

2005

Capacity Decisions at eCommerce Sites: A Competitive Analysis

Dengpan Liu

University of Texas at Dallas, dengpan.liu@student.utdallas.edu

Subodha Kumar

University of Washington, subodha@u.washington.edu

Vijay Mookerjee

University of Texas at Dallas, vijaym@utdallas.edu

Follow this and additional works at: <http://aisel.aisnet.org/amcis2005>

Recommended Citation

Liu, Dengpan; Kumar, Subodha; and Mookerjee, Vijay, "Capacity Decisions at eCommerce Sites: A Competitive Analysis" (2005).
AMCIS 2005 Proceedings. 376.

<http://aisel.aisnet.org/amcis2005/376>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISEL). It has been accepted for inclusion in AMCIS 2005 Proceedings by an authorized administrator of AIS Electronic Library (AISEL). For more information, please contact elibrary@aisnet.org.

Capacity Decisions at E-commerce Sites: A Competitive Analysis

Dengpan Liu

School of Management
University of Texas at Dallas
dengpan.liu@student.utdallas.edu

Subodha Kumar

School of Business, University of Washington
subodha@u.washington.edu

Vijay Mookerjee

School of Management
University of Texas at Dallas
vijaym@utdallas.edu

ABSTRACT

Capacity planning is a critical issue for most e-commerce web sites. We examine capacity decisions made in a duopolistic setting where customers may renege at one firm's web site to visit the competitor's web site. Independent capacity decisions made at each firm are analyzed and compared to those made by a central planner as well as to those made by a monopolistic firm that caters to the entire market demand. We find that competing firms under-invest in capacity (as compared to the central planner) when customers are relatively tolerant of web site delays, but over-invest when customer impatience level increases beyond a point. Also, when compared to the monopoly, the total profit earned by the duopoly can be larger or smaller depending on the customer tolerance level for delay. Several analytical and numerical results are presented.

Keywords

E-Commerce, Capacity planning, Optimization, Collaboration, Reneging.

INTRODUCTION

Although the U.S. online buying population continues to grow – reaching 79.3 million in 2001 (eMarketer, 2001) and expected to increase by 50% by 2008 (Evans *et al.*, 2004) – many online transactions are aborted before completion (Dennis, 2001). A report indicates that 69.4% of all potential online transactions in 2001 were abandoned (Datamonitor, 2001) and one of the biggest culprits was the poor experience in terms of processing speed. A new Forrester study shows that web site visitors leave if a page does not load within four seconds (Trotta, 2003), whereas the average response time for many retail web sites is 17 seconds (Dennis, 2001). Zona Research estimates the value of lost potential business due to dropped transactions to be more than \$25 billion a year and 82% of it is blamed on slow-loading web pages (Dennis, 2001).

Because poor response at a firm's web site clearly has serious business implications, many firms have increased their server capacity in past few years. A survey of 101 U.K. based firms reveals that 53% firms expect to increase their server capacity over the next year but 25% of the surveyed firms consider that up to half their present capacity is there just in case it is needed (Hollingsworth and Gooch, 2004). Average utilization rates for server capacity are running at a mere 10% and are as low as 4% for some firms (Vowler, 2003; Wright, 2004). Hence, better capacity planning is required for most of the e-commerce firms.

We consider the problem of choosing the optimal capacity at a firm's e-commerce site in order to maximize profit. Increasing capacity reduces the loss of customers but also increases the cost. Thus there is a trade-off between the cost of losing customers and the cost of increasing capacity. We first analyze the capacity optimization problem for a monopoly case and extend the analysis to a duopoly case. For the duopoly, we first consider a decentralized case where the two firms choose optimal capacity levels independently, i.e., with the objective of individual profit maximization. We then consider a central planner who chooses the optimal capacity level for both firms so that the total profit for the two firms can be maximized. Different numerical and analytical results are provided for the decentralized case and the central planner case. Finally, we compare the monopoly and duopoly results.

