Slide 0

A Router Architecture for Real-Time Point-to-Point Networks

Jennifer Rexford, John Hall, and Kang Shin

Real-Time Computing Laboratory
Electrical Engineering & Computer Science
University of Michigan, Ann Arbor

http://www.eecs.umich.edu/RTCL

Slide 1

Outline

- Motivation for real-time multicomputer networks
- Architecture for mixing time-constrained and best-effort traffic
- Traffic shaping and link scheduling for time-constrained traffic
- Performance evaluation and implementation complexity
- Conclusions and future work

Slide 2

Real-Time Parallel Applications

Real-time applications

- Industrial process control
- Computer-integrated manufacturing
- Fly-by-light/drive-by-wire
- Scientific visualization

→ large computational demands and tight timing constraints

Communication requirements

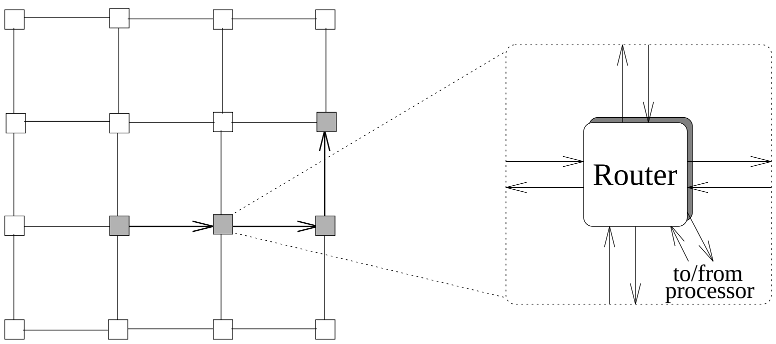
- Time-constrained traffic: (nearly) periodic; delay guarantees
- Best-effort traffic: aperiodic; good average latency

→ mixture of traffic with conflicting performance goals

Real-Time Computing Laboratory
The University of Michigan

Slide 3

Point-to-Point Multicomputer Networks



Real-Time Computing Laboratory
The University of Michigan

Slide 4

Real-Time Channels

***End-to-end* service for time-constrained connections**

- Traffic: rate r; maximum packet size; maximum burst size
- Guarantees: end-to-end delay $\leq D$; jitter bounds

$\rightarrow$ linear bounded arrival process with end-to-end delay guarantees

Admission control and route selection determine

- Feasible path through the network from source to destination
- Worst-case buffer requirements at each node
- Local delay bound $d \leq 1/r$ at each hop in the route

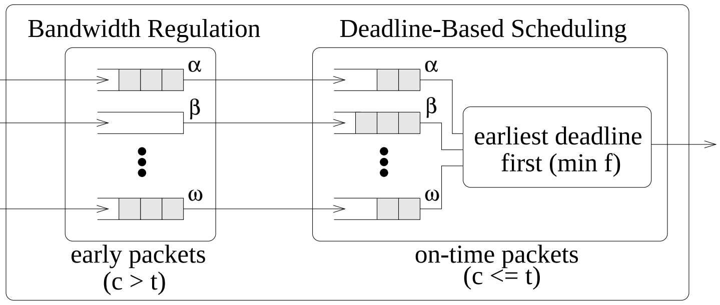
$\rightarrow$ performance guarantees compromising existing connections

Slide 5

Per-Link Run-Time Scheduling

Packet scheduling with rate r and local delay bound d

- Traffic shaping with conformance time $c = \max\{c + 1/r, t\}$
- Deadline-based scheduling with deadline $f = c + d$
- Per-link horizon parameter h for sending early traffic



Slide 6

Router Design Goals

- Wide range of end-to-end delay and bandwidth guarantees
- Low intrusion of best-effort traffic on time-constrained packets
- Good average performance for best-effort traffic
- Efficient hardware run-time link shaping/scheduling
- Simple software interface for connection establishment
- Integrated single-chip solution with multiple outgoing links

Slide 7

Router Architectural Parameters

- **Topology**: how nodes are interconnected
- **Routing**: which links/nodes a packet traverses
- **Switching**: how a packet flows between nodes
- **Buffer architecture**: where waiting packets are stored
- **Flow control**: how to limit traffic to avoid buffer overflow
- **Scheduling**: selecting which packet the link transmits next

