Understanding National Optical Networks

National Workshop on Cyberinfrastructure
Doubletree Hotel, Nashville TN
9:45AM, May 11, 2006

Joe St Sauver, Ph.D. (joe@uoregon.edu)

University of Oregon Computing Center

http://ww.uoregon.edu/~joe/national-optical-networks/

I. Introduction

Where I'm "Coming From"

- This talk is <u>not</u> about campus, metro, regional, or international optical networks. Issues of pivotal importance to national optical networks may be completely irrelevant to optical networks at other scales.

- My time horizon is two to three years. Wonderful things may happen farther out, but I'm primarily interested in what's happening in the immediately foreseeable future.

- I'm very concrete and applied: what's the <u>specific</u> real problem that we've identified which we're trying to solve?

- I believe in eating the pork chop that's already on your plate before you go back for 3 more from the buffet: If someone says they need OC192 (10Gbps) service, have they already demonstrated the ability to effectively load an OC48 (2.4Gbps)? If they already have an OC48 but it is largely idle, why not see what they can do with that, first?

Where I'm "Coming From" (continued)

- Ongoing projects are more interesting to me than brief one-off special projects or demonstrations. If you're going to work hard, I believe it makes sense to spend that effort building something strategic, something that will last. Create the Panama Canal, not an ice sculpture.

- Make decisions about projects with a twenty year duration carefully; you'll need to feed that baby until (s)he's an adult.

- Solutions must scale to handle anticipated target audiences (and more). Pay attention to step functions.

- Assume that budgets are limited, and money <u>does</u> matter. What's the business case?

- I like the simplest solution that will work.

- I tend to resist artificial urgency and ignore peer pressure.

My perspective may or may not be consistent with yours…

# Speaking of Perspectives: A Disclaimer

- The University of Oregon is <u>not</u> currently a member of National Lambda Rail, so my perspective with respect to that network is that of a 3rd party/outsider. We <u>are</u> a member of Internet2, and we do operate the Oregon Gigapop.

- That said, the views expressed in this talk are solely my own, and should NOT be taken as expressing those of Internet2, NLR, the University of Oregon, the Oregon Gigapop, the Abilene Network Technical Advisory Committee, or any other entity.

- National scale optical networking continues to be in flux. By the time this meeting is over, this talk will be outdated.

- Do not make any decisions based just on what I'll share during this talk; do your own due diligence and make up your own mind when it comes to the issues discussed.

5

II. Applications and Advanced Networks

Application "Fit" and Advanced Networks

- We believe that if you want to make effective use of advanced networks such as Abilene (or now NLR) you really should spend time thinking about how your prospective applications "fit" with those networks.

- If you <u>don't</u> think about application fit, you may build (or connect to) an absolutely splendid network only to see that facility lay unused.

- Those who remember the NSF HPC connections program will remember that a key component of applying for funding for a vBNS or Abilene connection was identification of specific <u>applications</u> that would actually use those new connections.

- "Applications should motivate new networks, and networks should enable new applications."

The Application-Driven
Network Deployment Process

8
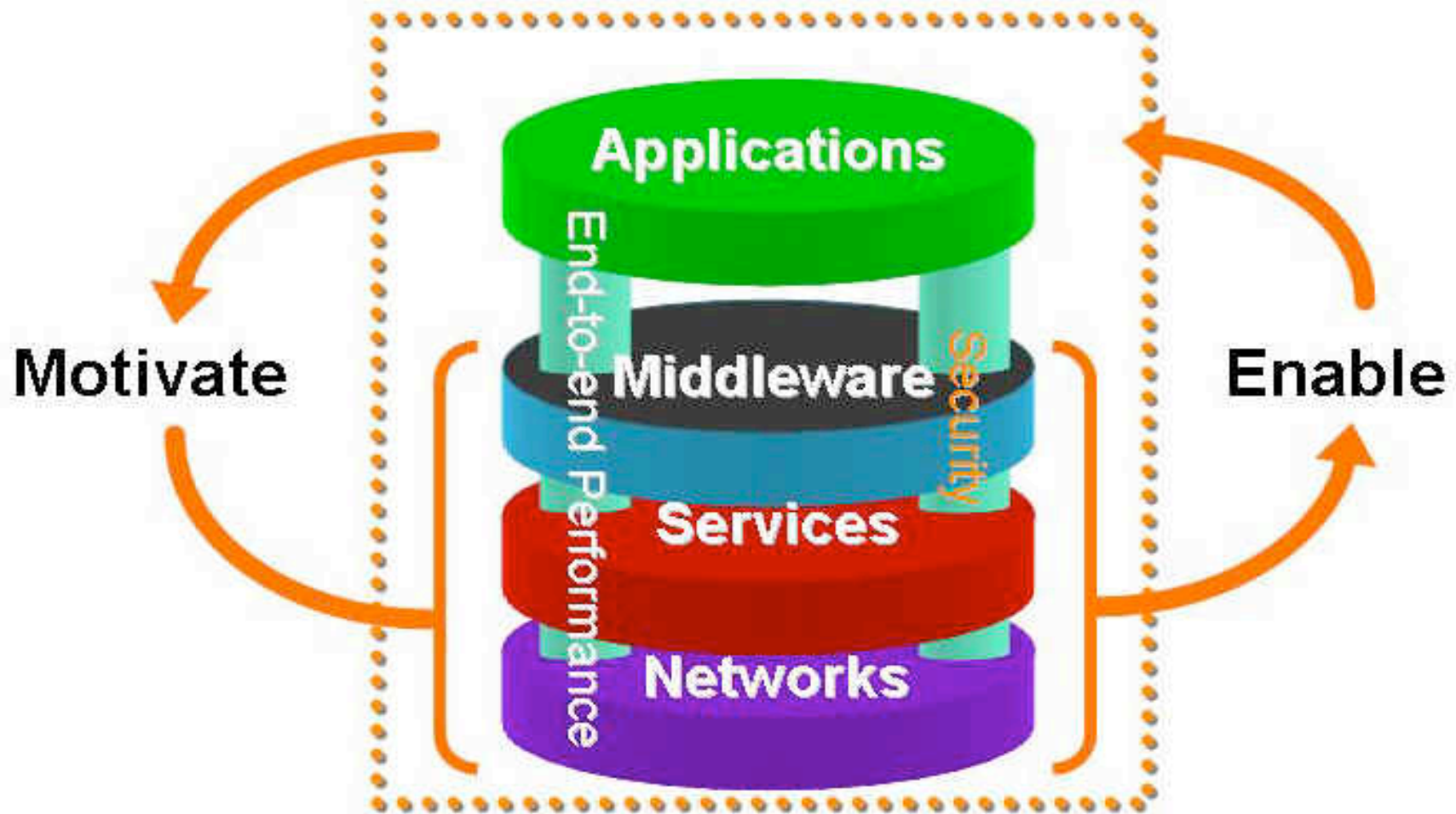
What Applications Do <u>You</u> Have Which Need Lambda-Based National Network Access?

- This is not a rhetorical question. :-)

- If you're <u>not</u> currently looking at your macroscopic traffic levels via **MRTG/RRDTool**, you should be. Moreover, you should also be looking at doing **Netflow** analysis of your network traffic, otherwise it will be hard for you to drill down and understand the traffic you're seeing.

- Beware of users who promise they'll need gigs worth of network capacity unless you're **already** empirically seeing substantial traffic from them in your flows. Be sure to also think about why *lambdas* are needed/why a traditional packet-based network **wouldn't** work.

- That said, it may be worth stepping back a little, beginning by reviewing some basic WDM concepts in the context of one national optical network, NLR.

III. WDM/National Lambda Rail Backgrounder

Lambdas Defined

- A lambda is a specific wavelength, or "color of light," in a wave division multiplexing (WDM) system, running over fiber optic links. Think of this as being kin to using a prism to break the white light that might normally flow over fiber into different colors, each of which can be used to carry information independently of what's going on "in" the other colors.

- By using WDM technology, the amount of traffic that a fiber optic link can carry is multiplied, perhaps to forty times its original capacity. Conceptually, where once a piece of fiber had room for only one channel of network traffic, you can now think of that same piece of fiber as supporting forty parallel independent channels of information, each on its own "lambda" or color of light, with the net result being that one pair of fiber can suddenly act as if it were forty.

