AN ANALYSIS OF PERFORMANCE VARIABILITY ON DRAGONFLY+ TOPOLOGY

Majid Salimi Beni, Biagio Cosenza
Department of Computer Science, University of Salerno, Italy
{msalimibeni , bcosenza}@unisa.it

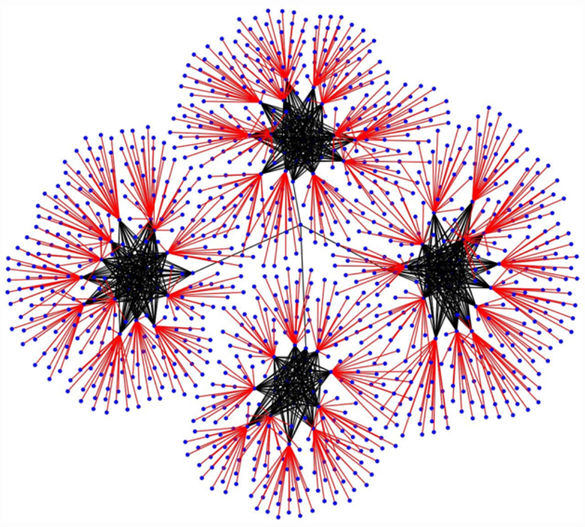
IEEE CLUSTER 2022
Heidelberg, Germany • 6–9 September

ABSTRACT

Large-scale compute clusters are highly affected by performance variability that originates from different sources. Among these sources, the network, as a shared resource between users and their jobs in a supercomputer, plays an important role. In this paper, we analyze the effect of some network-related sources on the performance variability of a modern compute cluster equipped with a Dragonfly+ interconnect. Specifically, we focus on the effects of job placement locality, communication patterns, routing strategy, and the network background traffic on performance variability of communication-intensive workloads.
To quantify the effect of network congestion (background traffic) on the performance variability, we propose a heuristic that can successfully estimate the amount of communication on the network produced by other jobs running on the cluster simultaneously. Then, we show how this network congestion contributes to the performance variability of different communication patterns and real-world communication-intensive applications.

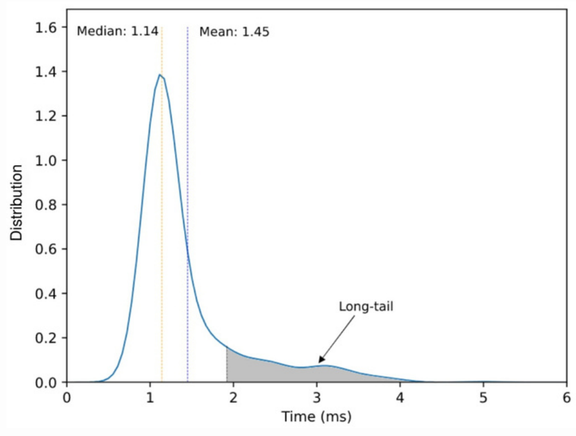
PROBLEM STATEMENT

Although Dragonfly+ topology provides a high network utilization, scalability, and router buffer utilization in comparison to other topologies like Dragonfly, it still suffers from performance variability. This performance variability originating from different sources degrades the application and system performance, and negatively affects the batch scheduler's decision makings.



- Dragonfly+ topology [1] of the Marconi100 supercomputer [2].
- It has 980 nodes, each is equipped with 2x16 cores IBM POWER9 and 4x NVIDIA Volta V100 GPUs
- This Dragonfly+ has 4 islands, each consisting of several groups of nodes.

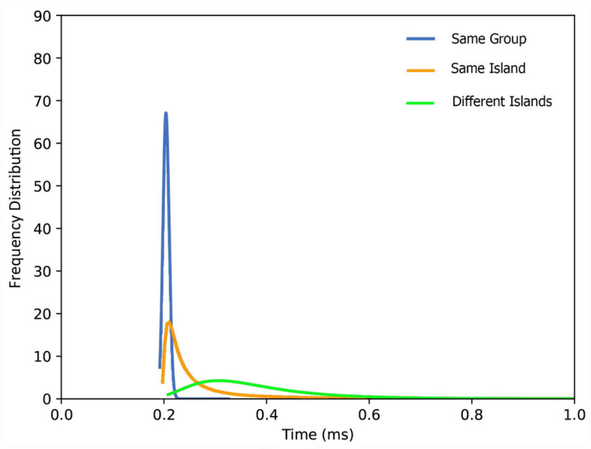
- The distribution of communication times of 1000 MPI_Reduce on 16 nodes of the cluster.
- More than 15% of the runs' latencies are larger than the 85th percentile (Belonging to long-tail).



