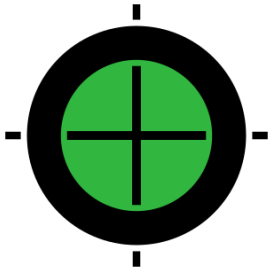
High Speed Testing

Doug Southworth · Indiana University · dojosout@iu.edu

perfSONAR is developed by a partnership of



Hardware Considerations

CPU

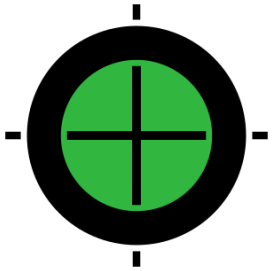
- Clock rate is king!
  - Bits have to make it from the NIC to RAM (and vice-versa), and the path between them is the CPU
  - Each TCP stream is single threaded, so multi-core alone won't help
  - Multiple high-speed cores are necessary for multi-thread transfers
- What to look for
  - Clock speed of at least 3.0 GHz per core for 40G testing, more for 100G
  - Enough cores to handle the anticipated number of simultaneous streams
  - Need physical cores, not hyperthreaded

RAM

- Massive amounts of RAM aren't necessary, but configuration counts
  - The goal is to maximize bandwidth between the CPU and RAM
  - This is usually achieved by occupying all RAM slots with identically sized modules
  - In practice this means using a greater quantity of smaller sized modules
- What to look for
  - Highest clock rate RAM your platform supports
  - Likely the smallest RAM modules you can find
  - This may result in more RAM than the system technically needs, but the goal is bandwidth, not strictly capacity

Storage

- Nothing on the market will cope with 100Gbps by itself
  - 100Gbps = 12,500MBps
  - SSD does about 600MBps on average
  - RAID arrays can help, but only so much
- Large distributed file systems could be the answer, but
  - What are you testing? Disk or Network?
  - Don’t allow scope creed to dominate your purchase decision



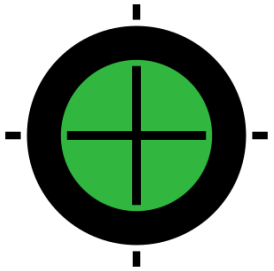
OS Considerations

Tuning is necessary

- Default OS network tunings aren't adequate
  - Can cope with 10G, but not 40/100G
  - Changing MTU, window size/TCP buffer, and even ring buffer are necessary
  - Packet pacing
- What to expect
  - MTU is pretty simple – 1500 is the default, 9000 is the goal, but for best performance the entire path needs to be 9000
  - 9000B packets can be fragmented down to 1500B, but at the cost of performance
  - Ring buffer tuning is a bit of an art. Trial and error will likely factor in.

Example: Arecibo data rescue

- Commodity hardware (Synology NAS) used to rescue and transfer irreplicable data
  - Tuned for 1G from the factory
  - 10G card was added, but initial results weren't promising (~800Mbps)
- Tuning was a big help
  - Ring buffers modified
  - TCP window modified
  - End result was ~4Gbps average, and the limiting factor was bus/CPU speed



