PRICING, LOCATION AND CAPACITY PLANNING WITH ELASTIC DEMAND AND CONGESTION

Dmitry Krass[1], Oded Berman, and Dehui Tong
krass@rotman.utoronto.ca, berman@rotman.utoronto.ca, dehui.tong@rotman.utoronto.ca
Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, ON, Canada M5S 3E6

INTRODUCTION

We focus on several of the most important strategic decisions for service facilities facing uncertain customer demand: finding the locations of the facilities, determining the service capacity, and choosing the price to charge for service. Utility-maximizing customers are assumed to reside at the nodes of the network, generating stochastic demand streams. Customer utility is assumed to be affected by the price, travel distance and waiting times at the facility. Note that most service-improving decisions generate positive first-order effects and negative second-order effects. For example, a decrease in price leads to an increase in demand (a positive first-order effect), which in turn leads to a higher congestion at the facility, longer wait times and a resulting decrease in demand (a negative second-order effect). These negative feedback loops complicate the determination of optimal price, location and capacity. We derive explicit conditions under which first-order effects are stronger than second-order effects.

Other complicating factors include different time frames for pricing, capacity, and location decisions (the former a typically tactical and short-term while the latter are strategic and long-term). While theoretical considerations require joint optimization of all three variables simultaneously, practical implementation is greatly simplifies when the decisions can be separated. One of key questions we investigate is deriving conditions under which this separation of decisions can be done without sacrificing the optimality of the outcome. For the cases where such optimality cannot be guaranteed, we develop a heuristic approach that leads to near-optimal outcomes while still separating pricing, capacity and location decisions; the near-optimality is demonstrated through computational testing as well as asymptotic theoretical results.

Finally, an important issue from the location aspect is whether optimal locations can be localized *a priori* to a finite subset of points on a network; the most obvious such subset is the set of nodes. We derive sufficient conditions for the optimality of nodal locations under rather general settings.

While the results described above are derived for the single-facility setting, most of them extend directly to the multi-facility case as well, assuming the decision-maker controls the assignment of customers to facilities. This control is reasonable to assume in some settings, but may not be present in others – customers often self-allocate to facilities in a way that maximizes their own utilities. While the model with customer-directed allocations is much harder to analyze, the directed-assignment version likely provides a good approximation as the goals of the decision-maker and customers are well-aligned: assigning customer away from the utility-maximizing facility causes a reduction in their expenditure and a resulting loss in revenue.

The only papers we are aware of that consider the price, location and capacity decisions simultaneously are Dobson and Stavrulaki (2006) [2], and Pangburn and Stavrulaki (2008) [3]. In both cases they studied a the problem on a line segment and assumed much more restrictive forms of the demand function, which leads to significantly different analysis and results. Please refer to the full version of the paper [1] for detailed analysis and results.

MODEL AND MAIN RESULTS

Let *V = (N, L)* be a network with node set *N* and link set *L*. Let $x \in V$ represent a facility location that can be both at nodes and along the links. Node *i* generates a stream of demands assumed to behave as a time-homogenous with the rate given by :

$$\lambda_{ix} = \lambda_i^{max} F(d_{ix}) G(p) H(w), \tag{1}$$

where λ_i^{max} is the maximum potential demand rate associated with node *i*, d_{ix} is the network distance between location *x* and node *i*, *w* is the expected waiting time at the facility, and functions *F(d), G(p),* and *H(w)* are assumed to be non-negative, non-increasing, and differentiable functions with values in [0, 1] representing the sensitivity of demand to the travel distance, price and waiting time, respectively. They can also be interpreted as "decay" functions, i.e., the percentage of maximum available demand lost due to increases in travel time, price and waiting time.

The total demand rate $\lambda(x,p,\mu) = \sum_i \lambda_{ix}$ where μ is the capacity of the facility. The facility is modeled as a G/G/1 queue with the waiting time $w(\lambda,\mu)$. We use standard approximations to obtain a closed-form expression for this quantity. The decision-maker selects location *x*, price *p* and capacity μ to optimize the net revenue

$$max_{x,p,\mu}\, R(x,p,\mu) = p\lambda(x,p,\mu) - c\mu - s(x), \tag{2}$$

where *c* is the unit capacity cost and *s(x)* is the fixed cost of locating a facility at *x*. Observe that the model contains a feed-back loops: the waiting time *w* is affected by λ, which itself is affected by *w;* increase in *w* leads to a decrease in λ ("first order" effect), which, in turn, leads to a decrease in *w*, and subsequent increase in λ ("second order" effect).

Some of our key results are:

1. For any values of (p, x, μ), there exists a unique equilibrium arrival rate $\lambda(p,x,\mu)$
2. If *F(d)* is convex, there is an optimal location at one of the nodes of the network
3. The equilibrium arrival rate $\lambda(p,x,\mu)$ is non-increasing in price *p*, and non-decreasing in capacity μ. Moreover, it is decreasing (increasing) in *x* when the distance component $\sum_i \lambda_i^{max} F(d_{ix})$ is decreasing (increasing) in *x*.

Note that third result demonstrates that the "first order" effects (i.e., improved service offering through a decrease in price, or increase in service capacity, or decrease in average travel distance) dominate "second order" effects. For a given location, computing optimal values of *p* and μ requires finding all roots of a system of non-linear equations, which may be difficult. For M/M/1 with exponential decay functions, the system decomposes and the optimal price is independent of the other decisions. Moreover, the optimal price for this case appears to be near-optimal for other cases as well, leading to a high-quality heuristic solution.

REFERENCES

[1] Berman, O., Krass, D., and D. Tong, Pricing, Location and Capacity Planning with Elastic Demand and Congestion, *working paper*, University of Toronto.
[2] Dobson, G., and E. Stavrulaki. Simultaneous price, location, and capacity decisions on a line of time-sensitive customers. *Naval Research Logistics*, v.54(1), 2007
[3] Pangburn M.S., and E. Stavrulaki. Capacity and price setting for dispersed time-sensitive customer

segments. *Europ. J. of Oper. Res.*, 184:1100, v1121, 2008.