Performance Analysis

Peter Harrison, Maria Vigliotti and Jeremy Bradley

Room 372. Email: jb@doc.ic.ac.uk

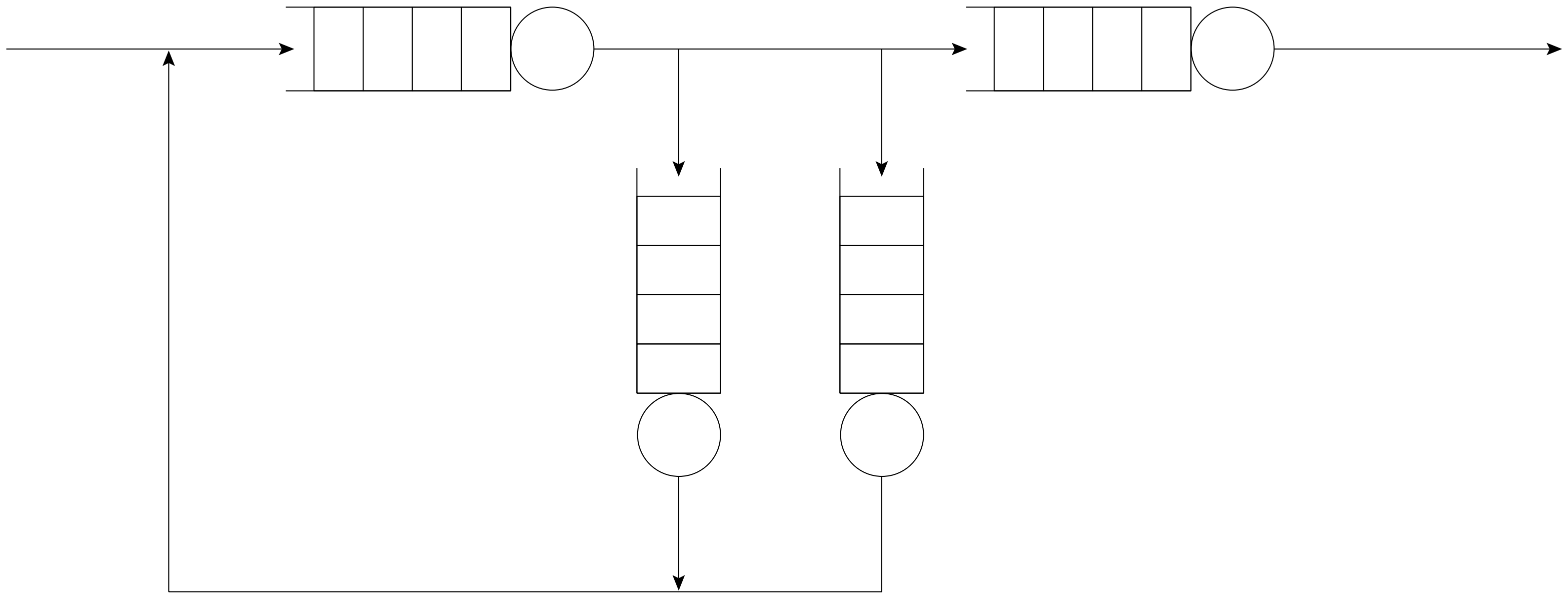
Department of Computing, Imperial College London

Produced with prosper and LaTeX

Useful facts...

- Little's Law: $L = \gamma W$
  - L – mean buffer length; γ – arrival rate; W – mean waiting time/passage time
  - only applies to system in steady-state; no creating/destroying of jobs
- For M/M/1 queue:
  - λ – arrival rate, μ – service rate
  - Stability condition, $\rho = \lambda/\mu < 1$ for steady state to exist
  - Mean queue length $= \frac{\rho}{1-\rho}$
  - $\mathbb{P}(n \text{ jobs in queue at s-s}) = \rho^n(1-\rho)$

