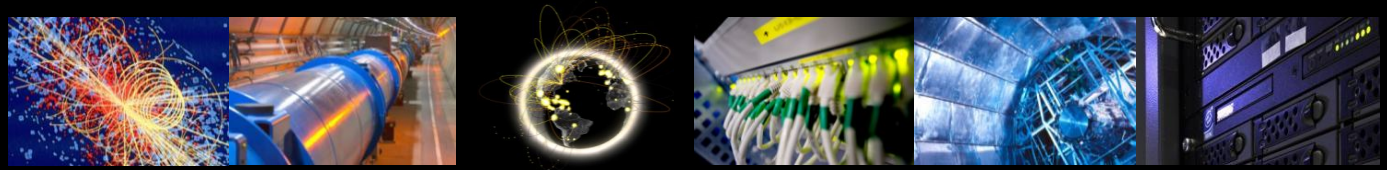
LHC[ONE|OPN] Network Monitoring Update

Shawn McKee / U Michigan, Marian Babik / CERN
on behalf of WLCG Network Throughput WG

October 24, 2022

https://indico.cern.ch/event/1146558/



WLCG Worldwide LHC Computing Grid

Open Science Grid

Outline

- OSG/WLCG Network Monitoring and WLCG Network Throughput WG
- perfSONAR community updates
- LHCOPN/LHCONE perfSONAR infrastructure status
- New Analytics and Tools
- Summary

perfSONAR Deployment Summary

223 Active perfSONAR instances

- **207 production endpoints**
- T1/T2 coverage
- Continuously testing over 5000 links
- Dedicated latency and bandwidth nodes at each site
- Testing coordinated and managed from central place
- LHC experiments, DUNE, BelleII, LSST
- LHCOPN/LHCONE, ARCHIVER, StashCache, SLATE, CC*

perfSONAR News

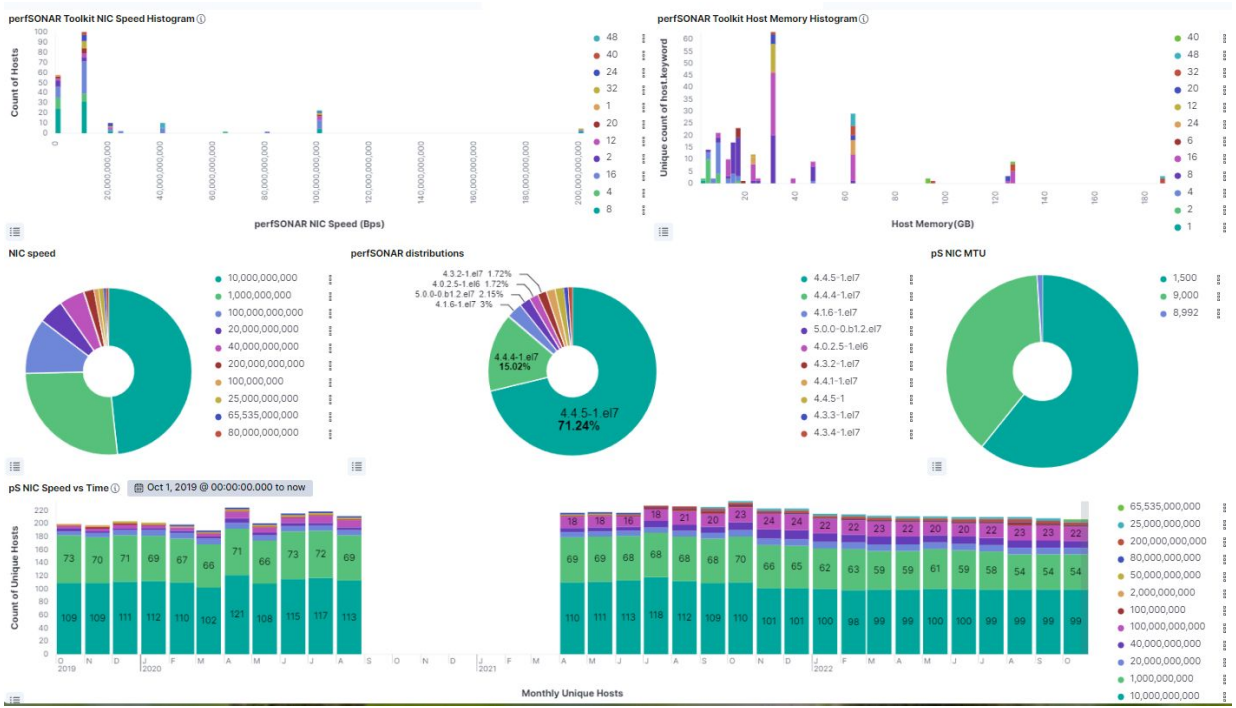
- perfSONAR 5 beta has been out since early summer
  - ElasticSearch as local archive (replacing esmond/Cassandra) + Logstash.
  - Grafana visualisations (dashboards).
  - Toolkit support for latest Debian, RHEL8 compatible systems (Alma).
    - CS8 not officially supported but will likely work fine.
    - Will require full reinstall (backup **not** needed).
- We still see various issues that impact our WLCG-wide deployment, even in the current beta of perfSONAR 5.
  - There are problems that seem to arise due to the scale of our deployment and the number of tests we are running.
  - Issues we have identified are open with the perfSONAR team.
  - Especially important for us is the ability to reliably configure the use of the HTTP archiver so we can directly send data from Toolkits to Elasticsearch.

perfSONAR deployment

223 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 should 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.

