CCAMP Working Group
Internet Draft
Intended status: Standards Track

V.Beeram (Ed)
I.Bryskin
W.Doonan
ADVA Optical Networking
J.Drake (Ed)
G.Grammel
Juniper Networks
Manuel Paul
Ruediger Kunze
Deutsche Telekom
Friedrich Armbruster
Cyril Magaria
NSN
Oscar González de Dios
Telefonica

Expires: September 12, 2012

March 12, 2012

GMPLS-UNI BCP
draft-beeram-ccamp-gmpls-uni-bcp-01.txt

Status of this Memo

Copyright Notice

Abstract

This document pools together the best current practices that are being used to apply the GMPLS Overlay model at the User-Network Interface (UNI) reference point (as defined in [G.8080])

Table of Contents

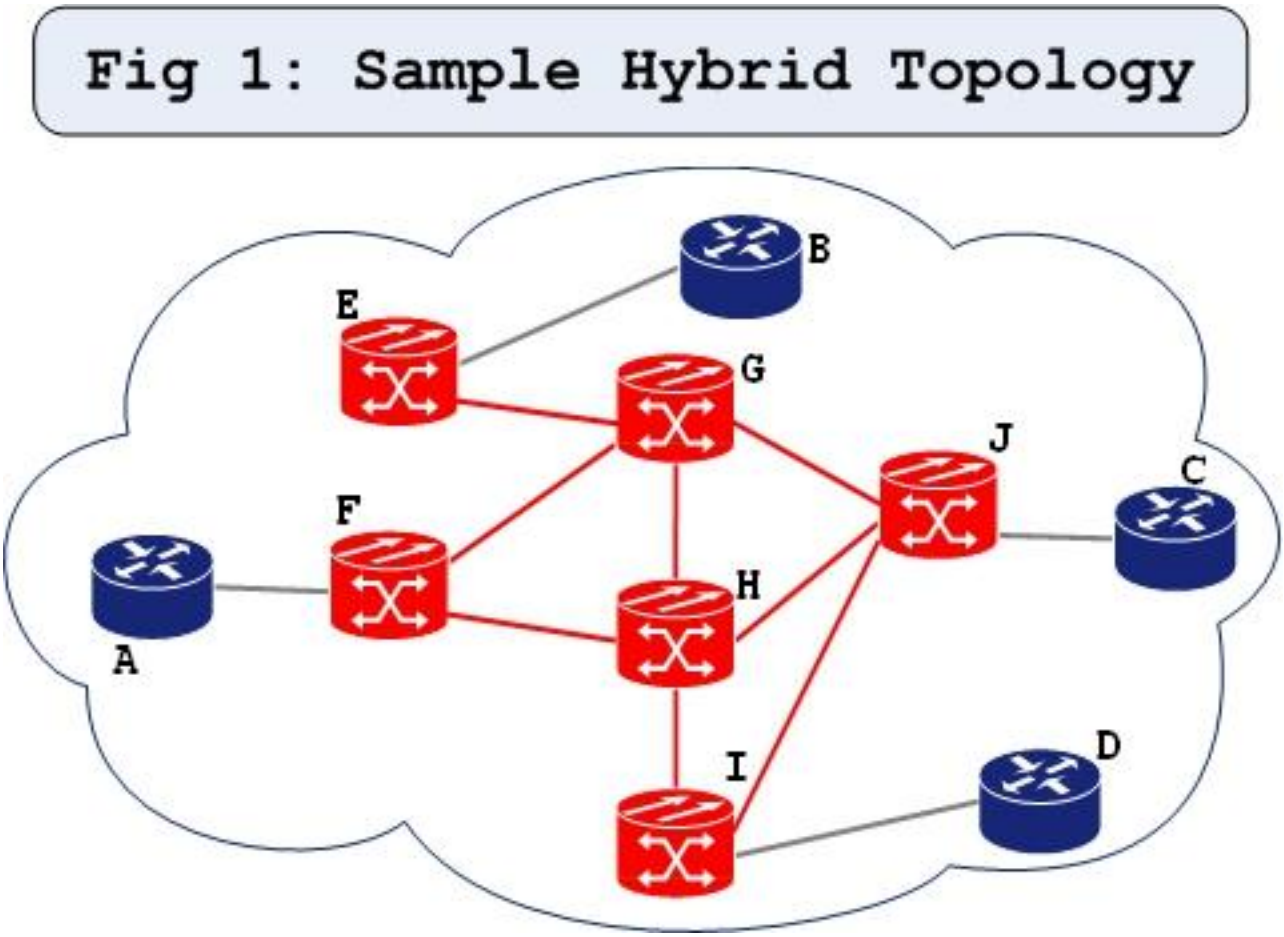
1. Introduction

Generalized Multiprotocol Label Switching (GMPLS) provides tools to create end-to-end services in various transport technologies. These tools can be used to support service management in different types of deployment models. [RFC 4208] discusses how GMPLS can be applied to the overlay model. There are a good number of implementations that have built on the basic concepts discussed in [RFC 4208] and have successfully demonstrated interoperability. This document is an attempt to pool together the best current practices that are being used to apply the GMPLS Overlay model at the User-Network Interface (UNI) reference point (as defined in [G.8080]).

[RFC 4208] recommends the use of hierarchical service activation when GMPLS is used for the core network and section 7.3.3 of [RFC4847], “Virtual Link Service Model” augments this by introducing a representation of server-layer network resources into a client-layer network topology. This memo explains how this augmentation enhances client-layer networking in an overlay model. The concepts discussed in this document are based primarily on experiences drawn from interoperating GMPLS-enabled IP routers with Optical Transport elements, but any GMPLS supported technology may be used in the client and server-layer networks.

2. Multi-Layered Approach

When an end-to-end service crosses a boundary between two regions of dissimilar transport technology, it is necessary to execute distinct forms of service activation within each region.



For example, in the hybrid network illustrated in Fig 1, provisioning a transport service between two GMPLS-enabled IP routers on either side of the optical WDM transport topology requires operations in two distinct layer networks; the client-layer network interconnecting the routers themselves, and the server-layer network interconnecting the optical transport elements in between the routers.

Activation of the end-to-end service begins with a path determination process, followed by the initiation of a signaling process from the ingress along the determined path, per the set of figures shown in Fig 2.