Queueing Networks



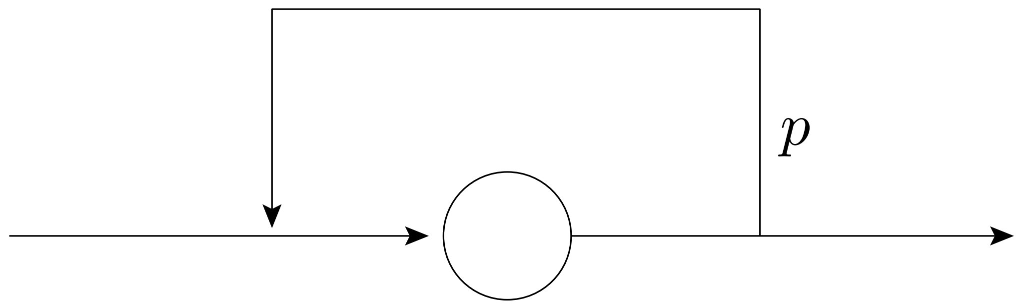
- Individual queue nodes represent contention for single resources
- A system consists of many inter-dependent resources – hence we need to reason about a *network* of queues to represent a system

Open Queueing Networks

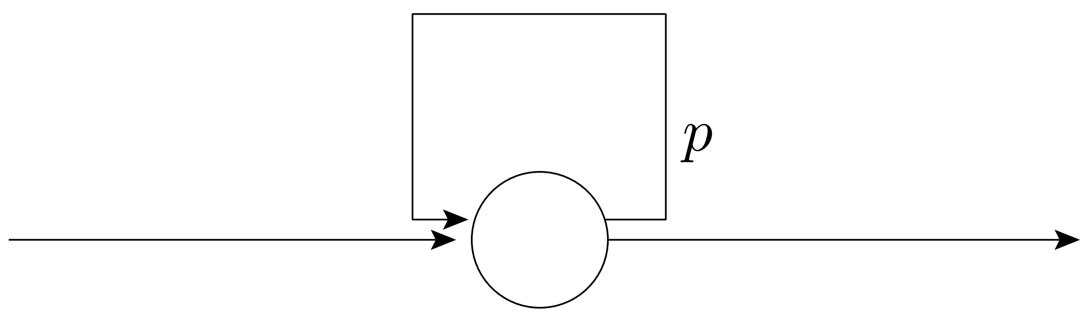
- A network of queueing nodes with inputs/outputs connected to each other
- Called an *open* queueing network (or OQN) because, traffic may enter (or leave) one or more of the nodes in the system from an external source (to an external sink)
- An open network is defined by:
  - γ_i, the exponential arrival rate from an external source
  - q_{ij}, the probability that traffic leaving node i will be routed to node j
  - μ_i exponential service rate at node i

OQN: Notation

- A node whose output can be probabilistically redirected into its input is represented as:



- or...



- probability p of being rerouted back into buffer

OQN: Network assumptions

In the following analysis, we assume:

- Exponential arrivals to network
- Exponential service at queueing nodes
- FIFO service at queueing nodes
- A network may be stable (be capable of reaching steady-state) or it may be unstable (have unbounded buffer growth)
- If a network reaches steady-state (becomes stationary), a single rate, λ_i, may be used to represent the throughput (both arrivals and departure rate) at node i

OQN: Traffic Equations

- The traffic equations for a queueing network are a linear system in λ_i
- λ_i represents the aggregate arrival rate at node i (taking into account any traffic feedback from other nodes)
- For a given node i, in an open network:

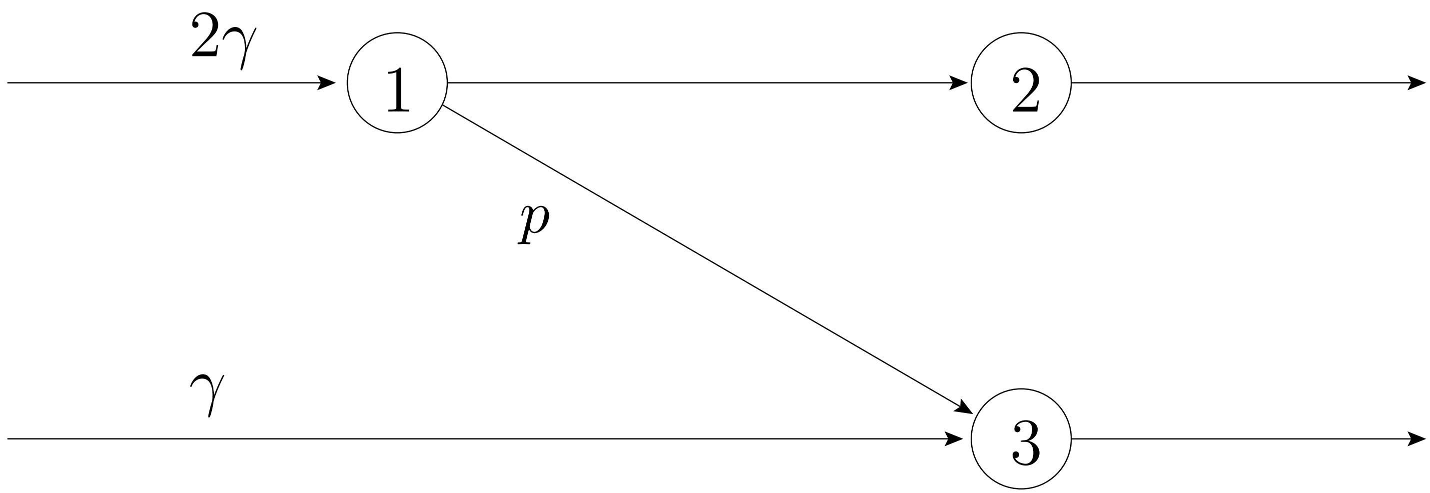
$$\lambda_i = \gamma_i + \sum_{j=1}^{N} \lambda_j q_{ji} \qquad : i = 1, 2, \ldots, N$$

OQN: Traffic Equations

- Define:
  - the vector of aggregate arrival rates $\vec{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N)$
  - the vector of external arrival rates $\vec{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_N)$
  - the matrix of routeing probabilities $Q = (q_{ij})$
- In matrix form, traffic equations become:

$$\begin{aligned} \vec{\lambda} &= \vec{\gamma} + \vec{\lambda} Q \\ &= \vec{\gamma}(I - Q)^{-1} \end{aligned}$$

OQN: Traffic Equations: example 1

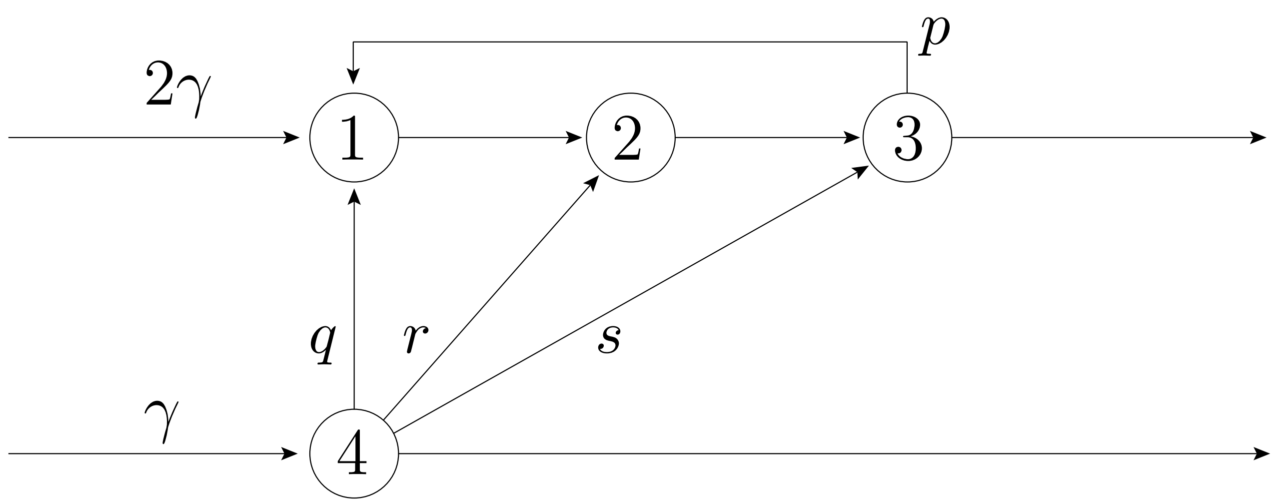


- Set up and solve traffic equations to find λ_i:

$$\vec{\lambda} = \begin{pmatrix} 2\gamma \\ 0 \\ \gamma \end{pmatrix} + \vec{\lambda} \begin{pmatrix} 0 & 1-p & p \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

