Queueing Systems: Lecture 6

Amedeo R. Odoni

November 1, 2004

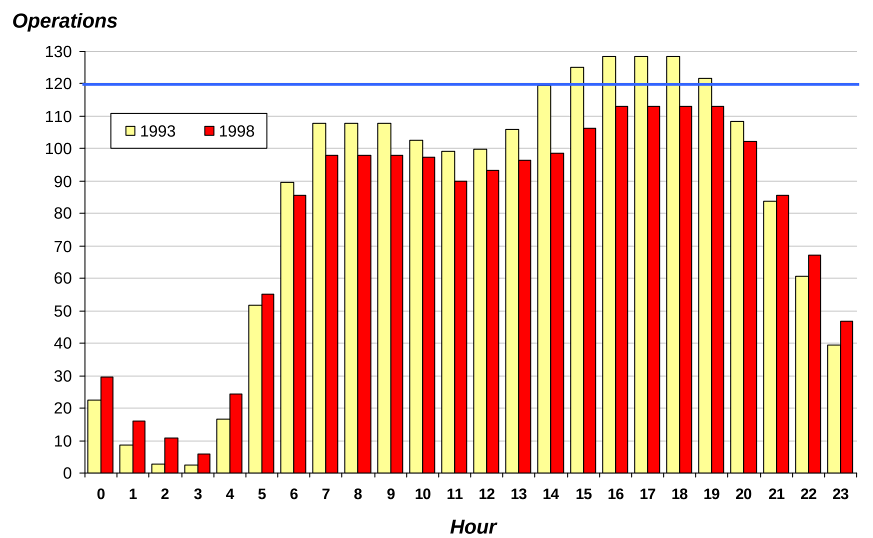
Lecture Outline

- Complete discussion of dynamic queues (qualitative obsrvations)
- Congestion pricing in transportation: the fundamental ideas
- Congestion pricing and queueing theory
- Numerical example
- A real example from LaGuardia airport
- Practical complications

Reference: Handout on "Congestion Pricing and Queueing Theory" (on course website)

Comparison of August Weekday Peaking Patterns

1993 vs. 1998 (3 Hour Average)



Two common "approximations" (??) for dynamic demand profiles

1. **Find the average demand per unit of time for the time interval of interest and then use steady-state expressions to compute estimates of the queuing statistics. [Problems?]**
2. **Subdivide the time interval of interest into periods during which demand stays roughly constant; apply steady-state expressions to each period separately. [Problems?]**

Problems
with the Approximate Methods

- **Problems with Approach 1:**

1. **For cases in which demand varies significantly (e.g., >10% from overall average value) the delay estimates can be VERY poor**
2. **Will underestimate overall average delay, possibly by a lot**

- **Problems with Approach 2:**

1. **May not have $\rho < 1$, for some intervals; then what?**
2. **Time to reach "steady state" is large for values of ρ which are close to 1; therefore "steady state" expressions may be very poor approximations when intervals are relatively short**
3. **Approach does not take into consideration the "dynamics" of the demand profile**

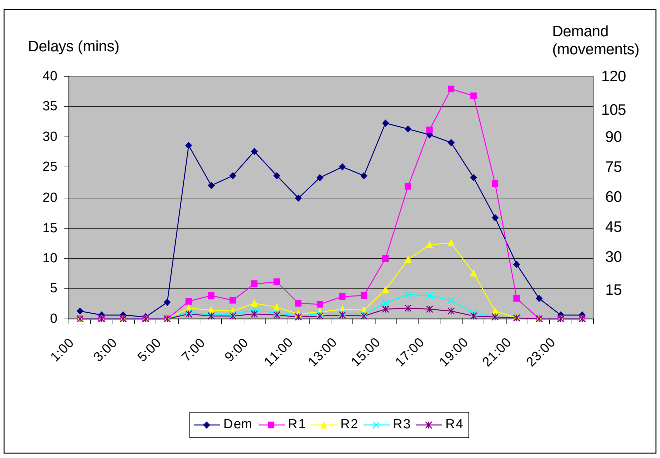
The Two Viable Approaches

1. **Simulation:**
   - High level of detail
   - May be only viable alternative for complex systems
   - Statistical significance of results?
2. **Numerical solution of equations describing the evolution of queueing system over time:**
   - Increasingly practical
   - May provide lots of information, such as $P_n(t)$