Fig 2: Hierarchical Service Activation

Fig 2a: Client-Layer Service setup is initiated

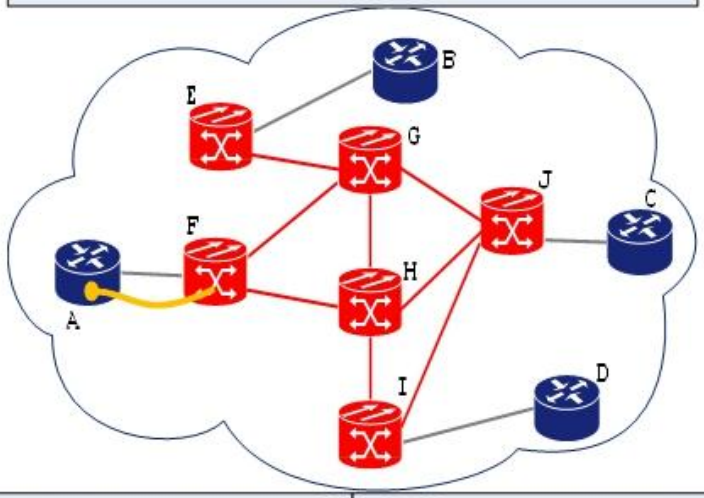


Fig 2b: Server-Layer WDM Service that caters to the client-layer service is established within the core

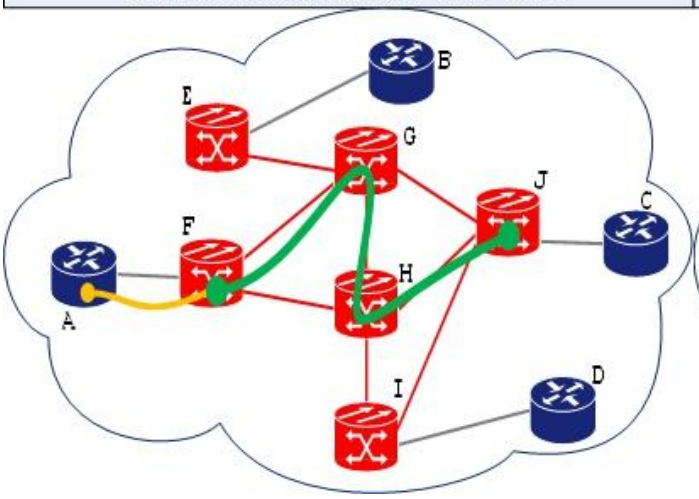
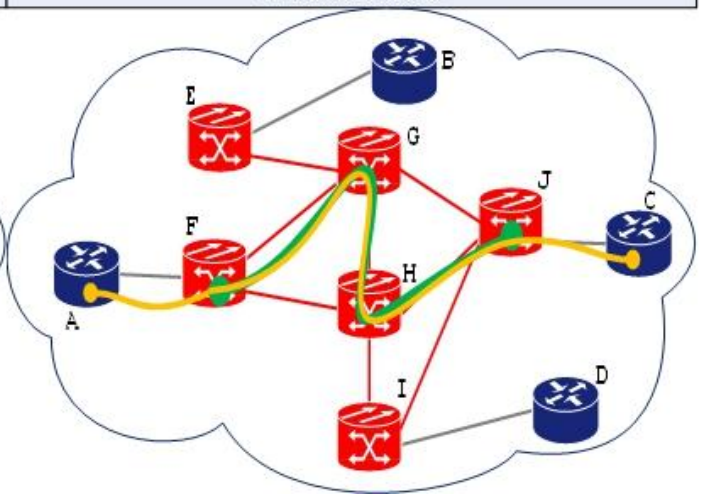


Fig 2c: Client-Layer Service setup is resumed and the end-to-end connection is established



3. Traffic Engineering

The previous section outlines the basic method for activating end-to-end services across a multi-layer network. As a necessary part of that process an initial path selection process was performed, whereby an appropriate path between the desired endpoints was determined through some means. Further, per expectations set through current practices with regard to service provisioning in homogeneous networks, operators expect that the underlying control plane system will provide automated mechanisms for computing the desired path or paths between network endpoints.

In particular, operators do not expect under normal circumstances to be required to explicitly specify the end-to-end path; rather, operators expect to be able to specify just the endpoints of the path and rely on an automated computational process to identify and qualify all the elements and links on the path between them. Hence when operating a hybrid network such as that described in Fig 1, it is necessary to extend existing traffic engineering and path computation mechanisms to operate in a similar manner.

Path computation and qualification operations occur at the path computation element (PCE) selected by ingress element of an end-to-end service. In order to be able to compute and qualify paths, the PCE must SHOULD be provided with information regarding the traffic engineering capabilities of the layer network to which it is associated, in particular the topology of the layer network and what layer-specific transport capabilities exist at the various nodes and links in that topology.

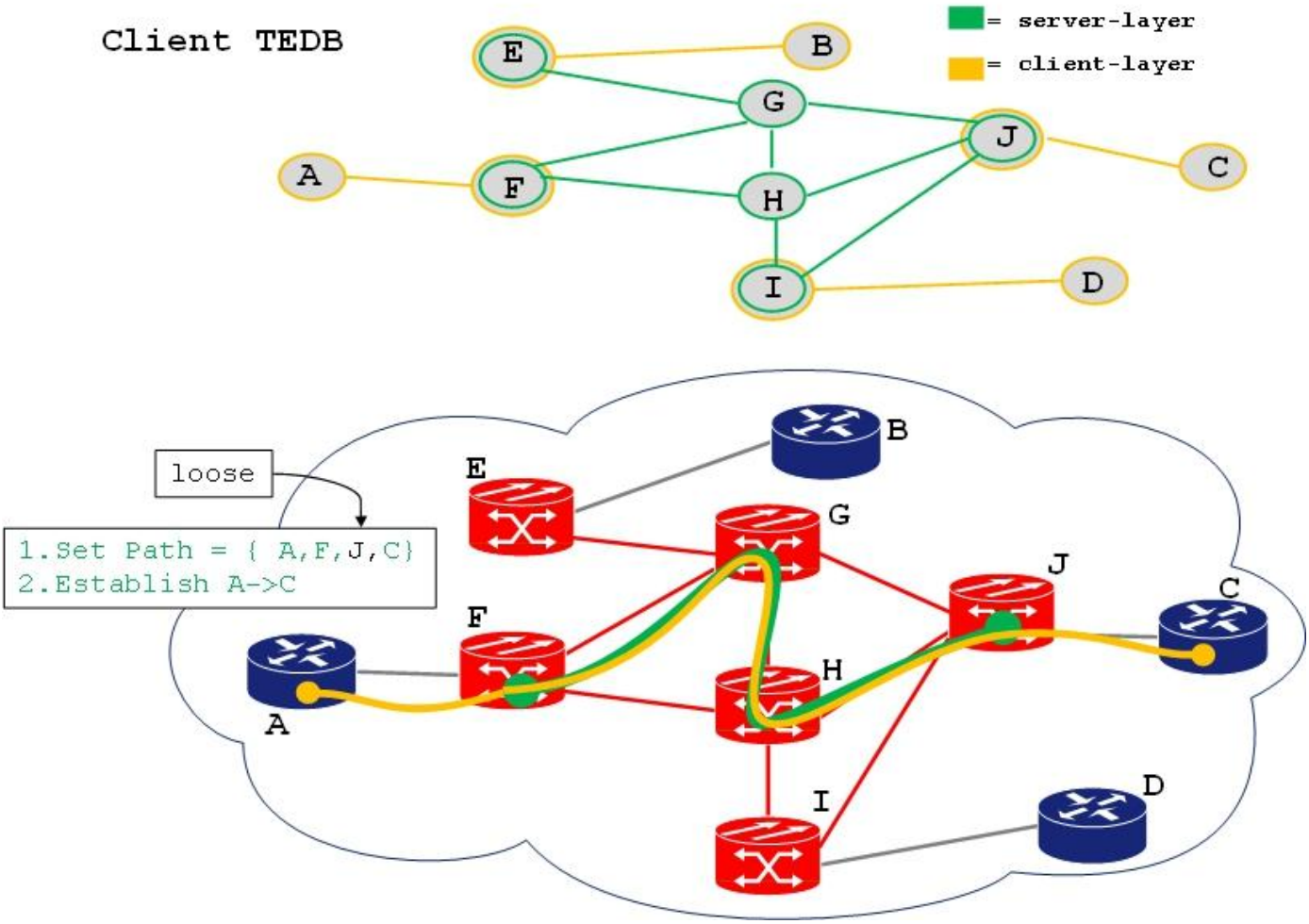
It is important to note that topology information is layer-specific; e.g. path computation and qualification operations occur within a given layer, and hence information about topology and resource availability are required for the specific layer to which the connection belongs. The topology and resource availability information required by elements in the client-layer is quite distinct from that required by the elements in the server-layer network. Hence, the server-layer traffic engineering links are of no importance for the client-layer network, and it is actually desirable to block their advertisements into the client TE domain by the server-layer border nodes.

