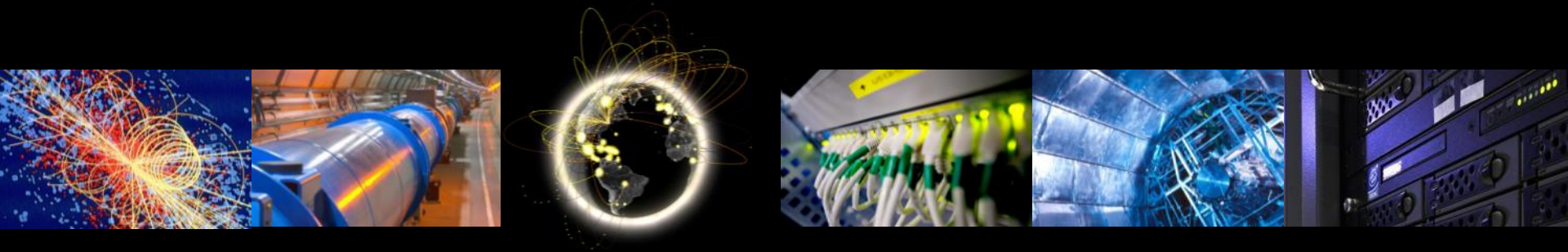


Update on OSG/WLCG perfSONAR infrastructure

Shawn McKee, Marian Babik

HEPIX Spring Workshop, Oxford
23rd - 27th March 2015



Network Monitoring in WLCG/OSG

- Goals:
 - Find and isolate “network” problems; alerting in time
 - Characterize network use (base-lining)
 - Provide a source of network metrics for higher level services
- Choice of a standard open source tool: perfSONAR
 - Benefiting from the R&E community consensus
- Tasks achieved by the perfSONAR TF:
 - Get monitoring in place to create a baseline of the current situation between sites
 - Continuous measurements to track the network, alerting on problems as they develop
 - Develop test coverage and make it possible to run “on-demand” tests to quickly isolate problems and identify problematic links

Network and Transfer Metrics WG

- Started in May 2015, bringing together network and transfer experts
- Follows up on the perfSONAR TF goals
- Mandate
 - Ensure all relevant **network** and **transfer metrics** are identified, collected and published
 - Ensure sites and experiments can better understand and fix networking issues
 - Enable use of network-aware tools to improve transfer efficiency and optimize experiment workflows
- Membership
 - WLCG perSONAR support unit (regional experts), WLCG experiments, FTS, Panda, PhEDEx, FAX, Network experts (ESNet, LHCOPN, LHCONE)

Network and Transfer Metrics WG

- Objectives
 - Coordinate commissioning and maintenance of WLCG network monitoring
 - Finalize perfSONAR deployment
 - Ensure all links continue to be monitored and sites stay correctly configured
 - Verify coverage and optimize test parameters
 - Identify and continuously make available relevant transfer and network metrics
 - Document metrics and their use
 - Facilitate their integration in the middleware and/or experiment tool chain
- Since inception, main focus was to finalize deployment and commissioning, extend the infrastructure, but also to jump start common projects with network and transfer metrics

perfSONAR Deployment

251 perfSONAR instances
registered in GOCDB/OIM
228 Active perfSONAR instances
223 Running latest version (3.4.1)
Instances at WIX, MANLAN,
GEANT Amsterdam



