Optical Networking in the Internet of the Future

Stanford University Networking Seminar

Joe Berthold

3 May 2007
Outline

→ Scaling the current Internet – challenges and opportunities
  → Traffic and its’ growth in the metro and core network
  → Implications of linearly scaling current design techniques
  → Approaches to cost-effective capacity scaling

→ Networking services outside of IP
  → Optical networking technology deployment today
  → High bandwidth services
  → Optical networking research
Scaling Internet Access - Issues
Metro example based on Internet, IPTV services
Broadcast Video – Traffic Load

→ Standard definition TV
  → 4 Mb/s per channel
  → For 200 channels – 800 Mb/s

→ High definition TV
  → 8 Mb/s per channel
  → For 100 channels – 800 Mb/s

→ Only 1 or 2 GE circuits need to be replicated across the region, possibly using multicast at the IP, Ethernet, or optical layers. This is a very modest traffic load
Typical Hierarchical Metro Office Architecture

Example of a region with 2 million users

10 Local Serving Office : 1 Metro Hub
10 Metro Hub : 1 Core Gateway

Local serving office
Metro Hub office
Core Gateway or Super Hub office
Future Traffic Estimate – Broadcast video, Internet, Voice

Assumptions:
- Video traffic is dominated by broadcast.
- Internet traffic engineering is 1 Mb/s per subscriber.
- Voice traffic evolves to VOIP
  - 20% simultaneous usage and half remains in Metro.
Personalized Video – Traffic Load
Unique data flow per viewer with interactive control

→ Presumes video is served from the Core gateway or Internet
→ Local service office with 20,000 subscribers
  → Upper bound – 20,000 X 3 channels X 8 Mb/s = ~ 500 Gb/s
  → Switch capacity that would be required at the local office

→ More reasonable assumption would be 10% simultaneous use, 50 Gbit/sec
→ At Metro Hub traffic is 10X larger, at Core Gateway 100X larger
Traffic Upper Bound – Video becomes personalized

Assumption: Each subscriber provisioned for 25 or 2.5 Mb/s traffic capacity from Core Gateway. Core Gateway is where Video Hub and Internet Gateway are located. Traffic estimates are independent of whether video comes from the Internet or local servers.
IPTV Metro Network Architecture
Ethernet and IP layer connectivity including resiliency

Primary Path Interfaces
- 4 @ 10G IP router
- 2 @ 10G Ethernet switch
- 2 @ 1G Ethernet switch

Alternate Path Interfaces
- 8 @ 10G IP router
- 2 @ 10G Ethernet switch
- 2 @ 1G Ethernet switch

Access Aggregation

Metro Core
## IP and Ethernet Equipment Costs of IPTV Architecture

IP and Ethernet layer switching only; excludes WDM transport, access equipment

Each port must have bandwidth reserved for each service instance

<table>
<thead>
<tr>
<th>Interface Cost Assumption A</th>
<th>Unit Cost ($)</th>
<th>Number</th>
<th>Total</th>
<th>$ per Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 G IP Router Port</td>
<td>100K</td>
<td>12</td>
<td>1200K</td>
<td>120</td>
</tr>
<tr>
<td>10 G Ethernet Switch Port</td>
<td>20K</td>
<td>4</td>
<td>80K</td>
<td>8</td>
</tr>
<tr>
<td>GE switch Port</td>
<td>2K</td>
<td>4</td>
<td>8K</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total Cost per engineered Mb/s</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>136</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interface Cost Assumption B</th>
<th>Unit Cost ($)</th>
<th>Number</th>
<th>Total</th>
<th>$ per Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 G IP Router Port</td>
<td>50K</td>
<td>12</td>
<td>600K</td>
<td>60</td>
</tr>
<tr>
<td>10 G Ethernet Switch Port</td>
<td>10K</td>
<td>4</td>
<td>40K</td>
<td>4</td>
</tr>
<tr>
<td>GE switch Port</td>
<td>1K</td>
<td>4</td>
<td>4K</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total Cost per engineered Mb/s</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>68</strong></td>
</tr>
</tbody>
</table>

*IP and Ethernet layer switching equipment cost per Mb of engineered bandwidth, assuming 100% utilization*
Additional Data Points

→ Presentations from IEEE HSSG meetings

→ Core Network - Level3 OIF presentation - April 2007
  → NYC to Washington DC link today – 280 Gb/s
  → Long term growth observation ~ 70% per year
  → Projecting 280 Gb/s at CAGR 70%
    → 5 years – 4 Tb/s; 10 years – 56 Tb/s

→ Optical channel rates
  → Just deploying 40G today
  → Growing to 100G in several years

→ Architecture
  → We can’t afford to grow to capacities anywhere near those above with our current internet backbone architecture
What needs to change?

