

Technická zpráva CESNETu číslo 21/2005

Realizace optických sítí pro aplikace zpracování dat ve fyzice částic, aktualizace v roce 2005

Petr Beneš, Pavel Čermák, Tomáš Davídek, Lukáš Fiala, Josef Hampl, Ondřej Chvála, Jiří Král, Miloš Lokajíček, Karel Sandler, Jan Švec
Prosinec 2005

Úvod

Ve Fyzikálním ústavu AV ČR (FZÚ) Na Slovance bylo zřízeno Regionální výpočetní centrum pro fyziku částic, které bylo slavnostně otevřeno 1. 11. 2004. Centrum zajišťuje výpočetní a úložnou kapacitu pro náročné výpočty experimentů D0 na urychlovači TEVATRON ve FERMILAB a ATLAS a ALICE na budovaném urychlovači LHC v CERN, který zahájí provoz v roce 2007. Pro běžící experiment D0 se provádějí v Centru počítačové simulace činnosti detektoru D0 a zpracování experimentálních dat a pro experimenty ATLAS a ALICE probíhají počítačové simulace v rámci přípravy fyzikálního programu pro tyto experimenty. Součástí centra je výpočetní farma Goliáš, která má v současné době 200 procesorů a 40 TB diskového prostoru. Centrum je integrováno do prostředí mezinárodního gridu LCG (LHC Computing Grid).

V roce 2005 bylo v rámci projektu optických sítí dokončeno propojení pražských laboratoří spolupracujících na experimentech fyziky částic vyhrazenými optickými spoji 1 Gb/s (ve spolupráci s projektem rozvoje PASNET). Podrobný popis CWDM propojení užitých v pražské akademické síti pro některá pracoviště fyziky částic uvádí příloha této technické zprávy.

Současně došlo k připojení dalších dvou institucí v zahraničí a to Fermiho národní laboratoře v USA a ASGCC (Academica Sinica Grid Computing Center) na Taiwanu. Spoj do Fermiho národní laboratoře byl intenzivně využíván pro přenos vstupních dat a zpětně i výsledků simulací a rekonstrukcí pro experiment D0, spoj do ASGCC byl zprovozněn až v prosinci a bude využíván pro přenos dat mezi FZÚ a Tier1 centrem v ASGCC v rámci projektu LCG.

Zařízení na jednotlivých pracovištích

Fyzikální ústav AV ČR

FZÚ je nyní připojen k síti CzechLight přes dedikované SM 1GE vlákno. To je zakončeno v prostoru nového Regionálního centra, kde je připojeno do switchu HP ProCurve 4108GL do 1000LX miniGBIC. Do tohoto switchu je připojen PC router HP DL360 (2x XEON 2,8GHz, 2GB RAM). Tento router má 3 metalické 1GE síťové karty. Dynamicky routuje pomocí BGP mezi sítí CzechLight a vnitřní sítí farmy a mezi připojením k internetu a vnitřní sítí farmy (v případě výpadku CzechLight posílá veškerý provoz přes internet). Mezi sítí CzechLight a internetem neroutuje, protože síť CzechLight není určena pro přístup k internetu.

ČVUT, ÚTEF

Počátkem roku 2005 bylo dokončeno a proměřeno optické spojení mezi CESNET a Ústavem technické a experimentální fyziky (ÚTEF) ČVUT. Optické vlákno je na straně ÚTEF ukončeno v patch-panelu v hlavním ústavním racku. Spoj je dále veden optickým patchem do routeru (2 x Xeon, 6 x 64bit PCI-X) vybaveného Gb síťovou kartou (zapůjčena od CESNET). Přes druhé Gb rozhraní je routována doména utef.cvut.cz do CzechLight. V listopadu 2005 byly ve spolupráci s FZÚ provedeny zátěžové testy na úrovni IP.

V budoucnosti je plánováno pořízení serveru pro centrální ukládání dat v rámci společného fyzikálního projektu s University of Alberta (Kanada) a jeho propojení s Kanadou prostřednictvím sítě CzechLight.

