DOMA Data Challenge debrief

L. Duflot

Data Challenge

- Semaine du 4 octobre
- ATLAS et CMS ont injecté du trafic supplémentaire dédié
  - T0 – T1, T1 -T2
- Ont du “faire de la place” sur les disques pour cela
  - Pas de grandes volumétries libérées (surtout CMS)
  - Mode alternatif transfert / effacement donc pas de charge constante

- ATLAS:
  - production RSEs and nuclei T2s (+1 ex)
  - datasets/files selected by (in collaboration with) DDM
- CMS:
  - “_Test” RSEs and T2s pre-selected by CMS
  - datasets/files selected by transfer team (Data Management/CompOps)

Contraintes opérationnelles plus fortes pour CMS, démarrage plus compliqué

T1s second look

Alessandra GDB

T1	Minimal Scenario 2027	Flexible scenario 2027	Minimal scenario ingress/egress targets 2021	Ingress (hourly avg/max)	Egress (hourly avg/max)	comments
CA-TRIUMF	200	400	10/10	17/49	25/70	ok
DE-KIT	600	1200	30/30	33/77	52/143	ok
ES-PIC	200	400	10/10	11/18	11/17	ok
FR-CCIN2P3	570	1140	30/30	35/70	41/80	ok
IT-INFN-CNAF	690	1380	30/30	25/57	43/87	sum ok
KR-KISTI-GSDC	50	100	0	0	0	Alice T1
NDGF	140	280	10/10	26/49	27/82	ok
NL-T1 (NIKHEF)	-	-	10/10	10/37	12/53	ok
NL-T1 (SARA)	180	360	10/10	13/51	16/79	ok
RU-JINR-T1	200	400	10/10	11/26	12/31	ok
RU-NRC-KI-T1	120	240	10/10	9/18	12/34	sum ok
TW-ASGC	-	-	10/10	8/16	10/13	explain
UK-T1-RAL	610	1220	30/30	16/41	25/43	explain
US-FNAL-CMS	800	1600	40/40	16/49	19/49	explain
US-T1-BNL	450	900	20/20	29/75	38/117	ok
Atlantic link	1250	2500	60/60			
Sum	4810	9620	240/240	259 avg	343 avg	ok

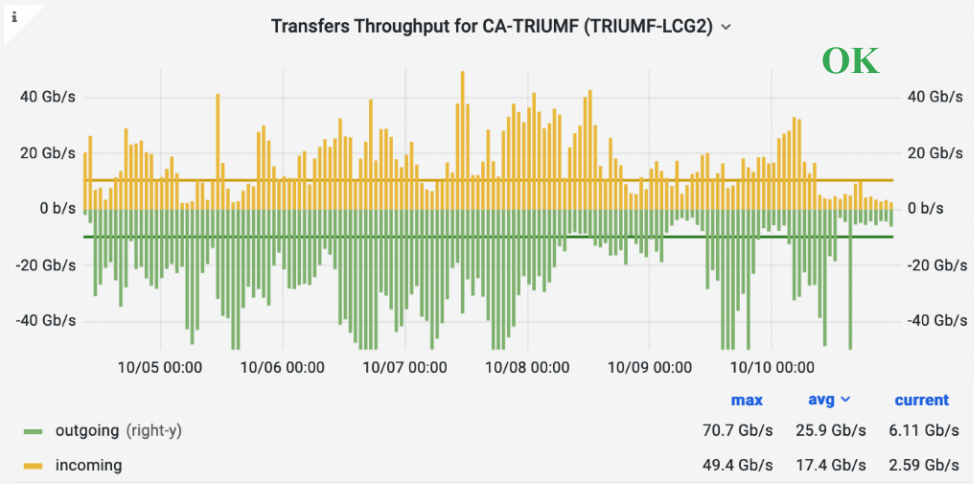
But du DC : 2 x minimal scenario

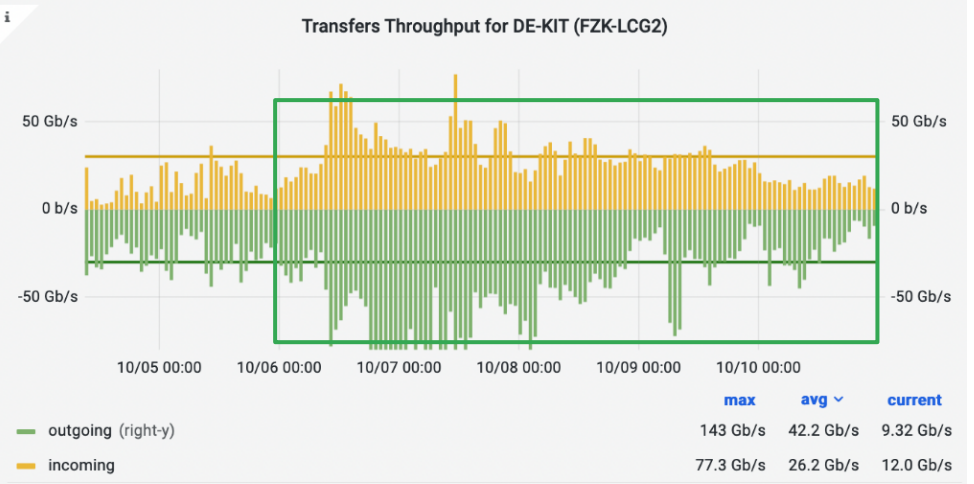
T1s further look

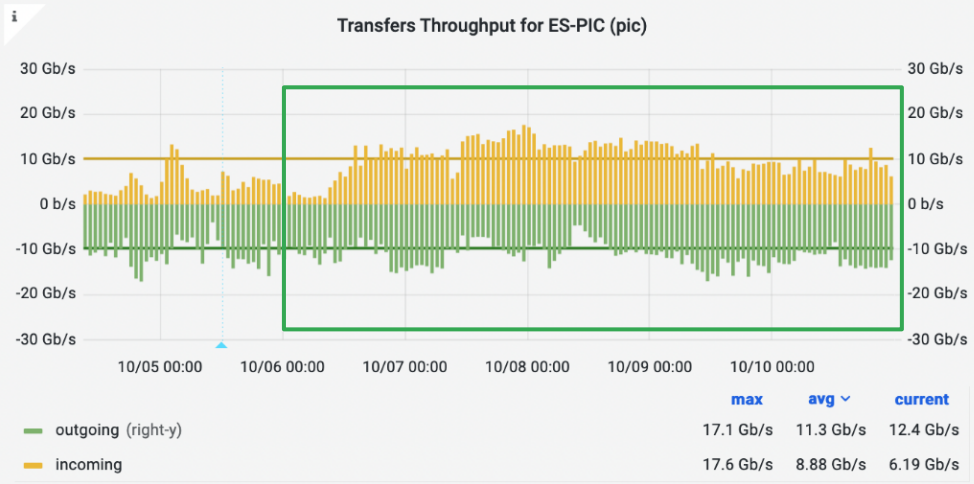
- TW/CNAF/RRC-KI
  - Close to the target for ingress and ok for egress
  - Sum of ingress + egress exceeds sum of targets
  - TW is far away, small and doesn't have tape, not worrying
  - CNAF, RRC-KI ingress not clear
    - Could be not enough data sent their way? Not enough production traffic?
- FNAL
  - Not enough external production traffic
  - Traffic injected wasn't enough to backfill 40Gbs+40Gbs
    - Badly affected by being full
- RAL
  - HTTP-TPC known performance problems on ECHO
    - Particularly affected ingress traffic
    - xrootd-ceph plugin being fixed
  - Network being upgraded but what is there is sufficient
    - Seen in production with gridftp
  - Repeat tests when echo code is fixed?

