Přenosové služby pro spojení se zahraničními laboratořemi

Jiří Chudoba, Marek Eliáš, Lukáš Fiala, Tomáš Kouba, Jan Švec

elias@fzu.cz

Nové typy přenosových služeb a jejich aplikace, 28. 05. 2014

Fyzikální ústav AV ČR, v. v. i. (FZÚ)

Obsah:

1. VS FZÚ — prehľad
2. prenosy dát do zahraničia
3. záťaž a využitie prenosových trás
4. IPv6 a jumboframe prenosy
5. problémy s centrálnym routrom

Výpočetní středisko FZÚ

Výpočty

- fyzika vysokých energií (HEP), experimenty:
  - ATLAS, ALICE (CERN)
  - D0, NOVA (FNAL)
- astročasticová fyzika
  - Auger (Pierre Auger Observatory)
  - CTA (Cherenkov Telescope Array)
- fyzika pevných látok

Clustery

- Golias: 4100 CPU cores, Ethernet (HEP + astro experimenty)
- Dorje: Altix ICE8200, 1.5 racku, Infiniband, 512 CPU cores
- Luna2013: SUPERMICRO X9DRW-iF, 1.5 racku, Infiniband, 800 CPU cores

- požiadavky na prenosové služby hlavne pre cluster Goliáš

Organizácia clusteru Goliáš

- 288 výpočtových uzlov (WN)
- hlavné storage systémy:
  - DPM: 13 datových uzlov, 2 PB (ATLAS, AUGER)
  - XROOTD: 4 datové uzly, 900 TB
  - NFS: 70TB
  - SAM cache: 6TB
  - dáta niekedy smerujú priamo na výpočtové uzly

Externé kapacity

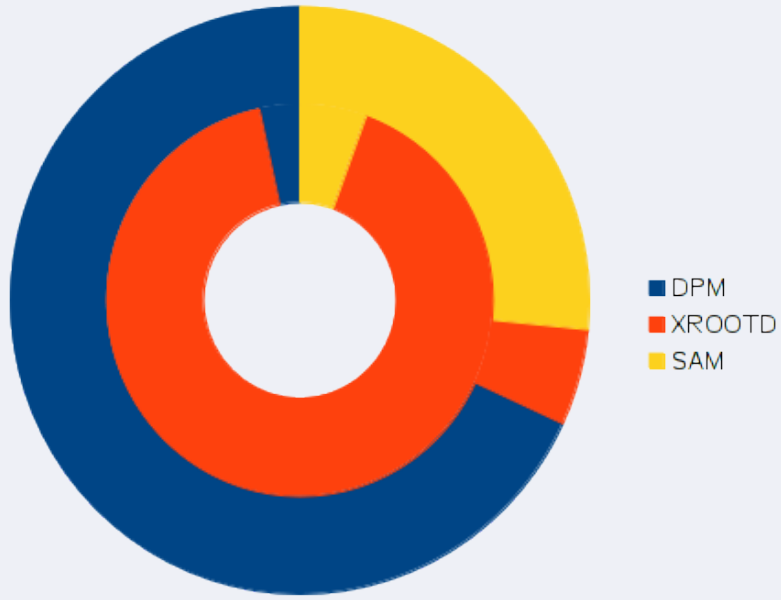
- UJF Řež: XROOTD datové uzly, ~100TB
- CESNET-DU Plzeň: dCache, 30TB na diskoch
  - využitie: 17TB na diskoch + 28TB na páskach, 40k súborov
  - vítané rozšírenie počas nedostatku kapacít vo FZU minulý rok
  - momentálne sa snažíme využívať hlavne lokálnu kapacitu

Datové prenosy vnútri FZÚ

Fyzikální ústav
Akademie věd ČR, v. v. i.