The remainder of this paper is organized as follows. In section 2, we describe the model. The monopoly case is described in Section 3. Section 4 presents the decentralized case and the central planner case for the duopoly and discusses some analytical results. Some numerical results are provided in Section 5. Section 6 summarizes and concludes the paper.

ASSUMPTIONS AND NOTATIONS

Table 1 provides a detailed list of model parameters and decision variables. We next summarize the assumptions of the study.

SYMBOL	DEFINITION	REMARKS
h	Average value per session	<i>Measured in dollars</i>
γ_0	Fixed capacity cost	<i>Measured in dollars per second</i>
γ_1	Marginal capacity cost	<i>Measured in dollars</i>
v_1	Level of impatience of the new arrival (<i>It is v for monopoly</i>)	<i>Takes the unit of number per second</i>
v_2	Level of impatience of the deflected customers	<i>Takes the unit of number per second</i>
λ	Arrival rate of customers	<i>Takes the unit of number per second</i>
ρ	Fraction of first-time renege customers who leaves the system without going to other firm	
μ_i	Server capacity (i.e., processing rate) for firm i , $i = 1, 2$. (<i>Subscript is dropped for monopoly case</i>)	Decision Variable <i>Takes the unit of number per second</i>
L_{ij}	Expected rate of customers renege from firm i to firm j , $i = 1, 2$; $j = 1, 2$; $i \neq j$. (<i>Subscript is dropped for monopoly case</i>)	Derived Variable <i>Takes the unit of number per second</i>
R_{i1}	Expected rate of new customers renege from firm i and leaving the system without going to other firm, $i = 1, 2$	Derived Variable <i>Takes the unit of number per second</i>
R_{i2}	Expected rate of customers renege from firm i out of those who renege and came from other firm, $i = 1, 2$	Derived Variable <i>Takes the unit of number per second</i>

Table 1. Model Parameters and Decision Variables

- (i) Customers arrive according to a Poisson distribution with constant mean λ . This assumption is consistent with the real-world (Moe and Fader, 2002).
- (ii) The server operates in a processor sharing mode. Each session requires a processing time that follows a Generic distribution with mean $1/\mu_i$ ($i = 1, 2$ for duopoly case and there is no subscript for monopoly). Since the processing requirements can vary widely across sessions, allowing the processing time distribution for a session to be generic makes the analysis robust.
- (iii) Customers are impatient during a session and can renege randomly anytime during the session after spending an amount of time that is assumed to be uniformly distributed between $[0, 1/v_i]$ ($i = 1, 2$ for duopoly case and there is no subscript for monopoly case). For the duopoly case, v_1 represents the level of impatience for customers who have not renege from any of the two firms and v_2 represents the level of impatience for deflected customers (those who have already renege from a firm). In practice, these two patience levels need not be the same; hence we consider two patience levels to make the analysis general.
- (iv) The customers' expected waiting time is larger than the expected response time, i.e.,

$$v^{-1} > (\mu - \lambda)^{-1}.$$
- (v) A customer visits a firm at most once. After renege from a firm, customers either leave the system (R_{i1}) or go to other firm (L_{ij}). Thus customers never return to a firm they renege from. This assumption closely represents the real-world as observed by Trotta (2003). If customers renege a second time, they leave the system (R_{i2}).
- (vi) The cost of capacity is linear ($\gamma_0 + \gamma_1\mu$) with a fixed and variable component. Although we can assume other functional form, we just use this linear form to simplify the model.

MONOPOLY CASE

The monopolist chooses an optimal level of capacity to maximize profit. Since the maximum number of allowable sessions is typically large ($K \rightarrow \infty$), the loss of customers per unit time can be written as:

$$L(\lambda, \mu) = v[\lambda / (\mu - \lambda)] \tag{1}$$

The total revenue per unit time (S) is given by: $S \equiv h(\lambda - L)$ and hence the profit function can be written as:

$$\pi = h\left(\lambda - \frac{v\lambda}{\mu - \lambda}\right) - (\gamma_0 + \gamma_1\mu) \tag{2}$$

Taking first-order and second-order conditions of π with respect to μ , we get the optimal capacity level $\mu^* = \lambda + \sqrt{hv\lambda/\gamma_1}$, and the optimal profit $\pi^* = (h - \gamma_1)\lambda - 2\sqrt{hv\lambda\gamma_1} - \gamma_0$. The optimal server capacity increases as the customer becomes more impatient or the ratio h/γ_1 increases.

