

Vad kan man använda ett bra nät till?

Börje Josefsson
SUNET

bj@sUNET.se

Internetdagarna 2004

Vad menas med bra?

- Bra fråga 😊

- Bra tillgänglighet
- Bra överföringskvalitet
- Bra hastighet

GigaSunet

Rikstäckande 10 Gbit/s nätverk

- Stomnät (22 städer):
 - 5.270 km DWDM.
 - 27 ortssammanbindande segment med 10 Gbit/sek.
 - 49 routrar (Cisco 124xx).
- Accessnät:
 - 25 svartfiber-ringar.
 - 4 segment med 2,5 Gbit/sek. DWDM.
 - 70 routrar (Cisco 10720).



Kvalitetsparametrar i GigaSunet

Dynamic metrics, **end-to-end** in core network:

Jitter (during 15 minutes)	≤ 5 ms
RTT	≤ 60 ms
Packet reordering	≤ 1 %
Packet loss	≤ 0.05 %

Static metrics, or metrics measured per PoP

Component	Core	Access (DPT layer)
Burst absorption capacity (BW delay component)	■ 10 Gbps / 410 ms ■ 2.5 Gbps / 1 220 ms	■ 2.5 Gbps / 220 ms
MTU	8 192 Bytes	4 096 / 8 192 Bytes
Forwarding outages (against each PoP)	< 7 min/month	< 14 min/month

Varför ska det då vara så svårt?

Ljushastigheten är begränsande

- För TCP – överföringar måste sändaren lagra varje paket tills en ACK från mottagaren har tagits emot.
- På grund av ljushastighetens begränsning så tar detta en stund, även i ett litet land som Sverige.
- RTT Luleå–Stockholm är cirka 20 ms.
- TCP "*window size*" för att klara 1 Gbit/sek måste då vara $(1000/8)*0,02 = 2,5$ Mbyte.
 - De flesta operativsystem har standardinställd storlek på 32–64 kbyte(!), dvs. 40–80 ggr för lite!
 - Med standardinställningar blir farten Luleå–Stockholm bara 3,5 Mbit/sek ☹
 - Det hjälper alltså inte bara med att ha ett gigabitnät!

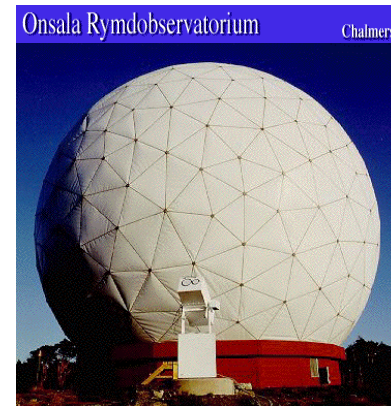
Paketförluster

- På en höghastighetsförbindelse över långa avstånd kommer paketförluster i ett TCP-flöde att mycket drastiskt sänka överföringshastigheten.
- Om sändaren går in i “*congestion avoidance*” efter en paketförlust (> 2 paket) så kommer återhämtningen till full fart igen att bli mycket långsam.
- Med RTT på 185 ms och “window size” på 25 Mbyte så kommer det då att ta cirka 50 minuter(!) att nå full fart igen.
- Paketförluster måste alltså undvikas till varje pris!

Användningsområden

Radioastronomer över hela världen

- "Sammankoppling" av radioteleskop över hela klotet ger betydligt bättre upplösning än vad de enskilda teleskopen har.
 - Very Long Baseline Interferometry – VLBI.
 - Måste veta avståndet mellan teleskopen på delar av millimeter när...
- Vitt brus – går inte att komprimera.
- Onsala rymdobservatorium utanför Göteborg är bäst i Europa (åtminstone...)
 - Idag: 512 Mbit/sek.
 - Om två år: 4 Gbit/sek.
 - Om 5–6 år: 32 Gbit/sek.



SweGRID

- Sammankoppling av beräkningscentra runt om i Sverige.
 - Umeå, Uppsala, Stockholm, Linköping, Göteborg, Lund.
- Idag: 1 Gbit/sek till vardera, kopplat till GigaSunet.
- 2007 kommer CERN att starta sin nya anläggning (LHC), då kommer dessa beräkningscentra att behöva mer kapacitet.

