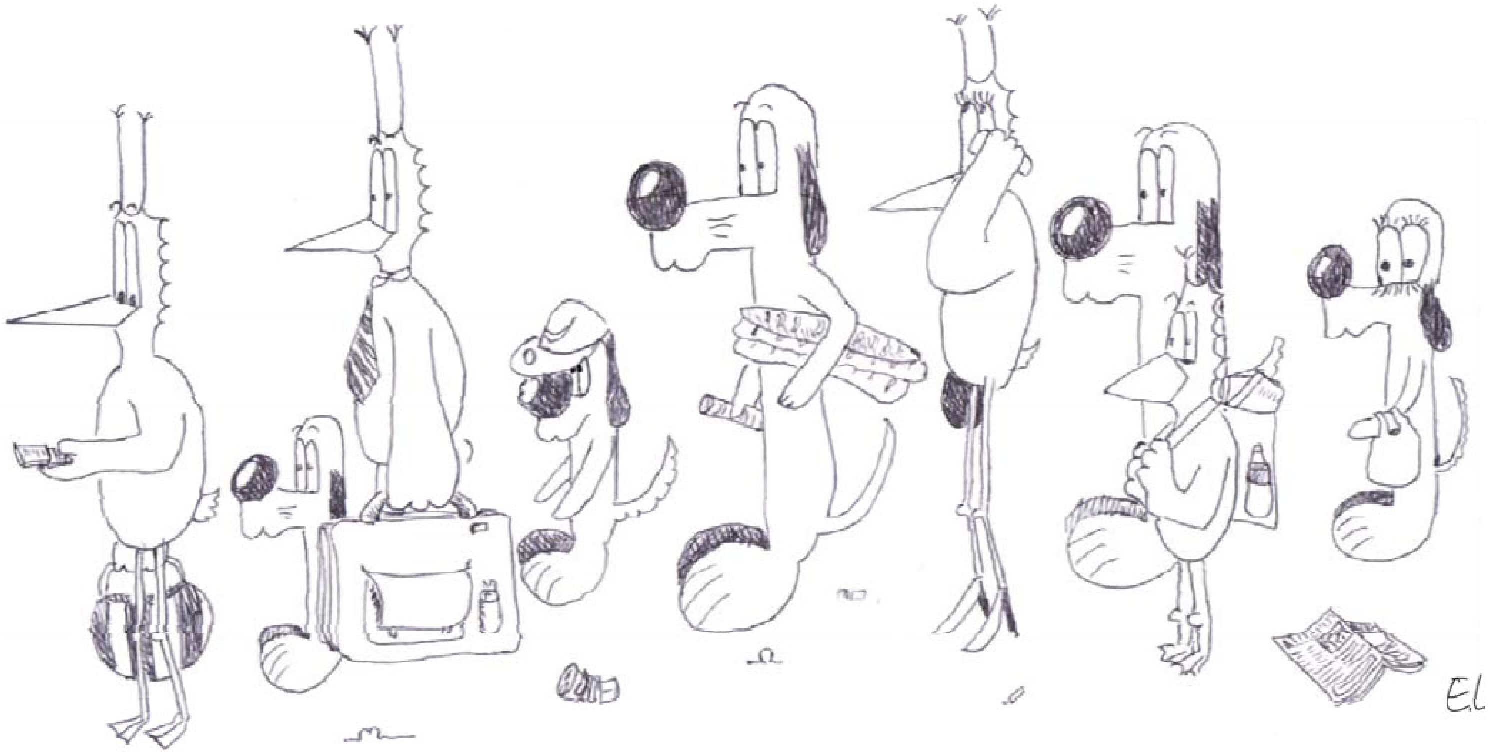
Queuing Networks



Jean-Yves Le Boudec

Networks of Queues
Stability

- Queuing networks are frequently used models
- The stability issue may, in general, be a hard one

- Necessary condition for stability (**Natural Condition**)

  server utilization < 1

  at every queue

Instability Examples

IIE Transactions (1997) **29**, 213–219

Simulation studies of multiclass queueing networks

J. BANKS and J. G. DAI

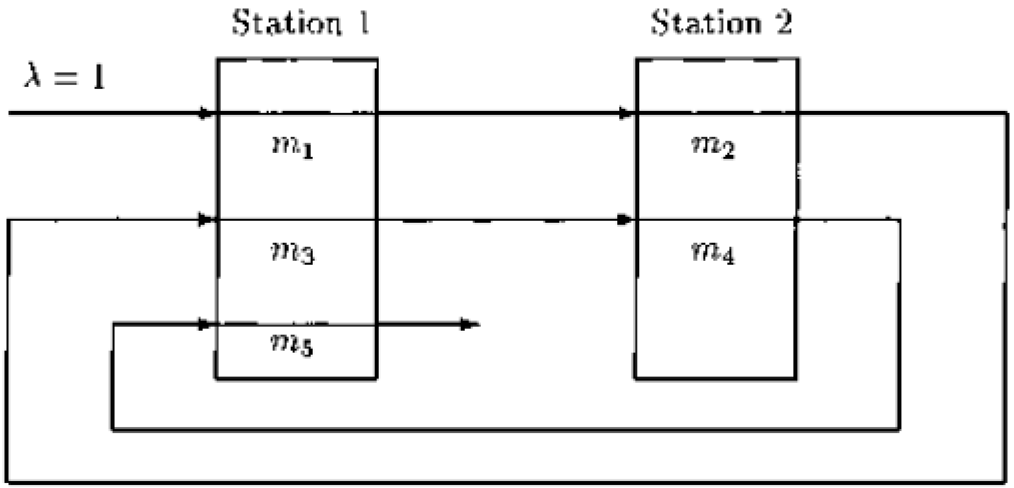


Fig. 1. An example of reentrant lines.