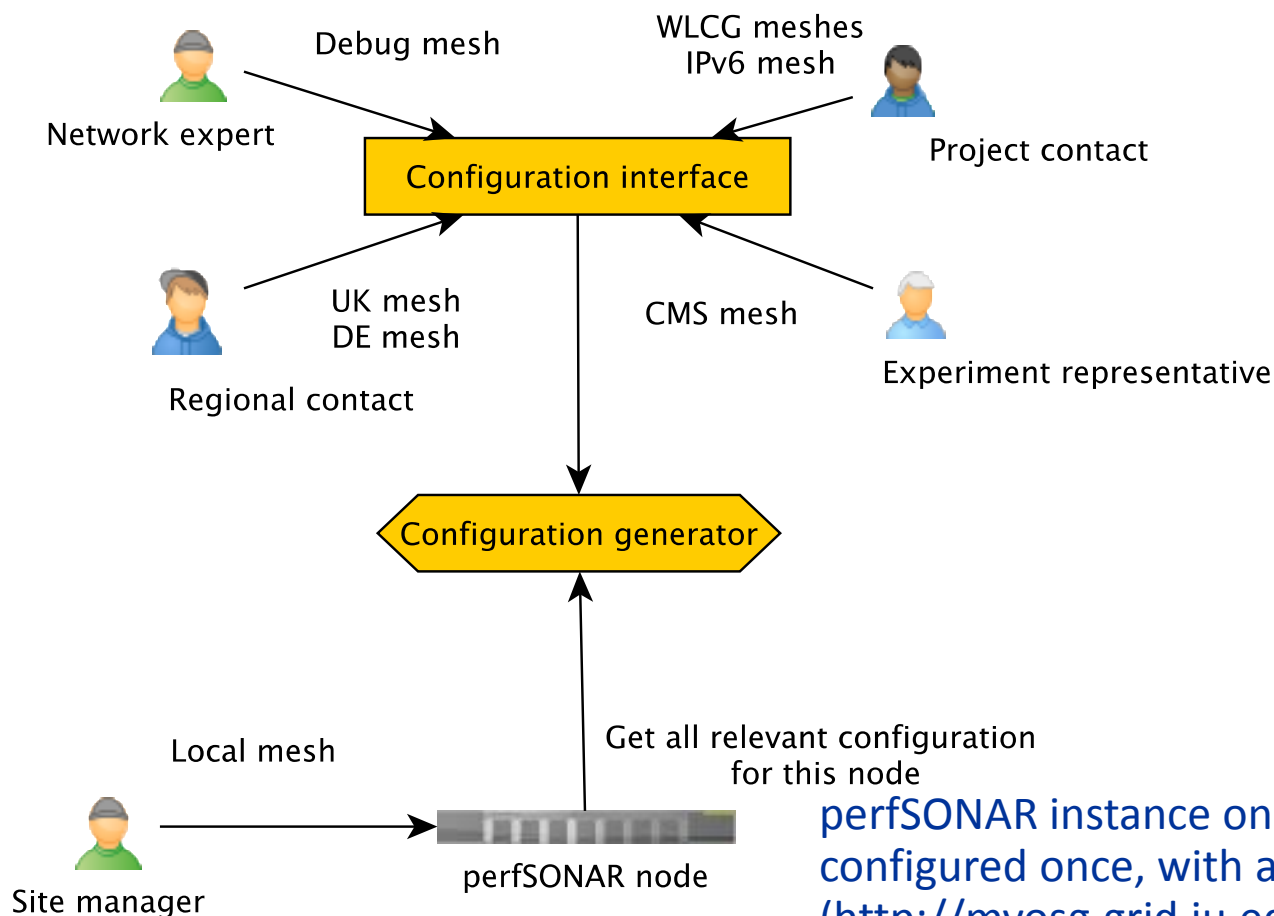
- Initial deployment coordinated by WLCG perfSONAR TF
- Commissioning of the network followed by WLCG Network and Transfer Metrics WG

perfSONAR Metrics and Meshes

- Tests are organized in meshes – set of instances that test to each other
- perfSONAR regular tests currently configured
 - Traceroute: End to end path, important to understand context of other metrics (Full WLCG mesh/hour)
 - Throughput: Notice problems and debug network, also help differentiate server problems from path problems (Full WLCG mesh/week)
 - Latency: Notice route changes, asymmetric routes and watch for excessive Packet Loss (Regional meshes, Continuous, 10Hz)
- perfSONAR is a testing framework, new tests and tools can be integrated as they become available
 - From iperf to iperf3, traceroute to tracepath
- Dynamic reconfigurations now possible
 - Creation, modification of meshes
 - Test frequency and parameters
- Additional perfSONAR nodes inside local network, and/or at periphery still needed (on LHCONE: MANLAN, WIX, GEANT)
 - Characterize local performance and internal packet loss
 - Separate WAN performance from internal performance

Configuration Interface


- perfSONAR instance can participate in more than one mesh
- Configuration interface and auto-URL enables dynamic re-configuration of the entire perfSONAR network



perfSONAR instance only needs to be configured once, with an auto-URL (http://myosg.grid.iu.edu/pfmesh/hostname/<node_id>)

Configuration Interface

← → ↻ <https://oim.grid.iu.edu/oim/meshconfig>

 OIM ▾ Home Certificate Topology Downtimes Virtual Organizations Support Centers Campus Grids Projects

