SWITCH

The Swiss Education & Research Network

"Circuit-based vs. connectionless" and other design choices for research networks

Simon Leinen, SWITCH
<simon@switch.ch>

Disclaimer

SWITCH
The Swiss Education & Research Network

This talk represents my personal opinions only.
It should not be construed as a SWITCH position.

Scope: Networks (“layer 3”)

SWITCH
The Swiss Education & Research Network

Transport connections and connections between application entities are out of scope.

What happens inside network nodes is out of scope – an IP router that uses MEMS-based optical switches in its internal switching fabric doesn't change the nature of the network.

However:

The choice of what functions are put in the network has strong impact on what “upper layers” can/must do.

Context: Research Network Operators

SWITCH
The Swiss Education & Research Network

- Serve research/academic/... communities
- Typically only one per region
  - “hierarchies” of continental/national/regional/municipal operators
- Usually offers “value added” services beyond connectivity
- Coexists with commercial network operators
- This coexistence may takes various forms:
  - Open competition
  - Long-term competitive threat
  - Symbiotic relationship
- Often several forms are present simultaneously.

Research Network Operators

SWITCH
The Swiss Education & Research Network

- Varying scope of connectivity offerings
  - “one-stop shop”(NREN/regional) vs. research-only traffic (GEANT, Internet2)
- Wide variety of organizational structures and funding sources
  - leads to differences in focus and operation
    - e.g. do “the New Thing” every couple years, vs. continuous evolution
    - “Sell” to funding agencies based on technological innovation merits, vs. sell to connected organizations for perceived usefulness.
- Good cooperation between RN's, especially within regions
  - This richness is an asset!

Clouds or circuits re(re(re))visited

SWITCH
The Swiss Education & Research Network

The Internet is a cloud – you inject packets and (sometimes) they come out on the other end.

Some people aren't satisfied by this.

They want “pipes” or circuits with well-defined properties.

This is an old discussion; I only witnessed the last ten years or so

- X.25
- ATM
- SONET/SDH
- MPLS
- **“Lightpaths”/GMPLS/G^{2}MPLS/...**

Why should anyone be against circuits?

SWITCH
The Swiss Education & Research Network

We heard several times why “end-to-end services” are useful.

Circuits such as lightpaths seem to provide such end-to-end services.

Optical switching will unlock huge potential capacity.

So why my aversion to circuits?

Traumatic experiences in formative years

SWITCH
The Swiss Education & Research Network

- Got interested in Internet in 1987
  - because it had all the software that our U.S. partner brought on QIC tapes
  - but there was no Internet (thanks, BITFTP@PUCC!)
- Had to do Lisp programming on a Cray in batch mode (COS)
  - very efficient resource-wise, but unsuited for my “bursty” needs
  - group got (slow) time-sharing server, never looked back.
- Internet finally did arrive, but it was not good
  - like, 2000-3000ms RTT for pings over 1.5km distance
  - people explained that this was due to underlying X.25 network

Traumatic experiences...

SWITCH
The Swiss Education & Research Network

- Moved to Switzerland, Internet connectivity suddenly was good
  - but somehow we weren't really supposed to use it
  - it would be replaced by International Standards (OSI) soon.
- Experimented with Gopher/WAIS/WWW
  - still supposed to wait for OSI
- Several years later at SWITCH, had to build ATM network
  - removed all hopes for IP QoS
  - but necessary for high-quality videoconferencing
  - configured connections once, never changed
  - only other users of e2e connections: QoS researchers
- Moved to pure IPv4/v6-over-fiber, never looked back
- But Peter Löthberg (Sprint) still greets me with the formula

  “How's life in X.25-land?”

Do we have to choose one over the other...

SWITCH
The Swiss Education & Research Network

...or is this an artificial problem?

- Packet switching and circuit switching can coexist
  - but “convergence” will be a strong force to avoid duplicate infrastructure
- So there's a “cart/horse” question
- Resource allocation conflict
  - This may be less apparent with “innovation”-based funding
- Strategic question – where do we see long-term benefits?

In favour of connection-oriented networks

SWITCH
The Swiss Education & Research Network

"Researchers want full control/dedicated end-to-end connections."

Problem: Perceived Lack of Control - many underlying reasons:

- Disappointment with actual throughput over shared networks
  - That's what the PERT has set out to solve
- Lack of transparency/unclear expectations
  - PerfSONAR etc. can help here
- Firewall circumvention
  - Conventional security mechanisms sometimes inappropriate
- Perceived insecurity of mixing with other people's traffic
  - End-to-end alternatives (encryption)?

