# DWDM-RAM:

## DARPA-Sponsored Research for Data Intensive Service-on-Demand Advanced Optical Networks

DWDM-RAM **demonstration** sponsored by
Nortel Networks and iCAIR/Northwestern University

| | |
|---|---|
| Dates & Times: | Monday Oct 6 at 4pm & 6pm |
| | Tuesday Oct 7 at 12Noon, 2pm & 4pm |
| | Wednesday Oct 8 at 10am & 12Noon |

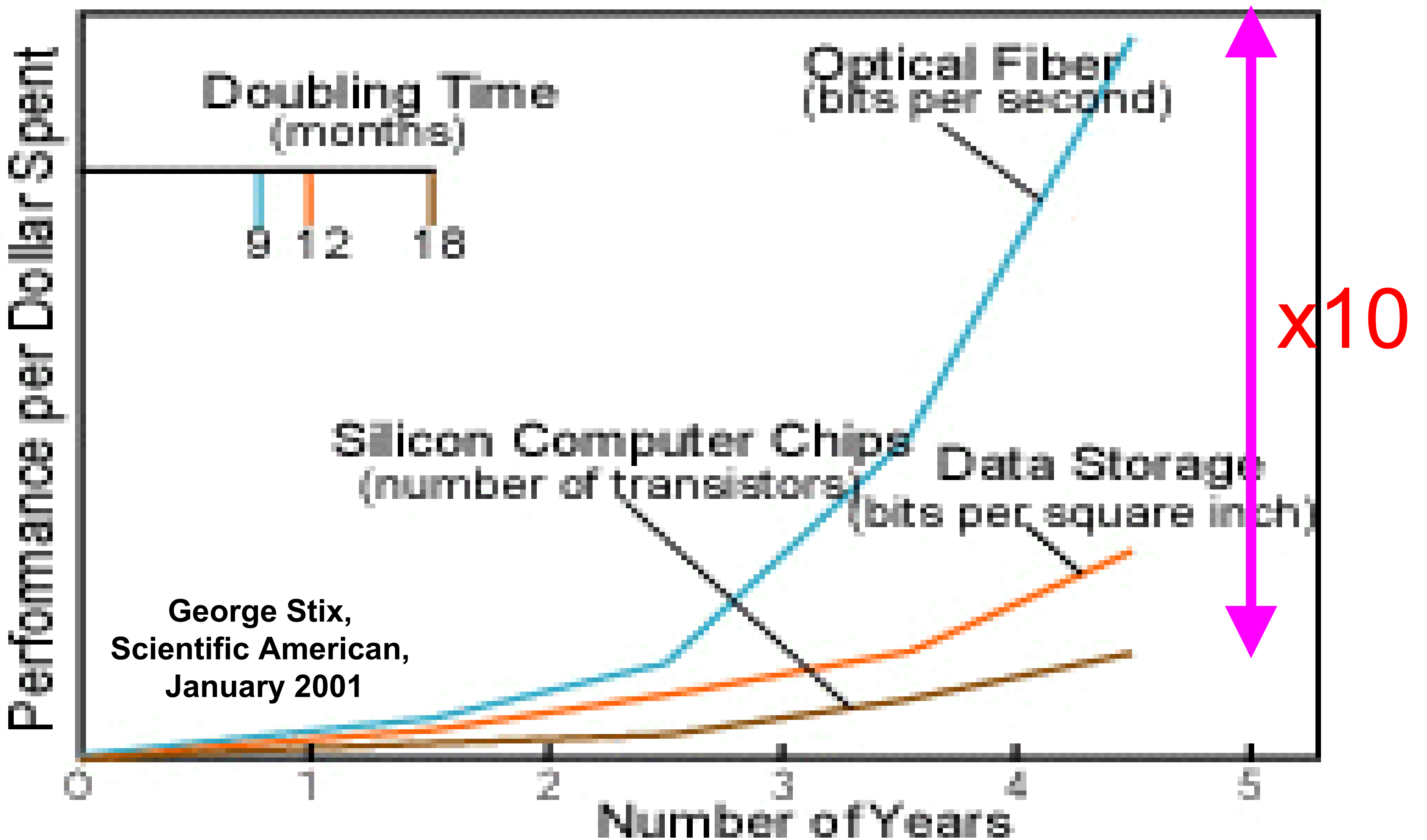
# SLAC Agenda

- DWDM-RAM Overview
  - The Problem
  - Our Architecture & Approach
- Discussion with HEP Community

# Problem: More Data Than Network

- Application-level network scheduling
  - Application must see dedicated bandwidth as a managed resource
  - Advance scheduling of network from application
  - Optimization is important
    - Rescheduling with under-constrained requests
- Data transfers require service model
  - Scheduled network and host data services combined
  - Co-reservation of storage, data, and network
  - Requires scheduling

# Optical Networks Change the Current Pyramid




DWDM- fundamental misbalance between computation and communication

# Radical mismatch: L1 – L3

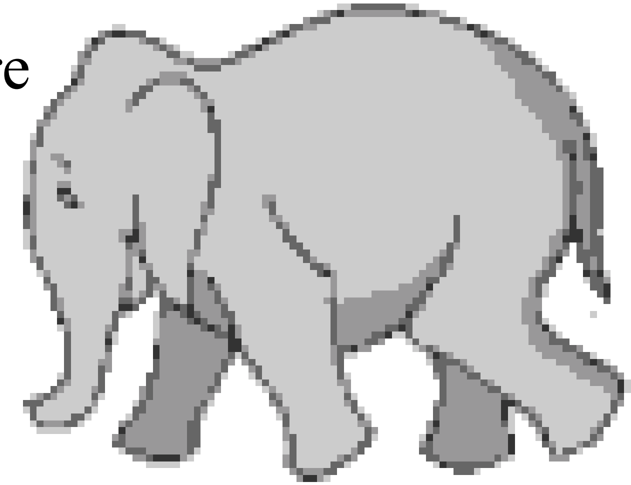
- Radical mismatch between the optical transmission world and the electrical forwarding/routing world.
- Currently, a single strand of optical fiber can transmit more bandwidth than the entire Internet core

- 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 (x$10^6$, x$10^9$)
  - 1Mbs & 10Gbs are different (x$10^6$)

