Performance Modeling of Cellular Networks

Traffic Engineering

Villy B. Iversen

Department of Photonic Engineering
Technical University of Denmark

Eighth FRUCT Conference
November 11, 2010

Outline

History

- Alexander Popov, Valdemar Poulsen, Guglielmo Marconi
- Example of Poulsen's Arc-Transmitter inaugurated 1927
  Between The Hague and Bandung/Java for one connection
  Transmitter: 2000 KW
  Wavelength: 17.850 m
  Antenna: 1800 m
  Distance 12.000 km
  Largest transmitter 3500 KW

Modeling Communication (Man/Machine) Systems

Three Elements in a Model:

- System (hardware, Deterministic)
  Number of channels, Accessibility
- Strategy (Software, Deterministic)
  Acceptance Control, Dealing with blocked calls
- Traffic (Stochastic)
  Arrival processes, Service processes, Human behavior

Purpose: to evaluate:

- Grade-of-Service
- Cost

Outline

Introduction and Background

Circuit Switched Networks
  Single Cell System
  Multi-Cell Systems

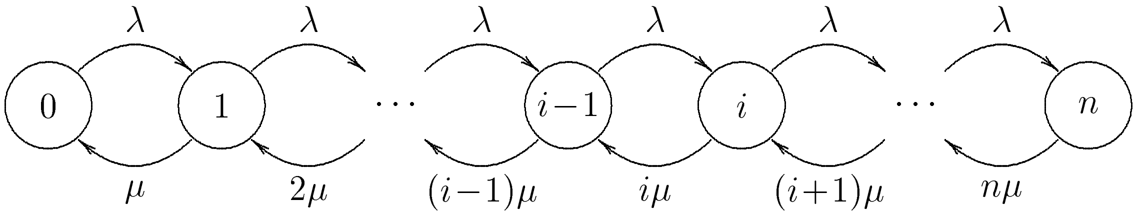
Packet Switched systems
  Single Cell Systems
  Multiple Cell Systems

Summary & Conclusions

Erlang's Model

- System: *n* channels
- Strategy: Full accessibility, Blocked Calls Cleared
- Traffic: One one service requiring one channel
  Offered traffic $A = \lambda/\mu$
  Poisson arrival process (rate λ calls per time unit)
  Exponential service time rate (mean value $1/\mu$)

State transition diagram:

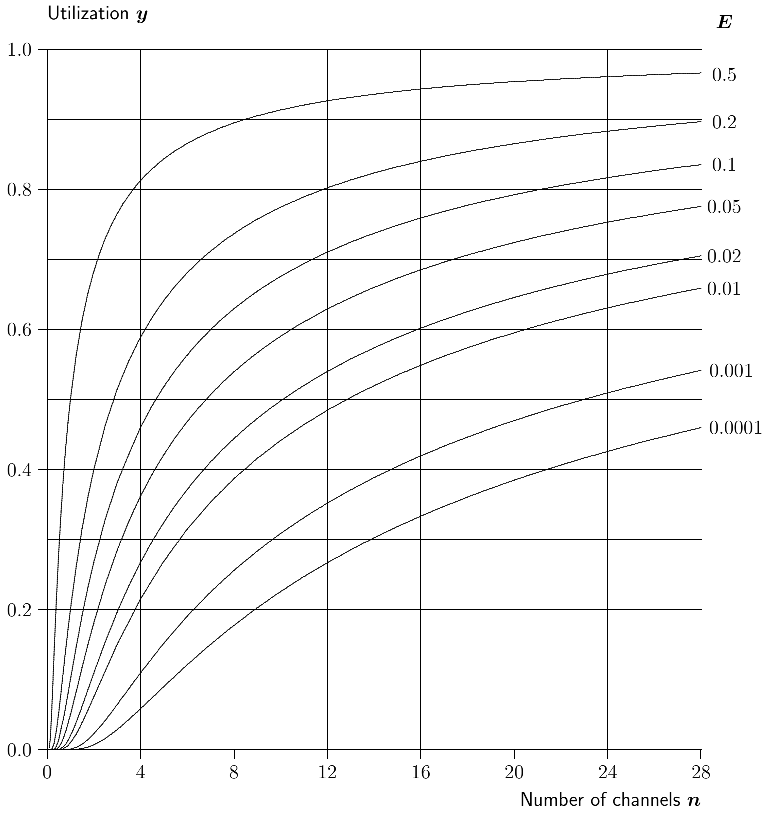


Erlang's B-formula

The system has $n + 1$ states $\{0, 1, 2, \ldots, n\}$
Blocking Probability = probability that all channels are busy:

$$B_n(A) = \frac{\frac{A^n}{n!}}{1 + A + \frac{A^2}{2!} + \cdots + \frac{A^n}{n!}} .$$

- The model is insensitive to service time distribution
- Accurate recursive algorithms to find B, n or A.
- Successful for POTS (Plain Old Telephone Service)



Average channel utilization for fixed blocking probability.

BPP traffic paradigm

Erlang's model describe traffic by mean value A
BPP-model describes traffic by A and $Z = V/M$, $V =$ Variance.

- $Z < 1$ **B** Binomial (Engset) Smooth traffic
- $Z = 1$ **P** Poisson (Erlang) Random traffic
- $Z > 1$ **P** Pascal (Palm/Wallström) Peaked traffic

