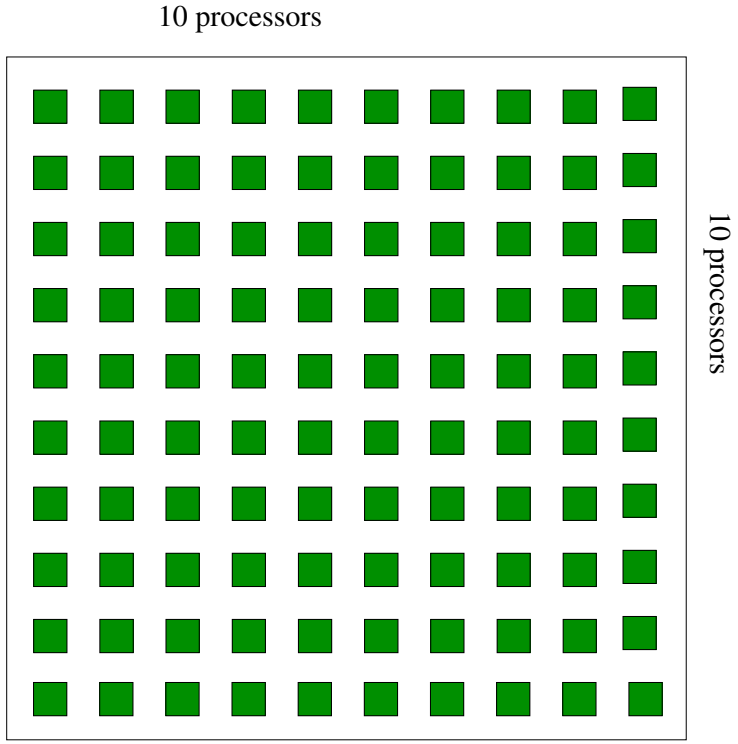
The Nostrum Network on Chip



Mikael Millberg, Erland Nilsson, Richard Thid, Johnny Öberg, Zhonghai Lu, Axel Jantsch

Royal Institute of Technology, Stockholm

December 1, 2004

Overview

Topology and Structure

Protocol Stack

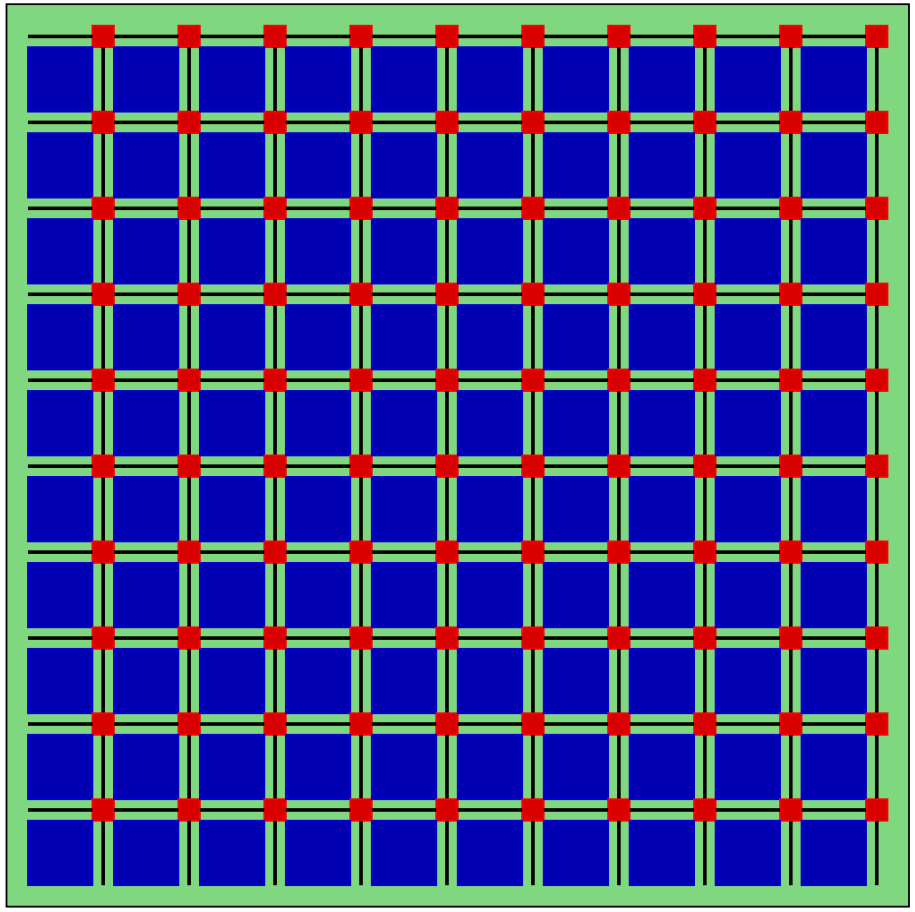
The Network Layer and the Switch

Data Protection

Simulation Environment

Clocking

Nostrum Topology: Mesh



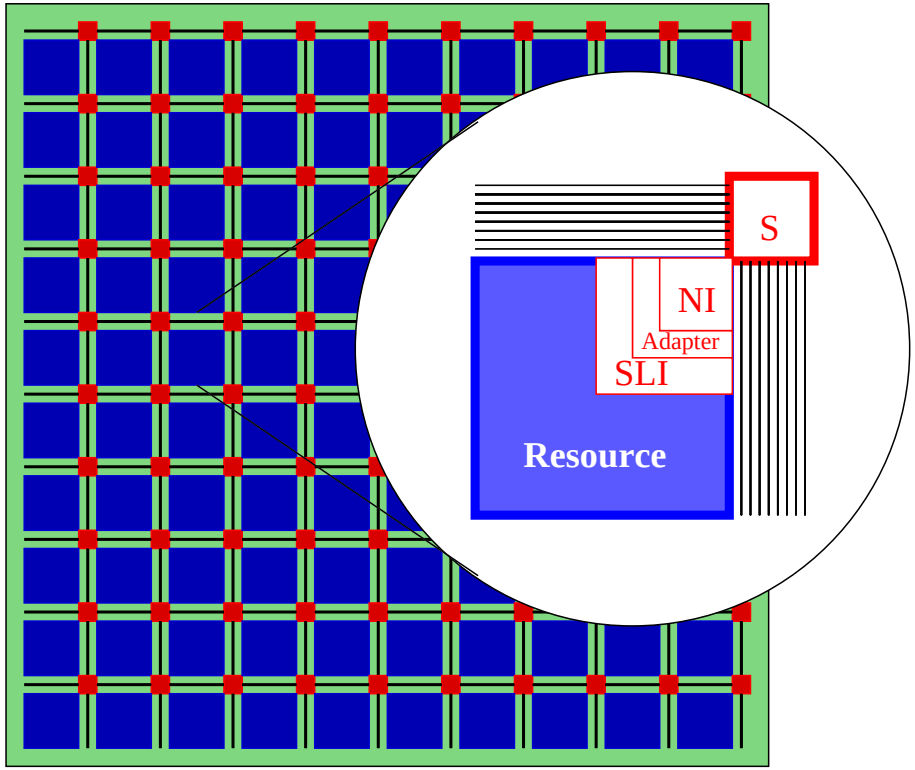
Characteristics:

- Resource-to-switch ratio: 1
- A switch is connected to 4 switches and 1 resource
- A resource is connected to 1 switch
- Max number of hops grows with $2n$

Motivation:

- Regularity of layout; predictable electrical properties
- Expected locality of traffic

The Node in a Mesh



NI: Network Interface:

- Compulsory
- HW
- Implements the network layer protocol

Adapter: Resource specific interface circuit;

SLI: Session Layer Interface:

- Optional
- Hardware and/or software
- Implements the session layer protocol

Overview

Topology and Structure

Protocol Stack

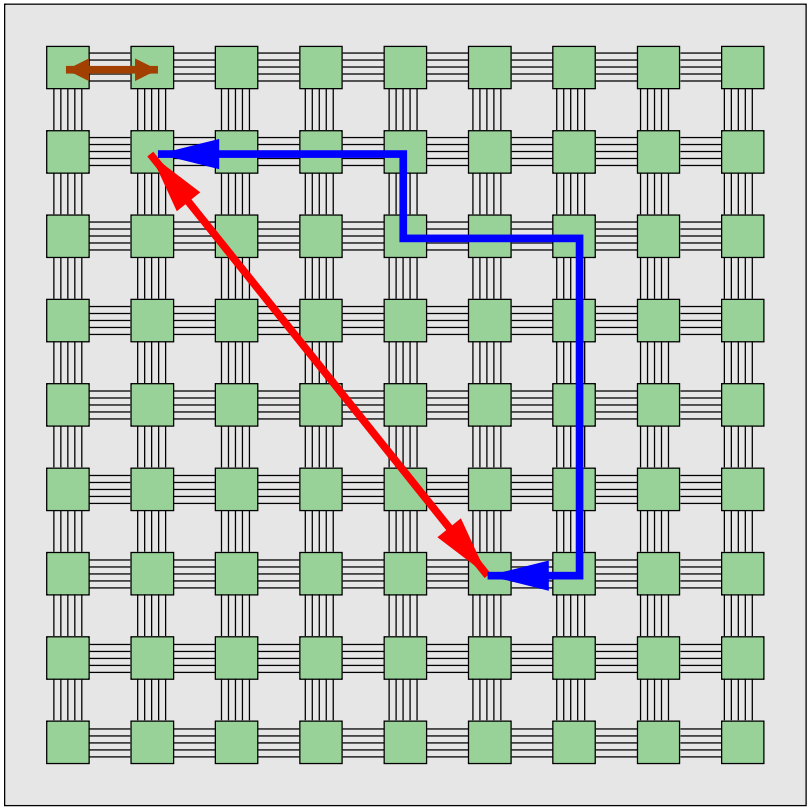
The Network Layer and the Switch

Data Protection

Simulation Environment

Clocking

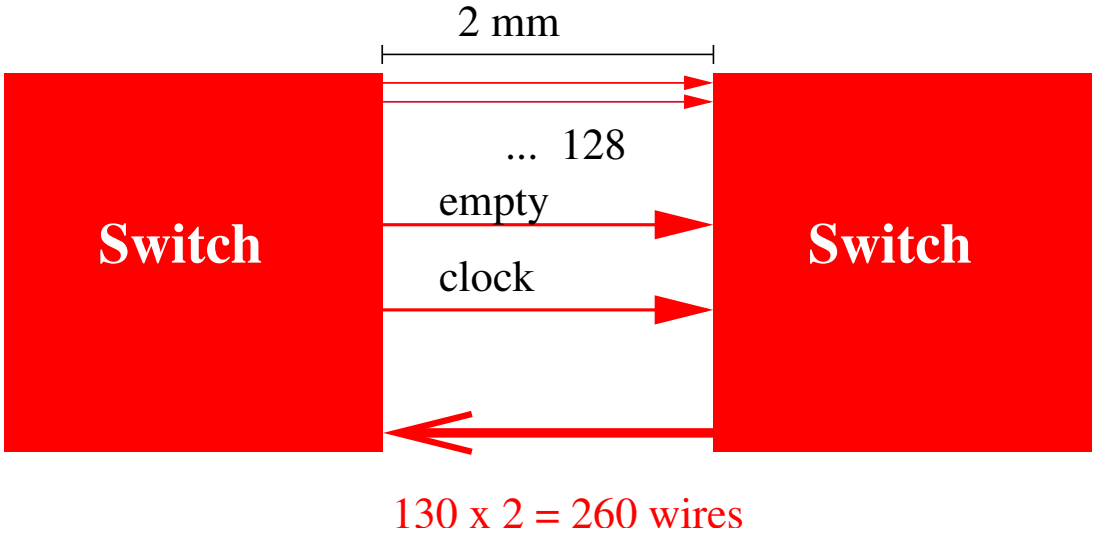
Communication is Key



Communication Layers:

