Deployment and Usage of perfSONAR Networking tools for non-HEP communities

Brian Davies (STFC)
Cristina DelCano Novales (STFC)
Alex White (DLS Ltd)

ISGC 2015 Taipei Taiwan
What is perfSONAR

• End to End testing and measurement infrastructure for multi domain network monitoring
  – www.perfsonar.net
• Various types of tests
  – Network throughput Rates (BWCTL), Latency, packet loss and routing
    • Using iperf, ping and traceroute
• Helps schedule and co-ordinate regular multi domain tests.
User Community 1: GridPP Tier1

- Supports part of the Worldwide LHC Computing Grid (WLCG) requirements for the Large Hadron Collider (LHC) within the UK.
- Main interest in results from BWCTL tests so as ensure fast throughput rates.
  - Users can use information to determine workflow methodology.
    - (Local vs Remote data access / Scheduling)
User Community: GridPP Tier1

• Sites use it to help discover network hardware/middleware problems
• ~170 sites within WLCG
• 0.224 EBytes (604M files)
  – In one year across WLCG
• Filesize varies from ~100kB to ~10GB
  – Log files, Raw and Processed data files
    • Disk/Tape
• 0.2-200ms latency
User Community 2: Diamond Light Source (DLS)

X-ray tomography and Crystal Diffraction Facility

- ~3k academic and industrial users
  - Fossils, jet engines, viruses, vaccines
  - Structural biology, energy, nanoscience, environmental science

- 100MB-300TB of data per user
  - High throughput needed.

- using NPAD tool perfSONAR to support remote users via ad-hoc troubleshooting

Co-Located at Rutherford Appleton Laboratory
Use Community: Diamond Light Source (DLS)

Similar sites but different communities/resources to WLCG

- **STRUBI**
  - Structural Biologists
    - Oxford UK
- **Membrane Protein Laboratory**
  - Imperial College London UK /Wellcome Trust
- **Rolls Royce** – Engine Fan Blades

With External partners
Climate Science Research

- International Climate Network Working Group
  - Under the Earth System Grid Federation
  - Network connections to allow scientists to work with petabytes of climate data across the world
  - Collaboration between Australia, Germany, Netherlands, UK and US
  - Aiming to achieve site-to-site data transfer rates of 4 to 8Gbps
  - 16Gbps by 2016
Monitoring a Worldwide system

Transfers to 6/7 continents.

73 endpoints require 250 Layer 3 Routers to connect to just one site.
Site Network Map
Test Configuration and Resource Discovery

Easy to Configure

• GUI setup of individual tests
• External MESH configuration control
  – Central Co-ordination for many sites
• CLI also available
• Separation of Latency and Throughput services

Easy Resource Discovery

• Communities
  – Easy to setup
  – Easy to find
• %Usage helpfully given
• GUI very Intuitive
  – Test host(s), type, interval, periodicity and protocol all configurable
**Easy to Configure Services**

The image shows a screenshot of a web page with a configuration tool for enabling services. The tool allows users to select which services to enable, such as 'Automatic Updates', 'BWCTL', 'OWAMP', 'NDT', and 'NPAD'. Each service has a description indicating its purpose, such as running Throughput, Traceroute, and Latency tests. The page also includes options to save or reset the configuration.
Easy to Configure Tests
Easy to Discover Hosts

Communities This Host Participates In

- ALICE ATLAS CMS GridPP Harwell Campus IANET Users LHCOPN LHCb
- WLCG

Other Communities As Of 2015-02-26 06:47

- 10G 50G AARNet AARNet-Edge ACORIN-NS AMLight APAN APAN-IP ARCON asnet-as am Atlas ATLASIT
- Bafle Blue Waters Brazil BRAA CAARN CAIRN CAIRN-HPC CANADIE CENIC Cenic CERNET CERNET
- CIC-CERNET 10G CNIC-CN CNRs ComputeCanada ComputeCanada ComputeCanada p5-NPToolkit-3.3
- Connecticut Education Network CERNET Diamond Light Source DOE Sites DOF-SC-LAB
- Emory University Enet Enet Enet FAU HPC Fermilab FIESR FIBRE-IR Florida LambdaLight
- FSQ Funsnet GAMMON GIANT GEIN Georgia Tech GEISA GPN and iam GridPP HEPI IF2 tested
- HIPPI-Canada ICCN ICNP ITRENAID ITPM IHEP IIWongNet ITPM INFN inteleaF Internet 2
- Internet2 Internet2 CIP TS Engineering IGN-X Kanaren LEARN LHC LHCOPN LHCONIC
- LHCNet LHCNet L2 UCC Los Netzos LSST LSU MAHC MAINZ MAINZ MAX MCNC Merit MCTP MIB
- MINTZ MITNET NASA NEREN NCSA NGI NHCC NISM NHIC NIH NIH NIE Octet NOAA NOAA-Boulder
- Net_archive Northern-Lights Notre Dame NOSE NOYNET ORNL OpenPOP OpenIP Open Source Grid
- OpenStack OSG out channels PennREN PennState PortClub port McCabe testing portSONAR portSONAR-PS
- PNNL PNNL PNNL PNNL-PNNL-PS p5-NPToolkit-3.3 p5-NPToolkit-3.1.1 p5-NPToolkit-3.3.2
- p5-NPToolkit-3.4 p5-NPToolkit-LIVSCO Purdue Qatar-Education.City REANIN RACIOLA RENATER
- RNP RNP RNP Russia SANSAN ScienceDMZ ScienceDMZ SAPP SfSC SfSC SIO SOUTH Dakota - REED
- SWY SWY South Carolina State RONI Stanford Southwest UNIV TEST Texas A&M University Thailand
External monitoring

- Can check to see if services are running
Shared Tools for meta-Monitoring
Single site Throughput Analysis

Endpoints within single site can have vastly different results

Must know your LAN and WAN (And test periodicity)

1.6Gbps

970Mbps
Single plot Analysis

Packet loss, throughput, latency
Results I

- Discovery of unexpected fragmentation of Jumbo frames in site router
- Firewall configuration issues
- Packet Loss due to congested link
- Internal broken/dirty fibres found and fixed
  - End user don’t (and shouldn’t) have network/computing knowledge. IE Application layer only knowledge
Results II

• Achieved 1.6Gbps transmission for DLS (BNL)
  – User can get dataset within a day
    • Current method is “DHLcopy”
• “Surprise” from Singapore user that rtt was 200ms
  – Corrected asymmetric routing
    • Reduce rtt by 40ms
• Improved intergroup communication and understanding of LAN topology
Results III

- Observe latency change in different routes
- Observe route changes
- 1G/10G inter-operation

Latency step change due to change in route measurable.
Latency vs rtt vs Throughput

Latency vs Throughput

Geography vs Internet vs physics
Summary

• Simple to deploy and exploit technology
• Extension from initial HEP community usage to non HEP communities has allowed for discovery and fixing of previously undiagnosed issues.
• Helps to increase understanding of intra-site/ inter-community network issues.

Performant Network -> Faster Transfers-> Reduce Latency in Workflows-> Faster Research Publications-> Happy User Community
Questions, Thanks and Contact

• Any questions
• Thanks to STFC Colleagues
  – Chris Brew, Philip Garrad, Ian Collier, Alex Dibbo
• Thanks to WLCG Collaborators
  – Shawn McKee, Duncan Rand
• For further information contact:
  – brian.davies@atsc.ac.uk
  – cristina.del-cano-novales@stfc.ac.uk
  – Alex.white@diamond.ac.uk
• Dual stacked IPv6 enabled. Over to Dave…