FERMILAB-SLIDES-24-0094-ITD

Fermilab U.S. DEPARTMENT OF ENERGY | Office of Science

Fermilab Site Report

Spring 2024 ESCC meeting, Berkeley,CA,

April 30 - May 2, 2024

Site Coordinators: Andrey Bobyshev, Phil Demar

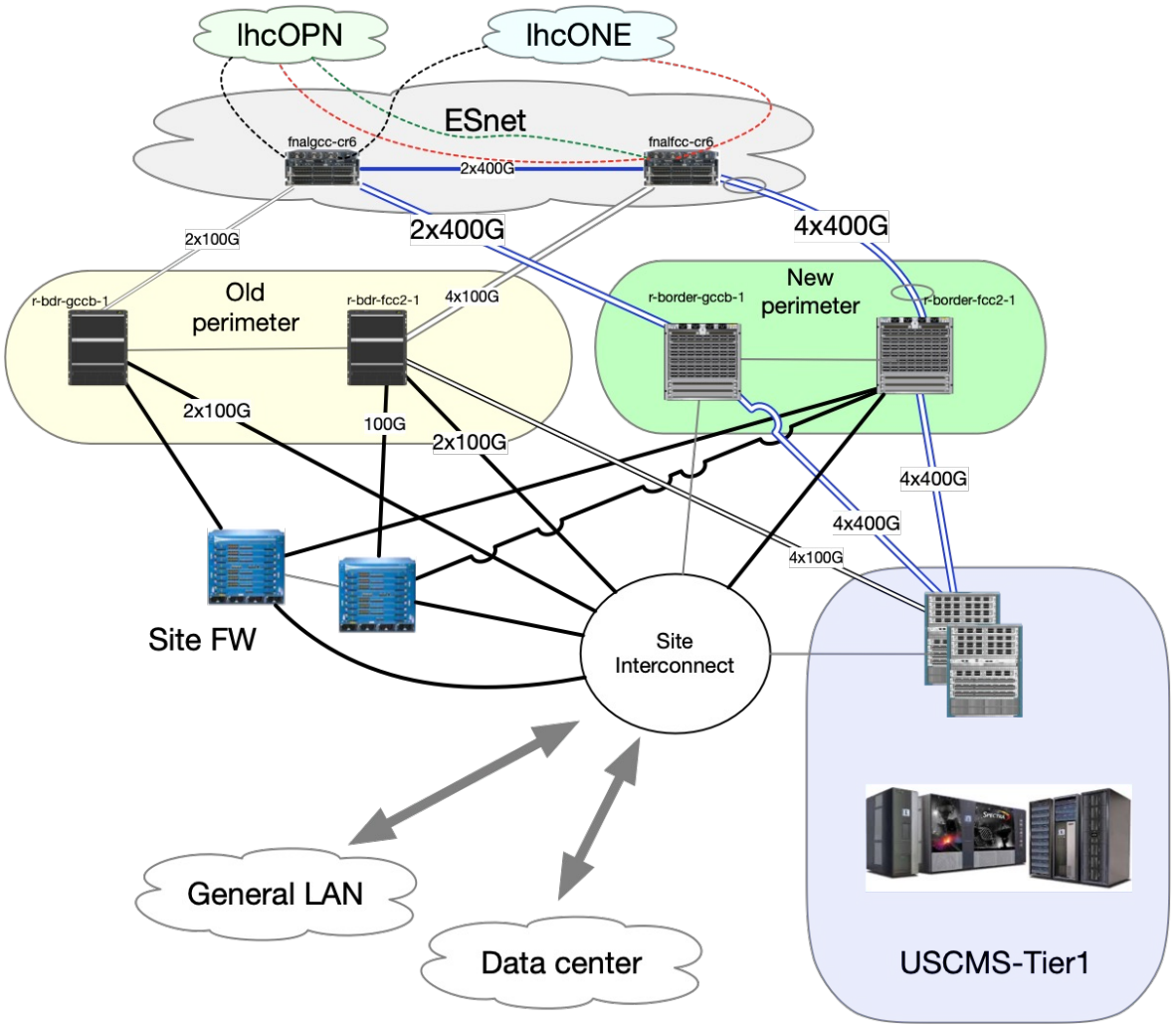
Outline

- Major upgrades
  - New border routers
  - Palo-Alto Site Firewall 100G
- LHC DC24
- Efforts on securing Controlled Unclassified Information
- IPv6 status and updates
  - IPv6 intermittent connectivity problem
- Answers to the questions posted in the Site Report Template

Network perimeter upgrades

- Currently:
  - Two Brocade MLXe16, 2x 100G and 4x 100G aggregated BW
- New deployment in-progress
  - Two Arista 7800R3
    - 2x 48-port 100G LC
    - 2x 36-port 400G LC
    - Deep buffers
    - 800G enabled
    - 2x 400G and 4x 400G off-site connections

A simplified perimeter diagram



Brocade MLXe16 border routers

It is a good platform, and we should praise our current Brocade MLXe16 border routers. We deployed them back in 2013, and since then, we have not experienced any major issues. I recall just a few minor ones that we have quickly addressed.

