

Results from US Mini-Challenges

Diego Davila / University of California San Diego & Shawn McKee / University of Michigan
LHCOPN/LHCONE Meeting #54 (<https://indico.cern.ch/event/1479019/>)
March 18, 2025



Fall 2024 US Mini Challenges

As previewed in the Nov 13, 2024 [WLCG DOMA](#), both USATLAS and USCMS undertook some capacity mini-challenges, designed to benchmark our current infrastructure.

These were simple load-tests where we wanted to evaluate the capacity limits for our various sites.

We were not trying to identify where we might adversely interact with other activities, as we do when we run the regular data challenges.

The fall challenges were orchestrated by Hiro Ito / BNL (for USATLAS) and Diego Davila / UCSD (for USCMS).



Open Science Grid



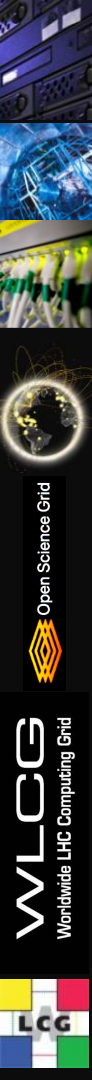
WLCG
Worldwide LHC Computing Grid



Original plan: USATLAS Mini Data Challenge Fall 2024

- Use [load test](#)¹ from T1 to each T2s at full T2's network capacity
 - To check if there are any changes from the results from the last test.
 - Network capabilities of US T2s: AGLT2(200 Gbps) MWT2 (200 Gbps), NET2 (expected to be 400 Gbps). SWT2 (100 Gbps)
 - Individually as well as simultaneously
 - Simultaneous test might present “choke” point in the path.
- T2s to T1 at full wan disk capacity.
 - Not capable to reach the full network capability of BNL at 1.6 Tbps due to the storage layout of T1 storage
- T1 Tape staging and readout test.
 - Check the staging throughput and readout throughput of staged data from BNL.
- Check and validate the accuracy of the various monitor at the site as well as the central ones at CERN, ESNet, BNL,...

¹ The program is found at the following BNLBox folder <https://bnlbox.sdcc.bnl.gov/index.php/s/XGs6LJEGNzf69zK>



USATLAS Testing Summary

- Summary of injection testing to US ATLAS sites (See backup slides)
 - **AGLT2** has achieved **150 Gbps**.
 - **MWT2** has achieved **200 Gbps**.
 - **BNL** has achieved **125 Gbps**.
 - In addition to the disk throughput, the analysis of tape system in terms of the staging throughput is on going.
 - BNL tests will be redone using AGLT2, MWT2 and NET2 to better identify bottlenecks.
 - **SWT2** UTA has achieved **30 Gbps**.
 - **NET2** was not quite ready for the testing.
 - It is waiting for the completion of the network upgrade to 400Gbps
 - It was eventually tested during the Jumbo Frames tests in February 2025. (200+ Gbps)
- As a facility we plan to continue “benchmarking” our sites on a regular basis.
 - We will continue inter-leaving capacity and capability tests moving forward.



USATLAS Capability Testing

In addition to capacity tests, we want to also investigate new tools, methods and technologies that might improve our **capabilities** for the HL-LHC era.

USATLAS targeted February 2025 as a “Capabilities” testing month. We started from the list of topics being tracked in a [WLCG DOMA Google folder](#).

NOTE: the WLCG DOMA plan is to have advocates self-organize (bottom up approach) rather than forced from the top.

USATLAS has focused on a few specific capabilities to start: **Jumbo Frames, SciTags, IPv6-only sites, VPN overlays and SDN via SENSE/Rucio.**

The only capability we managed to test during February was Jumbo Frames



(US)ATLAS Jumbo Frame Testing

Originally the USATLAS facility was planning jumbo frame testing between our sites.

However we were contacted by colleagues at CERN who were interested in testing ATLAS jumbo frames on long distance paths and this seemed to be a better test.

The overall results of the testing involving CERN are shown in the WLCG DOMA [slides](#) from Maria Arsuaga-Rios.

ATLAS-RUCIO: [CERN](#) to [NET2](#) and [BNL](#) transfers via RUCIO.

- Both NET2 and BNL already have JUMBO frames.
- Transfer 104 TB from CERN to NET2 and from CERN to BNL to evaluate performance with 3-4GB file sizes with and without jumbo frames (CERN end)

Results summary (details in [Google doc](#)):

- **Jumbo frames didn't adversely affect site performance or operations**
- **NET2:** a **12% throughput improvement** with Jumbo, **BNL:** no change in performance
- BNL likely has other bottlenecks than the network (unusual storage/net topology)



USATLAS Plans

Our current focus is on various capabilities we want to investigate including:

- Netbird (multipoint VPN) (MWT2)
- SciTags (flow labeling) but will require dCache support
- IPv6-only sites (AGLT2, NET2)
- SENSE/Rucio (NET2)

We would like to schedule a **capacity mini-challenge** (perhaps in conjunction with other experiments in our region) sometime in June or July.