Dynamic Behavior of Queues

1. The dynamic behavior of a queue can be complex and difficult to predict
2. Expected delay changes non-linearly with changes in the demand rate or the capacity
3. The closer the demand rate is to capacity, the more sensitive expected delay becomes to changes in the demand rate or the capacity
4. The time when peaks in expected delay occur may lag behind the time when demand peaks
5. The expected delay at any given time depends on the "history" of the queue prior to that time
6. The variance (variability) of delay also increases when the demand rate is close to capacity

The dynamic behavior of a queue; expected delay for four different levels of capacity



(R1= capacity is 80 movements per hour; R2 = 90; R3 = 100; R4 = 110)

Two Recent References on Numerical Methods for Dynamic Queuing Systems

- Escobar, M., A. R. Odoni and E. Roth, "Approximate Solutions for Multi-Server Queuing Systems with Erlangian Service Times", with M. Escobar and E. Roth, *Computers and Operations Research, 29*, pp. 1353-1374, 2002.
- Ingolfsson, A., E. Akhmetshina, S. Budge, Y. Li and X. Wu, "A Survey and Experimental Comparison of Service Level Approximation Methods for Non-Stationary M/M/s Queueing Systems," Working Paper, July 2002.

  http://www.bus.ualberta.ca/aingolfsson/working_papers.htm

Congestion pricing: The basic observation

- The congestion costs due to any specific user have 2 components:

  (1) Cost of delay to that user (<u>internal</u> cost)

  (2) Cost of delay to all other users caused by that user (<u>external</u> cost)

- At congested facilities, this second component can be very large

- A congestion toll can be imposed to force users to experience this cost component (to "internalize the external costs")

Economic principle

Optimal use of a transportation facility cannot be achieved unless each additional (marginal) user pays for all the additional costs that this user imposes on all other users and on the facility itself. A congestion toll not only contributes to maximizing social economic welfare, but is also necessary to reach such a result. (Vickrey, 1967, 1969; Carlin + Park, 1970)

Two hard technical problems

- In practice it is very hard to:

(1) Estimate external marginal delay costs (extensive data analysis and/or simulation have been typically needed – subtle issues);

(2) Determine equilibrium congestion tolls (trial-and-error approach that may take long time to converge)

- Queueing theory has much to offer (especially with regard to the first problem) under certain conditions.

Computing Internal and External Costs

Consider a queueing facility with a single type of users in steady-state. Let

c = delay cost per unit time per user

C = total cost of delay per unit time incurred in the system

Then:

$$C = cL_q = c\lambda W_q$$

and the marginal delay cost, *MC* , imposed by an additional ("marginal") user is given by:

$$MC = \frac{dC}{d\lambda} = c\,W_q + c\lambda \frac{dW_q}{d\lambda}$$

Marginal cost — Internal cost — External cost

Numerical Example

- Three types of aircraft; Poisson; FIFO service
  - Non-jets: λ_1 = 40 per hour; c_1 = \$600 per hour
  - Narrow-body jets: λ_2 = 40 per hour; c_2 = \$1,800 per hour
  - Wide-body jets: λ_3 = 10 per hour; c_3 = \$4,200 per hour
  - Total demand is: $\lambda = \lambda_1 + \lambda_2 + \lambda_3$ = 90 per hour
- pdf for service times is uniform
  - U[25 sec, 47 sec]
  - *E*[*S*] = 36 sec = 0.01 hour; μ = 100 per hour

$$\sigma_S^2 = \frac{22^2}{12} = 40.33\ \text{sec}^2 = 3.11213 \times 10^{-6}\ \text{hours}^2$$

- Note: We have a M/G/1 system

Numerical Example [2]

$$W_q = \frac{\lambda \cdot [E^2[S] + \sigma_S^2]}{2 \cdot (1-\rho)} = \frac{90 \cdot [(0.01)^2 + 3.11213 \times 10^{-6}]}{2 \cdot (1 - 90/100)} \approx 0.0464 \text{ hours} \approx 167 \text{ sec}$$

Define: $c = c_1 \frac{\lambda_1}{\lambda} + c_2 \frac{\lambda_2}{\lambda} + c_3 \frac{\lambda_3}{\lambda}$

$$C = c \cdot L_q = c \cdot \lambda \cdot W_q = (c_1 \cdot \lambda_1 + c_2 \cdot \lambda_2 + c_3 \cdot \lambda_3) \cdot W_q = \bar{c} \cdot W_q$$