Site firewalls PA-7050 100G upgrade (Complete)

- Two boxes, each chassis has SMC, LFC, 4x NPCs with 4x100G and 8x 10G Interfaces
- According to Palo-Alto, just a few other organizations completed 100G upgrades so far
- Instructions are not complete
- Issues we experienced
  - Swap of FANs had to be done under 45 secs
  - We had to downgrade Pan-OS to make the configuration backup transferrable into new HW
  - One 100G NPC failed and needed to be RMA-ed.
  - Log Forwarding Card (LFC) required us to create a new log forwarding interface otherwise we could not boot the new system after importing the backed-up configuration

LHC Data Challenge 2024 (12 – 23 February)

- Projected data volume for HL-LHC is ~350PB / year
- The required data rates from T0 to USCMS-T1
  - Minimal scenario in 2027: 800Gbps
  - Flexible scenario in 2027: 1600Gbps
- Target for DC24 is 25%
  - CMS goal T0-> USCMS-Tier1 is 250Gbps
  - USCMS-Tier1 -> Tier2s is 250Gbps
- **DC 2024 report from CMS:**
  - Successful test for CMS both networking wise and moving data to disks and tapes
- **DC2025: 60%** (480Gbps - 1000Gbps)
- **DC2027: 100%** (800Gbps - 1600Gbps)

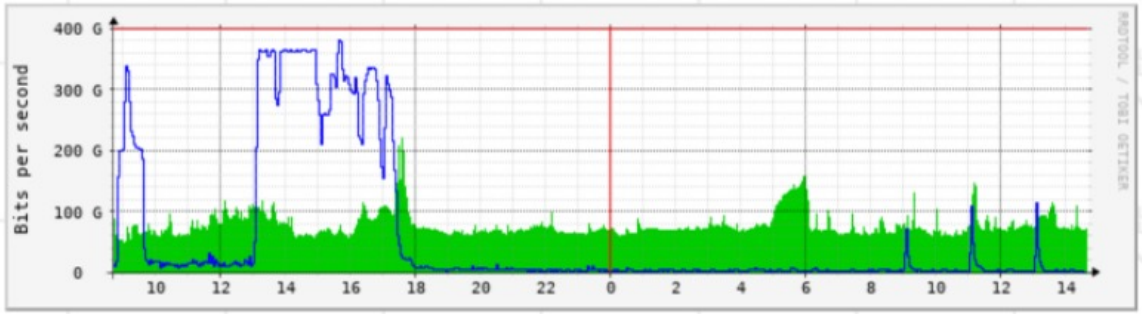
LHC Data Challenge 2024 (cont.)

Traffic Analysis for (1-min. interval) LAG50 r-cms-fcc2-2

Switch:	r-bdr-fcc2-1
Location:	FCC-2-1138
Maintainer:	net@fnal.gov
Interface Type:	Unknown Interface Type
Interface Name:	LAG50
Connected To:	r-cms-fcc2-2
Max Speed:	400.0 Gbits/s