Prenesené dáta medzi WN a dátovými uzlami za rok 2013

	WN →	→ WN
DPM:	70TB	13PB
XROOTD:	1.9PB	1PB
SAM:	115TB	5PB
Spolu:	2.1PB	19PB

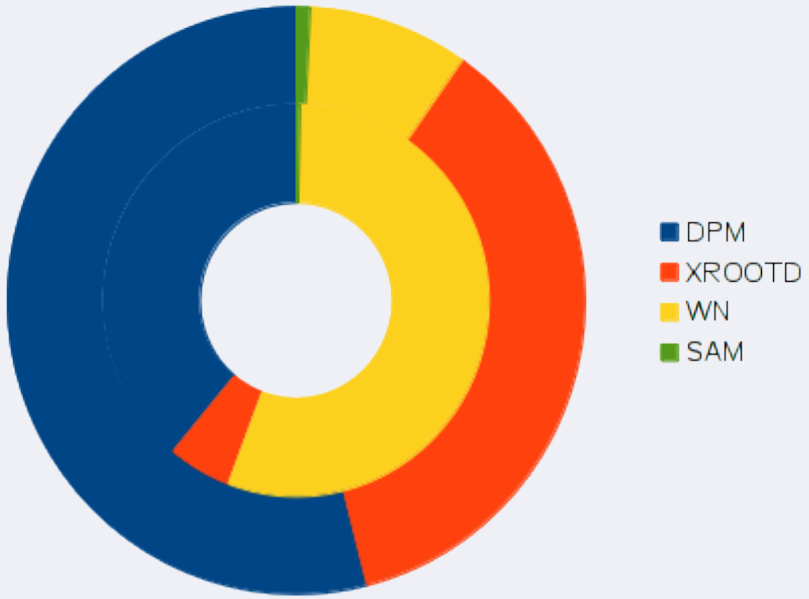


Externé datové prenosy

Fyzikální ústav
Akademie věd ČR, v. v. i.

Prenesené dáta za rok 2013:

	in	out
DPM:	1.6PB	1.5PB
XROOTD:	0.2PB	1.0PB
WN:	2.3PB	0.2PB
SAM:	12TB	22TB
Spolu:	4.0PB	2.9PB