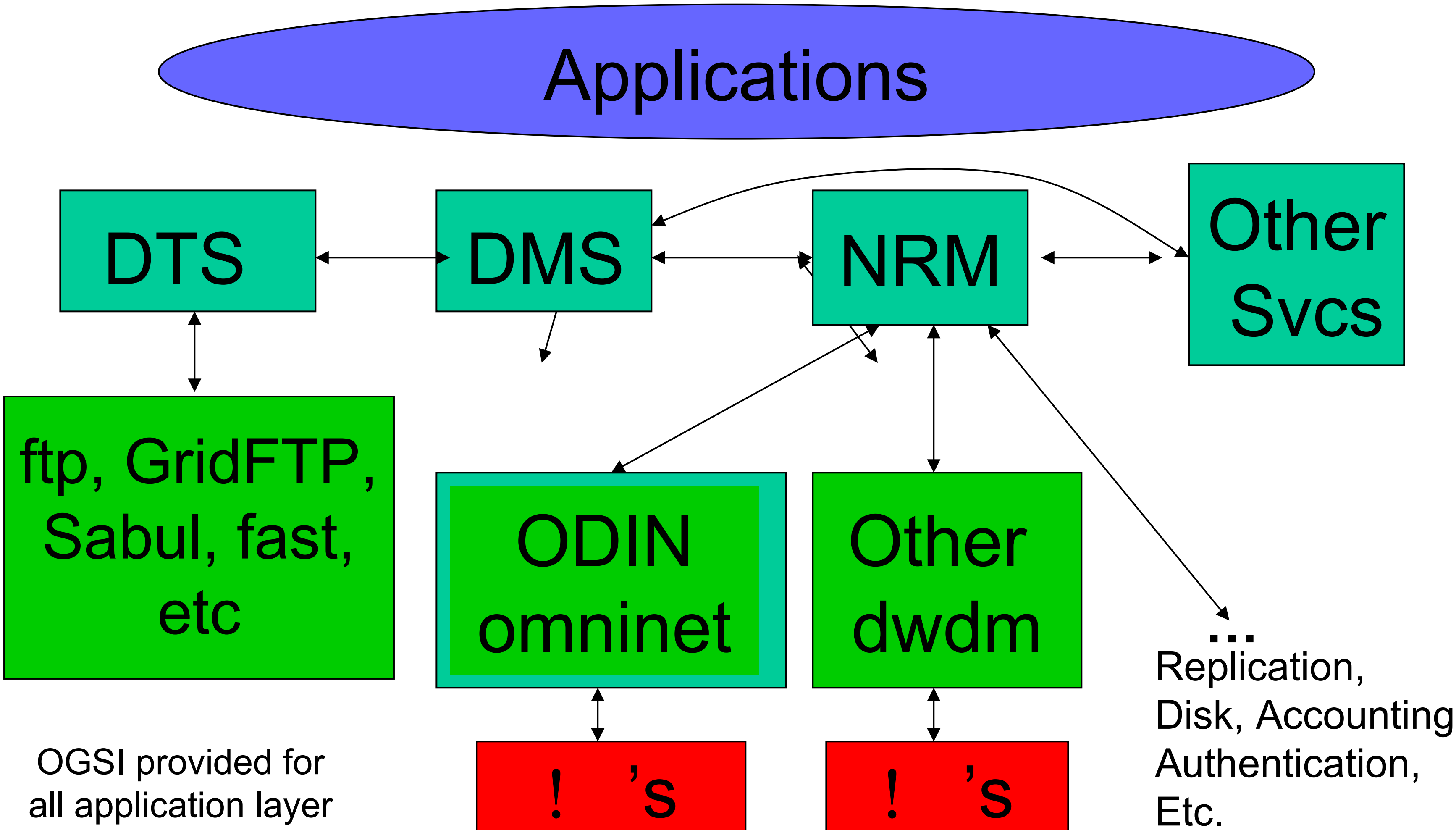


**Challenge**: Emerging data intensive applications require:

- 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

# Architecture




OGSI provided for all application layer interfaces

# 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

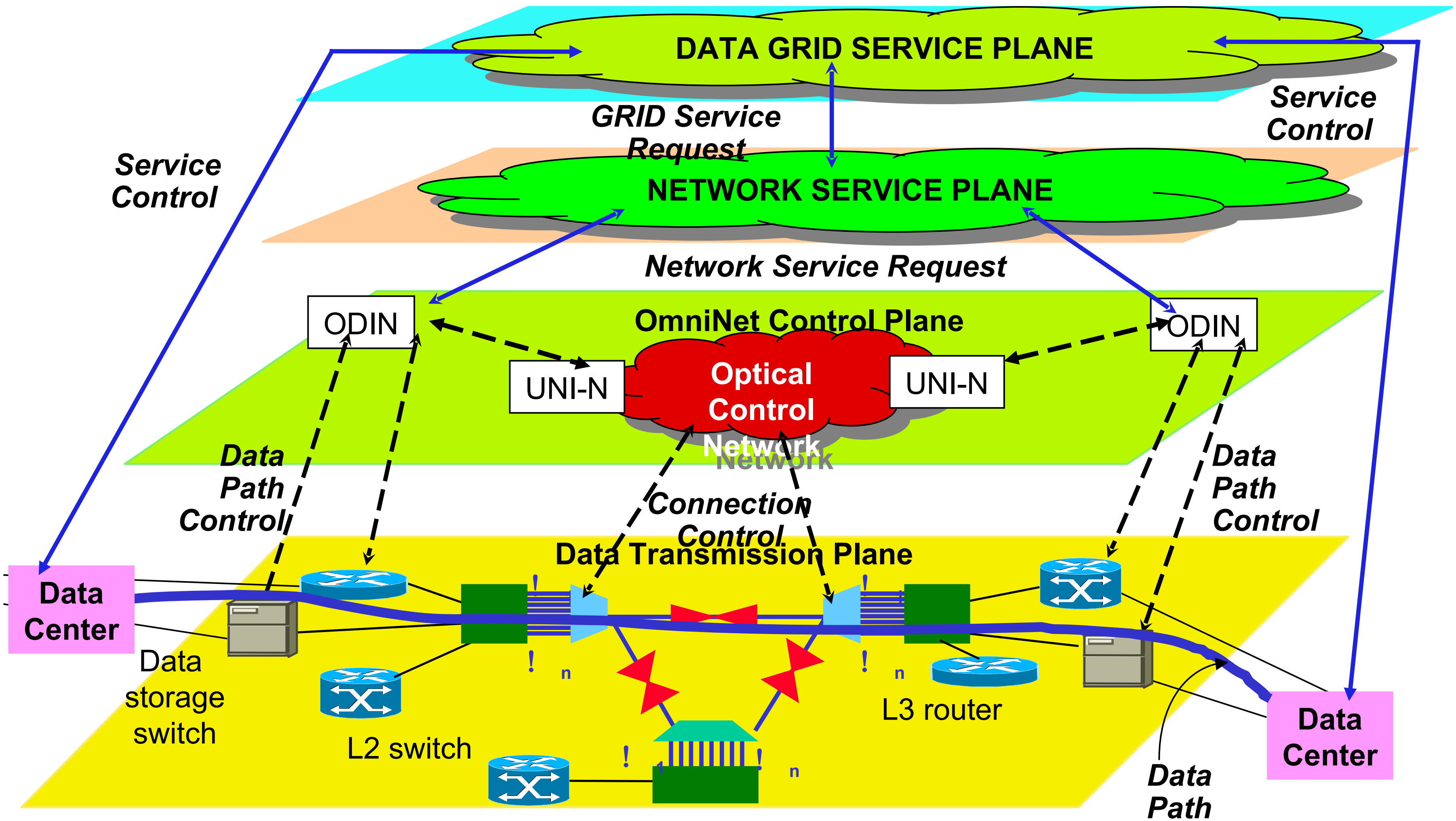
## Intelligent Middleware for Adaptive Optical Networking

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

# DWDM-RAM Service Control Architecture



# 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

# New Concepts

- Many-to-Many vs. Few-to-Few
- Apps optimized to waste bandwidth
- Network as a Grid service
- Network as a scheduled service
- New transport concept
- New control plane
- Cloud bypass

# 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

End of SLAC mini-overview
3 March 2004

# Key Terms

DTS – Data Transfer Service
Effects transfers
NRM – Network Resource Management
Interface to multiple physical/logical network types
Consolidation, topology discovery, path allocation, scheduler, etc.
DMS – Data Management Service
Topology discovery, route creation, path allocation
Scheduler/optimizer
Other Services
Replication, Disk, Accounting, Authentication, Security, etc.

# Possible Extensions

Authentication/Security
- Multi-domain environments

Replication for optimization
- May help refine current Grid file system models
- May Use existing replica location services

Priority models

Rule-based referees
- Allow local and policy-based management
- Add domain specific constraints

# Extending Grid Services

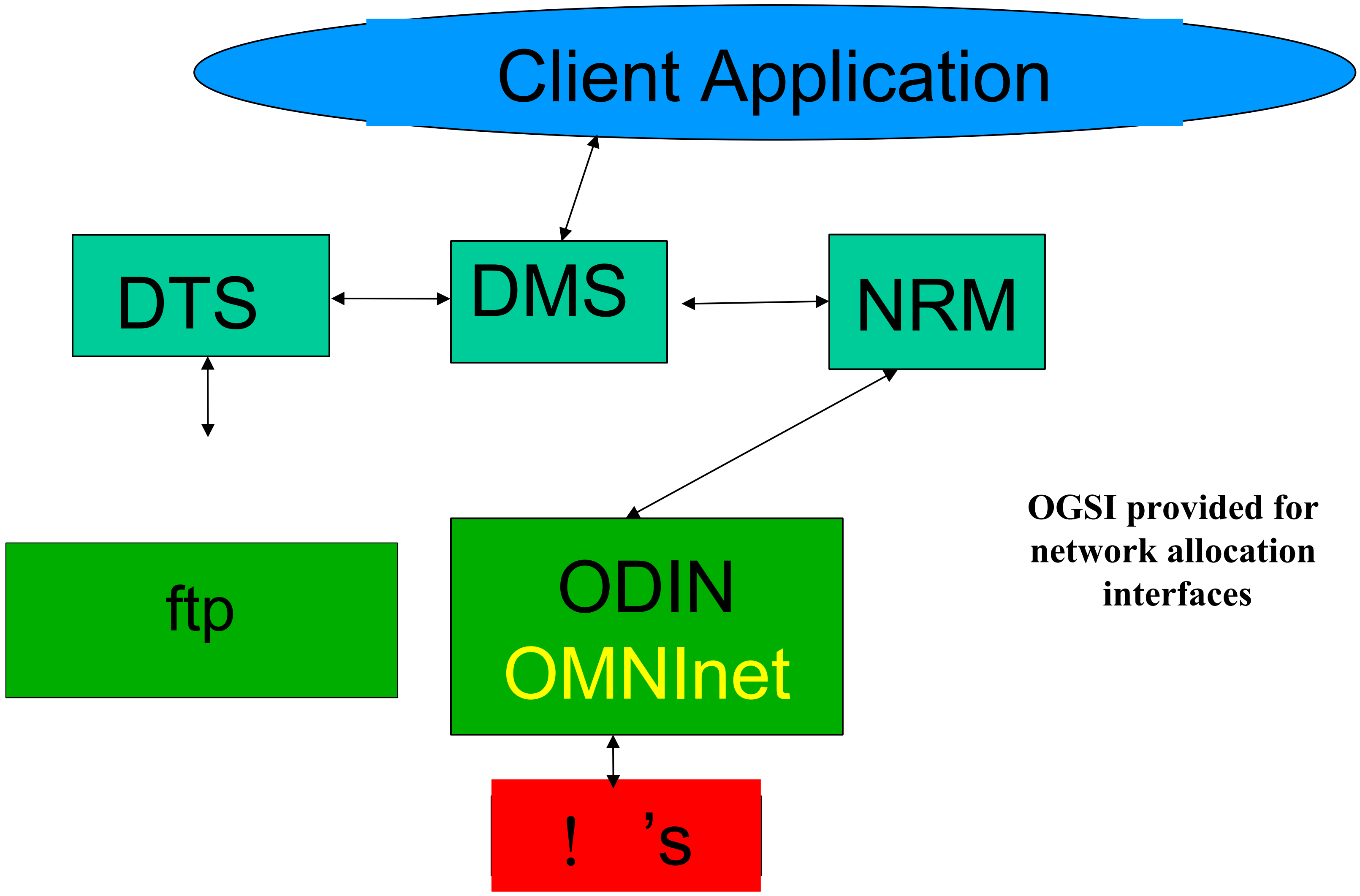
- OGSI interfaces
  - Web Service implemented using SOAP and JAX-RPC
  - Non-OGSI clients also supported
- GARA and GRAM extensions
  - Network scheduling is new dimension
  - Under-constrained (conditional) requests
  - Elective rescheduling/renegotiation
- Scheduled data resource reservation service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