- Poisson arrivals ; jobs go through stations 1,2,1,2,1 then leave
- A job arrives as type 1, then becomes 2, then 3 etc
- Exponential, independent service times with mean m_i
- Priority scheduling
  - Station 1 : 5 > 3 >1
  - Station 2: 2 > 4

- Q: What is the natural stability condition ?
- A: $\lambda\,(m_1 + m_3 + m_5\,) < 1$

  $\lambda\,(m_2 + m_4) < 1$

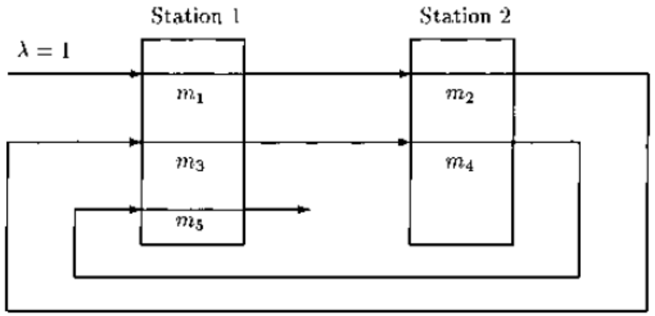


Fig. 1. An example of reentrant lines.

- $\lambda = 1$
  $m_1 = m_3 = m_4 = 0.1$
  $m_2 = m_5 = 0.6$
- Utilization factors
  - Station 1: 0.8
  - Station 2: 0.7
- Network is unstable !

- If $\lambda\,(m_1 + \ldots + m_5) < 1$ network is stable; why?