Engset traffic: Finite number of users

Pascal traffic: Batched Poisson arrival

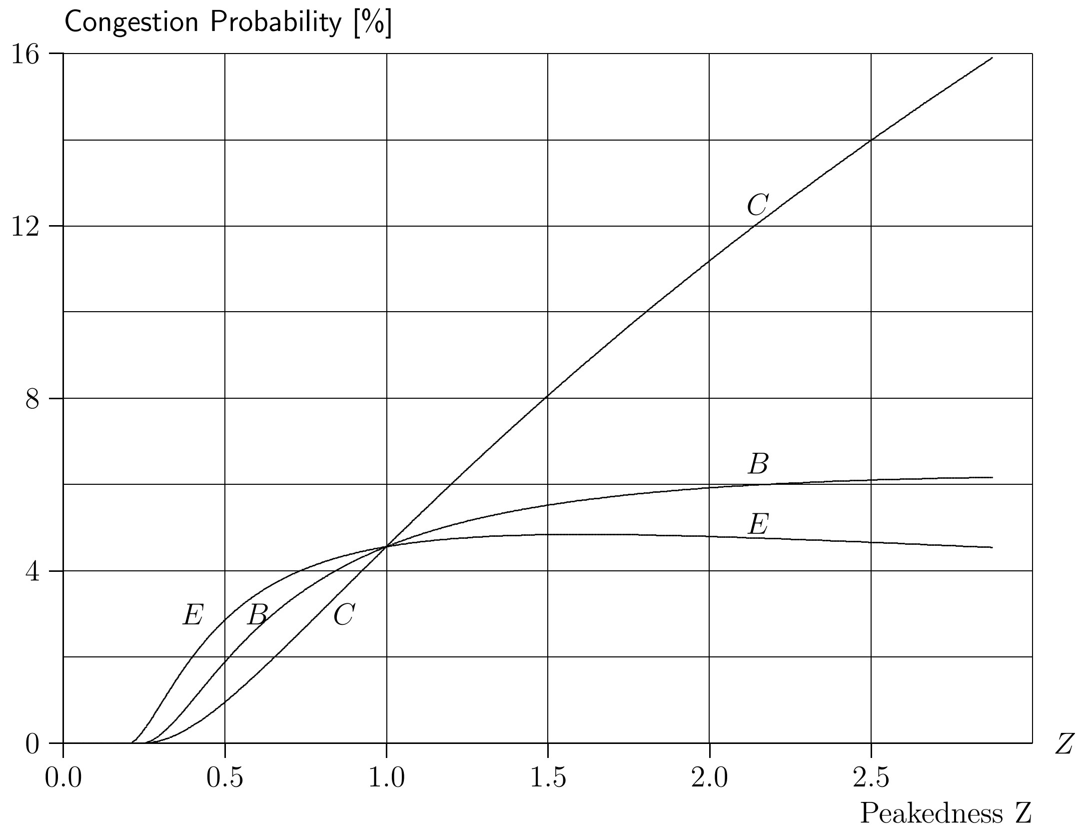
Performance measures

- $E =$ Time congestion: Proportion of time all channels busy
- $B =$ Call congestion: Proportion of call attempts blocked
- $C =$ Traffic congestion: Proportion of offered traffic lost

Only traffic congestion C is relevant for Grade-of-Service

E and B are misleading (but widely used)

- $Z < 1$: $E > B > C$
- $Z = 1$: $E = B = C$
- $Z > 1$: $E < B < C$



Time E, Call B and Traffic congestion C ($n = 20$, $A = 15$)

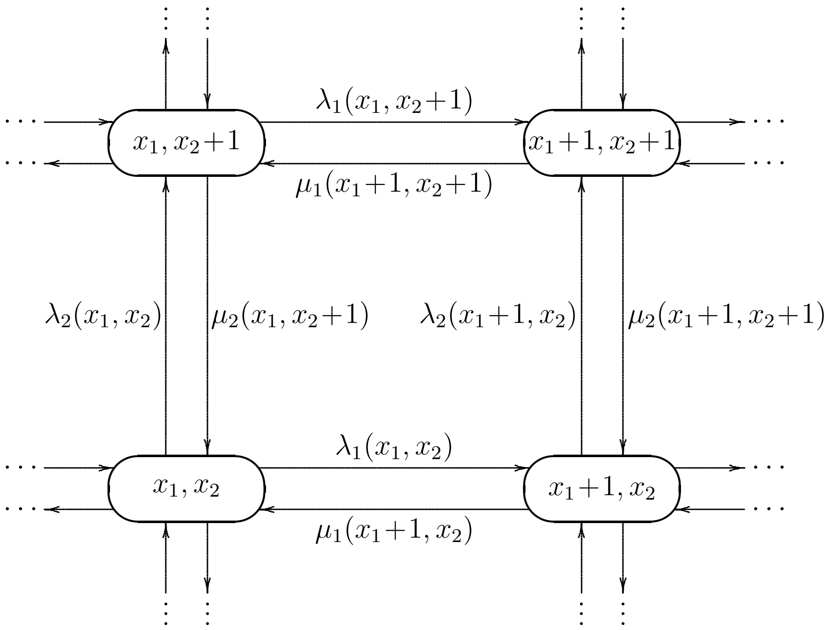
Summary: Single Cell systems with Single Service

- BPP traffic models with two parameters
  Mean and variance
- Insensitive to service time distribution
  Only mean service time is relevant
- Accurate algorithms for evaluation
  Also for inverse algorithms
- Consider only Traffic congestion
  Time and Call congestion used in literature
- Economy-of-scale (statistical multiplexing) Fixed blocking:
  higher utilization per channel in big systems

Multi-rate traffic

- Multi-service systems: N services
- Service i requires d_i BBU = Basic Bandwidth Units
- Each service described by 3 parameters $\{A_i, Z_i, d_i\}$
- State space: $p(x_1, x_2, \ldots, x_N)$
  $N = 10$ services, $d_i = i$
  $n = 100$: 82 million states
  $n = 150$: 2316 million states
  n (granularity) should be at least say 1024
  State space explosion!
  How to find state probabilities?

