Pricing and Delivery Time Differentiation Using Shared Capacity

SACHIN JAYASWAL, ELIZABETH JEWKES
Department of Management Sciences
University of Waterloo, Canada

SAIBAL RAY
Desautels Faculty of Management,
McGill University, Canada

CORS/Optimization Days Joint Conference
May 12-14, 2008
Quebec City, Canada
Motivation

• UPS Express Early A.M.
• UPS Express
• UPS Express Saver
• UPS Expedited

• FedEx Next Flight
• FedEx First Overnight
• FedEx Priority Overnight
• FedEx 2Day
Motivation contd...

• Different firms, different operations strategies

According to UPS, “It is their integrated air and ground network that enhances pickup and delivery density and provides them with the flexibility to transport packages using the most efficient and cost-effective transportation mode or combination of modes.”

“The optimal way to serve very distinct market segments is to operate highly efficient independent networks with different facilities, different cut-off times and different delivery commitments.”
— Fredrick W. Smith, chairman and CEO, FedEx

• How does the operations strategy of a firm affect its price and delivery time differentiation strategy?
Problem Statement

• Market sensitive to Price and Time
• Different segments of customers have different sensitivities to Price and Delivery Time.

• Decisions (Dedicated vs. Shared capacity setting):
  • Delivery time guarantee for each segment?
  • Price charged to each segment?

• Objective:
  • Maximize profit
Model Description

- High priority (h)/express customers
- Low priority (l)/regular customers

- A 2-class pre-emptive priority queue
- Class h served in priority over class l, and charged a premium for shorter guaranteed delivery time
Notations

- $p_i$: price for class $i$
- $L_i$: delivery time for class $i$
- $\lambda_i$: demand rate (exponential) for class $i$
- $\mu$: service rate (exponential)
- $m$: unit operating cost
- $\Pi$: profit per unit time for the firm
- $A$: marginal capacity cost
- $W_i$: waiting time (in queue + service) of class $i$
- $S_i$: delivery time reliability level, $P(W_i \leq L_i)$
- $\alpha$: delivery time guarantee
Model Description

• Demand:
  • Exponential with rates $\lambda_1$ and $\lambda_2$
  • Price and delivery time sensitive
    \[ \lambda_h = a - \beta_p p_h + \theta_p (p_l - p_h) - \beta_L L_h + \theta_L (L_l - L_h) \]
    \[ \lambda_l = a - \beta_p p_l + \theta_p (p_h - p_l) - \beta_L L_l + \theta_L (L_h - L_l) \]

$2a$ : potential market size for the product
$\beta_p$ : price sensitivity of demand
$\beta_L$ : delivery time sensitivity of demand
$\theta_p$ : sensitivity of switches towards price difference
$\theta_L$ : sensitivity of switches towards guaranteed delivery time difference
Mathematical Model

Pricing and Delivery Time Decision Problem [PDTDP]

\[ [PDTDP]: \text{Maximize} \Pi(p_h, p_l, L_h, \mu) \]
\[ = (p_h - m)\lambda_h + (p_l - m)\lambda_l - A\mu \quad (1) \]

subject to:

\[ S_h(L_h) = P(W_h \leq L_h) = 1 - e^{-(\mu - \lambda_h)L_h} \geq \alpha \quad (2) \]
\[ S_l(L_l) = P(W_l \leq L_l) \geq \alpha \quad (3) \]
\[ \lambda_h + \lambda_l - \mu < 0 \quad (4) \]
\[ p_h \geq m, \ p_l \geq m, \ L_h \leq L_l, \ \lambda_h \geq 0, \ \lambda_l \geq 0 \quad (5) \]
Solution Approach

Pricing Decision Problem [PDP]

\[ [PDP]: \text{Maximize} \Pi(p_h, p_l, \mu) \]
\[ = (p_h - m)\lambda_h + (p_l - m)\lambda_l - A\mu \quad (1) \]

subject to:
\[ \mu - \lambda_h \geq -\frac{\ln(1-\alpha)}{L_h} \]
\[ S_h(L_h) = P(W_h \leq L_h) = 1 - e^{-(\mu - \lambda_h) L_h} \geq \alpha \quad (2) \]
\[ S_l(L_l) = P(W_l \leq L_l) \geq \alpha \quad (3) \]
\[ \lambda_h + \lambda_l - \mu < 0 \quad (4) \]
\[ p_h \geq m, \quad p_l \geq m, \quad L_h \leq L_l, \quad \lambda_h \geq 0, \quad \lambda_l \geq 0 \quad (5) \]