**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

# Current Implementation




OGSI provided for network allocation interfaces

# 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

# NRM Web Services Compliance

Accessible as Web Service for non-OGSI callers

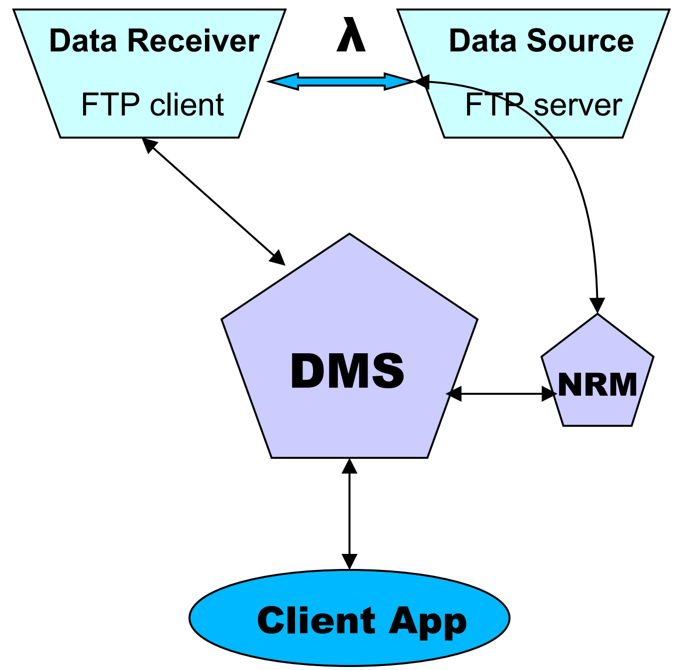
Fits Web Service model:

- Single-location always-on service
- Atomic message-oriented transactions
- State preserved where necessary at the application level

No OGSI extensions, such as service data and service factories

# Data Management Service

- Uses standard ftp (jakarta commons ftp client)
- Implemented in Java
- Uses OGSI calls to request network resources
- Currently uses Java RMI for other remote interfaces
- Uses NRM to allocate lambdas
- Designed for future scheduling

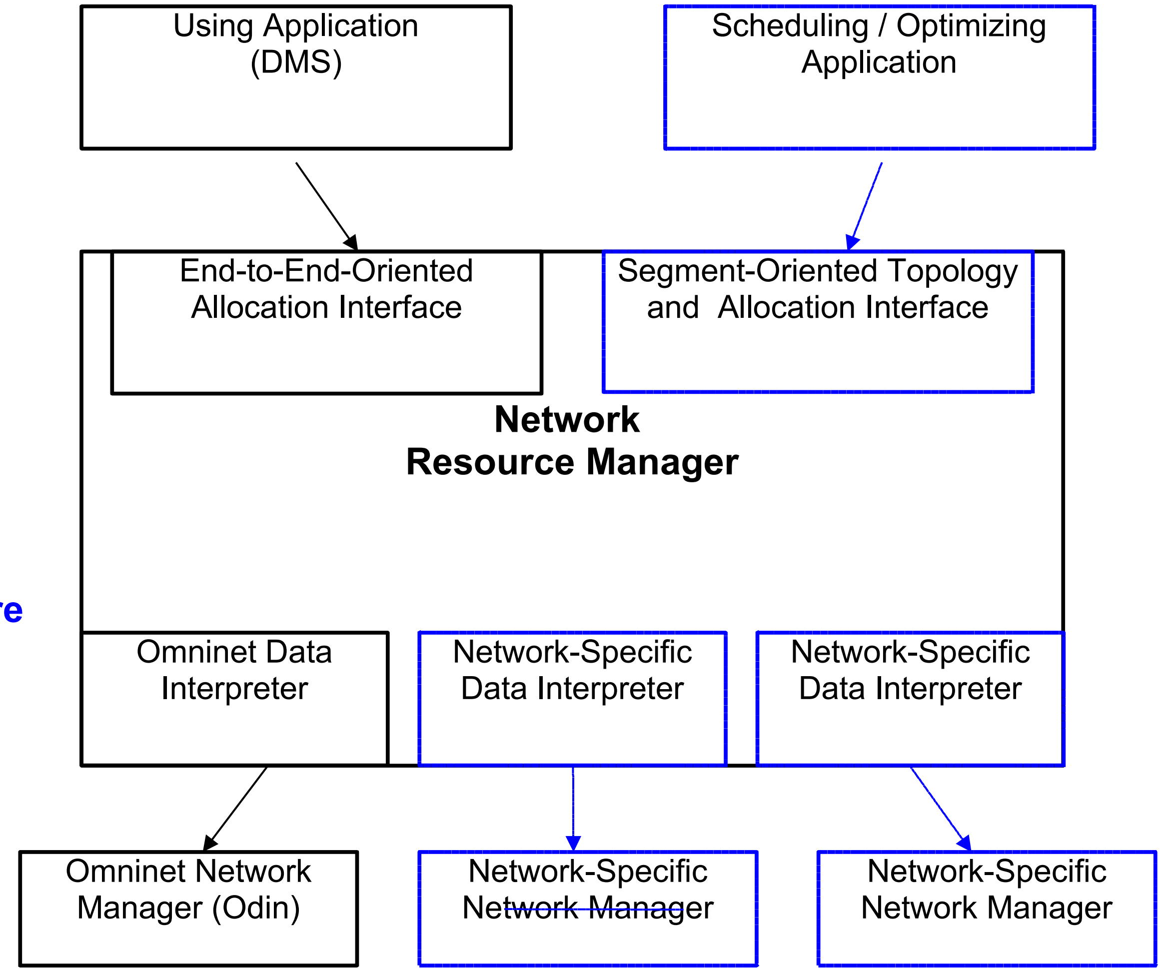


# Network Resource Manager

- 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

# Network Resource Manager




**Items in blue are planned**

# 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

# ODIN

Optical Dynamic Intelligent Networking Services:
An Architecture Specifically Designed to Support Large Scale,
Data Intensive, Extremely High Performance, Long-Term Flows

OGSA/OGSI Compliant Service
Dynamic Lambda Provisioning Based on DWDM
Beyond Traditional Static DWDM Provisioning
Scales to Gbps, Terabits Data Flows with
Flexible, With Fine-Grained Control

Lightpaths: Multiple Integrated Linked Lambdas, Including
One to Many and Many to One, Intradomain/Interdomain

# Terms

**ODIN Server** – A server software that accepts and fulfills requests (eg, allocates and manages routes, paths)

**Resource** – A host or other hardware that provides a service over the optical network, OGSA/OGSI compliant

**Resource Server** – Server software running on a Resource that provides the service

**Resource Config. Server** – Server software that receives route configuration data from the ODIN Server

**Client** – A host connecting to a Resource through the optical network, in this demonstration, Grid clusters

**Network Resource** – Dynamically allocated network resource, in this demonstration, Lightpaths

# Lightpath Provisioning Processes

Specialized Signaling

Request Characterization, Resource Characterization, Optimization, Performance, and Survival/Protection, Restoration, Characterization

Basic Processes Are Directed at Lightpath/!   !   Management:

Create, Delete, Change, Swap, Reserve

And Related Processes:

Discover, Reserve, Bundle, Reallocate, etc.