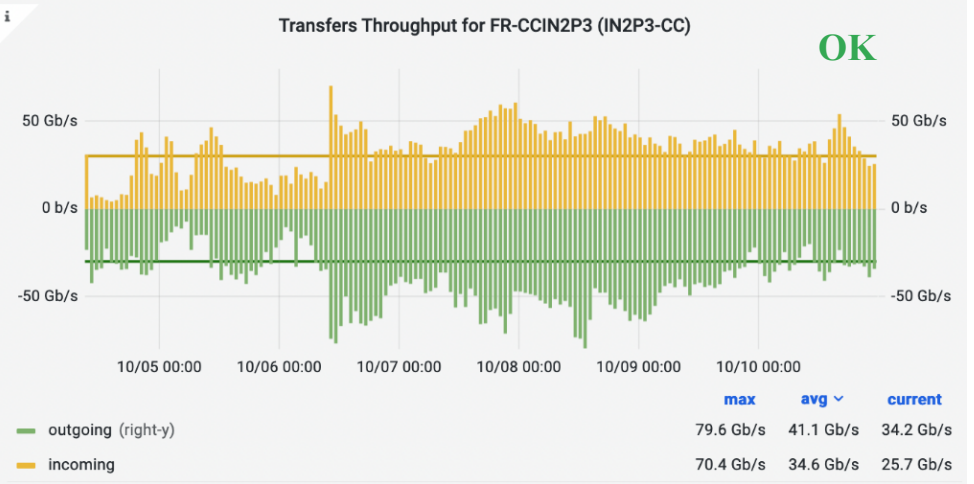
Sites Windows 1/4

Alessandra GDB



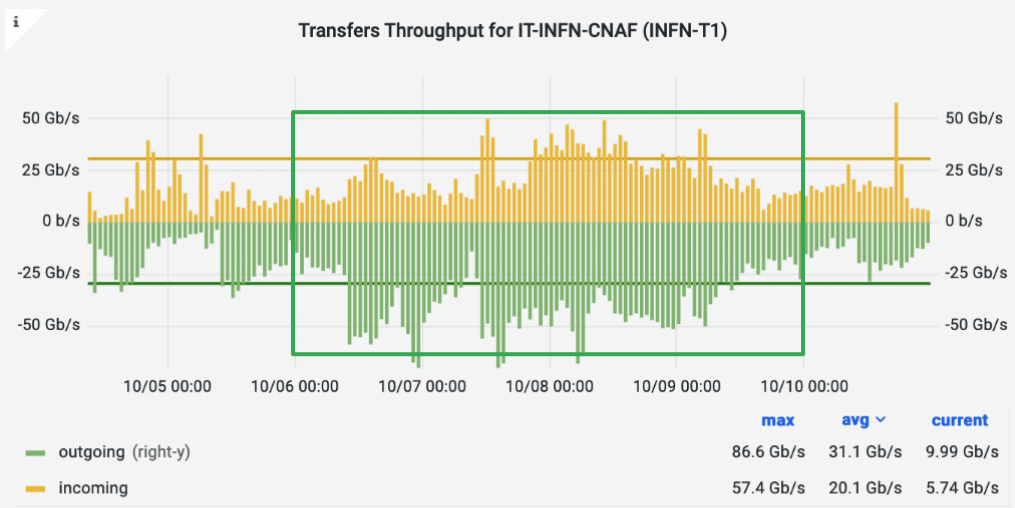


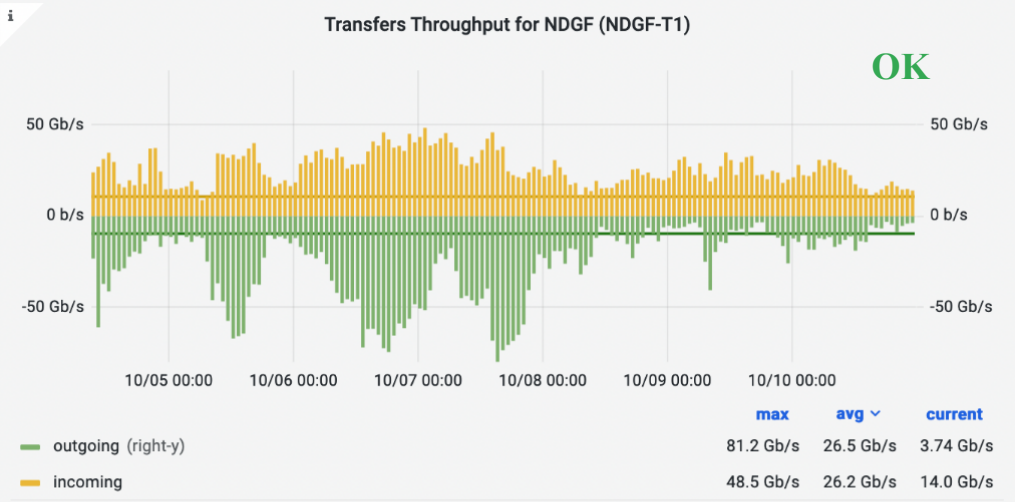


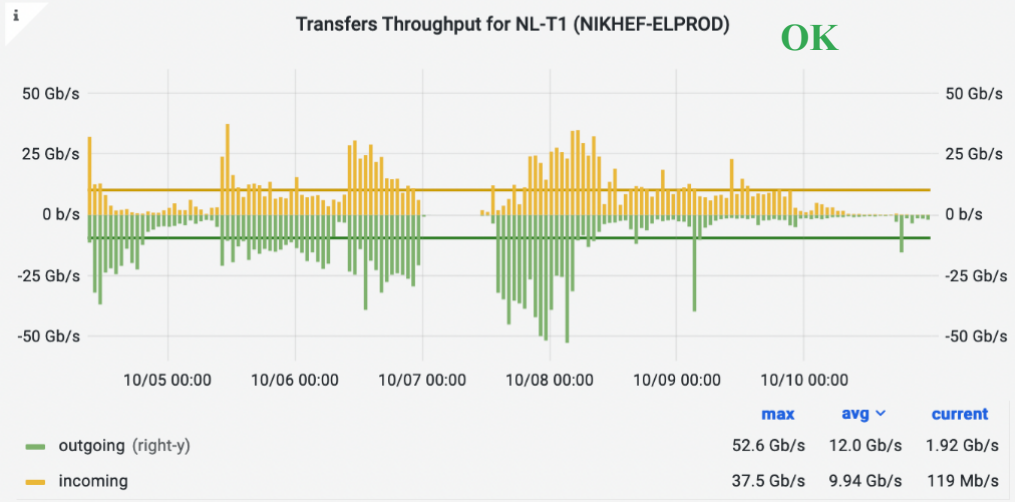


