DWDM RAM

Data@LIGHTspeed

An Architecture for Data Intensive Service Enabled by Next Generation Optical Networks

Tal Lavian : Nortel Networks Labs

Nortel Networks

International Center for Advanced Internet Research (iCAIR), NWU, Chicago

Santa Clara University, California

University of Technology, Sydney

DARPA NTONC NORTEL NETWORKS iCAIR NSF

Agenda

- Challenges
  - Growth of Data-Intensive Applications
- Architecture
  - Lambda Data Grid
- Lambda Scheduling
- Result
  - Demos and Experiment
- Summary

Radical mismatch: L1 – L3

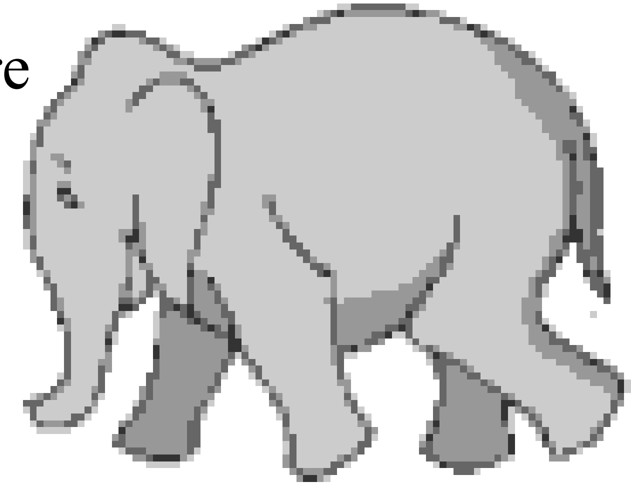
- Radical mismatch between the optical transmission world and the electrical forwarding/routing world.
- Currently, a single strand of optical fiber can transmit more bandwidth than the entire Internet core

- Current L3 architecture can't effectively transmit PetaBytes or 100s of TeraBytes
- Current L1-L0 limitations: Manual allocation, takes 6-12 months - Static.
  - Static means: not dynamic, no end-point connection, no service architecture, no glue layers, no applications underlay routing

Growth of Data-Intensive Applications

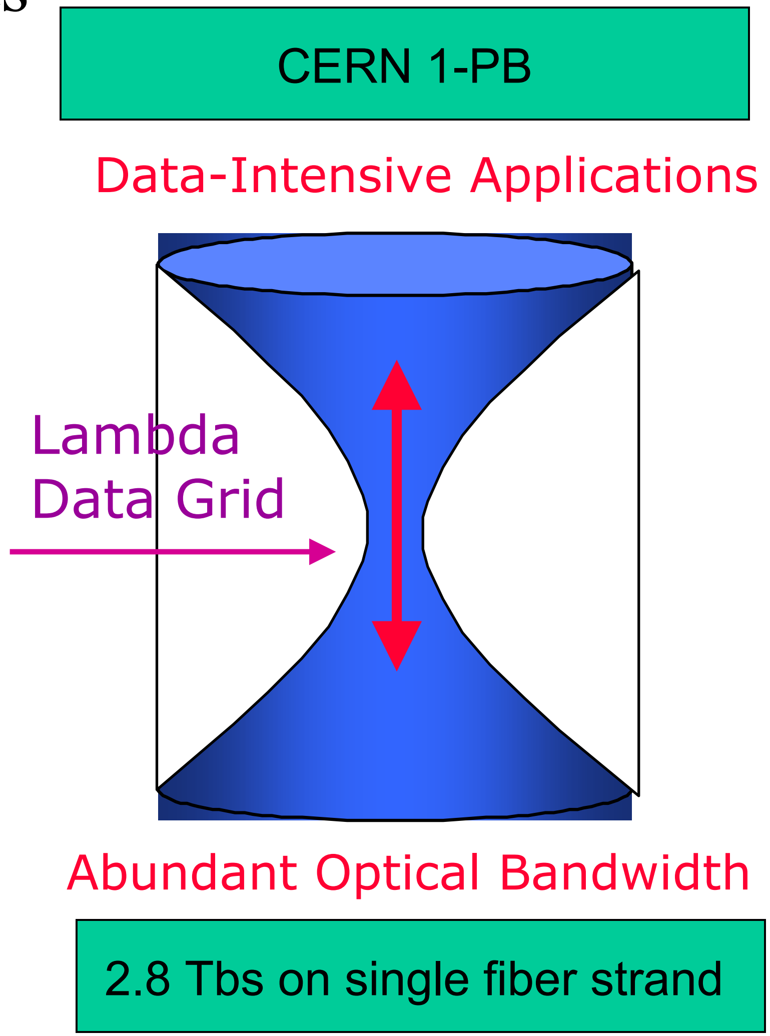
- **IP data transfer: 1.5TB (10^{12}) , 1.5KB packets**
  - Routing decisions: 1 Billion times (10^9)
  - Over every hop
- Web, Telnet, email – small files

- Fundamental limitations with data-intensive applications
  - multi TeraBytes or PetaBytes of data
  - Moving 10KB and 10GB (or 10TB) are different ($\times 10^6$, $\times 10^9$)
  - 1Mbs & 10Gbs are different ($\times 10^6$)



Lambda Hourglass

- Data Intensive app requirements
  - HEP
  - Astrophysics/Astronomy
  - Bioinformatics
  - Computational Chemistry
- Inexpensive disk
  - 1TB < $1,000
- DWDM
  - Abundant optical bandwidth
- One fiber strand
  - 280 λs, OC-192



DWDM
RAM
Data@LIGHTspeed

Challenge: Emerging data intensive applications require:

Extremely high performance, long term data flows
Scalability for data volume and global reach
Adjustability to unpredictable traffic behavior
Integration with multiple Grid resources

Response: DWDM-RAM - An architecture for data intensive Grids enabled by next generation dynamic optical networks, incorporating new methods for lightpath provisioning

DWDM RAM

Data@LIGHTspeed

DWDM-RAM: An architecture designed to meet the networking challenges of extremely large scale Grid applications.
Traditional network infrastructure cannot meet these demands, especially, requirements for intensive data flows