The statistics were last updated **Saturday, 10 February, 14:41:39 CST**

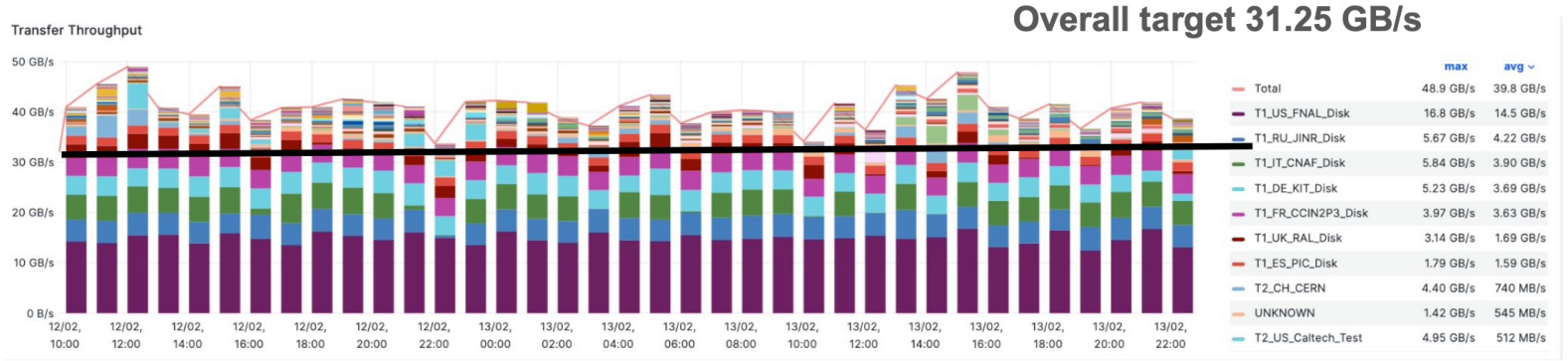
`Daily' Graph (01 Minute Average)



Max In: 222.2 Gb/s (55.5%) Average In: 78.5 Gb/s (19.6%) Current In: 68.8 Gb/s (17.2%)
Max Out: 380.2 Gb/s (95.1%) Average Out: 57.7 Gb/s (14.4%) Current Out: 1352.0 Mb/s (0.3%)