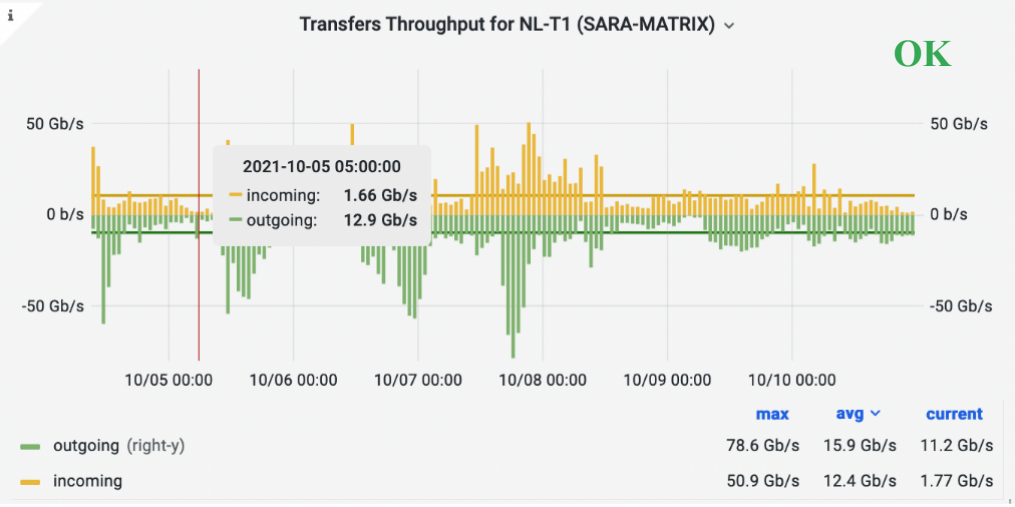
Sites Windows 2/4

Alessandra GDB



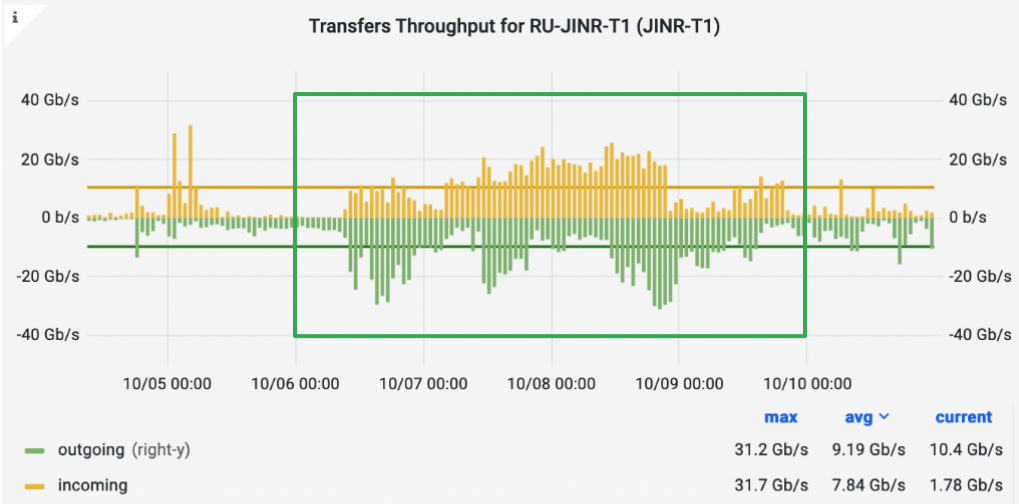


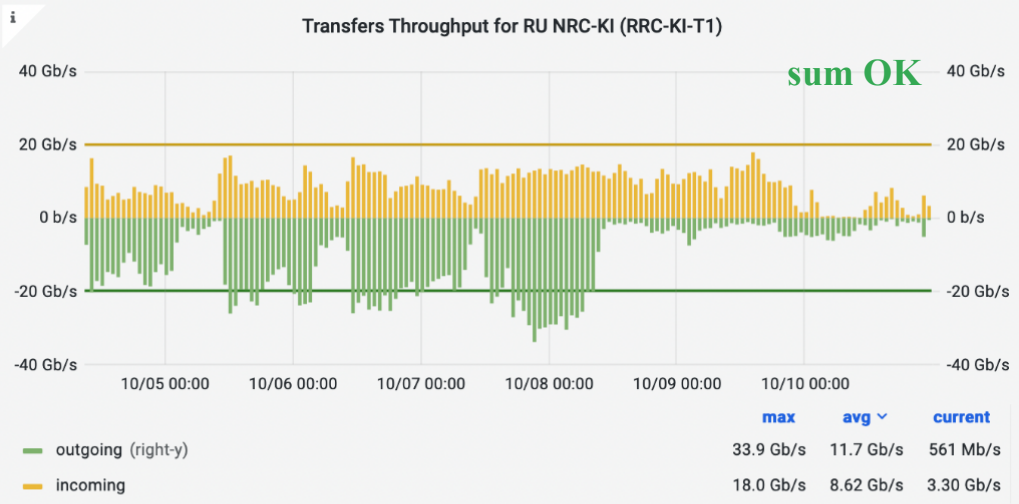


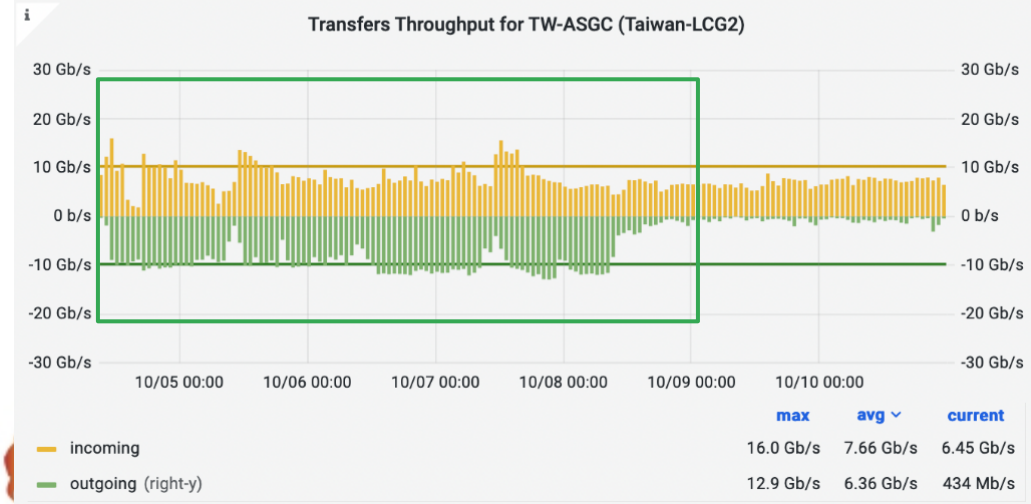


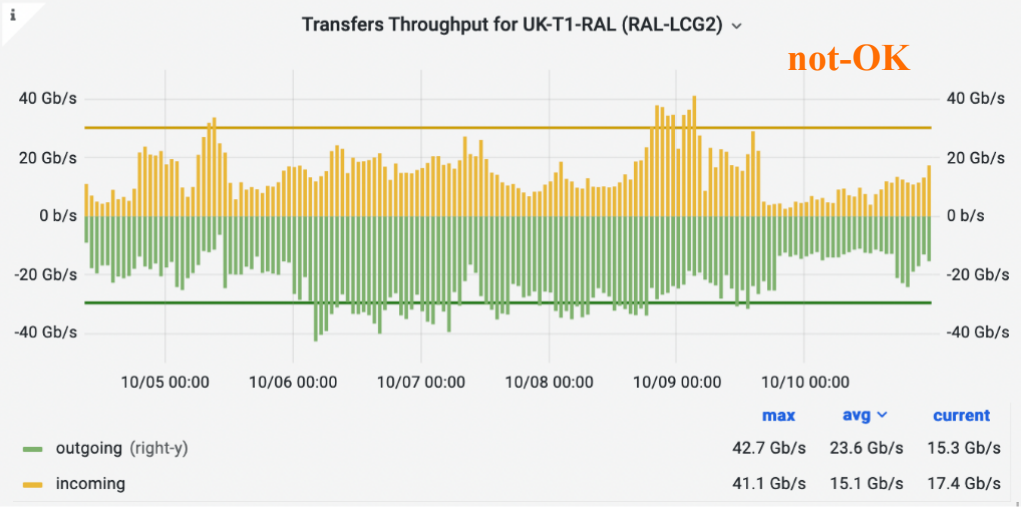
