

Presné a jednoduché merení kvalitativních parametru síte

Dr. Sven Ubik, ubik@cesnet.cz

Pri posuzování kvality prenosu dat v pocitacové síti se casto setkáváme s potrebou presne zmerit jednotlivé kvalitativní parametry (propustnost, zpoždění, atd.). Komerční zařízení pro tento účel jsou nákladná, volne dostupné nástroje mívají omezené možnosti a malou presnost. V tomto článku popisujeme řešení presného a soucasne jednoduchého merení kvalitativních parametru síte.

Kvalitativní parametry síte potrebujeme merit predevším ze dvou duvodu. Prvním duvodem je overování dodržování tzv. SLS (Service Level Specification), což je dohoda mezi uživatelem síte a poskytovatelem síťových služeb. Dohoda popisuje jak službu poskytovanou sítí, tak i povolené chování uživatele, to znamená, jaký objem dat muže uživatel poslat do síte. SLS zahrnuje zejména následující informace:

- Vymezení hranicních bodu, mezi kterými je služba poskytována
- Vymezení toku dat, na které se služba vztahuje (například pouze na pakety oznacené urcítým stupnem priority)
- Popis tzv. obálky dat urcující velikost a tvar povoleného datového toku
- Vlastní kvalitativní parametry zajištované službou
- Popis reakce síte v prípade, kdy uživatel pošle datový tok prekracující obálku

Pro správné poskytování služby popsané pomocí SLS je potreba overit jednak zda sít zajištuje predepsané kvalitativní parametry a jednak zda uživatel dodržuje povolenou obálku dat.

Druhým duvodem merení je sledování vlivu ruzných mechanismu pro řízení toku dat v síti (například pridělování šířky pásma, tvarování datového toku, tzv. traffic shaping, atd.) na ruzné vzorky datových toku. Pri tomto sledování jde jednak o obecné jevy, například o chování jednotlivého datového toku pri aplikaci urcítých mechanismu na celou agregaci datových toku (typicky v sítích typu diffserv), tak i o vlastnosti konkrétní implementace, například presnost tvarování datového toku na urcitém smerovací. Pro tato merení je potreba vysoká presnost naměřených hodnot, s casovými odchylkami okolo 4 mikrosekund. To je totiž doba potrebná k odeslání jednoho paketu minimální délky v sítích typu Fast Ethernet nebo Gigabit Ethernet (sít Gigabit Ethernet je 10x rychlejší, ale nejkratší paket je 10x delší, proto je doba jeho odeslání stejná). Dosažení této presnosti je obtížné zejména u jednosmerného zpoždění, protože je potreba presne synchronizovat cas mezi vysílacím a přijímacím bodem.

Existující nástroje a vymezení požadavku

Komerční merící zařízení, krome toho, že bývají velmi drahá, mají casto odesílající port urcený pro generování paketu a přijímající port urcený pro odchytávání paketu umístěné na stejném zařízení. Tím se eliminuje potreba presné casové synchronizace, ale není možné merit mezi dvema fyzicky vzdálenými body, například pres celou sít. Navíc je casto obtížné použít výsledky merení pro další zpracování následným programem.

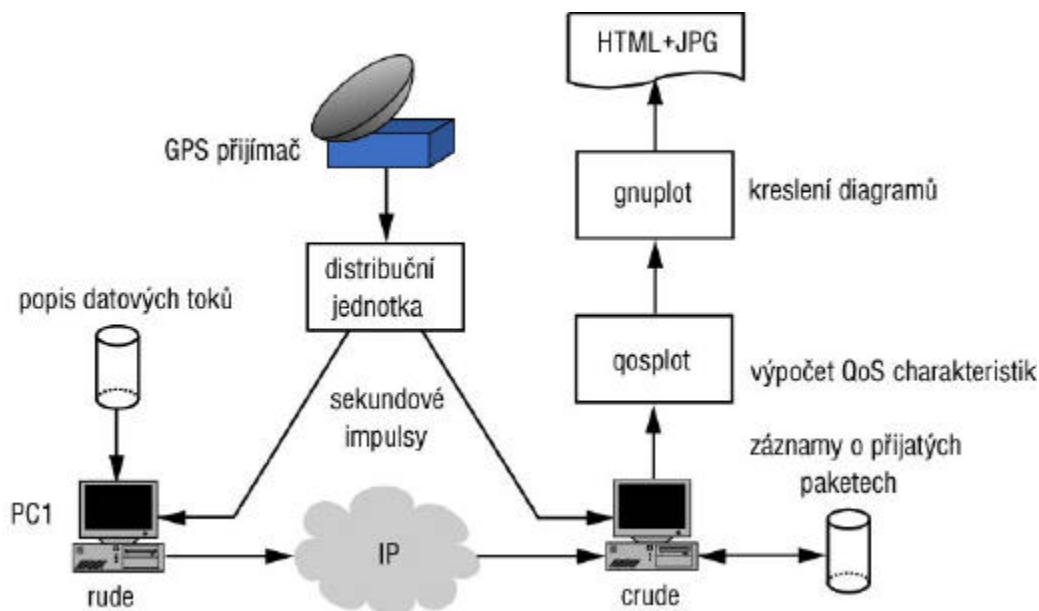
Volně dostupné programy integrující generování paketu s měřením základních kvalitativních charakteristik často umožňují generovat pouze datový tok o maximální průchodnosti (test jaký objem dat je možné přenést přes síť) a uvádějí průměrné hodnoty propustnosti [1], obousměrného zpoždění (round-trip) [2], případně dalších parametrů za dobu trvání testu. Detailní zobrazení průběhu hodnot v čase není k dispozici.

