

The Ongoing Evolution from Packet Based Networks to Hybrid Networks in Research & Education Networks

Olivier H. Martin^{1*}

¹ CERN, Geneva, Switzerland

Disclaimer: The views expressed herein are not necessarily those of CERN.

Abstract

The ongoing evolution from packet based networks to hybrid networks in Research & Education (R&E) networks, or what are the fundamental reasons behind this fundamental paradigm shift and the resulting growing gap between commercial and R&E Internet networks?

As exemplified by the Internet2 HOPI initiative, the new GEANT2 backbone, the GLIF¹ initiative and projects such as Dragon and Ultralight, National Research and Education Network (NREN) infrastructures are undergoing several very fundamental evolutions moving from conventional Packet based Internet networks to Hybrid networks while also moving from commercial Telecom Operator networks to Customer Empowered, dark fiber based, networks.

By hybrid networks, we mean the combination of conventional packet based Internet networks coupled with the capability to dynamically establish high speed End-to-end circuits, i.e. Bandwidth on Demand (BoD), also referred to sometimes as "lambda Grids",

This paper is attempting to explain the fundamental reasons behind this very significant paradigm shift and to assess its likely impact on National R&E, while also giving a very brief overview on what next generation Optical Transport Networks (OTN) may look like in a few years time with the advent of Ethernet over SONET/SDH (EoS), Generic Framing Procedures (GFP), Virtual Concatenation (VCAT) and Link Capacity Adjustment Scheme (LCAS).

Key words: Gigabit/s Wide Area networks, High speed Optical Networks, "Lambda Grids", Ethernet over SONET/SDH, Bandwidth on Demand (BoD).

Introduction

New classes of scientific users and applications, e.g. Very Long Baseline Interferometry (VLBI), High Energy Physics (HEP), are emerging with very large inter-domain bandwidth requirements in the 10-100Gbit/s range that are equivalent or even higher than the capacity of existing National Research and Education Networks and cannot therefore be handled by the existing hierarchical multi-domain networking infrastructures. Furthermore, the Telecom industry is in a "stalled" state and the prospect of having access to the next generation optical transport networks at bandwidth of 40 or 160 Gbit/s is extremely slim in the near to medium term. Therefore new innovative ways of interconnecting the national research & education networks, in order to meet the requirements of these new type of applications, are urgently required. This paper is organized as follows: first, we have a look at the Internet and at the evolution of Ethernet, SONET/SDH and WDM technologies. Next, we consider the Telecom Operator situation following the European Telecom de-regulation back in 1998 and the resulting "debacle" of the years 2000-2001 that led to a lasting over-supply of bandwidth situation and a fast downwards spiral of Telecom prices. We then look at the taxonomy of Research & Education users proposed by Cees de Laat (University of Amsterdam) and how the requirements of the, so called, "Class C" users will be taken care of

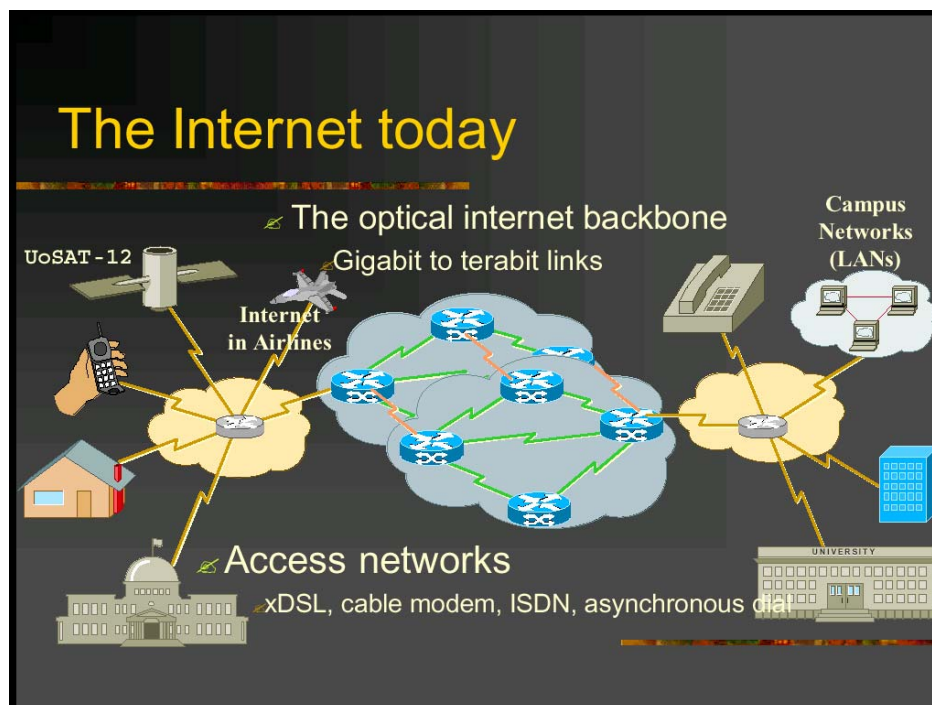
* Corresponding author: Olivier Martin, IT Dept., CERN, 1211 Geneva 23, Switzerland
E-mail addresses: Olivier.Martin@cern.ch

¹ Global Lambda Integrated Facility

by National Research and Education Networks in future, which implies a major overhaul of today's hierarchical organization of research networks with national domain boundaries as it cannot provide a cost-effective solution to the requirements of the emerging applications both in terms of aggregate bandwidth but also Quality of Service. Finally, we explain the reasons behind the demise of conventional packet based networks in the R&E community and the advent of community managed dark fiber networks with "Bandwidth On-Demand" aka "Lambda Grids" capabilities and we have a short look at the emerging Global Grid and its associated Wide Area Networking challenges.