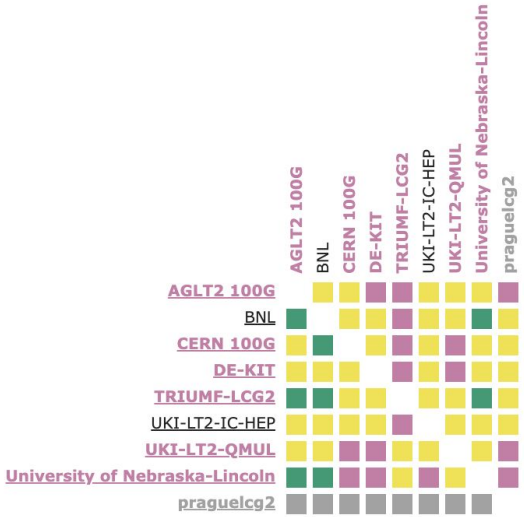
100Gbps Testing 28th March 2022

- LHCOPN/LHCONE 100Gbps mesh
  Mesh changes: QMUL replaced pic; thresholds updated

WLCG 100G Mesh - WLCG 100G IPv4 Bandwidth - Throughput

Throughput >= 10Gbps | Throughput < 10Gbps | Throughput <= 1Gbps | Unable to fir

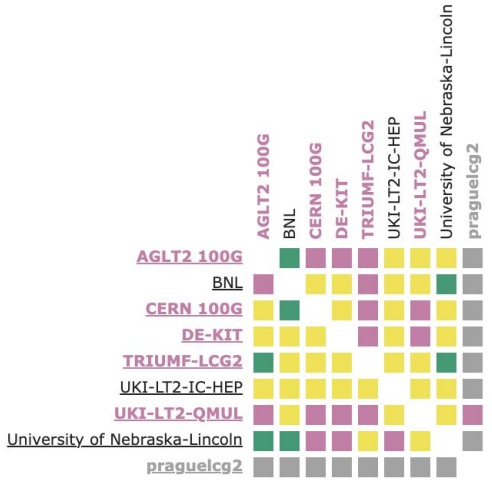
Found a total of 9 problems involving 7 hosts in the grid



WLCG 100G Mesh - WLCG 100G IPv6 Bandwidth - Throughput

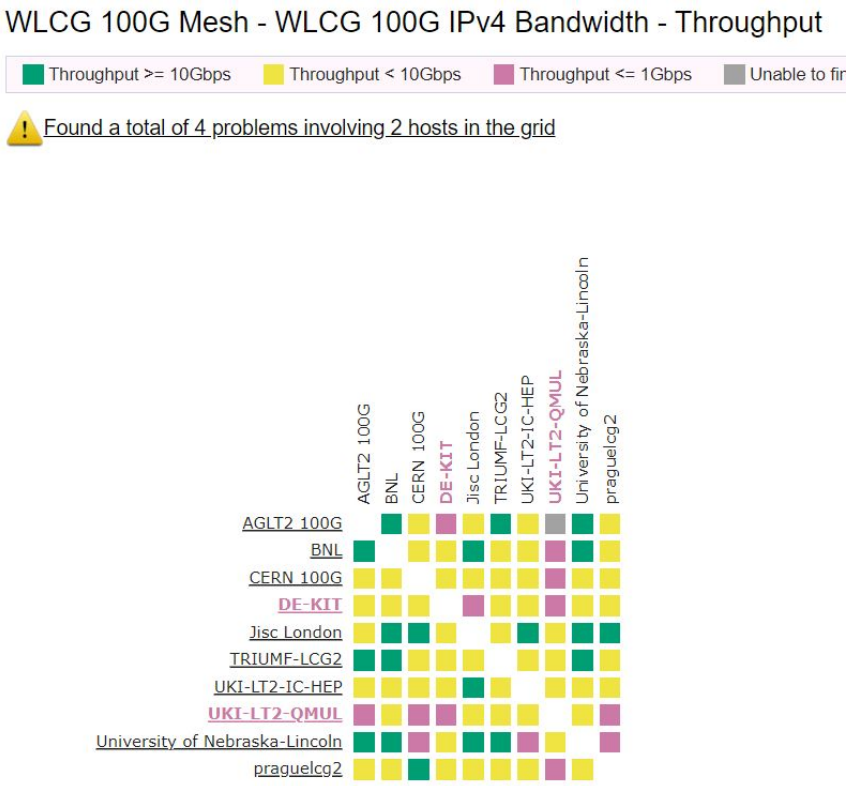
Throughput >= 10Gbps | Throughput < 10Gbps | Throughput <= 1Gbps | Unable to fi

