Monitorovanie sietí na rýchlosti 100 Gb/s

(Internet a Technologie 12)

Lukáš Kekely, Viktor Puš, Štěpán Friedl
(kekely, pus, friedl@cesnet.cz)

Praha, 24. 11. 2012

Obsah

(Štandardná) Motivácia

- rýchly nárast objemu dát tečúcich sieťou
  - stúpa počet služieb využívajúcich siete
  - rastie počet zariadení schopných komunikácie
  - zväčšuje sa objem prenášaných dát
- nutnosť zvýšenia prenosovej rýchlosti!

(Štandardná) Motivácia

- rýchly nárast objemu dát tečúcich sieťou
  - stúpa počet služieb využívajúcich siete
  - rastie počet zariadení schopných komunikácie
  - zväčšuje sa objem prenášaných dát
- nutnosť zvýšenia prenosovej rýchlosti!

Ako rýchlo skutočne rastie vyťaženie sietí?

(Štandardná) Motivácia

- rýchly nárast objemu dát tečúcich sieťou
  - stúpa počet služieb využívajúcich siete
  - rastie počet zariadení schopných komunikácie
  - zväčšuje sa objem prenášaných dát
- nutnosť zvýšenia prenosovej rýchlosti!

Ako rýchlo skutočne rastie vyťaženie sietí?

⇒ IEEE 802.3^{TM} Industry Connections Ethernet Bandwidth Assessment (júl 2012)
http://www.ieee802.org/3/ad_hoc/bwa/BWA_Report.pdf

Nárast rýchlosti (Dátové centrá)

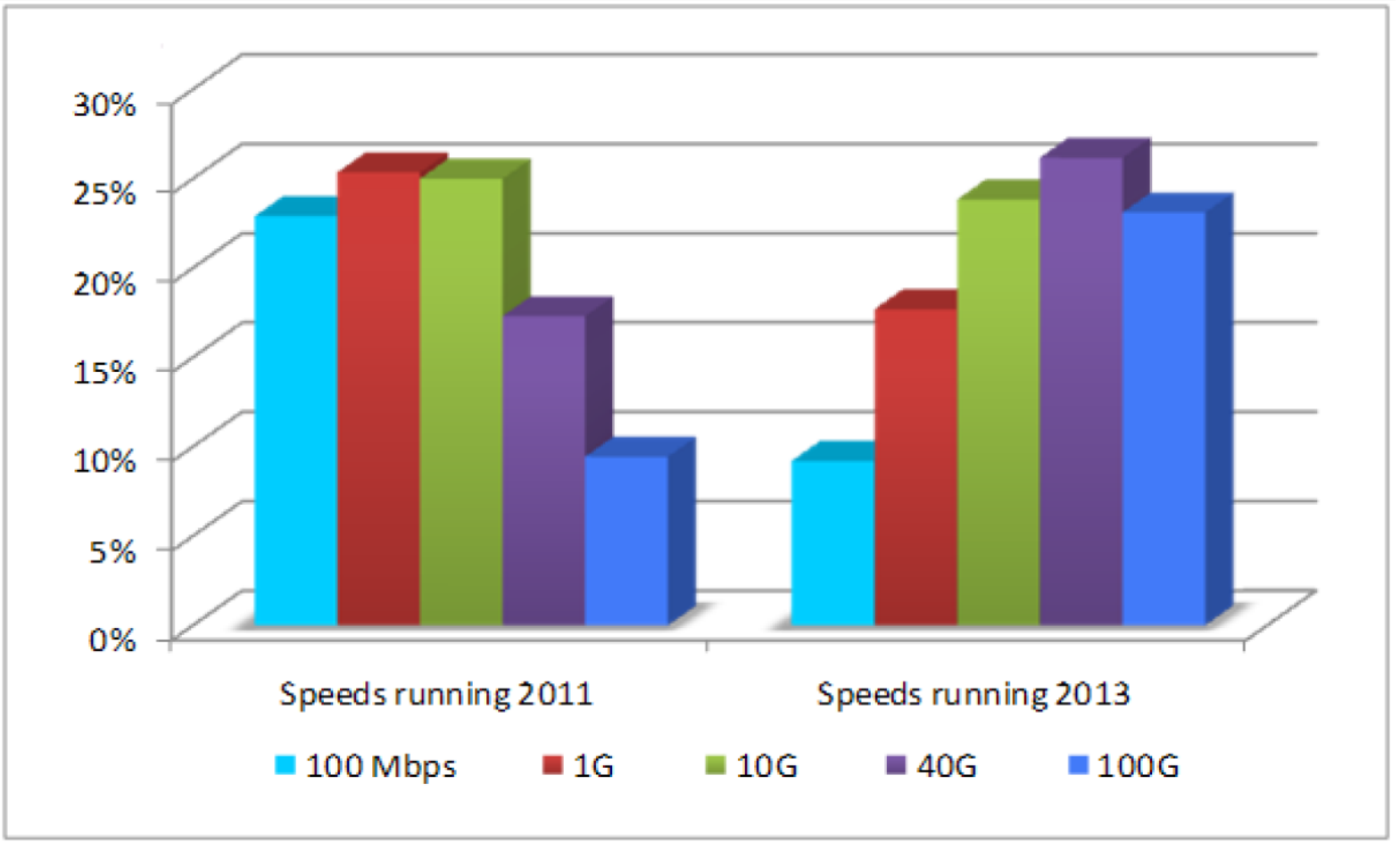


Figure 11—Data center study - percentage of links by speed

- v dátových centrách budú už budúci rok prevládať 40 a 100 Gb/s technológie

Nárast rýchlosti (Koncoví užívatelia)