A view of the Internet

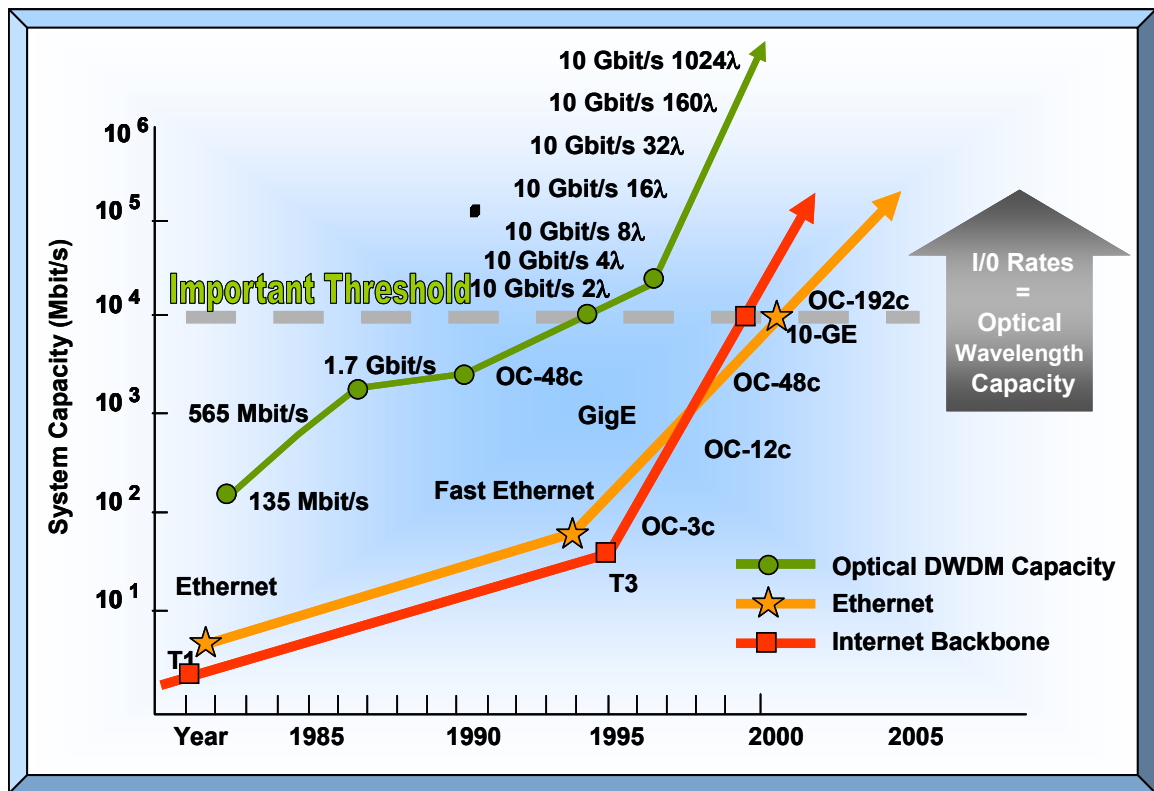
Although the slide below, dating back to 2001 or even earlier, is already fairly old it nonetheless shows extremely well the ubiquitousness of the Internet and, in particular, its capability of being accessed from nearly everywhere around the world at fairly high speed, i.e. ranging from 56/64Kilobit/s (analog modems, ISDN), Megabit/s (ADSL), 10/100/1000/10000 Megabit/s (Ethernet). The only "completely failed" prediction is actually the expected availability of Terabit/second links. Indeed, there is a "de-facto" 10Gbit/s bandwidth limit in today's commercially available optical circuits/wavelengths, often called "lambdas" because of the underlying Wave Division Multiplexing (WDM) optical infrastructure. Given, the Telecom Operator crisis² this situation is unlikely to evolve in any significant matter before many years unless there is some "dramatic" explosion of the demand.



² Since the early 2000, there is a consolidation of the Telecom industry leading in the best cases to mergers and/or acquisitions, otherwise to bankruptcies. Given the bandwidth "glut" leading to prices going down rapidly and profit margins becoming very slim, there is NO commercial justification and therefore little incentive for the remaining Telecom Operators to make any significant investments, this means that the existing Telecom infrastructure which is very far from being saturated is likely to stay "as is" for quite some years, unless some new "killer application" revolutionize the Internet, in the same way as the Web did back in 1992. In the meantime, the reality is that, so far, "bandwidth greedy" applications such as Data intensive Grids have been very slow to materialize and therefore have had extremely limited traffic impact on existing data networks.

Evolution of Ethernet, SONET/SDH and WDM Technologies

The slide below, courtesy of Arie van Praag (CERN), shows several interesting facts and trends, some of which are actually very little known:

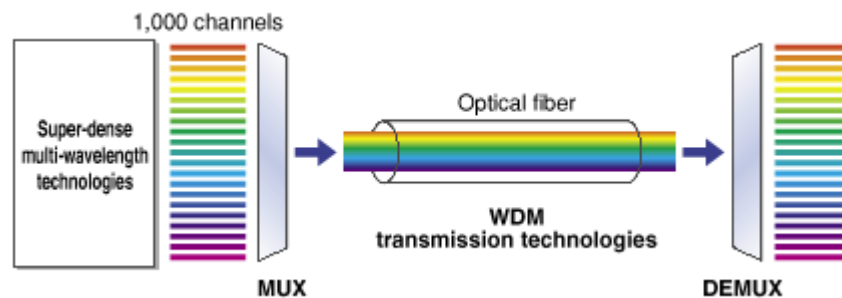


- 1) Local Area Network (LAN) technology has been lagging behind Wide Area Network (WAN) technology, i.e. SONET/SDH, for many years.
- 2) Whereas 10Gbit/s is widely deployed inside Commercial as well as Research & Education Internet backbones since year 2000, large scale deployment of 10 Gigabit Ethernet (10GigE) in LANs is just starting to happen as 10GigE interfaces are finally becoming affordable!
- 3) The 40Gbit/s SONET/SDH standard has been defined several years ago already and can be considered as mature³ although, to the best of our knowledge, there has been no operational deployment in production networks⁴. In contrast, there is still nothing concrete beyond 10GigE for LANs, i.e. will next generation Ethernet be 40Gbit/s or 100Gbit/s?
- 4) Next generation optical transport networks with up to 40Gbit/s capabilities are expected to be based on ITU-T's G.709 recommendation [2], often known as "*digital wrapper*". Unlike today's long-distance telecommunication networks, which can only transport SONET/SDH frames, these new WANs should also be able to transport 1Gbit/s Ethernet, 10Gbit/s Ethernet and several other types of frames transparently.

³ There has been numerous demonstrations already, for example Cisco has a prototype 40Gbit/s interface on its new CRS-1 backbone router

⁴ The only possible exception is the Sprint/Cisco technology trial

- 5) The Generic Framing Procedure (GFP) [3], defined by the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T), specifies a standard low-latency method to transport 1/10 Gigabit Ethernet signal transparently across a SONET/SDH (Synchronous Optical Network/Synchronous Digital Hierarchy) network. Back in 2002, as there was still no suitable GFP-capable multiplexer for phase 3 of the DataTAG⁵ testbed, the project had to resort to using an Alcatel 1670 instead, i.e. a multiplexer that could encapsulate 1Gbit/s Ethernet frames over SONET/SDH frames using a proprietary pre-GFP encapsulation scheme. Since the end of 2005 the situation has changed drastically with a number of vendors coming up with GFP compliant products (e.g. Alcatel, ADVA, Ciena, Lucent, Nortel)
- 6) The capabilities of Wave Division Multiplexing (WDM) equipment, i.e. the number of wavelengths per fiber, continues to evolve at a very rapid pace. Indeed, 128*10Gbit/s channels have been commercially available for quite some time, and NTT research laboratories⁶ demonstrated in 2005 across the Japan Gigabit Network testbed (JGN-II⁷) ten times more WDM channels with eight times the density (one eighth the wavelength spacing) of today's commercial systems. The 1,000-channel WDM transmission required two breakthroughs: one in super-dense multi-wavelength generation and the other in the development of a super-dense WDM multiplexer/de-multiplexer, with 1000 channels spaced at 6.25 instead of 50 GHz)



The Telecom Operator situation

- 1) As already explained, the Telecom Operators are slowly recovering from the Telecom “bubble” of the post de-regulation period, i.e. after 1998, and the heavy investments made in dark fiber infrastructure, purchasing of mobile 3G licences, etc.
- 2) Despite many bankruptcies and mergers there is still fierce competition in all the market segments, i.e. Voice, Data, Internet and Mobile services. As a result, prices continued to drop at a very impressive rate and profit margins have become very slim

⁵ The Data TransAtlantic Grid (DataTAG) testbed was jointly financed by European and U.S. government agencies. The European contribution was funded by the FP5/IST Program of the European Commission (DataTAG project, grant IST-2001-32459) [1]. This project ran from 1 January 2002 to 31 March 2004 and brought together five leading research organizations in the Grid networking community: the European Organization for Nuclear Research (CERN) in Switzerland, the National Institute for Nuclear Physics (INFN) in Italy, the National Institute for Research in Computer Science and Control (INRIA) in France, the Particle Physics and Astronomy Research Council (PPARC) in UK, and University of Amsterdam (UvA) in The Netherlands. The testbed was operated jointly by CERN and Caltech.

⁶ <http://www.ntt.co.jp/RD/OFIS/active/2005pdf/hot/nw/07.html>

⁷ <http://www.kyushu.jgn2.jp/en/index.shtml>

to the extent that some Operators are seriously considering to withdrawing completely from some non-profitable market segments, e.g. voice or basic Internet access in order to focus on “value added services” instead.

- 3) Because of the wide use of WDM technology in Telecom Operators networks worldwide, there is still far more potential bandwidth capacity available than is required to satisfy current as well as near to medium term customer demand. This unhealthy situation is unlikely to change in the foreseeable future!
- 4) As a result, there is very little economic justification to invest in next generation optical transport network, i.e. G.709 and/or to provide new services such as 40Gbit/s SONET/SDH circuits and/or wavelengths.
- 5) In other words, unlike the Internet “boom” of the 1990s that was largely due to the very fast deployment of the Web, we now live in a “frozen” Telecom world where Internet backbones being far from saturated, thanks, in particular, to major advances in techniques to efficiently replicate content, i.e. keeping content at the edges of the network, there is absolutely no commercial justification for the deployment of expensive new network capabilities and services.

Facts and Conclusions, so far

- 1) Internet is everywhere
- 2) Ethernet is everywhere
- 3) The advent of next generation G.709 Optical Transport Networks is very unsure, in other words “we”, the user community, are stuck with 10Gbit/s circuits for some, probably many, years!
- 4) Hence, users must learn how to best live with existing network infrastructures.
- 5) This may well explain all the “hype” about “bandwidth on-demand” as well as “lambda” Grids as NRENs cannot cope anymore with the new, circuit oriented, type of applications!
- 6) For the first time in the history of the Internet, the Commercial and the Research & Education Internet appear to follow different routes.
 - Will they ever converge again?
- 7) Dark fiber based, customer owned long haul, networks appear to have increasingly become the “norm” in R&E backbones, in other words:
 - R&E network operators are becoming their own Telecom Operator!
 - Is it a good or a bad thing, in the short to medium term?
 - What about the long term?