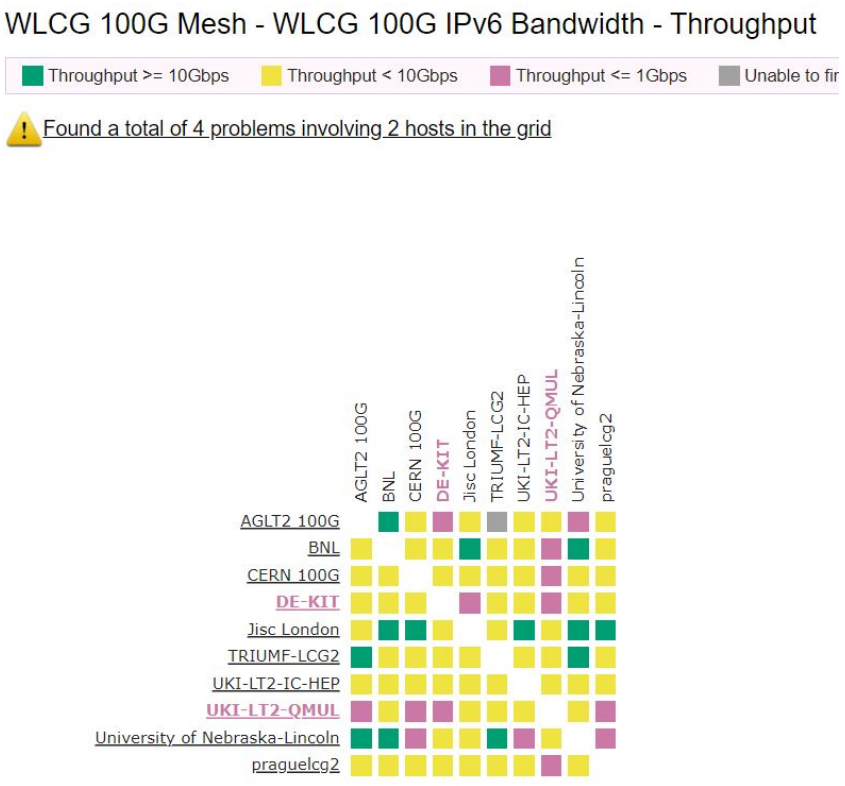
Found a total of 8 problems involving 6 hosts in the grid



100Gbps Testing 21st October 2022

- LHCOPN/LHCONE 100Gbps mesh: all sites responding but not great results





100Gbps Testing

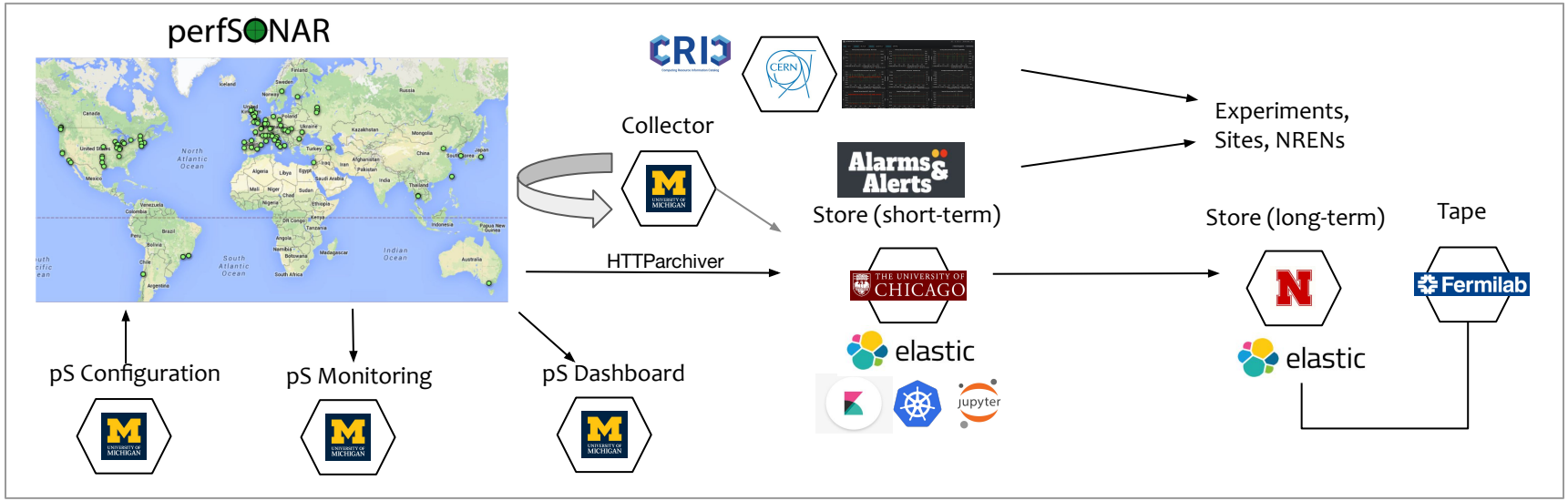
- It has been a while since our last meeting; trying to schedule ~now
  - Aim to achieve 10% of avail. capacity (~10Gbps) on a regular basis
  - Discussing ways to tune the nodes and improve stability
  - wlcg-perfsonar-100g mailing list (join)
- Various issues found and fixed at different sites
  - Since last time all sites are at least working and making measurements
- Tunings
  - Used psetf along with ES/Kibana dashboards to check status
  - TCP buffers and MTU appear to have made the biggest difference
    - TCP buffers by default at ~ 200MB, need to be increased to 1GB
  - References:
    - https://fasterdata.es.net/host-tuning/linux/100g-tuning/
  - Tried FQ but that actually decreased the throughput in tests (not work-conserving)
  - NIC interrupts/core sync only possible via manual tests
- Can check our host-based Grafana dashboard to see data graphs