IETF GMPLS As Wavelength Implementation Tools

Utilizes New Photonic Network Protocols

# Core Processes

O-UNI, Specialized Interfaces, eg, APIs, CLIs
Wavelength Distribution Protocol
Auto-Discovery of Optical Resources
Self-Inventorying
Constraint Based Routing
Options for Path Protection, Restoration
Options for Optical Service Definitions

## Addressing and Identification

Options for Interface Addressing

Options for VPN IDs

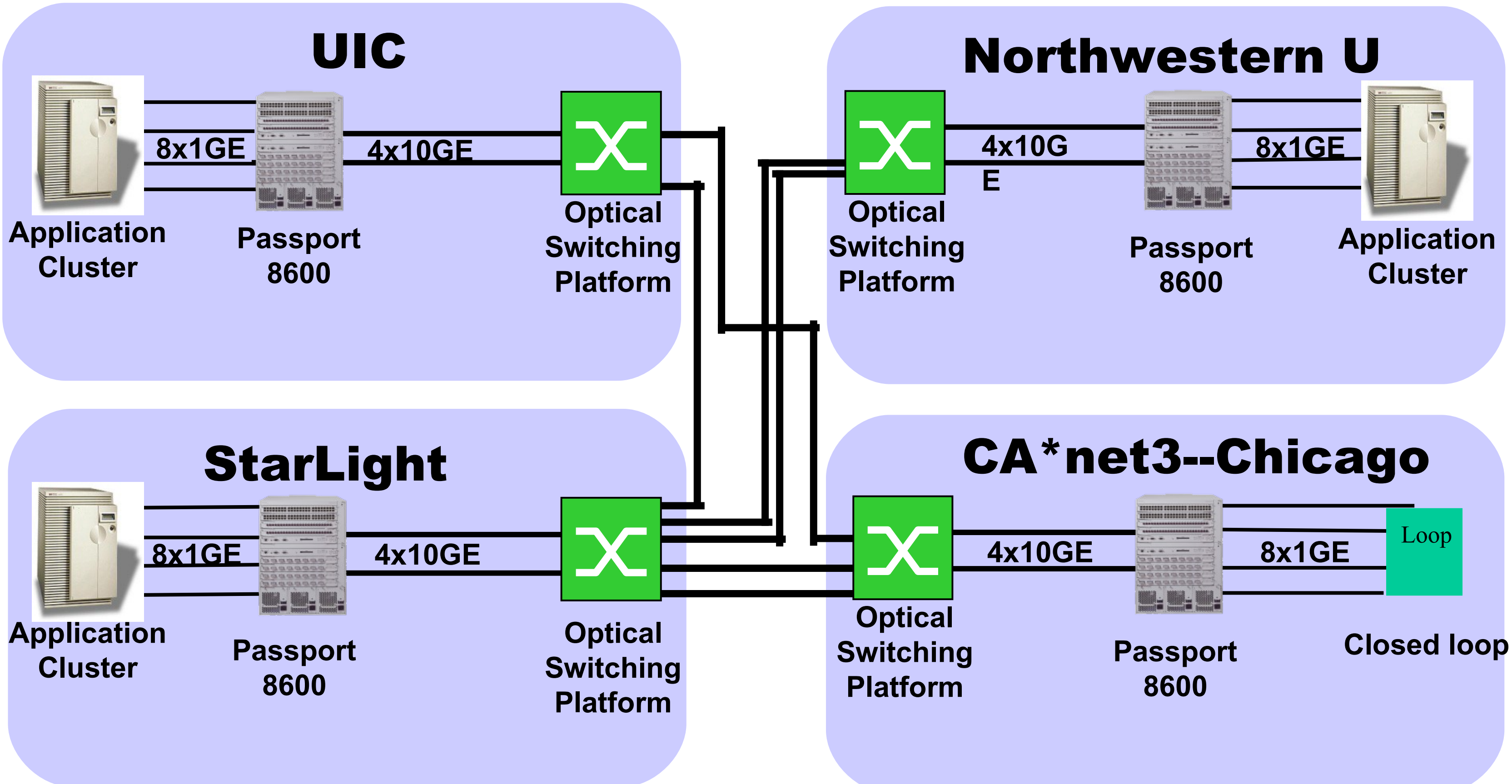
Port, Channel, Sub-channel IDs

Routing Algorithm Based on Differentiated Services

Options for Bi-directional Optical Lightpaths, and Optical Lightpath Groups

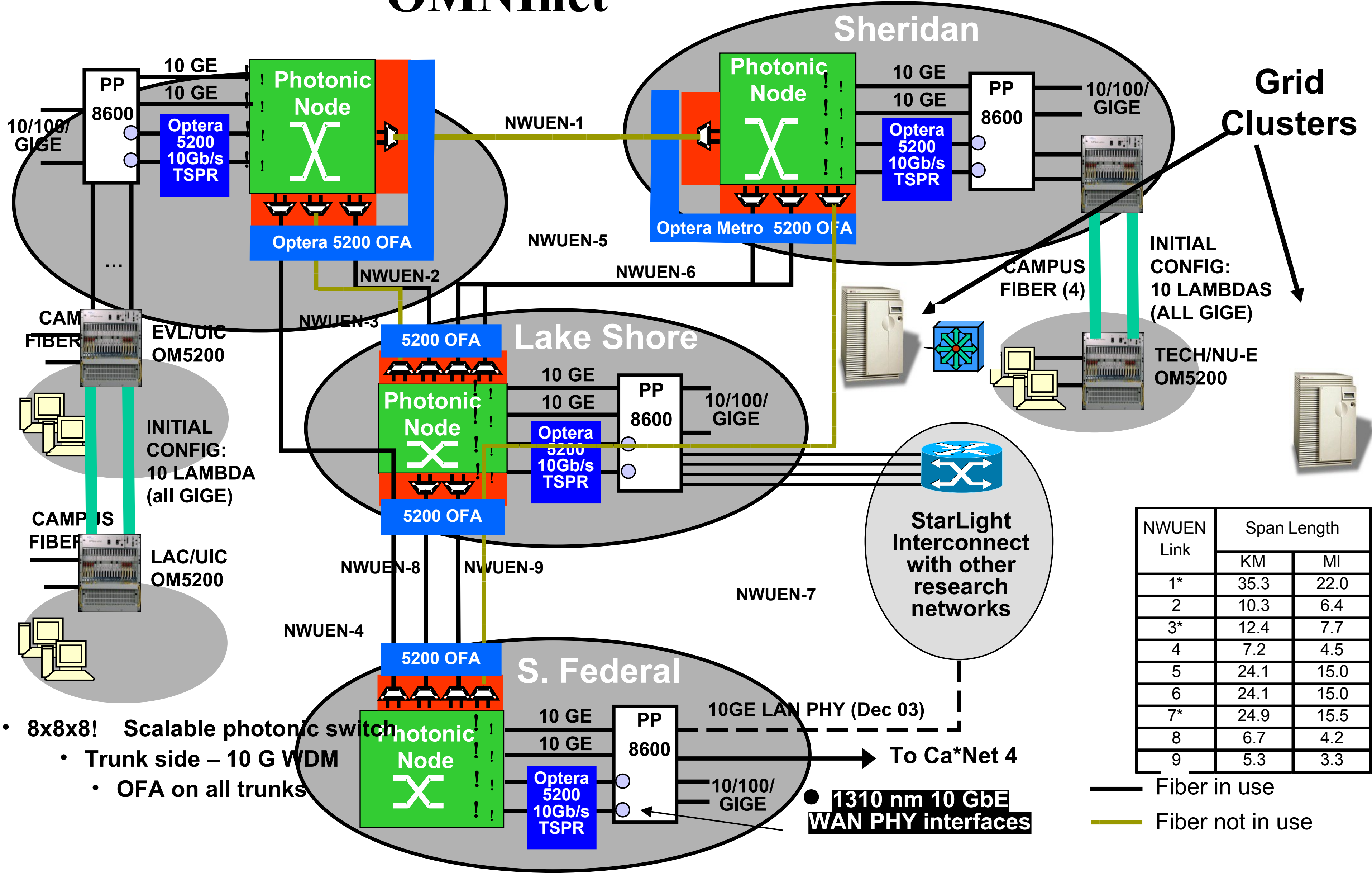
Optical VPNs

# OMNInet Core Nodes




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

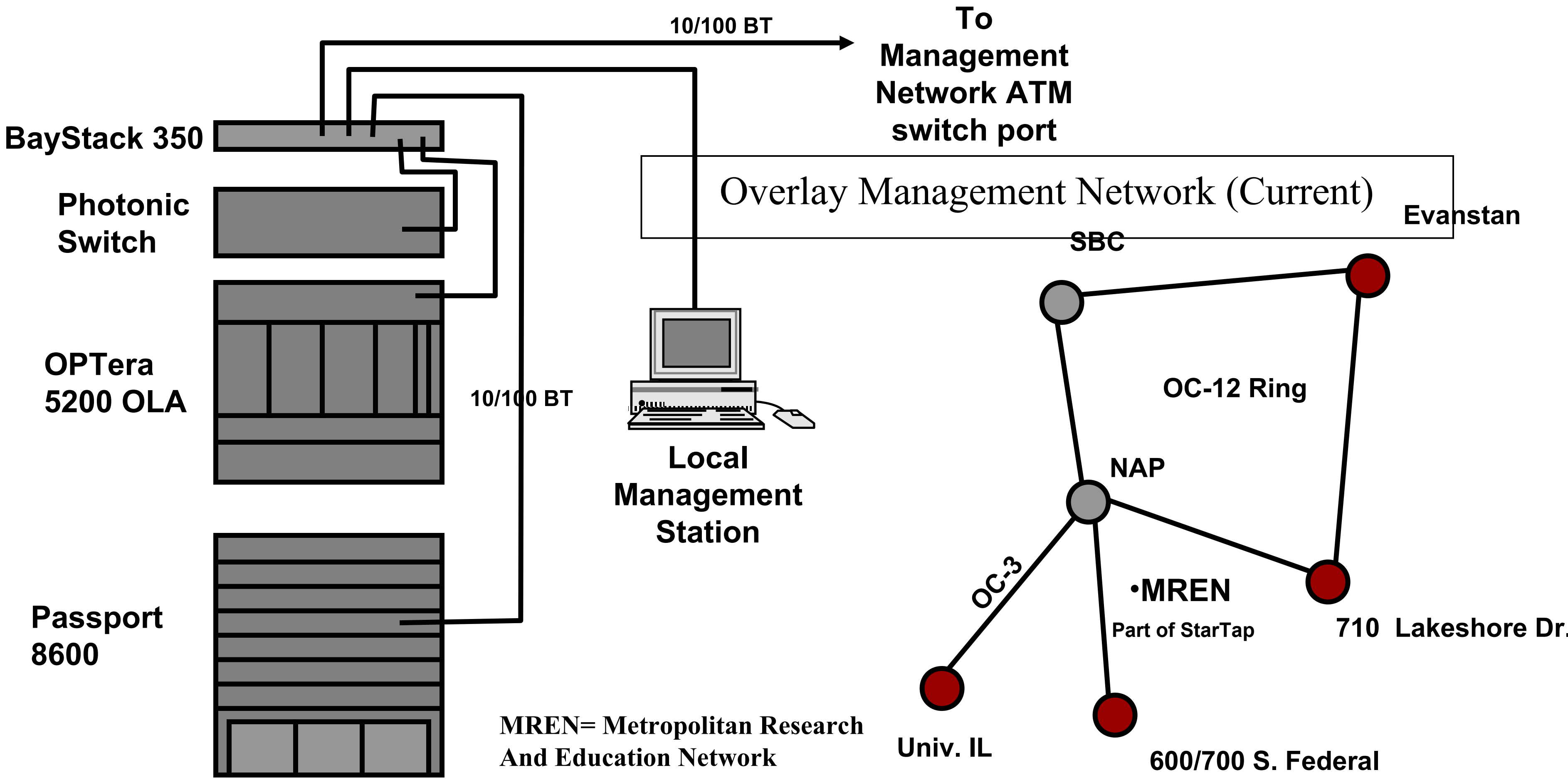