Andra krävande tillämpningar

- Institutet för Rymdfysik i Uppsala
 - Projektet LOIS – a LOFAR Outrigger In Scandinavia
 - Transmission av data från sensornätverk i Växjö till beräkningsdatorer i Uppsala.
 - Sensorerna generar data med upp till 2 Gbit/s per sensor
 - Det är hundratal sensorer...
- Patologi – överföring av bilder från mikroskop
 - 320–400 Tbyte per fall.
 - 20–30 fall per dag för en patolog.
 - 1 Pbyte (10^{15}) per dygn $\approx$ 100 Gbit/s per person(!)

Produktionskvalitetsvideo

- Forskningsprojekt 2004–2005.
- Deltagare:
 - LTU (Akademiskt mediacentrum i Piteå)
 - SUNET
 - CERNET (Kinas "SUNET")
 - Tsinghua University i Peking
 - Cisco
- Videoöverföring på långdistans över Internet
 - "Hem till studion", inte ut till konsument.
 - Inför olympiaden 2008...



... samt naturligtvis allt annat

Distansundervisning *Databaser* *IM*
SMTP
IP-telefoni *Web* *Chat* *RFC1149*
Fjärrstyrning *NTP* *Radio* *SNMP*
FTP
Musik *Backup* *Fildelning* *NFS*
Fjärrövervakning *Annat* *Megacom*
Fjärranalys *Tidssynkronisering*
NetNews *Videokonferens* *Video* *Multicast*
Plugh *TV* *Mail* *Datainsamling* *MIME*
Google
Laborationer *Outsourcing*
Samkörning

Internet Land Speed Record

Internet Land Speed Record

*"To realize Internet's full potential,
end-to-end network performance needs to
take a huge leap forward"*

– Gordon Bell

LSR - bakgrund

- Det blir allt svårare att skicka data (med TCP) över Internet när avståndet ökar.
- Internet2 i USA instiftade pris för att uppmuntra försök och forskning.
- Resultaten mäts i (peta-) bitmeter per sekund.
 - Avstånd (enkel väg) * hastighet
 - $1000 \text{ km} * 1 \text{ Gbit/sek} = 1 \cdot 10^6 * 1 \cdot 10^9 = 10^{15} = 1 \text{ Pbmps}$

Vill du också ha ett världsrekord?



... och hamna i Guinness rekordbok?

Science and Technology/Internet/Internet2 Speed - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.guinnessworldrecords.com/content_pages/record.asp?recordid=58445

Kalix: Säljes

Additional plugins are required to display all the media on this page. Install Missing Plugins...



HUMAN BODY AMAZING FEATS NATURAL WORLD SCIENCE & TECH ARTS & MEDIA MODERN SOCIETY TRAVEL & TRANSPORT SPORTS & GAMES

SCIENCE AND TECHNOLOGY << INTERNET << INTERNET2 SPEED

Internet2 Speed Record

On 14 April 2004 a team of engineers from Sprint Corporation in San Jose, California, USA, and the Swedish National Research and Education Network successfully transmitted 838.86 gigabytes of data over a distance of 16,343 km (10,155 miles) in 1,588 seconds. The resulting record value of 69,073 terabit metres-per-second was achieved over public networks using IPv4. The start and end points of the transmission were San Jose, California, USA, and the University of Luleå, Luleå, Sweden.

WHO: Luleå University of Technology & Sprint

WHEN: April 14, 2004

WHERE: Luleå, Sweden

RELATED RECORDS

- MOST QUESTIONS ASKED ONLINE
- MOST CONDOLENCES ON THE INTERNET
- LARGEST DOMAIN OWNERSHIP
- SMALLEST WEB SERVER
- MOST CYBERSTAR VARIATIONS
- BEST-SELLING MP3 PLAYER
- LARGEST FREE EMAIL PROVIDER

NEW

... Enkelt – Gör bara så här:

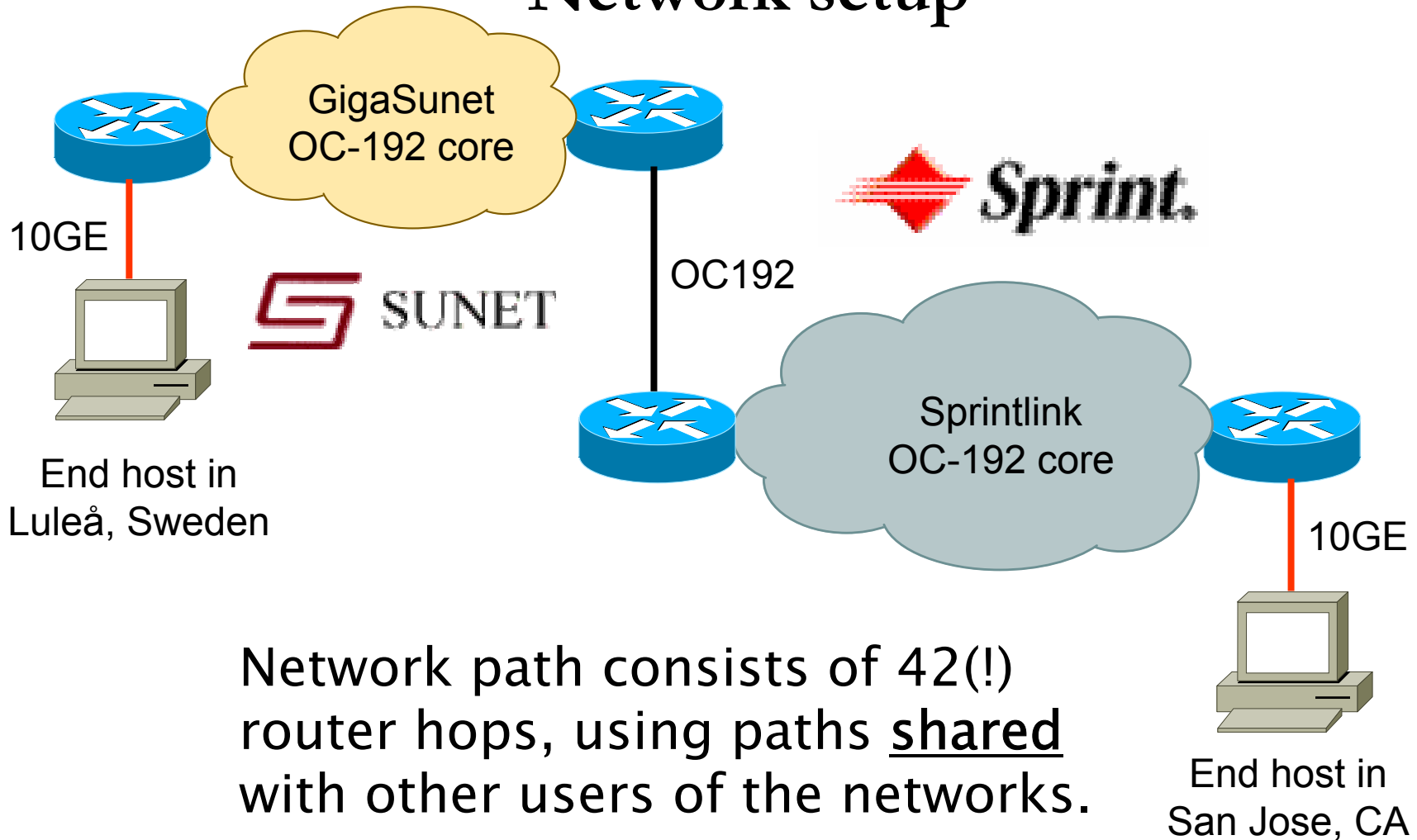
■ Recept:

- ❑ Skaffa minst 30 000 km fiberkabel.
- ❑ Anslut utrustning av riktigt hög kvalitet.
- ❑ Använd datorer kapabla att sända och ta emot data tillräckligt fort.
- ❑ Välj ett operativsystem utan fåniga flaskhalsar.
- ❑ Hitta entusiaster som spenderar kvällar och helger för att sända ohemula mängder data mellan t.ex. Luleå och San Jose ☺