Or: $C = \bar{c} \cdot W_q = (\$138{,}000) \cdot (0.0464) = \$6{,}400$

$$\frac{dW_q}{d\lambda} = \frac{E^2[S] + \sigma_S^2}{2 \cdot (1-\rho)} + \frac{\lambda \cdot [E^2[S] + \sigma_S^2]}{2 \cdot (1-\rho)^2} \cdot \frac{1}{\mu} \approx 5.1556 \times 10^{-6} \text{ hours} \approx 18.6 \text{ sec}$$

Numerical Example [3]

$$\frac{dC}{d\lambda_1} = c_1 \cdot W_q + \bar{c} \cdot \frac{dW_q}{d\lambda} \approx \$28 + \$711 = \$739$$

internal cost (↑ $28)

external cost= congestion toll (↑ $711)

$$\frac{dC}{d\lambda_2} = c_2 \cdot W_q + \bar{c} \cdot \frac{dW_q}{d\lambda} \approx \$85 + \$711 = \$796$$

$$\frac{dC}{d\lambda_3} = c_3 \cdot W_q + \bar{c} \cdot \frac{dW_q}{d\lambda} \approx \$198 + \$711 = \$909$$

Generalizing to *m* types of users…

- **Facility users of type *i*: arrival rate λ_i ;**

  service time S_i with $\mu_i^{-1} = E[S_i]$;

  cost per unit of time in the system c_i

- **For entire set of facility users, we have**

$$\lambda = \sum_{i=1}^{m} \lambda_i \qquad \frac{1}{\mu} = E[S] = \sum_{i=1}^{m} \left(\frac{\lambda_i}{\lambda} \times \frac{1}{\mu_i} \right)$$

$$\rho = \frac{\lambda}{\mu} = \sum_{i=1}^{m} \rho_i = \sum_{i=1}^{m} \frac{\lambda_i}{\mu_i} \qquad c = \sum_{i=1}^{m} \left(\frac{\lambda_i}{\lambda} c_i \right)$$

Generalization (continued)

- **As before:** $C = cL_q = c\lambda W_q$

giving:

$$MC(i) = \frac{dC}{d\lambda_i} = c_i W_q + c\lambda \frac{dW_q}{d\lambda_i}$$

- **When we have explicit expressions for W_q, we can also compute explicitly the total marginal delay cost *MC(i)*, the internal (or private) cost and the external cost associated with each additional user of type *i***

Example

For an M/G/1 system:

$$MC(i) = \frac{dC}{d\lambda_i} = c_i \frac{\lambda \cdot E[S^2]}{2(1-\rho)} + c\lambda \frac{(1-\rho)E[S_i^2] + \frac{\lambda}{\mu_i}E[S^2]}{2(1-\rho)^2}$$

- Can extend further to cases with user priorities

Finding Equilibrium Conditions and Optimal Congestion Tolls!

We now generalize further: let x_i be the *total* cost perceived by a user of type *i* for access to a congested facility and let $\lambda_i(x_i)$ be the demand function for type *i* users.

$$\mathbf{x_i = IC_i + CT_i + K_i}$$

***IC_i* = internal private cost; it is a function of the demand rates, $\lambda_i(x_i)$**

***CT_i* = congestion toll imposed; equal to 0 in absence of congestion tolls; can be set arbitrarily or can be set as a function of the $\lambda_i(x_i)$under congestion pricing schemes**

***K_i* = any other charges that are independent of level of congestion**

Finding Equilibrium Conditions and Optimal Congestion Tolls! [2]

- With *m* types of users, the equilibrium conditions for any set of demand functions, can be found by solving simultaneously the *m* equations:

$$x_i = c_i \cdot W_q[\hat{\lambda}(\hat{x})] + \left(\sum_{j=1}^{m} c_j \cdot \lambda_j(x_j) \right) \cdot \frac{dW_q[\hat{\lambda}(\hat{x})]}{d\lambda_i(x_i)} + K_i \qquad \forall i$$

where $\hat{\lambda}(\hat{x}) = \{\lambda_1(x_1), \lambda_2(x_2), \ldots, \lambda_m(x_m)\}$.

The missing piece: Demand functions can only be roughly estimated, at best!

