}, \forall j \in E \cup E' \quad (29)$$

The effective arrival rate in this case is the same as Equation (21) where probabilities Pr_k are computed from Equations (23)-(24).

Obviously, the right-hand side of Equations (28) and (29) is a function of λ_js ($j \in E \cup E'$). Therefore, it can be written as a system of equations,

$$\lambda_j = \varphi(\vec{\lambda}), \forall j \in E \cup E' \quad (30)$$

where $\vec{\lambda}$ is the vector of facilities' arrival rates. Since $\varphi(.)$ is a non-linear function and there are $(q + q')$ facilities, Equation (30) indicates a non-linear system of $(q + q')$ equations and $(q + q')$ variables.

For solving this system of equations we employ a procedure similar to fixed point iteration approach [23]. This procedure has the following steps:

1) Compute $x_{ij}, \forall i \in N, \forall j \in E \cup E'$ using Equations (7) and (14);

2) Set $t = 0$ and $\vec{\lambda}^{(t)} = 0$. Compute $\mathrm{Pr}_k(\lambda_j^{(t)})$, $\forall k = 0, 1, 2, ..., K$ using Equations (19)-(20) or Equations (23)-(24);

3) Compute $\varphi(\lambda_j^{(t)})$ using Equation (28) or (29);

4) Compute a new value

$$\vec{\lambda}^{(t+1)} = \theta.\vec{\lambda}^{(t)} + (1 - \theta).\Gamma(\vec{\lambda}^{(t)}), \quad 0 < \theta < 1 \quad (31)$$

where Γ is the vector of right hand side functions of Equation (28) or (29), $\varphi(.)$ and θ is a problem-dependant factor.

5) Check the convergence condition. If it holds, stop with the current solution else set $t = t + 1$ and go to step 3. Convergence is reached when the value of two successive results for $\vec{\lambda}$ become close together *i.e.* $\left| \vec{\lambda}^{(t+1)} - \vec{\lambda}^{(t)} \right| < \varepsilon$, where ε is a nonnegative small real number.

Having defined the possible reactions of impatient customers in congestible facilities, we can analyze them in a competitive planning model. The utilized measure for this purpose is the market share of firms or their facilities. However, other measures could also be derived. The market share of our firm is defined as

$$MS = \sum_{j \in E} \bar{\lambda}_j / \sum_{i=1}^{n} \gamma_i \qquad (32)$$

In the next section we test the approaches through an illustrative example.

3. An Illustrative Example

Suppose that an area is formed as a network that includes 50 demand nodes. Three firms are competing with each other for customers' purchasing power. They have just established some facilities. The deployment outline of the demand nodes and also the firms' facilities is exhibited by **Figure 4**.

Note that all nodes in the network indicate a demand node. An oval node indicates that a facility of firm 1 has been located in that node. A square node indicates that a facility belonging to firm 2 has been established in the node and a diamond node shows a facility of firm 3. The circles show the demand nodes with no established facility.

The length of available direct paths in the network is known and the shortest distance between each pair of nodes could be determined.

Table 1 gives the demand generating rates for demand nodes. **Table 2** gives the queuing parameters for the competing firms. The price charged for customers is assumed to be $p = 12$, the same for all three firms. The values for other parameters are given in **Table 3**.

Now we apply different congestion-sensitivity reactions and their relevant modeling approaches on the defined problem. The obtained results are given in **Table 4**. The table gives the market share of competing firms and their facilities from demand nodes. The last column gives the percentage of total captured demand of the market. **Table 4** illustrates also the results for the case which disregards congestion effects (Subsection 2.1). It is given only for comparison purposes.
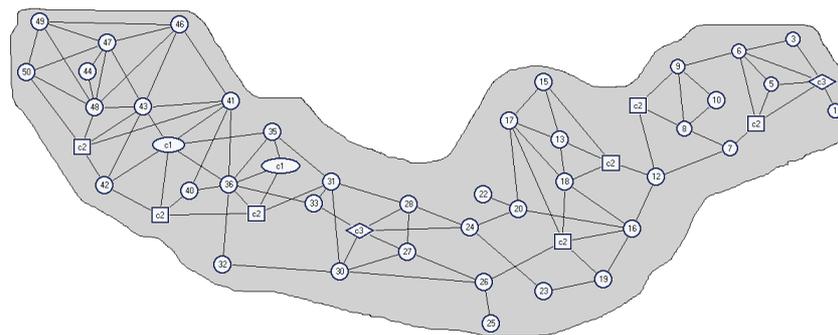