ČVUT, Fakulta Jaderná a Fyzikálně inženýrská

Připojení budovy FJFI v Břehové ulici do CzechLight bylo zprovozněno v listopadu 2005. Optický kabel vede do CISCO SM 1000Base-CWDM 1550nm GBIC v EM316-2GBIC, v druhém GBIC slotu je 1000Base-T síťová karta, která je dočasně pro testování připojena přímo do switchu cílového clusteru. Do konce roku 2005 bude mezi switch clusteru a EM316 zařazen router spojující CzechLight, síť FJFI a síť clusteru. Cílový cluster se skládá z 10 strojů (2xXEON, 2,0GHz, 1GB).

ÚJF Řež

V Řeži je optický kabel přiveden do budovy F1 (oddělení jaderné spektroskopie) a je zakončen v samostatném optickém rozvaděči. Propojen je přes patchcord (3m, konektory E2000 a SC) na 1000LXE GBIC (OptoWay) do switchu CISCO WS-C3550-12T (10x10/100/1000 + 2xGBIC). Tento switch je provozován pouze pro síť CzechLight. V současné době jsou připojeny čtyři servery.

Univerzita Karlova v Praze, Matematicko-fyzikální fakulta

Během první poloviny roku 2005 došlo na Ústavu Částicové a Jaderné Fyziky MFF UK v Tróji (ÚČJF) k plánovanému nahrazení zapůjčeného hardware (CISCO 2950, 1000LX GBIC) novým serverem s gigabitovým optickým rozhraním Intel 82545GM. Kvalita spoje byla testována a gigabitová propustnost linky byla potvrzena.

Nový server obsahuje 64-bitovou architekturu (duální motherboard MSI K8D, procesory Opteron 2.4 GHz, 2 GB RAM, SCSI hot-swap disky), má tedy dostatečnou propustnost na směrování gigabitů i dostatečný výkon, aby se stal základem lokální Tier-3 gridové struktury (LCG, EDG). V současnosti je server využíván také na numerické simulace pracovníků ÚČJF.

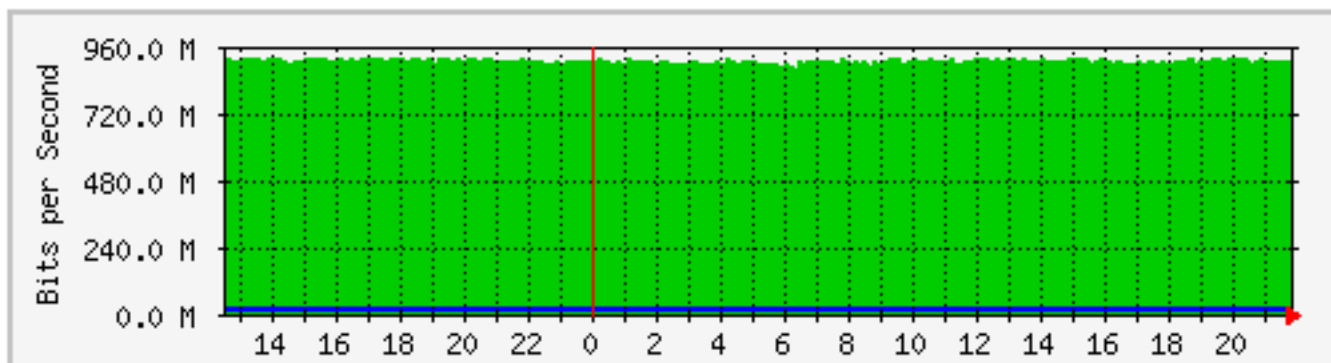
Za účelem výměny směrovacích záznamů v rámci sítě CzechLight běží na serveru směrovací BGP daemon Quagga. Kompatibilita s produkty CISCO a dlouhodobá stabilita provozu byly ověřeny.

Zátěžové testy

K otestování průchodnosti sítě mezi jednotlivými institucemi v Praze byl použit program iperf verze 1.7.0, který byl vždy spuštěn na jednom ze serverů clusteru Goliáš ve FZÚ a současně na serveru, umístěném za routerem v každé připojené instituci (případně přímo na routeru). Ideální velikost TCP okna byla ve všech případech 256kb, zvětšování okna nad tuto hodnotu již nepřineslo žádné významné zvýšení rychlosti. Testy běžely vždy minimálně 48 hodin a v tomto časovém úseku byla ověřena vysoká propustnost, pohybující se dlouhodobě mezi 91,2 – 92,7% nominální kapacity linky 1Gb/s. Na obrázcích níže je výstup z programu MRTG, který přes SNMP získává statistická data ze switchu Cisco Catalyst C6503 a který potvrzuje tyto výsledky.

