Contents

1 Introduction

What is the best network topology for simultaneous all-to-all communication?

Currently implemented topologies, i.e., Butterfly, are better for certain applications compared to other options, i.e., Mesh. However, no network design has been analytically shown to be optimal for any one task, much less a combination of objectives. In this study, a method for searching for an optimal topology, with respect to multiple goals, is carried out. An evolutionary search technique, simulated annealing, is used to determine network designs which are tailored to the needs of a specific application and set of user-defined goals. These objectives can include cost, time to solution, and reliability.

2 Method

In order to find network topologies specifically designed for narrow cases, two approaches are taken. In the first, compute nodes are distingished from router nodes on an undirected graph. Rather than considering network traffic and router algorithms, only graph-based metrics are used: the hop count distribution and minimum bisection count. Each of these two must be defined in the context of a graph with two distinct types of nodes – compute nodes from which information is sent and received, and router nodes which only redirect information.

In the second part of the study, a more realistic approach is taken. Rather than randomly generating the network of compute and router nodes, it is assumed that compute nodes and first router share a common design. This assumption is based on normal manufacturing procedures of producing homogeneous components.

Since the time-to-solution is the most direct measure of quality for a network topology, a more detailed simulation is needed. The need for timing requires that traffic congestion be included, which necessitates routing algorithms be specified. A cycle-accurate network simulator, SST/macro, is used to carry out the bucket sort algorithm.

3 Questions

Choice of fitness function: Does the minimum bisection count and average hop count yield the same solution as the average time-to-solution?

To measure this, we set up a highly asymmetric network composed of two connected routers. One router is connected to one computer, and the other router is connected to 7 other computers.

Average hop count ($\langle hc \rangle$) is simpler to measure (using Dijkstra's algorithm for single-source shortest path on each pair of compute nodes) compared to average time to solution ($\langle tts \rangle$). The Floyd-Warshall algorithm for all-pairs shortest path is not useful since we do not include routers as sources of traffic generation. The simplification of $\langle hc \rangle$ is due to not accounting for congestion. Congestion on a graph is a measure of how many shortest routes use the same edge.

Minimum bisection was studied for use as a fitness function, but the cost of correct measurement is high. Bisection bandwidth for communication graphs is defined (Ref [1], page 52) as half the compute nodes being separated.

$$U = 2^M \sum_{x=1}^{(N+2)/2} x \tag{1}$$

$$p = \frac{\log(1-c)}{\log(1-(1/U))} \tag{2}$$

Choice of protocol: Does the use of MPI versus OpenSHMEM make a difference in the outcome?

If the information being transmitted is small, then the overhead associated with two-way communication (MPI) could be detrimental.

Choice of initial network topology: If the initial network is random/lattice/scale-free/Butterfly/Clos, does it always converge towards the same pattern?

4 Results

5 Conclusion

The objective of this research is to help network designers make informed decisions on optimization for specific tasks.

Multi-objective optimization takes the form
fitness function = $a\cdot$(cost)+$b\cdot$(time to solution)+$c\cdot$(reliability).

In the work presented the network cost can be calculated from component prices, and time to solution was the explicit goal. We envision the capability to include reliability based on practical feedback from existing networks: how often does a router fail? How often does a cable fail? What are the failure modes?

6 Future work

Although routing algorithms were not part of this study, they are integral to the performance of the topology. We plan to investigate dynamic routing algorithms and the use of out-of-band information.

References

[1] C. Thompson *A complexity theory for VLSI.* PhD dissertation, (1980)