Available at www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

ELSEVIER

Future Generation Computer Systems 19 (2003) 999–1008

FGCS
FUTURE GENERATION COMPUTER SYSTEMS

www.elsevier.com/locate/future

The rationale of the current optical networking initiatives

Cees de Laat [a,*], Erik Radius [b], Steven Wallace [c]

[a] *Informatics Institute, University of Amsterdam, Kruislaan 403, 1098SJ Amsterdam, The Netherlands*
[b] *SURFnet, Utrecht, The Netherlands*
[c] *Advanced Network Management Lab, Indiana University, Bloomington, IN, USA*

Abstract

The future of networking is to move to an entirely optical infrastructure. Several leading National Research Networking organizations are creating test-beds to pilot the new paradigm. This paper explores some thoughts about the different usage models of optical networks. Different classes of users are identified. The services, required by the Internet traffic from those different classes of users, are analysed and a differentiated Internet architecture is proposed to minimize the cost per transported packet for the whole architecture.
© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Lambda networking; Optical networking; Switching; Routing; DWDM; High performance; High throughput; Bandwidth on demand

1. Introduction

In various places in the research networking world test-beds are launched to study and deploy lambda networking. In this contribution we discuss a differentiated architecture in which we can deliver different transport services for different classes of users. The current lambda networking initiatives tend to only connect routers via SONET circuits. On the physical layer those circuits are mapped to colours of light on the fibre (sometimes four or more circuits are merged using time division multiplexing (TDM) in one wavelength). While prices of those SONET circuits are rapidly dropping and the speeds are increasing, the main cost is going to be in the router infrastructure in which the circuits are terminated. Full Internet routers must be capable of finding the long prefix match for each routed packet in a next-hop database that now contains in the order of 120.000 entries. At 10 Gb/s speeds an IP packet takes in the order of 100 ns to arrive, so the logic for doing routing is becoming increasingly complex and, therefore, expensive. We will assume that a significant amount of the backbone traffic in fact does not need to be routed and, therefore, should stay at the optical or switching layer.

2. Optical networking

2.1. History

Optical transport technology has been around since the 1970s [4,5], as it provided a reliable way of transporting high-bitrate signals over long distances. Its primary use was in telecom networks, for inter-city and international transport networks. An important milestone in the development of optical networking can be attributed to the erbium-doped fibre amplifier (EDFA) in the late 1980s. This EDFA provided

* Corresponding author. Tel.: +31-20-525-7590; fax: +31-20-525-7490.
E-mail addresses: delaat@science.uva.nl (C. de Laat), ssw@indiana.edu (S. Wallace).

0167-739X/03/$ – see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0167-739X(03)00077-3