An illustrative example from airports

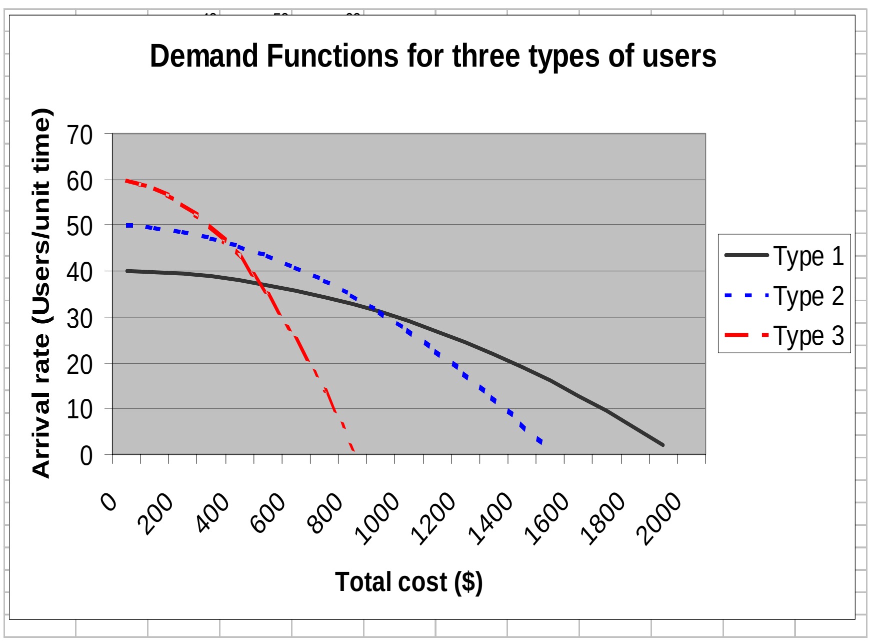
	Type 1 (Big)	Type 2 (Medium)	Type 3 (Small)
Service rate (movements per hour)	80	90	100
Standard deviation of service time (seconds)	10	10	10
Cost of delay time ($ per hour)	$2,500	$1,000	$400

Hypothetical Demand Functions

$$\lambda_1(x_1) = 40 - 0.001 \cdot x_1 - 0.00001 \cdot x_1^2$$

$$\lambda_2(x_2) = 50 - 0.003 \cdot x_2 - 0.00002 \cdot x_2^2$$

$$\lambda_3(x_3) = 60 - 0.01 \cdot x_3 - 0.00008 \cdot x_3^2$$

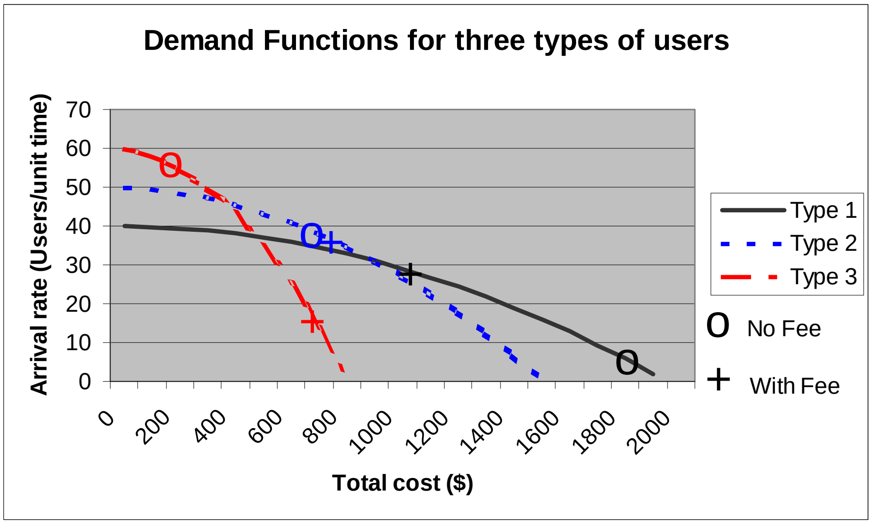


Case 1: No Congestion Fee

	Type 1	Type 2	Type 3
No Congestion Fee			
(1) Delay cost (IC) per aircraft	$1,802	$721	$288
(2) Congestion fee	$0	$0	$0
(3) Total cost of access [=(1)+(2)]	$1802	$721	$288
(4) Demand (no. of movements per hour)	5.7	37.4	50.5
(5) Total demand (no. of movements per hour)	93.6		
(6) Expected delay per aircraft	43 minutes 15 seconds		
(7) Utilization of the airport (% of time busy)	99.2%		

Case 2: Optimal Congestion Fee

Optimal Congestion Fee			
(8) Delay cost (IC) per aircraft	$135	$54	$22
(9) Congestion fee (CF)	$853	$750	$670
(10) Total cost of access [=(1)+(2)]	$988	$804	$692
(11) Demand (no. of movements per hour)	29.2	34.6	14.9
(12) Total demand (no. of movements per hour)	78.7		
(13) Expected delay per aircraft	3 minutes 15 seconds		
(14) Utilization of the airport (% of time busy)	89.9%		