"Why Does WDM Gear Always Generate 40 Waves?"

- Sometimes the question comes up of, "Why does WDM gear always provide 40 wavelengths?" The answer, of course, is that it doesn't.

- You can purchase dense wave division multiplexing (DWDM) gear that can yield 80 or 160 or even 320 wavelengths from a piece of fiber, or coarse wave division multiplexing (CWDM) gear that only gives you a 8 or even fewer channels.

- The higher density gear -- because it allows you to cram more channels onto a piece of fiber and because it is built to tighter tolerances -- generally costs more than the coarse, lower channel count, WDM gear.

- The optronics used for NLR, however, does happen to be 40 channel gear (or 32 channel in some cases).

Dedicated Circuits vs. Shared Capacity

- The relative abundance that's associated with WDM makes it possible for us to begin potentially thinking on a national or International scale about <u>dedicated circuits</u> rather than just the shared (or "<u>statistically multiplexed</u>") network capacity that's typical of packet switched networks such as the Internet, or Abilene.

- While it would not make sense for you to set up a lambda just to distribute a web page from someone's web server in New York to a browser in Texas, or to use a lambda to distribute an email message from someone in California to someone in Florida, <u>maybe</u> there will be times when it might make sense to give someone "their own lambda" rather than having them share network capacity with other users. We'll see!

- So how about NLR in particular?

NLR: Born in the Golden State

- Understanding NLR means understanding its roots and original role… CENIC's CALREN, the California research and education network, envisioned three tiers of network service for its constituencies:

  1) Ubiquitous regular/commodity Internet service,
  2) High performance production research and education network access, needed by/of interest to a smaller set of users, such as physical scientists working with large datasets, and
  3) Experimental access to a "breakable" cutting-edge network, offering services needed by an even smaller set of <u>extremely</u> advanced users, such as computer scientists doing bleeding edge network research.

The Three-Tier CENIC CALREN Pyramid
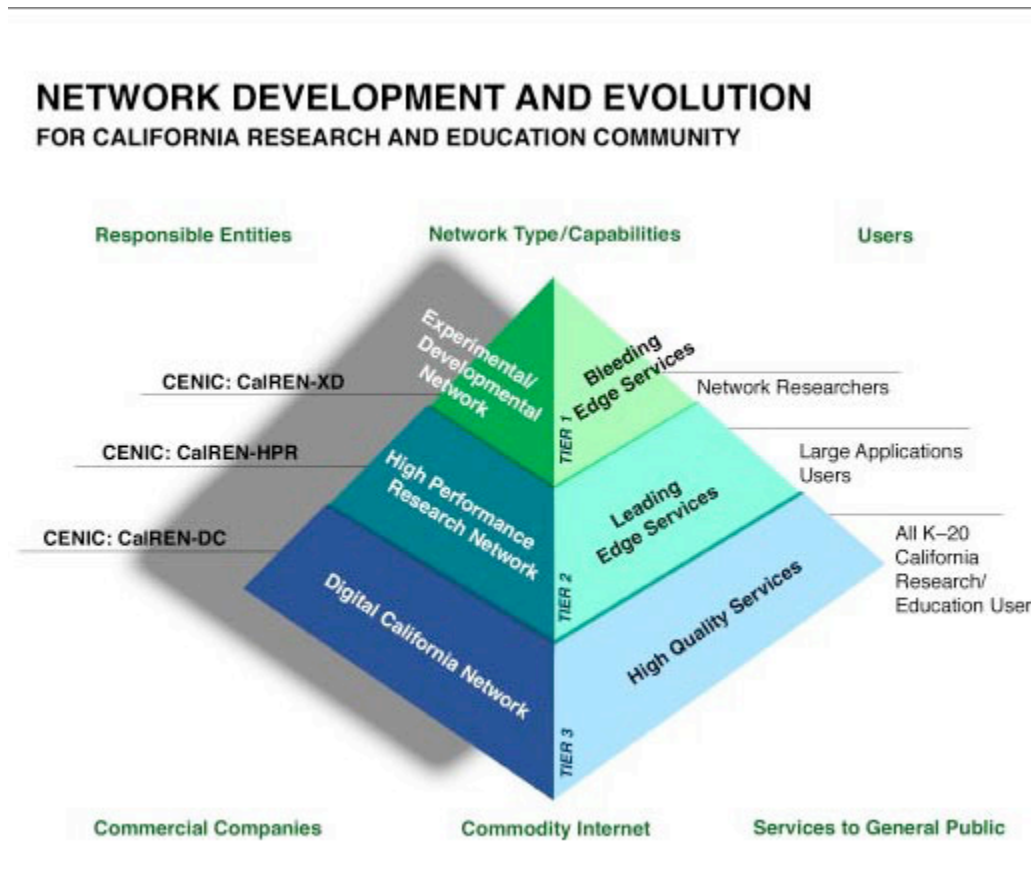
Source: http://www.cenic.org/calren/index.htm
used with permission

Traditional Mapping of Those Three Functions to Educational Networking Organizations

- -- Commodity Internet Service:
  TheQuilt (http://www.thequilt.net/)

  -- High Performance Production R&E Network Service:
  Abilene (http://abilene.internet2.edu/)

  -- *Experimental Access to A Breakable Network: National Lambda Rail (http://www.nlr.net/)*

  Please note: that's the traditional/historical mapping, and should not be taken as being representative of what's happening today.

- So given that interest in/participation in lambda-based networks vastly exceeds the number of folks who need experimental access to a breakable network for network research, is there something else that motivates interest in national optical networks?

Factors Motivating Interest in Optical Networks

- CANARIE, the Canadian research and education network, became an articulate advocate for the simplicity and cost-effectiveness of customer-owned fiber networks

- Gigapops continued to add customers, including state K12 networks ("SEGP"'s), which incented both upgrades to Abilene connections and the creation of regional optical networks, key components of the current NLR model

- More regional fiber was deployed than was needed; wave division multiplexing caused a national bandwidth surplus

- It became possible to swap excess capacity in one region to get capacity on another route for just the cost of hardware

- By purchasing a few additional fiber links, you could tie all those regional networks into a unified national network

- The Internet financial bubble burst, making the needed residual fiber potentially cheap to acquire

Additional Motivating Factors (cont.)

- The Cisco GSR routers that were originally used on Internet2 got replaced with Juniper T640's; after a bit, Cisco released <u>its</u> new uber-router, the CRS-1, and wanted to re-engage the higher ed R&E networking community

- TheQuilt drove commodity Internet prices down about as low as they could go; the only thing that would be cheaper would be settlement free peering. Settlement free peering required the ability to cost-effectively haul commodity Internet traffic to multiple locations nationally.

- Abilene's conditions of use foreclosed some opportunities; for example, Internet2 was limited in its work with federal mission networks. A new network could be AUP free.

- There was concern over being "locked in" to one network provider (Qwest) for all high performance R&E networking.

Additional Motivating Factors (cont. 2)

- The supercomputing community hit a slump and needed to reinvent themselves; grids were born. High performance links were integral to interconnecting those clusters (much as the original vBNS linked traditional supercomputer sites)

- "Big science" embarked on projects which would generate prodigious amounts of data, data which would need to be wheeled around the country and to/from overseas.

- The engineering folks wanted to do something new and fun

- Some folks who were "late to the party" when Internet2 first got started were highly interested and motivated and determined to not miss out the second time around.

- The U.S. developed a "lambda gap" vis-à-vis Europe

- Abilene lost its "elite" cachet (even K12 had access!) and no longer served a winnowing function for research funding

And So NLR Was Born…

- An optical network that was to be many things to many different constituencies, including coming to have some roles far-removed from it's original Californian pyramid capstone niche.