Infrastructure: Network Data Pipeline Transition

- As you may recall from last time, we are updating our network data pipeline.
- This is primarily motivated by two things
  - The operational experience with using RabbitMQ (high latency; too many parts)
  - The upcoming perfSONAR v5 change to the “default” Measurement Archive (MA), the location that hosts the measurement results
- We are in the process of transitioning away from using a message bus and instead sending data directly to Elasticsearch from each perfSONAR toolkit.
  - **Pros**: Reduces latency, simplifies pipeline with less components required, matches direction the perfSONAR developer have chosen allowing us to utilize and benefit from their work, should be easier to operate and monitor
  - **Cons**: All dependencies on the RabbitMQ bus must be transformed to use the HTTP archiver. To maintain the ability to grab data for those toolkits that cannot “push” also requires updating the perfsonar_collector to send data directly to Elasticsearch.

Network Measurement Platform Evolution

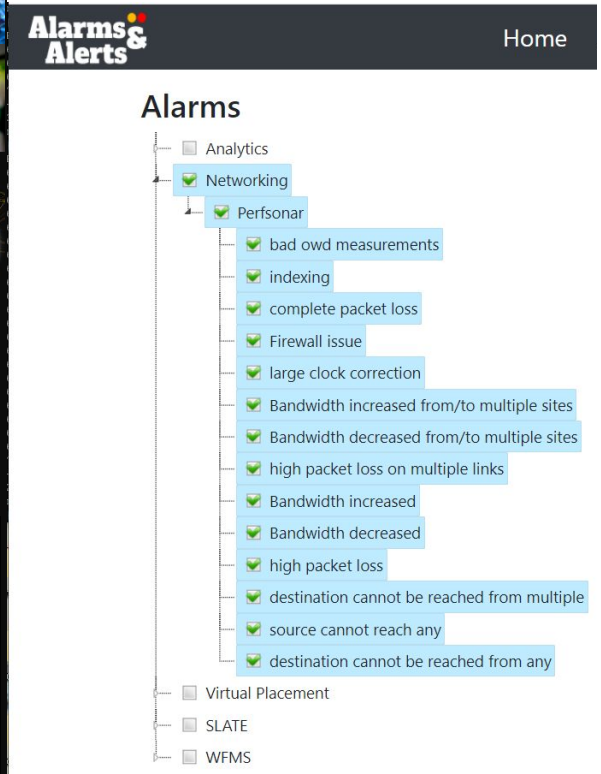
- Our platform collects, stores, configures and transports all network metrics
  - Distributed deployment - operated in collaboration
- Planned evolution based on the perfSONAR 5 already partially implemented.
  - Directly publishing results from perfSONARs to ES@UC (have test_* indices)
  - High-level services provided to the experiments/users (Alerting, Dashboards)



Tools and Applications for Network Data

- To organize access to all the various resources we recommend using our Toolkitinfo page: https://toolkitinfo.opensciencegrid.org/
- Reminder: we already have Kibana dashboards looking at
  - Bandwidth
  - Traceroute
  - Packetloss / Latency
  - Infrastructure
- For this meeting we want to present our recent work towards a user subscribable alerting and alarming service
  - User interface to subscribe is **AAAS** (ATLAS Alerting and Alarming Service)
  - Tool to explore alerts is **pS-Dash** (Plotly base perfSONAR dashboard UI tool)

