Multi-class Queues and Stochastic Networks

LNMB - 2016

Richard J. Boucherie & Werner Scheinhardt

department of Applied Mathematics
University of Twente
Multi-class Queues and Stochastic Networks

Detailed content:
1. reversibility, stationarity, basic queues, output theorem, feedforward networks
2. partial balance, Jackson network, Kelly-Whittle network, arrival theorem
3. quasi-reversibility, customer types, BCMP networks, bandwidth sharing networks
4. blocking, aggregation, decomposition
5. loss networks, insensitivity via supplementary variables
6. sojourn time distribution in networks
7. MVA, AMVA, QNA
8. fluid queues, basic models
9. feedback fluid queues, networks of fluid queues
Multi-class Queues and Stochastic Networks
Today (lecture 4): implications partial balance

Nelson, sec 10.3.5—10.6

• Blocking
• Norton’s theorem: decomposition and aggregation
Last time on MQSN ….  

- Kelly Whittle network  
- Partial balance  
- Quasi reversibility
Theorem: The equilibrium distribution for the closed Jackson network containing \( N \) jobs is

\[
\pi(n) = B_N \prod_{j=1}^{J} \rho_j^n \quad \rho_j = \gamma_j / \mu_j \quad n \in S_N = \{n: \sum_j n_j = N\}
\]

and satisfies partial balance

\[
\sum_{k=1}^{J} \pi(n)q(n, n-e_j + e_k) = \sum_{k=1}^{J} \pi(n-e_j + e_k)q(n-e_j + e_k, n)
\]

traffic equations

\[
\sum_k \gamma_j p_{jk} = \sum_k \gamma_k p_{kj}
\]

\[
q(n, n-e_j + e_k) = \mu_j p_{jk}
\]

\[
q(n, n-e_j) = \mu_j p_{j0}
\]

\[
q(n, n+e_k) = \lambda_k
\]
closed network: equilibrium distribution

\[
\sum_{j=1}^{J} \sum_{k=1}^{J} \pi(n)q(n, n-e_j + e_k) = \sum_{j=1}^{J} \sum_{k=1}^{J} \pi(n-e_j + e_k)q(n-e_j + e_k, n)
\]

\[
\sum_{j=1}^{J} \sum_{k=1}^{J} \prod_{s=1}^{J} \rho_s^{n_s} \mu_j p_{jk} = \sum_{j=1}^{J} \sum_{k=1}^{J} \prod_{s=1}^{J} \rho_s^{n_s} \frac{\rho_k}{\rho_j} \mu_k p_{kj}
\]

\[
\sum_{j=1}^{J} \prod_{s=1}^{J} \rho_s^{n_s} \frac{1}{\rho_j} \sum_{k=1}^{J} \rho_j \mu_j p_{jk} = \sum_{j=1}^{J} \prod_{s=1}^{J} \rho_s^{n_s} \frac{1}{\rho_j} \sum_{k=1}^{J} \rho_k \mu_k p_{kj}
\]

\[
\sum_{j=1}^{J} \pi(n-e_j) \sum_{k=1}^{J} \gamma_j p_{jk} = \sum_{j=1}^{J} \pi(n-e_j) \sum_{k=1}^{J} \gamma_k p_{kj}
\]
Theorem: The equilibrium distribution for the open Jackson network is

\[ \pi(n) = B \prod_{j=1}^{J} \rho_j^{n_j} \rho_j = \frac{\gamma_j}{\mu_j} n \in \{ n : n_j \geq 0, j = 1, \ldots, J \} \]

and satisfies partial balance

\[ \sum_{k=0}^{J} \pi(n)q(n, n-e_j + e_k) = \sum_{k=0}^{J} \pi(n-e_j + e_k)q(n-e_j + e_k, n) \]

• Where (traffic equations)

\[ \gamma_j = \lambda_j + \sum_k \gamma_k p_{kj} \]