LHC Data Challenge 2024 (cont.)

Example: Day 1 and 2 T0 export



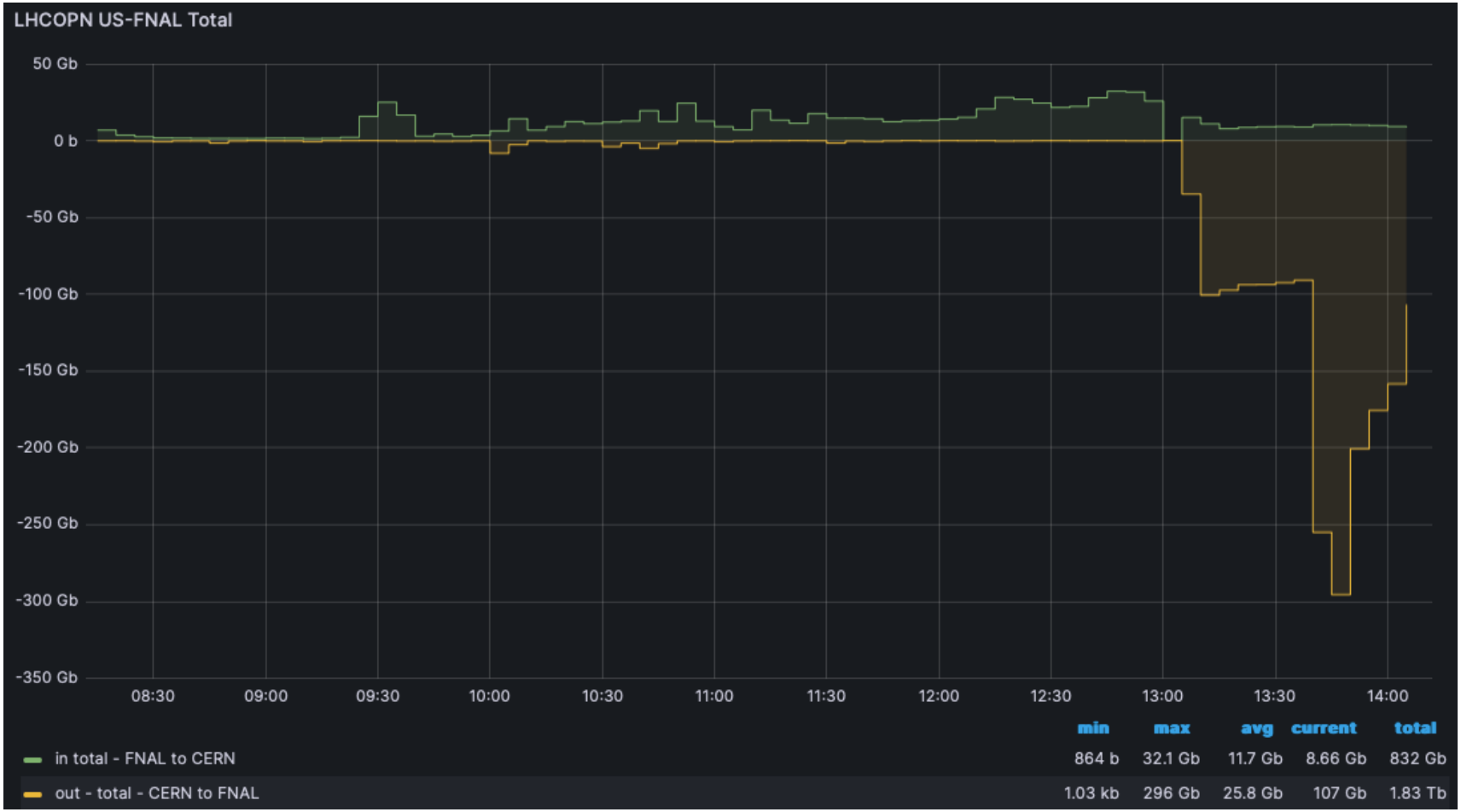
Dest	Target Rate (GB/s)	Avg rate (GB/s)
T1_DE_KIT_Disk	3.252	3.69
T1_ES_PIC_Disk	1.301	1.59
T1_FR_CCIN2P3_Disk	3.349	3.63
T1_IT_CNAF_Disk	4.227	3.90
T1_RU_JINR_Disk	3.602	4.22
T1_UK_RAL_Disk	2.513	1.69
T1_US_FNAL_Disk	13.007	14.5