Logically separate end-to-end connections can be built on top of IP

- Ethernet-over-IP (L2TPv3)
- or EoMPLS (more widely supported in hardware)

“Packet switching is wasteful”

SWITCH
The Swiss Education & Research Network

At high speeds, it is easier to switch traffic based on connections.

- “Beyond <x> Mb/s packet switching is no longer economically feasible and you really have to use <circuit-switching-technology-of-the-day y>
  - x=45, y=ATM/SONET
  - x=10000, y=wavelength switching

It is true that optical switching is more efficient than electrical switching

- In particular, energy-efficient
- Optical *packet* switching is very attractive and not totally infeasible.

But there's a trade-off

- Circuit switching scales well to large bandwidths
- Packet switching scales well to large numbers of “connections”
- (Best-effort, unicast) Packet switching scales well over administrative domains

Routers always seem to catch up somehow

- 92 Tb/s core routers, 2.4 Tb/s L3 switches

Speaking of “wasteful”

SWITCH
The Swiss Education & Research Network

LHC Tier-0/Tier-1 network:

- Data is generated at Tier-0
- ~12 identical copies are delivered to Tier-1 centers
  - over ~12 dedicated links (frequently sharing underlying infrastructure)
  - over ~12 *symmetrical* links (although data flow is strictly unidirectional)

Alternative approaches:

- (reliable) multicast
- BitTorrent
- Distributed filesystem with replication/migration à la GoogleFS

But router interfaces are expensive!

SWITCH
The Swiss Education & Research Network

Especially if you look at carrier class routers

- 1997-1999 designs
- low-volume products
- carrier-class margins
- “deep packet buffers”
- light competition (~two vendors left after .com bust)

Consider embracing commodity technology

- glorified data center/campus switches
- Cisco OSR, Foundry, Force10, Extreme... (Broadcom...)
- modular optics nicely fit fiber/WDM infrastructures
  - too many standards (GBIC/SFP/XENPAK/X2/XPAK/XFP/SFP+/...)
  - too little competition (this is likely to change)
- supported by 10GE inroads into storage networking (iSCSI etc.)

