Queueing Systems: Lecture 5

Amedeo R. Odoni

October 27, 2004

Lecture Outline

- Bounds for G/G/1 systems
- A fundamental result for queueing networks
- State transition diagrams for Markovian queueing systems and networks: examples
- Analysis of dynamic systems
- Qualitative behavior of dynamic systems

Reference: Sections 4.8.3, 4.10, 4.11 + material in handout

A general upper bound for G/G/1 systems

- A number of bounds are available for very general queueing systems (see Section 4.8)
- A good example is an upper bound for the waiting time at G/G/1 systems:

$$W_q \le \frac{\lambda \cdot (\sigma_X^2 + \sigma_S^2)}{2 \cdot (1-\rho)} \quad (\rho < 1) \qquad \textbf{(1)}$$

where *X* and *S* are, respectively, the r.v.'s denoting inter-arrival times and service times

- Under some fairly general conditions, such bounds can be tightened and perform extremely well

Better bounds for a (not so) special case

- For G/G/1 systems whose inter-arrival times have the property that for all non-negative values of t_0,

$$E[X - t_0 \mid X > t_0] \le \frac{1}{\lambda}$$

(what does this mean, intuitively?)

it has been shown that:

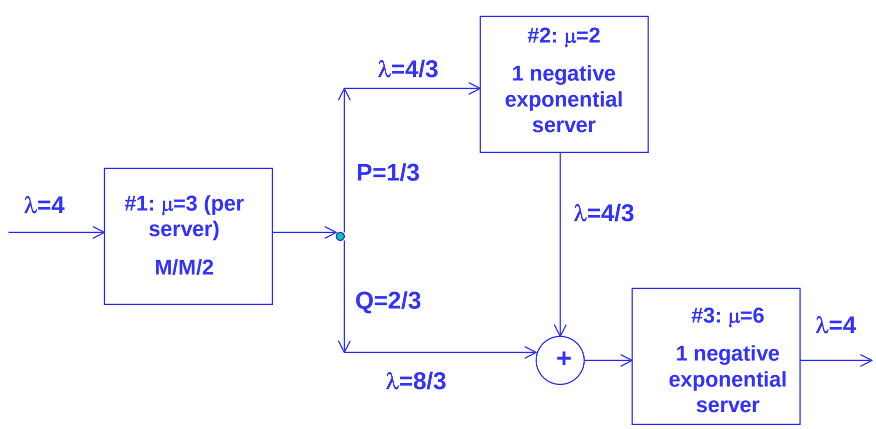
$$B - \frac{1+\rho}{2\lambda} \le W_q \le B = \frac{\lambda \cdot (\sigma_X^2 + \sigma_S^2)}{2 \cdot (1-\rho)} \quad (\rho < 1) \qquad \textbf{(2)}$$

- Note that the upper and lower bounds in (1) differ by, at most, $1/\lambda$ and that the percent difference between the upper and lower bounds decreases as ρ increases!

A result which is important in analyses of queuing networks

- Let the arrival process at a M/M/m queuing system with infinite queue capacity have parameter λ. Then, under steady state conditions ($\lambda < m\mu$) the departure process from the queuing system is also Poisson with parameter λ.
- Implication: greatly facilitates analysis of open acyclic networks consisting of M/M/m queues with infinite queue capacities.
- The bad news: result holds only under exact set of conditions described above.

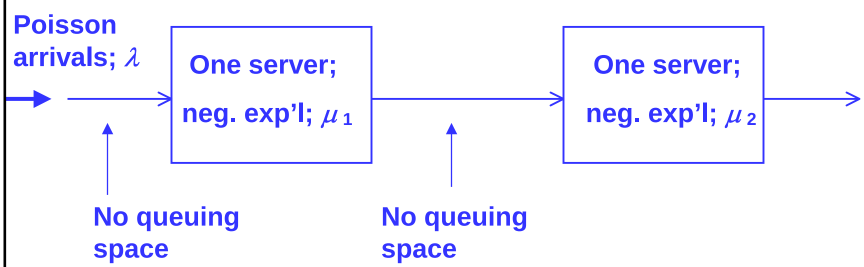
Open acyclic network of M/M/. systems



State transition diagrams for queuing systems and networks

- When external arrivals are Poisson and service times are negative exponential, many complex queuing systems and open acyclic queuing networks can be analyzed, even under dynamic conditions, through a judicious choice of state representation.
- This involves writing and solving (often numerically) the steady-state balance equations or the Chapman-Kolmogorov first-order differential equations.
- The "hypercube model" (Chapter 5 of Larson and Odoni) is a good example.

