

## Control of Lightpaths in an Optical Network

### Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC2026](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

### Abstract

This document details requirements and mechanisms for optical bandwidth management and restoration in a dynamically reconfigurable optical network. A management approach is described where IP algorithms and mechanisms are used to control optical resources, paving the way for the optical Internet. The proposal is specifically intended for optical internetworking in which IP routers are connected by the reconfigurable optical layer using lightpaths. However, it is assumed that the same methodology will be used for non-IP traffic as well.

## **1. Introduction**

This document describes an approach for optical bandwidth management in a dynamically reconfigurable optical network. The optical network consists of optical layer cross-connects (OLXCs) that switch high-speed optical signals (e.g. OC-48, OC-192) from input ports to output ports. These OLXCs are interconnected via WDM links. The OLXCs may be purely optical or electrical or a combination. The

network is assumed to be within a single domain of authority (or trust), with the inter-domain capability to be addressed in a future document.

Every node in the network consists of an IP router and an OLXC. This document is only concerned with the functions of the router as they relate to the control of the optical layer. In general, the router may be traffic bearing as proposed in [1], or it may function purely as a controller for the optical layer and carry no IP data traffic. The node may be implemented using a stand-alone router interfacing with the OLXC through a defined interface, or may be an integrated system, in which the router is part of the OLXC system. The policies and mechanisms proposed within this document for optical bandwidth management and restoration are applicable whether the router carries data or not.

In the networks considered, it is assumed that the physical hardware is deployed, but that network connectivity is not defined until lightpaths are established within the network. A lightpath is a constant bit-rate data stream connected between two network elements such as IP routers. An example is one direction of an OC-48/STM-16 (2.5 Gbit/s) or an OC-192/STM-64 (10 Gbit/s) established between two client routers through the OLXCs with or without Multiplex / regenerator Section Overhead termination.

Lightpaths may be requested by client IP aware network elements, or by external operations systems used for IP-ignorant network elements. Such requests may be for uni-directional or bi-directional lightpaths of a given bandwidth and with specified restoration requirements. The lightpaths are provisioned by choosing a route through the network with sufficient available capacity. The lightpath is established by allocating capacity on each link along the chosen route, and appropriately configuring the OLXCs. Restoration is provided by reserving capacity on routes that are physically diverse to the primary lightpath.

This document is a contribution to the on-going discussion on the provisioning and management of optical networks. Specifically, we propose a framework for the management of optical layer resources and restoration. We identify a set of services that we foresee offered by an optical network, and derive the requirements on functionality offered by the network. We define an addressing and naming scheme, which is required to facilitate distributed information maintenance, and separate the connectivity management from higher levels concerns, such as global network and customer management. The main part of the document specifies in detail the mechanisms and information requirements for fast provisioning, diverse routing and restoration. In this part, we discuss the state required and the mechanisms for the maintenance of this state, and propose a new model for restoration.

The approach proposed in this document complements that proposed by Awduche et al. on Multi-Protocol Lambda Switching [2], which is



based on Multi-Protocol Label Switching (MPLS). While our proposal does not require MPLS, it is consistent with the use of MPLS and specifically of the signaling protocols proposed for MPLS [3,4]. As such, this document is a contribution into the work on Multi-Protocol Lambda Switching. The objective of our contribution is to incorporate the optical layer requirements in the selection and extension of the proposals under discussion. In this document we illustrate how IP algorithms and mechanisms can be used to implement these requirements. While our architectural assumptions are congruent with those in [2], we analyze optical services and networks in greater detail, thereby addressing some of the issues raised in [2].

Neither this document nor the proposal in [2] analyzes the efficiency of network utilization with respect to a specific protocol choice for provisioning and restoration. In addition to the functional requirements advanced in this document we believe that the capacity efficiency is an important issue to be considered in developing the algorithms and protocols for lightpath provisioning and restoration in the optical network.

The rest of the document is organized as follows. In [Section 2](#) we provide background and definitions needed for the rest of the document. [Section 3](#) provides a brief discussion of the network architecture. In [Section 4](#) we analyze network services and outline how they translate into requirements for the optical layer functionality. Naming and addressing are discussed in [Section 5](#). Sections [6](#) and [7](#) contain the embodiment of the implementation of the requirements for provisioning and restoration at the optical layer. Section 8 discusses periodic resource reconfiguration policies. Sections [9](#), [10](#) and [11](#) specify in detail the information requirements and interface primitives.

## **2. Background**

In order to facilitate the discussion we define the following network objects:

- Wavelength Division Multiplexer (WDM). A system which takes multiple optical inputs, converts them into narrowly spaced wavelength optical signals within an optical amplification band and couples them onto a single fiber. The amplified signal is received at the receive end, demultiplexed and converted to multiple channels of standard wavelength to interface with other equipment. It is, however, possible to take the wavelength specific signals directly as the inputs. In that case no wavelength conversion is necessary at the WDM system. The WDM system may or may not be integrated with an OLXC.
- Channel. A channel is a uni-directional optical tributary connecting two OLXCs. Multiple channels are multiplexed optically at the WDM system. One direction of an OC-48/192 connecting two immediately neighboring OLXCs is an example of a channel. A



single direction of an Optical channel (Och) as defined in ITU-T G.872 [5] between two OLXCs over a WDM system is another example of a channel. A channel can generally be associated with a specific wavelength in the WDM system. However, with a WDM system with transponders the interfaces to the OLXC would be a standard single color (1310 or 1550 nm). In addition, a single wavelength may transport multiple channels multiplexed in the time domain. For example, an OC-192 signal on a fiber may carry four STS-48 channels. For these reasons we define a channel which is different from wavelength although in many applications there is a one-to-one correspondence.

