12th ANNUAL WORKSHOP 2016

InfiniBand Topologies and Routing in the Real World

Susan Coulter, Jesse Martinez, Dominic Manno,
Mike Mason, Brad Settlemyer

Los Alamos National Laboratory – LA-UR-16-21830

[April 8th, 2016]

Conclusions

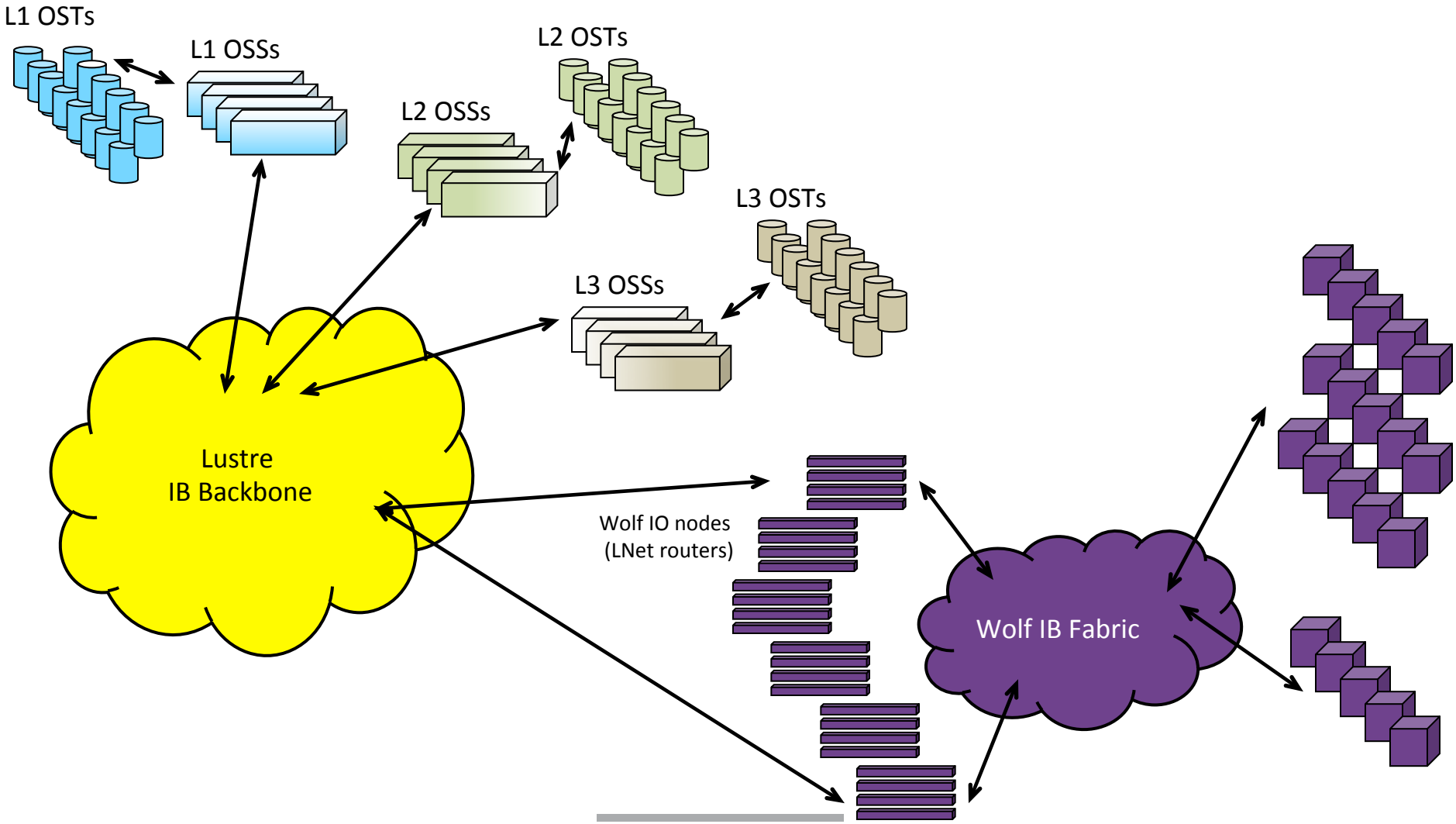
The high speed network (HSN) internal to a compute cluster is a significant contributor to external file system performance and throughput.

- Interconnect technology is extremely complicated, with very few simple assumptions holding true in the real world.
- Testing and experience at scale is critical to understanding and verifying HSN performance.
- Avoid putting more than one IO node per leaf switch if at all possible

The Problem

- LANL had 2 distinct Lustre file systems deployed in the Turquoise Network, L1 and L2.
- A third file system, L3, was being deployed.
- All 3 of these file systems used the same IB backbone.
- As part of the L3 deployment, bandwidth testing was being done from an existing compute cluster, Wolf, using a reservation of 64 contiguous compute nodes.
- The bandwidth test results were ~19 GB/s, while the expected bandwidth was ~55GB/s.
  - These results were consistent, regardless of the amount of the file system being targeted. (a few OSSs or the entire set of 40 OSSs)

Lustre Backbone / Wolf Compute Cluster connectivity



Wolf IB fabric details

Intel True Scale 12800-360

- 36 line cards with 18 external ports and 18 internal ports each
  - 18 host ports
  - 18 spine-connected ports
- 18 spine cards with 36 ports each
- 36x18 = 648 possible internal spine ports for routes
- IB Routing is not symmetric, (A->B != B->A)
- Number of routes = (hosts**2) - hosts)
  - Wolf has 408,960 routes
- OpenSM is the Subnet Manager
  - SM runs on wf-master
  - Fat Tree routing with Up/Down as secondary fall back
  - port_shifting set to TRUE
  - All 24 IO nodes plugged in to port 18 on 24 different line cards

The Problem in more detail

- As part of the trouble shooting, the existing L1 and L2 file systems were tested.
  - L1 and L2 test results showed possible degradation.
  - The degradation appeared to happen after the configurations for L3 were added to various systems in the Turquoise Network
  - Typical performance for L2 was considered to be ~35-40 GB/s
  - Regular and consistent bandwidth test results were not available for either L1 or L2.
  - Due to scheduling and end-of-year constraints, L3 was added to the production IB backbone before it was fully tested. This resulted in the lack of independent bandwidth data.

Investigation

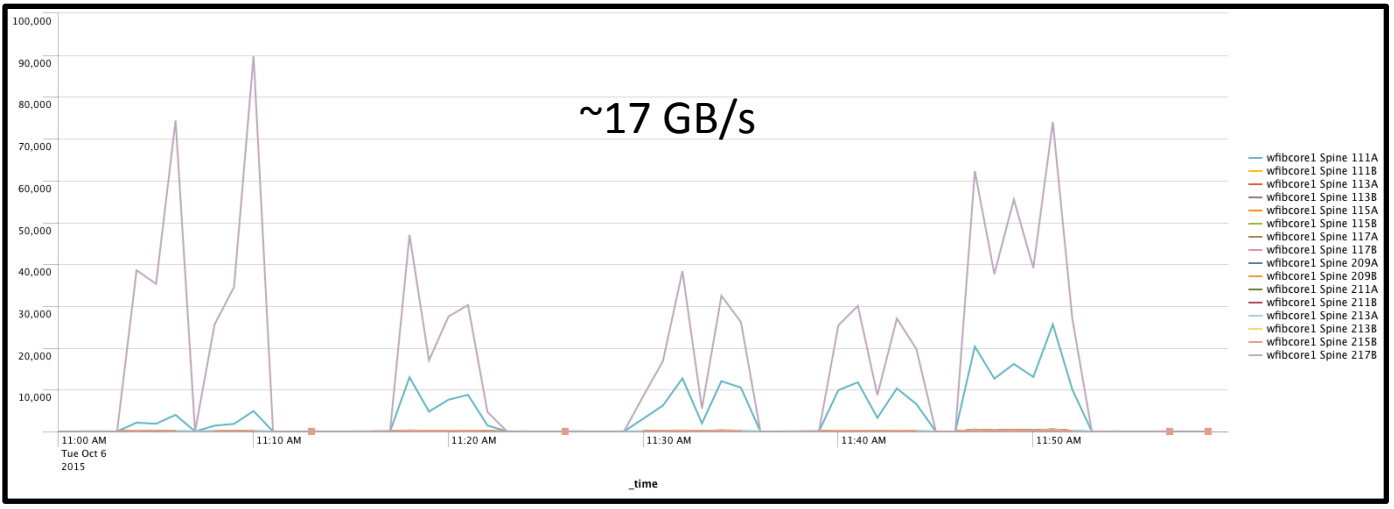
- The focus shifted to testing L2, partly because of the possible degradation in the performance of a production file system, and partly because some amount of previous data existed to compare results to.
- The first L2 tests were done using the same contiguous allocation used to test L3.
- Subsequent tests were run at larger sizes to look for differences and/or patterns. These subsequent tests became spread across the cluster as larger allocations were dynamically created, which resulted in these larger tests running on non-contiguous allocations.
  - The contiguous allocations consistently performed at ~17 GB/s
  - The non-contiguous allocations consistently performed at ~35-40 GB/s
- These results held true even when the number of nodes in both the contiguous and non-contiguous tests were equal.

