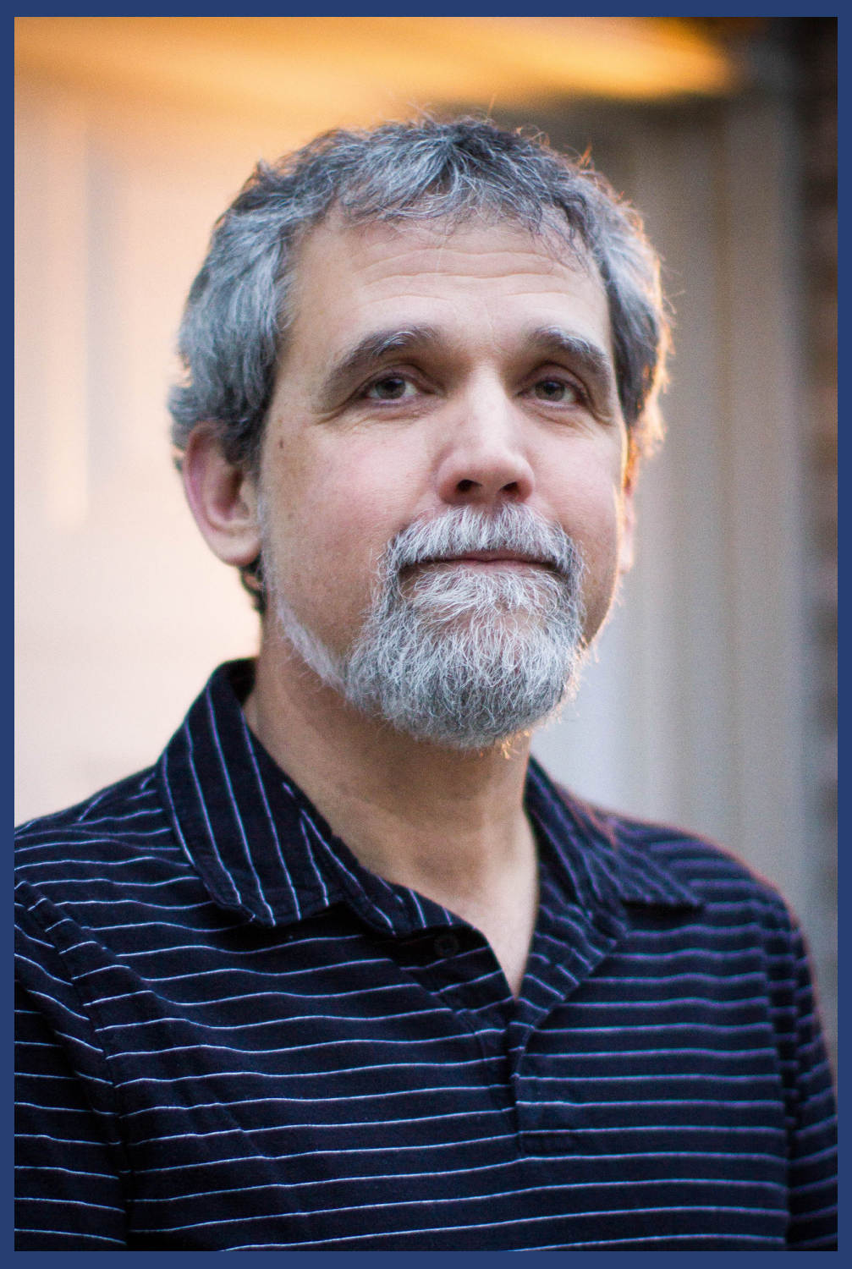


ANALYZING, IDENTIFYING & ALERTING ON NETWORK ISSUES

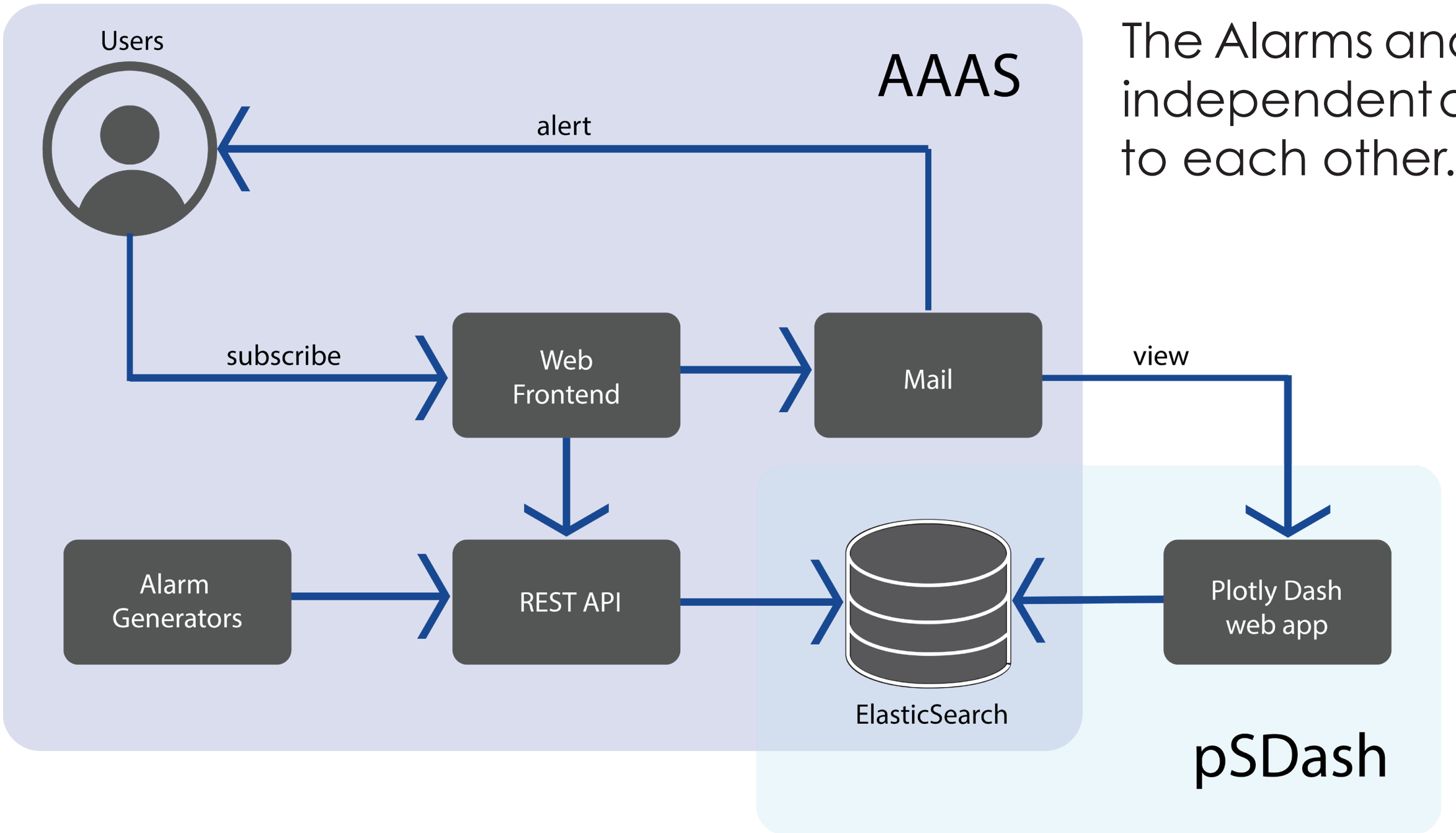
Shawn McKee[1], Marian Babik[2], Petya Vasileva[1], Ilija Vukotic[3]
[1]University of Michgan, [2]CERN, [3]University of Chicago

MOTIVATION

WLCG relies on the network as a critical part of its infrastructure and therefore needs to guarantee effective network usage and prompt detection and resolution of any network issues, including connection failures, congestion and traffic routing. The OSG Networking Area, in partnership with the WLCG Throughput working group, has created a monitoring infrastructure that gathers metrics from the global WLCG perfSONAR deployment and centrally stores the measurements in Elasticsearch.

The collected data is rich and bulky, which makes it hadrer to distinguish real issues from noise. Furthermore, in order to resolve network problems, responsible people should be allowed to be automatically notified whenever their site is involved. This work adresses both concerns by providing two complementary applications - the Alarm & Alerts service (AAAS) and pSDash. AAAS generates and sends notifications of recent network problems, while pSDash provides further details and visual tools based on those problems.

ARCHITECTURE



The Alarms and Alerts Service and pSDash are independent applications but complimentary to each other.