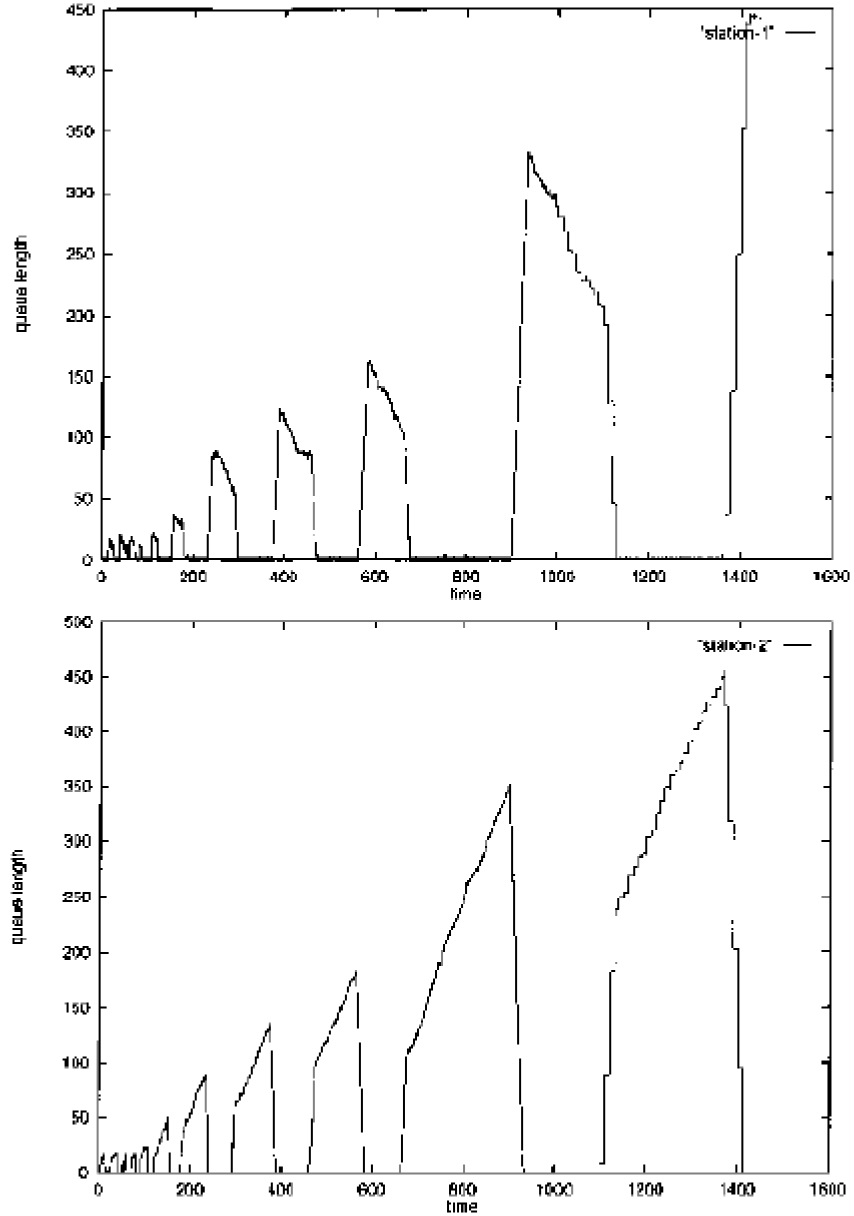


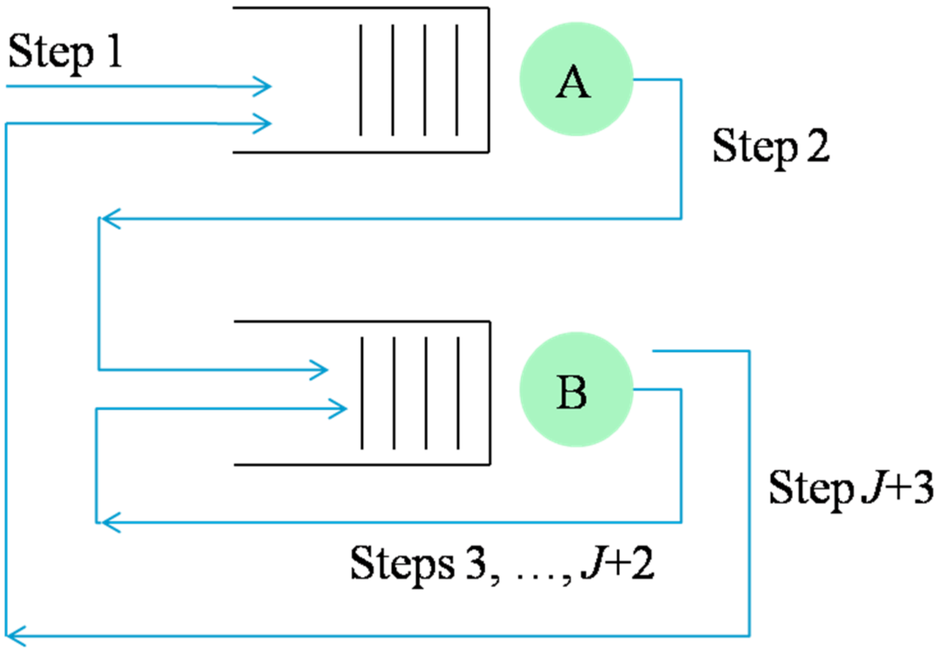
Fig. 2. Job size plots at stations 1 and 2.

Bramson's Example 1: A Simple FIFO Network

The Annals of Applied Probability
1994, Vol. 4, No. 2, 414–431

INSTABILITY OF FIFO QUEUEING NETWORKS

BY MAURY BRAMSON[1]



- Poisson arrivals; jobs go through stations A, B,B…,B, A then leave
- Exponential, independent service times
  - Steps 2 and last: mean is L
  - Other steps: mean is S

- Q: What is the natural stability condition ?
- A: $\lambda (L + S) < 1$
  $\lambda ((J\text{-}1)S + L) < 1$
- Bramson showed: may be unstable whereas natural stability condition holds

Bramson's Example 2
A FIFO Network with Arbitrarily Small Utilization Factor

The Annals of Applied Probability
1994, Vol. 4, No. 3, 693–718

INSTABILITY OF FIFO QUEUEING NETWORKS WITH QUICK SERVICE TIMES[1]

BY MAURY BRAMSON

$$1 \to 2 \to \cdots \to 2 \to 3 \to \cdots \to 3 \to \cdots \to m \to \cdots \to m$$
$$1 \to 1 \to 2 \to \cdots \to 2 \to 3 \to \cdots \to 3 \to \cdots \to m \to \cdots \to m \to 1$$

S *S* *L* *S* *S* *L* *S* *S* *L* *S* *S* *L*

- Utilization factor at every station ≤ 4 λ *S*
- Network is unstable for
  S ≤ 0.01
  L ≤ S^8
  m = floor(-2 (log *L*)/*L*)

- *m* queues
- 2 types of customers
- λ = 0.5 each type
- routing as shown,
  … = 7 visits
- FIFO
- Exponential service times, with mean as shown

Take Home Message

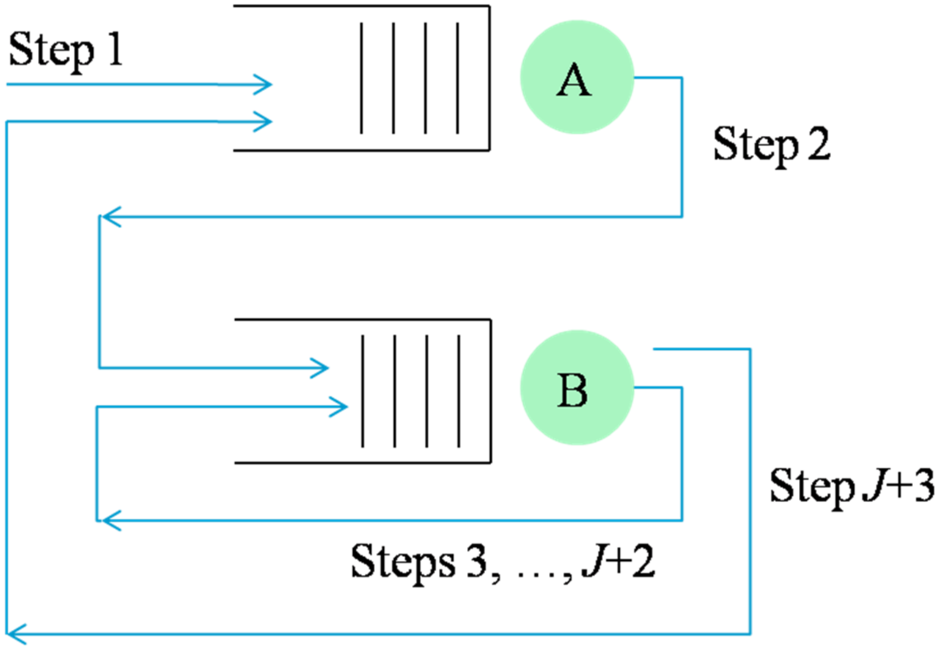
- The natural stability condition is necessary but may not be sufficient
- There is a class of networks where this never happens. Product Form Queuing Networks