Reversible Systems



Kolmogorov cycles: Flow clockwise equals flow counter-clockwise

Local Balance & Product form

1. Solve linear balance equations?
   General method, but too many states
2. Reversibility = local balance (additional restriction)
   - Insensitivity to service time distribution
   - Express all states by state zero and normalize
     Too many states!
3. Product Form (additional restriction)

$$p(x_1, x_2, \dots, x_N) = p(x_1) \cdot p(x_2) \cdot \ldots \cdot p(x_N)$$

Reduction of number of states from $O(n^N)$ to $O(N \cdot n)$

Local Balance & Product form

1. Solve linear balance equations?
   General method, but too many states
2. Reversibility = local balance (additional restriction)
   - Insensitivity to service time distribution
   - Express all states by state zero and normalize
     Too many states!
3. Product Form (additional restriction)

$$p(x_1, x_2, \dots, x_N) = p(x_1) \cdot p(x_2) \cdot \ldots \cdot p(x_N)$$

Reduction of number of states from $O(n^N)$ to $O(N \cdot n)$

Local Balance & Product form

1. Solve linear balance equations?
   General method, but too many states
2. Reversibility = local balance (additional restriction)
   - Insensitivity to service time distribution
   - Express all states by state zero and normalize
     Too many states!
3. Product Form (additional restriction)

   $$p(x_1, x_2, \ldots, x_N) = p(x_1) \cdot p(x_2) \cdot \ldots \cdot p(x_N)$$

   Reduction of number of states from $O(n^N)$ to $O(N \cdot n)$

Algorithms for BPP Multi-rate traffic

- Convolution Algorithm (Iversen 1987)
  Product form $\Leftrightarrow$ Convolution
  Aggregate $N-1$ services by repeated convolutions
  End up with two-dimensional state space:

  $$p(x_1, x_2, \ldots, x_N) \curvearrowright p(x_{1,2,\ldots i-1,i+1,\ldots N}, x_i)$$

  - Find detailed performance measures for stream i
  - Change order of convolution for other streams
  - Complexity: $N \cdot n^2$

  For each service:

  - Guaranteed minimum capacity n_i
  - Maximum allowed capacity m_i

Algorithms for BPP Multi-rate traffic

- State-based recursive algorithm (Iversen 2006)
  Global states from contributions from each service
  Special cases:
    - Recursive formulæ for Erlang and Engset models
    - Kaufman & Roberts algorithm (Poisson traffic)
    - Delbrouck's algorithm

  Complexity: $N \cdot n$ (linear in both n and$N$)
  Based on Global States:
    - Allows for trunk/channel reservation with individual levels (only exact when all mean holding times are equal)

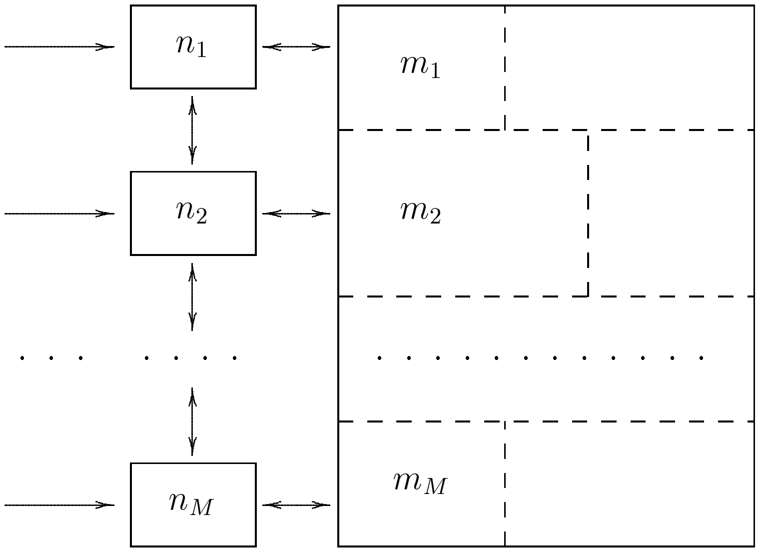
Summary: Single Cell Systems with Multi-Services