AAAS provides a simple way to store alarms, handles users subscriptions and generates alerts. The alarms are logically grouped in types and generated periodically. A REST API allows for storing new alarms. The users can specify a period, tag sites and select from a list of multiple types of alarms on which to be alerted. Once subscribed, they will receive an e-mail with a list of alarms and links to pSDash for further details.

pSDash is a web application based on Plotly Dash. It provides some interactive visualizations of the network problems and lets users search through the alarms.

ALARM TYPES

The perfSONAR framework delivers hundreds of thousands of results per day by actively measuring various network metrics on more than 5000 links. The amount, quality and complexity of the received data are challenging to analyse and understand. Therefore, our AAAS + pSDash implementation helps in extracting statistically significant cases worth investigating. Bellow are the main types of alarms, where each type uses a specific perfSONAR test.

Traceroute

- ! path changed
- ! destination cannot be reached (from any/multiple hosts)
- ! source cannot reach (any/multiple hosts)

Packet loss

- ! complete packet loss
- ! firewall issues
- ! high packet loss

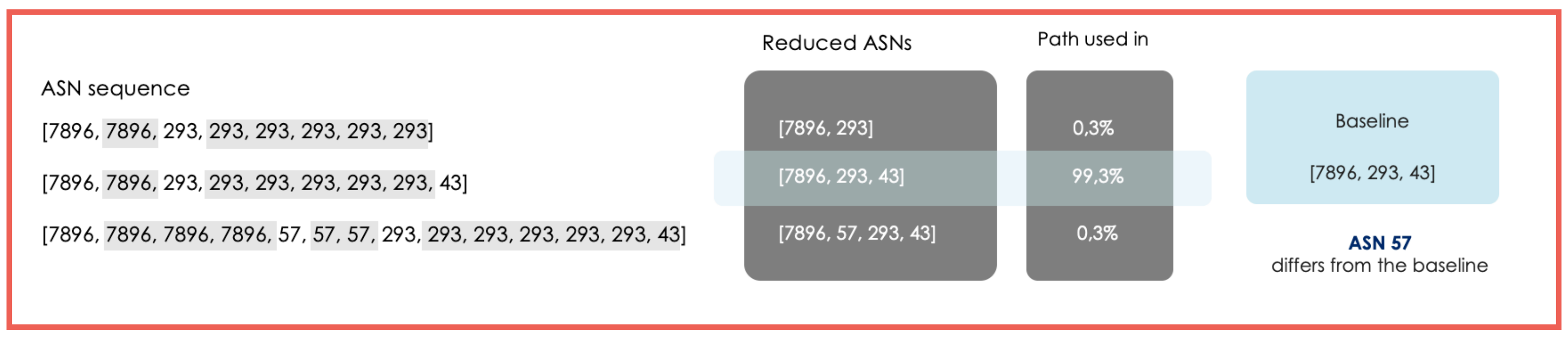
Throughput

- ! bandwidth decreased (from/to multiple sites)
- ! bandwidth increased (from/to multiple sites)

One-Way delay

- ! bad owd measurements
- ! large clock correction

Some of the alarms, such as the traceroute alarms, follow an inductive approach. Traceroute tests measure the network path, which can be quite dynamic in times and tracking changes on the path is not trivial. However, routes which suddenly change trajectory and (for example) utilise general purpose internet providers instead of the R&E networks, can be detected by looking at the ASNs (Autonomous System Numbers). Bellow it is described that we first find a baseline path, then we compare all alternative paths to the norm and return a list of ASNs unusual for the checked site pair.