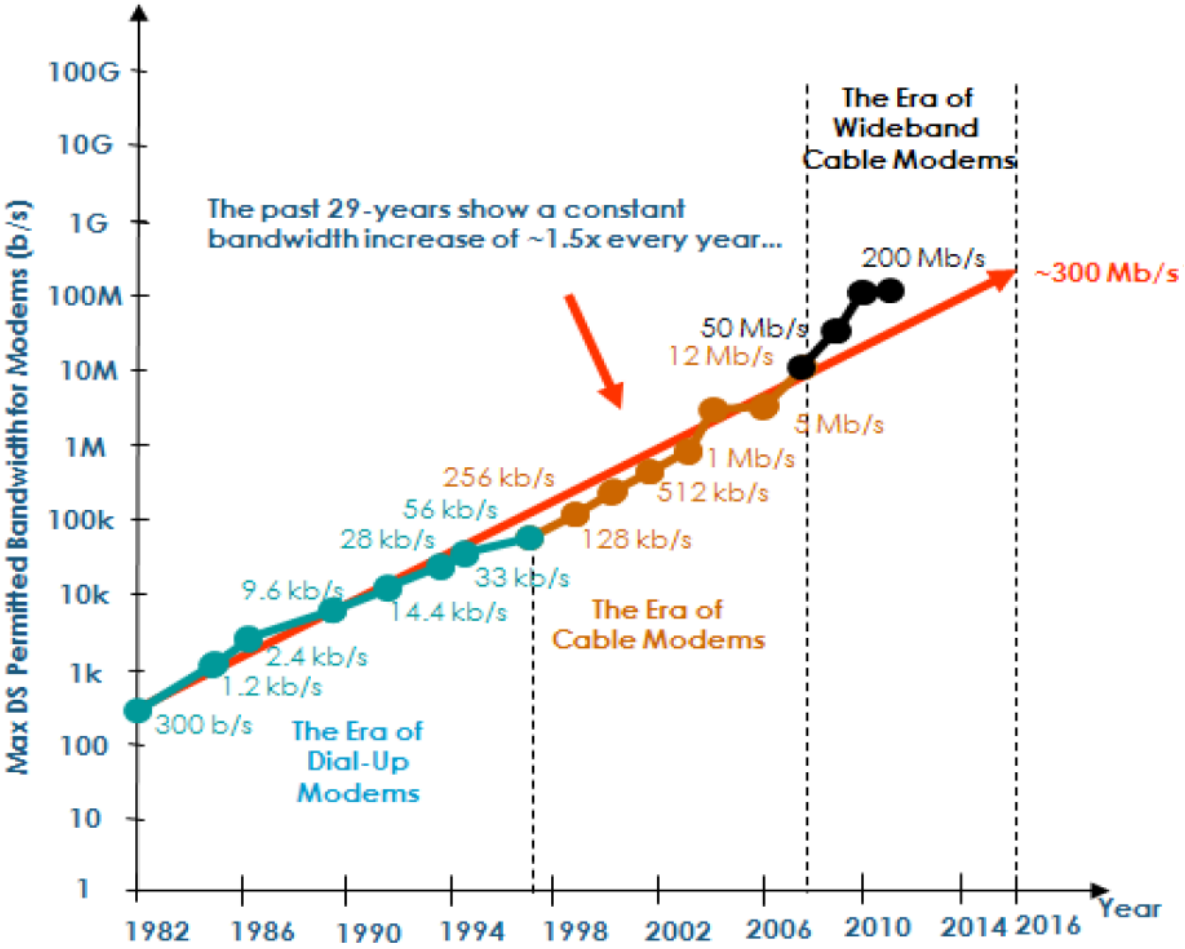


Figure 19—Maximum permitted downstream bandwidth trend

- *Nielsen's Law of Internet bandwidth:* rýchlosť pripojenia high-end užívateľov vzrastie o 50 % ročne

Nárast rýchlosti (Vysokorýchlostné jadro)

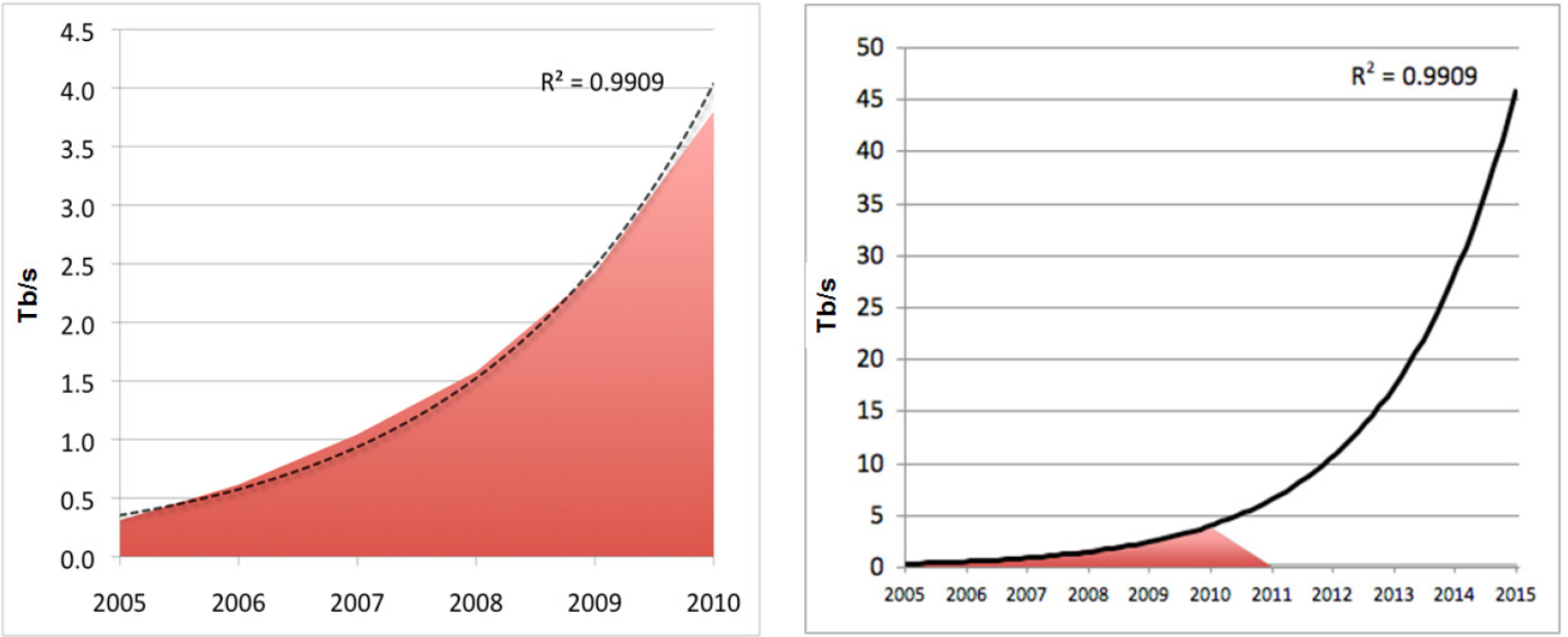


Figure 39—Five year peak European IXP traffic projection

- nárast prenosových rýchlosti priemerne o asi 57 % ročne

Typ prenášaných dát

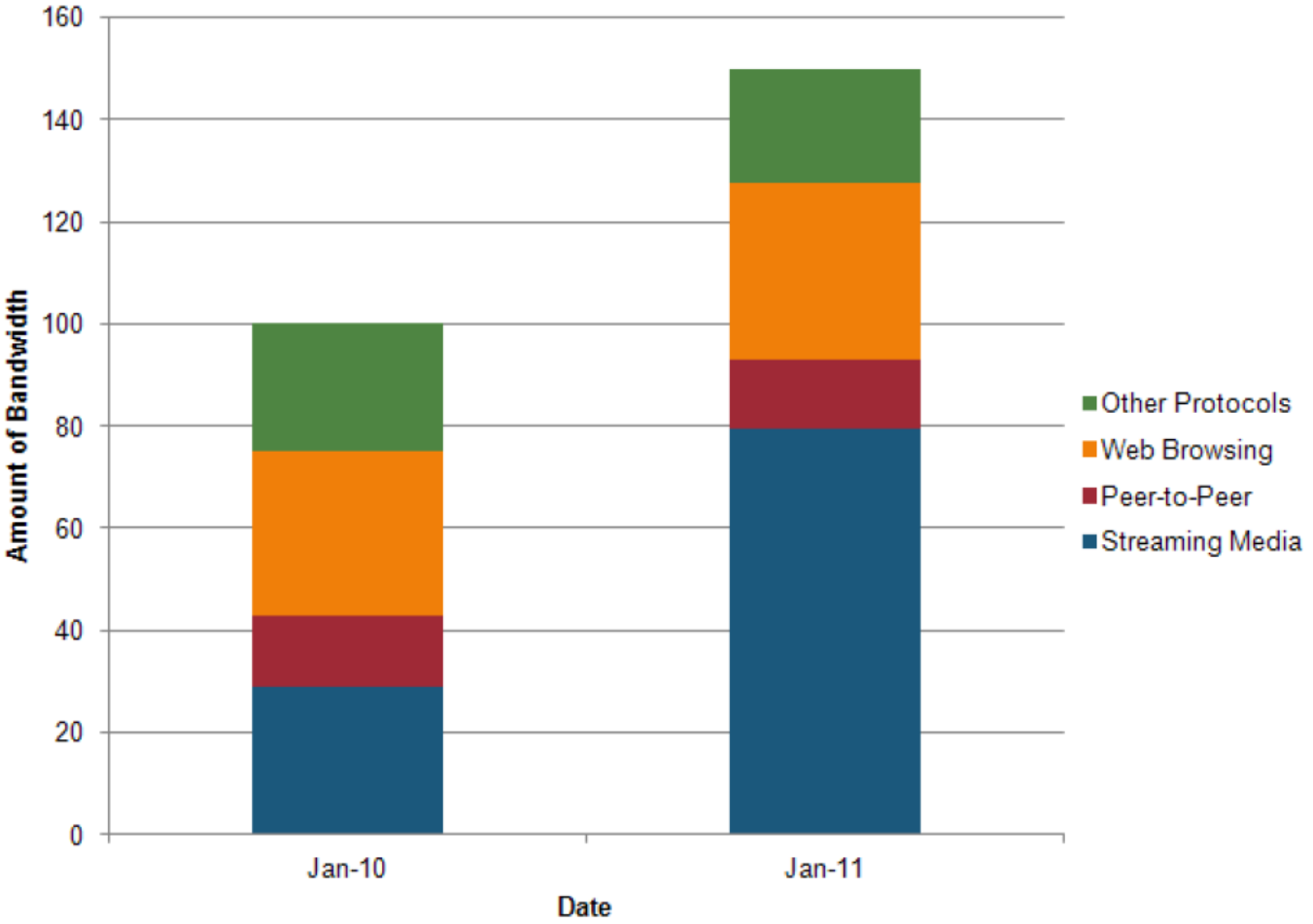


Figure 23—Mix of traffic type vs. time

- výrazne rastie pomer multimediálneho obsahu
- CESNET predviedol technológiu na prenos 8K videa

