DragonView: Toward Understanding Network Interference in Dragonfly-based Supercomputers

Yarden Livnat, Abhinav Bhatele, Nikhil Jain, Peer-Timo Bremer, Valerio Pascucci

Overview

The dragonfly topology[1] is becoming a popular choice for building high-radix, low-diameter networks with high-bandwidth links.

Preliminary experiments[2] on Edison at NERSC suggest that network congestion and job interference impact communication-heavy applications.

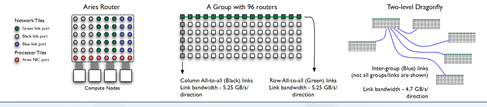
DragonView is a multi-window web-based visualization system for studying network congestion and job interference in dragonfly-based supercomputers. Facilitates investigation of the roles and impact of

- Job placement policies
- Routing algorithms
- Machine configuration.

Dragonfly Topology

The Cray Cascade[3] implementation uses 48-port Aries routers arranged in logical groups of 16×6 routers that are connected:

- All-to-all in each row (so called **green** links)
- All-to-all in each column (**black** links)
- **Blue** links connect routers from different groups



Challenges

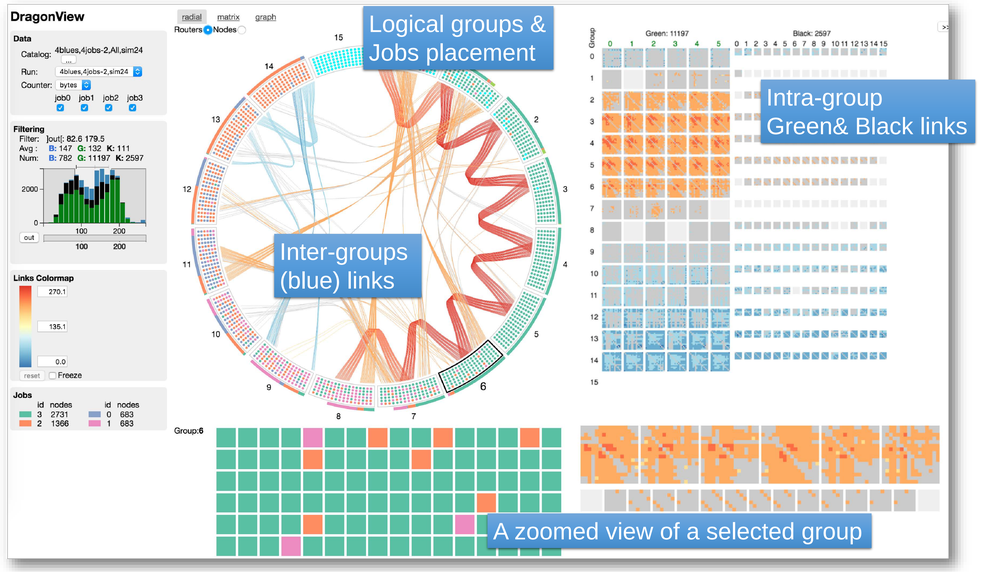
- **Routing:** The randomized global routing makes *quantitative* one-to-one link comparison between two runs meaningless
- **Global effects:** A local hot spot can affect unrelated jobs on the other side of the machine
- **Sparseness:** Hardware counters can be collected only from routers associated with the monitoring application

References

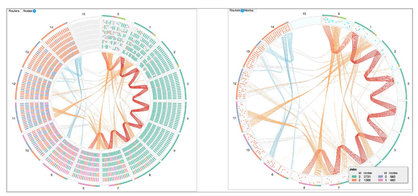
1. J. Kim et al. Technology-driven, highly- scalable dragonfly topology. SIGARCH Comput. Archit. News, 36:77– 88, June 2008
2. A. Bhatele et al. Analyzing network health and congestion in dragonfly-based systems. In Proceedings of the IEEE International Parallel & Distributed Processing Symposium, IPDPS '16. May 2016.
3. G. Faanes et al. Cray cascade: A scalable hpc system based on a dragonfly network. In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, SC '12, Los Alamitos, CA, USA, 2012