A taxonomy of Research & Education Internet users

Back in 2002, during the iGRID conference in Amsterdam, Cees de Laat from the University of Amsterdam proposed the following categorization of National Research and Education Network (NREN) users in an attempt to justify, mostly for economical but also for Quality of Service (QoS), reasons an “all-optical” end to end network approach for a very specific category of users the, so called, “Class C” users.

The taxonomy goes as follows:

- 1) Class A: Lightweight users, Internet browsing, electronic mail, news, file transfer, home use. Class A users require full Internet routing and “one to many”, i.e. client-server mode of operation. Class A users are no different from residential commercial Internet users and one may wonder why Class A only users need to be connected through NREN infrastructures⁸ rather than through regular commercial Internet Service Providers (ISP) and whether this situation will last for ever?
- 2) Class B: Business applications, multicast, streaming, IP telephony, Peer to Peer, Virtual Private Networks (VPN), mostly LANs, using MPLS layer2 or layer3 technology. Class B users also require full Internet routing and server to server, i.e. peer-to-peer capabilities. Class B, IPv4 only, users could as well be connected through commercial ISPs. One of the main differences between commercial ISPs and NRENs is that many NRENs do support IPv6 whereas very few ISPs do it. Unfortunately, as there is still a lack of IPv6 only applications, the really decisive factor is the effectiveness of the quality of service implementation which can vary greatly between network operators, including NRENs. One should also add that, in practice, very few R&E users are connected directly to their NREN as there is usually at least one regional network or some form of traffic “aggregator” in between. Furthermore, international NREN traffic is, by definition, multi-domain, which makes very problematic the provision of advanced network services, such as Service Level Agreements (SLA) with respect to availability, Quality of Service (QoS), etc.
- 3) Class C: Special scientific applications (e.g. eVLBI, HEP), data-intensive and/or computing Grids, virtual-presence, etc. Class C users typically require very high speed circuits, “fat pipes”, i.e. 1/10 Gigabit Ethernet, and have stringent QoS requirements, i.e. zero packet losses because of the well known weaknesses of the standard Internet Transport Protocol (TCP) in the presence of packet losses, especially over high speed long distance networks). Class C users typically use Grid technology and are grouped in separate Virtual Organizations (VO) with their own Certification Authorities (CA) in order to authenticate users, and the required connectivity is “few to few”.

How do Research & Education networks deal with Class A/B/C users?

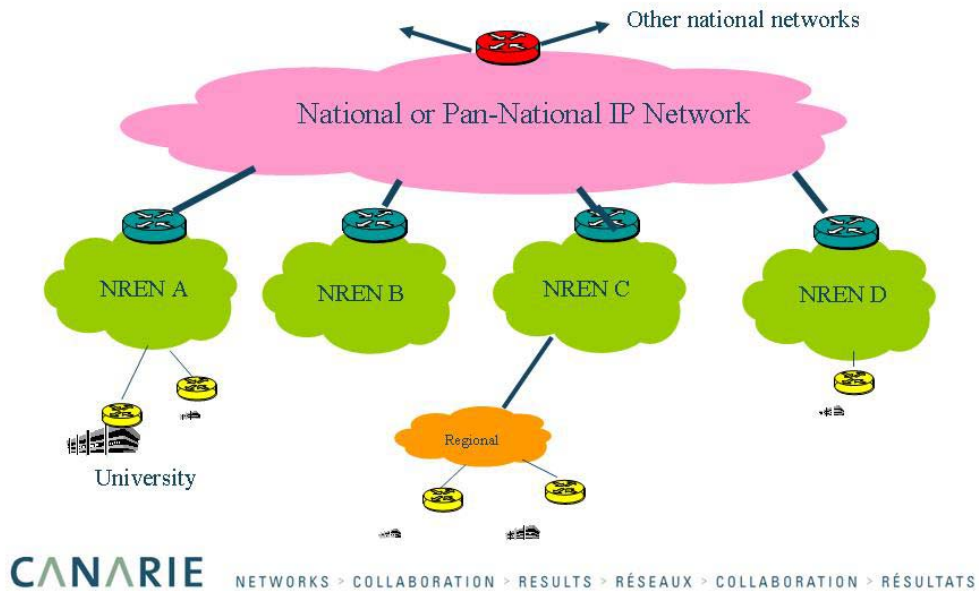
Whereas Cees de Laat was the first one to provide a clear categorization of NREN users and to propose a hybrid architecture with conventional packet based IP services for Class A & B users, and on-demand “fat pipes” or “lambdas” for class C users, Bill St Arnaud from Canarie went one step further as clearly explained in several of his excellent presentations by questioning the hierarchical nature of NRENs and proposing “Peer to Peer” and “Application specific” Virtual Private Networks (VPN) networks across the international multi-domain Dense Wave Division Multiplex (DWDM) infrastructure in order to allow end-end “light paths” or “lambdas” to be easily built and possibly dynamically established, although it is doubtful whether this is realistic from an economic perspective?

- 1) Although NRENs were initially built for Class A, B and C users, the fact is that Class C users have been very slow to come up with the real, long promised, “bandwidth greedy” applications. Therefore, until very recently, NREN infrastructures were more than adequate to satisfy the needs of their user community.

⁸ The only known exception is the Abilene backbone, the US NREN, also known as Internet2.