- Physical layer: switch-to-switch and switch-to-resource
- Data link layer: switch-to-switch and switch-to-resource
- Network layer: resource-to-resource
- Session layer: process-to-process
- Application layer: application-to-application

Physical Layer



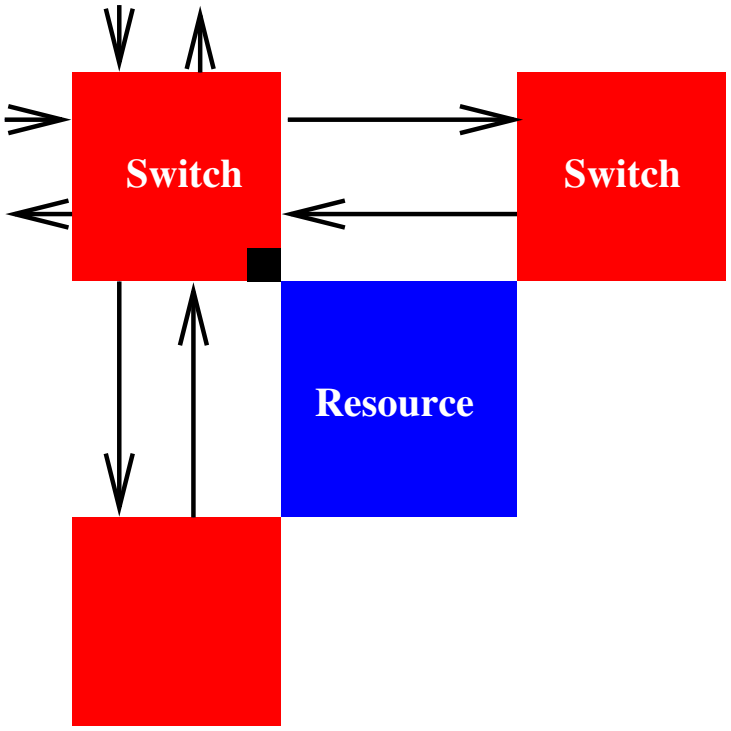
Parameters:

- Physical distance
- Number of lines
- Activity control
- Buffers and pipelining

Nostrum status:

- Channel dimension: $2mm \times 100\mu m$
- 128 data lines in each direction on 4 metal layers
- No pipelining
- On/off control for power saving

Data Link Layer



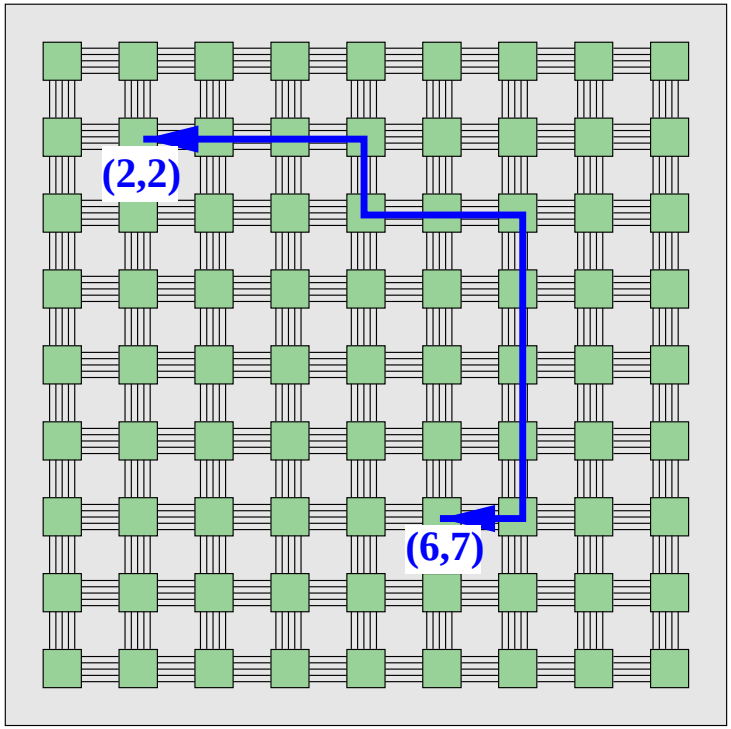
Parameters:

- Line frequency versus switch frequency
- Buffering
- Error correction
- Power optimization encoding

Nostrum status:

- Physical packet = data link packet
- Physical clock = data link clock
- Single packet input buffer
- Error correction
- On/off activity control

Network Layer



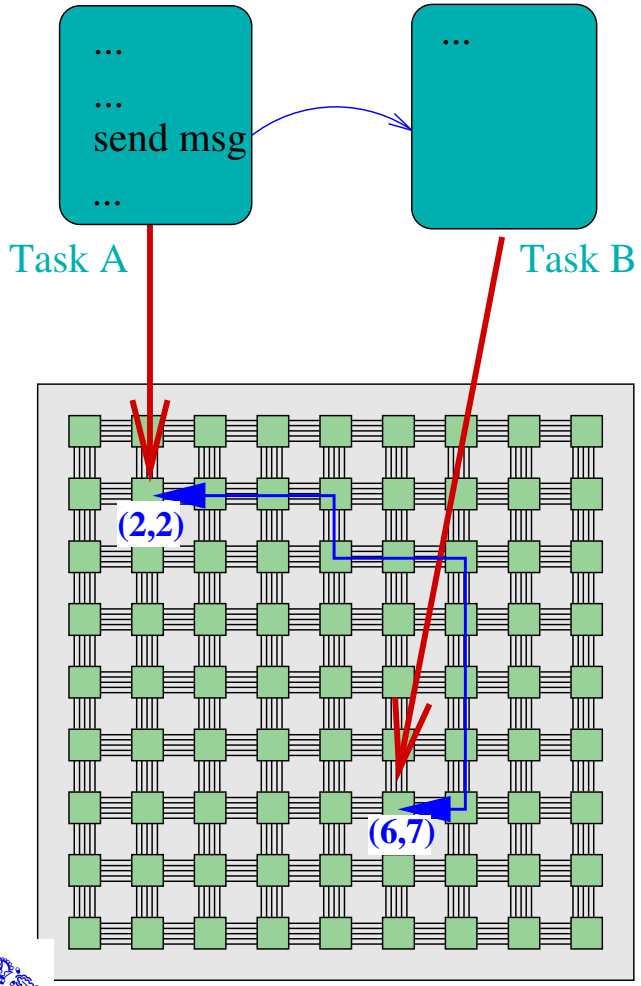
Parameters:

- Link layer cell size vs. network layer packet size
- Network address scheme
- Routing algorithm
- Priority classes
- Error correction

Nostrum status:

- Link layer packet = network layer packet
- Relative x-y addresses
- Deflective routing with no buffers and no routing tables
- Virtual circuits with guaranteed bandwidth and delays
- No error protection

Session Layer



Parameters:

- Task level communication primitives
- Message passing
- Shared memory based communication
- Synchronization
- Error correction

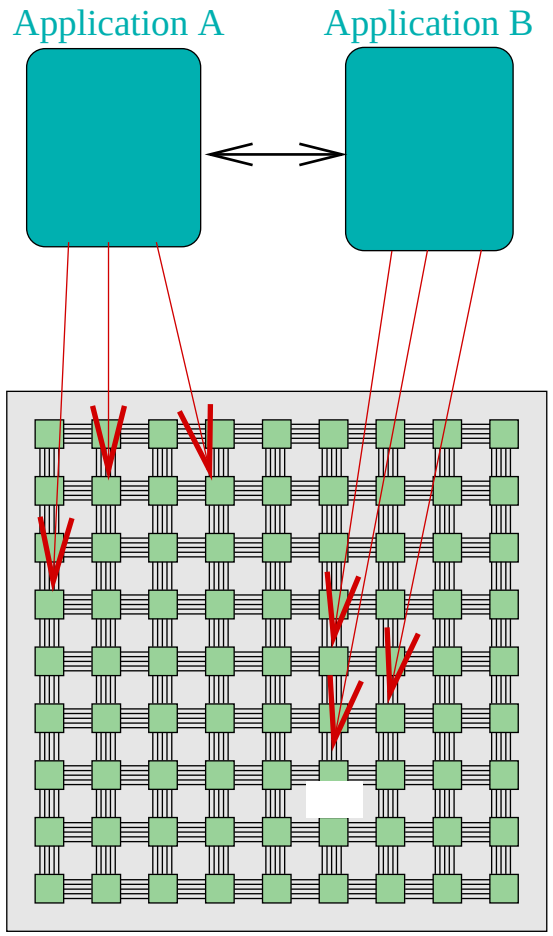
Nostrum status:

- Set of communication primitives defined
- Both message passing and shared memory
- User controlled synchronization
- Optional end-to-end data protection

Session Layer Communication

- Message passing communication:
  - `open/listen/accept/bind` primitives to open a channel
  - `send/receive` to communicate
  - `close` to tear down the channel
  - blocking/non-blocking send/receive
- Shared memory communication:
  - allocation
  - read/write
  - free
  - interruptible/non-interruptible
- VHDL,C and SystemC libraries under development

Application Layers



Application specific communication services;
E.g. the NoC operating system could use:

- Task/resource database access protocol
- Task migration protocol

Overview

Topology and Structure

Protocol Stack

The Network Layer and the Switch

Data Protection

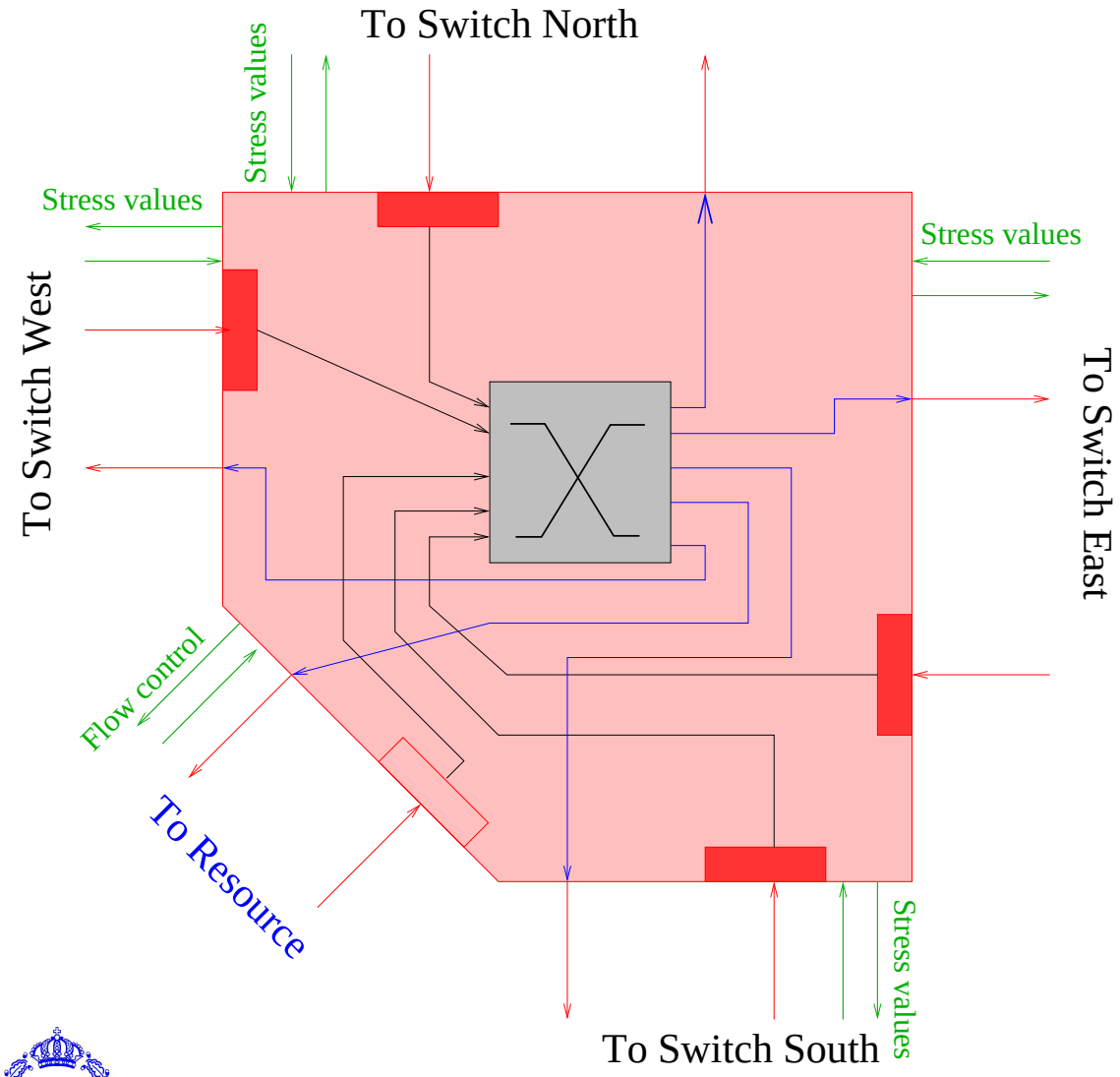
Simulation Environment

Clocking

The Network Layer

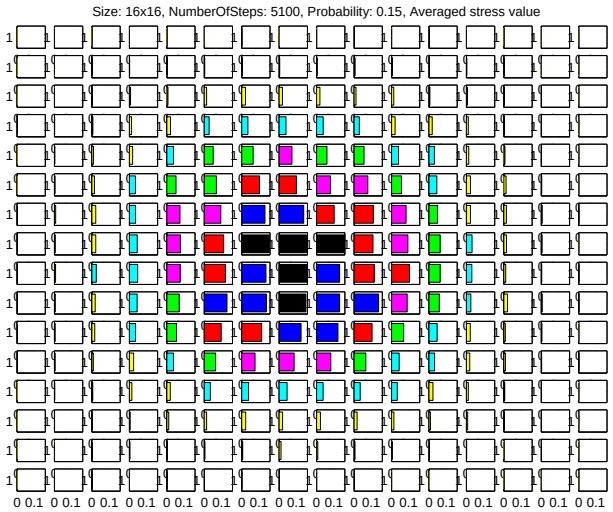
- Packet switched best effort service
  - ⋆ Packets are guaranteed to arrive
  - ⋆ Packet payload may be protected (4 levels of protection)
  - ⋆ Load dependable delay in the network
  - ⋆ Load dependable delay at the network access point
- Virtual circuit service
  - ⋆ Guaranteed bandwidth
  - ⋆ Guaranteed maximum delay
  - ⋆ Multicast circuits
  - ⋆ Based on packet switching service

The Bufferless Switch

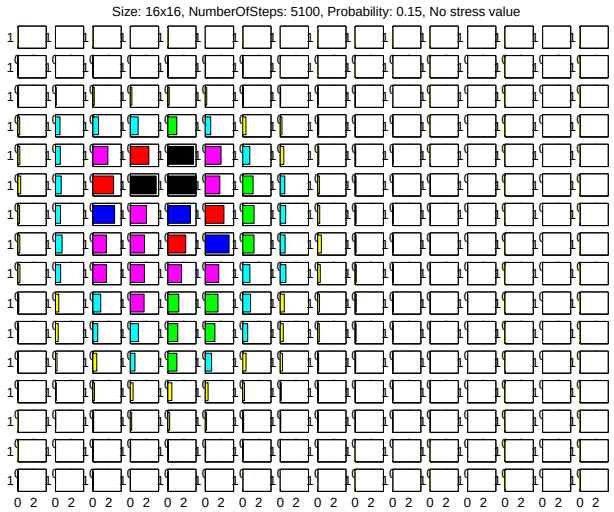


- \+ No buffers
- \+ No routing table
- \+ Small area
- \+ Short delay
- \+ Low power consumption
- – Non-shortest path
- – Header overhead due to destination address

Stress Value Effect on Buffer Sizes and Delays

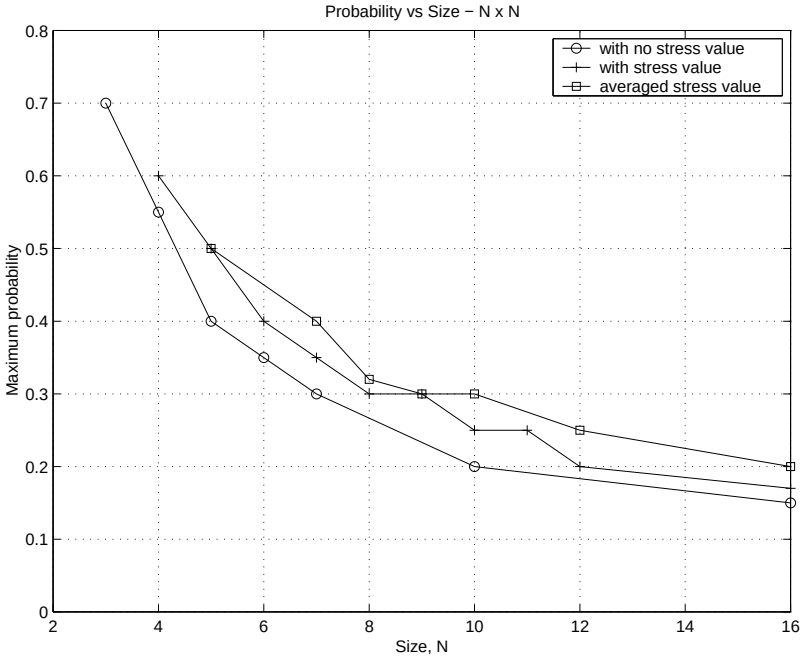


Largest average buffer size: 3.2



Largest average buffer size: 0.1

Stress Value Effect on Maximum Load



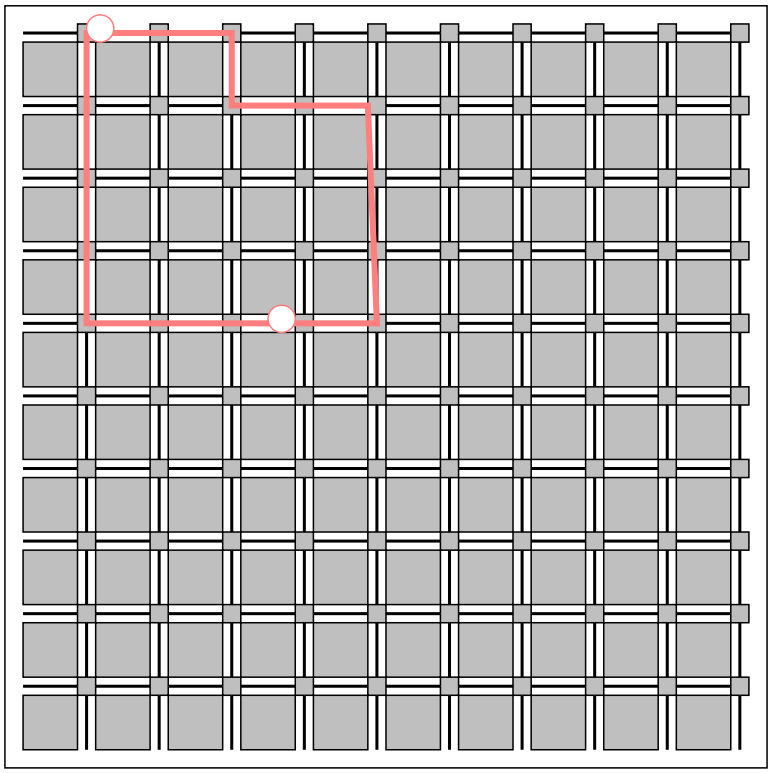
Looped Container based Virtual Circuit

- A container packet loops between two or more end points
- The looping container establish a closed virtual circuit
- The virtual circuit allows multicast and bus protocol emulation
- Possible bandwidth allocation:

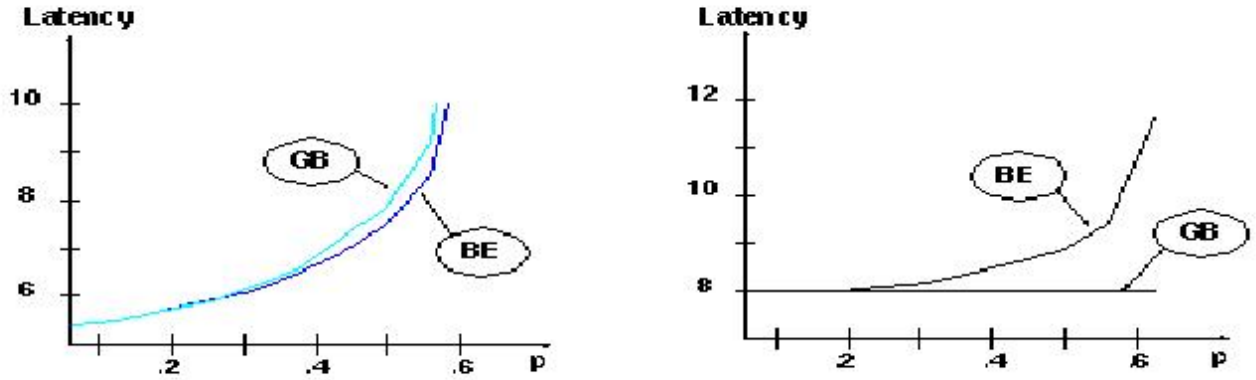
$$2^{j-d}B$$

where B = link bandwidth, d = length of the container loop, $1 \leq j \leq d$

- Examples:
  $d = 2$: possible allocations: 100% and 50%
  $d = 4$: possible allocations: 100%, 50%, 25%, .5%



Best Effort and Guaranteed Bandwidth Traffic



The background traffic and the AB traffic

Overview

Topology and Structure

Protocol Stack

The Network Layer and the Switch

Data Protection

Simulation Environment

Clocking

Data Protection

- Two level protection: Link layer and session layer
- Data link layer protection:
  - ⋆ SEC-DED header protection (16/26 bits)
  - ⋆ Four levels of payload protection:
    - ∗ Maximum bandwidth - no protection (102/102 bits)
    - ∗ Guaranteed integrity - DED protection (90/102 bits)
    - ∗ Minimum latency - SEC protection (90/102 bits)
    - ∗ High reliability - SEC-DED protection (81/102 bits)
- Session layer:
  - ⋆ Normal mode: Send-and-Forget (SaF) service
  - ⋆ Reliability mode: Acknowledgement-and-Retransmit (AaR) service
    - ∗ window size $N, 1 \leq N \leq 64$
    - ∗ $2N$ packets are buffered in sender and receiver
    - ∗ End-to-end flow control mechanism
- in total 8 modes available

Overview

Topology and Structure

Protocol Stack

The Network Layer and the Switch

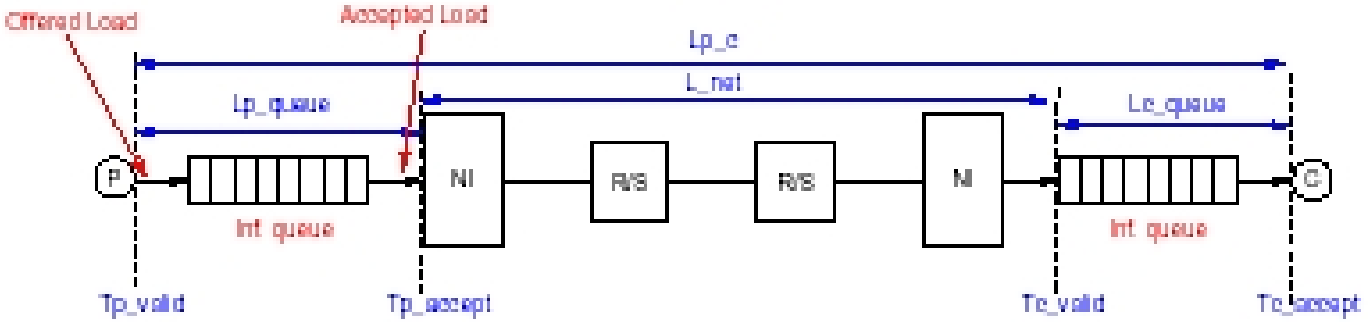
Data Protection

Simulation Environment

Clocking

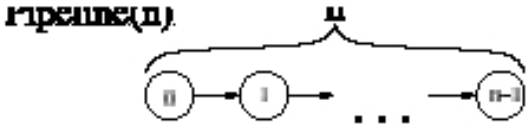
Measurement Framework

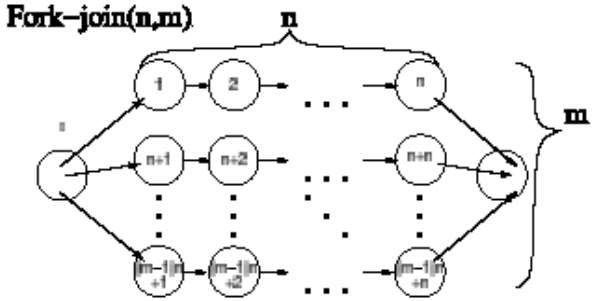
- Ponts of measurement
- Level of abstraction
- Service type

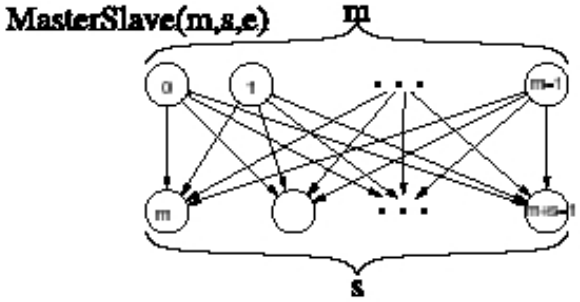


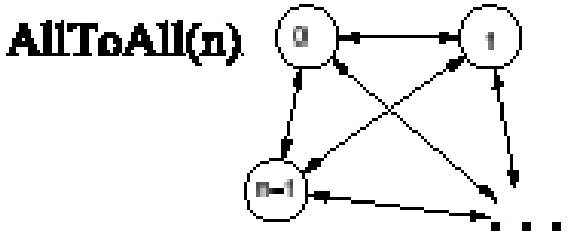
Workload Models

- Spatial patterns
- Spatial probabilit distributions
- Temporal probability distributions