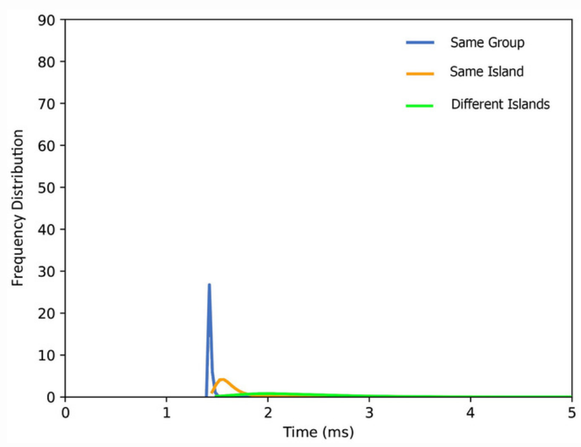
NODE ALLOCATION LOCALITY

According to the Dragonfly+ topology, we define 3 node allocation hierarchies:
1. All nodes on the same group (Only 1 hop between every 2 nodes)
2. all nodes on the same island
3. all nodes on different islands

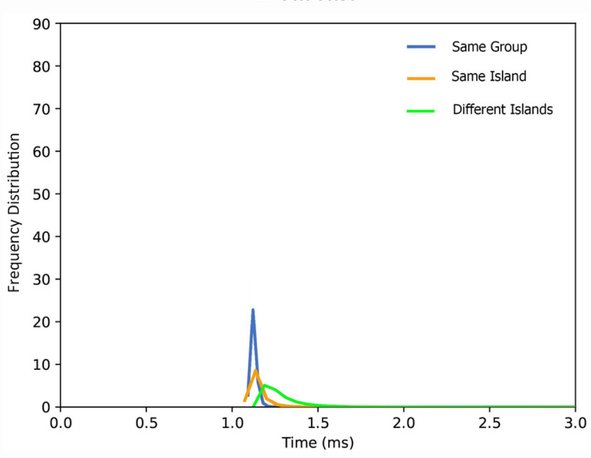
Then, we show the latency distribution of 3 collective communications:



Broadcast



Reduce



Alltoall

- When all the nodes are allocated on a single group, the communication is minimally affected by the global background traffic.
- Broadcast has the shortest tail and higher peak than Reduce and Alltoall for both the same group and the same island.
- Alltoall is the one with the longest tail when less locality is expressed, due to its communication intensity.

It is not always possible to allocate all nodes to the same group due to the limitation of the number of nodes in each group as well as the penalty of staying a long time in the job queue of the job scheduler.

BACKGROUND TRAFFIC ESTIMATION

While allocating nodes on different islands, the communication-intensive jobs' execution time is highly affected by the background traffic on the network. For this purpose, we propose the following heuristic, in which *b* is the estimated background traffic that is between 0 and 100.

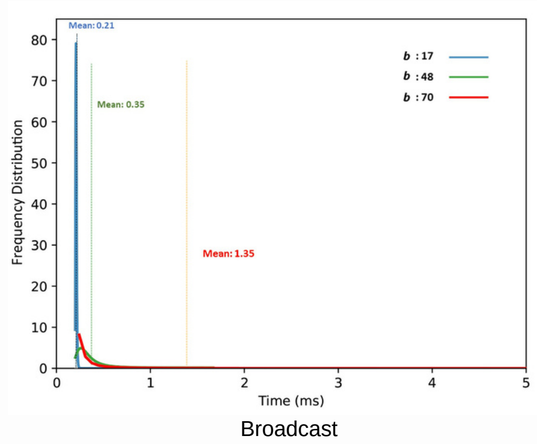
$$b = \frac{N_c}{N_t} * \frac{N'_c}{N_a} * 100$$

Nc: the number of unique nodes which are involved in communication,
Nt: the total number of cluster physical nodes (980 in Marconi100),
$N'c$: the number of nodes that contribute to communication (some nodes are shared among different jobs, we count them as many times they are allocated),
Na: all of the allocated running nodes (containing shared nodes).

