Setting Cutoff Times for Picking Systems with Capacity Degradation

Kevin R. Gue
Department of Industrial & Systems Engineering
Auburn University, Auburn, AL, USA

Abstract

In the new landscape of e-commerce distribution, firms must offer increasingly aggressive delivery promises and then make good on them. These promises often take the form of a cutoff time, such that orders placed before the cutoff time receive premium service (next-day, same-day, etc.) and those placed afterward do not. Later cutoff times are stronger, of course, but the fulfillment system might not be able to process the order before the deadline. How late is too late? We develop a deterministic model to answer this and related questions when the order fulfillment system batches orders for efficient picking operations and therefore exhibits a phenomenon we call capacity degradation.

1 A New Objective

The retail industry is in the midst of a sea change in distribution models. What was once a simple choice of in-store or catalog delivery (to arrive "in a few days") has given way to multiple channels of distribution, including time-definite, guaranteed delivery of the form "Order by this time, and we'll deliver it by tomorrow." Guaranteed two-day and next-day delivery are now staples of the world's largest retailers and distributors, and same-day delivery is already being offered in limited markets.

The most common type of service promise in order fulfillment has the form "Order by 2:00pm and receive your shipment the next day." Orders arriving before 2:00pm must be processed by the order fulfillment center before the last truck departure (the deadline) or else the customer does not receive the promised service and the retailer suffers a penalty such as refunded shipping charge or loss of goodwill.

Naturally, the retailer would like to offer the latest possible cutoff time in order to garner the maximum possible revenue from the premium service. But offering a cutoff time too close to the deadline (last truck departure) risks orders missing the truck. How to compute the best time? Could the firm offer a later cutoff time, say, by hiring more order pickers? What exactly is the relationship between throughput capacity and optimal cutoff time?

In this paper, we develop a theoretical model of order fulfillment systems that offer a cutoff time guarantee and that feature a manual order picking area in which workers must traverse aisles to pick order lines in a batch of orders. The model illustrates fundamental relationships among the deadline, the cutoff time, throughput capacity, and the order arrival rate. It also introduces a novel concept called *capacity degradation* that can be used to model reduced efficiency of workers as the deadline approaches.

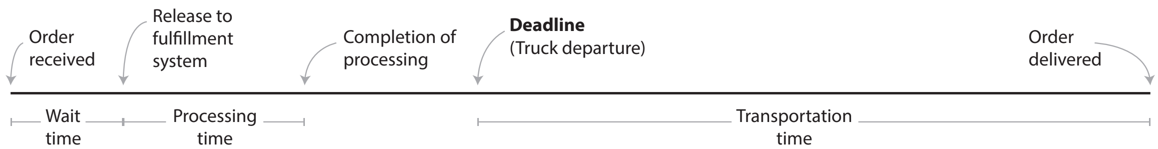


Figure 1: Timeline of a typical order in a fulfillment system. Only a small portion of the total order-to-receipt time seen by the customer happens in the fulfillment center. However, missing a deadline (truck departure) by just 5 minutes typically delays final delivery by an entire day.

2 Literature Review

Manufacturing and service systems that make due date or delivery promises have been the subject of study for many years [12, 13]. In general, the literature addresses both *customer-specific* and *uniform* delivery promises. A customer-specific promise is made dynamically upon customer order, and may be allowed to vary according to operational or other conditions. Nearly all existing due-date quotation literature is of this type. Research in due date quotation has considered both infinite and finite capacity production and service systems [8]. Other authors have explored uniform delivery promises, which are consistent across customers irrespective of the time of the order or the state of the order fulfillment system [16, 22, 23].

Common among these papers is the assumption that delivery lead time begins at the time of order and ends after "processing," where processing time is modeled as fixed or (more often) as a random variable. In e-commerce and other order fulfillment environments, however, it is not the time between the order arrival and completion of processing that matters but rather whether or not the order made it onto the target truck by the time of last truck departure (Figure 1) [7].

All existing literature of which we are aware assumes a stationary distribution of service capacity. As we describe below, effective capacity in many order fulfillment environments (manual order picking in particular) degrades as time approaches the deadline. We are not aware of any due date quotation literature that addresses such changes in service capacity.

Several authors have recently addressed the relationship between capacity and promised delivery time. Ho and Zheng [10] consider service delivery promises and their effect on capturing market share in competitive environments. Shang and Liu [21] investigate the relationship between promised delivery times, quality of service, and capacity in a oligopoly game.

