Research topics

Samuli Aalto

Background

- Ph.D. Thesis: “Studies in Queueing Theory”, University of Helsinki, Finland, 1998
- Supervisors: Prof. E. Nummelin and Ph.D. T. Lehtonen
  - [1] S. Aalto (1997) Optimal control of batch service queues with Poisson arrivals and finite service capacity, Rep Dep Mathematics 166, University of Helsinki, 1997
  - [2] S. Aalto (1998) Optimal control of batch service queues with compound Poisson arrivals and finite service capacity, Math Meth Oper Res 48, 3, 317-335
  - [3] S. Aalto (1998) Characterization of the output rate process for a Markovian storage model, J Appl Prob 35, 1, 184-199
  - [4] S. Aalto (1998) Output of a multiplexer loaded by heterogeneous on-off sources, Stoch Models 14, 4, 963-1005

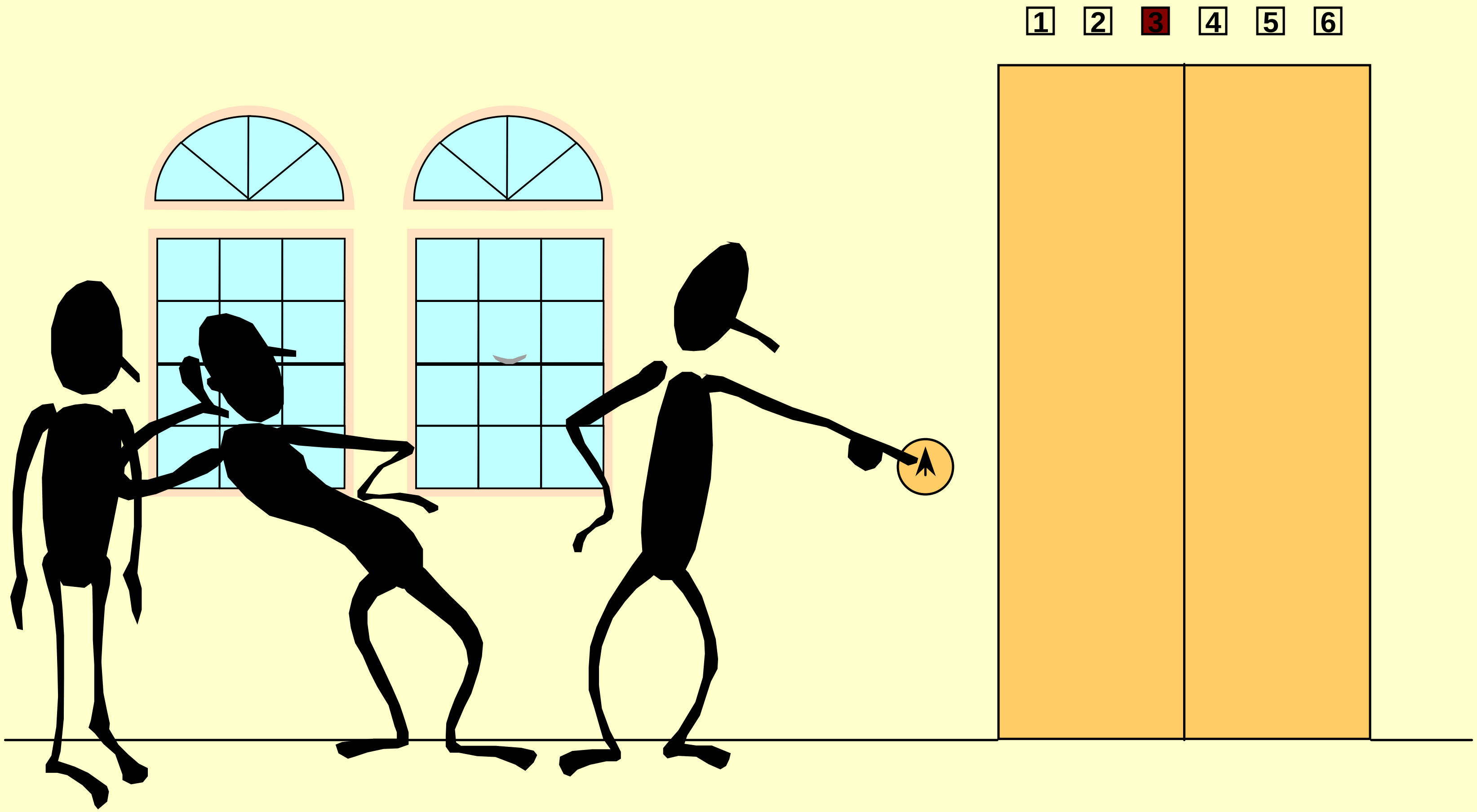
Contents

- Optimal control of batch service queues
- Characterization of the output process for some fluid queues
- Blocking probabilities in dynamic multicast networks
- Other topics