Router vs. switch

SWITCH
The Swiss Education & Research Network

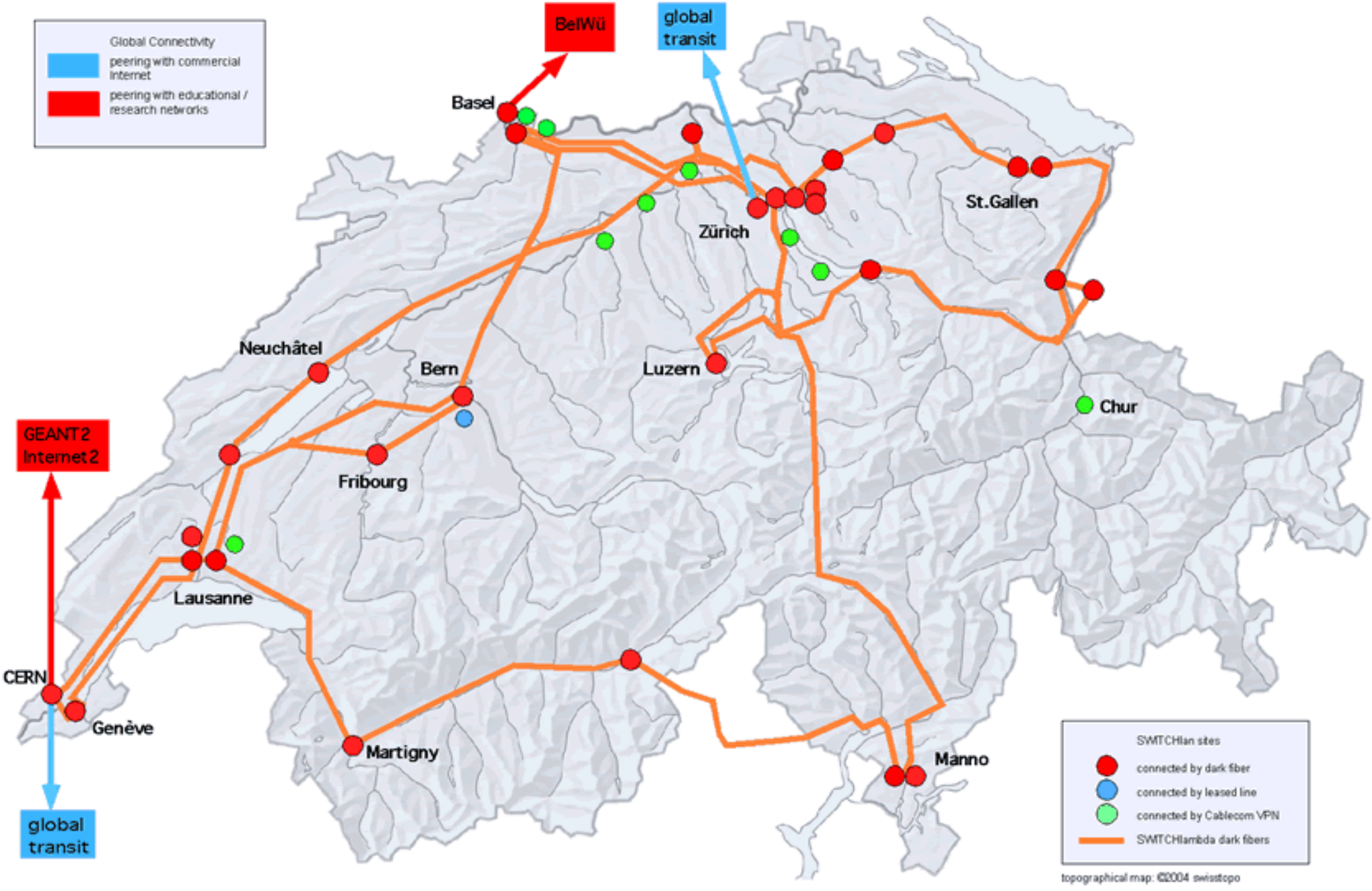
(Deliberate?) confusion

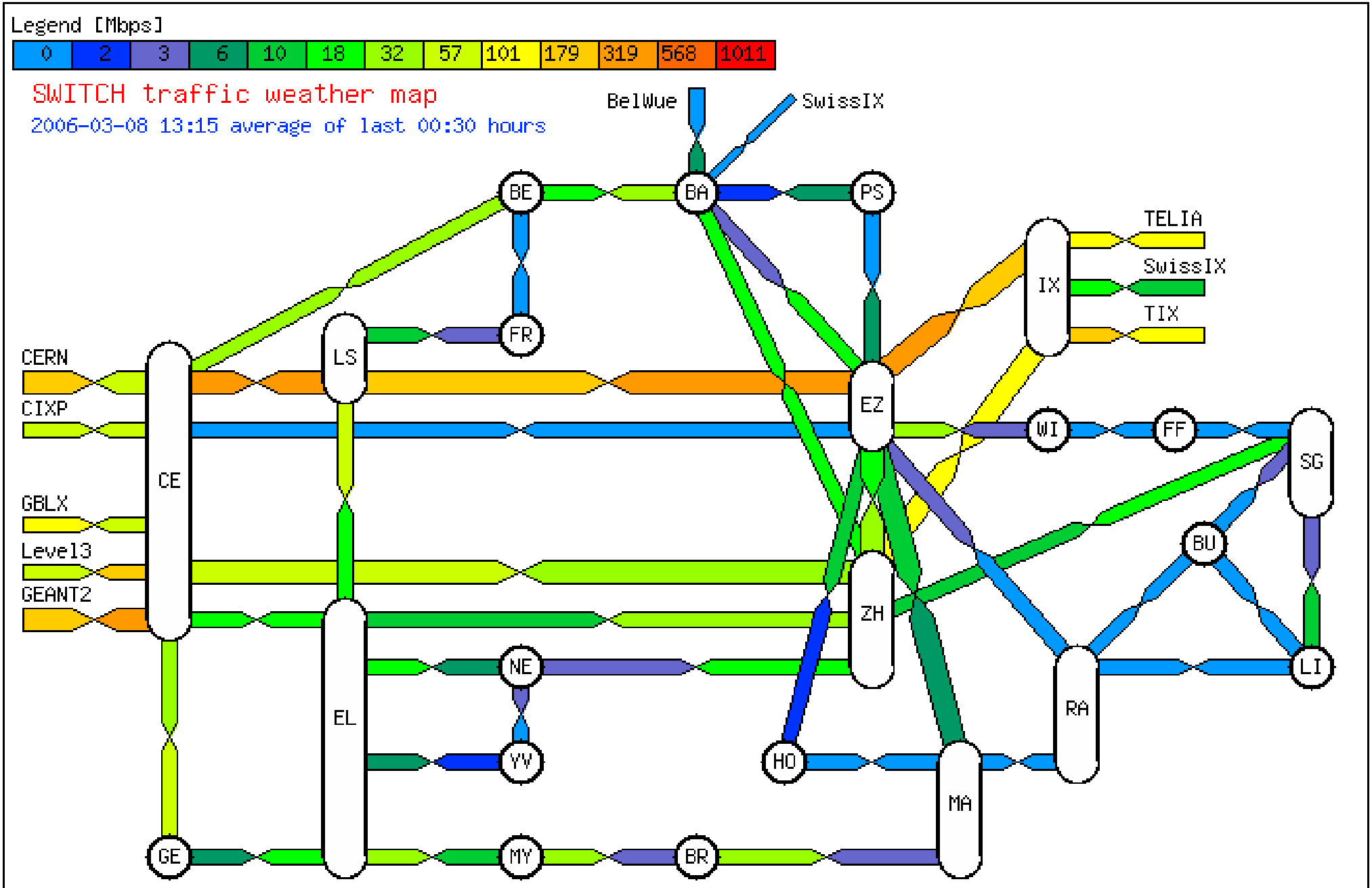
- What is a router?
  - box that runs routing protocols and forwards packets based on L3 address?
  - box where high-speed interfaces are expensive?
- What is a switch?
  - virtual-circuit switch (ATM/TDM/optical)?
  - box where high-speed interfaces are cheap?
  - a router that is used to router based on L2 headers and spanning tree?