# OMNInet




- 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 |

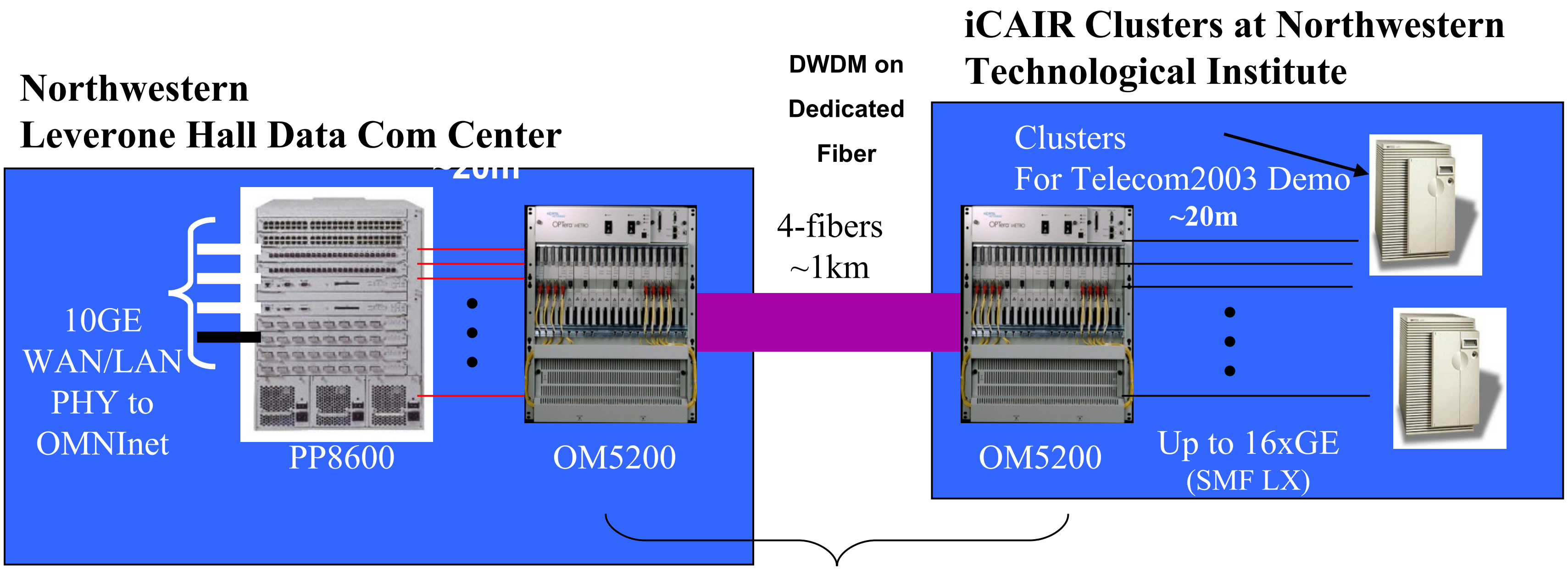
# OMNInet Control Plane Overlay Network




MREN= Metropolitan Research
And Education Network

- Uses ATM PVC with 2 Mb/s CIR from existing network (MREN + OC12)
- Hub and spoke network from 710 Lakeshore Dr.

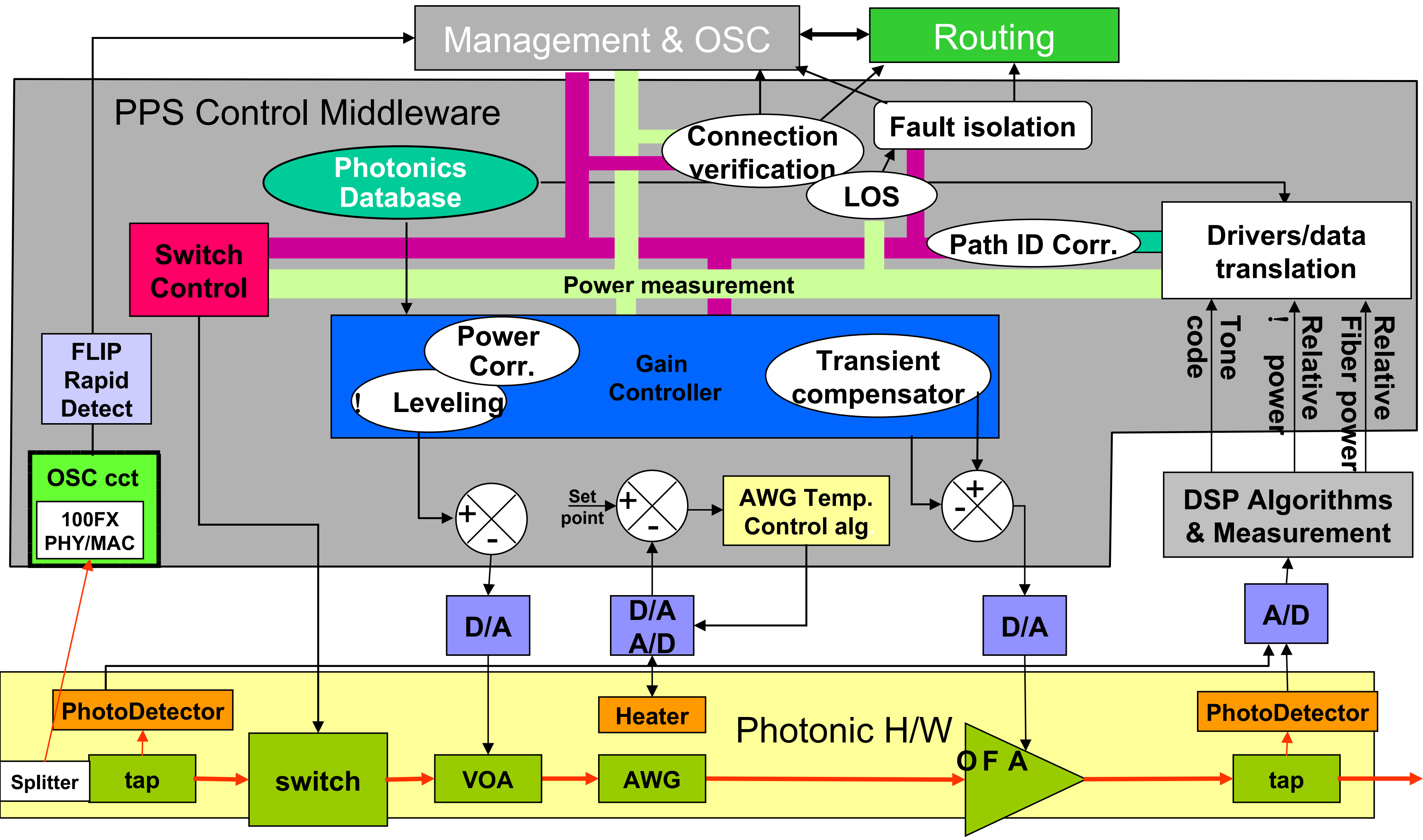
# OMNInet Optical Grid Clusters




DWDM Between Cluster Site and OMNInet Core Node at iCAIR sites at Northwestern in Evanston

- The implementation is lambdas (unprotected).
- Installed shelf capacity and common equipment permits expansion of up to 16 lambdas through deployment of additional OCLD, and OCI modules.
- A fully expanded OM5200 system is capable of supporting 64 lambdas (unprotected) over the same 4-fiber span.