- Each service described by 3 parameters
- The model is insensitive to service time distribution
- Detailed performance measures for each service
- Accurate & fast implementation of algorithms
- Min & Max allocation parameters for each service (convolution)
- Trunk (= Channel) reservation parameter for each service (state-based)

Outline

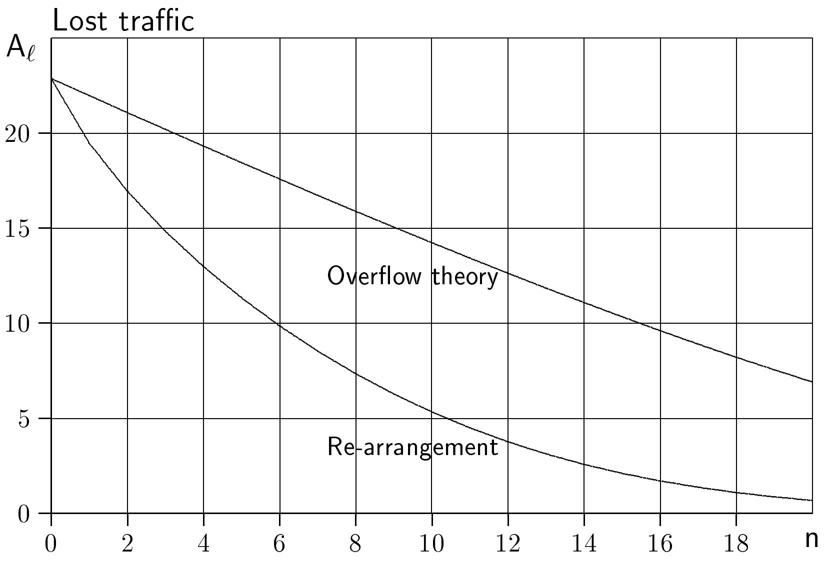
- Introduction and Background
- Circuit Switched Networks
  - Single Cell System
  - Multi-Cell Systems
- Packet Switched systems
  - Single Cell Systems
  - Multiple Cell Systems
- Summary & Conclusions

Femto-cell System



Cellular system with M micro-cells and one macro-cell.
$\equiv$ to a link offered M traffic streams with min/max allocation.
Product form: take-back of calls from macro- to micro-cell.

Femto-cell System



30 micro cells, each offered 15.25 erl to 20 channels ($B = 5\%$). By using take-back (re-arrangement) one channel in the macro cell increases carried traffic by 3.39 erl. Single-rate traffic.

Circuit Switched Networks

Link	Route				Number of channels
	1	2	$\cdots$	N	
1	d_{11}	d_{21}	$\cdots$	d_{N1}	n_1
2	d_{12}	d_{22}	$\cdots$	d_{N2}	n_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
K	d_{1K}	d_{2K}	$\cdots$	d_{NK}	n_K

A circuit switched network with direct routing.
d_{ij} is the bandwidth demand of route j on link i.

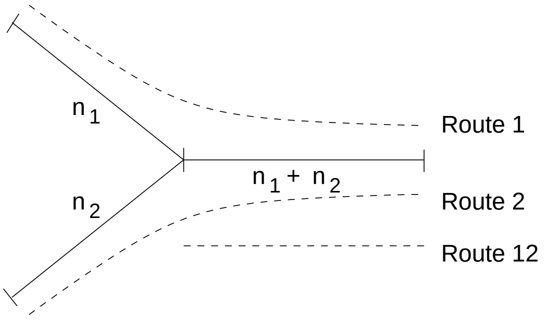
Circuit Switched Networks

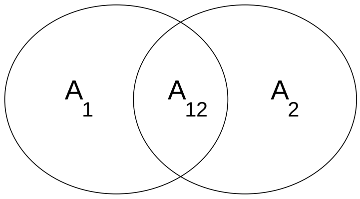
We have Product Form between routes
Convolution algorithm:

- To truncate state space we keep record of either
  - Number of connections per route (compact networks), or
  - Number of busy channels per link (sparse networks)
- We convolve multi-dimensional vectors of size:
  $\prod_{i=1}^{K}(n_i + 1)$ (worst case, in reality much less)
- A link corresponds to a restriction on a set of routes (maximum number of channels used)

State Space Explosion!

