

# Scalable and Low Cost Interconnection Topologies

Professor Alex Mohr

Asha Andrade

Ungzu Mun

Dong Hwan

## Abstract

*Information drives businesses. Storage systems play a significant role in today's enterprise. We have seen a recent flurry of technologies to satisfy this growing demand. As we shall see in this short write up, each of these cater to a very specific subset of clients and the costs vary widely depending on the service class. We aim to closer the gap in these service guarantees while keeping the cost moderate.*

## 1. Introduction

Today's high-performance computing places a stringent demand on the storage capacity and performance of its components. Fulfilling these growing demands motivates the need for designing efficient interconnects [1]. While the best performance can be gained using expensive interconnects, affordability is the primary concern.

To cite a simple example, an expensive way to setup a 46-node network will be to use one 48-port switch. As an alternative, a cheaper and less efficient way will be to use 7 8-port switches. One of the solutions we propose is to use 2 24-port switches. The cost structure is given in the table below [2]:

Switch	Vendor specs	Switch Cost \$	Cost to connect 46 nodes \$	
8-port	Power Connect 2708; 8 port Gigabit Ethernet Switch; 8 10/100/1000 BASE-T ports	109	763	
24-port	Power Connect 2324; 24 Port Fast Ethernet Switch with Gigabit; 24 10/100 BASE-T ports	149	298	√
48-port	Power Connect 3448; High Performance Fast Ethernet Switch; 48 10/100BASE-T Fast Ethernet ports	749	749	

## 2. Simulation Setup

The performance of a network can be evaluated by knowing the aggregate bandwidth utilized and the average path length across it. To aid in our process of determining the optimal interconnects for a given node-switch pair, we simulate a reasonable number of random flows across the commonly used graph types, in addition to



<b>TORUS (4096)</b>				<b>64 X 64</b>				<b>8 X 8 X 8 X 8</b>				<b>32 X 32 X 4</b>				<b>16 X 16 X 16</b>			
Node Count	Avg Rate	Avg Fath		Avg Rate	Avg Fath		Avg Rate	Avg Fath		Avg Rate	Avg Fath		Avg Rate	Avg Fath					
50000	535452.64	10.71	32.06	3367454.60	67.35	8.00	1083795.88	21.68	17.01	1873214.97	37.46	12.01							
100000	560772.11	5.61	31.92	3764544.81	37.65	8.01	1168282.28	11.68	17.05	2018733.30	20.19	11.99							
125000	565429.04	4.52	31.98	3847759.52	30.78	8.01	1191052.45	9.53	17.01	2061576.48	16.49	11.99							
150000	565746.32	3.77	32.03	3950403.22	26.34	8.00	1224337.40	8.16	16.99	2091086.91	13.94	12.00							
175000																			
200000																			
	6.91%			24.11%			9.96%			17.02%									
<b>TORUS (8192)</b>																			
<b>64 X 64 X 2</b>				<b>8 X 8 X 8 X 8 X 2</b>				<b>32 X 32 X 8</b>				<b>16 X 16 X 16 X 2</b>							
Node Count	Avg Rate	Avg Fath		Avg Rate	Avg Fath		Avg Rate	Avg Fath		Avg Rate	Avg Fath		Avg Rate	Avg Fath					
50000	1029628.10	20.59	32.45	5737470.71	114.75	8.49	1971226.85	39.42	18.05	3316381.80	66.33	12.51							
100000	1086783.67	10.87	32.55	6699788.46	67.00	8.49	2165028.80	21.65	17.98	3729133.33	37.29	12.50							
125000	1096932.66	8.78	32.53	7002171.91	56.02	8.51	2222623.09	17.78	18.01	3839945.04	30.72	12.50							
150000	1124143.00	7.49	32.50	7202907.17	48.02	8.50	2258315.49	15.06	17.99	3912673.34	26.08	12.52							
175000																			
200000																			
	5.49%			19.54%			9.19%			13.65%									
<b>TORUS (16384)</b>																			
<b>64 X 64 X 4</b>				<b>8 X 8 X 8 X 8 X 4</b>				<b>32 X 32 X 16</b>				<b>16 X 16 X 16 X 4</b>							
Node Count	Avg Rate	Avg Fath		Avg Rate	Avg Fath		Avg Rate	Avg Fath		Avg Rate	Avg Fath		Avg Rate	Avg Fath					
50000	1875398.54	37.51	33.00	9063575.83	181.27	9.00	3457966.51	69.16	19.97	5591109.55	111.82	13.01							
100000	2045517.94	20.46	33.05	11343154.09	113.43	8.99	3907989.65	39.08	19.99	6661565.86	66.62	12.99							
125000	2108145.66	16.87	32.99	12021495.08	96.17	9.00	4029444.09	32.24	19.99	6924984.93	55.40	13.02							
150000	2125504.01	14.17	33.02	12573609.40	83.82	9.00	4104586.28	27.36	20.02	7160145.72	47.73	13.00							
175000																			
200000																			
	4.32%			15.35%			8.35%			10.93%									

Shown above are some of the results obtained by running the simulation on a large set of nodes. It essentially serves to indicate that higher the dimension, the better is the throughput.

#### 4. References

- [1] Interconnection Architecture for Petabyte-Scale High-Performance Storage Systems – Andy D. Hospodor and Ethan L. Miller
- [2] <http://www.dell.com>

#### 5. Seawulf Cluster

The next and most important step is to gather sufficient data on the customer requirements and the current network setups and to be able to come up with topologies which can be utilized in the real world.

We have run a few simulations on the Seawulf cluster as part of data gathering and would like to be able to run many more. We have currently exceeded the quota assigned to us and this is a request to provide us 10000 hours/month \* 3 months.