Časté datové prenosy

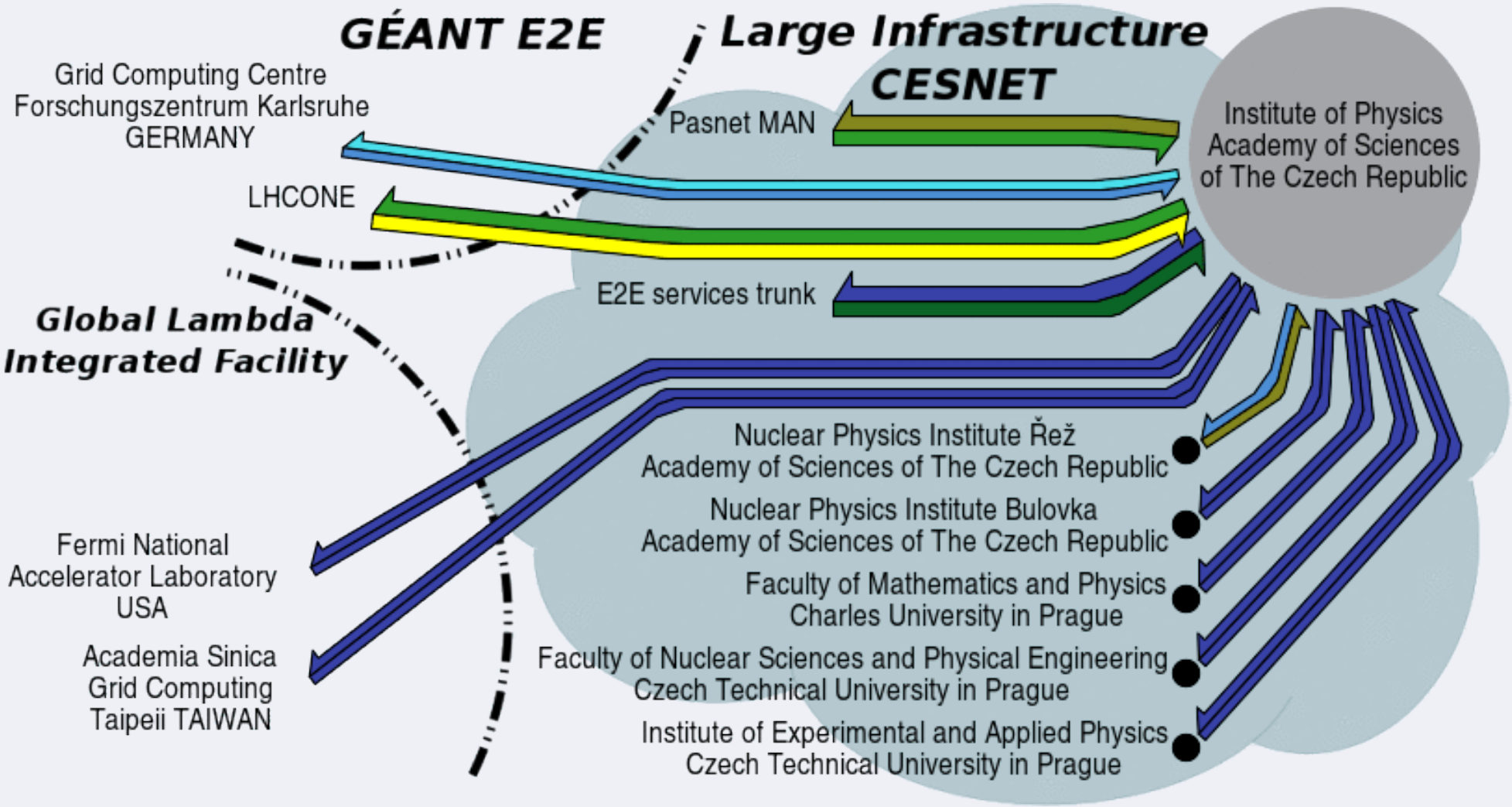
- KIT (Karlsruhe), naše nadradené Tier1 centrum (ATLAS)
- CERN (ALICE)
- ďalšie výpočtové centrá patriace pod KIT Tier1 (ATLAS)
- FNAL (D0, NOVA)

Externé datové prenosy

Prenesené dáta podľa inštitúcie:

	in	out
KIT Karlsruhe (DE):	792TB	403TB
CERN (CH):	695TB	1078TB
UJF Řež (CZ):	505TB	33TB
CESNET-DU Plzeň (CZ):	260TB	27TB
DESY Zeuthen (DE):	41TB	51TB
FNAL (USA):	18TB	26TB
ASGC Taipei (TW):	16TB	16TB
BNL (USA):	NA	NA

Zahraničné linky



WLCG

- globálna kolaborácia
- viac než 150 výpočtových stredísk z takmer 40 krajín
- jej úlohou je poskytovať výpočtové prostriedky uchovávanie, distribúciu a analýzu dát z detektoru LHC v CERNe