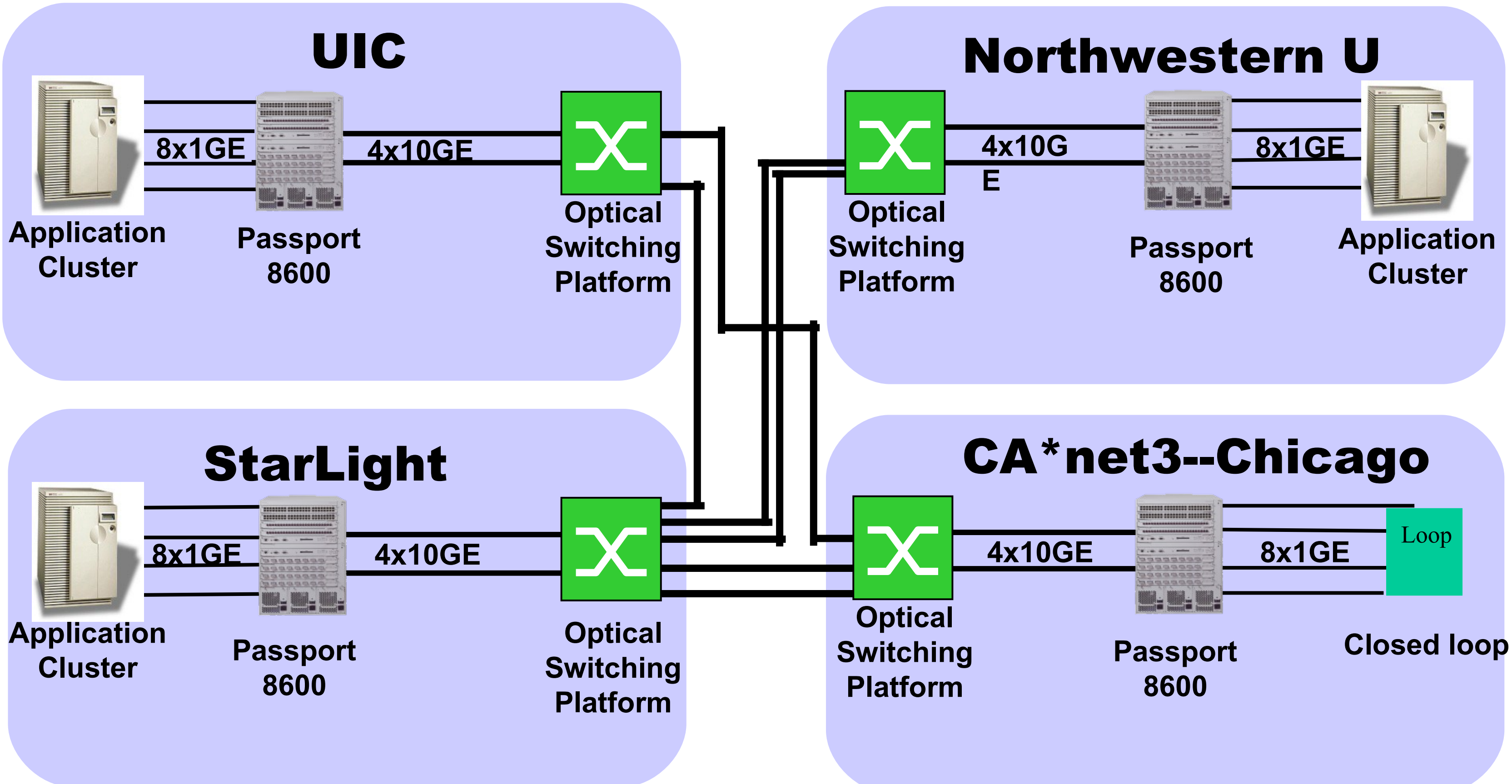
DWDM-RAM Components Include:

- Data management services
- Intelligent middleware
- Dynamic lightpath provisioning
- State-of-the-art photonic technologies
- Wide-area photonic testbed implementation

Agenda

- Challenges
  - Growth of Data-Intensive Applications
- Architecture
  - Lambda Data Grid
- Lambda Scheduling
- Result
  - Demos and Experiment
- Summary