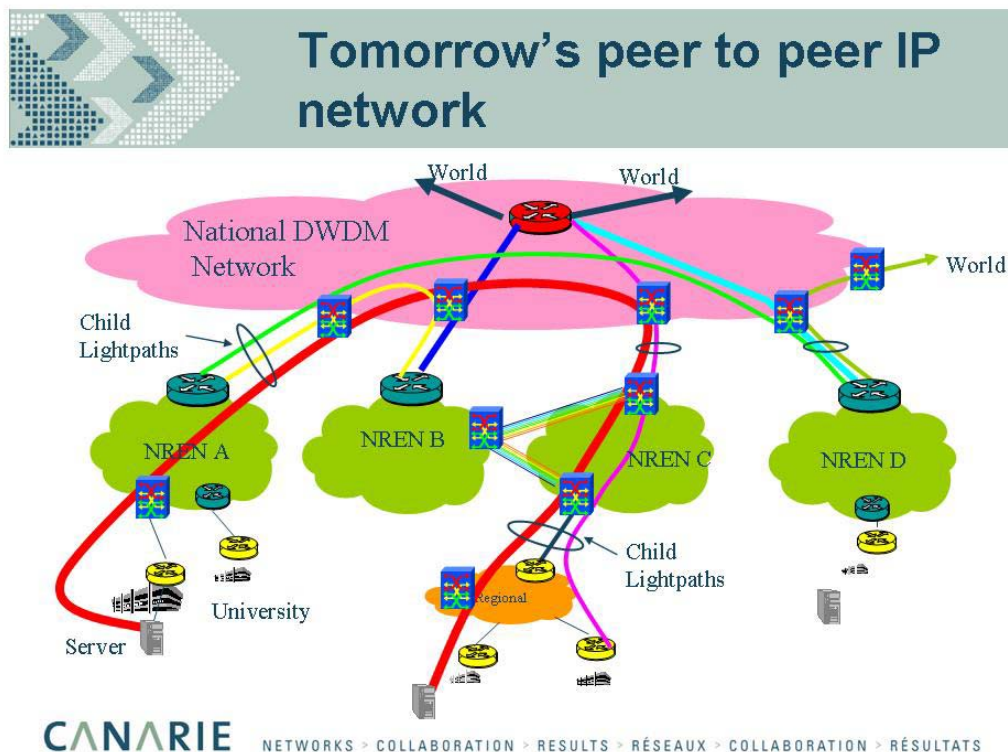
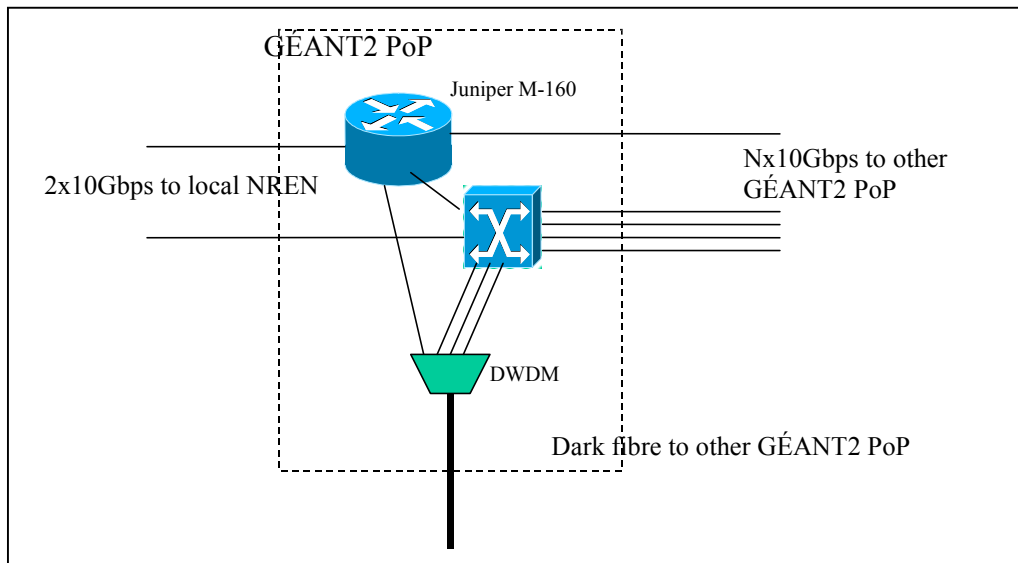
Today's hierarchical IP network



- 2) Unfortunately, it recently became very clear that emerging class C applications, e.g. data-intensive Grids such as the LHC⁹ Computing Grid (LCG), could not be efficiently dealt with across existing European NRENs interconnected through GEANT, the pan-European backbone, for many obvious reasons, e.g. de-facto 10Gbit/s circuit limit, overall circuit as well as interfaces costs¹⁰, lack of end to end traffic engineering because of the multi-domain hierarchical organization of R&E networks, related cost sharing issues, etc.
- 3) Therefore a new architecture similar to the one proposed by Bill St Arnaud on a national scale needs to be implemented in Europe which is being done by the GEANT2 project, in cooperation with the European NRENs, with the acquisition of massive amounts of dark fibers and the deployment of multi-layered Point of Presence (PoP) as depicted below:

⁹ CERN's Large Hadron Collider that will become operational in 2007

¹⁰ Whereas large commercial ISPs, i.e. Tier1, often have fairly global coverage, this is NOT the case of National R&E Networks (NREN) who are, by definition, national. Connecting Lyon to Geneva, for example, at 10Gbit/s may imply, many more 10Gbit/s circuits and interfaces than a direct connection because of the topology of independently managed networks such as RENATER (France), GEANT and SWITCH (Switzerland), especially as the route between Paris and Geneva may go through Francfort and Milano. The reason being that provisioning of GEANT1 was mostly done on minimizing the overall costs without taking much account of the expected traffic flows!



So what are the economical and commercial facts?