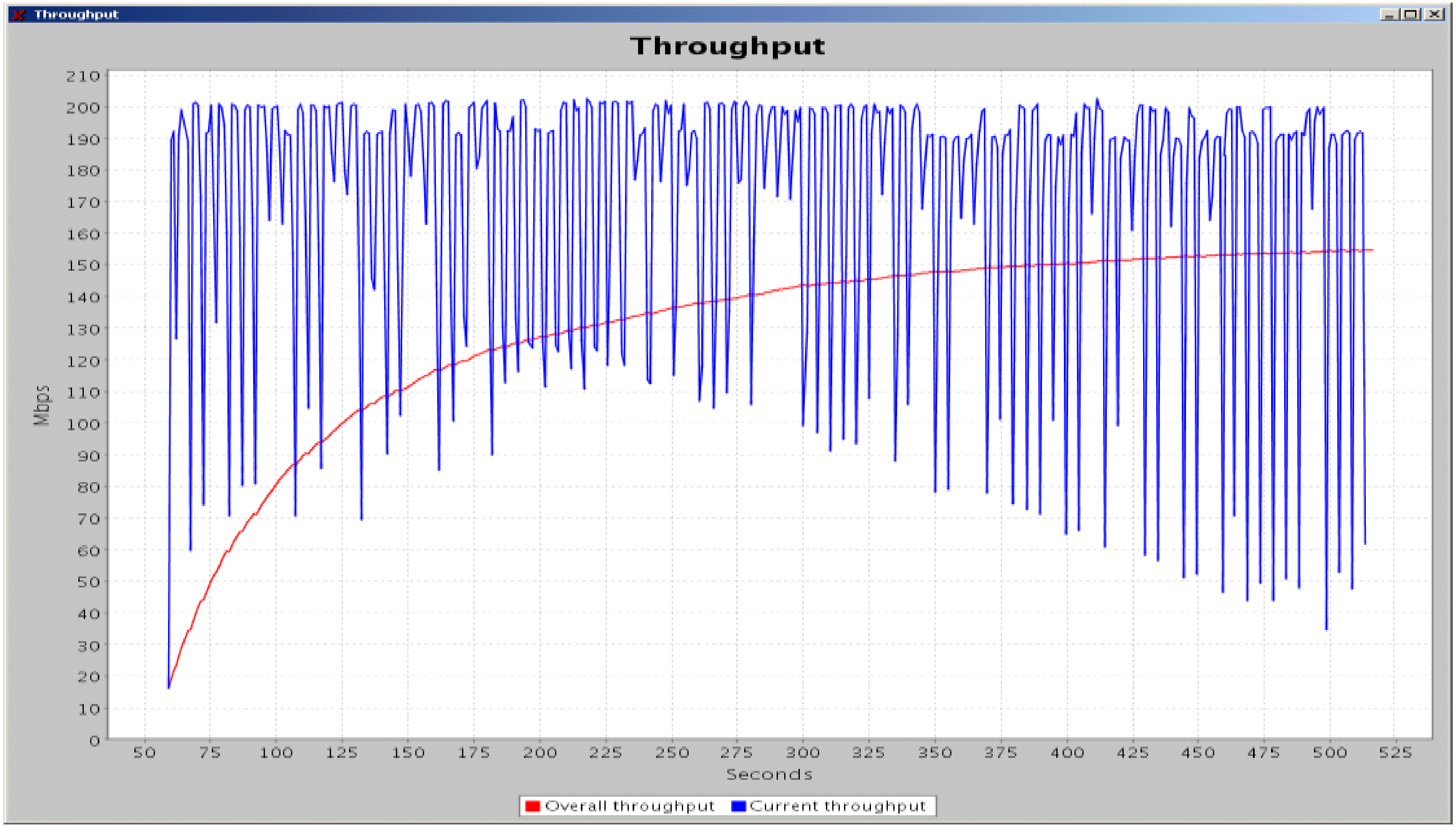
# Physical Layer Optical Monitoring and Adjustment



# Application level measurements

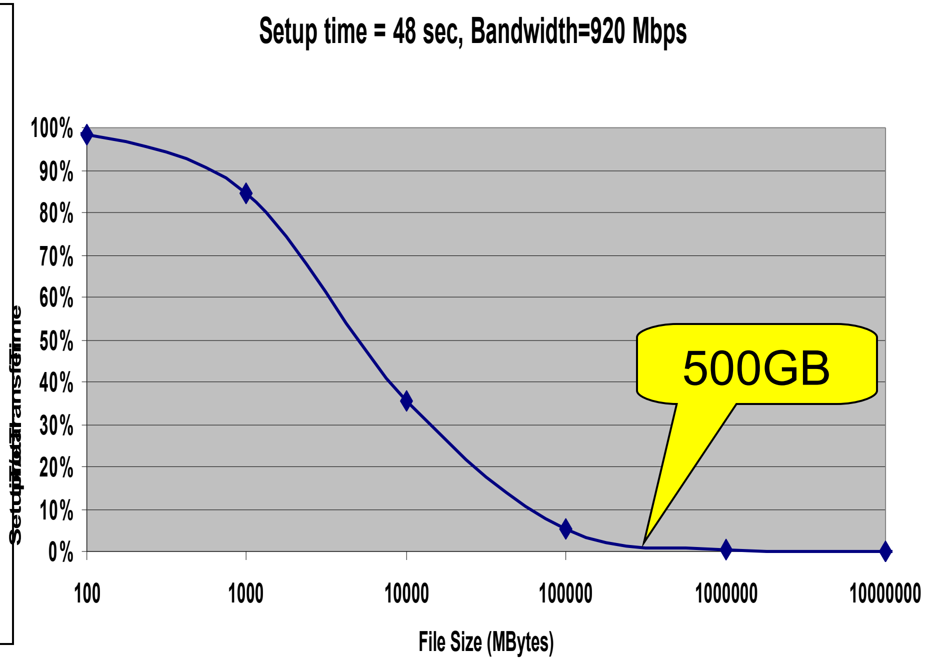
| | |
|---|---|
| Path allocation: | 48.7 secs |
| Data transfer setup time: | 0.141 secs |
| FTP transfer time: | 464.624 secs |
| Effective transfer rate: | 156 Mbits/sec |
| Path tear down time: | 11.3 secs |
| File size: | 10 GB |