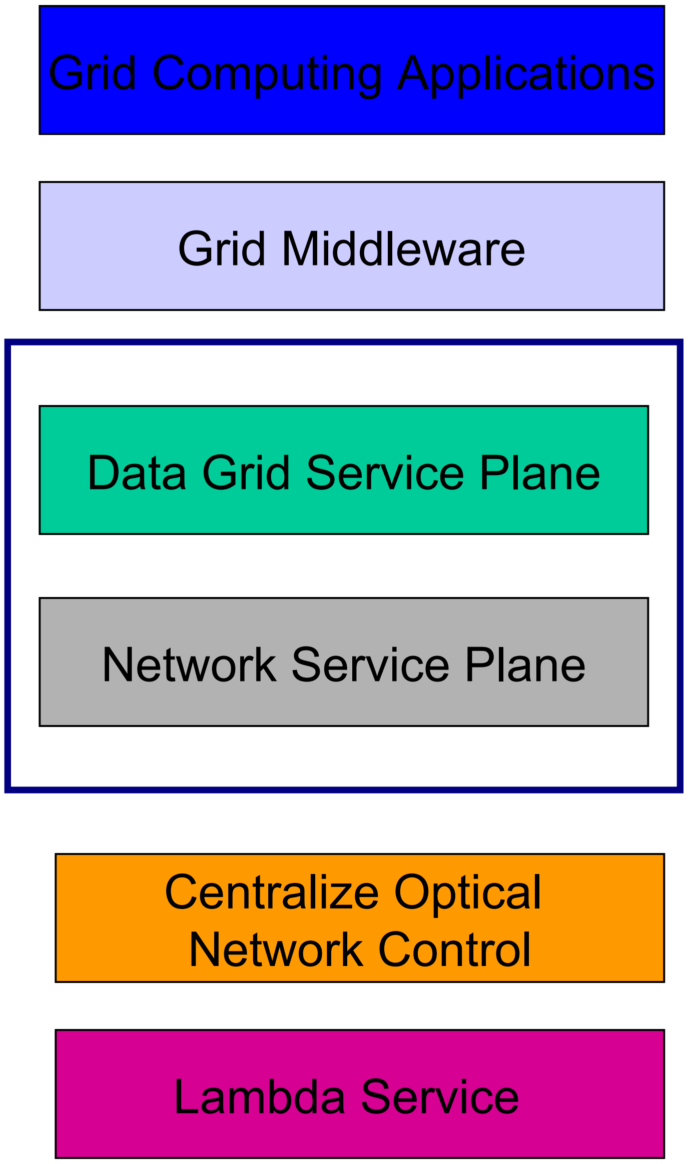
OMNInet Core Nodes



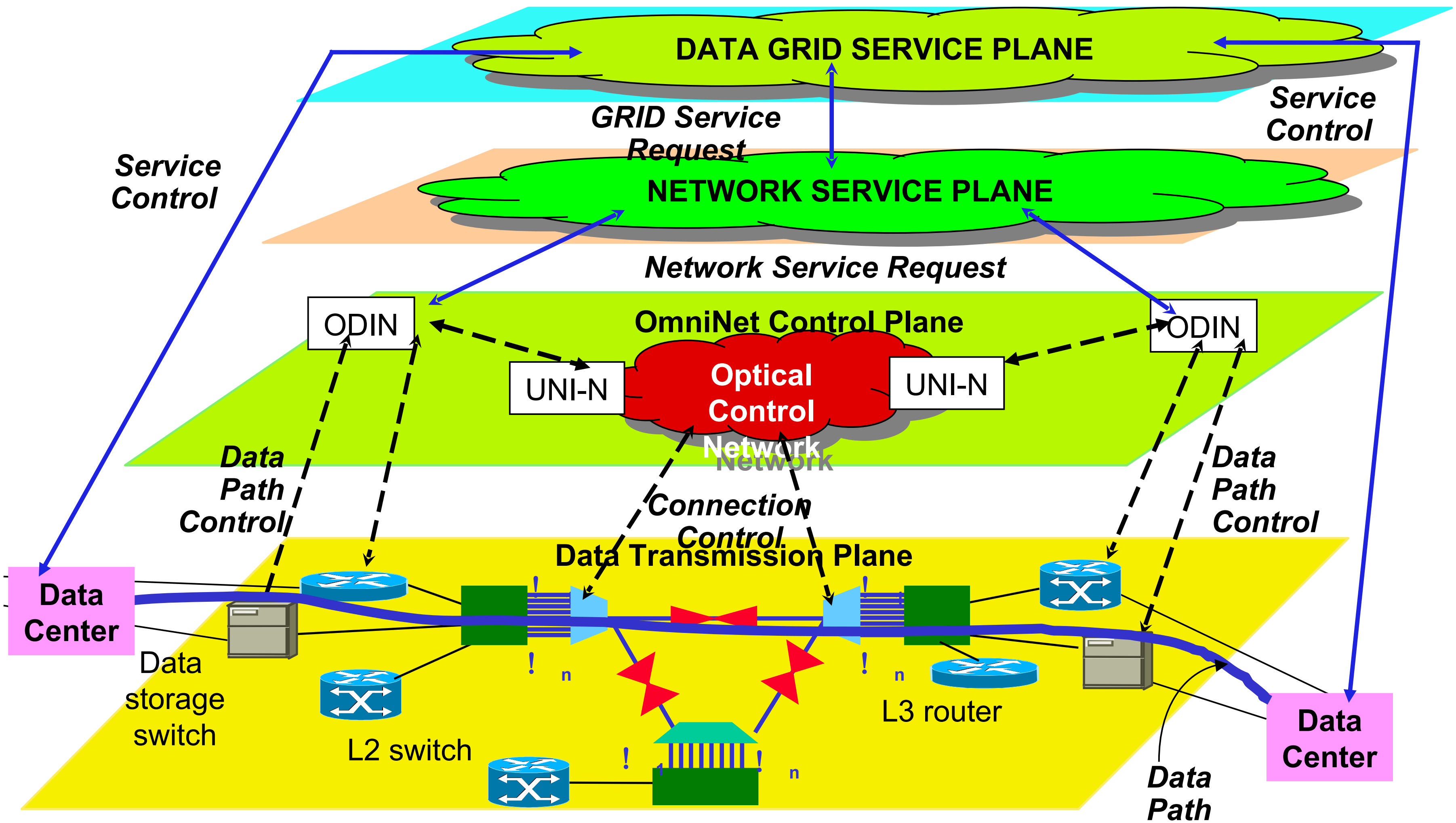
- A four-node multi-site optical metro testbed network in Chicago -- the first 10GE <u>service</u> trial!
- A test bed for all-optical switching and advanced high-speed services
- OMNInet testbed Partners: SBC, Nortel, iCAIR at Northwestern, EVL, CANARIE, ANL

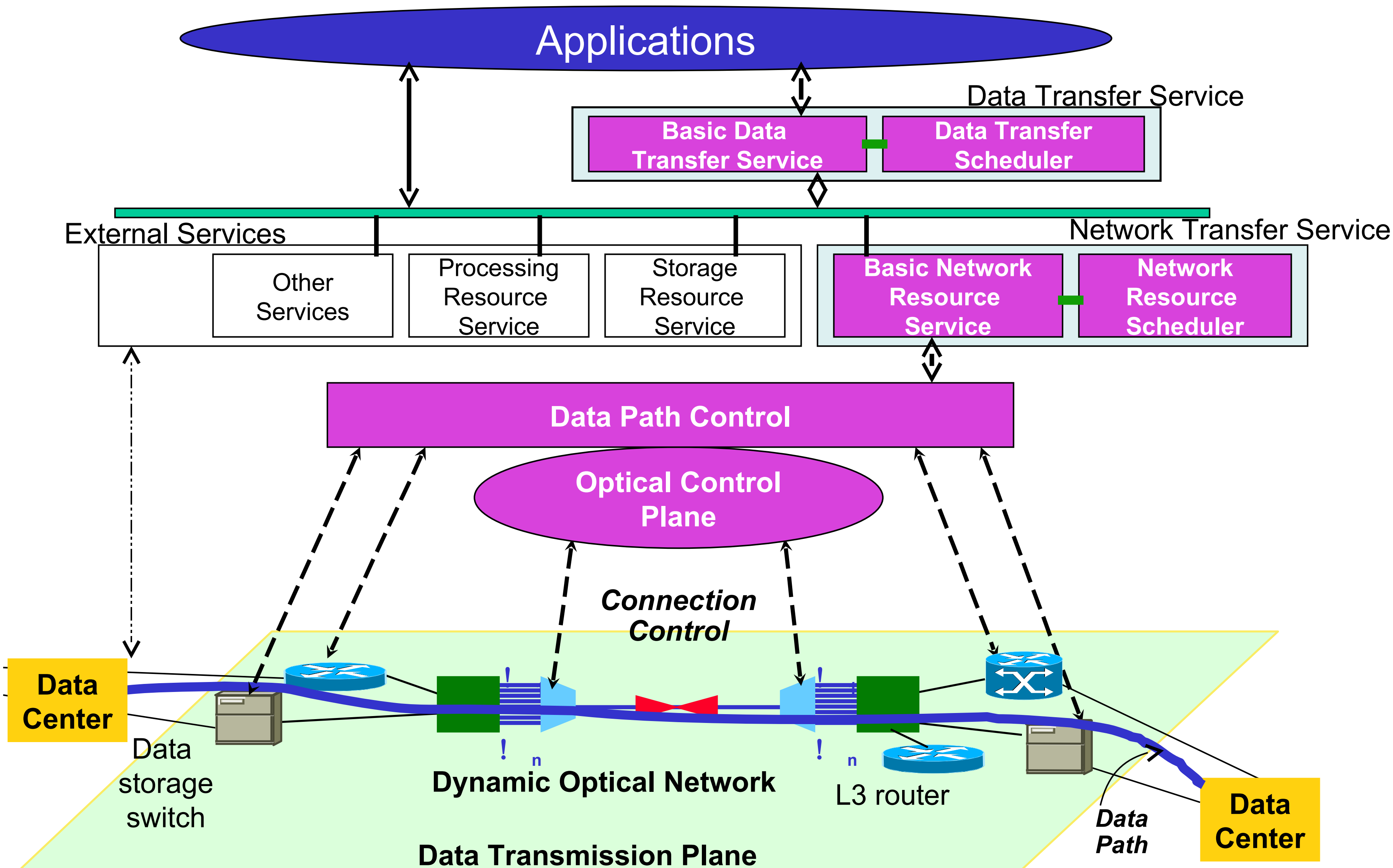
What is Lambda Data Grid?

- A service architecture
  - comply with OGSA
  - Lambda as an OGSI service
  - on-demand and scheduled Lambda
- GT3 implementation
- Demos in booth 1722



DWDM-RAM Service Control Architecture




Applications
Data Transfer Service
Basic Data Transfer Service
Data Transfer Scheduler
External Services
Network Transfer Service
Other Services
Processing Resource Service
Storage Resource Service
Basic Network Resource Service
Network Resource Scheduler
Data Path Control
Optical Control Plane
Connection Control
Data Center
Data storage switch
Dynamic Optical Network
L3 router
Data Path
Data Center
Data Transmission Plane

Data Management Services

- OGSA/OGSI compliant
- Capable of receiving and understanding application requests
- Has complete knowledge of network resources
- Transmits signals to intelligent middleware
- Understands communications from Grid infrastructure
- Adjusts to changing requirements
- Understands edge resources
- On-demand or scheduled processing
- Supports various models for scheduling, priority setting, event synchronization

Intelligent Middleware for Adaptive Optical Networking

- OGSA/OGSI compliant
- Integrated with Globus
- Receives requests from data services
- Knowledgeable about Grid resources
- Has complete understanding of dynamic lightpath provisioning
- Communicates to optical network services layer
- Can be integrated with GRAM for co-management
- Architecture is flexible and extensible