Figure 4. An outline of the market area in the example

Table 1. The nodes' demand generating rates

Node	Rate	Node	Rate	Node	Rate	Node	Rate	Node	Rate
1	1.2	11	3.9	21	1.9	31	1.1	41	3.8
2	1.7	12	1.8	22	2.7	32	4	42	3.7
3	4.1	13	4	23	1.2	33	2.3	43	0.1
4	1.5	14	0.8	24	1.1	34	2.1	44	4
5	3.3	15	0.7	25	1	35	2.4	45	1
6	2	16	1.5	26	2.2	36	1.5	46	2.6
7	1.5	17	1.7	27	0.9	37	2.9	47	2.2
8	0.8	18	1.3	28	1.7	38	1.7	48	1.3
9	3.1	19	0.7	29	0.1	39	0.4	49	3.3
10	2.8	20	1	30	2.3	40	1.2	50	0.2

Table 2. The queuing specifications of the firms

Firm	Number of Facilities	Facility Nodes (No. of Servers)	System Capacity (K)	Service Rate (μ)
Firm 1	2	37 (2), 39 (3)	10	5
Firm 2	7	2 (2), 11 (1), 14 (2), 21 (2), 34 (1), 38 (1), 45 (3)	8	4
Firm 3	2	4 (2), 29 (2)	15	5

Table 3. Other parameters of the network

Measure	Value
Customers' behavior uniformity (v)	0.1
Cost of travelling and waiting time (f)	1
Reneging rate (α)	5

As it can be seen from **Table 4**, in the cases which customers are congestion-insensitive or decide at origin based on their knowledge on waiting time levels, the whole available demand of the market is captured. This is because that; in these two cases the customers don't escape from congestion but accept it as a usual phenomenon.

The case of "Balking and reneging" results in the least market capture because the congestion-sensitive or impatient customers leave highly congested facilities and return to their origins. When a part of those leaving customers doesn't return and decides to being served by other facilities, the overall capture increases. This is reflected by "Balking, reneging and veering" case. A similar analysis could be stated for "Balking" and "Balking and veering" cases.

It is interesting to note that facility F2 of firm 1 captures maximum share of the market except for congestion-insensitivity case. This is because of its better location and also its larger number of servers. In the contrast, facility F2 of firm 2 captures the minimum share of the market except for congestion-insensitivity case. This is because that it has only one server and its system capacity and service rate are smaller than other facilities. In the congestion-insensitivity case, since the congestion effect is disregarded, the only parameters affecting customers' behavior are price and facilities' location. Since price is assumed to be the same for all facilities, their locations play the main role in determining market share. Therefore it is expected that a facility located at a dense area would capture a larger share of the market.

In the second set of experiments we analyze the effect of different parameters such as the default number of servers, mean service rate and system capacity on the firms' market shares. The results are given by **Figures 5-8**.

Figure 5 presents the analysis with respect to the firm 3's mean service rate which changes by −80% to +80% (in steps of 40%) around its base value ($\mu^0 = 5$).

Table 4. The market share of firms and facilities (percentage)

| Customer behavior	Firm 1			Firm 2								Firm 3			Overall Capture
	F1	F2	Total	F1	F2	F3	F4	F5	F6	F7	Total	F1	F2	Total	
Congestion-insensitive	9.3	9.9	19.1	9.1	8.7	7.9	10.0	9.8	9.7	8.8	63.9	8.0	8.9	17.0	100.0
Learning to revise	9.6	10.4	20.0	9.3	8.0	8.1	10.1	8.9	8.9	9.3	62.6	8.3	9.1	17.4	100.0
Balking	8.4	9.4	17.9	8.0	6.4	7.0	8.7	7.1	7.1	8.3	52.4	7.6	8.4	16.0	86.3
Balking and veering	9.4	10.6	20.0	8.7	7.0	7.7	9.5	8.1	8.0	9.1	58.1	8.2	9.2	17.4	95.4
Balking and reneging	7.4	9.0	16.3	6.7	4.3	6.0	7.2	4.7	4.6	7.8	41.2	6.6	7.2	13.9	71.4
Balking, reneging and veering	9.3	11.1	20.4	8.2	5.8	7.5	9.0	6.9	6.9	9.4	53.7	7.8	8.8	16.6	90.8

(a)