Product Form Networks

- Customers have a *class* attribute
- Customers visit stations according to *Markov Routing*

  routing matrix $Q = \left(q_{c,c'}^{s,s'}\right)_{s,s',c,c'}$

- External arrivals, if any, are Poisson



2 Stations
Class = step, J+3 classes

Can you reduce the number of classes ?

Chains

- Customers can switch *class*, but remain in the same *chain*

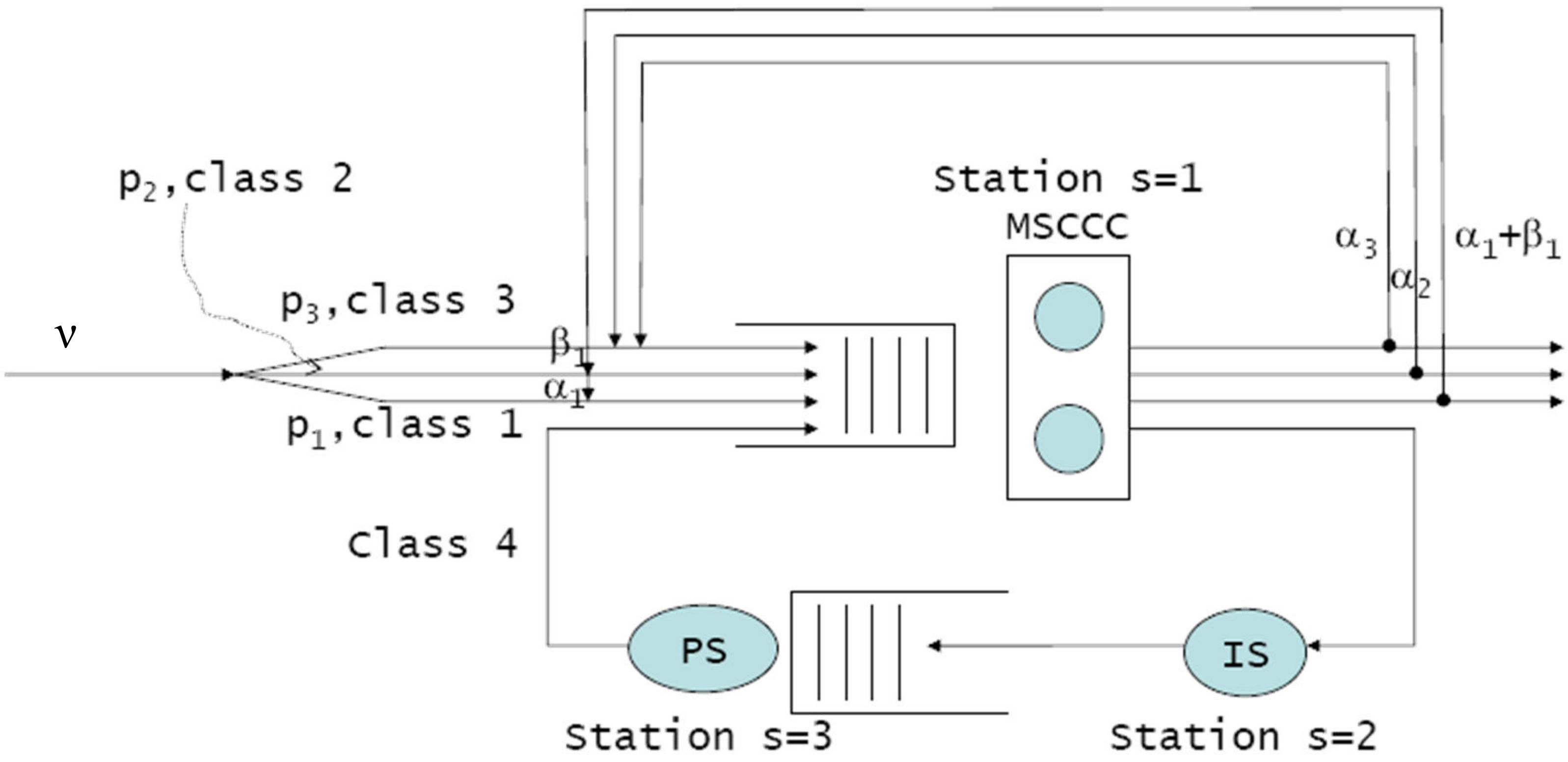


Figure 8.11: A Simple Product Form queuing network with 2 chains of customers, representing a machine with dual core processor. Chain 1 consists of classes 1, 2 and 3. Chain 2 consists of class 4.