- For the record, NLR's official goals were/are:
  - Support experimental and production networks
  - Foster networking research
  - Promote next generation applications
  - Facilitate interconnectivity among high performance research and education networks

  www.nlr.net/presentations/SC2004_TWW_Slides.htm
  (slide 31)

Current NLR Higher Ed Members (Mostly Consortial)

- Case Western Reserve University
- CENIC
- CIC
- Cornell
- Duke Univ, representing a coalition of NC universities
- Florida Lambda Rail
- Lonestar Education and Research Network
- Louisiana Board of Regents
- Mid-Atlantic Terascale Partnership and the VA Tech Foundation
- Oklahoma State Board of Regents
- Pittsburgh Supercomputing Center and the Univ of Pittsburgh
- PNW Gigapop
- Southern Light Rail
- SURA
- UCAR, representing a coalition of universities and government agencies from Colorado, Wyoming, and Utah
- Univ of New Mexico, on behalf of the State of New Mexico

However NOT "Everyone" Has Joined NLR

- Sometimes there's a misconception that "everyone" has joined NLR…

- By my count, based on the list of participants available at the NLR and I2 web sites, there are <u>108</u> universities which are part of Internet2 but which are NOT NLR participants (more than half of Internet2's membership).*

- Nor would it be correct to assert that the 108 non-members are all fairly inactive Internet2 members, or universities which are disinterested in research. For example, that list of 108 **non**-members includes Carnegie Mellon, Harvard, Johns Hopkins, MIT, Princeton and Yale…

----

* Compare http://www.nlr.net/members_participants.html with
http://members.internet2.edu/university/universities.cfm

Internet2 Schools <u>NOT</u> Part of NLR

Arkansas State University
Auburn University
Binghamton University
Boston College
Boston University
Bowling Green State University
Bradley University
Brandeis University
Brown University
Carnegie Mellon University
Catholic University of America
The City University of New York
Clemson University
Cleveland State University
Dartmouth College
DePaul University
Drexel University
East Carolina University
Emory University
George Washington University
Georgetown University
Harvard University
Iowa State University
Jackson State University
Johns Hopkins University
Kansas State University
Kent State University Main Campus
Lehigh University
Loyola University of Chicago
Marquette University
Massachusetts Institute of Technology

Mayo Clinic College of Medicine
Medical University of South Carolina
Michigan Technological University
Mississippi State University
New Jersey Institute of Technology
North Dakota State University
Northeastern University
Northern Illinois University
Ohio University Main Campus
Oregon State University
Portland State University
Princeton University
Rutgers, The State University of New Jersey
Saint Louis University
Seton Hall University
South Dakota School of Mines and Technology
South Dakota State University
Southern Illinois University at Carbondale
Stephen F. Austin State University
Stony Brook University, State University of New York
Syracuse University
Temple University
Texas Tech University
University of Akron Main Campus
University of Alabama - Tuscaloosa
University of Alabama at Birmingham
University of Alabama in Huntsville
University of Arkansas at Little Rock
University of Arkansas for Medical Sciences
University of Arkansas Main Campus

23

Internet2 Schools <u>NOT</u> Part of NLR (cont)

University of Cincinnati Main Campus
University of Connecticut
University of Delaware
University of Kansas Main Campus
University of Kentucky
University of Louisville
University of Maine
University of Maryland Baltimore County
University of Maryland Baltimore
University of Maryland College Park
University of Massachusetts
University of Medicine and Dentistry of New Jersey
The University of Memphis
University of Mississippi
University of Missouri-Columbia
University of Missouri-Kansas City
University of Missouri-Rolla
University of Missouri-Saint Louis
University of Nebraska
University of New Hampshire
University of North Dakota Main Campus
University of North Texas
University of Notre Dame
University of Oregon
University of Pennsylvania
University of Puerto Rico
University of Rhode Island
University of South Carolina, Columbia
University of South Dakota
University of South Florida
University of Southern Mississippi

University of Tennessee
University of Toledo
University of Tulsa
University of Vermont
University of Wisconsin-Milwaukee
Vanderbilt University
Wake Forest University
Washington University
Wayne State University
West Virginia University
Western Michigan University
Wichita State University
Widener University
Worcester Polytechnic Institute
Wright State University
Yale University

Being an NLR Participant vs. Actually Pushing Traffic

- In thinking about NLR, it is also important to distinguish between being an NLR participant, and actually pushing traffic (whether that's via one or more dedicated NLR lambdas or via NLR's shared infrastructure).

- We believe that there are currently at least some NLR participants who do NOT exchange traffic over the NLR infrastructure, either because:
  -- they don't currently need those unique capabilities, or
  -- because they have infrastructure issues that need to be worked out before they can physically do so.

- Assuming users do want lambda-based networks to actually move bits, what general network attributes might they be hoping to get?

IV. General Capabilities

Network Availability/SLAs?

- For example, would a lambda-based network like NLR give us greater network availability/tighter service level agreements vis-à-vis Abilene?

- That would actually be pretty hard to accomplish given that Abilene's network architecture and protection mechanisms have resulted in core node network availability that has averaged 99.9978% (see http://www.internet2.edu/presentations/ spring06/20060425-abilene-cotter.pdf at PDF slide 5). This means that a dual-homed Abilene connector should effectively have NEVER have seen a loss of Abilene reachability. [In general, lambda based networks often offer LESS protection or longer restoration times than traditionally engineered SONET-based networks.]

Premium Quality of Service (QoS)?

- Or maybe traffic sent cross-country via a dedicated lambda is somehow "better" than best-effort traffic sent via an uncongested (but shared) Abilene connection?
-- Will we see lower latency?
-- Less jitter?
-- Less packet loss?
-- Higher throughput?
Is NLR at root a <u>wide area premium QoS project</u>?
[Y'all may know how much I just "love" QoS, particularly for interdomain applications across a lightly loaded/over provisioned core network… screen door on a submarine, folks.]

If <u>Not</u> Better-Than-Best-Effort Traffic, Maybe We're Looking for Bandwidth That's Above What Abilene Offers?

- If NLR is not about better-than-best-effort service, then what <u>is</u> it about?

- Is it about providing relief for traffic levels that cannot be accommodated by the already available Abilene connections, including 10GigE/OC192 connections? For example, will the "default" NLR connection not be a single 10Gig pipe, but some aggregate of two, three or more? Are traffic levels necessitating those sort of pipes already discernable, or known to be coming in the foreseeable future?

- During the Spring '06 Internet2 Member Meeting, it was mentioned that the fiber and optronics vendors envisioned for the new Abilene will allow it to go to 40Gbps & 100Gbps interfaces when those are needed…

Or Is It Just About Cost/Bit Carried?

- Or is it a matter of carrying that sort of bulk traffic over lambda-based connections at a lower cost than current Abilene 10 gigabit connections? Currently Abilene 10gig connections cost $480K/year, or ~$20.83/Mbps/month (assuming that connection is fully loaded).

- If you just want cheap bandwidth, in 2003 Cogent publicly disclosed pricing with TheQuilt as low as $10/Mbps/month (see http://www.net99.net/htdocs/press.php?func=detail&person_id=27), and more recently there have been some ISPs that have offered $10/Mbps/month pricing commercially (e.g., see http://www.askwebhosting.com/special/981/He.net_Badwidth_Promotion.html)

- The price you get depends on <u>where</u> you are, <u>how much</u> you want to buy, traffic patterns, and other factors.

COU-Related Motivations?

- Is an important role for NLR the carrying of traffic that can't be carried over Abilene for **policy** reasons?

- For example, the Abilene Conditions of Use ("COU") (see http://abilene.internet2.edu/policies/cou.html) states "Abilene generally is not for classified, proprietary, unrelated commercial, recreational, or personal purposes."

- If that's the key motivator behind moving to NLR for you, note that Internet2 announced at the Spring 06 Member Meeting that Abilene will be changing its conditions of use to allow commercial traffic.

'Mission Network' Traffic?

- Related to commodity internet/commercial traffic (in terms of having COU-limited access to Abilene) is mission network traffic. [Mission networks are the high-performance networks run by federal agencies in support of their scientific research programs such as the Department of Energy's ESNet, DOD's DREN, NASA's NREN, etc.] Mission networks connecting to Abilene historically/traditionally did NOT see the full set of routes that regular higher ed connectors got (see http://abilene.internet2.edu/policies/fed.html).

- That restrictive routing policy limits the usefulness of Abilene for mission-network-connected agencies, and may have motivated interest by at least some of those agencies in AUP-free alternatives such as NLR, but again, that's something that Abilene is in the process of fixing.

Lambda-based Networks and <u>Local</u> Policy Issues

- The commodity Internet constraint and the mission network constraint just mentioned are examples of policy-driven <u>Internet2-level</u> network limitations, but they may not be the only policy-driven problems which NLR may be used to overcome -- there may also be local policy artifacts.