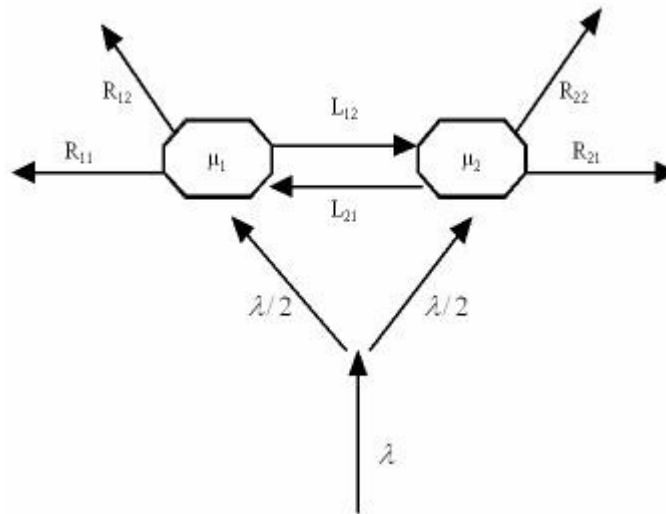


Figure 1. Duopoly Case

DUOPOLY CASE

The duopoly case is depicted in Figure 1 using the notations described in Table 1. Customers randomly choose a firm and the arrival rate for each firm is $\lambda/2$. The loss functions for firm 1 can be written as: $L_{12} = \frac{(\lambda/2)(1-\rho)v_1}{\mu_1 - (\lambda/2 + L_{21})}$;

$$R_{11} = L_{12} \frac{\rho}{1-\rho}; R_{12} = \frac{L_{21}v_2}{(\lambda/2)(1-\rho)v_1} L_{12}.$$

We relate the capacity for firm 1 to the “cross traffic” terms (L_{12} and L_{21}) as:

$$\mu_1 = \frac{(\lambda/2)(1-\rho)v_1}{L_{12}} + \lambda/2 + L_{21} \tag{3}$$

Expressions for the loss functions and the capacity for firm 2 can be similarly obtained. We will use L_{12} and L_{21} as surrogate decision variables to simplify the presentation.

Decentralized Case

In the decentralized case, firms choose their optimal capacity levels with the objective of individual profit maximization. It is easy to write the profit of firm 1 (given L_{21}) and the profit of firm 2 (given L_{12}) as:

$$\pi_1(L_{12}) = h[\lambda/2 + L_{21} - (R_{11} + R_{12} + L_{12})] - (\gamma_0 + \gamma_1\mu_1) \quad (4)$$

$$\pi_2(L_{21}) = h[\lambda/2 + L_{12} - (R_{21} + R_{22} + L_{21})] - (\gamma_0 + \gamma_1\mu_2) \quad (5)$$

The first order conditions (FOCs) obtained from (4) and (5) are symmetric, and we can prove that these two FOCs equations have only one symmetric solution. We therefore set $L_{12} = L_{21} = L$ and obtain:

$$-h \left[\frac{1}{1-\rho} + \frac{Lv_2}{(\lambda/2)(1-\rho)v_1} \right] + \frac{\gamma_1(\lambda/2)(1-\rho)v_1}{L^2} = 0 \quad (6)$$

From symmetry, (3) can be rewritten as:

$$\mu^* = \frac{(\lambda/2)(1-\rho)v_1}{L^*} + \lambda/2 + L^* \quad (7)$$

In (7) L^* is the solution to (6). The expression for L^* is complex and is not presented here. Using (7) we can get the Nash Equilibrium capacity pair (μ_D^*, μ_D^*) , where the subscript D indicates that this is a solution for the decentralized case.