Slide 8

Switching Schemes

- **Circuit switching**
  → guaranteed network access; limited bandwidth sharing
- **Packet switching**
  → control over scheduling and bandwidth allocation
- **Virtual cut-through switching**
  → low latency; reduced buffer requirements
- **Wormhole switching**
  → low latency; low buffer requirements; some unpredictability

→ packet switching for time-constrained packets; wormhole switching for best-effort traffic

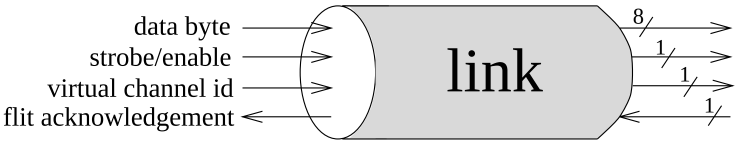
Real-Time Computing Laboratory
The University of Michigan

Slide 9

Mixing Time-Constrained and Best-Effort Traffic

- Tailor policies to time-constrained and best-effort traffic
- Provide separate *virtual channels* (lanes) for each traffic class
- Divide wormhole packets into small *flits* (flow control units)
- Enforce fine-grain arbitration between virtual channels

→ best-effort traffic can use wormhole switching and large packets



Real-Time Computing Laboratory
The University of Michigan

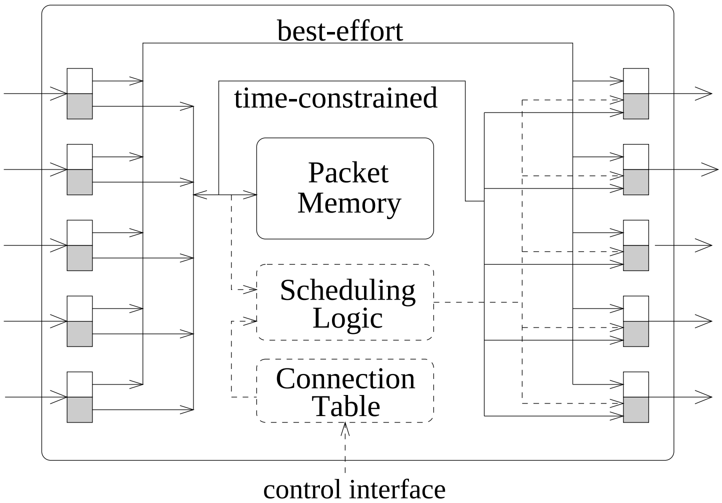
Slide 10

Real-Time Router Architecture

	Time-Constrained	Best-Effort
Switching	Packet switching	Wormhole switching
Packet size	Small, fixed size	Variable length
Scheduling	Deadline-driven	Round-robin on inputs
Routing	Table-driven multicast	Dimension-ordered
Buffers	Shared output queues	Flit buffers at inputs
Flow control	Rate-based	Flit acknowledgements

Slide 11

Real-Time Router Architecture



Real-Time Computing Laboratory
The University of Michigan

Slide 12

Memory and Control

Packet memory

- Fixed-size packet buffers
- Shared memory amongst incoming and outgoing links
- Wide bus and memory for high bandwidth
- Free-list of available memory addresses

Connection table

- Bit mask of outgoing links/ports
- Local delay bound d
- Outgoing connection id

Real-Time Computing Laboratory
The University of Michigan

Slide 13

Scheduling Time-Constrained Traffic

Implementation goals:

- Overlap scheduling with packet transmission
- Limit size of clock t and handle clock rollover
- Limit complexity of shaping/scheduling logic

Architectural techniques:

- Integrate traffic shaping (sorting by conformance time c)
  with link scheduling (sorting by deadline f)
- Integrate scheduling logic for the multiple outgoing links/ports
- Pipeline access to the shared scheduling unit

→ efficient, single-chip solution

Real-Time Computing Laboratory
The University of Michigan

Slide 14

Traditional Link-Scheduling Model

- **Queue 1:** on-time time-constrained packets ($c \leq t$)
  $\rightarrow$ priority queue (by deadline f)
- **Queue 2:** best-effort packets
  $\rightarrow$ first-in-first-out queue
- **Queue 3:** early time-constrained packets ($c > t$)
  $\rightarrow$ priority queue (by conformance time c)

