

---

# Crashing Spacecraft and Crossing Continents: Summary of a Network Performance Investigation

aka  
"What's up with PERT"

Pekka Savola (CSC/FUNET) \*, John Chevers (DANTE), Toby Rodwell (DANTE)

\*) Visiting SWITCH at the time of investigation



# What's PERT?

---

## ■ Why PERT is needed?

- Long distance links used to be the bottleneck in a network
- The capacity of long distance links has significantly increased
- End-to-end performance bottle-necks may be anywhere
  - End-system (application, OS, hardware), LAN or WAN
- As such, it is becoming more and more difficult for a non-expert end-user to diagnose their network performance issues

## ■ The PERT Concept

- A support structure to investigate and resolve problems in the performance of applications over computer networks
  - Education, measurement infrastructure, human support
- Operating philosophy similar to the CERT
- First idea in 2001: Internet2 End-to-End Performance Initiative
  - <http://e2epi.internet2.edu>

# The European PERT

---

## ■ Project GÉANT

- Trial PERT, 2002-4: 6 NRENs plus DANTE

## ■ Project GÉANT2

- Pilot PERT 2004-5: Part time operations
- Full production 2005- : Full time operations

## ■ PERT organization now

- Virtual team of NREN engineers rotate as Duty Case Manager
- Volunteer Subject Matter Experts (SEs)
  - From NRENs, academia and industry
- A PERT Ticket System for tracking cases
- A mail list for Case Managers and SEs
- A Knowledge Base (public Wiki) for advice on performance issues
  - <http://kb.pert.geant2.net>

# The European PERT - Future

---

## ■ The Future of the PERT

- PERT will continue beyond GÉANT2
- PERT will become a federation
  - ▶ National and regional PERTs to be established
  - ▶ Central PERT for co-ordinating multi-domain cases and assisting regions with no PERT of their own
  - ▶ Already done: a public report on tradeoffs of centralized/distributed PERT
  - ▶ TBD: a detailed policy for federated PERT; workshop on setting up a local PERT
  - ▶ Transition to federated PERT after August 2008.