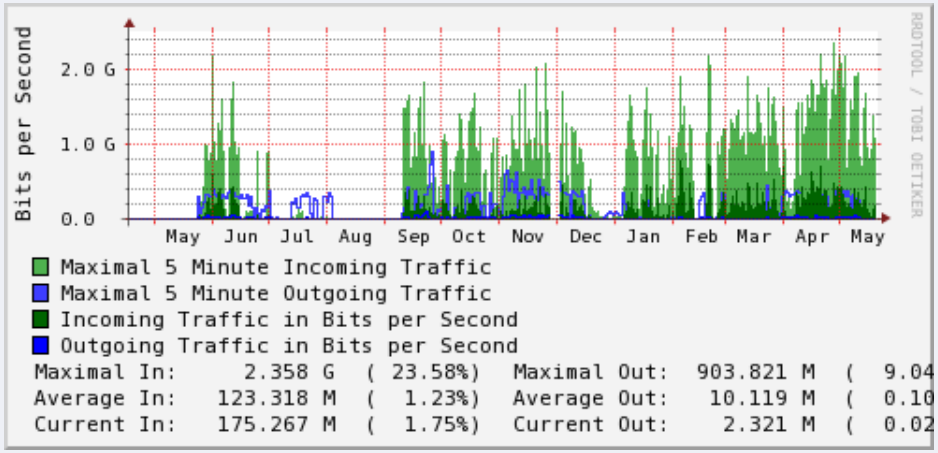
LHCONE

- sieť spájajúca vybrané výpočtové strediská WLCG
- cieľ: minimalizovať hop-count
- cez BGP inzerovaných 144 rôznych prefixov
- FZÚ nemá vlastný AS, BGP nám zaisťuje router CESNETu

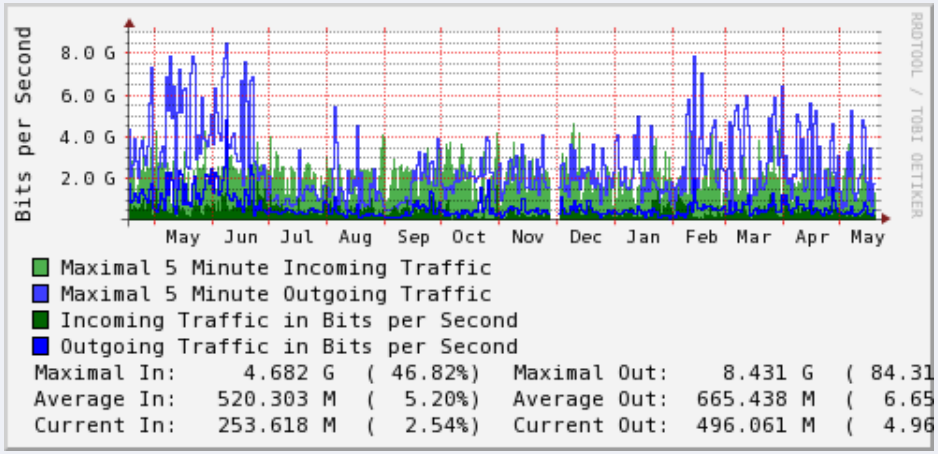
Záťaž 10Gbps liniek

Fyzikální ústav
Akademie věd ČR, v. v. i.