Na základě potřeb uvedených v úvodu článku jsme stanovili následující požadavky na systém pro měření kvalitativních parametrů sítě:

- Měření všech primárních kvalitativních charakteristik, tj. ztrátovost, propustnost, zpoždění, změna zpoždění a distribuce zpoždění a jeho změny
- Výpočet a zobrazení průběhu všech charakteristik v čase s libovolně nastavitelnou časovou granularitou
- Přesnost měření jednosměrného zpoždění lepší než 10 mikrosekund
- Možnost dalšího zpracování naměřených hodnot
- Cenově dostupné řešení

Architektura systému

Architektura systému je znázorněna na obrázku 1. Pro generování a odchyťování paketu jsme použili dvojici programu RUDE/CRUDE [3] umožňující posílání testovacích datových toků s nastavitelnou délkou paketu a počtem paketu za sekundu. Tyto parametry se mohou v průběhu testu libovolně měnit a je možné posílat více různých datových toků současně. Program CRUDE o každém přijatém paketu uloží do souboru krátký záznam s číslem datového toku, číslem paketu a časem odeslání a přijetí paketu.



Obrázek 1: Architektura systému

Důležitým rozhodnutím bylo řešení synchronizace času. Tu lze řešit v principu třemi způsoby. Prvním je použití přesných autonomních hodin v každém uzlu měření, například atomových hodin. Takové hodiny zajišťují vysokou přesnost, jsou však nákladné. Druhým způsobem je použití signálu z GPS přijímače. Ten je určen primárně pro stanovení přesné

polohy na zemi, ale protože k tomu účelu využívá časový signál získaný z přesných atomových hodin v družicích, poskytuje většina GPS přijímaců na svém výstupu i časový signál včetně přesných sekundových impulsů PPS (Pulse Per Second). Třetím způsobem je použití vzájemné synchronizace přes síť protokolem NTP.

Vzhledem k výše uvedeným požadavkům a k potřebě řešení zaručeně nezávislého na aktuálním stavu měřené sítě jsme zvolili použití GPS přijímaců. Dražší přijímaci poskytují impulsy PPS s přesností řádu desítek nanosekund. Pro náš účel postačuje levnější model, například Garmin 35-LVS poskytující impulsy PPS s přesností 1 mikrosekunda v úrovni TTL. Tento signál je převeden na úroveň rozhraní RS-232 a přiveden na sériový port počítače. Počítač používá operační systém Linux s procesem protokolu NTP, který je určen pro příjem časového signálu z různých vnějších zdrojů včetně impulsů PPS. V případě, že všechny uzly měření jsou umístěny v jedné budově (například měření přes lokální síť) a vzdálenost mezi GPS přijímacem a počítačem je řádu desítek metrů, není nutné, aby každý uzel měl svůj vlastní GPS přijímač. Postačí jeden přijímač s tím, že časový signál můžeme rozvést k jednotlivým počítačům prostřednictvím běžně dostupné síťové kabeláže kategorie 5. Poslední propojovací kabel nepřipojíme do síťové karty počítače nýbrž na jeho sériový port. Tato kabeláž sice není určena pro přenos signálu rozhraní RS-232, ale při vzdálenosti řádu desítek metrů je signál na straně počítače dostatečně kvalitní pro jeho bezpečné zpracování.

Skutečnost, kterou je potřeba zvážit, je rozlišení systémových hodin počítače. Standardní PC používá krystalový oscilátor s kmitočtem 14.31818 Mhz, který je vydělen dvanácti na kmitočet 1.193182 Mhz. Tento kmitočet je vstupem do čítače, který je možné programově číst s rozlišením 1/1.193182 MHz, tedy asi 0.838 mikrosekundy. Pro jemnější rozlišení je potřeba rozšířit jádro operačního systému o nanokernel [4]. Ten používá obsah registru TSC (Time Stamp Counter) procesoru Pentium, který počítá hodinové impulsy procesoru. Se současnými procesory je tak možné dosáhnout rozlišení řádu 1 nanosekunda.

