TOPOLOGIES AND ADAPTIVE ROUTING ON LARGE-SCALE INTERCONNECTS

Shafayat Rahman (rahman@cs.fsu.edu)

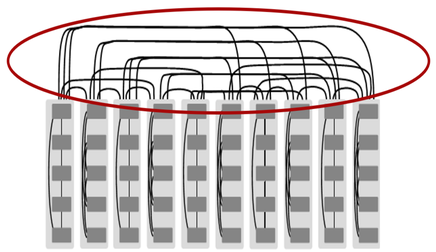
As a PhD student in the FSU CS EXPLORER (EXtreme-scale comPuting, modeLing, netwORking & systEms Research) lab under the supervision of Dr. Xin Yuan, my research activity revolves around the analysis, improvement and performance evaluation of a number of topology and adaptive routing schemes widely used in the field of HPC.

Load-Balanced Slim Fly Networks

Slim Fly:
- A recently proposed diameter-two topology
- Reaches within 88% of the optimum degree-diameter graph

Our contributions:
- Analyzed link-usage probability
- Showed that inter-plane links are more likely to be used for common traffic patterns
- Proposed two strategies to ensure load-balance in Slim Fly networks

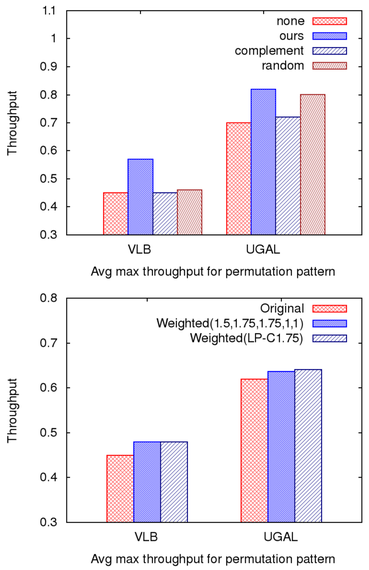
SF(q)	N_r	Probability ratio
SF(11)	242	1.24 : 1
SF(13)	338	1.27 : 1
SF(17)	578	1.32 : 1
SF(19)	722	1.34 : 1
SF(23)	1058	1.36 : 1
SF(29)	1682	1.42 : 1



Solution 1: b/w provisioning:
- Increase the b/w of the inter-plane links proportionally

\+ Completely eliminates load-imbalance

\- Implementation issues.



Solution 2: Weighted-VLB routing:
- Divert some traffic from the over-used links to the under-used ones
- So essentially, assign "weights" to paths

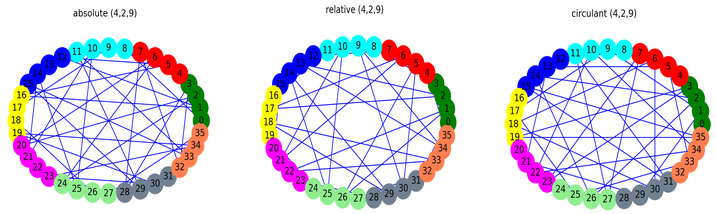
\+ More feasible implementation

\- Reduces load-imbalances, but does not remove it completely.

Dragonfly Design Space: Link Arrangement and Path Diversity

Dragonfly:
- Routers are grouped together in clusters
- Clusters are connected to form a diameter-three topology
- If each group has **a** routers, and each router has **h** global connections, then maximum number of groups in the system, $\mathbf{g_{max} = a*h + 1}$

Different Dragonfly arrangements:



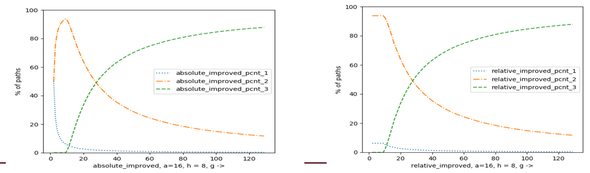
Open research questions:

What if $g < g_{max}$?
- What to do with the extra ports?
- Which group should be connected to which port?

Can randomly assigning ports help in topology creation?
- Fully random connections?
- Some greedy heuristic?
- Reinforcement learning?

What are some good performance metrics?

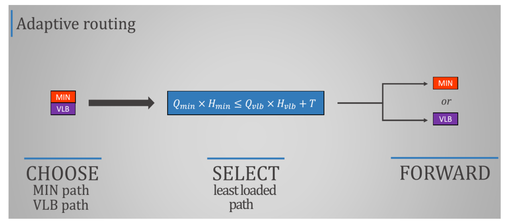
How do the minimum-path lengths change with increasing group number, g?
- Can they be exploited in routing?

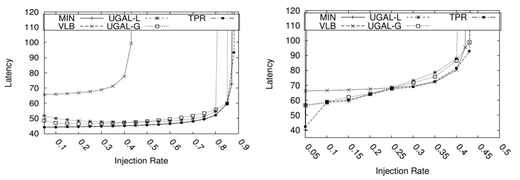


Traffic-Pattern Based Adaptive Routing for Dragonfly

Investigated the performance of adaptive routing in the Dragonfly used in Cray Cascade

Our contributions:
- Inferred traffic pattern by observing the packet destinations within a certain window
- Tuned the **T** value accordingly to bias towards either minimal or non-minimal routing





Performance Modeling Studies

Modeling UGAL on the Dragonfly Topology
- Modeled the UGAL routing over Dragonfly topology to get a better theoretical understanding on how the routing works

Throughput Models of Interconnection Networks: the Good, the Bad, and the Ugly
- Evaluated a number of commonly-used throughput models and identified similar and contradictory trends in their performance

A Comparative Study of Topology Design Approaches for HPC Interconnects
- Studied the performance characteristics of a number of topologies that provide either low diameter or high path diversity