- For example, it is easy to overlook the extent to which local perimeter firewalls (or other mandated "middleboxes") can cause problems for some applications, particularly if you're trying hard to go fast or do something innovative. It will often be virtually impossible to get an exemption from site-wide security policies for conventional connections.

- On the other hand, if you're bringing in a *lambda*, that lambda will both have a different security risk profile and may not even be <u>able</u> to be handled by available firewalls. Thus, it may be exempted from normal security mandates.

Coverage in Tough-to-Reach Areas?

- NLR could have been a way to tackle other issues, too.

- For example, NLR might have been a solution for some Internet2 members in geographically challenged parts of the country (e.g., our Northern Tier friends in the Dakotas, for example).

- Hmm… maybe, but remember that in NLR's case, the network footprint closely follows the existing Abilene map, with access network issues generally remaining the responsibility of a regional networking entity rather than being handled directly. NLR wasn't meant to fix the "Northern Tier" problem (although who knows what may become possible in the future).

- See http://www.ntnc.org/default.htm for more information about the Northern Tier Network Consortium.

Research Conducted Via the Network vs. Networking Research

- I would be remiss if I did not acknowledge that NLR does not exist solely for the purpose of serving those doing research <u>via</u> the network (such as those working with supercomputers, or physicists moving experimental data). Another major role is support for research <u>about</u> networking.

  Quoting Tom West:
  > "NLR is uniquely dedicated to network research. In fact, in our bylaws, we are committed to providing at least half of the capacity on the infrastructure for network research."

  http://www.taborcommunications.com/hpcwire/hpcwireWWW/04/1110/108776.html

Experimenting on Production Networks

- Most computer science networking experiments can be run on the Internet (or over Abilene) without disrupting normal production traffic. <u>Some</u> experiments, however, are radical enough that they have the potential to go awry and interfere with production traffic.

- When Abilene was first created, there was hope among computer scientists that it might remain a "breakable" network capable of supporting extreme network experimentation, but Abilene quickly became a production network upon which we all depended, and thus too mission-critical to potentially put at risk.

- Given that, one possible niche for a national lambda-based network would be as breakable infrastructure upon which risky experimentation can (finally) occur.

- Recall NLR's original role in the CALREN service pyramid₃₆

But Is A <u>National Scale</u> Breakable <u>Lambda-Based</u> Experimental Network What's Needed?

- When thinking about a breakable network testbed, the question that needs to be asked is, "Does such a network need to actually have a <u>national</u> footprint? Or could the same experiments be done in a testbed lab located at a single site, or perhaps on a state-scale or regional-scale optical network? Does that testbed need to be <u>in the ground/at real facilities</u> or could that sort of work be handled satisfactorily with reels of fiber looped back through WDM gear in a warehouse, instead?

- Is it sufficient for a national scale network testbed facility to be at the <u>lambda</u> level, or are we still "too high up the stack"? Will critical research involving long haul optics, for example, actually require the ability to work at layer 0, in ways that (once again) might be incompatible with production traffic running over that same glass?

General Possibilities vs. Specific Applications

- The preceding are all general possibilities relating to national optical networking.

- While it is fine to talk about general possibilities for NLR, when access to NLR becomes more broadly available, how, <u>specifically</u>, will lambda-based architectures likely end up being used?

- One approach to seeing what's well-suited to NLR is to take a look at how NLR is <u>currently</u> being used by early adopters, looking perhaps for common application themes or characteristics.

V. Current NLR Layer 1 ("WaveNet") Projects

Public NLR Layer 1 Projects

- There are a number of publicly identified NLR layer one (lambda-based) testbed projects at this time (see http://www.nlr.net/supported.html). They are:
  1) The Extensible TeraScale Facility (TeraGrid)
  2) OptIPuter
  3) DOE UltraScience Net
  4) Pacific Wave Extensible Peering Project
  5) Internet2 HOPI project
  6) Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA)

- Some additional projects not mentioned on that page include Cheetah and regional initiatives using NLR waves

- NLR also provided wavelengths for SC2004- and SC2005-related activities

The Sept 12th-14th 2005 NASA Meeting

- With respect to information about current applications, there was an invitation-only NASA meeting at which roadmaps for many NLR-related projects were discussed. See: "Optical Networks Testbed Workshop 2" http://www.nren.nasa.gov/workshop8/

- **If you end up looking at only one presentation from that workshop, make it Robert Feurstein (Level3)'s: "A Commercial View of Optical Networking In the Near Future," http://www.nren.nasa.gov/workshop8/pps/ 17.F.Level3_Feuerstein.ppt (also known as the "Poppycock/Forgeddabout It/ Hooey/Malarkey" talk)**

VI. NLR Native L2 ("FrameNet") and L3 ("PacketNet") Services

The NLR L2 ("FrameNet") and L3 ("PacketNet") Services

- In addition to the specific special projects mentioned in the preceding section (all basically L1 based), NLR also offers ubiquitous NLR layer two and layer three services to NLR participants. Those services represent a minimum commitment of two of the five pre-defined full footprint NLR waves:
  1) NLR Layer 2 service
  2) NLR Layer 3 service
  3) HOPI wave
  4) hot spare
  5) Wave in support of network research projects (being equipped by Cisco's Academic Research and Technology Group)

  www.nlr.net/docs/NLR.quarterly.status.report.200503.pdf

The Commonly Seen Map of NLR: Many L1 POPs
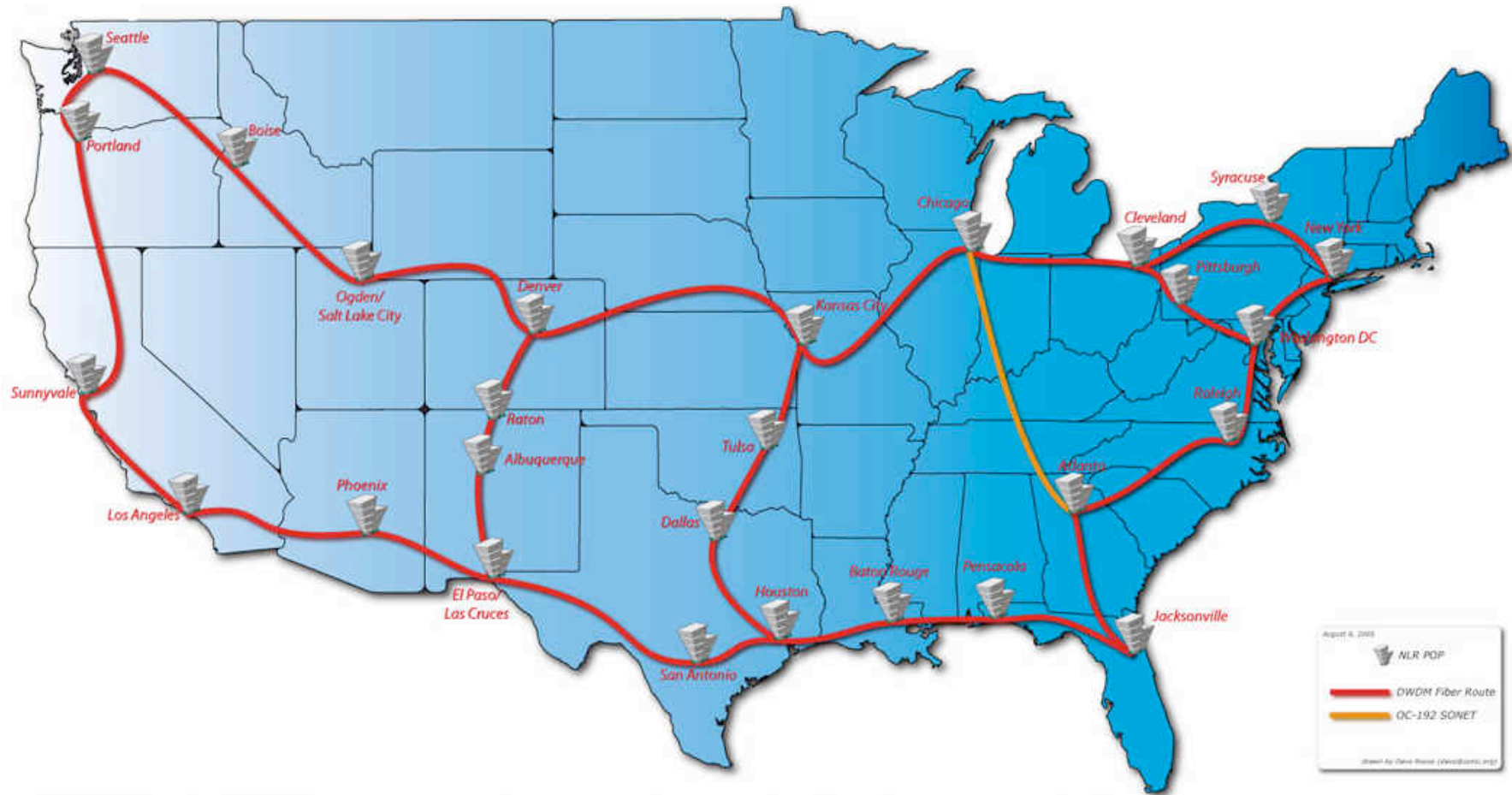
© 2005 National LambdaRail™   For more information regarding NLR see http://www.nlr.net or contact info@nlr.net