Výsledná přesnost synchronizace času mezi jednotlivými počítači je tedy ovlivněna následujícími faktory [5]: přesností impulsů PPS poskytovaných GPS přijímacem, zpožděním v konvertoru TTL/RS-232, dobou šíření signálu v kabelech, dobou zpracování impulsů PPS v počítači, krátkodobým kolísáním kmitočtu krystalového oscilátoru a stabilitou smyčky fázového závěsu. Pokud všechny počítače používají stejný zdroj impulsů PPS rozvedený po kabelech srovnatelné délky, jsou první tři faktory eliminovány. Významné jsou tedy zbývající faktory.

Pro výpočet průběhu kvalitativních charakteristik v case jsme vytvořili program qosplot. Jako vstup přijímá záznamy o přijatých paketech vytvořené programem CRUDE. Jako výstup vytváří souhrnou tabulku údajů za celou dobu měření a především datový a řídicí soubor pro grafický program gnuplot [6]. Ten lze následně použít pro vytvoření sady grafů znázorňujících průběh všech primárních kvalitativních charakteristik v case. Program qosplot může pracovat pouze s vybranými částmi jednotlivých datových toků, posouvat jejich průběh v case, měnit amplitudu propustnosti (pro přepočítání propustnosti mezi dvěma vrstvami sítě, například ze síťové do linkové vrstvy) a počítat kvalitativní charakteristiky s libovolně nastavitelnou časovou granularitou. To je užitečné například při testování tvarování datového toku (traffic shaping), kde u reálných smerovaců je možné nastavit pouze určité diskrétní a neprůliš zaokrouhlené délky tvarovacího intervalu dané vnitřní implementací smerovace.

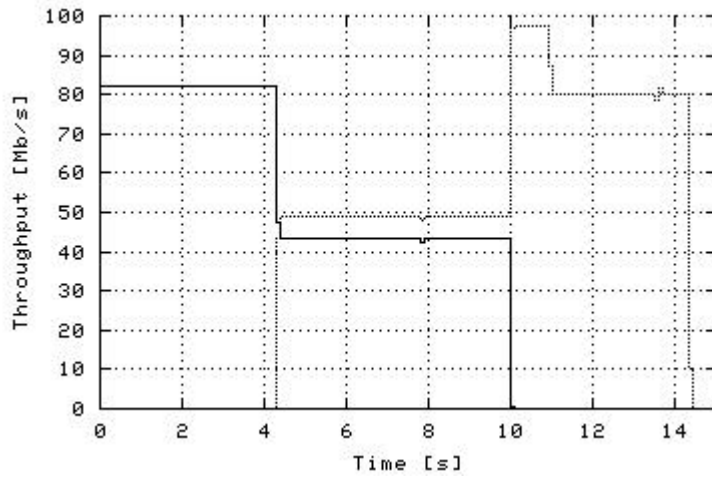
Praktické zkušenosti

Při přímém propojení dvou počítačů krátkým kabelem v technologii Gigabit Ethernet jsme naměřili zpoždění paketu minimální délky (tedy 512 bajtu + 8 bajtu preamble) v rozmezí 22 až 26 mikrosekund v průběhu jedné hodiny. Počítače byly vyhrazeny pro měření, neprobíhaly tedy na nich současně další výpočetní nebo diskové náročné operace. Konstantní část zpoždění je tvořena dobou potřebnou pro zaslání paketu do sítě (asi 4 mikrosekundy) a zpracováním v programech RUDE/CRUDE, v implementaci protokolu UDP/IP a v síťových adaptérech. Proměnlivá část zpoždění je způsobena kolísáním doby zpracování impulsu PPS v počítači a nestabilitou oscilátoru a smyčky fázového závěsu. Konstantní část zpoždění zjištěnou pro danou konfiguraci hardware a software lze od naměřených hodnot odečíst. Z rozsahu proměnlivé části zpoždění vyplývá, že je možné změřit jednosměrné zpoždění s přesností přibližně 2 mikrosekundy.