Dynamic Lightpath Provisioning Services

- Optical Dynamic Intelligent Networking (ODIN)
- OGSA/OGSI compliant
- Receives requests from middleware services
- Knowledgeable about optical network resources
- Provides dynamic lightpath provisioning
- Communicates to optical network protocol layer
- Precise wavelength control
- Intradomain as well as interdomain
- Contains mechanisms for extending lightpaths through
- E-Paths - electronic paths

Agenda

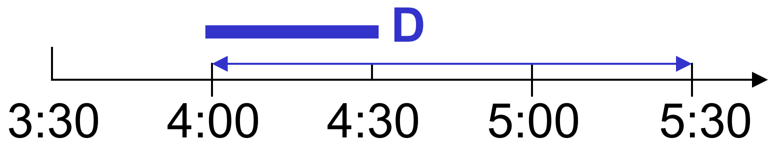
- Challenges
  - Growth of Data-Intensive Applications
- Architecture
  - Lambda Data Grid
- Lambda Scheduling
- Result
  - Demos and Experiment
- Summary

Design for Scheduling

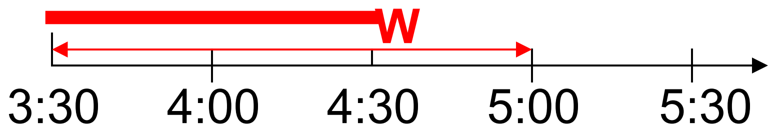
- Network and Data Transfers scheduled
  - Data Management schedule coordinates network, retrieval, and sourcing services (using their schedulers)
  - Network Management has own schedule
- Variety of request models
  - Fixed – at a specific time, for specific duration
  - Under-constrained – e.g. ASAP, or within a window
- Auto-rescheduling for optimization
  - Facilitated by under-constrained requests
  - Data Management reschedules
    - for its own requests
    - request of Network Management

Example 1: Time Shift

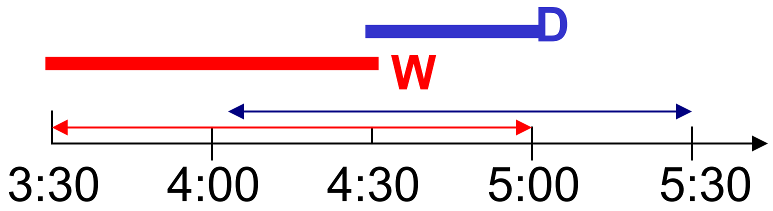
- Request for 1/2 hour between 4:00 and 5:30 on Segment D granted to User W at 4:00



- New request from User X for same segment for 1 hour between 3:30 and 5:00



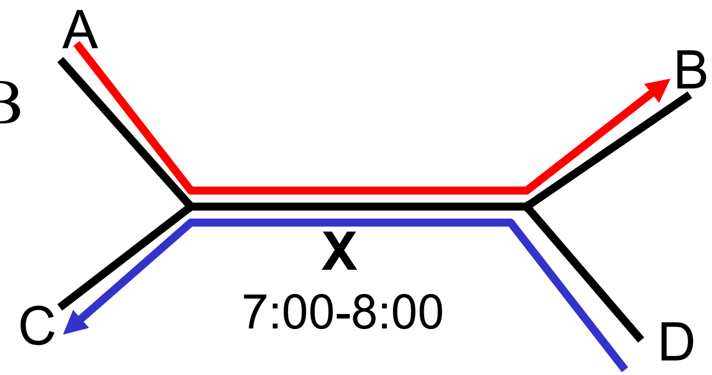
- Reschedule user W to 4:30; user X to 3:30. Everyone is happy.

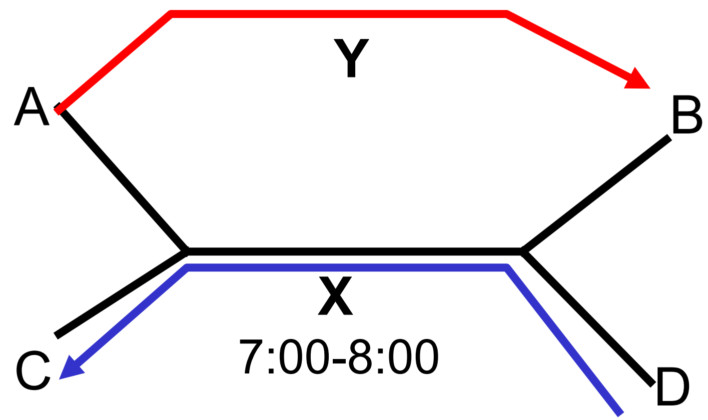


Route allocated for a time slot; new request comes in; 1st route can be rescheduled for a later slot within window to accommodate new request