http://www.nlr.net/images/NLR-Map-large.jpg
Image credit: National Lambda Rail, used with permission.

Less Commonly Seen: The FrameNet Traffic Map

- To see the current NLR Layer 2 (FrameNet) topology, see the NLR Layer 2 Network Status Weathermap at: http://weathermap.grnoc.iu.edu/nlrmaps/layer2.html

Just what Is the NLR L2 Service?

- Caren Litvanyi's talk "National Lambda Rail Layer 2 and 3 Networks Update" (http://www.internet2.edu/presentations/ jtvancouver/20050717-NLR-Litvanyi.ppt) is excellent and provides the best description… Excerpts include:

- "Provide circuit-like options for users who can't use, can't afford, or don't need, a 10G Layer1 wave."

- "MTU can be standard, jumbo, or custom"

- "Physical connection will initially be a 1 Gbps LX connection over singlemode fiber, which the member connects or arranges to connect."

- "One 1GE connection to the layer 2 network is part of NLR membership. Another for L3 is optional."

What Is the NLR L2 Service? (cont.)

- Continuing to quote Litvanyi…
"Initial Services:

   "--Dedicated Point to Point Ethernet – VLAN between 2 members with dedicated bandwidth from sub 1G to multiple 1G.

   "--Best Effort Point to Multipoint – Multipoint VLAN with no dedicated bandwidth.

   "--National Peering Fabric – Create a national distributed exchange point, with a single broadcast domain for all members. This can be run on the native vlan.
This is experimental, and the service may morph."

- Litvanyi's talk includes a list of NLR L2 street addresses (can be helpful in planning fiber build requirements)

Some Thoughts About NLR L2 Service

- NLR L2 service is likely to be the most popular NLR production service among the pragmatic folks out there:
  -- it is bundled with membership at no additional cost
  -- the participant-side switch will be affordable
  -- the L2 service has finer grained provisioning that is most appropriate to likely load levels

- Hypothetical question: assume NLR participant wants to nail up point to point L2 VLAN with participant at CHI with dedicated 1Gbps bandwidth. Later, ten additional participants ALSO want to obtained dedicated 1 Gbps VLANs to CHI across some common part of the NLR L2 shared wave. What's the plan? Will <u>multiple</u> NLR lambdas be devoted to handle that shared L2 service load? Will some of that traffic get engineered off the hot link? Will additional service requests just be declined?

NLR Transit and Peering Project

- NLR has now announced their Peering and Transit project; see http://www.nlr.net/20060420-PR.htm
As the first part of that project, CENIC, PNW Gigapop, Front Range GigaPoP, the Mid- Atlantic Terascale Partnership, and Pittsburgh Supercomputing Center will endeavor to shift commodity transit (and peering) traffic onto NLR's "TransitRail" service; after 9 months, all NLR participants will be able to participate. (Note that CENIC and PNWGP were already involved with a west coast distributed peering project)

Another Map: The NLR <u>PacketNet</u> Traffic Map

- To see the current NLR Layer 3 (PacketNet) topology, see the NLR Layer 3 Network Status Weathermap at: http://weathermap.grnoc.iu.edu/nlrmaps/layer3.html
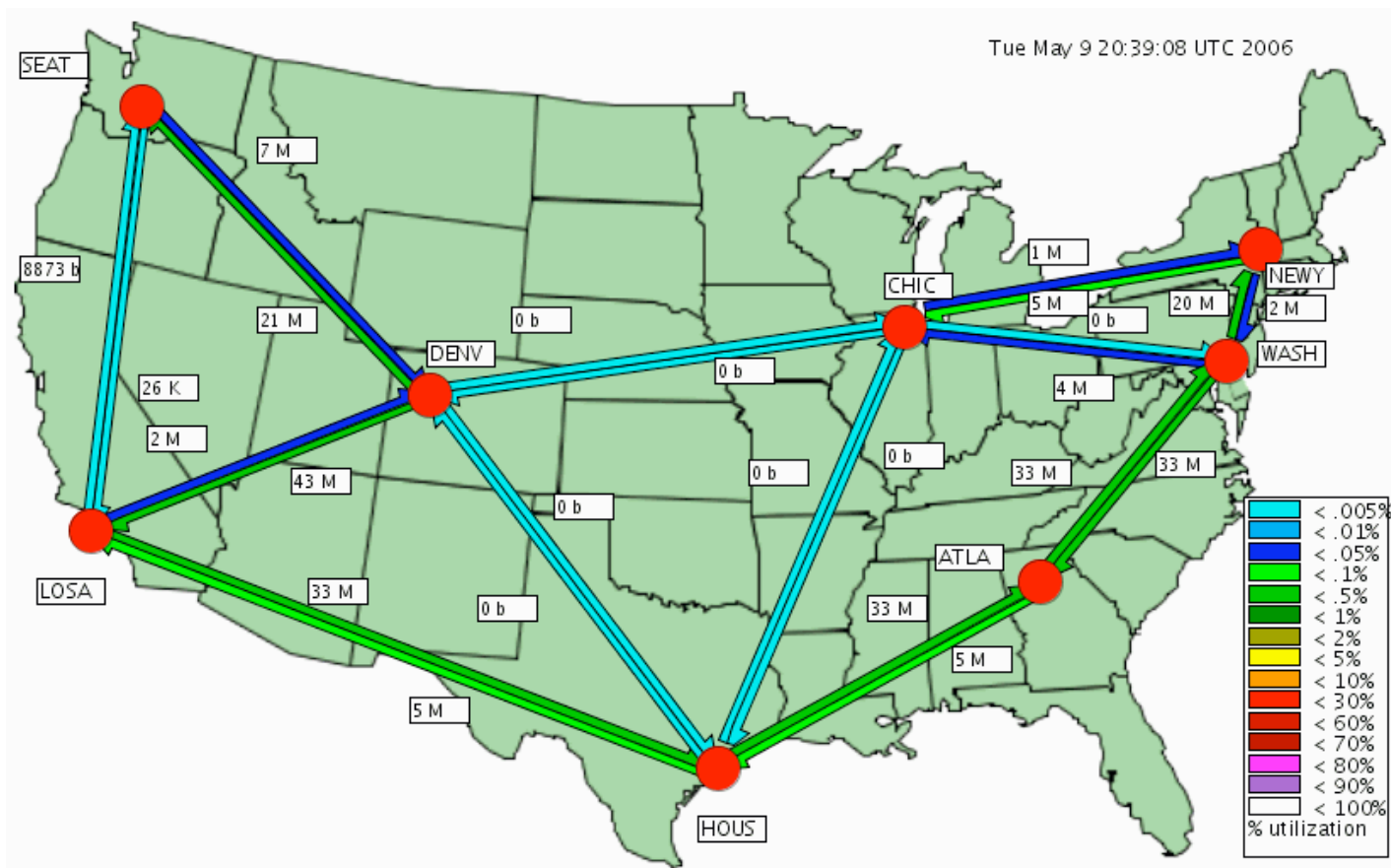
What Is NLR L3 Service?