(b)
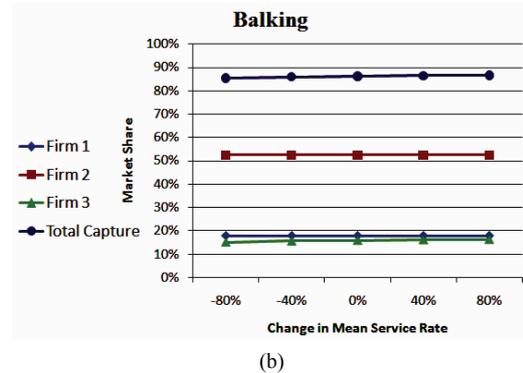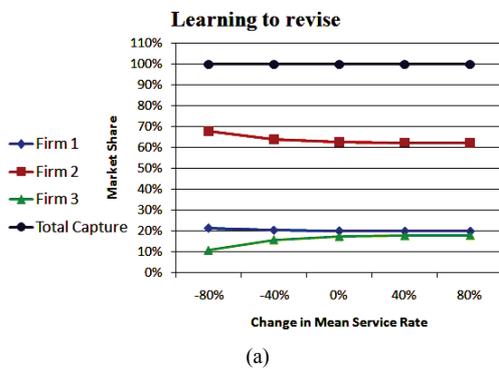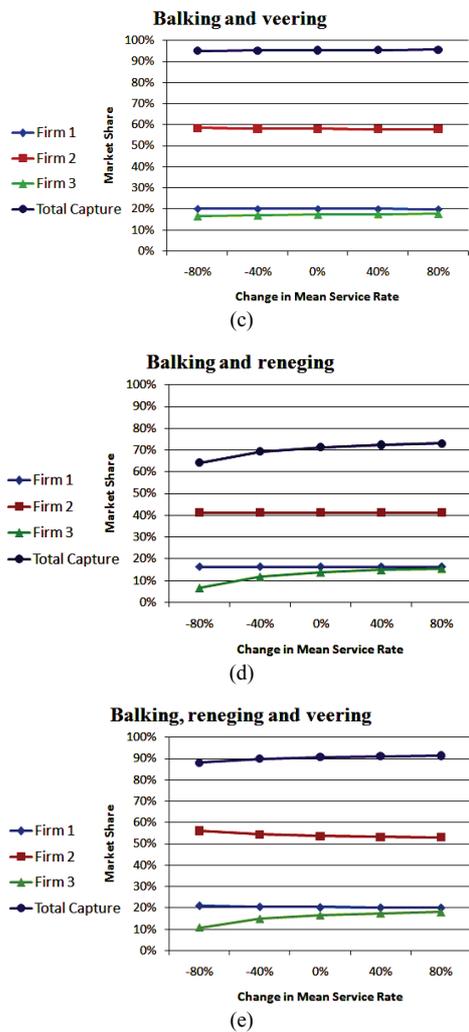
Figure 5. Sensitivity of firms' market share to the change in mean service rate

Similarly, **Figure 6** presents the analysis with respect to the default number of servers for firm 3's facilities and **Figure 7** presents the analysis with respect to the default capacity of firm 3's facilities.

Figure 8 shows the results of analyzing the effect of reneging rate of customers on the firms' market share. We change its base value ($\alpha = 5$) by −80% to +80% (in steps of 40%). This test could be applied only on two cases which deal with reneging customers.

We summarize our observations of the sensitivity analyses as the following:

• From **Figures 5** and **6**, we conclude that more market demand will be captured by firm 3 when the servers' number assigned to its facilities or the mean service rate of its servers is high. The market shares of other firms decrease except for "Balking" and "Balking and reneging" cases because in these two cases, the parameters and also the arrival rates of other firms are not changed.
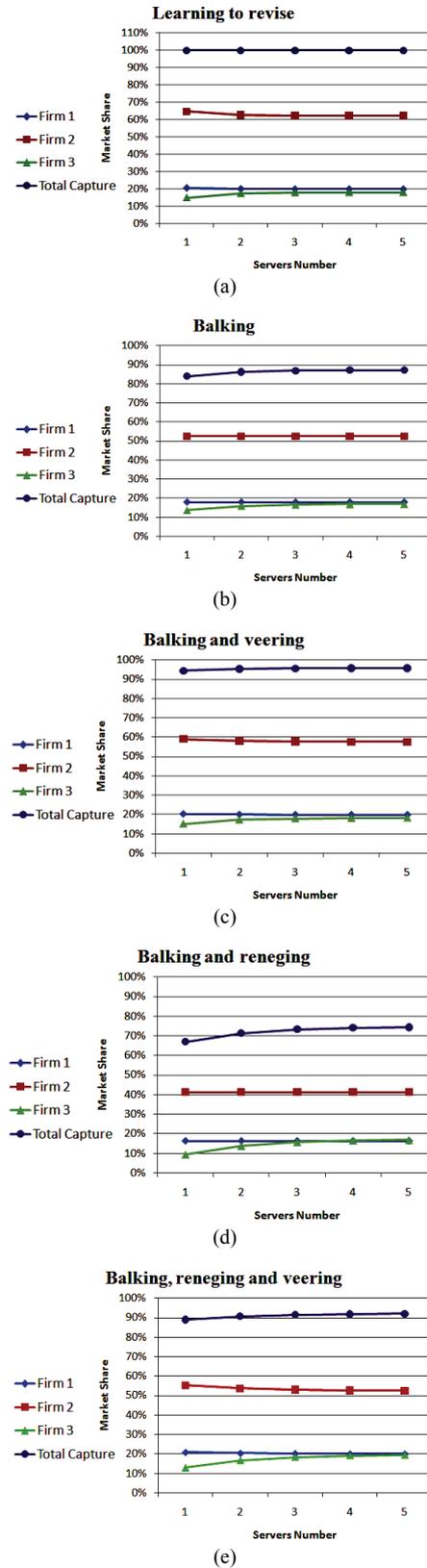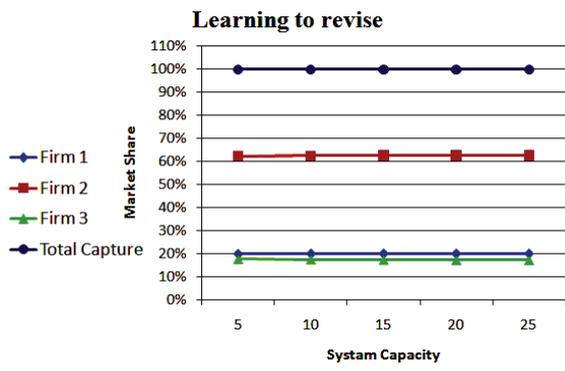

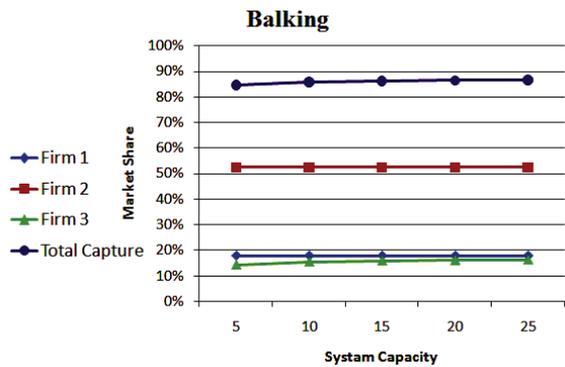**Figure 6. Sensitivity of firms' market share to the servers' number**

The result achieved from analyzing the effect of system capacity (**Figure 7**) is similar to the mean service rate and servers' number except for the case of "Learning to revise" in which a larger system capacity has a negative effect on firm 3's market share. This can be reasoned regarding the fact that a larger system capacity will cause longer waiting time.

• From **Figure 8**, we conclude that high reneging rates lower the market share of all firms. The decrement slope in the case of "Balking and reneging" is high because all reneged customers return to their homes.