Príklad z Londýna

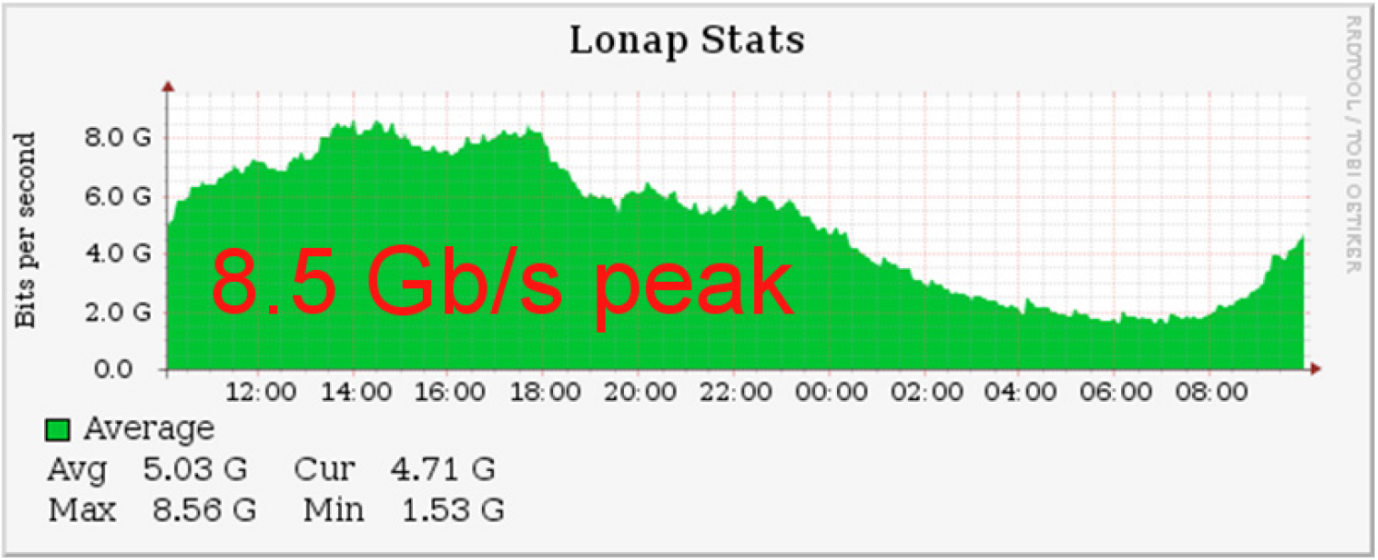


Figure 34—LONAP (London) traffic on a 'normal' weekday

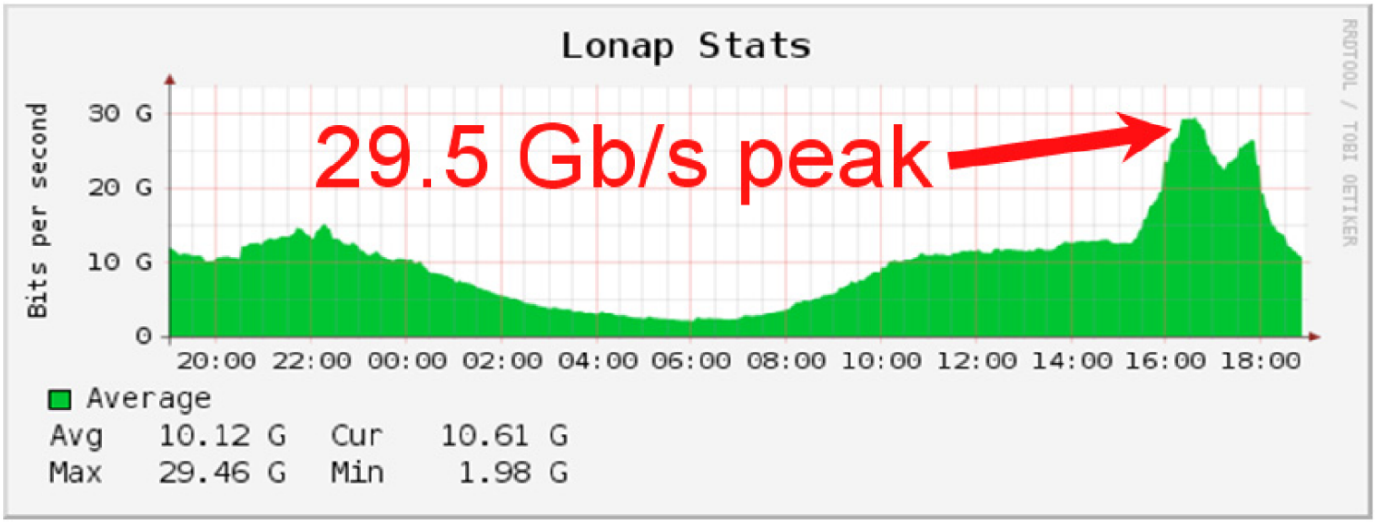


Figure 35—LONAP (London) traffic during the World Cup 2010 England vs. Slovenia

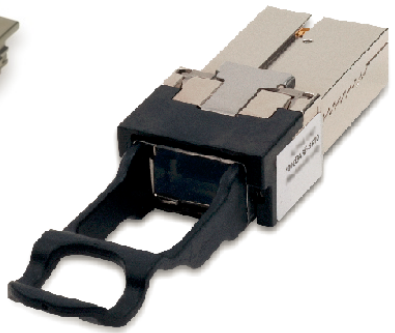
Základné vlastnosti

- štandard 802.3ba prijatý IEEE v júni 2010
- dve prenosové rýchlosti – 40 a 100 Gb/s
- na spojovej vrstve (L2) zachováva väčšinu vlastností
  - enkapsulácia dát a formát rámca
  - minimálna a maximálna veľkosť rámca
  - adresovanie pomocou MAC adries
  - detekcia chýb pomocou CRC
- na fyzickej vrstve (L1) nastali zmeny
  - niekoľko paralelných fyzických ciest
  - rámec je prenášaný naraz všetkými cestami
  - pridané značky na zarovnanie posunu ciest
  - $\Rightarrow$ potreba novej realizácie fyzickej vrstvy