- Optical layer cross-connect (OLXC). A switching element which connects an optical channel from an input port to an output port. These devices are also often referred to as optical cross-connects (OXC). Note that an optical add-drop multiplexor (OADM) is viewed here as a simple OLXC. The switching fabric in an OLXC may be either electronic or optical. If the switching fabric is electronic, then switching would occur at a given channel rate, but the interface ports may in fact be at higher rates (i.e. multiplex multiple channels onto a single wavelength). It is important to note that because of the multiplexing function assumed in the OLXC, we do not restrict the lightpaths to be identical to the Och defined in ITU-T G.872 [2]. If the WDM systems contain transponders or if electronic OLXCs are used, then it is implied that a channel associated with a specific wavelength in the WDM input can be converted to an output channel associated with a different wavelength in the WDM output (i.e. wavelength conversion is inherent). However, if the switching fabric is optical and there is no transponder function in the WDM system, then wavelength conversion is only implemented if optical to electronic conversion is performed at the input or output ports, or if optical wavelength converters are introduced to the OLXC. Also, we assume that the rates in the input and output channels in an all-optical OLXC are identical, implying that Time Division Multiplexing (TDM) is not offered within the OLXC.
- Link. A link is a set of channels in a given direction connecting a particular pair of OLXCs and routed along the same physical route. Multiple links may exist between the same OLXCs, for example if route diversity is implemented between two OLXCs. Note that links defined this way are uni-directional. There can be multiple WDMs within a link. A single WDM can be divided into multiple links (i.e. between different OLXCs). The link is thus not necessarily a union of WDMs, and there is not necessarily a one-to-one correspondence between WDM systems and links.
- Fiber Span. A fiber span consists of a collection of fiber cables that are located in the same conduit or right of way. If there is a cut in the fiber span, then failures would potentially be experienced on all fibers within the fiber span.



- Shared Risk Link Group (SRLG). For restoration and diverse routing purposes it may be necessary to associate links within a fiber span in a Shared Risk Link Group (SRLG). A SRLG is a union of all links that ride on a fiber span. Links may traverse multiple fiber spans, and thus be in multiple SRLGs.
- Drop Port. An OLXC port that connects to the end client network element (NE). The drop interface connects the client port to the OLXC drop port. This is essentially a User Network Interface (UNI) connecting the end devices to the optical layer. The drop port terminates the user network interface between the client NE and the optical network. It is necessary to distinguish this type of interface from others to identify network requests originating from a client NE.
- Network Port. An OLXC port not directly interfacing with an end client NE. A Network Port in an OLXC would always interface with another Network Port via a WDM system or directly via optical fibers.
- Lightpath. The elementary abstraction of optical layer connectivity between two end points is a uni-directional lightpath. A lightpath is a fixed bandwidth connection (e.g. one direction of a STM-N/OC-M payload or an Och payload) between two network elements established via the OLXCs. A bi-directional lightpath consists of two associated lightpaths in opposite directions routed over the same set of nodes. Note that if the OLXC is an electronic SONET/SDH line terminating equipment, the entire path need not be OC-48 for an OC-48 path. Note also that an OC-N and Och are by definition bi-directional, whilst lightpaths are by default uni-directional (anticipating asymmetric loads). Therefore it is assumed that independent lightpaths in opposite directions may use a bi-directional OC-48 or Och span.
- Source and Source Address. A source can be a client router physically connected to an OLXC by one or more OC-48/192 interfaces. A source can also be a non-IP NE connected to the OLXC via an OC-48/192 interface. In the case of an IP router source, the router will have an IP address and the physical interfaces to the OLXC are identified with some set of addresses (potentially a single IP address, or a unique address per port). In the case of a non-IP NE, either the NE will be assigned an IP address, or the OLXC port connecting the NE will have an IP address. For non-IP aware equipment interfacing the OLXC, any connection request must be originated externally via craft or external OS interfaces.
- Destination and Destination Address. The destination is essentially the same as the source from the physical interface perspective. When a request is generated from one end, the other end client or end OLXC interface becomes the destination.





- First-hop router. The first router within the domain of concern along the lightpath route. If the source is a router in the network, it is also its own first-hop router.
- Last-hop router. The last router within the domain of concern along the lightpath route. If the destination is a router in the network, it is also its own last-hop router.
- Mediation device (MD). A vendor specific controller used to control the OLXC. The mediation device provides the interface between external sources and the OLXC, translating logical primitives to and from the proprietary controls of the OLXC. If the router is integrated with the OLXC, then the mediation device is merely a function within the integrated entity, and not an explicit device.

### 3. Network architecture

The salient feature of the network architecture is that every node in the network consists of an IP router and a reconfigurable OLXC. The IP router is responsible for all non-local management functions, including the management of optical resources, configuration and capacity management, addressing, routing, traffic engineering, topology discovery, exception handling and restoration. In general, the router may be traffic bearing as proposed in [1], or it may function purely as a controller for the optical network and carry no IP data traffic. The mechanisms and requirements discussed within this document are applicable regardless of whether data traffic traverses through the routers or not. Although the IP router performs all management and control functions, lightpaths may carry arbitrary types of traffic.