Investigation continued …

- The results for contiguous and non-contiguous compute nodes were consistently reproducible.
- SM verified to be running Fat Tree during the tests (port_shifting and scatter_ports only applies to Up/Down routing engine)
- Refined the granularity of one of our existing monitoring processes to capture throughput on the switch ports.
- Monitored throughput via Splunk during testing.

IB throughput data from PerfManager / Splunk

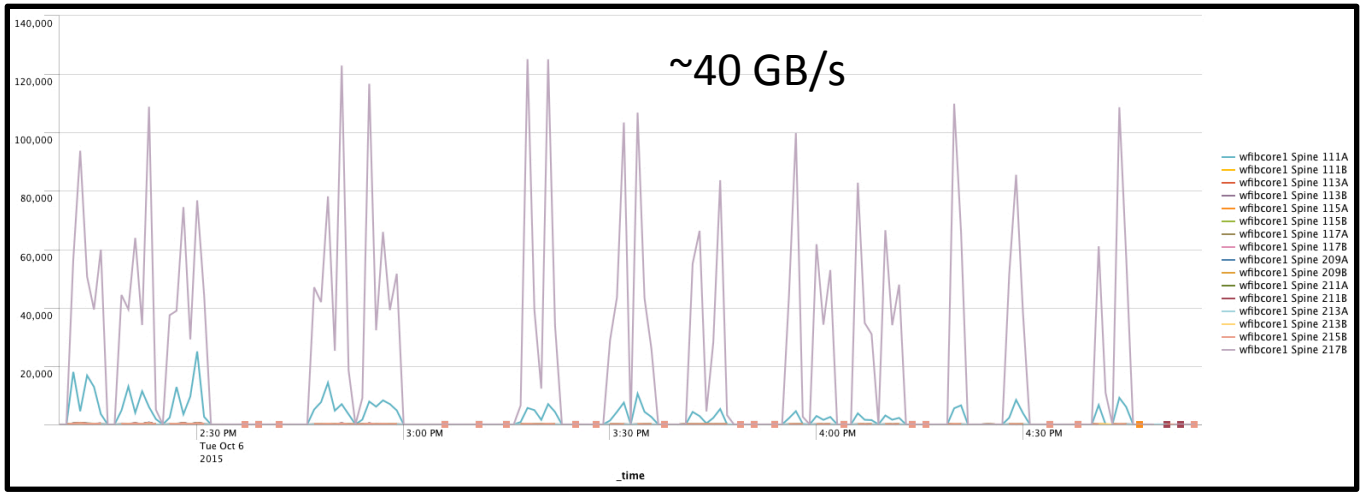
Wolf IB Spine Switch Activity



Contiguous node allocation

63 nodes

96 nodes



Non-contiguous node allocation

IB throughput data further refined

Wolf IB Spine Port Route Distribution

[port number]

14	[32] "wfibcore1 Spine 111A"
15	[33] "wfibcore1 Spine 111A"
17	[21] "wfibcore1 Spine 111A"
17	[31] "wfibcore1 Spine 111A"
322	[19] "wfibcore1 Spine 217B"
345	[21] "wfibcore1 Spine 217B"
391	[27] "wfibcore1 Spine 217B"
391	[33] "wfibcore1 Spine 217B"