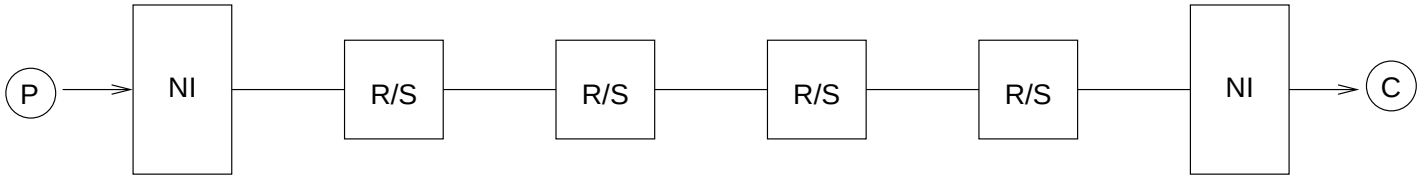




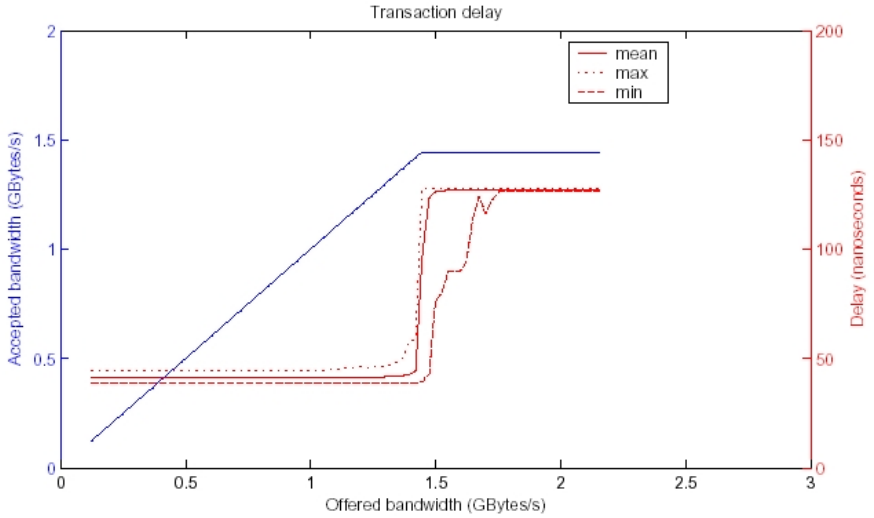


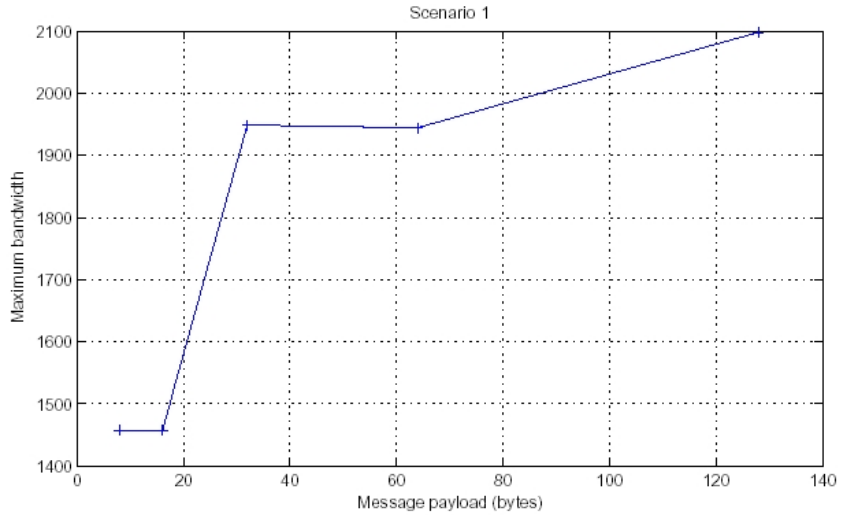


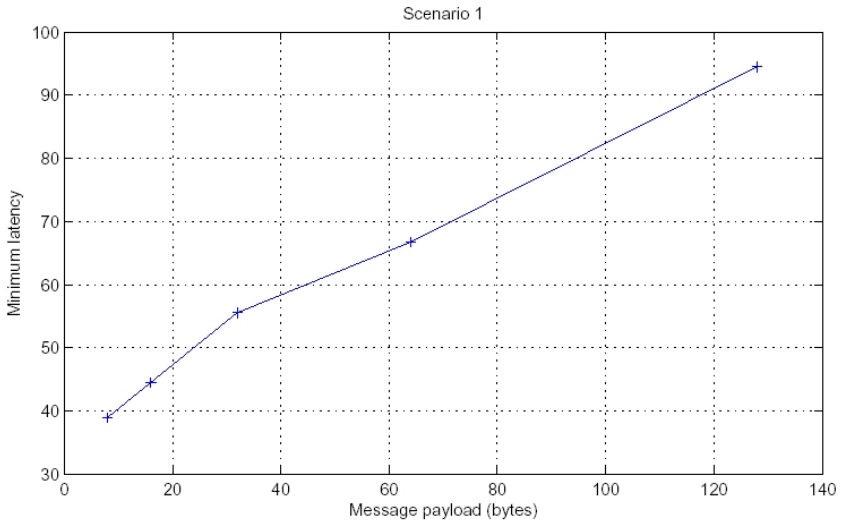
Simulation Scenario 1



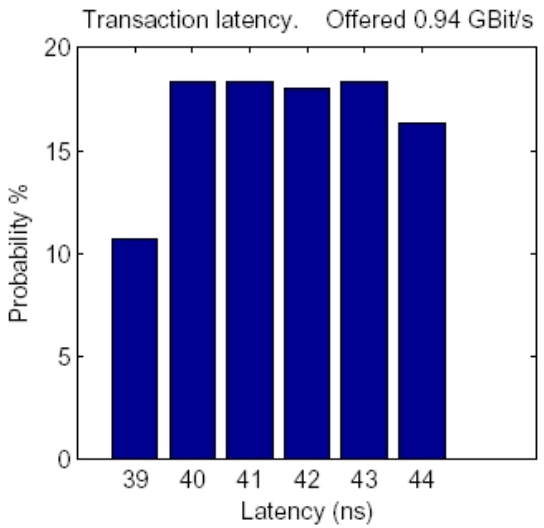
Simulation Scenario 1 - cont'd

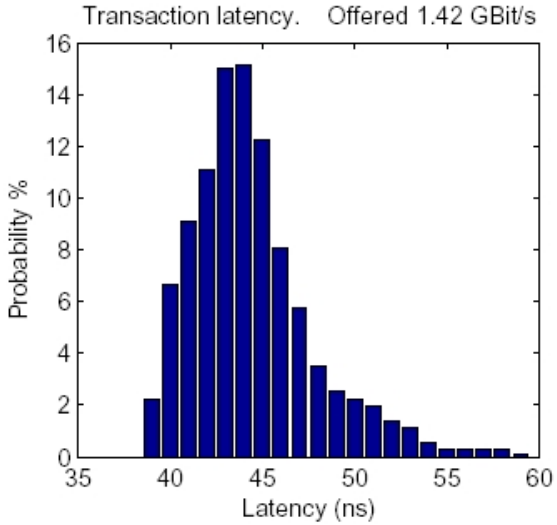


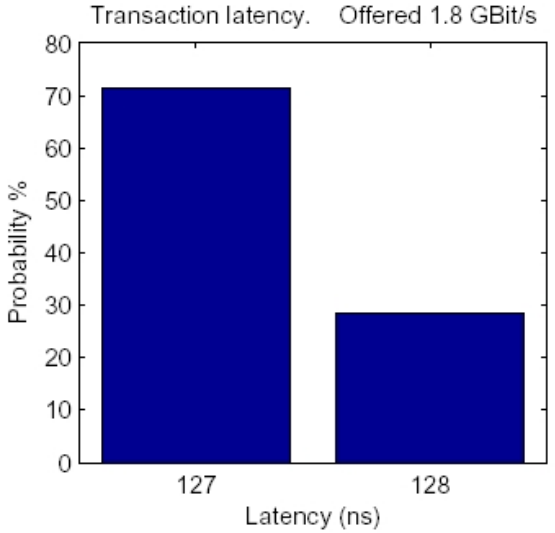




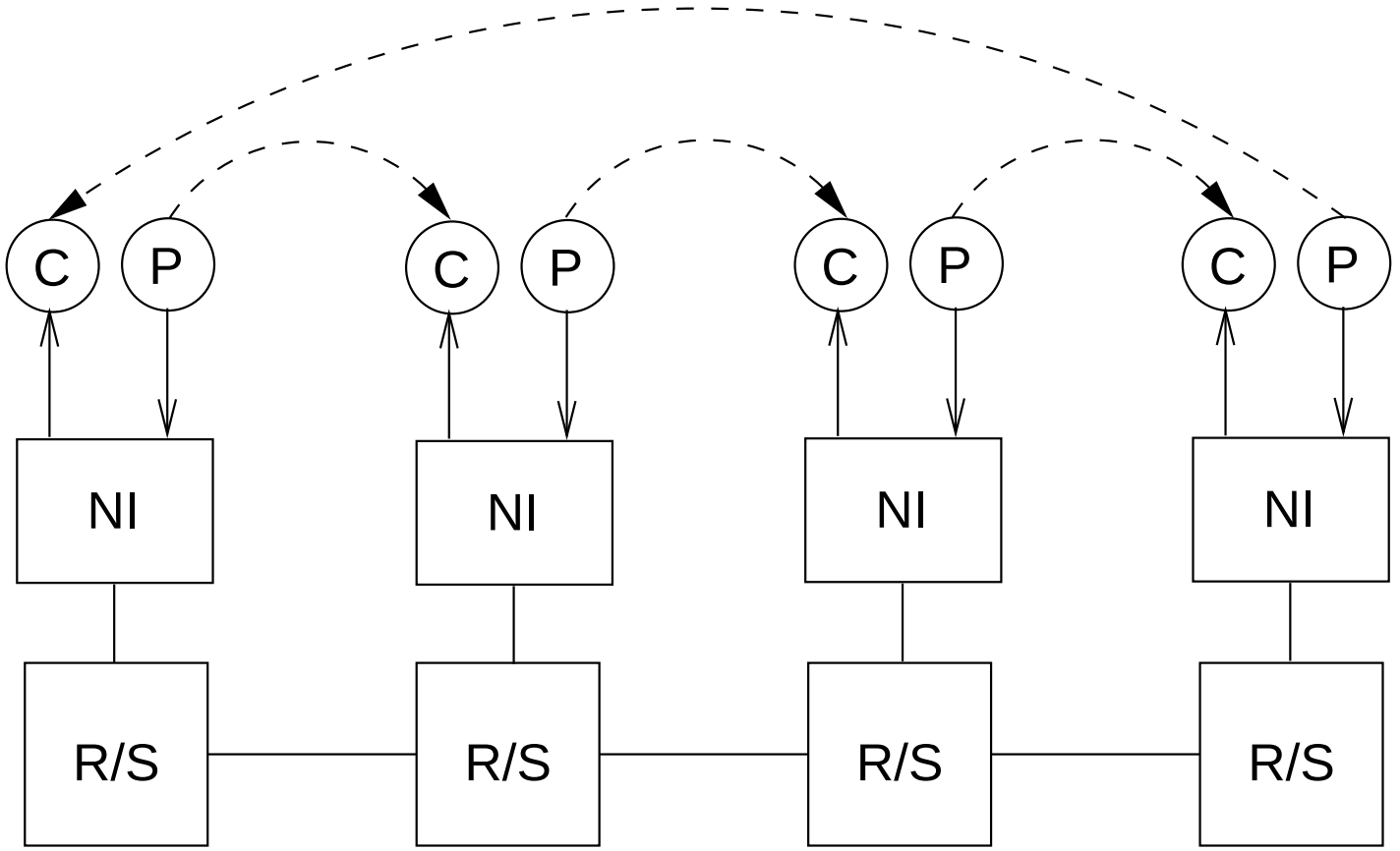
Simulation Scenario 1 - cont'd



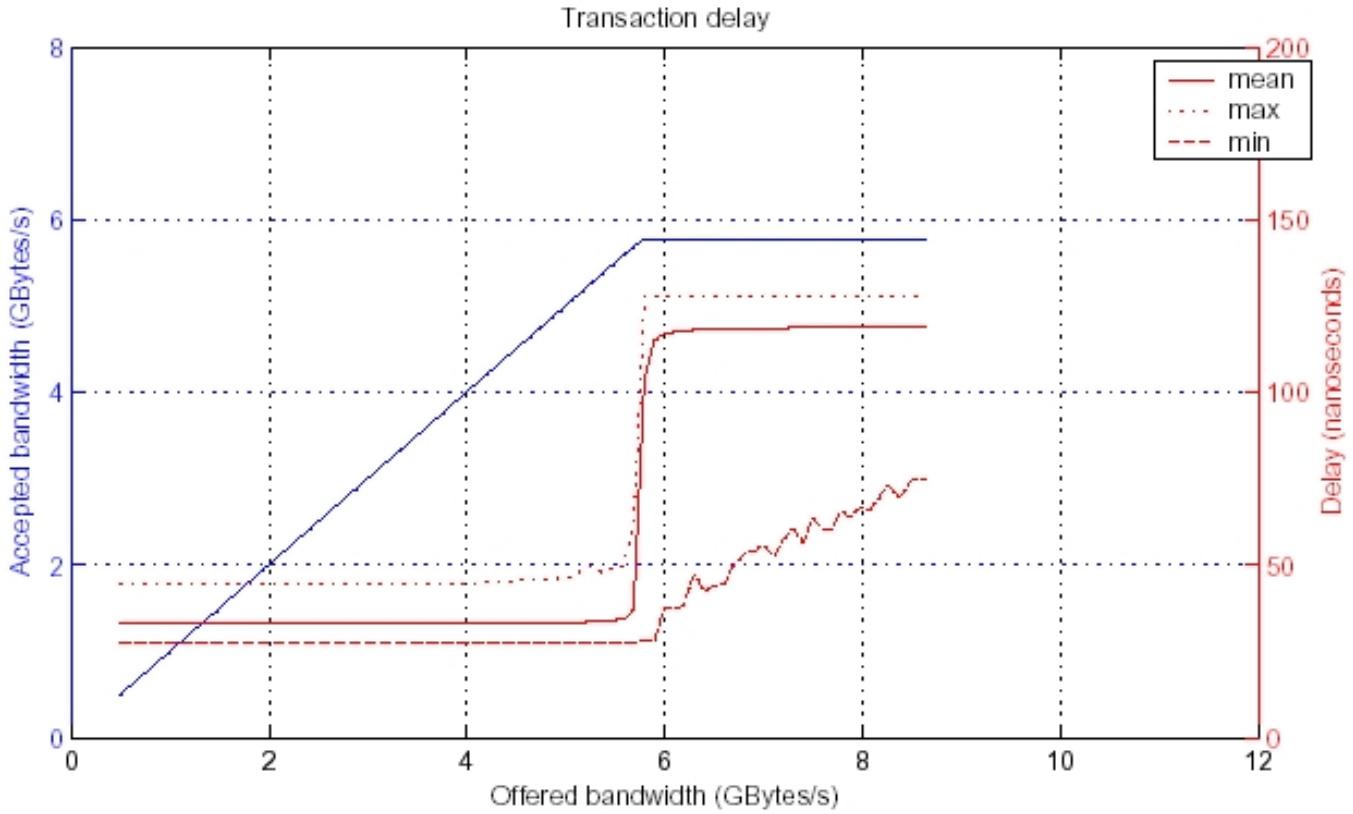




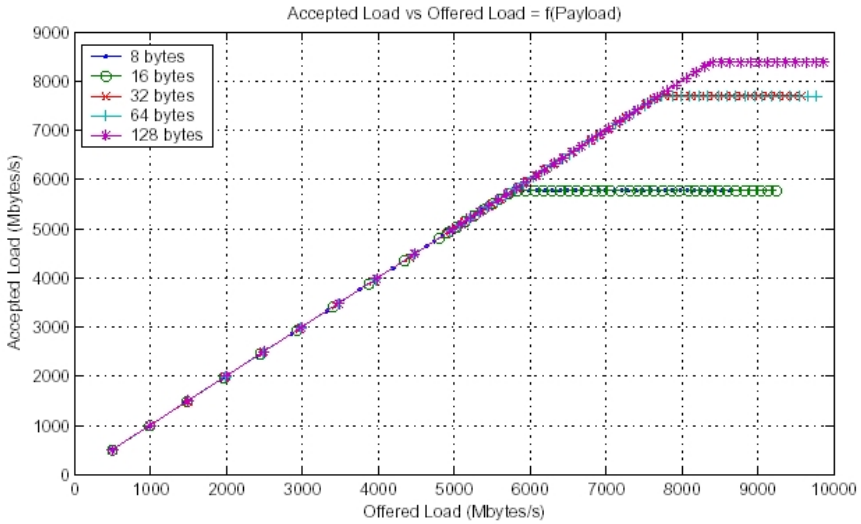
Simulation Scenario 2

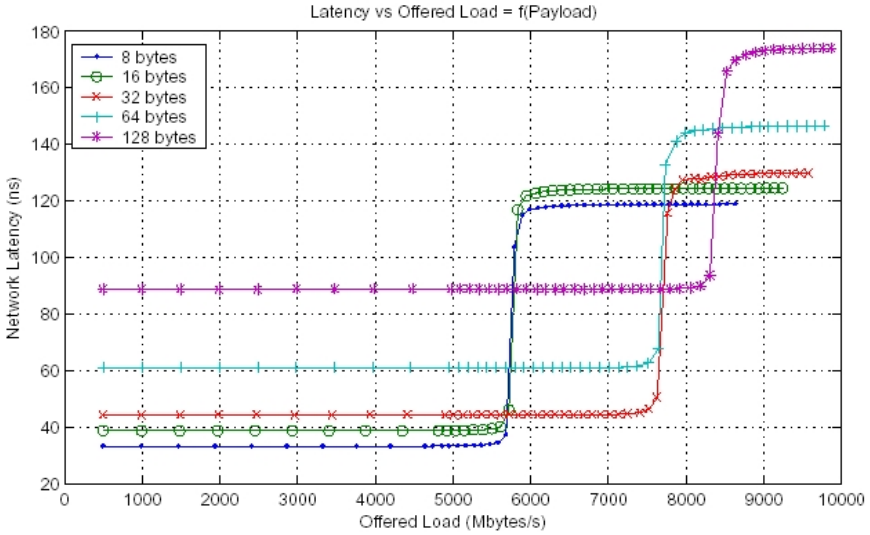


Simulation Scenario 2 - cont'd

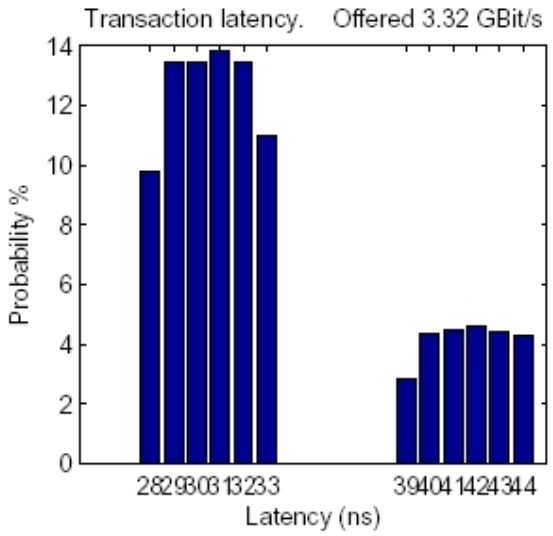


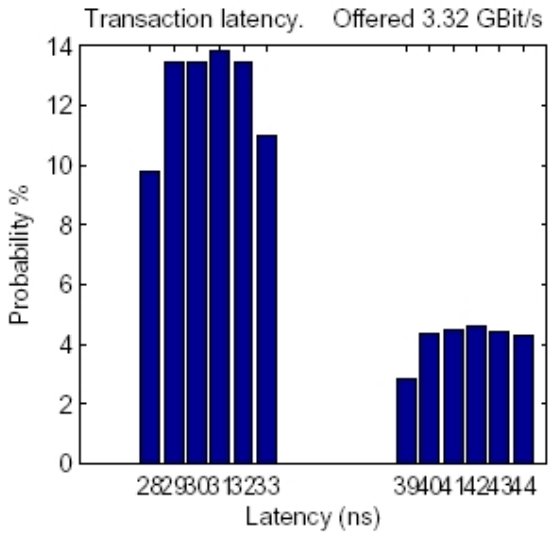
Simulation Scenario 2 - cont'd

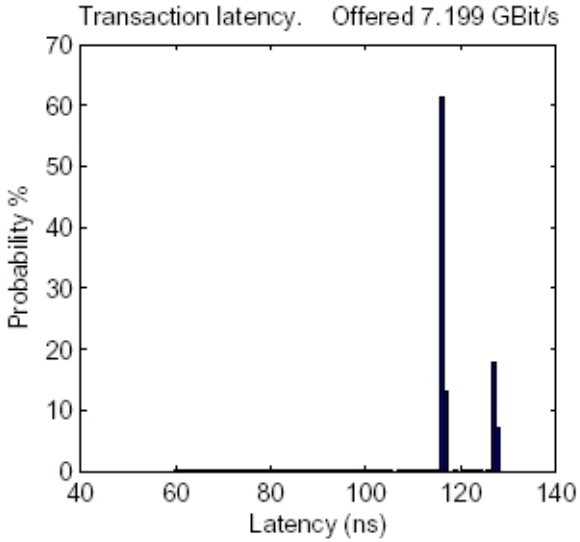




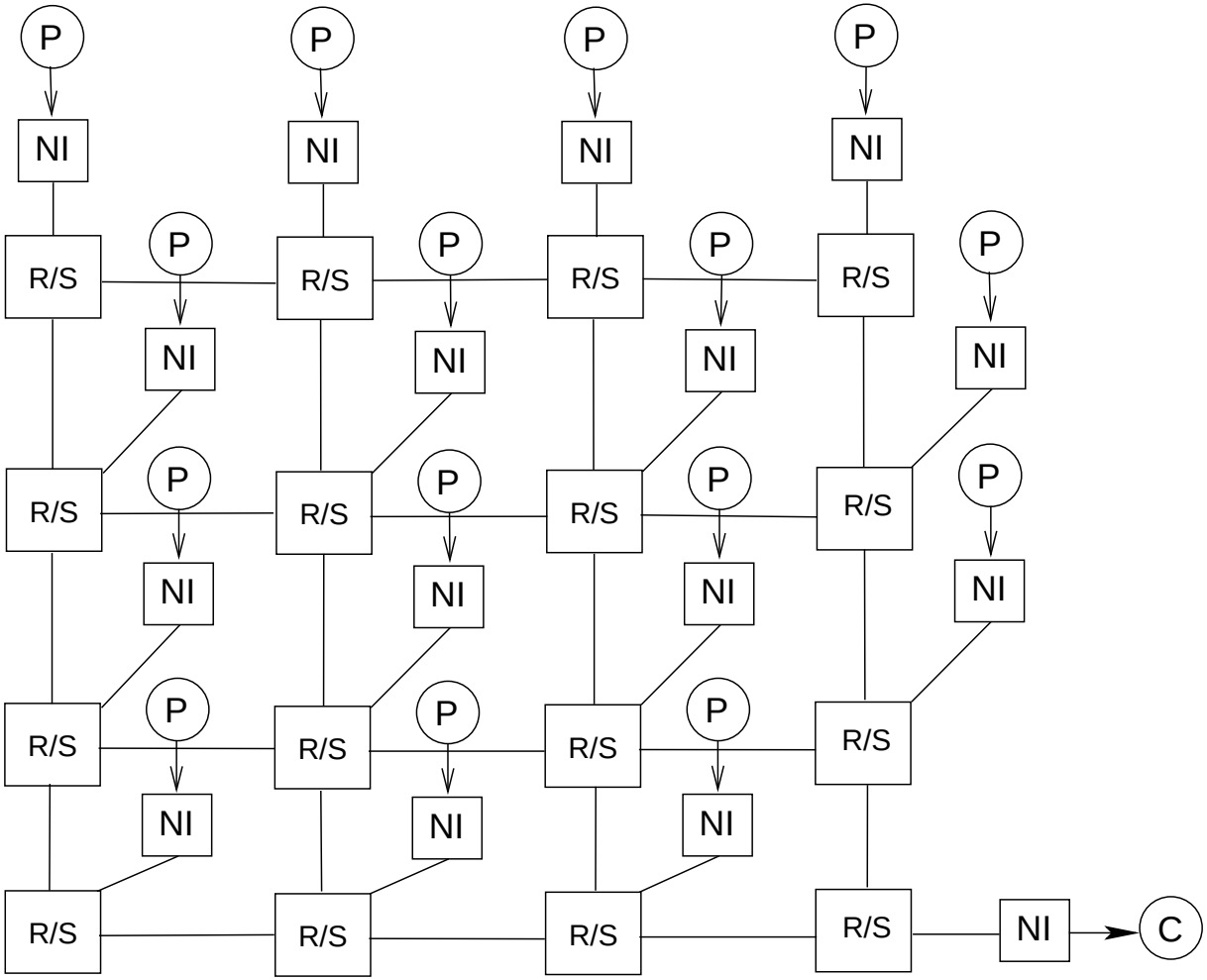
Simulation Scenario 2 - cont'd