For example, in the sample hybrid network (Fig 1) there are multiple optical transport elements supporting the connection between the GMPLS-enabled IP routers, and hence the physical topology between them includes several nodes and links. However, the optical

elements between the IP routers are not able to switch traffic within the client-layer network of routers (e.g. IP/MPLS), as the optical elements are lambda switches, not IP/MPLS switches. Hence while the intervening optical elements may physically exist along the path, they are not a part of the topology available to the IP/MPLS routers for the purposes of traffic engineering in the client-layer network.

An example of what the client-layer Traffic Engineering topology would look like for the sample hybrid network is shown in the top half of Fig 3.

Fig 3:Traffic Engineering - ERO with "loose hop"



In this example, the TE topology associated with the client-layer network is indicated by the links and nodes colored yellow, whereas the TE topology associated with the server-layer network is indicated by the links and nodes colored green. The nodes at the edge of the server-layer network are visible in both the topologies. The yellow topology is capable of switching traffic within the client-layer, whereas the green topology is capable of switching traffic within the server-layer.

In this example, if the "B" router attempts to determine a path to the "D" router it will be unable to do so, as the yellow topology to

which the B and D routers is connected does not include a fully-yellow path between them. The only way to setup an end-to-end path in this case is to use an ERO with a "loose hop" across the server-layer domain as illustrated in Fig 3. This would cause the server-layer to create the necessary link in the client-layer topology on the fly. However, this approach has a few drawbacks - [a] the necessity for the operator to specify the ERO with the "loose" hop; [b] potential sub-optimal usage of server-layer network resources; and [c] unpredictability with regard to the fate-sharing of the new link (that is created on the fly) with other links of the client-layer topology.

In order to be able to compute an end-to-end path between the two client-layer endpoints, the yellow topology MUST be sufficiently augmented to indicate where there are paths through the green topology which can provide connectivity between nodes in the yellow topology. In other words, in order for a client to compute path(s) across the server-layer network to other clients, the feasible paths across the server-layer network SHOULD be periodically computed by the server-layer network and made available (in terms of TE links and nodes that exist in the client-layer network) to all the clients. This is discussed in detail in the next section.

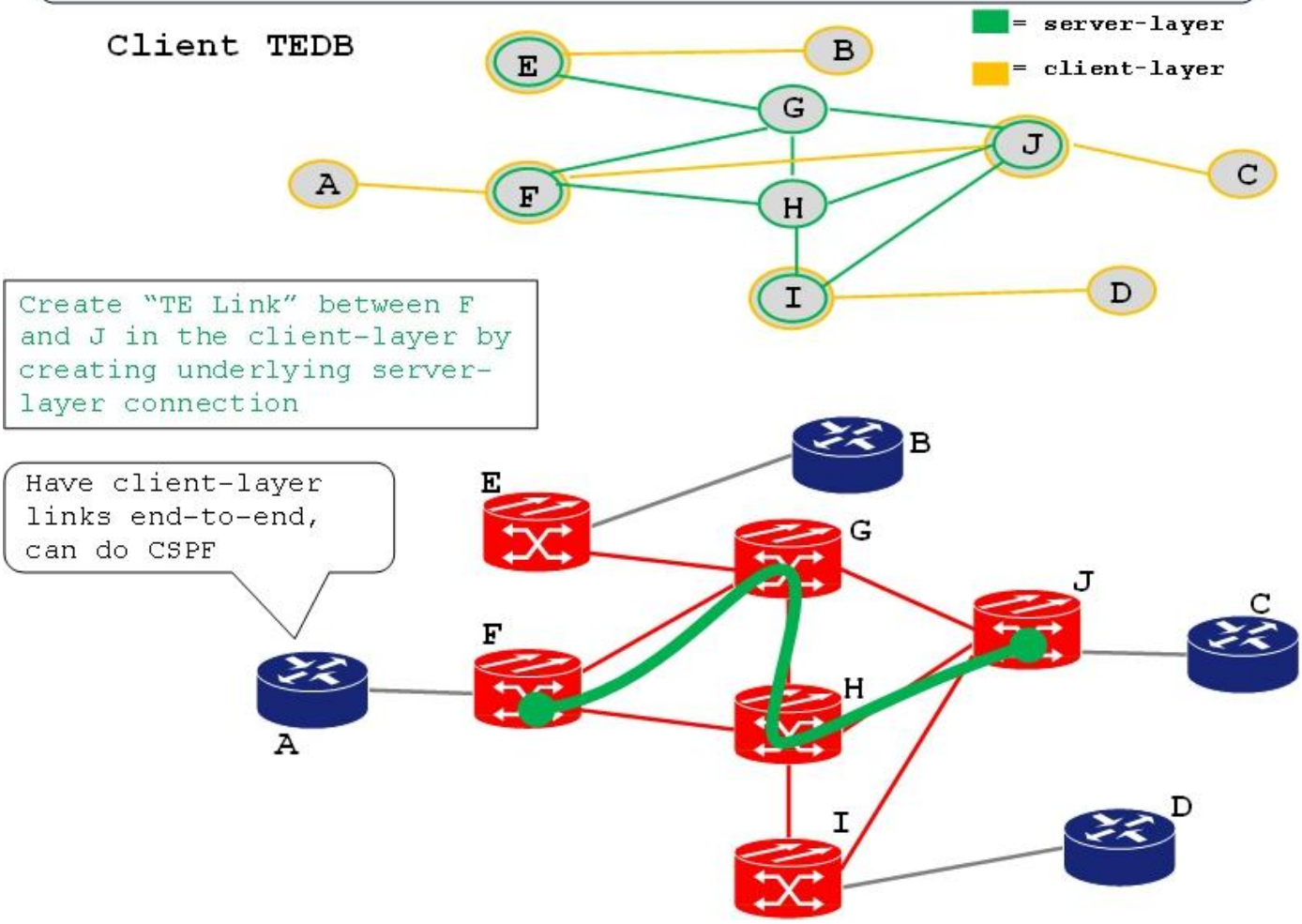
In the overlay model the client and network domains, generally speaking, exist in separate layer networks. One important use case, however, is when the client and network topologies are in the same layer network. For example, IP routers that are connected via GMPLS UNI to a WDM network may be capable of terminating optical trails that are lambda switched by the network. Because the network domain normally would not want to leak its actual topology information into the client domain, clients would not be able to compute end-to-end paths across the network domain despite that client and network links belong to the same (WDM) layer network. The method described in the following sections of this document solves the problem of partitioned client topology for this case as well.