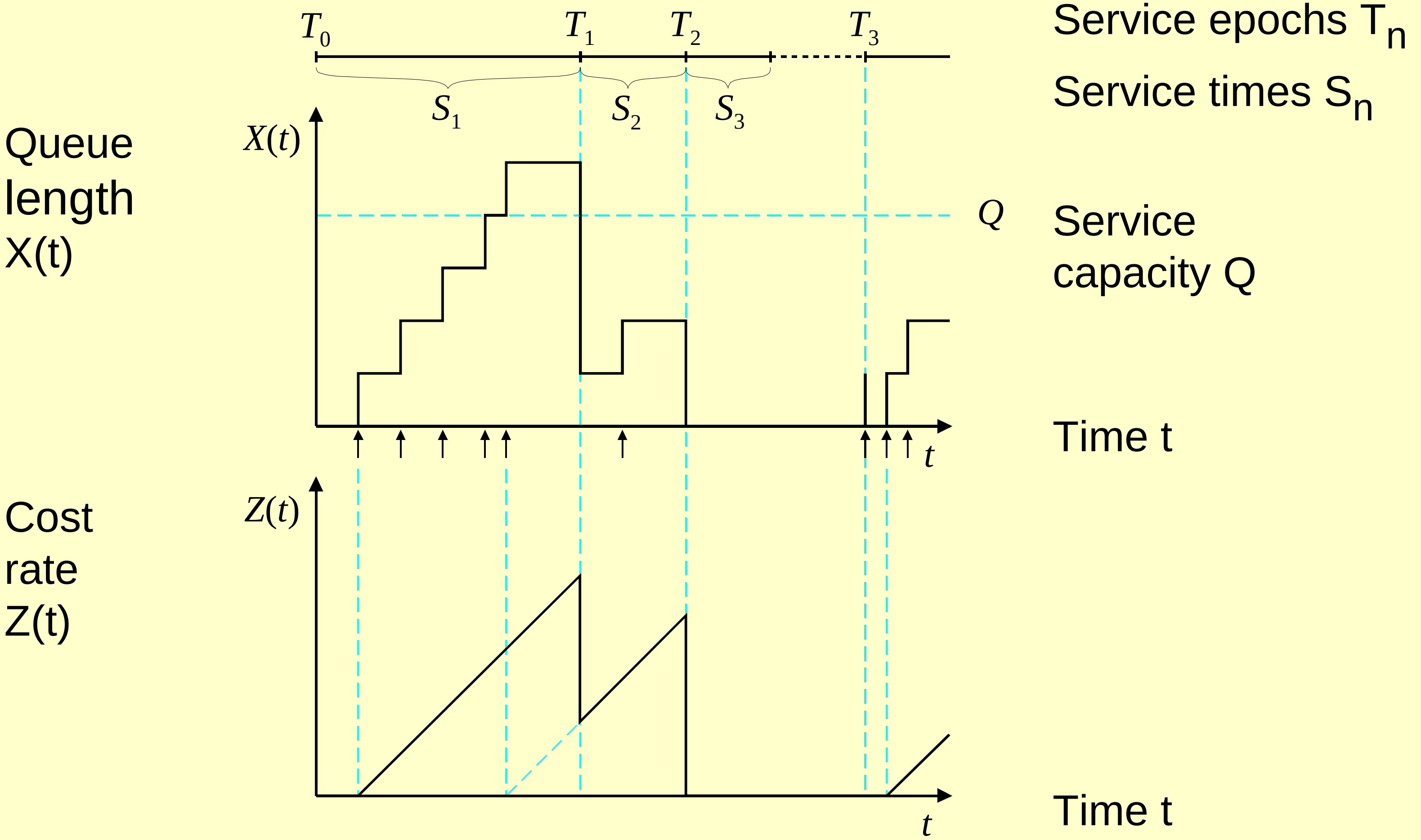
Batch service queue

- In an ordinary queue
  - customers are served individually
- In a batch service queue
  - customers are served in batches of varying size
- Additional parameter needed:
  - Q = service capacity = max nr of customers served in a batch

Application



Evolution


Queue
length
X(t)
Cost
rate
Z(t)
T0
T1
T2
T3
S1
S2
S3
X(t)
Q
t
Z(t)
t
Service epochs Tn
Service times Sn
Service
capacity Q
Time t
Time t

Control problem

- **Given**
  - arrival process $A(t)$ and
  - service times S_n
- **Determine**
  - service epochs T_n
  - service batches B_n
- Operating policy $\pi = ((T_n),(B_n))$
  - should be admissible
- Usual operating policy:
  - after a service completion, a new service is initiated as soon as

$$X(t) \geq 1$$

  - a service batch includes as many customers as possible

Cost structure

- **Holding costs**: Z(t)
  - described by the **cost rate** process Z(t)
  - cost rate depends on
    - the nr of waiting customers, X(t), and
    - the times they have been waiting, $W_1(t),\ldots,W_{X(t)}(t)$
  - called **linear** if

$$Z(t) = h(X(t))$$

- **Serving costs**: $K + cB_n$
  - K per each service batch
  - c per each customer served

Queueing models considered

- M/G(Q)/1
  - Poisson arrivals
  - generally distributed IID service times
  - single server with service capacity Q
- M^X/G(Q)/1
  - compound Poisson arrivals
  - generally distributed IID service times
  - single server with service capacity Q

Known results

	Infinite service capacity $Q = \infty$	**Finite** service capacity $Q < \infty$
Linear holding costs $z = h(x)$	**Case A**: - Deb & Serfozo (1973) - Deb (1984)	**Case B**: - Deb & Serfozo (1973)
General holding costs $z = h(x,w)$	**Case C**: - Weiss (1979) - Weiss & Pliska (1982)	**Case D**

A	B

Cases A and B: linear holding costs

- Deb & Serfozo (1973)
  - Poisson arrivals
  - average cost & discounted cost cases
- Deb (1984)
  - compound Poisson arrivals
  - discounted cost case only
- Result:
  - h(x) is “uniformly increasing”
    => a **queue length threshold policy** is optimal
- Note: Optimal threshold is always less or equal to Q

A	B

Queue length threshold policies

- **Queue length threshold policy** π_x with threshold x:
  - after a service completion, a new service is initiated as soon as

$$X(t) \geq x$$

  - a service batch includes as many customers as possible
- Note: the usual operating policy = π_1

Case C:
general holding costs & infinite service capacity

C	

- Weiss (1979),
  Weiss & Pliska (1982)
  - compound Poisson arrivals
  - average cost case only
- Result:
  - Z(t) is increasing (without limits when service is postponed forever)
    => a **cost rate threshold policy** is optimal

C	

Cost rate threshold policies

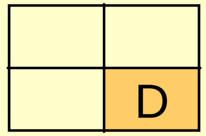
- **Cost rate threshold policy** $\pi(z)$ with threshold z:
  - after a service completion, a new service is initiated as soon as

$$Z(t) \geq z$$

  - a service batch includes as many customers as possible
    - infinite capacity => all waiting customers

New results

	Infinite service capacity $Q = \infty$	**Finite** service capacity $Q < \infty$
Linear holding costs $z = h(x)$	**Case A**: - Deb & Serfozo (1973) - Deb (1984)	**Case B**: - Deb & Serfozo (1973) - Deb (1984)
General holding costs $z = h(x,w)$	**Case C**: - Weiss (1979) - Weiss & Pliska (1982)	**Case D**: - Aalto (1997) [1] - Aalto (1998) [2]



Case D:
General holding costs & finite service capacity (1)

- Aalto (1997) [1]
  - Poisson arrivals
  - average cost & discounted cost cases
- Result:
  - FIFO queueing discipline,
  - consistent holding costs and
  - no serving costs included (K = c = 0)