Looking for **additional proponents** to help define, execute and document capabilities.



USCMS Fall-2024 mini-challenge - Overview

- Carried out during the Fall of 2024 (Nov 25th - Dec 17th)
- Two main goals:
 - Test all USCMS Sites at the highest proposed rate of DC24 (FNAL: 400, T2s: 100 Gbps)
 - Make sure all Sites report to the WLCG monitoring:
<https://monit-grafana-open.cern.ch/d/Mwuxgoglk/wlwg-site-network?orgId=16>
- For the load test we used the same tool used in DC24: dc_inject:
https://gitlab.cern.ch/wlwg-doma/dc_inject
- Some sites were excluded due to different reasons:
 - FNAL: Weeks before this test, a Production campaign drove FNAL above 700 Gbps
 - Caltech(*): Similarly had demonstrated performing at ~300Gbps
 - Mostly Storage and not via ESnet (just to UCSD via CENIC-dev)
 - Currently moving their infrastructure to reach ESnet via 400 Gbps
 - UCSD: We had an ongoing network upgrade... believe it or not, we are still in it.



USCMS Fall-2024 mini-challenge - Testing Summary

		OUT (Site → FNAL)		IN (FNAL → Site)	
Site	Max Target	Max Achieved	Best day Average	Max Achieved	Best day Average
Nebraska	125	100	80	100	90
Wisconsin	150	160	100	75	60
Florida	300	190	100	200	135
Vanderbilt	100	100	90	100	80
MIT	80	70	65	30	28
Purdue	100	100	100	70	60

Max Achieved: more or less stable Maximum

Best day Average: more or less sustained Average

Full report available here:

https://docs.google.com/document/d/1rtfhxfsvHYfqc5xQXsFQot-Vu_Ek83WGJMEfKCiVs1/edit?usp=sharing

USCMS Fall-2024 mini-challenge - New data

		OUT (Site → FNAL)		IN (FNAL → Site)	
Site	Max Target	Max Achieved	Best day Average	Max Achieved	Best day Average
Nebraska	125	100	80	100	90
Wisconsin	150	160	100	75	60
Florida	300	190	100	200	135
Vanderbilt	100	100	90	100	80
MIT	80	70	65	30	28
Purdue	100	100	100	70	60

Latest tests show
120 Gbps [2]

Latest tests show
180 Gbps [1]

[1]<https://monit-grafana-open.cern.ch/d/Mwuxgoglk/wlwg-site-network?orgId=16&from=1740498856914&to=1740501966953&var-site=UFlorida-HPC&var-bin=1m>

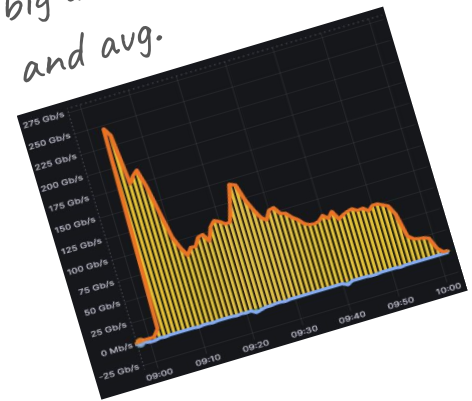
[2]<https://monit-grafana-open.cern.ch/d/Mwuxgoglk/wlwg-site-network?orgId=16&from=1740593954294&to=1740598760934&var-site=Nebraska&var-bin=1m>

USCMS Fall-2024 mini-challenge - Highlights

		OUT (Site → FNAL)		IN (FNAL → Site)	
Site	Max Target	Max Achieved	Best day Average	Max Achieved	Best day Average
Nebraska	125	100	120	100	90
Wisconsin	150	160	100	75	60
Florida	300	190	100	200	180
Vanderbilt	100	100	90	100	80
MIT	80	70	65	30	28
Purdue	100	100	100	70	60

big diff between read/write rates

big diff between max and avg.



*Reads are great,
writes not so much*

*Recently upgraded
needs re-test*

USCMS Fall-2024 mini-challenge - Highlights

		OUT (Site → FNAL)		IN (FNAL → Site)	
Site	Max Target	Max Achieved	Best day Average	Max Achieved	Best day Average
Nebraska	125	100	120	100	90
Wisconsin	150	160	100	75	60
Florida	300	190	100	200	180
Vanderbilt	100	100	90	100	80
MIT	80	70	65	30	28
Purdue	100	100	100	70	60