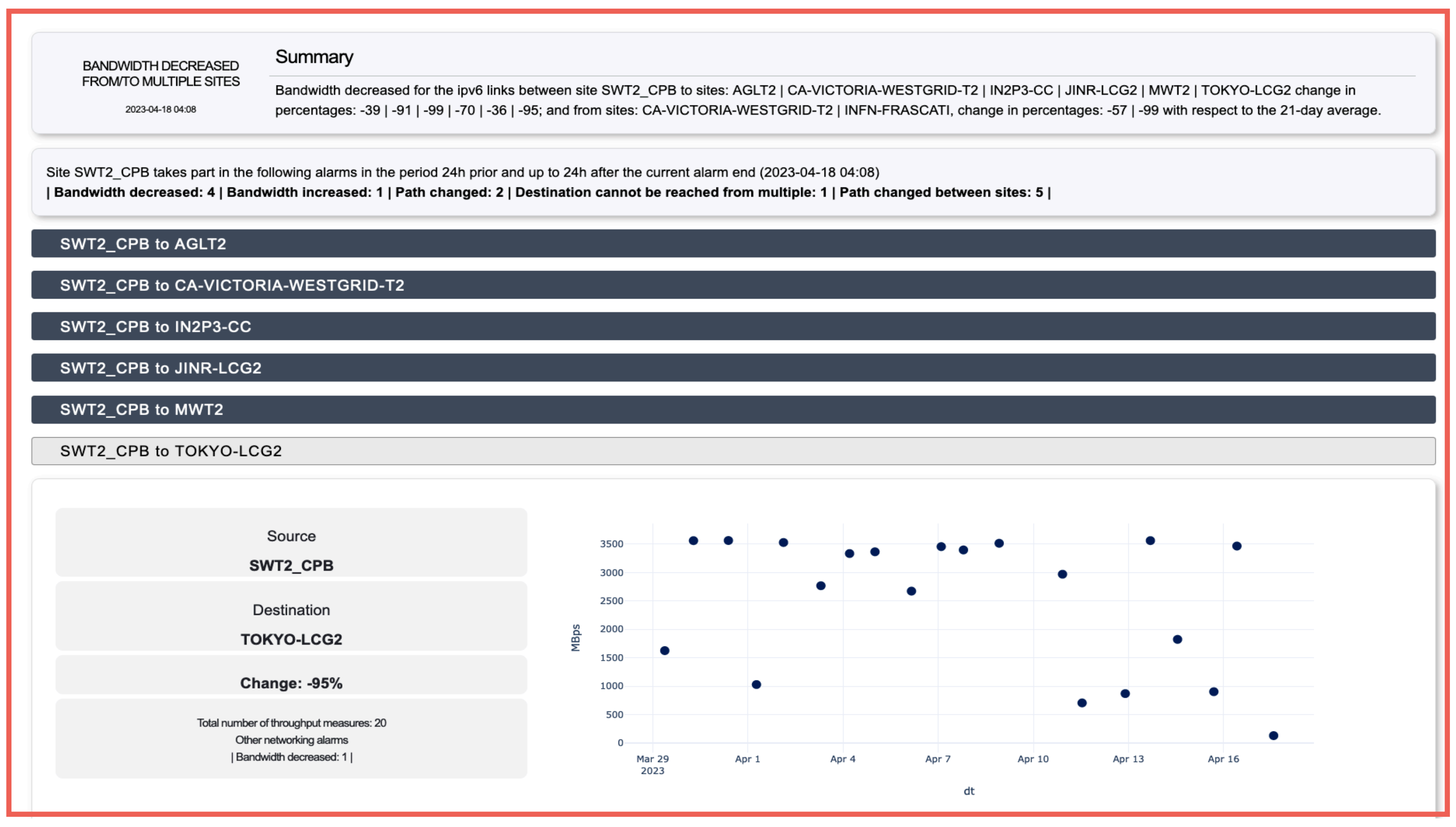


ASN sequence	Reduced ASNs	Path used in
[7896, 7896, 293, 293, 293, 293, 293, 293]	[7896, 293]	0,3%
[7896, 7896, 293, 293, 293, 293, 293, 293, 43]	[7896, 293, 43]	99,3%
[7896, 7896, 7896, 7896, 57, 57, 57, 293, 293, 293, 293, 293, 293, 43]	[7896, 57, 293, 43]	0,3%

Baseline
[7896, 293, 43]

ASN 57
differs from the baseline

Bandwidth alarms consider a period of up to 3 weeks. This is neccessary due to the limitted number of measurements, which can vary from a couple a day to a test every few days. An alarms is generated if recent throughput drops more than 2 sigma from the average over the whole period.



BANDWIDTH DECREASED FROM/TO MULTIPLE SITES
2023-04-18 04:08

Summary

Bandwidth decreased for the ipv6 links between site SWT2_CPB to sites: AGLT2 | CA-VICTORIA-WESTGRID-T2 | IN2P3-CC | JINR-LCG2 | MWT2 | TOKYO-LCG2 change in percentages: -39 | -91 | -99 | -70 | -36 | -95; and from sites: CA-VICTORIA-WESTGRID-T2 | INFN-FRASCATI, change in percentages: -57 | -99 with respect to the 21-day average.