    => a **cost rate threshold Q-policy** is optimal

D

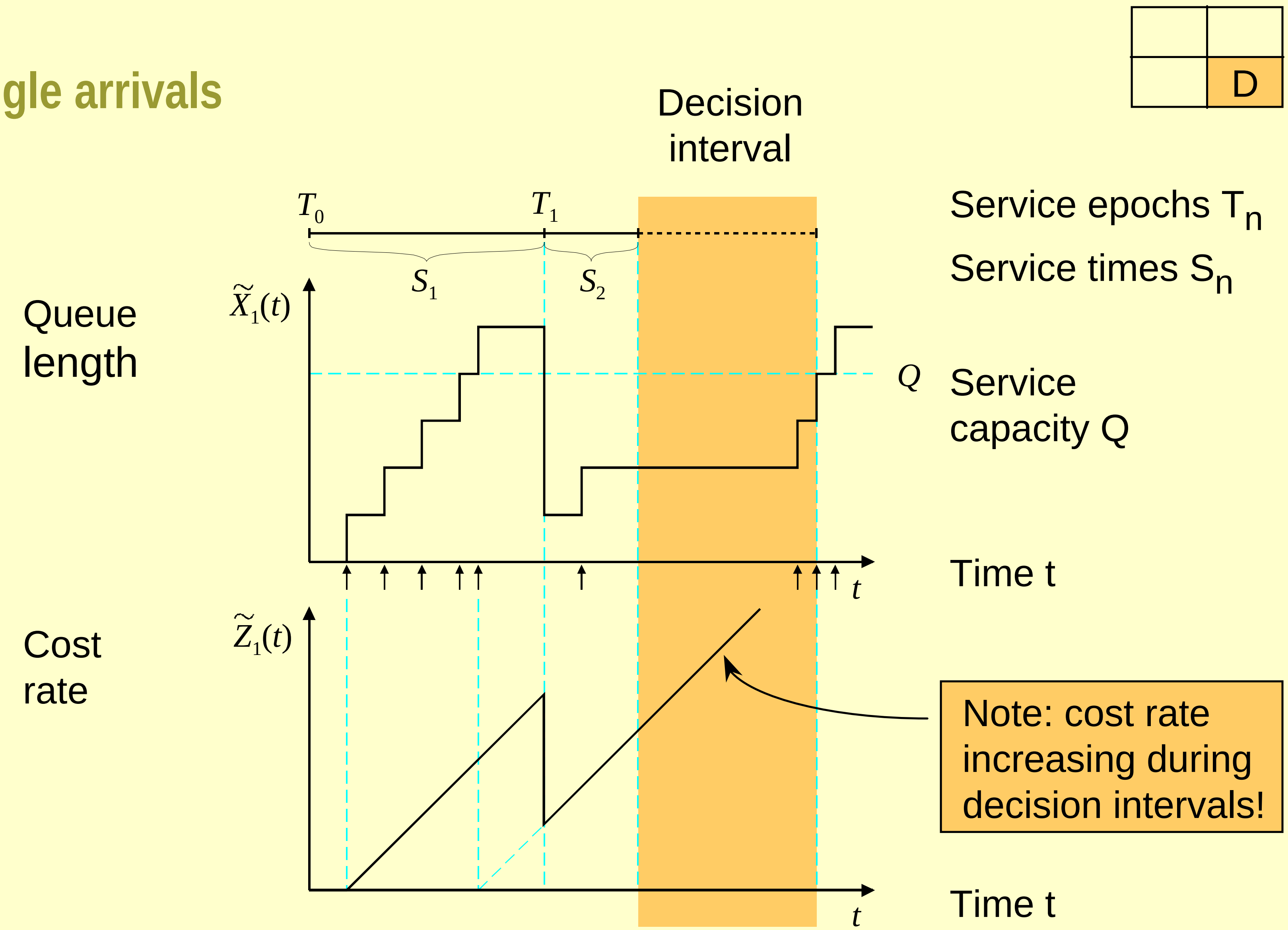
Cost rate threshold Q-policies

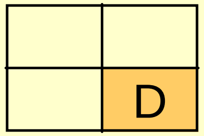
- **Cost rate threshold Q-policy** $\pi_Q(z)$ with threshold z:
  - after a service completion, a new service is initiated as soon as

$$Z(t) \geq z \quad \text{or} \quad X(t) \geq Q$$

  - a service batch includes as many customers as possible
    - finite capacity => min{X(t),Q}

Single arrivals





Case D: General holding costs & finite service capacity (2)

- Aalto (1998) [2]
  - **compound** Poisson arrivals
  - discounted cost case only
- Result:
  - FIFO queueing discipline
  - consistent holding costs,
  - no serving costs included (K = c = 0) and
  - bounded arrival batches

  => a **general threshold Q-policy** is optimal

D

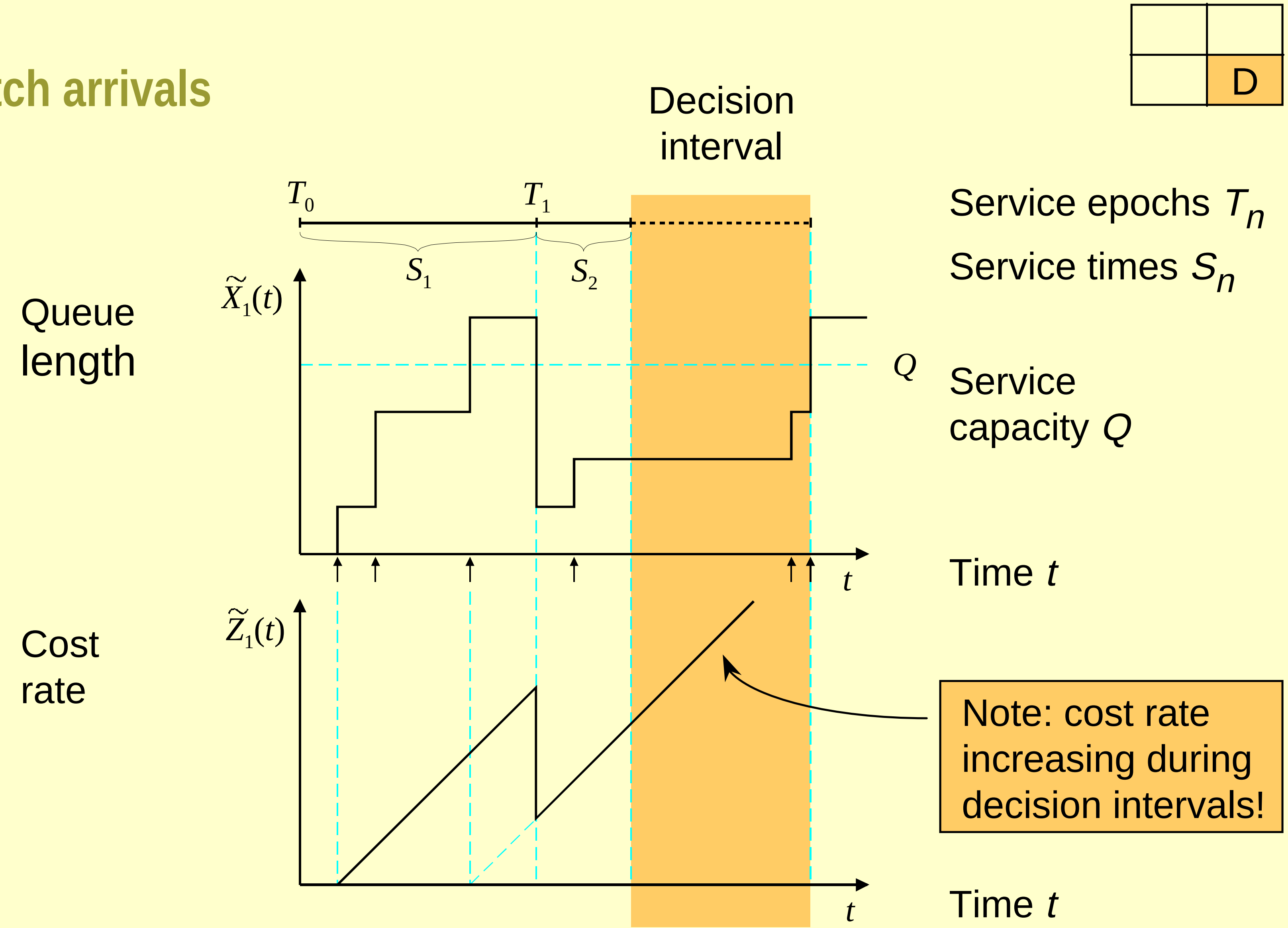
General threshold Q-policies

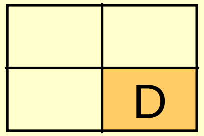
- **General threshold Q-policy** $\pi_Q(z,\zeta)$ with threshold z and (increasing) value function ζ:
  - after a service completion, a new service is initiated as soon as

$$Z(t) + \zeta(X(t)) \geq z \quad \text{or} \quad X(t) \geq Q$$

  - a service batch includes as many customers as possible
    - finite capacity => min{X(t),Q}