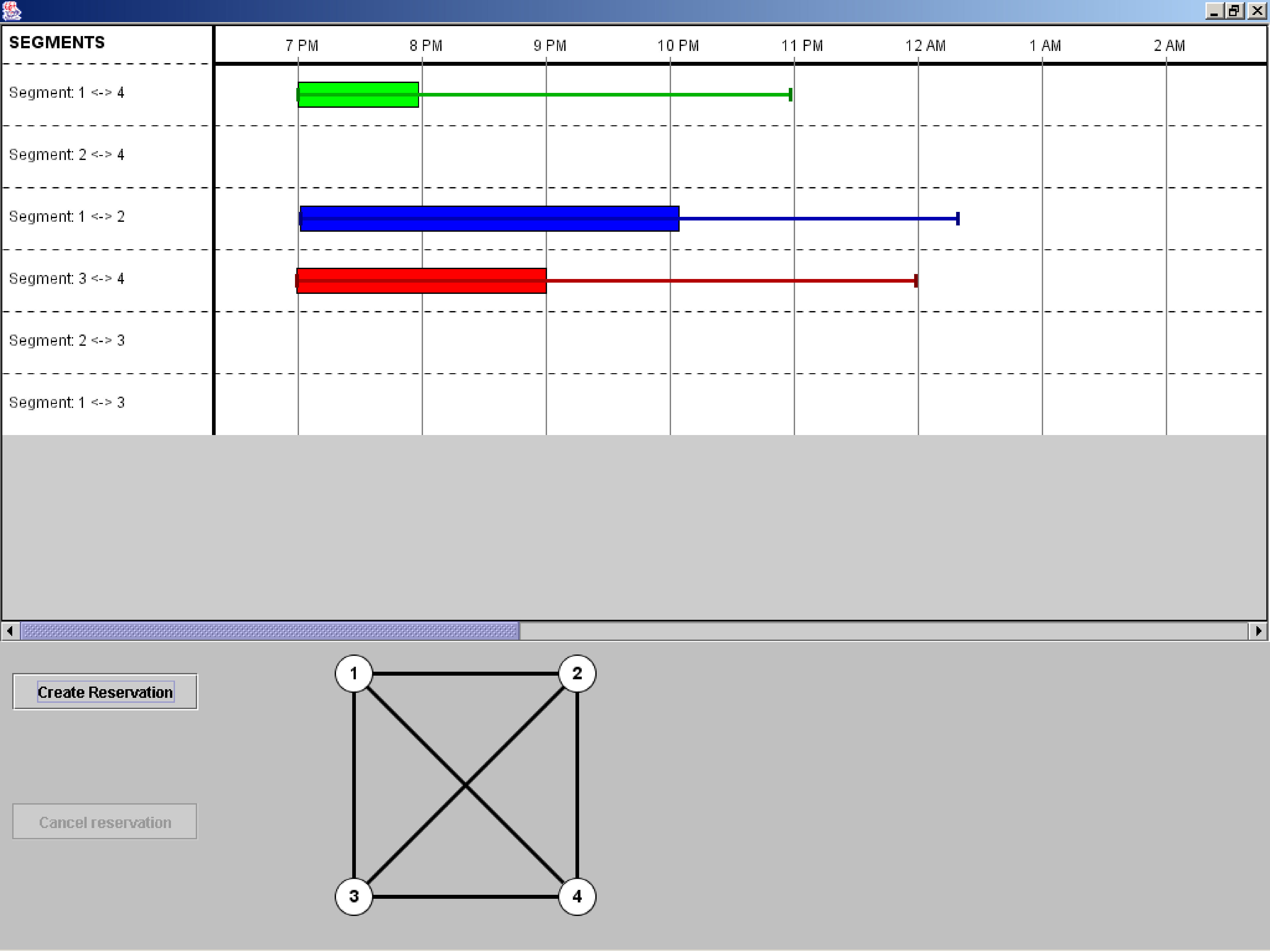
Example 2: Reroute

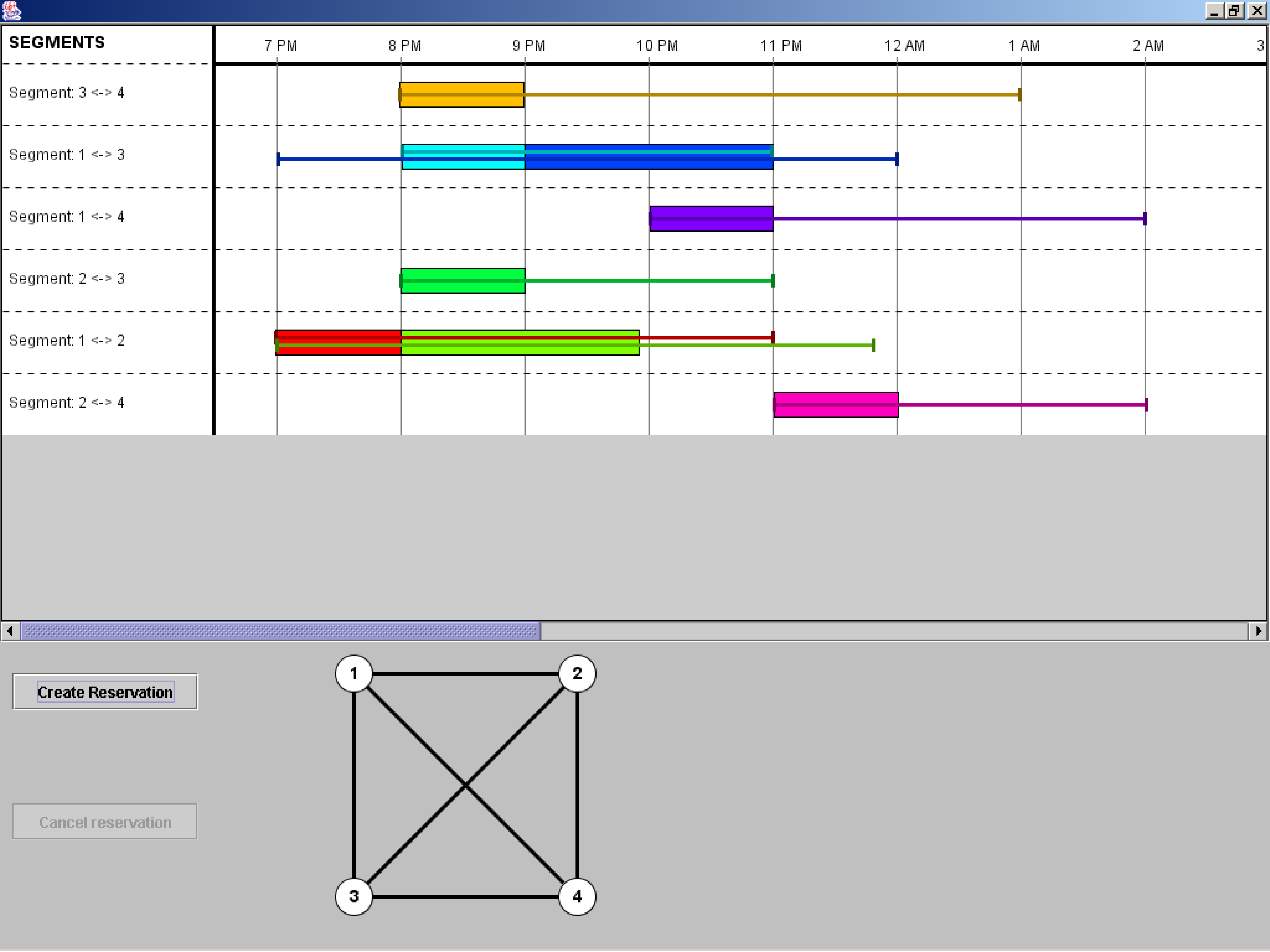
- Request for 1 hour between nodes A and B between 7:00 and 8:30 is granted using Segment X (and other segments) for 7:00
- New request for 2 hours between nodes C and D between 7:00 and 9:30 This route needs to use Segment E to be satisfied
- Reroute the first request to take another path thru the topology to free up Segment E for the 2nd request. Everyone is happy





Route allocated; new request comes in for a segment in use; 1st route can be altered to use different path to allow 2nd to also be serviced in its time window


SEGMENTS
7 PM
8 PM
9 PM
10 PM
11 PM
12 AM
1 AM
2 AM
Segment: 1 <-> 4
Segment: 2 <-> 4
Segment: 1 <-> 2
Segment: 3 <-> 4
Segment: 2 <-> 3
Segment: 1 <-> 3
Create Reservation
Cancel reservation
1
2
3
4

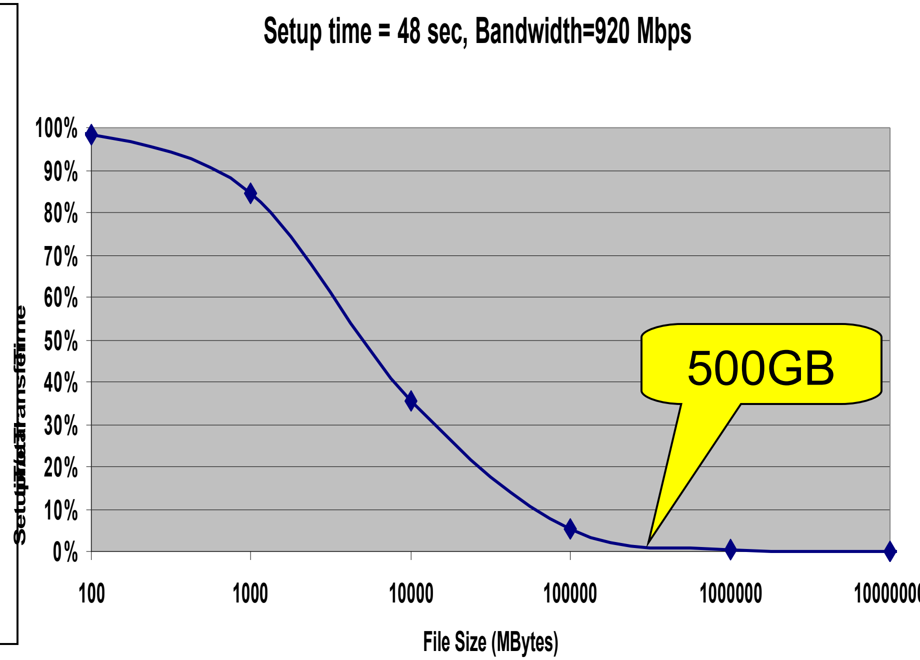

SEGMENTS
7 PM
8 PM
9 PM
10 PM
11 PM
12 AM
1 AM
2 AM
3
Segment: 3 <-> 4
Segment: 1 <-> 3
Segment: 1 <-> 4
Segment: 2 <-> 3
Segment: 1 <-> 2
Segment: 2 <-> 4
Create Reservation
Cancel reservation
1
2
3
4