Sites Windows 3/4

Alessandra GDB



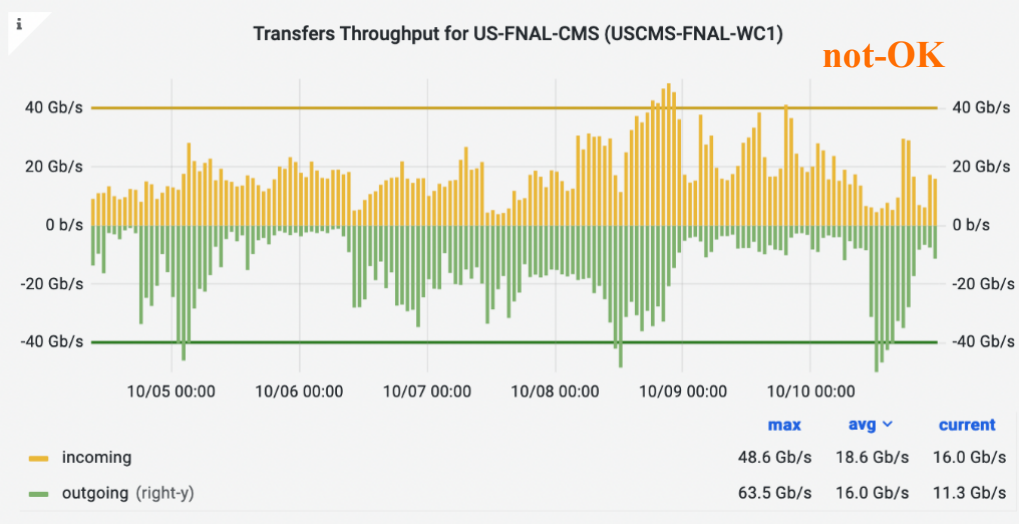


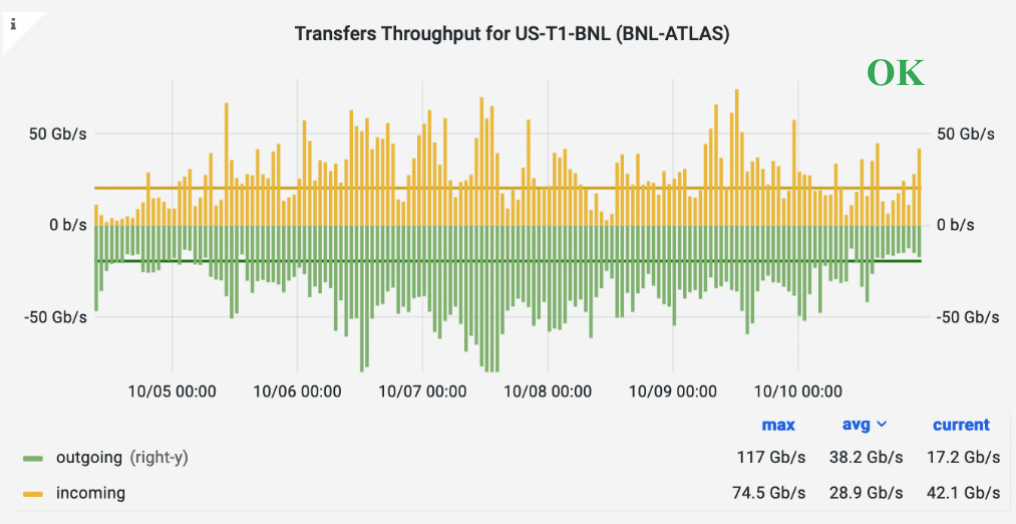




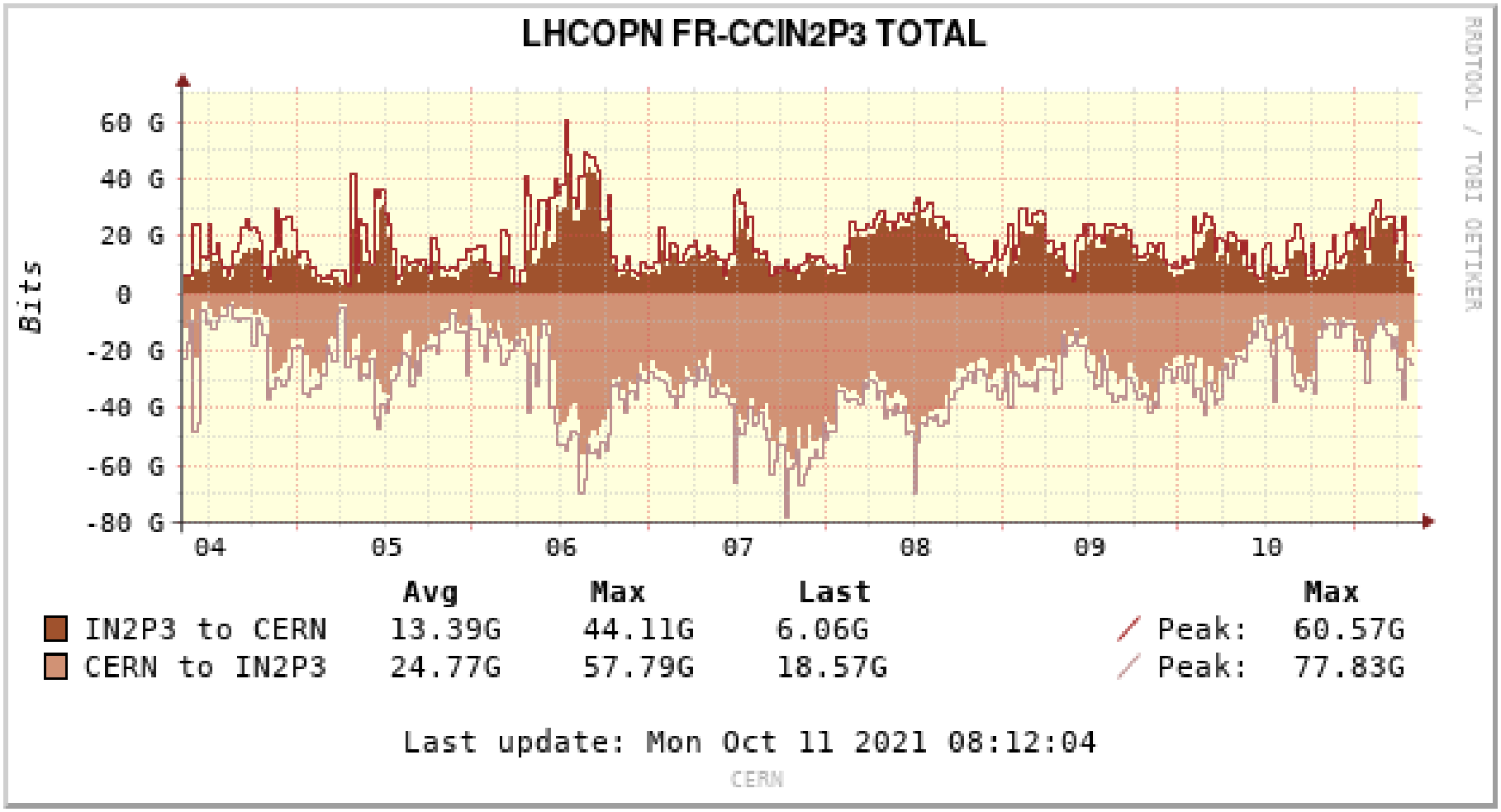
Sites Windows 4/4

Alessandra GDB