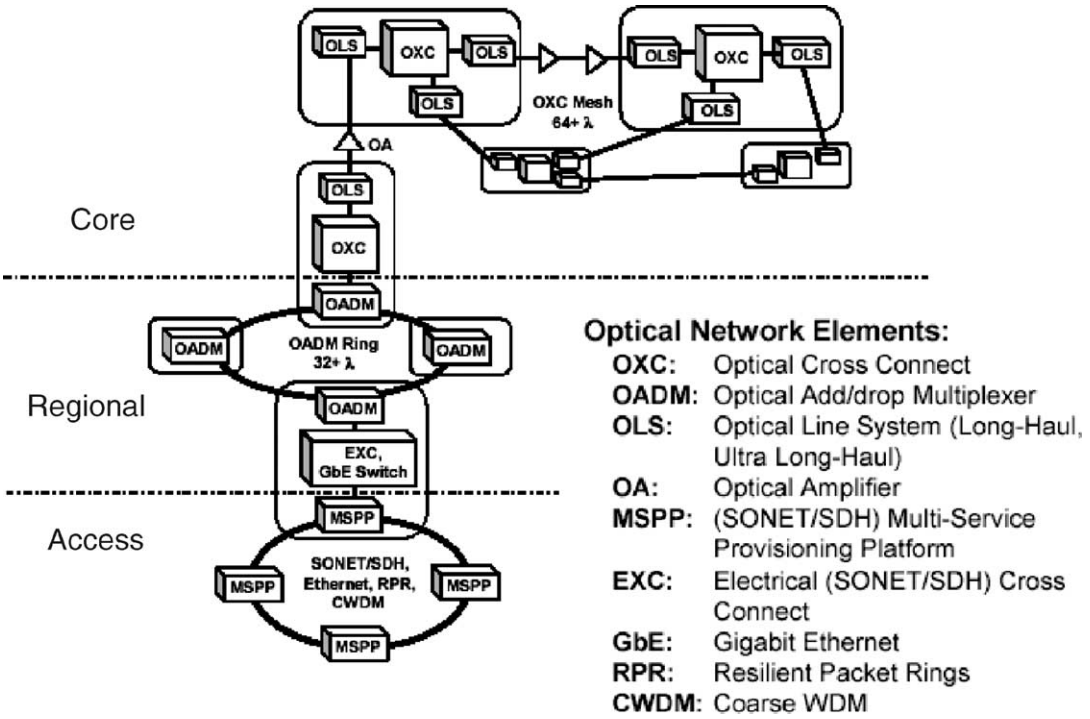

Fig. 1. Emerging optical transport network architecture.

signal amplification in a wide optical spectrum. The EDFA together with the development of the narrow line-width laser diode brought dense wavelength division multiplexing (DWDM) and terabits per second transport per fibre within reach. As mankind discovered the Internet, the hunger for bandwidth started to explode. As a result of this need for raw bandwidth, optical fibres were increasingly being deployed in regional and metro area networks, providing more bandwidth than the few megabits per second that traditional (copper-based) local loops in the access network can carry. While the cost of installing fibre into every household is still too high, it holds a clear promise for replacing the copper access line in the years to come. In the mean time, the abundance of dark fibres, hastily put into the ground since the early 1990s in cities and regional business districts (but never lit since) provide a good opportunity for anyone with high bandwidth needs to get a cheap optical access loop into a carrier point of presence (POP). Transport networks are typically divided into different categories: core, regional and access, each demanding diverse functions and optical networking elements. See Fig. 1 for a typical representation of this architecture as defined by Bonenfant and Rodriguez Moral [1].

This optical network representation is typical for a next-generation telecom network, primarily SDH or SONET based, with some native Ethernet transport capabilities. The ‘true’ optical domain is in the core and regional network, shielded from the end-user in the access network by multi-service provisioning platforms (MSPPs).

By contrast, the Internet community is increasingly using optical technology for short-haul purposes, e.g. for interconnecting switches and routers with gigabit Ethernet interfaces. This technology is now very popular and due to the massive market relatively cheap compared to more traditional long haul telecom SONET equipment. It is successful in regional networks for dark fibre links up to approx. 100 km. So whenever it is feasible to get dark fibre on suitable distances it seems advantageous to apply gigabit Ethernet equipment for those connections. There are standards emerging for more flexible transport of

Ethernet frames in SONET/SDH; e.g. generic framing procedures [6], link capacity adjustment scheme (LCAS) [7]. LCAS is a protocol to synchronize the adjustment of the bandwidth parameter of a circuit in the network and thus allows adaptation of already provisioned circuits.

The recently finalized 10 Gb/s Ethernet (10GbE) WAN-phy standard specifies a SONET-framed physical interface type, becoming an alternative to 10 Gb packet-over-SONET transport.

2.2. What is in store: some trends in optical networks

Now that EDFA and DWDM have boosted the usable fibre capacity into the terabit/s realm, we observe optical element developments gearing towards optical switching techniques. Core switching techniques using lambda or fibre switching cross connects move the current static optical networks into a dynamical on the fly reconfigurable transport infrastructure. DWDM is nowadays used to provide cost-effective traffic grooming for the regional (metro) network, which may aggregate many disparate traffic types to a core POP.

On a line systems level, 10 Gb/s technology has matured and 40 Gb/s is the next step. However, the current slowdown in the worldwide economy may stall the adoption of 40 Gb/s systems for several years.

For cost-effective and modest capacity increase in access rings, coarse WDM (CWDM) is gaining ground as a non-amplified, no-frills multiplexing technology. Further on the horizon, optical subsystems for optical packet switching may one day enable all-optical data networking and computing.