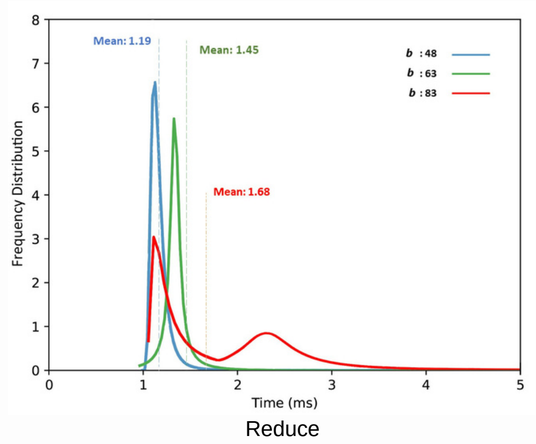
Using *Pearson Correlation Coefficient*, we showed that the correlation of the proposed metric with communication time becomes stronger when the data size is larger.

IMPACT OF BACKGROUND TRAFFIC

Here, we show how background traffic affects the distribution of 1000 iterations of 3 collective communications. For each collective, we show the distribution while there are 3 different background traffics.

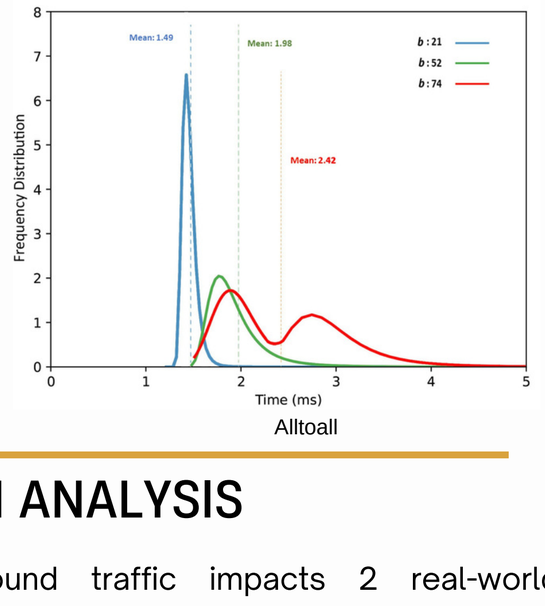


Broadcast



Reduce

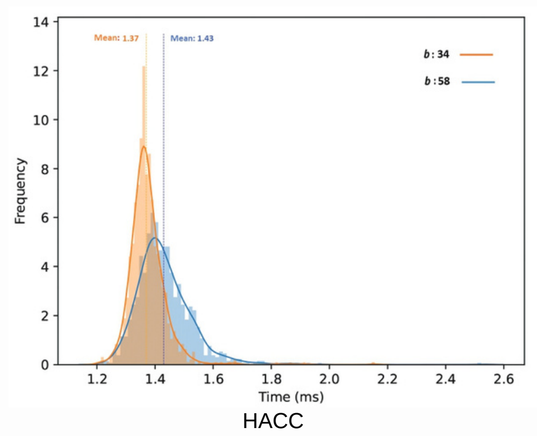
- Overall, the larger the b, the longest the tail of distributions.
- In Reduce and Alltoall, the distribution of the highest b is dual that is because of the **Adaptive routing algorithm**, choosing a non-minimal path.
- Alltoall possesses the longest tail that is because of its higher communication intensity.



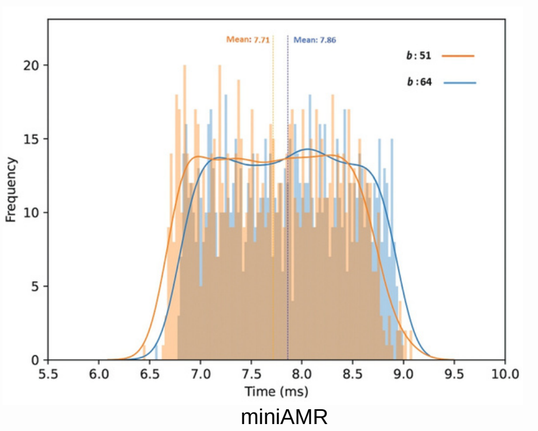
Alltoall

APPLICATION ANALYSIS

Finally, we show how the background traffic impacts 2 real-world communication-intensive applications, HACC, and miniAMR.



HACC



miniAMR

The overall average execution time increases with increasing the network background traffic. Applications' distribution is mostly affected by their dominant communication pattern. Although, the distribution can be affected by the routing algorithm, as well.

[1] Alexander Shpiner, Zachy Haramaty, Saar Eliad, Vladimir Zdornov, Barak Gafni, and Eitan Zahavi. 2017. Dragonfly+: Low cost topology for scaling datacenters. In 2017 IEEE 3rd International Workshop on High-Performance Interconnection Networks in the Exascale and Big- Data Era (HiPINEB). IEEE, 1–8.
[2] Marconi100, new accelerated system. https://www.hpc.cineca.it/hardware/marconi100. Accessed: 2022–06–22.