• Variation of other parameters such as price, demand rates and time to cost parameter has not considerable impacts on the final results.
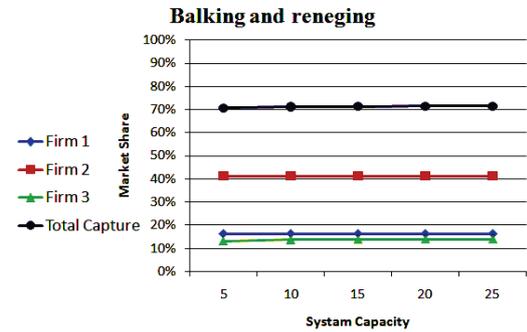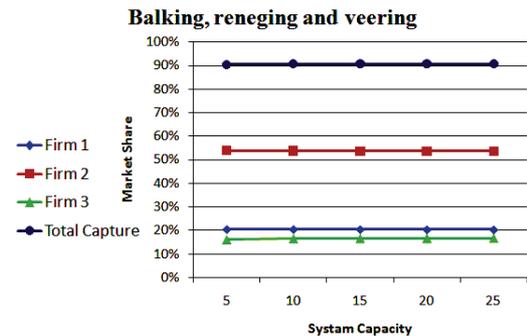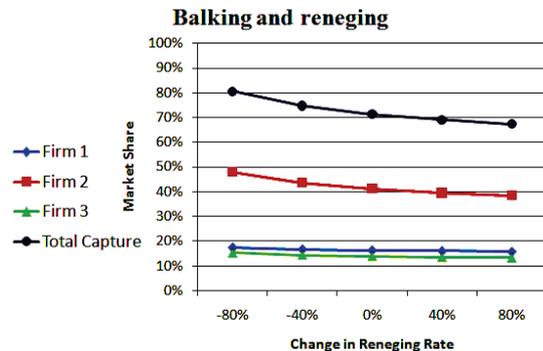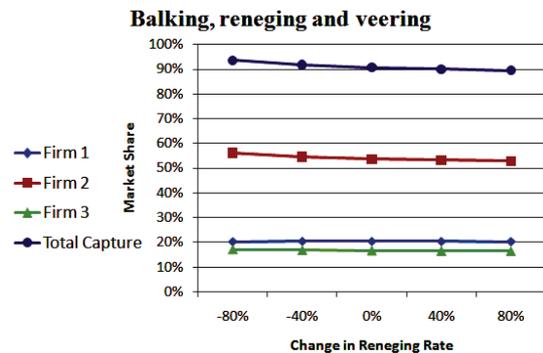

(a)

(b)
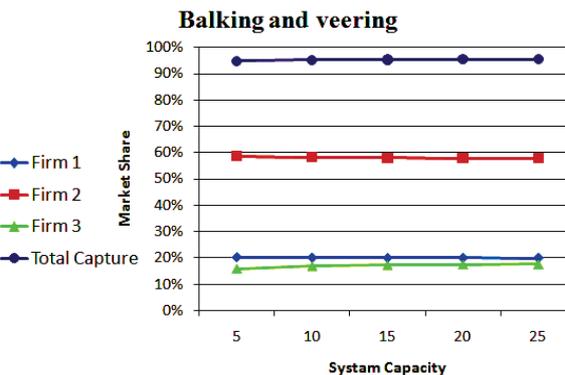
(c)

(d)

(e)

Figure 7. Sensitivity of firms' market share to the system capacity

(a)

(b)

Figure 8. Sensitivity of firms' market share to reneging rate

4. Conclusions and Future Research

In this paper we have considered customers' patronizing behavior in a competitive market. It has been concluded that the better approach for formulating customers' choice behavior in spatial competitive modeling is a probabilistic model based on three variables, distance, waiting time and price. With emphasis on congestion effects, we have also studied customers' reactions to congested facilities. These are especially balking, reneging and veering. This is the first paper considering congestion-sensitivity reactions in competitive congested systems and the first work studying veering as a usual event in congested systems. By veering we mean the case in which after a customer balked or reneged from a facility, he/she may decide to patronize another facility rather than coming back to his/her origin.

Although the prevailing approach in the literature assumes that customers take congestion into account at their origins, it has been claimed that they initially don't know a lot about facilities' congestion level. Our proposed approaches retain customers unaware until they reach at the facilities. The first approach assumes that customers amend their future decisions according to the waiting time faced by them at the previous experiences. The four other approaches assume that customers react to the congestion when they reach at the facilities. They may balk, renege, veer or divulge a combination of them.

An illustrative example has also given to demonstrate differences between the outcomes of proposed approaches. We have seen that congestion-sensitivity of customers has a considerable effect on the firms' market share. Therefore, a much attention must be paid for formulating the congestion-sensitivity of customers in spatial planning models.

We have tried to study all possible reactions to the congestion. However, a special type of queues has been considered. It will be interesting to study other types of queuing systems.

REFERENCES

[1]   T. Drezner and Z. Drezner, "Finding the Optimal Solution to the Huff Based Competitive Location Model," *Computational Management Science*, Vol. 1, No. 2, 2004, pp. 193-208.

[2]   S. L. Hakimi, "Optimum Locations of Switching Centres and the Absolute Centres and Medians of a Graph," *Operations Research*, Vol. 12, No. 3, 1964, pp. 450-459.