Cellular Network $\equiv$ Circuit Switched-Network

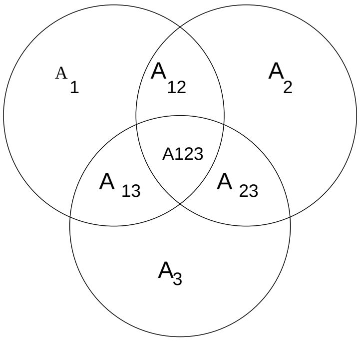




		Route			
		1	2	12	Capacity
	1	1	0	0	n_1
Link	2	0	1	0	n_2
	12	1	1	1	$n_1 + n_2$

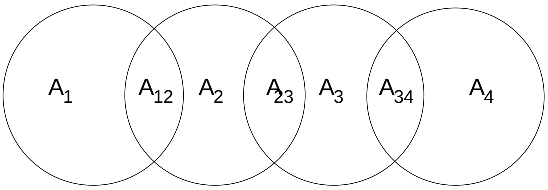
Equivalence of a cellular network with two overlapping cells and a circuit-switched network with direct routing. Optimal Packing

Cellular Network ≡ Circuit Switched-Network



	Route							
Link	1	2	3	12	13	23	123	Capacity
1	1	0	0	0	0	0	0	n_1
2	0	1	0	0	0	0	0	n_2
3	0	0	1	0	0	0	0	n_3
12	1	1	0	1	0	0	0	$n_1 + n_2$
13	1	0	1	0	1	0	0	$n_1 + n_3$
23	0	1	1	0	0	1	0	$n_2 + n_3$
123	1	1	1	1	1	1	1	$n_1 + n_2 + n_3$

Cellular Network ≡ Circuit Switched-Network



	Route							
Link	1	2	3	4	12	23	34	C
1	1	0	0	0	0	0	0	30
2	0	1	0	0	0	0	0	30
3	0	0	1	0	0	0	0	30
4	0	0	0	1	0	0	0	30
12	1	1	0	0	1	0	0	60
23	0	1	1	0	0	1	0	60
34	0	0	1	1	0	0	1	60
123	1	1	1	0	1	1	0	90
234	0	1	1	1	0	1	1	90
1234	1	1	1	1	1	1		120

Numerical Studies

- Cellular Systems: Many small cells.
  Small independent groups ⇔ low utilization
- Overlap: By 20% overlap and dynamic Hand-Over same utilization as in full accessible groups
- Hierarchical system with call packing:
  - high utilization of all micro-cells
  - Guarantee a high Grade-of-Service

Outline

Packet Switched Systems

- Traffic:
  - CBR = Constant Bit Rate traffic $\sim$ circuit switched traffic
  - VBR = Variable Bit Rate traffic
    - Use effective bandwidth concept
    - Use BPP multi-rate model Engset traffic = on/off traffic
- System: Fixed capacity $= n$ BBU)
- Strategy:
  - Delay or delay/loss system (finite buffer k)
  - Sharing of bandwidth, service at reduced rate during overload

Classical queueing models are very sensitive to service time distribution!
First queueing paper by Erlang 1909: constant service time.