The IP router implements the necessary IP protocols and uses IP for signaling to establish lightpaths. Specifically, optical resource management requires resource availability per link to be propagated, implying link state protocols such as OSPF. In subsequent discussions we assume OSPF. However, other link state algorithms, for example that used in PNNI [6], may be equally applicable.

On each link within the network, one channel is assigned as the default routed (one hop) lightpath. The routed lightpath provides router to router connectivity over this link. These routed lightpaths reflect (and are thus identical to) the physical topology. The assignment of this default lightpath is by convention, e.g. the 'first' channel. All traffic using this lightpath is IP traffic and is forwarded by the router. All control messages are sent in-band on a routed lightpath as regular IP datagrams, potentially mixed with other data but with the highest forwarding priority. We assume multiple channels on each link, a fraction of which is reserved at any given time for restoration. The default routed lightpath is restored on one of these channels. Therefore we can assume that as long as the link is functional, there is a default routed lightpath on that link.



In resource constrained parts of the network, such as the link connecting the customer premise to the network, it may not be economically feasible to reserve a channel and the associated IP interface for the default routed lightpath. Within the network, where each link has multiple channels carrying traffic from many customers, the overhead of the routed wavelength is amortized over the channels on that link. In contrast, the link connecting the customer premise to the network may typically have only a single traffic bearing channel. In this case, unless the routed lightpath is also used for IP data traffic, the overhead of an optical channel dedicated for control may be excessive. If electronic line terminating OLXCs are used, an alternative to dedicating an optical channel as the routed lightpath is to transport the IP datagrams within the framing overheads of the signals (e.g. SONET Multiplex and/or Regenerator Section Overhead).

The IP router communicates with the OLXC mediation device (MD) through a logical interface. The interface defines a set of basic primitives to configure the OLXC, and to enable the OLXC to convey information to the router. The mediation device translates the logical primitives to and from the proprietary controls of the OLXC. Ideally, this interface is both explicit and open. We recognize that a particular realization may integrate the router and the OLXC into a single box and use a proprietary interface implementation. The crucial point is that this proprietary interface must still provide equivalent functionality to the interface described herein.

Another interface of importance is the service interface between the customers and the network. This interface determines the set of services that the optical network provides. In [Section 11](#) we discuss this interface.

#### **4. Optical Network Requirements**

It is important to identify the services that an optical network should offer, and the functionality that must be implemented by the optical infrastructure to support these services. Within the same domain of trust, servers and other network management systems may have access to the network information available to routers, and may actively interact with the network by requesting lightpaths. These servers may for example provide authentication, risk analysis and management, and more. While this document defines mechanisms that would be used by these higher layer systems, the specifics of these advanced services are not discussed herein. The following outlines the optical network services and functionality.

##### **4.1. Optical network services**

Lightpath services. Lightpath requests between a source and destination with the following attributes:

- Lightpath identifier. A globally unique identifier.



- Bandwidth: A limited set of bandwidth allocations are available (e.g. OC-48, OC-192).
- Uni-directional or bi-directional lightpath.
- Diversely routed lightpath group identifier(s). A globally unique group identifier defined for diversely routed lightpath groups (see below). A convenient way to create one is by concatenating the IP address of the first-hop router, and a sequence number unique at the router. If the diversely routed lightpath group is not coordinated by the first-hop router (see [Section 6.3](#)) but instead by an external operations system, the address of the coordinating entity would be used instead.
- Restoration class: one of (i) restored lightpath, (ii) restored IP connectivity, (iii) not restored, (iv) not restored and preemptable. For Class (i) the lightpath must be restored using another lightpath, whose route is different from the primary. IP restored (Class (ii)) assumes that the traffic transported on the lightpath is IP, and may be restored by routing through the network routers if needed and given that routing capacity is available [1]. Clearly, the network will attempt to restore all lost connectivity if and when possible. This is however done on a best effort basis.

Diversely routed lightpath groups. A set of diversely routed non-restored lightpaths so that for any single failure, at most a given number of lightpaths out of the group fail. A lightpath belongs to one or more diversely routed lightpath group(s).

The simplest form of diversely routed lightpaths is a group originating at the same first hop router. This case is handled by the first hop router. More generally, the lightpaths of a group may potentially have different sources and destinations, and may be required to satisfy other more stringent requirements, such as ensuring that particular end-points are always connected. The implementation of these more elaborate risk management services is outside the scope of this document and would typically be provided by higher level management system(s) external to the network nodes.

#### **4.2. Requirements on optical network functionality**

To cope with decreasing provisioning time scales, and to enhance scalability, it is necessary to maintain the network state in a distributed manner. This need drives most other system requirements and implementation choices, and the service requirements above imply the need for the following information and algorithms:

- 1) Information on topology and inventory of physical resources (e.g. channels).
- 2) Information about shared risk link groups (SRLGs). This is necessary for routing of restoration lightpaths, and for diverse routing of primary lightpaths.
- 3) Information regarding the current resource allocations must be propagated throughout the network. For scalability, details of individual wavelength allocations are not distributed.
- 4) An addressing and naming scheme.



- 5) Algorithms for distributed state maintenance of the above.
- 6) Algorithms and mechanisms for the allocation of bandwidth resources to new lightpaths, and for the reservation of restoration capacity. These algorithms and mechanisms must be able to support diversely routed lightpaths as described above.
- 7) Algorithms for the management and optimizations of resource allocation; and the minimization of resources reserved for restoration. Established lightpaths may occasionally be reconfigured to optimize resource allocations.
- 8) Algorithms and mechanisms to ensure diversity in routes among a set of lightpaths.
- 9) Algorithms and mechanisms for fault detection and recovery (i.e., notification and exception handling).
- 10) Specification of interfaces between the external systems (including client) and the network.
- 11) Specification of interfaces between the router and the OLXC mediation device.