Almost all sites can sustain reads at ~100 Gbps

Significant variation in write rates

USCMS Fall-2024 mini-challenge - WLCG Monitoring

Site	Status	Observations
Caltech	OK	
Florida	OK	Fixed by Swapping IN and OUT in the SNMP script
MIT	Missing	They haven't contacted their Network team to request SNMP access
Nebraska	OK	
Purdue	Missing	Was working until last month. Needs reconfiguration after a Router migration
UCSD	OK	Fixed by filtering unwanted connections. Https version cannot handle more than 1 connection at a time
Vanderbilt	OK	Deployed the go version. They had issues with how the output was being handled but this has been fixed by monIT as of this week.
Wisconsin	OK	
FNAL	OK	Reconfigured after upgrading their Border Router

USCMS Fall-2024 mini-challenge - Outcomes

- Found an issue at Wisconsin (under investigation)
 - Reads go on a 200 Gbps path whilst writes go over a 100 Gbps one.
- Found a bottleneck at Nebraska. (upcoming upgrade should take care)
 - They are technically connected via 2x100G but logically traffic into their storage system is always assigned to a single link which caps Writes to 100G
- Found a problem with the go implementation of the WLCG Monitoring
 - It's been fixed
- Found missing checksums at Florida
 - This was fixed the same day it was found
- Increased FTS limits to 300+ from the 200 default for most T2s
- Switched: Florida and Wisconsin to use FNAL's FTS instead of CERN's
- Found an asymmetry within ESnet for Purdue: reads/writes over different interfaces.
 - This issue hasn't been solved yet.



USCMS Capabilities mini-challenge - Preview

As opposed to the usual “Load Tests” this “Capabilities” mini-challenge focus on testing new technologies on production.

In this first round USCMS picked the following Capabilities to test:

1. Jumbo Frames (ongoing)
2. Scitags (ongoing)
3. File Sizes (coming soon)
4. SENSE (delayed until June)

USCMS Capabilities mini-challenge - Jumbo Frames

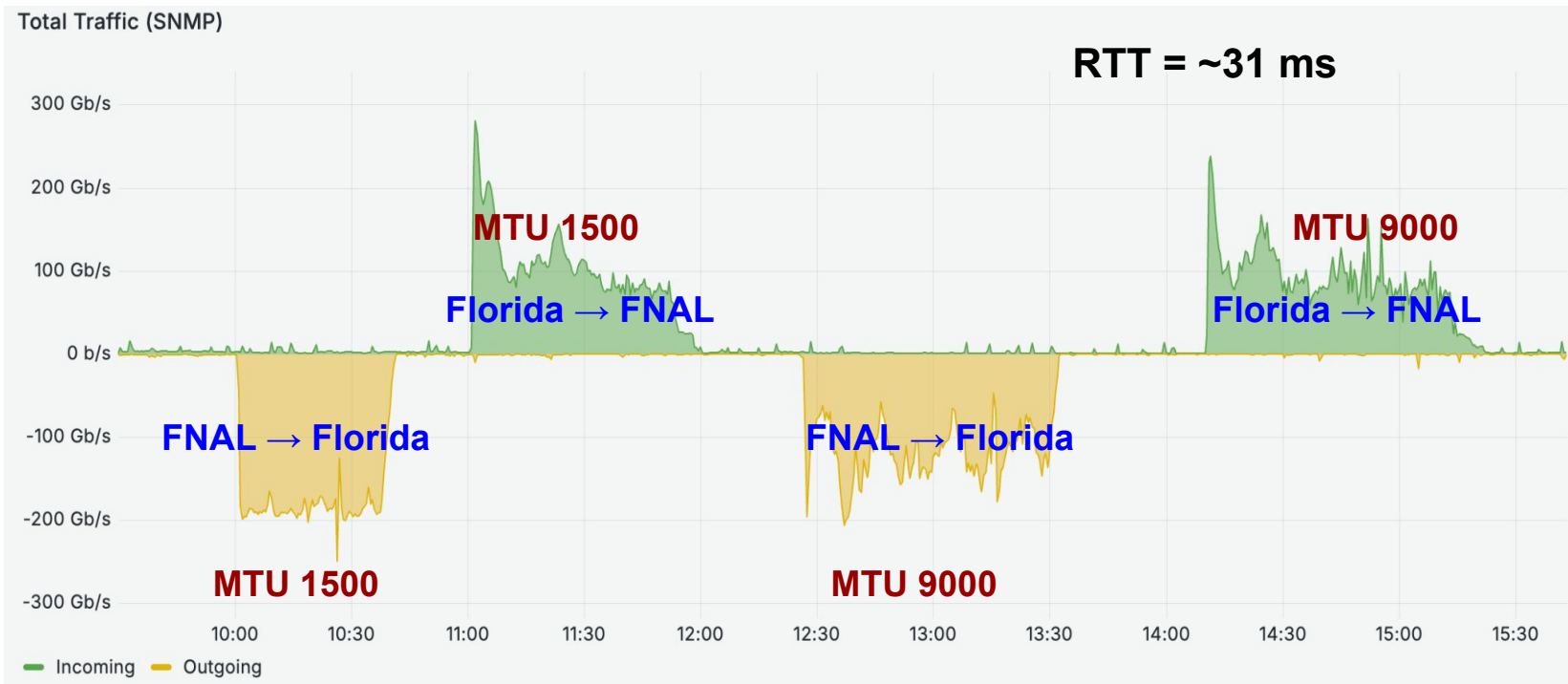
[This is just a preview, more analysis is needed]