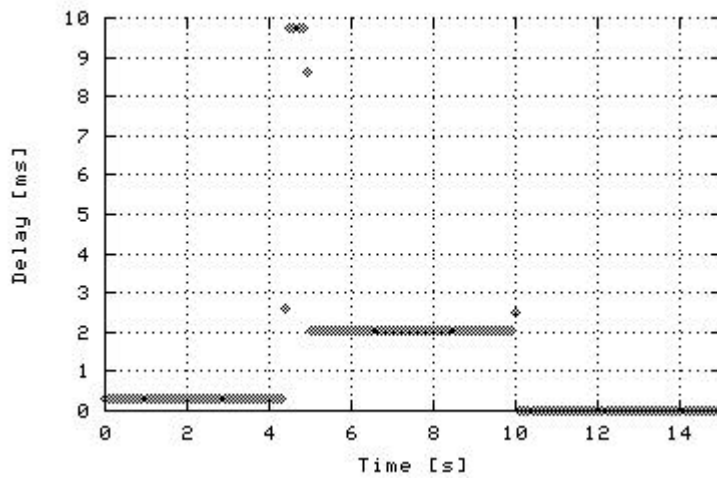
Jako příklad zobrazení charakteristik uvádíme na obrázku 2 průběh propustnosti dvou datových toků 80 Mb/s sdílejících stejnou linku o kapacitě 100 Mb/s. První tok s pakety o délce 256 bajtu byl spuštěn v case 0 až 10 sekund a je znázorněn plnou čarou. Druhý tok s pakety o délce 1500 bajtu byl spuštěn v case 5 až 15 sekund a je znázorněn prerušovanou čarou. V case 5 až 10 sekund tedy oba toky sdílely kapacitu linky. Na výstupním portu příslušného směrovace byl zapnut mechanismus WRR (Weighted Round Robin), kterým byla každému datovému toku přidělena kapacita 50 Mb/s. Mechanismus WRR se postaral o to, že každý tok skutečně obdržel kapacitu přibližně 50 Mb/s. Bez tohoto mechanismu by propustnost obou toků pseudo-náhodně kolísala. Je zároveň vidět, že tok s pakety o délce 1500 bajtu získal o něco větší kapacitu než tok s pakety o délce 256 bajtu. To je vlastností mechanismu WRR, který zvýhodňuje toky s většími pakety. Tento problém je odstraněn například u mechanismu MDRR (Modified Deficit Round Robin), který však u daného směrovace nebyl k dispozici. Obrázek 3 znázorňuje průběh zpoždění prvního toku.

Obrázek 4 znázorňuje distribuci zpoždění v impulzním datovém toku o objemu 5 Mb/s posílaném v dávkách o objemu 20 Mb/s v délce 100 ms s přestávkami v délce 300 ms, na který bylo aplikováno tvarování (traffic shaping) pro vyrovnání impulzního charakteru datového toku. Každá vstupující dávka dat byla směrovacem převedena na 4 menší dávky dat, kterým odpovídají shluky zpoždění znázorněné v distribučním grafu.

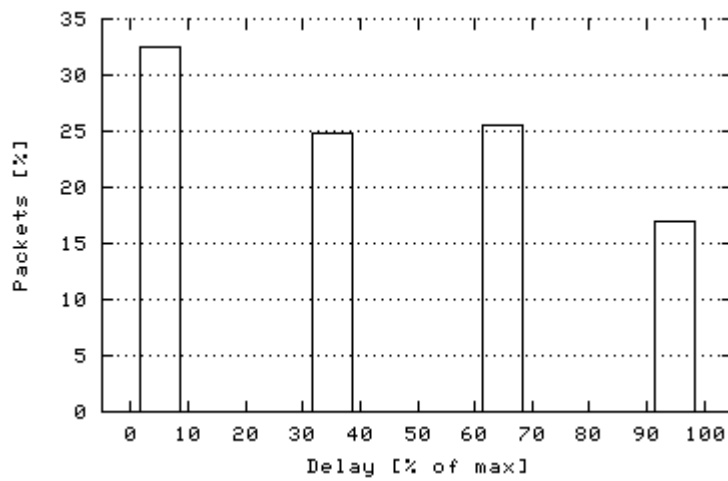
Popsaný systém byl použit pro celou řadu dalších měření v projektu QoS v IP [7].



Obrázek 2: Prubeh propustnosti



Obrázek 3: Prubeh zpoždění



Obrázek 4: Prubeh distribuce zpoždění

- [1] „Netperf: A Network Performance Benchmark“, <http://www.netperf.org>.
- [2] „ping“. <http://www.ping127001.com/pingpage.htm>.
- [3] Juha Laine, Sampo Saaristo, Rui Prior. “RUDE & CRUDE: Real-time UDP Data Emitter and Collector”, <http://www.atm.tut.fi/rude/>.
- [4] D.L. Mills, P.-H. Kamp. “The nanokernel”, Proc. Precision Time and Time Interval (PTTI) Applications and Planning Meeting, Reston, VA, 2000.
- [5] Sven Ubik, Vladimír Smotlacha (Cesnet), Sampo Saaristo (Tampere University of Technology), Juha Laine (Soon Communications). „Low-cost Precise QoS Measurement Tool“, technická zpráva Cesnet 7/2001, <http://www.cesnet.cz/doc/techzpravy/2001/07>.
- [6] Thomas Williams, Colin Kelley. „gnuplot – plotting program“, <http://www.gnuplot.info>.
- [7] Projekt QoS v IP. <http://www.cesnet.cz/english/project/qosip>.