Contiguous
63 node allocation
1,121 routes

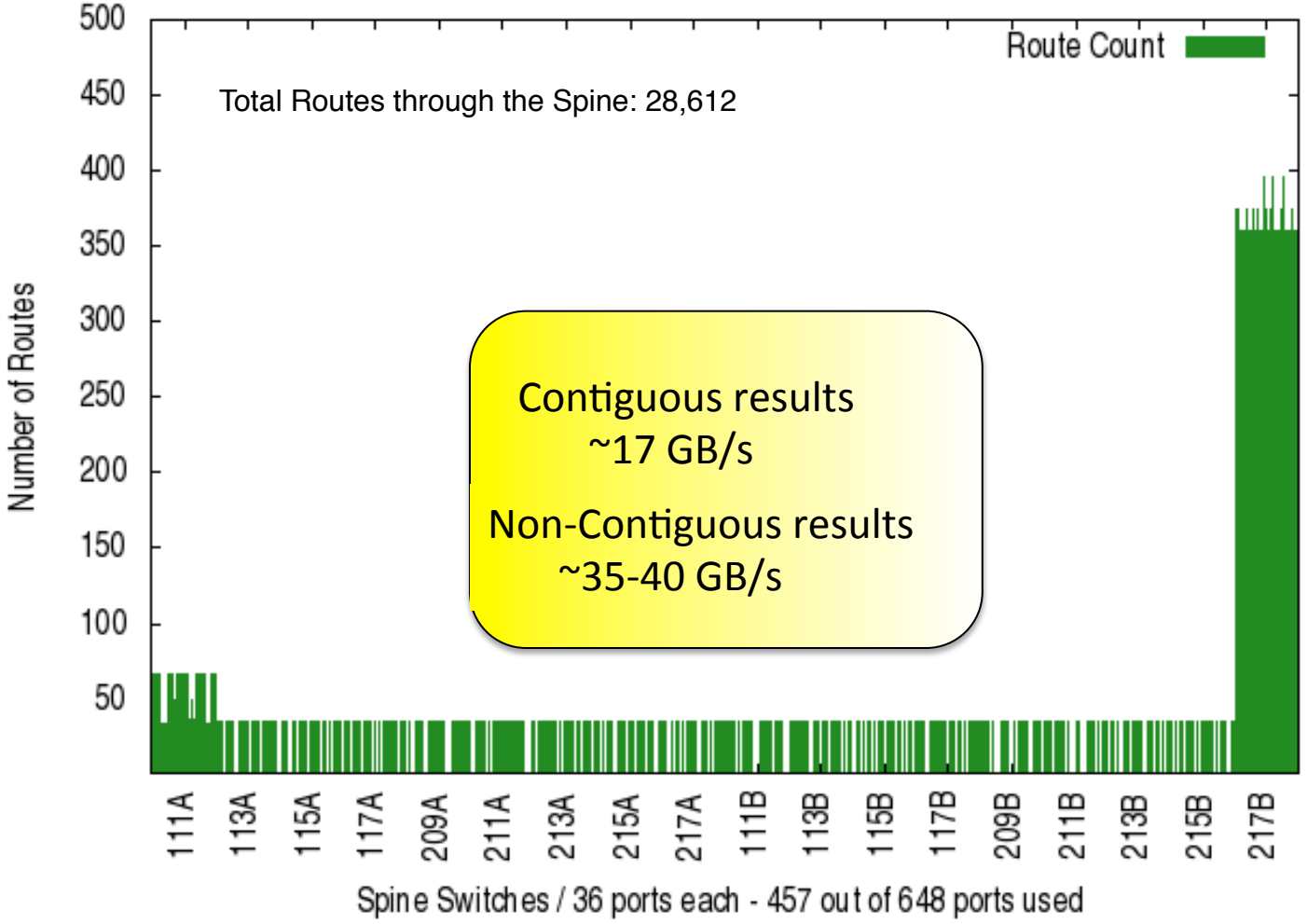
Number of routes through each port

2	[19] "wfibcore1 Spine 111A"	21	[13] "wfibcore1 Spine 217B"
2	[26] "wfibcore1 Spine 111A"	21	[15] "wfibcore1 Spine 217B"
2	[2] "wfibcore1 Spine 111A"	21	[31] "wfibcore1 Spine 217B"
2	[3] "wfibcore1 Spine 111A"	21	[32] "wfibcore1 Spine 217B"
2	[7] "wfibcore1 Spine 111A"	22	[3] "wfibcore1 Spine 217B"
2	[8] "wfibcore1 Spine 111A"	22	[8] "wfibcore1 Spine 217B"
4	[14] "wfibcore1 Spine 111A"	42	[9] "wfibcore1 Spine 217B"
4	[22] "wfibcore1 Spine 111A"	44	[28] "wfibcore1 Spine 217B"
6	[10] "wfibcore1 Spine 111A"	63	[1] "wfibcore1 Spine 217B"
6	[16] "wfibcore1 Spine 111A"	63	[5] "wfibcore1 Spine 217B"
6	[1] "wfibcore1 Spine 111A"	63	[6] "wfibcore1 Spine 217B"
6	[24] "wfibcore1 Spine 111A"	66	[22] "wfibcore1 Spine 217B"
8	[12] "wfibcore1 Spine 111A"	84	[12] "wfibcore1 Spine 217B"
8	[13] "wfibcore1 Spine 111A"	84	[18] "wfibcore1 Spine 217B"
8	[18] "wfibcore1 Spine 111A"	105	[29] "wfibcore1 Spine 217B"
10	[36] "wfibcore1 Spine 111A"	126	[30] "wfibcore1 Spine 217B"
12	[29] "wfibcore1 Spine 111A"	147	[23] "wfibcore1 Spine 217B"
14	[30] "wfibcore1 Spine 111A"	147	[36] "wfibcore1 Spine 217B"
14	[35] "wfibcore1 Spine 111A"	176	[7] "wfibcore1 Spine 217B"

Non-contiguous
63 node allocation
1,456 routes

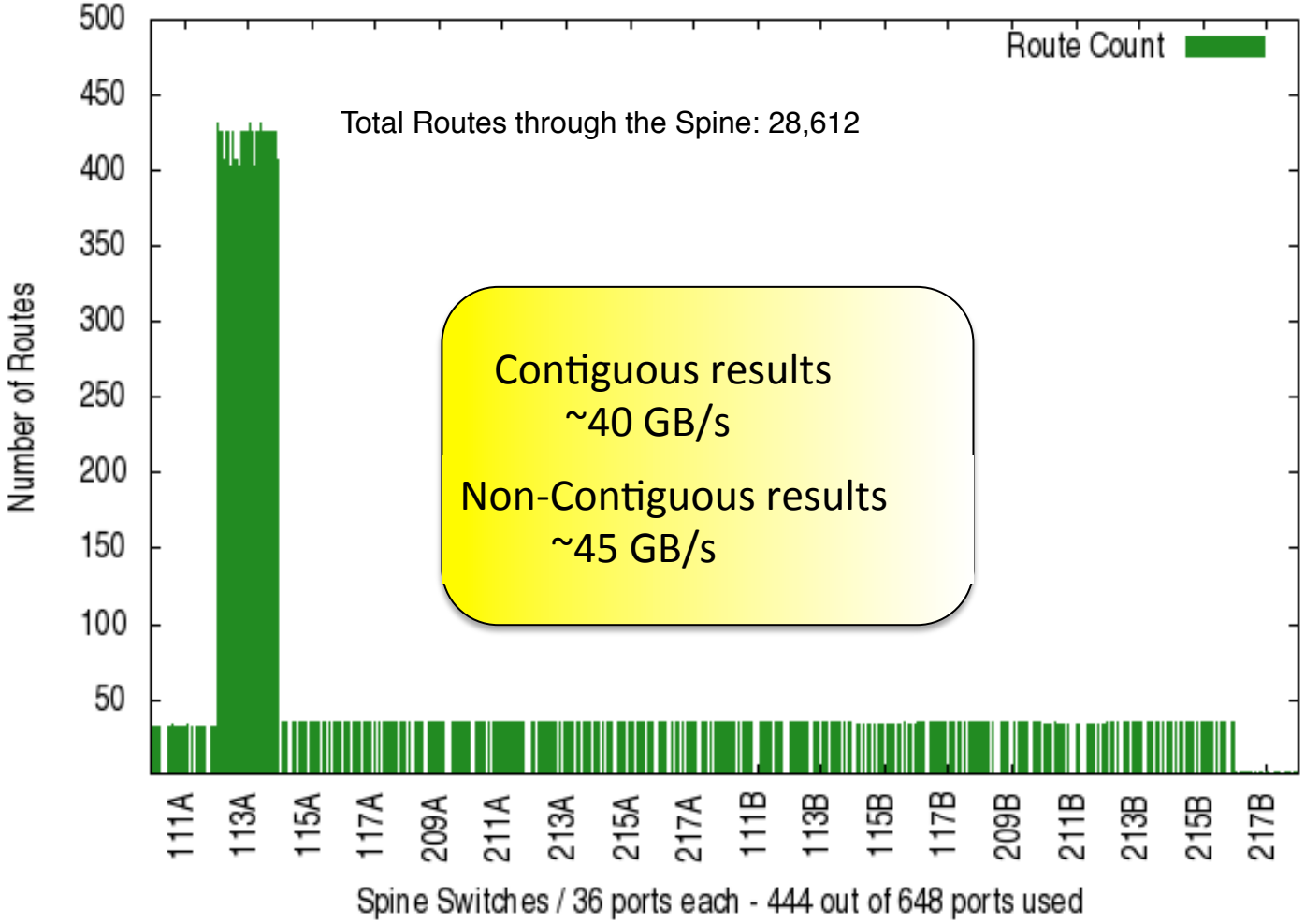
Wolf: All Compute Nodes to All IO Nodes
Spine Route Distribution

