Networking Topics for Science: Activities and Plans

May 2, 2022

IRIS-HEP Topical Session: What's going on in Networking?

Shawn McKee / University of Michigan OSG-LHC

https://indico.cern.ch/event/1138837/

OSG Networking Introduction

To start, it is important to provide a quick overview of what this area does:

Primarily network monitoring to support distributed science across OSG and beyond

- We have developed and operate a global network metrics data pipeline, gathering **perfSONAR** and other metrics
- We provide tools and user support to find and localize network problems.
- We track future networking technologies and plan for possible use cases supporting OSG and WLCG

Overview / Review

- OSG has been working on networking for its constituents and collaborators for more than 9 years
  - We have a complete infrastructure to reliably measure, gather and store important network metrics
- We have acquired a **rich dataset** including **perfSONAR** and **HTCondor and Xrootd** transfer metrics, SNMP data from ESnet and FTS transfer results
  - **Beneficially using this data is our primary focus**
- In this presentation I will cover the current activities, related collaborations and our near-term plans for the OSG networking area
- **Petya** will then cover some of our work to extract value from our data and next steps to utilize ML to help in this task.

Current Network Status

Regarding our networking efforts, I want to state a few things about our current status up front

- Our networks continue to perform very well for our community
- Most users are happy with the networking we have but see areas for improvement.
- Primary concerns exist around our ability to fully utilize **existing** networks
- Visibility is key to understanding, maintaining and fixing our networks

So there continues to be **near-term** work regarding our networking in optimizing, monitoring and fixing network problems, **but** we also must think longer term regarding how the situation may evolve and what that might mean for us.

Laundry List of Network-related Activities

Snowmass CompF4 has dedicated network section

RNTWG (Research Networking Technical Working Group)

Packet marking / flow labeling, packet pacing, orchestration

Network Data Pipeline gathers network metrics and transfer data

Network user-facing tools (more on this later)

Network analytics is beginning to explore our data to identify & localize issues

perfSONAR v5 the next version of perfSONAR is due out this month

WLCG DOMA (Data Organization, Management and Access) includes networking

- **WLCG DOMA BDT** Bulk Data Transfer working group has network component

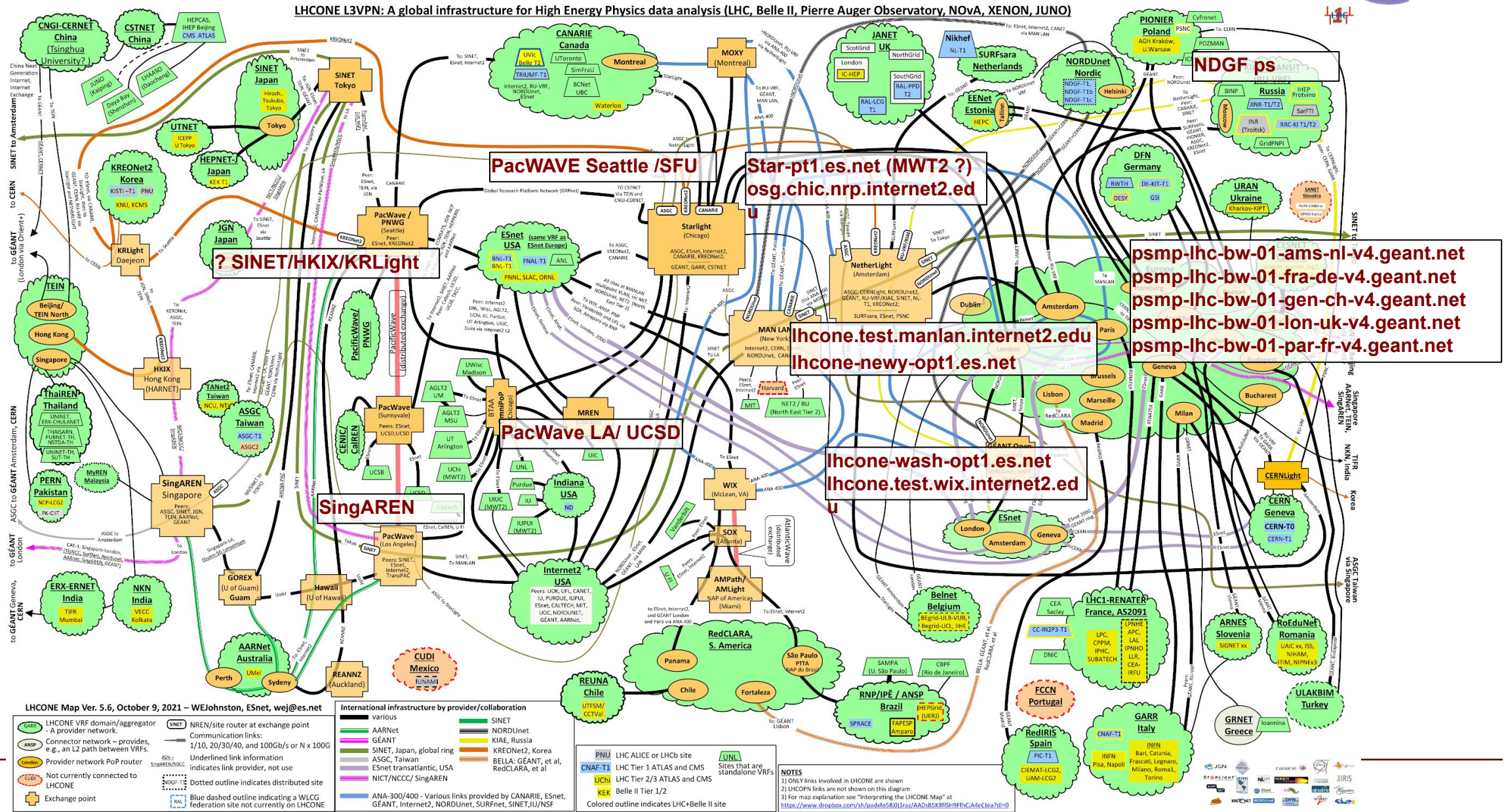
WLCG Monitoring Task Force has activities to make site network more visible.