→ Scaling core routers by 2 orders of magnitude isn’t feasible
  → Cost, space and power per unit bandwidth are high
  → Scaling with capacity is linear, or perhaps super-linear

→ We need to move to a new core internet architecture
  → Dynamic router interconnect
  → Scalable, dynamic, adaptive optical core network
Scaling Internet Access - Opportunities
Optimal application of multilayer architecture to lower cost
Why a Multilayer Architecture?

- Although applications are moving to IP, carrier service offerings are needed at multiple layers.
- Large business customers want control, security.
- Layers have different costs and capabilities.

### Advantages

- Express Cost
- Interface Cost
- DWDM
- TDM

### Relative Switching Costs

<table>
<thead>
<tr>
<th></th>
<th>IP/MPLS</th>
<th>TDM</th>
<th>DWDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface Cost</td>
<td>10</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Express Cost</td>
<td>20</td>
<td>2</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Cost Relationships for 10G Technology

- Core Router – SONET interface $1.00
- Ethernet switch – 10G LAN PHY (short reach) $0.10
- Optical Switch – OEO – 10G SR interface $0.10
- DWDM Transceiver (Ultra-long haul) $0.10
- ROADM/WSS (Express path) per 10G channel $0.02
- Amplifier (Ultra-long haul) per 10G channel $0.005
Scalable Metro Network Architecture

Local Office A

Legacy SDH

OTN

Ethernet

Video Hub and IP Gateway Office

Video Servers

Ethernet

Metro Hub A

VOIP
Private IP
Public IP
TDM

IP
IP
IP
ADM

OTN

Metro Hub B

IP
IP
IP
ADM

Access Aggregation

Metro Core
Scalable Metro Network Architecture

Video distribution example – as traffic grows, cost/bit decreases with optical bypass

Traffic Scenarios

- Shared wave end-to-end
- Dedicated wave to local office
- Dedicated wave to access node
Observations

The previous example assumed a centralized server architecture

→ At low traffic levels several stages of packet switching provide an efficient distribution mechanism

→ As utilization grows, intermediate switching nodes can be replaced by bulk forwarding mechanisms at selected nodes

→ At highest utilization direct lightpaths are justified from server to final distribution nodes

Peer-to-peer distribution mechanism would drive the architecture in a different direction

→ Fewer centralized servers as users provide storage

→ Continued growth of router capacity
Optical Switching in Today’s Networks
Expanding beyond transport infrastructure to a services platform
Intelligent Optical Networks - Commercial

Intelligent Optical Network Element

Transport link

connection provisioned
Dynamic Optical Services are here today!

Dynamic Optical Services are here today!

Optical Mesh Service. Delivered.

Configure your Network for your Business

Service Description
- Adaptive Bandwidth with Zero-touch provisioning and near real-time setup of connections
- Application Agnostic, Layer 1 Transport Networking
- Carrier Grade Network
  - Availability & Restorability
  - Mesh Architecture with re-routing capabilities
  - Network and Equipment maintenance

Customer Applications
- Key industries: Financial, Media & Entertainment
- Corporate Utility Backbone Networks
- Management and distribution of real-time content (e.g., Video)

Customer Benefits
- Flexibility: bandwidth where it’s needed, when it’s needed
- Speed: “zero touch” provisioning
- Control: configure your network for your business

Internet2 - Intelligent Optical Switching Introduction

http://www.internet2.edu/files/Internet2-New-Network-Tech-v0.9.pdf
**Infiniband in the WAN**

- Infiniband is an effective data transport protocol for storage networks (few meters).
- TCP is not easily extended or not optimal for such data transfers.