Chains may be open or closed

- Open chain = with Poisson arrivals. Customers must eventually leave
- Closed chain: no arrival, no departure; number of customers is constant

- Closed network has only closed chains
- Open network has only open chains
- Mixed network may have both

3 Stations
4 classes
1 open chain
1 closed chain

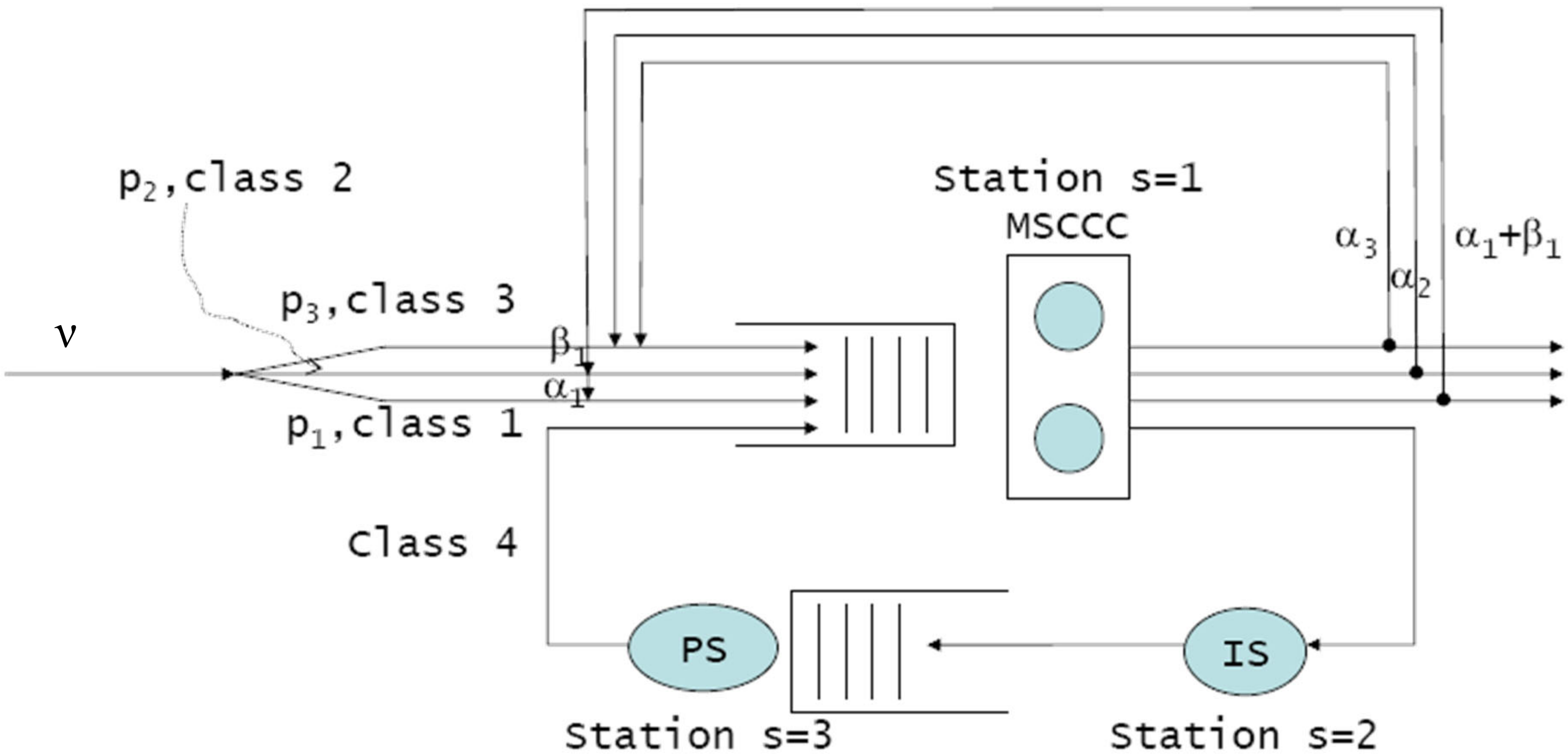


Figure 8.11: A Simple Product Form queuing network with 2 chains of customers, representing a machine with dual core processor. Chain 1 consists of classes 1, 2 and 3. Chain 2 consists of class 4.

Bramson's Example 2
A FIFO Network with Arbitrarily Small Utilization Factor

The Annals of Applied Probability
1994, Vol. 4, No. 3, 693–718

INSTABILITY OF FIFO QUEUEING NETWORKS WITH QUICK SERVICE TIMES[1]

By Maury Bramson

$$1 \to 2 \to \cdots \to 2 \to 3 \to \cdots \to 3 \to \cdots \to m \to \cdots \to m$$

$$1 \to 1 \to 2 \to \cdots \to 2 \to 3 \to \cdots \to 3 \to \cdots \to m \to \cdots \to m \to 1$$

$S \quad S \quad L \quad S \quad S \quad L \quad S \quad S \quad L \quad S \quad S \quad L$

2 Stations
Many classes
2 open chains
Network is open

Visit Rates

We define the numbers θ_c^s (*visit rates*) as one solution to

$$\theta_c^s = \sum_{s',c'} \theta_{c'}^{s'} q_{c',c}^{s',s} + \nu_c^s \tag{8.24}$$

If the network is open, this solution is unique and θ_c^s can be interpreted[12] as the number of arrivals per time unit of class-c customers at station s. If c belongs to a closed chain, θ_c^s is determined only up to one multiplicative constant per chain. We assume that the array $(\theta_c^s)_{s,c}$ is one non identically zero, non negative solution of Eq.(8.24).