- 1) Whereas some Telecom Operators are willing to rent dark fibers through Indefeasible Right of Use (IRU), i.e. the effective long-term lease (temporary ownership) of a portion of the capacity of an international cable, Telecom Operators are unwilling to lease wavelengths on a “cost based” manner, in other words the price of two 10Gbit/s wavelength is twice the price of a single 10Gbit/s wavelength, which clearly does not reflect at all the real costs incurred by the Telecom Operators.

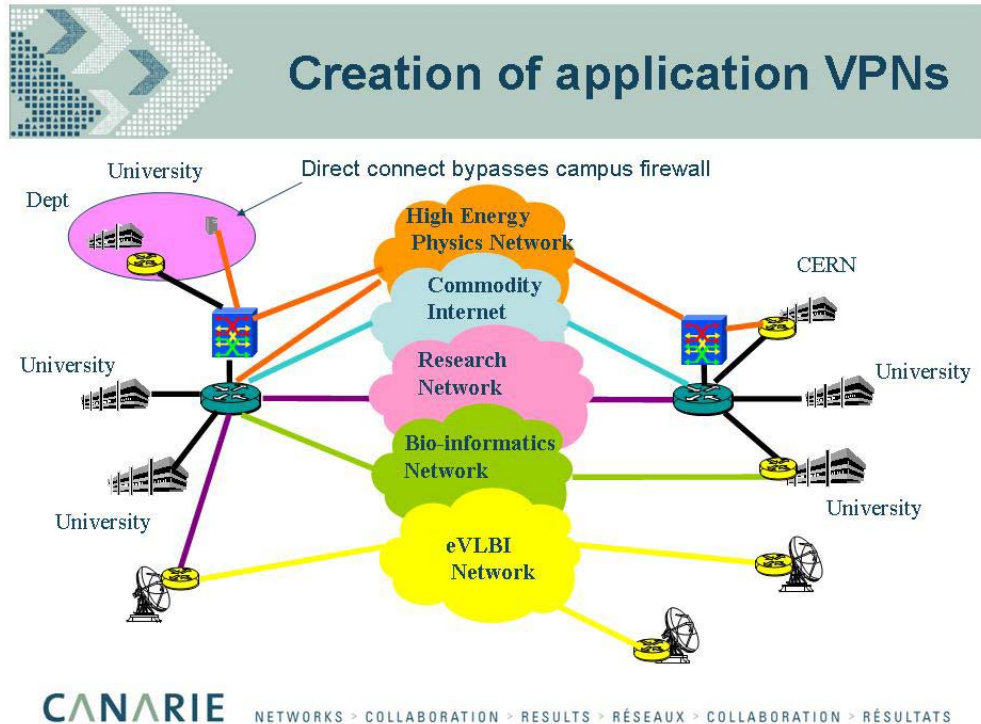
- 2) As a result many national R&E networks as well as GEANT2, the new pan-European interconnection backbone, are moving towards dark fiber based backbones, despite all the hassle of lighting the fiber, operating the optical equipment, monitoring the layer1 service and so on!
- 3) In other words, National R&E networks are becoming their own Telecom Operator!
- 4) Given the price difference, this is probably the only viable way to provide the required multiple 10Gbit/s wavelengths in the short term, but the real question is whether this is a viable option in the long term?
 - i. The consensus within the R&E network community is yes,
 - ii. However, I am personally doubtful but I am well aware that I am of the very few “dissidents”!
- 5) Indeed, the deployment of 40Gbit/s wavelengths is likely to require different type of fibers and optical equipment than the one used on today’s 10Gbit/s infrastructure, unless the NTT laboratories proposal to use Virtual Concatenation (VCAT), i.e. inverse multiplexing technology, in order to implement 40Gbit/s end-to-end circuits using 4*10Gbit/s wavelengths, receives wide acceptance which, at this stage, is very unlikely.
- 6) At this point it is important to differentiate between long haul optical networks which require optical amplifiers as well as full optical re-generation equipment and short haul or metropolitan networks which do not.
 - i. Therefore, I am only questioning the wisdom of acquiring IRUs on long haul routes.
- 7) Independently of whether national networks own the underlying optical infrastructure or not, it is quite clear that the costs of layer1 and layer2 equipment is much lower than the cost of layer3 equipment, therefore next generation networks MUST be multi-layer, multi-services networks. Indeed, high-end backbone routers are far more evolved than yesterday’s supercomputers due to the complexity of supporting most of the functionality in the hardware in order to reach wire-speed performance, e.g. IPv4 & IPv6 forwarding, MPLS layer2&3, access lists, etc.

In conclusion, “Big Science” need fat pipes, “Big Science” is, by nature, multi-national, at least, or even Global, therefore, a hybrid architecture serving all users in one coherent and cost effective way, without being bound by the national NREN boundaries, is the only way forward.

Lambda Grids, Bandwidth on Demand, Why & How?

As already explained, conventional layer 3 technology is no longer “fashionable” because of the prohibitive costs of high-end backbone routers, high speed interfaces, the implied use of shared, hierarchically organized, network backbones, and the 10Gbit/s bandwidth limit which means that parallel circuits would need to be deployed in order to meet the requirements of the “Class C” users which is neither desirable nor economically feasible.

Using layer 1 or layer 2 technologies for such users is therefore very attractive as it allows to solve a number of problems, e.g. protocol transparency. The minimum functionality of layer 1 & layer 2 equipment combined with the use of direct end to end circuits also allows to drastically reduce the overall networking costs.



Fortunately, the best known “Class C” user communities, e.g. VLBI, HEP, have very static requirements that can therefore easily be met across “flat” single-domain Internet backbones with simple traffic engineering techniques, i.e. provisioning of direct end-to-end circuits at best, or minimizing the number of hops, otherwise.