Before Changes



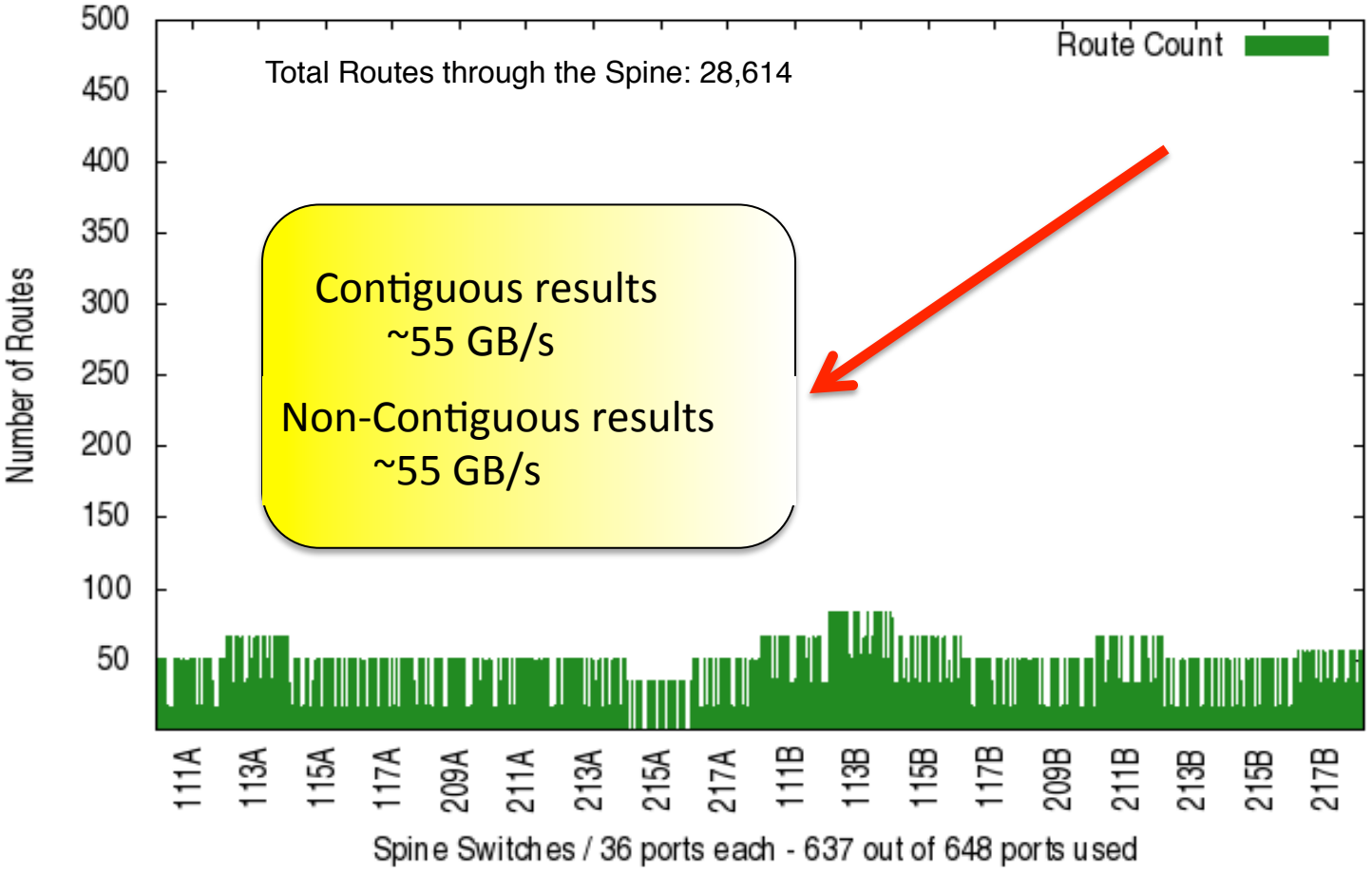
Wolf: All Compute Nodes to All IO Nodes
Spine Route Distribution