Mesh Config Administrator

Host Groups Parameter Sets **Configs** Tests

Configuration to be part of

* Required

Service Type * Required

Name * Required

Parameters * Required

Mesh Type * Required

Host Group A

☐ Disable

Infrastructure Monitoring

- Based on OMD/check_mk and MadDash
- MadDash developed as part of perfSONAR – 1.2 version released recently
- OMD/check_mk extended to cover WLCG perfSONAR needs
- Developed bootstrapping and auto-configuration scripts
 - Synchronized with GOCDB/OIM and OSG configuration interface
- Packaged and deployed in OSG
- Developed new plugins - core functionality
 - Toolkit Version, Regular Testing, NTP, Mesh configuration, Esmond (MA), Homepage, Contacts
 - Updated to perfSONAR 3.4 information API/JSON
- High level functionality plugins
 - Esmond freshness – checks if perfSONAR node's local MA contains measurements it was configured to perform
 - Extremely useful during commissioning

Infrastructure Monitoring

- Auto-summaries are available per mesh
- Service summaries per metric type

https://psomd.grid.iu.edu/WLCGperfSONAR/check_mk/

Check **MK** 1.2.4p5 Hostgroup OPN

Availability

Tactical Overview

Hosts	Problems	Unhandled
251	30	30

Services

Services	Problems	Unhandled
3084	720	720

Quicksearch

Views

- Dashboards
 - Host & Services Problems
 - Main Overview
 - Network Topology
- Hosts
 - All hosts
 - All hosts (Mini)
 - All hosts (tiled)
 - Favourite hosts
 - Host search

Hostgroup OPN

state	Host	Icons	Alias	OK	Wa	Un	Cr	Pd
UP	ccperfsnar1.in2p3.fr		ccperfsnar1.in2p3.fr	10	2	0	0	0
UP	ccperfsnar2.in2p3.fr		ccperfsnar2.in2p3.fr	10	2	0	0	0
UP	lcgps01.gridpp.rl.ac.uk		lcgps01.gridpp.rl.ac.uk	10	2	0	0	0
UP	lcgps02.gridpp.rl.ac.uk		lcgps02.gridpp.rl.ac.uk	10	2	0	0	0
UP	lhcbandwidth.twgrid.org		lhcbandwidth.twgrid.org	10	2	0	0	0
UP	lhclatency.twgrid.org		lhclatency.twgrid.org	9	0	0	3	0
UP	lhcmn.bnl.gov		lhcmn.bnl.gov	9	2	0	1	0
UP	lhcmperfmon.bnl.gov		lhcmperfmon.bnl.gov	10	2	0	0	0
UP	perfsnar-bw.cern.ch		perfsnar-bw.cern.ch	9	2	0	1	0
UP	perfsnar-de-kit.gridka.de		perfsnar-de-kit.gridka.de	9	2	0	1	0
UP	perfsnar-it.cern.ch		perfsnar-it.cern.ch	10	2	0	0	0
	perfsnar-ow.cnaf.infn.it		perfsnar-ow.cnaf.infn.it	10	2	0	0	0
	perfsnar-ps.cnaf.infn.it		perfsnar-ps.cnaf.infn.it	9	2	0	1	0
	perfsnar-ps.ndgf.org		perfsnar-ps.ndgf.org	10	2	0	0	0
	perfsnar-ps2.ndgf.org		perfsnar-ps2.ndgf.org	10	2	0	0	0
	perfsnar2-de-kit.gridka.de		perfsnar2-de-kit.gridka.de	10	2	0	0	0
	ps-bandwidth.lhcmn.triumf.ca		ps-bandwidth.lhcmn.triumf.ca	10	2	0	0	0
	ps-gsdc01.sdfarm.kr		ps-gsdc01.sdfarm.kr	10	2	0	0	0
	ps-gsdc02.sdfarm.kr		ps-gsdc02.sdfarm.kr	10	2	0	0	0
	ps-latency.lhcmn.triumf.ca		ps-latency.lhcmn.triumf.ca	10	2	0	0	0
	ps.lhcopn-ps.sara.nl		ps.lhcopn-ps.sara.nl	10	2	0	0	0
	ps2.lhcopn-ps.sara.nl		ps2.lhcopn-ps.sara.nl	10	2	0	0	0
	psb01.pic.es		psb01.pic.es	10	2	0	0	0
	psl01.pic.es		psl01.pic.es	10	2	0	0	0
	psnar3.fnal.gov		psnar3.fnal.gov	10	2	0	0	0
	psnar4.fnal.gov		psnar4.fnal.gov	10	2	0	0	0