We exceed the expected rates for most of the sites on the first day for T0 export

7

Fermilab

LHC Data Challenge 2024 (cont.)



Fermilab

Controlled Unclassified Information (CUI)

- Restriction on email forwarding
  - Fermi email must stay on FNAL O365 system or **G**overnment **F**urnished **E**quipment (GFE)
- Restriction on using BYOD for email
- Restriction on using Cloud Documentation Services
  - Use only FNAL O365 provided tools for originating documents
  - Can participate on other platforms if other entities are hosting the document

CUI: Problems we are trying to address

- Separation of GFE vs non-GFE equipment at the lab
  - Will help with CUI compliance solutions being investigated
- GFE devices will be connecting to regular Fermi network
- Non-GFE devices will be connecting to a new “Collaboration” network with certain restrictions

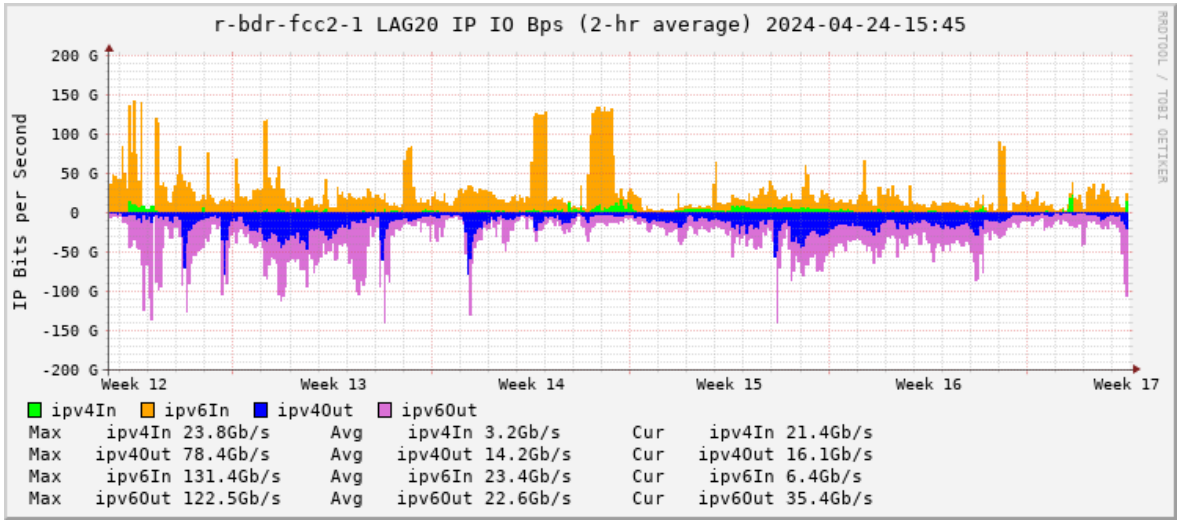
CUI: Tests and Evaluations

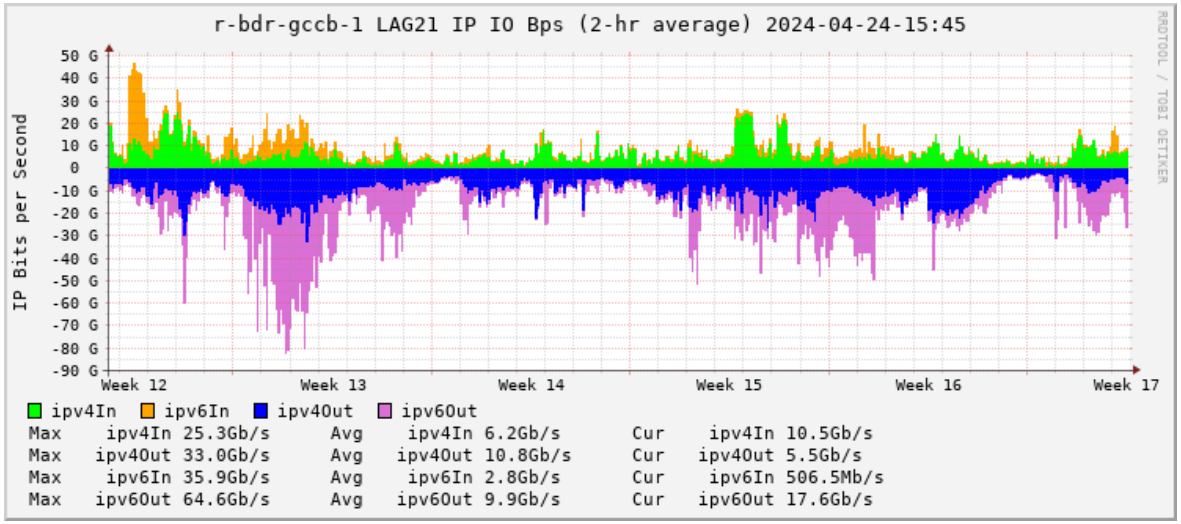
- Lots of investigations of different products (Zscaller, NetScope, Island.IO, CITRIX)
- On Networking side (GFE vs non-GFE):
- Started from VPN (Cisco AnyConnect)
  - Two certificates, one cert for device and another for user (Yubikey)
  - Issues in MacOS
    - Managed to make AnyConnect working with certs from keychain.
  - *Testing in progress*
    - *Jamf can deploy this certificate properly*
    - *Certificate cannot be copied over to another system.*

IPv6 Status: No major updates in last 6 months

- DHCPv6 deployed
- Moving VoIP subnets into IPv6
- Moving printer subnets into IPv6
- Moving security cameras into IPv6
- WiFi: A pilot with IPv6-only SSID in FCC building
  - Cyber Team has concerns for a wider deployment due to lack of MAC-based controls
- USCMS-Tier1 ~ 2300 nodes (dual-stack)
- General DC ~ 1800 nodes (dual-stack)
- Campus ~ 620 nodes (dual-stack) – Desktops, printers, VoIP phones

IPv6/IPv4 General vs Science Traffic stats





Fermilab

A serious issue with IPv6 in data center

- Intermittent loss of IPv6 connectivity between production dcache and STK (tape storage) nodes in same layer2 networks
- Broken neighbor discovery protocol, a router at the middle does not pass Neighbor Solicitation Messages
- Affected 1000+ nodes
- A temporary mitigation by preferring IPv4 over IPv6
- Worked with Cisco TAC for 1.5 months – not much progress towards a resolution
- Upgraded to next NX-OS release ~~– fixed~~. Don't observe this issue so far

A serious issue with IPv6 in data center (cont.)