Warehousing and order picking systems have an enormous literature [6]. With only a few exceptions, the research community has focused on methods and models of warehouse and order picking operations with objectives such as maximizing throughput capacity, minimizing cycle time, or minimizing cost. As argued in Doerr and Gue [7], none of these objectives is appropriate when the order fulfillment system operates against a daily deadline. Nevertheless, operations involving a deadline still must pick, pack, and ship orders, and therefore papers showing how to estimate the length of a picking tour are important for our purposes. In their seminal paper, Ratliff and Rosenthal [17] developed a dynamic programming solution to develop optimal solutions for a one block warehouse. This work was extended by Roodbergen and de Koster [18] to warehouses with multiple blocks, and recently to the fishbone warehouse design by Çelik and Süral [2].

Our interest, however, is not in optimal solutions to particular picking tours but in expected distance for expected orders, a subject which has also received significant treatment in the literature. Many papers have addressed warehouses with a single block of picking aisles [5, 9, 11, 14, 19]. In general, these papers build upon one another and are increasingly more accurate at estimating the length of a tour using a particular routing heuristic (S-shape or Largest Gap, for example). In papers addressing product assignment, Caron et al. [1] and Le-Duc and de Koster [15] both develop travel distance estimation models for two block warehouses in which the middle aisle is vertical. Roodbergen et al. [20] develop a very general and accurate estimate for the travel distance of S-shape tours in a warehouse having multiple cross aisles. They offer evidence that their model is superior to those of Kunder and Gudehus [14] and Hall [9].

Three papers have been written on order fulfillment environments with a daily deadline. Doerr and Gue [7] introduced a metric called "Next Scheduled Deadline," which measures the fraction of orders arriving between consecutive cutoff times that make it onto the next departuring truck. They use goal setting theory to show that establishing the cutoff time with regard to worker motivation leads to improved system performance compared to doing so without considering motivation. Çeven and Gue [3] show how to set the cutoff time in an order fulfillment system with daily deadlines for a single class of orders when orders are released to the system in waves. Çeven and Gue [4] extend this line of research to the case of multiple order classes. The operation we model below assumes a continuous flow of work rather than wave releases.

3 Model

In addition to estimating the travel distance of tours with certain heuristics, Hall [9] developed the following lower bound for the expected length of a tour with N stops:

$$E_N[D^*] = \sqrt{A/r}[2(N-1)/(N+1)] + M\sqrt{A}\sqrt{r}\sum_{i=0}^{N}\binom{N}{i}\left[\frac{1}{M}\right]^i\left[\frac{M-1}{M}\right]^{N-i}[1-0.5^i].$$

A plot of this expression as a function of N suggests approximately a "square root increase" in the number of stops for sufficiently low values of N (see the "Lower Bound" in Figure 2). Assuming the nominal values $M = 25$, $A = 1$ and $r = 0.5$ used by Hall, we suppose an approximation function $L(N) = c\sqrt{N}$. To find an appropriate coefficient

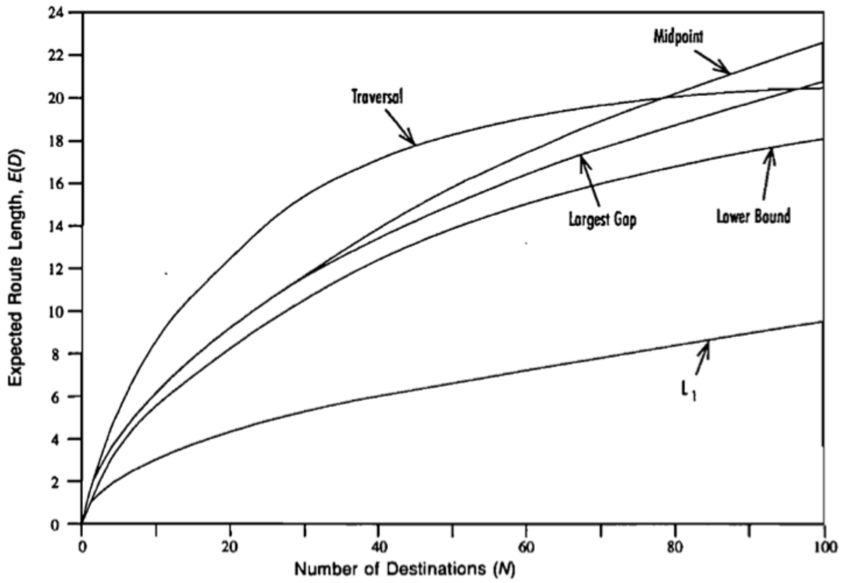


Figure 2: [from 9] Expected distance of a picking tour versus the number of stops in the tour in a warehouse of 25 aisles for common routing heuristics.