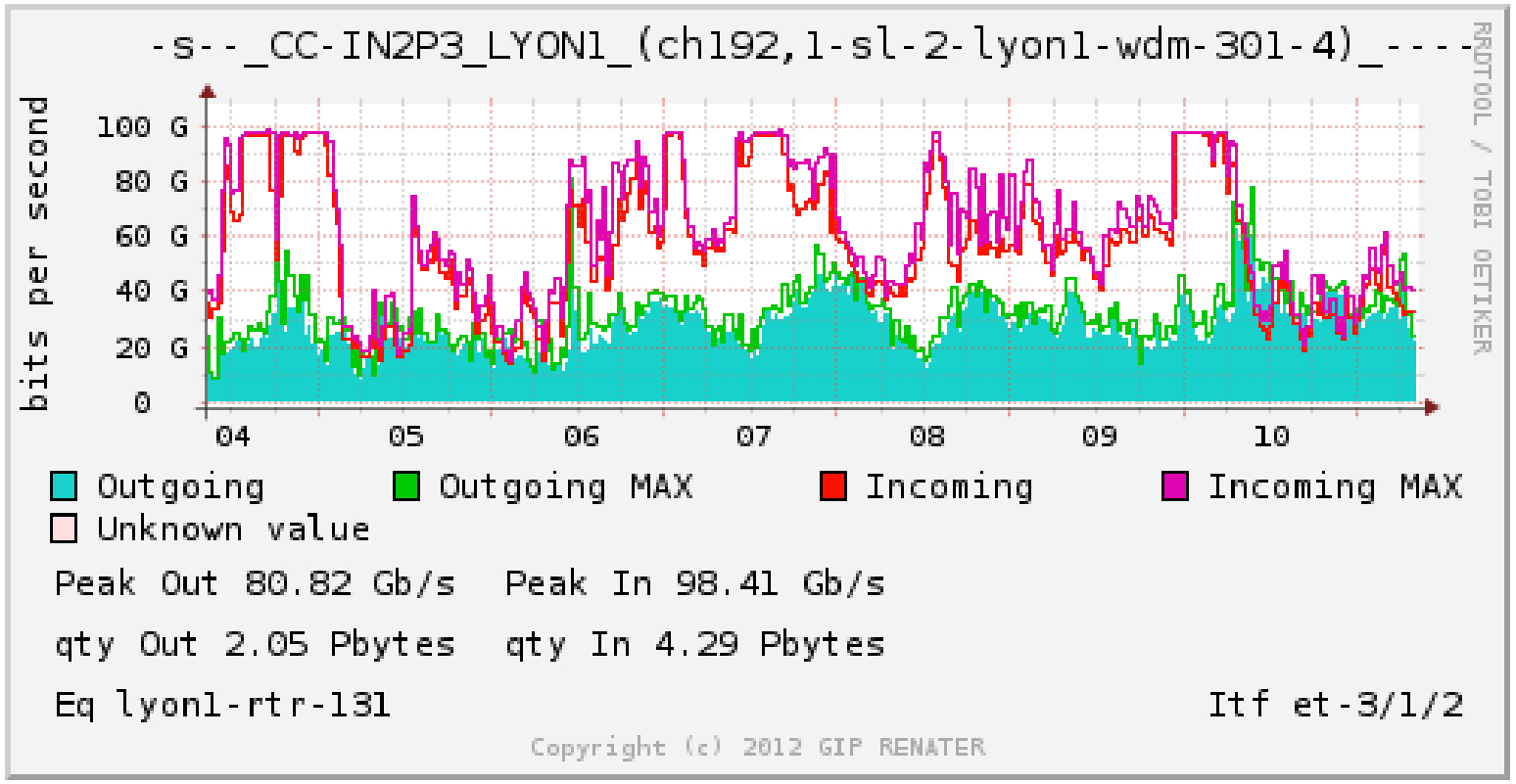


Data Challenge en France: CC-IN2P3



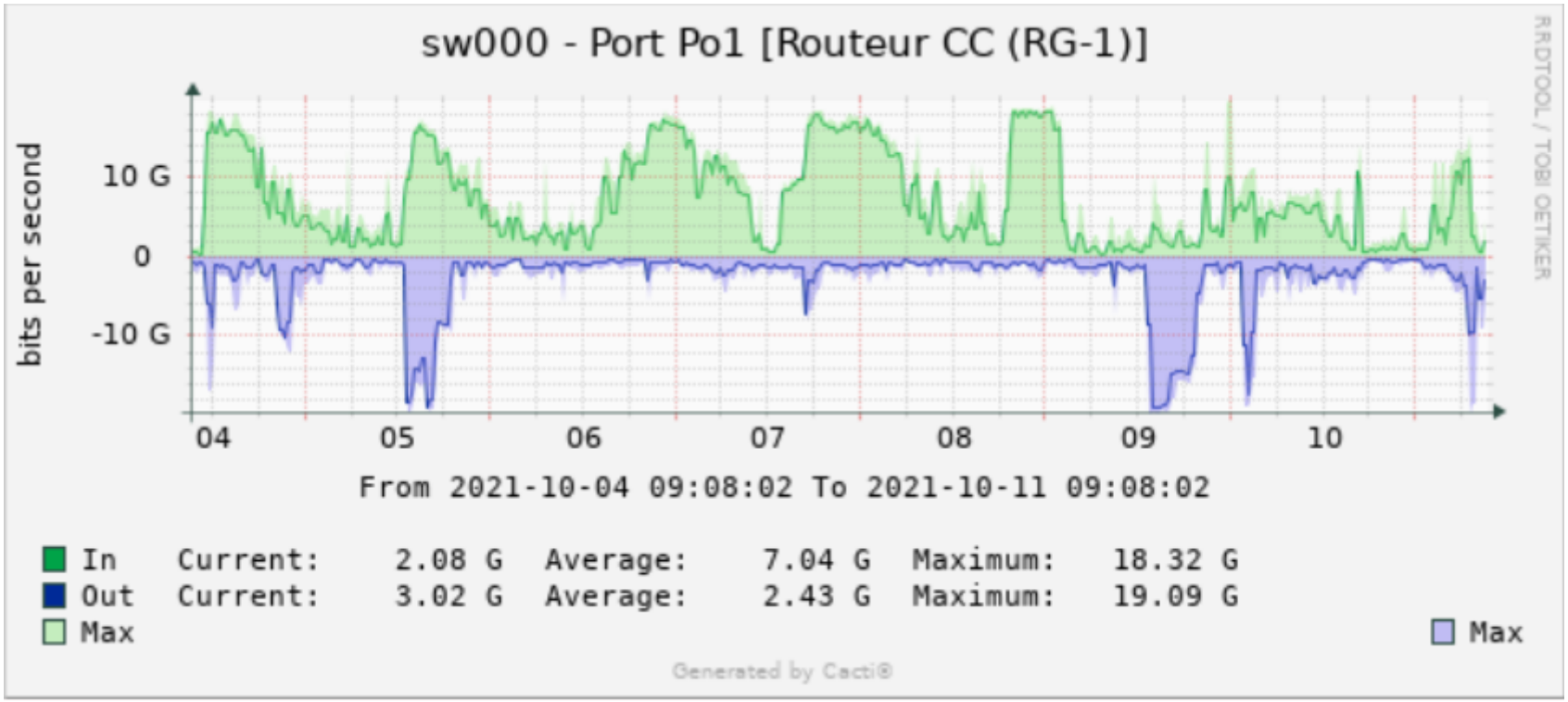
Pas assez de pression de transfert: on ne sature jamais le lien 100G
Pourquoi encore autant de traffic ipv4?