IPv6 ND is not working as expected 9500 series with FX/GX/EX LCs

CSCwj73213

Customer Visible Notifications Save Bug Open Support Case

Description

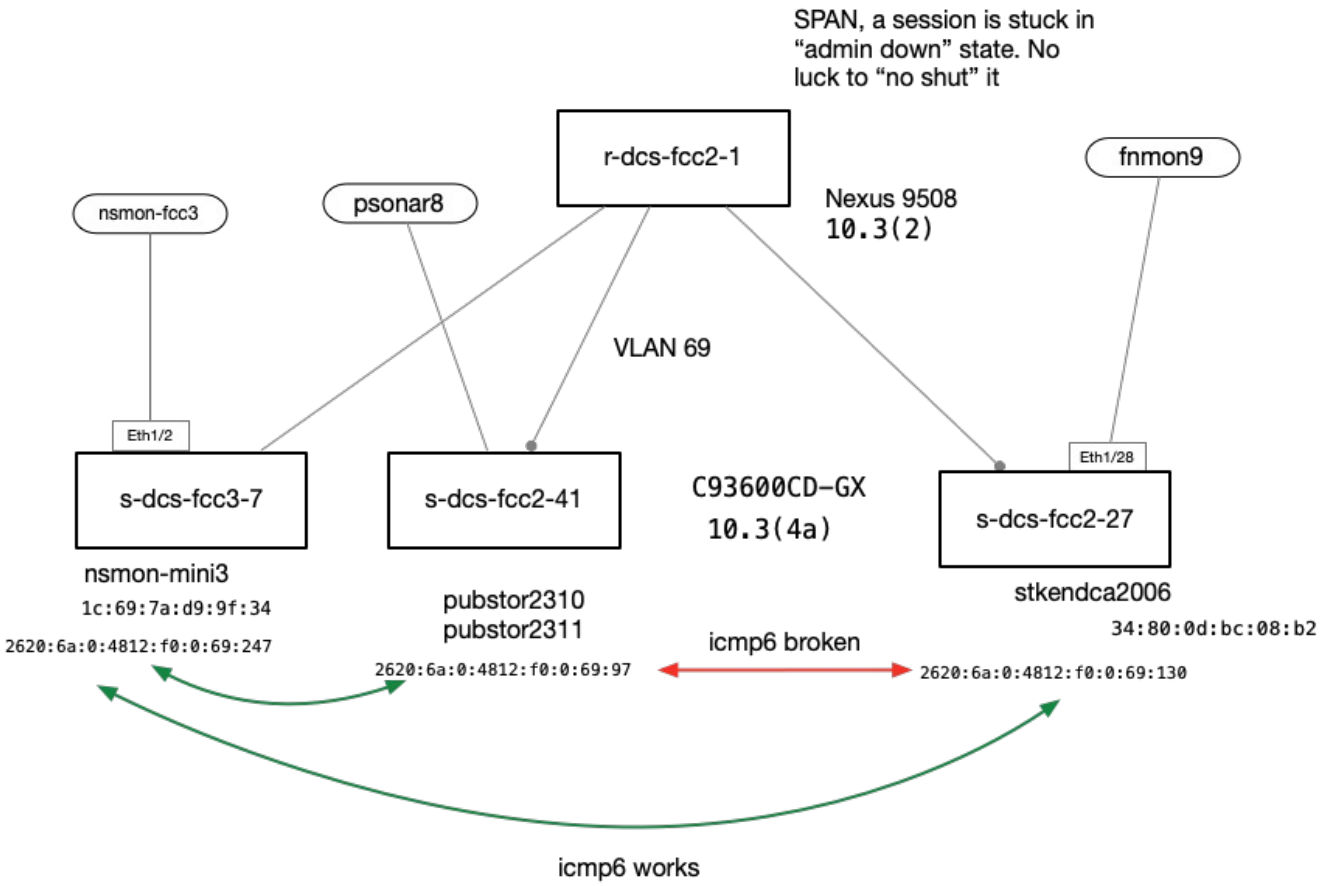
Ipv6 neighbor discovery is not working as expected resulting in no communication between the IPv6 hosts

Symptom:
Ipv6 neighbor discovery is not working as expected Nexus is not forwarding Neighbor solicitation messages.
Resulting in not connectivity between the Ipv6 Hosts.

Conditions:
IPv6 with HSRP v6 on N9500 series with FX/GX/EX LCs
N9500 chassis with FX/GX/EX LC

```
r-dcs-fcc2-1# show mod
Mod Ports             Module-Type                    Model
  Status
--- ----- ---------------------------------------- ---------------------
 --------
1    16   16x400G Ethernet Module                  N9K-X9716D-GX
 ok
22   0    8-slot Fabric Module                     N9K-C9508-FM-G
 ok
23   0    8-slot Fabric Module                     N9K-C9508-FM-G
 ok
```

A serious issue with IPv6 in data center (cont.)



Fermilab

2. IPv6 Status Updates

- What pain points and challenges has your site faced in moving to native IPv6 since our last meeting?