If \( S_l \) is \textit{concave}, it can be approximated using a set of \textit{tangent hyperplanes}, and the resulting problem can be solved using a \textit{cutting plane method} (Kelley 1990)
Solution Approach [PDP]

Linear approximation of constraint (3) at \((p_h^k, p_l^k, \mu^k)\)

\[
S_l^k(\cdot) \approx \left( p_h - p_h^k \right) \left( \frac{\partial S_l^k(\cdot)}{\partial p_h} \right) + \left( p_l - p_l^k \right) \left( \frac{\partial S_l^k(\cdot)}{\partial p_l} \right) + \left( \mu - \mu^k \right) \left( \frac{\partial S_l^k(\cdot)}{\partial \mu} \right) \alpha \quad \forall k \in K
\]  

(6)
Markov Chain Representation

**State Variables:**
- \( N_h(t) \): Number of *high* priority customers in the system
- \( N_l(t) \): Number of *low* priority customers in the system

\[ N(t) = \{N_l(t), N_h(t), t \geq 0\} \] 2-dimensional Markov process

**State space:** \( \{n = (n_l, n_h) | n_l \geq 0, n_h \leq M; l = l, h \} \)

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Rate</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>((n_l, n_h))</td>
<td>((n_l, n_h + 1))</td>
<td>(\lambda_h)</td>
<td>for (n_l \geq 0, n_h \geq 0)</td>
</tr>
<tr>
<td>((n_l, n_h))</td>
<td>((n_l + 1, n_h))</td>
<td>(\lambda_l)</td>
<td>for (n_l \geq 0, n_h \geq 0)</td>
</tr>
<tr>
<td>((n_l, n_h))</td>
<td>((n_l, n_h - 1))</td>
<td>(\mu)</td>
<td>for (n_l \geq 0, n_h &gt; 0)</td>
</tr>
<tr>
<td>((n_l, n_h))</td>
<td>((n_l - 1, n_h))</td>
<td>(\mu)</td>
<td>for (n_l &gt; 0, n_h = 0)</td>
</tr>
</tbody>
</table>
Rate Matrix

\[
Q = \begin{pmatrix}
B_0 & A_0 \\
A_2 & A_1 & A_0 \\
& A_2 & A_1 & A_0 \\
& & & & \ddots & \ddots & \ddots & \ddots
\end{pmatrix}
\]

\(B_0, A_0, A_2, A_2\) are square matrices of size \(M + 1\)
Matrix Geometric Method

<table>
<thead>
<tr>
<th>Obtain stationary distribution of ( {N_l(t), N_h(t)} ) using Matrix Geometric Method (Neuts, 1981)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniformize Markov Chain to obtain waiting time distribution of low priority customer ( s^k_L(L_l) ) (Latouche and Ramaswami, 1999)</td>
</tr>
<tr>
<td>Obtain gradients of ( s^k_L(L_l) ) using finite difference</td>
</tr>
</tbody>
</table>
| Obtain a tangent hyperplane:  
\[
S^k_L(L_l) + (p_h - p_h^k) \left( \frac{\partial S^k_L(L_l)}{\partial p_h} \right) + (p_l - p_l^k) \left( \frac{\partial S^k_L(L_l)}{\partial p_l} \right) + (\mu - \mu^k) \left( \frac{\partial S^k_L(L_l)}{\partial \mu} \right) \geq 0
\] |
Cutting Plane Algorithm

1. Start
2. $k = 0$, i.e., constraint set (6) is initially empty.
3. Solve $[PDP(K)]$ to obtain $p^*_k, p^*_t$ and $\mu^k$.
4. Given $p^*_k, p^*_t$ and $\mu^k$, compute $S^*_t(\cdot)$.
5. $k = k + 1$
6. Is $S^*_t(\cdot) \geq \alpha$?
   - Yes: Stop
   - No: Compute gradients $\frac{\partial S^*_t(\cdot)}{\partial p^*_k}, \frac{\partial S^*_t(\cdot)}{\partial p^*_t}, \frac{\partial S^*_t(\cdot)}{\partial \mu^k}$ using finite difference method. Generate a cut of the form (6).
   - Add a cut to $[PDP(K)]$. 
Solution Approach [PDTDP]

[PDTDP] can be rewritten

\[ \text{[PDTDP]: } \max_{L_h \in (0, L_f)} f(L_h) \]

where \( f(L_h) \) is a [PDP] for a fixed \( L_h \).