- We used 2 pairs of sites for these tests:
[FNAL ↔ Florida] and [FNAL ↔ UNL]
- Using the dc_inject tool we triggered transfers totaling 45TB (1hr @ 100 Gbps) first in one direction i.e. FNAL => Florida and then the opposite.
- The above was repeated after switching the T2's DTNs to MTU 1500 (default is 9k)

Thanks to Asif Shah for running these tests!!

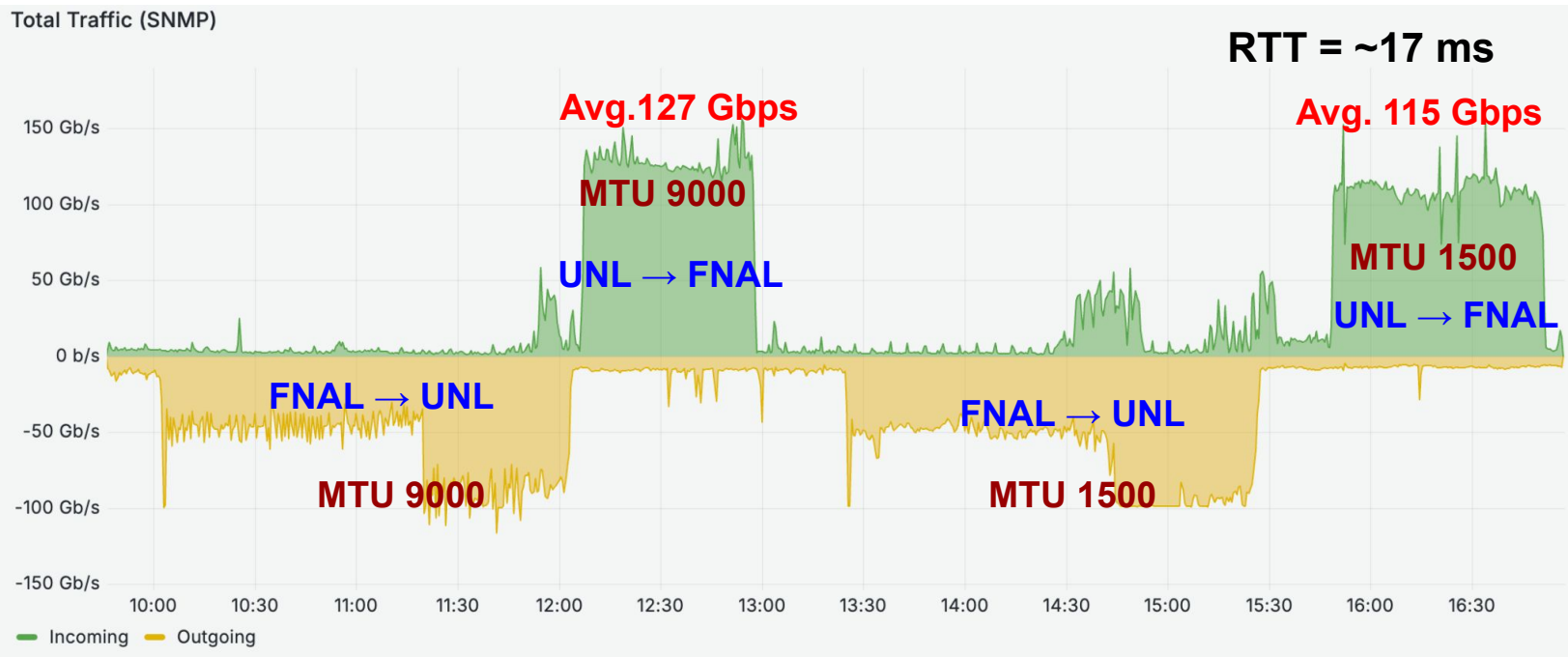
USCMS Capabilities mini-challenge - Jumbo Frames

