A. Cover Page

Submitted To: U.S. Department of Energy

Project Title: ISOGA: Integrated Services Optical Grid Architecture for Emerging E-Science Collaborative Applications

Report Number: FG02-03ER25566

Project type: Career Award

Principal Investigator: Oliver Yu
Department of Electrical & Computer Engineering
University of Illinois at Chicago
851 S. Morgan St., 1020 SEO, Chicago, IL 60607-7053
Phone: (312) 996-2308; Fax: (312) 996-6465; Email: oyu@uic.edu

Summary

This final report describes the accomplishments in the ISOGA (Integrated Services Optical Grid Architecture) project. ISOGA enables efficient deployment of existing and emerging collaborative grid applications with increasingly diverse multimedia communication requirements over a wide-area multi-domain optical network grid; and enables collaborative scientists with fast retrieval and seamless browsing of distributed scientific multimedia datasets over a wide-area optical network grid. The project focuses on research and development in the following areas: the polymorphic optical network control planes to enable multiple switching and communication services simultaneously; the intelligent optical grid user-network interface to enable user-centric network control and monitoring; and the seamless optical grid dataset browsing interface to enable fast retrieval of local/remote dataset for visualization and manipulation.

.

B. Table of Contents

C. ISOGA Goals and Objectives

The ISOGA (Integrated Services Optical Grid Architecture) project develops optical network control planes, optical grid user-network interface, and optical grid dataset browsing interface. These deliverables are integrated in a multi-domain optical network grid to enable efficient deployment of existing and emerging collaborative grid applications with increasingly diverse multimedia communication requirements; and to enable collaborative scientists with fast retrieval and seamless browsing of scientific multimedia datasets for manipulation and visualization. The project consists of the following deliverables:

- The Multi-granular Integrated Services Optical Network (MISON) control plane enables a polymorphic photonic network with multi-granular all-optical switching and communication services providing efficient data traffic transports with diverse communication requirements.
- The Secure Photonic Interdomain Negotiator (SPIN) enables secure dynamic optical connection establishment between the source and destination domains over a multi-domain optical network with heterogeneous intra-domain control planes.
- The intelligent optical grid user interface (IOGUI) includes the user-network interface to enable users/applications of controlling optical network connections and monitoring traffic operations; and the dataset browsing interface to enable users of retrieving local/remote dataset to any collaborating grid cluster for dataset manipulation and visualization.

C.1. Multi-granular Integrated Services Optical Network (MISON)

C.1.a. Background and Motivation

Wavelength division multiplexing (WDM) optical network architectures with diverging optical cross-connect (OXC) data planes and traffic control planes have been designed to meet specific communication and traffic bandwidth granularity requirements of applications. Optical circuit switching (OCS) networks are designed to efficiently support constant wavelength stream traffic of multi-gigabit rate through a lightpath of wavelength-switched channels. OXC networks employ a data plane of all-optical wavelength switches with optional wavelength-converters; and a control plane enabling two-way reservation of static or dynamic lightpaths. The OCS-based wavelength-routed optical networks (WRONs) enable lightpath setup on demand from users, or from network providers during network reconfiguration. Existing OCS-based control plane includes IETF Internet-based generalized multi-protocol label switching (GMPLS) [1] and ITU telecommunication-based automatically switched optical networks (ASON) [2].

All-optical waveband switches [3] enable switching of wavelength-groups as a whole, which allow minimization of the number of optical ports required for given traffic loading. Waveband switching (WBS) networks can be used to support multi-wavelength stream traffic through a waveband-switched lightpath-group. However, to successfully establish an end-to-end waveband-switched lightpath-group, the number of wavelengths available for each connecting optical link should be considerably larger than the waveband granularity. This condition may not be met since each optical link accommodates only a small number of wavelengths in current WDM networks. This paper extends OCS to multi-wavelength OCS, which supports multi-wavelength stream traffic through a wavelength-switched lightpath-group with constituent lightpaths allowed to take on different routes while satisfying end-to-end delay-variation requirement within the group.

Optical burst switching (OBS) [4] networks are designed to efficiently support dynamic sub-wavelength bursty traffic through statistical multiplexing of lightpaths and burst buffering at the network edge. OBS networks employ a data plane of fiber delay lines (FDLs) and all-optical switches with optional wavelength-converters; and a control plane enabling one-way or fast two-way reservation of dynamic switching of sub-wavelength bursts. Existing OBS control planes include just-in-time (JIT) [5] and just-enough-time (JET) [6] schemes.

As emerging applications demand more diverse communication and bandwidth granularity requirements, it would be desirable to have a multi-service optical network architecture or polymorphic virtual optical networks [7] that can provide heterogeneous network services to efficiently transport Internet, telecom or scientific applications traffic. Internet traffic is primarily connectionless and elastic in nature, while high performance computing (HPC) [8] scientific applications require connection-oriented high bandwidth connectivity for aggressive data traffic transport.

Furthermore, parallel HPC grid applications require multi-wavelength connectivity between computing clusters to transport synchronized parallel data streams with minimum end-to-end delay variations among the data streams.

Multi-service optical network architectures include the followings: heterogeneous OXCs with multiple control planes, hybrid or multi-layer OXCs with unified control plane, and single-layer OXCs with multiple control planes or a unified control plane. Employing heterogeneous OXCs to support multi-service optical networks, it will result in network resource partition and sub-optimal resource utilization as each OXC-type is service specific.

Hybrid OXC enables multi-layer traffic switching at different granularities by combining two or more switching fabrics such as fiber, waveband and wavelength and sub-wavelength. For example, the multi-granular optical cross-connect (MG-OXC) [9] enables traffic switching at different granularities of fiber, waveband and wavelength. To enable sub-wavelength switching, hybrid OXCs employ optical-electrical-optical (OEO) switching with electrical switching fabrics. The hybrid OXCs in [10] [11] combines all-optical waveband and OEO-based time-division multiplexed (TDM) switching fabrics. Multi-service optical networks can be enabled over hybrid OXCs through a unified control plane. However, reconfigurations of the port interfaces between different switching fabrics in hybrid OXCs could not be performed dynamically to adapt to traffic loading changes.

Multi-service optical networks could be realized over single-layer OXCs via multiple control planes. For example, OCS and OBS based services can be combined [12][13] with respective control planes over a physical topology of all-optical wavelength switches. To avoid the complexity of maintaining multiple control planes, polymorphic multi-service optical network (PMON) [14] employs a unified GMPLS control plane to combine OCS and labeled-OBS based services over labeled optical-burst switches, which are enabled with FDLs and orthogonal optical modulation of label and data payload. Overspill routing in optical network (ORINON) [15] similarly employs unified control to combine OCS and labeled-OBS based services, with the added feature of congestion-based deflection routing to optimize wavelength utilization.

C.1.b. Problem Formulation and Approach

From user's and network provider's perspectives, the design objective of multi-service optical network would be to maximize the number of offered data transport services based on mature single-layer OXCs and a unified control plane, while minimizing accessory devices such as FDLs and OEO interfaces. Based on commercially available all-optical wavelength switches with no requirement for FDLs and OEO interfaces, we have designed Multi-Granular Optical Switching (MISON) network architecture that offers four stream (sub-wavelength, wavelength, multi-wavelength) and bursty data transport services through a unified control plane. The MISON control plane consists of a unified multi-granular resource reservation signaling protocol over the network core, and service-specific controls at the network edge.

Sub-wavelength stream traffic transport is efficiently supported by the novel Synchronous Stream Optical Switching (SS-OS) service, which shapes sub-wavelength traffic streams into periodic burst-trains at the edge node, and employs proactive periodic reservation to minimize signaling overhead and scheduling complexity. Unlike the scheme in [16], SS-OS does not require FDLs and fast all-optical switches with sub-nanosecond switching time.

Sub-wavelength burst traffic transport is efficiently supported by the novel Adaptive Robust Optical Burst Switching (AR-OBS) service; which aggregates burst traffic at the edge node, and employs fast two-way reservation (with reservation signaling triggered before the burst is fully aggregated) and robust recovery control for reservation blocking. Unlike the centralized-control wavelength-routed OBS (WR-OBS) [17] scheme, AR-OBS employs delayed reservation and distributed-control to avoid scalability and reliability problem.

Multi-wavelength stream traffic transport is efficiently supported by the novel multi-wavelength optical switching (MW-OS) service, which decomposes a multi-wavelength data stream into parallel data streams at the edge through inverse-multiplexing, and transports them through a wavelength-switched lightpath-group while minimizing end-to-end delay-variation within the group. MW-OS employs a novel group-routing and wavelength assignment (GRWA) scheme to select the routing paths and wavelength-channels for a lightpath-group setup request with specified delay-variation requirement. On the other hand, wavelength stream traffic transport is efficiently supported by the defaulted wavelength-routed OCS service, which is enhanced with the reservation signaling protocol to enable fast and robust lightpath setup.

Wavelength-routed OCS and AR-OBS require wavelength reservation, but AR-OBS demands fast and robust reservation to minimize burst blocking. SS-OS service requires reservation of a wavelength at periodic time slots, while MS-OS service requires multi-wavelength reservation. MISON network employs the novel Multi-Granular

Robust Fast Optical Reservation Protocol (MG-RFORP), which extends the GMPLS-based RSVP-TE reservation

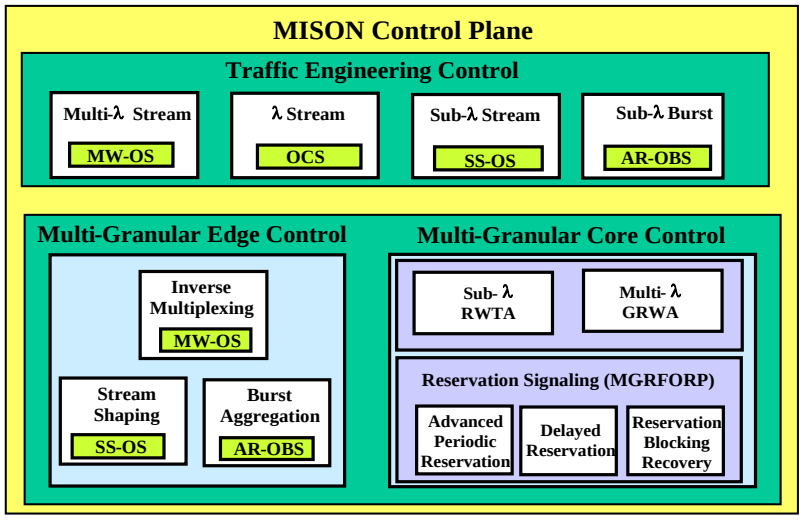


Fig. 1. MISON Control Plane Architecture

protocol [18] to allow fast and robust two-dimensional reservation of time slot and wavelength. MG-RFORP maps the physical topology into a logical topology of wavelength-constrained path segments [19], along which a lightpath must be assigned with the same wavelength as all-optical switches are incapable of wavelength conversion. MG-RFORP employs segmented-based parallel reservation processing to enable fast reservation, and reservation blocking recovery to enable robust reservation.

C.1.c. MISON Control Plane Architecture

MISON control plane (Fig. 1) integrates four data transport services: SS-OS for sub-wavelength stream traffic, wavelength-routed OCS for wavelength stream traffic, MW-OS for multi-wavelength stream traffic, and AR-OBS for sub-wavelength burst traffic. The four transport services are enabled by the multi-granular core control functions and service-specific edge control functions.

Sub-wavelength and wavelength/multi-wavelength data traffic are transported respectively as light-rate (matched to wavelength data-transfer rate) bursts and streams. SS-OBS edge controller shapes an incoming sub-wavelength stream into a periodic light-rate burst train. AR-OBS edge controller aggregates incoming sub-wavelength bursts into light-rate bursts. Wavelength-routed OCS edge controller simply converts incoming wavelength stream from electrical-rate into light-rate streams; while MW-OS edge controller inverse-multiplexes an incoming multi-wavelength stream into a group of light-rate streams with timing relation.

The core nodes are all-optical switches with traffic control unit to support scheduling of wavelength reservation over time-division frames of fixed-sized time slots. Through the control unit and the all-optical switching fabrics, light data is routed transparently at each node through the dimensions of port, wavelength and optional time-slot. Light-rate bursts are carried in slotted-time lightpath at the sub-wavelength level; and they are routed in each core node according to the current time-slot reservation status, incoming wavelength and port. Light-rate streams are carried in lightpath at the wavelength level; and they are routed in each core node according to the incoming wavelength and port.

AR-OBS edge-controller signals for slotted-time lightpaths to transport light-rate bursts of sub-wavelength burst traffic. SS-OS edge-controller signals for proactive periodic scheduling of slotted-time lightpaths to transport light-rate burst train of sub-wavelength stream traffic. Wavelength-routed OCS edge-controller signals for lightpaths to transport light-rate streams. MW-OS edge-controller signals for a lightpath-group to transport multiple light-rate streams with timing relation.

MISON core control can employ any existing decoupled routing, wavelength and time-slot assignment (RWTA) algorithms [20] to select the routing path, the wavelength-channels and the time-slots for a requested slotted-time lightpath. In our simulation network testbed, MISON core control employs a simple RWTA algorithm that is based on open-shortest-path-first (OSPF) routing and first-fit wavelength and time-slot assignment. On the other hand,

MISON core control employs a novel group-routing and wavelength assignment (GRWA) algorithm to select the group of routing paths and the wavelength-channels for a requested lightpath-group with specified end-to-end delay-variation tolerance. The proposed GRWA algorithm will operate as a simpler RWA algorithm when the granularity of a lightpath-group is reduced to a single lightpath.

MISON core control employs a novel unified two-dimensional resource reservation signaling protocol of MG-RFORP to discover/reserve time-slots and wavelength-channels during setup of lightpaths, slotted-time lightpaths and lightpath-groups. Besides advanced periodic reservation and delayed reservation, MG-RFORP also enables reservation blocking recovery controls, which include wavelength-level alternative selection and link-level localized rerouting. In our simulation network testbed, wavelength resources are partitioned between wavelength/multi-wavelength level transport services (wavelength-routed OCS and MW-OS) and sub-wavelength level transport services (SS-OS and AR-OBS); network wavelength resources are completely shared within each partition.