Site SWT2_CPB takes part in the following alarms in the period 24h prior and up to 24h after the current alarm end (2023-04-18 04:08)
| Bandwidth decreased: 4 | Bandwidth increased: 1 | Path changed: 2 | Destination cannot be reached from multiple: 1 | Path changed between sites: 5 |

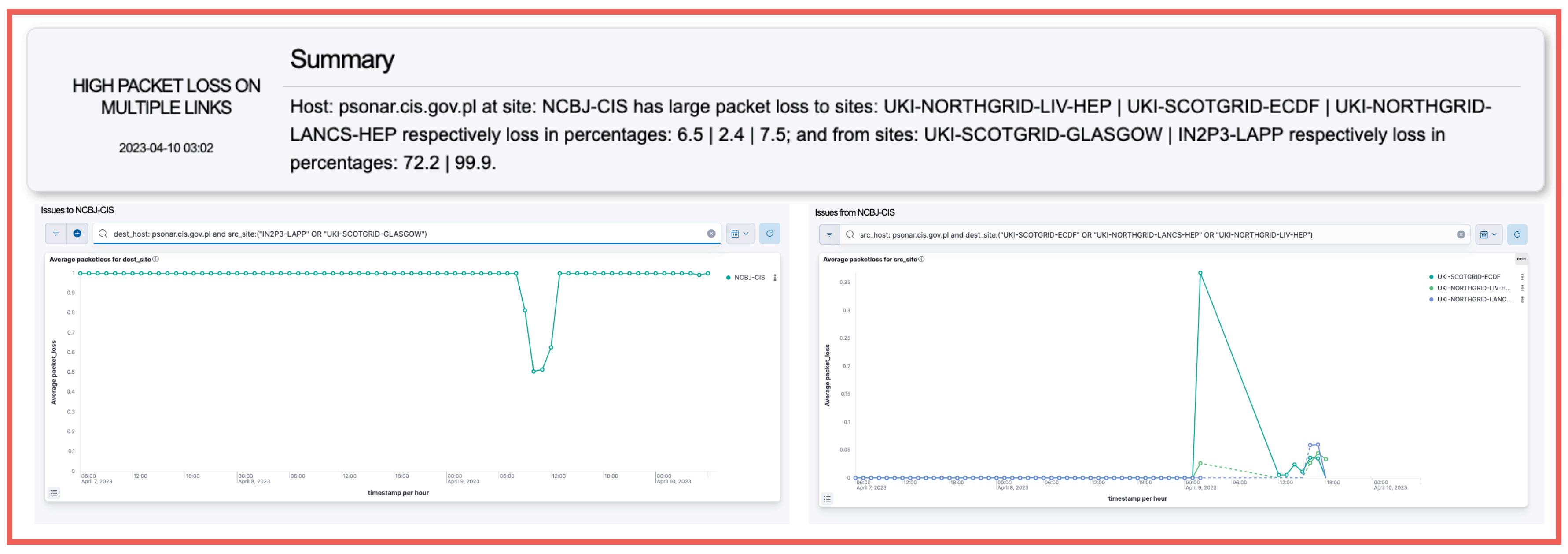
SWT2_CPB to AGLT2
SWT2_CPB to CA-VICTORIA-WESTGRID-T2
SWT2_CPB to IN2P3-CC
SWT2_CPB to JINR-LCG2
SWT2_CPB to MWT2
SWT2_CPB to TOKYO-LCG2

Source
SWT2_CPB

Destination
TOKYO-LCG2

Change: -95%

Total number of throughput measures: 20
Other networking alarms
| Bandwidth decreased: 1 |



HIGH PACKET LOSS ON MULTIPLE LINKS
2023-04-10 03:02

Summary

Host: psonar.cis.gov.pl at site: NCBJ-CIS has large packet loss to sites: UKI-NORTHGRID-LIV-HEP | UKI-SCOTGRID-ECDF | UKI-NORTHGRID-LANCS-HEP respectively loss in percentages: 6.5 | 2.4 | 7.5; and from sites: UKI-SCOTGRID-GLASGOW | IN2P3-LAPP respectively loss in percentages: 72.2 | 99.9.