Preliminar results: FNAL \leftrightarrow Florida



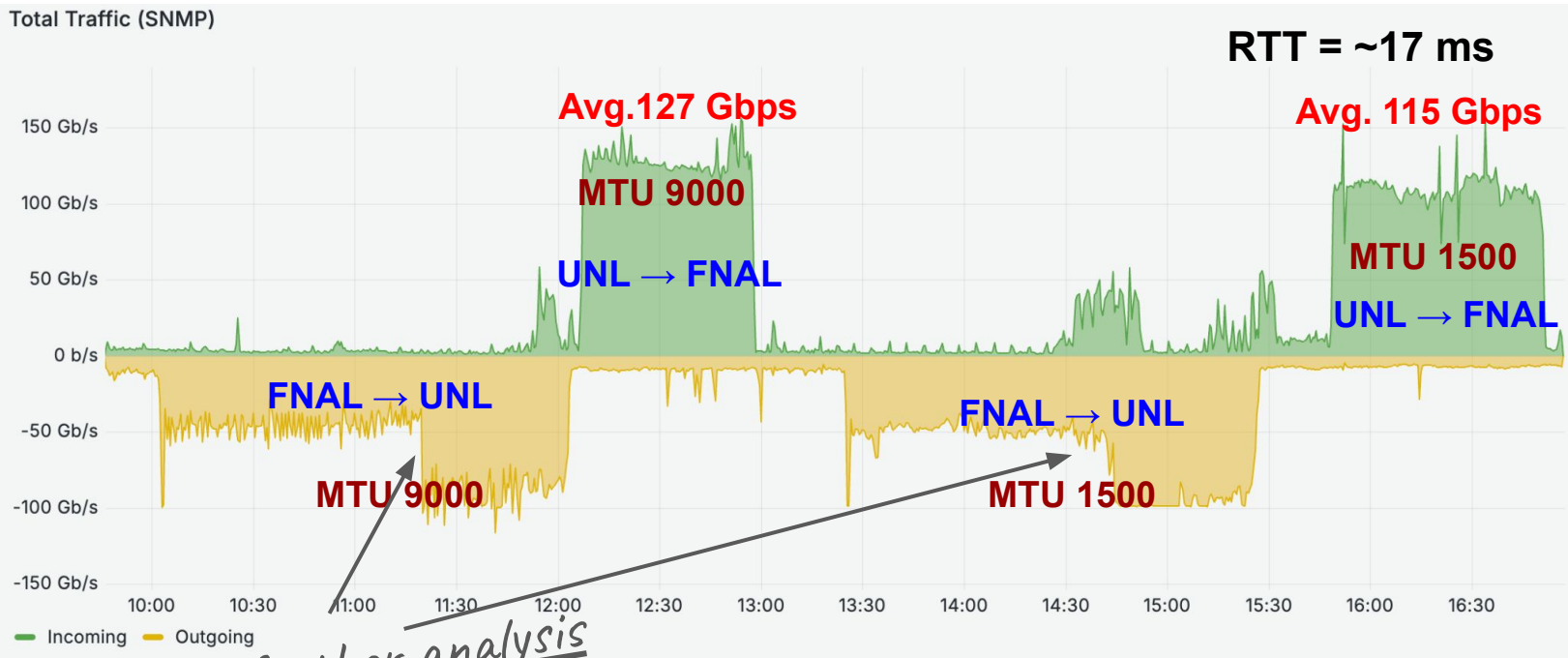
USCMS Capabilities mini-challenge - Jumbo Frames

Preliminar results: FNAL \leftrightarrow UNL



USCMS Capabilities mini-challenge - Jumbo Frames

Preliminary results: FNAL \leftrightarrow UNL



USCMS Capabilities mini-challenge - Scitags

[This is just a preview, more analysis is needed]

- Two main goals:
 - Get more sites configured to send fireflies
 - Validate the data collected
- So far we have 5 out of 8 USCMS sites configured: Caltech, UCSD, Purdue, Wisconsin and Florida
- We performed a quick-validation by transferring 1 dataset marked as “Data Challenge” from UCSD to Wisconsin, then pulling the relevant scitag data and comparing the stats.

Thanks to Garhan Attebury for helping with this activity!

USCMS Capabilities mini-challenge - Scitags

```
import requests
import json

headers = {'Content-Type': 'application/json'}
query='meta.esdb.src.org.short_name:UWMADISON AND meta.activity_name:"Data Challenge"'

#March 5th, 2:30pm (pacific) in miliseconds
start=1741213813000

data={ "size":10000,"query":{"bool":{"filter":[{"range":{"start":{"gte":start,"lte":(start+3600000)}}, {"query_string":{"analyze_wildcard":"true","query":query}}]}}}

data_string = json.dumps(data)
response = requests.post('https://e1.gc1.prod.stardust.es.net:9200/stardust_firefly-*/_search', headers=headers, data=data_string)

d = response.json()
records_list=[]
for record in d['hits']['hits']:
    records_list.append(record["_source"])

#Every file transfer produces 4 fireflies, this number should match the number of files trasferred and it does
num_files = len(records_list)/4
print("Number of files: "+str(num_files))

d = response.json()
records_list=[]
total_received=0
for record in d['hits']['hits']:
    total_received+=record['_source']['values']['usage']['received']

print("Total data: "+str(total_received/(10**9))+ " GB")
```