- Again quoting Litvanyi's "National Lambda Rail Layer 2 and 3 Networks Update"…

- "Physical connection will be a 10 Gbps Ethernet (1310nm) connection over singlemode fiber, which the member connects or arranges to connect."

- "One connection directly to the layer 3 network is part of NLR membership, a backup 1Gbps VLAN through the layer 2 network is optional and included."

Random Notes About NLR L3 Service

- Probably obvious, but….
  Total $ Cost to NLR for each L3 routing node >>
  Total $ Cost to NLR for each L2 switching node >>
  Total $ Cost to NLR for each L1 lambda access POP
  (e.g., higher layer site also have the lower layer equipment)

- Demand for L3 service may be limited: 10Gbps routers and router interfaces don't come cheap.

- L3 participant backhaul will burn incremental lambdas

- Default L3 access link speed (10Gbps) is equal to the core network speed (10Gbps); implicitly, any L3 participant has sufficient access capacity to saturate the shared L3 core.

- NLR was assigned AS19401 for its use on 2005-05-31

Abilene and NLR L2/L3 Geographical Matrix

Site	Abilene Router	NLR CSR-1 Node	L3 Stub	L2 Node
Atlanta	X	X	n/a	X
Chicago	X	X	n/a	X
DC	X	X	n/a	X
Denver	X	X	n/a	X
Houston	X	X	n/a	X
Indianapolis	*X*	*NO*	*NO*	*NO*
Kansas City	*X*	*NO*	*NO*	*X*
LA	X	X	n/a	X
New York	X	X	n/a	X
Seattle	X	X	n/a	X
Sunnyvale	*X*	*NO*	*NO*	*X*
Albuquerque	NO	NO	X	X
Baton Rouge	NO	NO	X	X
Jacksonville	NO	NO	X	X
Pittsburgh	NO	NO	X	X
Raleigh	NO	NO	X	X
Tulsa	NO	NO	X	X
Cleveland	NO	NO	NO	X
El Paso	NO	NO	NO	X

AS19401 Routes (Routeviews.Org, 2006-05-09-2000)

- 5050 1206
  5078
  5078 3923
  5078 30116
  7066 225
  10764
  19718
  32361

- Those Autonomous System Numbers belong to:
  AS225:   University of Virginia
  AS1206:   Pittsburgh Supercomputing Center
  AS3923:   FAA
  AS5050:   Pittsburgh Supercomputing Center
  AS5078:   OneNet (Oklahoma)
  AS7066:   Network Virginia
  AS10764   NCSA (Illinois Urbana Champaign)
  AS19718:   NCNI (Research Triangle NC)
  AS30116:   Weathernews Americas, Inc.
  AS32361:   UltraLight (CalTech)

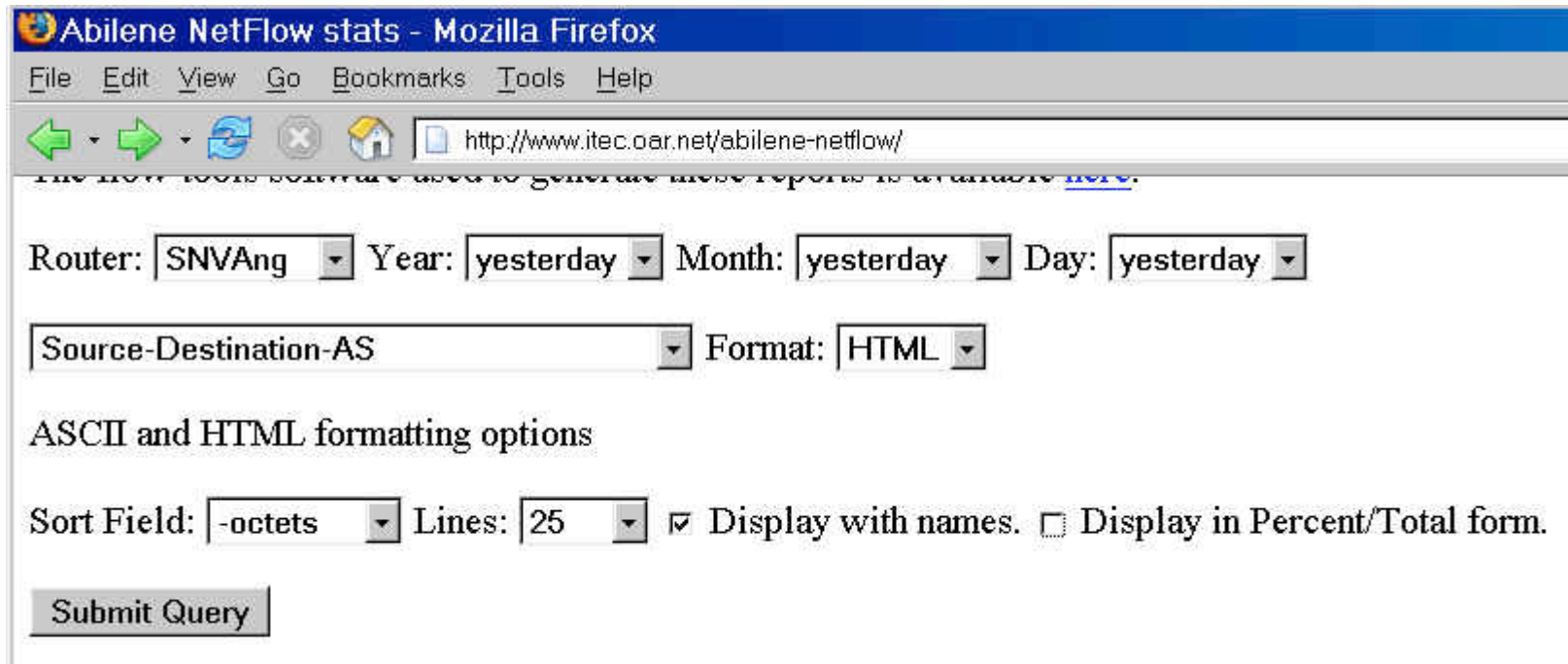# VII. So Let's Come Back to The Classic High Bandwidth Point-to-Point Traffic Scenario

Sustained High Bandwidth
Point-to-Point Traffic

- If you're facing sustained high bandwidth point-to-point traffic, <u>that</u> is usually pointed to as the classic example of when you might want to use a dedicated lambda to bypass the normal Abilene core.

- Qualifying traffic is:
  -- **NOT** necessarily the FASTEST flows on Abilene (why? because those flows, while achieving gigabit or near gigabit speeds, may only be of short duration)
  -- **NOR** are you just looking for a SINGLE large flow that transfers the most data per day (some applications may employ multiple parallel flows, or be "chatty," repeatedly opening and closing sessions, or there may be multiple applications concurrently talking between two sites, flows which <u>when aggregated</u> represent more traffic than any individual large flow).

Identifying Potential Site Pairs for Lambda Bypass

- Okay then… so how <u>do</u> we spot candidate traffic which we might want to move off the Abilene core?

- First step in the process is basically the same one involved in hunting for commodity peering opportunities: analyze existing source X destination traffic matrices, looking for the hottest source-destination traffic pairs.

- Internet2 kindly provides netflow data, including per-node top source-destination aggregates.