Data Challenges to prepare for HL-LHC have network data challenges ~2 years

Importance of Measuring & Monitoring Our Networks

- End-to-end network issues are difficult to spot and localize
  - Network problems are **multi-domain**, complicating the process
  - Performance issues involving the network are complicated by the number of components involved end-to-end
  - Standardizing on specific tools and methods focuses resources more effectively and provides better self-support.
- **Network problems can severely impact experiment's workflows and have taken weeks, months and even years to get addressed!**
- **perfSONAR** provides a number of standard metrics we can use
  - Latency, Bandwidth and Traceroute
  - These measurements are **critical for network visibility**
- **Without measuring our complex, global networks we wouldn't be able to reliably use those network to do science**

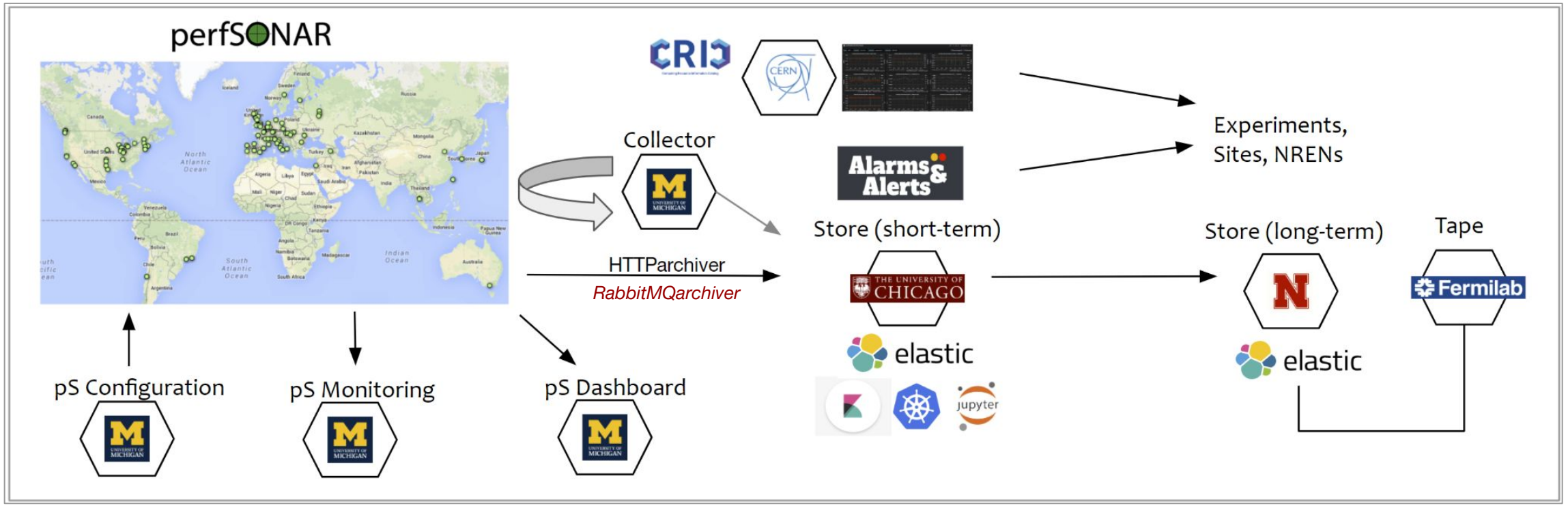
LHCONE Map and perfSONAR Toolkits



The OSG/WLCG Network Data Pipeline

The focus in OSG Network Monitoring for many years has been the deployment and configuration of our global **perfSONAR** monitoring framework

- The **deployment** and **operations** of **perfSONAR** has been stable
- The network pipeline has been hardened and is operating reliably.