Batch arrivals





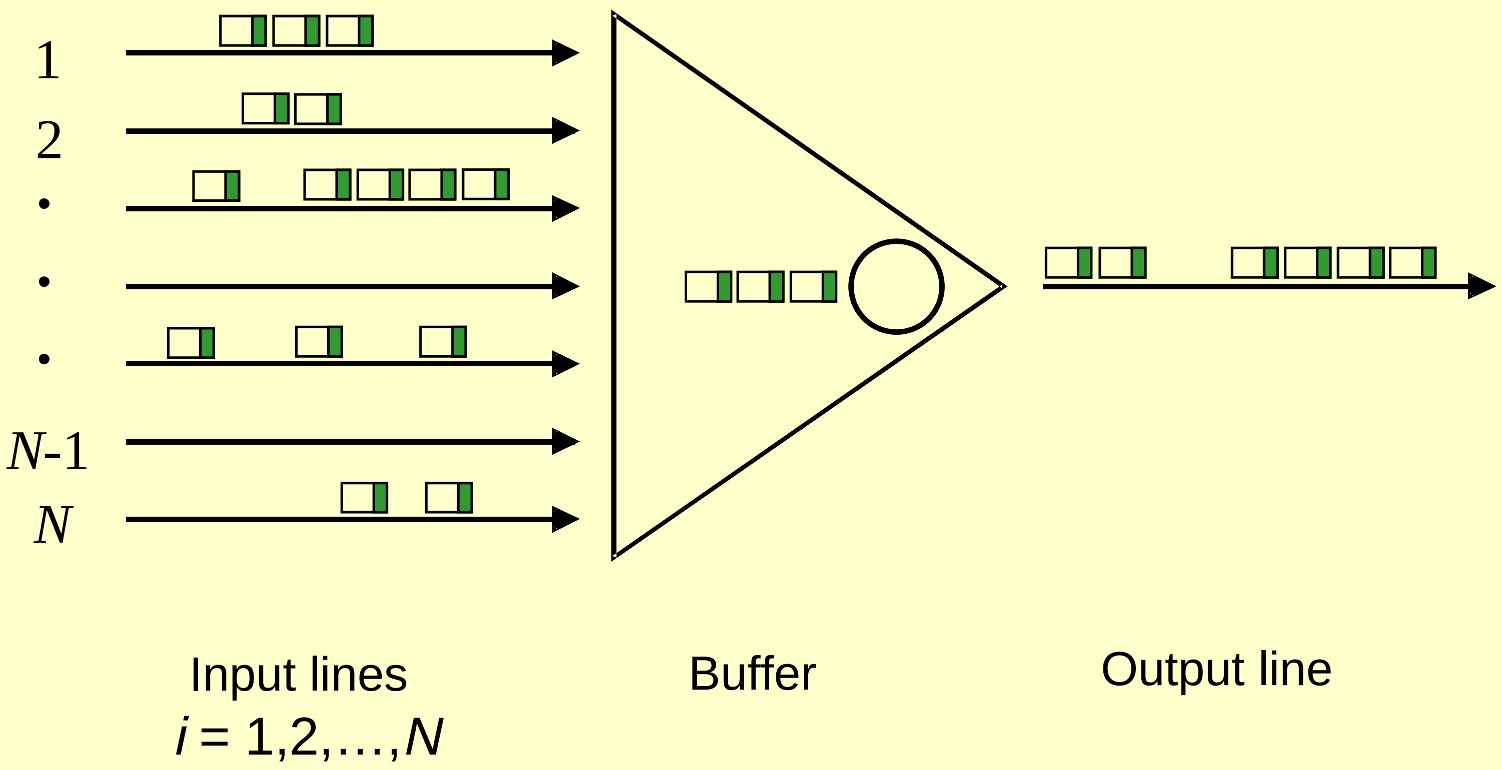
Open problems in Case D

- As regards the arrival batches,
  - How to get rid of the boundedness assumption?
- If no serving costs are included (K = 0, c = 0),
  - Is it true that similar results are valid in the average cost case as in the discounted cost case?
- If serving costs are included (K > 0, c > 0),
  - What is the optimal policy in the average cost or discounted cost sense?

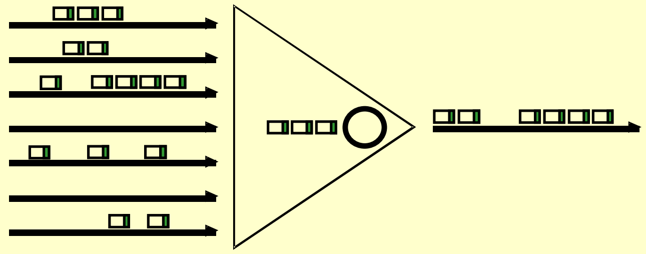
Contents

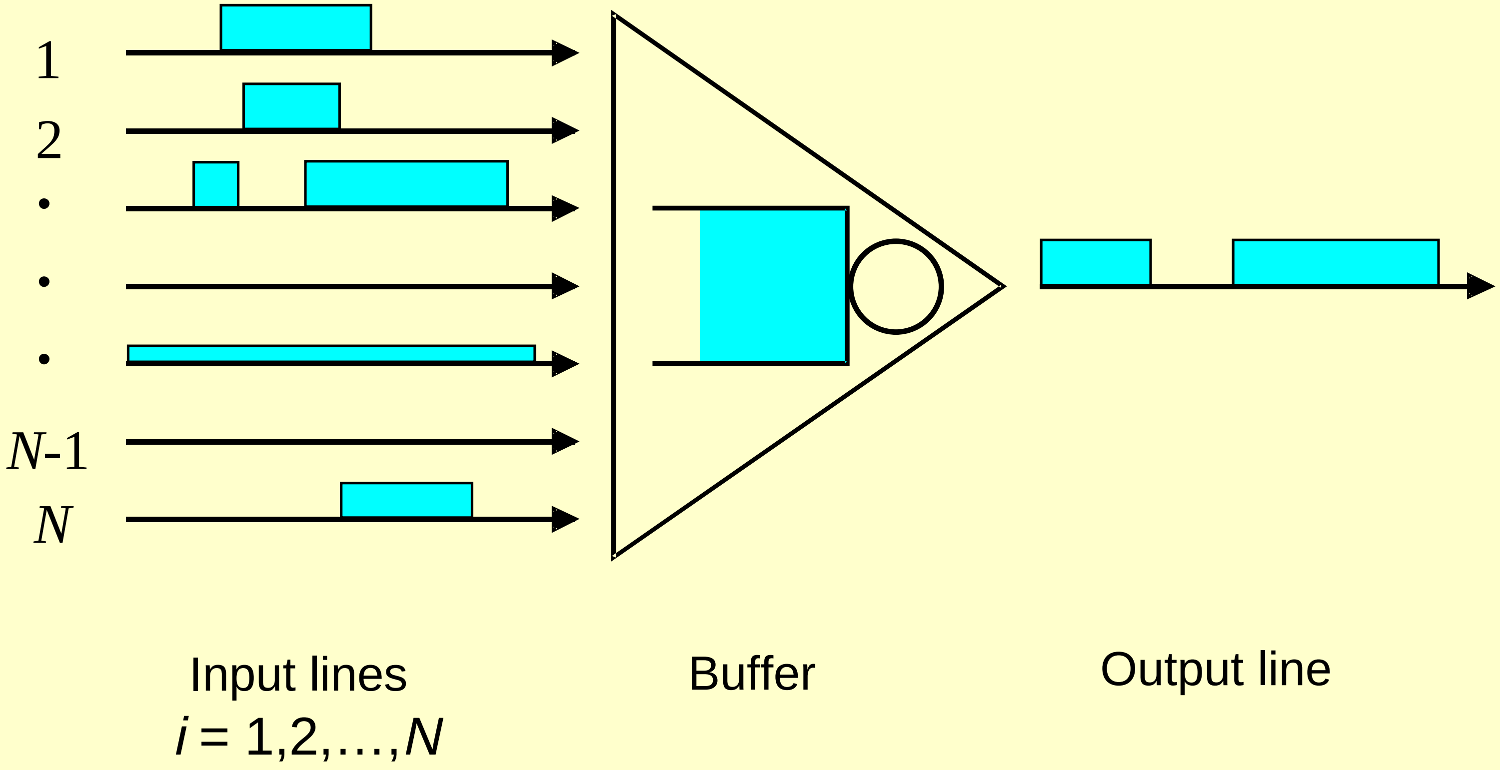
- Optimal control of batch service queues
- Characterization of the output process for some fluid queues
- Blocking probabilities in dynamic multicast networks
- Other topics

