Open Queueing Networks

Himanshu Singh Vikaskumar Kalsariya Aditya Kulkarni Kiran Gullapalli
Keshava V

March 2, 2024

Queueing Networks

- A queueing network is a connected directed graph, whose nodes represent servers (each with their own queue), and edges between nodes represent channels through which jobs can be routed.

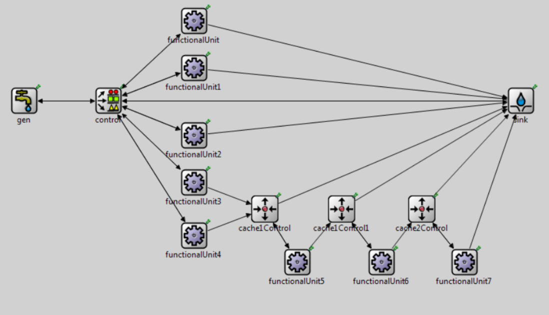


Figure: Intel Ivy Bridge queue model [2]

Open Networks

- A network is said to be open if there is at least one edge for an ‘external’ arrival, and from every node there is a path for an ‘external’ departure.

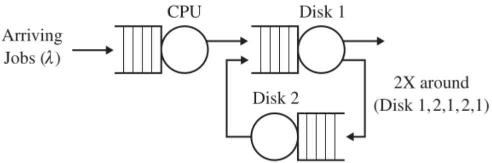


Figure: An open queueing network

Jackson Network

- A general form of open network
  - Unbounded queues, with FCFS service order
  - External arrivals $\sim$ Poisson(r_i)
  - Service rates $\sim$ Exp(μ_i)
  - Probabilistic routing - P_{ij}
  - States: $(n_1, n_2, \ldots, n_k)$

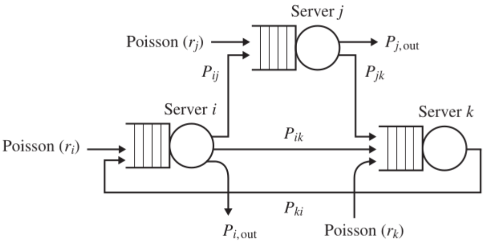
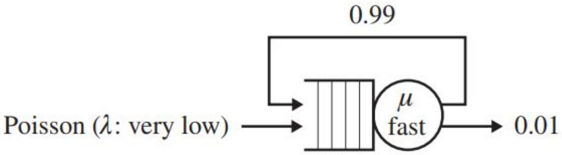


Figure: A Jackson network

Poisson Arrivals?

- $\lambda_i = r_i + \sum_j \lambda_j P_{ji}$
- Consider a single server network with properties:
  - Low arrival rate
  - High service rate
  - High P_{ii}



Violates independent increment property!

Balance Equations

- Rate of jobs leaving the state = Rate of jobs entering the state.

$$\pi_{n_1,n_2,\ldots,n_k} \cdot \left[\sum_{i=1}^{k} r_i + \sum_{i=1}^{k} \mu_i (1 - P_{ii})\right] =$$

$$\underbrace{\sum_{i=1}^{k} \pi_{n_1,\ldots,n_i-1,\ldots,n_k} \cdot r_i}_{\text{outside arrival}} + \underbrace{\sum_{i=1}^{k} \pi_{n_1,\ldots,n_i+1,\ldots,n_k} \cdot \mu_i P_{i,\text{out}}}_{\text{departure to outside}}$$

$$+ \underbrace{\sum_{i=1}^{k} \sum_{j \neq i} \pi_{n_1,\ldots,n_i-1,\ldots,n_j+1,\ldots,n_k} \cdot \mu_j P_{ji}}_{\text{internal transition from server } j \text{ to server } i,\ j \neq i}$$

Balance Equations (Contd.)

- In general, very hard to solve.
- Local Balance Approach
  - Break both sides of the equation into $k + 1$ matching components and equate each component.
  - Define A = Rate of leaving state due to an external arrival.
  - Define A' = Rate of entering state due to an external departure.
  - Define B_i = Rate of leaving state due to a depature from server i.
  - Define B_i' = Rate of entering state due to an arrival at server i.

Balance Equations (Contd.)