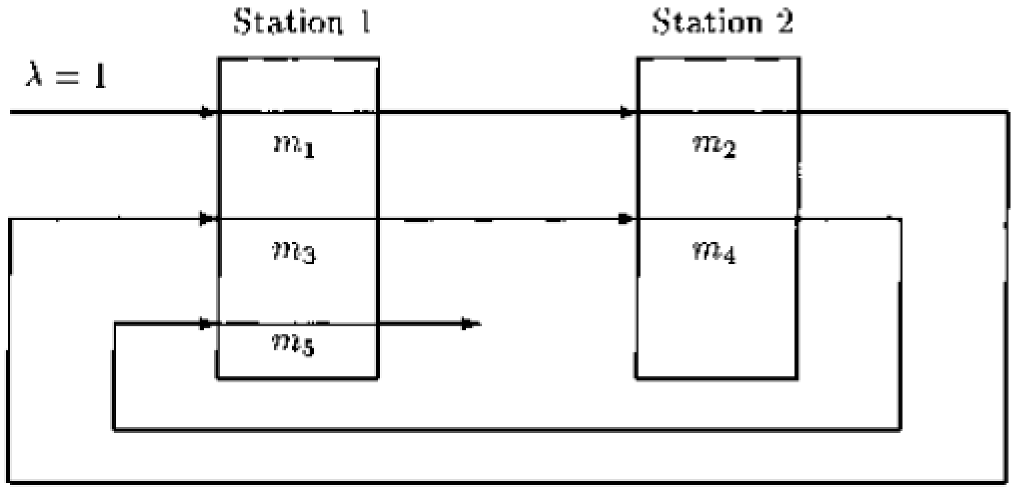


Fig. 1. An example of reentrant lines.

2 Stations
5 classes
1 chain
Network is open

Visit rates
$\theta^1_1 = \theta^1_3 = \theta^1_5 = \theta^2_2 = \theta^2_4 = \lambda$
$\theta^s_c = 0$ otherwise

$$
\begin{array}{llllllllll}
\text{Class 1:} & \theta_1^1 & = & \nu\frac{p_1}{1-\alpha_1}; & \theta_1^2 & = & 0; & \theta_1^3 & = & 0; \\
\text{Class 2:} & \theta_2^1 & = & \nu\left(p_2+\beta_1\frac{p_1}{1-\alpha_1}\right); & \theta_2^2 & = & 0; & \theta_2^3 & = & 0; \\
\text{Class 3:} & \theta_3^1 & = & \nu\frac{1}{1-\alpha_3}\left(p_3+\alpha_2 p_2+\alpha_2\beta_1\frac{p_1}{1-\alpha_1}\right); & \theta_3^2 & = & 0; & \theta_3^3 & = & 0; \\
\text{Class 4:} & \theta_4^1 & = & 1; & \theta_4^2 & = & 1; & \theta_4^3 & = & 1.
\end{array}
$$