C.1.d. MISON Core Control

The transport services of SS-OBS and AR-OBS are enabled by the sub-wavelength routing control of MISON-RWTA to select resources for slotted-time lightpaths. The transport services of wavelength-routed OCS and MW-OS are enabled by the wavelength/multi-wavelength routing control of MISON-GRWA to select resources for lightpath-groups or single lightpaths. On the other hand, all four transport services are enabled by the unified resource reservation control of MG-RFORP.

C.1.d.1. Sub-wavelength Routing: MISON-RWTA

MISON network architecture does not require FDLs, and assumes wavelength-converters are optional for the all-optical switches. MISON-RWTA takes into account of wavelength-continuity constraint (i.e., sparse or null wavelength-conversion), which restricts a lightpath to be assigned with the same wavelength along path segments where all-optical switches are incapable of wavelength conversion. MISON-RWTA also takes into account of the time-slot continuity constraint (i.e., null FDLs), which restricts a slotted-time lightpath to be assigned end-to-end with the same time-slot.

Fig. 2 shows an example of two-dimensional wavelength and time-slot allocation along two connecting links. Along the x-axis, four wavelengths (λ_1, λ_2, λ_3, λ_4) are multiplexed into one fiber link through WDM. Each wavelength is further divided into time-slots (t_1, t_2, t_3, t_4) as shown along the y-axis. Two traffic flows *A* and *B* occupy one slot each of λ_1. Traffic flow *C* occupies four slots of λ_2. Traffic flow *D* occupies three out of four time-slots of λ_3, while traffic flows *E* and *F* share λ_4.

To reduce computational complexity over large-scale networks, the MISON-RWTA decouples route selection and the assignment of wavelength and time-slots. Route selection is based on OSPF to minimize the end-to-end propagation delay; and simple first-fit is used for wavelength and time-slot assignment.

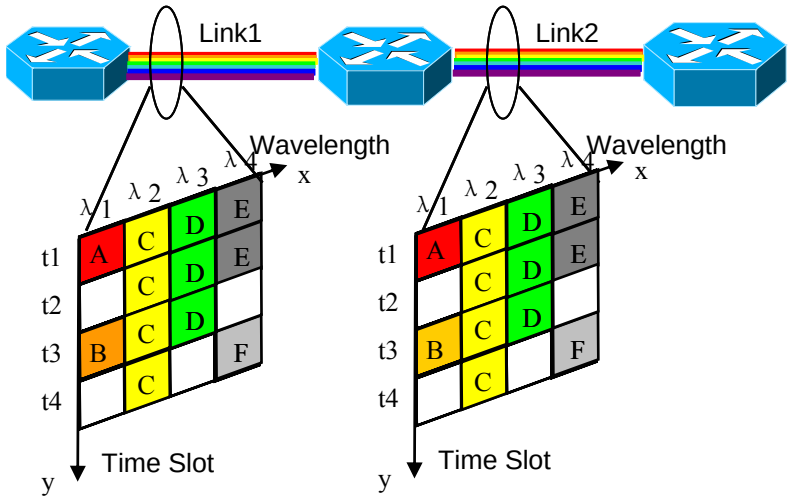


Fig. 2. Two-dimensional Wavelength and Time-slot Allocation

C.1.d.2. Wavelength/Multi-wavelength Routing: MISON-GRWA (ILP formulation)

Multi-wavelength stream traffic is transported through a wavelength-switched lightpath-group with individual lightpaths allowed to take on different routes while satisfying end-to-end delay-variation requirement. The GRWA problem could be formulated as an Integer Linear Programming (ILP) [21] optimization algorithm. The optimization objective would be to select the lightpath-group routes that minimize the weighted cost function in terms of the total number of hops and the total per-hop propagation delays subject to the end-to-end delay variation constraint; and to select the wavelength channels subject to the wavelength-continuity constraint. The ILP optimization problem is formulated as follows.

Notation:

E	set of links;
Z	total number of requests for lightpath-groups;
V	set of nodes;
V_C	set of nodes with wavelength conversion capability;
V_{NC}	set of nodes without wavelength conversion capability; $V_C \cup V_{NC} = V$;

Link(*m*,*n*) a directed link from node *m* to node *n*;

Parameters:

Q^i_k	a binary parameter specified with 1 if the established lightpath-group for request *k* is still active when the new setup request *i* arrives; otherwise it is specified with 0;
R_i	the number of grouped lightpaths needed for the *i*-th request;
s_i	the source node of the *i*-th request;
d_i	the destination node of the *i*-th request;
$C_{m,n}$	the propagation delay of the *link*(*m*,*n*);
M	a route-selection weighting factor that is used to favor routes minimizing total number of hops over routes minimizing end-to-end propagation delay;
T_i	the delay variance requirement for the *i*-th request;

Variables:

$L^{i,j}_{m,n}(\lambda)$	a binary variable that is set to 1 if the wavelength λ of *link*(*m*,*n*) has been reserved for the *j*th path of the requested lightpath-group i , otherwise it is set to 0.

The constraints are listed as follows:

$$\sum_{j=1}^{R_i} L^{i,j}_{m,n}(\lambda) + \sum_{k=1}^{i-1}\sum_{j=1}^{R_i} L^{k,j}_{m,n}(\lambda)\cdot Q^i_k \le 1 \qquad \lambda \in \Lambda\ ,(m,n)\in E\ ,i\in[1,Z] \qquad (1)$$

The link occupation constraint (1) ensures that any wavelength channel within each link can be reserved at most once at any given time. The left hand side sums up the number of successful reservations for wavelength λ of *link*(*m*,*n*) due to the current incoming request i and any past requests with pending releases. Hence this constraint dictates that the summation to be no more than 1.

$$\sum_{m\in V}\sum_{\lambda\in\Lambda} L^{i,j}_{m,s_i}(\lambda) = 0 \qquad j\in[1,R_i], i\in[1,Z] \qquad (2)$$

Constraint (2) ensures that the selected route will not include link leading to the source s_i

$$\sum_{\lambda\in\Lambda}\sum_{m\in V} L^{i,j}_{m,d_i}(\lambda) = 1 \qquad j\in[1,R_i], i\in[1,Z] \qquad (3)$$

Constraint (3) ensures that the destination is reached once by each individual lightpath of the *i*-th request

$$\sum_{m\in V} L^{i,j}_{m,n}(\lambda) = \sum_{m\in V} L^{i,j}_{n,m}(\lambda) \qquad n\in V_{NC}, n\ne s_i, n\ne d_i\ ,\ \lambda\in\Lambda\ , i\in[1,Z] \qquad (4)$$

$$\sum_{\lambda\in\Lambda}\sum_{m\in V} L^{i,j}_{m,n}(\lambda) = \sum_{\lambda\in\Lambda}\sum_{m\in V} L^{i,j}_{n,m}(\lambda) \qquad n\in V_C, n\ne s_i, n\ne d_i\ ,\ i\in[1,Z] \qquad (5)$$

Constraint (4) and (5) are the flow constraints for wavelength convertible and non-convertible intermediate nodes respectively. Constraint (4) ensures that the total numbers of lightpaths terminated at the input and output ports of a

wavelength-convertible node are equal. Constraint (5) ensures that the numbers of lightpaths with a given wavelength λ terminated at the input and output ports of a non-convertible node are equal.

$$\sum_{\lambda\in\Lambda} L^{i,j}_{m_1,m_2}(\lambda) + \sum_{\lambda\in\Lambda} L^{i,j}_{m_2,m_3}(\lambda) + \ldots + \sum_{\lambda\in\Lambda} L^{i,j}_{m_p,m_1}(\lambda) < p \qquad p = 1,2,\ldots|V|, \forall(m,n)\in E, \forall m, n \neq s_i, \forall m, n \neq d_i \qquad (6)$$

Constraint (6) prevents looping during route selection by ensuring that the selected route will not include all the connecting wavelength channels of any *p*-hop loop.

$$\max_j\{\sum_{(m,n)\in E}\sum_{\lambda\in\Lambda} L^{i,j}_{m,n}(\lambda)\cdot C_{m,n}\} - \min_j\{\sum_{(m,n)\in E}\sum_{\lambda\in\Lambda} L^{i,j}_{m,n}(\lambda)\cdot C_{m,n}\} \le T_i \qquad i\in[1,Z] \qquad (7)$$

Constraint (7) is the lightpath-group delay variation constraint, which ensures that the maximum delay variance among individual lightpaths of the requested lightpath-group is bounded by the specified delay variance T_i.

The optimization function is formulated as:

$$Minimize:\ M\cdot\sum_{j=1}^{R_i}\sum_{(m,n)\in E}\sum_{\lambda\in\Lambda} L^{i,j}_{m,n}(\lambda) + \sum_{j=1}^{R_i}\sum_{(m,n)\in E}\sum_{\lambda\in\Lambda} L^{i,j}_{m,n}(\lambda)\cdot C_{m,n}$$

The optimal lightpath-group minimize the cost as a weighted function of the total number of hops and the total per-hop propagation delays subject to the end-to-end delay variation and the wavelength-continuity constraints. Once the routes and wavelength for a lightpath-group are selected, the edge node initiates reservation signaling for the lightpath-group.

C.1.d.3. Wavelength/Multi-wavelength Routing: MISON-GRWA (Heuristic Algorithm)

The optimal GRWA ILP algorithm computes the optimal solution of assigning the routes and wavelengths to a lightpath-group for a given network topology and traffic loading condition. The computation complexity increases exponentially with the number of wavelengths and links. The MISON-GRWA heuristic algorithm is proposed to avoid the high computation complexity. The heuristic always tries to minimize the blocking probability of future lightpath-group setup requests by minimizing the total number of wavelength channels allocated for a requested lightpath-group. The procedures of the heuristic are described as follows:

Step 1: Find the set of *k* shortest routes (in terms of number of hops) for the given source-destination pair.
Step 2: Generate route-groups out of the *k* routes to satisfy the end-to-end delay variation requirement. The maximum difference in end-to-end delays within each route-group should be less than the end-to-end delay variation requirement.
Step 3: Select the routes within the route-group that minimize total the number of wavelength channels to setup the requested lightpath-group, providing that wavelength resources are available along those routes.

The value *k* determines the heuristic performance and the computation complexity. In the extreme case when *k* approaches the total number of possible routes for the given source-destination pair, the heuristic performance will approach to that of the ILP.

C.1.d.4. Multi-granular Resource Reservation

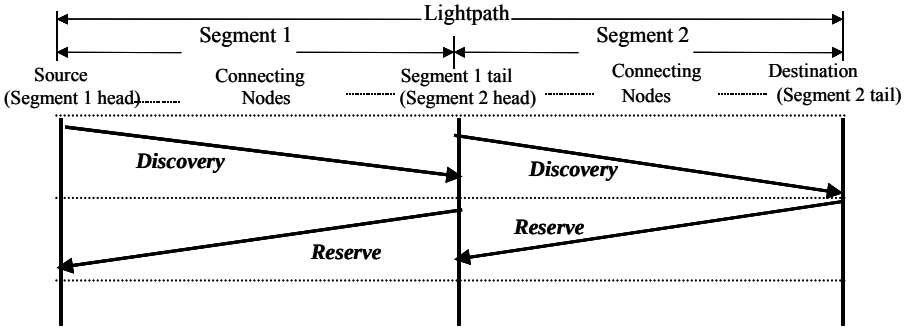
The MG-RFORP reservation protocol supports two-dimensional (wavelength and time-slot) resource discovery and reservation controls. MG-RFORP divides each lightpath into one or more wavelength-constrained segments. Each segment begins at a wavelength convertible nodes (segment *Head*) and ends at the next wavelength convertible nodes (segment *Tail*). Within each segment, wavelength continuity constraint demands the same wavelength to be assigned to each optical link since none of the intermediate nodes has wavelength conversion capability. MG-RFORP employs two-phase reservation procedure including wavelength discovery phase and wavelength reservation phase. At segment level, parallel signaling operations are employed in both discovery and reservation phases to minimize the lightpath setup delay. Within each segment, serialized link-based reservation and discovery are adopted.

Resource Discovery and Reservation

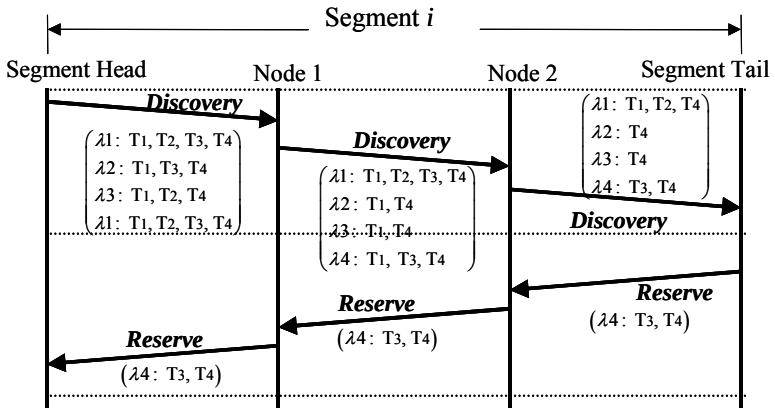
Each lightpath setup request specifies the required length of the time interval based on the size of data payload. The time interval contains one or more time-slots. In the resource discovery phase, source node sends out the resource probe requests to each segment head simultaneously. Link-based serialized resource discovery is performed within

each segment. Each node gets the resource information from upstream node, AND the resource information with its own resource information, and pass the new information to downstream node. Once the information reaches the segment tail, the tail node passes the information back to the destination. The destination node computes the available wavelengths and time intervals based on the wavelength and time-slot continuity constraint; and then it triggers the reservation phase. In the reservation phase, each segment tail gets the reservation request from the destination containing the assigned wavelength and time interval. Link-based serialized resource reservation is performed within the segment.