The Alerting and Alarming Tools



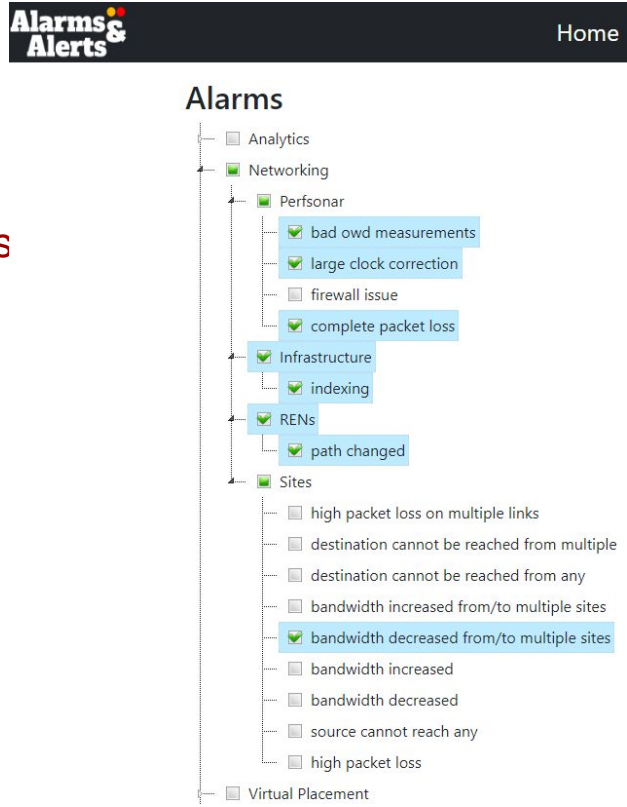
https://aaas.atlas-ml.org/

(Uses EDUGain/InCommon)

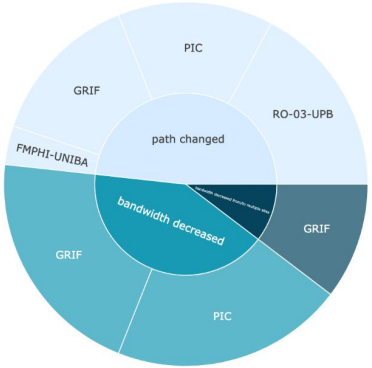
Purpose: provides user-subscribable alerting for specific types of network issues found by analyzing perfSONAR data

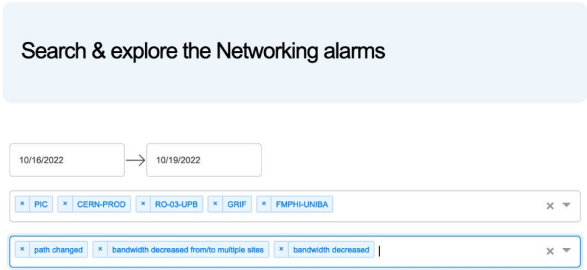
Updates and Activities

- Working to add "acknowledgements" for known issues.
- Have reorganized Alarms
  - Old **left**, new **right**
- Continuing tuning based upon use-cases (site admin, network admin, user)



Updates on pS-Dash (1/2)





List of alarms

BANDWIDTH DECREASED

src_site	dest_site	ipv	ipv6	last3days_avg	%change	from	to	alarm_id	tag_str
filter data...									
JINR-LCG2	GRIF	ipv6	true	558	-53	2022-09-25 04:08	2022-10-16 04:08	VIEW	JINR-LCG2, GRIF
RRC-KI-T1	pic	ipv4	false	99	-80	2022-09-25 04:08	2022-10-16 04:08	VIEW	RRC-KI-T1, PIC
pic	INFN-T1	ipv4	false	320	-92	2022-09-25 04:08	2022-10-16 04:08	VIEW	PIC, INFN-T1
FZK-LCG2	GRIF	ipv4	false	1516	-61	2022-09-25 04:08	2022-10-16 04:08	VIEW	FZK-LCG2, GRIF