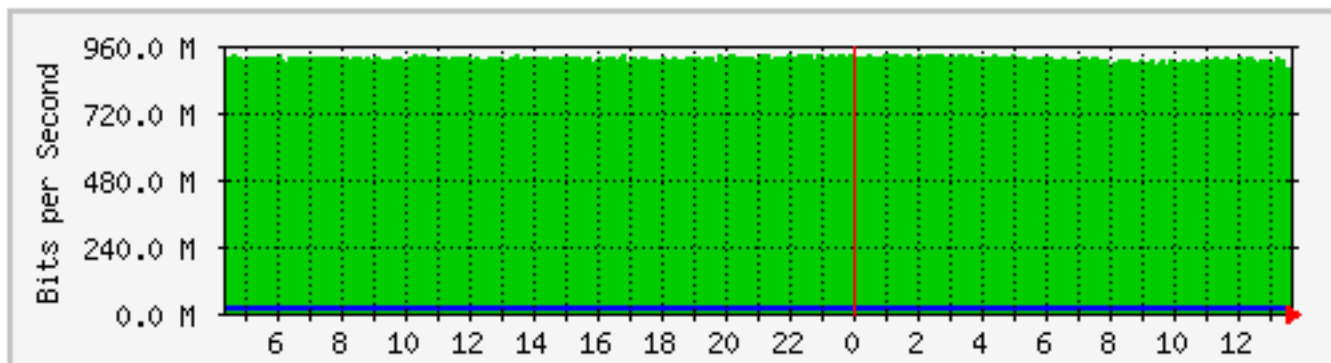
FZÚ – ÚTEF (5 minutový průměr)

Max: 929.9 Mb/s (93.0%) Average 912.4 Mb/s (91.2%)



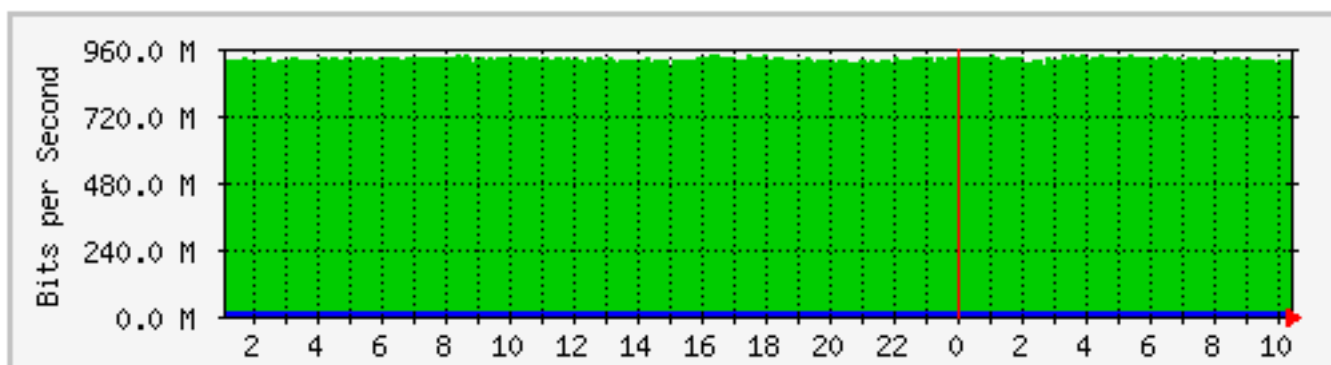
FZÚ – FJFI (5 minutový průměr)

Max: 938.1Mb/s (93.8%) Average 919.9 Mb/s (92.0%)



