Indirect Adaptive Routing on Large Scale Interconnection Networks

Nan Jiang, William J. Dally
Computer System Laboratory
Stanford University

John Kim
Korean Advanced Institute of Science and Technology
Overview

• Indirect adaptive routing (IAR)
  – Allow adaptive routing decision to be based on local and remote congestion information

• Main contributions
  – Three new IAR algorithms for large scale networks
  – Steady state and transient performance evaluations
  – Impact of network configurations
  – Cost of implementation
Presentation Outline

• Background
  – The dragonfly network
  – Adaptive routing
• Indirect adaptive routing algorithms
• Performance results
• Implementation considerations
The Dragonfly Network

- High Radix Network
  - High radix routers
  - Small network diameter

- Each router
  - Three types of channels
  - Directly connected to a few other groups

- Each group
  - Organized by a local network
  - Large number of global channels (GC)

- Large network with a global diameter of one
Routing on the Dragonfly

- Minimal Routing (MIN)
  1. Source local network
  2. Global network
  3. Destination local network

- Some Adversarial traffic congests the global channels
  - Each group $i$ sends all packets to group $i+1$

- Oblivious solution: Valiant’s Algorithm (VAL)
  - Poor performance on benign traffic
Adaptive Routing

- Choose between the MIN path and a VAL path at the packet source [Singh'05]
  - Decision metric: path delay
  - Delay: product of path distance and path queue depth

- Measuring path queue length is unrealistic
- Use local queues length to approximate path
  - Require stiff backpressure
Adaptive Routing: Worst Case Traffic
Indirect Adaptive Routing

- Improve routing decision through remote congestion information
- Previous method:
  - Credit round trip [Kim et. al ISCA’08]
- Three new methods:
  - Reservation
  - Piggyback
  - Progressive
Credit Round Trip (CRT)

- Delay the return of local credits from the congested router
- Creates the illusion of stiffer backpressure

**Drawbacks**
- Remote congestion is still inferred through local queues
- Information not up to date

[Kim et. al ISCA’08]
Reservation (RES)

- Each global channel track the number of incoming MIN packets
- Injected packets creates a reservation flit
- Routing decision based on the reservation outcome

Drawbacks
- Reservation flit flooding
- Reservation delay
Piggyback (PB)

- Congestion broadcast
  - Piggybacking on each packet
  - Send on idle channels
- Congestion data compression

- Drawbacks
  - Consumes extra bandwidth
  - Congestion information not up to date (broadcast delay)
Progressive (PAR)

- MIN routing decisions at the source are not final
- VAL decisions are final
- Switch to VAL when encountering congestion

- Draw backs
  - Need an additional virtual channel to avoid deadlock
  - Add extra hops
Experimental Setup

• Fully connected local and global networks
  – 33 groups
  – 1,056 nodes
• 10 cycle local channel latency
• 100 cycle global channel latency
• 10-flit packets
Steady State Traffic: Uniform Random

- Piggyback
- Credit Round Trip
- Progressive
- Reservation
- Minimal

Packet Latency (Simulation cycles)

Throughput (Flit Injection Rate)
Steady State Traffic: Worst Case

Throughput (Flit Injection Rate) vs. Packet Latency (Simulation cycles) for different traffic patterns:
- Piggyback
- Credit Round Trip
- Progressive
- Reservation
- Valiant’s
Transient Traffic: Uniform Random to Worst Case

**Average Packet Latency per Cycle - UR to WC**

- **Progressive**
- **Piggyback**

**% Packets Routing Non-minimally per Cycle - UR to WC**

- **Progressive**
- **Piggyback**
Network Configuration Considerations

• Packet size
  - RES requires long packets to amortize reservation flit cost
  - Routing decision is done on per packet basis

• Channel latency
  - Affects information delay (CRT, PB)
  - Affects packet delay (PAR, RES)

• Network size
  - Affects information bandwidth overhead (RES, PB)

• Global diameter greater than one
  - Need to exchange congestion information on the global network
Cost Considerations

• Credit round trip
  – Credit delay tracker for every local channel

• Reservation
  – Reservation counter for every global channel
  – Additional buffering at the injection port to store packets waiting for reservation

• Piggyback
  – Global channel lookup table for every router
  – Increase in packet size

• Progressive
  – Extra virtual channel for deadlock avoidance
Conclusion

- Three new indirect adaptive routing algorithms for large scale networks
- Performance and design evaluation of the algorithms

- Best Algorithm?
  - Piggyback performed the best under steady state traffic
  - Progressive responded fastest to transient changes

  - Network configurations will affect some algorithm performance
  - Cost of implementation
Thank You!

• Questions?
Adaptive Routing: Uniform Traffic

Packet Latency - Simulation cycles

Throughput - Flit Injection Rate

VAL
MIN
Adaptive
Transient Traffic: Worst Case to Uniform Random

Average Packet Latency per Cycle - WC to UR

Packet Latency vs. Cycles After Transition

% Packets Routing Non-minimally per Cycle - WC to UR

% of Packets Routing Non-minimally vs. Cycles After Transition
Transient Traffic: Worst Case 1 to Worst Case 10

Average Packet Latency per Cycle - WC1 to WC10

- PAR
- PB

% Packets Routing Non-minimally per Cycle - WC1 to WC10

- PAR
- PB
1000 Random Permutation Traffic
Effect of Packet size on RES: Worst Case Traffic

Latency - Simulation cycles

Throughput - Flit Injection Rate

- 1 Flit
- 2 Flits
- 4 Flits
- 8 Flits
Large local network: Uniform Random

Throughput - Flit Injection Rate

Packet Latency - Simulation cycles

PB
CRT
MIN
PAR
RES
Large local network: Worst Case

Packet Latency - Simulation cycles

Throughput - Flit Injection Rate

- PB
- CRT
- PAR
- RES
- VAL