• Proof
\[
\sum_{k=0}^{J} \pi(n)q(n, n-e_j + e_k) = \sum_{k=0}^{J} \pi(n-e_j + e_k)q(n-e_j + e_k, n)
\]

\[
\sum_{k=1}^{J} \pi(n)q(n, n+e_k) = \sum_{k=1}^{J} \pi(n+e_k)q(n+e_k, n), \quad j = 0
\]

\[
\sum_{k=0}^{J} \pi(n)q(n, n-e_j + e_k)
\]

\[
= \pi(n-e_j)q(n-e_j, n) + \sum_{k=1}^{J} \pi(n-e_j + e_k)q(n-e_j + e_k, n), \quad j \neq 0
\]

\[
\pi(n)\sum_{k=0}^{J} \lambda_k = \pi(n)\sum_{k=1}^{J} \gamma_k p_{k0}, \quad j = 0
\]

\[
\pi(n-e_j)\sum_{k=0}^{J} \rho_j \mu_j p_{jk} = \pi(n-e_j)\left\{\lambda_j + \sum_{k=1}^{J} \rho_k \mu_k p_{kj}\right\}, \quad j \neq 0
\]
Partial balance

\[ \pi(n)q(n, n-e_j + e_k) = \pi(n-e_j + e_k)q(n-e_j + e_k, n) \]

\[ \sum_{k=0}^{J} \pi(n)q(n, n-e_j + e_k) = \sum_{k=0}^{J} \pi(n-e_j + e_k)q(n-e_j + e_k, n) \]

\[ \sum_{j=0}^{J} \sum_{k=0}^{J} \pi(n)q(n, n-e_j + e_k) = \sum_{j=0}^{J} \sum_{k=0}^{J} \pi(n-e_j + e_k)q(n-e_j + e_k, n) \]

• Detailed balance: Prob flow between each two states matches
• Partial balance: prob flow out of state n due to departure from queue j is balanced by prob flow into state n due to arrival to queue j, for each queue j, j=0,...,J
• Global balance: total prob flow out of state n equals total prob flow into state n
Kelly Whittle network

- State space $S$
- Transition rates

$$q(n, n - e_j + e_k) = \frac{\psi(n - e_j)}{\phi(n)} \mu_j p_{jk}$$

$$q(n, n - e_j) = \frac{\psi(n - e_j)}{\phi(n)} \mu_j p_{j0}$$

$$q(n, n + e_k) = \frac{\psi(n)}{\phi(n)} \lambda_k$$

- Where $\psi$ is non-negative function, and $\phi$ positive function
- notation

$$\lambda_k = \mu_0 p_{0k}$$
Kelly Whittle network

\[ q(n, n-e_j + e_k) = \frac{\psi(n-e_j)}{\phi(n)} \mu_j p_{jk} \]

\[ q(n, n-e_j) = \frac{\psi(n-e_j)}{\phi(n)} \mu_j p_{j0} \]

\[ q(n, n+e_k) = \frac{\psi(n)}{\phi(n)} \lambda_k \]

Theorem: The equilibrium distribution for the Kelly Whittle network is

\[ \pi(n) = B \phi(n) \prod_{j=1}^{J} \rho_j^{n_j} \quad \rho_j = \gamma_j / \mu_j \quad n \in S \]

where

\[ \gamma_j = \lambda_j + \sum_{k} \gamma_k p_{kj} \]

and \( \pi \) satisfies partial balance

\[ \sum_{k=0}^{J} \pi(n)q(n, n-e_j + e_k) = \sum_{k=0}^{J} \pi(n-e_j + e_k)q(n-e_j + e_k, n) \]
Insert equilibrium distribution and rates in partial balance

\[ \sum_{k=0}^{J} \pi(n)q(n, n-e_j + e_k) = \sum_{k=0}^{J} \pi(n-e_j + e_k)q(n-e_j + e_k, n) \]

\[ \sum_{k=0}^{J} \phi(n) \prod_{s=1}^{J} \rho_s^{n_s} \frac{\psi(n-e_j)}{\phi(n)} \mu_j p_{jk} \]

\[ = \sum_{k=0}^{J} \phi(n-e_j + e_k) \prod_{s=1}^{J} \rho_s^{n_s} \frac{\rho_k}{\rho_j} \frac{\psi(n-e_j)}{\phi(n-e_j + e_k)} \mu_k p_{kj} \]

\[ \psi(n-e_j) \prod_{s=1}^{J} \rho_s^{n_s-\delta_{sj}} \sum_{k=0}^{J} \rho_j \mu_j p_{jk} \]

\[ = \psi(n-e_j) \prod_{s=1}^{J} \rho_s^{n_s-\delta_{sj}} \sum_{k=0}^{J} \rho_k \mu_k p_{kj} \]