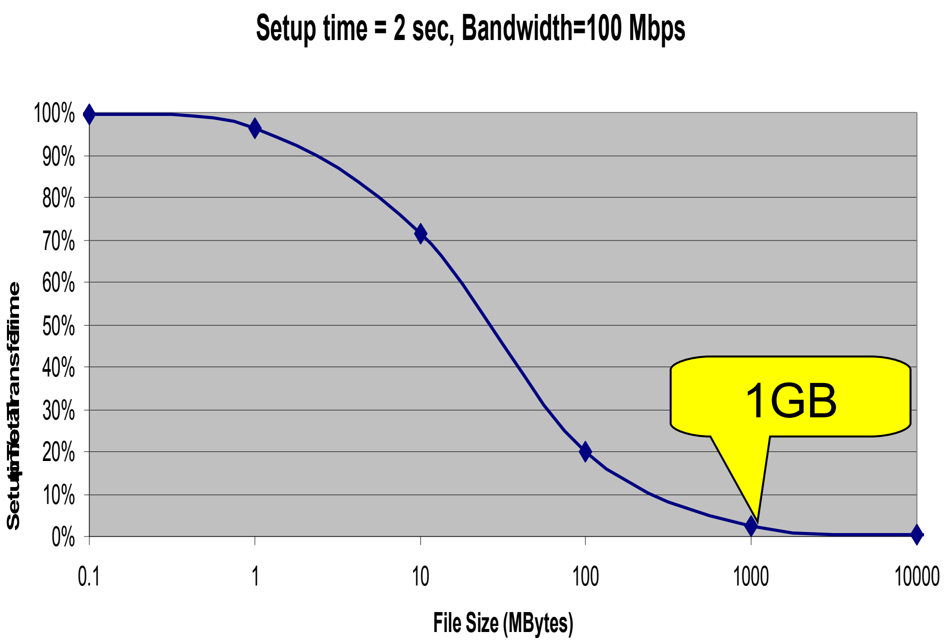
# 10GB file Transfer

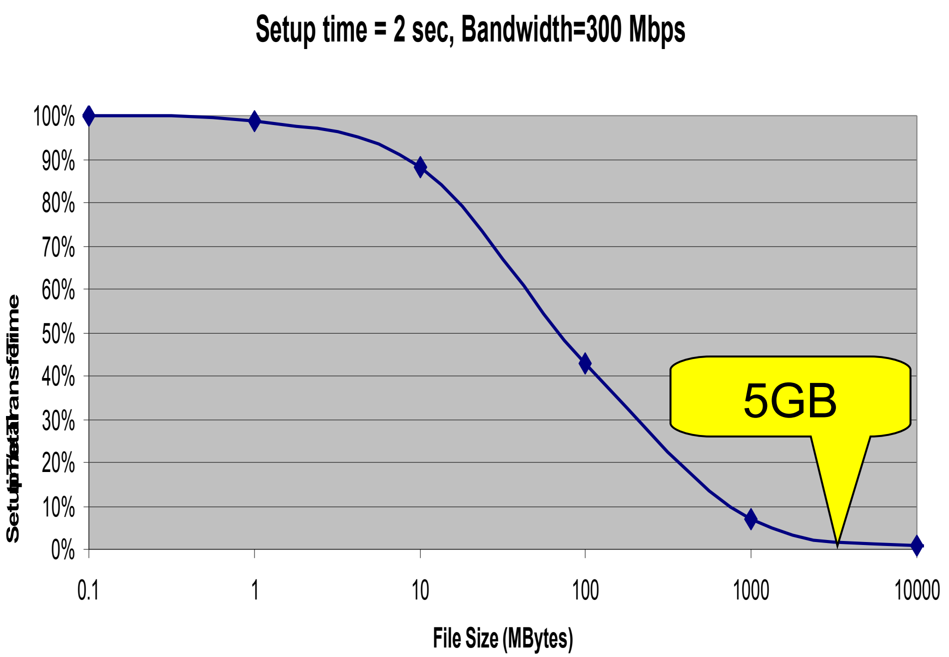


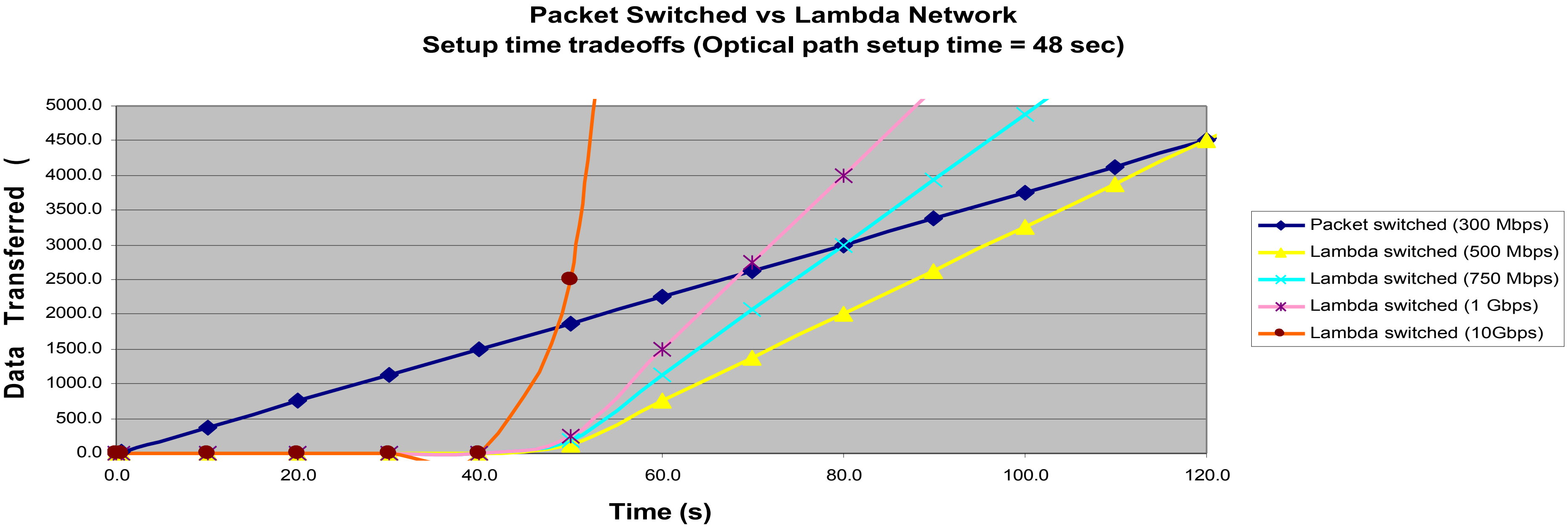
# Path Allocation Overhead as a % of the Total Transfer Time