As explained above, increased collaboration between NRENs and GEANT is required in order to allow end-to-end circuits to be established for use by static “Class C” users, the question which is still outstanding is whether there is any serious need for “on-demand lambdas” and although the majority view is that there is a real requirement, I am, personally, extremely skeptical, especially as the use cases are still very loosely defined, to say the least!

Furthermore, moving back to a “circuit oriented” model like the telephone network seems very odd, despite some clear advantages, e.g. simpler transport protocols, lower equipment costs (i.e. « in theory » a factor 2 or 3 per layer), path towards “all-optical” networks. In addition, « end-to-end » is still very loosely defined, does it mean “site-to-site”, “cluster-to-cluster”, “host-to-host” or “peer-to-peer”? What about the costs of high speed circuits in standby mode most of the time, the scalability, the additional middleware to deal with circuit set up/tear down, etc!

One of the main unresolved issues in very high-speed networking, i.e. multi Gbit/s, is the inability of the standard Internet transport protocol, namely TCP new Reno, to deal efficiently with packet losses or what looks like packet losses to the transport layer, e.g. out of order packets¹¹.

¹¹ This is still the case across the GEANT layer 3 backbone because of the presence of outdated Juniper M160 routers which are, in effect, preventing the deployment of very high speed applications!

Although “site to site” connectivity helps to reduce the probability of packet losses, because of the shorter path, it does not guarantee by itself zero packet losses, the same holds for “cluster-to-cluster” connectivity. In other words the only safe way to guarantee “near to zero” packet loss rates is to implement “host-to-host” or “peer-to-peer” connectivity which is not practical given the number of parallel high-speed circuits potentially required!

So, either new modes of operations need to be validated, e.g. GridJIT¹², that require some form of switch, e.g. all optical switch¹³, between the long haul wavelength(s) and the hosts in order to switch the circuit(s) between the hosts, or alternatively non-standard TCP implementations such as FAST¹⁴ or Westwood¹⁵ must be used. In practice, use of non-standard TCP implementations is more likely to happen as it may provide an acceptable, low cost and easy to deploy, solution to the lack of resilience of TCP new-Reno to packet losses.

As to the availability of high-speed switched services, I cannot see how it can be economically provided from a service provider perspective, but also how they can be used from a user/application perspective, unless new middleware is developed to make the applications “switched circuits” aware.

All together, this means, in my opinion, that switched services are unlikely to become commercially available in the near to medium term, unless the cost of bandwidth really becomes dramatically low!

Additional technical challenges Local Area Networks (LAN)

We already mentioned that unprecedented cooperation between NRENs and GEANT2, the new pan-European backbone, will be required in order to meet the requirements of the “Class C” users, i.e. to achieve seamless interconnection between the various “Class C” sites.

Likewise, Local Area Networks (LAN), i.e. Campus Networks, will need to be upgraded in order to be able to cope with “discipline specific network applications”.

In particular, the practice of having a firewall between the wide area and the local area network will need to be re-visited as commercial firewalls are usually unable to deal with 10-100 Gbit/s applications.

Furthermore, given the nature of these applications, i.e. only “few-to-few” connectivity required, there is no reason for having a possibly very expensive firewall in-between, assuming such firewalls would actually exist.

In other words, campuses will need to deploy parallel networks for high end users, i.e. have different security zones with different level of outside connectivity.

This new approach to local area networking has a number of advantages:

¹² <http://www.mcnc.org/events/mcncopticalworkshop/presentations/DanGRIDJITControlPlane.PPT>

¹³ CERN as well as Internet2 and University of Amsterdam already use an all-optical switch from Glimmerglass

¹⁴ <http://netlab.caltech.edu/FAST>

¹⁵ <http://kb.pert.switch.ch/cgi-bin/twiki/view/PERTKB/WestwoodTCP>

- 1) Reduces costs by providing high end network capability to only those who need it
- 2) Eliminates the limitations of campus firewall and border router

but also raises a number of new issues with respect to security, back door routing, etc.

Last but not least, end-to-end performance does not only depend on networks but also on local host architecture (bus), the network interface cards (NIC), etc. Mastering the whole chain involved, i.e. disk-bus-memory-bus-network-bus-memory-bus-disk is actually more than “acrobatic”, to say the least!

Transport Control Protocol (TCP)

It is very well known that the standard Internet Transport Protocol, TCP new-Reno, does not scale well to high speed, high latency and noisy environments and has unfair behaviour with respect to: Round Trip Time (RTT), Frame size (MSS), Access Bandwidth.

Hence the widespread use of multiple streams in order to compensate for inherent TCP/IP limitations (e.g. Gridftp, BBftp) which are clearly more of a bandage than a cure!

New TCP/IP proposals are required in order to restore performance in single stream high-speed environments (i.e. “Class C” users), however, the IETF standardization process is extremely slow, furthermore the problem is not deemed to be general enough to require an urgent solution.

In the mean time, there is an absolute requirement for “packet loss and packet re-ordering” free communications which, as demonstrated above, reinforces the case for some form of “lambda Grids”.