3.1. Augmenting the Client-Layer Topology

In the example hybrid network shown below in Fig 4, consider a scenario where each GMPLS-enabled IP router is connected to the optical WDM transport network via a transponder. Further consider the situation where the transponder at node F can be connected to the transponder in node J via the optical path F-G-H-J. A lambda LSP can be provisioned in the server-layer along this path, and then

advertised as a TE link into the client-layer. With the availability of this link, the path computation function at node A is able to

Fig 4: Traffic Engineering - End to End Path Computation



compute an end-to-end path from A to C.

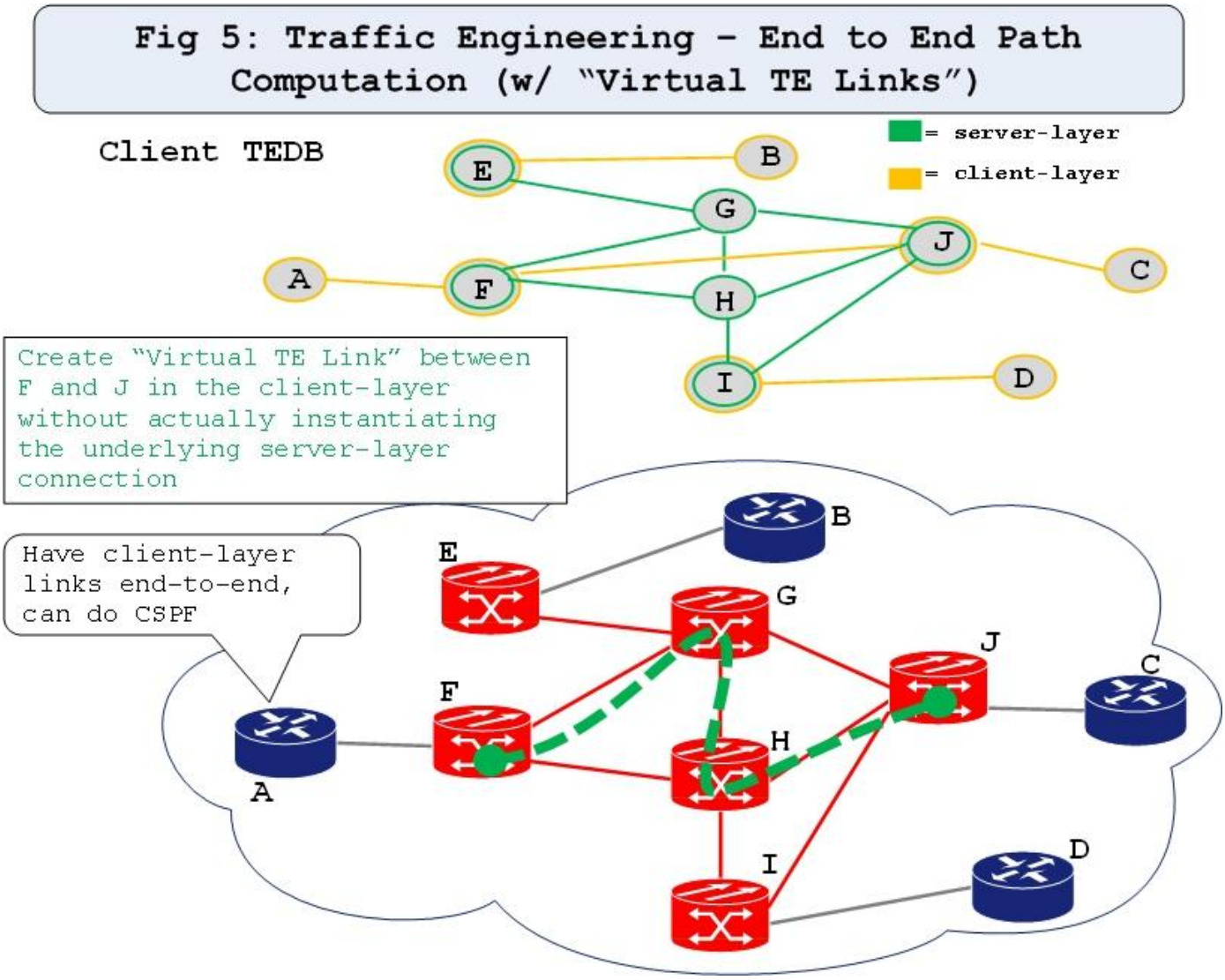
In this case, in order for the TE link to be made available in the client-layer network topology, the network resources corresponding to the underlying server-layer LSP MUST be fully provisioned beforehand.