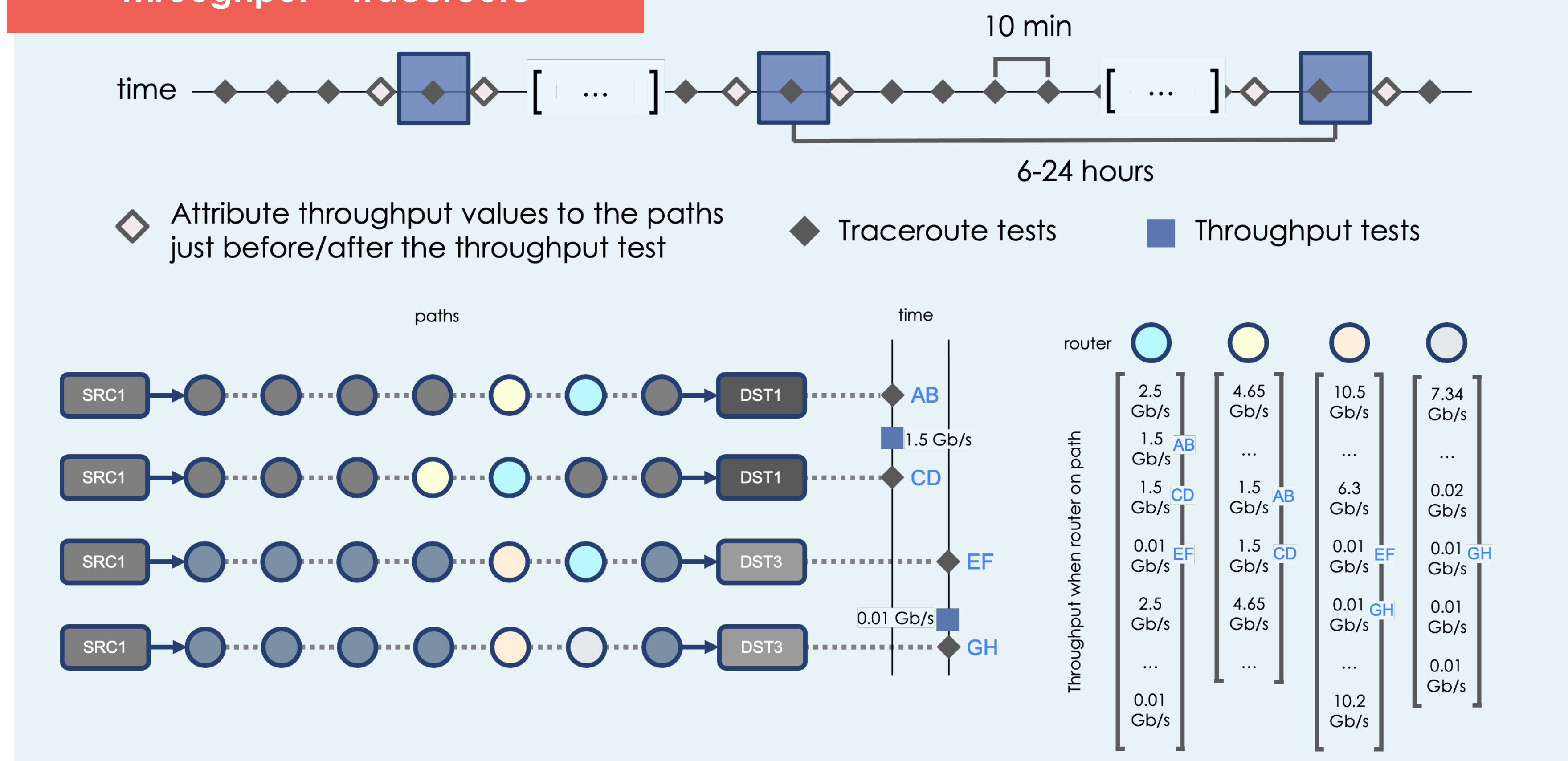
Issues to NCBJ-CIS
dest_host: psonar.cis.gov.pl and src_site:("IN2P3-LAPP" OR "UKI-SCOTGRID-GLASGOW")

Issues from NCBJ-CIS
src_host: psonar.cis.gov.pl and dest_site:("UKI-SCOTGRID-ECDF" OR "UKI-NORTHGRID-LANCS-HEP" OR "UKI-NORTHGRID-LIV-HEP")