Fig. 3 shows the discovery and reservation signaling procedure. Fig. 3(a) shows that in the discovery phase, the source node notifies the head of each segment to initiate parallel segment-based discovery processes. Similarly, in the reservation phase, the destination node notifies the tail node of each segment to initiate parallel segment-based reservation of the selected wavelengths. Fig. 3(b) shows the discovery and reservation signaling within a segment. The two-dimensional resource available information is shown along each link. At the first link, time-slots T1 to T4 are available for wavelength λ1, T1, T3 and T4 are available for λ2, T1 T2 T4 are available for λ3 and all four time-slots are available for λ4. Then the resource information is passed to the second link, and second link AND this information with its own resource information to form the new common resource available information. When the



(a) Parallel Segment-based Discovery and Reservation



(b) Serialized link-based Discovery and Reservation

Fig. 3. Discovery and Reservation Signaling Procedure

message reaches the segment tail, the discovered common resource information for segment *i* is T1 T2 T4 for λ1, T4 for λ2 and λ3, T3 T4 for λ4. Then the tail node assigns the wavelength and time interval according to the request requirement. In this example λ4 for T3 T4 is selected, and segment tail initiate the reservation along the links.

Reservation Blocking Recovery

During the reservation phase, blocking occurs when the discovered time-slots for a particular wavelength have been allocated for other request. Reservation blocking will trigger recovery controls include link-level localized rerouting and wavelength-level alternative wavelength selection. Alternate wavelength selection will choose the same relative

time-slots to satisfy the time continuity constraint. However, localized rerouting will change the length of the route, and then break the time continuity constraint. Thus, localized rerouting is only employed by wavelength-routed OCS and MW-OS, where no time-slot assignment is considered. If all the recovery attempts fail, the failure is reported back to the source node.

Fig. 4(a) illustrates a scenario of reservation blocking recovery via *alternate wavelength selection* in segment *i*. Node-H is the head node, and node-T is the tail node. The discovered common wavelengths for segment *i* are λ1 and λ4 for time interval T. Node-1 tried to reserve λ1 for link 1-2, and found that λ1 was unavailable for that link since another connection had already reserved it. Node-1 reported the blocking failure to the tail node-T via Reserve_FAIL message. The tail node then selects λ4 as the alternate wavelength, and initiates reservation for the segment by repeating the same procedure as normal reservation. If the candidate wavelengths are exhausted, the node-T will initiate the second blocking recovery process of localized rerouting.

Fig. 4(b) illustrates a scenario of reservation blocking recovery via *localized rerouting* in segment *i*. A reservation failure occurs at the link between node 1 and 2 when the tail initiate the Reserve_LT to reserve a discovered wavelength λ3 with localized rerouting recovery. The failure is reported to Node T, and a detour from Node1-Node3-Node2 is tried as localized rerouting. If all the recovery attempts fail, failure status is signaled to both source

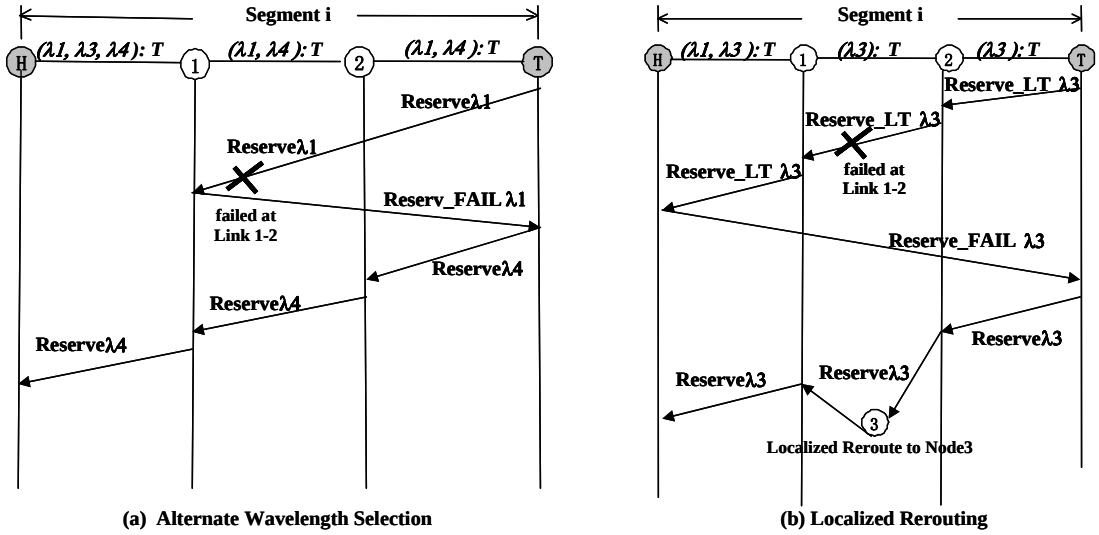


Fig. 4. MG-RFORP Reservation Recovery Control

and destination, and reserved wavelength is released.

C.1.e. MISION Edge Control

As illustrated in Fig. 5, MISON edge node consists of four service-specific edge control blocks, admission control module, traffic monitor, and MG-RFORP signaling module. The admission control module deals with on-user-demand requests for connection-oriented provisioning of wavelength and multi-wavelength stream traffic. These requests are processed and mapped into detailed wavelength reservation requirements. Then, the admission control module informs MG-RFORP to setup OCS or MW-OS based connection. If the connection is successfully established, admission control will inform user application to start data transmission, otherwise the request is rejected. Once the stream traffic is admitted, the incoming data is passed to corresponding edge control module.

SS-OS and AR-OBS deal with connectionless sub-wavelength traffic. A traffic monitor routes the incoming traffic to corresponding edge control modules according to the traffic characteristics. The stream traffic is characterized by long duration flow, while the burst traffic is characterized by short-term traffic of variable bit rate. Once a traffic stream is detected, traffic monitor will route the data to SS-OS edge control, otherwise the data is routed to AR-OBS edge control. While the burst train is being shaped at the SS-OS edge buffer, signaling request for sub-wavelength stream transport is sent to MG-RFORP module to setup cyclic slotted-time lightpaths. While a data burst is being aggregated at the AR-OBS edge buffer, signaling request for sub-wavelength burst transport is sent to MG-RFORP

module to setup a slotted-time lightpath. The detailed discussion of AR-OBS edge control is presented in 4.A. The edge controls of SS-OS and MW-OS are presented in 4.B and 4.C respectively. The OCS edge control is not discussed in this paper.

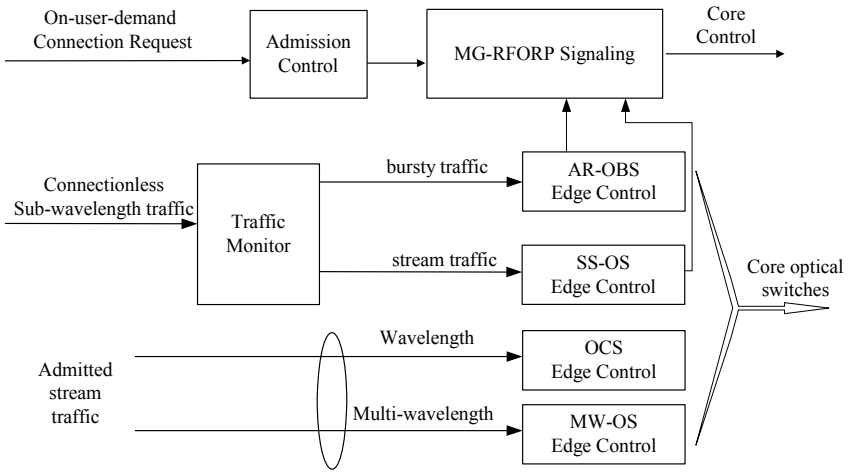


Fig. 5. MISON Edge Node Control Structure

C.1.e.1. Edge control for AR-OBS

Fig. 6 shows the edge control structure in AR-OBS, which provides burst scheduling and offset delay controls. Incoming data is aggregated into different buffers based on its destination via the Destination Classifier. To reduce the associated signaling and switching overhead of each optical burst that reduces the wavelength utilization, we always try to generate an optical burst as large as possible.

To avoid edge buffer overflow and aggregated data timeout, the Estimator monitors the data-incoming rate of each Buffer, and predicts the burst transmitting time of each Buffer. It will inform the Scheduler to trigger lightpath setup signaling for an optical burst via Lightpath Setup module. When it is time for the burst to be sent out, the Scheduler triggers the Burst Generator to dump the data from that buffer into the optical core network as an optical burst. Meanwhile, the buffer is emptied in order to accept new incoming data.

The bursty traffic can be classified into two types based on the ratio of data incoming rate to available edge buffer size: fast bursty traffic and slow bursty traffic. For fast bursty traffic, the data incoming rate is high and the edge buffer will be filled up before any of the aggregated data expires. The optical burst should be scheduled to avoid buffer overflow. The Estimator dynamically predicts the average data incoming rate in the near future based on history data. According to the estimated incoming rate, when the edge buffer will be filled up can be predicted, and that instant is defined as buffer full time. The burst should be sent out at the buffer full time to maximize the wavelength utilization and avoid the data dropping caused by buffer overflow. Lightpath signaling procedure will be triggered in advance based on the predicted buffer full time. In the signaling command, the lightpath setup time is set to be the predicted buffer full time, and the burst transmission time is fixed as B_{size}/B_{core}, where B_{size} and B_{core} are the buffer size and core network bandwidth respectively.

For the slow bursty traffic, the data incoming rate is slow and some of the aggregated data will expire before the buffer is filled up. The optical burst should be scheduled to avoid data expiration at the edge buffer. The signaling procedure is triggered based on the earliest expiring time of any buffered data. Based on the estimated traffic rate, the aggregated data size at the scheduled burst sending time can be predicted. In the signaling command, the wavelength holding time is set according to predicted burst size, and lightpath setup time is set to be the earliest data expiring time.

The Estimator in AR-OBS keeps monitoring the data-incoming rate and updating the burst transmitting time until the lightpath signaling has to be sent out. If the estimator is not perfect, mismatch or estimation error always occurs. Some bandwidth is wasted since the buffer is not fully loaded if the data incoming rate is overestimated. On the other hand, some of the data will be lost due to buffer overflow if the data incoming rate is underestimated. For our implemented estimator, the estimation error decreases if the estimator can use the more recent history data. Thus, for a given scheduling criteria, a smaller deviation of the estimation error requires the AR-OBS utilizes the history data

as fresh as possible.

The offset delay is defined as the time between signaling for a lightpath setup to transport an aggregating burst and sending out the aggregated burst. The offset delay should be equal to or larger than the signaling delay to ensure the acknowledgement returns before the burst transmission time. Since the built-in recovery mechanism of MG-RFORP may bring additional delay to the total signaling delay, the offset delay should also take account of such potential delay. For a given route, the maximum additional delay due to each trial of blocking recovery is a deterministic value. We call the blocking recovery trials to be blocking recovery retries.

Increasing offset delay for a given route allows longer time for lightpath signaling (more blocking recovery retries are allowed) and may decrease the blocking probability. However, a larger offset delay prevents the estimator to utilize more recent history data for its estimation and results in larger estimation error, which deteriorates the performance of wavelength utilization or increases edge overflow blocking. In order to reduce estimation error, offset delay should be as small as possible. Therefore, offset delay is set to allow just enough blocking recovery retries to satisfy the reservation-blocking requirement for a given route. Based on the status of core network, edge

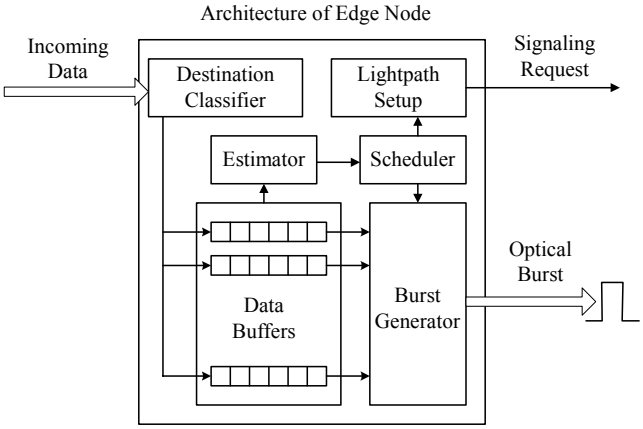


Fig. 6. Edge Control of AR-OBS

node determines the number of blocking recovery retries allowed for a route, and sends out such information along the signaling message.

C.1.e.2. Edge control for SS-OS

Fig. 8 shows the edge control structure of SS-OS. Data packets are buffered in the stream buffers. Upon successful bandwidth reservation with light-path scheduler, all data belonging to that traffic stream in the buffer will be shaped into a periodic burst train. Since periodic time-slots are reserved in the network core for each incoming stream, data bursts are dumped to the core network at assigned time-slot/time-slots periodically. To reduce the computational complexity, SS-OS divides the time scale into fixed size time frames (scheduling frame) and shapes the aggregated data into one or more bursts for each frame. The burst size of different streams can be different due to different bandwidth requirements.

Each sub-wavelength constant-bit-rate (CBR) traffic stream can be shaped into a periodic burst train with fixed-size bursts. For variable-bit-rate (VBR) traffic streams, in order to take the advantage of periodic scheduler, the traffic stream will also be supported with a periodic fixed-sized burst train. Note that in the VBR case, the actual data aggregation size may be larger or less than the specified reserved burst capacity. Therefore the appropriate burst capacity based on the traffic profile could greatly improve the performance. If the reserved burst capacity is too large, there would be too much vacancy in the reserved bandwidth; if the reserved burst capacity is too small, a lot of packets would be dropped due to insufficient reserved burst capacity. To decrease the data loss rate due to insufficient reservation, two traffic streams can statistically multiplex the reserved burst capacity if they share the same source-destination. This implies that multiple traffic streams with the same source and destination could share their reserved bandwidth.

