An Effective Queuing Scheme to Provide Slim Fly topologies with HoL Blocking Reduction and Deadlock Freedom for Minimal-Path Routing

Pedro Yébenes¹, Jesús Escudero-Sahuquillo¹,
Pedro J. García¹, Francisco J. Quiles¹, Torsten Hoefler²

¹: University of Castilla – La Mancha, Spain ; ²: ETH Zurich, Switzerland
Outline

- Motivation
- Slim Fly topology
- Proposal Description
- Evaluation
- Conclusion
Motivation
HPC Systems

- Interconnection networks are **key elements** in HPC systems and datacenters.
  - Thousands of processing and/or storage nodes (Exascale challenge).
  - Applications demand increasing computing power.

- The interconnection network may become the **system bottleneck** if not properly designed and configured.

*Achieving high network performance is mandatory.*

---

Sunway TaihuLight
41,000 nodes - Cores 10,649,600
1st **Top500** (November 2016)
Motivation

Interconnection networks

• Network designers try to optimize the network resources.

• The lower average distance, the lower the resources needed.
  • High-radix switches available in the market.

• New topologies minimize the network diameter: Dragonfly, Flattened Butterfly, KNS, etc.
  • **Slim Fly**: a high-performance cost-effective network topology.
Motivation

Congestion appearance

- The working zone may be near the saturation point.
  - Power management techniques may reduce network bandwidth.
- Applications traffic may lead to hotspots.

![Graph showing latency and injected traffic related to working and saturation zones.](image)
Motivation

Head-of-Line (HoL) Blocking

- The real problem derived from congestion.
- Network performance may degrade significantly.
Motivation
Queuing Schemes

• Several queues, supporting Virtual Channels (VCs), or Virtual Lanes (VLs) are used at each port to separate traffic flows, reducing the HoL-blocking produced among them.

• A static criterion is used to map packets to queues.

The most efficient queuing schemes are tailored to a specific network topology and a specific routing algorithm.
Motivation

Queuing Schemes

- Some schemes are topology agnostic:
  - **VOQnet**: one queue per each destination in the network
  - **VOQsw**: one queue per output port in the switch
  - **DBBM**: maps packets to queues using the formula:
    - $Queue = Packet\_destination \% \#Queues\_per\_Port$
- However, the most efficient ones are tailored to a specific
  network topology and a specific routing algorithm:
  - **Flow2SL, vftree** for fat-trees.
  - **BBQ** for KNS topology.
  - **H2LQ** for Dragonfly.
Motivation
Design a queuing scheme

• Tailored to Slim Fly topology using minimal path routing.
  • Deadlock freedom.
• Effectively reduce HoL blocking by using the lower amount of queues.
Outline

• Motivation
• **Slim Fly topology**
• Proposal Description
• Evaluation
• Conclusion
Slim Fly topology

Benefits

- Network diameter is close to the theoretically optimal.
  - Connection pattern is based in the MMS graphs to ensure diameter 2.
- High bandwidth and resiliency.
- Low latency.
- Reduced cost and power consumption in comparison with other topologies.

Slim Fly topology
Connection

• Not intuitive connection pattern:
  • Find a prime number $q$
  • Constructing the Galois field $F_q$
  • Constructing the generator sets $X$ and $X'$

Slim Fly topology

Connection

- Switches are labeled: $\{0,1\} \times F_q \times F_q$

1. Switch $(0,x,y) \rightarrow (0,x,y')$ iff $y - y' \in X$

2. Switch $(1,m,c) \rightarrow (1,m,c')$ iff $c - c' \in X'$

3. Switch $(0,x,y) \rightarrow (1,m,c)$ iff $y = mx + c$

There are cycles in the channel dependency graph.

Slim Fly topology
HoL-blocking problem

An Efficient Queuing Scheme to Provide Slim Fly with HoL Blocking
Reduction and Deadlock Freedom for Minimal-Path Routing
Pedro Yébenes
Feb 5th, 2017
Austin, USA
Slim Fly topology
HoL-blocking problem
Slim Fly topology
HoL-blocking problem
Outline

- Motivation
- Slim Fly topology
- **Proposal Description**
- Evaluation
- Conclusion
Proposal Description

Benefits

• **Slim Fly Two-Level Queuing** (SF$_2$LQ).

• Two **virtual networks** (VNs) consisting of disjoint sets of queues to prevent deadlocks:
  - **Standard** Virtual Network (SVN).
  - **Escape** Virtual Network (E VN).

• **HoL-Blocking is reduced** in both VNs by applying different and independent mapping policies.
Proposal Description
SF2LQ mapping policy

• At Standard Virtual Network (SVN):
  • \( SVC = (\text{Destination}/p) \% \#\text{Standard\_VCs} \)
  • Maximum VCs: \( k' \) (number of ports connected to other switches)

• At Escape Virtual Network (EVN):
  • \( EVC = \text{Destination} \% \#\text{Escape\_VCs} \)
  • Maximum VCs: \( p \) (number of ports connected to nodes)
Proposal Description

**SF2LQ reducing HoL blocking**

- 3 VCs in the SVN and 2 VCs in the EVN
Proposal Description
SF2LQ reducing HoL blocking

- 3 VCs in the SVN and 2 VCs in the EVN
Proposal Description

SF2LQ reducing HoL blocking

• 3 VCs in the SVN and 2 VCs in the EVN

An Efficient Queuing Scheme to Provide Slim Fly with HoL Blocking Reduction and Deadlock Freedom for Minimal-Path Routing

Pedro Yébenes
Feb 5th, 2017
Austin, USA
Outline

• Motivation
• Slim Fly topology
• Proposal Description
• Evaluation
• Conclusion
Evaluation
Simulation Tool

OMNeT++-based simulator:

• Different topologies.
• Different routing algorithms.
• Different queuing schemes.
• Quality of Service support.