As another scenario, consider a network configuration where the transponders at nodes E, F, J and I are connected to each other via directionless ROADM components. It is physically possible to connect any transponder to any other transponder in the server-layer network. As there are transport capabilities available in the server-layer network between every element containing an adaptation function to the client-layer network, the operator in this case would not wish to reserve any network resources in the server-layer network until a client LSP is signaled. The next section proposes a method to address this common operational requirement.

3.1.1. Virtual TE Links

A “Virtual TE Link” as defined in section 7.3.3 of [RFC4847] is a TE link that is advertised into the client-layer network, with the available but not necessarily reserved/commuted resources in the server-layer network necessary to support that TE link. In other words, “Virtual TE Links” represent specific transport capabilities available in the server-layer network which can support the establishment of LSPs in the client-layer network.

The two fundamental properties of a Virtual TE Link are: [a] it is advertised just like a real TE link and thus contributes to the buildup of the client-layer network topology; and [b] it does not require allocation of resources at the server-layer until used, thus allowing the sharing of server-layer network resources with other Virtual TE links.



In the example shown in Fig 5, the availability of a lambda channel along the path F-G-H-J results in the advertisement by nodes F and J of a Virtual TE Link between F and J into the client-layer network topology (yellow line). With the advertisement of this Virtual TE Link, the path computation function at node A is able to compute an end-to-end path from A to C.

Whenever a Virtual TE Link gets selected and signaled in the ERO of a client-layer connection, it ceases temporarily to be "virtual" and transforms into a regular TE-link. When this transformation takes

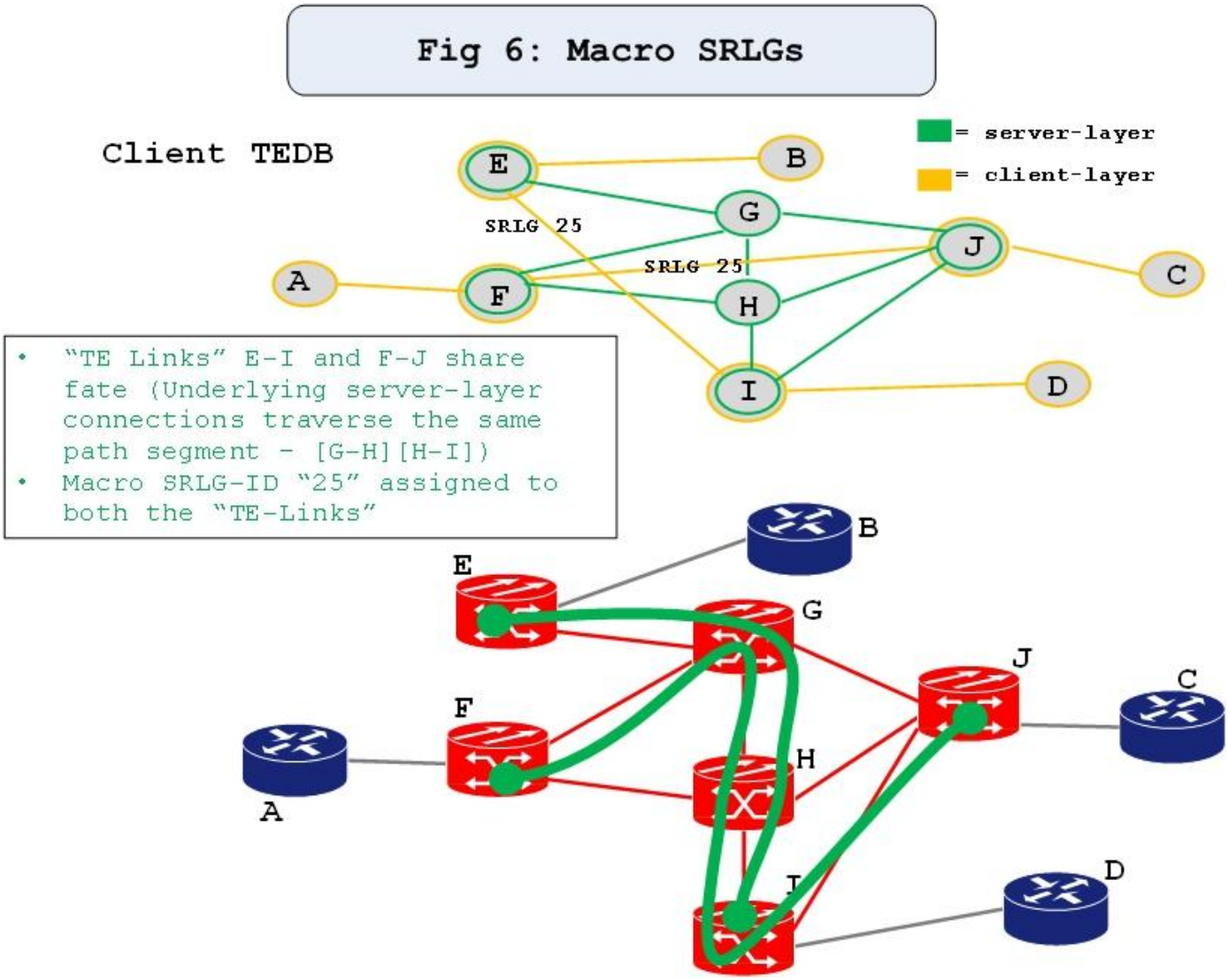
place, the clients will notice the change in the advertised available bandwidth of this TE-link. Also, all other Virtual TE links that share resources with the TE-link in question start advertising “zero” available bandwidth. Likewise, the TE network image reverts back to the original form as soon as the last client-layer connection, going through the TE link in question, is released, i.e. Virtual TE Link becomes “virtual” again

3.2. Macro SRLGs

The Virtual TE links that are advertised into the client-layer network topology cannot be assumed to be totally independent. It is quite possible for a given Virtual TE Link to share fate with one or more other Virtual TE Link(s). This is because the underlying server-layer LSPs (real or potential) can traverse the same server-layer network link and/or node, and failure of any such shared link/node would make all such LSPs inoperable (along with the Virtual TE Links supported by the LSPs). If diverse end-to-end paths for client-layer LSPs are to be computed, the fate-sharing information of the Virtual TE Links needs to be taken into account. The standard way of addressing this problem is to use SRLGs as a part of Virtual TE Link advertisements.

A traditional SRLG represents a shared physical network resource upon which normal function of a link depends. Such SRLGs can also be referred to as physical SRLGs. Zero, one or more physical SRLGs could be identified and advertised for every TE link in a given layer network. However, there is a scalability issue with physical SRLGs in multi-layer environments. For example, if a WDM layer LSP serves an IP layer link, every WDM link and node traversed by the LSP MUST be considered as a separate SRLG. The number of SRLGs to be advertised to client (e.g. IP) layer per TE link would be directly proportional to the number of hops traversed by the underlying server-layer LSP.