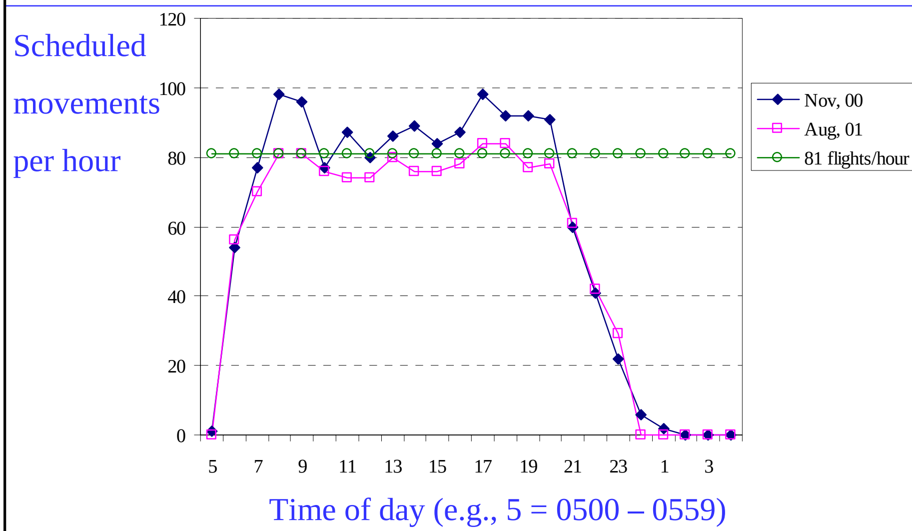
Important to note…

- The external costs computed, in the absence of congestion pricing, give only an upper bound on the magnitude of the congestion-based fees that might be charged
- These are *not* "equilibrium prices"
- Equilibrium prices may turn out to be considerably less than these upper bounds
- Equilibrium prices are hard to estimate, absent knowledge of demand functions

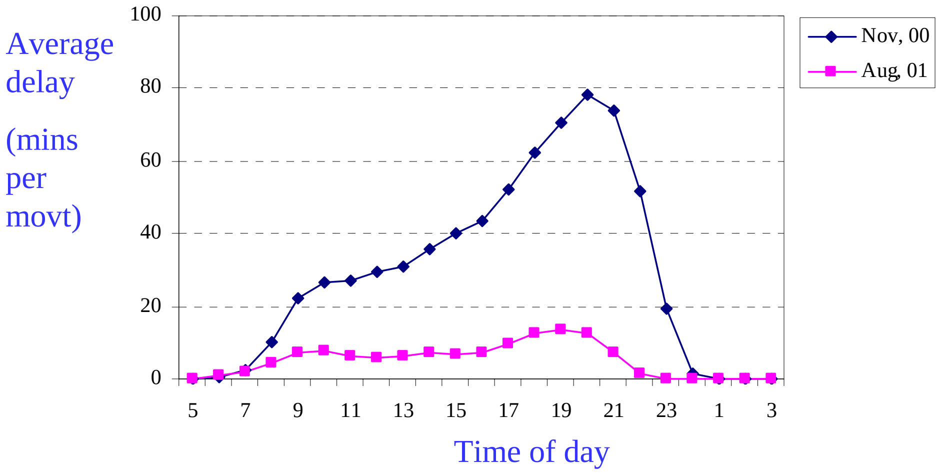
Case of LaGuardia (LGA)

- Since 1969: Slot-based High Density Rule (HDR)
  - DCA, JFK, LGA, ORD; "buy-and-sell" since 1985
- Early 2000: About 1050 operations per weekday at LGA
- April 2000: Air-21 (Wendell-Ford Aviation Act for 21st Century)
  - Immediate exemption from HDR for aircraft seating 70 or fewer pax on service between small communities and LGA
- By November 2000 airlines had added over 300 movements per day; more planned: virtual gridlock at LGA
- December 2000: FAA and PANYNJ implemented slot lottery and announced intent to develop longer-term policy for access to LGA
- Lottery reduced LGA movements by about 10%; dramatic reduction in LGA delays
- June 2001: Notice for Public Comment posted with regards to longer-term policy that would use "market-based" mechanisms
- Process stopped after September 11, 2001; re-opened recently

Scheduled aircraft movements at LGA before and after slot lottery



Estimated average delay at LGA before and after slot lottery in 2001



LGA: Marginal delay caused by an additional operation by time of day