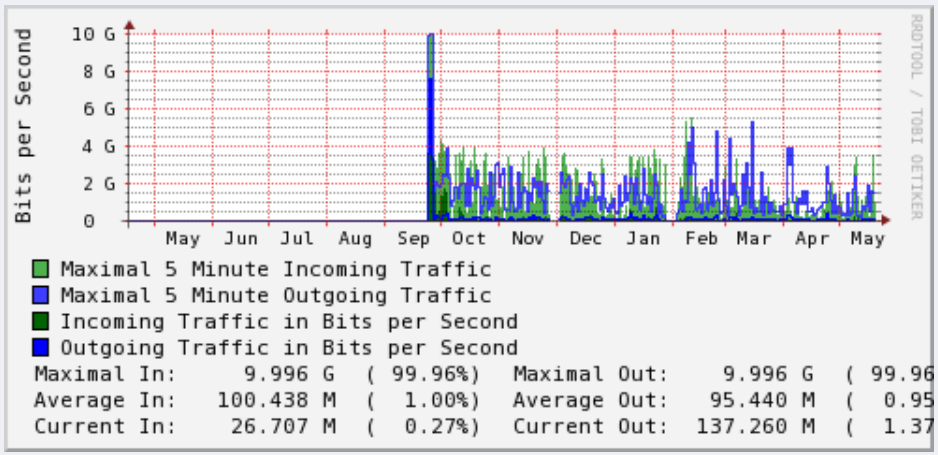
CESNET E2E trunk



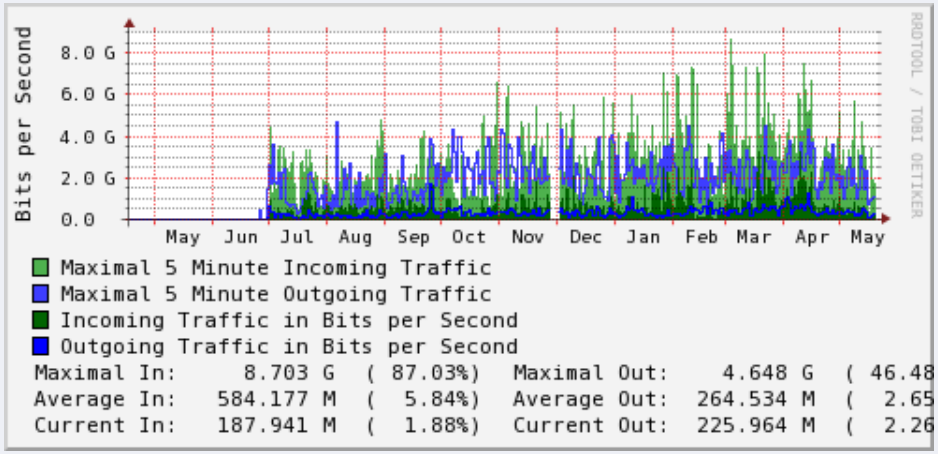
Pasnet



KIT (DE)



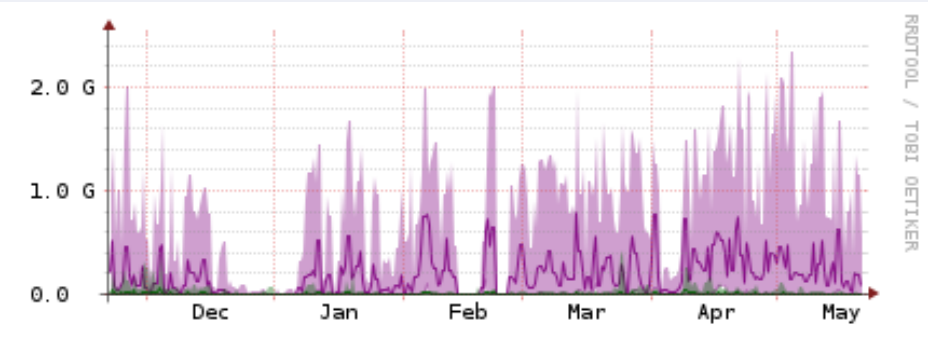
LHCONE



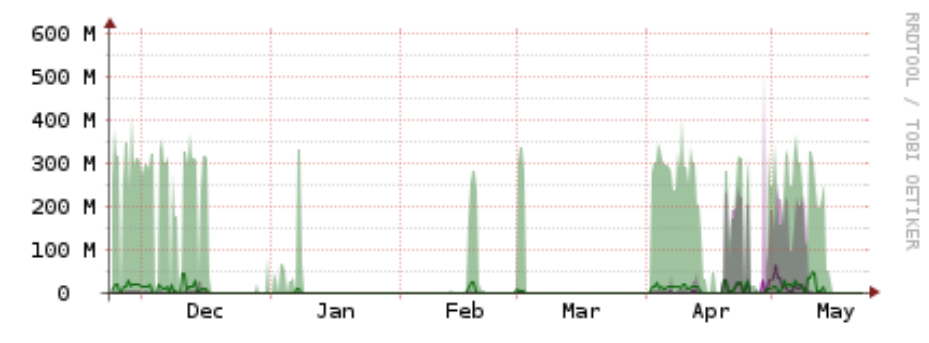
Záťaž liniek v E2E trunku

Fyzikální ústav
Akademie věd ČR, v. v. i.

