Network Working Group   K. Kompella (Juniper Networks)
Internet Draft   Y. Rekhter (Cisco Systems)
Expiration Date: January 2001   A. Banerjee (Calient Networks)
                                         J. Drake   (Calient Networks)
                                         G. Bernstein (Ciena)
                                         D. Fedyk   (Nortel Networks)
                                         E. Mannie   (GTS Network)
                                         D. Saha   (Tellium)
                                         V. Sharma   (Tellabs)

IS-IS Extensions in Support of MPL(ambda)S

draft-kompella-isis-ompls-extensions-00.txt

1. Status of this Memo

2. Abstract

This document specifies extensions to the IS-IS routing protocol in support of Multiprotocol Lambda Switching (MPL(ambda)S).

3. Introduction

This document specifies extensions to the IS-IS routing protocol in support of carrying link state information for Multiprotocol Lambda Switching (MPL(ambda)S). For motivations and overall architecture of MPL(ambda)S see [1]. The set of required enhancements to IS-IS are outlined in [2]. This document enhances the routing extensions [3] required to support MPLS Traffic Engineering. Some of these enhancements also need to be carried in the signaling protocols [6].

The organization of the remainder of the document is as follows. In Section 4, we describe the types of links that may be advertised in IS-IS TE LSAs (we call Link State PDUs LSAs to avoid confusion with Label Switched Paths). In Section 5, we define a new set of Type/Length/Value (TLV) triplets, as well as extend a proposed TLV, and describe their formats.

4. MPL(ambda)S Links

In this section we describe the various types of links that can be announced in IS-IS TE LSAs, namely, normal links, non-packet links, and forwarding adjacencies.

4.1. Normal links

If the nodes on both ends of a link can send and receive on a packet-by-packet basis, this link is a normal link. Control packets (IS-IS protocol packets and signaling packets) and data packets can be sent over the link, so nothing special needs to be done.

4.2. Non-packet links

If either node on the end of a bi-directional link cannot multiplex/demultiplex individual packets on that link (see [5]) then that link cannot be used for sending IS-IS hellos and LSAs. In this case, a proxy control channel is needed for sending and receiving control packets. From IS-IS's point of view, the combination of the data link (in the context of this document, a non-packet link is also called a bearer channel) and the control channel is a single link;

the TE attributes associated with this link are those of the bearer channel, but are sent over the control channel.

The association of a bearer channel with a control channel is by configuration. Note that for a bearer channel D to be associated with a control channel C, D and C should have the same end points, and that the same association must be made at both ends. The means by which it is verified that the two ends have the same associations is outside the scope of this document, however, see [7].

If there are several non-packet links between the same pair of nodes, the associated control channels may be logical interfaces over the same physical control link.

4.2.1. Excluding data traffic from control channels

The control channels between nodes in an MPL(ambda)S network, such as optical cross-connects (OXCs) (see [1], [2]), SONET cross-connects and/or routers, are generally meant for control and administrative traffic. These control channels are advertised into IS-IS as normal IS links as mentioned in the previous section; this allows the routing of (for example) RSVP messages and telnet sessions. However, if routers on the edge of the optical domain attempt to forward data traffic over these channels, the channel capacity will quickly be exhausted.

If one assumes that data traffic is sent to BGP destinations, and control traffic to IGP destinations, then one can exclude data traffic from the control plane by restricting BGP nexthop resolution. (It is assumed that OXCs are not BGP speakers.) Suppose that a router R is attempting to install a route to a BGP destination D. R looks up the BGP nexthop for D in its IGP's routing table. Say R finds that the path to the nexthop is over interface I. R then checks if it has an entry in its Link State database associated with the interface I. If it does, and the link is not packet-switch capable (see [5]), R installs a discard route for destination D. Otherwise, R installs (as usual) a route for destination D with nexthop I. Note that R need only do this check if it has packet-switch incapable links; if all of its links are packet-switch capable, then clearly this check is redundant.

Other techniques for excluding data traffic from control channels may also be needed.