$\Rightarrow$ 100 GE nie je len 10-krát rýchlejší 10 GE

Fyzická vrstva

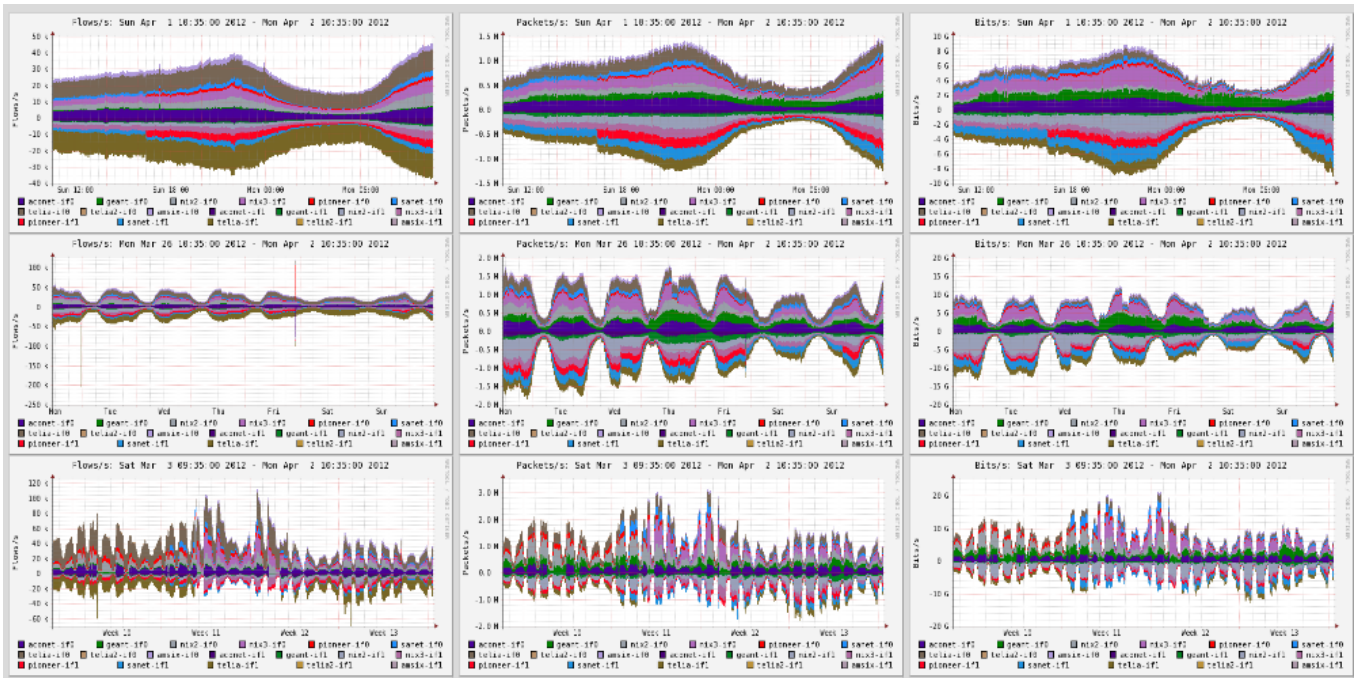
- dostupné optické moduly a optické médiá
- paralelný prenos rámca viacerými cestami
  - 1 vlákno, 4 vlnové dĺžky po 25 Gb/s (= 100 Gb/s)
  - 1 vlákno, 4 vlnové dĺžky po 10 Gb/s (= 40 Gb/s)
  - 10 vlákien po 10 Gb/s (= 100 Gb/s)
  - 4 vlákna po 10 Gb/s (= 40 Gb/s)
  - 1 vlákno, 10 vlnových dĺžok po 10 Gb/s (= 100 Gb/s)
- rôzne realizácie optických transcieverov v 2 generáciách



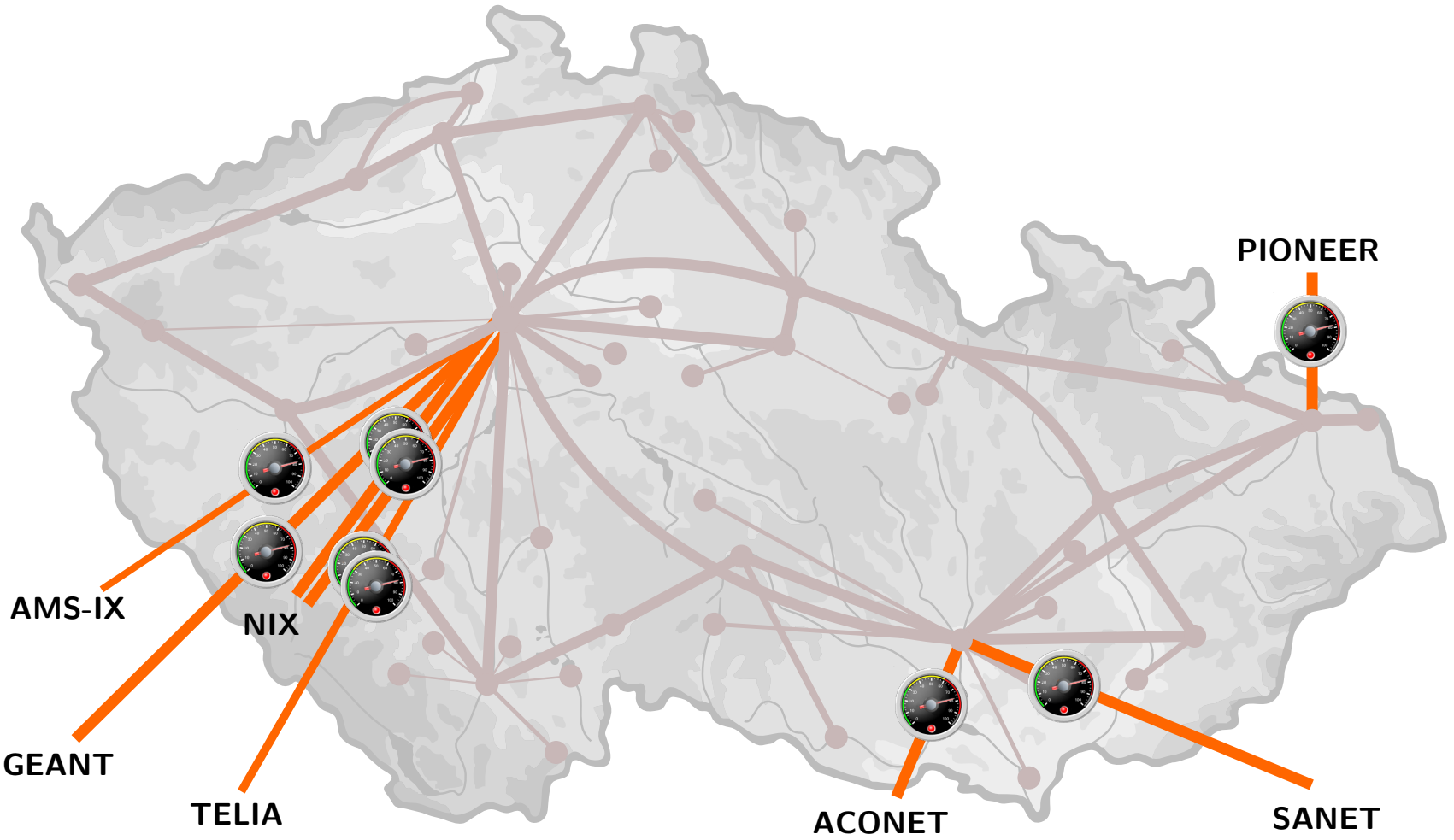
* obrázky sú zo stránok firmy Finisar

Prehľad

- *Oddělení nástrojů pro monitoring a konfiguraci*
- monitorovacia infraštruktúra na zber informácií o IP tokoch
- celkovo 9 meracích bodov na zber dát
- dáta spracovávané na kolektore
- pokrytie všetkých hraničných liniek siete CESNET2
- podpora monitorovania na plnej rýchlosti liniek
- okrem zberu dát aj vyhodnotenie, reprezentácia a dalšie