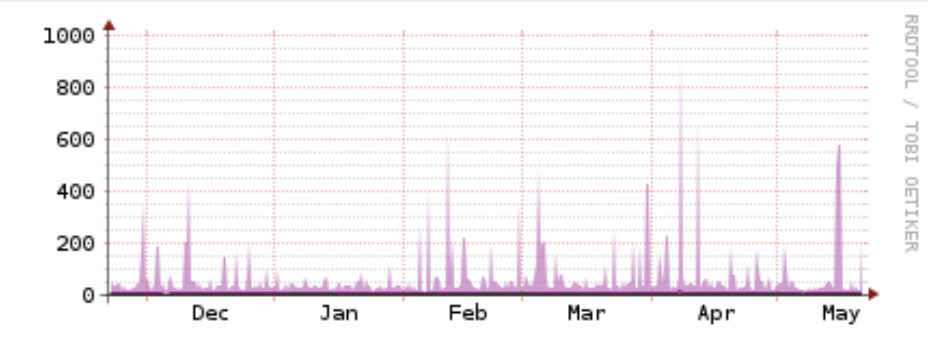
UJF Řež



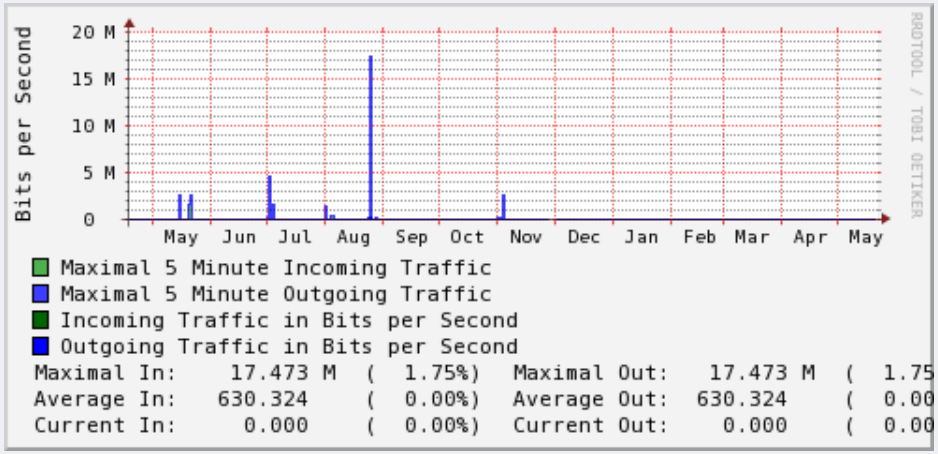
FNAL



ASGC Taipei



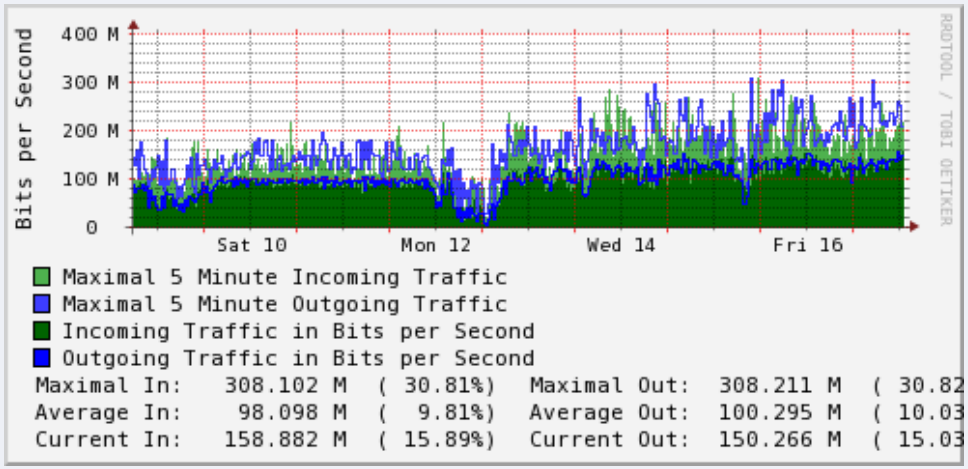
BNL (zrušená)



Fyzikální ústav
Akademie věd ČR, v. v. i.