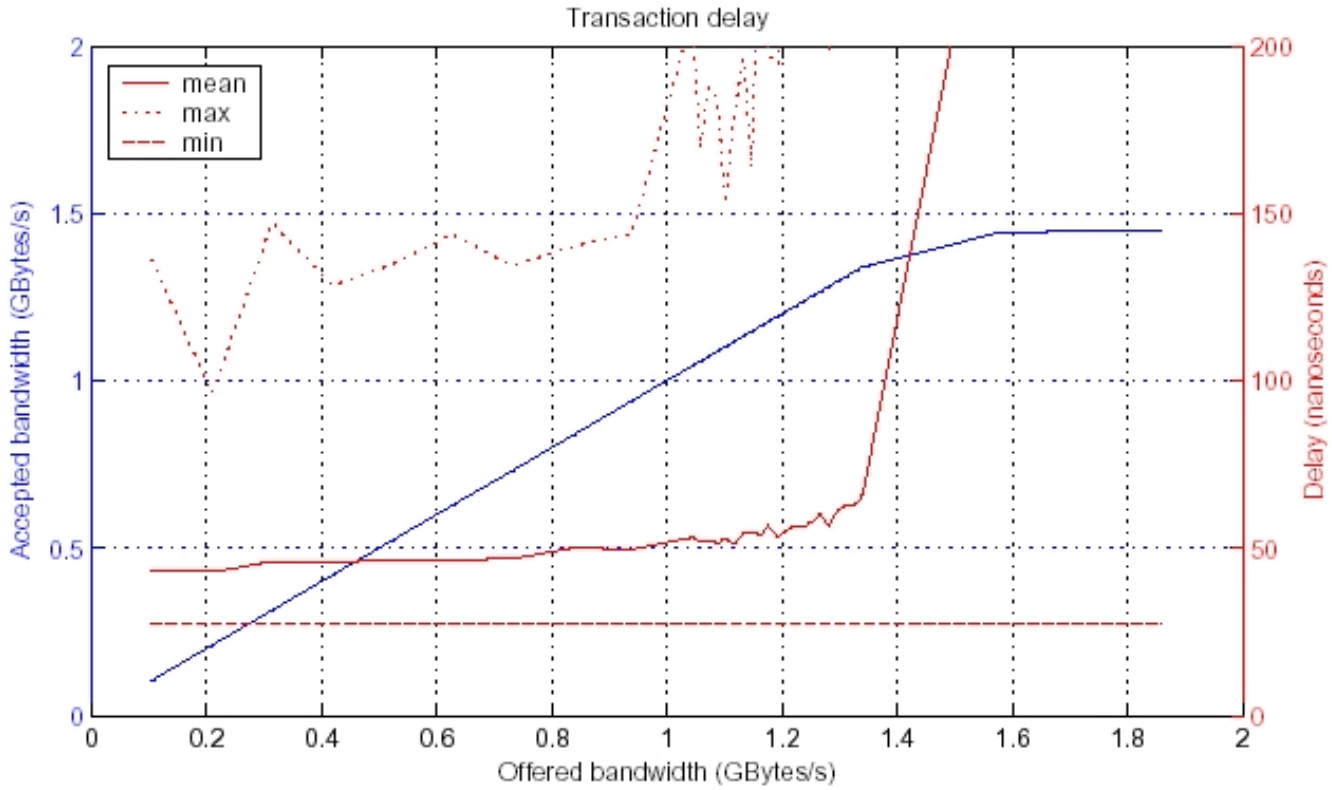




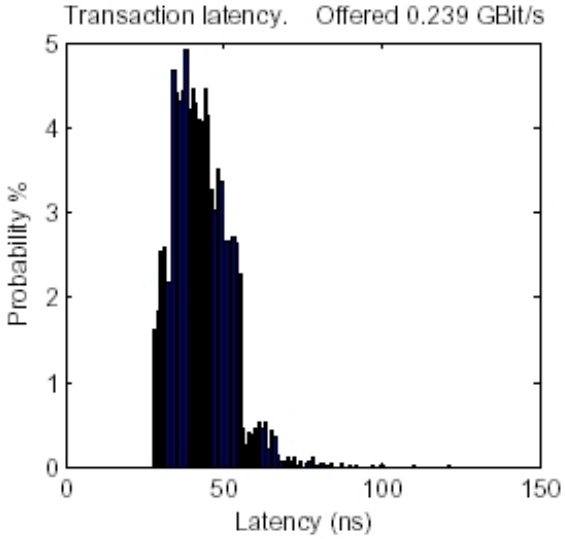
Simulation Scenario 3

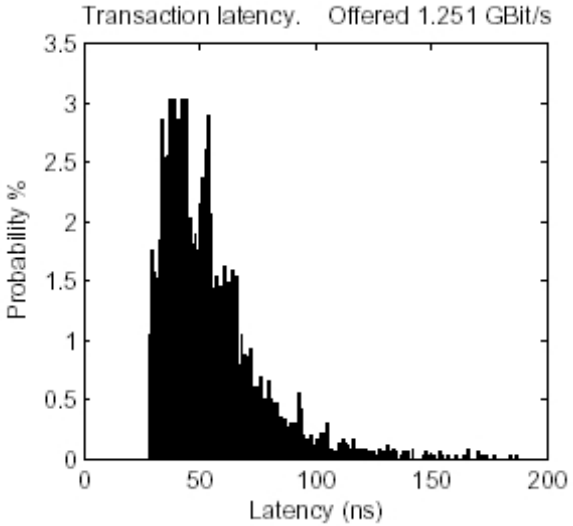


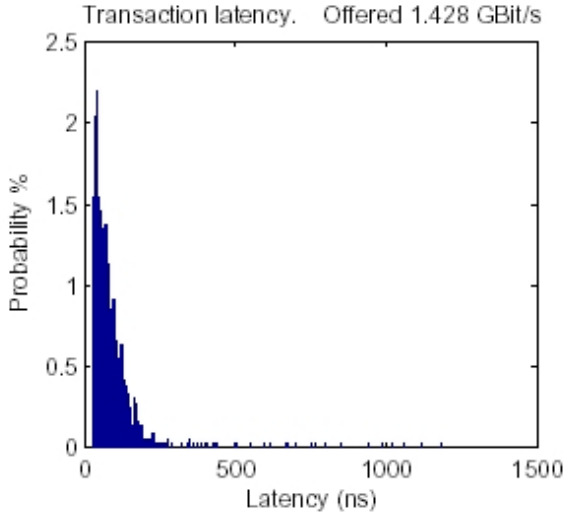
Simulation Scenario 3 - cont'd



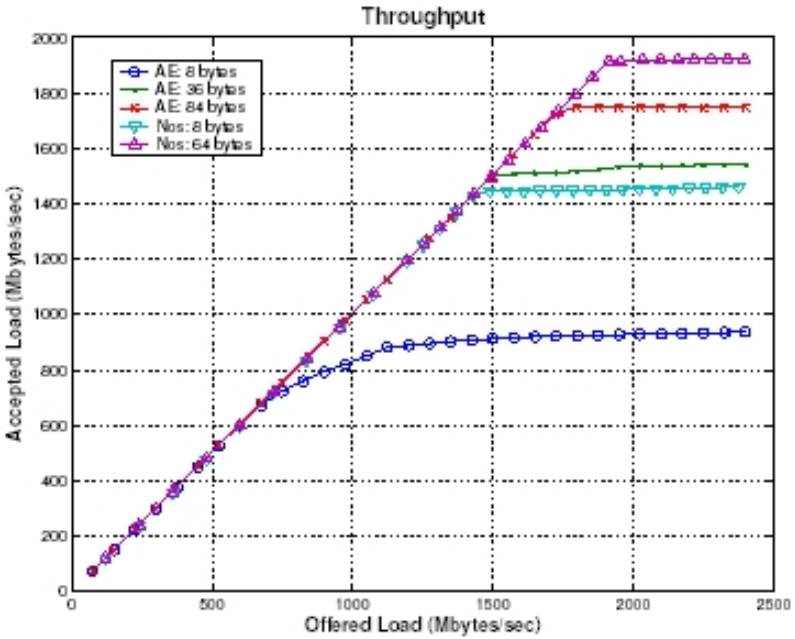
Simulation Scenario 3 - cont'd

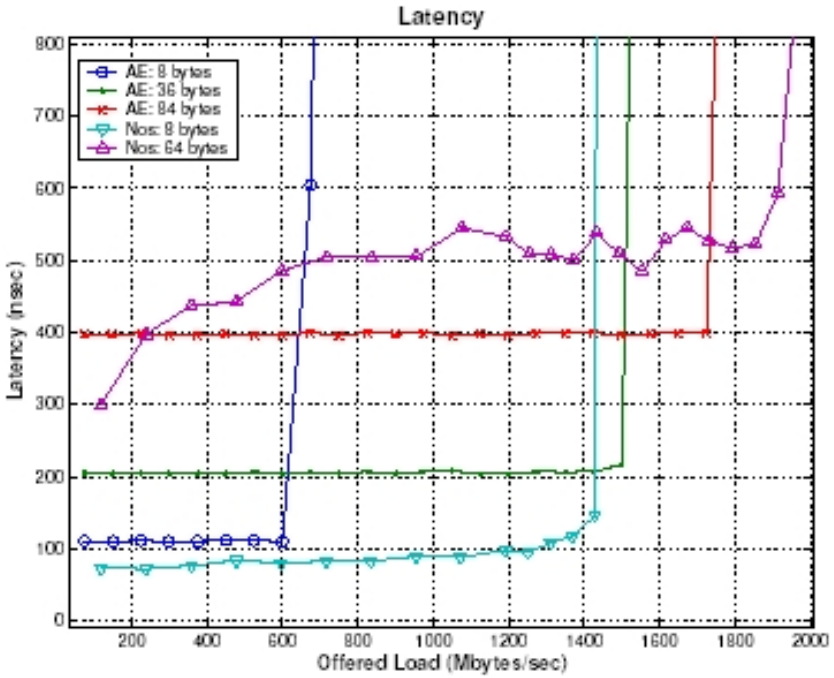






Aethereal and Nostrum





Overview

Topology and Structure

Protocol Stack

The Network Layer and the Switch

Data Protection

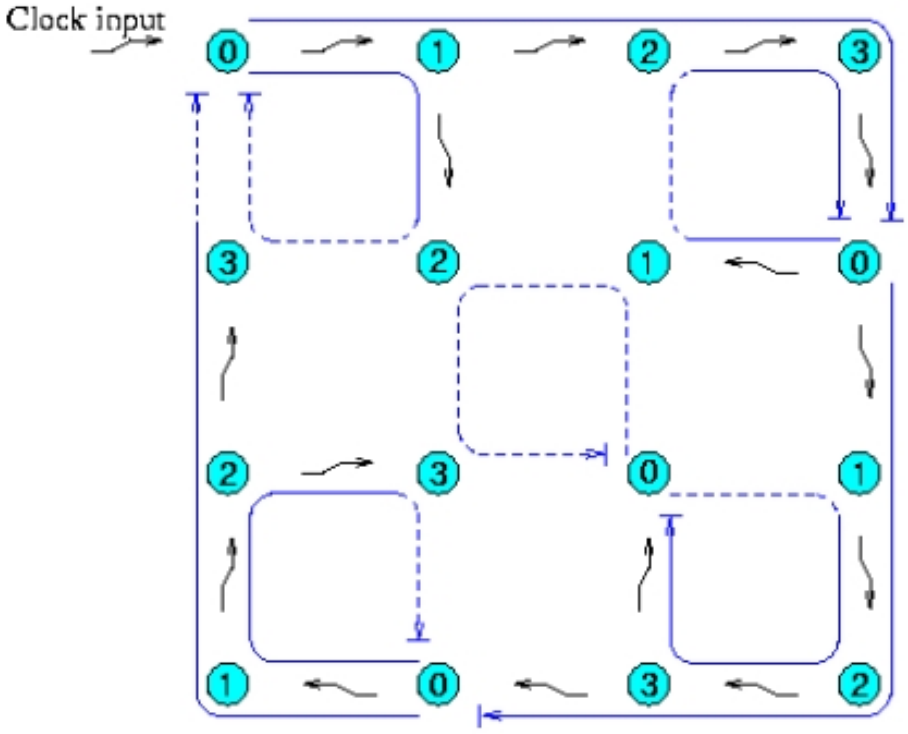
Simulation Environment

Clocking

Globally Pseudosynchronous - Locally Synchronous Clocking

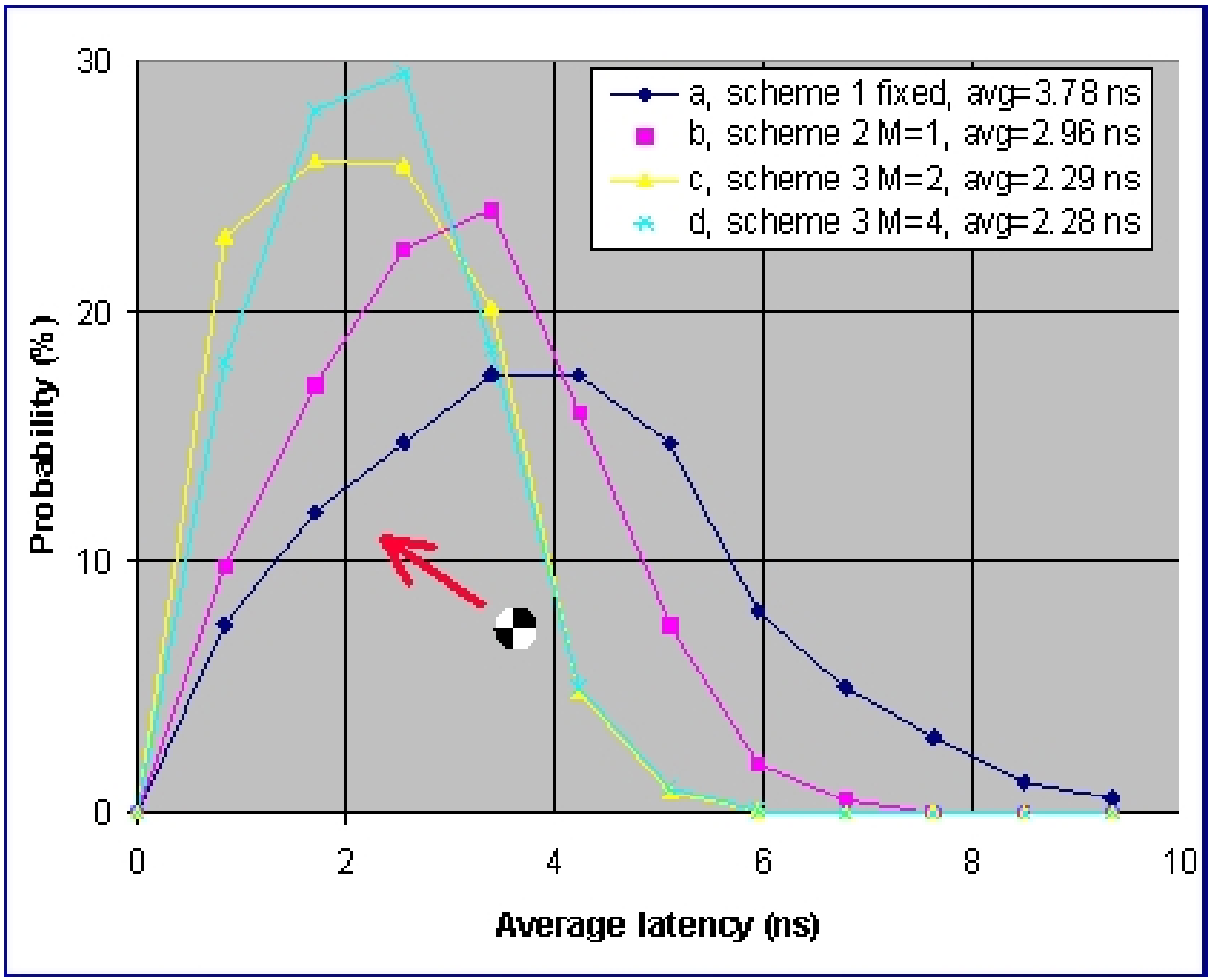
- Latency reduce with 29% at low load; 40% at high load
- Can handle 10% higher load
- More skew tolerant
- Clock skew and jitter is depending only on local constraints
