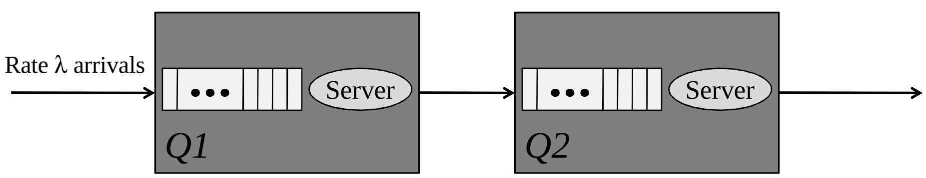
Return to Queueing Theory

- Build on the previous material
  - From single-hop queueing delay to end-to-end queueing delay
  - Overall delay experienced in routing a packet from source to destination node

- Queueing Networks
  - Section 3.6 in text
  - Each queue represents a node which may have multiple input streams of "arrivals" from other nodes
  - Also, output stream may be split to multiple next-hop queues
  - Arrival process at one queue may depend on departure processes of multiple other queues, so probably is not Poisson
  - System is complex! What can be done?

1

Example



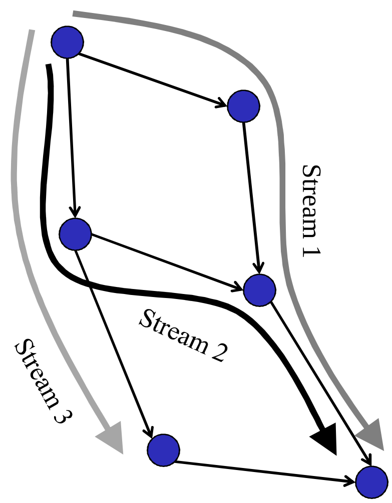
- "Transmission line" model
  - Rate λ arrivals over one link, fixed length packets
    - Service times are constant (equal to transmission delay $1/\mu$), so *Q1* is *M/D/1* (*D* means deterministic) Arrivals into *Q2* separated by at least $1/\mu$
    - If rate of second link is ≥ that of first link, there is ***NO waiting in the second queue EVER!***
      - If packet lengths are iid exponential, this becomes *M/M/1*
    - How to characterize/analyze the second queue?

2

How to Evaluate Delay?

– Setup:

- Assume multiple packet streams each following a unique path
- Let x_s be the arrival rate (in packets/second) for stream *s*
- For circuit switching:
  – Arrivals on a link *(i,j)* occur with rate λ_{ij} equal to summation of x_s for each stream *s* traversing *(i,j)*
- For packet switching:
  – Include fraction $f_{ij}(s)$ of packets in stream *s* traversing *(i,j)*

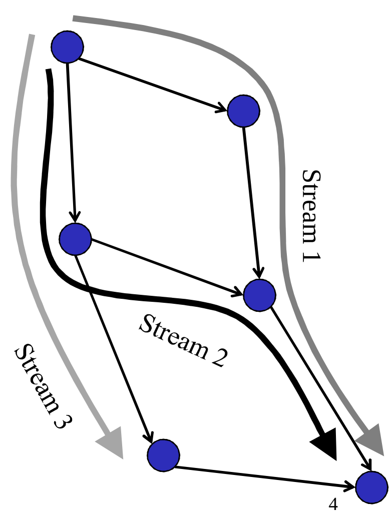


3

Kleinrock Independence Approximation

- 1964: Kleinrock suggests that random merger of several packet streams into a single stream restores independence of arrival times and packet lengths

- In particular, for densely connected networks with moderate-to-heavy traffic, inputs are approximately Poisson processes



4

Analysis under KIA

- Average number of packets in queue or service:
  - For link *(i,j)*, $N_{ij} = \lambda_{ij} / (\mu_{ij} - \lambda_{ij})$ where $1/\mu_{ij}$ is average transmit delay
  - Total packets in queue or service N is summation of N_{ij} over all links *(i,j)*
- Average queueing delay:
  - From Little's Theorem, $T = N/\gamma$, where γ is sum rate of all x_s arrival streams
- Including proc/prop delay d_{ij} at each *(i,j)*:
  - Add the term $\lambda_{ij} d_{ij}$ in the summation