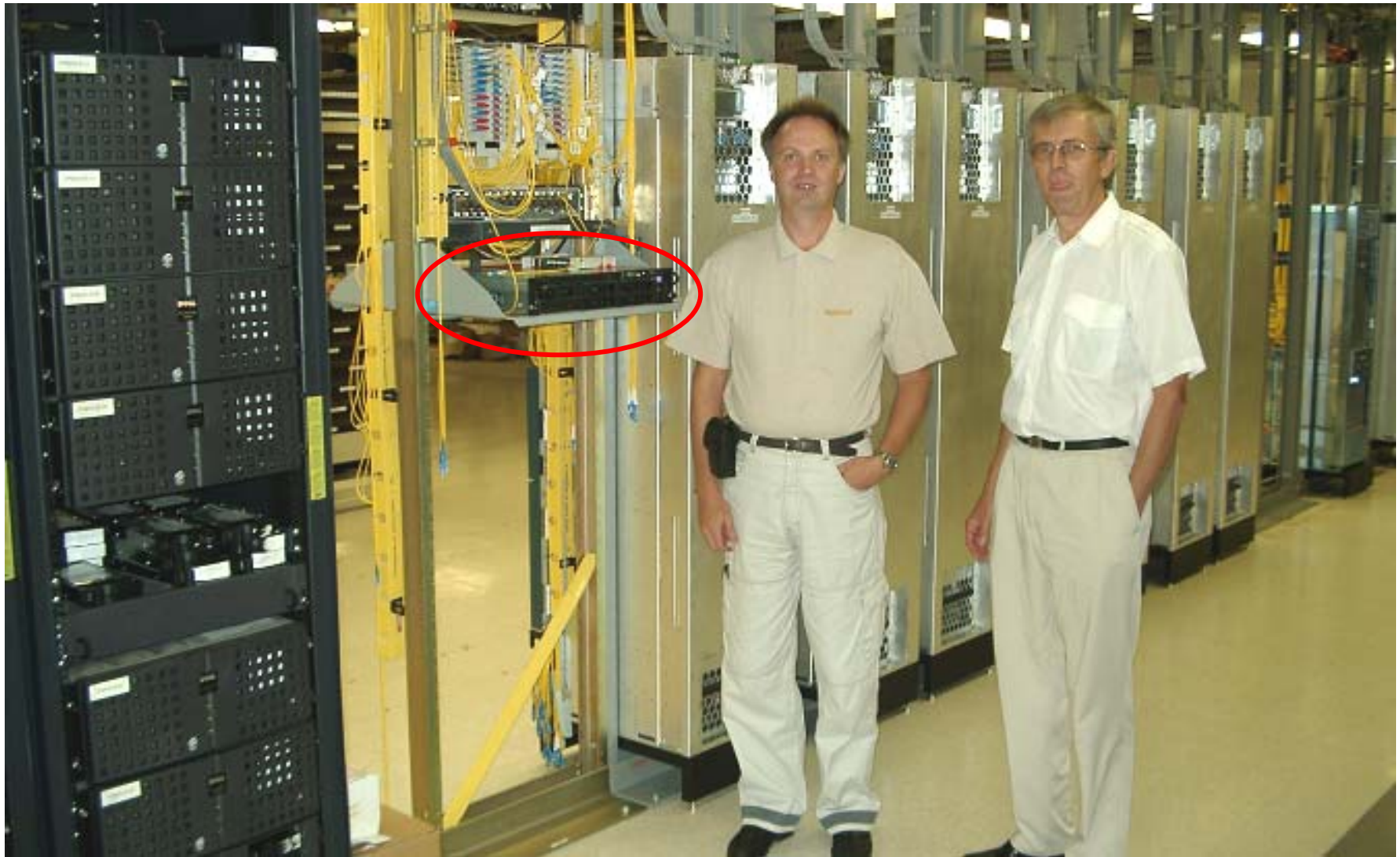
Så här gjorde vi:

SUNET Internet Land Speed Record

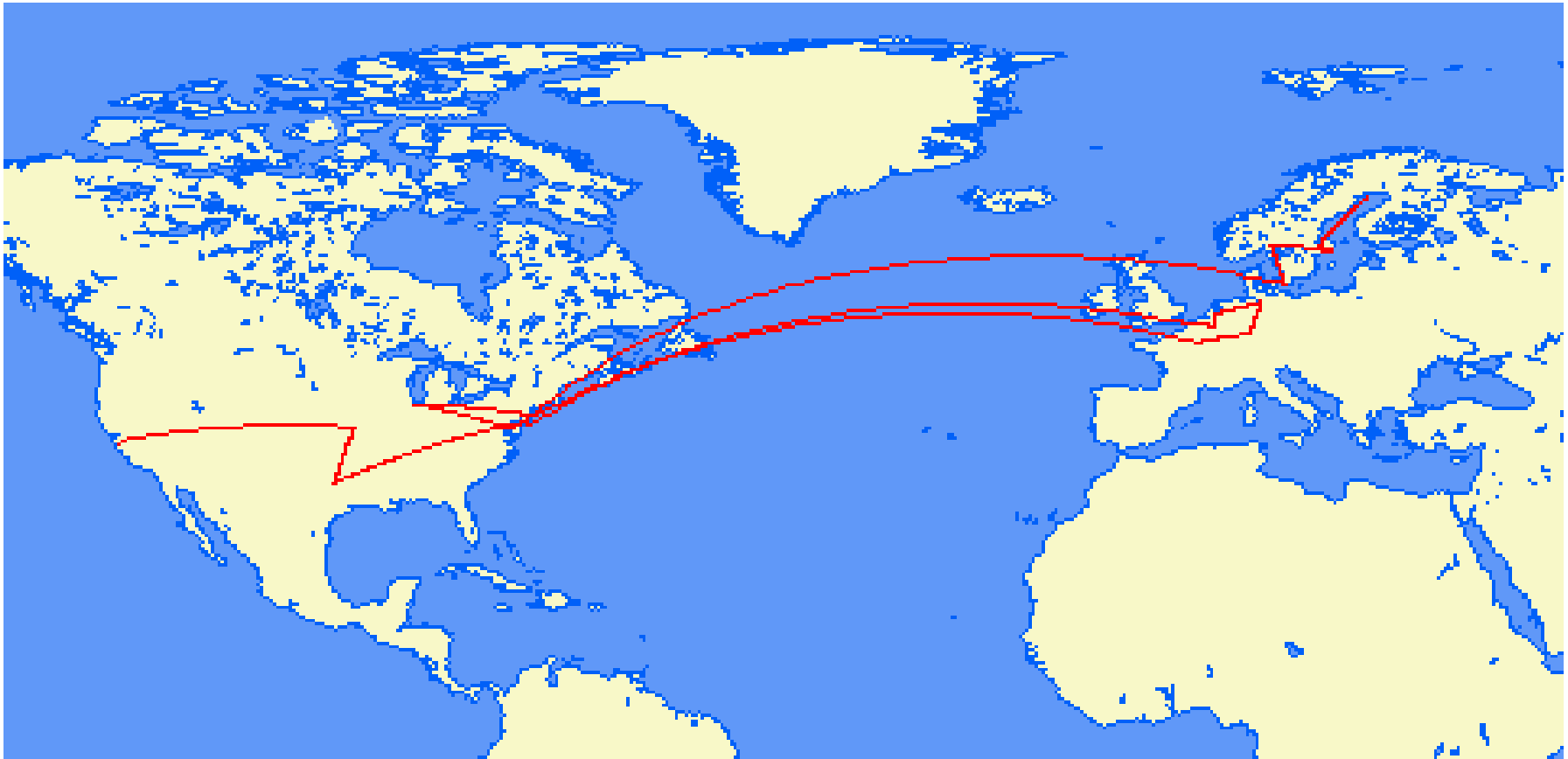
Network setup



Installation i San Jose



Vägval (”traffic engineering” ☺)