c, we use the `NMinimize` function in MATHEMATICA to minimize the sum of absolute differences between the lower bound and the approximation:

$$\sum_{N=1}^{100} |E_N[D^*] - c\sqrt{N}|.$$

The optimal $c \approx 1.9177$ (see Figure 3). Because the approximation is less than the lower bound for some values of N and because the lower bound is fairly tight (see curve for the Largest Gap heuristic in Figure 2), we could use $c = 2$ to approximate the length of a tour $L(N) = 2\sqrt{N}$. The correct value of c depends on a number of instance-specific parameters such as area of the order picking area and how close one's routing method is to the lower bound, so we retain the coefficient simply as c in expressions that follow.

It is important to note that the square root approximation breaks down for large values of N: the function $L(N) = c\sqrt{N}$ continues to increase, whereas the actual length of a tour converges to a fixed value equal to the number of aisles times the depth of the warehouse (the length of an aisle). That is, for very high values of N, the optimal picking tour simply traverses every aisle. Other functional approximations could be used in the models we present below, so long as they are integrable (if only numerically).

Given the approximation of the length of a tour, we can develop a throughput capacity expression for an individual worker. If the worker's velocity is v, then the time T to pick N items is $T(N) = c\sqrt{N}/v$ and the worker is able to pick N items in that time. That is,

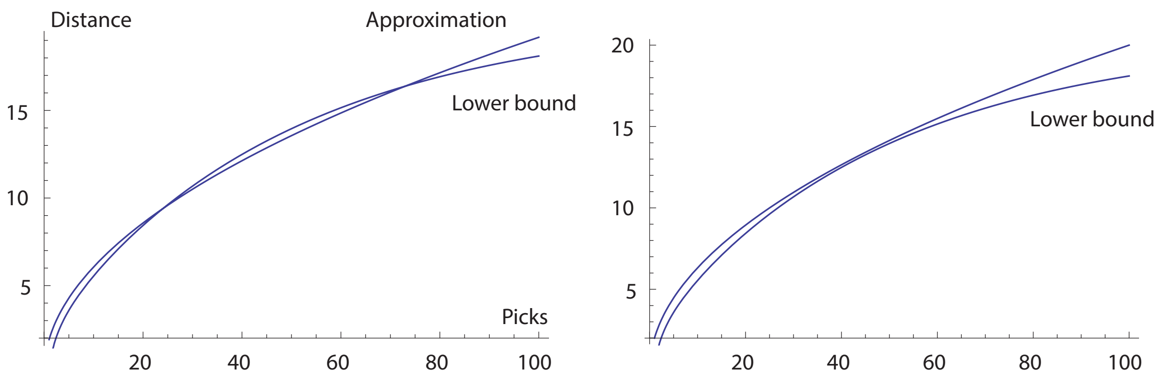


Figure 3: Approximations of the lower bound developed by Hall [9]. On the left, the approximation uses $c = 1.9177$; on the right, $c = 2$, which places the entire curve slightly above the lower bound, as desired.

the worker's throughput capacity

$$\mu = \frac{\text{picks per tour}}{\text{time per tour}} = \frac{N}{c\sqrt{N}/v} = \frac{v}{c}\sqrt{N}. \tag{1}$$

In most order fulfillment applications with manual picking, workers make picking tours with a picking cart having storage capacity N_{max} (in a book warehouse, $N_{max} \approx 100$). This implies that the worker's maximum throughput capacity $\mu_{max} = (v/c)\sqrt{N_{max}}$ and the maximum throughput capacity of the picking system would be approximately this value times the number of workers.

In the presence of a daily deadline, there is a significant problem: picking tours within a certain window of the deadline might not finish before the deadline, and orders in those tours would miss the delivery promise. It is typical when the deadline is near to assign workers some $N < N_{max}$ items to pick, such that the tour will finish before the deadline. Because workers are picking fewer items per tour, their effective throughput capacity goes down. For manual order picking systems, throughput capacity is governed by Equation 1 above.

