Analyzing Interconnect Congestion on a Production Dragonfly-based System

Joy Kitson[1,2], Sudheer Chunduri[2], Abhinav Bhatele[1]

[1]University of Maryland, College Park [2]Argonne National Laboratory

Introduction

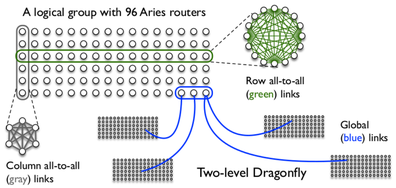
We set out to investigate how congestion manifests across Theta in both time and space. In particular, we sought to answer two questions:

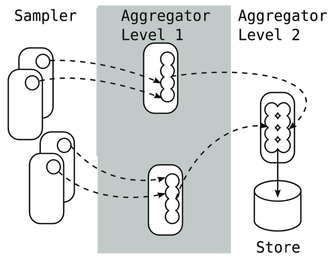
1 Is network congestion persistent?

2 Is congestion widespread or localized?

Theta:

- Production system at Argonne Leadership Computing Facility (ALCF)
- Cray XC machine with 4,392 compute nodes
- Uses Aries routers in a Dragonfly topology





LDMS:

- Lightweight Distributed Metric Service
- Collects data from all routers in the system

(Figure from [1])

Methodology: Data Collection and Analysis Pipeline

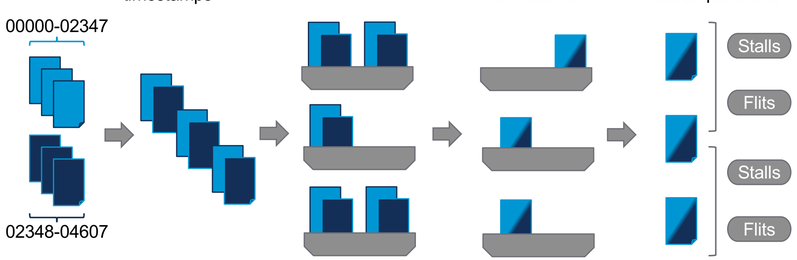
1 Find log files

2 Pair up logs with the closest timestamps

3 Separate pairs into hour-long bins

4 Combine the first valid pair from each bin

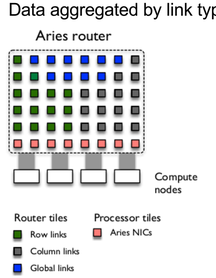
5 Compute metrics between subsequent bins



Flit and stall values calculated from four LDMS counters

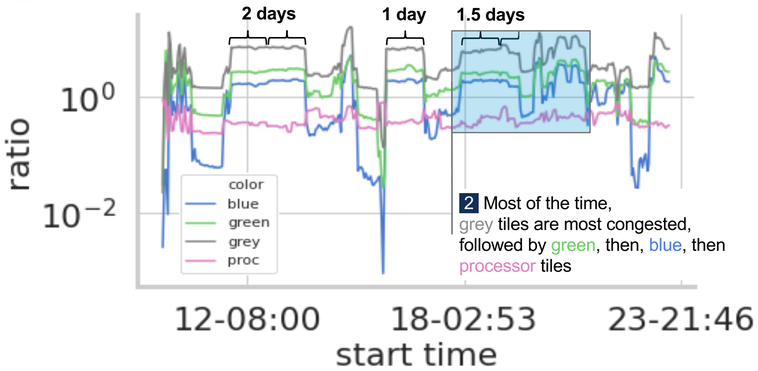
Counter	Metric	Tile Type
AR_RTR_<r>_<c>_INQ_PRF_INCOMING_FLIT_VC{0-8}	Flits	Network
AR_RTR_<r>_<c>_PT_INQ_PRF_INCOMING_FLIT_VC{0,4}	Flits	Processor
AR_RTR_<r>_<c>_INQ_PRF_ROWBUS_STALL_CNT	Stalls	Network
AR_RTR_<r>_<c>_PT_PRF_ROWBUS_STALL_CNT	Stalls	Processor

Data aggregated by link type (color)