This is the beauty of partial balance!
Multi-class Queues and Stochastic Networks
Today (lecture 4): implications partial balance

Nelson, sec 10.3.5—10.6

• Blocking
• Norton’s theorem: decomposition and aggregation
Kelly Whittle network with state dependent routing

- State space $S$
- Transition rates

\[ q(n, n-e_j + e_k) = \frac{\psi(n-e_j)}{\varphi(n)} \mu_j p_{jk}(n-e_j) \]

\[ q(n, n-e_j) = \frac{\psi(n-e_j)}{\varphi(n)} \mu_j p_{j0}(n-e_j) \]

\[ q(n, n+e_k) = \frac{\psi(n)}{\varphi(n)} \mu_0 p_{0k}(n) \]

- Where $\psi$ is non-negative function, and $\varphi$ positive function
Kelly Whittle network

\[ q(n, n-e_j + e_k) = \frac{\psi(n-e_j)}{\varphi(n)} \mu_j p_{jk}(n-e_j) \]

\[ q(n, n-e_j) = \frac{\psi(n-e_j)}{\varphi(n)} \mu_j p_{j0}(n-e_j) \]

\[ q(n, n+e_k) = \frac{\psi(n)}{\varphi(n)} \mu_0 p_{0k}(n) \]

Theorem: The equilibrium distribution for the Kelly Whittle network is

\[ \pi(n) = B \phi(n) \prod_{j=1}^{J} \rho_j^{n_j} \]  \[ \rho_j = \frac{\gamma_j}{\mu_j} \quad n \in S \]

where

\[ \sum_{k} \gamma_j p_{jk}(n-e_j) = \sum_{k} \gamma_k p_{kj}(n-e_j) \]

and \( \pi \) satisfies partial balance

\[ \sum_{k=0}^{J} \pi(n) q(n, n-e_j + e_k) = \sum_{k=0}^{J} \pi(n-e_j + e_k) q(n-e_j + e_k, n) \]
Insert equilibrium distribution and rates in partial balance

\[
\sum_{k=0}^{J} \pi(n)q(n,n-e_j+e_k) = \sum_{k=0}^{J} \pi(n-e_j+e_k)q(n-e_j+e_k,n)
\]

\[
\sum_{k=0}^{J} \phi(n) \prod_{s=1}^{J} \rho_s^{n_s} \frac{\psi(n-e_j)}{\phi(n)} \mu_j p_{jk}(n-e_j)
\]

\[
= \sum_{k=0}^{J} \phi(n-e_j+e_k) \prod_{s=1}^{J} \rho_s^{n_s} \frac{\rho_k}{\rho_j} \frac{\psi(n-e_j)}{\phi(n-e_j+e_k)} \mu_k p_{kj}(n-e_j)
\]

\[
\psi(n-e_j) \prod_{s=1}^{J} \rho_s^{n_s-\delta_{sj}} \sum_{k=0}^{J} \rho_j u_j p_{jk}(n-e_j)
\]

\[
= \psi(n-e_j) \prod_{s=1}^{J} \rho_s^{n_s-\delta_{sj}} \sum_{k=0}^{J} \rho_k u_k p_{kj}(n-e_j)
\]

This is the beauty of partial balance!
Blocking in tandem networks of M/M/1 queues (1)

- M/M/1 queues, exponential service queue \( j, j=1,\ldots,J \)

- Transition rates
  
  \[ q(n,n - e_j + e_{j+1}) = \mu_j \]
  
  \[ q(n,n - e_j) = \mu_j \]
  
  \[ q(n,n + e_1) = \lambda \]

- Traffic equations
  
  \[ \rho_j \mu_j = \rho_{j-1} \mu_{j-1} \quad , \quad j = 2,\ldots,J \]
  
  \[ \rho_1 \mu_1 = \lambda \]