With the io_guid_file

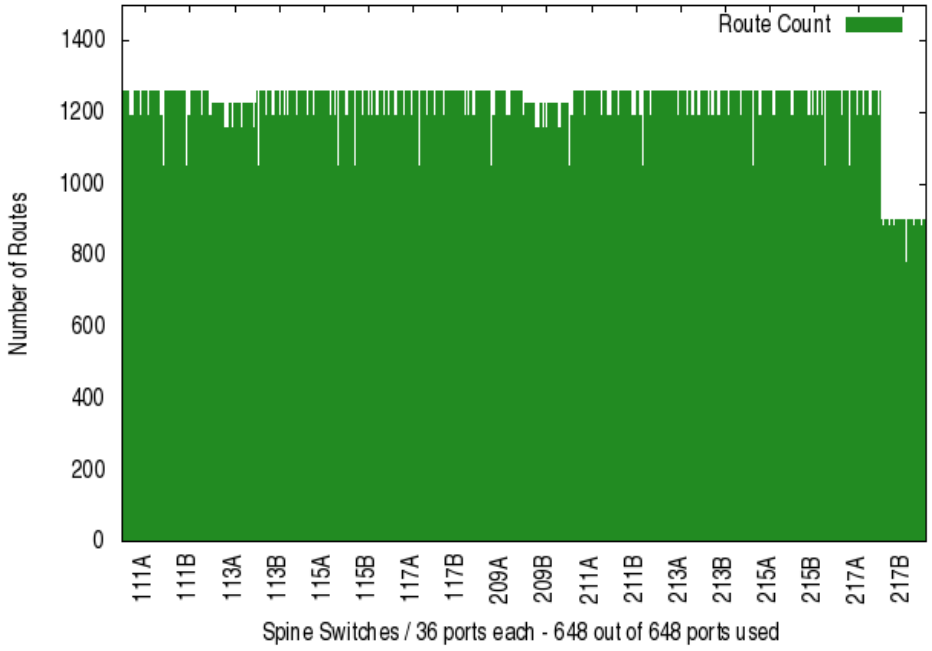


Wolf: All Compute Nodes to All IO Nodes Spine Route Distribution

With IO nodes physically moved

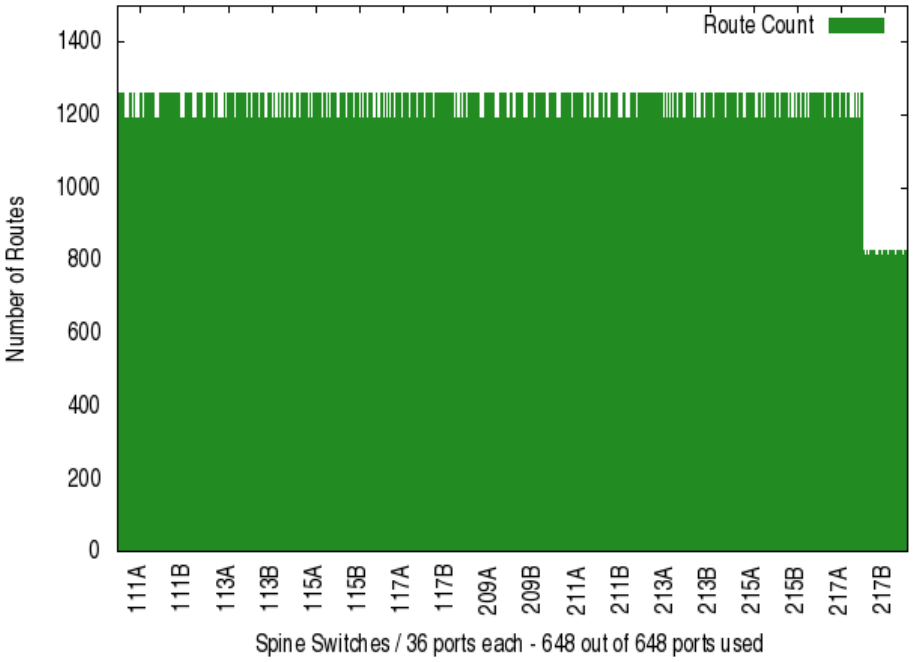


Entire Wolf Spine Route Distribution
All Compute / IO ⟷ All Compute / IO



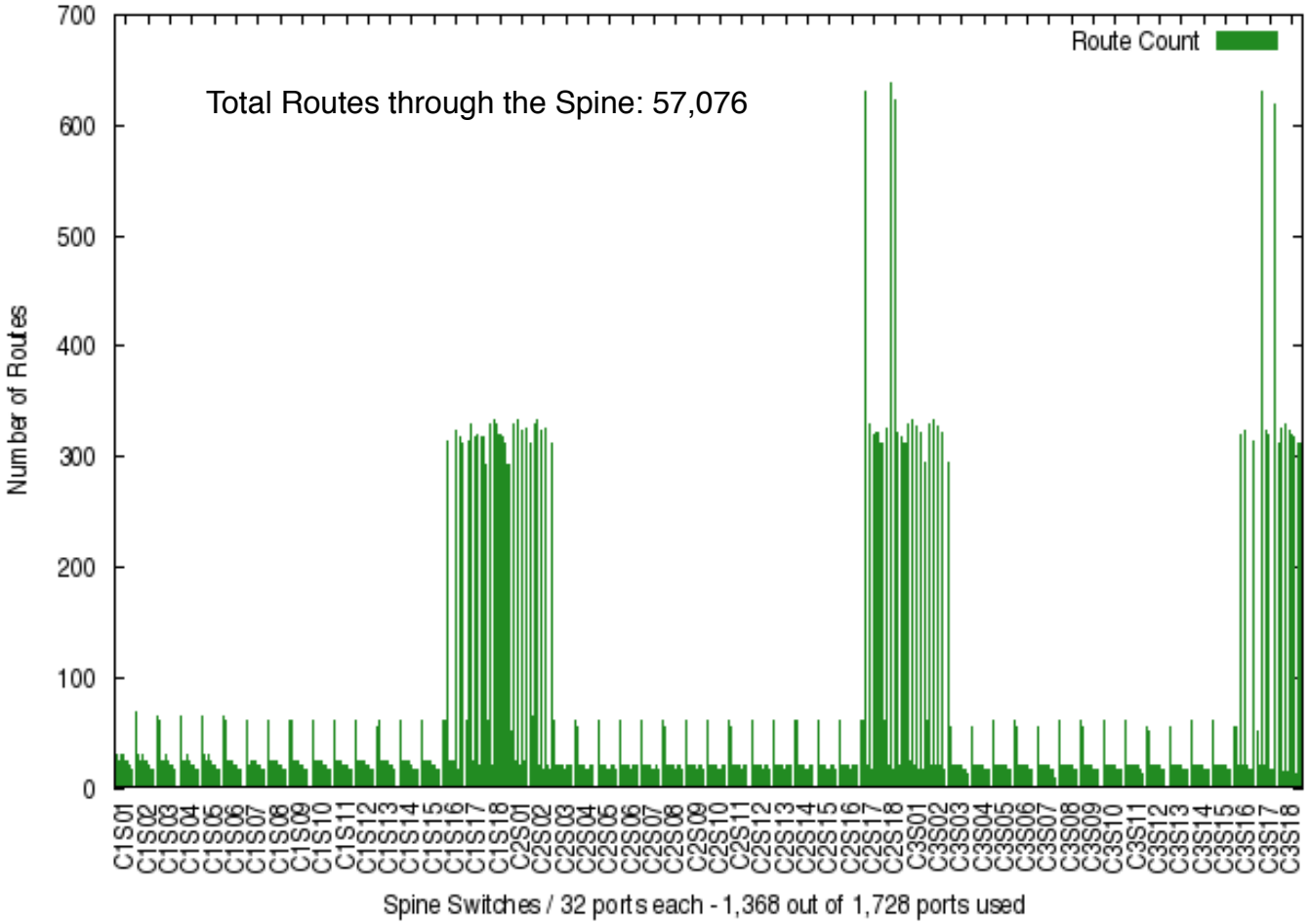
Before Changes

After IO nodes moved



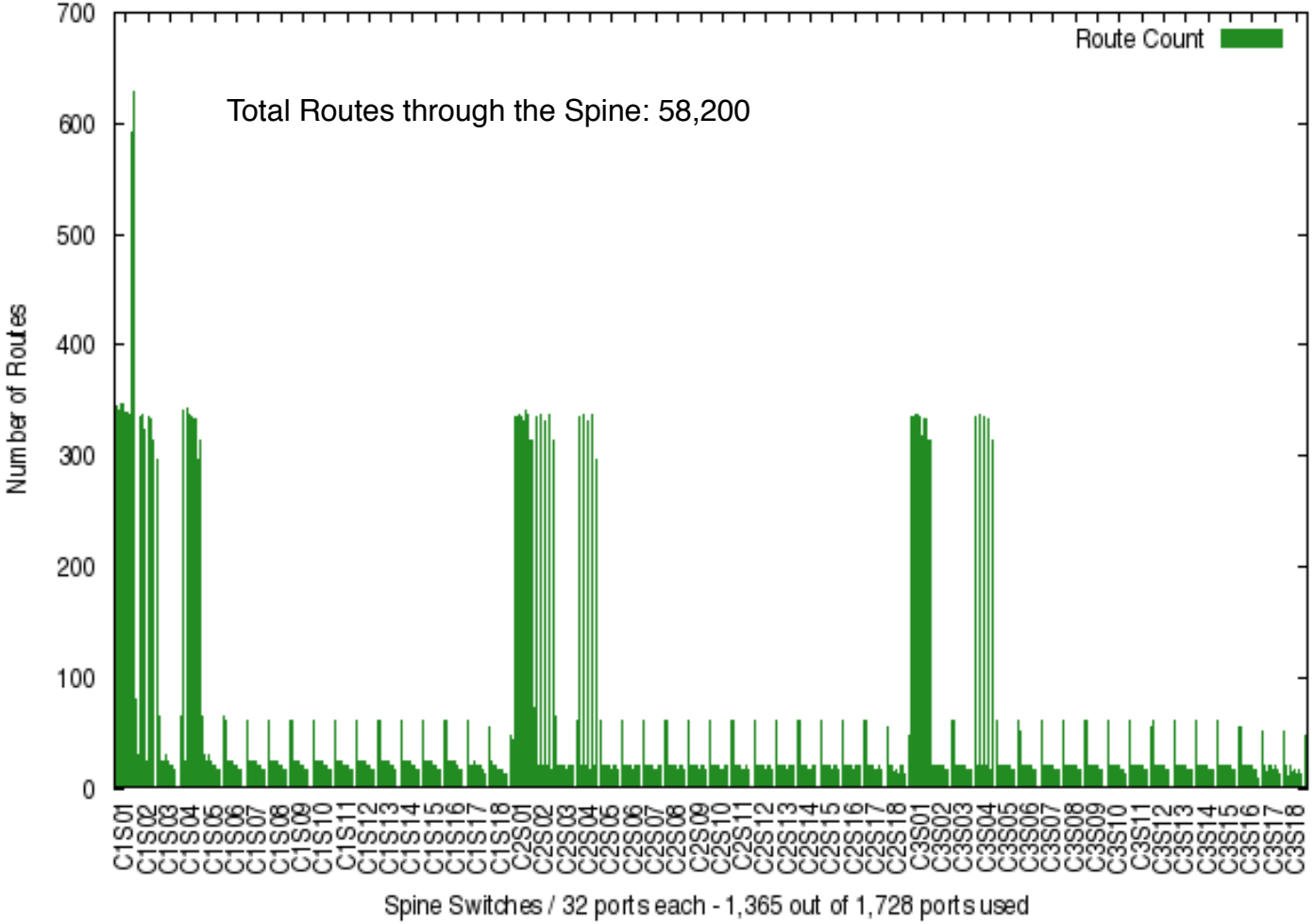
Mustang: All Compute Nodes to All IO Nodes Spine Route Distribution