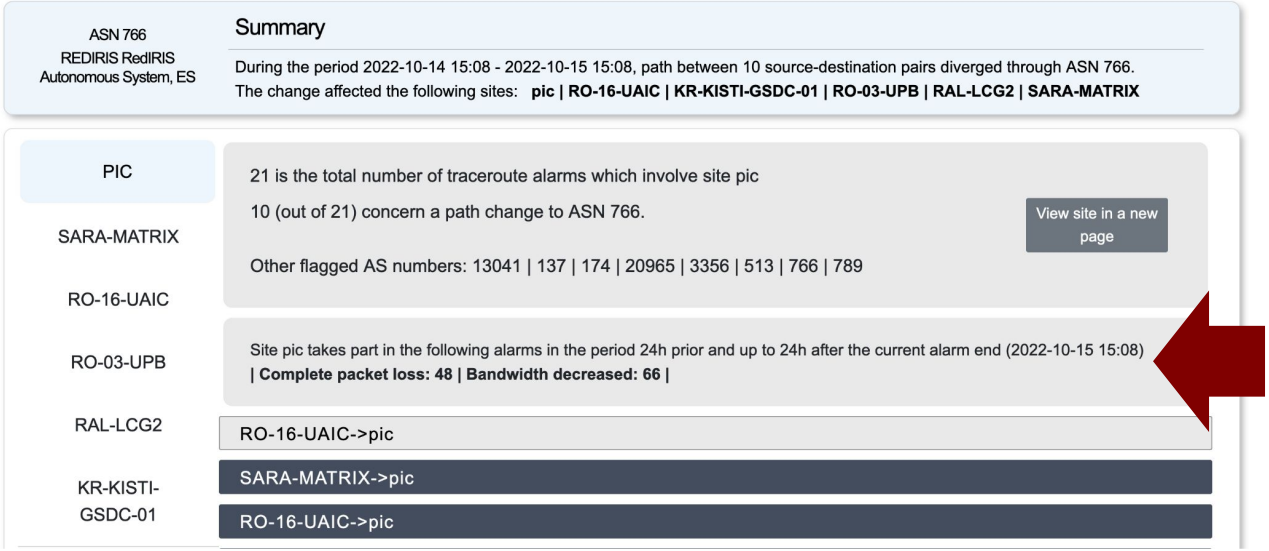
Reimagined as a tool to explore alarms

New features:

- Search alarms based on site names (tagged in the alarm)
- Bandwidth and “path changed” alarms can be explored in a dedicated page

See pS-Dash dashboard at: https://ps-dash.uc.ssl-hep.org/

Updates on pS-Dash (2/2)



New features:

- "Path changed" dedicated page - provides details on the typical paths and unusual changes of ASN
- Alarms can be correlated by looking for other events/alarms happening around the time of the selected alarm

Next steps: Use the data for building an ML/AI model and understand better and/or localize the causes

Example: LHCOPN/LHCONE Load Balancing

Source

INFN-T1

2001:760:4205:254::11

Destination

SARA-MATRIX

2001:610:108:203a::31

Total number of traceroute measures: 280
Other networking alarms: None found

BASELINE PATH

Taken in 50% of time — Always reaches destination: NO

137 513 1103

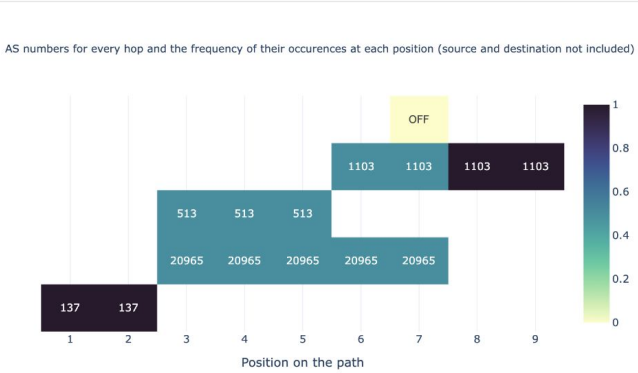
ALTERNATIVE PATHS

Taken in 50.0% of time — Always reaches destination: NO

137 20965 1103

Taken in 0.0% of time — Always reaches destination: NO

137 20965



At position	Typically goes through	Changed to
3	**513** CERN, CH	**20965** GEANT The GEANT IP Service, NL
4	**513** CERN, CH	**20965** GEANT The GEANT IP Service, NL
5	**513** CERN, CH	**20965** GEANT The GEANT IP Service, NL
6	**1103** SURFNET-NL SURFnet, The Netherlands, NL	**20965** GEANT The GEANT IP Service, NL

Example: LHCOPN Alternate via ESnet

USCMS-FNAL-WC1->CERN-PROD

Source

USCMS-FNAL-WC1

131.225.205.23

Destination

CERN-PROD

128.142.208.134

Total number of traceroute measures: 248

Other networking alarms: None found

BASELINE PATH

Taken in 99% of time | Always reaches destination: NO

3152 513

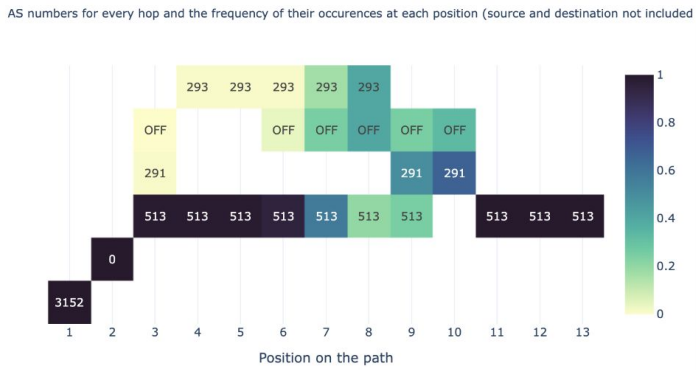
ALTERNATIVE PATHS

Taken in 0.0% of time | Always reaches destination: NO

3152

Taken in 1.0% of time | Always reaches destination: NO

3152 291 293 513



At position	Typically goes through	Changed to
3	**513** CERN, CH	**291** ESNET-EAST, US

At position	Typically goes through	Changed to
4	**513** CERN, CH	**293** ESNET, US

At position	Typically goes through	Changed to
5	**513** CERN, CH	**293** ESNET, US

Example: FNAL Incident (BW drop)