**Question:** Is Infiniband effective over wide-area? - Yes

**Diagram:**
- ORNL to Sunnyvale, 700 miles
- ORNL to Chicago, 3300 miles
- ORNL to Seattle, 4300 miles

**Infiniband Over SONET – Joint with ORNL/NLGF**
- **collaborators:** Makia Minich, Steven Carter

- **Linux host**
- **longbow IB/S**

- **ORNL**
- **Chicago CDCL**
- **Seattle CDCL**
- **Sunnyvale CDCL**

**Link Rates:**
- **ORNL loop - 0.2 mile:** 7.5 Gbps
- **ORNL - Chicago loop – 1400 miles:** 7.46 Gbps
- **ORNL - Chicago - Seattle loop – 6600 miles:** 7.23 Gbps
- **ORNL – Chicago - Seattle - Sunnyvale loop – 8600 miles:** 7.20 Gbps

Applications and Optical Networks: How to Bridge the Performance Gap?
Nagi Rao, Qishi Wu, Steven Carter, Bill Wing, Oak Ridge National Laboratory - OFC 2007
DARPA CORONET Program
Submissions Nov 17, 2006 – No awards announced yet

→ The Defense Advanced Research Projects Agency's Strategic Technology Office (STO)

→ Dynamic Multi-Terabit Core Optical Networks: Architecture, Protocols, Control and Management


→ Objective

→ “… to develop the architecture, protocols, and control and management software for highly dynamic, multi-terabit global core optical networks with greatly enhanced performance, survivability and security.”

→ Ultimate Goal

→ “… to transition the CORONET technology, at the end of Phase 2, to commercial telecommunications carriers.
## CORONET Network Context

### High-Level CORONET Program Goals

<table>
<thead>
<tr>
<th>Network Requirements</th>
<th>Today’s State-of-the-Art Networks</th>
<th>High-Level CORONET Program Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Fiber Capacity</td>
<td>Up to 1.6 Tb/s</td>
<td>Up to 16 Tb/s</td>
</tr>
<tr>
<td>Bit Rate per Wavelength</td>
<td>10 or 40 Gb/s</td>
<td>40 or 100 Gb/s, or more</td>
</tr>
<tr>
<td>Maximum Bit Rate per Stream</td>
<td>40 Gb/s</td>
<td>Up to 1 Tb/s</td>
</tr>
<tr>
<td>Aggregate Network Demand</td>
<td>Up to 10 Tb/s</td>
<td>Up to 100 Tb/s</td>
</tr>
<tr>
<td>End-to-End Network Services</td>
<td>IP, SONET</td>
<td>IP (75%±) and λ-Services (25%±)</td>
</tr>
<tr>
<td>Optical-Layer Multicasting</td>
<td>Typically Not Possible</td>
<td>Basic Requirement</td>
</tr>
<tr>
<td>Performance Monitoring</td>
<td>Mostly in Electrical Layer</td>
<td>In Electrical and Optical Layers</td>
</tr>
<tr>
<td>Optical-Layer Configurability</td>
<td>Slow, Often Manual</td>
<td>Fast, Fully Automatic</td>
</tr>
<tr>
<td>Max Speed of Service Setup</td>
<td>Hours to Weeks</td>
<td>≤ 100 msec (CONUS)</td>
</tr>
<tr>
<td>Speed of Dedicated Protection</td>
<td>≤ 50 msec</td>
<td>≤ 250 msec (Global)</td>
</tr>
<tr>
<td>Speed of Shared Restoration*</td>
<td>50 to 100’s msec (Ring)</td>
<td>≤ 100 msec (CONUS)</td>
</tr>
<tr>
<td></td>
<td>Sec’s to Min’s (Mesh)</td>
<td>≤ 250 msec (Global)</td>
</tr>
<tr>
<td><em>From How Many Failures:</em></td>
<td>Typically, One Failure</td>
<td>Up to Three Failures</td>
</tr>
<tr>
<td>IP Services with Differentiated</td>
<td>Very Limited</td>
<td>Basic Requirement</td>
</tr>
<tr>
<td>End-to-End QoS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Next generation GIG-BE network
- 100 Network Nodes Globally
- IPoWDM, largely supporting IP/MPLS (~75% of traffic)
- Dynamically switched λ services (~25% of traffic)
- Optical layer mesh restoration
- Aggressive processing speeds
- Aggressive BW utilization
What does this mean for the future Internet?

→ Scaling the current Internet – Metro
  → Concentrate IP routing equipment at customer edge and service edge, using Ethernet, OTN and photonic technologies in aggregation/distribution

→ Scaling the current Internet – Core
  → Introduce dynamic router interconnect topologies to optimize use of expensive resources
  → Create express paths across the core with Ethernet, OTN and photonic technologies

→ New services for the future Internet
  → Dynamic Ethernet lightpaths
  → Fibre channel, Infiniband across the WAN
Thank You