Statistical multiplexer ...



… as seen at burst level

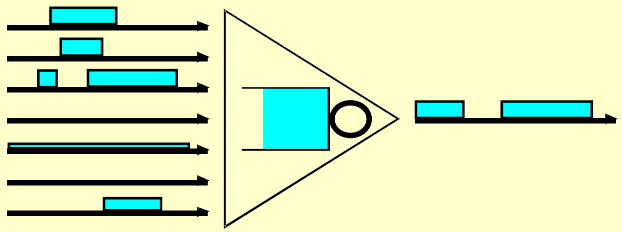




Fluid queue = fluid flow storage model

- Input rate $r_0(t)$
  - varying randomly
- Buffer size
  - finite or infinite
- Leak rate c_1
  - max output rate

- Buffer content process Z(t)
- Output rate process $r_1(t)$



$$Z(t) = Z(0) + \int_0^t r_0(u)du - \int_0^t r_1(u)du$$

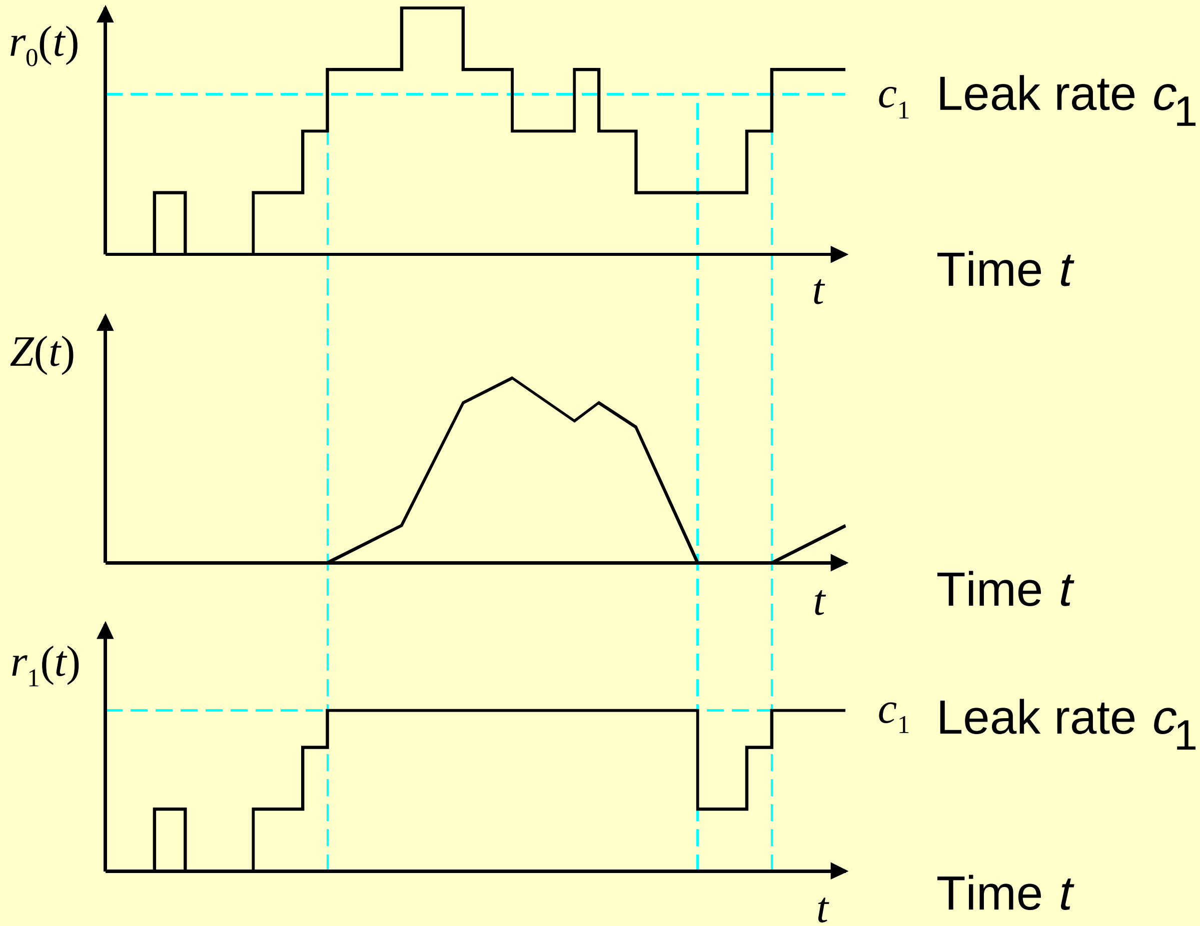
$$r_1(t) = \begin{cases} \min\{r_0(t), c_1\}, & \text{if } Z(t) = 0 \\ c_1 & \text{, if } Z(t) > 0 \end{cases}$$

Evolution

Input rate
$r_0(t)$

Buffer content
$Z(t)$

Output rate
$r_1(t)$



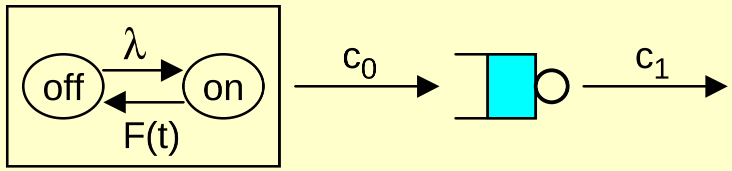
Fluid queue driven by a single on-off source

- On-off source with rate c_0
  - Idle periods ~ Exp(λ)
  - Active periods ~ F(t)
- $c_0 \leq c_1$ => input = output
- Assumption:

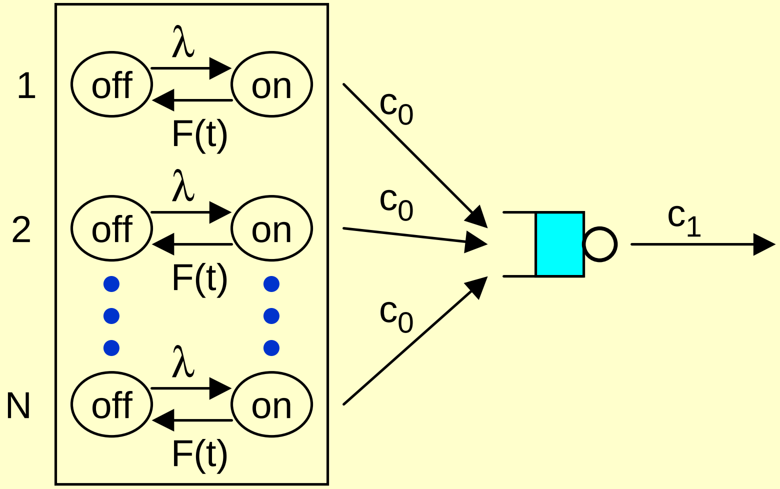
$$c_0 > c_1$$

=> output looks like another on-off source (with rate c_1)

- Result:
  - active periods on the output line ~ busy periods in M/G/1 queue with arrival rate (1 - c)λ and service time distribution function F(ct)



Fluid queue driven by multiple homogeneous on-off sources



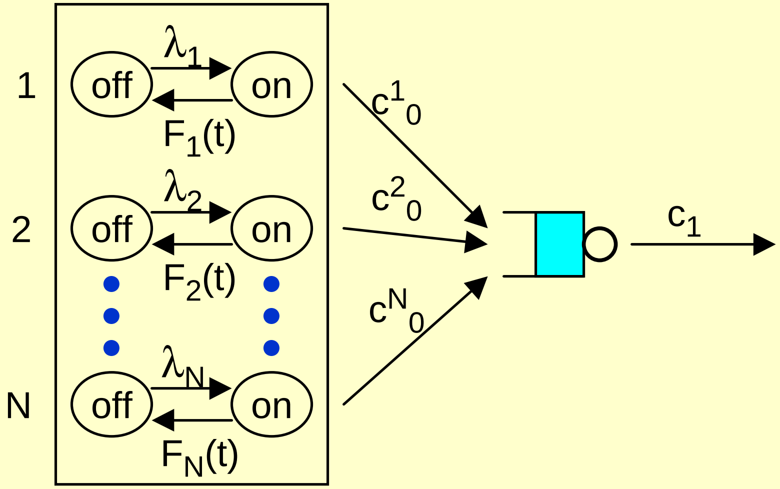
- Assumption:

$$c_0 \geq c_1$$

=> output looks like another on-off source (with rate c_1)

- Rubinovitch (1973): $c_0 = c_1$ =>
  - active periods on the output line ~ busy periods in M/G/1 queue with arrival rate (N - 1)λ and service time distribution function F(t)
- Boxma & Dumas (1998), Aalto (1998) [4]:
  - active periods on the output line ~ busy periods in M/G/1 queue with arrival rate (N - c)λ and service time distribution function F(ct)

Fluid queue driven by multiple heterogeneous on-off sources



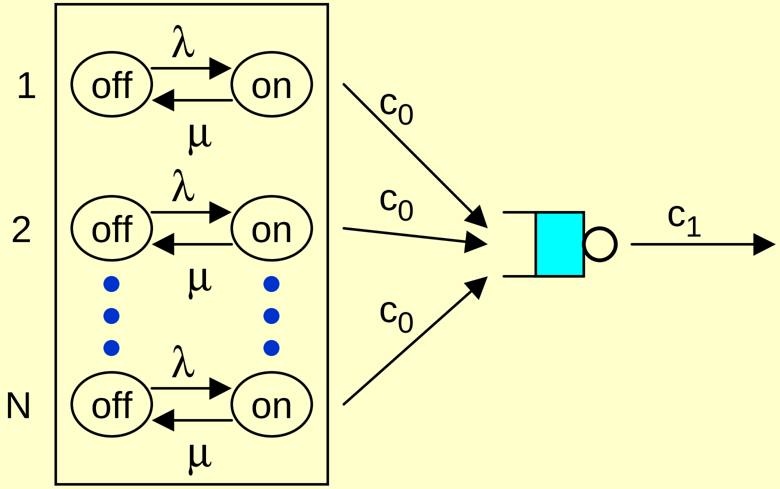
- Assumption:

$$c_0^i \geq c_1 \quad \text{for all } i$$

=> output looks like another on-off source (with rate c_1)

- Kaspi & Rubinovitch (1975): $c^i_0 = c_1$ for all i =>
  - characterization of the active periods on the output line by Laplace transforms
- Boxma & Dumas (1998), Aalto (1998) [4]:
  - characterization of the active periods on the output line by Laplace transforms

Fluid queue driven by multiple exponential on-off sources (1)



- **A-M-S model** first analysed by Anick, Mitra & Sondhi (1982)
- Assumption:

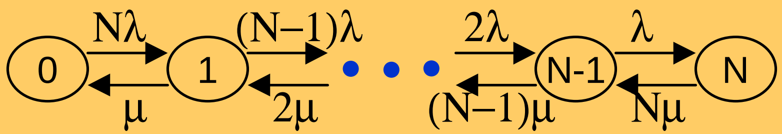
$$c_0 \neq c_1$$

- Input rate is modulated by a finite state Markov birth-death process J(t): $r_0(t) = J(t)c_0$
- Aalto (1994) ITC-14:
  - characterization of the output rate process by constructing another Markov jump process, which modulates the output rate:

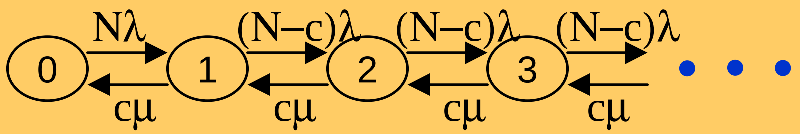
$$r_1(t) = d(\tilde{J}(t))$$

Fluid queue driven by multiple exponential on-off sources (2)

- Input rate is modulated by the following birth-death process J(t):



- In general, the ouput rate is modulated by a 3-dimensional Markov jump process (with an infinite state space)
- However, if $c_0 > c_1$, then the output looks like an on-off source and it is modulated by the following birth-death process:



  – note that, the active periods of output line ~ busy periods in M/M/1 queue with parameters (N-c)λ and cμ

Fluid queue driven by a Markov jump process (1)

- Input rate modulated by a (general) Markov jump process J(t)

$$r_0(t) = a(J(t))$$

- Assumption 1:

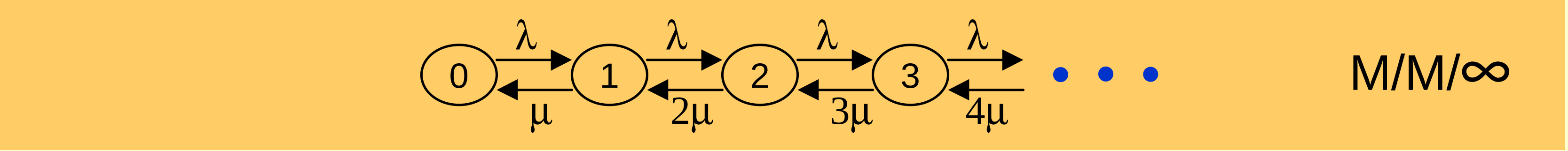
$$a(j) \neq c_1 \quad \text{for all } j$$