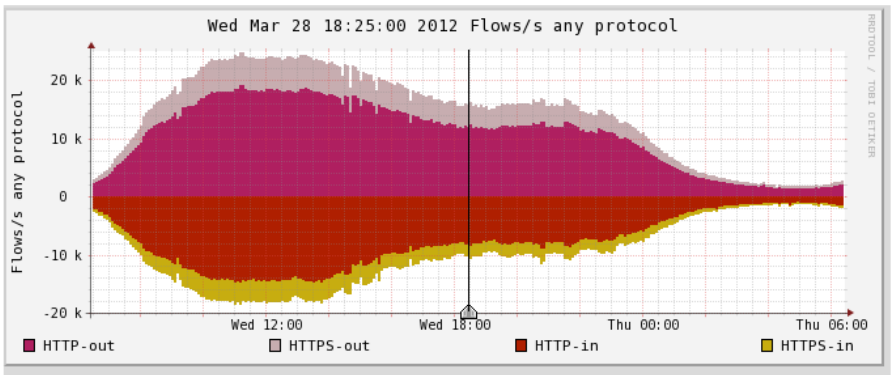


Topológia

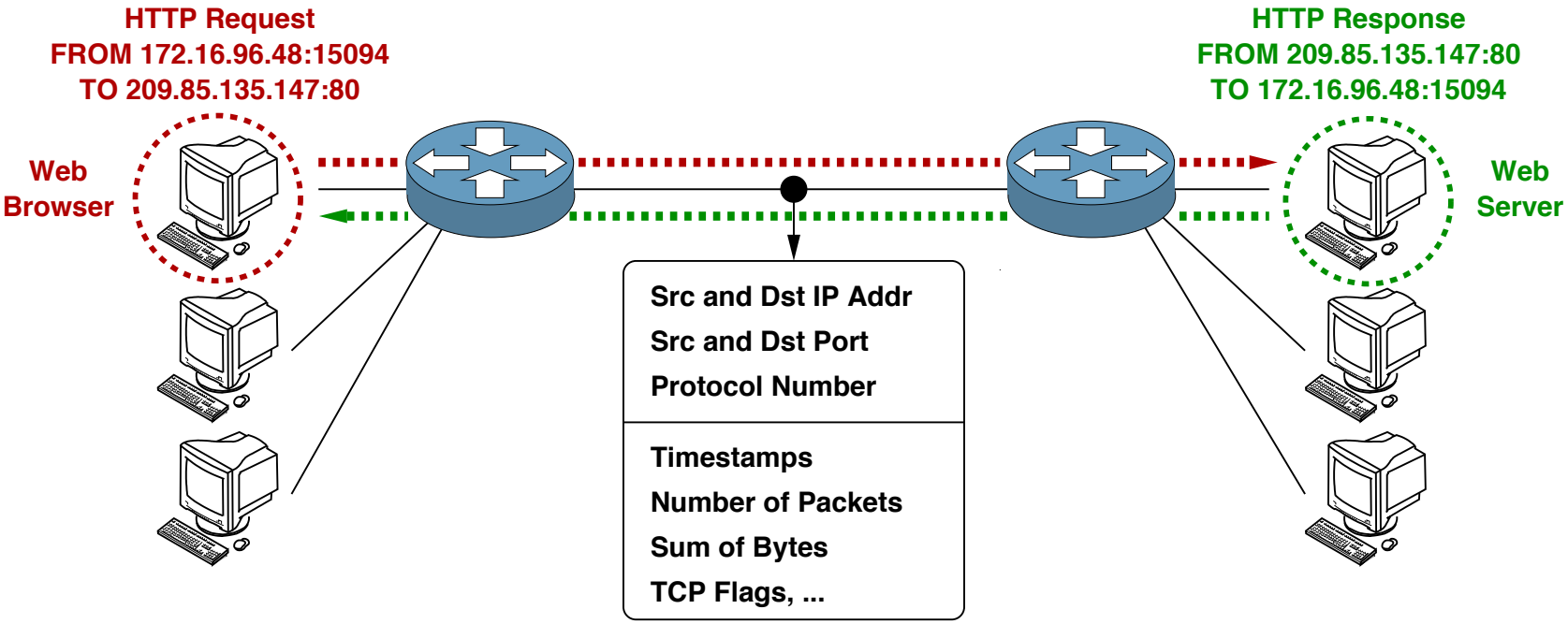


Monitorovanie

- *kto* komunikuje *s kým*, *ako dlho*, akým *protokolom* atď
- založené na CISCO *NetFlow v5/v9* a IETF *IPFIX*
  - *IPFIX* prináša flexibilitu záznamov $\Rightarrow$ možnosť rozšírenia o pokročilejšie informácie (napr. L7)
- monitoring a analýza sieťových tokov v reálnom čase
  - detekcia anomálií, útokov, výpadkov, nakazených PC atď
- ukladanie a prezentácia dlhodobých štatistík
  - agregácia na 5 minútové intervaly (raw cca 140 GB/deň)
  - sledovanie dlhodobého vývoja siete
  - dohľadanie podrobností k riešeniu incidentov



NetFlow príklad



Flow start	Duration	Proto	Src IP Addr:Port		Dst IP Addr:Port	Flags	Packets	Bytes
09:41:21.763	0.101	TCP	172.16.96.48:15094	->	209.85.135.147:80	.AP.SF	4	715
09:41:21.893	0.031	TCP	209.85.135.147:80	->	172.16.96.48:15094	.AP.SF	4	1594

Merací bod – Prehľad

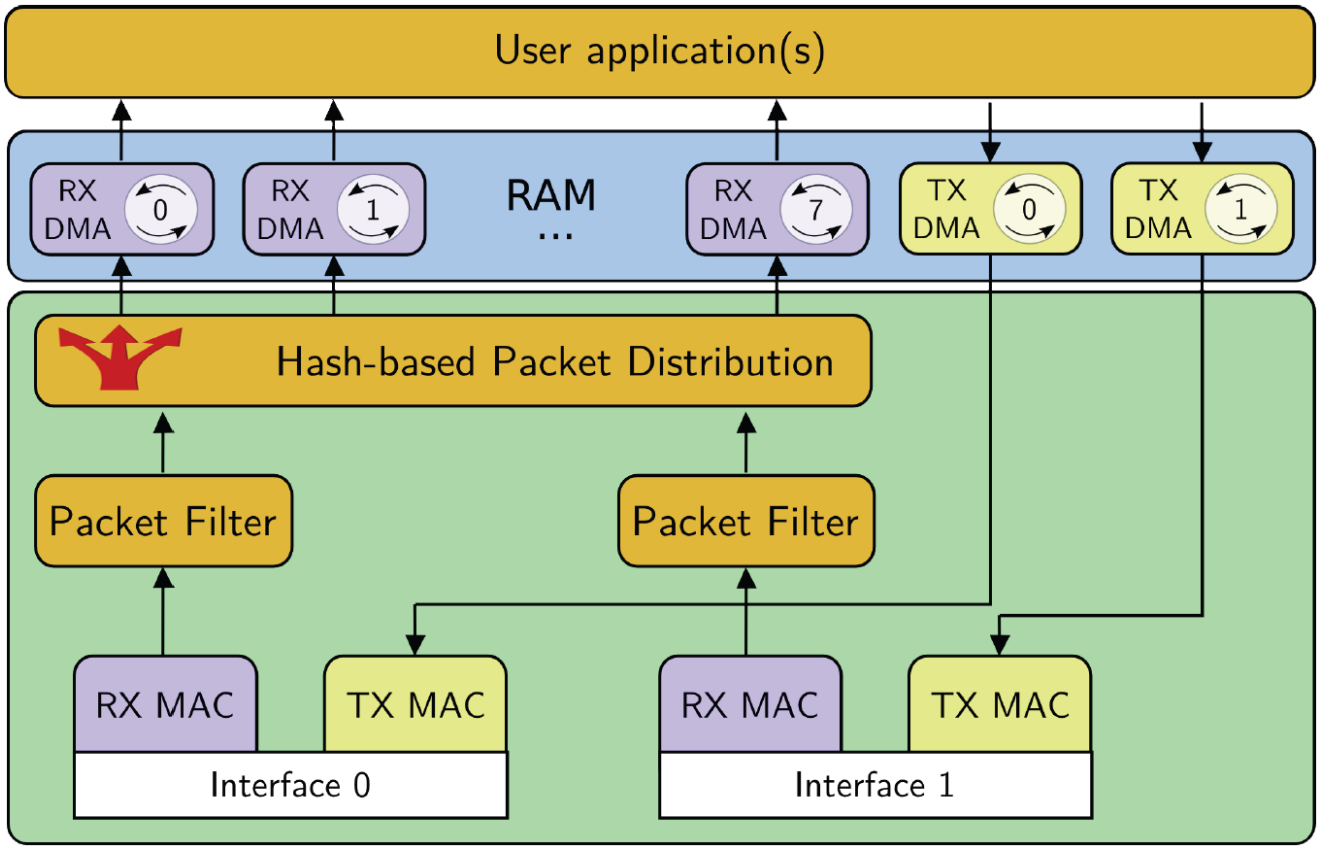
- komoditný server pod Linuxom
- vlastný viacvláknový SW exportér (NetFlow/IPFIX)
- hardvérová sonda COMBOv2 vlastnej výroby
  - PCI Express karta s FPGA Virtex-5