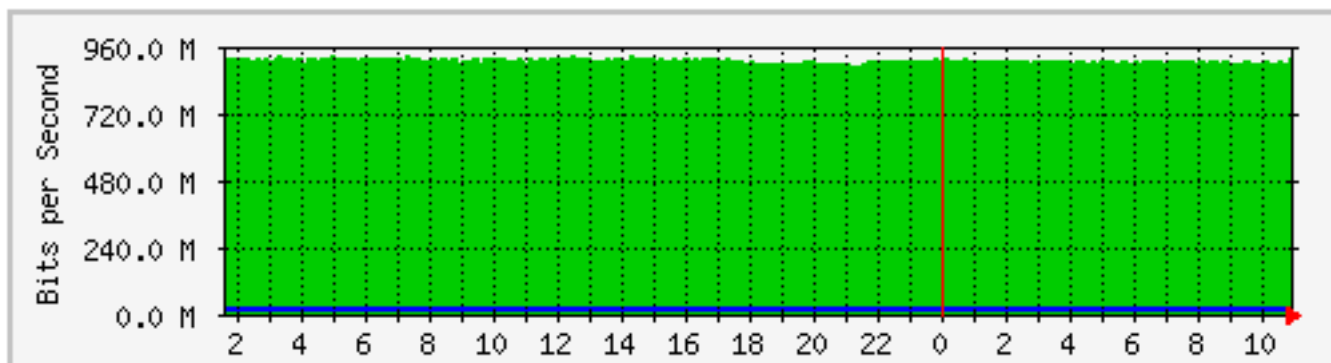
FZÚ – MFF (5 minutový průměr)

Max: 946.4Mb/s (94.6%) Average 927.3 Mb/s (92.7%)



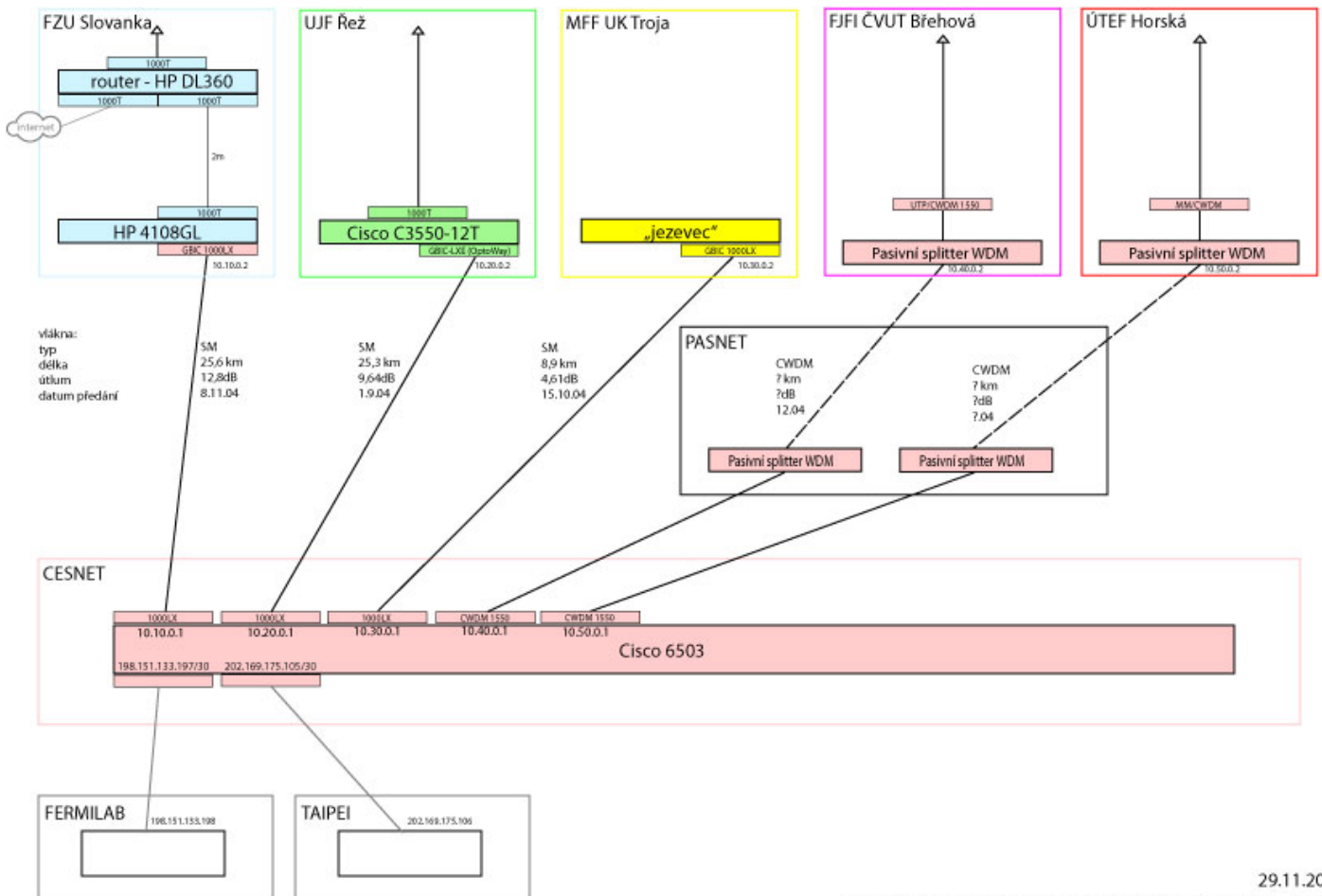
FZÚ – Řež (5 minutový průměr)