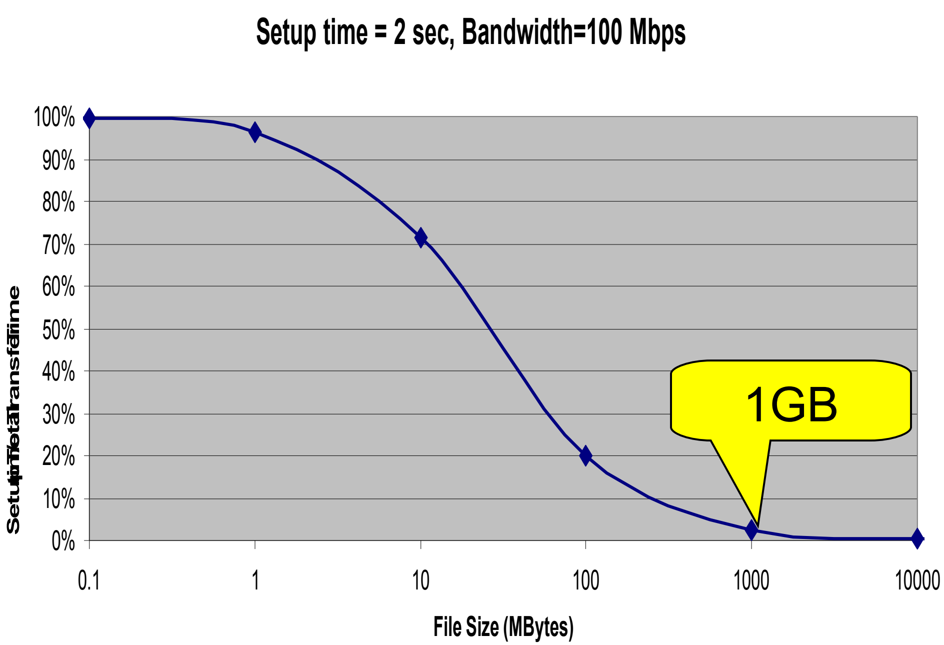
Agenda

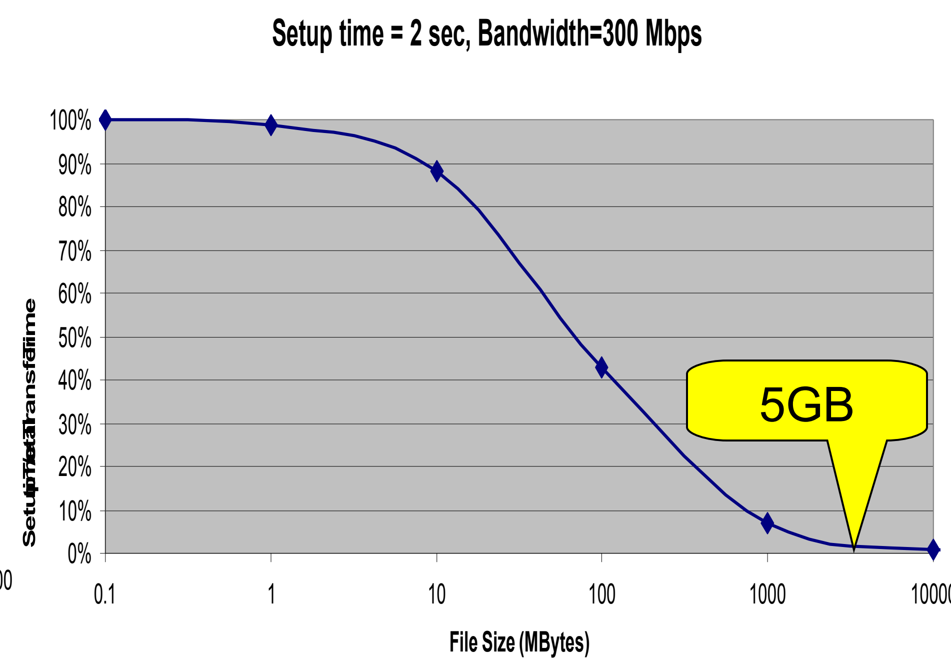
- Challenges
  - Growth of Data-Intensive Applications
- Architecture
  - Lambda Data Grid
- Lambda Scheduling
- Result
  - Demos and Experiment
- Summary

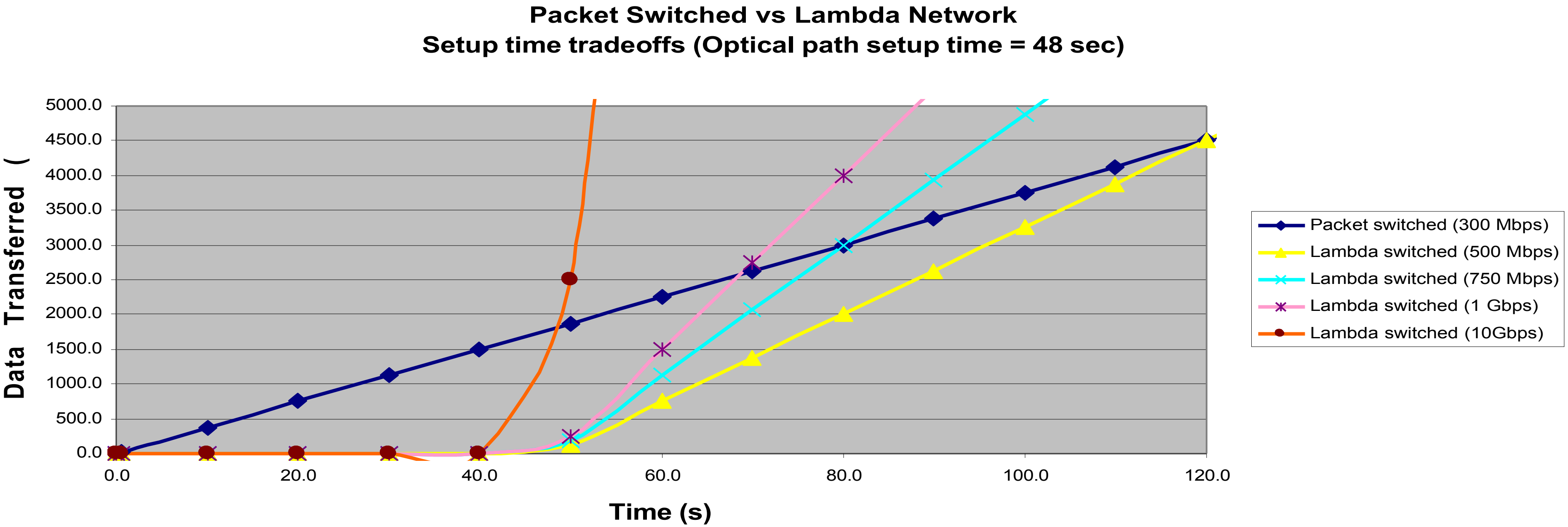
Path Allocation Overhead as a % of the Total Transfer Time