TCP dynamics over a 10Gbps path with 100ms Round Trip Time (RTT) & 1500 bytes packets)

Window size (W) = Bandwidth*Round Trip Time

- Wbits = 10Gbps*100ms = 1Gb
- Wpackets = 1Gb/(8*1500) = 83333 packets

Standard Additive Increase Multiplicative Decrease (AIMD) mechanisms:

- $W=W/2$ (halving the congestion window on loss event)
- $W=W + 1$ (increasing congestion window by one packet every RTT)

Time to recover from $W/2$ to W (congestion avoidance) at 1 packet per RTT:

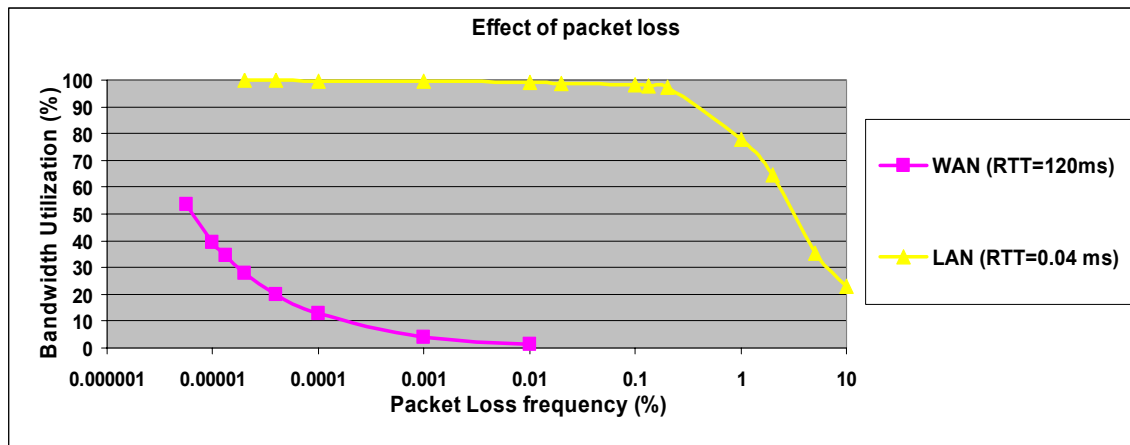
- $RTT*Wp/2 = 1.157$ hour
- In practice, 1 packet per 2 RTT because of delayed acks, i.e. 2.31 hour

Packets per second:

- $RTT*Wpackets = 833'333$ packets

Single TCP stream performance under periodic losses

(by courtesy of S. Ravot/Caltech)



Conclusions

The, so called, “Class C” users, i.e. the “bandwidth greedy” applications with 10-100 Gbit/s aggregate bandwidth requirements are emerging.

In order to meet their requirements in a cost effective manner, unprecedented collaboration between National Research & Education Networks (NREN) and GEANT, the pan-European backbone, as well as a major overhaul of Local Area Networks (LAN) architectures is required.

The advent of next generation G.709 Optical Transport Networks with 40Gbit/s capabilities being very unsure, the R&E user community is actually “stuck” with 10Gbit/s circuits for some years!

Hence, users must learn how to best live with existing 10Gbit/s limited network infrastructures.

For the first time in the history of the Internet, the Commercial and the Research & Education Internet appear to follow different roads, it is far from clear whether and when they will converge again?

There is a very clear trend towards community managed dark fiber networks. As a result, National Research & Education Networks are becoming Telecom Operators, is it right?

- In the short term, almost certainly YES
- In the longer term, probably NO
- In countries, where there is NO other way to have affordable access to multi-Gbit/s capable networks, this is clearly the right move

Whereas static requirements, such as those of the VLBI and HEP communities, are fairly easy to accommodate, despite the fact that new operational models may be required in order to circumvent the deficiencies of TCP “new Reno”, the standard Internet transport protocol, “on-demand” lambdas are unlikely to be economically viable!

Acknowledgments

Many thanks to Cees de Laat (University of Amsterdam), Bill S^t Arnaud (Canarie), Harvey Newman & Sylvain Ravot (CALTECH) for their significant contributions to this paper.

References

- [1] DataTAG Project, <http://www.datatag.org/>
- [2] ITU-T, *Interfaces for the Optical Transport Network*, Recommendation G.709, March 2003.
- [3] ITU-T, *Generic Framing Procedure*, Recommendation G.7041/Y.1303, December 2003.
- [4] iGrid 2002 Conference, <http://www.igrid2002.org/>
- [5] Internet2 Project, <http://www.internet2.edu/>
- [6] DataTAG Project, Work Package 1, <http://www.datatag.org/wp1/>
- [7] C. Meirosu, P. Golonka, A. Hirstius *et al.*, "Native 10 Gigabit Ethernet Experiments over Long Distances", *Future Generation Computer Systems*, Dec 2004.
- [8] R. Hughes-Jones, P. Clarke and S. Dallison, "Performance of 1 and 10 Gigabit Ethernet Cards with Server Quality Motherboards", *Future Generation Computer Systems*, Dec 2004.
- [9] GÉANT Website, <http://www.dante.net/server/show/nav.007>
- [10] The DataTAG Transatlantic testbed: O. Martin, J.P. Martin-Flatin, E. Martelli, P. Moroni (CERN), H. Newman, S. Ravot, D. Nae (CALTECH), Elsevier's FGCS (Future Generation Computing Systems) Journal Volume 21, Issue 4, April 2005, Pages 443-456 [pdf].
- [11] M. Rio, A. di Donato, F. Saka, N. Pezzi, R. Smith, S. Bhatti and P. Clarke, "Quality of Service Networking for High Performance Grid Applications", *Journal of Grid Computing*, Vol. 1, No. 4, pp. 329-343, 2003.
- [12] DataTAG Publications, <http://www.datatag.org/papers/>

Biography



Olivier Martin is the Project Leader of the DataTAG project. He received an M.Sc. degree in EE from École Supérieure d'Électricité (Supélec), Paris, France in 1962. He joined CERN in 1971, held various positions in the Software Group of the Data Handling Division, and then moved to the Communications Group of the Computing & Networks Division in 1984, where he has been Head of the External Networking Section from 1989 until 2004. Prior to the DataTAG project, he was involved in several European projects (including BETEL, BETEUS and STEN) in the framework of the RACE, ACTS and TEN programs. His research interests include high-speed networking, transport protocols and Grids.