The notion of Macro SRLGs addresses this scaling problem. Macro SRLGs have the same protocol format as their physical counterparts and can be assigned automatically for each Virtual TE Link that is advertised into the client-layer network as a result of the existence of an underlying server-layer LSP (instantiated or otherwise). A Macro SRLG represents a set of shared path segments that are traversed by two or more of the underlying server-layer LSPs. Each shared path segment can be viewed as a sequence of shared resources where each individual resource has a physical SRLG associated with it (example depicted in Fig 6). The actual procedure for deriving these Macro SRLGs is beyond the scope of this document.



3.3. MELGs

If two or more Virtual TE Links share fate, it means that the links could be concurrently activated and used by client LSPs with a caveat that the links could be taken out of service by a single network failure, and, thus, cannot be used in the same protection scheme. There could be a stronger (than fate sharing) relationship between two or more Virtual TE Links. Because a set of Virtual TE Links could be mapped onto the same uncommitted network resources, the situation can arise when only one Virtual TE Link from the set

could be activated at any given time. In other words, two or more Virtual TE Links could be mutually exclusive.

One example of mutually exclusive Virtual TE Links is when the paths for the network domain LSPs supporting the Virtual TE Links not only intersect, but also require usage of the same resource (e.g. lambda channel) on the intersection. Another example is when the said paths depend on a common physical resource (e.g. transponder, regenerator, wavelength converter, etc.) that could be used only by one LSP at a time.

For a client path computation function (especially a centralized one capable of concurrent computation of multiple end-to-end paths) it is important to know about such mutually exclusive relationship between Virtual TE Links. This memo introduces a concept of Mutually Exclusive Link Group (MELG) and suggests a new sub-TLV – MELGs sub-TLV – to be added to the top level TE Link TLV. The purpose of the MELGs sub-TLV is:

- To indicate via a separate network unique number (MELG ID) an element or a situation that makes the advertised Virtual TE Link to belong to one or more mutually exclusive link groups. Path computer will be able to decide on whether two or more Virtual TE Links are mutually exclusive or not by finding the overlap of advertised MELGs (similar to deciding on whether two or more TE Links share fate or not by finding common SRLGs)
- To indicate whether the advertised Virtual TE link is committed or not at the moment of the advertising. Such bit of information is important for a path computer: committing new Virtual TE links (vs. re-using committed ones) has a consequence of committing more network resources and disabling other Virtual TE links that have common MELGs with newly committed Virtual TE Link

Exact format of the MELGs sub-TLV is described in [MELG]

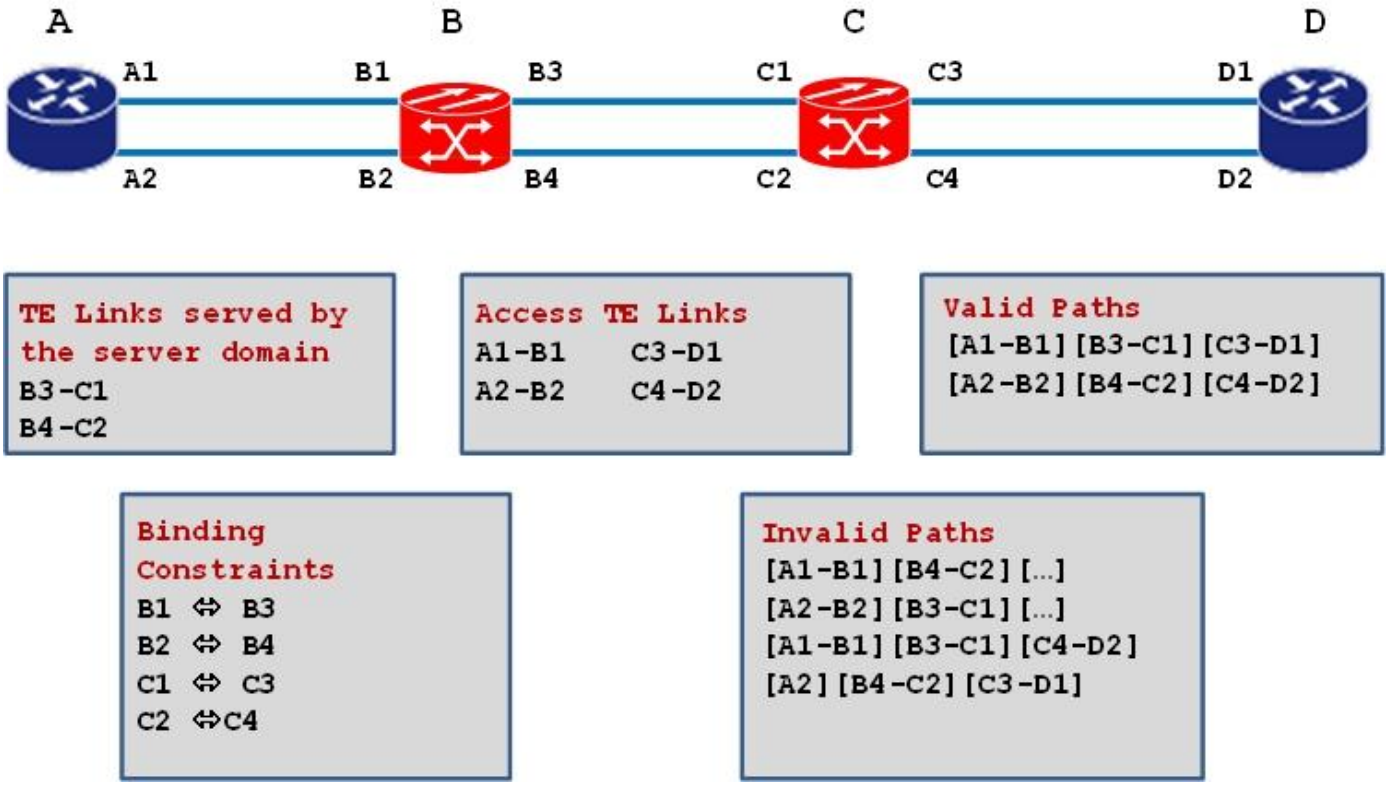
[TBD: MELG Figure/Example]

3.4. Switching Constraints

Certain types of network configurations necessitate the specification of connectivity constraints in the Virtual TE Link advertisements. If the switching constraints associated with the binding of Virtual and access TE links terminated on a given network border node do not get advertised into the client domain, there is a

risk of an invalid path being computed (Fig 7). This document recommends the use of the extensions specified in [GEN_CNSTR] to address this issue.