Network Data Volume (2018-2021)

Net Pipeline	Data Type	Total Tests	Tests/day	Storage Size
perfSONAR	Latency	6.91B	7.95M	3.1TB
	Packet Loss	7.00B	8.08M	2.4TB
	Retransmits	14.7M	18.8k	6.3GB
	Throughput	15.6M	19.2k	7.0GB
	Network Path	1.28B	2.14M	1.5TB
ESnet	Traffic	1.1B	44.7M	1.74TB
	Interfaces	3.2M	11.8k	530MB
HTCondor	Job Transfers	734M	446k	610GB
Total		**17.1B**	**64.4M**	**9.4TB**

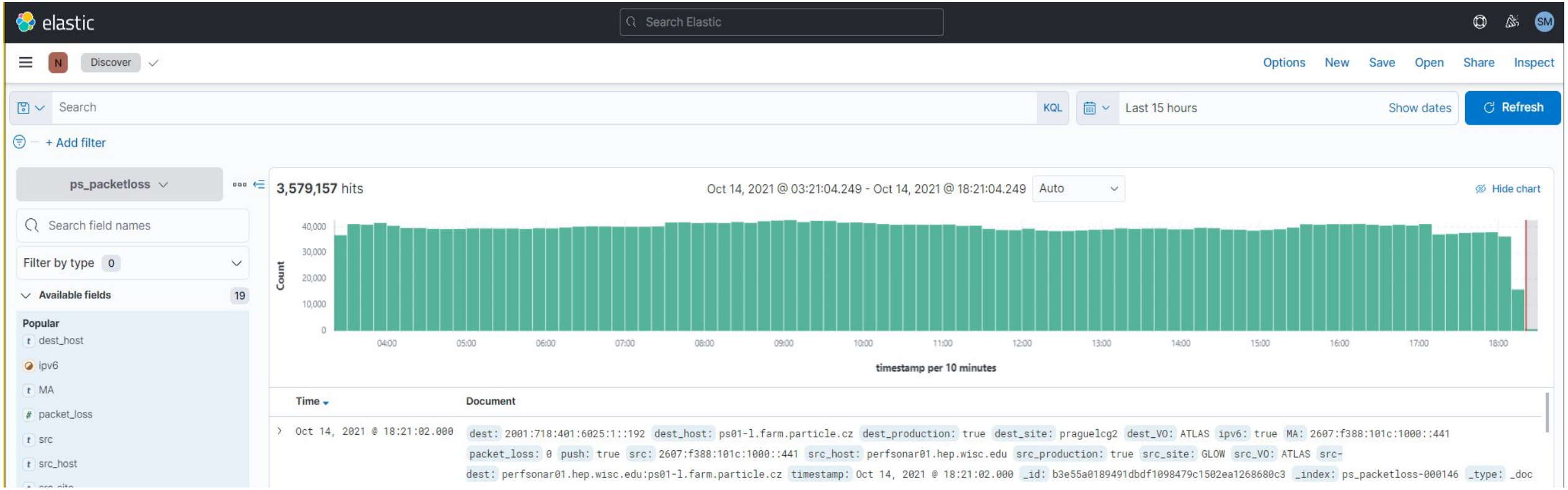
User Interfaces to Data

Net Data Access and Visualization

In the next few slides we show some examples of how the data is accessed and visualized.

For **exploring the data**, perhaps the easiest means is by using our Kibana interface:

https://atlas-kibana.mwt2.org/s/networking/goto/3c4e90deab15fd901289035a9227de62?auth_provider_hint=anonymous1