4.3. Forwarding Adjacency

An LSR uses MPLS TE procedures to create and maintain an LSP. The LSR then may (under its local configuration control) to announce this LSP as a link into IS-IS. When this link is advertised into the same instance of IS-IS as the one that determines the route taken by the LSP, we call such a link a "forwarding adjacency" (FA). For details about FAs (for example, their TE properties, the methodology for their use in constrained SPF path computation, etc) see [5].

4.4. Link Bundling

For each type of link just described, it is possible to "bundle" several links together, i.e., treat them as a single link from IS-IS's point of view. For example, several normal links can be advertised as a single normal link.

The mechanisms for bundling, including the restrictions on when links can be bundled and the TE attributes of the bundle are described in [4].

5. IS-IS Routing Enhancements

In this section we define the routing enhancements on the various types of links that can be announced in IS-IS TE LSAs, namely, normal links, non-packet links, and forwarding adjacencies. In this document, we enhance the sub-TLVs for the extended IS reachability TLV (see [3]) in support of MPL(ambda)S. Specifically, we add the following sub-TLV:

Sub-TLV Type	Length	Name
20	1	Link Protection Type

We further add two new TLVs.

TLV Type	Length	Name
136 (TBD)	variable	Link Descriptor
138 (TBD)	variable	Shared Risk Link Group

5.1. Link Protection Type sub-TLV

The Link Protection Type sub-TLV represents the protection capability that exists on a link. It is desirable to carry this information so that it may be used by the path computation algorithm to set up LSPs with appropriate protection characteristics. This is a sub-TLV (of type 20) of the extended IS reachability TLV (type 22) as defined in [3].

If the link has 1+1 protection, it means that a disjoint backup bearer channel is reserved and dedicated for protecting the primary bearer channel. This backup bearer channel is not shared by any other connection, and traffic is duplicated and carried simultaneously over both channels.

If the link has 1:N protection, it means that for N primary bearer channels, there is one disjoint backup bearer channel reserved to carry the traffic. Additionally, the protection bearer channel MAY carry low-priority preemptable traffic. The bandwidth of backup bearer channels will be announced in the unreserved bandwidth sub-TLV at the appropriate priority.

If the link has ring protection, it means that the primary bearer channel is protected by the presence of an alternate link possibly using other links and nodes in the network.

The Link Protection Type sub-TLV has a length of two octets, the first of which can take one of the following values:

Value	Link Protection Type
0	Reserved (see signaling draft [6])
1	None
2	1+1
3	1:N
4	Ring

The second octet gives a priority value, such that a new connection with a priority value higher than this value is guaranteed to be setup on a primary (or working) channel, and not on a secondary (or protect) channel.

The Link Protection Type sub-TLV is optional and if an LSA does not carry the TLV, then the Link Protection Type is unknown.

5.2. Link Descriptor TLV

The Link Descriptor TLV represents the characteristics of the link, in particular the link type and the bit transmission rate. These characteristics represent one of the constraints on the type of LSPs that could be carried over a link.

These characteristics should not be confused with the physical link encoding or multiplex structure (if any). For example there are systems where four OC-48s are switched and transported over a single fiber via wavelength division multiplexing (WDM) technology. There are systems where four OC-48s are transported in a transparent OC-192 time division multiplex (TDM) structure, i.e., all the overheads of the OC-48's are persevered. In both these cases the essential information from a routing perspective is that both of the links can transport media of type OC-48.

The proposed Link Descriptor TLV (of type 136 TBD) contains a new data structure consisting of:

    7 octets of System ID and Pseudonode Number
    4 octets of IPv4 interface address
    4 octets of IPv4 neighbor address

and a list of Link Descriptors, where each element in the list has 10 octets. The length of the TLV is thus 15+10*n octets, where n is the number of elements in the list.