- Solution
  
  \[ \rho_j = \frac{\lambda}{\mu_j} \quad , \quad j = 1,\ldots,J \]
Blocking in tandem networks of simple queues (2)

- M/M/1 queues, exponential service queue \( j, j=1,\ldots,J \)

- Equilibrium distribution

\[
\pi(n) = \prod_{j=1}^{J} (1 - \rho_j) \rho_j^{n_j} \quad n \in S = \{n : n \geq 0\}
\]

- Partial balance

\[
\pi(n)q(n, n - e_j + e_{j+1}) = \pi(n - e_j + e_{j-1})q(n - e_j + e_{j-1}, n) \quad j = 1, \ldots, J - 1
\]

\[
\pi(n)q(n, n + e_1) = \pi(n + e_j)q(n + e_j, n) \quad j = 0
\]

\[
\pi(n)q(n, n - e_J) = \pi(n - e_j + e_{J-1})q(n - e_j + e_{J-1}, n) \quad j = J
\]

- PICTURE J=2
Blocking in tandem networks of simple queues (3)

- Simple queues, exponential service queue \( j, j=1,\ldots,J \)
- Suppose queue 2 has capacity constraint: \( n_2 < N_2 \)
- Transition rates
  
  \[
  q(n, n - e_j + e_{j+1}) = \mu_j, \ j = 2,\ldots,J
  \]
  
  \[
  q(n, n - e_1 + e_2) = \mu_1 1(n_2 < N_2)
  \]
  
  \[
  q(n, n - e_j) = \mu_j
  \]
  
  \[
  q(n, n + e_1) = \lambda
  \]
- Partial balance?
  \[
  \pi(n) = \prod_{j=1}^{J} (1 - \rho_j) \rho_j^{n_j} \quad n \in S = \{n : n_2 < N_2\}
  \]
- PICTURE J=2
- Stop protocol, repeat protocol, jump-over protocol
Blocking in tandem networks of simple queues (4)

- Stop protocol: when queue is blocked stop all other queues including the “external queue” (= arrival process)
- Recirculate protocol: when queue is blocked recirculate departures from all other queues including the external queue
- Jump-over protocol: when queue is blocked let customers arriving to that queue jump over the queue and continue their route (not the same as letting the service rate increase to infinity)

- Yield product form due to preservation of partial balance
Multi-class Queues and Stochastic Networks

Today (lecture 4): implications partial balance

Nelson, sec 10.3.5—10.6

• Blocking

• Norton’s theorem: decomposition and aggregation
Quasi-reversibility

• Multi class queueing network, class $c \in C$

• A queue is quasi-reversible if its state $x(t)$ is a stationary Markov process with the property that the state of the queue at time $t_0$, $x(t_0)$, is independent of
  (i) arrival times of class $c$ customers subsequent to time $t_0$
  (ii) departure times of class $c$ customers prior to time $t_0$.

• Theorem
  If a queue is QR then
  (i) arrival times of class $c$ customers form independent Poisson processes
  (ii) departure times of class $c$ customers form independent Poisson processes.
Quasi-reversibility

- Multi class queueing network, class \( c \in C \)
- \( S(c,x) \) set of states in which queue contains one more class \( c \) than in state \( x \)
- Arrival rate class \( c \) customer independent of state \( x \)
  \[ \alpha(c) = \sum_{x' \in S(c,x)} q(x, x') \]
- Departure rate class \( c \) customer independent of state \( x \)
  \[ \pi(x)q^r(x, x') = \pi(x')q(x', x) \]
  \[ \alpha(c) = \sum_{x' \in S(c,x)} q^r(x, x') \]
- Operational definition QR
  \[ \alpha(c) = \sum_{x' \in S(c,x)} q(x, x') = \sum_{x' \in S(c,x)} q^r(x', x) \]
Quasi-reversibility