Data Challenge en France: CC-IN2P3



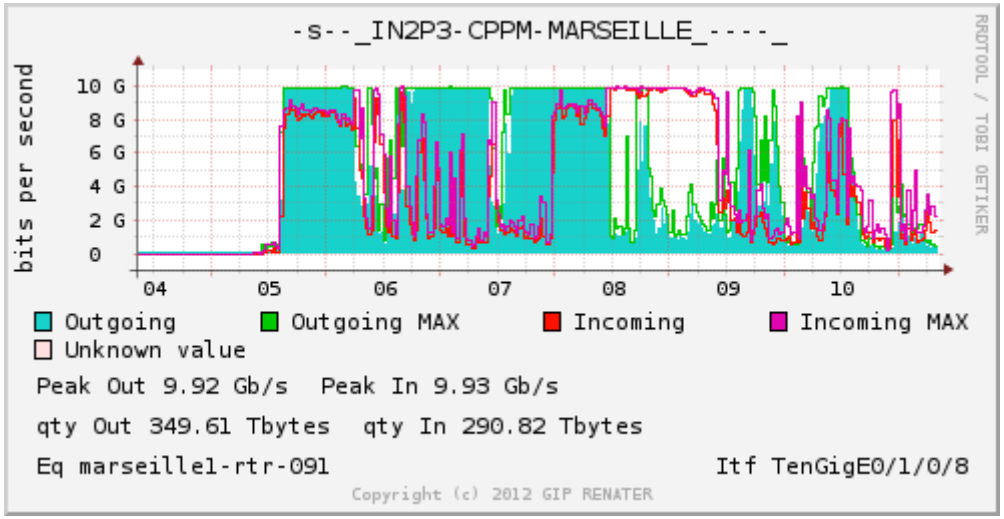
On sature régulièrement le lien 100G “Sud”

Data Challenge en France: LAPP



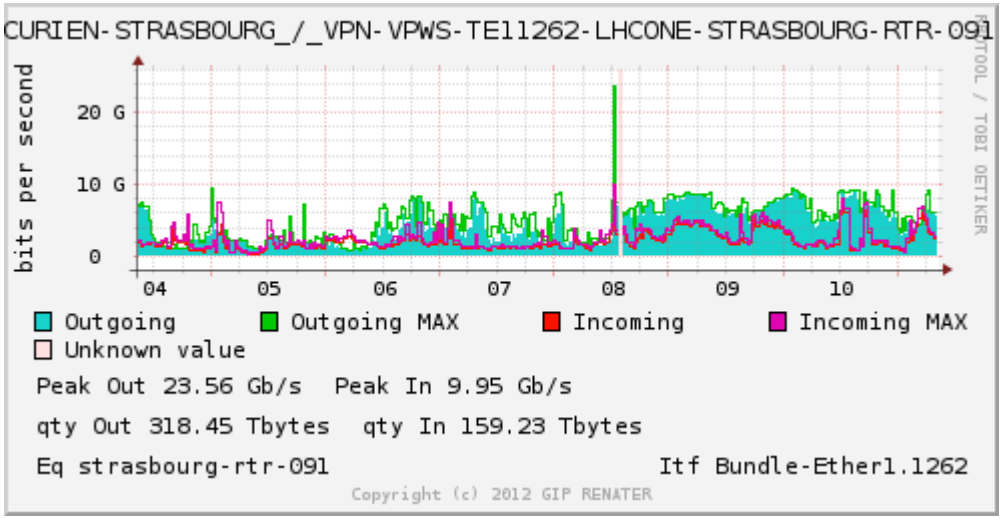
Proche des 20G plusieurs fois pendant le challenge
Ce traffic s'ajoute à celui pure CC sur LHCOne Sud-CC

