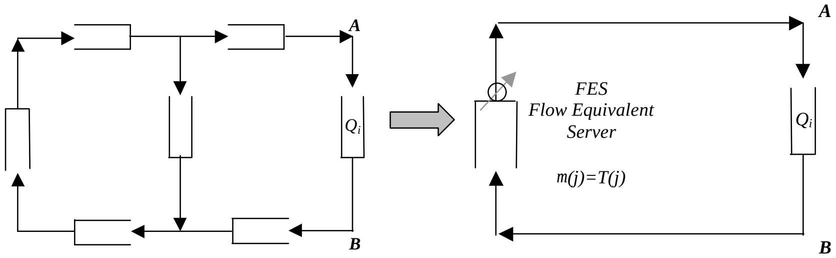
Queueing Networks - II

Norton's Theorem for Closed Queueing Networks



Original Closed Queueing Network (before reduction)

Equivalent Queueing Network (after reduction using Norton's Theorem)

Theorem is called "Norton's Theorem" in analogy with Norton's Theorem for reducing electronic circuits to a simpler equivalent form

• May be used for an arbitrarily connected closed queueing network (say with *K* queues and *M* circulating jobs) where the product-form solution holds (i.e. BCMP networks)

• All queues except those in a *designated sub-network* may be replaced by a single *Flow Equivalent Server (FES)*

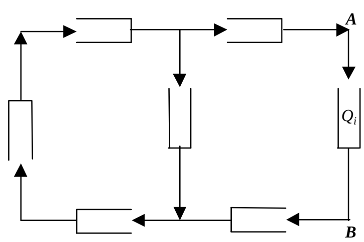
• With the proper choice of FES, the behaviour of the equivalent network will be exactly the same as that of the original network

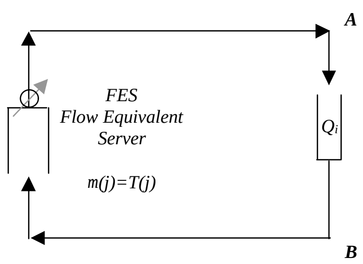
• May also extended to networks with multiple classes of customers

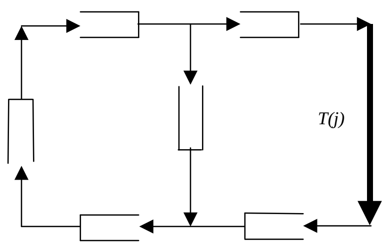
• Does not hold for non-BCMP networks except as an approximation

For non-BCMP networks, put the non-product form type queues in the designated network. Get the FES corresponding to the product form components and use this with Norton's Theorem to study the overall network either through an approximation or using simulations.

The Flow-Equivalent Server (FES)





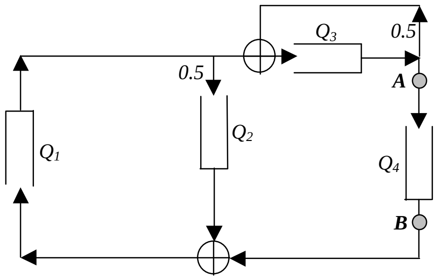


• *Short* the queues in the designated network by making their service times zero

• With this, calculate the throughput *T(j)* of the network with *j* jobs where *j=1,……, M*

• *m(j)=T(j)* for the required FES

Example (Application of Norton's Theorem)



• All the queues are single server queues with infinite capacity

• Number of Jobs $M=5$

• The service rates are -

$m_1 = 1 \quad m_3 = 0.5$

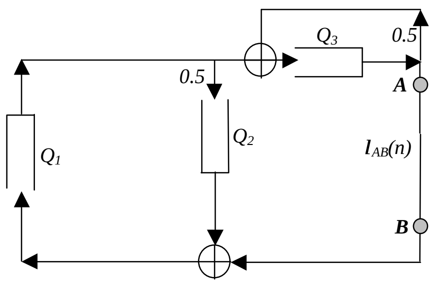
$m_2 = 1 \quad m_4 = 0.5$

• Designated Sub-network: Queue Q_4

• Flow Equivalent Server required for the rest of the network

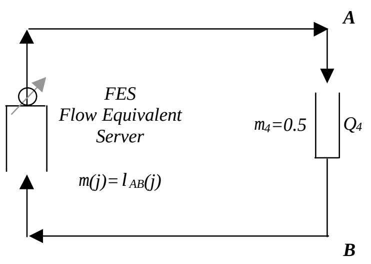
To find the FES to be used short the network between A and B (i.e. by making the service time of Q_4 to be zero)

Let $l_{AB}(n)$ be the throughput through the shorted path when there are n jobs in the network



We can obtain that -

$$l_{AB}(1) = 0.143$$
$$l_{AB}(2) = 0.2$$
$$l_{AB}(3) = 0.226$$
$$l_{AB}(4) = 0.238$$
$$l_{AB}(5) = 0.244$$



$m(1) = 0.143$
$m(2) = 0.2$
$m(3) = 0.226$
$m(4) = 0.238$
$m(5) = 0.244$

Solving this (analysis or simulation), we can get the performance measures for Q_4 to be -

$$N=0.796 \quad W=3.402 \quad \text{and} \quad l=0.234$$

Mixed Queueing Networks — *Multiple job classes, network open for some job classes and closed for the others*

May be analyzed using an approximate algorithm by Lazowska et al. for the case where the service times are exponentially distributed and the external arrivals are Poisson in nature.