With optical switching components inside the network, optical circuits may be setup and torn down dynamically and on demand. These dynamical features will only be used if adequate policy systems are developed and installed to control these new resources. One main reason of the failure of ATM 6 years back was the lack of this control, which inhibited the providers to offer Switched Virtual Circuits to Universities and end users.

2.3. Related activities on optical networking

The research and education network community has expanded on the telecommunications-based connotation of “lambda networking” to include technologies and services that have one or more of the following attributes in common with this new optical technology:

(1) Transmission capacities of 2.5 and 10 Gb/s. These capacities represent the typical provisioned capacity of a wavelength in a DWDM system.
(2) The circuit nature of individual wavelength provisioned capacity. These individual wavelengths are provisioned as constant bit rate circuits. The term “light paths” has been coined to describe end-to-end circuits.
(3) The lower cost of high capacity circuits in both long haul and metro systems.
(4) The ability to more directly interface high-speed local area network technologies (e.g. 1 and 10 Gb Ethernet) to telecommunications services.
(5) The ability to provision new services in a more automated fashion.

Examples of lambda network initiatives in this context include the StarLight facility, NetherLight and Teragrid. The Chicago Starlight facility is designed to provide interconnection services and co-location space for high-speed national and international research and education networks. Interconnection will initially use 1 and 10 Gb Ethernet technology over layer-two switching. Starlight’s mantra, “bring us your lambda’s”, implies connecting to Starlight with sufficient capacity and appropriate equipment to interconnect at gigabit speeds.

NetherLight consists of SURFnet’s transoceanic 10 Gb/s lambda between Amsterdam and Chicago and its 10 Gb/s lambda between Amsterdam and Geneva and a multilayer switching infrastructure in Amsterdam, described in more detail in Chapter 5. These lambda’s are interfaced at both ends as unprotected transparent OC192c, but TDM equipment is installed at each end so that engineers may experiment with customer-provisioned Ethernet circuits to support traffic engineering and quality of service. CA*net4’s current design document also describes a similar approach of using TDM-based equipment as a functional model for future customer-provisioned lambda networking. The distributed Teragrid backbone will consist of unprotected transparent OC192c SONET circuits terminating in Juniper M160 routers. One of its goals is to allow for a “virtual machine room”, where over provisioning of bandwidth in the

wide area network allows for transparent placement of devices across the multi-location machine room.

In each of the above examples institutions are deploying and investigating various attributes of lambda network services, but there are no concrete plans to access the analogue lambda's themselves. Given the various overloaded meanings of "lambda networks", the authors recommend the following extended nomenclature to more precisely identify particular networking techniques: Redefine a lambda from the pure physics interpretation as being the wavelength of light to: "a lambda is a pipe where you can inspect packets as they enter and when they exit, but principally not when in transit. In transit one only deals with the parameters of the pipe: number, colour, bandwidth". This redefinition allows studying the concepts of optical networking using the current available technology and when true optical components become available, the older components can just be replaced.

3. Factorizing the problem space

3.1. Motivation

Optical networking adds another dimension to the field of data-transport. The challenge is to use it as close as possible to the edge of the infrastructure because it promises sheer throughput capacity. The advantages seem numerous: high quality of service, protocol independent, cheap when calculated on throughput capacity. An estimation shows that at the moment of writing this article a 32 lambda optical switch costs about $ 80,000, while a 32×10 Gb/s switch costs about $ 800,000 and a full router with that capacity several millions. This leads to an optimization strategy in the provider backbone where the number of routers is minimized in favour of (optical) switching equipment and thus trying to minimize the combined costs of transport (lambda's and fibres) and active equipment (switches and routers). However, pure optical networks in its current form also have drawbacks such as: static point-to-point connection oriented, telephone system like, management overhead, etc. To understand where optical networking currently makes sense we need to decompose the problem space. We do that by first analysing the current user constituency of the Internet, then defining scales and last determining what services are needed in which circumstances.

3.2. User classification

When looking into the usage of the networks we can clearly classify three user groups. The first group (class A) are the typical home users. Those users can live with ADSL speeds (which is a moving target in time, currently order 1 Mb/s). They typically use that for WWW, mailing, streaming, messaging and peer-to-peer applications. They typically need full Internet routing and flows are generally very short-lived.

