An Analysis of Network Congestion in the Titan Supercomputer's Interconnect

Jonathan Freed
University of South Carolina
902 Sumter Street
Columbia, SC 29208
(+1) 803.777.1609
freedjm@email.sc.edu

Saurabh Gupta (advisor)
Oak Ridge National Laboratory
1 Bethel Valley Road
Oak Ridge, TN 37831
(+1) 865.576.7658
guptas1@ornl.gov

Devesh Tiwari (advisor)
Oak Ridge National Laboratory
1 Bethel Valley Road
Oak Ridge, TN 37831
(+1) 865.576.7658
tiwari@ornl.gov

ABSTRACT

The Titan supercomputer is used by computational scientists to run large-scale simulations. These simulations often run concurrently, thus sharing system resources. A saturated system can result in network congestion—negatively affecting interconnect throughput. Our project analyzed data collected by measuring the throughput between different node pairs in Titan. In particular, we searched for correlations when the throughput was low. We investigated the direct path between the two test nodes, as well as the neighborhood of nodes that are one connection away from the direct path. For each set of nodes, we analyzed the effects of the distance between the nodes, the number of "busy" nodes (nodes currently allocated to applications), and the applications that were running. By understanding application interference, we can develop job-scheduling strategies that lower such interference and lead to more efficient use of Titan's resources and faster computations for researchers.

1. INTRODUCTION

Nodes in the Titan supercomputer are connected via the Cray Gemini interconnect. The Gemini routers are connected to form a 3D Torus topology [1] and each Gemini router hosts two compute nodes. A 3D Torus topology reduces the number of connections and minimizes the maximum length of the connections. However, a 3D Torus topology is more prone to network congestion. This research focuses on investigating how the available bandwidth is impacted by the activity in the node neighborhood. We first analyze it without distinguishing between applications and then investigate further if certain applications are more likely to cause congestion than others.

2. METHODOLOGY

2.1 Data Collection

We conducted a series of experiments on Titan under real operating conditions. Each experiment tested the throughput between several pairs of nodes on Titan. For each node pairing we choose for this study, between 20-25 samples of throughput were collected. This provides us over 1400 such samples to study the correlation of throughput with the amount of busy nodes in the neighborhood and specific applications.

We also simultaneously collected the allocation status (the status of all of the compute nodes in the system). The allocation status allows us to examine what applications were running at each node in Titan. This data can be used to determine the amount of activity and specific details about the applications running near the location of each throughput measurement.

2.2 Data Analysis

We conducted three different analyses on the data that we collected. First, we analyzed the effect that the path distance had on the observed throughput. Second, we searched for correlations between the number of "busy" nodes and the throughput. We quantified these correlations using Pearson's and Spearman's coefficients. Third, we investigated the possibility of certain applications being the cause of network congestion.

3. RESULTS

3.1 Impact of Path Distance

From Figure 1, we can clearly see the anisotropic nature of throughput where the X- and Z-dimensions have significantly higher throughput compared to the Y-dimension. This is due to lower number of physical links and hence lower bandwidth of the Y-dimension in the 3D Torus on Titan. From our analysis, we found that greater distance between nodes results in lower throughput. This is expected due to the topology and sharing of bandwidth in the network often referred to as geometric degradation in bandwidth. This is why close packing of compute nodes belonging to an application is beneficial for performance. One interesting characteristic to note is that the throughput does not reduce with distance in a canonical geometric ratio, instead it shows stepwise behavior at several places as the distance is

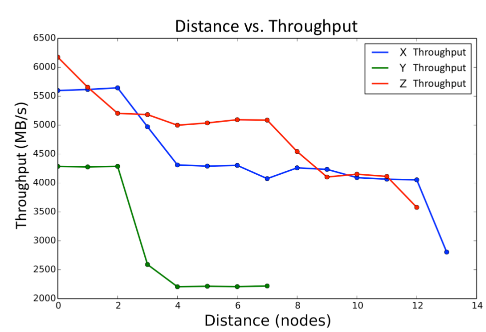

Figure 1. Distance versus throughput for each dimension.

increased. In the case of the Y- and Z-dimensions, the location of the steps can be seen directly related to where the next link is a lower bandwidth cable instead of the faster *backplane* or *mezzanine* links [2].