Fig. 8. gives a simple example of how SS-OS supports sub-wavelength CBR and VBR streaming traffic. As illustrated in the figure, three traffic streams are sharing one wavelength along the same route. A is a CBR stream; B and C are VBR streams. Suppose the burst trains for traffic A, B and C are reserved with the same capacity R*T, which is equal to the data aggregated in one time frame of length T with the average data incoming rate of R. Traffic A has a data rate of R, so it will use up its the reserved burst capacity; while B and C have different data rate in these three time frames. In time frame 1, the aggregate of traffic B exceeds its reserved burst capacity and results in traffic dropping. In time frame 2, the aggregated data of traffic B does not use up its reserved burst capacity since its data rate is less than R. The unused capacity for traffic B will be shared by the aggregated data from traffic C. In time frame 3, neither traffic B nor traffic C use up the corresponding reserved capacity. In order to fill up the reserved burst capacity, padding will be added. In this case, there will be some bandwidth waste in the reserved bursts.

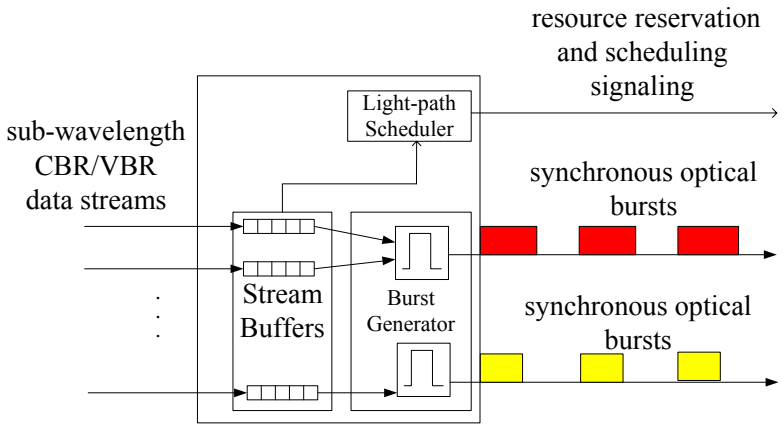


Fig. 8. Edge Control of SS-OS

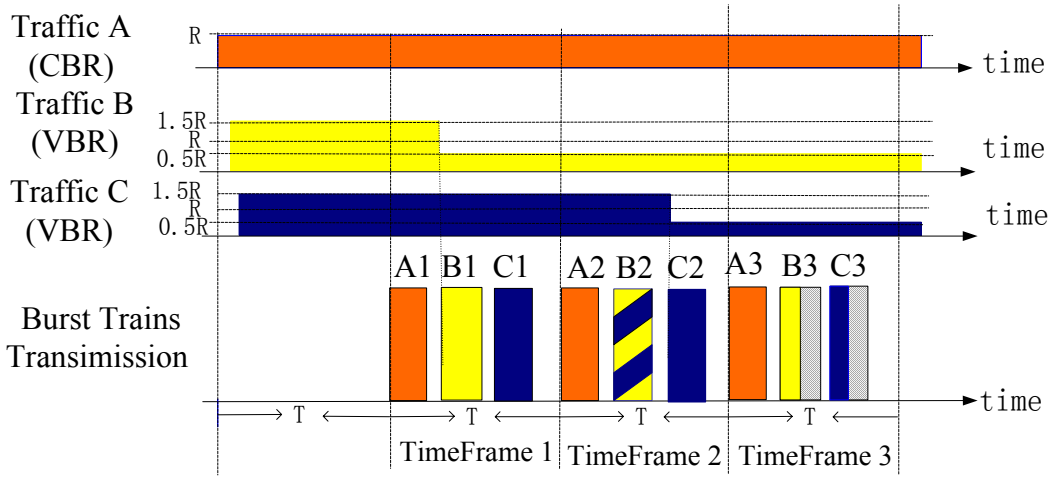


Fig. 7. TDM and Statistical multiplexing of SS-OS

C.1.e.3. Edge control for MW-OS

The input data of MW-OS is ultra-high bandwidth traffic flow that needs more than one wavelength to support its transmission in the core optical networks. Theoretically, the data flow is inversely multiplexed into multiple wavelengths that generally have the restriction of end-to-end delay variations among the grouped wavelengths. Fig. 9(a) illustrates the edge control functions of MW-OS in theoretical case. Inverse-multiplexer breaks down an ultra-high-bandwidth traffic into a group of wavelength stream traffic that can be dispersed over a range of channels to be carried over the network. However, current applications usually cannot provide the ultra-high bandwidth traffic flows such as terabits/sec that need to be supported by multi-wavelengths in the core networks.

A practical application of inverse multiplexing is shown in Fig. 9(b). Considering a scenario of high performance computing (HPC), the cluster accepts logic input from the applications and generates multi-wavelengths traffic to be sent to other clusters. Thus, the cluster logically inverse multiplexes the input to multi-wavelengths. If the cluster is treated as a part of edge control of MW-OS, the cluster itself serves as the inverse multiplexer for input traffics, though the input is not an ultra-high bandwidth one.

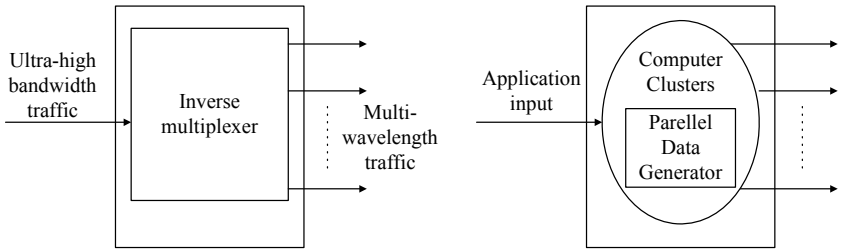


Fig. 9. Edge control of MW-OS

C.1.f. Simulation Results

In this section we present our simulation experiments and results. The simulation is conducted in our own developed simulator written in C. The performance is evaluated for four traffic types including sub-wavelength stream, sub-wavelength bursty, wavelength stream and multi-wavelength stream traffic respectively. Simulation results show that our proposed SS-OS and AR-OBS could support sub-wavelength traffic efficiently. Their performances are compared to JET-OBS [6] and OCS in terms of data loss rate and wavelength utilization regardless of the wavelength convertibility ratio of the network. The blocking probability of our proposed heuristic of MW-OS is evaluated under the NSF-14 node and 7-node topologies, and compared to the numerical results of the ILP formulation under the simplified 7-node topology.

The network used for simulation is the NSF 14-node as shown in Fig. 10. The link propagation delay is shown along the link in the unit of millisecond. The parameters of the topology model are: number of links E=42 (bi-direction links assumed); number of wavelength per link W=4. The normalized traffic load per input wavelength and per link shown in the following figures is defined as:

$$TrafficLoad = \frac{R_{data} * H_{avg}}{|E| * W * B_{core}}$$

where R_{data} denotes the total data incoming rate to the network. For bursty traffic, R_{data} is calculated based on the total number of incoming packets per sec and the average packet size. For streaming traffic, R_{data} is defined as

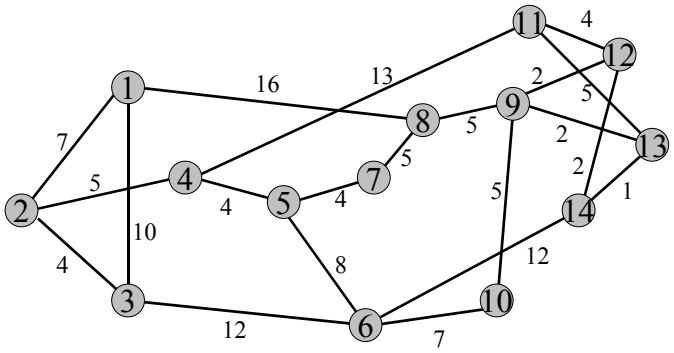


Fig. 10. 14-node NSF network topology

$R_{data} = R_{arrival} * T_{WHT} * B_{request}$. $R_{arrival}$ is the average traffic arrival rate, which is defined as the number of

incoming traffic streams per sec. T_{WHT} is the average holding time for incoming traffic streams. $B_{request}$ is the average bandwidth of traffic streams. H_{avg} is the average number of hops per route. Assuming the traffic is even distributed among edge nodes and shortest path routing is adopted, H_{avg} is set to be 2.2 for NSF-14 topology. B_{core} is the core bandwidth per wavelength.

The simulation assumptions are: 1Gbps bandwidth per wavelength; each node has a 8 ms processing delay for wavelength discovery or reservation; switching delay on each optical switch is set to be 3ms (The mechanic-optical switch usually has switching delay between 1~5 ms).

Fixed routing is assumed for SS-OS, AR-OBS and OCS. Assume no FDL in core network for JET-OBS. RSVP-TE is implemented as the distributed signaling protocol in traditional OCS networks, which is denoted as OCS (RSVP). The performance is evaluated for three different wavelength convertibility ratios of the core network as NULL conversion, SPARSE conversion with 50% wavelength convertibility (Node 1,3,5,7,9,11, 13 are wavelength convertible nodes) and FULL conversion. In the simulation, all presented data points are averaged over 100 runs, so that they have a 95% confidence interval with 10% (or better) precision based on student's t-distribution. The 95% confidence intervals are shown in the figures for all sample points in Fig. 12,13,14,16 in the following sections.

C.1.f.1. Sub-wavelength stream traffic

Simulations are conducted for both constant bit rate (CBR) and variable bit rate (VBR) stream traffic. The traffic arrival rate and holding time of incoming streams are Poisson distributed. In our simulation, $R_{arrival}$ is set within the range of 2~4 request/sec, and the average holding time is 30 sec. Within each traffic stream, packets are fixed sized at 10Kbit. For CBR traffic, assume each traffic stream requests a deterministic bandwidth of 125Mbps. For VBR traffic, assume traffic follows exponential ON/OFF distribution with average bandwidth of 125Mbps. Packets are sent at a fixed rate during ON periods, and no packet is sent during OFF periods. Both ON and OFF periods are taken from an exponential distribution. The length of scheduling frame T is set to be 300 ms.

Fig. 11 and Fig. 12 present the effect of traffic load on data loss rate of CBR and VBR traffic respectively in different wavelength convertibility scenarios. As expected, the performance of JET-OBS deteriorates greatly as wavelength convertibility ratio decreases, since JET-OBS relies on wavelength converter to solve wavelength reservation contention. On the other hand, SS-OS prevents such contention by employing two-way reservation. Therefore, it is observed that SS-OS can achieve satisfactory results regardless of wavelength convertibility ratio. In the NULL and SPARSE wavelength conversion scenarios, SS-OS outperforms JET-OBS because the number of wavelength convertible nodes is not enough. In the FULL conversion scenario, JET-OBS gives the best performance. It is observed that data loss rate is reduced greatly compared to (RSVP) under medium traffic load by employing sub-wavelength transport services JET-OBS and SS-OS, since OCS cannot support sub-wavelength traffic efficiently.

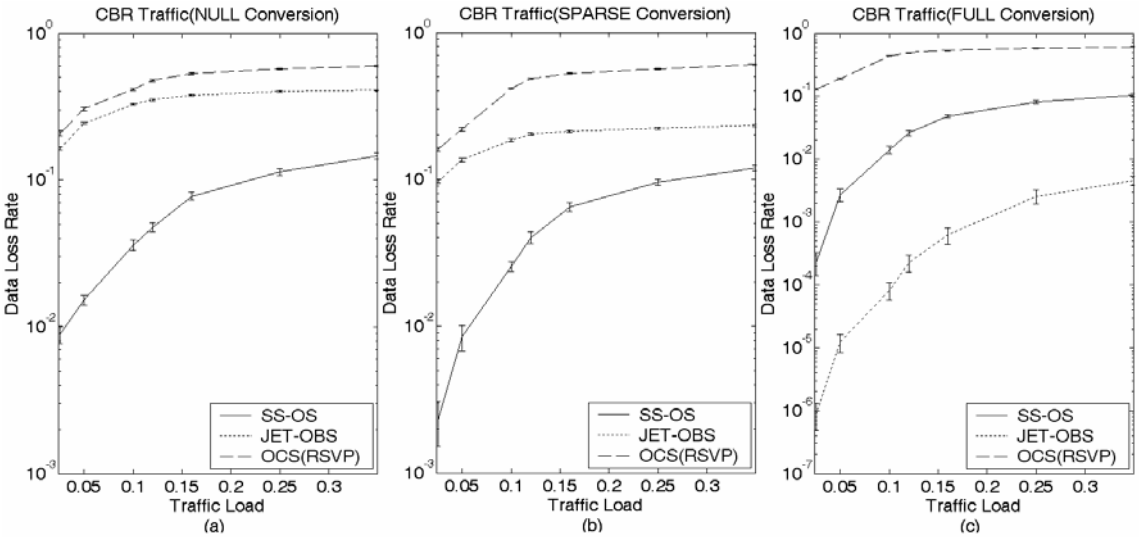


Fig. 11 Effect of Traffic Load on Data Loss Rate (Sub-Wavelength CBR Stream Traffic)

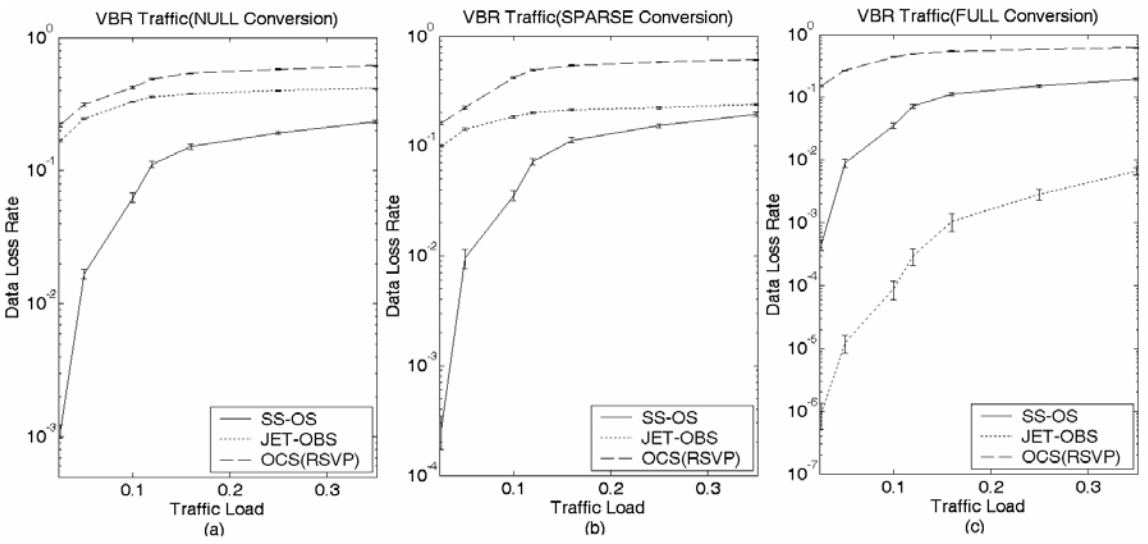


Fig. 12 Effect of Traffic Load on Data Loss Rate (Sub-Wavelength VBR Stream Traffic)

In addition, SS-OS will suffer from the fact that the aggregated data may be higher than the reserved burst size in the VBR case. Therefore, the degradation of SS-OS from the CBR to VBR case is higher than it of JET-OBS. However, due to the fact that the performance of JET-OBS is greatly affected by the wavelength convertibility ratio, SS-OS still outperforms JET-OBS in the NULL and SPARSE conversion scenarios.