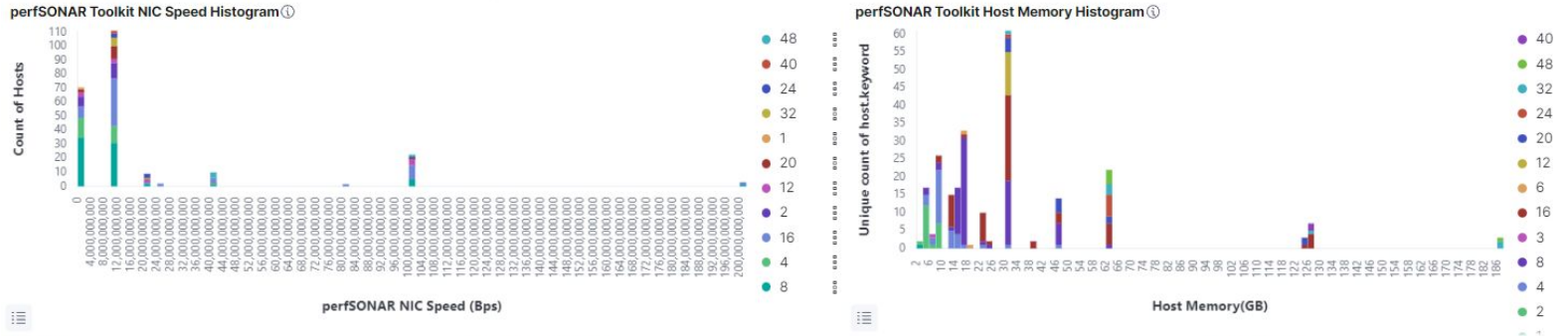


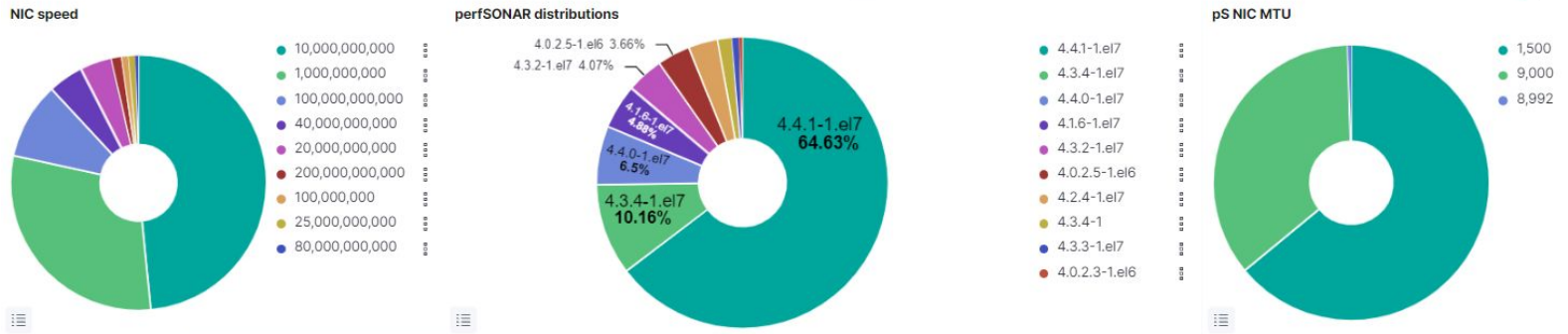
Infrastructure Dashboard

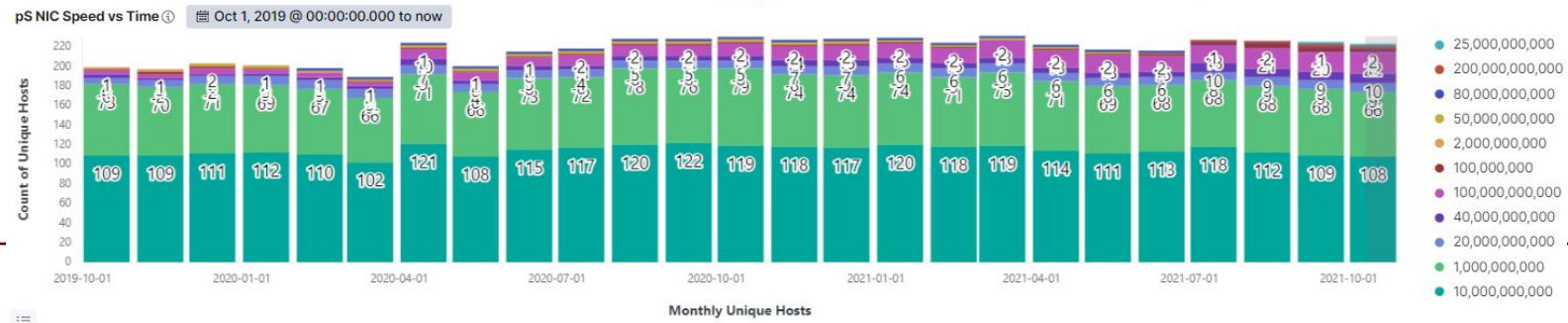
240 Active perfSONAR instances - **207 production endpoints** - T1/T2 coverage

- Continuously testing over 5000 links - testing coordinated and managed from central place

- Dedicated latency and bandwidth nodes at each site - **Open platform (testing and data)**