- ~~*NO pain, it is mostly a fun !!!!*~~
- *A lot of pain due to an intermittent IPv6 issue in DC that I described earlier*
- ~~*No IPv6 specific challenges – Typical ones with software bugs, different platforms compatibility and implementations and etc…*~~

1. Site-specific information

- Major upgrades, accomplishments, or other changes since your previous ESCC report
- New initiatives, testing, product evaluations, or other items that might interest the ESCC community
- Future opportunities/challenges in the remainder of FY24 or beyond

- *Border routers and Site Firewall PA-7050 upgrades*

2. IPv6 Status Updates

- Do you have a category of devices that are ‘mostly native’ IPv6, where just about everything runs natively but still requires v4 for certain functions based on vendor limitations? Please share details that would be of benefit to the group.
- *DHCPv6 is not yet officially in production due to a Cyber team’s concern of losing MAC-based controls. Basically, most desktops with SLAAC assigned IPv6 addresses, except a pilot group with DHCPv6 assigned addresses, still need IPv4 for DNS.*
- *An interesting case: Logitech Videoconference Controller TAP-IP/ Android based for Zoom rooms - NO plans for DHCPv6 support, only SLAAC. Currently, it does not allow to type DNS servers manually – promised to fix, but still no plans for DHCPv6*

2. IPv6 Status Updates

- What is your overall staff readiness to support, troubleshoot, complete cyber forensics, etc, in native IPv6? Has staff training been provided? Are there plans to do so?
  - Have IPv6 enabled for 10+ years. I believe that we have developed some proficiency
  - Distribute IPv6 related tasks amongst all members of Network Services
- What tools are available to identify dual stack devices moving from v6 to v4? Would you be aware if v6 stopped working for a given server?
  - *Not at this point. We do monitor some services, interfaces by names, but these are not yet in IPv6 environment*

3. Other Topics of Interest to the Community

- What are you using for MFA? How has your experience been with PIV and PIV-I? What percentage of your staff is using a PIV based solution? Do you expect to roll out MFA across the board (if not in place already)?
  - Yubikey/RSA -
    - Network Services staff 100%
    - System administrators – 100%
    - We are supposed to move to HSPD-12 PIV within a year

Questions ?

HL-LHC Network bandwidth needs per T1

T1	%ATLAS	%CMS	% Alice	% LHCb	ATLAS+CMS Network Needs (Gbps) Minimal Scenario in 2027	Alice Network Needs (Gbps) Minimal Scenario in 2027	LHCb Network Needs (Gbps) Minimal Scenario in 2027	LHC Network Needs (Gbps) Minimal Scenario in 2027	LHC Network Needs (Gbps) Flexible Scenario in 2027
CA-TRIUMF	10	0	0	0	200	0	0	200	400
DE-KIT	12	10	21	17	450	80	70	600	1200
ES-PIC	4	5	0	4	180	0	20	200	400
FR-CCIN2P3	13	10	14	15	450	60	60	570	1140
IT-INFN-CNAF	9	15	26	24	480	110	100	690	1380
KR-KISTI-GSDC	0	0	12	0	0	50	0	50	100
NDGF	6	0	8	0	110	30	0	140	280
NL-T1	7	0	3	8	140	10	30	180	360
NRC-KI-T1	3	0	13	5	50	50	20	120	240
UK-T1-RAL	15	10	3	27	490	10	110	610	1220
RU-JINR-T1	0	10	0	0	200	0	0	200	400
US-T1-BNL	23	0	0	0	450	0	0	450	900
US-FNAL-CMS	0	40	0	0	800	0	0	800	1600
(atlantic link)					1250	0	0	1250	2500
Sum	100	100	100	100	4000	400	410	4810	9620

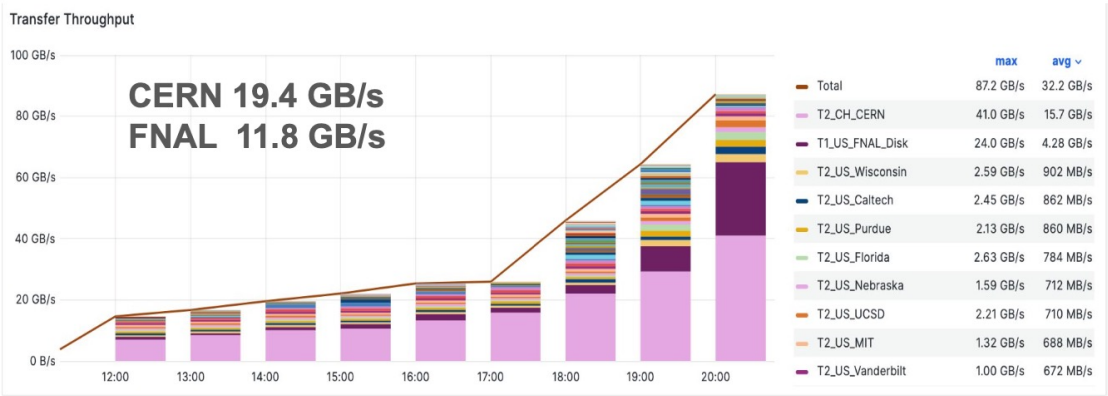
Table 1: network bandwidth needs per T1 (or region)