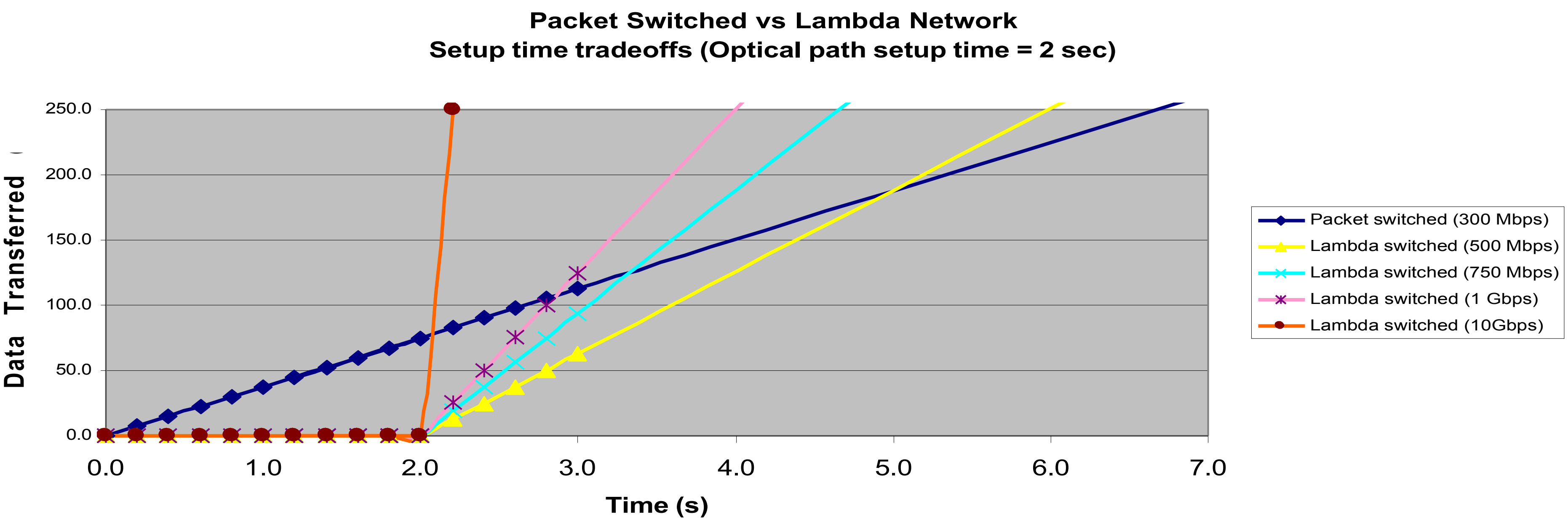
- **Knee point shows the file size for which overhead is insignificant**








Packet Switched vs Lambda Network
Setup time tradeoffs (Optical path setup time = 48 sec)
Data Transferred (
5000.0
4500.0
4000.0
3500.0
3000.0
2500.0
2000.0
1500.0
1000.0
500.0
0.0
0.0
20.0
40.0
60.0
80.0
100.0
120.0
Time (s)
Packet switched (300 Mbps)
Lambda switched (500 Mbps)
Lambda switched (750 Mbps)
Lambda switched (1 Gbps)
Lambda switched (10Gbps)


Packet Switched vs Lambda Network
Setup time tradeoffs (Optical path setup time = 2 sec)
Data Transferred
250.0
200.0
150.0
100.0
50.0
0.0
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
Time (s)
Packet switched (300 Mbps)
Lambda switched (500 Mbps)
Lambda switched (750 Mbps)
Lambda switched (1 Gbps)
Lambda switched (10Gbps)

Agenda

- Challenges
  - Growth of Data-Intensive Applications
- Architecture
  - Lambda Data Grid
- Lambda Scheduling
- Result
  - Demos and Experiment
- Summary

Summary

- Next generation optical networking provides significant new capabilities for Grid applications and services, especially for high performance data intensive processes

- DWDM-RAM architecture provides a framework for exploiting these new capabilities

- These conclusions are not only conceptual – they are being proven and demonstrated on OMNInet – a wide-area metro advanced photonic testbed

Thank you !

NRM OGSA Compliance

OGSI interface

GridService PortType with two application-oriented methods:

allocatePath(fromHost, toHost,...)
deallocatePath(allocationID)

Usable by a variety of Grid applications

Java-oriented SOAP implementation using the Globus Toolkit 3.0

Network Resource Manager

- Presents application-oriented OGSI / Web Services interfaces for network resource (lightpath) allocation
- Hides network details from applications
- Implemented in Java

Items in blue are planned

Scheduling : Extending Grid Services

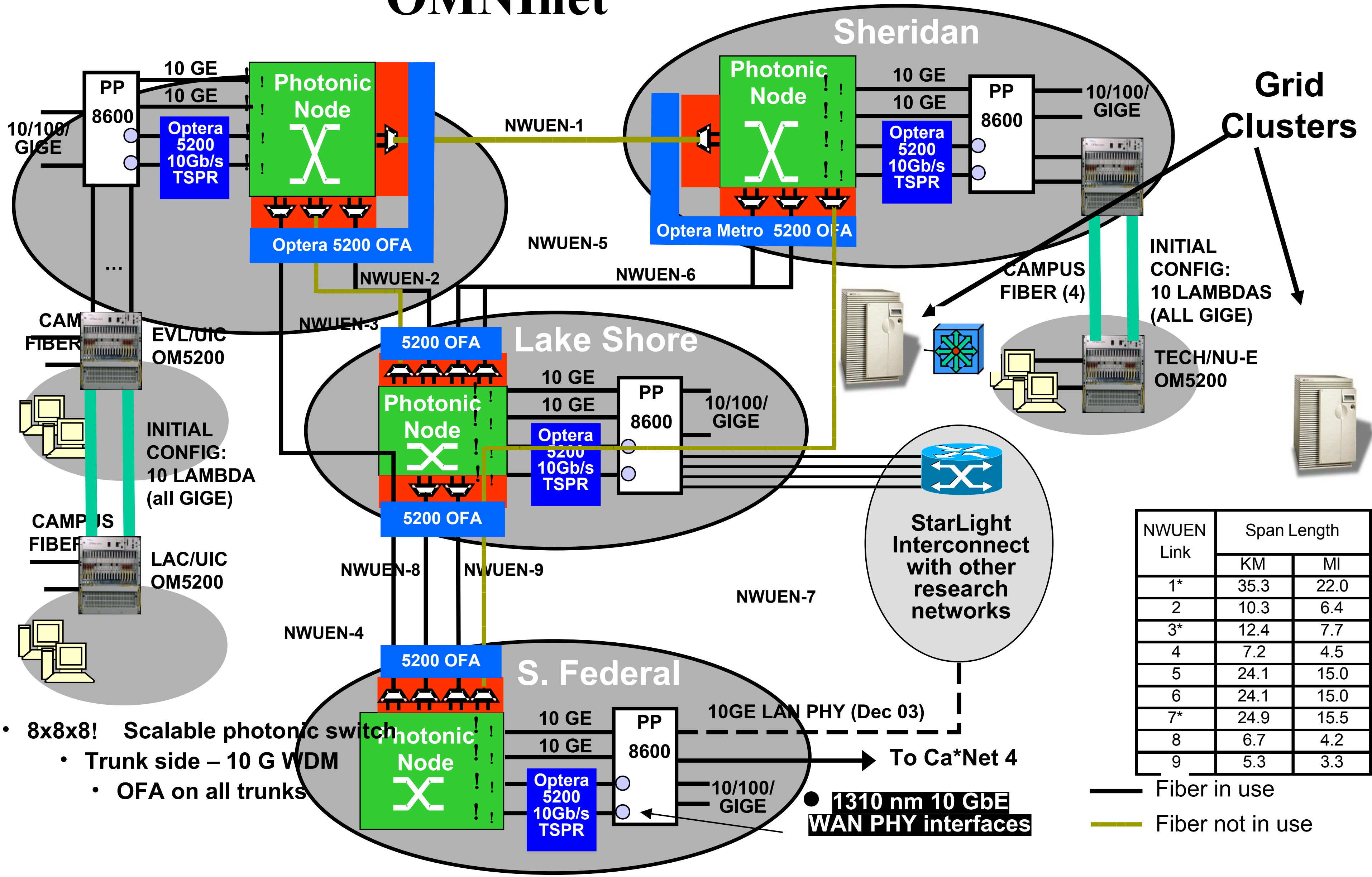
- OGSI interfaces
  - Web Service implemented using SOAP and JAX-RPC
  - Non-OGSI clients also supported
- GARA and GRAM extensions
  - Network scheduling is new dimension
  - Under-constrained (conditional) requests
  - Elective rescheduling/renegotiation
- Scheduled data resource reservation service ("Provide 2 TB storage between 14:00 and 18:00 tomorrow")