Each Link descriptor element consists of the following fields: the first field is a one-octet value which defines the link encoding type, the second field is a one-octet value which defines the lowest priority at which that link encoding type is available, the third field is four-octets and specifies the minimum reservable bandwidth (in IEEE floating point format, the units being bytes per second) for this link encoding type, and the last field is four-octets and specifies the maximum reservable bandwidth (in IEEE floating point format, the units being bytes per second) for this link encoding type. Link encoding type values are taken from the following list:

Value	Link Encoding Type
1	Standard SONET
2	Arbitrary SONET
3	Standard SDH
4	Arbitrary SDH
5	Clear
6	GigE
7	10GigE

A link having Standard SONET (or Standard SDH) link encoding type can switch data at a minimum rate, which is given by the Minimum

Reservable Bandwidth on that link, and the maximum rate is given by the Maximum Reservable Bandwidth on that link. In other words, the Minimum and Maximum Reservable Bandwidth represents the leaf and the root of one branch within the structure of the SONET (or SDH) hierarchy. An LSP on that link may reserve any bit transmission rate that is allowed by the SONET (or SDH) hierarchy between the minimum and maximum reservable values (the spectrum is discrete). For example, consider a branch of SONET multiplexing tree : VT-1.5, STS-1, STS-3c, STS-12c, STS-48c, STS-192c. If it is possible to establish all these connections on a OC-192 link, it can be advertised as follows: Minimum Reservable Bandwidth VT-1.5 and Maximum Reservable Bandwidth STS-192.

A link having Arbitrary SONET (or Arbitrary SDH) link encoding type can switch data at a minimum rate, which is given by the Minimum Reservable Bandwidth on that link, and the maximum rate is given by the Maximum Reservable Bandwidth on that link. An LSP on that link may reserve any bit transmission rate that is a multiple of the Minimum Reservable Bandwidth between the minimum and maximum reservable values (the spectrum is discrete).

To handle the case where a link could support multiple branches of the SONET (or SDH) multiplexing hierarchy, one could use multiple LSP descriptors. For example, if a link supports VT-1.5 and VT-2 (which are not part of same branch of SONET multiplexing tree), then it could advertise two LSP descriptors one for each one.

For a link with Clear encoding, the minimum and maximum reservable bandwidth will imply that the optical path is tuned to carry traffic within those range of values. (Note that it should be possible to carry OC-x as well as GigE or any other encoding format, as long as the bit transmission rate of the data to be carried is within this range).

For other encoding types, the minimum and maximum reservable bandwidth should be set to have the same values.

Link Descriptors present a new constraint for LSP path computation.

On a bundled link we assume that either the whole link is configured with the Link Descriptor Types, or each of its component links are configured with the Link Descriptor Types. In the latter case, the Link Descriptor Type of the bundled link is set to the set union of the Link Descriptor Types for all the component links.

It is possible that Link Descriptor TLV will change over time, reflecting the allocation/deallocation of component links. In general, creation/deletion of an LSP on a link doesn't necessarily

result in changing the Link Descriptor of that link. For example, assume that STS-1, STS-3c, STS-12c, STS-48c and STS-192c LSPs can be established on a OC-192 link whose Link Type is SONET. Thus, initially in the Link Descriptor the minimum reservable bandwidth is set to STS-1, and the maximum reservable bandwidth is set of STS-192. As soon as an LSP of STS-1 size is established on the link, it is no longer capable of STS-192c. Therefore, the node advertises a modified Link Descriptor indicating that the maximum reservable bandwidth is no longer STS-192, but STS-48. If subsequently there is another STS-1 LSP, there is no change in the Link Descriptor. The Link Descriptor remains the same until the node can no longer establish a STS-48c LSP over the link (which means that at this point more than 144 time slots are taken by LSPs on the link). Once this happened, the Link Descriptor is modified again, and the modified Link Descriptor is advertised to other nodes.

Note that changes to the Link Descriptor TLV will also affect the Unreserved Bandwidth sub-TLV with respect to bandwidth available on the link.

5.3. Shared Risk Link Group TLV

A set of links may constitute a 'shared risk link group' (SRLG) if they share a resource whose failure may affect all links in the set. For example, two fibers in the same conduit would be in the same SRLG. A link may belong to multiple SRLGs. Thus the SRLG TLV describes a list of SRLGs that the link belongs to. An SRLG is identified by a 32 bit number that is unique within an IGP domain.