WORK IN PROGRESS

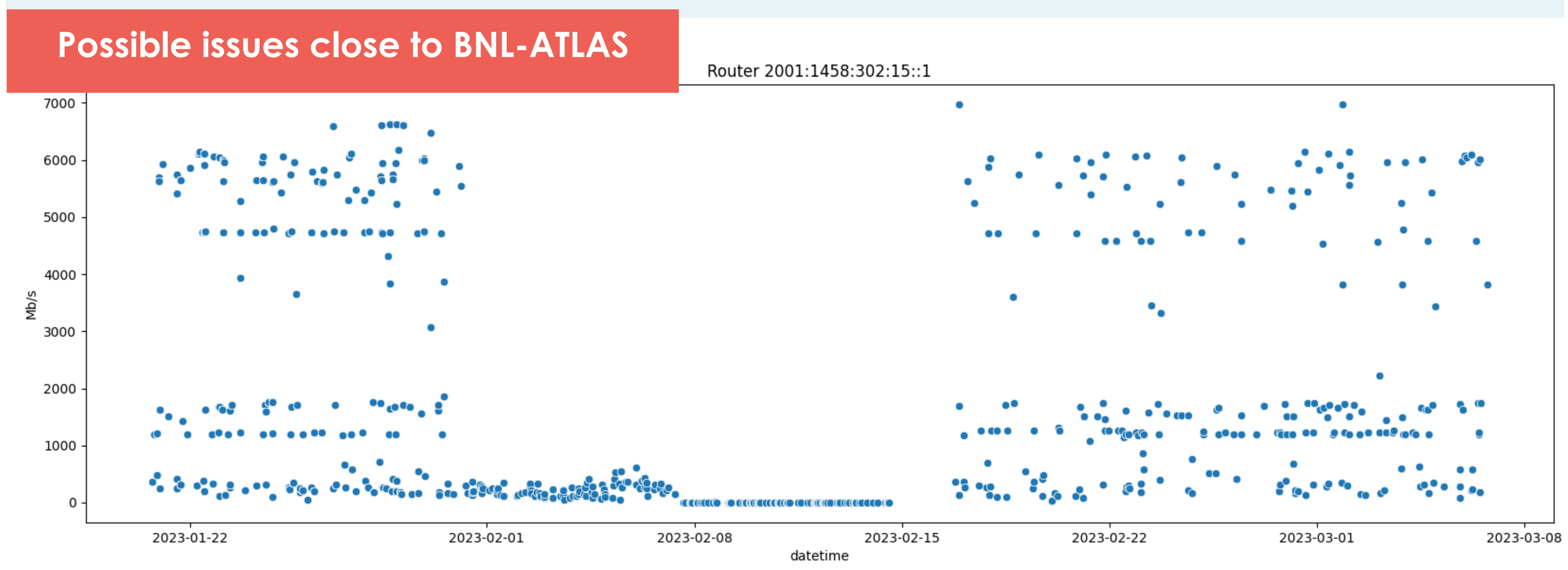
Key role into understanding the R&E networks is to be able to correlate network metrics. However, each perfSONAR test category has a different execution time range. To overcome that obsticle, recent developments attempt to combine traceroute data with bandwidth values. The idea is to use the tracepaths (just before and just after the throughput test) as valid paths and attribute the bandwidth to both.

Throughput + Traceroute



As seen on the figures above, the method assigns one bandwidth value for all routers on the two paths - before the test started and after it completed. The result is a vector of throughput values for each of the routers. Finally, we look for a downtrend, considering the number of values, number of paths and the maximum throughput seen on the paths passing through the router.

Possible issues close to BNL-ATLAS



NEXT STEPS

Future developments are focused on localisation possible source of network issues by correlating the Traceroute measurements with other types of statistics and tests. To achieve that, we will rely on machine learning as a tool for finding non-obvious patterns in our data. In addition, we will build annotated datasets, which will include known and already fixed issues. This will provide additional insights on how the network reacts when a specific problem occurs.

HOW TO ACCESS

Subscribe for alerts at:
> http://psa.osg-htc.org/

Visit pSDash at:
> https://ps-dash.uc.ssl-hep.org/

ACKNOWLEDGMENTS

We want to explicitly acknowledge the support of the National Science Foundation which provided long-term support for this work via:
OSG: NSF MPS-1148698
NSF SAND Project: NSF OAC-1827116
IRIS-HEP: NSF OAC-1836650