- Assumption 2:
  - visits to underloaded ($a(j) < c_1$) and overloaded ($a(j) > c_1$) states constitute an alternating renewal process
- Aalto (1998) [3]:
  - characterization of the output rate process by constructing another Markov jump process, which modulates the output rate:

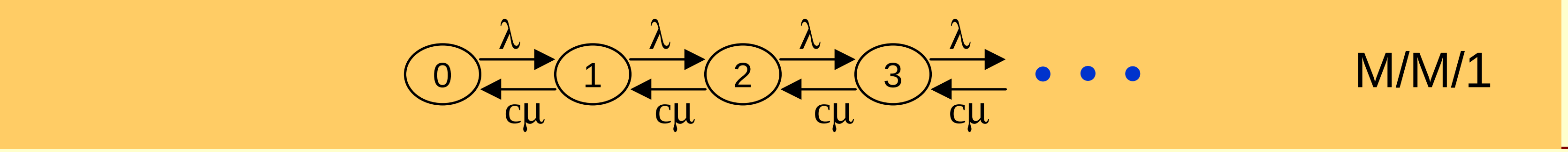
$$r_1(t) = d(\tilde{J}(t))$$

Fluid queue driven by a Markov jump process (2)

- In general, the ouput rate is modulated by a 2d+1-dimensional Markov jump process (with an infinite state space), where d refers to the dimension of J(t)
- **Example** (Fluid queue driven by an M/M/∞ queue):
  - Input rate is modulated by the following birth-death process J(t):



  - $r_0(t) = J(t)c_0$
  - If $c_0 > c_1$, then the output looks like an on-off source and it is modulated by the following birth-death process:



Open problems

- Fluid queue driven by on-off sources:
  - generalization of paper [4]
  - characterization of the output rate when $c_0 < c_1$ and active periods are non-exponential (even heavy tailed)
- Fluid queue driven by an M/G/∞ queue
  - some results exist for the buffer content process even in the case $c_0 < c_1$ (Jelenkovic)
  - is it possible to characterize output rate process in this case?
- Fluid queue driven by a Semi-Markov process:
  - generalization of paper [3]
  - is it possible to characterize the output process in this case?
- Networks of fluid queues?

Contents

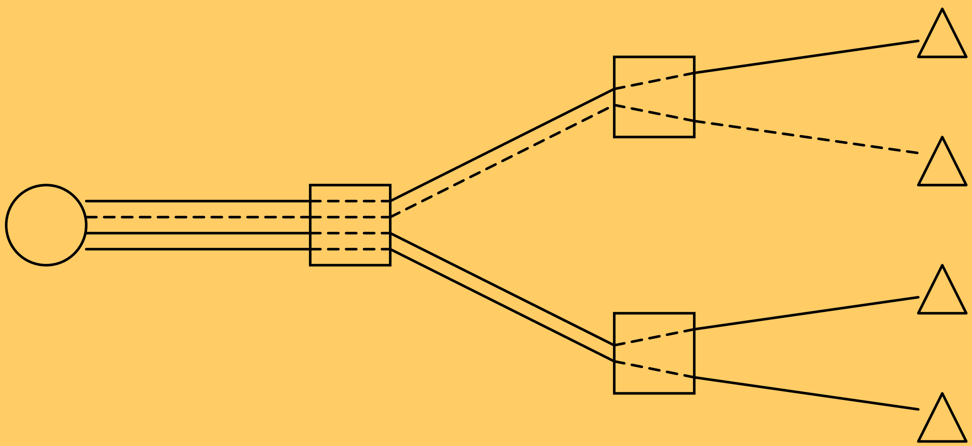
- Optimal control of batch service queues
- Characterization of the output process for some fluid queues
- Blocking probabilities in dynamic multicast networks
- Other topics

Blocking probabilities in dynamic multicast networks

- Co-operation with Jorma Virtamo, Jouni Karvo and Olli Martikainen from Helsinki University of Technology
- Application:
  - TV or radio delivery via a telecommunication network
- Multicast communication can be implemented by
  - point-to-point connections,
  - static multicast connections, or
  - dynamic multicast connections