COMBOI-10G2

COMBOI-10G4

Merací bod – Hanic

- firmvér pre FPGA na Combov2, akcelerovaná NIC s možnosťou doplnenia vlastného predspracovania paketov
- nastaviteľná distribúcia dát na jadrá
  - základom je hash vybraných políčok z hlavičiek
  - distribúcia zachováva toky

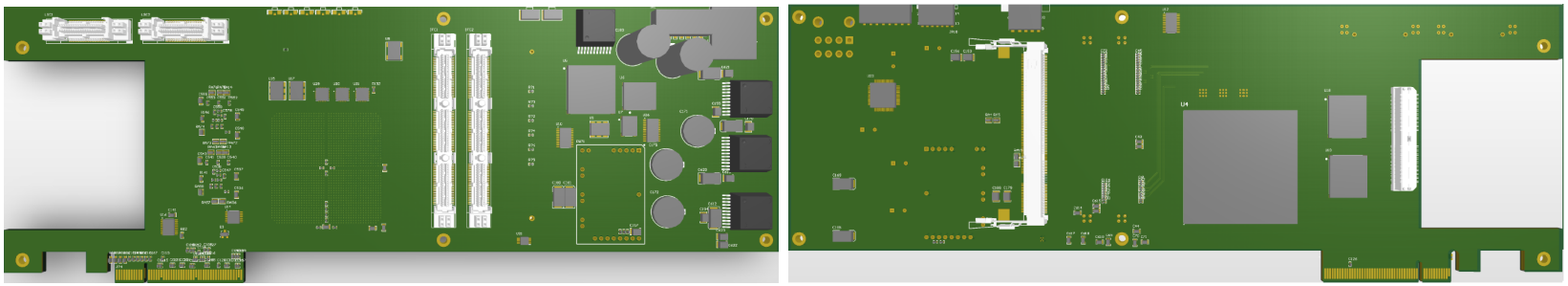


Prehľad

- zachovanie systému meracích bodov a kolektoru
- monitorovanie založené na IETF *IPFIX*
- nutnosť vytvorenia novej karty do meracieho bodu
- rozšírenie monitorovania o ďalšie informácie
- nový model HW/SW kodizajnu v meracom bode

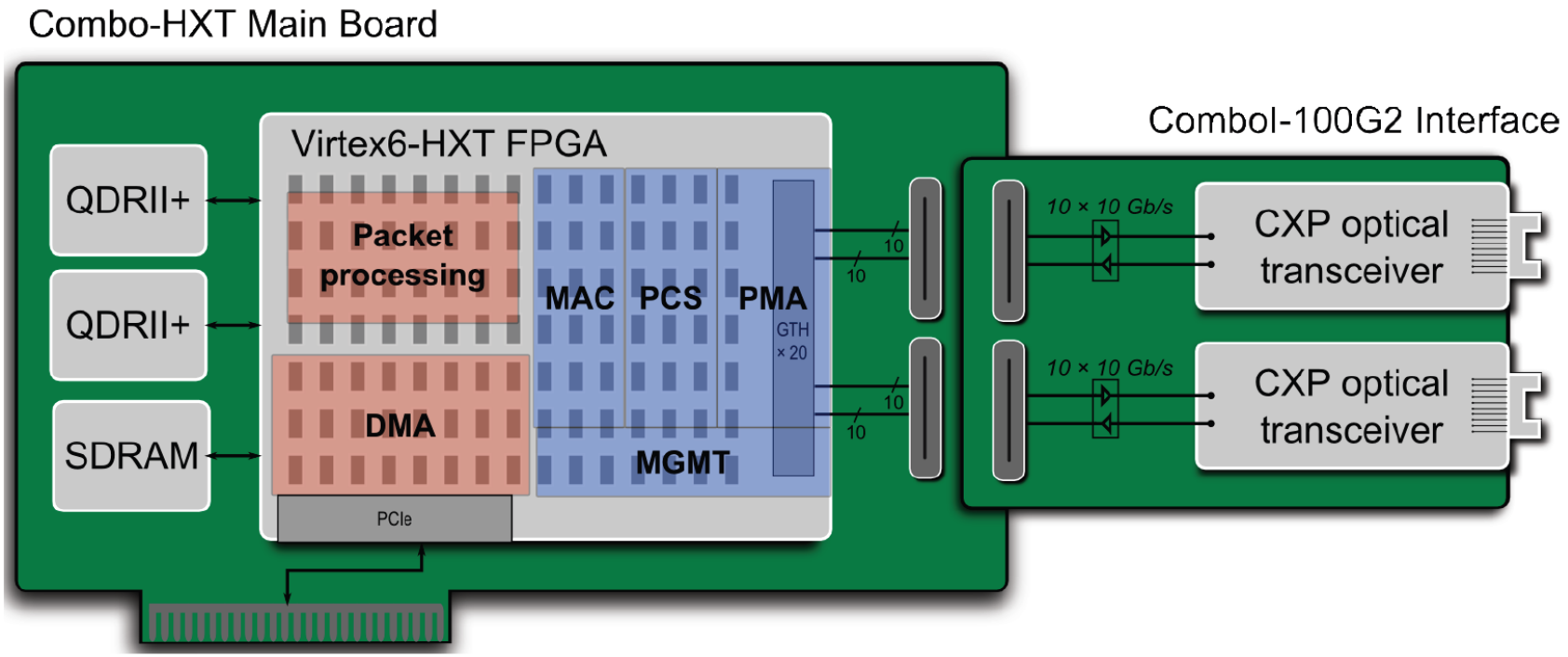
Nová karta – Combo-HXT

- vynútená zmenou fyzickej vrstvy a nárastom rýchlosti
- základňová PCI Express karta s FPGA Virtex6-HXT
  - 24x rýchly GTH transceiver (11,3 Gb/s každý)
  - 2x PCIe x8 rozhranie generácie 2 (64 Gb/s)
  - 2x QDRII+ pamäť (80 Gb/s)
  - SODIMM slot na DDR3 pamäť
- možnosť pripojiť rôzne interface karty



Nová karta – Interface karty

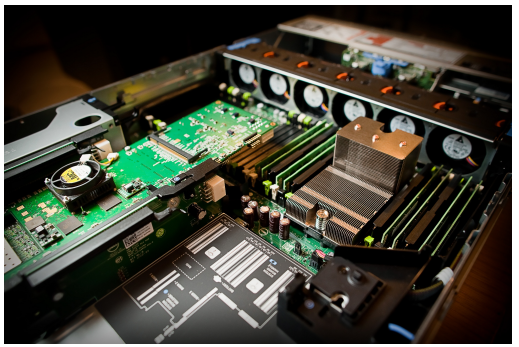
- jednoduchá zmena fyzických portov karty
- plánované rôzne varianty pre 40 aj 100 Gb/s
  - 1x 40 GbE, 2x 40 GbE
  - 2x 100 GbE
  - rozdeľovací kábel na 6x 40 GbE alebo 24x 10 GbE