Fig. 13 shows how optical switching controls affect the wavelength utilization, which is defined as the ratio of the time used to transmit the payload to the time reserved for the whole burst (switching time is counted in) at core switches. For CBR traffic, since SS-OS reserves exact bandwidth according to the data rate, SS-OS achieves comparable wavelength utilization as the JET-OBS as shown in Fig. 13(a). For VBR traffic (Fig. 13(b)), SS-OS has to reserve more time-slots in each time frame to accommodate the variable bit stream, so the wavelength utilization is degraded about 30% compared to CBR traffic. In both cases, the wavelength utilization of OCS (RSVP) is only around 10% of SS-OS and JET-OBS.

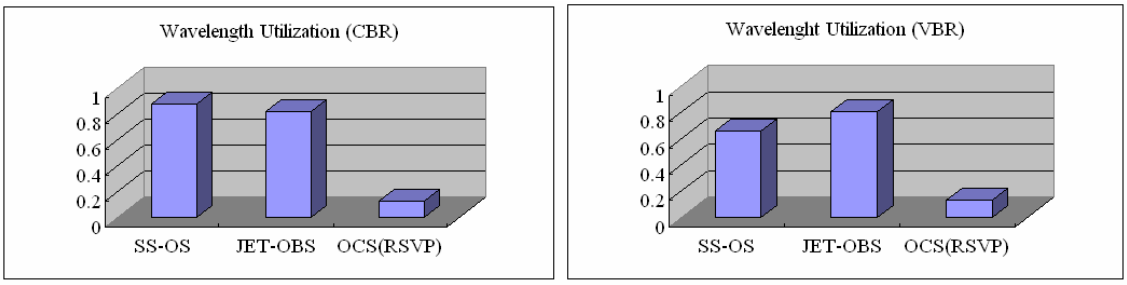


Fig. 13 Wavelength Utilization for Sub-Wavelength Stream Traffic

C.1.f.2. Sub-wavelength Bursty traffic

For sub-wavelength bursty traffic, we assume both the packet size and packet inter-arrival time follow Pareto distribution with shape parameter 1.5. Average packet size is 100Kbit, and the edge buffer size is 10Mbit. For AR-OBS, we use exponential weighted moving average estimator to estimate the data incoming rate.

Fig. 14 illustrates the effect of traffic load on data loss rate with different wavelength convertibility. As shown in Fig. 14 (a-b), JET-OBS performs worse than AR-OBS when the network is not a fully wavelength convertible one, since JET-OBS implements one-way reservation which brings the random burst contention blocking. When the network is full wavelength convertible AR-OBS achieves comparable performance as the JET-OBS. We also evaluate the data loss rate on OCS (RSVP) network as a reference. For OCS, we assume maximum static

connections are established between edge nodes, and data is transmitted through these connections. The wavelength utilization is shown in Fig. 15. AR-OBS and JET-OBS achieve comparable wavelength utilization due to the employed delayed reservation. The difference between AR-OBS and JET-OBS is less than 5%. The wavelength utilization of OCS is less than 15% of AR-OBS and JET-OBS.

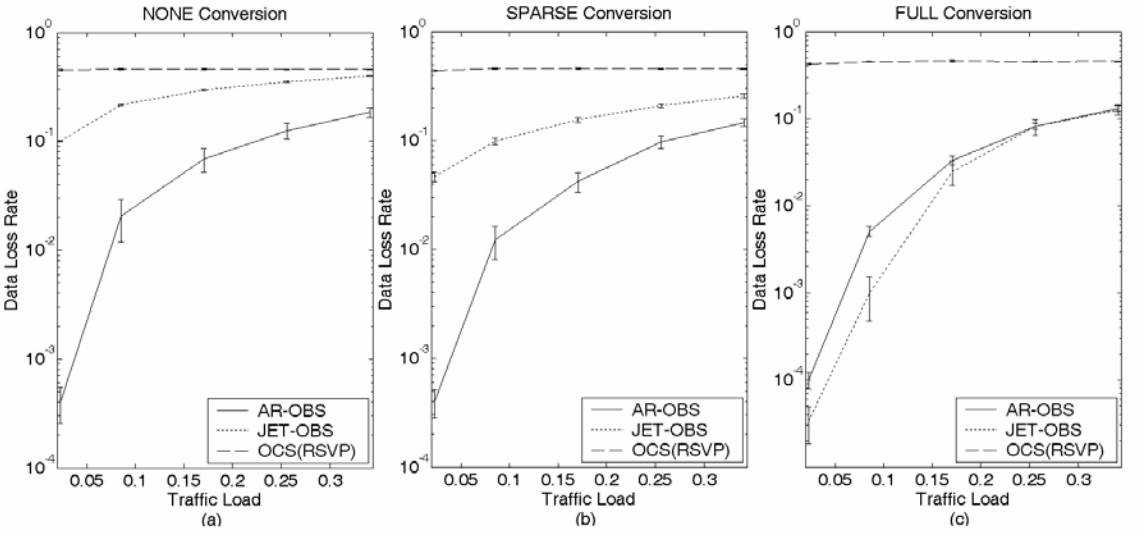


Fig. 14 Effect of Traffic Load on Data Loss Rate (Sub-wavelength Bursty Traffic)

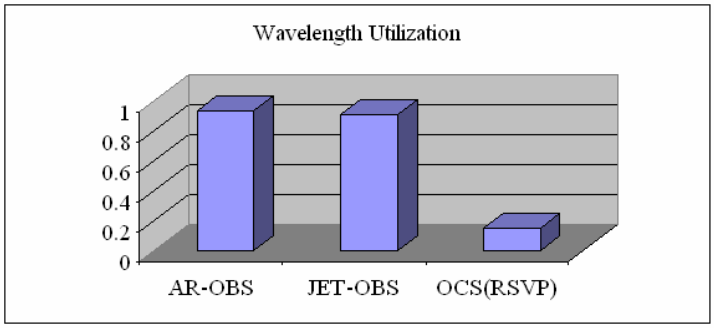


Fig. 15 Wavelength Utilization for Sub-wavelength Burst Traffic

C.1.f.3. Multi-Wavelength Stream Traffic

Due to computation complexity, the MW-OS ILP algorithm is executed over a simple 7-node topology model as shown in Fig. 16. The heuristic algorithm is executed over the simple 7-node and the NSF 14-node topology models. The 7-node topology model consists of two wavelength-convertible nodes (nodes 2 and 6). The propagation delay of each link is shown along the links (in millisecond unit). For each source-destination pair, the input traffic arrivals are Poisson distributed with holding time equals to twice the inter-arrival time. The number of lightpaths required by each requested lightpath-group is chosen randomly from 1 to 4 following a uniform distribution. A request will be blocked if not enough wavelength channel resources are available or the delay variation requirement cannot be met.

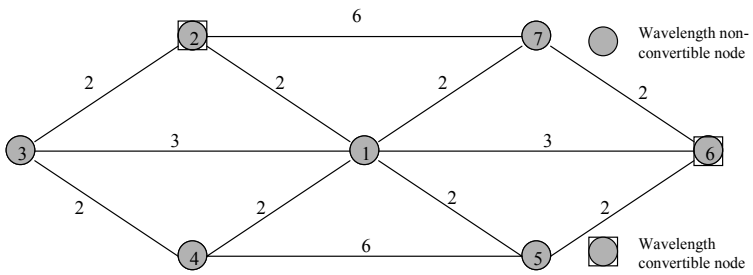


Fig. 16 Simulation topology for ILP formulation

Fig. 17 (a) shows the effect of delay variation tolerance on request blocking probability under different wavelength capacity per link for both ILP and heuristic algorithms. The number of candidate routes for each source and destination pair is set up to 15 for the heuristic algorithm. The blocking probability decreases as the delay variation tolerance increases for both ILP and heuristic algorithms. Larger delay variation tolerance allows more choices of potential routes for the requested lightpath-group. For a given delay variation tolerance, the blocking probability decreases as the wavelength capacity increases. By increasing the wavelength capacity from 1 to 2 and 3, the blocking probability decreases by more than 50% and 80% respectively.

Fig. 17 (b) shows the effect of delay variation tolerance on request blocking probability under different number of candidate routes *k* for the heuristic algorithm. The wavelength capacity is fixed at 3. As *k* increases, the blocking probability approaches to that of the ILP algorithm. By increasing *k* from 2 to 5 and 10, the blocking probability decreases by more than 50% and 75% respectively. When *k* increases to 15, the difference in results between the ILP and heuristic algorithms is negligible.

The heuristic algorithm is executed over the NSF 14-node topology model (**Fig. 10**). The system parameters are configured similarly as the simple 7-node topology model; except that the number of lightpaths required by each requested lightpath-group is chosen randomly from 1 to 6 instead of 4. Fig. 18(a) shows the effect of delay variation tolerance on request blocking probability under different wavelength capacity for the heuristic algorithm. As expected, the blocking probability decreases as the delay variation tolerance increases. For a given delay variation tolerance, the blocking probability decreases by more than 25%, 50% and 65% respectively by increasing the wavelength capacity from 1 to 2, 3 and 4.

Fig. 18(b) shows the effect of delay variation tolerance on request blocking probability under different number of candidate routes *k* for the heuristic algorithm. The wavelength capacity is fixed at 3. By increasing *k* from 5 to 10, 15and 20, the blocking probability decreases respectively by about 10%, 13% and 15%. The decrease in blocking probability slows down as *k* reaches over 10.

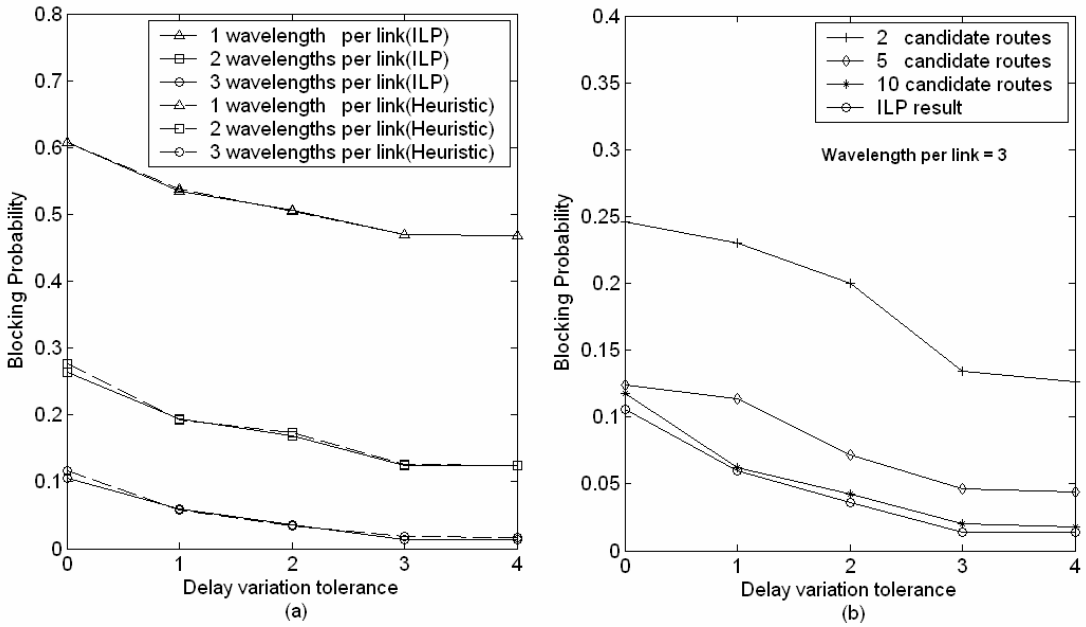


Fig. 17 The Effect of Delay Variation Tolerance on Blocking Probability (7 Nodes Topology)