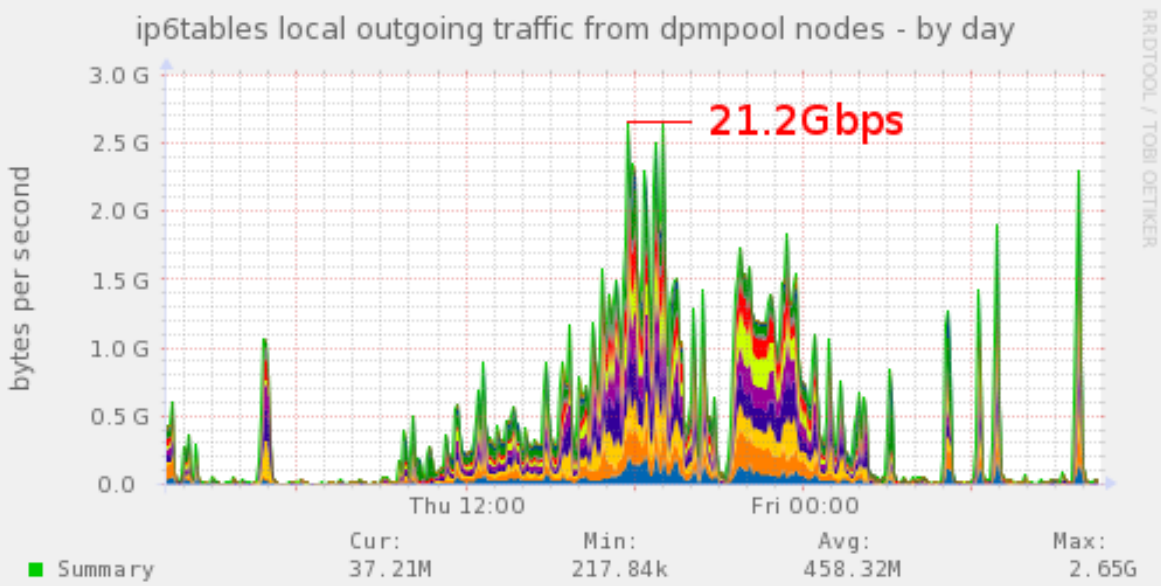
IPv6

- spolupráca so zahraničnými laboratóriami na zavedení protokolu IPv6
- laboratóriá: IC, RAL, QMUL (GB), KIT, DESY (DE), PIC (ES), IN2P3 (FR), INFN (IT), FNAL (USA), NDGF, ...
- pravidelné dátové prenosy pomocou GridFTP