• Multi class queueing network, class $c \in C$
• $S(c,x)$ set of states in which queue contains one more class $c$ than in state $x$
• Arrival rate class $c$ customer independent of state $x$
• Departure rate class $c$ customer independent of state $x$

\[ \alpha(c) = \sum_{x' \in S(c,x)} q(x, x') \]

\[ \pi(x)q^r(x, x') = \pi(x')q(x', x) \]

\[ \alpha(c) = \sum_{x' \in S(c,x)} q^r(x, x') \]

• Characterise QR, combine

\[ \sum_{x' \in S(c,x)} \pi(x)q(x, x') = \sum_{x' \in S(c,x)} \pi(x')q(x', x) \]

• A form of partial balance
Quasi-reversibility and reversibility

- Birth death process with state dependent arrivals is reversible but not quasi reversible
- Ex 10.14 quasi-reversible is not reversible

- Note QR is sum of detailed balance for a subset of the transitions:

\[
\sum_{x' \in S(c,x)} \pi(x)q(x,x') = \sum_{x' \in S(c,x)} \pi(x')q(x',x)
\]

- A form of partial balance
Network of quasi reversible nodes

- Construct network by multiplying rates for individual queues
- Transition rates
  - Arrival of type $i$ causes queue $k = r(i,1)$ to change at
    \[ q_k(x_k, x_k') \quad x_k' \in S_k(i,1,x_k) \]
  - Departure type $i$ from queue $j = r(i,S(i))$
    \[ q_j(x_j, x_j') \quad x_j \in S_j(i,S(i),x_j') \]
- Routing
  \[ q_j(x_j, x_j') \frac{q_k(x_k, x_k')}{\sum_{x' \in S_k(i,s+1,x_k)} q_k(x_k, x')} = q_j(x_j, x_j') \frac{q_k(x_k, x_k')}{\alpha_k(i,s+1)} \]
  \[ x_k' \in S_k(i,s+1,x_k) \quad x_j \in S_j(i,s,x_j') \]
- Internal $q_j(x_j, x_j')$
Multi-class Queues and Stochastic Networks
Today (lecture 4): implications partial balance

Nelson, sec 10.3.5—10.6

• Blocking

• Norton’s theorem: decomposition and aggregation
Norton’s theorem

State $n= (n_1, \ldots, n_N)$
Clusters $C_1, \ldots, C_R$
Stations $i \in C_r$
State of cluster $n^{(r)} = (n_i, i \in C_r)$
Global state $N_r = \sum_{i \in C_r} n_i$
$N = (N_1, \ldots, N_R)$

$$\sum_{j \in C_r} q_{0j}^{(r)}(n^{(r)}) = 1.$$  

$$q(n, n-e_i + e_j) = \begin{cases} 
q_{i0}^{(r)}(n^{(r)})q_{R}^{rs}(\bar{N})q_{0j}^{(s)}(n^{(s)}), \\
q_{ij}^{(r)}(n^{(r)})\mu^{(r)}(\bar{N}) + q_{i0}^{(r)}(n^{(r)})q_{R}^{rs}(\bar{N})q_{0j}^{(r)}(n^{(r)} - e_i^{(r)}), & i, j \in C_r, \\
q_{i0}^{(r)}(n^{(r)})q_{R}^{r0}(\bar{N}), \\
q_{R}^{0s}(\bar{N})q_{0j}^{(s)}(n^{(s)}), & i \in C_r, j \in C_0, \\
q_{R}^{rs}(\bar{N}) = \mu^{(r)}(\bar{N})p^{rs}(\bar{N}) & i \in C_0, j \in C_s, 
\end{cases}$$  

(2.2)
Norton’s theorem (2)

\[
q(n, n - e_i + e_j) = \begin{cases} 
q_{i0}^{(r)}(n^{(r)})q^{rs}_{R}(\bar{N})q_{0j}^{(s)}(n^{(s)}), & i \in C_r, j \in C_s, \\
q_{ij}^{(r)}(n^{(r)})\mu^{(r)}(\bar{N}) + q_{i0}^{(r)}(n^{(r)})q^{rr}_{R}(\bar{N})q_{0j}^{(r)}(n^{(r)} - e_i^{(r)}), & i, j \in C_r, \\
q_{i0}^{(r)}(n^{(r)})q^{r0}_{R}(\bar{N}), & i \in C_r, j \in C_0, \\
q^{0s}_{R}(\bar{N})q_{0j}^{(s)}(n^{(s)}), & i \in C_0, j \in C_s, 
\end{cases}
\]