Suppose a day of unit length and assume the deadline is at time 1. Further, assume there is only one worker (what follows is easily generalized to multiple workers). Let $t_r = 1 - t$ be the remaining time available to pick a batch of orders. Assuming we want the picker to gather as many items as possible, $t_r = c\sqrt{N}/v$, and the maximum number of items is

$$N = \left(\frac{vt_r}{c}\right)^2 = \left(\frac{v}{c}\right)^2 (1-t)^2$$

We now have a complete expression for the throughput capacity of the worker as a function of time:

$$\mu(t) = \min\left\{\frac{v}{c}\sqrt{N_{max}},\ \left(\frac{v}{c}\right)^2 (1-t)^2\right\}.$$

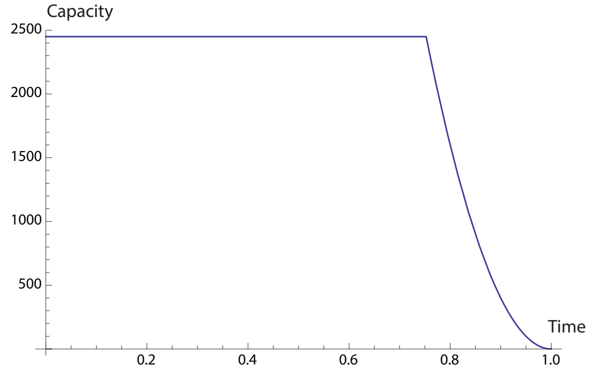


Figure 4: Effective capacity for a worker under deadline driven operations. As the deadline approaches ($t = 1$), workers must make tours with fewer and fewer stops in order to finish by the deadline, thus reducing effective capacity. Just before the deadline, there is insufficient time to retrieve even one item, so effective capacity is zero.

Figure 4 shows $\mu(t)$ for an entire day. Well before the deadline, workers can make the most efficient tours possible (thus keeping costs low), but as the deadline gets close their effective capacity decreases until, just before the deadline, they are unable to fill any orders at all in time.

The time at which capacity degradation must begin (when workers are given fewer picks) is when

$$\frac{v}{c}\sqrt{N_{max}} = \left(\frac{v}{c}\right)^2 (1-t)^2,$$

which has solution

$$t = 1 - \sqrt{\frac{c\sqrt{N_{max}}}{v}}.$$

Figure 5 illustrates the time period $[0.7, 1.0]$ in a system whose capacity begins to degrade just after $t = 0.75$. The shaded regions under the curve represent the number of backlogged orders (magenta) and the remaining capacity (blue). At the time that these two regions have equal area, the remaining capacity exactly matches the backlog and newly arriving orders would miss the deadline. This is the latest possible cutoff time.

The optimal cutoff time is the time at which backlogged work equals remaining capacity. That is, the optimal cutoff time t^* satisfies

$$\int_{t=0}^{t^*} \max\{0, \lambda(t) - \mu(t)\}\, \mathrm{d}t = \int_{t=t^*}^{1} \mu(t)\, \mathrm{d}t.$$

The difficulty in practice, of course, is that the arrival rate function $\lambda(t)$ is not precisely known. Nevertheless, historical data could suggest such a function, or at least a rough approximation.