Sharing Strategies

- PS = Processor Sharing (say $n = 1$)
  - All users share the capacity equally
  - If only one user in system, he obtains the whole capacity
- GPS = Generalized Processor Sharing (say $n > 1$)
  - If at most n users in the system,
    each user get one BBU (user access rate)
  - If more than n users, the share the capacity equally
- RS = Reversible Sharing.
  - Generalization of GPS to multi-rate traffic

All models are insensitive to the service time distribution
No Product Form

Sharing Strategies

- PS = Processor Sharing (say $n = 1$)
  - All users share the capacity equally
  - If only one user in system, he obtains the whole capacity
- GPS = Generalized Processor Sharing (say $n > 1$)
  - If at most n users in the system,
    each user get one BBU (user access rate)
  - If more than n users, the share the capacity equally
- RS = Reversible Sharing.
  - Generalization of GPS to multi-rate traffic

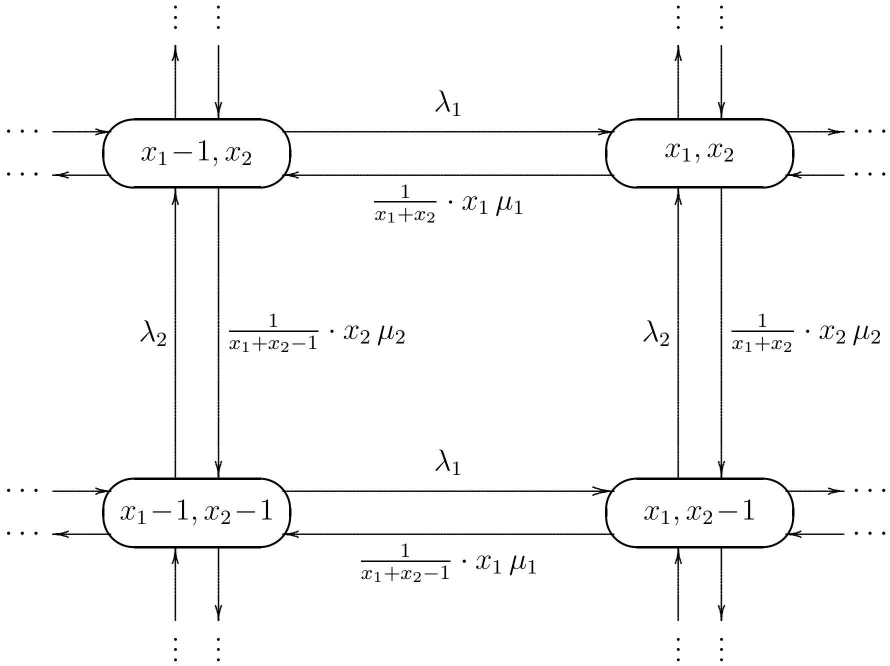
All models are insensitive to the service time distribution
No Product Form

Sharing Strategies

- PS = Processor Sharing (say $n = 1$)
  - All users share the capacity equally
  - If only one user in system, he obtains the whole capacity
- GPS = Generalized Processor Sharing (say $n > 1$)
  - If at most n users in the system,
    each user get one BBU (user access rate)
  - If more than n users, the share the capacity equally
- RS = Reversible Sharing.
  - Generalization of GPS to multi-rate traffic