Stationary distributions: \(\sum_{i,j \in C_r \cup \{0\}} R \sum_{i,j \in C_r \cup \{0\}} \{\pi^{(r)}(n^{(r)})q^{ij}_{R}(n^{(r)}) - \pi^{(r)}(n^{(r)} - e_i^{(r)} + e_j^{(r)})q^{ij}_{R}(n^{(r)} - e_i^{(r)} + e_j^{(r)})\} = 0\)

\(\sum_{s,r=0} R \sum_{s,r=0} \{\pi_R(\bar{N})q^{rs}_{R}(\bar{N}) - \pi_R(\bar{N} - \bar{E}_r + \bar{E}_s)q^{sr}_{R}(\bar{N} - \bar{E}_r + \bar{E}_s)\} = 0.\)

Quasi-reversibility of clusters \(\sum_{i \in C_r} \{\pi^{(r)}(n^{(r)})q^{0j}_{0j}(n^{(r)}) - \pi^{(r)}(n^{(r)} + e_j^{(r)})q^{0j}_{0j}(n^{(r)} + e_j^{(r)})\} = 0\)

Partial balance global process \(\sum_{s=0} R \sum_{s=0} \{\pi_R(\bar{N})q^{rs}_{R}(\bar{N}) - \pi_R(\bar{N} - \bar{E}_r + \bar{E}_s)q^{sr}_{R}(\bar{N} - \bar{E}_r + \bar{E}_s)\} = 0.\)

Theorem: \(\pi(n) = B\pi_R(\bar{N}) \prod_{r=1}^R \pi^{(r)}(n^{(r)})\)
Norton’s theorem (3)

\[ q(n, n - e_i + e_j) = \begin{cases} 
q_{i0}^{(r)}(n^{(r)})q_{R}^{s}(\bar{N})q_{0j}^{(s)}(n^{(s)}), & i \in C_r, \ j \in C_s, \\
q_{ij}^{(r)}(n^{(r)})\mu^{(r)}(\bar{N}) + q_{i0}^{(r)}(n^{(r)})q_{R}^{sr}(\bar{N})q_{0j}^{(r)}(n^{(r)} - e_i^{(r)}), & i, j \in C_r, \\
q_{i0}^{(r)}(n^{(r)})q_{R}^{0i}(\bar{N}), & i \in C_r, \ j \in C_0, \\
q_{R}^{0s}(\bar{N})q_{0j}^{(s)}(n^{(s)}), & i \in C_0, \ j \in C_s, 
\end{cases} \]

Global process:

\[ M^{(r)}(\bar{N}) = \mu^{(r)}(\bar{N})\frac{\pi^{(r)}(N_r - 1)}{\pi^{(r)}(N_r)}, \quad r = 0, \ldots, R \]

\[ Q(\bar{N}, \bar{N} - E_r + E_s) = M^{(r)}(\bar{N})p^{rs}(\bar{N}) \]

First order equivalent:

\[ \Pi(\bar{N}) = \sum_{n: \sum_{i \in C_r} n_i = N_r, \ r = 1, \ldots, R} \pi(n). \]

\[ \Pi(\bar{N})Q(\bar{N}, \bar{N} - E_r + E_s) = \sum_{n: \sum_{i \in C_r} n_i = N_r, \ r = 1, \ldots, R} \sum_{i \in C_r, \ j \in C_s} \pi(n)q(n, n - e_i + e_j) \]

Theorem: global process is first order equivalent, and

\[ \Pi(\bar{N}) = B_R \pi_R(\bar{N}) \prod_{r=1}^{R} \pi^{(r)}(N_r). \quad \pi(n|\bar{N}) = \prod_{r=1}^{R} \pi^{(r)}(n^{(r)}|N_r) \]
Multi-class Queues and Stochastic Networks
Today (lecture 4): implications partial balance

Nelson, sec 10.3.5—10.6

• Blocking
• Norton’s theorem: decomposition and aggregation