Before Changes



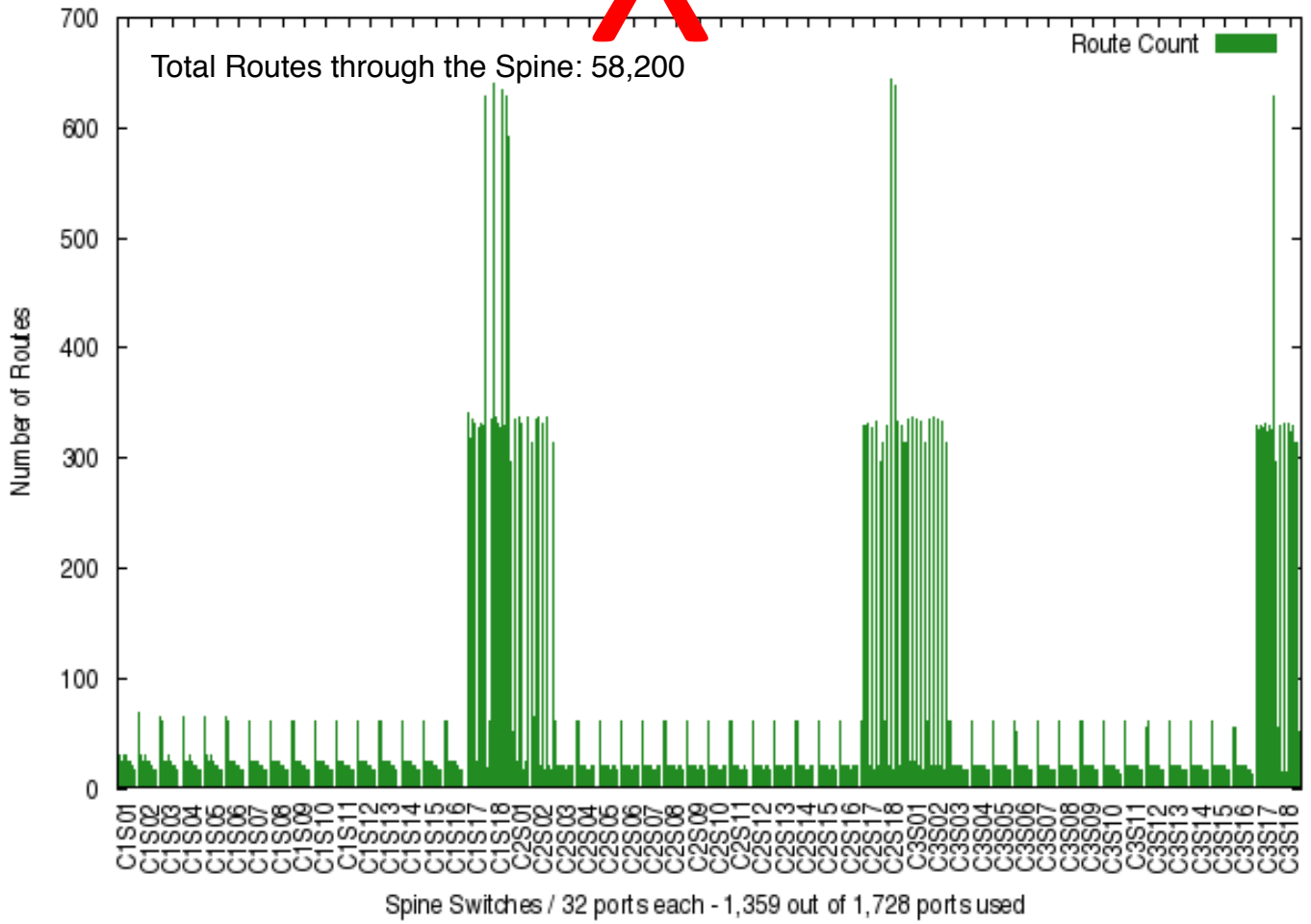
Mustang: All Compute Nodes to All IO Nodes Spine Route Distribution

With io_guid_file



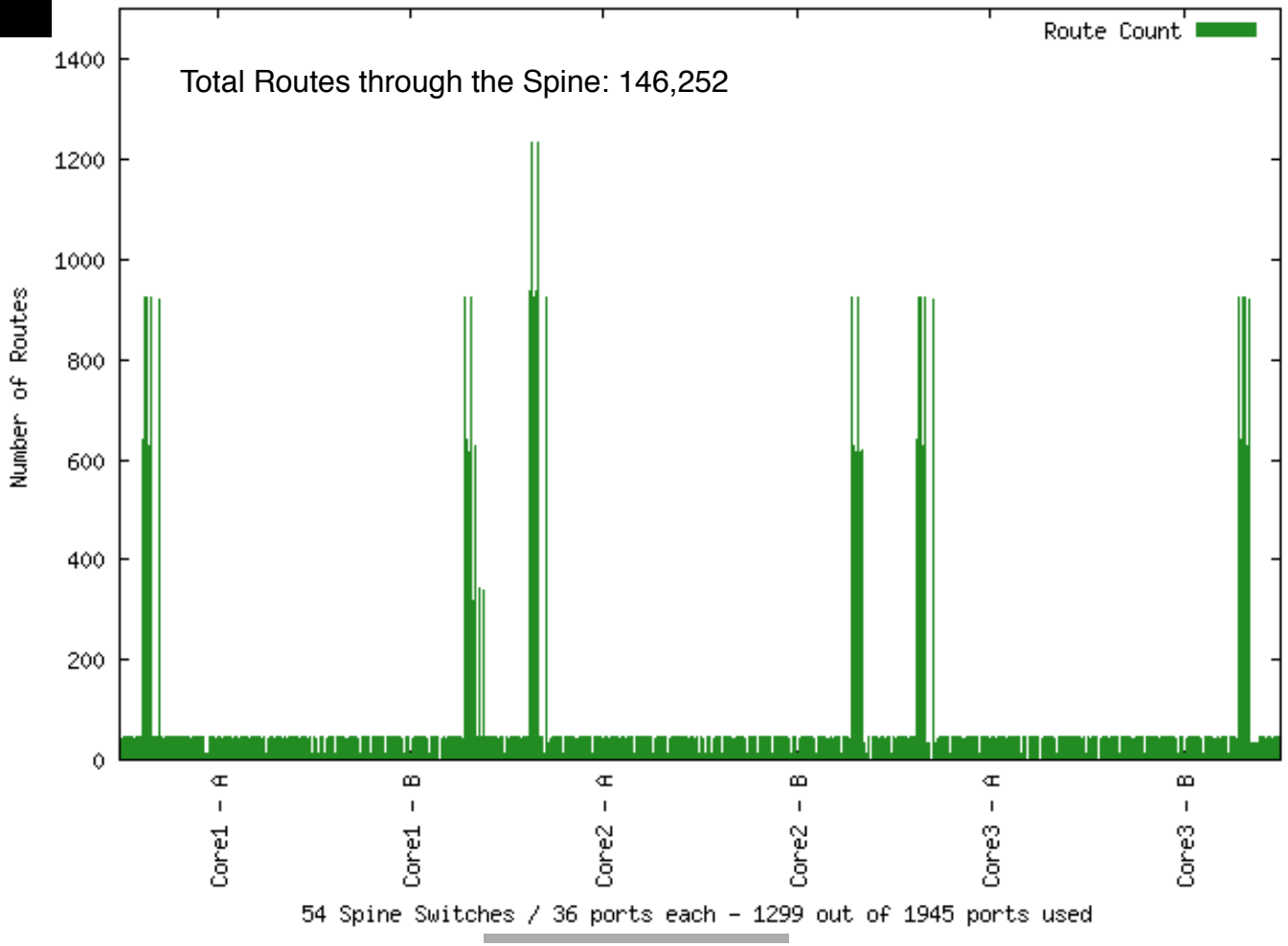
Mustang: All Compute Nodes to All IO Nodes Spine Route Distribution