Central Planner Case

Here we choose the optimal capacity level for both the firms with the objective of maximizing the total profit. The total profit function can be written as: $\Pi(L_{12}, L_{21}) = \pi_1(L_{12}) + \pi_2(L_{21})$. We can prove that the global optimum is symmetric

(i.e., $\mu_1^* = \mu_2^* = \mu_C^*$), and $\mu_C^* = \frac{(\lambda/2)(1-\rho)v_1}{L^*} + \lambda/2 + L^*$ where L^* is the solution of:

$$\frac{\partial \Pi}{\partial L} = \left(\frac{-2h\rho}{1-\rho} - 2\gamma_1 \right) - \frac{8hv_2L}{\lambda(1-\rho)v_1} + \frac{\gamma_1\lambda(1-\rho)v_1}{L^2} = 0 \quad (8)$$

Analytical Results

Proposition 1: *In the decentralized and centralized cases, the optimal capacity has the following properties: $\frac{\partial \mu_i^*}{\partial X} > 0$,*

where $i = D, C$; $X = v_2$.

As “deflected” visitors become more impatient (v_2 increases), the capacity increases in both cases. In the decentralized case, each firm has an incentive to increase its capacity as deflected visitors become more impatient. In the central planner case, in addition to the need to accommodate more impatient deflected visitors, there is another for the firms to increase capacity that can be explained as follows. If $v_2 > v_1$, the central planner has an incentive to increase capacity to suppress this “conversion” of more to less patient customers. Thus as v_2 increases, μ_C^* increases faster than μ_D^* because there is an additional reason for the central planner to increase capacity. When $v_2 = 0$, i.e., the renege customers become perfectly patient, the optimal capacity is at its lowest value. However, μ_{\min}^{*D} is greater than μ_{\min}^{*C} because the central planner benefits from the ability to get a second chance to serve customers.

In the following proposition we contrast the optimal capacity levels in the decentralized and centralized cases.

Proposition 2: *There exists a critical value for v_2 , denoted as v_{2C} . The optimal capacity $\mu_C^* > \mu_D^*$, if $v_2 > v_{2C}$ and $\mu_C^* \leq \mu_D^*$, if $v_2 \leq v_{2C}$, where $v_{2C} = \frac{(h - \gamma_1)\lambda v_1}{2h} \sqrt{\frac{2((h - \gamma_1)(1 - \rho) + h)}{\gamma_1 \lambda v_1}}$.*

There are two opposing effects that underlie Proposition 2. First, because the central planner can influence the conversion of customers from being more patient to less patient, the capacity level chosen by the central planner is higher than the level chosen in the decentralized case. Second, because the central planner gets a second chance to serve a renegeing customer, the optimal capacity is lower. Below v_{2C} , a critical value associated with the patience level of renegeing customers, the capacity in the central planner case is lower than the decentralized case. In this region, the inherent efficiency of the central planner dominates. However, when v_2 exceeds the critical value, the customer conversion effect starts to dominate and therefore, the capacity in the central planner case becomes greater than that of the decentralized case.

NUMERICAL ANALYSIS

In this section we numerically explore the phenomenon discussed in the previous section. In this simulation work, to illustrate Proposition 2, we set $v_1 = 0.1, \lambda = 1, \gamma_0 = 1, \gamma_1 = 0.5, h = 10$ and $\rho = 0.5$. In Figure 2, we compare the optimal capacity as a function of v_2 . We observe the property stated in Proposition 2: there exists a critical value for v_2 , below which, the optimal capacity in decentralized case is higher and above which, the optimal capacity in central planner case is higher.

We also compare the total profit in the duopoly and monopoly cases as the impatience level of renegeing customers increases. In the monopoly case, (which has a demand rate λ), the optimal profit has a simple expression: $\pi^* = (h - \gamma_1)\lambda - 2\sqrt{h\nu\lambda\gamma_1} - \gamma_0$. However, in the duopoly cases, the expressions for the optimal profit are very complex, and an analytical comparison is difficult. Hence, we numerically compare the two cases as illustrated in Figure 3. For this comparison, we set $v_1 = 0.1, \lambda = 1, \gamma_1 = 0.5, \gamma_0 = 0, h = 10$, and $\rho = 0.2$.