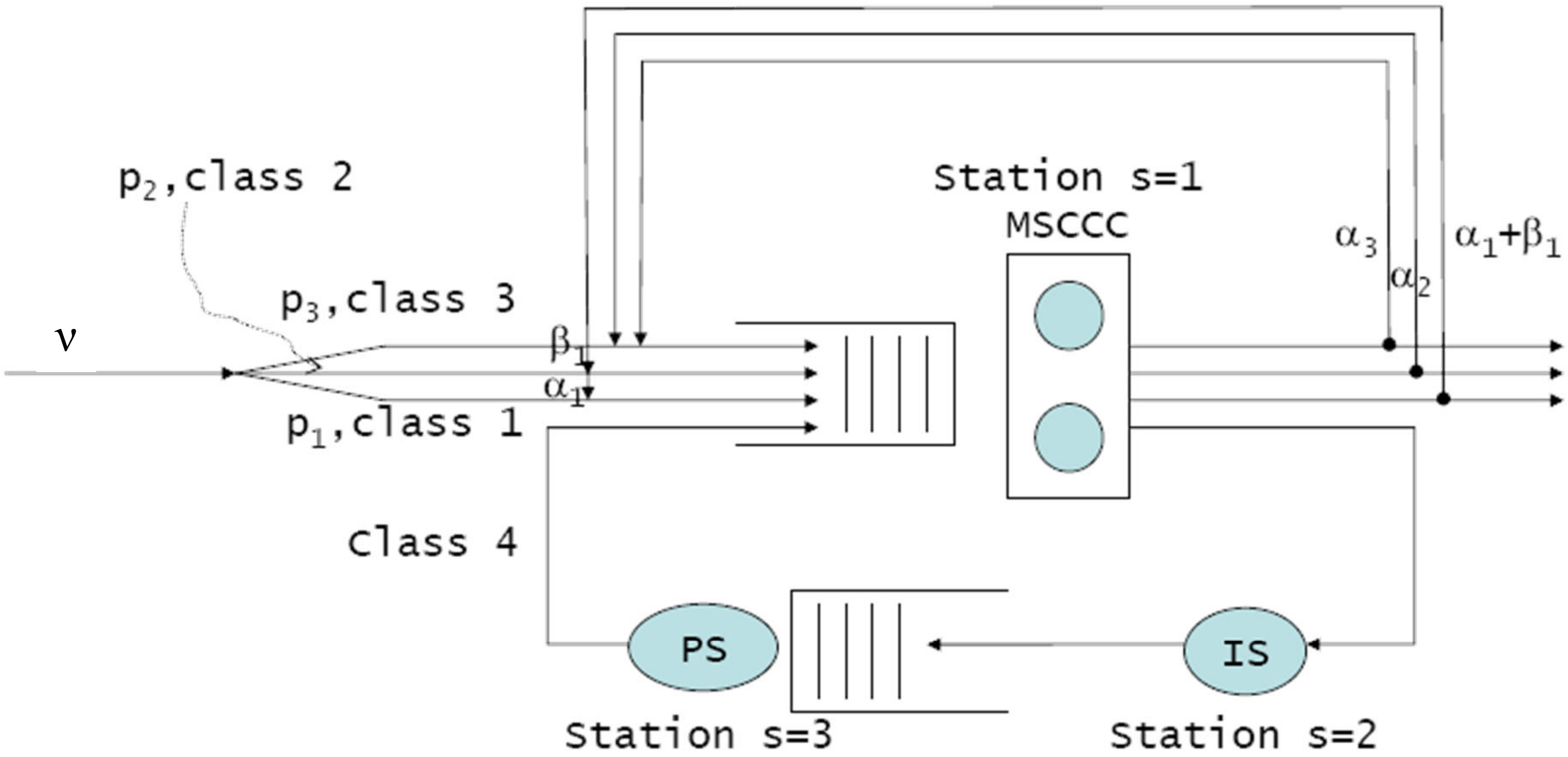


Figure 8.11: A Simple Product Form queuing network with 2 chains of customers, representing a machine with dual core processor. Chain 1 consists of classes 1, 2 and 3. Chain 2 consists of class 4.

Constraints on Stations

- Stations must belong to a restricted catalog of stations
- See Section 8.4 for full description
- We will give commonly used examples
- Example 1: *Global Processor Sharing*
  - One server
  - Rate of server is shared equally among all customers present
  - Service requirements for customers of class c are drawn iid from a distribution which depends on the class (and the station)
- Example 2: *Delay*
  - Infinite number of servers
  - Service requirements for customers of class c are drawn iid from a distribution which depends on the class (and the station)
  - No queuing, service time = service requirement = residence time

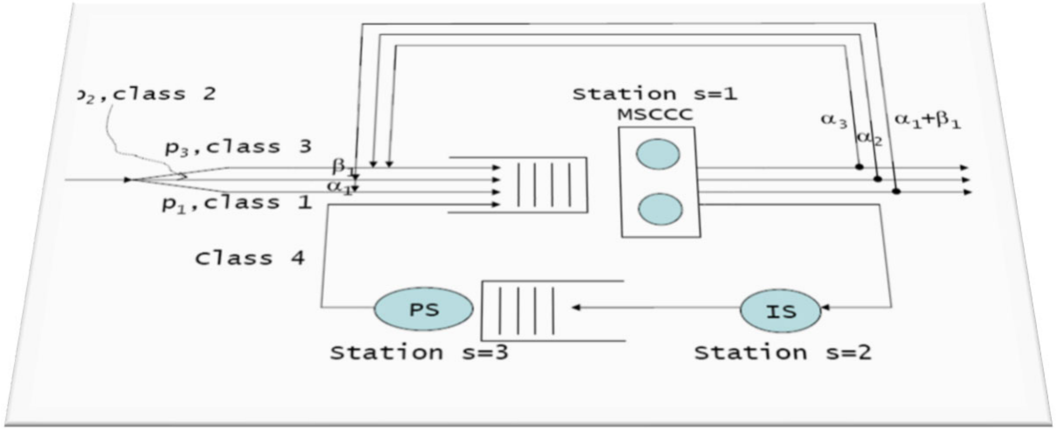
- Example 3 : *FIFO with B servers*
  - *B* servers
  - FIFO queueing
  - Service requirements for customers of class *c* are drawn iid from an *exponential* distribution, *independent of the class* (but may depend on the station)

- Example of Category 2 (MSCCC station): *MSCCC with B servers*
  - *B* servers
  - FIFO queueing with constraints