Nový merací bod

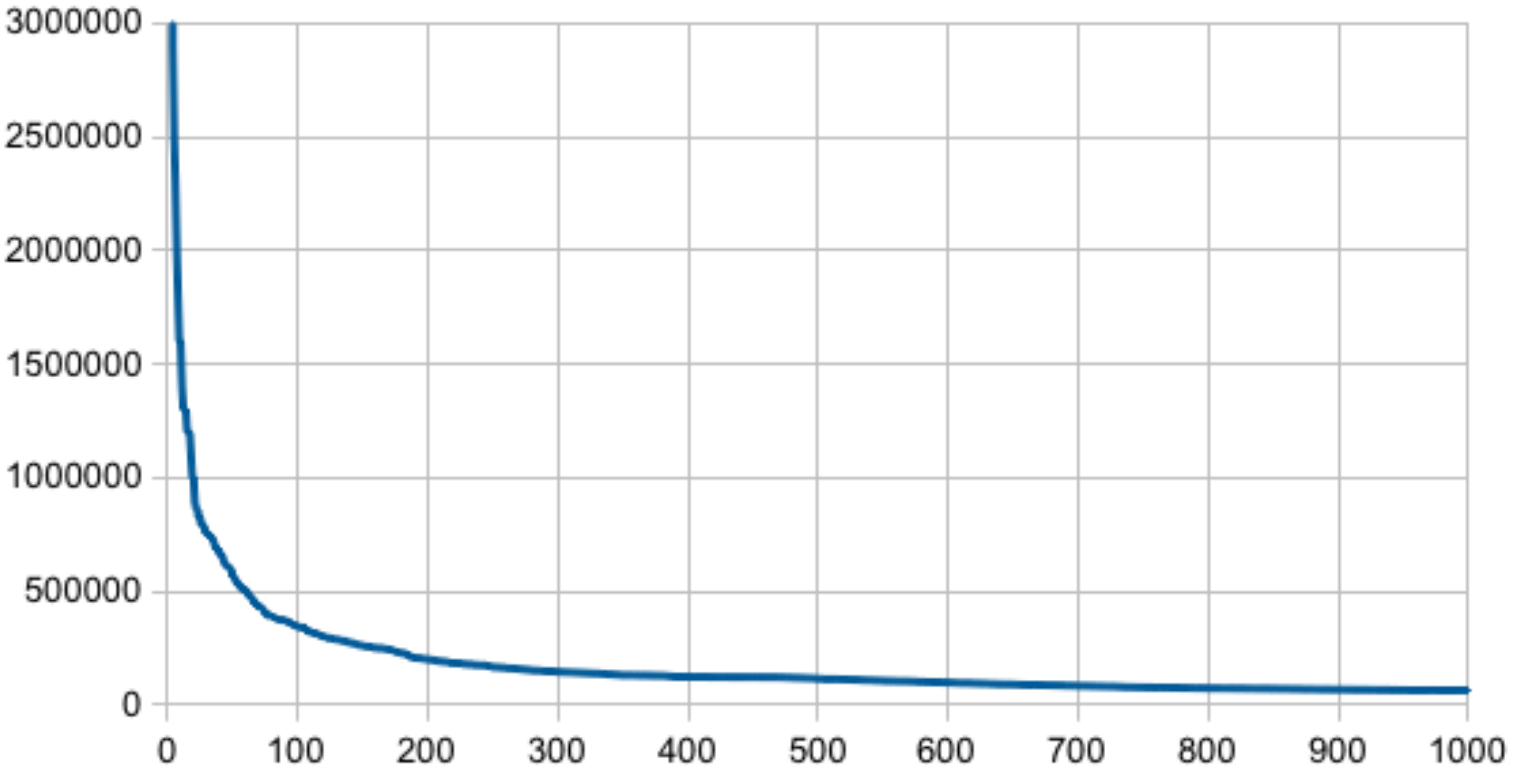
- komoditný server pod Linuxom s Combo-HXT
- hlavné problémy vynucujúce zmenu
  - PCIe kritickým bodom na priepustnosť (len 64 Gb/s)
  - podpora pokročilejšieho monitorovania (detekcia)
  - snaha odľahčiť SW spracovanie paketov
- využiteľné vlastnosti sieťových tokov
  - *heavy-tail* rozloženie tokov
  - nezaujímavosť dátovej časti väčšiny paketov
- riešenie inšpirované *Software Defined Networking*
  - užšie previazanie SW riadenia a HW akcelerácie
  - SW sa stará o riadenie a využíva HW na odľahčenie





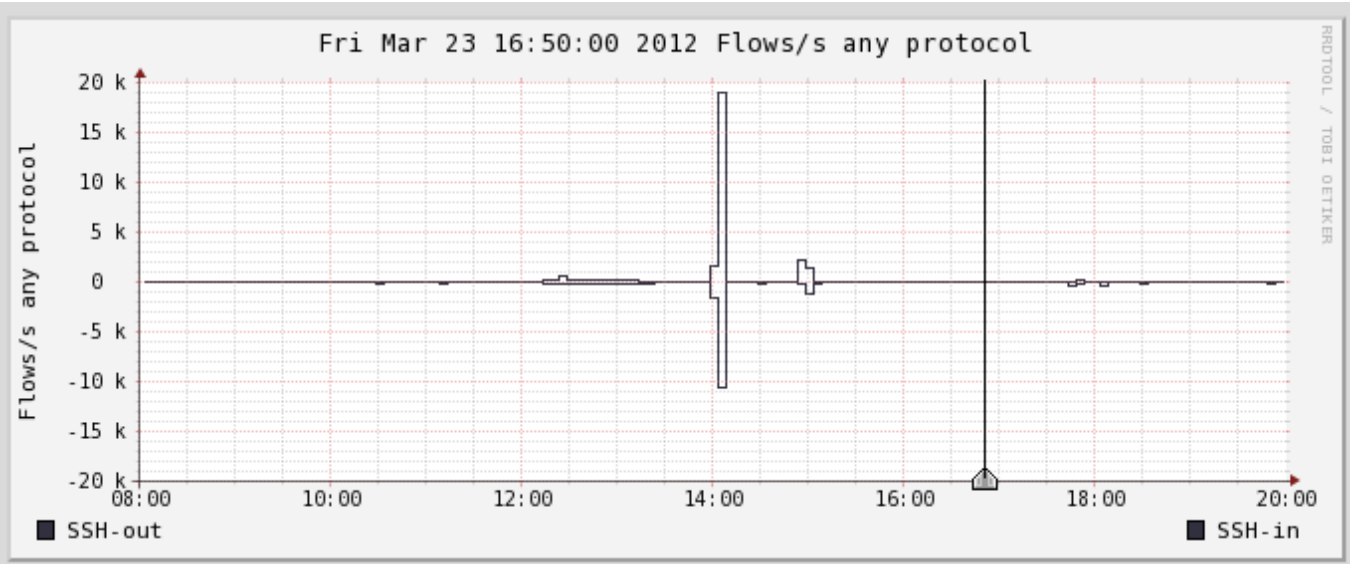
Nový merací bod – Heavy-tail

- pozorované správanie veľkosti sieťových tokov na linke
- málo najväčších tokov tvorí veľkú časť z toku linkou
- meranie v sieti CESNET2
  - 1 000 najväčších tokov obsahuje cca 30 % paketov
  - 10 000 najväčších tokov obsahuje cca 60 % paketov