USCMS Capabilities mini-challenge - Scitags

```
import requests
import json

headers = {'Content-Type': 'application/json'}
query='

#March
start=1

data={

data_st
respons

d = res
records
for rec

#Every
num_fil
print("

d = res
records
total_received=0
for record in d['hits']['hits']:
    total_received+=record['_source']['values']['usage']['received']

print("Total data: "+str(total_received/(10**9))+ " GB")
```

- We compared 2 metrics: Number of files and total data transferred
- Knowing that each file transferred produces 4 fireflies (2 per site), we can see that we get exactly 4 times (968) the number of files in the datasets (242)
- Total data transferred doesn't match exactly (673.92 vs 673.01GB) but it is very close. This is because fireflies include overheads (e.g. protocol headers).

Summary & Plans

We have successfully measured the current USATLAS and USCMS capacities during the Fall 2024 and also have identified some issues that require further work.

Capability testing is ongoing. Stay tuned!

We (IRIS-HEP/OSG-LHC/US-LHC) need to further **clarify** and **document** existing plans, **mini-challenges** and goals for the next year and onwards to DC27.

These efforts should help us improve our infrastructure, drive technology deployment and demonstrate capabilities at scale.

Questions or Discussion?



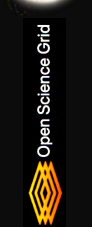
Acknowledgements

Thanks to **Hiro Ito** for his contributions to the slides and for running the USATLAS tests!! and to **Asif Shah** for helping running the USCMS tests.

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:

- **IRIS-HEP: NSF OAC-1836650 and PHY-2323298**



Background Material

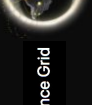
Here are some resources we know about:

Presentations

- [WLCG Data Challenge 2024 \(DC24\) Status and Plans Related to ATLAS DDM](#) (Jun 2023)
- [DC24 Planning and Near Term Activities](#) (Jul 2023)
- [USATLAS Data Challenge 2024 Take-aways](#) (Feb 2024)
- [Medium to Long Term Network Plans for ATLAS and CMS](#) (Mar 2024)
- [DC24 Network Activities & Results](#) (May 2024)

Some Google Docs

- [WLCG/DOMA Data Challenge 2024: Final Report](#)
- [USATLAS Milestones/MiniChallenges for Next WLCG Data Challenge in 2024](#)
- [Planning Mini-Challenges for US ATLAS Facilities and Distributed Computing](#)
- [NOTES: USATLAS Facility Status and Evolution Discussion](#)



DC24 Links

Official DC24 report

<https://zenodo.org/records/11402618>

DC24 Network Activities and Results:

<https://docs.google.com/presentation/d/1s0VvbXEpj1PN9umFT8wgsHsHmG9EYucymbalKNrvuKQ/edit#slide=id.p1>

Katy Ellis LHCONE/LHCOPN DC24 presentation:

https://docs.google.com/presentation/d/1Tm3pCMkfHj5KHTW3PXbgS7mdHf72lr27qr1JgMbrnRg/edit#slide=id.g1ea89411ecb_0_4

Next Steps Towards DC26:

https://docs.google.com/presentation/d/1mMx6QaihWJWpbVEQgxNjZXRT5_s4SkBTXu0SpELtuvl/edit#slide=id.gd170caf633_1_0

DC24 ATLAS Retrospective:

https://docs.google.com/presentation/d/1Lh_D57BvWn13AFCIhhucz-m-j-tKV-yMez_oD4yYUtBo/edit#slide=id.gd170caf633_1_0

Backup Slides

OSG-LHC/IRIS-HEP Current Plans

At the [IRIS-HEP retreat](#) in September 2024, we discussed how to prepare for DC26
As mentioned, mini-challenges are an important tool that we want to enable

Goals for the next DC:

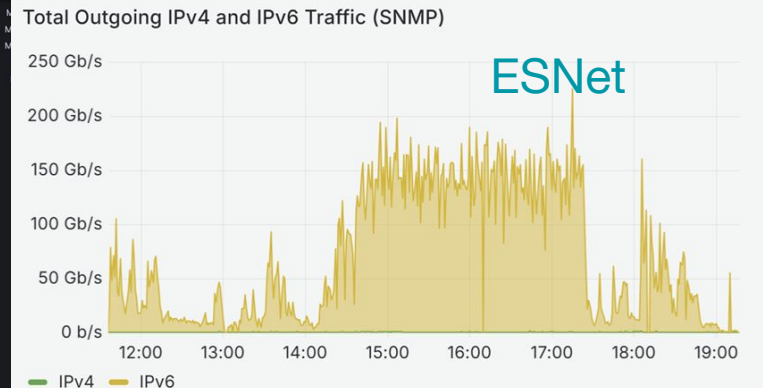
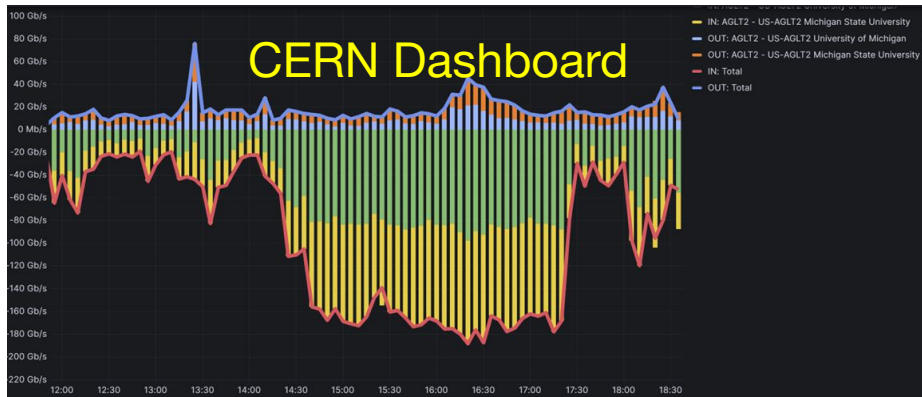
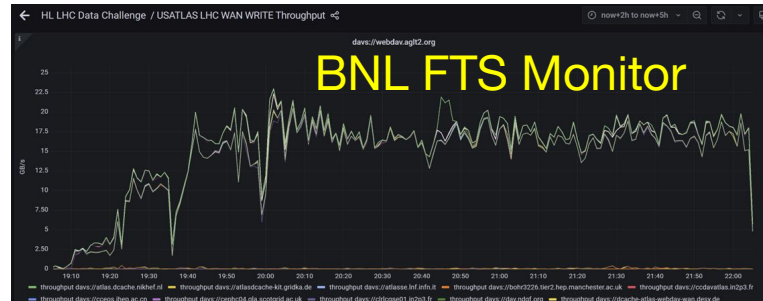
- Move the majority of our data via IPv6 and have one or more sites **IPv6-only**
- Have 80%+ of our traffic identified by SciTags
- Have SENSE/Rucio used in production at one or more sites
- Improve site network monitoring to identify traffic by LHCONE, LHCOPN, R&E and commodity

The plan:

- (DONE) Before the end of 2024 rerun capacity tests for US sites to determine current values
- (NEXT) February 2025, execute a joint USATLAS-USCMS **capabilities** mini-challenge: scitokens, SciTags, SENSE, jumbo frames
- Early-to-mid Summer 2025, execute a joint USATLAS-USCMS **capacity** mini-challenge

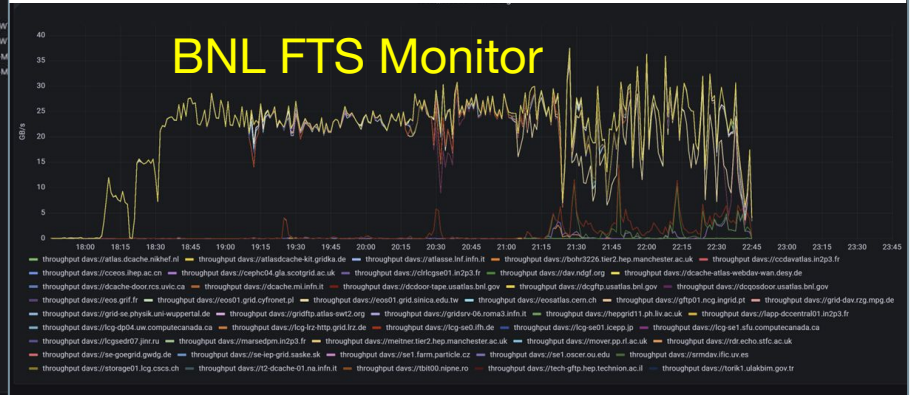
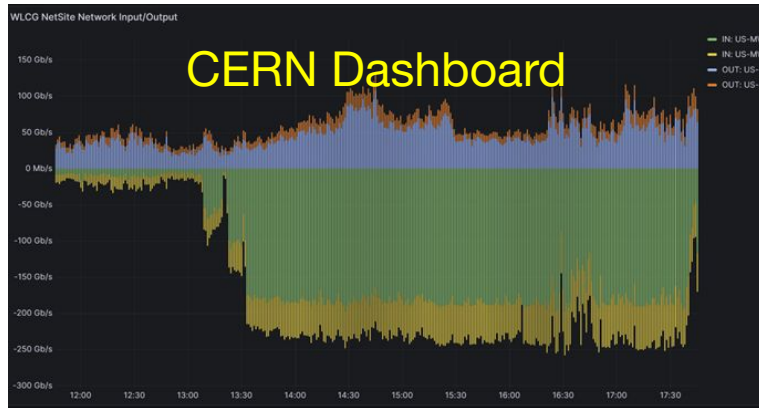
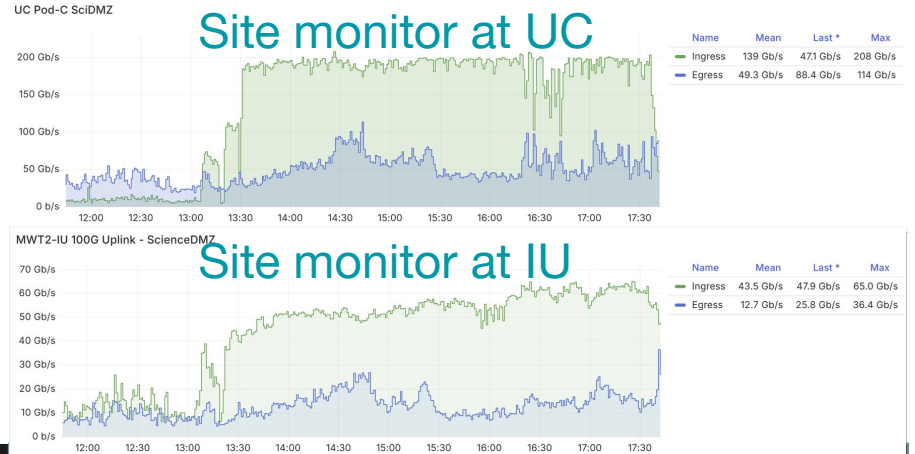
AGLT2 Ingestion Testing