3.2 Correlation with Number of Busy Nodes

In Figure 2, we show the results of our correlation analysis between the number of busy nodes and the observed throughput. This analysis is oblivious of what applications are scheduled in the neighborhood and only concerns with how many nodes are running an application (busy) when the throughput is being measured. We found that a moderate correlation exists and the correlation is higher for the X- and Y-dimensions. The Z-dimension has higher bandwidth and hence is less affected by congestion.

Neighborhood	Dimension	Pearson Coefficient	Pearson p-value	Spearman Coefficient	Spearman p-value
Direct Path	X	-0.421876	0.000000	-0.519125	0.000000
	Y	-0.467197	0.000000	-0.713708	0.000000
	Z	-0.269941	0.000000	-0.099382	0.017526
1 hop	X	-0.421416	0.000000	-0.505862	0.000000
	Y	-0.404840	0.000000	-0.615761	0.000000
	Z	-0.384263	0.000000	-0.388042	0.000000

Figure 2. Pearson's and Spearman's coefficients and corresponding p-values.

3.3 Correlation with Specific Applications

We investigated if there are certain applications that may be causing significantly more network congestion than others. At the application level the network impact can be quantified. But, the challenge here is to identify the applications that should be selected for deeper analysis because Titan runs hundreds of scientific applications everyday. Therefore, our analysis aims for identifying some selected applications that often cause low throughput in our tests.

We calculated how often the applications were in the node neighborhood when we observed low throughput. This is signified by the *confidence* in Figure 3. The *presence* is the average percentage of the neighborhood that a given application occupies. Several applications can be identified here as potentially aggressive and they impact the network negatively very often. Therefore, deeper understanding of these applications' behavior is critical for avoiding interference with them. For example, application 1 occupies only about 50% of the neighborhood on average and still shows high confidence in the ability to affect the throughput negatively. On the other hand, application 19 shows over 80% presence in our experiments and only shows 50% confidence in the ability to affect the throughput negatively. In this manner, this analysis guides us in choosing the applications, which may be good candidates for deeper analysis of communication behavior to improve scheduling and minimize the impact of the application on its neighbors. We have replaced the actual application names to keep the identity of the users/applications anonymous in the figure.

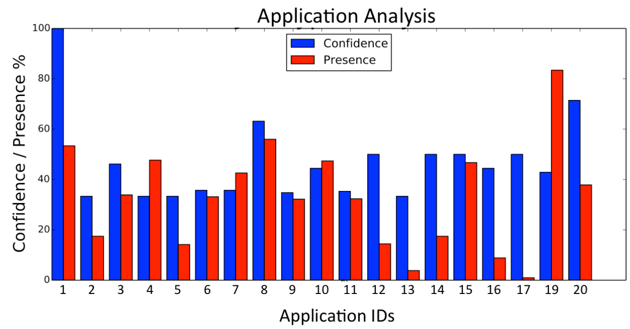

Figure 3. Applications suspected of affecting available network bandwidth by causing network congestion. (Names of applications are replaced with Application ID here)

4. CONCLUSION AND FUTURE WORK

Exploratory analysis on the network congestion data collected on Titan shows some interesting insights. We found that further distance and increased activity both correlate to lower throughput. We also found some suspected applications that may cause network congestion. In order to mitigate network congestion, further investigation into application behavior will be performed and we believe that scheduling policies should take this into account to make better application placement decisions.

5. REFERENCES

[1] Matt Ezell, "Understanding the Impact of Interconnect Failures on System Operation," *In Cray User Group (CUG) 2013.*

[2] Kevin Pedretti, Courtenay Vaughan, Richard Barrett, Karen Devine, K. Scott Hemmert, "Using the Cray Gemini Performance Counters", *In Cray User Group (CUG) 2013.*