The SRLG of a LSP is the union of the SRLGs of the links in the LSP. The SRLG of a bundled link is the union of the SRLGs of all the component links. The SRLG values are an unordered list of 4 octet numbers that the link belongs to.

If an LSR is required to have multiple diversely routed LSPs to another LSR, the path computation should attempt to route the paths so that they do not have any links in common, and such that the path SRLGs are disjoint.

The proposed SRLG (of type 138 TBD) contains a new data structure consisting of:

    7 octets of System ID and Pseudonode Number
    4 octets of IPv4 interface address
    4 octets of IPv4 neighbor address

and a list of SRLG values, where each element in the list has 4 octets. The length of the TLV is thus 15+4*n octets, where n is the number of elements in the list.

The SRLG TLV starts with a configured value and does not change over time, unless manually reconfigured. The SRLG TLV is optional and if an LSA doesn't carry the SRLG TLV, then it means that SRLG of that link is unknown.

6. Security Considerations

The sub-TLVs proposed in this document does not raise any new security concerns.

7. Acknowledgements

The authors would like to thank Suresh Katukam, Jonathan Lang and Quaizar Vohra for their comments on the draft.

8. References

[1] Awduche, D., Rekhter, Y., Drake, J., Coltun, R., "Multi-Protocol Lambda Switching: Combining MPLS Traffic Engineering Control With Optical Crossconnects", draft-awduche-mpls-te-optical-01.txt (work in progress)

[2] Basak, D., Awduche, D., Drake, J., Rekhter, Y., "Multi-protocol Lambda Switching: Issues in Combining MPLS Traffic Engineering Control With Optical Crossconnects", draft-basak-mpls-oxc-issues-00.txt (work in progress)

[3] Smit, H., Li, T., "IS-IS Extensions for Traffic Engineering", draft-ietf-isis-traffic-01.txt (work in progress)

[4] Kompella, K., Rekhter, Y., Berger, L., "Link Bundling in MPLS Traffic Engineering", draft-kompella-mpls-bundle-02.txt (work in progress)

[5] Kompella, K., Rekhter, Y., "LSP Hierarchy with MPLS TE", draft-ietf-mpls-lsp-hierarchy-00.txt (work in progress)

[6] Optical MPLS Group, "MPLS Optical/Switching Signaling Functional Description" (work in progress)

[7] Lang J., Mitra K., Drake J., Kompella K., Rekhter Y., Saha D., Berger L., and Basak D., "Link Management Protocol", draft-lang-mpls-lmp-00.txt (work in progress)

9. Authors' Information

Kireeti Kompella
Juniper Networks, Inc.
1194 N. Mathilda Ave
Sunnyvale, CA 94089
Email: kireeti@juniper.net

Yakov Rekhter
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134
Email: yakov@cisco.com

Ayan Banerjee
Calient Networks
5853 Rue Ferrari
San Jose, CA 95138
Phone: +1.408.972.3645
Email: abanerjee@calient.net

John Drake
Calient Networks
5853 Rue Ferrari
San Jose, CA 95138
Phone: (408) 972-3720
Email: jdrake@calient.net

Greg Bernstein
Ciena Corporation
10480 Ridgeview Court
Cupertino, CA 94014
Phone: (408) 366-4713
Email: greg@ciena.com

Don Fedyk
Nortel Networks Corp.
600 Technology Park Drive
Billerica, MA 01821
Phone: +1-978-288-4506
Email: dwfedyk@nortelnetworks.com

Eric Mannie
GTS Network Services
RDI Department, Core Network Technology Group
Terhulpsesteenweg, 6A
1560 Hoeilaart, Belgium
Phone: +32-2-658.56.52
E-mail: eric.mannie@gtsgroup.com

Debanjan Saha
Tellium Optical Systems
2 Crescent Place
P.O. Box 901
Ocean Port, NJ 07757
Phone: (732) 923-4264
Email: dsaha@tellium.com

Vishal Sharma
Tellabs Research Center
One Kendall Square
Bldg. 100, Ste. 121
Cambridge, MA 02139-1562
Phone: (617) 577-8760
Email: Vishal.Sharma@tellabs.com