[3]   R. L. Church and C. ReVelle, "The Maximal Covering Location Problem," *Papers of Regional Science Association*, Vol. 32, No. 1, 1974, pp. 101-118,

[4]   H. Hotelling, "Stability in Competition," *Economic Journal*, Vol. 39, No. 153, 1929, pp. 41-57.

[5]   C. ReVelle, "The Maximum Capture or Sphere of Influence Location Problem: Hotelling Revisited on a Network," *Journal of Regional Science*, Vol. 26, No. 2, 1986, pp. 343-358.

[6]   R. Aboolian, O. Berman and D. Krass, "Competitive Facility Location Model with Concave Demand," *European Journal of Operational Research*, Vol. 181, No. 1, 2007, pp. 598-619.

[7]   M. D. Garcia Perez, P. F. Hernandez and B. P. Pelegrin, "On Price Competition in Location-Price Models with Spatially Separated Markets," *TOP*, Vol. 12, No. 2, 2004, pp. 351-374.

[8]   F. Plastria, "Static Competitive Facility Location: An Overview of Optimisation Approaches," *European Journal of Operational Research*, Vol. 129, No. 3, 2001, pp. 461-470.

[9]   D. Huff, "Defining and Estimating a Trading Area," *Journal of Marketing*, Vol. 28, No. 3, 1964, pp. 34-38.

[10] J. R. Roy and J. C. Thill, "Spatial Interaction Modeling," *Papers in Regional Science*, Vol. 83, No. 4, 2004, 339-361.

[11] A. Jain and V. Mahajan, "Evaluating the Competitive Environment in Retailing Using the Multiplicative Competitive Interactive Model," In: J. Sheth, Ed., *Research in Marketing*, JAI Press, Greenwich, Vol. 2, 1979, pp. 217-235.

[12] D. McFadden, "Conditional Logit Analysis of Qualitative Choice Behaviour," In: P. Zarembka, Ed., *Frontiers in Econometrics*, Academic Press, New York, 1974.

[13] R. Aboolian, O. Berman and D. Krass, "Competitive Facility Location and Design Problem," *European Journal of Operational Research*, Vol. 182, No. 1, 2007, pp. 40-62.

[14] H. Lee and M. Cohen, "Equilibrium Analysis of Disaggregate Facility Choice System Subject to Congestion-Elastic Demand," *Operations Research*, Vol. 33, No. 2, 1985, pp. 293-311,

[15] F. Silva and D. Serra, "Incorporating Waiting Time in Competitive Location Models," *Networks and Spatial Economics*, Vol. 7, No. 1, 2007, pp. 63-76.

[16] V. Marianov, M. Rios and M. J. Icaza, "Facility Location for Market Capture When Users Rank Facilities by Shorter Travel and Waiting Times," *European Journal of Operational Research*, Vol. 191, No. 1, 2008, pp. 32-44.

[17] A. M. Kwasnica and E. Stavrulaki, "Competitive Location and Capacity Decisions for Firms Serving Time-Sensitive Customers," *Naval Research Logistics*, Vol. 55, No. 7, 2008, pp. 704-721.

[18] R. Aboolian, Y. Sun and G. J. Koehler, "A Location–Allocation Problem for a Web Services Provider in a Competitive Market," *European Journal of Operational Research*, Vol. 194, No. 1, 2009, pp. 64-77.

[19] M. L. Brandeau, S. S. Chiu, S. Kumar and T. A. Grossman, "Location with Market Externalities," In: Z. Drezner, Ed., *Facility Location*, Springer-Verlag, 1995, pp. 121-150.

[20] V. Marianov, "Location of Multiple-Server Congestible Facilities for Maximizing Expected Demand, When Services are Non-Essential," *Annals of Operations Research*,

Vol. 123, No. 4, 2003, pp. 125-141.

[21] D. Gross and C. M. Harris, "Fundamentals of Queuing Theory," 3rd Edition, John Wiley and Sons, New York, 2002.

[22] D. Q. Yue and Y. P. Sun, "Waiting Time of M/M/C/N Queuing System with Balking, Reneging, and Multiple Synchronous Vacations of Partial Servers," *Systems Engineering-Theory & Practice*, Vol. 28, No. 2, 2008, pp. 89-97.

[23] C. T. Kelley, "Iterative Methods for Linear and Nonlinear Equations," *No.* 16 *in Frontiers in Applied Mathematics, SIAM*, Philadelphia, 1995.