- The observed throughput for injecting AGLT2 was about 150Gbps.
- Various monitors were checked against each other to evaluate their accuracies.
- Although all monitor shows the similar number, CERN Dashboard seems a bit higher the other two? **However we must note that CERN Site Network is ALL traffic**



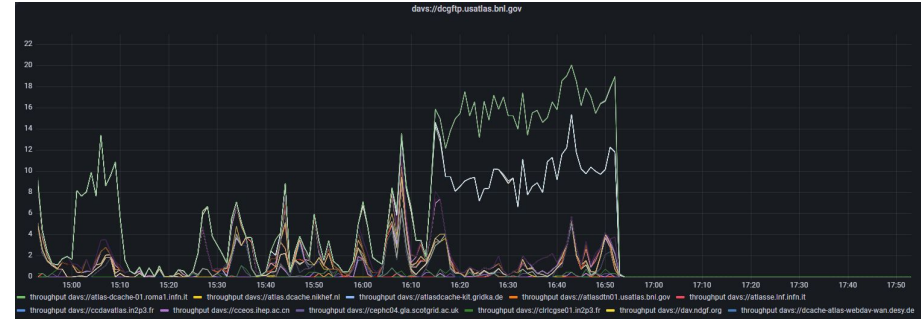
MWT2 Ingestion Testing

- The observed throughput for MWT2 was about 200Gbps.
- CERN Dashboard shows again a bit higher values.
- NOTE: ESNet monitor only shows UC.

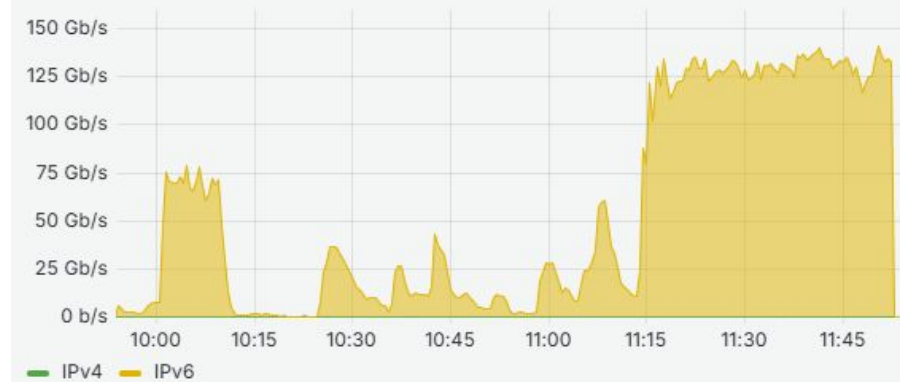


BNL T1 Ingestion Testing

- **Complication**
 - It requires multiple sites to drive BNL to its bandwidth capacity
 - AGLT2 encountered storage issue at the time of BNL testing.
 - Cause of delay
 - Some shorter testing after AGLT2 became operational.
- It achieved ~125Gbps.
- It requires additional testing to investigate the actual current limitation. (Redo in February?)



Total Outgoing IPv4 and IPv6 Traffic (SNMP)



SWT₂

- SWT2 (UTA) has achieved 30 Gbps.
- This is still the limit of the network at the site.
 - The flatness of the plot indicates that it is indeed the network limit.
- Discussion with the network engineer is under way to increase the bandwidth. (Needs both bandwidth and DTN capacity increases)