A second group (class B) consists of the corporations, enterprises, Universities, virtual organisations and laboratories. Those operate obviously at LAN speeds, currently in the order of 1 Gb/s. Those need mostly layer 2 services, virtual private networks and full Internet routing uplinks. For the business part they typically need many to many connectivity and collaborative support. Usually most of the traffic stays within the virtual organisation.

The third type of users (class C) are the really high end applications, which need transport capacities far above 1 Gb/s. Examples in the science world are the radio astronomers who want to link radio telescopes around the world to correlate the data in real time to improve accuracy to pinpoint sources. Other examples are the data replication effort in the high energy physics field, data base correlation in biology and earth observation data handling. Those applications tend to have long ($\gg$minutes) lasting flows originating from a few places destined to a few places or just point to point. That traffic does not require routing, it always needs to take the same route from source to destination. If we estimate that the backbone load of the total sum of class A is of the same order of magnitude as the class B traffic being around a Gb/s in a country like the Netherlands, then the appearance of a 5 Gb/s class C user on the backbone of a provider is a disturbing event. Typically providers get nervous when their lines and interfaces get regularly populated with traffic loads of several 10's of percents and they then start to discuss upgrading their costly infrastructure. But as seen above it does not seem to make sense to invest in another round of full Internet routers if the major disturbing load on a backbone is coming from

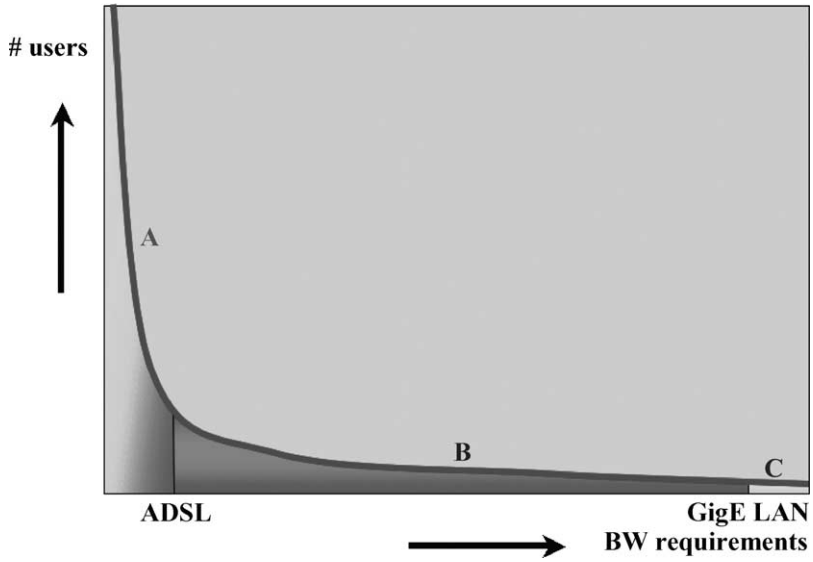

Fig. 2. User classification according to the typical amount of bandwidth usage and type of connectivity used. Class A users are the low bandwidth full Internet access needing email and browser users, class B are VPNs, corporate networks, LANs which need Internet Uplink, class C are the applications with the few big flows.

class C users, which do not need full Internet routing or even layer 2 switching (Fig. 2).

3.3. Scales

Taking the scale of the environment in account makes a big difference when discussing optical networking. We see three typical scales: the metro area, the national or regional scale and the transatlantic or worldwide scale. We discuss here the scale in round trip times in milliseconds (see Fig. 3). A 1 ms RTT equals light transport through 100 km fibre and back. The worldwide scale would then translate to 200 ms, since that computes to 20.000 km or half the circumference of the Earth. Taking orders of magnitude we get a scale of 20 ms or about half of the size of the USA or about the size of western Europe. The 2 ms scale is around the size of Chicago or about half of the Netherlands.