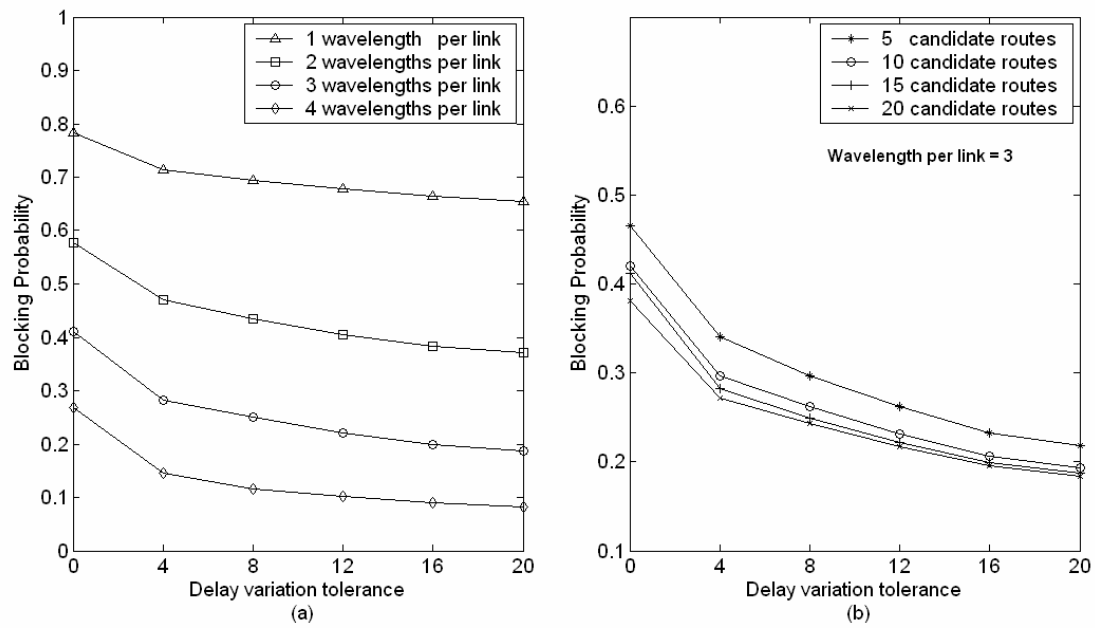


Fig. 18 The Effect of Delay variation Tolerance on Blocking Probability (NSF 14 Topology)

C.2. Secure Photonic Interdomain Negotiator (SPIN)

The Secure Photonic Interdomain Negotiator (SPIN) enables a secure multi-domain control plane to setup dynamic lightpaths on demand between the source and destination domains over a multi-domain optical grid. As illustrated in Fig. 19, each SPIN-based domain is enabled by a Control Server that provides interdomain traffic engineering functions, and a Security Server that provides interdomain security management. Each Control Server consists of a routing controller for topology discovery; and a signaling controller for lightpath setup and wavelength reservation. Each domain defines security policies for user authentication, service authorization and network resource access. The Security Server maintains the policies database while the Control Server enforces these policies, and controls lightpath management in a multi-domain environment.

C.2.a. Interdomain Security Management

SPIN interdomain security management mechanism enables secure communication and resources access in multi-domain Lambda systems. Distributed SPIN Security Servers implement AAA-based secure access policies. A security policy defines allowable operation and access privileges for given systems entities (i.e., users, services and resources). Each domain can define its own local security policies, while multiple domains can negotiate and agree on federation security policies. For each domain, the SPIN Security Server acts as the policy decision point (PDP) where policy decisions are made, while the SPIN Control Server acts as the policy enforcement point (PEP) where policy decisions are actually enforced. Each Security Server maintains the policies database to enable secure communication and access of distributed Grid computing and optical networking resources in a multi-domain environment. In accordance with the security policies, each Control Server enforces the policies and invokes the corresponding intra-domain control plane services for lightpath setup and maintenance.

For security policy relating to optical networking resource access, the Security Server would translate the policy into an interdomain routing or signaling policy. For example, a Lambda Grid domain may specify that it is unavailable at certain times as a connecting domain so as to avoid interfering performance-sensitive applications scheduled during those times. The associated Security Server would map this security policy into an interdomain routing policy that prohibits any multi-domain lightpath to transit through this domain. The corresponding Control Server would instruct its routing controller to enforce this routing policy by advertising that all other domains could not be reachable via this domain.

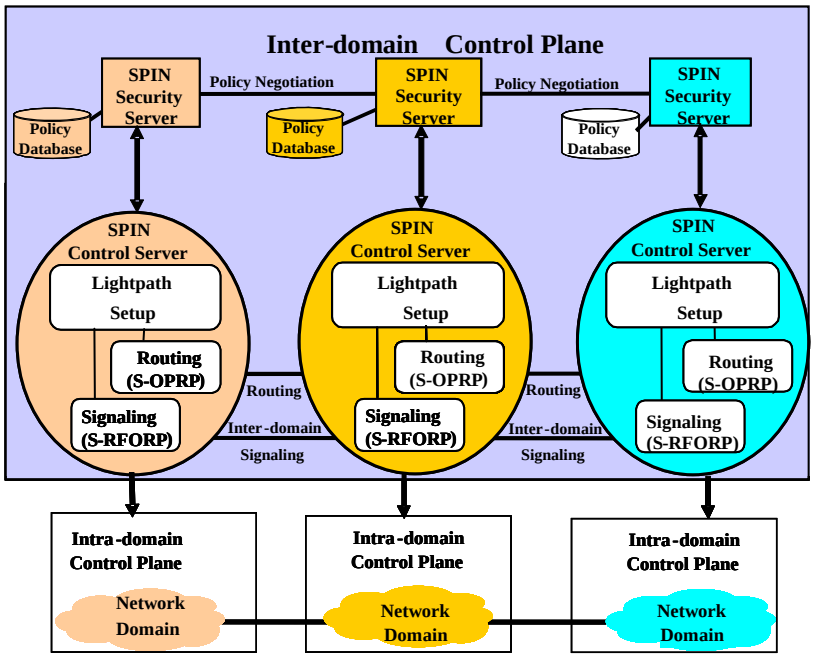


Fig. 19. SPIN Architecture

C.2.b. Interdomain Routing and Wavelength Assignment

SPIN interdomain routing and wavelength assignment mechanism enables robust dynamic lightpath setup in multi-domain Lambda systems. It includes topology and wavelength state information discovery, route selection, wavelength assignment and reservation. Current wide-area multi-domain optical network architecture is characterized with border nodes employing optical-electrical-optical (OEO) switches for long distance interdomain transmission, and core nodes employing either OEO switches or photonic all-optical switches for protocol-transparent transmission. While OEO switches inherently support wavelength conversion, current all-optical switches are incapable of that. Consequently, wavelength-continuity constraint (portions of lightpath have to be assigned with the same wavelength) must be applied if all-optical switches are deployed in a multi-domain optical network.

SPIN specifies a hierarchical interdomain routing and wavelength assignment protocol that takes into account of wavelength continuity constraint. As shown in **Fig. 20**, SPIN maps the network into two hierarchical layers of routing areas and wavelength-constrained segments. The lower routing layer consists of independent local routing areas of the domains. SPIN allows each domain to independently employ proprietary routing protocols or to adopt IETF-based routing protocols such as link-state Open Shortest Path First (OSPF), or distance-vector Routing Information Protocol (RIP).

The upper routing layer maps the multi-domain optical network into a layer of wavelength-constrained segments, with each segment being a series of links and non-converting all-optical switches enclosed between two OEO switches that support wavelength conversion. SPIN Control Server employs the Segment-based Optical Path-vector Routing Protocol (S-OPRP), which adapts the path-vector routing Border Gateway Protocol (BGP) to operate in an interdomain optical routing environment with wavelength-constrained segments.

S-OPRP allows the domain-route computation of a multi-domain lightpath to be performance-based or policy-based relating to secure access of wavelength/link resources. Each interdomain routing controller maintains a path-vector topology database with information consisting of possible routes to destination domains, and advertises path-vector information (the list of domains along a path to a given destination domain) to other distributed controllers. Unlike the BGP, the S-OPRP allows the advertisement of multiple routes to the same destination domain, and each route calculated based on performance constraints and security policies.

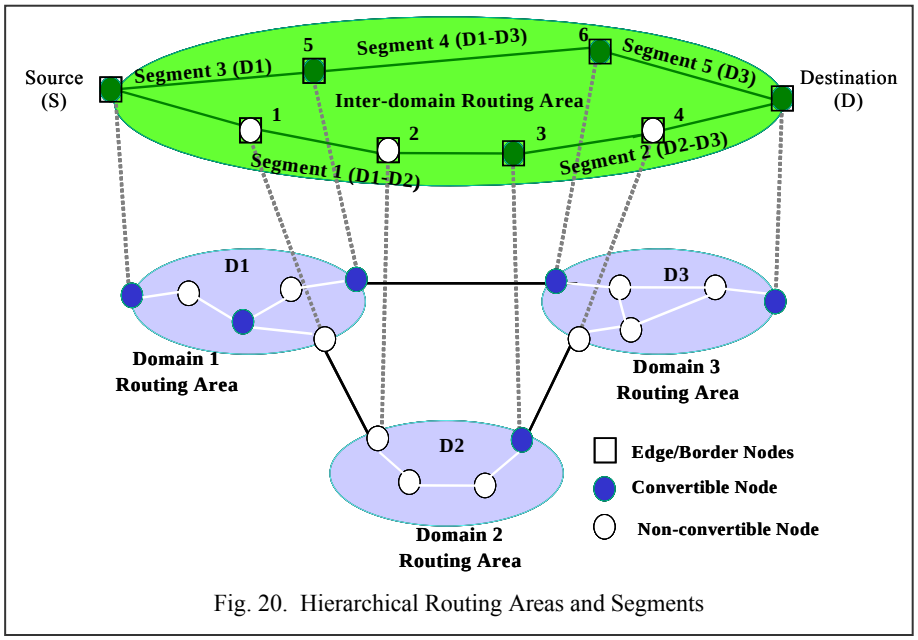


Fig. 20. Hierarchical Routing Areas and Segments

C.2.c. Interdomain Reservation Signaling and Lightpath Setup

SPIN Control Server employs the Segment-based Robust Fast Optical Reservation Protocol (S-RFORP) for interdomain lightpath reservation signaling. Operating over wavelength-constrained segments, S-RFORP extends RFORP [22] with interdomain reservation capability while optimizing reservation delay and robustness. As an intra-domain reservation protocol, RFOPR improves over the IETF-based RSVP by employing fast parallel link-based signaling operations to minimize reservation delay, and wavelength-blocking recovery mechanism to optimize reservation robustness.

S-RFORP employs fast parallel segment-based signaling operations in both wavelength discovery and reservation phases to minimize multi-domain lightpath setup delay. **Fig. 21**a shows that in the wavelength discovery phase, the source node notifies the head of each segment to initiate parallel segment-based discovery processes. Similarly in the wavelength reservation phase, the destination node notifies the tail node of each segment to initiate parallel segment-based reservation of the selected wavelength. Within each segment, **Fig. 21**b shows that serialized link-based wavelength reservation and discovery can be adopted.

A scenario example is presented for interdomain lightpath setup over the three-domain topology as shown in **Fig. 20**. SPIN invokes S-OPRP determines the two routes between source Domain 1 and destination Domain 2. The first route consists of intermediate border nodes (1, 2, 3, 4), while the second route consists of intermediate border nodes (5, 6). These two routes are mapped into wavelength-constrained segments in the upper routing layer. The first route includes Segment 1 and Segment 2; while the second route includes Segment 3, Segment 4 and Segment 5. Assume that S-OPRP selects the first route (S-1-2-3-4-D) based on the performance constraints and security policies. After the route is selected, SPIN invokes S-RFORP to signal each domain for interdomain wavelength discovery and reservation. The parallel segment-based discovery process is carried out in Segment 1 and Segment 2 via the *Discovery* message signaling. In the reservation phase, the selected available wavelength is allocated via the *Reserve* message signaling within Segment 1 and Segment 2 respectively. If wavelength blocking occurs during reservation phase, recovery controls including alternative wavelength selection and localized rerouting will be triggered.

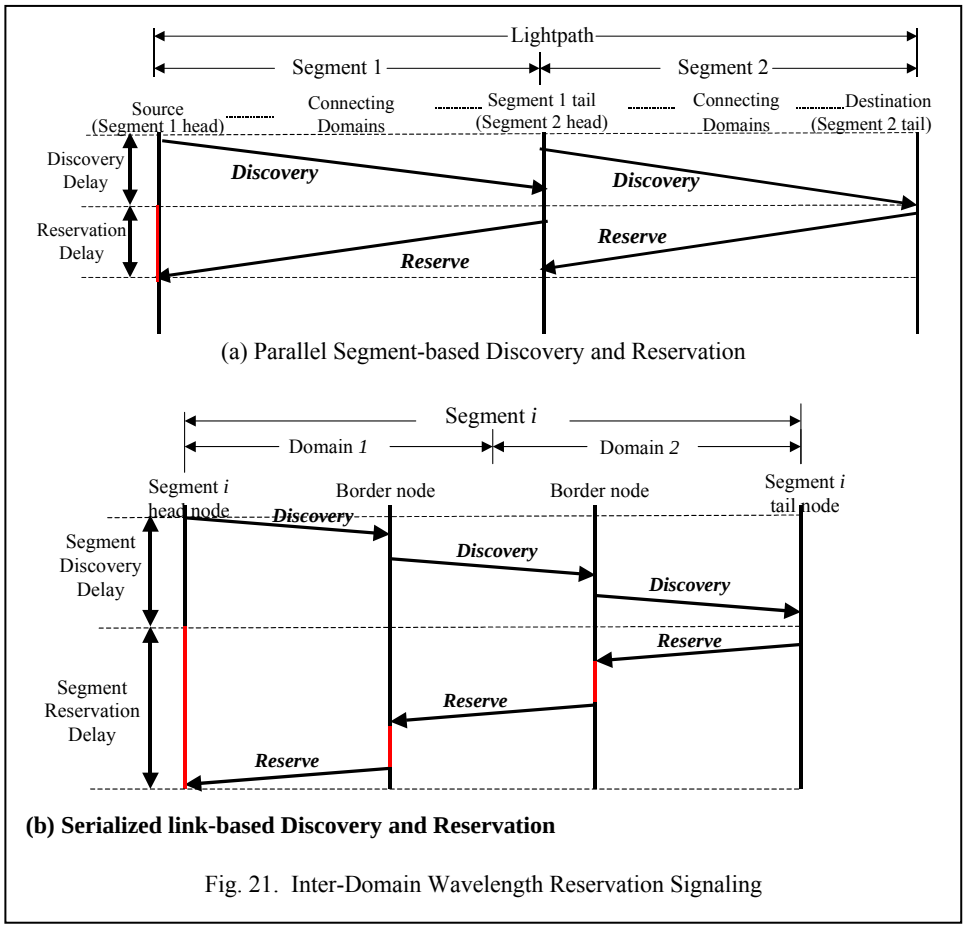


Fig. 21. Inter-Domain Wavelength Reservation Signaling

C.3. Intelligent Optical Grid User Interface (IOGUI)

The intelligent optical grid user interface (IOGUI) consists of user-network interface to enable users/applications of controlling optical network connections and monitoring traffic operations; and dataset browsing interface to enable users of retrieving and visualizing local/remote dataset to any collaborating grid cluster for dataset manipulation.