USCMS-FNAL-WC1 to NDGF-T1

Source

USCMS-FNAL-WC1

Destination

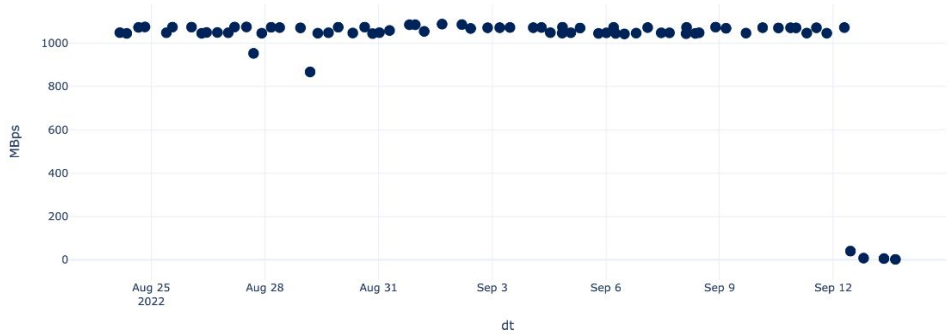
NDGF-T1

Change: -100%

Total number of throughput measures: 71

Other networking alarms

| High packet loss: 2 | High packet loss on multiple links: 1 |



push	MA	src	dest	src_host	dest_host	ipv6	src_site	src_VO	dest_site	dest_VO	src_production	dest_production	timestamp	throughput	pair	dt	MBps
filter da																	
false	131.225.205.23	131.225.205.23	109.105.124.88	psonar3.fnal.gov	perfsonar-ps2.ndgf.org	false	USCMS-FNAL-WC1	UNKNOWN	NDGF-T1	ATLAS	true	true	1661561263000	1048161726	131.225.205.23->109.105.124.88	2022-08-27T00:47:43	1048.16
false	131.225.205.23	131.225.205.23	109.105.124.88	psonar3.fnal.gov	perfsonar-ps2.ndgf.org	false	USCMS-FNAL-WC1	UNKNOWN	NDGF-T1	ATLAS	true	true	1661618925000	952946516	131.225.205.23->109.105.124.88	2022-08-27T16:48:45	952.95
false	131.225.205.23	131.225.205.23	109.105.124.88	psonar3.fnal.gov	perfsonar-ps2.ndgf.org	false	USCMS-FNAL-WC1	UNKNOWN	NDGF-T1	ATLAS	true	true	1662626239000	1045220096	131.225.205.23->109.105.124.88	2022-09-08T08:37:19	1045.22
false	131.225.205.23	131.225.205.23	109.105.124.88	psonar3.fnal.gov	perfsonar-ps2.ndgf.org	false	USCMS-FNAL-WC1	UNKNOWN	NDGF-T1	ATLAS	true	true	1661678463000	1072068304	131.225.205.23->109.105.124.88	2022-08-28T09:21:03	1072.07
false	131.225.205.23	131.225.205.23	109.105.124.88	psonar3.fnal.gov	perfsonar-ps2.ndgf.org	false	USCMS-FNAL-WC1	UNKNOWN	NDGF-T1	ATLAS	true	true	1662439905000	1072905581	131.225.205.23->109.105.124.88	2022-09-06T04:51:45	1072.91
false	131.225.205.23	131.225.205.23	109.105.124.88	psonar3.fnal.gov	perfsonar-ps2.ndgf.org	false	USCMS-FNAL-WC1	UNKNOWN	NDGF-T1	ATLAS	true	true	1661659004000	1073324325	131.225.205.23->109.105.124.88	2022-08-28T03:56:44	1073.32
false	131.225.205.23	131.225.205.23	109.105.124.88	psonar3.fnal.gov	perfsonar-ps2.ndgf.org	false	USCMS-FNAL-WC1	UNKNOWN	NDGF-T1	ATLAS	true	true	1662672411000	1074163359	131.225.205.23->109.105.124.88	2022-09-08T21:26:51	1074.16
false	131.225.205.23	131.225.205.23	109.105.124.88	psonar3.fnal.gov	perfsonar-ps2.ndgf.org	false	USCMS-FNAL-WC1	UNKNOWN	NDGF-T1	ATLAS	true	true	1662902418000	1071231326	131.225.205.23->109.105.124.88	2022-09-11T13:20:18	1071.23
false	131.225.205.23	131.225.205.23	109.105.124.88	psonar3.fnal.gov	perfsonar-ps2.ndgf.org	false	USCMS-FNAL-WC1	UNKNOWN	NDGF-T1	ATLAS	true	true	1662093921000	1085912472	131.225.205.23->109.105.124.88	2022-09-02T04:45:21	1085.91
false	131.225.205.23	131.225.205.23	109.105.124.88	psonar3.fnal.gov	perfsonar-ps2.ndgf.org	false	USCMS-FNAL-WC1	UNKNOWN	NDGF-T1	ATLAS	true	true	1662696230000	1068710540	131.225.205.23->109.105.124.88	2022-09-09T04:03:50	1068.71