Data Challenge en France: CPPM et IPHC

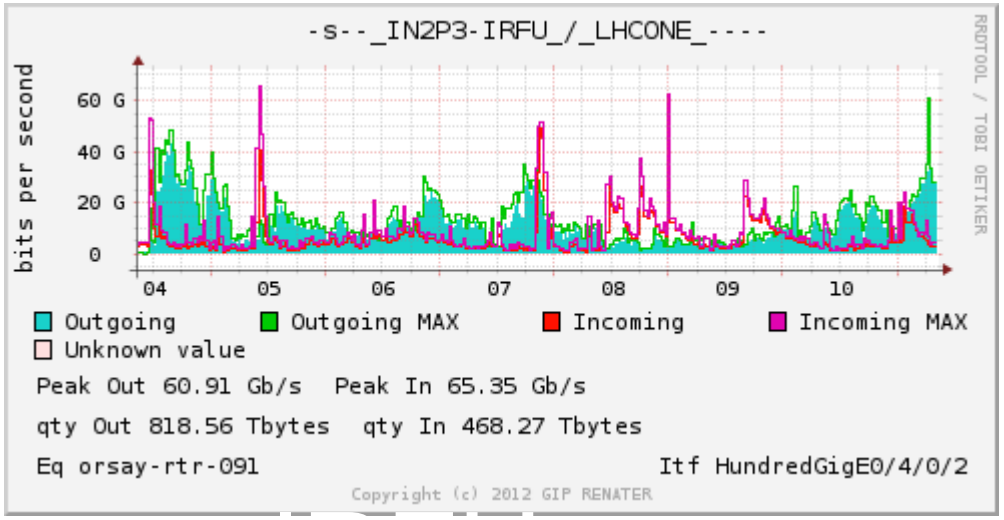


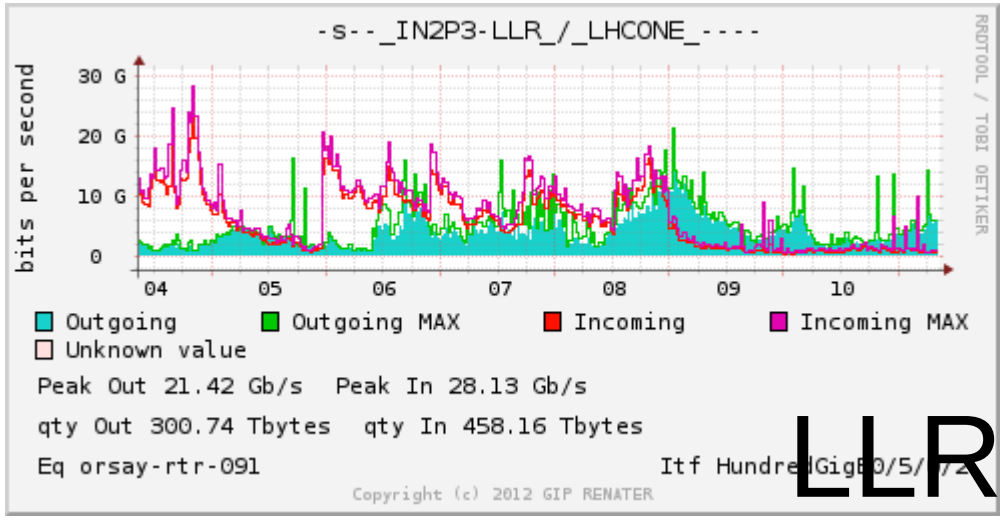
CPPM: clairement limité par le réseau (10G)

IPHC: pas assez de pression de transfert?
Espace ‘Test’ sur un seul serveur ?



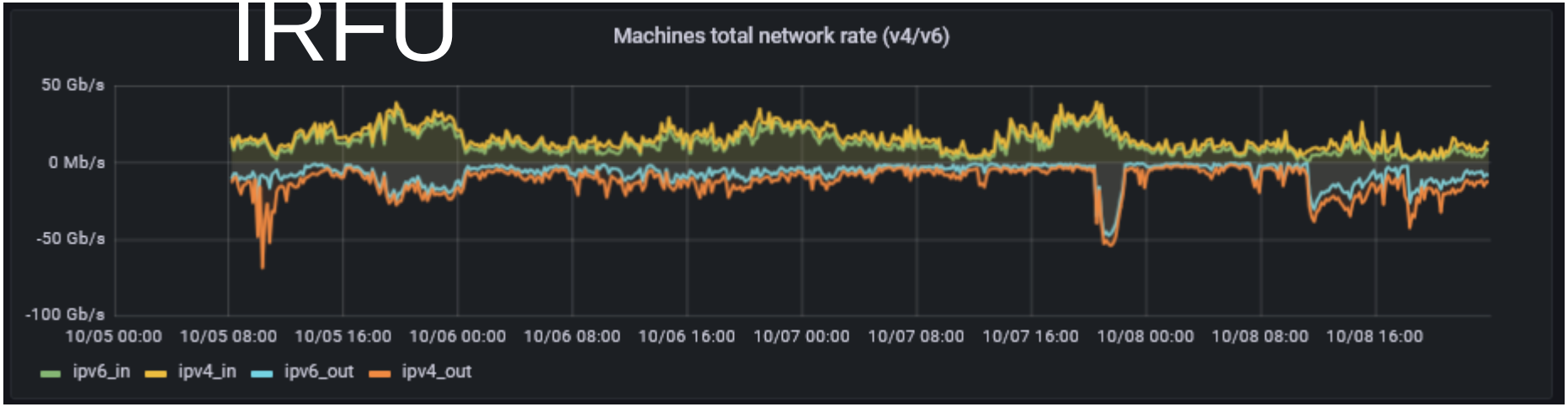
Data Challenge en France: IRFU et LLR





LLR

IRFU



Très loin des capacités théoriques du réseau: 100G+100G → 100G vers Paris
Pas assez de pression de transfert ?

Dommage n'a pas permis d'évaluer les limites du système (réseau? Stockage?)

Data Challenge en France: debrief

- À part des problèmes hors de notre contrôle au début (CPPM, IRFU), le challenge s'est bien passé en France
- LHCone Sud saturé mais operations sans trop de problèmes:
  - Un peu plus de timeout de transferts observés au CC
  - Pas de grosse charge sur le stockage au CC
- CPPM clairement limité avec 10G
- CMS a eu plus de mal à démarrer le challenge et a moins transféré en volume ensuite:
  - pas assez de pression pour bien tester IPHC et LLR?
  - Spacetoken de test sur un seul serveur disque?
- Pas assez de retours sur le stockage (charge..) ou "doors" du CC

Site Monitoring

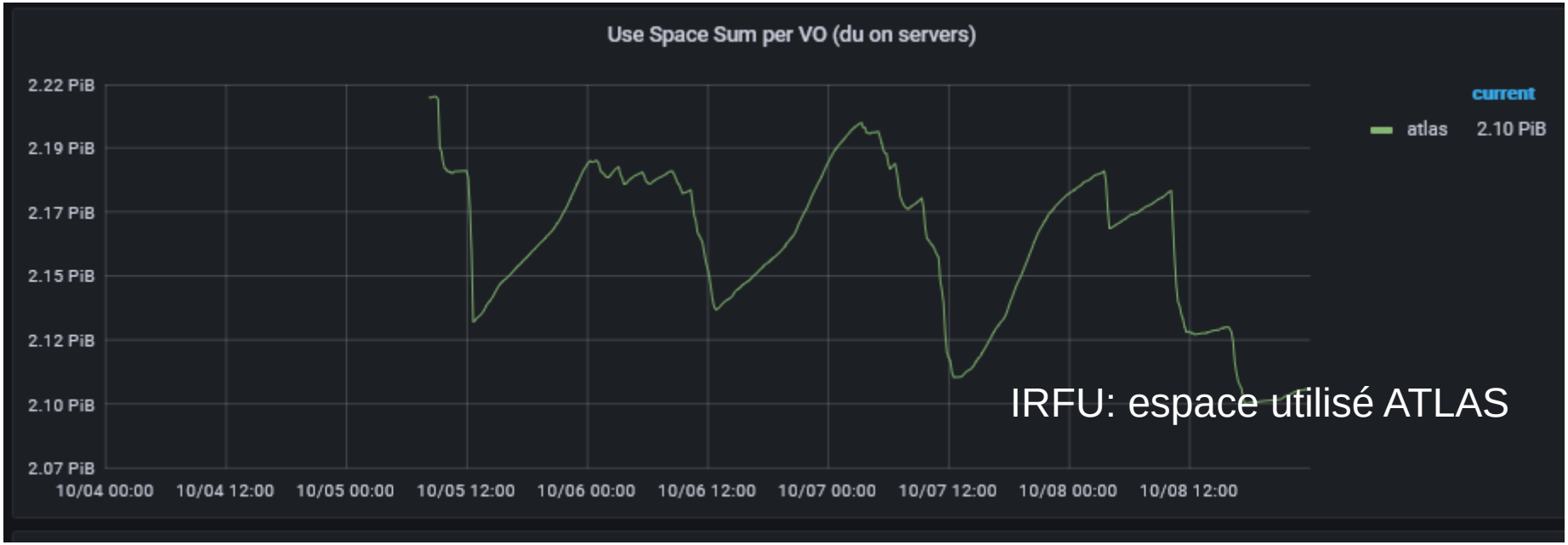
- Difficult already said, but
  - Network and tape challenge both missed it
  - Work to create a json for tape monitoring and push it in MONIT already done
    - Work could be brushed up
    - Same mechanism could be used for network
  - Belonging to LHCONE-LHCOPN makes at least WLCG traffic identifiable
    - Do T1s have similar monitoring?
    - Few T1 inspected have a good match between FTS plots and LHC* network plots
  - Network people looking at other more general monitoring like NetSage
    - Not the time series we are used to though more network maps and throughput matrices
  - Avoid extending to T2s until there is a more defined monitoring

Monitoring general

- Priority must be given to work already ongoing on experiment common dashboard
  - General comments on working on existing data to add xrootd and FTS traffic to a common dashboard already presented
- More details in the post-mortem
- And tomorrow dedicated presentation at the WLCG Ops Coord (Borja and Rizart)

Monitoring: comment aller plus loin?