t2ps-bandwidth.physics.ox.ac.uk

State	Service	Status detail	Age	Checked	Icons	Perf-O-Meter
OK	perfSONAR 3.4+ Toolkit Version	OK toolkit version found 3.4.1	2015-02-17 07:22:26	27 sec		
OK	perfSONAR Administrator Details	OK - Administrator is Ewan Mac Mahon, email e.macmahon1@physics.ox.ac.uk (cached:0)	2014-12-11 19:57:58	3 hrs		
OK	perfSONAR BWCTL Bandwidth Test Controller	TCP OK - 0.139 second response time on 163.1.5.211 port 4823	2014-12-11 19:58:23	29 min		139.213 ms
WARN	perfSONAR esmond Freshness Bandwidth Direct	WARNING Found stale hosts for certain events, time-range: 3700	2015-02-17 21:47:47	3 hrs		
WARN	perfSONAR esmond Freshness Bandwidth Reverse	WARNING Found stale hosts for certain events, time-range: 3700	2015-02-17 21:48:10	3 hrs		
OK	perfSONAR esmond Measurement Archive	OK esmond reachable	2014-12-11 19:56:42	3 hrs		
OK	perfSONAR Homepage	OK homepage reachable	2015-01-27 19:58:50	3 hrs		
OK	perfSONAR Latitude/Longitude Configured	OK - Latitude is 51.81806, Longitude is -1.30489 (cached:1)	2014-12-11 19:54:37	3 hrs		
OK	perfSONAR Mesh Configuration	OK auto-URL configured	2015-01-26 13:55:06	3 hrs		
OK	perfSONAR NTP Service	OK NTP synchronized	2015-01-29 20:16:35	28 min		
OK	perfSONAR Regular Testing Service	OK Regular Testing enabled and running	2015-01-29 20:17:03	28 min		
OK	perfSONAR Toolkit Version	OK - Version 3.4.1 OK (cached:1)	2014-12-11 19:56:17	3 hrs		

Stale services

- Addons
- Search Graphs
- Other
- Comments

UP psnar3.fnal.gov

UP psnar4.fnal.gov

OSG perfSONAR Datastore

- All perfSONAR metrics should be collected into the OSG network datastore
 - This is an Esmond datastore from perfSONAR (postgresql+cassandra backends)
 - Loaded via RSV probes; currently one probe per perfSONAR instance every 15 minutes.
- Validation and testing ongoing in OSG
 - Plan is to have it production ready by Q3
- Datastore on psds.grid.iu.edu
 - JSON at <http://psds.grid.iu.edu/esmond/perfsonar/archive/?format=json>
 - Python API at http://software.es.net/esmond/perfsonar_client.html
 - Perl API at <https://code.google.com/p/perfsonar-ps/wiki/MeasurementArchivePerlAPI>

Datastore API

← → ↺ psds.grid.iu.edu/esmond/perfsonar/archive/?format=json&limit=10&time-range=3600&event-type=packet-trace

```
{
  "destination": "129.107.255.30",
  "event-types": [
    {
      "base-uri": "/esmond/perfsonar/archive/c72f59c8c2104da5b0820fed0dbf82ab/packet-trace/base",
      "event-type": "packet-trace",
      "summaries": [],
      "time-updated": 1427193876
    },
    {
      "base-uri": "/esmond/perfsonar/archive/c72f59c8c2104da5b0820fed0dbf82ab/failures/base",
      "event-type": "failures",
      "summaries": [],
      "time-updated": null
    },
    {
      "base-uri": "/esmond/perfsonar/archive/c72f59c8c2104da5b0820fed0dbf82ab/path-mtu/base",
      "event-type": "path-mtu",
      "summaries": [],
      "time-updated": null
    }
  ],
  "input-destination": "http://localhost",
  "input-source": "http://lpsc-perfsonar.in2p3.fr",
  "measurement-agent": "193.48.83.97",
  "metadata-key": "c72f59c8c2104da5b0820fed0dbf82ab",
  "org_metadata_key": "9962f9dc3364490e881d52674bd714a3",
  "source": "193.48.83.97",
  "subject-type": "point-to-point",
  "tool-name": "bwctl/tracepath,traceroute",
  "uri": "/esmond/perfsonar/archive/c72f59c8c2104da5b0820fed0dbf82ab/"
},
```

