How PERTs Can Help With Network Performance Issues

Simon Leinen (SWITCH)
<simon@limmat.switch.ch>

End-to-End Performance Issues

SWITCH
The Swiss Education & Research Network

Discrepancy between user expectations and perceived performance

- You tell me we have gigabit networks, yet I only get <low number> Mb/s?

Networks span many administrative domains

- Finger pointing loops
  - “my network is fine, see [ping shows no loss] so it must be that other network”
  - “our network is also fine [all lights are green!] so it must be ...”
  - “... is fine so it must be you!”

Performance results from multi-layer interactions

- Application and protocol design
- End-system issues (OS, I/O, other internal bottlenecks)
- Network path (delay, loss, reordering)

The interactions are often difficult to understand!

Approaches to Improve E2E Performance

SWITCH
The Swiss Education & Research Network

- Training (users, programmers, admins, network eng/ops)
  - Examples: Internet2 Performance Workshops
- Instrumentation (network & end-system stacks)
  - Examples: RIPE TTM, PerfSONAR, Web100, NDT, NPAD
- Support infrastructure:
  PERT (Performance Enhancement/Response Team)

PERT History

SWITCH
The Swiss Education & Research Network

2001: Thought experiment in the Internet2 community

2003: Trial PERT (some European NRENs)

2004: GN2 Pilot PERT

Since March 2005: GN2 PERT as “production” service

Several individual NRENs have PERT groups and services emerging.

GN2 PERT Organization

SWITCH
The Swiss Education & Research Network

GN2 is a (partly) EU-funded project that includes the GEANT2 network.

The “SA3” (Service Activity 3) work package on “End-to-end Quality of Service” includes PERT activities.

Participants include DANTE and many NRENs: GARR (IT), RENATER (FR), RCCN (PT), SWITCH (CH), PSNC (PL), CESNET (CZ),. ..

Case Managers are responsible for progress with open issues (cases):

- *Duty Case Managers* change weekly, handle new cases, track progress
- *Special Case Managers* adopt a specific case and handle it until closure

Subject Matter Experts support the PERT with their expertise

- *Mostly by commenting on the pert-discuss mailing list*

Support systems:

- *PERT Ticket System (PTS) on https://www.pert.geant2.net/pert/*
- *Mailing lists: pert-report@geant2.net, pert-discuss@geant2.net*
- *PERT Knowledge Base Wiki on http://kb.pert.switch.ch/*

GN2 PERT Ticket System (PTS)

SWITCH
The Swiss Education & Research Network

- Custom system developed for the needs of the GN2 PERT
- J2EE Web application with SQL back-end database
  - Access through X.509 client certificate for Cms/SMEs
  - Username/password administration for cust^H^H^H^Husers
- Developed at PSNC (Poznań, PL)
- Operational since February 2005, PTSv2 deployed in August 2006

GN2 PERT Ticket System (PTS)

SWITCH
The Swiss Education & Research Network

GÉANT2
PERT Ticket System

GÉANT2

Information Society and Media

Managed by POZNAN

Logged user: **Simon Leinen** Logout

PTS Home | PERT Cases | PERT Diary | PERT Schedule | PERT Users | Settings | Search

- Create a Ticket
- View Tickets
- Assigned Tickets
- Active tickets
- Updated Tickets
- Favourites
- Pending Tickets
- Keywords
- Advanced Search

PTS Home > PERT Cases > Active tickets

Active tickets

↑ id ↓	↑ Subject ↓	↑ State ↓	↑ Date ↓	↑ Author ↓	Notes	↑ Last Modified ↓	
4	Performance Problems between FNAL and DESY	waiting for customer action	2006-01-24	Francois-Xavier Andreu	117/117	2006-02-06	[Show]
14	Below expected throughput, Budapest to New York [summary]	waiting for third party action	2006-08-29	Toby Rodwell	11/11	2006-08-29	[Show]
18	Slow file transfers between TIGO (CL) and JIVE (NL)	waiting for customer action	2006-09-18	Toby Rodwell	24/24	2006-09-18	[Show]