Given the current situation one can easily acquire dark fibre in the 2 ms scale. The number of lambda's is then just limited by the DWDM equipment one deploys. Therefore, tens or hundreds of lambda's on that scale are within reach. On the 20 ms scale owning dark fibre may still be out of reach but owning a number of lambda's on a fibre of a provider is very well possible. On the world scale of 200 ms owning a very few lambda's is currently the limit. The hand waiving arguments estimation is that the number of lambda's to work with is approximately

$$\#\lambda \approx \frac{200\,e^{(t-2002)}}{\text{rtt}}$$

where t is the year and rtt the distance in millisecond round trip time light speed through fibre.

The usage of lambda's and the way of multiplexing, therefore, depends on the scale at hand. Also the regional and world scales will usually involve more administrative domains that require bandwidth on demand and authorization authentication and accounting architectures, which are the topic of another paper [3].

4. Architecture

Given the classification of the users, the necessary services for their packets and the three different scales we can discuss what architectures are optimal in the different situations. For that we make a table listing the scale versus the class of user (Table 1). Given the

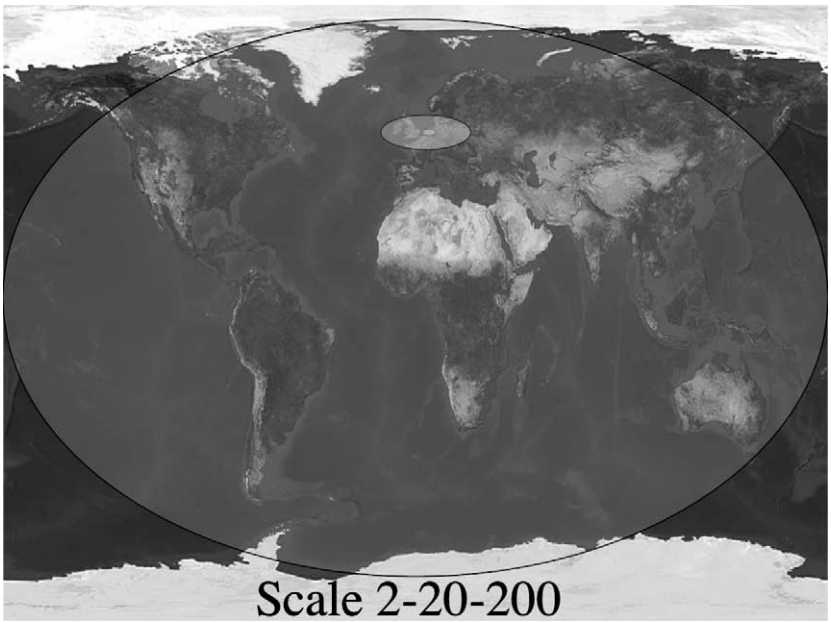

Fig. 3. Scales of networking in round trip times on fibre. A RTT of 1 ms equals 100 km. Therefore, a 200 ms should be enough to reach the other side of the Earth.

different classes of users and the current state of the technology for the different 2, 20 and 200 ms scales we envision to build a provider backbone architecture in which we minimize the amount of routers. The routers are currently the most expensive parts. The traditional model in which the provider places an edge router at each university interconnecting via one or few interfaces to the central router of the University becomes too costly at high speeds. The model is now to provide a DWDM device and to "transport" the A, B, C type connections from the University to either one of the more central core routers for A and B class traffic, switches for class B type traffic and/or directly to the destination if it consists of class C traffic. This means that at customer sites mostly optical equipment will be placed if the scale of the network makes it possible for the provider to own enough lambda's or dark fibres and if the distances are suitable.

At several strategic places in the core network switches and routers can be placed to act as VPN/VLAN or distributed Internet exchange islands, while routers can take care of the traffic needing full Internet routing. The proposed model architecture is shown in Fig. 4. This architecture seems suitable for the metro and most probably the national scale where the provider owns many lambda's. On the edges connectivity needs to be established with the worldwide scale where the amount of lambda's to destinations might be few or just single. The choices then are either router-to-router connection, switch to switch or a completely optical path. The disadvantage of these solutions is the discontinuity for the class C traffic. A fourth solution, which currently is being

Table 1
Applicable technologies for different scales for different classes of users