5

Time Reversibility

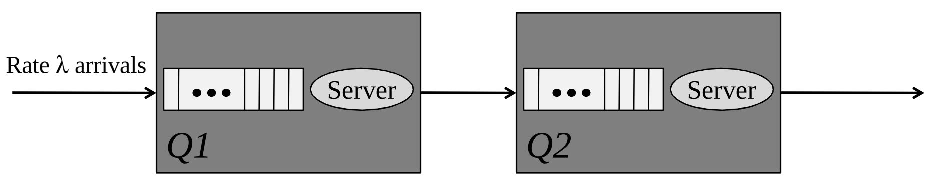
- Let X be a DTMC, and let X_n denote the state at time $n>>0$. Let's look at the behavior backward in time, instead of looking forward in time.
- Let P^*_{ij} be the reverse transition probability given by $\Pr[X_m = j \mid X_{m+1} = i]$.
- Show that $\pi_i P^*_{ij} = \pi_j P_{ji}$
- *Def:* A DTMC X is ***time reversible*** if $P^*_{ij} = P_{ij}$. In this case, the stationary distribution of the reversed DTMC X^* is the same as the forward chain X.
  - We immediately see that a DTMC X is time reversible *if and only if* the detailed balance equations hold.
- This extends in a straightforward way to CTMCs.

6

Burke's Theorem

- *Theorem (Burke)*: Consider an *M/M/1*, *M/M/k*, or *M/M/∞* queue with arrival rate λ, and suppose the system starts in steady-state. Then:
  - The departure process is Poisson with rate λ.
  - The number of customers in the system at time t is independent of the sequence of departure times prior to t.
- Why is this true?
  - *M/M/1*, *M/M/k*, and *M/M/ ∞* are time reversible CTMCs.
  - Departures prior to time t are arrivals after time t in reversed process, which is Poisson, so future arrivals do not depend on system occupancy.

Example – Revisited



- Returning to the previous example:
  - Suppose *Q1* is *M/M/1* with exp. service rate μ_1.
  - From Burke's Theorem, arrivals to *Q2* are Poisson with rate λ and exp. service rate μ_2 → *M/M/1*.
  - → $\pi_n^{(i)} = \rho_i^n(1-\rho_i)$, $\rho_i = \lambda/\mu_i$, $i=1,2$
  - From 2nd part of Burke's Theorem, number of customers in each queue is independent, so Pr$[n \text{ in } Q1, m \text{ in } Q2] = \pi_n^{(1)}\, \pi_m^{(2)}$
  - This is a simple example of an ***acyclic queueing network***

Queueing Networks

- Setup:
  - A network of K queues, each with a single server, service rate μ_i at queue i
  - Arrivals into the network at queue i follow a Poisson process with rate r_i. At least one r_i is positive.
  - Customer served by queue i moves to queue j with probability P_{ij} – leaves with residual probability
  - Total arrival rate into queue j is

$$\lambda_j = r_j + \sum_{i=1}^{K} \lambda_i P_{ij}, \qquad \text{for } j = 1, \ldots, K$$

  - Let $\rho_j = \lambda_j / \mu_j$

9

Queueing Network State Space

- Let $n = (n_1,\ldots,n_K)$ denote a state in the state space $\mathcal{X}^K$, where $\mathcal{X}$ is the state space of each queue (here, the non-negative integers), i.e. n_i = number of customers in queue i.
- State transitions (with very high probability):
  - New arrival at queue j (with rate r_j)
    - State $n(j^+) = n + e_j = (n_1,\ldots,n_{j-1},n_j+1,n_{j+1},\ldots,n_K)$
  - Exit system from queue j (with rate $\mu_j \left(1 - \sum_{i=1}^{K} P_{ji}\right)$)
    - State $n(j^-) = n - e_j = (n_1,\ldots,n_{j-1},n_j-1,n_{j+1},\ldots,n_K)$
  - Customer moves from j to i (with rate $\mu_j P_{ji}$)
    - State $n(j^-, i^+) = n - e_j + e_i$

10

Jackson's Theorem

- *Theorem (Jackson)*: For such a queueing network, if $\rho_j < 1$ for $j=1,\ldots,K$, then for all $n=(n_1,\ldots,n_K)$, $n_j >= 0$, we have

$$\pi_n = \prod_{j=1}^{K} \pi_{n_j}^{(j)}$$

$$\pi_{n_j}^{(j)} = \rho_j^{n_j}(1-\rho_j)$$

- In other words, queueing network behaves as **independent collection of *M/M/1* queues** even though the arrival to each queue **is not Poisson**.

11

Closed Queueing Networks

- Same setup as before, only $r_i=0$ and P_{ij} sum to 1 over j for all $i=1,\ldots,K$, i.e. nobody enters and nobody leaves, so

$$\lambda_j = \sum_{i=1}^{K} \lambda_i P_{ij}, \qquad \text{for } j = 1, \ldots, K$$

  and total number of customers is fixed at M.
- Here, π_n can be positive only if

$$n_1+\ldots+n_K = M.$$

12