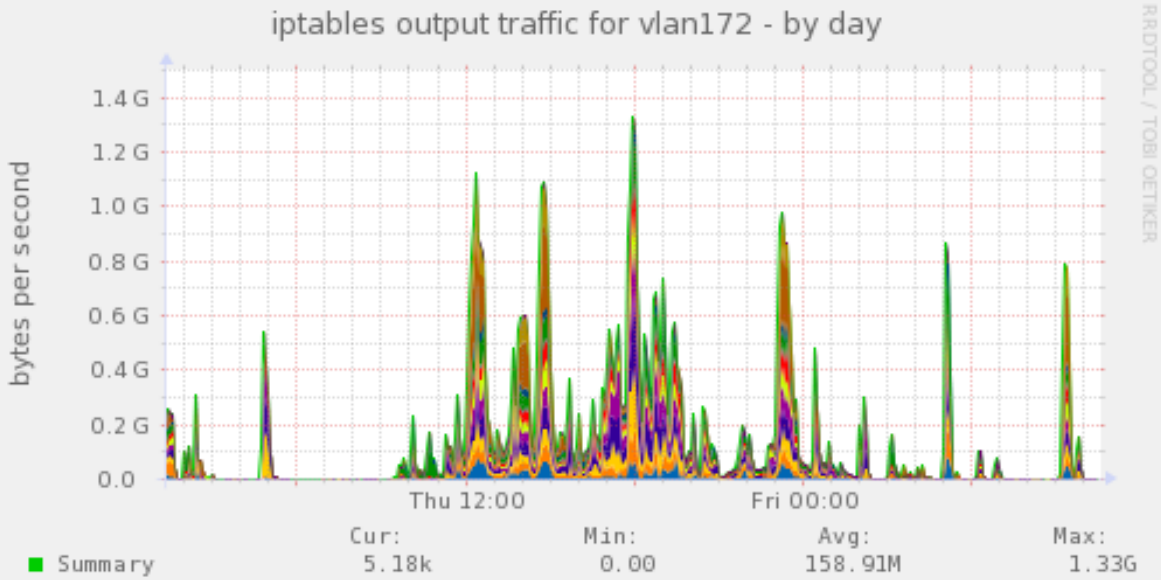


- testovanie middleware a HEP-specific aplikácií
- vytváranie best-practices pre HEP komunitu

Lokálne IPv6 prenosy v produkcii



DPM → WNs:
IPv6 traffic



DPM → WNs:
IPv4 traffic

IPv6 a jumboframes

Zapojenie IPv6

- na komoditnej linke do Pasnetu
- BGP peering s KIT cez dedikovanú linku

Jumboframes

- zapnuté na linkách do Pasnetu a KIT
- cez IPv6 používame jumboframes
- spolupráca s QMUL (GB) na testovaní jumboframes
  - flapujúca cesta FZU $\rightarrow$ QMUL
  - cesta cez Geant podporuje jumboframy
  - obcas sa routovanie presmerovalo cez Tiscali, to ich nepodporuje
  - pri problémoch sme vždy kontaktovali Pasnet
  - problém s routovaním kvôli chýbajúcemu AS Pasnetu

Centrálny router C6500

Úpravy centrálneho routra VS FZÚ

- upgrade z 5U na 9U chassis a nový hypervisor (CESNETom)
- na radu Pasnetu odpojené FWSM (malý výkon na 10G linku)
- filtrovacie pravidlá FWSM sme previedli do ACL routra

Hardwarový problém

- pri odpojení a opätovnom zapojení port nenabehol
- nutný reštart routra
- po pár týždňoch sa problém objavil znova
- po >6 mesiacoch riešenia vymenený celý hardware

CPU load problém

- bug v IOS: problém s kombináciou netflow a NAT
- po pár týždňoch dostupná oprava

Výhľady

LHC run 2 (2015 – 2017)

- zvýšenie energie a luminozity (počtu zrážok)
- zvýšenie dátového objemu
- zvýšenie prenosu do ČR
  - v závislosti na zvyšovaní úložných a výpočtových kapacít
  - predpoklad, že kapacity sa budú postupne navyšovať

experiment NOVA (FNAL, USA)

- naberanie novych dat, simulacne ulohy

experiment Belle II (KEK, Japonsko)

- diskusia s CESNET-DU o uložení niekoľkých PB dát na pásky

Auger (Pierre Auger Observatory)

- prenosy dát v rámci existujúch stredísk, stovky TB ročne

Cherenkov Telescope Array (CTA)

- prenosy simulačných dát

Ďakujem za pozornosť.

Marek Eliáš

elias@fzu.cz

http://www.farm.particle.cz

Prácu na IPv6 čiastočne podporil Fond rozvoje CESNETu, číslo projektu 482/2013
Výpočty pre LHC experimenty sú podporované projektami MŠMT INGO LG13031 a LG13009