## 5. Naming and Addressing

Every network addressable element must have an IP address. Typically these elements include each node and every optical link and IP router port. When it is desirable to have the ability to address individual optical channels those are assigned IP addresses as well. The IP addresses must be globally unique if the element is globally addressable. Otherwise domain unique addresses suffice.

Local naming schemes can be used to identify channels within fibers, or to identify fibers within links. However, globally unique names will be required to specify routes through the network. A possible naming convention for uniquely identifying the channels used along a route through a network is proposed. This convention identifies a channel according to the OLXC from which it is sourced, the link within the OLXC and the channel within the link. How these values are used depends on what elements are assigned IP addresses. If only the OLXC has a unique IP address, then the naming scheme uses a pre-defined convention to identify links and channels within the OLXC (i.e. OLXC IP address : link number : channel number). Alternatively, if the link is also assigned an IP address, then the channel is uniquely defined by the link IP address, and the channel identifications within that link (i.e. link IP address : NULL identifier : channel number). The NULL identifier is used to indicate that a given field is invalid. For example, in the identifier associated with the link IP address, the second field contains a NULL identifier, which is used to indicate that a link number is not required, because the IP address corresponds to a unique link. Thus, the first non-NULL identifier can be used to denote what the IP address corresponds to (i.e. OLXC or link). The same applies for addresses assigned at finer granularities, e.g., for each channel. Clearly, other variants on the above naming scheme are possible.





A client must also have an IP address by which it is identified. However, optical lightpaths could potentially be established between devices that do not support IP (i.e. are not IP aware), and consequently do not have IP addresses. This could be handled by either assigning an IP address to the device, or alternatively assigning an address to the OLXC port to which the device is attached. Whether or not a client is IP aware can be discovered by the network using traditional IP mechanisms.

## **6. Provisioning at the Optical Layer**

### **6.1. Provisioning lightpaths in a network with wavelength converters**

In an optical network with wavelength conversion, channel allocation can be performed independently on different links along a route. However, if wavelength converters are not available, then a common wavelength must be located on each link along the entire route, which requires some degree of coordination between different nodes in choosing an appropriate wavelength. We commence this section by outlining how lightpath provisioning may be performed in a network with wavelength converters. Networks and sub-networks without wavelength converters are considered in [Section 6.5](#).

A lightpath request from a source is received by the first-hop router. The first-hop router creates a lightpath setup message and sends it towards the destination of the lightpath where it is received by the last-hop router. If the originator of the request is not the source, the originator tunnels the request to the first-hop router. The lightpath setup is sent from the first-hop router on the default routed lightpath as the payload of a normal IP packet with router alert. A router alert ensures that the packet is processed by every router in the path. A channel is allocated for the lightpath on the downstream link at every node traversed by the setup. The identifier of the allocated channel is written to the setup message. If no channel is available on some link, the setup fails, and a message is returned to the first-hop router informing it that the lightpath cannot be established. We propose to use the 'destination not reachable' ICMP (Internet Control Messaging Protocol) message for this, but any comparable mechanism would suffice. For example, if all routers are MPLS capable one could use the appropriate CR-LDP (Constraint-based Routing - Label Distribution Protocol) message. If the setup fails, the first-hop router issues a release message to release resources allocated for the partially constructed lightpath. Upon failure, the first-hop router may attempt to establish the lightpath over an alternate route, before giving up on satisfying the original user request. Note that the lightpath is established over the links traversed by the lightpath setup packet. Moreover, when electronic line terminating OLXCs are used it is possible to alternatively use the channel overheads of the chosen lightpath channels to carry the lightpath setup. After a channel has been allocated at a node, the router communicates with the OLXC to reconfigure the OLXC to provide the desired connectivity.



After processing the setup, the destination (or the last-hop router) returns an acknowledgement to the source. The acknowledgment indicates that a channel has been allocated on each hop of the lightpath. It does not, however, confirm that the lightpath has been successfully implemented (i.e. the OLXCs have been reconfigured). It may be desirable to have the acknowledgement confirm that every hop has completed the OLXC configuration. However, to verify that end-to-end connectivity has been established requires that additional mechanisms be implemented. These could for example be tandem connection identification verification, as defined in ITU-T SONET/SDH and OTN. Either way, the channel becomes available immediately after the request is sent, at the discretion of the user. Once established, the lightpath may carry arbitrary traffic, such as ATM, Frame Relay or TDM circuit.

If the user requests a restored lightpath, then capacity must be reserved within the network. This reserved capacity is shared over multiple failures and only allocated (i.e., configured in the OLXC) upon failure. The capacity reservation is performed independently of the setup of the primary lightpath albeit perhaps simultaneously. It may take a significantly longer time than the lightpath setup. The first-hop router is responsible for ensuring that restoration capacity is reserved for all restorable failures. The first-hop router informs the source once this is completed. The establishment of a restored lightpath is completed when the primary capacity is allocated and the restoration capacity is reserved.