Network Architecture

Switches

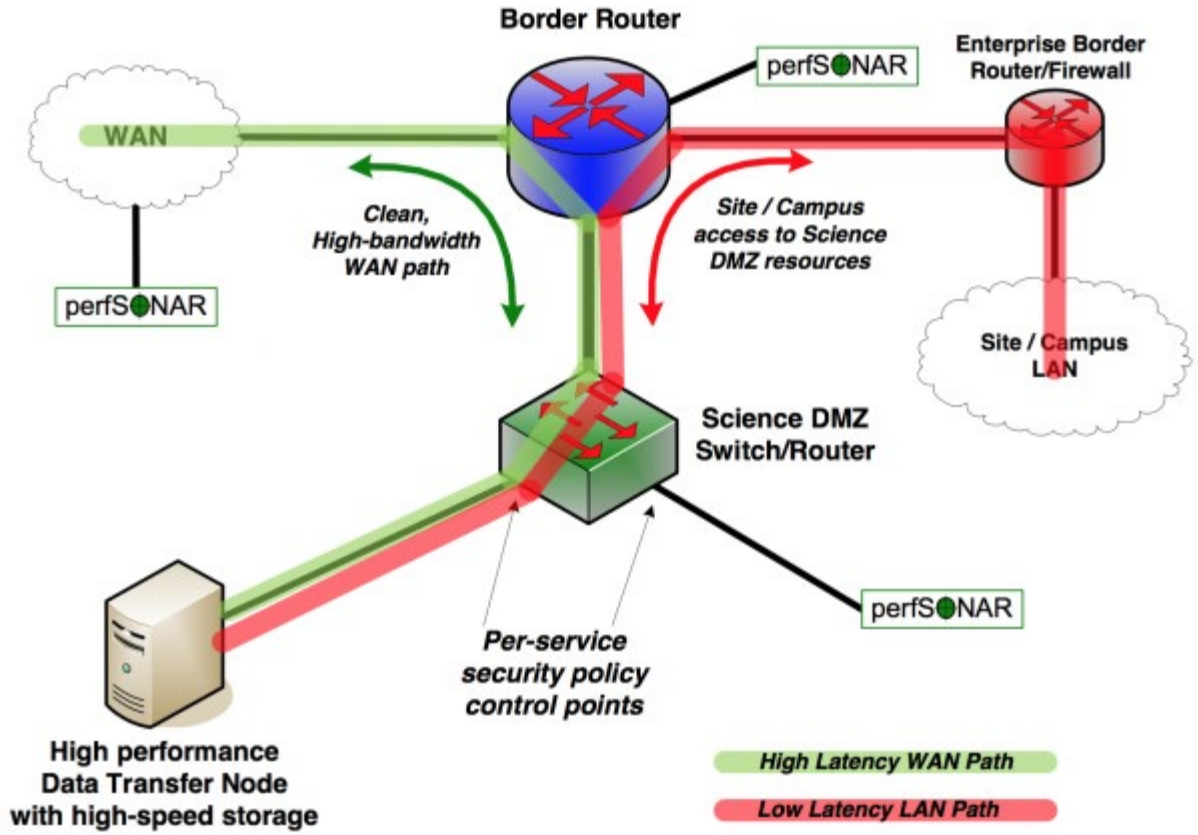
- Shallow buffers are everywhere
  - Shallow is a relative term when you approach 100G speeds
  - BDP of a 100G transfer over 50ms latency is 600MB!
- While deep buffered switches are necessary for high speed transfers, they can be detrimental to enterprise traffic
  - Understanding the use case is critical
  - Networks are not one-size-fits-all

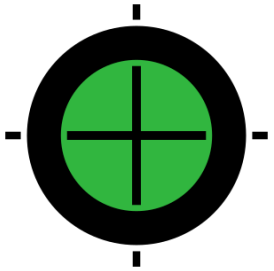
Firewalls and UTMs

- Often hamper performance
  - Not enough processor to maintain line speed on high speed transfers
  - Designed for “herds of mice”, not “column of elephants”
- Security must be carefully planned
  - Installing a device doesn’t mean your network is secure
  - Not having one in your network doesn’t mean it’s vulnerable

Science DMZ

- Enterprise networks aren't designed for large file transfers over long distances
  - Firewalls are rarely performant
  - Shallow buffer switches are prevalent
  - Packet loss is generally acceptable
- The Science DMZ is specifically designed to address these concerns while maintaining security





Some things to think about

Useful metrics

- Packet loss
  - The goal is as close to zero as possible
  - Bandwidth only tells part of the story
- Latency and routing
  - Is there a more direct path?
  - Can latency be lowered?
  - Is a higher latency path “cleaner?”
  - Have there been changes recently?

How the network is used

- 100G single stream doesn't happen in the wild, so testing that way has limited usefulness
  - 10G multi-stream will generally yield more realistic results
  - In many cases, the goal is to model end-user performance expectations
- Participation in a mesh is useful
  - …but only if it's monitored!
  - Set realistic thresholds and alarms

Human networking

- Routing Working Group
  - As the name would imply, route efficiency focused
  - Large overlap with high performance use cases
- Conferences
  - Hallway conversations are always valuable
  - More problems get solved over dinner than email
- Partners
  - Who are you (or your researchers) doing a lot of transfers with?

Data Mobility Exhibition

- Models real world achievable performance
  - Well tuned resources
  - Not perfSONAR centric per se, but participants are very active in the measurement and monitoring community
- https://fasterdata.es.net/science-dmz/learn-more/2019-2020-data-mobility-workshop-and-exhibition/

Question and answer icon by iconosphere from The Noun Project

Questions and Answers

High Speed Testing

Doug Southworth · Indiana University · dojosout@iu.edu

perfSONAR is developed by a partnership of