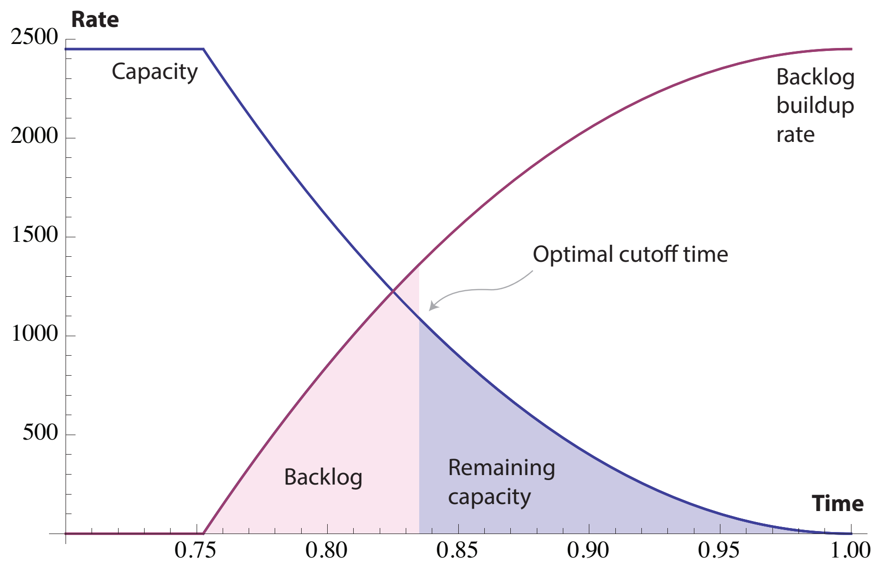


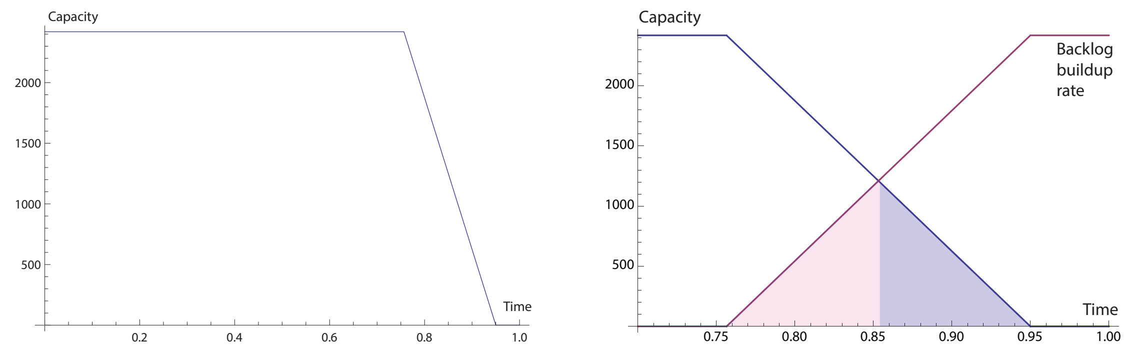
Figure 5: How to determine the optimal cutoff time. The light blue region indicates the number of items that can be picked before the deadline; the light magenta region reflects the number of items backlogged. When these two regions have equal area, all backlogged items can be picked on time, but not more.

An approaching deadline necessarily creates capacity degradation in other order picking systems as well. In a flow rack order picking area, for example, orders are assembled by moving a tote or carton along an aisle of order picking locations. Each picking "tour" is comprised of time to travel along a fixed length of rack and time to make picks. Suppose the time to build the order is $T(n) = a + bn$, where a is the fixed time to walk the picking line, b is the time per pick, and n is the number of picks in the tour. Figure 6 shows the corresponding capacity degradation curve and curves to determine the optimal cutoff time in this case. The capacity function is different, but the approach is the same as before.

It is also interesting to consider a pure flow order picking system in which orders are not batched for picking. Capacity degradation for such a system would be absolute and immediate, as depicted in Figure 7.

4 Conclusions

The models in this paper are not an easy solution to existing problems in practice, but are intended rather to provide important insight and a basis for more complex and realistic models—in the same way that the EOQ formula is useful in inventory modeling. We have ignored many important aspects of "the real problem," including randomness, nonstationary arrivals, and order fulfillment processes beyond picking such as packing and shipping. Future research will incorporate these complications.



(a) Capacity degradation.

(b) Optimal cutoff time.

Figure 6: Capacity degradation for a flow rack or other order picking system in which tours contain a fixed time component.

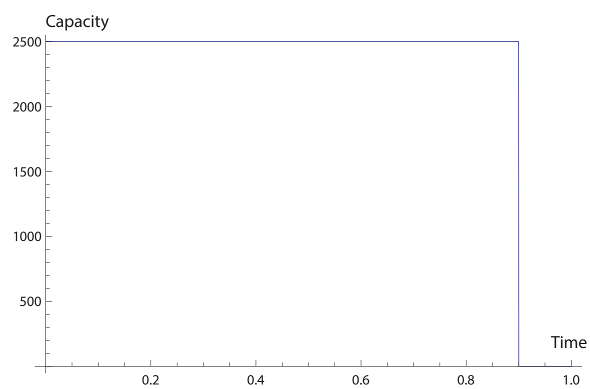


Figure 7: Capacity degradation for an order picking system without batching.