## **6.2. Softness of State**

To simplify exception handling, all network state is assumed to be soft unless otherwise stated. This applies in particular to lightpath and restoration state. Soft state has an associated time-to-live, and expires and may be discarded once that time is passed. To avoid expiration the state must be periodically refreshed. To reduce the overhead of the state maintenance, the expiration period may be increased exponentially over time to a predefined maximum. This way the longer a state has survived the fewer the number of refresh messages that are required.

For lightpaths this implies that the source must periodically resend the lightpath request. Similarly, the first-hop router must resend the lightpath setup. If the state of a lightpath expires at a particular node, the state is locally removed and all resources allocated to the lightpath are reclaimed.

## **6.3. Lightpath Routing**

To satisfy the requirements of diverse routing and restoration we assert that it is necessary to use explicit routing for constructing lightpaths. In addition, explicit routes may be valuable for traffic engineering and load optimizations in the network. The route on which a new lightpath is to be established is specified in the



lightpath setup message. This route would typically be chosen by the first-hop router, but could be determined by a pre-authenticated higher level network management system. Through routing protocols the first-hop router has a representation of the full physical network topology and the available resources on each link. These are obtained and updated via OSPF link state advertisements. The explicit route might be carried directly in the IP datagram using the IP source route option, or might be carried in the packet payload as would be the case if RSVP were used for signaling lightpath requests. The route may be specified either as a series of nodes (routers / OLXCs), or in terms of the specific links used (as long as IP addresses are associated with these links).

Numerous policies can be used to route lightpaths through the network, such as constraint-based routing algorithms. It is expected that using a good routing algorithm will produce better route selection and improve network resource utilization.

To ensure diversity in routes, each diversely routed lightpath group is coordinated by a single network entity. To create a diversely routed lightpath group, a user registers with a coordinator, and receives the group identifier. For groups originating through the same first-hop router, this router would typically act as the coordinator. To ensure diversity in routes, K SRLG and node disjoint routes through the network are selected, where K represents the number of diverse routes required. The corresponding lightpaths are then established independently. When a router receives a diversely routed lightpath request coordinated by another network entity, the router uses the address in the diversely routed lightpath group identifier to retrieve the explicit route for the new path from the coordinator.

#### **6.4. Provisioning bi-directional lightpaths**

The construction of a bi-directional lightpath differs from the construction of a uni-directional lightpath above only in that upon receiving the setup request, the last-hop router returns the setup message using the reverse of the explicit route of the forward path. Both directions of a bi-directional lightpath share the same characteristics, i.e., set of nodes, bandwidth and restoration requirements. For more general bi-directional connectivity, a user simply requests multiple individual lightpaths.

#### **6.5. Provisioning lightpaths in a (sub-)network without wavelength converters**

The provisioning techniques proposed earlier in this section apply to optical networks with wavelength conversion. However, future all-optical OLXCs may not have the ability to convert an incoming wavelength to a different outgoing wavelength (i.e. do not implement wavelength conversion). Such OLXCs may be used throughout an optical network, or may be used in only some nodes, creating all-optical sub-networks. Sections of a network that do not have



wavelength converters are thus referred to as being wavelength continuous. A common wavelength must be chosen on each link along a wavelength continuous section of a lightpath. Whatever wavelength is chosen on the first link defines the wavelength allocation along the rest of the section. A wavelength assignment algorithm must thus be used to choose this wavelength. It is plausible, although unlikely, that wavelength conversion could also be eliminated between the client and the network. Wavelength selection within the network must be performed within this subset of client wavelengths.

Optical non-linearities, chromatic dispersion, amplifier spontaneous emission and other factors [7] together limit the scalability of an all-optical network. Routing in such networks will then have to take into account noise accumulation and dispersion to ensure that lightpaths are established with adequate signal qualities. In the following discussion we assume that the all-optical (sub-)network considered is geographically constrained so that all routes will have adequate signal quality, and physical layer attributes can be ignored during routing and wavelength assignment. However, the policies and mechanisms proposed here can be extended to account for physical layer characteristics.

One approach to provisioning in a sub-network without wavelength converters would be to propagate information throughout the network about the state of every wavelength on every link in the network. However, the state required and the overhead involved in maintaining this information would be excessive. By not propagating individual wavelength availability information around the network, we must select a route and wavelength upon which to establish a new lightpath, without detailed knowledge of wavelength availability.

We propose in this case to probe the network to determine an appropriate wavelength choice. We use a probe message to determine available wavelengths along wavelength continuous routes. A vector of the same size as the number of wavelengths on the first link is sent out to each node in turn along the desired route. This vector represents wavelength availability, and is set at the first node to the wavelength availability on the first link along the wavelength continuous section. If a wavelength on a link is not available or does not exist, then this is noted in the wavelength availability vector (i.e. the wavelength is set to being unavailable). Once the entire route has been traversed, the wavelength availability vector will denote the wavelengths that are available on every link along the route. The vector is returned to the source OLXC, and a wavelength is chosen from amongst the available wavelengths using an arbitrary wavelength assignment scheme, such as first-fit [8]. Note that wavelength assignment is performed here using wavelength usage information from only the links along the chosen route. Also, multiple lightpaths can be simultaneously established using the same wavelength availability information.

Alternative techniques can be used for selecting a wavelength, such as attempting to establish a lightpath on successive wavelengths in





turn, or simultaneously attempting to allocate the lightpath on all wavelengths that are available at the source.

The key point is that extensions of the provisioning techniques proposed in this document for optical networks with wavelength converters can be used to implement fast provisioning in networks without wavelength converters, and that the two techniques can interwork in a network with OLXCs with and without wavelength conversion.