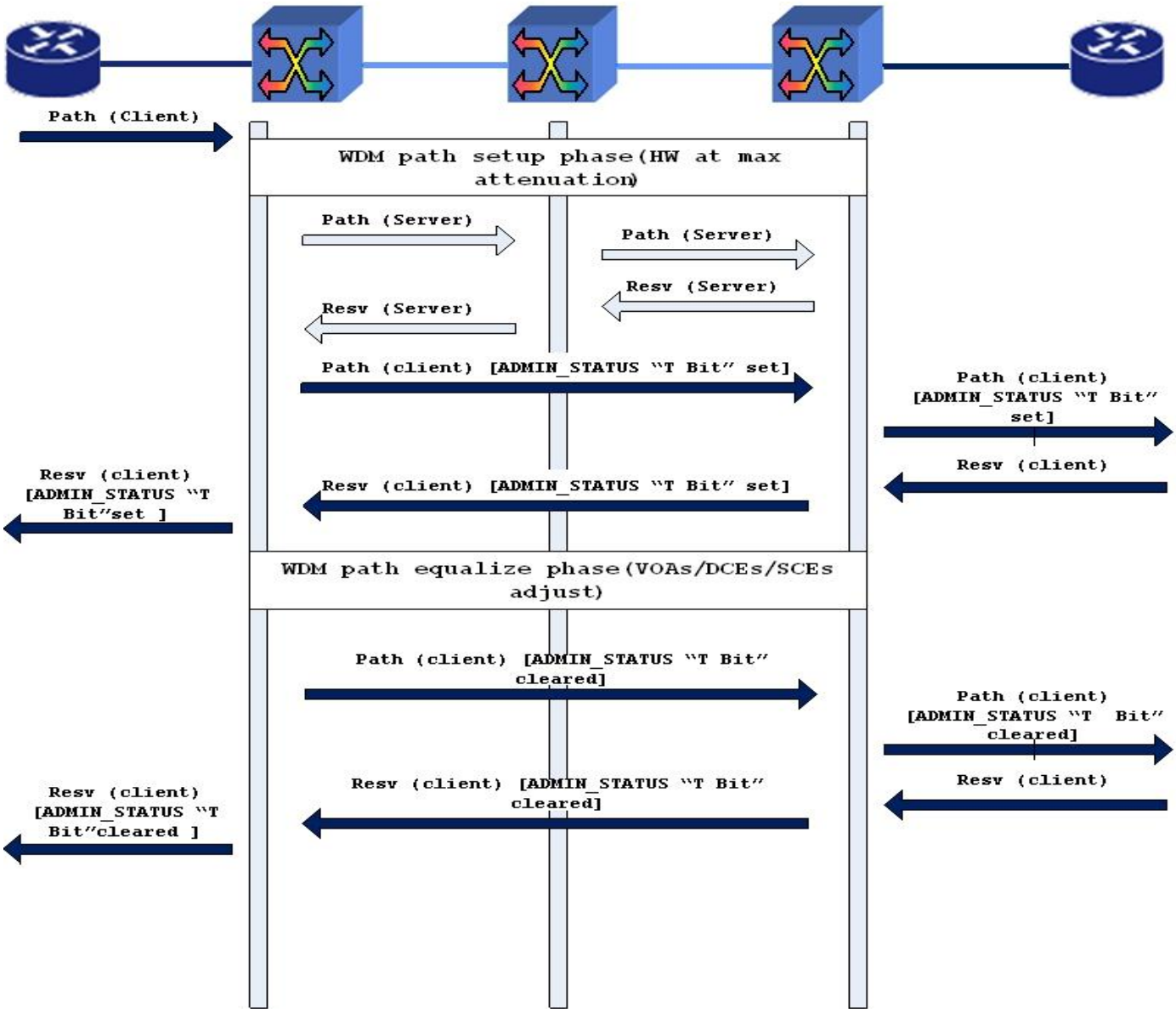
Fig 7: Switching Constraints



4. Connection Setup

Experience with control plane operations in multi-layer networks indicates there are benefits to coordinating certain signaling operations, in the following manner. Consider the scenario where the network domain is a WDM layer topology comprising of ROADMs. The set-up time for a service at the WDM layer can be fairly long, as it can involve time-consuming power-equalization procedures, amongst other layer-specific operations. This means that at very least, the setup timers for the client-layer service would need to be somehow coordinated with that of the server-layer service. To avoid this operationally awkward issue, a phased connection setup process as depicted in Fig 8 is proposed.

Fig 8: Phased Connection Setup



As long as the LSP segment across the server-layer network is not completely "UP" (e.g., Fully Power Equalized), the nodes at the edge of the server-layer network through which the LSP passes would signal the client-layer PATH/RESV messages with the T (Testing) bit set in the ADMIN_STATUS. The T bit would be cleared in these

messages only after the LSP segment across the server-layer network is deemed fully operable.

5. Path computation aspects

It is assumed that a client domain path computation function makes use of advertised client domain TE links as well as Virtual TE Links while computing end-to-end paths for client LSPs. The said path computation function could be local (i.e. located on client LSP ingress nodes, (Corresponding to RFC4655 Composite PCE node) or remote (i.e. network/External PCEs). Path computations could be triggered by client nodes or NMS. Generally speaking, the responsibility of the client domain path computation function is to compute one or two paths for each source-destination pair of the TE-LSPs. Path computation SHOULD be subject to one or more path optimization criterions (such as shortest path, minimal latency, etc.) and path computation constraints (e.g. link unreserved bandwidth, link colors, layer-specific constraints, explicit exclusions, etc.)

As the augmented topology does hide server layer links and nodes, it is RECOMMENDED to support SRLG diverse path computation.

Furthermore the path computation SHOULD consider the connectivity and switching constraint in addition to all usual TE path computation constraints (e.g. unreserved bandwidth, link colors, layer-specific constraint) when available.

When using PCE architecture and PCEP protocol those aspects are covered by RFC5440, RFC5521 and RFC5541.

As described in section 3.3. Virtual TE link may not only share risk but may also depend on the same also server layer resources, thus creating mutual exclusivity between Virtual TE Links. Therefore, network topologies containing Virtual TE links have an increased probability of LSP setup failures. In such topologies concurrent path computation that takes in consideration MELG will reduce signaling failures (Not considering MELGs may result, for example, in two LSPs routed on two Virtual TE-Links sharing the same server layer resource). PCEP supports concurrent path computation per RFC5440, expressing MELG constraint is out of scope of this document (defined in [MELG])

Core domain path computation and Inter-PCE path computation is out of scope for this document.