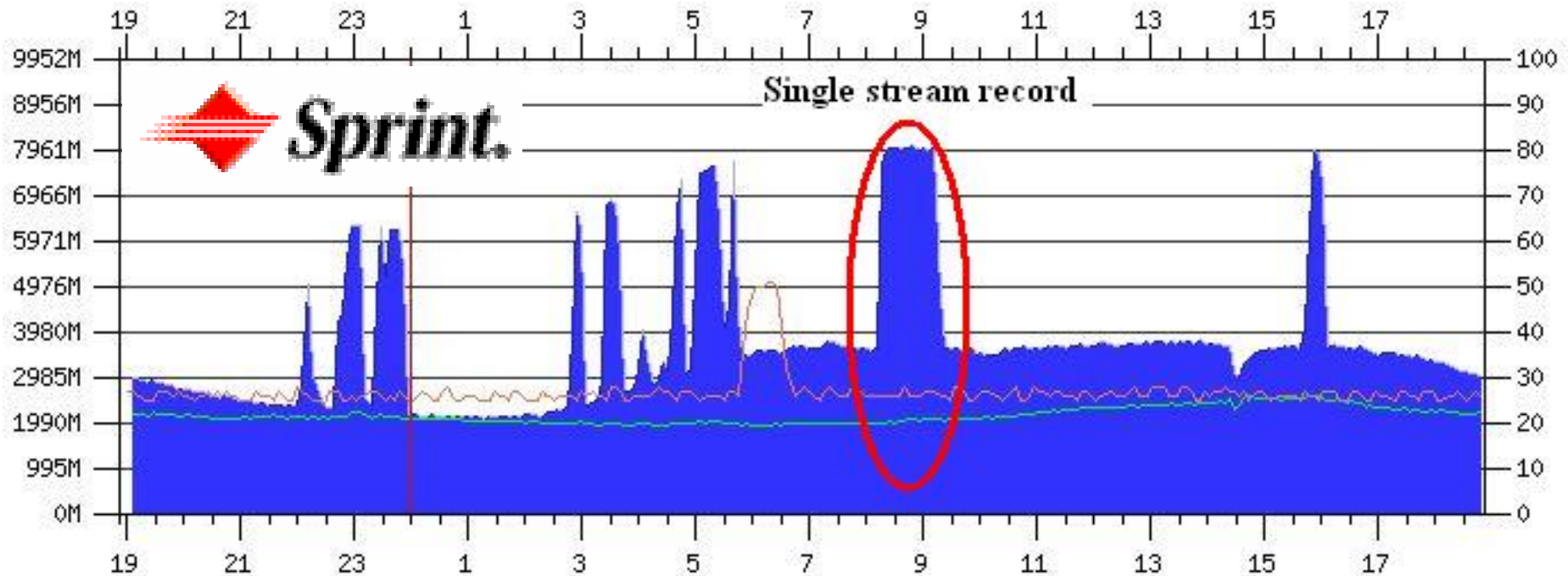


Avståndet från Luleå till San Jose är (enligt regelverkets sätt att räkna) 28 983 km .

SUNETs rekord från 12 september:

- 1 966 080 000 000 bytes på 3 648,81 sekunder = 4 310,62 Mbit/sekund.
- $\approx$ 1 831 Gbyte (2 800 CD-skivor) på en timme.
- 120 000 paket/sekund med MTU på 4 470 bytes.
- Avstånd: 28 983 km.
- Vårt rekord i IPv4 single-stream-klassen blir då 123,93 petabitmeter per sekund.

Nätverkslast



Jämfört med tidigare rekord

- ❑ Mindre kraftfulla datorer i ändarna.
- ❑ 200%(!) längre avstånd.
- ❑ Mindre än halva MTU-storleken (medför högre CPU-last på datorerna).
- ❑ Vi använde inga “privata” förbindelser, utan dom normala produktionsnäten GigaSunet och Sprintlink.

Paketförluster?

42 IP-router-hopp.

Inte ett enda paket tappades under en timme!

I själva verket så hade vi **inte ett enda bitfel** på denna timme, dvs. 15 728 640 000 000 bits (!) överfördes korrekt (då räknar jag ändå inte annan trafik på nätet).