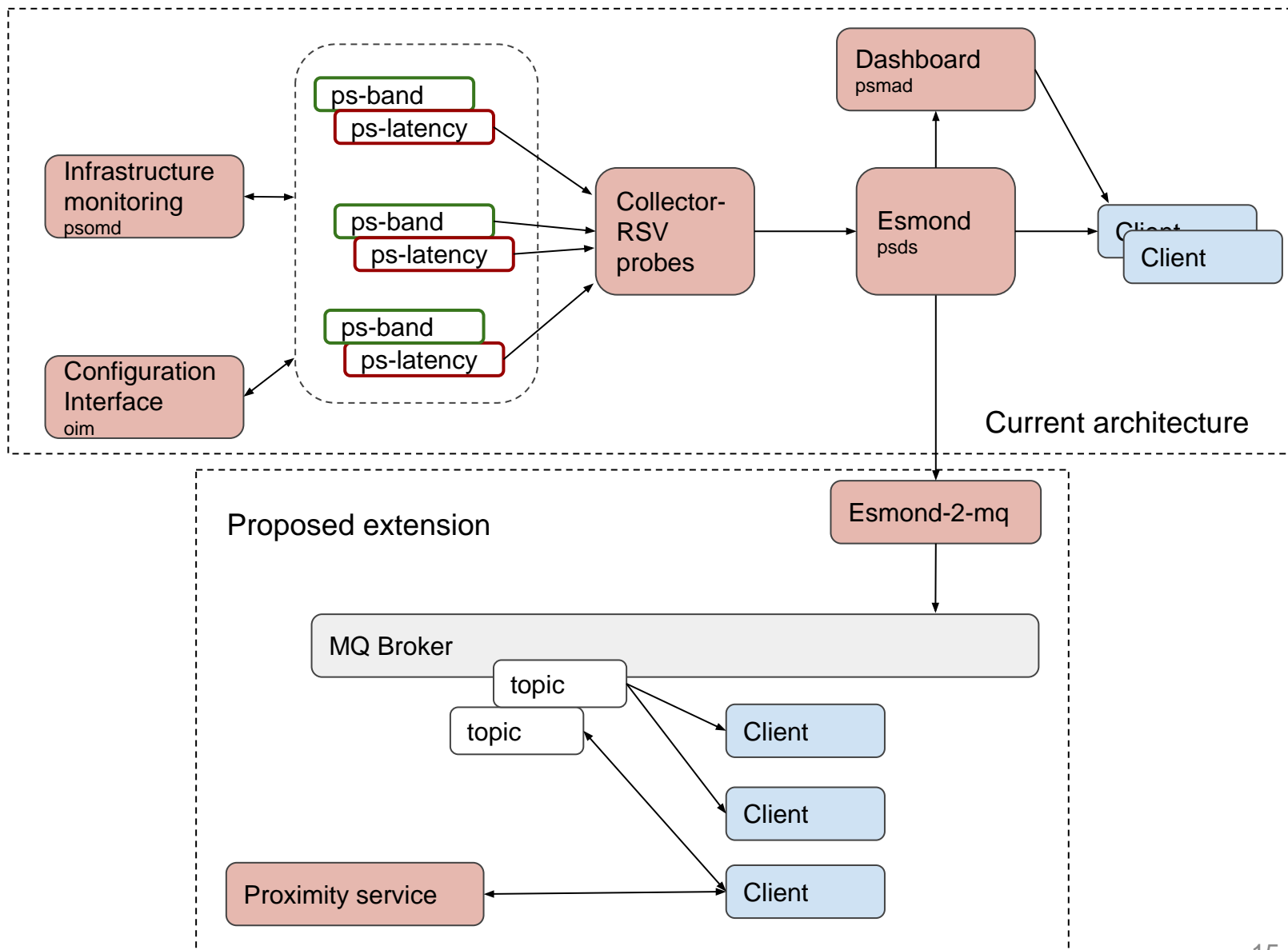
Integration Projects

- Goal
 - Provide platform to integrate network and transfer metrics
 - Enable network-aware tools (see ANSE <http://cern.ch/go/M9Sj>)
 - Network resource allocation along CPU and storage
 - Bandwidth reservation
 - Create custom topology
- Plan
 - Provide latency and trace routes and test how they can be integrated with throughput from transfer systems
 - Provide mapping between sites/storages and sonars
 - Uniform access to the network monitoring
- Pilot projects
 - FTS performance – adding latency and routing to the optimizer
 - Experiment's interface to datastore