Max: 934.8 Mb/s (93.5%) Average 914.8 Mb/s (91.5%)



Propojení jednotlivých pracovišť

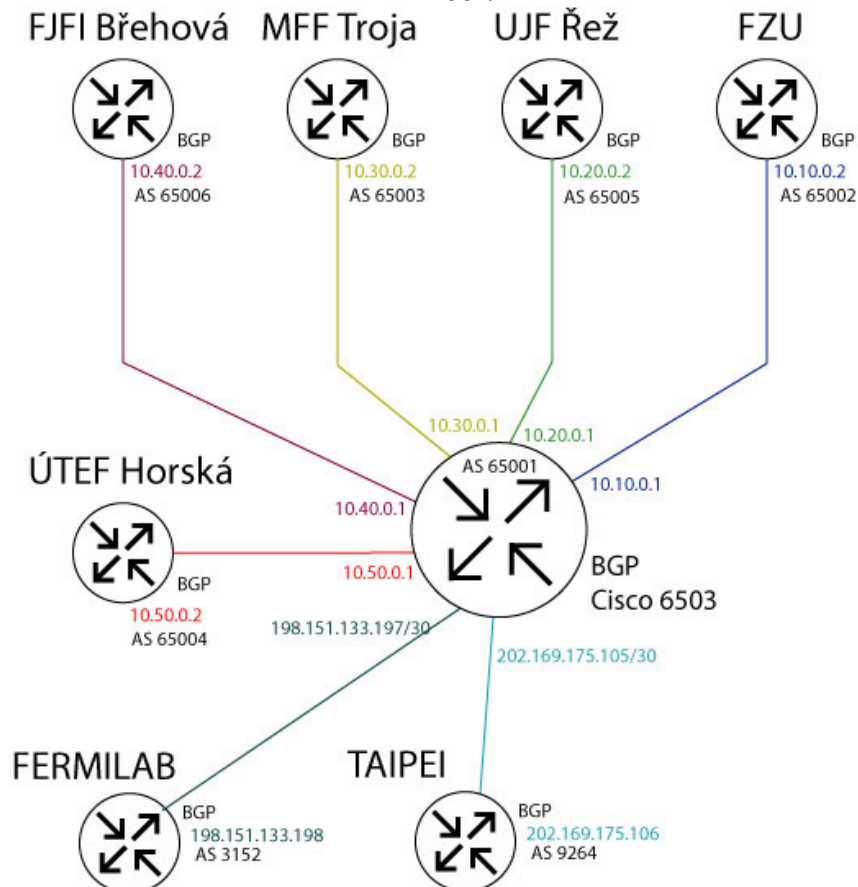
V průběhu roku 2005 bylo v rámci aktivity optické sítě dokončeno zprovoznění vyhrazených SM vláken mezi CESNETem (středem hvězdy) a pracovišti FZÚ Slovanka, ÚJF Řež, UK MFF Troja. Dále byla připojena pracoviště FJFI Břehová a ÚTEF Horská pomocí CWDM ve spolupráci s PASNETem. Na obr.1 je znázorněno stávající propojení jednotlivých pracovišť a také použitý síťový HW. Na obr. 2 je podrobněji rozepsáno nastavení BGP a IP adres.