## **6.6. Lightpath removal**

A lightpath must be removed when it is no longer required. To achieve this, an explicit release request is sent by the first-hop router along the lightpath route. Each router in the path processes the release message by releasing the resources allocated to the lightpath, and removing the associated state. It is worth noting that the release message is an optimization and need not be sent reliably, as if it is lost or never issued (e.g., due to customer premise equipment failure) the softness of the lightpath state ensures that it will eventually expire and be released.

## **7. Restoration plan**

### **7.1. Restoration in a network with wavelength conversion**

When a restored lightpath is requested, the primary lightpath is established as described above, and the restoration capacity must be reserved. The extent to which a network provider chooses to protect the network depends on which failures can be recovered from. In this discussion we assume that recovery is guaranteed for all individual channel, link and single fiber span failures (i.e., links in a common SRLG). Recovery from node or multiple fiber span failures is not guaranteed.

There are three aspects to restoration: reservation of restoration capacity, failure detection and exception handling. We treat each of these separately, as discussed in the following. We propose a distributed approach to the restoration management.

#### **7.1.1. Failure detection and exception handling**

We treat the handling of failures in an optical network as equivalent to exception handling in advanced programming languages. We equate failures to exceptions. When a component receives an exception (at the lowest level detects a failure), it either handles the exception or throws it up the chain of control. Locally, the chain of control goes from the router to the OLXC. For a lightpath the chain of control goes downstream through the routers. This means that exceptions get thrown from the OLXC to the local router, from there to the upstream router, and then recursively to the router further upstream until the exception is handled.



This approach separates the mechanisms of exception propagation from the policy of deciding who and how the exception is handled, yielding great flexibility in the management of restoration capacity. In general, each lightpath is recovered independently. However, in some situations it may be desirable to handle multiple exceptions as a single unit. For example, if a fiber is cut, all channels may be restored in a single action.

It is worth stressing that restoration capacity is reserved, and not allocated. The capacity reserved for restoration is therefore shared and not dedicated to any particular lightpath. The restoration capacity is either idle or is used for preemptable lightpaths. The use of preemptable lightpaths enables the use of a larger percentage of the total capacity albeit for secondary services. This is particularly attractive for adaptable services, as are common in the Internet, which would benefit from exploiting the restoration capacity under normal operating conditions, but would gracefully adapt to the reduction in capacity during failure.

Since restoration capacity is only reserved, handling the exception translates into allocating the restoration lightpath on failure. This requires efficient setup mechanisms for the construction and allocation of the restoration lightpath to meet the tight restoration timing constraints. Ideally the basic lightpath setup would be suitable for this purpose. Otherwise a separate mechanism must be devised for this purpose. In either case, we believe that it is essential to pre-compute and store the restoration routes. The advantage of using a fast lightpath setup is that a normal setup would be issued from the exception handler, allowing all lightpath specific state, specifically the restoration state, to be stored only at the nodes traversed by the primary lightpath. This significantly reduces the maintenance of the soft restoration state. However, other considerations may dictate which mechanisms are used for setting up the primary lightpath even if those mechanisms are poorly suited for restoration. For example, the processing of explicitly routed RSVP messages may be acceptable to setup primary lightpaths, but appears too costly for meeting restoration timing guarantees. To cope with this, the state for the restoration path may be pre-established along the restoration route, leaving out only the OLXC configuration. This way a simple allocation notification (a touch message) along the restoration path is sufficient to trigger the OLXC configuration. The notification can be forwarded by the router before it is processed, thus avoiding accumulating the processing overhead of each node, allowing for very rapid restoration setup. Data can then be transmitted on the restoration path immediately, with insignificant data loss. Such a router notification is described in [9].

Note that the lightpath establishment message must distinguish between a restoration lightpath and a new lightpath request, so that restoration lightpaths allocate resources out of the preemptable capacity reserved for restoration.



### **7.1.2. Management and reservation of restoration capacity**

The first-hop router selects the restoration route(s), and is responsible for reserving restoration capacity. Numerous policies may be used for determining the lightpath restoration routes. The choice of a good restoration policy is a tradeoff between simplicity, utilization and restoration speed. The simplest approach is to restore only at the first-hop router using a single end-to-end route completely SRLG and node disjoint from the primary lightpath. Such a disjoint route is sufficient for all failures along the primary route. Even if restoring only from the first-hop router, it may be preferable to use different restoration routes depending on which hop of the primary lightpath failed. However for longer lightpaths the delay in exception propagation from the point of failure to the first-hop router may be too excessive, and thus it may be desirable to perform the restoration (handle the exception) at intermediate nodes along the path. The mechanisms above support all of these options.

The first-hop router stores all of the restoration routes for which it is responsible (i.e. for which it is the first hop of the primary lightpath) and calculates the total restoration resources required for these routes on each link in the network and for each different link failure, taking into account risk groups and available resources. This calculation can be performed on-line using a greedy algorithm, thus optimizing the choice of restoration routes conditional on the existing lightpath allocations and reserved restoration capacity. Restoration capacity is reserved on a link for the failure of each single SRLG within the network. Thus, the number of lightpaths that use a given link for restoration will differ depending on which SRLG failure is considered. Restoration resources on a given link must thus be independently reserved for each different link failure within the network. The resources required by a first-hop router,  $s$ , on a given link,  $l$ , for restoration of a failed link  $i$  is denoted here by  $r[s][i](l)$ . The  $r[s][i](l)$  values are transmitted to the links ( $l$ ) at regular intervals and when restoration resource requirements are altered (i.e. for each arriving and departing restored lightpath). In a network with  $L$  links, this requires that  $O(L)$  values be transmitted to link  $l$  from first-hop router  $s$ . The resources reserved on a link for restoration are stored locally at that link. This implies the equivalent of storing a two dimensional array of information for each link  $l$  which documents the number of channels reserved at link  $l$  for each first-hop router and every possible link failure (i.e. requires that  $O(NL)$  values be stored, where  $N$  is the number of nodes / sources, and  $L$  is the number of links in the network). The total number of resources reserved on link  $l$  for restoration is the maximum over all possible fiber span failures (risk groups) of the sum over all first-hop nodes of restoration resources required on each link within the risk group.