DragonView

Single-run View

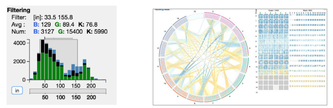


Nodes or routers views

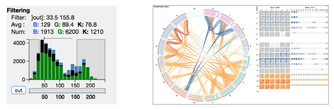


Filtering

In a range

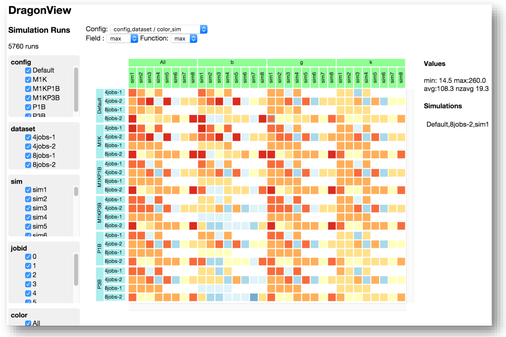


Outside a range: Show extremes



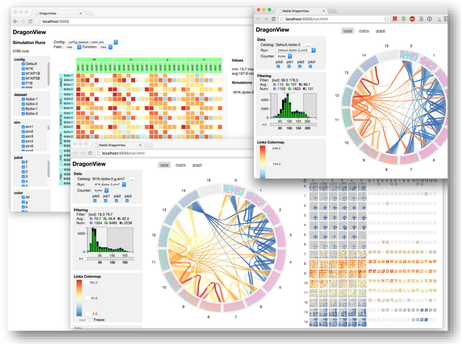
Ensemble view

Summary over multiple runs using a pivot table and filtering



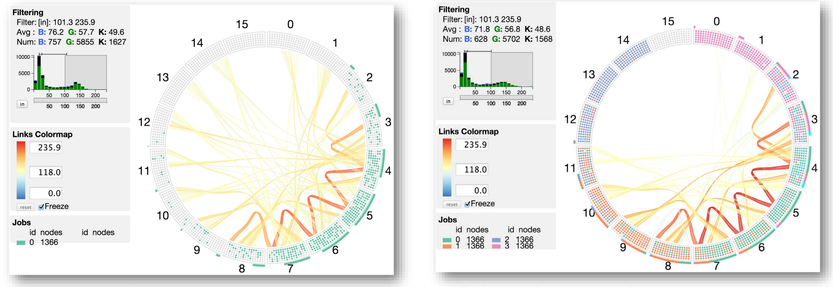
Single program with multiple windows

Open multiple single-run views from the ensemble view



Analysis of Simulation Runs

Inter-job Interference



A 4D Stencil job running on an empty machine (left) and in a workload (right). The number of blue links with high traffic decreases but the overall maximum traffic increases. The job's traffic is confined to fewer blue links in order to share bandwidth with other jobs.

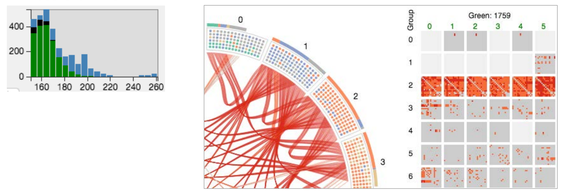
Network wiring

A) Configuration with different number of blue links per router

B) Examining by link color shows that blue links are affected the most

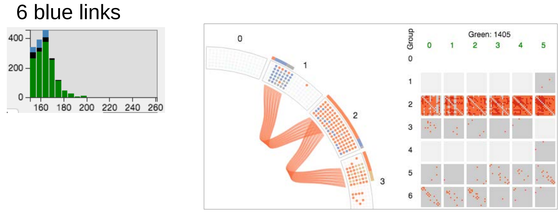
C) Impact (run 8jobs2-sim6)

- 2 blue links



Routing algorithm uses both direct and indirect routers in an effort to spread the load.

- 6 blue links



Blue links: Large reduction likely associated with use of shortest paths

Green (black) links: Small improvement

D) Traffic patterns (run 4jobs2-sim19)