# PERT - topical challenges

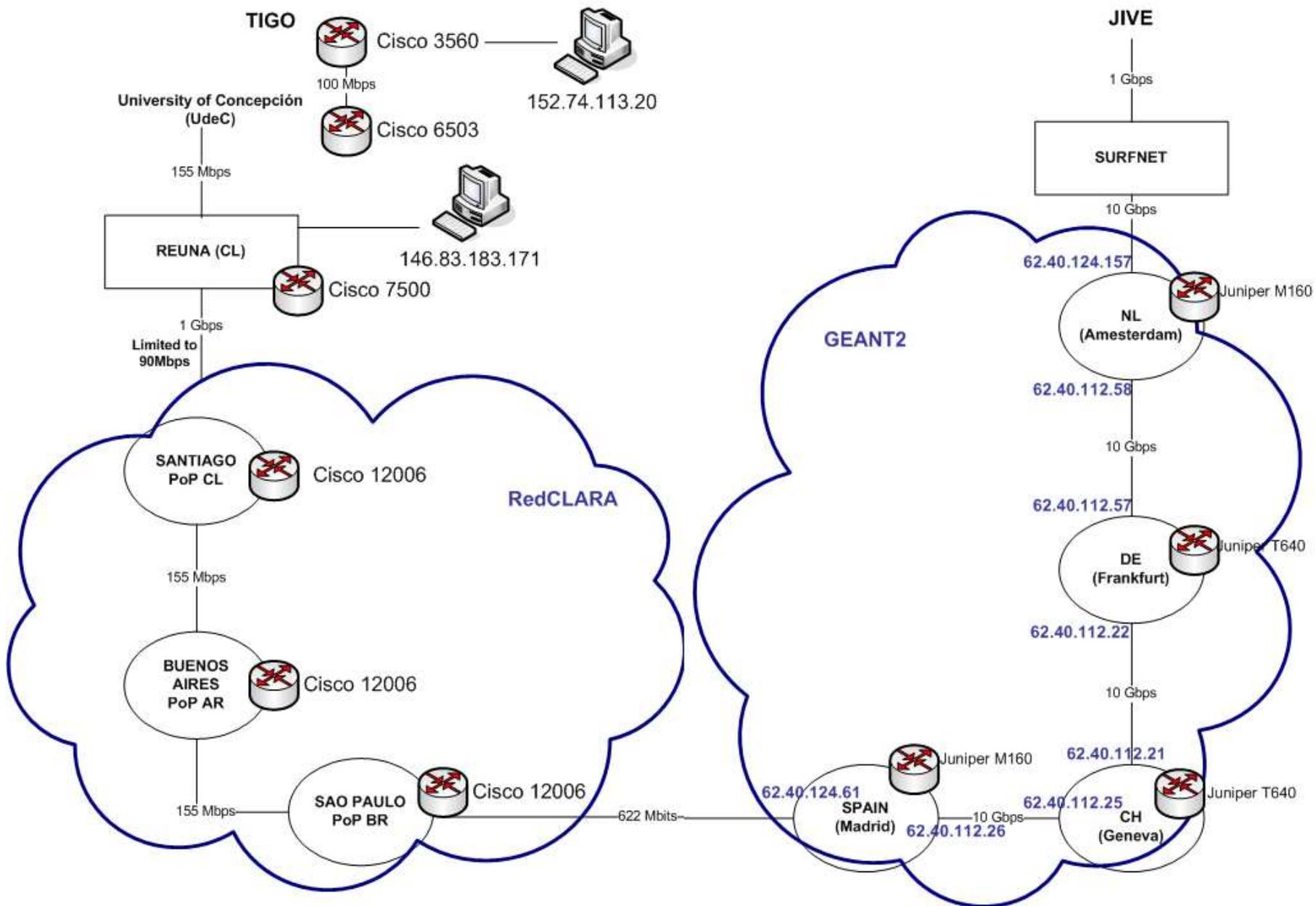
---

- Interdomain lambdas -- not much experience yet
  - IP networks have a number of troubleshooting tools
    - ▶ Ping, traceroute, looking glass, iperf/bwctl servers, ...
    - ▶ Problem analysis is often possible even without access to endpoints (or the middle)
  - Lambdas: ~2 end-points, N semi-transparent networks in the middle
    - ▶ The networks in the middle have limited debugging capabilities
    - ▶ Connect a test PC to the lambda?? O&M remote hands&feet nightmare..
    - ▶ When there's a problem, external people may need access to either endpoint to debug
  - Getting a lambda to work at all
    - ▶ Which debug tools are available to the enduser? ('ping 10.0.0.1 doesn't work!')
    - ▶ Operator of a specific part of the lambda in the middle? ('Our power levels are ok..')
    - ▶ All operators of all the parts of the lambda? ('Our part seems to be working OK..')
  - Getting a lambda to perform to the satisfaction
    - ▶ If you see IP packet loss, jitter, etc., how do you narrow it down?
    - ▶ IP routed network vs Ethernet lambda

# A PERT Case: eVLBI from Chile to the NL

---

- A spacecraft was to crash on the moon in Sep/06
- Telescope in Chile best placed to observe the event
- Observation data transfers to the NL (JIVE) with TCP
- JIVE contacted PERT in Aug/06 to check connectivity
- Adequate TCP performance could not be established
  - Initial performance: 2-7 Mbit/s
- Characteristics of the path
  - Very long RTT (300+ ms)
  - A large number different networks and operators
  - South America has bandwidth challenges



# Some of the challenges faced..

---

## ■ Issues of particular note

- Very long TIGO-JIVE path - ~310 ms
  - ▶ With long RTT, packet losses have dramatic performance on TCP congestion control
  - ▶ If 4 MB has to be transferred every second (32 Mbit/s), with 30ms latency you may need 34 Mbit/s to compensate for loss; with 300 ms latency, you need 80 Mbit/s
- Path: TIGO-UdeC-REUNA-RedClara-GEANT2-SURFNET-JIVE
  - ▶ Many admin. domains increases the difficulty of investigation and communication
- Mismatch of expectations vs network capacity
  - ▶ Initial requirement was 100 Mbit/s TCP transfer
  - ▶ The whole REUNA (CL) backbone's Internet2/GEANT rate was limited to 90 Mbit/s!
  - ▶ Gigabit (or 10G) networking in Europe is commonplace, but not necessarily in some other parts of the world..
- Network had/has intentional constraints to ensure fairness
  - ▶ Rate-limits at UdeC/REUNA (5 Mb/s) and Reuna/RedClara borders (90 Mb/s)