What makes routers more expensive than (L2/L3) switches?

- Volume
  - bleeding edge/custom technology vs. standard components
  - software amortization (including advanced and custom features)
- Competition
- Buffer sizes

My view: many cheap “L3 switches” better than few “real” routers


Global Connectivity
peering with commercial internet
peering with educational / research networks
BelWü
global transit
Basel
Zürich
St.Gallen
Neuchâtel
Bern
Luzern
Chur
Fribourg
GEANT2
Internet2
Lausanne
CERN
Genève
Martigny
Manno
global transit
SWITCHlan sites
connected by dark fiber
connected by leased line
connected by Cablecom VPN
SWITCHlambda dark fibers
topographical map: ©2004 swisstopo


Legend [Mbps]
0
2
3
6
10
18
32
57
101
179
319
568
1011
SWITCH traffic weather map
2006-03-08 13:15 average of last 00:30 hours
BelWue
SwissIX
BE
BA
PS
FR
LS
IX
TELIA
SwissIX
TIX
CERN
CIXP
GBLX
Level3
GEANT2
CE
EZ
WI
FF
SG
BU
ZH
LI
RA
NE
EL
YV
HO
MA
GE
MY
BR

Long-term cost perspective

SWITCH
The Swiss Education & Research Network

“Prices of optical components are falling!”

- Yes, but that's true for packet switching components as well

“Bandwidth is free compared to router interfaces”

- See above for router interface costs
- Some claim WAN bandwidth prices will stabilize (e.g. Feuerstein, ONT'05)

Evaluating network architectures

SWITCH
The Swiss Education & Research Network

Research network operators, like all organizations, have a natural tendency to perpetuate themselves...

...but primarily, they should serve the needs of their customers (members, constituents...)

So how do we weigh different customer needs against each other?

- Follow the lead of the PetaByte crowd! This is just the beginning of a trend!
- Some of us have become sceptical from past experience.

Vertically integrated / shared infrastructure

SWITCH
The Swiss Education & Research Network

Vertically integrated networks

- user (customer) sees higher-level *services*
- *infrastructure* not directly accessible – it is owned by the service provider and operated (and optimized) to support the services offered.

Examples: railroad network, telco network

Shared-infrastructure networks

- different users/customers access the infrastructure for their own applications/services.
- “Value is created at the edges.” (Web, Akamai, eBay (users), Google)

Example: road network, Internet.