- For example, we can look at what's happening at Sunnyvale (we'll only look at one day's worth of data; in reality, you'd obviously want to look at a much longer period to develop baselines)…

The Abilene Netflow Web Interface

Sample Output

NetFlow Report - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.itec.oar.net/cgi-bin/abileneproc3?router=SNVAng&year=00&month=00&day=00&report=Source-D

\#

source-as	destination-as	flows	octets	packets	duration
ABILENE	ABILENE	154607	4215952461900	697720500	1644201160
UMDNET	APNIC-AS-X-BLOCK	7155	528495456200	1060376900	336041856
APNIC-AS-X-BLOCK	UMDNET	7211	521387009800	1006966300	335594560
FERMILAB	ANBR-AS	34239	232552191400	165580500	1033410396
UCLA	UNIV-ARIZ	18836	229356128000	162890500	859981988
UONET	CONCERT	92451	200343316100	164249700	2088545549
UONET	CSUNET-NW	35520	168177299400	119901700	964113119
UONET	0	43881	159725673300	144528700	2041478200
UONET	WASHINGTON-AS	38786	157740526700	111336200	858185773
UONET	BCNET-AS	21190	155945265800	106094100	477494034
UONET	STANFORD	20169	142686890600	97952300	484137734
CSUNET-NW	0	3187	131341194900	89113000	171663645
UONET	UCLA	15927	116639378800	99552200	470124248
PENN-STATE	USC-OBERON	8228	116503094300	78162800	341723981
ORST-AS	UMN-AGS-NET-AS	20308	110994418700	75666000	885087182
FRGP	UONET	30465	104179028000	83800700	531979157
UONET	ALASKA	16422	100908184500	68076500	556786799

59

Percents Rather Than Really Big Numbers…

```
# records:                189178
# first-flow:             1126051201 Wed Sep  7 00:00:01 2005
# last-flow:              1126137597 Wed Sep  7 23:59:57 2005
# now:                    1126170674 Thu Sep  8 09:11:14 2005
#
# mode:                   streaming
# compress:               off
# byte order:             little
# stream version:         3
# export version:         5
#
```

| source-as	destination-as	flows	octets	packets	duration
11537	11537	0.380791	26.558490	3.930691	0.612359
27	2500	0.017622	3.329269	5.973759	0.125154
2500	27	0.017760	3.284489	5.672864	0.124987
3152	1251	0.084329	1.464968	0.932817	0.384879
52	1706	0.046392	1.444834	0.917663	0.320288
3582	81	0.227703	1.262067	0.925320	0.777849
3582	2152	0.087484	1.059437	0.675480	0.359070
3582	0	0.108077	1.006196	0.814220	0.760319
3582	73	0.095528	0.993690	0.627226	0.319619
3582	271	0.052190	0.982381	0.597694	0.177836
3582	32	0.049675	0.898859	0.551826	0.180310

Some Thoughts on That Sample Traffic Data…

- For Sunnyvale, for this day, the top source-destination pair (>26% of octets) is obviously intra-Abilene traffic (presumably iperf measurement traffic).

- It would probably <u>not</u> be a good idea to move traffic that's specifically designed to characterize the Abilene network onto a network other than Abilene. Some things you just need to leave where they are. :-)

- Excluding measurement traffic, nothing else jumps out at us at the same order of magnitude… ~3% of traffic seen at that site (the next highest traffic pairing) is probably <u>not</u> enough to justify pulling that traffic out of the shared Abilene path for those nodes, especially since the Abilene backbone itself is still uncongested.

- The lack of promising opportunities for bypass shouldn't be surprising since traffic normally isn't highly localized.

And Even <u>10%</u> of 3Gbps Wouldn't Be All That Much

- If you assume that…
  -- the Abilene core as shown on the Abilene weather map is running *maybe* 3Gbps on its hottest leg
  -- an absurdly high estimate for the level of flow locality (or point-to-point concentration) might be 10% of that, excluding iperf traffic (remember, reality is ~3%)
  -- the unit of granularity for bypass circuits is a gigabit…
  THEN you really don't have much hope for discovering a set of ripe existing gigabit-worthy bypass opportunities: 10% of 3Gbs is just 300 Mbps

- Yeah, 300 Mbps isn't peanuts, but it also isn't anything that the existing Abilene core can't handle, and it seems a shame to "waste" a gig (or even 10gig!) circuit on just 300Mbps worth of traffic when the existing infrastructure can handle it without breaking a sweat.

Current Abilene Traffic Levels
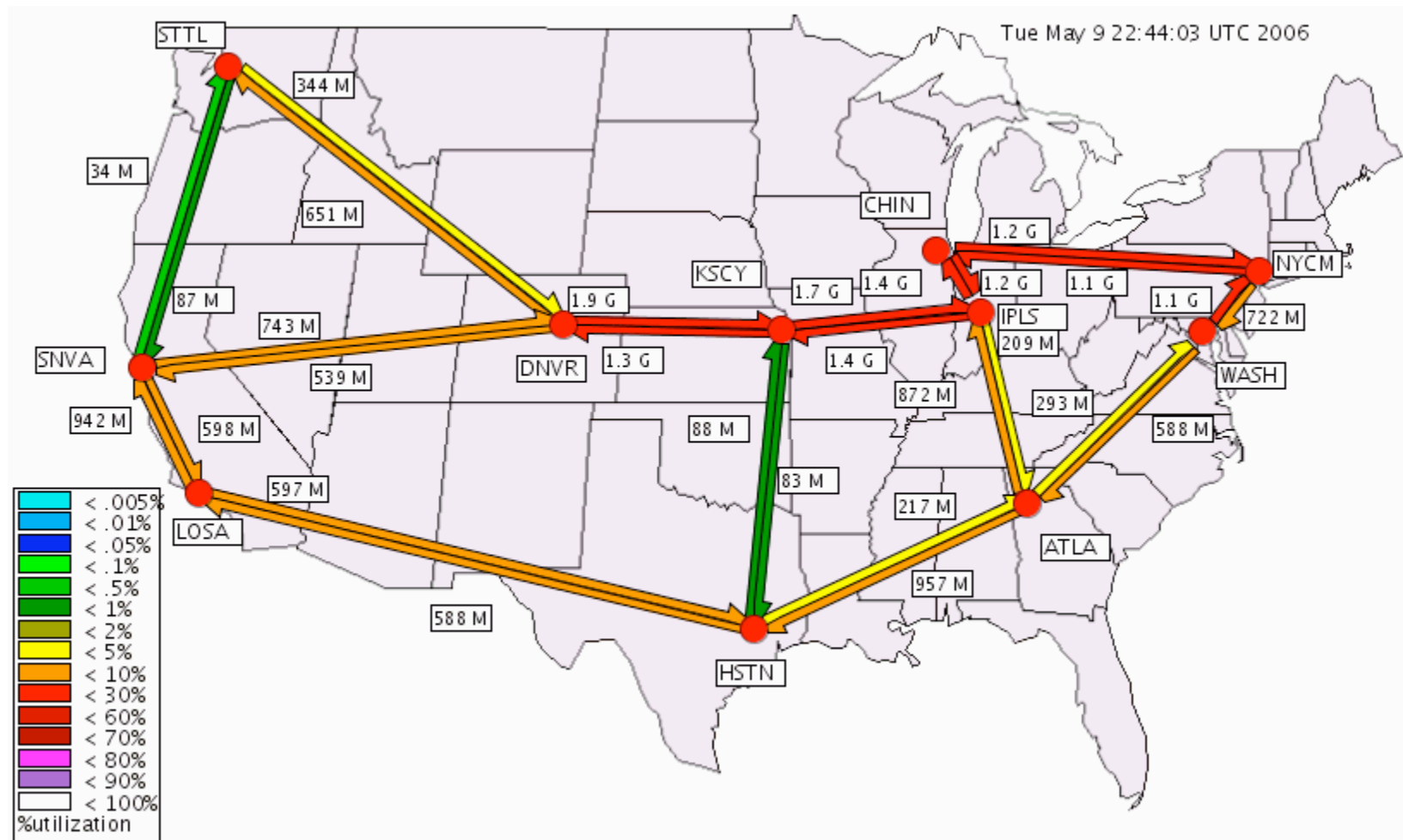
What About From The Perspective of an Individual Connector?

- Even if it doesn't make sense from Abilene's point of view to bother diverting a few hundred Mbps onto NLR, what about from the perspective on an individual connector? For example, what if an Abilene OC12 (622 Mbps) connector was "flat-topping" during at least part of the day? Should <u>they</u> try diverting traffic onto NLR, bypassing/offloading their hypothetical current Abilene OC12 connection, *or* should they upgrade that regular Abilene connection to GigE, OC48, or 10GigE/OC192?

- The issue is largely economic – NLR costs a minimum of $5 million over 5 years, while the incremental cost of going to even 10GigE/OC192 from OC12 is just ($480,000/yr-$240,000/yr), or $1.2 million over 5 years. If you as a connector need more capacity, just upgrade your existing Abilene circuit.

ASNs vs. Larger Aggregates

- The analysis mentioned on the preceding pages was done on an autonomous system by autonomous system (ASN x ASN) basis. [If you're not familiar with ASNs, see http://darkwing.uoregon.edu/~joe/one-pager-asn.pdf for a brief overview.] At least in the case of NLR lambdas, ASNs may be too fine a level of aggregation.