Plot of \( f(L_h) \) vs. \( L_h \) suggests \( f(L_h) \) is unimodal.

Hence, [PDTDP] can be solved using *Golden Section Search* method.
Market Scenarios

• Price and time difference sensitive (PTD) market:

\[ \beta_p \theta_L > \theta_p \beta_L \]

• Time and price difference sensitive (TPD) market:

\[ \beta_p \theta_L < \theta_p \beta_L \]

• Neither PTD nor TPD market:

\[ \beta_p \theta_L = \theta_p \beta_L \]
Observations & Insights [PDP]

**Observation 1:** A shared capacity results in (i) a generally higher optimal price for express customers (ii) a lower price for regular customers and (iii) a higher price differentiation between the two.
Observations & Insights [PDP]

**Observation 2:** Sharing capacity results in (i) a lower demand for express customers (ii) a higher demand for regular customers and (iii) a higher or lower total demand, depending on the delivery time differentiation.
Observations & Insights [PDP]

- **Observation 3**: Sharing capacity results in a larger profit. The relative gain in profit increases with delivery time differentiation.

% Profit Gain vs. Delivery Time of high priority

- **Ideal condition for a shared capacity setting is when the firm has identified markets for very different delivery time options.**
## Observations & Insights [PDTDP]

<table>
<thead>
<tr>
<th>Market Type</th>
<th>SC</th>
<th>DC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTD</td>
<td>0.597</td>
<td>0.112</td>
</tr>
<tr>
<td>Neither</td>
<td>0.594</td>
<td>0.367</td>
</tr>
<tr>
<td>TPD</td>
<td>0.612</td>
<td>0.575</td>
</tr>
</tbody>
</table>

Price and Delivery Time Differentiation

- **Observation 4:** The differences in the delivery times quoted and the prices charged to express and regular customers are higher in a shared capacity setting.
Observations & Insights [PDTDP]

- **Sharing capacity allows a firm to maintain a greater price and delivery time differentiation compared to dedicated capacities.**

<table>
<thead>
<tr>
<th>FedEx</th>
<th>UPS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Service</strong></td>
<td><strong>Guaranteed Delivery</strong></td>
</tr>
<tr>
<td>FedEx First Overnight</td>
<td>by 9:00 AM</td>
</tr>
<tr>
<td>FedEx Priority Overnight</td>
<td>by 12:00 PM</td>
</tr>
</tbody>
</table>

Price and Delivery Time Differentiation by FedEx vs. UPS
Observations & Insights [PDTDP]

Effect of Capacity Cost on (a) Delivery Time Differentiation (b) Price Differentiation

- **Observation 5:** With an increase in capacity cost ($A$), (i) the optimal delivery time differentiation decreases in both dedicated as well as shared capacity setting (ii) the optimal price differentiation decreases in a dedicated capacity setting but increases in a shared capacity setting
Observations & Insights [PDTDP]

Effect of Target Delivery Time Reliability on (a) Delivery Time Differentiation (b) Price Differentiation

- **Observation 6**: With an increase in the firm’s target delivery time reliability ($\alpha$), (i) the optimal delivery time differentiation decreases in both dedicated as well as shared capacity setting (ii) the optimal price differentiation decreases in a dedicated capacity setting but increases in a shared capacity setting
Implications of Rising Fuel Prices

- Reduce product differentiation both in terms of delivery time and price.
- Reduce delivery time differentiation and yet increase the price differentiation.
Future Research

- Extend the model to a competitive setting
- Computing Nash equilibrium a major challenge
Questions
Solution Approach: Literature Review

- Atalson, Epelman & Henderson (2004): *Call center staffing with simulation and cutting plane methods*

- Henderson & Mason (1998): *Rostering by integer programming and simulation*

- Morito, Koida, Iwama, Sato & Tamura (1999): *Simulation based constraint generation with applications to optimization of logistic system design*
Rate Matrix

\[
A_0 = \begin{pmatrix}
\lambda_l & & \\
& \ddots & \\
& & \lambda_l
\end{pmatrix}
\]

\[
A_2 = \begin{pmatrix}
\mu & & \\
& 0 & \\
& & \ddots
\end{pmatrix}
\]

\[
B_0 = \begin{pmatrix}
\ast & \lambda_h & \\
\mu & \ast & \lambda_h \\
\mu & \ast & \ddots & \ddots & \\
& \ddots & \ddots & \ddots & \\
\mu & \ast & & & \\
\end{pmatrix}
\]

\[
A_1 = B_0 - A_2
\]

where \( \ast \) is such that \( A_0 e + B_0 e = 0 \)
Matrix Geometric Method

• Stationary probability vector $x$:

$$x = [x_{00}, x_{01}, \ldots, x_{0M}, x_{10}, x_{11}, \ldots, x_{1M}, \ldots, \ldots, x_{l1}, \ldots, x_{lM}, \ldots, \ldots]$$