Experiments Interface to Datastore

- Aim
 - Develop publish/subscribe interface to perfSONAR
 - Enable possibility to subscribe to different events (filter) and support different clients - integration via messaging, streaming data via topic/queue
 - Provide mapping/translation btw sonar infrastructure and experiment's topology
- Components
- esmond2mq – prototype already exists
 - Retrieves all data (meta+raw) from esmond depending on existing mesh configs
 - Publishes to a topic
- Proximity/topology service
 - Handle mapping/translation of services (service to service; storage to sonar), service to site (sonar to site)
 - Test different algorithms (site mapping, traceroutes, geoip)
 - Evaluate if existing tools can be reused for this purpose

Proposed Extension



PuNDIT Project

Georgia Tech College of Computing
UNIVERSITY OF MICHIGAN
pundit.gatech.edu

Pythia Network Diagnosis Infrastructure (PuNDIT)

PIs: Shawn McKee (smckee@umich.edu) and
Constantine Dovrolis (dovrolis@cc.gatech.edu)
Members: Jorge Batista and Danny Lee

National Science
Foundation

Award No. 1440571 and 1440585
SI2 Project Type: SSE



Problem Statement

- Monitoring of the network infrastructure is key to efficient distributed collaborations
- Currently, intermittent network problems are manually identified
 - operator dashboards
 - user trouble tickets
- PuNDIT aims to automate the detection and localization of network problems

Objectives

- Key idea: Convert complex network metrics into easily understood diagnoses in an automated way
- Integrate with the de-facto standard perfSONAR network measurement infrastructure



- Work with paris-traceroute developers to create accurate localization in perfSONAR
- PuNDIT aims to:
 - Identify short-term events
 - Produce results in near real time
 - Scale to large number of agents
 - Visualize useful summaries



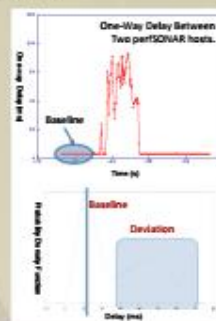
How PuNDIT Works



- Uses a lightweight process on each perfSONAR agent for detection
- Uses a central server for problem event repository and for localization algorithm

Detection

- Find significant deviations from the baseline for loss, latency and reordering



Localization

- Uses Range (latency) and Boolean Tomography (loss, reordering)

Example of Range Tomography:



- Measured lossy links:
 - 15% loss between (1,5)
 - 5% between (2,5)
 - 7% between (3,5)
- Plausible solution:
 - Link (4,5) has loss rate [5%-7%]
 - Link (1,4) has loss rate [8%-10%]

Current Progress

- New project that started in September 2014
- Set up the PuNDIT testbed with perfSONAR nodes at seven sites spanning the country
 - Allows us to test our prototype versions in realistic conditions
- PuNDIT testbed participants:
 - Logos of participating institutions: UIowa, Michigan, State of Ohio, ESnet, and others.
- To document and describe our project, we have setup a PuNDIT website at: <http://pundit.gatech.edu/>
- PuNDIT code and documentation is hosted on GitHub: <https://github.com/pundit-project>



- Development is underway, managed using OpenProject
- PuNDIT prototyping infrastructure using VMware:
 - Two test VMs running perfSONAR 3.4 for agent development
- Prototype central PuNDIT server instantiated as a VM
 - Gathers PuNDIT agent data from our deployments
 - Used to estimate the required hardware profile needed for a future PuNDIT production server

