Evaluating Adaptive Routing Performance on Large Scale Megafly Topology

[1]Md Nahid Newaz, [2]Md Atiqul Mollah, [3]Peyman Faizian & [4]Zhou Tong
[1,2]Oakland University, [3]Florida State University, [4]Wheaton College

ABSTRACT

- Megafly is a recently proposed efficient and hierarchical interconnect topology
- Adopted by Niagara, Canada's fastest supercomputer
- Currently Progressive Adaptive Routing (PAR) is the best known routing scheme for Megafly networks
- In this work we have developed two new practical adaptive routing schemes KAPR and KU-GCN.
- Using event-driven simulator (SST-Macro) we shows that our new routing schemes outperform PAR by 43.5% on a topology with large number of groups

MEGAFLY TOPOLOGY

Megafly Single Group Architecture

- Every group consists of two levels of routers
- One level known as spine routers, are connected to global channels
- Another level known as leaf routers, are connected to compute nodes
- Intra-group connection in megafly is a complete bipartite graph

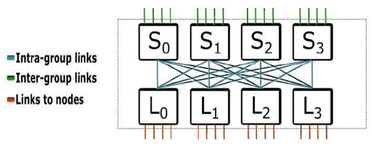

Figure : Group Architecture of Megafly

Megafly Network Architecture

- Multiple megafly groups can be connected using global links
- Global link arrangement can produce multiple variations on megafly network

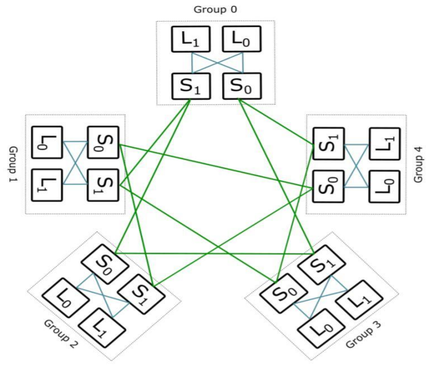

Figure : Megafly Topology with 2 routers in each level of a group and total 5 groups that is *mfly(2,5)*.

ADAPTIVE ROUTING IN MEGAFLY

Progressive Adaptive Routing (PAR)

i. Adaptive Routing decision is made on global spine routers
ii. If there is a congestion in minimal path global link it routes packet in a nonminimal path

K-path Adaptive Piggyback Routing (KAPR)

i. Routing decision is made on source leaf routers
ii. Global link congestion is piggybacked to all connected source routers and a spread of K minimal paths are sampled
iii. If there exists no globally uncongested route among the K paths, the packet is forwarded non-minimally.
iv. Otherwise minimally

K-path UGAL with Global Congestion Notification (KU-GCN)

i. First 3 steps are same as KAPR
ii. Otherwise, an Adaptive Load balancing (UGAL) comparison is made between the globally uncongested minimal path with lowest local output buffer occupancy with a globally uncongested minimal of lowest buffer occupancy towards a randomly selected destination obtained from another K paths sampling.

EVALUATION SETUP

Topology	Global Connectivity	# of Groups	Router Radix	# of Nodes	Minimal Path Diversity
mfly(12,9)	High	9	24	1296	18
mfly(10,11)	Medium	11	20	1100	10
mfly(8,17)	Low	17	16	1088	4

SYNTHETIC TRACE RESULTS

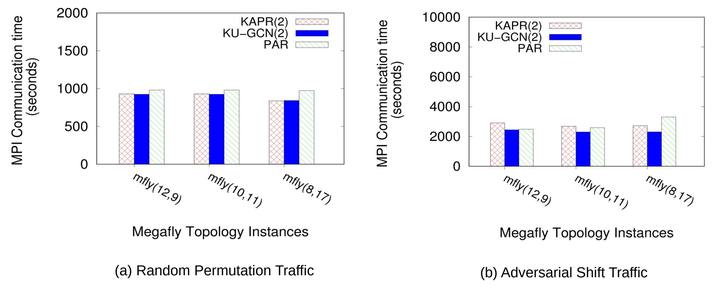

Figure : Communication Performance of 1024-rank synthetic workloads on Megafly topologies

APPLICATION BENCHMARK RESULTS

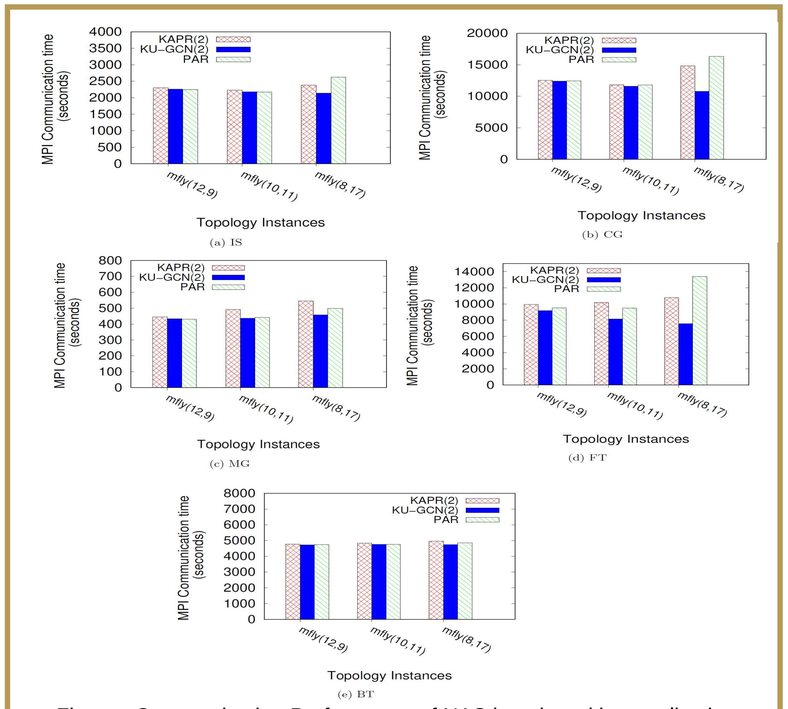

Figure : Communication Performance of NAS benchmarking applications on Megafly topology networks

EFFECT OF NETWORK LATENCY ON ROUTING

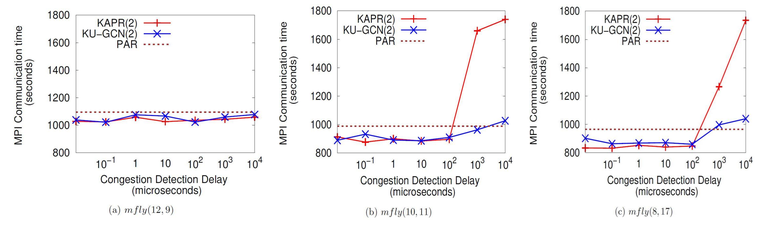

Figure : Routing performances trend of Megafly topologies on a random permutation workload over varied delay of congestion detection. The horizontal axis is base-10 exponential.

SUMMARY

- We introduced two new adaptive routing schemes
- The new routing schemes perform significantly better than the currently widely adopted progressive adaptive routing (PAR) , especially on low path diversity topologies
- Our KU-GCN routing scheme performs best among the three in the majority of workloads and topology sizes.