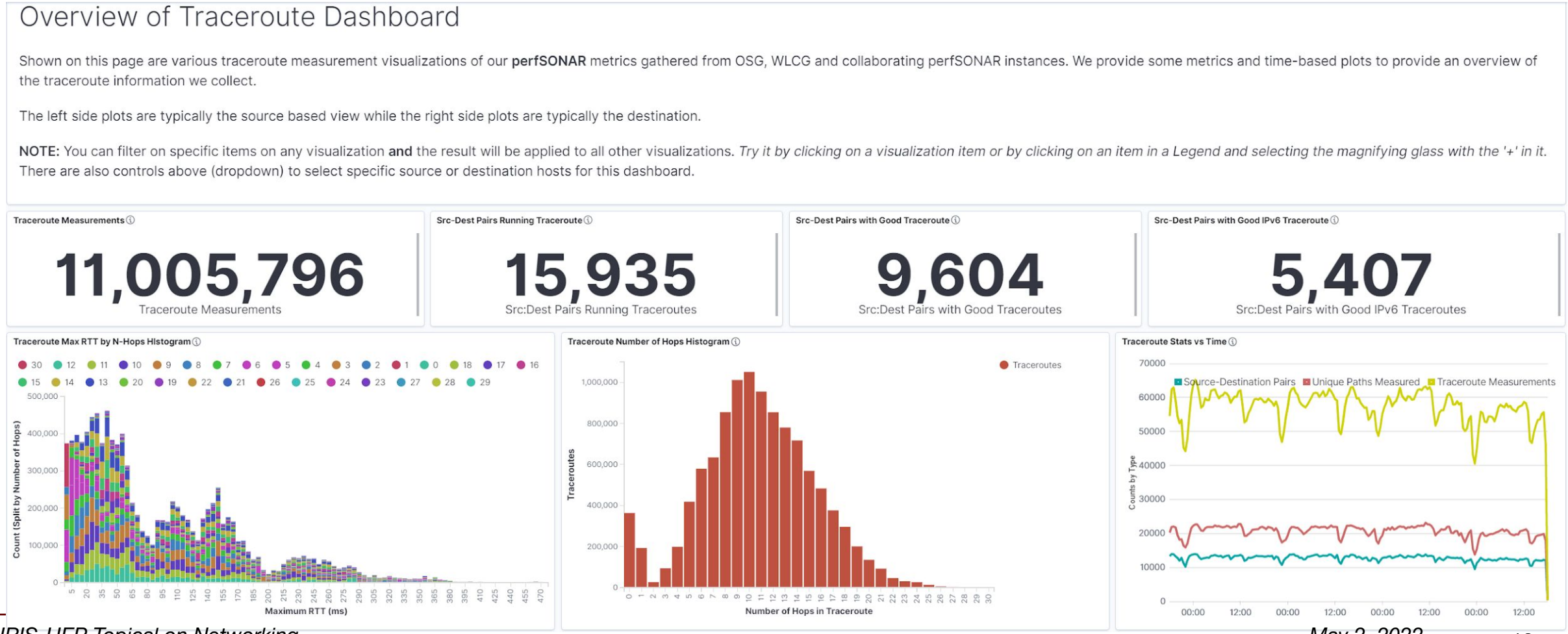
Our global toolkit deployment has a **range** of systems in terms of age and capability.

Dashboard in ELK

Sites need to remember to not **only** upgrade perfSONAR software but also the underlying **hardware,** as nodes become too old or are unable to test at the site storage speed.

Kibana Dashboard for Traceroute Data

For specific data we have various Kibana dashboards. Below is one for **traceroute data** we process through our infrastructure. Similar dashboards exist for packet-loss, throughput, latency.



Research Networking Technical WG

Research Networking Technical WG

- HEPiX Network Functions Virtualisation Working Group
  - Working Group Report was published at the end of 2019 with three chapters
    - Cloud Native DC Networking
    - Programmable Wide Area Networks
    - Proposed Areas of Future Work
- LHCOPN/LHCONE workshop (spring 2020)
  - Requirements on networks from the WLCG experiments
- Research Networking Technical Working Group
  - Formed after the workshop in response to the requirements discussion
  - 98 members from ~ 50 organisations have joined
- The RNTWG is working on three areas:
  - **Network visibility** via Packet Marking / Flow Labeling
  - **Network usage optimization** via Packet Pacing / Traffic Shaping
  - **Network management** via Network Orchestration / GNA-G DIS / SENSE / NOTED

Network Visibility and Scitags

Scientific Network Tags (scitags) is an initiative promoting identification of the science domains and their high-level activities at the network level.