References

[1] F. Caron, G. Marchet, and A. Perego. Routing policies and COI-based storage policies in picker-to-part systems. *International Journal of Production Research*, 36(3):713–732, 1998.

[2] Melih Çelik and Haldun Süral. Order picking under random and turnover-based storage policies in fishbone aisle warehouses. *IIE Transactions*, 43(3):283–300, 2014.

[3] Erdem Çeven and Kevin R. Gue. Wave Release Stategies for Order Fulfillment Systems with Deadlines. Submitted for publication, 2013.

[4] Erdem Çeven and Kevin R. Gue. Wave Release Policies for Order Fulfillment Systems with Multiple Customer Classes. Submitted for publication, 2014.

[5] Ek Peng Chew and Loon Ching Tang. Travel time analysis for general item location assignment in a rectangular warehouse. *European Journal of Operational Research*, 112:582–597, 1999.

[6] R. de Koster, T. Le-Duc, and K. J. Roodbergen. Design and Control of Warehouse Order Picking: A Literature Review. *European Journal of Operational Research*, 182:481–501, 2007.

[7] Kenneth H. Doerr and Kevin R. Gue. A Performance Metric and Goal-Setting Procedure for Order Fulfillment Operations. *Production and Operations Management*, 22 (3):691–710, 2013.

[8] Izak Duenyas and Wallace J. Hopp. Quoting Customer Lead Times. *Management Science*, 41(1):43–57, 1995.

[9] Randolph W. Hall. Distance Approximations for Routing Manual Pickers in a Warehouse. *IIE Transactions*, 25(4):76–87, 1993.

[10] Teck Ho and Yu-Sheng Zheng. Setting Customer Expectation in Service Delivery: An Integrated Marketing-Operations Perspective. *Management Science*, 50(4):479–488, 2004.

[11] Jay M. Jarvis and Edward D. McDowell. Optimal Product Layout in an Order Picking Warehouse. *IIE Transactions*, 23(1):93–102, 1991.

[12] Philip Kaminsky and Dorit Hochbaum. Due date quotation models and algorithms. In Joseph Y. Leung, editor, *Handbook of Scheduling: Algorithms, Models, and Performance Analysis*, chapter 20. CRC Press, Inc., 2004.

[13] Pinar Keskinocak and Sridhar Tayur. Due date management policies. In David Simchi-Levi, S. David Wu, and Zuo-Jun (Max) Shen, editors, *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*, pages 485–554. Springer, 2004.

[14] R. Kunder and T. Gudehus. Mittlere wegzeiten beim eindimensionalen kommissionieren. *Zeitschrift für Operations Research*, 19:B53–B72, 1975.

[15] T. Le-Duc and R.(M.)B.M. de Koster. Travel distance estimation and storage zone optimization in a 2-block class-based storage strategy warehouse. *International Journal of Production Research*, 43(17):3561–3581, 2005.

[16] Uday S. Rao, Jayashankar M. Swaminathan, and Jun Zhang. Demand and Production Management with Uniform Guaranteed Lead Time. *Production and Operations Management*, 14(4):400–412, 2005.

[17] H. Donald Ratliff and Arnon S. Rosenthal. Order-Picking in a Rectangular Warehouse: A Solvable Case of the Traveling Salesman Problem. *Operations Research*, 31(3):507–521, 1983.

[18] Kees Jan Roodbergen and René de Koster. Routing order pickers in a warehouse with a middle aisle. *European Journal of Operational Research*, 133:32–43, 2001.

[19] Kees Jan Roodbergen and Iris F. A. Vis. A Model for Warehouse Layout. *IIE Transactions*, 38(10):799–811, 2006.

[20] Kees Jan Roodbergen, Gunter P. Sharp, and Iris F. A. Vis. Designing the layout structure of manual order picking areas in warehouses. *IIE Transactions*, 40:1032–1045, 2008.

[21] Weixin Shang and Liming Liu. Promised Delivery Time and Capacity Games in Time-Based Competition. *Management Science*, 57(3):599–610, 2011.

[22] Kut C. So and Jing-Sheng Song. Price, delivery time guarantees and capacity selection. *European Journal of Operational Research*, 111(1):28–49, 1998.

[23] Timothy L. Urban. Establishing delivery guarantees. *European Journal of Operational Research*, 196:959–967, 2009.