29.11.2005

http://www.particle.cz/farm/czechlight/dokumenty/CL_hw.pdf

obr. 1



29.11.2005

http://www.particle.cz/farm/czechlight/dokumenty/CL_IP.pdf

Obr. 2.

CWDM v pražské akademické síti

Ing. Kamil Šmejkal, ČVUT-VIC

Při realizaci spojů CWDM v pražské akademické síti PASNET bylo podstatnou podmínkou zachování stávajících propojení, která komunikují na vlnové délce 1310 nm. Tuto možnost současného využití vlnové délky 1310nm zahraniční dodavatelé standardně neposkytovali. To se týkalo nejen dodavatele pro CISCO, ale i dalších renomovaných výrobců. Navíc jejich zařízení, jak multiplexory, tak i OADM, byly mnohonásobně dražší. Naším požadavkům vyšli vstříc tuzemští dodavatelé a podle zadaného požadavku byly vyrobeny potřebné prvky rozdělující vlnové délky. To nám umožnilo realizovat další požadavky na okruhy GE. Byla odzkoušena zařízení a kompatibilita od různých dodavatelů, CISCO systems (CWDM moduly GBIC a SFP), MRV (CWDM GBIC), Optokon Jihlava (MUX), RLC a.s. (MUX, OADM) a DataCOMMS (MUX a OADM). Ukázalo se, že nedošlo k žádné nekompatibilitě, při nasazování je jen nezbytné brát v úvahu rozdílné parametry (u modulů rozhraní v citlivosti a výkonu laseru, u multiplexorů pak rozdílné vložné útlumy). Byla zjištěna jen nefunkčnost u výrobku CISCO TP GBIT při převodu z optiky na kabel TP, který byl nahrazen výrobkem od firmy ProLabs. Použitý modul od firmy NBASE (dva sloty pro GBIC), použito i pro přechod z CWDM na MM. V případě propojení optiky na krátkou vzdálenost přijímač není schopen od určitého výkonu správně vyhodnocovat přenášený signál (příliš velký výkon na vstupu do aktivního prvku). V tomto případě je nutné pro zajištění spolehlivého přenosu použít útlumového prvku. Podstatnou podmínkou pro realizaci spoje před zapojením do provozu je testování a odzkoušení jednotlivých prvků na pracovišti, popřípadě simulace provozu. To má význam v tom, že se předejde velmi náročnému hledání problémů jak v propojení, tak v instalaci zařízení CWDM.

Výsledkem bylo ověření funkčnosti CWDM v pražské metropolitní síti, kde celková přenosová kapacita na jeden pár optických vláken činila 18Gbytu (1x10GE a 8xGE). Pro využití takové kapacity v optické síti je potřeba sdílet infrastrukturu optické sítě, a to z důvodu zajištění větší flexibility, a tím také dosažení větší účinnosti ve vztahu k daným požadavkům na propojení GE okruhů. Nasazením technologie CWDM se v některých případech zvýšila provozní spolehlivost.

Uvedený systém má v současné době dostatečné rezervy do doby, než intenzita provozu bude požadovat další technologii pro zvyšující se datový přenos. V tomto směru lze v pásmu 1530 až 1570 nm nasadit systém DWDM, který může koexistovat v rámci CWDM. Vývoj také ukazuje, že v pásmu okolo 1310 nm mohou být zrealizovány čtyři 10GE okruhy. Nejnověji v Glasgow byl uvedeno 4 Gbit/s CWDM. Další směr vývoje ukazuje na prosazování SDH v metropolitní síti pod názvem Next Generation. Rychlý postup ve vývoji technologie v optických sítích a trendy na principu WDM se tak uplatňují ve sféře akademické sítě. V tomto směru se jedná o technologii propojování datových relací na různých úrovních, jako je **end-to-end** pro vzdálené řízení projektů nebo **point-to-point** na propojování vzdálených lokalit. Služby **LAN-to-LAN** (propojování LAN) mohou tedy být zajištěny jen sítěmi, kde jsou pro to vytvořené podmínky.