- Calculate utilizations for each of the open classes first
- Use these to compute net utilization of each node because of all the open classes
- Eliminate the open classes and analyze the network for only the closed classes by appropriately inflating their service demands
- Get performance of the open classes by accounting for the actual mean queue lengths for the closed classes at each node

The QNA Approach (Ward Whitt)

Approximate Analysis of Open Networks of *GI/G/m* Queues

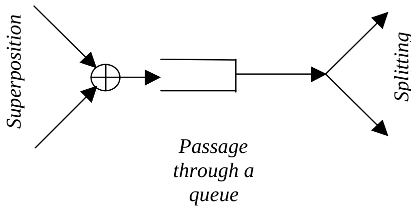
• A *2-moment* approach which requires the means and SQVs of both the inter-arrival times (for each external arrival process) and the means and SQVs of the service times at each queue

• Incorporates the concept of a *multiplication factor* applied to the output of a node to create multiple jobs for each outgoing job

• Whitt's QNA used with *single class jobs with probabilistic routing* or *multiple job classes with deterministic class dependent routing*

• Approach has been extended to other types of queueing networks by other authors

Basic Modelling Assumptions of QNA



• Each of these operations maintains the *GI* nature of the job stream

• Immediate feedback is removed from around a queue

For an arrival process following a GI stream, the inter-arrival times are i.i.d random variables following any general distribution.

Process is Poisson when the inter-arrival times have the exponential distribution

Moreover, ………………………….

•The queueing network is an *open* one

•The individual queues may have one or more servers and has infinite waiting space so that there is no blocking or loss anywhere in the system

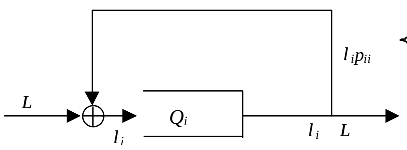
•The service discipline is FCFS in nature

•The approach is described here for a single class of customers with probabilistic routing. (Whitt's original paper also discusses using this for multi-class systems where the routing is deterministic in nature.

•The routing matrix is static and does not change over time.

QNA also incorporates a *multiplication factor* that can be applied at a node. For example, a job sent by a node may get split into several jobs independently on departure from the queue.

Removal of Immediate Feedback



Original Queue with Immediate Feedback

Original Parameters

Mean Service Time = $t_{i,U}$

SQV of Service Time = $c_{si,U}{}^2$

$p_{ij,U}$ $j=1,.....,K$ with $p_{ii,U}>0$

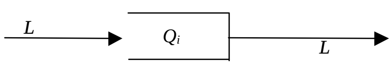
Modified Parameters

$$t_{i,M} = \frac{t_{i,U}}{1-p_{ii,U}}$$

$$c^2_{si,M} = p_{ii,U} + (1-p_{ii,U})c^2_{si,U}$$

$$p_{ij,M} = \frac{p_{ij,U}}{1-p_{ii,U}} \qquad j \neq i$$

$$= 0 \qquad j = i$$



Equivalent Queue without Immediate Feedback

Removal of Immediate Feedback

• Results are obtained by analyzing the network with the *modified queues* (i.e. ones from which, immediate feedback, if any have been removed)

• The results obtained for a *modified queue* are then changed back as follows -

Actual Mean Queueing Delay

$$W_{qi,U} = (1 - p_{ii,U})W_{qi,M}$$

Actual Mean Arrival Rate to the queue

$$l_{i,U} = \frac{l_{i,M}}{(1 - p_{ii,U})}$$

The Solution Approach of QNA

1. Remove immediate feedback, if any, in the queue

2. Use flow balance equations and approximations (*Sec. 6.2.2*) to solve for the mean and SQV of the total flow entering each queue of the network

Mean Flow Rate entering Q_i

$$l_i = l_{0i} + \sum_{j=1}^{K} l_j n_j p_{ji} \qquad j=1,\ldots\ldots, K$$

l_{0i} = Average job arrival rate from outside the network to Q_i

n_j = Multiplication Factor for jobs leaving Q_j

The Solution Approach of QNA……..*continued*……..

The approximate equivalent SQV c_{ai} of Q_i (i.e.of the interarrival times of jobs entering Q_i) may be found by solving the set of simultaneous equations given in Sec. 6.2.2. (This is more complex than solving flow balance!)

3. The queues in the network can now be treated independently using the mean and SQV of the flow entering each queue and the mean and service time of the queue. This is done for each queue to obtain the queueing performance parameters for each queue and therefore of the overall queueing network. (See *Sec. 6.2.3*)

This gives the mean waiting time in queue and its SQV for each queue in the "modified" network (after feedback removal)