Example 1: Two M/M/1 Queuing Systems, Each with Finite Queue Capacity

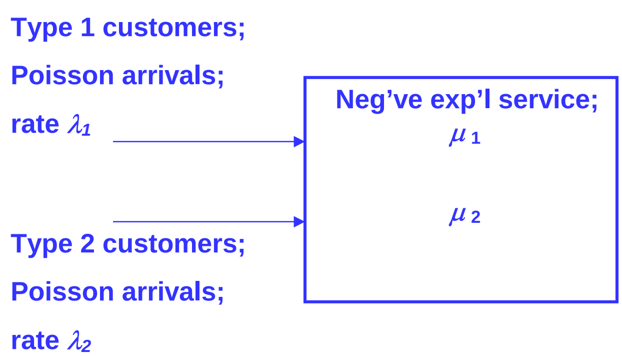


Note: The queuing system on the right may "block" the one on the left.

Example 2: M/E_K/1 System, with system capacity for total of N users

See distributed notes.

Example 3: Two Types of Users and Non-Preemptive Priorities

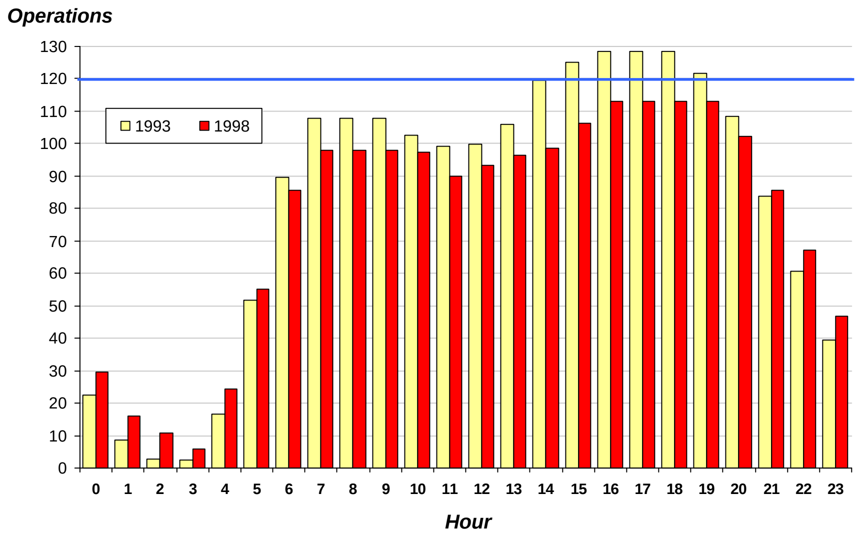


Type 1 customers have non-preemptive priority

See distributed notes.

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**

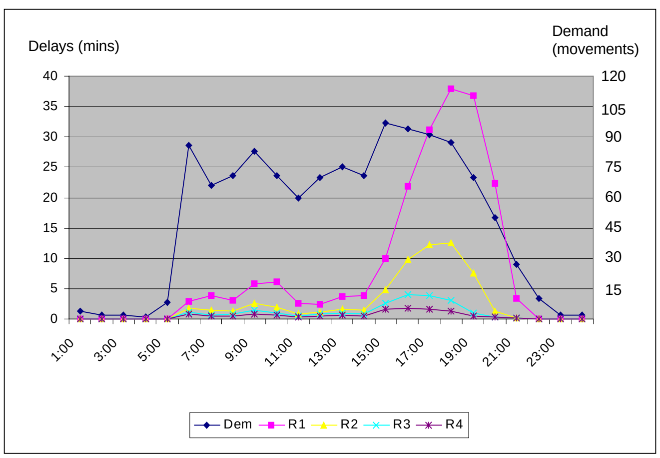
The Two Viable Approaches

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 2003.

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