With IO nodes physically moved



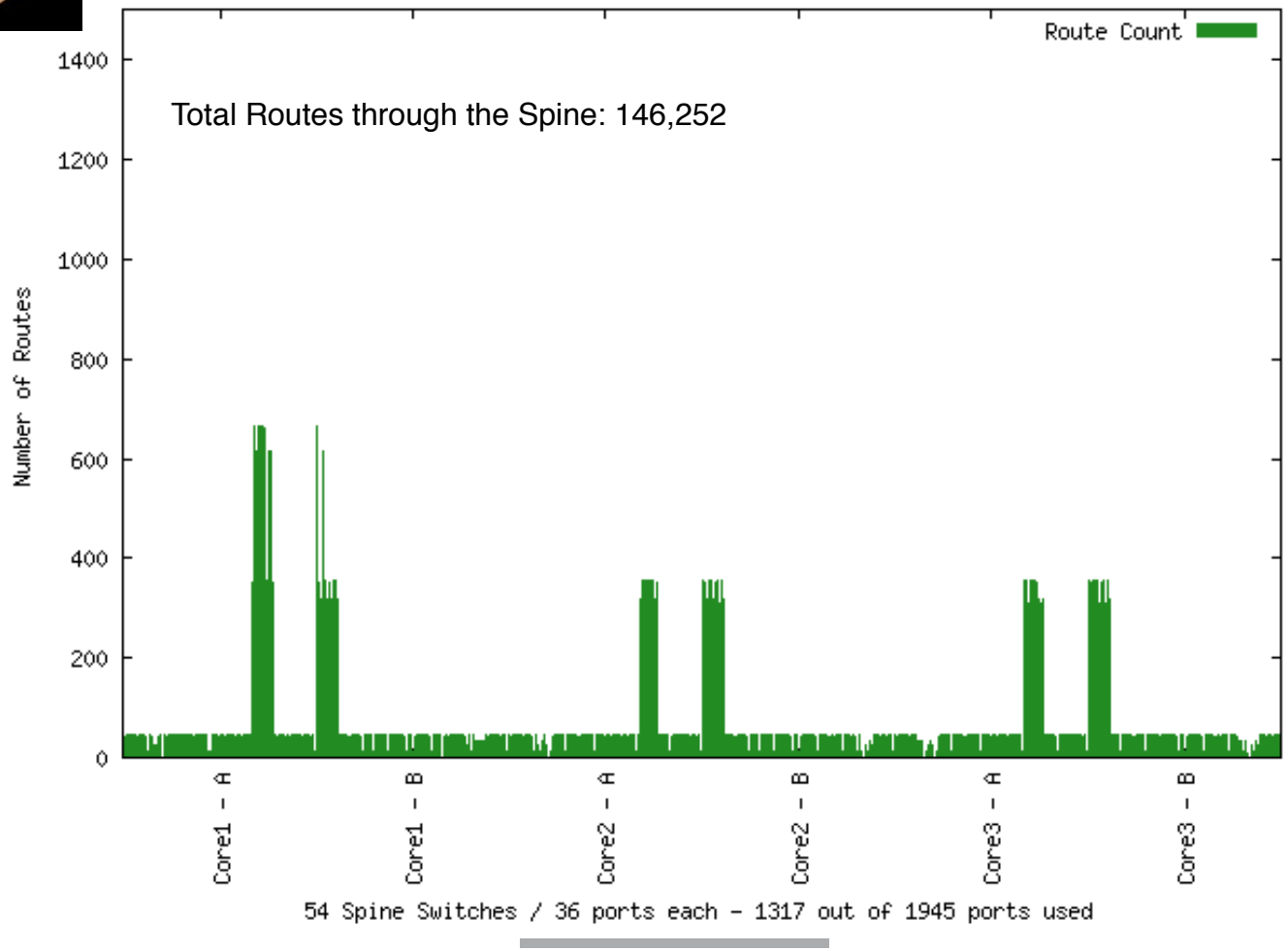
Luna: All Compute Nodes to All IO Nodes Spine Route Distribution

Before Changes



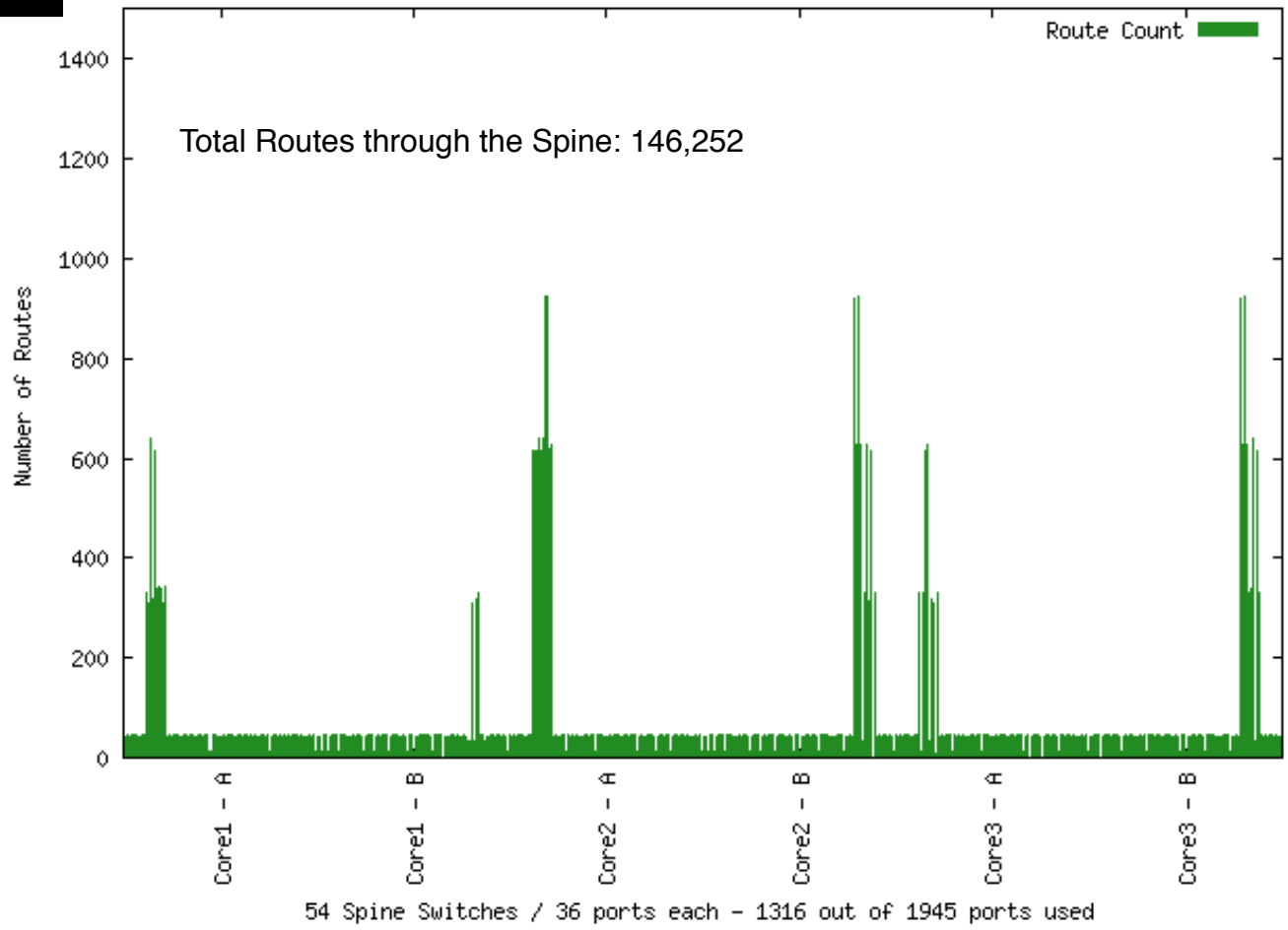
Luna: All Compute Nodes to All IO Nodes Spine Route Distribution