- i.e. $\lambda_1 = 2\gamma$, $\lambda_2 = (1-p)\lambda_1$, $\lambda_3 = \gamma + p\lambda_1$

OQN: Traffic Equations: example 2



- Set up and solve traffic equations to find λ_i:

$$\vec{\lambda} = \begin{pmatrix} 2\gamma \\ 0 \\ 0 \\ \gamma \end{pmatrix} + \vec{\lambda} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ p & 0 & 0 & 0 \\ q & r & s & 0 \end{pmatrix}$$

OQN: Network stability

- Stability of network (whether it achieves steady-state) is determined by utilisation, $\rho_i < 1$ at every node i
- After solving traffic equations for λ_i, need to check that:

$$\rho_i = \frac{\lambda_i}{\mu_i} < 1 \quad : \forall i$$

Recall facts about M/M/1

- If λ is arrival rate, μ service rate then $\rho = \lambda/\mu$ is utilisation
- If $\rho < 1$, then steady state solution exists
- Average buffer length:

$$\mathbb{E}(N) = \frac{\rho}{1-\rho}$$

- Distribution of jobs in queue is:

$$\mathbb{P}(n \text{ jobs is queue at steady-state}) = \rho^n(1-\rho)$$

OQN: Jackson's Theorem

- Where node i has a service rate of μ_i, define $\rho_i = \lambda_i / \mu_i$
- If the arrival rates from the traffic equations are such that $\rho_i < 1$ for all $i = 1, 2, \ldots, N$, then the steady-state exists and:

$$\pi(r_1, r_2, \ldots, r_N) = \prod_{i=1}^{N} (1 - \rho_i) \rho_i^{r_i}$$

- This is a *product form* result!

OQN: Jackson's Theorem Results

- The marginal distribution of no. of jobs at node i is same as for isolated M/M/1 queue: $\rho^n(1-\rho)$
- Number of jobs at any node is independent of jobs at any other node – hence *product form* solution
- Powerful since queues can be reasoned about separately for queue length – summing to give overall network queue occupancy

OQN: Mean Jobs in System

- If only need mean results, we can use Little's law to derive mean performance measures
- Product form result implies that each node can be reasoned about as separate M/M/1 queue in isolation, hence:

$$\text{Av. no. of jobs at node } i = L_i = \frac{\rho_i}{1 - \rho_i}$$

- Thus total av. number of jobs in system is:

$$L = \sum_{i=1}^{N} \frac{\rho_i}{1 - \rho_i}$$

OQN: Mean Total Waiting Time

- Applying Little's law to whole network gives:

$$L = \gamma W$$

  where γ is total external arrival rate, W is mean response time.

- So mean response time from entering to leaving system:

$$W = \frac{1}{\gamma} \sum_{i=1}^{N} \frac{\rho_i}{1 - \rho_i}$$

OQN: Intermediate Waiting Times

- r_i represents the the average waiting time from arriving at node i to leaving the system
- w_i represents average response time at node i, then:

$$r_i = w_i + \sum_{j=1}^{N} q_{ij} r_j$$

- which as before gives a vector equation:

$$\begin{aligned} \vec{r} &= \vec{w} + Q\vec{r} \\ &= (I - Q)^{-1}\vec{w} \end{aligned}$$

OQN: Average node visit count

- v_i represents the average number of times that a job visits node i while in the network
- If γ represents the total arrival rate into the network, $\gamma = \sum_i \gamma_i$:

$$v_i = \frac{\gamma_i}{\gamma} + \sum_{j=1}^{N} v_j q_{ji}$$

- so for $\vec{\gamma}' = \vec{\gamma}/\gamma$:

$$\begin{aligned} \vec{v} &= \vec{\gamma}' + \vec{v}Q \\ &= \vec{\gamma}'(I - Q)^{-1} \end{aligned}$$

OQN: Average node visit count

- Compare average visit count equations with traffic equations:

$$\vec{v} = \vec{\gamma}'(I - Q)^{-1}$$
$$\vec{\lambda} = \vec{\gamma}(I - Q)^{-1}$$

- We can see that: $\vec{v} = \vec{\lambda}/\gamma$, so if we have solved the traffic equations, we needn't perform a separate linear calculation