- Idealement nous devrions avoir les outils pour facilement collecter les charges réseau, service (transfert, stockage) et serveurs de stockage
  - # de connexions, load des serveurs, DB DPM/dCache etc...



Monitoring: comment aller plus loin?

- Réseau:
  - Alarmes (perfsonar)
  - Comment intercepter le flow marking (paquets UDP envoyés en // du transfert) et packet marking (ipv6 header) et l'utiliser pour monitorer le traffic d'une site?
    - Des collègues ASR spécialisés réseau des labos pourraient-ils aider?
  - Intérêt pour centraliser le collecte sur un site dédié e.g. netSAGE ?

Research Network Technology WG - update

Main current activity: understand scientific traffic flows in detail
More participants joined the activity, now ~90
New website https://www.scitags.org/

Proposed two approaches:
- Flow marking using UDP fireflies (works for both IPv4 and IPv6)
- Packet marking using IPv6 flow label and/or header extensions

Very good progress on the firefly part. Implemented in few sites, being tested with the ongoing WLCG data challenge

Developed a reference implementation: flowd

https://indico.cern.ch/event/1022426/contributions/4492285/attachments/2326491/3963219/Research%20Network%20Technical%20WG%20update.pdf

15

NetSAGE to monitor LHC data

NetSAGE is a tool to make network monitoring stats easily accessible and understandable
It can ingest data from many sources, like SNMP counters, net/s/flow, perfSONAR…
It is committed to privacy, fully GDPR compliant
NetSAGE instances can also be installed locally

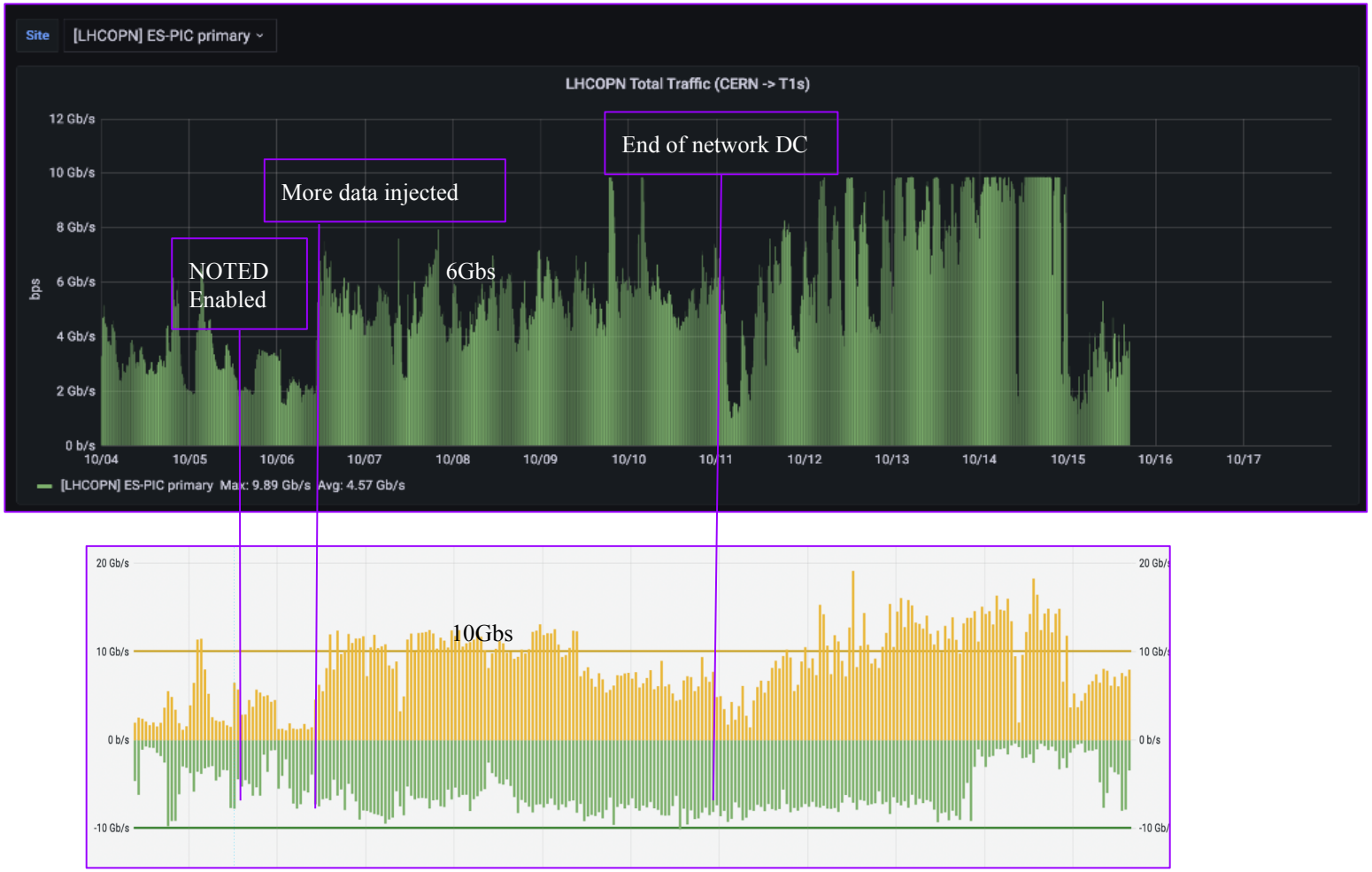
NetSAGE instance at Indiana currently has access to few sources, like NEA3R, GEANT, TransPAC, SingAREN. Not enough for a clear view.
The NetSAGE team would like to get feedback on existing features and new needed ones
The NetSAGE team offers to collect and show stats for LHCONE. It would need access to data, though. They would also need IP mapping to better understand the traffic. The CRIC database could help.

Action: discuss the possibility to use NetSAGE for LHCOPN and LHCONE monitoring

https://indico.cern.ch/event/1022426/contributions/4453394/attachments/2326591/3963394/DC LHCOPN 2021_10_12.pdf 13

Backup

CERN-PIC NOTED



- NOTED (SDN) enabled between 5/10 13:00 and 19/10 13:00 to increase PIC network capability
  - When LHCOPN saturated NOTED controller added LHCONE channel 6Gb/s -> 10Gb/s T0-T1
  - NOTED uses FTS traffic to estimate need of network and orchestrate traffic
- Example of exercise that can be carried out in future challenges
  - Will need more integrated monitoring to make more than qualitative comparisons