1

Entries per page 10

top of page

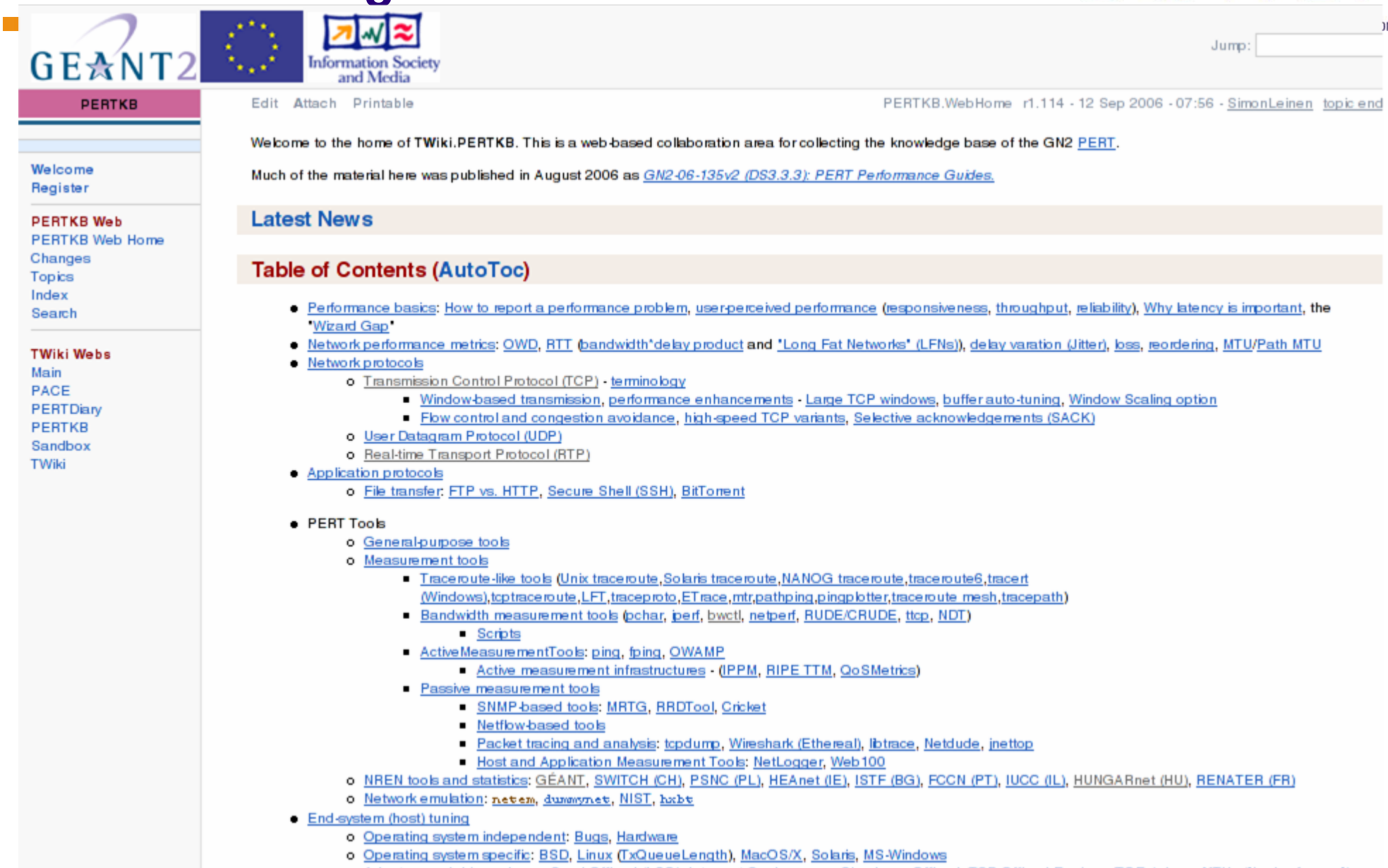
Credits | Legal | Contact | Help

PERT Knowledge Base

SWITCH
The Swiss Education & Research Network

- Tool for managing knowledge gathered by the PERT
- Currently implemented as a Wiki (open registration, anti-spam)
  - http://kb.pert.switch.ch/ aka http://pace.geant2.net/cgi-bin/twiki/view/PERTKB/WebHome
- Should become useful resource for network users/admins
- Content from the PERT KB sometimes extracted to “deliverables”
  - DS3.3.3 (August 2006) -

PERT Knowledge Base



GÉANT2 Information Society and Media

Jump:

PERTKB

Edit Attach Printable

PERTKB.WebHome r1.114 - 12 Sep 2006 - 07:56 - SimonLeinen topic end

Welcome
Register

PERTKB Web
PERTKB Web Home
Changes
Topics
Index
Search

TWiki Webs
Main
PACE
PERTDiary
PERTKB
Sandbox
TWiki

Welcome to the home of TWiki.PERTKB. This is a web-based collaboration area for collecting the knowledge base of the GN2 PERT.

Much of the material here was published in August 2006 as *GN2-06-135v2 (DS3.3.3): PERT Performance Guides.*

Latest News

Table of Contents (AutoToc)