WHAT: Enable tracking and correlation of science data transfers with Research and Education Network Providers (R&Es) network flow monitoring

HOW: By either **marking packets** or **labeling flows**

WHY:

1. Experiments can better understand how network flows perform along the path
   a. Improve visibility into how network flows perform (per activity) within R&E segments
   b. Get insights into how experiment is using the networks, get additional data from R&Es on behaviour of our transfers (traffic, paths, etc.)
2. Sites can get visibility into how different network flows perform
   a. Network monitoring per flow (with experiment/activity information)
      i. E.g. RTT, retransmits, segment size, congestion window, etc. all per flow

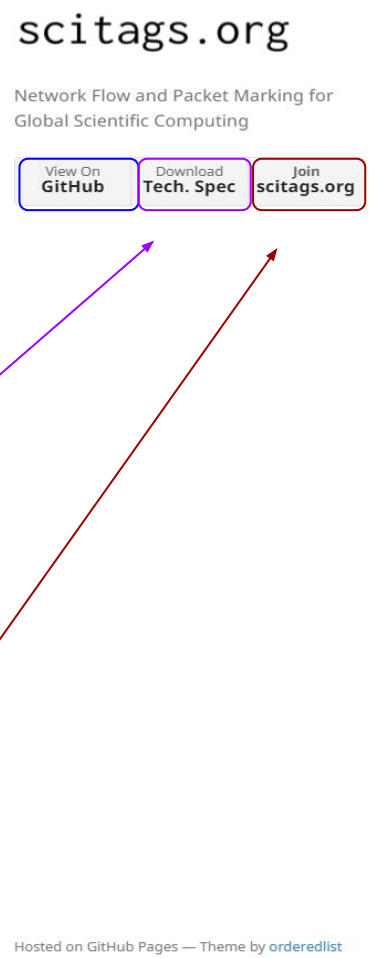
Finding More Information: https://scitags.org

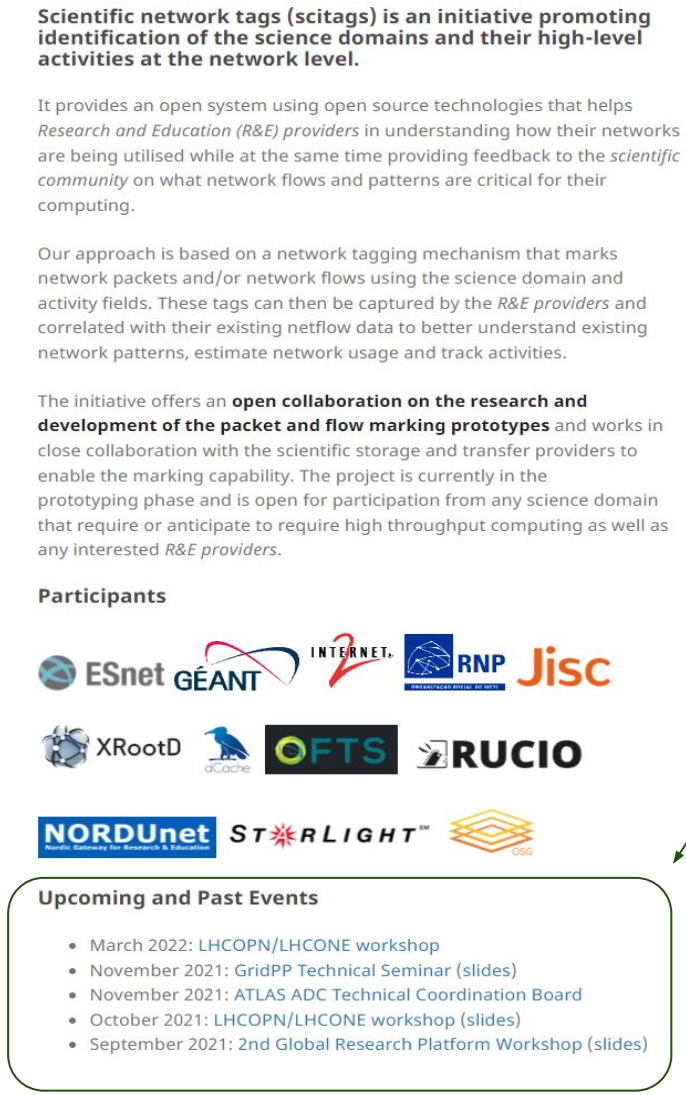
Code

Tech Specs

Mailing List

Presentations





Pacing/Shaping WAN data flows

A challenge for HEP storage endpoints is to utilize the network efficiently and fully.

- An area of interest for the experiments is **traffic shaping/pacing**.
  - Without traffic pacing, network packets are emitted by the network interface in bursts, corresponding to the wire speed of the interface.
    - **Problem**: microbursts of packets can cause buffer overflows
    - The impact on TCP throughput, especially for high-bandwidth transfers on long network paths can be **significant**.
- Instead, pacing flows to match expectations [min(SRC,DEST,NET)] smooths flows and significantly reduces the microburst problem.
  - An important extra benefit is that these smooth flows are much friendlier to other users of the network by not bursting and causing buffer overflows.
  - Broad implementation of pacing could make it feasible to run networks at much higher occupancy before requiring additional bandwidth

This work has yet to have much effort; we plan to begin work during this summer!