- Given the consortial nature of many NLR connections, it may make more sense to analyze traffic data at the NLR-connection X NLR-connection level instead.

- We keep coming back to the problem, though, that core Abilene traffic levels, while non-trivial, just aren't high enough to justify the effort of pruning off existing flows.

"What About Those Anticipated Huge Physics Data Flows I Keep Hearing About?"

- If you're thinking of the huge flows that are expected to be coming in from CERN, those will be handled by NLR all right, but via the DOE Science Data mission network described earlier in this talk. I'm fully confident that they've got things well in hand to handle that traffic, ditto virtually any other commonly mentioned mega data flows.

- If you know an example of one that's NOT already being anticipated and provided for, I'd love to hear about it.

VIII. The Paradox of Relative Resource Abundance

One Wavelength? <u>Plenty</u>.
Forty Wavelengths? <u>Not Enough</u>.

- Abilene currently runs on just one wavelength – 10 Gbps -- and that's enough, at least for now.

- NLR, on the other hand, has forty wavelengths -- 400 Gbps -- but because of the way those wavelengths may get allocated, that may not be "enough" (virtually from the get go).

- It would thus be correct, in a very Zen sort of way, to talk about it being both very early, and possibly in some ways already "too late," when it comes to getting involved with NLR.

Do The Math…

- We start with 40 waves, half reserved for network research

- Of the remaining 20, at LEAST four were allocated "at birth" (shared L2 service, shared L3 service, HOPI, 1 hot spare) -- 16 are left after that. (I say "at least 4" because L2 service may be so popular that it could need multiple lambdas.)

- There are 15 known NLR members already. If each participant wanted even *one* full-footprint non-research lambda for its own projects, well…

- Some projects use <u>multiple parallel waves</u> across a common path, or long resource-intensive transcontinental waves; other participants need to have L3 connections backhauled to the nearest L3 router node, etc.

- Add additional new Fednet/Int'l/Commercial participants…

- Before you know it, you're <u>out</u> of waves, at least at some locations, and you're just getting going.

"What About The Southern Route?"

- Whenever things look tight this way, folks always look at the redundant connectivity engineered into the system – in NLR's case, "What about the Southern Route?" I assert that it would be a really bad idea to book your backup capacity for production traffic. Gear fails. Backhoes eat fiber. Hurricanes flood POPs. Disgruntled employees burn down data centers. You really want redundant capacity to handle misfortunes.

- So, if my capacity analysis is correct, I believe NLR should either be looking at higher density WDM gear (to get more waves onto their existing glass), higher bandwidth interfaces (so they can avoid parallel 10 gig link scenarios) or if it is cheaper, they should be thinking about preparing to acquire and light additional fiber.

- Or you could redefine what's "network research" :-)

NLR *Is* Looking at Its Needs and Options…

- I'd encourage folks to read "Assessment of Optical Network System Technology and Services for National Lambda Rail," a report prepared by Dr. Kristin Rauschenbach, Technical Director of Optical Networking, BBN Technologies, March 17, 2006 (see: http://www.nlr.net/pubs/NLR-TechReport-BBN.pdf) and its discussion of NLR's options moving forward in terms of higher speed interfaces, higher density WDM gear, equipment replacement, etc.

NLR May Have Pricing Issues, Too

- I suspect NLR might run into pricing issues, too. It is really hard to get pricing right so that capacity get efficiently used.

- Too high? Capacity lies idle. No one uses the resource.

- Too low? Capacity gets allocated inefficiently and gobbled up prematurely (and in extreme cases, you don't generate enough revenue to purchase the next increment of capacity)

- NLR may have a tough price point to hit:
  -- assume NLR costs $100 million invested over 5 years
    to build, or $20 million/year
  -- ($20 M/yr) / 40 waves ==> $500K/wave/yr (asset value)
  -- **But you can get an Abilene 10Gig for less, $480K/year**

- Complications: $480K/year is ongoing; NLR investment probably has a life > 5 years; time value of money isn't considered; not all lambdas are in use; etc.

IX. Recent Next Generation Abilene Developments

Internet2's NewNet

- The April 2006 Internet2 Meeting in Arlington VA had a number of interesting disclosures which strongly impact the national optical networks environment. I would encourage you to listen to:
  -- Doug Van Houweling's and Larry Faulkner's session on "The Future of Internet2":
  http://events.internet2.edu/2006/spring-mm/ netcast-archive.cfm?session=2576
   -- Steve Cotter's Abilene Update:
  http://www.internet2.edu/presentations/spring06/ 20060425-abilene-cotter.pdf
  -- Rick Summerhill's HOPI Update session:
  http://events.internet2.edu/2006/spring-mm/ netcast-archive.cfm?session=2536

Another Question: "Will Internet2 Peer With NLR?"

- During the Member Meeting, the question was asked, "Will Internet2 peer with NLR?"

- This seemingly innocuous question raises some interesting issues, including:
  -- are the two networks effectively equal? (peering normally occurs only between similarly-sized networks with roughly symmetric traffic volumes)
  -- would NLR-Abilene peering competitively advantage (or competitively harm) either party?
  -- if NLR and Abilene peered, would it practically matter any more which network someone connected to?

- Let's start with that last question first.

Peering: You Get Customer Routes (<u>ONLY</u>)

"If NLR and Abilene peered, would it practically <u>matter</u> any more which network someone connected to?"

This is an interesting question because when you peer, you exchange customer routes, and ONLY customer routes, while much of the aggregate value of Abilene to I2 participants comes from the **OTHER** networks with which Abilene currently peers (e.g., International MOU partners, FedNet partners, state K12 educational networks connecting as SEGPs, corporate participants, etc.).

If "Abilene customer routes" were narrowly defined to be JUST the routes associated with I2's 206 core university members, universities connecting only to NLR would only see a fraction of the routes they'd see if they were connecting via Abilene, and as a result they'd likely sink and source far less traffic through Abilene. NLR could work to peer with the other networks, but that would take time…[76]

What If NLR Became An Abilene <u>Customer</u>?

- **In this scenario, NLR would effectively function as a "National Gigapop."** That, too, has some implications…

- **Financial:**

  -- If NLR were to become an Abilene customer, they'd probably need at least 3 10gig interconnection points with Abilene, each $480K/year. *Ideally*, NLR'd want to interconnect *all* 8 NLR PacketNet nodes with Abilene, for a total of $3,840,000/year plus one time costs. (And if NLR bought less than 8 interconnection points, where would those interconnection points be sited?)

  -- Virtually all NLR participants *already have* Abilene connectivity; the only scenario where having NLR buy connectivity from Abilene on behalf of its members would be a situations where at least "some" NLR customers do *NOT* also have an Abilene connection. Hmmm….

Problems With NLR Being An Abilene Customer (2)

- **Technical:**
  -- adds another layer/ASN, with increased traffic opacity, increased probability of asymmetry/misrouted traffic, increased latency due to limited interconnectivity, etc.
  -- depending on AUP/COU congruence, unacceptable NLR traffic (e.g., commercial traffic) might need to be specially tagged/handled for some Abilene R&E only peers
  -- experimental traffic sourced from NLR has the potential to affect the stability/availability of Abilene (in the hypothetical 3x10gig interconnect case, those connections could potentially generate enough traffic to swamp both northern and southern routes across Abilene until Abilene does its next gen network deployment)
- **==>** I doubt that NLR would become an Abilene customer.

X. Conclusion

Lots to Think About Right Now….

- At the risk of stating the obvious: the national optical network scene is very fluid and complex right now.

- You should strive to let empirical application requirements drive your network connectivity choices. As far as I can tell, the current Abilene backbone meets all existing and immediately foreseeable application requirements, and should continue to do so until the new Abilene backbone comes online within 18 months.

- The new Internet2 backbone will provide both the future capabilities and the future capacity that the higher education community will eventually need, and it will also address the policy-related issues that made using Abilene awkward for some federal and commercial participants.

- Finally, it is worth recognizing that I2 is a lot more than just Abilene… I'd encourage you to seek out opportunities to participate and get involved with <u>your</u> Internet2.