With io_guid_file



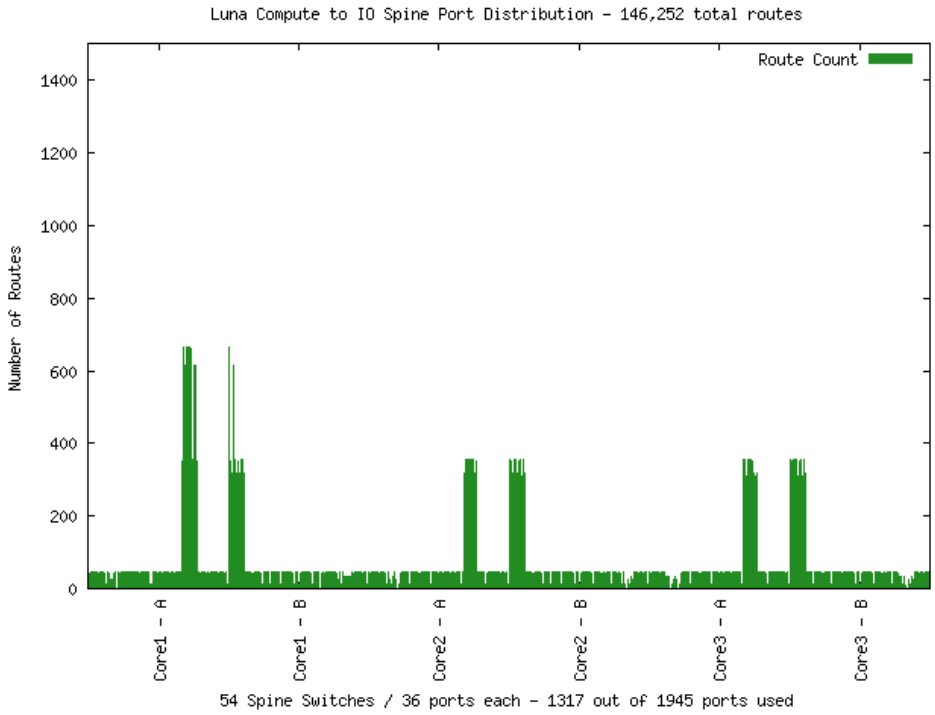
Luna: All Compute Nodes to All IO Nodes Spine Route Distribution

With IO nodes physically moved

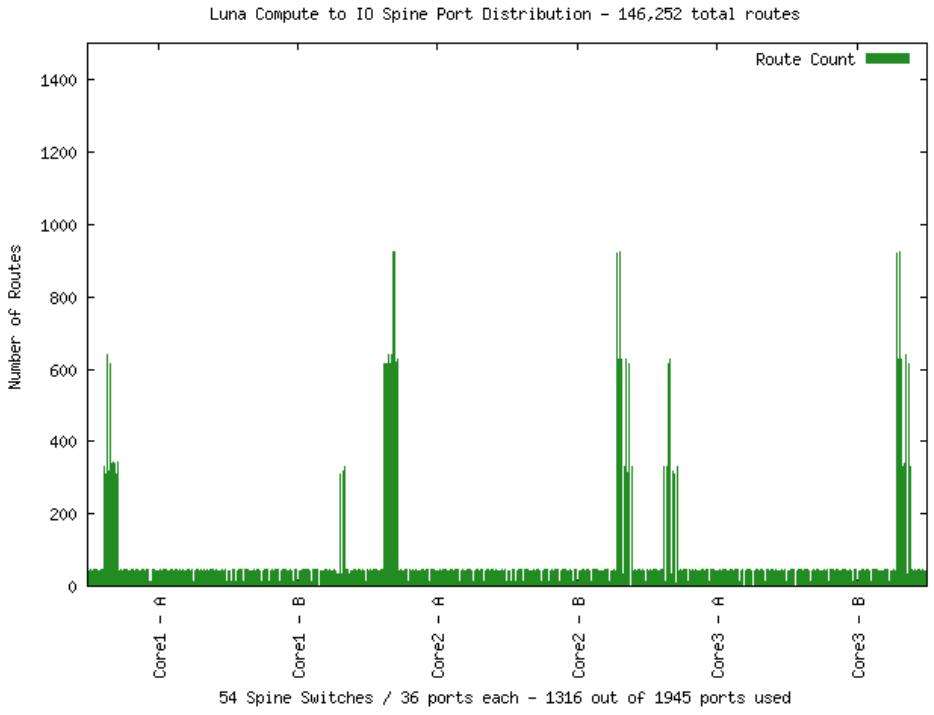


Luna: All Compute Nodes to All IO Nodes Spine Route Distribution

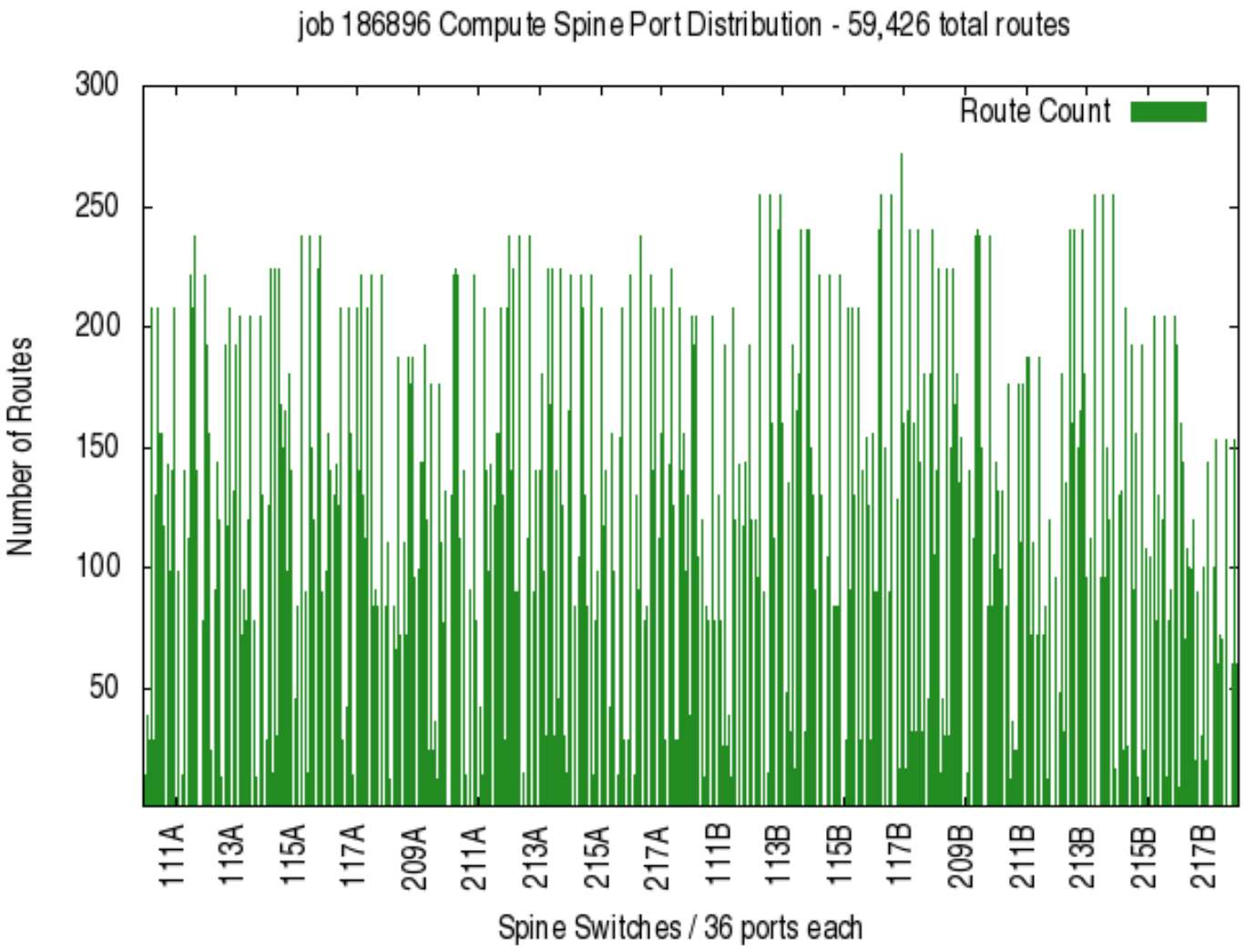
With io_guid_file



With IO nodes physically moved



Ability to graph routes by job now available via a script



Conclusions

- The high speed network (HSN) internal to a compute cluster is a significant contributor to external file system performance and throughput.
- Interconnect technology is extremely complicated, with very few simple assumptions holding true in the real world.
- Testing and experience at scale is critical to understanding and verifying HSN performance.
- Avoid putting more than one IO node per leaf switch if at all possible

OPENFABRICS
ALLIANCE

12th ANNUAL WORKSHOP 2016

THANK YOU

Jesse Martinez

Los Alamos National Laboratory