Network orchestration

- OpenStack and Kubernetes are being leveraged to create very dynamic infrastructures to meet a range of needs.
  - Critical for these technologies is a level of **automation** for the required networking using both software defined networking and network function virtualization.
  - For HL-LHC, it is important to find tools, technologies and improved workflows that may help bridge the anticipated gap between the resources we can afford and what will actually be required
- The ways we organize our computing / storage resources will need to evolve.
- This area is being led by the **GNA-G** (Global Network Advancement Group; https://www.gna-g.net/) and is exploring many options for traffic engineering, resource management and network-application interfaces.
  - The **SENSE** project is serving as a reference implementation
- The NOTED project is also an example of a practical way to effectively utilize available paths to better distribute network load.

WLCG Data Challenges

WLCG Data Challenges

The **WLCG** has significant and growing requirements for its infrastructure.

- The biggest challenge is coming in **Run-4**, the High Luminosity LHC run, which will require more resources than WLCG is forecasted to have.

To bridge the gap, WLCG (and IRIS-HEP) are exploring new ways to organize and analyze data, as well as new tools/technologies to support data access/movement.

WLCG has planned regular **Network Data Challenges** every two years

2021 - **10%** of HL-LHC, **2023** - **30%** of HL-LHC, **2025** - **70%** of HL-LHC and finally **2027** - **100%** of HL-LHC required end-to-end rates.

Site specific monitoring: one of the **main deficiencies** identified in the first WLCG Network Data Challenge was missing site specific network traffic data

- The WLCG Monitoring TF is working to provide a template site network description

Most important part is for site's to provide network monitoring JSON data (IN/OUT traffic)

perfSONAR Support for Future Data Challenges

We plan to continue to use **perfSONAR** and associated tools and analytics to help us evolve our infrastructure to prepare for HL-LHC.

IRIS-HEP/OSG-LHC has an important role to play here!

perfSONAR allows us to define and track the current baseline between our sites and identify bottlenecks to focus on improving.

The Network Data Challenges are really end-to-end and involve all our infrastructure components including software, technologies, hardware and architecture.

Outreach and Collaboration Activities

We have collaborated with WLCG on data challenges to prepare the infrastructure to monitor/measure the network.

There have been presentations on networking at LHCONE/LHCOPN, HEPiX and for the IPv6 working group.

The RNTWG (Research Networking Technical WG) has meet many times, focused on implementing **packet/flow marking**.

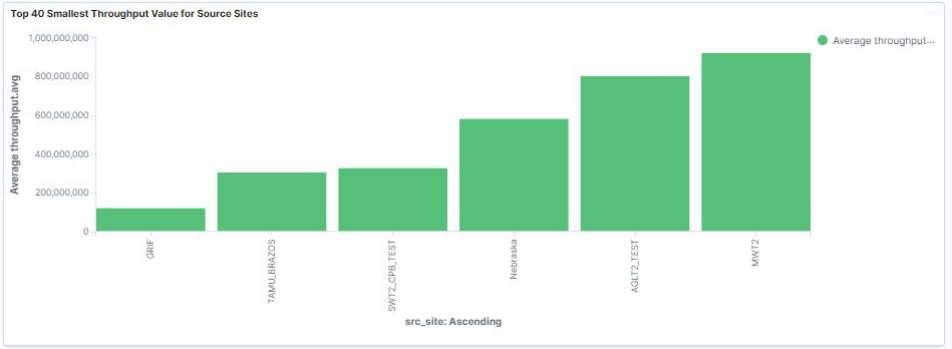
We finished up with the **SAND** project on Analytics (HEPiX presentation https://indico.cern.ch/event/810635/contributions/3593348/) in July 2021

Students: We have one undergraduate (Tommy) and one graduate student (Petya) looking at the OSG/WLCG network metrics (weekly meetings on Fridays 10-11 AM Eastern)