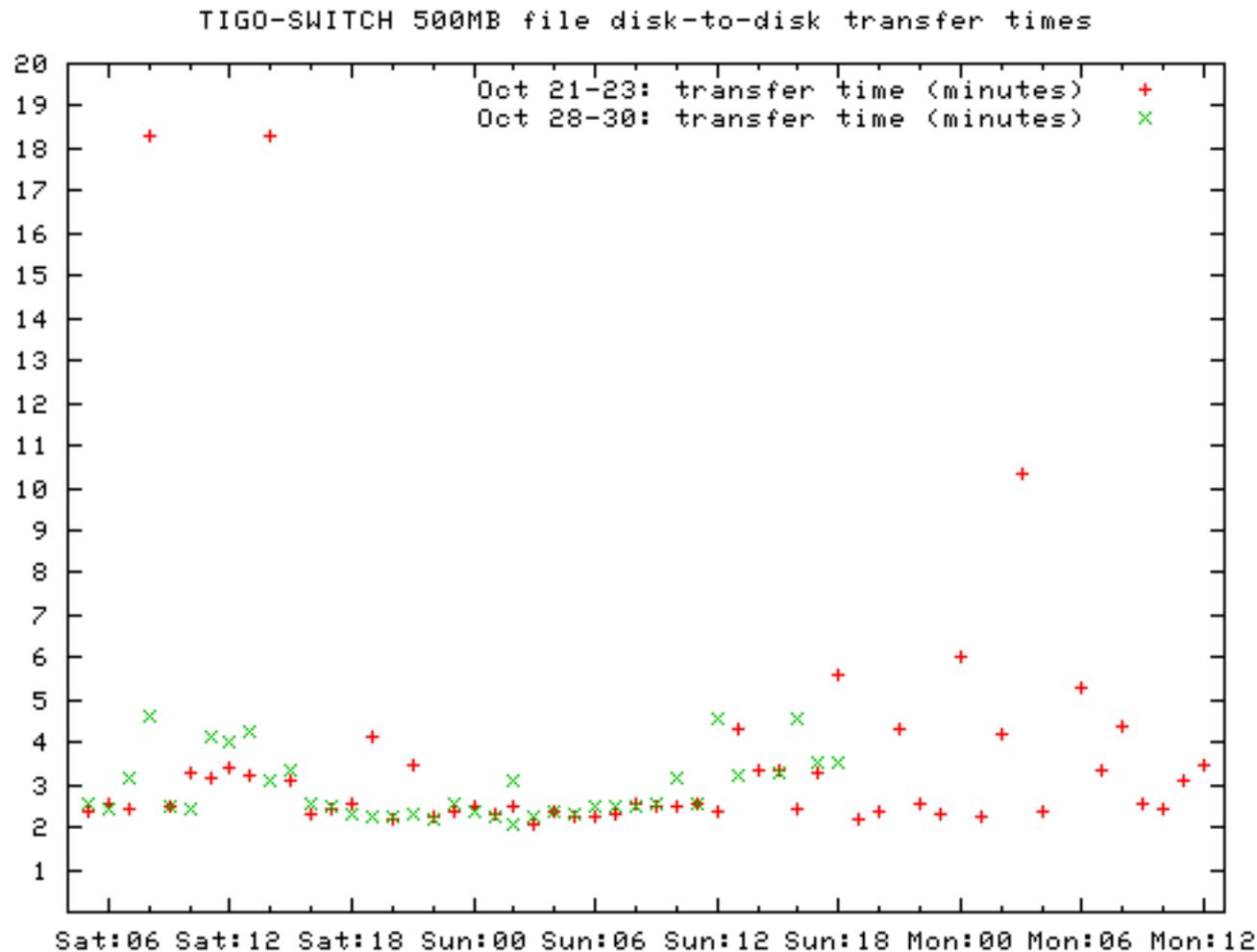
# Improvements from Aug/06 to Dec/06

---

## ■ Performance Improvements

- Initial performance ~5 Mbit/s (due to rate-limiter)
- Removing the limiter and adjusting TCP settings allowed ~15 Mbit/s
  
- Periodic test transfers since October; performance varies in Chile
  - ▶ During weekend or at night, 20-35 Mbit/s disk-to-disk rate achieved reliably
  - ▶ During weekend, there are periods of lossless UDP even at 70 Mbit/s
  - ▶ During day, only 5-10 Mbit/s can be expected
  - ▶ A network upgrade doubled Chile backbone capacity in Dec 2006
  
- The last-hop fiber link was upgraded in December
  - ▶ Forced-100M to 1000M
  - ▶ TCP performance jumped to 60-70 Mbit/s
  - ▶ Yet similar UDP loss as before -- this should not have affected performance
  - ▶ Short-term issue resolved, PERT personnel changed, root cause was not found
  
- A more general question of performance remains

# Disk-to-disk results before the upgrade



- Time in CET/CEST, subtract 5 to get Chile time
- Speed reference
  - 2 minutes = 35 Mbit/s, 3 minutes = 23 Mbit/s, 4 minutes = 17 Mbit/s
- Steady performance during "quiet hours"

# Things to think about

---

## ■ Funding model - expectations vs capacity

- Does the remote site/NREN get funded for providing (higher-than-average) capacity?
  - If not, a value judgment of what traffic is most important..
- Rate-limiters are commonplace in a bandwidth-challenged networks

## ■ Long RTT challenges

- TCP buffer tuning only helps if there is no or very little packet loss
- UDP/DCCP may be need to be used or packet loss eliminated
  - However, must be careful not to starve the network

## ■ Would a lightpath have solved the problem?

- No funding to provide a lightpath
- Lightpath performance debugging could be .. interesting
- Money better spent on improving <GE network bottlenecks?