Once restoration routes have been determined, a restoration reservation message (in IP packets) is sent to reserve the



restoration capacity on the links along the chosen routes. This is performed in a manner similar to lightpath allocations using explicit routing, with the difference that while capacity is reserved, the OLXCs are not reconfigured. Instead, counts of reserved restoration capacity are updated at each of the links along the route.

As long as provisioning time-scales remain long, it is alternatively viable to do restoration management in a centralized fashion, where a centralized Risk Management Center assumes the responsibility for selecting and maintaining restoration routes. This center would subscribe to routing updates but would in addition need to be informed about the routes used for every lightpath established within the network. This last part becomes infeasible as time-scales shrink.

### **7.1.3. Repair and return to primary lightpaths**

Once a failed link or resource has been repaired, the restoration lightpath is released and the lightpath is restored on the original route. This responsibility is also delegated to the first-hop router, which periodically repeats the original lightpath request until it succeeds. For extended outages, the first-hop router may eventually give up on the primary path, and compute and allocate a new restorable primary route. Reverting back to the primary lightpath route after a failure requires that this capacity remain allocated during the time that the lightpath uses the restoration capacity. The proposal here assumes soft connection states, so that if a lightpath refresh is not periodically received for an established lightpath, then its capacity will be de-allocated. This causes a problem in that these refresh messages will not be received along a primary route downstream of the failure. An explicit notification to the closest node downstream of the failure is needed to temporarily reduce the available capacity to ensure that this capacity is not allocated to new lightpaths during the failure.

### **7.2. Restoration in a network without wavelength converters**

End-to-end restoration is proposed for all-optical networks or sub-networks. If no wavelength conversion is used in the network and on the client / network interface, then the same wavelength will be required for the primary and restoration lightpaths if the client cannot retune its wavelength on failure. Whether or not the client can provide this retuning can be passed as a parameter in the lightpath request.

Wavelength selection on the primary and restoration lightpaths should be simultaneously performed if the same wavelength is required on both of these lightpaths. This requires that the wavelengths available on both of the lightpaths be returned to the first-hop router, and a decision made before either lightpath is established. It also requires that specific wavelengths be reserved for restoration at each node, significantly increasing the state





information required. The issue becomes even more complex in a hybrid transparent and opaque OLXC environment. However, we believe that we should focus on opaque OLXC environment on the first phase while keeping in mind that in the future it may be required to incorporate transparent and mixed optical networks.

## **8. Network reconfiguration**

The above proposal performs the calculation of primary and restoration lightpath routes on-line as the individual requests arrive. The lightpath routes are thus chosen conditional on the existing lightpath allocations. A more optimal set of lightpath routes could be calculated off-line, with all of the requests known and their routes simultaneously calculated. However, as the lightpaths vary over time, the implementation of the "optimal" route choices would likely result in the reconfiguration of lightpath routes being required. Although a large number of lightpath reconfigurations may not be acceptable, it is possible that a limited number of lightpath reconfigurations could dramatically improve the network state, freeing up resources for future lightpath allocations.

For restored lightpaths, rerouting would generally have to be performed within the time limits set for restoration. The lightpath allocation schemes would either be fast enough to make this achievable, or additional mechanisms would be employed to hide the delay in lightpath construction. The number of reconfigurations that a given lightpath experiences should be limited, to ensure that lightpaths don't suffer a constant route fluttering. Lightpath reconfigurations should also be confined only to those lightpaths that are rearrangeable (as identified in the lightpath requests).

## **9. Resource discovery and maintenance**

Topology information is distributed and maintained using standard routing algorithms. On boot, each network node goes through neighbor discovery. By combining neighbor discovery with local configuration, each node creates an inventory of local resources and resource hierarchies, namely: channels, channel capacity, wavelengths, links and SRLGs.

We expect that most of these parameters would be automatically discovered. However, some parameters, such as the SRLG information, may need to be inserted by external means. Once the local inventory is constructed, the node engages in the routing protocol.

### **9.1. Information requirements**

The following information should be stored at each node and must be propagated throughout the network as OSPF link-state information:

- Representation of the current network topology and the link states (which will reflect the wavelength availability). This can be



achieved by associating the following information with the link state:

- total number of active channels (note that if a laser fails, for example, then the channels using this laser become inactive, and are not counted in the total number of active channels)
- number of allocated channels (non-preemptable)
- number of allocated preemptable channels
- number of reserved restoration channels (maximum allocated over all potential SRLG failures within the network)
- Risk groups throughout the network (i.e. which links share risk groups)
- Optional physical layer parameters for each link. These parameters are not expected to be required in a network with 3R signal regeneration, but may be used in all-optical networks.