Class	2 ms (metro)	20 ms (national/regional)	200 (world)
A	Switching, routing, ADSL/ATM	Routing	Router$
B	LAN, VPNs, (G)MPLS	Routing, VPNs, (G)MPLS	Routing
C	Dark fibre, optical switching, Ethernet	Lambda switching	Sub-lambda's, SONET/SDH

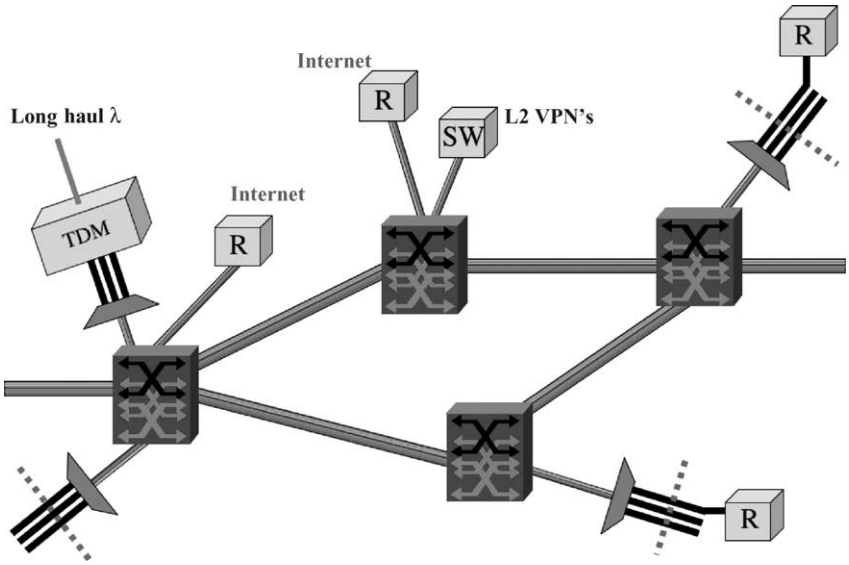

Fig. 4. A differentiated architecture including optical switching, packet switching and routing. The dashed lines denote administrative domain boundaries.

investigated in our test-bed (see Section 5), is to terminate such a worldwide lambda in a TDM switch and to use the sub-channels as a kind of lambda's. Some of the sub-channels can be dedicated for the router-to-router connectivity to service the class A and some B traffic. Other sub-channels can be dedicated to extended VLAN support to provide protected networking environments for class B users or grid virtual organizations. The class C channels can use the remaining channels in a bandwidth on demand fashion; see [3] for a multidomain authorization and provisioning model.

Several questions still need to be answered:

(1) If streams are very long lasting, why have a dynamical optical switching core instead of static patching?
(2) What is the cost model when including all interfaces on each layer?
(3) What is the cost model given a certain amount of over-provisioning of lambda's so that there is something to switch?
(4) What optical switches can be used for which types of signals in different scales?
(5) How to connect different scales infrastructures together?

It is our intention to use our test-beds in collaboration with StarLight and others to investigate these novel new architectures, which are hopefully leading to an optical world.

5. Test-bed

5.1. Architecture

In defining the next-generation Dutch lambda test-bed we want to achieve the following goals:

- Dark fibre infrastructure experiments with full optical components,
- Lambda-based Internet Exchange prototype,
- International protected bandwidth connectivity for virtual organizations spanning multiple domains, scales and architectures.

To achieve these goals we construct a facility in Amsterdam, NetherLight, which serves as a proof of concept next-generation Lambda-based Internet Exchange. On a national scale we create a test-bed where the DWDM and various optical components can find their place to test the models described in

Chapter 4. In the near future we will incorporate pure optical switching elements with millisecond setup times to study true lightpath provisioning.

5.2. NetherLight test-bed activities

For the near future we defined a test-bed to prototype the abovementioned architectural ideas. The Dutch test-bed NetherLight is a next-generation Internet exchange facility with lambda switching capabilities in Amsterdam combined with a national differentiated transport infrastructure for classes A, B and C type traffic. NetherLight focuses on the assessment of optical switching concepts and bandwidth provisioning for high-bitrate applications. In order to achieve these goals, NetherLight deploys TDM circuit switching and Ethernet VLAN switching today and lambda switching in the very near future. Connectivity to similar advanced test-beds is established by means of international lambda circuits to StarLight in Chicago, IL, USA and the Teragrid test-bed, CERN, Geneva, Switzerland. Connections to other grid test-beds are in preparation. International collaborations include those with OMNInet in Chicago, Teragrid, DataTAG at CERN, Canarie.