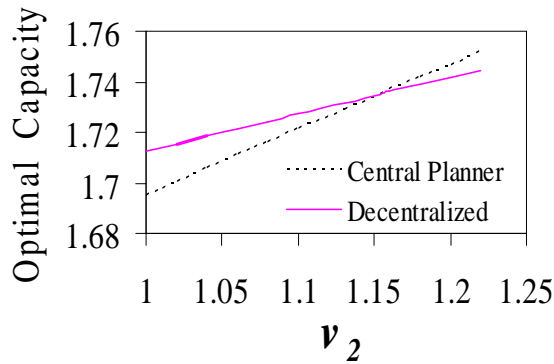


Figure 2. Capacity comparison

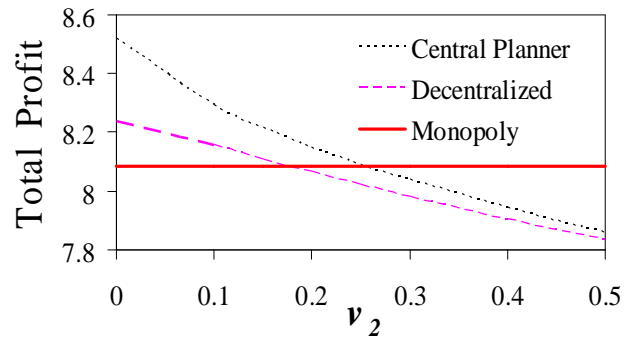


Figure 3. Profit comparison with zero fixed capacity

It is interesting to observe that the monopolist’s profit can be higher than the profit of centrally planned duopoly even when the fixed cost of capacity is set to zero. When renegeing customers are very impatient the main advantage of the duopoly is lost: in a duopoly there is a second chance to capture the sale from a renegeing customer. However, the monopolist gains operational efficiencies from consolidating the capacity into a single server.

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

In this paper, we analyze the capacity decisions of firms in monopoly and duopoly cases. In duopoly case, we consider both the decentralized and centralized optimization. We find that the optimal capacity level in the centralized case might be greater or smaller to the optimal level in the decentralized case. While a monopolist loses from not being able to convert cross traffic visitors to a sale, the consolidation of capacity provides operational benefits.

We examined the capacity planning problem in the context of one or two firms making capacity decisions. With some modification our analysis may also be applicable when a choice needs to be made between using a high capacity site versus two lower capacity mirror sites. Having a single high capacity site may be advantageous when reneging customers are extremely impatient. However, as reneging customers become more patient, it may be better to distribute the capacity across two or more mirror sites so that these customers can be served better. In this work, only simulated data is used to illustrate some propositions, and it would be better if real data could be utilized to verify these results, however, unavailability of real data is the difficulty we are facing. We are also exploring the best way to estimate those parameters in the model.

REFERENCES

1. Datamonitor (January 1, 2001) Online Customer Service.
2. Dennis, S. (May 8, 2001) Almost 50 Percent of Online Purchases Aborted, www.ComputerUser.com.
3. eMarketer (September,2001) eCommerce: B2C Report.
4. Evans, P. F., Berk, M., Schatskv, D. and Sehgal, V. (January 23, 2004) Market Forecast: U. S. Retail, 2004-2008, *Jupiter Research*.
5. Hollingsworth, T. and Gooch, A. (March 25, 2004) New survey shows UK firms are set to increase server capacity, *M2 Presswire*, www.m2.com.
6. Moe, W. W. and Fader, P. S. (May, 2002) Modeling Online Store Visit Patterns as a Measure of Customer Satisfaction, *Wharton Marketing Department Working Paper*.
7. Trotta, G.(July 7,2003) WebSphere Tricks, and Customers Who Stick, *ebizQ*, www.ebizq.net.
8. Vowler, J.(July 29,2003) Pulling It All Together, *Computer Weekly*, pp. 22.
9. Wright, J. (July 20, 2004) Buy Servers Only If You Really Need Them, *Computer Weekly*, pp. 24.