All of the above information is obtained via OSPF updates, and is propagated throughout the network. Note that we do not inform nodes of which channels are available on a link. Thus, in networks with OLXCs without wavelength converters, decisions at the first-hop router are made without knowledge of wavelength availability. This is done to reduce the state information that needs to be propagated within the network.

In addition to this, extra information would be stored locally (i.e., in the router), including the following list (note that this is not exhaustive):

- IP routing tables
- Additional routing table information containing currently active lightpaths passing through, sourced or destined to this node and the channels that they are allocated
- For each link exiting the OLXC:
  - total capacity (number of channels and their bandwidth)
  - available capacity
  - preemptable capacity
  - number of channels reserved for restoration on this link for each potential link failure within the network and for each first-hop router (if distributed restoration capacity calculations are being done). Thus, if there are L links within the network and N nodes, then there are must be L.N unique values stored here.
  - association between channels and fibers / wavelengths. This is particularly important for OLXCs without wavelength converters and for OLXCs in which lower rate channels are multiplexed onto a common higher rate channel on a common fiber (e.g. four OC-48s multiplexed onto a single OC-192 for transmission).
- The first-hop router maintains for each client:
  - client identification
  - associated lightpath IDs for every established lightpath for this client
  - set of primary and restoration routes associated with each lightpath ID



## **10. Attributes for a lightpath request**

The information conveyed in a client request for lightpath connectivity should include the following parameters:

- globally unique lightpath identifier
- diversely routed lightpath group identifier(s)
- destination address
- source address
- bandwidth requirements (e.g. OC48 or OC192)
- uni-directional / bi-directional
- security object  $\hat{u}$  for authentication
- restoration class: one of (i) restored lightpath, (ii) restored IP connectivity, (iii) not restored, (iv) not restored and preemptable. For Class (i) the lightpath must be restored using another lightpath. IP restored (Class (ii)) assumes that the traffic transported on the lightpath is IP, and may be restored by routing through the network routers if needed and given that routing capacity is available [1].
- wavelength rearrangeability (optional parameter required only for client / network interfaces without wavelength conversion).

Note that the unique lightpath identifier can be assigned by the customer when the lightpath is requested, or can be assigned by the network once the lightpath has been established.

## **11. Interface primitives for IP router and OLXC**

We propose the following interface primitives for communication between the router and the OLXC within a node.

- connect(input link, input channel, output link, output channel): commands sent from the router to the OLXC requesting that the OLXC cross-connect input channel on the input link to the output channel on the output link. Note that one end of the connection can also be a drop port. This is true for the following connection primitives as well.
- disconnect(input link, input channel, output link, output channel): command sent from the router to the OLXC requesting that it disconnect the output channel on the output link from the connected input channel on the input link.
- bridge(input link, input channel, output link, output channel): command sent from the router controller to the OLXC requesting the bridging of a connected input channel on input link to another output channel on output link.
- switch(old input link, old input channel, new input link, new input channel, output link, output channel): switch output port from the currently connected input channel on the input link to the new input channel on the new input link. The switch primitive is equivalent to atomically implementing a disconnect(old input channel, old input link, output channel, output link) followed by a connect(new input link, new input channel, output link, output channel).



- alarm(exception, object): command sent from the OLXC to the router informing it of a failure detected by the OLXC. The object represents the element for which the failure has been detected.

Note that IP packets are also passed by the OLXC to the router in the network when the control packets from clients are transmitted within the framing overheads.

## **12. Summary**

This document outlined how IP algorithms and mechanisms can be used as the basis for a control plane for an optical network. This contribution provides the optical layer requirements that can be the basis for the selection and extension of the proposals on algorithms and protocols. The document illustrated how optical lightpath management, and particularly rapid lightpath provisioning and restoration can be implemented using IP control.

## **13. Acknowledgments**

The authors wish to thank John Strand, Albert Greenberg, Bob Tkach, Bob Doverspike, Evan Goldstein and Jerry Ash for their contributions to this proposal.

## **14. References**

- [1] A. Greenberg, G. Hjßlmt<sup>2</sup>sson and J. Yates, "Smart Routers û Simple Optics: A Network Architecture for IP over WDM," accepted for publication at OFC 2000.
- [2] D. Awduche, Y. Rekhter, J. Drake, R. Coltun, "Multi-Protocol Lambda Switching: Combining MPLS Traffic Engineering Control with Optical Crossconnects," IETF Internet draft.
- [3] D. Awduche, L. Berger, D. Gan, T. Li, G. Swallow, and V. Srinivasan, "Extensions to RSVP for LSP Tunnels," IETF Internet Draft, Work in Progress, 1999.
- [4] B. Jamoussi et al, "Constraint-Based LSP Setup using LDP," IETF Internet Draft, Work in Progress, 1999.
- [5] ITU-T G.872, "Architecture for Optical Transport Networks," 1999.
- [6] ATM Forum, "Private Network-Network Interface Specification: Version 1.0," March 1996.
- [7] R. Tkach, E. Goldstein, J. Nagel and J. Strand, "Fundamental Limits of Optical Transparency," Optical Fiber Communication Conf., pp. 161-162, Feb. 1998.
- [8] I. Chlamtac, A. Ganz and G. Karimi, "Lightpath Communications: An Approach to High Bandwidth Optical WANs," IEEE Trans. On Comms., vol. 40, pp. 1171-1182, July 1992.
- [9] G. Hjßlmt<sup>2</sup>sson, "IPv6 Courtesy-Copy Extension Headers," IETF Internet Draft.