- Solving A = A'

$$\sum_{i=1}^{k} \pi_{n_1,\ldots,n_i,\ldots,n_k} \cdot r_i = \sum_{i=1}^{k} \pi_{n_1,\ldots,n_i+1,\ldots,n_k} \cdot \mu_i P_{i,\text{out}}$$

- Can we make a guess for $\pi_{n_1,\ldots,n_k}$ such that it satisfies the above equality?
- Observe that $\pi_{n_1,\ldots,n_i,\ldots,n_k}$ term in A and the $\pi_{n_1,\ldots,n_i+1,\ldots,n_k}$ term in A' only differ in n_i.
- Let us assume that

$$\pi_{n_1,\ldots,n_i,\ldots,n_k} \cdot c_i = \pi_{n_1,\ldots,n_i+1,\ldots,n_k}$$

- Substituting, we get

$$\sum_{i=1}^{k} r_i = \sum_{i=1}^{k} (c_i \cdot \mu_i) P_{i,\text{out}}$$

Balance Equations (Contd.)

- If we further assume $c_i \cdot \mu_i = \lambda_i$, we get

$$\sum_{i=1}^{k} r_i = \sum_{i=1}^{k} \lambda_i P_{i,\text{out}}$$

- By our assumption, $c_i = \frac{\lambda_i}{\mu_i} = \rho_i$. Substituting this we get

$$\pi_{n_1}, \ldots, \pi_{n_i}, \ldots, \pi_{n_k} \cdot \rho_i = \pi_{n_1}, \ldots, \pi_{n_i+1}, \ldots, \pi_{n_k}$$

- Repeating the argument yields the following equation

$$\pi_{n_1}, \ldots, \pi_{n_k} = C\rho_1^{n_1} \ldots \rho_k^{n_k}$$

Balance Equations (Contd.)

- We still need to check if our assumption yields a valid solution to $B_i = B_i'$

$$
\begin{aligned}
B_i &= B_i' \\
C\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_k^{n_k}\mu_i(1-P_{ii}) &= \sum_{j\neq i} C\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_k^{n_k}\left(\frac{\rho_j}{\rho_i}\right)\mu_j P_{ji} \\
&\quad + C\rho_1^{n_1}\rho_2^{n_2}\cdots\rho_k^{n_k}\left(\frac{1}{\rho_i}\right) r_i \\
\mu_i(1-P_{ii}) &= \sum_{j\neq i}\frac{\rho_j\mu_j P_{ji}}{\rho_i} + \frac{r_i}{\rho_i} \\
\rho_i\mu_i(1-P_{ii}) &= \sum_{j\neq i}\rho_j\mu_j P_{ji} + r_i \\
\lambda_i(1-P_{ii}) &= \sum_{j\neq i}\lambda_j P_{ji} + r_i
\end{aligned}
$$

Balance Equations (Contd.)

- Lastly, we need to find the normalizing constant C.

$$\sum_{n_1,\ldots,n_k} \pi_{n_1,\ldots,n_k} = 1$$

$$C \sum_{n_1,\ldots,n_k} \rho_1^{n_1} \cdots \rho_k^{n_k} = 1$$

$$C \left(\sum_{n_1} \rho_1^{n_1}\right) \left(\sum_{n_2} \rho_2^{n_2}\right) \cdots \left(\sum_{n_k} \rho_k^{n_k}\right) = 1$$

$$C \left(\frac{1}{1-\rho_1}\right) \left(\frac{1}{1-\rho_2}\right) \cdots \left(\frac{1}{1-\rho_k}\right) = 1$$

$$C = (1-\rho_1)(1-\rho_2)\cdots(1-\rho_k).$$

$$\boxed{\pi_{n_1,\ldots,n_k} = \rho_1^{n_1}(1-\rho_1)\rho_2^{n_2}(1-\rho_2)\cdots\rho_k^{n_k}(1-\rho_k)}$$

Balance Equations (Contd.)

- What about the jobs at a given server?

$$
\begin{aligned}
P\{n_1 \text{ jobs at server } 1\} &= \sum_{n_2,\ldots,n_k} \pi_{n_1,\ldots,n_k} \\
&= \sum_{n_2,\ldots,n_k} \rho_1^{n_1}(1-\rho_1)\rho_2^{n_2}(1-\rho_2)\cdots\rho_k^{n_k}(1-\rho_k) \\
&= \rho_1^{n_1}(1-\rho_1)
\end{aligned}
$$

$$
\boxed{P\{n_i \text{ jobs at server } i\} = \rho_i^{n_i}(1-\rho_i)}
$$

- Thus, all servers behave like M/M/1 queues in terms of their stationary queue length distributions, even though their arrival processes are not necessarily Poisson.