- Petya will present after this, next slide covers Tommy's work

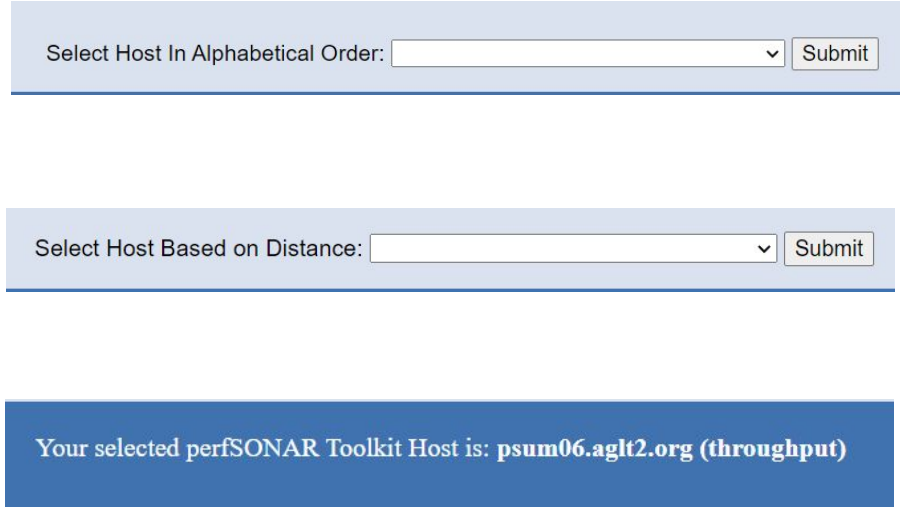
Thomas (Tommy) Shearer

IRIS-HEP Fellow Working on OSG Network Data
Fellowship dates: May-August 2020
Home Institution: University of Michigan, Ann Arbor
Undergraduate (Senior) UM



Improving the User Interface to OSG-LHC Network Metrics

The OSG-LHC Network Monitoring collects and manages many different types of metrics related to network performance. My work has created a system of user interfaces that are organized by data type, and filterable by the parameters of interest. Additionally, my work on the OSG Toolkit Information Page (http://toolkitinfo.opensciencegrid.org) has allowed users to easily maneuver these interfaces dynamically with queried toolkit hosts, organization by name or distance, and dynamical linking to user interfaces. Furthermore, I have worked on a new web-app that provides customized self-subscription for Alarms and Alerts within the OSG-LHC Network (https://aaas.atlas-ml.org/)



New Applications and Analytics

Motivation and History

As we have constructed our **perfSONAR** monitoring infrastructure we have gotten significant feedback from our users.

One of the most frequent requests is a mechanism to alert users when network issues identified by perfSONAR and involving their infrastructure happen.

- This is more challenging than you might think
- We want to alert users to issues **they** can address
- Need to ensure it is really a **network problem** and have it **localized**

One of the most important components is a new Alerting and Alarming Service which tries to address this need. **Petya** will cover this in her presentation.

Next Steps

Future Work Areas

- More work on network topology - cleaning, analyzing and visualizing.
- Exploring the value of network flow labeling and packet marking (with RNTWG)
  - Add in new perfSONAR tests for flow-labels and/or packet marking
- Creating / enabling **site specific monitoring** (from CRIC URLs)
  - Data will need to be "harvested" and integrated into our pipeline to augment perfSONAR
- Improving the latency and reliability of our pipeline by continuing the transition from a primarily "pull" data model (collectors) to a secure "push" model.
- Broadening the engagement with the Global research community
  - Access to R&E monitoring and integration with our tools and datastores
- We responded to the new *Internet Measurement Research: Methodologies, Tools, and Infrastructure (IMR)* solicitation with a proposals (ML on net data)

Others? Please suggest additional areas or ideas :)

Summary and Conclusion

The **collaboration of OSG, WLCG and various research projects** have created an extensive, reliable infrastructure to monitor our networks via **perfSONAR** and provide associated analytics and visualization.

While making progress in evolving our tools and infrastructure, we need to continue to monitor and maintain what we have built.

There are a number of challenges remaining, including the difficult meta-challenge of making effective insights available for everyone to easily use.

Questions or Comments?

Useful URLs to Reference

- IRIS-HEP OSG Networking project https://iris-hep.org/projects/osg-networking.html
- Current OSG network documentation
  - https://osg-htc.org/networking/
- pS Toolkit Info Page: https://toolkitinfo.opensciencegrid.org/
- See recent presentations at the LHCONE/LHCOPN meeting, the HEPiX Meeting or the WLCG DOMA BDT meeting presentation.
- The psetf check_mk pS monitoring page: https://psetf.opensciencegrid.org/etf/check_mk/index.py?start_url=%2Fetf%2Fcheck_mk%2Fdashboard.py
- SAND: http://sand-ci.org

Backup Slides

Videos Available

We created a set of videos that describe the perfSONAR Toolkitinfo app, the Alerting and Alarming app and the ps-dash app.

They are available on the SAND web page: https://sand-ci.org/

Let us know if you have any feedback on these!

Update on the 2021 WLCG Data Challenge

First WLCG data challenge was Oct 04-08, 2021

Goal was 240 Gbps from T0 to T1's and from T1's to T2's involving primarily ATLAS and CMS

The network was NOT the bottleneck in general!