Closing remarks

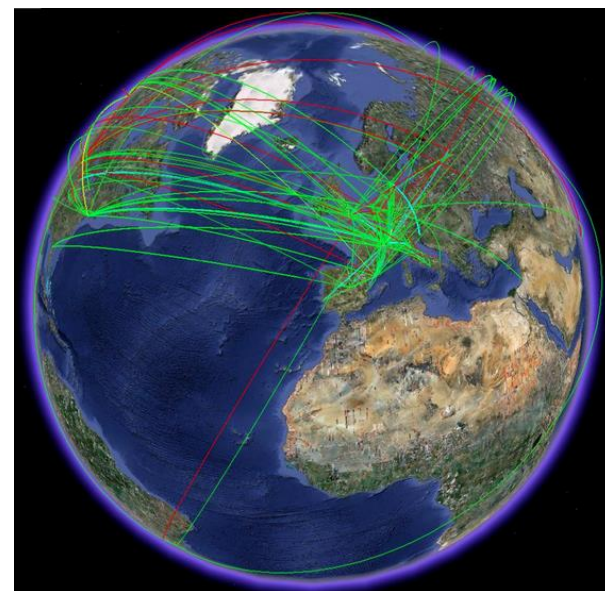
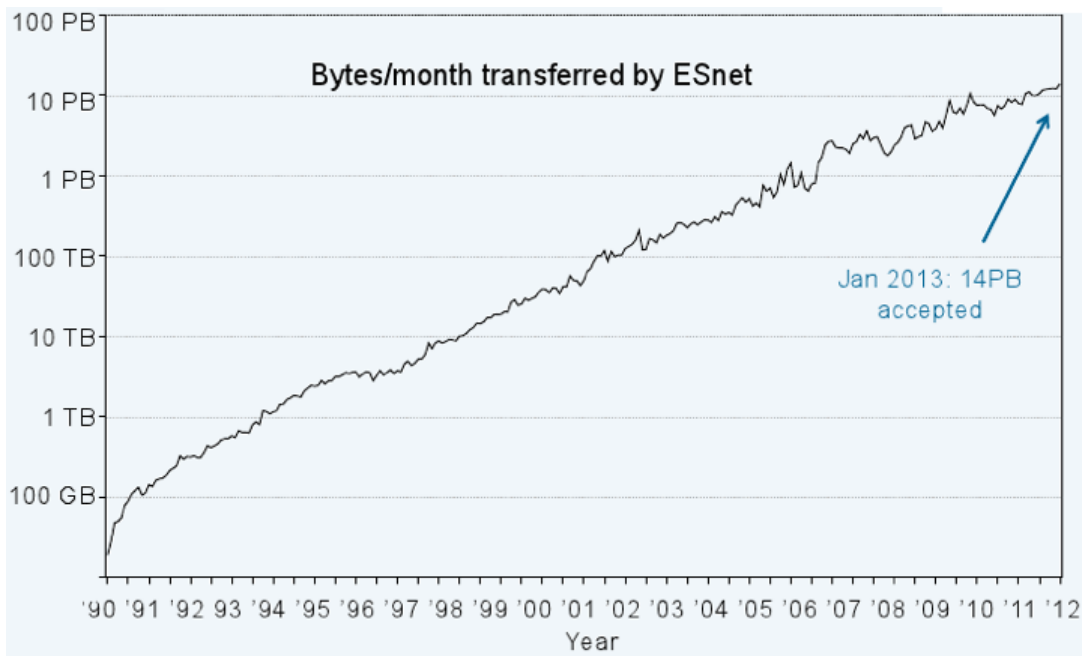
- perfSONAR widely deployed and already showing benefits in troubleshooting network issues
 - Additional deployments by R&E networks still needed
- Significant progress in configuration and infrastructure monitoring
 - Helping to reach full potential of the perfSONAR deployment
- OSG datastore – community network data store for all perfSONAR metrics – planned to enter production in Q3
- Integration projects aiming to aggregate network and transfer metrics
 - FTS Performance
 - Experiment's interface to perfSONAR
- Advanced network monitoring - diagnosis and alerts based on perfSONAR, developed within NSF funded PuNDIT project



Backup slides

Introduction

- OSG/WLCG critically depend upon the network
 - Interconnect sites and resources
- Traffic grows at a rate of factor 10 every 4 years
- Progressively moving from tier-based to peer to peer model
- Emergence of new paradigms and network technologies