Bit error rate på hela systemet var alltså **bättre** än $6,35 \cdot 10^{-14}$

1 sl-bbh-sj.sprintlink.net (198.67.129.33) 2.554 ms
2 sl-bbh-stk-1-0.sprintlink.net (144.232.8.30) 3.927 ms
3 sl-bb22-stk-13-0.sprintlink.net (144.232.4.45) 2.362 ms
4 sl-bb21-stk-15-0.sprintlink.net (144.232.4.241) 2.496 ms
5 sl-bb22-kc-2-0.sprintlink.net (144.232.20.168) 42.365 ms
6 sl-bb21-kc-6-0.sprintlink.net (144.232.2.133) 42.429 ms
7 sl-bb26-fw-13-0.sprintlink.net (144.232.8.63) 53.072 ms
8 sl-bb27-fw-15-0.sprintlink.net (144.232.11.86) 53.176 ms
9 sl-bb20-pen-13-0.sprintlink.net (144.232.8.64) 103.180 ms
10 sl-bb21-pen-14-0.sprintlink.net (144.232.16.34) 103.314 ms
11 sl-bb23-rl-0-0.sprintlink.net (144.232.20.32) 105.730 ms
12 sl-bb27-rl-10-0.sprintlink.net (144.232.14.142) 105.512 ms
13 sl-bb23-chi-12-0.sprintlink.net (144.232.20.184) 125.224 ms
14 sl-bb24-chi-15-0.sprintlink.net (144.232.26.101) 125.547 ms
15 sl-bb25-nyc-5-0.sprintlink.net (144.232.9.157) 148.753 ms
16 sl-bb24-nyc-10-0.sprintlink.net (144.232.13.181) 148.658 ms
17 sl-bb23-nyc-4-0.sprintlink.net (144.232.13.169) 148.627 ms
18 sl-bb20-par-11-0.sprintlink.net (144.232.20.44) 219.872 ms
19 sl-bb21-fra-13-0.sprintlink.net (213.206.129.66) 369.752 ms
20 sl-bb20-fra-15-0.sprintlink.net (217.147.96.33) 229.642 ms
21 sl-bb21-ham-14-0.sprintlink.net (213.206.129.62) 403.593 ms
22 sl-bb20-ham-15-0.sprintlink.net (217.147.96.45) 239.534 ms
23 sl-bb21-ams-14-0.sprintlink.net (213.206.129.49) 254.658 ms
24 sl-bb20-ams-15-0.sprintlink.net (217.149.32.33) 245.388 ms
25 sl-bb21-bru-14-0.sprintlink.net (213.206.129.46) 248.710 ms
26 sl-bb20-bru-15-0.sprintlink.net (80.66.128.41) 261.456 ms
27 sl-bb22-lon-13-0.sprintlink.net (213.206.129.41) 456.846 ms
28 sl-bb23-lon-15-0.sprintlink.net (213.206.128.161) 253.476 ms
29 sl-bb21-lon-13-0.sprintlink.net (213.206.128.55) 390.525 ms
30 sl-bb21-tuk-10-0.sprintlink.net (144.232.19.69) 322.420 ms
31 sl-bb20-tuk-15-0.sprintlink.net (144.232.20.132) 322.506 ms
32 sl-bb20-msq-10-0.sprintlink.net (144.232.20.172) 324.196 ms
33 sl-bb21-msq-15-0.sprintlink.net (144.232.9.110) 324.173 ms
34 sl-bb20-cop-14-0.sprintlink.net (144.232.19.30) 401.493 ms
35 sl-bb21-olo-13-0.sprintlink.net (213.206.129.72) 409.121 ms
36 sl-bb20-olo-15-0.sprintlink.net (80.77.104.32) 409.234 ms
37 sl-bb20-sto-14-0.sprintlink.net (213.206.129.30) 416.407 ms
38 sl-tst1-sto-0-0.sprintlink.net (213.206.131.10) 416.737 ms
39 stockholm1.POS14.sunet.se (130.242.94.221) 416.692 ms
40 vasteras1-pos4.sunet.se (130.242.82.10) 418.342 ms
41 gavle1-pos4.sunet.se (130.242.81.49) 421.155 ms
42 lulea1-pos0.sunet.se (130.242.81.42) 436.836 ms
43 dino.dc.ltu.se (130.242.94.246) 436.656 ms

Routerhopp som besvärar?

Fördröjning i de inblandade komponenterna:

~17 800 km undervattensfiber	95,4 ms
~18 020 km fiber på land	96,6 ms
~ 4 800 km dispersionskompensering	25,7 ms
42 routrar @ 12 μ s (worst case, normalt 7 μ s)	0,5 ms
Total fördröjning (enkel väg)	218,2 ms

Routrarna svarar alltså för mindre än 0,2% av fördröjningen...

Mer att läsa...

- <http://proj.sunet.se/LSR/>
 - Describes how the Land Speed Records were achieved
- <http://proj.sunet.se/E2E/>
 - About end-to-end performance in GigaSunet

Frågor?

Börje Josefsson

<bj@sunset.se>