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

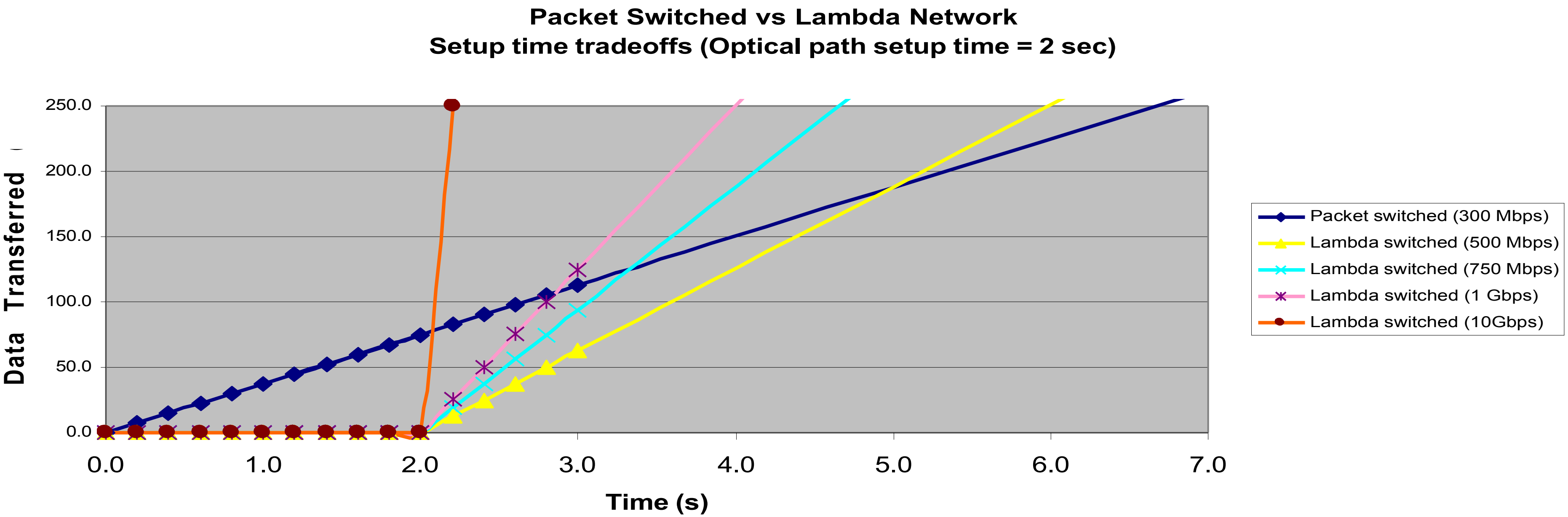


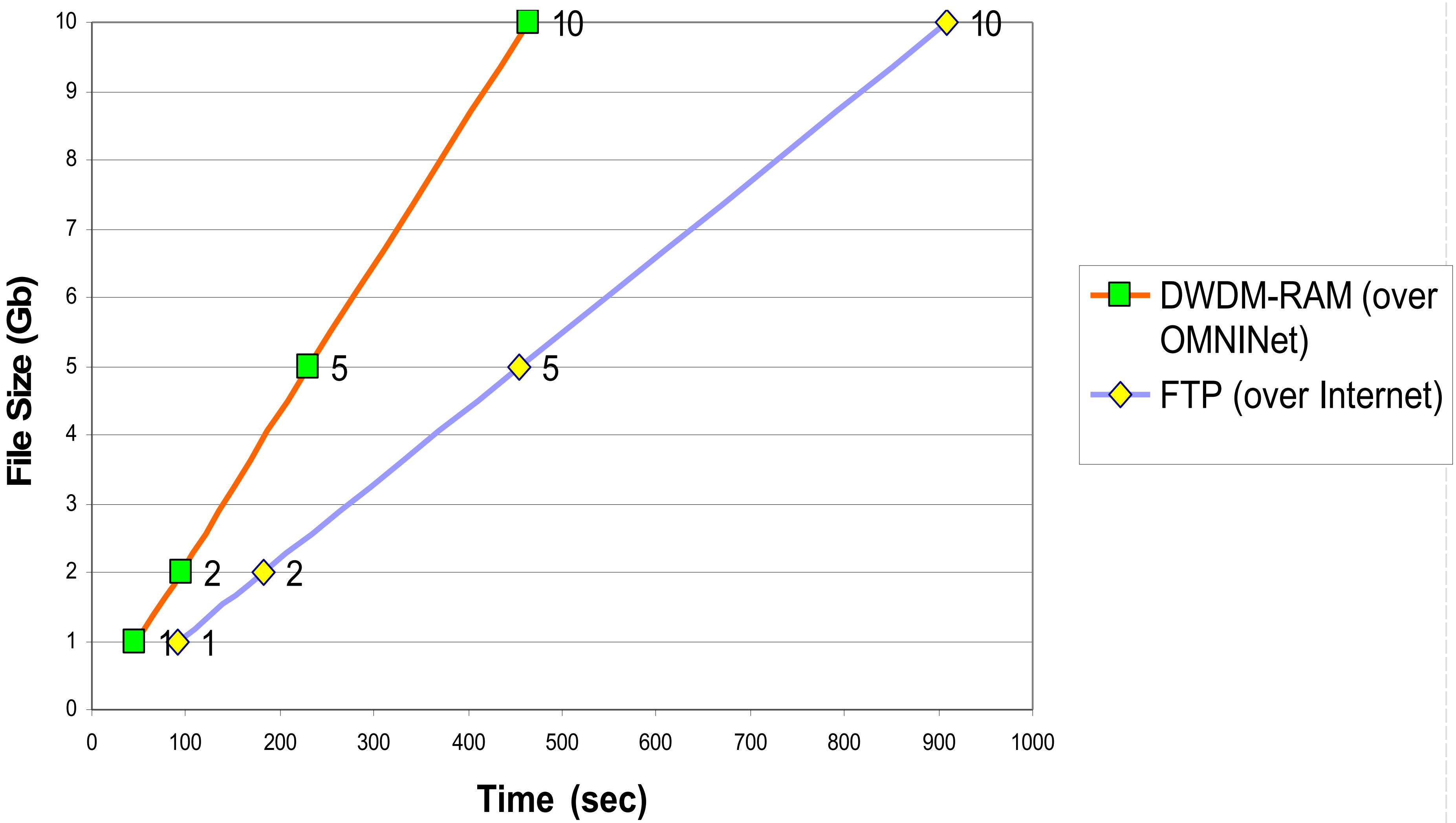


Packet Switched vs Lambda Network
Setup time tradeoffs (Optical path setup time = 48 sec)
Data Transferred (
Time (s)
0.0
20.0
40.0
60.0
80.0
100.0
120.0
0.0
500.0
1000.0
1500.0
2000.0
2500.0
3000.0
3500.0
4000.0
4500.0
5000.0
Packet switched (300 Mbps)
Lambda switched (500 Mbps)
Lambda switched (750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

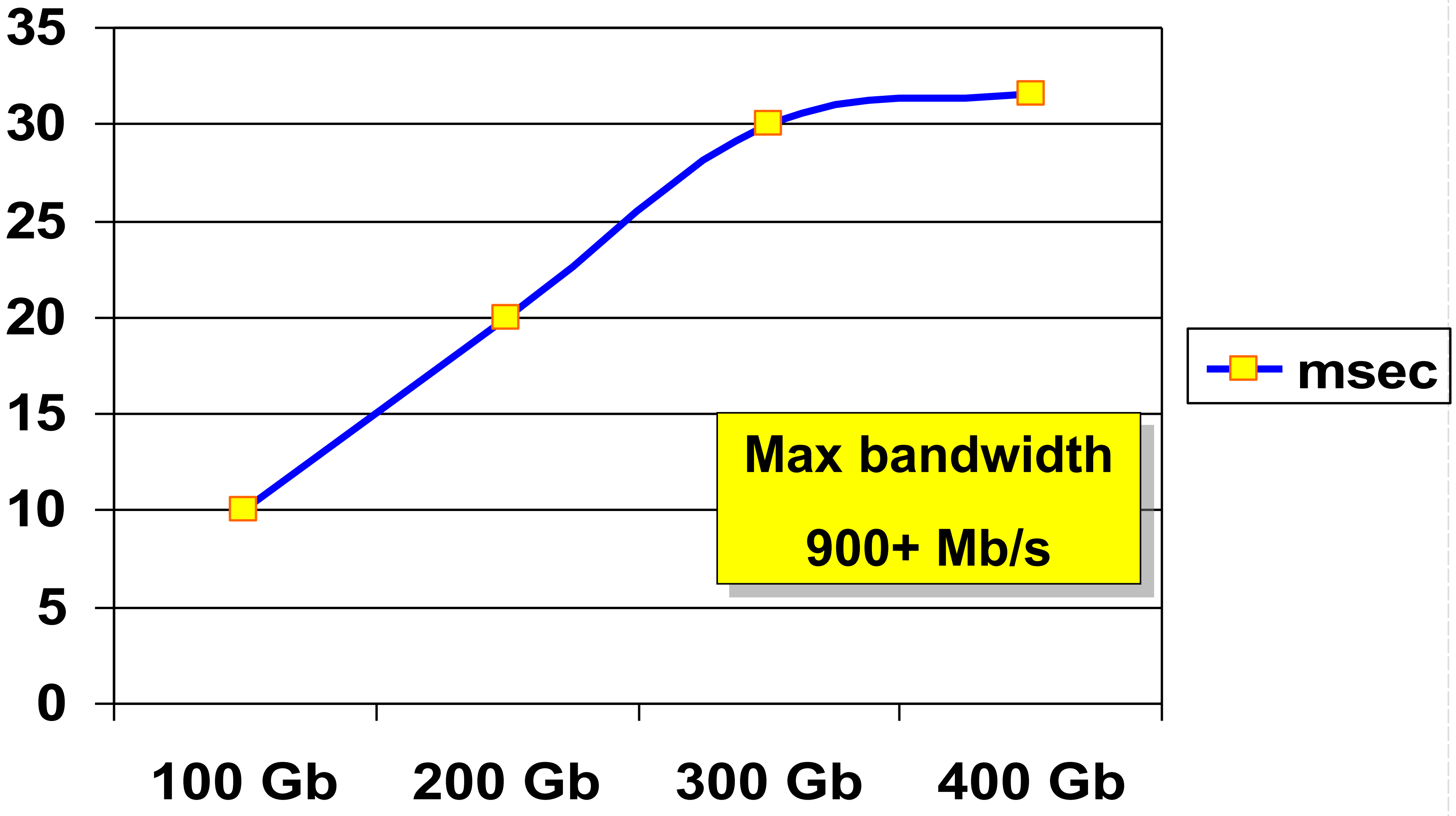



Packet Switched vs Lambda Network
Setup time tradeoffs (Optical path setup time = 2 sec)
Data Transferred
Time (s)
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
0.0
50.0
100.0
150.0
200.0
250.0
Packet switched (300 Mbps)
Lambda switched (500 Mbps)
Lambda switched (750 Mbps)
Lambda switched (1 Gbps)
Lambda switched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

# File transfer times



# File Transfer Times



# Optical level measurements

| | |
|---|---|
| Time to set up an individual X-connect: | secs |
| UNI-N processing time for request: | secs |
| Time taken by the routing card to send command to control card: | secs |
| Time taken by the routing card to forwarding request to next hop in control plane: | secs |
| Time taken by the control card to drive the switch card : | secs |
| End-to-end light path setup : | secs |

## Enhanced Optical Dynamic Intelligent Network Services

Additional OGSA/OGSI development
Enhanced signaling
Enhanced integration with optical component addressing methods
Extension of capabilities for receiving information from
L1 process monitors
Enhanced capabilities for establishing optical VPNs
New adaptive response processes for dynamic conditions
Explicit segment specification

# Enhanced Middleware Services

Enhanced integration with data services layer
Enhanced understanding of L3-L7 requirements
Awareness of high performance L3/L4 protocols
Enhanced integration with edge resources
Middleware process performance monitoring and analysis
New capabilities for scheduling
Security

## Expanded Data Management Service

New methods for scheduling
New methods of priority setting
Enhance awareness of network resources
Technique for forecasting demand and preparing responses
Replication services
Integration with metadata processes
Integration with adaptive storage services
Enhanced policy mechanisms

# Photonic Testbed - OMNInet

Implementation of RSVP methods
Experiments with parallel wavelengths
Experiments with new types of flow aggregation
Experiments with multiple 10 Gbps parallel flows
Enhancement of control plane mechanisms
Additional experiments with interdomain integration
Enhanced integration with clusters and storage devices

# Additional Topics

Enhanced security methods
Optimization heuristics
Integration with data derivation methods
Extended path protection
Restoration algorithms
Failure prediction and fault protection
Performance metrics, analysis and reporting
Enhanced integration of optical network information flows
with L1 process monitoring