Nový merací bod – Nezaujímavosť dát

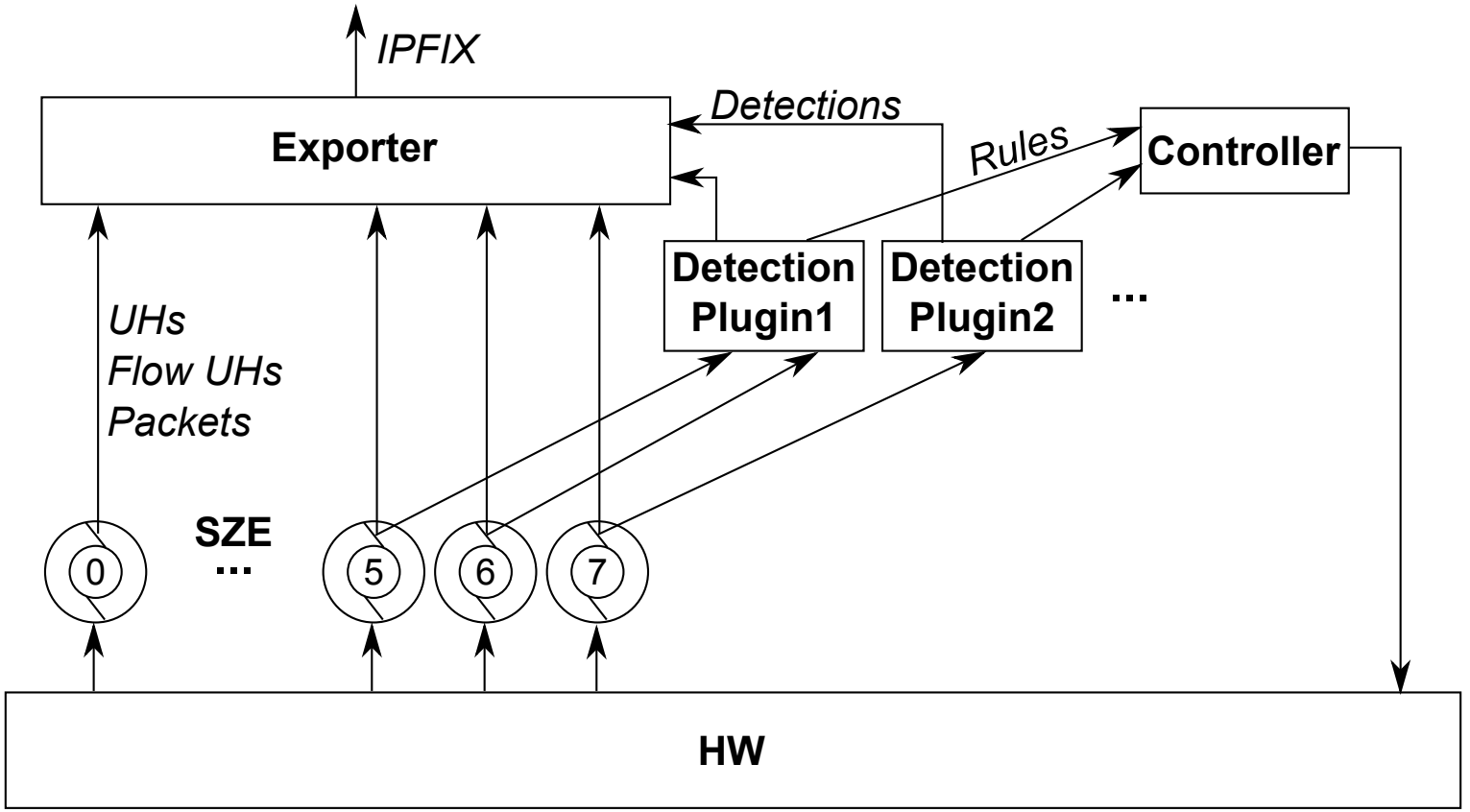
- priemerná veľkosť paketu je cca 600-800 B
- základné monitorovanie potrebuje len metadáta
  - adresy, porty, protokol, časová značka, veľkosť
  - veľkosť metadát je podstatne menšia ako väčšina paketov
- pokročilé monitorovanie nemusí mať všetky pakety
  - nespracovateľné protokoly (napr. šifrované) tvoria asi 10 %
  - detekcia SPAMu potrebuje len SMTP (menej ako 1 %)
  - detekcia DNS útokov potrebuje len DNS (menej ako 1 %)



SSH sken z Prahy do sveta

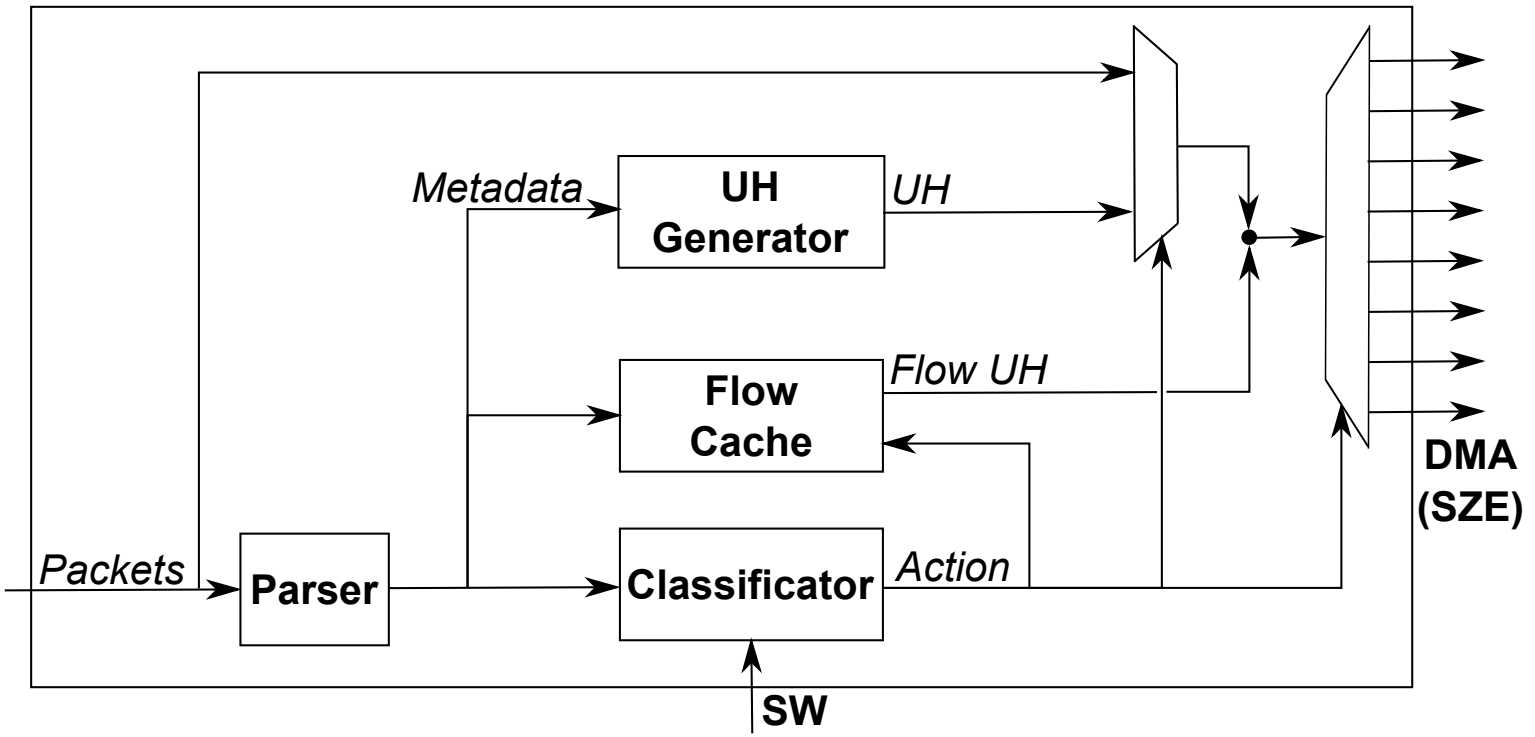
Nový merací bod – Návrh SW

- počítanie a export celkového monitoringu tokov
- jednoduchá rozšíriteľnosť merania (pluginy)
- konfigurácia klasifikácie a akcií v HW
  - dátovo nezaujímavé toky (stačí UH)
  - *heavy-tail* nezaujímavé toky (stačí Flow UH)



Nový merací bod – Návrh HW

- rozšírenie architektúry Hanic
- nastaviteľná klasifikácia paketov s akciou
- schopnosť extrahovať a posielať len metadáta do SW (UH)
- počítanie čiastkového monitoringu tokov (Flow UH)



Zhrnutie

- monitoring nesmie zaostávať za vývojom infraštruktúry
- poskytovanie ešte lepších služieb aj na vyššej rýchlosti

Zhrnutie

- monitoring nesmie zaostávať za vývojom infraštruktúry
- poskytovanie ešte lepších služieb aj na vyššej rýchlosti

“Optická síť CESNET2 je plně připravena na nasazení přenosové technologie 100 Gb/s, kterou poskytují jen nejpokročilejší infrastruktury.”

Ďakujem za pozornost.