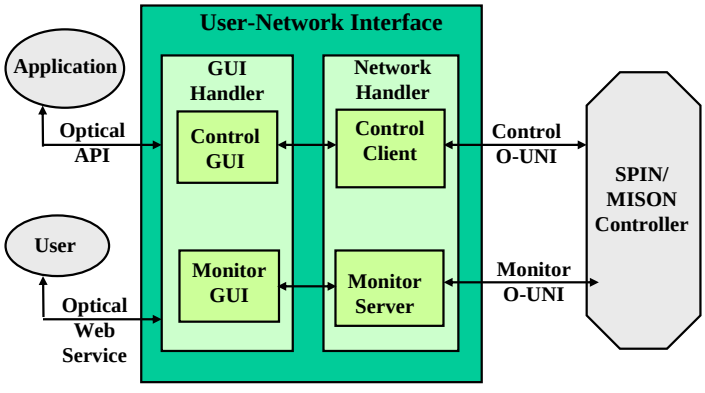


Fig. 22. IOGUI User-Network Interface (UNI)

C.3.a. User-Network Interface

Users interact with the optical network grid through the IOGUI user-network interface (UNI) as shown in Fig. 22. It consists of the user-centric network handler to enable users to interact with the interdomain SPIN and the intra-domain MISON control plane; and of the Generic User Interface (GUI) handler to enable users to control and monitor lightpath provisioning. The IOGUI enables users with visibility of the multi-domain optical network topology. Thus, users are capable of selecting the route for lightpath setup, and to pre-select the route for lightpath restoration via virtual protected lightpath.

The network handler consists of a control client and a monitor server. The control client signals lightpath provisioning requests to the SPIN/MSION controller. The monitor server collects network topology and lightpath status information from the SPIN/MISON control server. The network handler interacts with the SPIN/MISON controller via a proprietary optical user network interface (O-UNI).

The GUI handler supports lightpath provisioning control and network monitor GUIs. The control GUI enables users and applications to setup and release lightpaths on demand respectively though a web service interface and a Java/C++ programming interface. The interface request messages are forwarded to the control client of the network handler, which translates the user or application requests into signaling control messages to invoke the SPIN/MISON controller for lightpath setup or restoration. The monitor GUI enables users to monitor the operational status of multi-domain data plane transport and control plane signaling. It graphically displays updated network topology and existing active lightpaths, which are continuously updated and refreshed via the monitor server of the network handler.

As shown in Fig. 23, the IOGUI graphical display consists of two parts: network monitoring and user control panels. The network monitoring panel displays the topology of a target multi-domain lambda Grid, and the routes of currently active lightpaths. The user control panel allows users to setup and release lightpaths on demand. It accepts the following information: authentication inputs for secure network resource access; lightpath setup/release requests with specified bandwidth and optional user-defined explicit domain routes; and reservation requests at scheduled times with specified duration.

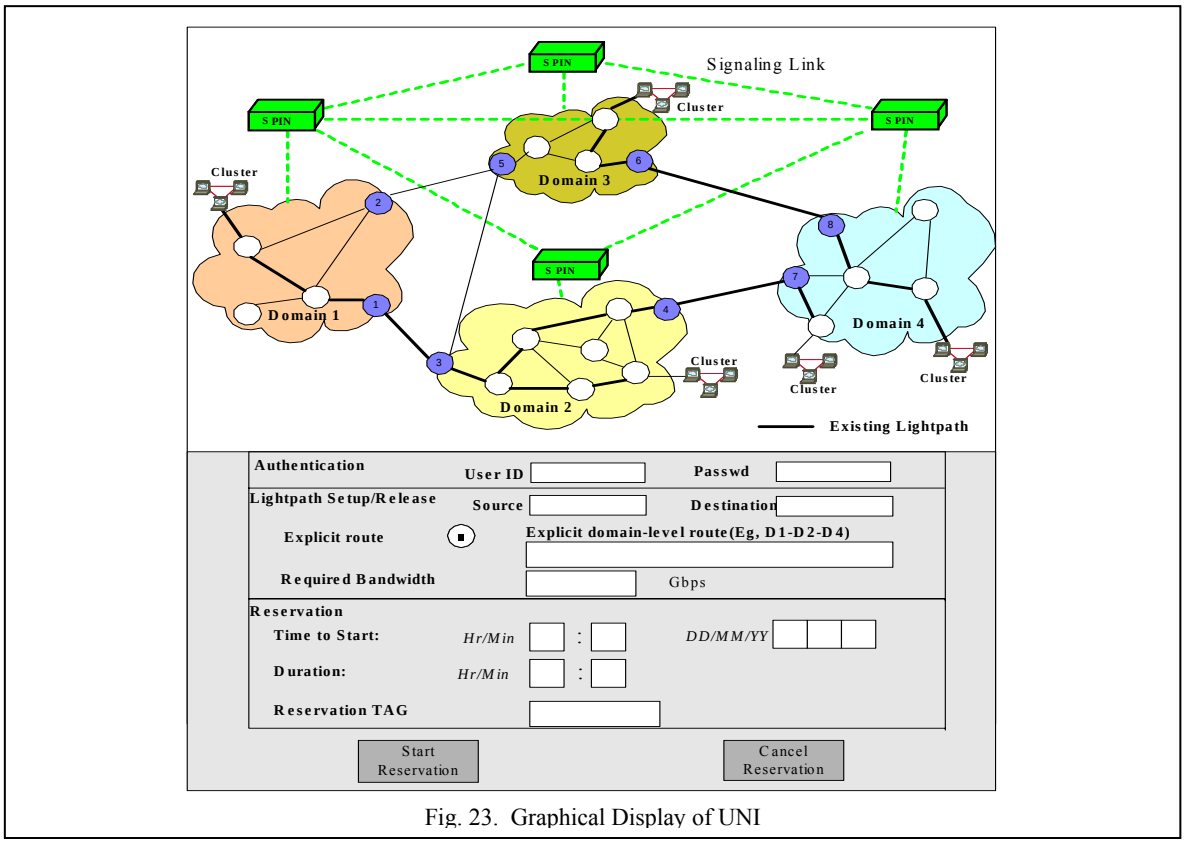


Fig. 23. Graphical Display of UNI

C.3.b. Dataset Browsing Interface

As shown in Fig. 24, the optical grid dataset browsing interface (DBI) enables users to retrieve and visualize a local/remote dataset to any collaborating grid cluster for dataset manipulation. Through the control GUI, users make the selection on datasets, visualization tools, and the cluster locations for manipulating the datasets. Involved dataset types include biomedical, geoscience and oceanology. Current tools for 2-D and 3-D visualization include Juxtaview that is designed specifically for scalable tiled displays to visualize extremely high-resolution geological and biomedical images, LambdaVision that is designed for ultra-high-resolution visualization to support interactive collaboration for various earth science research activities, and GeoWall that provides affordable 3D stereoscopic visualization of small-to-modest-sized geoscience datasets.

By selecting a dataset, visualization tool and manipulation location, the dataset browsing interface enables the dataset to be visualized regardless of its location of discovery. The browsing interface invokes SPIN/MISON controller to dynamically setup a lightpath between location of dataset discovery and desirable location of visualization. Then transmission of dataset is triggered over the established link. Then the DBI invokes the requested visualization tool to manipulate the dataset upon completing transmission.

Fig. 25 shows the graphical display of dataset browsing interface that allows users or collaborative grid applications to post and mine scientific datasets in a multi-domain optical grid. Users can select the visualization tools for displaying the selected datasets.

The UNI monitor displays the optical network topology, animates the movement of signaling information flows along out-of-band signaling links, and animates the movement of data information flows along current active lightpaths. As shown Fig. 23, a collaborative user in Domain 1 requests via the data portal to retrieve a dataset, and to manipulate it at a cluster in Domain 4. After the data portal discovers the dataset in Domain 2, SPIN 1 delegates SPIN 2 to control dataset delivery from Domain 2 (discovery location) to Domain 4 (manipulator location). Consequently, the Monitor will show the animation of lightpath reservation signaling between SPIN 2 and SPIN 4. Once the lightpath is setup successfully, the Monitor will show the animation of dataset transfer from Domain 2 to Domain 4.

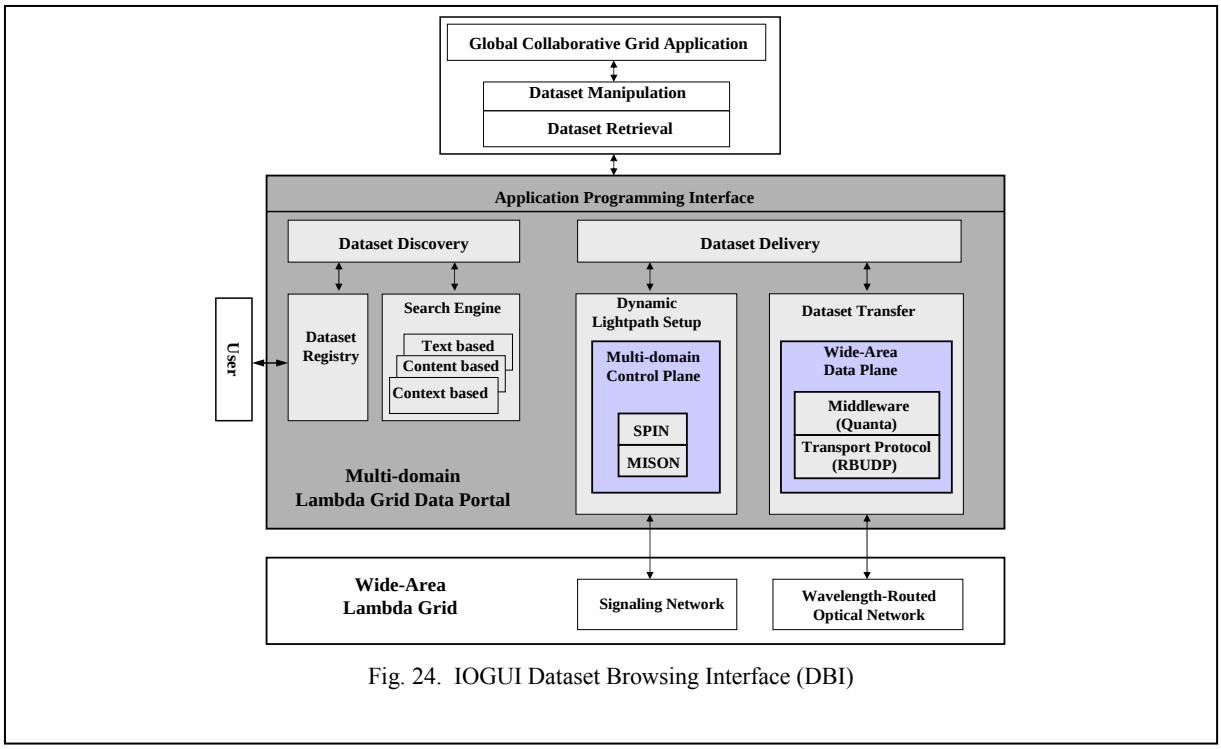


Fig. 24. IOGUI Dataset Browsing Interface (DBI)

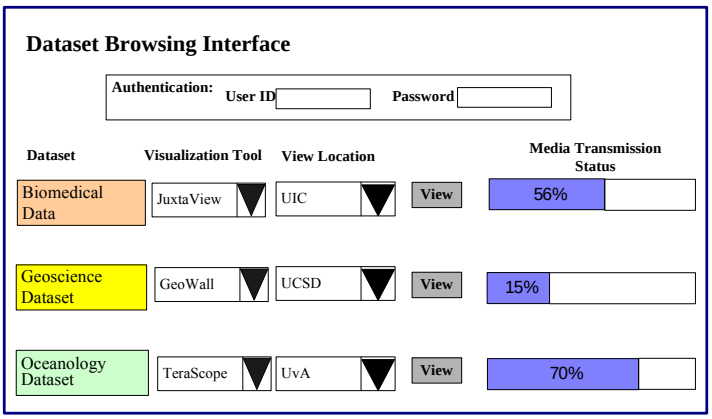


Fig. 25. Graphical Display of DBI

C.4. ISOGA Testbed

An ISOGA-based wide-area optical grid testbed (**Fig. 26**) has been implemented to support international collaborative Grid applications. Interconnecting U.S. and European research organizations. The testbed interoperates multiple optical grid domains based on all-optical local area network (LAN) and metropolitan area network (MAN) with photonic switches.