Fig. 5 shows the NetherLight topology, end of 2002. It contains TDM (SONET) multiplexers in Amsterdam, Chicago and Geneva, for the termination of international OC48c circuits (to be upgraded to OC192c by late December 2002). The TDM multiplexers (Cisco ONS15454 equipment) are used to map multiple GbE channels into the SONET circuits. In Chicago at StarLight, the ONS15454 connects at the LAN (GbE side) to the StarLight switch with 2xGbE ports, and to an ONS15454 from Canarie with 2xGbE for lightpath experiments. In Amsterdam at SARA, the ONS15454 LAN side connects to an Ethernet switch for layer 2 transport experiments by participating research groups. In CERN at Geneva, an ONS15454 multiplexer connects 2xGbE ports to an Ethernet switch for the DataTAG project at CERN.

Connected to NetherLight are:

- The University of Amsterdam,
- The Dutch Institute for Astronomical Research ASTRON/JIVE,
- StarLight, via the ONS15454 at Chicago,
- Canarie, via a direct peering at StarLight.

ASTRON/JIVE connected to the Amsterdam core of NetherLight via a DWDM line system carries several GbE lambda's.

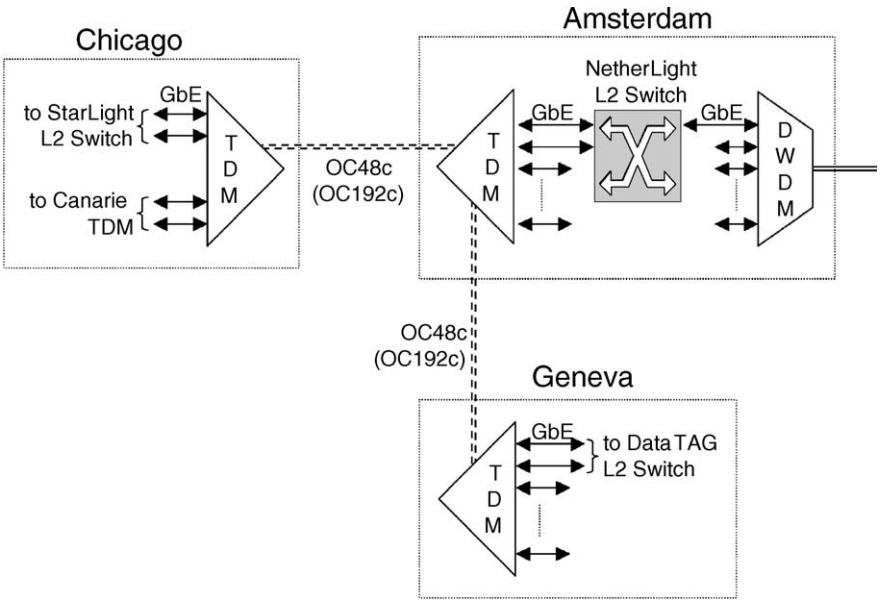

Fig. 5. NetherLight topology as being projected to be in place in beginning 2003.

6. Conclusions and lessons learned

We think that optical networking techniques have the potential of delivering huge amounts of cheap tailored bandwidth to applications. However, we think that for cost and complexity reasons this should not be done with routed backbones but that a differentiated infrastructure delivers the best transport service for different classes of users. First tests with SURFnet's for research only lambda from Amsterdam to Chicago learned that unexpected traffic behaviours surface when real applications are put on the new infrastructure [2]. The properties of the layer one and two infrastructure have profound impact on the layer 4 transport protocols. Multidomain path provisioning is still an open field [3].

Acknowledgements

We wish to thank the following persons for their stimulating discussions: Bill StArnaud, Kees Neggers, Tom DeFanti, Joe Mambretti, Erik-Jan Bos, Victor Reijs. This work is in large part funded by SURFnet and by the IST Program of the European Union (grant IST-2001-32459) via the DataTAG project.