Lightpath Services

Enabling High Performance Support for
Data-Intensive Services With On-Demand Lightpaths Created By Dynamic Lambda Provisioning, Supported by Advanced Photonic Technologies

OGSA/OGSI Compliant Service
Optical Service Layer: Optical Dynamic Intelligent Network (ODIN) Services
Incorporates Specialized Signaling
Utilizes Provisioning Tool: IETF GMPLS
New Photonic Protocols

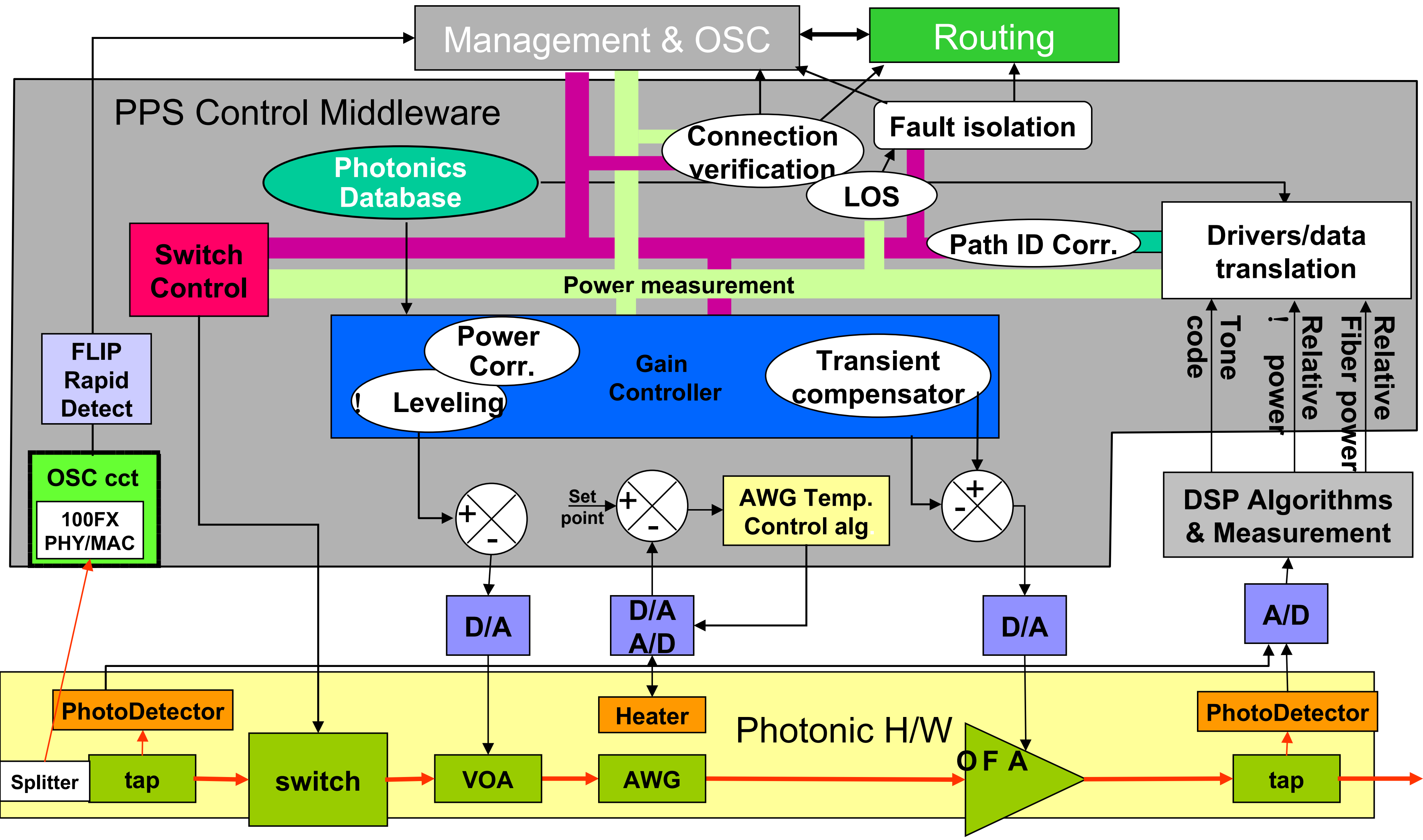
OMNInet



- 8x8x8! Scalable photonic switch
  - Trunk side – 10 G WDM
    - OFA on all trunks

NWUEN Link	Span Length	
	KM	MI
1*	35.3	22.0
2	10.3	6.4
3*	12.4	7.7
4	7.2	4.5
5	24.1	15.0
6	24.1	15.0
7*	24.9	15.5
8	6.7	4.2
9	5.3	3.3

Physical Layer Optical Monitoring and Adjustment



Summary (I)

- Allow applications/services
  - to be deployed over the Lambda Data Grid
- Expand OGSA
  - for integration with optical network
- Extend OGSI
  - interface with optical control
  - infrastructure and mechanisms
- Extend GRAM and GARA
  - to provide framework for network resources optimization
- Provide generalized framework for multi-party data scheduling

Summary (II)

- Treating the network as a Grid resource
- Circuit switching paradigm moving large amounts of data over the optical network, quickly and efficiently
- Demonstration of on-demand and advance scheduling use of the optical network
- Demonstration of under-constrained scheduling requests
- The optical network as a shared resource
  - may be temporarily dedicated to serving individual tasks
  - high overall throughput, utilization, and service ratio.
- Potential applications include
  - support of E-Science, massive off-site backups, disaster recovery, commercial data replication (security, data mining, etc.)

Extension of Under-Constrained Concepts

- Initially, we use simple time windows
- More complex extensions
  - any time after 7:30
  - within 3 hours after Event B happens
  - cost function (time)
  - numerical priorities for job requests

Extend (eventually) concept of under- constrained to user-specified utility functions for costing, priorities, callbacks to request scheduled jobs to be rerouted/rescheduled (client can say yea or nay)