The testbed consists of four optical grid domains: the optical LAN at University of Illinois at Chicago (UIC), the OMNInet optical MAN at Chicago, the optical LAN at University of California at San Diego (UCSD), and the optical LAN at University of Amsterdam (UvA). The gigabit Ethernet clusters of the grid domains are interconnected via the StarLight facility in Chicago and the NetherLight facility in Amsterdam, which act as regional peering exchange for optical networks and as optical points of presence (PoP) for international optical backbone networks. The trans-Atlantic OC192 (10Gbps) fiber link between StarLight and NetherLight enables 10 Gbps data transport.

The optical LANs of the UIC, UCSD and UvA domains are developed with Calient and Glimmerglass all-optical photonic switches, which are enabled by the MISON intra-domain control plane. The optical MAN of the OMNInet is developed with Nortel photonic switches, which are enabled by the proprietary ODIN intra-domain control plane with GMPLS-like signaling and routing protocols.

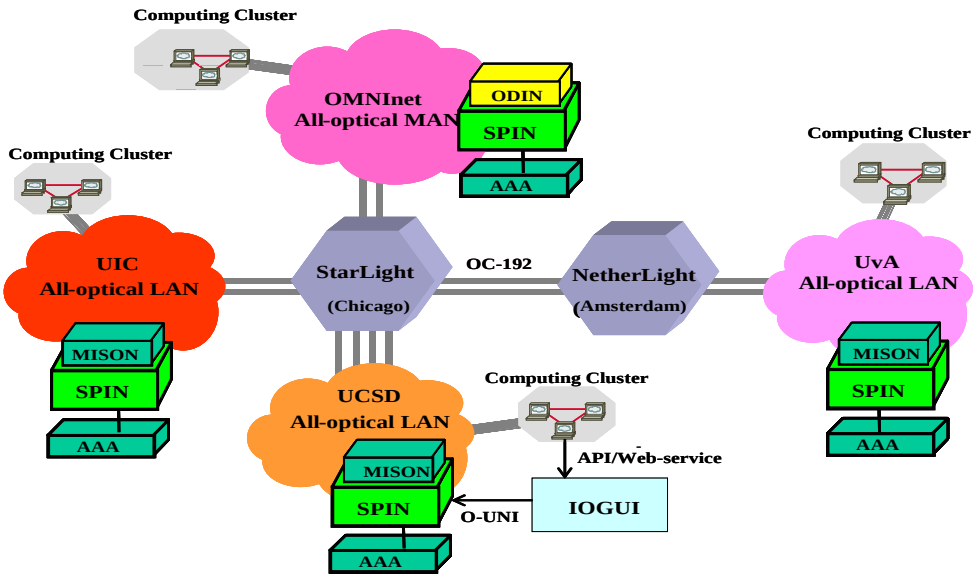


Fig. 26. Multi-domain ISOGA Testbed

D. Accomplishment

D.1. Milestones

The MISON/SPIN control planes and the IOGUI user-network and dataset browsing interfaces had been developed and integrated in Phases 1, 2 and 3 testbeds. Scientific collaborative applications had been successfully deployed and demonstrated over these testbeds.

The 2-domain Phase 1 ISOGA testbed consists of the optical LAN and computing clusters at the University of Illinois at Chicago (UIC) domain; and the OMNInet optical MAN with computing clusters at the Northwestern University (NW) domain.

The 3-domain Phase 2 ISGOA testbed is derived from the Phase 1 testbed by interconnecting the additional optical LAN and computing clusters at the University of Amsterdam (UvA) sites. These domains are interconnected via the optical interchange facilities of Chicago StarLight and Amsterdam NetherLight. At the Supercomputing (SC) 2004 conference, control of on-demand multimedia traffic provisioning by scientific application users had been demonstrated over the Phase 2 testbeds.

The 4-domain Phase 3 ISOGA testbed is derived from the Phase 2 testbed by interconnecting the additional optical LAN and computing clusters at the University of California at San Diego (UCSD) domain. At the iGrid 2005 conference in San Diego, global collaborative applications had been deployed successfully over the Phase 3 testbed. During the optical-grid browser demonstration, on-site UCSD scientists searched and discovered biomedical microscopy and geological bathymetry datasets posted in the UA computing clusters. These UA datasets were then automatically delivered via dynamic optical connections to both USCD and UIC sites, and visualized in the cluster display panels at these two sites. Effectively, the optical-grid browser allowed distant scientists to seamlessly browse and collaboratively manipulate the datasets.

D.2. Publications Supported by this Grant

2008

- **O. Yu**, H. Xu, "QoS differentiated Adaptive Scheduled Optical Burst Switching for Grid Networks," in *Proc. IEEE/CreateNet International Workshop on Networks for Grid (GridNets 2008),* Oct. 2008.

2007

- H. Xu, **O. Yu**, L. Yin, M. Liao, "Segment-Based Robust Fast Optical Reservation Protocol," in *Proc. IEEE INFOCOM 2007 High-Speed Networks Workshop*, May 2007, pp.36-40.
- **O. Yu**, H. Xu, L. Yin, "Robust Timely Scheduled Optical Burst Switching," in *Technical Digest of IEEE/LEOS/OSA Optical Fiber Communication Conf. (OFC/NFOEC 2007),* March 2007, pp. 1-12.

2006

- **O. Yu**, L. Yin, H. Xu, M. Liao, "Multi-Granular Integrated Services Optical Network," *Optical Society of America / Journal of Optical Networking*, vol. 5, issue 12, pp. 985-1001, December 1, 2006.
- Y. Cao and **O. Yu**, "Groupcast in Wavelength-Routed WDM Networks," *IEEE Journal of Lightwave Technology*, vol. 24, no. 11, pp. 4286-4295, November 2006.
- **O. Yu**, A. Li, Y. Cao, H. Xu, M. Liao, L. Yin, "Multi-domain Lambda Grid Data Portal for Collaborative Grid Applications," *Elsevier Journal of Future Computer Systems,* vol. 22, issue 8, pp. 993-1003, Oct. 2006.
- Y. Cao, **O. Yu**, "Optimization of Loss-balanced Multicast in All-Optical WDM Networks," *Springer Journal of Combinatorial Optimization*, vol. 12, no. 1-2, pp. 71-82, Sept. 2006.
- **O. Yu**, Y. Cao, "Placement of Light Splitters and Wavelength Converters for Efficient Multicast in All-Optical WDM Networks," *IEICE Transactions on Information and Systems*, vol. E89-D, no. 2, pp. 709-718, Feb. 2006.
- **O. Yu**, M. Liao, Y. Cao, "Multi-Granular Stream Optical Burst Switching," in *Proc. International Workshop on Optical Burst/Packet Switching (WOBS 2006)*, Oct. 2006.
- **O. Yu**, Y. Cao, "Dynamic Groupcast Traffic Grooming in WDM Networks," in *Proc. IEEE International Conf. on Communications (ICC 2006),* vol. 29, no. 1, June 2006, pp.2590-2596.

2005

- **Yu**, Y. Cao, "Mathematical Formulation of Optical Multicast with Loss-balanced Light-forest," in *Proc. IEEE Global Telecommunications Conf. (Globecom 2005)*, vol. 24, no. 1, Nov. 2005, pp. 1967-1971.
- **Yu**, M. Liao, Y. Cao, "Synchronous Stream Optical Burst Switching," in *Proc. IEEE/CreateNet International Workshop on Networks for Grid (GridNets 2005),* Oct. 2005, pp. 524-529.
- **Yu**, H. Xu, L. Yin, "Adaptive Robust Optical Burst Switching," in *Proc. IEEE/CreateNet International Workshop on Networks for Grid (GridNets 2005),* Oct. 2005, pp. 496-504.
- Y. Cao. **O. Yu**, "On the Study of Group Multicast in WDM Mesh Network," in *Proc. IEEE International Conf. on Communications (ICC 2005),* no. 1, May 2005, pp. 1625-1630.
- **O. Yu**, Y. Cao, "Optimal Placement of Light Splitters and Wavelength Converters for Multicast in WDM Networks," in *Proc. 3rd International Conf. on Communication, Circuits and Systems (ICCCAS 2005)*, vol. 1, May 2005, pp. 585-589.
- Y. Cao, **O. Yu**, "QoS-Guaranteed Routing and Wavelength Assignment for Group Multicast in Optical WDM Networks," in *Proc. 9th IFIP/IEEE Optical Network Design and Modeling (ONDM 2005),* pp. 175-184, Feb. 2005.

2004

- **O. Yu**, T. DeFanti, "Collaborative Lambda-Grid over All-Optical Metro Network with Intelligent Application Signaling," in *Proc. ACM/IEEE Conf. on Supercomputing (SC 2004)*, ISBN:0-7695-2153-3, Nov. 2004.
- **O. Yu**, "Intercarrier Interdomain Control Plane for Global Optical Networks," in *Proc. IEEE International Conf. on Communications (ICC 2004),* June 2004, vol. 27, no. 1, pp. 1679-1683.

D.3. PhD Students Participation

1. Yuan Cao (2006)
   - Graduated in summer 2006.
   - Thesis Title: "Optical Network Service Provisioning for Multicast and Groupcast Communications".

2. Huan Xu (2008)
   - Graduated in spring 2008.
   - Thesis Title: "Robust Differentiated-Services Wavelength Routed Optical Networks with Uncertain Link State Information".

3. Leping Yin (2008)
   - Graduated in fall 2008.
   - Thesis Title: "Reliable Traffic Control and Resource Provisioning in Multi-Granular Integrated Services Optical Network".

E. References

1. IETF RFC3471 “Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description,” Jan. 2003
2. ITU-T Draft Recommendation G.8080/Y.1304, “Architecture for the Automatically Switched Optical Network (ASON),” Nov. 2001, http://www.itu.int/ITU-T
3. X. Cao, V. Anand, Y. Xiong and C. Qiao, “Performance Evaluation of Wavelength Band Switching in Multi-fiber All-Optical Networks,” in Proc. IEEE Infocom 2003, San Francisco, CA, Apr. 2003. pp. 2251 –2261.
4. C. Qiao, and M. Yoo, “Optical burst switching (OBS)-A new paradigm for an optical internet,” J. High speed Networks, vol. 8, no. 1, pp. 69-84, 1999.
5. J. Y. Wei and R. I. McFarland, “Just-in-time Signaling for WDM Optical Burst Switching Networks,” J. Lightwave Tech., vol. 18, no. 12, pp. 2019-2037, Dec. 2000.
6. M. Yoo and C. Qiao, “Just-enough-time (JET): A high speed protocol for bursty traffic in optical networks,” in Proc. IEEE/LEOS Conf. on Technologies for a Global Information Infrastructure, 26–27, Aug. 1997.
7. C. Qiao, "Polymorphic Control for Cost-Effective Design of Optical Networks," in European Trans. On Telecommunications (ETT), Special issue on WDM Networks, Vol. 11, No. 1, pp. 17-26.
8. High performance cluster computing, R Buyya, Prentice Hall PTR Upper Saddle River, NJ, 1999
9. X. Cao, V. Anand, Y. Xiong, and C. Qiao, "A study of waveband switching with multilayer multigranular optical cross-connects," IEEE J. Sel. Areas Commun. 21(7), 1081–1095, 2003.
10. S. Yao, C. Ou and B. Mukherjee, “Design of Hybrid Optical Networks With Waveband and Electrical TDM Switching,” in Proc. IEEE GLOBECOM 2003, 5, pp. 2803-2808, Dec. 2003.
11. R. Izmailov, S. Ganguly and T. Wang, Y. Suemura, Y. Maeno and S. Araki “Hybrid Hierarchical Optical. Networks,” IEEE Communication Magazine, vol.40, no. 11, pp.88-94, Nov.2002
12. C. Xin, C. Qiao, Y. Ye and S. Dixit, “A hybrid optical switching approach,” in Proc. IEEE GLOBECOM 2003, 7, 3808–3812, Dec. 2003.
13. G. M. Lee, B. Wydrowski, M. Zukerman, J. K. Choi, and C. H. Foh, “Performance Evaluation of an Optical Hybrid Switching System,” in Proc. IEEE GLOBECOM 2003 5, 2508–2512, Dec. 2003.
14. I. de Miguel1 , J. C. González, T. Koonen, R. Durán1, P. Fernández and I. T. Monroy "Polymorphic Architectures for Optical Networks and their Seamless Evolution towards Next Generation Networks," Photonic Network Communications, vol. 8, no. 2, pp.177-189, Sept. 2004.
15. Van Breusegem, E.; Cheyns, J.; De Winter, D.; Colle, D.; Pickavet, M.; Demeester, P. and Moreau, J. A broad view on overspill routing in optical networks: a real synthesis of packet and circuit switching,” Optical Switching and Networking. Vol. 1: (1) 51-64, 2005.
16. N.F. Huang, G.H. Liaw, and C.P. Wang, "A novel all-optical transport network with time-shared wavelength channels," IEEE J. Sel. Areas Commun. 18, no.10, pp.1863–1875, Oct. 2000.
17. M. Duser, and P. Bayvel, “Analysis of a dynamically wavelength-routed optical burst switched network architecture,” J. Lightwave Technology, Vol. 20, no. 4, pp.574-585, 2002.
18. RFC 3473, “Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions”
19. Oliver Yu, Huan Xu, Leping Yin “Adaptive Robust Optical Burst”, in Proc. IEEE/Create-Net International Workshop on Networks for Grid (GridNets 2005), pp. 524-529, Oct. 2005.
20. B. Wen and K. M. Sivalingam, "Routing, wavelength and time-slot assignment in time division multiplexed wavelength-routed optical WDM networks," in Proc. IEEE INFOCOM, (New York, NY), June 2002.
21. Rajesh M. Krishnaswamy and Kumar N. Sivarajan,"Design of Logical Topologies: A Linear Formulation for Wavelength-Routed Optical Networks with No Wavelength Changers," IEEE/ACM Trans. Networking, Vol. 9, no. 2, April 2001.
22. O. Yu, “Intercarrier Interdomain Control Plane for Global Optical Networks,” in Proc. ICC 2004, IEEE International Communications Conference, vol.3, pp.1679-1683, 20-24 June 2004.