Summary

$$\lambda_i = r_i + \sum_j \lambda_j P_{ji}$$

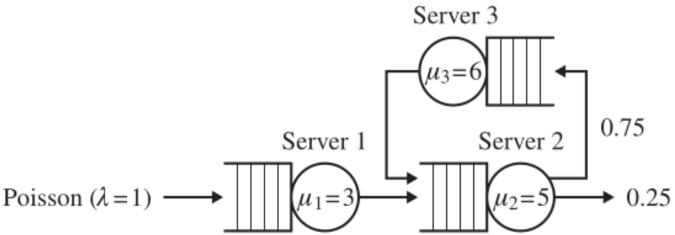
$$\rho_i = \frac{\lambda_i}{\mu_i}$$

$$\pi_{n_1,\ldots,n_k} = \rho_1^{n_1}(1-\rho_1)\rho_2^{n_2}(1-\rho_2)\cdots\rho_k^{n_k}(1-\rho_k)$$

$$P\{n_i \text{ jobs at server } i\} = \rho_i^{n_i}(1-\rho_i)$$

$$E\{N_i\} = \frac{\rho_i}{1-\rho_i}$$

Case 1


Server 3
$\mu_3=6$
0.75
Server 1
Server 2
Poisson ($\lambda=1$)
$\mu_1=3$
$\mu_2=5$
0.25

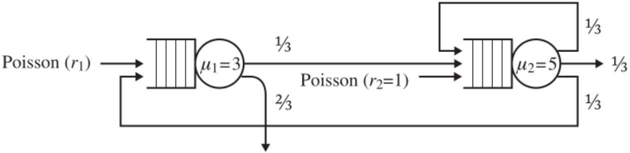
Case 1

- $\lambda_1 = 1$

  $\lambda_2 = \lambda_1 + \lambda_3$

  $\lambda_3 = 0.75\lambda_2$
- $\lambda_1 = 1, \lambda_2 = 4, \lambda_3 = 3$
- $\rho_1 = \frac{1}{3}, \rho_2 = \frac{4}{5}, \rho_3 = \frac{1}{2}$
- $E[N_1] = \frac{1}{2}, E[N_2] = 4, E[N_3] = 1$
- $E[T_1] = \frac{1}{2}, E[T_2] = 1, E[T_3] = \frac{1}{3}$

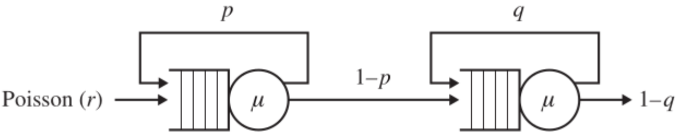
Case 2



Case 2

- $\lambda_1 = r_1 + \frac{1}{3}\lambda_2$
  $\lambda_2 = 1 + \frac{1}{3}\lambda_1 + \frac{1}{3}\lambda_2$
- $\lambda_1 = \frac{6r_1+3}{5}, \lambda_2 = \frac{3r_1+9}{5}$
- $\rho_1 = \frac{2r_1+1}{5}, \rho_2 = \frac{3r_1+9}{25}$
- For $\rho_1 < 1$, $r_1 < 2$. Let $r_1 = 1.5$.
- $\lambda_1 = 2.4, \lambda_2 = 2.7$
- $\rho_1 = 0.8, \rho_2 = 0.34$
- $E[N_1] = 4, E[N_2] = 1.17$
- $E[T_1] = 1.67, E[T_2] = 0.43$

Case 3



Case 3

- $\lambda_1 = r + p\lambda_1$
  $\lambda_2 = (1-p)\lambda_1 + q\lambda_2$
- $\lambda_1 = \frac{r}{1-p}, \lambda_2 = \frac{r}{1-q}$
- $\rho_1 = \frac{r}{\mu(1-p)}, \rho_2 = \frac{r}{\mu(1-q)}$
- $E[N_1] = \frac{r}{\mu(1-p)-r}, E[N_2] = \frac{r}{\mu(1-q)-r}$
- Let $r = 0.5, \mu = 3, p = 0.1, q = 0.2.$
- $\lambda_1 = \frac{5}{9}, \lambda_2 = \frac{5}{8}$
- $\rho_1 = \frac{5}{27}, \rho_2 = \frac{5}{24}$
- $E[N_1] = 0.23, E[N_2] = 0.26$
- $E[T_1] = 0.41, E[T_2] = 0.41$

References

[1] M. Harchol-Balter. *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*. Cambridge University Press, 2013.

[2] Damián Roca Marí. *High Level Queuing Architecture Model for High-end Processors*. 2014.

[3] Isi Mitrani. *Queueing networks*. Cambridge University Press, 1997.