Pedro Yébenes, Jesús Escudero-Sahuquillo, Pedro J. García, Francisco J. Quiles: Towards Modeling Interconnection Networks of Exascale Systems with OMNeT++. PDP 2013
Evaluation

Network Configurations

- Slim Fly configurations:

<table>
<thead>
<tr>
<th>Name</th>
<th>q</th>
<th>k'</th>
<th>p</th>
<th>Ports per SW</th>
<th>Switches</th>
<th>Endnodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SlimFly-19_10</td>
<td>13</td>
<td>19</td>
<td>10</td>
<td>29</td>
<td>338</td>
<td>3380</td>
</tr>
<tr>
<td>SlimFly-29_15</td>
<td>19</td>
<td>29</td>
<td>15</td>
<td>44</td>
<td>722</td>
<td>10830</td>
</tr>
</tbody>
</table>
Evaluation
Switch Architectures Evaluated

- Input Queued Switch Architecture.

- Input Queued Switch Architecture implementing **Virtual Output Queues (VOQs)**:
  - Buffers are divided at the same time into VCs and VOQs.
  - Flow control is performed at VC level.
Evaluation

Queuing Schemes Evaluated

- **DLA-1+1**: 1 VC in the SVN + 1 VC in the EVN = 2 VCs
  - Basic scheme to avoid deadlocks. No HoL Blocking prevention.
- **DBBM-6+2**: 6 VCs in the SVN + 2 VC in the EVN = 8 VCs
- **DBBM-12+4**: 12 VCs in the SVN + 4 VC in the EVN = 16 VCs
- **SF2LQ-6+2**: 6 VCs in the SVN + 2 VC in the EVN = 8 VCs
- **SF2LQ-12+4**: 12 VCs in the SVN + 4 VC in the EVN = 16 VCs
Evaluation
Traffic Patterns

• **Uniform traffic:**
  - 100% traffic addressed to random destinations
  - Low-order HoL blocking.

• **Hot-Spot traffic:**
  - 75% of endnodes generating traffic to random destinations.
  - 25% of endnodes generating traffic to one destination.
  - High-order HoL blocking.
Evaluation

Uniform Traffic Results

- Metric: Packet Latency vs. Normalized Efficiency
- 100% random traffic.
Evaluation

Uniform Traffic Results

- Metric: Packet Latency vs. Normalized Efficiency
- 100% random traffic.
- Virtual Output Queues.

![Graph showing average packet latency vs normalized accepted traffic for different queueing schemes.](image1)

**SlimFly-19_10**
3380 endnodes

![Graph showing average packet latency vs normalized accepted traffic for different queueing schemes.](image2)

**SlimFly-29_15**
10830 endnodes
Evaluation
Hotspot Traffic Results

- Metric: Normalized efficiency vs. Generated traffic.
- 75% random traffic. 25% addressed to a hotspot endnode.
Evaluation

Hotspot Traffic Results

• Metric: Normalized efficiency vs. Generated traffic.
• 75% random traffic. 25% addressed to a hotspot endnode.
• Virtual Output Queues.

<table>
<thead>
<tr>
<th>Generated Traffic</th>
<th>Normalized Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

SlimFly-19_10
3380 endnodes

SlimFly-29_15
10830 endnodes
Outline

• Motivation
• Slim Fly topology
• Proposal Description
• Evaluation
• Conclusion
Conclusion

• We have analyzed the **congestion dynamics** in Slim Fly networks using minimal-path routing.

• SF₂LQ is an **efficient** deadlock-freedom queuing scheme which reduces HoL blocking in Slim Fly topology.

• **Topology-aware queuing schemes**, like SF₂LQ, efficiently leverage the available queues to reduce HoL blocking.
Conclusion

Future work

• Testing SF2LQ with traffic based on real application communication patterns.

• Extending SF2LQ to fit adaptive routing.

• Implementing SF2LQ in a real system built from commercial networks elements.
An Effective Queuing Scheme to Provide Slim Fly topologies with HoL Blocking Reduction and Deadlock Freedom for Minimal-Path Routing

**Pedro Yébenes**¹, Jesús Escudero-Sahuquillo¹, Pedro J. García¹, Francisco J. Quiles¹, Torsten Hoefler²

¹: University of Castilla – La Mancha, Spain ; ²: ETH Zurich, Switzerland
Slim Fly topology

Description

• Symbols used to describe Slim Fly topology:
  • N: number of endnodes
  • p: number of endnodes attached to a switch
  • k': number of channels to other switches
  • k: switch radix (k'+p)