Algorithm

```
if (queue_1 is non-empty)
  transmit packet from head of queue_1 (minimum f);
else if (queue_2 is non-empty)
  transmit packet from head of queue_2;
else if ((queue_3 is non-empty) and (head has c-t<=h))
  transmit packet from head of queue_3 (minimum c);
else
  do not transmit any packet;
```

Real-Time Computing Laboratory
The University of Michigan

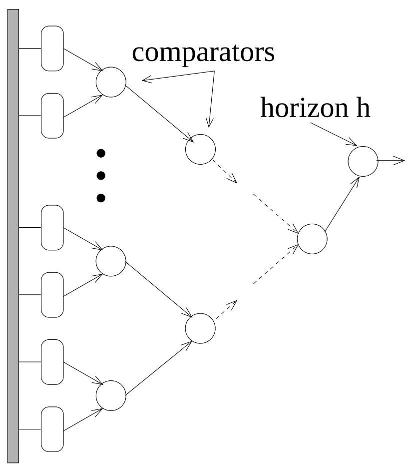
Slide 15

Scheduling Architecture

Packet sorting key

On-time	0	0	f - t
Early	0	1	c - t
Ineligible	1	–	——

Pipelined comparator tree

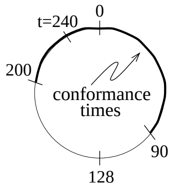


Real-Time Computing Laboratory
The University of Michigan

Slide 16

Handling Clock Rollover

Trade-off between clock range and flexibility in selecting d and h



Packets queued for link j after link $j-1$

- **Arrival:** link $j-1$ can transmit a packet with delay bound d_{j-1} up to $(h_{j-1}+d_{j-1})$ time units early
- **Departure:** link j can transmit a packet with delay bound d_j as late as its deadline $f = c + d_j$

$\rightarrow$ connection can have packets with $c \in [t - d_j, t + (h_{j-1} + d_{j-1})]$

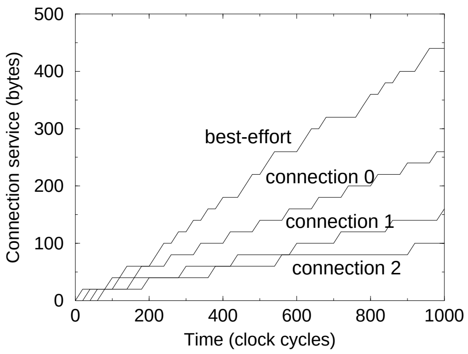
Real-Time Computing Laboratory
The University of Michigan

Slide 17

Performance Evaluation

Backlogged traffic:

- Best-effort packets
- 0: $d = 3$, $1/r = 4$
- 1: $d = 5$, $1/r = 7$
- 2: $d = 8$, $1/r = 9$



Real-Time Computing Laboratory
The University of Michigan

Slide 18

Router Chip Implementation

Architectural parameters

- Capacity: 256 connections and 256 20-byte packets;
  10-byte flit buffer for each link
- Range: 8-bit clock and 9-bit packet sorting keys
- Sorting: 2-stage pipeline in comparator tree

Chip complexity

- Tools: Verilog and Epoch Silicon Compiler
- Size: 123 pins and $905,104$ transistors
- Area: 8.1 mm $\times$ 8.7 mm in 0.5μm 3-metal CMOS process
- Speed: 50 MHz clock, power of 2.3 watts

Real-Time Computing Laboratory
The University of Michigan

Slide 19

Conclusions:

- Tailoring architecture to time-constrained and best-effort traffic
- Integrating traffic shaping and link scheduling
- End-to-end performance guarantees for time-constrained traffic

$\rightarrow$ efficient, single-chip design for real-time point-to-point networks

Ongoing work:

- Simulation model in `pp-mess-sim` multicomputer simulator
- Performance evaluation in multi-hop configurations
- Balance of space and time trade-offs in comparator tree
- Scalable integration of traffic shaping and link scheduling
  (Rexford, Bonomi, Greenberg, & Wong paper)

$\rightarrow$ balance between powerful algorithms and efficient architectures

Real-Time Computing Laboratory
The University of Michigan