Valuating networks

SWITCH
The Swiss Education & Research Network

Value of a network in relation to number of users

- **Metcalfe's Law: n^2**
- **Reed's Law: 2^n (group-forming networks)**
- **Odlyzko's Law: $n \log n$**

"Real Options" approach

- ***Where market uncertainty is high, networks that allow independent experiments create higher value (Gaynor and Bradner, SIGCOMM'03)***
- ***So can we predict what the "high-value" uses of the network will be?***

Ultimately, "willingness to pay"

- ***on the part of whoever funds the operator***

Stop talking about money!

SWITCH
The Swiss Education & Research Network

- Architecture of distributed systems
  - scheduling (central control) vs. adaptation
  - compare with batch vs. time-sharing
- Promising advances in adaptation
  - P2P systems
    - data replication (BitTorrent)
    - real-time applications (Skype)
  - A/V applications (streaming, BitTorrent)
- Could we give applications more adaptation possibilities?
  - maybe even without forcing them to build their own (overlay) network

Internet architecture – success and crisis

SWITCH
The Swiss Education & Research Network

Tremendous Internet growth enabled by dumb/transparent network

- Important factor: “end-to-end principles”

Prominent/original example of e2e principle: TCP/IP split

- endpoints responsible for retransmission of lost/corrupt segments
- endpoints responsible for rate control

Dumb+transparent best-effort packet network enabled:

- easy growth at the edges
- cheap experimentation at the edges

Recent issues:

- growth pains, notably IPv4 address shortage
- threats to neutrality in the core
- loss of transparency at the edges (NATs+firewalls)
- still perceived as insecure

Shall we try to fix this or abandon the entire concept?

Alternative – take stuff *out* of (IP) networks

SWITCH
The Swiss Education & Research Network

End-to-end arguments applied to the *path selection* function

- The network cannot know what the best path is for a given packet
- Only the application (in the endpoint) knows
- So shouldn't the endpoint choose the path(s)?
  - This is called “source routing”, which has a bad name

Historical aside:

The end-to-end folks (Saltzer/Reed/Clark) have thought about this

See “Source Routing for Campus-Wide Internet Transport"

(J. Saltzer, D.P. Reed, D.D. Clark, 1980)

Why “campus-wide”?

During the time it was considered unthinkable to apply this reasoning to WANs

(before telecom liberalization)

Source-Routing Internet

SWITCH
The Swiss Education & Research Network

Source-routing alternative Internet outline:

- One or multiple “path services” that a source uses to discover paths to dests.
  - Bootstrapping requirement for paths to an initial path service
  - Could be based on hierarchy (like DNS), DHT (P2P), ...
- The discovered paths could be decorated with QoS/cost attributes
- Sources are free to choose/switch/bundle paths
  - According to QoS requirements, perceived value, robustness requirements
- If full (strict) source routes are used, routers will be radically simplified
  - No routing tables necessary in the network
  - Local topology knowledge would be helpful to enable “local repair” of failures
    - But this is just an optimization – end systems can always use alternate paths

Worries about Source-Routed Networks

SWITCH
The Swiss Education & Research Network

Won't headers become very big with full source routes?

- Use next-hop index per hop; e.g. for a router with 8-15 neighbors, encode the next-hop in 4 bits (can use huffman coding to further optimize)
- An interesting variant is to use an encoding that allows route reversal.

Wouldn't hosts have to know the entire topology to select paths?

- No, they just use local preferences

Won't users be able to source-route around filters?

- Yes, if you filter based on topological assumptions
- Another argument for more effective security mechanisms

- Doesn't this change the economic structure of the Internet
  - Yes, but hopefully for the better.
  - It does mean that it will be hard to get there.

Networking Nirvana?

SWITCH
The Swiss Education & Research Network

The combination of:

- End-system choice of paths (source routes + powerful path service(s))
- Compensation and price announcements for all providers of infrastructure

...would provide a market environment encouraging investment in useful infrastructure.

Conclusions

SWITCH
The Swiss Education & Research Network

- Reports of the death of packet switching seem exaggerated
  - a “wait-and-see” attitude towards optical switching may be rational for some.
- Connection-oriented network services are suitable for some customers
  - there may be other ways of serving those customers (if you have them)
- Basing the network architecture on connections has some drawbacks
  - e.g. increased delays due to reduced number of routers
    - “But it's only legacy/commodity traffic anyway”
- There are interesting engineering issues for the Internet
- There are interesting research issues in packet-switched networks
  - Research networks seem less interested in solving those, which is a pity.


SWITCH
The Swiss Education & Research Network