Example: Fail-over to Commodity Network

Nebraska -> RAL-LCG2

Source

Nebraska

2600:900:6:1102:2eea:7fff:fef5:d140

Destination

RAL-LCG2

2001:630:58:1820::82f6:b06d

Total number of traceroute measures: 280
Other networking alarms: None found

BASELINE PATH

Taken in 99% of time — Always reaches destination: NO

7896 11537 20965 786

ALTERNATIVE PATHS

Taken in 1.0% of time — Always reaches destination: NO

7896 6939 786

Taken in 0.0% of time — Always reaches destination: YES

7896 7029 6461 3257 1299 786

At position	Typically goes through	Changed to
4	**11537** INTERNET2-RESEARCH-EDU, US	**7029** WINDSTREAM, US

At position	Typically goes through	Changed to
7	**11537** INTERNET2-RESEARCH-EDU, US	**6939** HURRICANE, US

At position	Typically goes through	Changed to
7	**6461** ZAYO-6461, US	**6939** HURRICANE, US

Activity to Gather Site IN/OUT Data for 2nd Data Challenge

- After the WLCG Network Data Challenge last October, we identified site network monitoring as one of the key missing components
- **We would like to have at least this information:**
  - Network information targeted for humans
  - Total network traffic IN/OUT targeted for central monitoring
- **Goal:** our Tier-1s and bigger Tier-2s provide URLs to both types of information and register those URLs into WLCG CRIC
- The WLCG Monitoring Task Force is working on achieving this goal.
- More details are available from my September 2022 GDB presentation.
- You can see the code and docs at https://gitlab.cern.ch/wlcg-doma/site-network-information
- **NOTE**: Wednesday DC discussion https://indico.cern.ch/event/1212782/

Other Activities

Planning to submit one or two abstracts to CHEP22 next month

Working with the RNTWG (see Marian's talk tomorrow) on identifying and monitoring network traffic details via the SciTags initiative.

Exploring other network monitoring activities in the perfSONAR space including ARGUS

Working to organize and annotate our data for ML/AI work (Petya Vasileva)

We will need a concerted effort to get our infrastructure updated and properly (re)configured, once perfSONAR v5 is released and we have all the known bugs addressed.

Summary

- OSG in collaboration with WLCG operates a comprehensive network monitoring platform
  - Provides data and feedback to LHCOPN/LHCONE, HEPiX, WLCG and OSG communities
- The IRIS-HEP project and completed SAND project have produced some new tools for exploring and utilizing our network data that we continue to develop.
- Developing high-level services based on perfSONAR measurements that will help sites, experiments and R&Es receive targeted alarms/alerts on existing issues in the infrastructure
- We have to continue to watch our network monitoring infrastructure as it is a complex system with lots of areas for issues to develop.

Questions / Discussion?

Acknowledgements

We would like to thank the **WLCG**, **HEPiX**, **perfSONAR** and **OSG** organizations for their work on the topics presented.

In addition we want to explicitly acknowledge the support of the **National Science Foundation** which supported this work via:

- OSG: NSF MPS-1148698
- IRIS-HEP: NSF OAC-1836650

Useful URLs

- OSG/WLCG Networking Documentation
  - https://opensciencegrid.github.io/networking/
- perfSONAR Infrastructure Dashboard
  - https://atlas-kibana.mwt2.org:5601/s/networking/goto/9911c54099b2be47ff9700772c3778b7
- perfSONAR Dashboard and Monitoring
  - http://maddash.opensciencegrid.org/maddash-webui
  - https://psetf.opensciencegrid.org/etf/check_mk
- perfSONAR Central Configuration
  - https://psconfig.opensciencegrid.org/
- Toolkit information page
  - https://toolkitinfo.opensciencegrid.org/
- Grafana dashboards
  - http://monit-grafana-open.cern.ch/
- ATLAS Alerting and Alarming Service: https://aaas.atlas-ml.org/
- The perfSONAR Dashboard application: https://ps-dash.uc.ssl-hep.org/
- ESnet WLCG Stardust Dashboard: https://public.stardust.es.net/d/XkxDL5H7z/esnet-public-dashboards?orgId=1

Backup Slides Follow

WLCG Network Throughput Support Unit

Support channel where sites and experiments can report potential network performance incidents:

- Relevant sites, (N)RENs are notified and perfSONAR infrastructure is used to narrow down the problem to particular link(s) and segment. Also tracking past incidents.
- Feedback to WLCG operations and LHCOPN/LHCONE community

Most common issues: MTU, MTU+Load Balancing, routing (mainly remote sites), site equipment/design, firewall, workloads causing high network usage

As there is no consensus on the MTU to be recommended on the segments connecting servers and clients, LHCOPN/LHCONE working group was established to investigate and produce a recommendation. (See coming talk :))

Importance of Measuring 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 experiments 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**