- No global clock distribution with associated power gains
- Reduced peak power with 50% at best
- Jitter reduced significantly

Globally Pseudosynchronous Clocking - cont'd

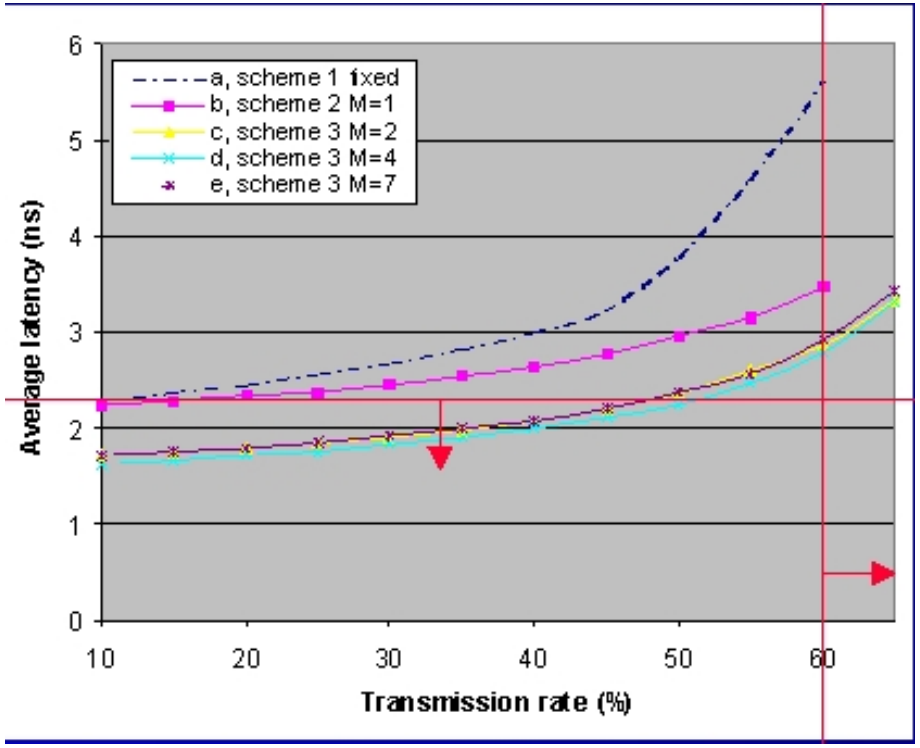


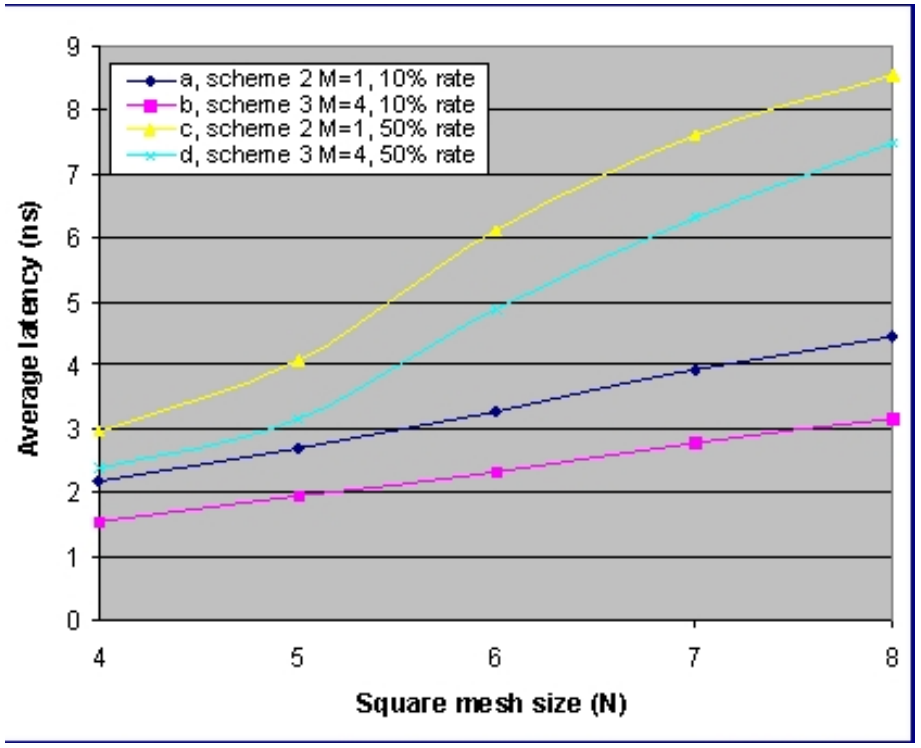
- Downstream data create low latency paths (Data Motorways)
  - ⋆ Guaranteed data motorways
  - ⋆ Phase related data motorways
- Periphery roundtrip:
  - ⋆ 14 cycles downstream
  - ⋆ 21 cycles upstream
  - ⋆ 24 cycles synchronous

Globally Pseudosynchronous Clocking - cont'd



Globally Pseudosynchronous Clocking - cont'd





Summary of Nostrum Status

- Nostrum defines a 2 D mesh topology;
- Protocol stack for link layer, network layer and session layer;
- Packet switched and virtual circuit communication services;
- Buffer-less, loss-less switch with no routing tables;
- 2 level data protection scheme;
- Session layer communication primitives;
- Flexible NoC Simulator;

Further information: www.imit.kth.se/info/FOFU/Nostrum/

Research Challanges

- Power management
  - ⋆ Error correction and fault tolerance
  - ⋆ Dynamic frequency and voltage scaling
  - ⋆ Integrated network-resource power management
- Reconciliation of dynamic power management with real-time constraints
- Admission protocol
- Communication refinement
- Application mapping and traffic planning
- Application designers interface and programmers model

Master Thesis Projects

- Comparison of wormwhole routing and virtual circuits
- Laypout planning with constraint logic programming
- Components in the Nostrum configuration tool
  - ⋆ Resources - VHDL and SystemC
  - ⋆ Switches - VHDL and SystemC
  - ⋆ Network interfaces - VHD and SystemC