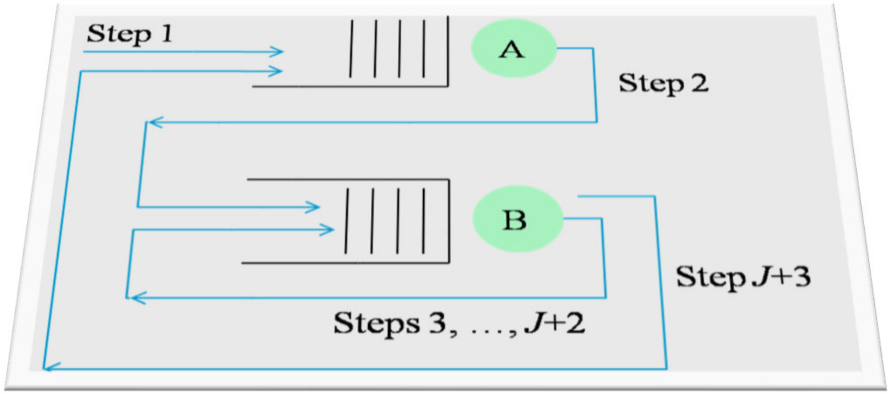
    At most one customer of each class is allowed in service
  - Service requirements for customers of class *c* are drawn iid from an *exponential* distribution, *independent of the class* (but may depend on the station)
- Examples 1 and 2 are *insensitive* (service time can be anything)
  Examples 3 and 4 are not (service time must be exponential, same for all)

- Say which network satisfies the hypotheses for product form

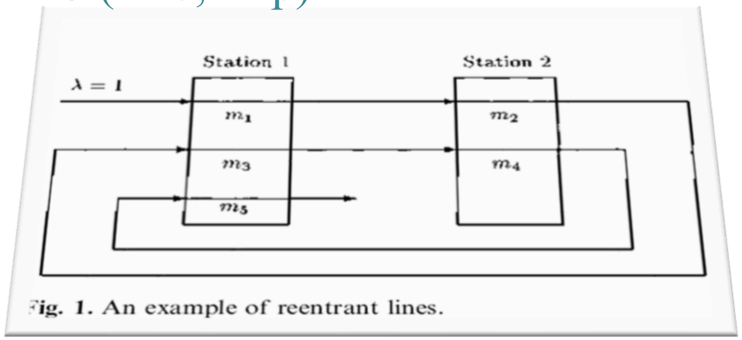
A



B (FIFO, Exp)



C (Prio, Exp)



The Product Form Theorem

- If a network satisfies the « Product Form » conditions given earlier
  - The stationary distrib of numbers of customers can be written explicitly
  - It is a product of terms, where each term depends only on the station
  - Efficient algorithms exist to compute performance metrics for even very large networks
  - For PS and Delay stations, service time distribution does not matter other than through its mean *(insensitivity)*
  - The natural *stability* condition holds

8.3.3 The Processor Sharing Queue, M/GI/1/PS

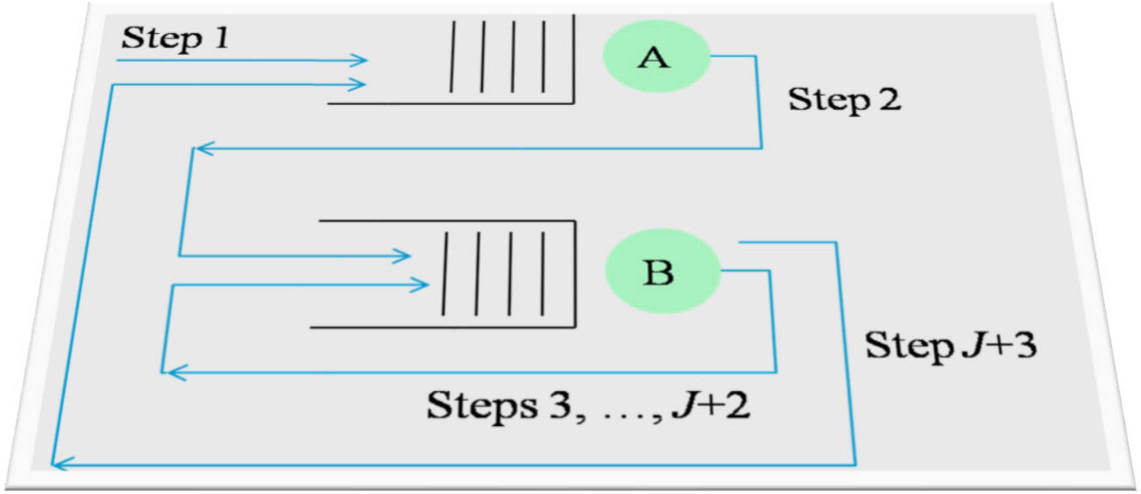
$$\mathbb{P}(N(t) = k) = (1 - \rho)\rho^k$$

M/M/B Queue

For more specific system, one can say more. A frequently used system is the M/M/B queue, i.e. the system with Poisson arrivals, B servers, exponential service times and FIFO discipline. The system can be studied directly by solving for the stationary probability. Here when $\rho < 1$ there is a unique stationary regime, which is also reached asymptotically when we start from arbitrary initial conditions; for $\rho \geq 1$ there is no stationary regime.

When $\rho < 1$ the stationary probability is given by

$$\mathbb{P}(N(t) = k) = \begin{cases} \eta \frac{(B\rho)^k}{k!} \text{ if } 0 \leq k \leq B \\ \eta \frac{B^B \rho^k}{B!} \text{ if } k > B \end{cases} \tag{8.21}$$



QUESTION 8.11.5. *In Section 8.4 we mention the existence of a network in [16] which is unstable with utilization factor less than* 1. *Can it be a product-form multi-class queuing network ? Why or why not ?* [25]

[25] It cannot be a product-form multi-class queuing network because they are stable when utilization is less than 1. It violates the assumptions because of FIFO stations with class-dependent service rates.