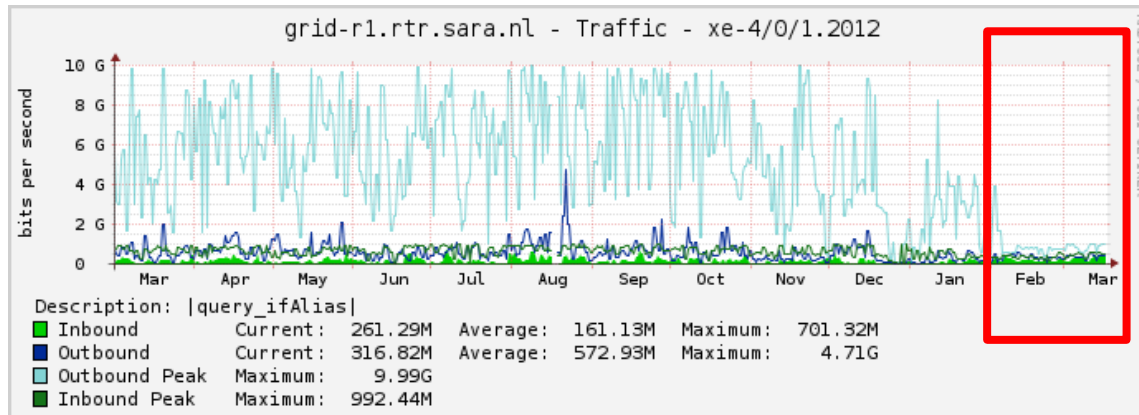


Network Troubleshooting

- End-to-end network issues difficult to spot and localize
 - Network problems are multi-domain, complicating the process
 - Standardizing on specific tools and methods allows groups to focus resources more effectively and better self-support
 - Performance issues involving the network are complicated by the number of components involved end-to-end.
- Famous “BNL-CNAF” network issue
 - 7 months, 72 entries in the tracking ticket, requiring work from many people

perfSONAR in Troubleshooting

- Recent problem reported SARA->AGLT2. FTS timed out because rate for large files < 2-300 Kbytes/sec
- perfSONAR tests confirmed similar results
- Opened ticket with I2, who opened ticket with GEANT
 - GEANT brought up LHCONe pS instance. Tests to AGLT2 showed 3 times BW vs (much closer) SARA
 - Problematic link identified within few days (0.2-0.5% packet loss)
- Currently establishing procedure to report on network performance issues in WLCG



Impacts overall link throughput.
Many fixes tried. No solution yet.

Where should we track/document cases?

Commissioning Challenges

- Installing a service at every site is one thing, but commissioning a NxN system of links is squared the effort.
 - This is why we have perfSONAR-PS installed but not all links are monitored.
- perfSONAR is a “special” service
 - Dedicated hardware and management is required
 - We understand this creates complications to some fabric infrastructure. Sharing your experience within WG might be the best way to help other sites
- 3.4 was a major release that introduce many new features and components
 - Some of them required follow up and bug fixing
 - 3.4.2rc is currently validated in a PS testbed by WLCG/OSG – monitored by the same monitoring infrastructure as production
- Security incidents had major impact on perfSONAR
 - wlcg-perfsonar-security mailing list was established to be a single point of contact in case of security incidents
 - New documentation was written to give guidelines to sites
- We still have issues with firewalls
 - New documentation very clear on port opening and campus/central firewall
 - Monitoring information now exposed by almost all sites – also thanks to migrating infrastructure monitoring to OSG

Infrastructure Monitoring

- We have 3 versions of perfSONAR Infrastructure monitoring
 - Prototype at maddash.aglt2.org
 - Testing at OSG's ITB instance
 - Production at OSG's production instance
- Main monitoring types are MaDDash and OMD/Check_MK
 - Prototype: <http://maddash.aglt2.org/maddash-webui>
https://maddash.aglt2.org/WLCGperfSONARcheck_mk
 - Testing: <http://perfsonar-itb.grid.iu.edu/>
 - Production: <http://psmad.grid.iu.edu>
<http://psomd.grid.iu.edu>
- Notes:
 - OSG instances rely upon OSG Datastore: <http://psds.grid.iu.edu>
 - X509 cert needed to view check_mk/OMD pages (any IGTF cert)
- Plan:
 - Develop additional plugins to check core functionality – IPv6, memory, etc.

FTS performance

- FTS - low level data movement service – used by majority of WLCG transfers
- Current granularity and coverage is a good match to perfSONAR network
 - Mapping SEs to sonars needed
- Goal: Adding latency and routing (tracepath) to the optimizer algorithm
 - Better tune number of active transfers
- Integration of traceroutes and FTS monitoring already attempted in the ATLAS FTS performance study (lead by Saul Youssef - <http://egg.bu.edu/atlas/adc/fts/plots/>)
 - Integrates FTS monitoring and tracepath to determine weak channels and identify problems
 - Proposed to extend it to CMS and LHCb