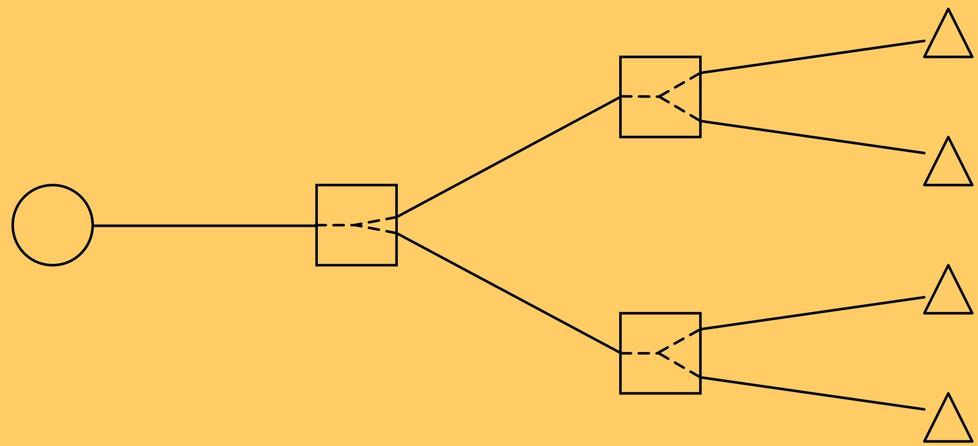
Point-to-point connections

- Flexible but …
- … wasting resources



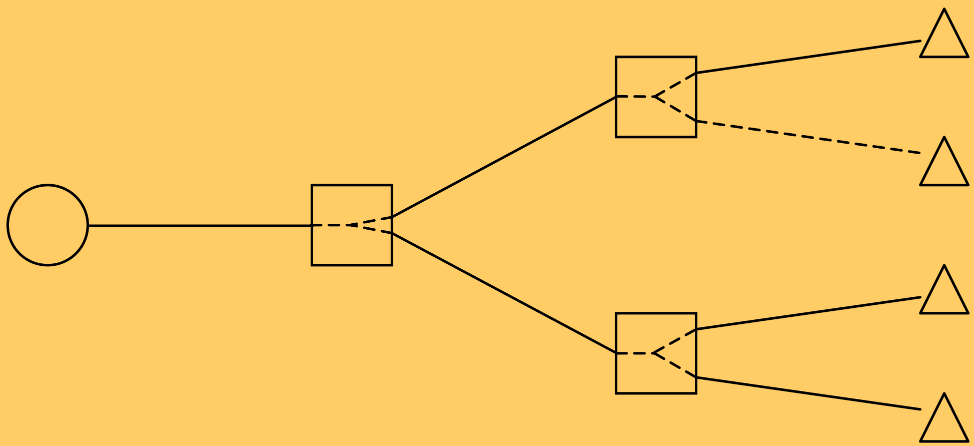
Static multicast connections

- Saving resources but …
- … inflexible



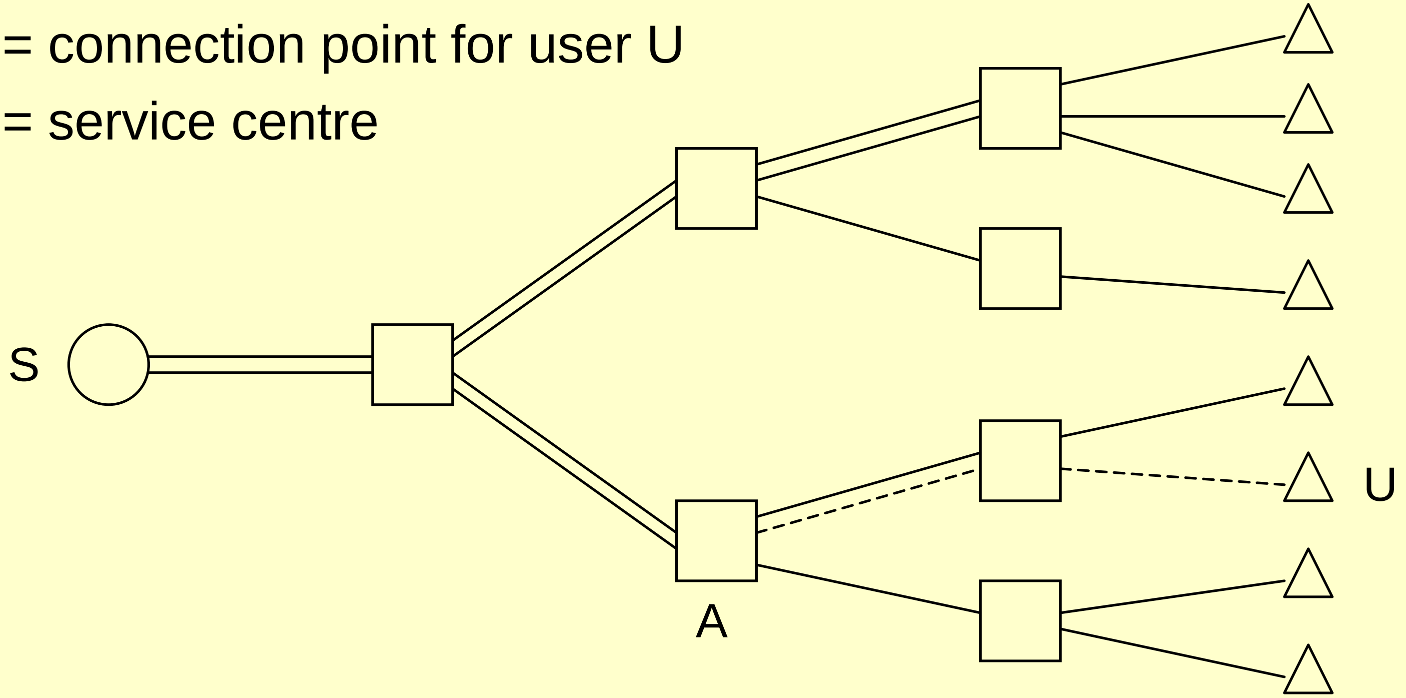
Dynamic multicast connections

- Flexible
- Saving resources



Dynamic multicast network (multicast tree)

- U = new user of green channel
- A = connection point for user U
- S = service centre



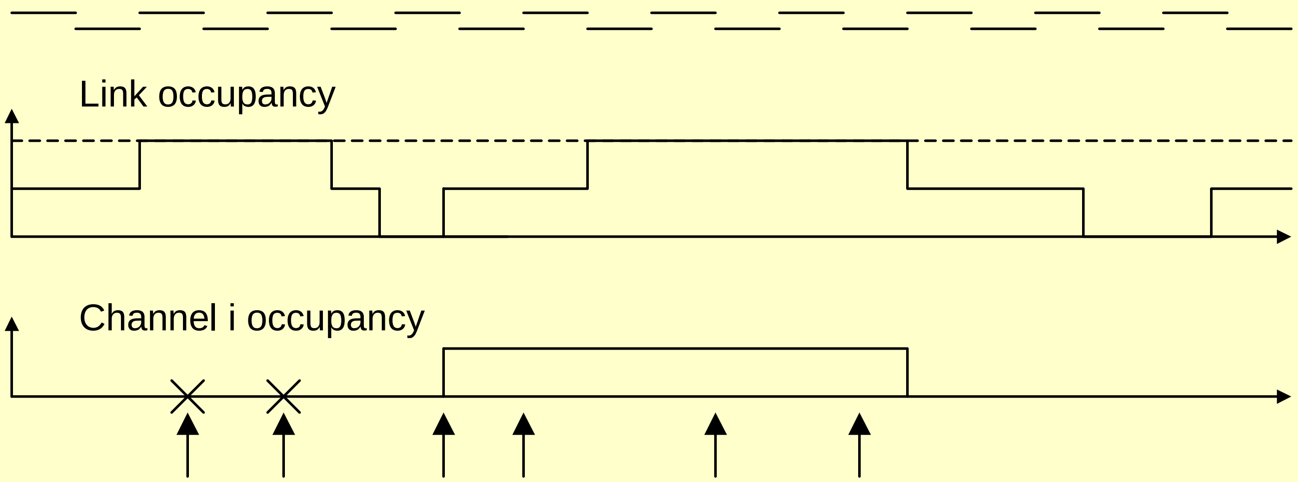
Blocking of dynamic multicast connections

- Networks with
  - point-to-point or
  - **static** multicast connections

  can be modelled as loss networks, but not a network with
  - **dynamic** multicast connections
- Thus, new methods are needed to calculate blocking probabilities in dynamic multicast networks

Different types of blocking

- Call blocking $(b^c_i = 2/6 = 33\%)$
- Channel blocking $(B^c_i = 2/3 = 67\%)$
- Time blocking $(B^t_i = 8/20 = 40\%)$



Blocking in a single link

- Karvo, Virtamo, Aalto & Martikainen (1998a) BC'98:
  - method to calculate link occupancy distribution and different types of blocking probabilities in a single link with finite capacity
- Assumptions:
  - other links with infinite capacity
  - user populations in the leaves of the multicast tree subscribe to different channels i according to independent Poisson processes
  - channel subscription times (of user populations) generally distributed with channel-wise means
- Ideas:
  - active periods of channel i ~ busy periods in $M/G/\infty$ queue
  - channel blocking as call blocking in a generalized Engset system
  - time blocking based on link occupancy distrib'n in an infinite system

End-to-end blocking in a network

- Karvo, Virtamo, Aalto & Martikainen (1998b) submitted:
  - calculation of end-to end call blocking probabilities by applying the Reduced Load Approximation method
  - verification by simulations
- Results:
  - RLA seems to give an upper bound
  - results of approximations and simulations on the same scale
  - difference about 10 - 50 % (not totally satisfactory)

Open problems

- Method to calculate the end-to-end call blocking probability exactly
- Improved simulations methods
- Improved approximation methods for the end-to-end blocking

Contents

- Optimal control of batch service queues
- Characterization of the output process for some fluid queues
- Blocking probabilities in dynamic multicast networks
- Other topics

Other topics

- Average signalling load in dynamic multicast networks
  - Co-operation with Jouni Karvo
- Workload and waiting time distributions in M/D/n queue
  - Co-operation with Jorma Virtamo
- Optimal configuration of delay lines in an optical switch


FIN