- Performance basics: How to report a performance problem, user-perceived performance (responsiveness, throughput, reliability), Why latency is important, the "Wizard Gap"
- Network performance metrics: OWD, RTT (bandwidth*delay product and "Long Fat Networks" (LFNs)), delay variation (Jitter), loss, reordering, MTU/Path MTU
- Network protocols
  - Transmission Control Protocol (TCP) - terminology
    - Window-based transmission, performance enhancements - Large TCP windows, buffer auto-tuning, Window Scaling option
    - Flow control and congestion avoidance, high-speed TCP variants, Selective acknowledgements (SACK)
  - User Datagram Protocol (UDP)
  - Real-time Transport Protocol (RTP)
- Application protocols
  - File transfer: FTP vs. HTTP, Secure Shell (SSH), BitTorrent
- PERT Tools
  - General-purpose tools
  - Measurement tools
    - Traceroute-like tools (Unix traceroute,Solaris traceroute,NANOG traceroute,traceroute6,tracert (Windows),tcptraceroute,LFT,traceproto,ETrace,mtr,pathping,pingplotter,traceroute mesh,tracepath)
    - Bandwidth measurement tools (pchar, iperf, bwctl, netperf, RUDE/CRUDE, ttcp, NDT)
      - Scripts
    - ActiveMeasurementTools: ping, fping, OWAMP
      - Active measurement infrastructures - (IPPM, RIPE TTM, QoSMetrics)
    - Passive measurement tools
      - SNMP-based tools: MRTG, RRDTool, Cricket
      - Netflow-based tools
      - Packet tracing and analysis: tcpdump, Wireshark (Ethereal), libtrace, Netdude, inettop
      - Host and Application Measurement Tools: NetLogger, Web100
  - NREN tools and statistics: GÉANT, SWITCH (CH), PSNC (PL), HEAnet (IE), ISTF (BG), FCCN (PT), IUCC (IL), HUNGARnet (HU), RENATER (FR)
  - Network emulation: netem, dummynet, NIST, hubt
- End-system (host) tuning
  - Operating system independent: Bugs, Hardware
  - Operating system specific: BSD, Linux (TxQueueLength), MacOS/X, Solaris, MS-Windows
  - Adapters and drivers: Large Send Offload (LSO), Interrupt Coalescence, Checksum Offload, TCP Offload Engines (TOEs), Jumbo MTUs ("jumbo frames")

Example Cases

“Downloads from microsoft.com are slow”

SWITCH
The Swiss Education & Research Network

- Customer claims they get 30-40 kb/s from download.microsoft.com
- Customer has a 100 Mb/s connection
- First tests from backbone:

```
$ wget http://download.microsoft.com/.../dotnetfx.exe
--13:44:16--  http://download.microsoft.com/.../dotnetfx.exe
          => `dotnetfx.exe'
Resolving download.microsoft.com... 195.176.255.136, 195.176.255.135
Connecting to download.microsoft.com[195.176.255.136]:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 24,265,736 [application/octet-stream]
100%[====================================>] 24,265,736     4.49M/s

13:44:22 (3.54 MB/s) - `dotnetfx.exe' saved [24265736/24265736]
```

“Downloads from microsoft.com are slow”

SWITCH
The Swiss Education & Research Network

- Obviously the customer was doing something wrong here...
- ...think of possible problems before calling them.
- But let's do another quick test:

```
$ wget http://download.microsoft.com/.../dotnetfx.exe
--13:53:29--  http://download.microsoft.com/.../dotnetfx.exe
          => `dotnetfx.exe.1'
Resolving download.microsoft.com... 212.162.0.30
Connecting to download.microsoft.com[212.162.0.30]:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 24,265,736 [application/x-msdownload]
12% [===>                          ] 2,968,483     7.37K/s    ETA 51:32
```

“Downloads from microsoft.com are slow”

SWITCH
The Swiss Education & Research Network

- Oops... what happened?
- It turns out Microsoft uses not just Akamai for content distribution, but another layer of load distribution which sometimes selects Akamai, sometimes other servers.
- It doesn't always do a good job at choosing these servers, at least when throughput is concerned.
- Passed the explanation to the customer, with two suggestions:
  - Complain to Microsoft as a customer (recommended, but not quick fix)
  - Hack local DNS to hardwire download.microsoft.com to Akamai (evil hack)
- Customer seems to be happy.

FNAL-DESY Throughput Limitation

SWITCH
The Swiss Education & Research Network

- Fermilab-DESY (7100km GC) TCP transfers “only” reach 150 Mb/s
- DESY-CERN (860km GC) reaches >900 Mb/s
- Initial investigation focused on TCP settings...
- After a lot of measurement, we found a problem in the backbone:
  - The London-Paris GEANT2 link was losing ~0.003% packets (SDH misconf.)
- Several months after that was fixed, another problem appeared:
  - Underprovisioned LAN switch card at DESY
- What we learned:
  - It is **very hard** to locate packet loss in the network
    - We managed using custom ACLs and ACL counters – takes a lot of coordination
  - Tiny amounts of packet loss (as in: nobody notices for months) prevent TCP from going fast over an LFN (Long Fat Network)

General PERT Issues

SWITCH
The Swiss Education & Research Network

- Is this a good idea at all? – Yes, I think so...
  - The potential of the (high-speed parts of the) Internet isn't used well
  - I believe that a shared network is a better substrate for research than dedicated networks for each researcher/project/... (the lightpath approach)
  - It works well as a vehicle for learning/teaching about network performance
- But to be useful, the PERT has to improve
  - Measurement infrastructure
    - Promising developments (PerfSONAR for Research Networks, RIPE TTM etc.)
  - Communication (PERT-customers, PERT-netadmins, intra-PERT)
- Can this be applied in a competitive multi-provider world?
  - Probably not all of it...
  - Breaks the telco-like concepts (illusions?) of service levels etc.

If you're interested in the PERT...

SWITCH
The Swiss Education & Research Network

- Test us if you suspect problems with GEANT etc.
- Contribute to the PERT Knowledge Base wiki
- Provide feedback