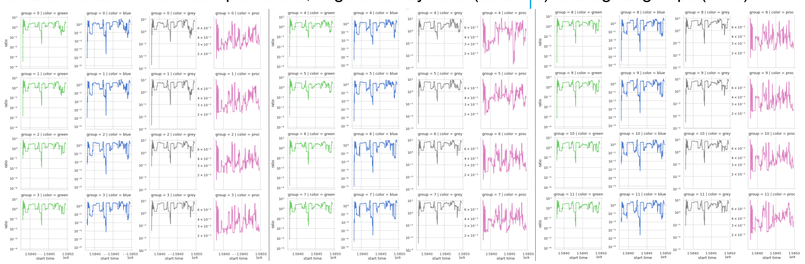
Analyzing Two Weeks of Network Data

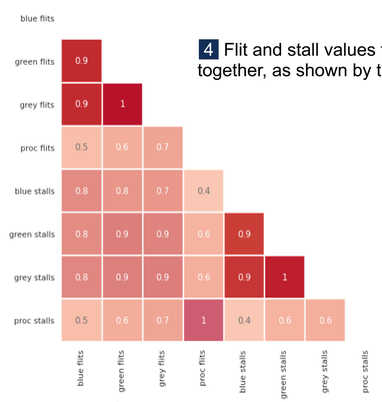
1 There were three periods of consistent congestion which lasted at least a day



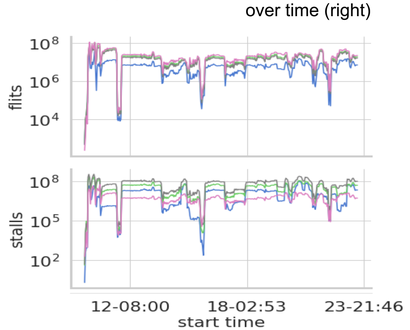
2 Most of the time, grey tiles are most congested, followed by green, then, blue, then processor tiles

3 The level of congestion is generally consistent across the logical groups of the system. Below are smaller versions of the ratio plot above, segmented by color (columns) and logical groups (rows).



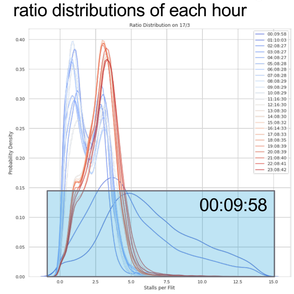


4 Flit and stall values for different colors of network tiles track up and down together, as shown by their correlations (left) and a plot of their mean values over time (right)

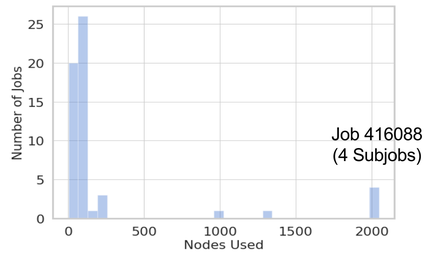


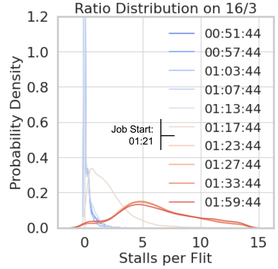
Case Study: Investigating a Period of High Congestion

1 High congestion in the first hour of the 17th, as shown by comparing the ratio distributions of each hour

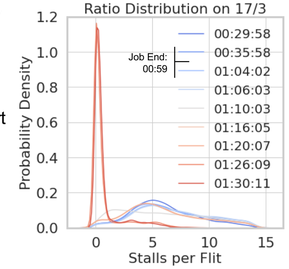


2 Job logs show that a large job was running at that time

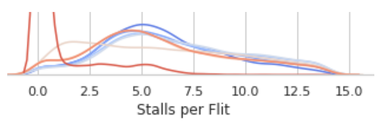




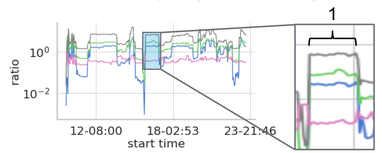
3 The periods immediately before and after the job ran show drastic shifts in the congestion of the system, as shown in the ratio distributions around the start (left) and end (right) of job 416088



4 It took about 25 minutes after the job ended for the system to return to a low congestion state



5 The runtime of Job 416088 lines up with a day-long period of congestion



Conclusions

1 Is network congestion persistent?

- The length of any single period of congestion is highly variable
- Some periods can last for days
- In at least one case, the congestion persists for some time after the apparent cause - a large job - is gone

2 Is congestion widespread or localized?

- High congestion periods seem to extend across all logical groups in the system
- Most congestion appears to occur on network tiles and to be focused on the column links within logical groups

References

[1] A. Agelastos, et al., "The lightweight distributed metric service: a scalable infrastructure for continuous monitoring of large scale computing systems and applications," in SC'14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE, 2014, pp. 154–165.

U.S. DEPARTMENT OF ENERGY · Argonne NATIONAL LABORATORY · UNIVERSITY OF MARYLAND