• Partition $x$ by levels into sub vectors $x_i$, where $x_i = [x_{i0}, x_{i1}, \ldots, x_{iM}]$

$$x = [x_0, x_1, x_2, x_3, \ldots, \ldots]$$

• $x$ can be obtained using the balance equations, given in matrix form by:

$$xQ = 0$$
Matrix Geometric Method

• The above set of equations can be solved recursively using the relation (Neuts, 1981):

\[ x_{i+1} = x_i R \]

• R is the minimal non-negative solution to the matrix quadratic equation:

\[ A_0 + RA_1 + R^2 A_2 = 0 \]

• The matrix R can be computed using well known methods (Latouche and Ramaswami, 1999):

\[ R(n + 1) = -[A_0 + R^2(n)A_2]A_1^{-1}; \quad R(0) = 0 \]
Matrix Geometric Method

• The probabilities $x_0$ are determined from the boundary condition:

$$x_0(R_0 + RA_2) = 0$$

subject to the normalization condition:

$$\sum x_1 e = x_0(I - R)^{-1}e = 1$$
Matrix Geometric Method for service level of low priority customers

- Time spent in the system by a low priority customer is the time until absorption in a modified Markov process $\tilde{N}(t)$, obtained by setting $\lambda_l = 0$. 

\[
\tilde{Q} = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
b_0 & \tilde{B}_0 & 0 & & & & \\
0 & A_2 & \tilde{A}_1 & 0 & & & \\
0 & & A_2 & \tilde{A}_1 & 0 & & \\
\vdots & & & & & \ddots & \ddots \\
\end{pmatrix}
\]

where, $\tilde{B}_0 = B_0 + A_0$; $\tilde{A}_1 = A_1 + A_0$; and $b_0 = [\mu \ 0 \ \cdots \ 0]^T_{M+1}$.
Matrix Geometric Method for service level of low priority customers

\[ S^k_\ell(y) = 1 - \overline{S}^k_\ell(y) = 1 - \sum_{i=0}^{\infty} x_i \overline{S}^k_\ell(y) e \]

- \( \overline{S}^k_\ell(y) \) can be computed more conveniently by uniformizing the Markov process \( \{ \overline{N(t)} \} \) with a Poisson process of rate \( \gamma \), where

\[
\gamma = \max_{0 \leq i \leq M} (-\hat{Q})_{ii} = \max_{0 \leq i \leq M} (-\hat{A}_1)_{ii} = \max_{0 \leq i \leq M} -(A_0 + A_1)_{ii}
\]
Service level of low priority customers

- For given prices and service rate \((p^k_i, p^k_i, \mu^k)\), service level for low priority customers can be computed as:

\[
S^k_i(L_1) = 1 - \overline{S^k_i}(L_1) = \sum_{n=0}^{\infty} e^{-\gamma L_1} \frac{(\gamma L_1)^n}{n!} x_0 (I - R)^{-1} H_n e
\]

\[
H_{n+1} = H_n \hat{A}_1 + RH_n \hat{A}_2; \quad H_0 = I
\]

\[
\frac{\partial S^k_i(\cdot)}{\partial p_h} = \frac{S^k_i(p^h_i + dp_h, p_i, \mu^k)(\cdot) - S^k_i(p^h_i - dp_h, p_i, \mu^k)(\cdot)}{2 dp_h}
\]

\[
\frac{\partial S^k_i(\cdot)}{\partial p_i} = \frac{S^k_i(p^h_i, p^k_i + dp_i, \mu^k)(\cdot) - S^k_i(p^h_i, p^k_i - dp_i, \mu^k)(\cdot)}{2 dp_i}
\]

\[
\frac{\partial S^k_i(\cdot)}{\partial \mu} = \frac{S^k_i(p^h_i, p_i, \mu^{k+2} + d\mu)(\cdot) - S^k_i(p^h_i, p_i, \mu^{k-2} - d\mu)(\cdot)}{2 d\mu}
\]