References

[1] P. Bonenfant, A. Rodriguez Moral, Regional optical transport networks, J. Opt. Netw. 1 (2001) 9–17. http://www.osa-jon.org/abstract.cfm?URI=JON-1-1-9.

[2] A. Antony, J. Blom, C. de Laat, J. Lee, W. Sjouw, Microscopic examination of TCP flows over trans-Atlantic links, in: Special Issue on IGRID2002, Amsterdam, The Netherlands, 2002, Fut. Gen. Comput. Syst. 19 (6) (2003) 1017–1029.

[3] L. Gommans, C. de Laat, B. van Oudenaarde, A. Taal, Authorization of a QoS path based on generic AAA, in: Special Issue on IGRID2002, Amsterdam, The Netherlands, 2002, Fut. Gen. Comput. Syst. 19 (6) (2003) 1009–1016.

[4] Optical communication, networks for the next-Generation Internet, Special Issue IEEE Network, vol. 14, No. 6, November/December 2000.

[5] IP-optical integration, Special Issue IEEE Network, vol. 15, No. 4, July/August 2001.

[6] Generic framing procedures. http://www.ewh.ieee.org/r1/njcoast/events/GFP071102.pdf.

[7] ITU, G.LCAS Link Capacity Adjustment Scheme (LCAS) for Virtual Concatenation, October 2001.

Cees de Laat is senior scientific staff member of the Informatics Institute at the University of Amsterdam. He received a PhD in physics from the University of Delft. He has been active in data acquisition systems, heavy ion experiments and virtual laboratories. Over the past 7 years he has been investigating applications for advanced research networks. Current projects include optical networking, lambda switching and provisioning, policy-based networking and authorization, authentication and accounting architecture research. He participates in the European DataGrid project and the Dutch ASCII DAS project. He is responsible for the research on the lambda switching facility ("NetherLight"), which is currently being built in Amsterdam as a peer to StarLight in Chicago. He implements research projects in the GigaPort Networks area in collaboration with SURFnet.

He currently serves as Grid Forum Steering Group member, Area Director for the Peer to Peer area and GGF Liaison towards the IETF. He is co-chair of the IRTF Authentication, Authorization and Accounting Architecture Research group and member of the Internet Research Steering Group (IRSG). http://www.science.uva.nl/~delaat.

Erik Radius is a manager of Network Services at SURFnet, the national computer network for higher education and research in the Netherlands. He received a MSc in experimental physics from the University of Amsterdam and since worked at the University of Twente and at KPN Research on optical network innovations for backbone, metro and access networks. After a brief stint in the Internet startup Bredband Benelux, during which he designed the optical part of a fibre to the home network infrastructure, he joined SURFnet to work on lambda networking and the closer integration of IP over optical networks. In that role, Erik is responsible for NetherLight, the experimental optical internet exchange in Amsterdam. http://www.surfnetters.nl/radius/.

Steven Wallace brings to the Advanced Network Management Lab more than 16 years of experience in data networking, computer programming and network management in both the public and private sectors. His most recent experience includes managing Internet2's Abilene Project for Indiana University and serving as manager and senior technical advisor for Bloomington Data Services at

Indiana University where he oversaw 35 employees and a $ 4 million budget. Wallace co-founded and served as vice president for a computer retail and system integration company with multiple locations, more than 20 employees, and annual sales exceeding $ 1 million. Other experience includes positions as director of networks, manager of network operations and principle applications programmer at Indiana University and several private Indiana firms.

Over the past 10 years, Wallace has focused on the design and operation of large high-speed data networks and the development of network-related management tools. His specific areas of interest and expertise include the optimal design of high-performance layer two networks (Ethernet broadcast domains), the creation of network management tools to support such environments, and research into end-to-end application performance.

Outside of his work in the ANML, Wallace, is an avid collector of early undersea telecommunications cables and a wide range of technical gadgets. He is a weather enthusiast and holds an amateur radio license. Wallace commutes daily by recumbent bicycle from his Bloomington, Indiana home.