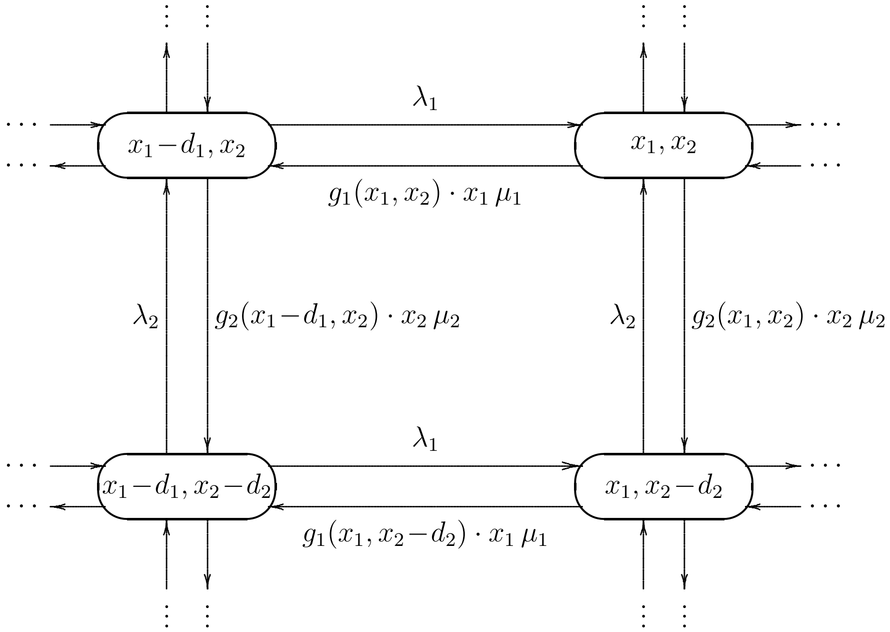
All models are insensitive to the service time distribution
No Product Form

Processor Sharing model



State transition diagram of PS-system is reversible (no product form).

Processor Sharing model



State transition diagram for a reversible multi-rate system with n servers and reduction factors $g_i(\overline{x})$

Reduction Factors

- Requiring for BPP multi-rate traffic that
    - The state transition diagram is reversible
    - All capacity is used if needed

  we find a unique recursive expression for reduction factors.
- The above simple "State-based recursive algorithm" is generalized to include this system (aggregation of states)
- Performance measures for each service:
    - Virtual waiting time
      (increased sojourn time due to finite buffer)
    - Loss probability

Processor Sharing model

Reversible sharing

- Includes PS and GPS for single rate
- Sharing mode for multi-rate
    - Underload: All users obtain capacity requested
    - Normalload: Broad-band traffic is reduced more than narrow-band
      (in percentage, still it obtain more bandwidth)
    - Overload: In the limit all obtain same bandwidth (PS)

Outline

Multi-rate Queueing Networks

- Due to reversibility nodes can be combined in networks
  Reversibility: arrival process $\sim$ departure process
- Flow balance equations give relative load of each node from each service
- From reduction factors calculate detailed state space
- Aggregate nodes by convolution algorithm
  (Usual Queueing Network approach)
- Truncate state space for closed chains
- Evaluate performance measures

The algorithm has been implemented in Java and C (linear in number of (servers + buffers($(k+n)$ and services N

Summary & Conclusions

- Each service described by 3 parameters
- Min and Max allocation guarantee end-to-end GoS
- All models are insensitive to service time distribution
- Includes both Circuit-switching and Packet-switching
- Generalization of Markovian Teletraffic Theory
- Effective accurate algorithms

Questions?

Bibliography

Iversen, V.B. & Benetis, V. & Hansen, P.D.: Performance of Hierarchical Cellular Networks with Overlapping Cells, Lecture Notes in Computer Science, ISSN 0302-9743, Issue: Volume 3427/2005 pages: 7 - 19. March 2005.

Iversen, V.B.: Reversible fair scheduling: the teletraffic revisited. Proceedings from 20th International Teletraffic congress, ITC20, Ottawa, Canada, June 17-21, 2007. Springer Lecture Notes in Computer Science. Vol. LNCS 4516 (2007), pp. 1135-1148.

Iversen, V.B.: Teletraffic Engineering and Network Planning. Technical University of Denmark, 2010. 376 pp. (Textbook)
http://www.com.dtu.dk/teletraffic

Iversen, V.B.: Evaluation of Multi-service CDMA Networks with soft Blocking. Proceedings Third conference on Smart Spaces, ruSMART 2010 and NEW2AN 2010, St. Petersburg, Russia, August 23-25, 2010. Springer Lecture Notes in computer Science, Vol. LNCS 6294, pp. 160-171.