100Gbps Networks
Network Engineering Perspective

Eli Dart, Joe Metzger

100 Gbps transatlantic science trials workshop at SC13

Denver, CO

November 18, 2013

Outline

Equipment limitations and debugging experiences

High-performance data transfers and workflow decomposition

Commonalities – more common than you think

Experience With 100G Equipment

ESnet experiences

- Advanced Networking Initiative
- ESnet5 production 100G network
- Helping other people debug their stuff

Important takeaways

- R&E requirements are outside the design spec for most gear
  - Results in platform limitations – sometimes can't be fixed
  - You need to be able to identify those limitations before you buy
- R&E requirements are outside the test scenarios for most vendors
  - Bugs show up when R&E workload is applied
  - You need to be able to troubleshoot those scenarios

Platform Limitations

We have seen significant limitations in 100G equipment from all vendors with a major presence in R&E

- 100G single flow not supported
  - Channelized forwarding plane
  - Unexplained limitations
  - Sometimes the senior sales engineers don't know!
- Non-determinism in the forwarding plane
  - Performance depends on features used (i.e. config-dependent)
  - Packet loss that doesn't show up in counters anywhere

If you can't find it, nobody will tell you about it

- Vendors don't know or won't say
- Watch how you write your procurements

Second-generation equipment has proven to be much better

Vendors have been responsive in rolling new code to fix problems

They Don’t Test For This Stuff

Most sales engineers and support engineers don’t have access to 100G test equipment

- It’s expensive
- Setup of scenarios is time-consuming

R&E traffic profile is different than their standard model

- IMIX (Internet Mix) traffic is normal test profile
  - Aggregate web browsers, email, YouTube, Netflix, etc.
  - Large flow count, low per-flow bandwidth
  - This is to be expected – that’s where the market is
- R&E shops are the ones that get the testing done for R&E profile
  - SCinet provides huge value
  - But, in the end, it’s up to us

New Technology, New Bugs

Bugs happen.

- Data integrity (traffic forwarded, but with altered data payload)
- Packet loss
- Interface wedge
- Optics flaps

Monitoring systems are indispensable

Finding and fixing issues is sometimes hard

- Rough guess – difficulty exponent is degrees of freedom
  - Vendors/platforms, administrative domains, time zones

Takeaway – don’t skimp on test gear (at least maintain your perfSONAR boxes)

Design For Easy Debug

International circuits often have special circumstances

- Undersea cables
- Multiple administrative domains for one circuit

These things can make debugging harder than for terrestrial circuits

TCP loss impact and other issues are more damaging

It must be easy to run tests on international circuits

- Regular monitoring with perfSONAR
- As-needed testing for debugging specific issues

Workflow Decomposition

Many people still think in terms of one program running inside one system image on one computer

Workflows that process tens of terabytes of data must work differently

What does your workflow look like?

- What produces the data?
- Where is the storage?
- What does the analysis? (What storage goes with analysis?)
- Where can data be reduced?
- What can be automated?

Different components have different requirements

Proper decomposition can have significant benefits

Component Reuse

Many people understand about software reuse

Not many people understand workflow component reuse

Do you really want to re-invent the wheel?

- High-speed data transfer (Globus)
- Integration of virtualized components (OpenStack)
- Volume rendering, feature detection, FFT, CFD, …

Many scientists/experiments think they are a unique snowflake

- In some ways they are
- However, there is a set of tasks common to many workflows

Find your commonalities and exploit them – we can’t scale otherwise

Questions?

Thanks!

Eli Dart - dart@es.net

http://www.es.net/

http://fasterdata.es.net/