HL-LHC Data challenge target rates

T1	LHC Network Needs (Gbps) Minimal Scenario in 2027	LHC Network Needs (Gbps) Flexible Scenario in 2027	Data Challenge target 2027 (Gbps)	Data Challenge target 2025 (Gbps)	Data Challenge target 2023 (Gbps)	Data Challenge target 2021 (Gbps)
CA-TRIUMF	200	400	100	60	30	10
DE-KIT	600	1200	300	180	90	30
ES-PIC	200	400	100	60	30	10
FR-CCIN2P3	570	1140	290	170	90	30
IT-INFN-CNAF	690	1380	350	210	100	30
KR-KISTI-GSDC	50	100	30	20	10	0
NDGF	140	280	70	40	20	10
NL-T1	180	360	90	50	30	10
NRC-KI-T1	120	240	60	40	20	10
UK-T1-RAL	610	1220	310	180	90	30
RU-JINR-T1	200	400	100	60	30	10
US-T1-BNL	450	900	230	140	70	20
US-FNAL-CMS	800	1600	400	240	120	40
(atlantic link)	1250	2500	630	380	190	60
Sum	4810	9620	2430	1450	730	240

DC24: Any data Anytime Anywhere

Day 8- AAA (CERN/FNAL to T1s, T2s)



Grouped by source rse

Rate was not achieved immediately because the injector tool used small files as input

When fixed, we achieved the rates

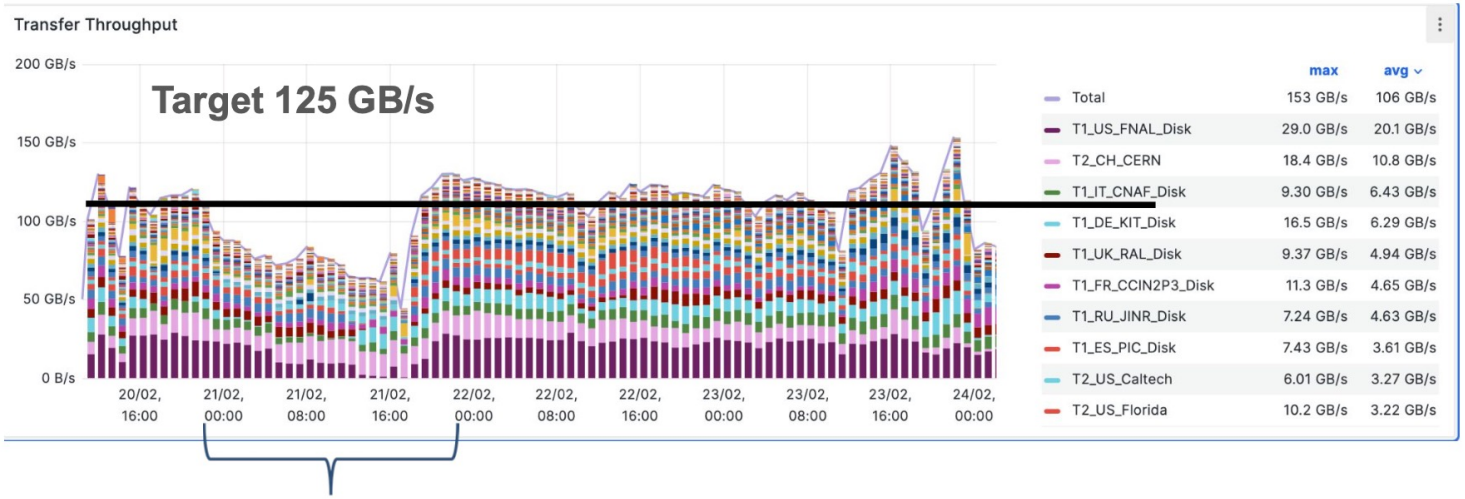
[dashboard]

10

Fermilab

DC24 Day 9 – 10 Maximum rate

Day 9-10 maximum rates



February 21st: Rucio could not handle deletions due to large backlog had to act to achieve target rates the next days

Dashboard

11

Fermilab

DC24 CMS Summary

Summary - Milestones Achievements

- **Successful test for CMS and other VOs**
- Tests our system to the maximum
- First challenge with token authentication for 25 sites!
- Had to monitor and fine tune parameters- we understand our system better
- Started to discuss the paper structure
- A lot post mortem analyses will follow, GDC meeting and DC workshop