6. L1VPNs

[RFC4208] implies that multiple independent sets of clients, located in the same or different layer networks, could be connected to the same network domain, providing the connectivity between the clients within each set, while blocking the connectivity between the clients from different sets (i.e. allows for the L1VPNs application).

This document suggests:

- New sub-TLV – VPN IDs sub-TLV – to be added to the top level TE Link TLV. Exact format of the VPN IDs sub-TLV is described in [GMPLS UNI RTG]
- Configuring on the network end of each access TE link zero, one or more network unique VPN IDs and adding the configured information as VPN IDs sub-TLV to the TE link advertisement;
- Configuring zero, one or more network unique VPN IDs for each Virtual TE Link and adding the configured information as VPN IDs sub-TLV to the TE link advertisement;
- Making the network responsible for proper filtering of the TE Link advertisements, so that the information pertinent to VPN X is leaked only to the clients that are members of the said VPN X

This approach would achieve the following:

- Automatic VPN member auto-discovery;
- Providing to the clients VPN specific view of the network ;
- Partitioning network resources between VPNs;
- Ensuring successful path computations (and therefore connectivity) only between members of the same VPN

[RFC4208] implies that access TE Links could be named from a single or a separate (per-VPN) name space. This draft takes the former approach, that is, regardless of the associated VPNs, all access and Virtual TE Links MUST be named from the same (specifically, network) name space. Apart from simplicity, one reason for such choice is the following consideration: a GMPLS LSP established between a pair of clients is likely to be advertised as a TE Link into the client's layer TE domain. For example, a GMPLS LSP established between a pair of IP routers is likely to be advertised as a TE Link into IP/MPLS layer TE domain. This means that neither access nor Virtual TE Links belong to the "real" client layer network. Hence assigning addresses for access and Virtual TE links from the network name space would not cause address collisions/re-configurations in the client layer.

[TBD: L1VPN Figure/Example]

7. Use cases

7.1. Service optimization and restoration in Multi-Layer Networks

Multi-layer networks, as described in this document, are a reality today and they are operated by different groups following different operational procedures.

This requires an independent optimization of the client and server layer networks, and this could lead to the situation where the re-routing of a client layer LSP fails because some of the resources on the selected alternate path share fate with some of the resources on the LSP's failed path. This would happen due to lack of knowledge of the server layer network when the client layer path computation function selects the alternative path.

The high percentage of IP traffic in operator networks today makes it necessary that client and server layer share sufficient information to enable an optimized transport for IP/MPLS services and address existing inefficiencies. One important point from the carrier perspective is that the usage of server-layer SRLG information by the client layer path computation is essential to address these issues.

In a typical multi-layer network, in which the IP/MPLS network is the client network and the WDM/OTN network is the server network, it is the client layer network that is responsible for the protection of the IP/MPLS traffic using mechanisms such as FRR and/or LFA. Regardless of the mechanism that is used, SRLG information from the server layer network helps to optimize the client layer network with respect to reduced link utilization and reliable and efficient protection of the client traffic.

Today server layer network SRLGs are used mainly to calculate diverse alternative paths for the IP/MPLS client layer network. Therefore the following procedure MUST be periodically performed:

- Build traffic matrix for the server layer network (based on IP links)
- Solve traffic engineering problems in the server layer network
- Calculate new SRLGs for the client layer network
- Simulate failure scenarios

GMPLS UNI reduces the OPEX costs of doing these procedures manually by providing:

- the advertisement of server layer network SRLG information into client layer network via common routing protocol
- the client layer network path computation function uses this SRLG information in selecting maximally diverse paths.

7.2. IP/MPLS Offloading with UNI automation

A typical application in multi-layer (IP/MPLS over optical) networks is termed 'IP Offloading', in which the network responds to the increase in traffic of a particular service or across a network segment in the IP network by placing IP traffic into GMPLS LSPs in the server layer network in order to reduce the load on intermediate IP routers. The increase in traffic is typically caused by an elevated number of high traffic flows/services traversing an IP network segment, which requires core routers to forward large IP traffic volumes.

The decision process driving IP offloading is complex and is constrained by a set of rules that reduce the cost of running the multi-layer network while ensuring that it remains stable.

Automation of IP Offloading poses a number of challenges. It must establish GMPLS LSPs in the server layer (e.g. optical) network and automatically assign them identifiers, either numbered or unnumbered, in the client layer network. This information can be automatically exchanged using the procedures from [RFC 4203]. However, such procedures are not always implemented in commercial equipment. Consequently, this information may need to be configured manually as part of the initial set-up/installation of these LSPs.

Later, when the GMPLS LSP tunnel needs to be established, the hierarchical TE Link addresses MUST be included in the UNI path request.

7.3. Use of PCE and VNTM in Multilayer Network Operation

Two key elements have been proposed to help in the management and coordination of multi-layer networks: the Path Computation Element (PCE) and the Virtual Network Topology Manager (VNTM). PCE is responsible for the calculation of paths between endpoints,

particularly in complex scenarios involving, for example, WDM layer physical impairments. VNTM is in charge of maintaining the topology of the client layer network by instantiating GMPLS LSPs, in the server layer network. I.e., in can be used to provide TE links to the client layer network in real time.

Several cooperation modes between PCE, VNTM and the NMS have been proposed in [RFC 5623]. For instance, the operator can request a new MPLS path via the NMS, which consults a PCE with information of the multi-layer network. The PCE, in case that there are enough resources in the MPLS layer, returns a path made of real TE links. On the other hand, if there is a lack of resources at the MPLS layer, the response may contain a path with one or more Virtual TE-Links. In this case, the NMS can cooperate with the VNTM to suggest the set-up of a GMPLS LSP(s) in the server layer network. The VNTM, based on the local policies, can accept the suggestion and cause the set-up of the GMPLS LSPs in the server layer network.

In order for the computation to be effective, the PCE needs knowledge of the augmented topology (SRLGs, MELGs, TE metrics of the Virtual TE-Links), which can be provided via GMPLS-UNI.

8. Security Considerations

TBD

9. IANA Considerations

This document has no actions for IANA.

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[G.8080] Architecture for the automatically switched optical network (ASON)

[RFC4208] G. Swallow, J.Drake, H. Ishimatsu, and Y. Rekhter, “GMPLS UNI: RSVP-TE Support for the Overlay Model”, RFC 4208, October 2005.

[MELG] Mutually Exclusive Link Group, [draft-<tbd>-melg-00.txt]

[GEN CNSTR] G.Bernstein, Y.Lee, D.Li, W.Imajuku, "General Network Element Constraint Encoding for GMPLS Controlled Networks" [draft-ietf-ccamp-general-constraint-encode-07.txt]

[GMPLS UNI RTG] GMPLS UNI Routing Extensions [draft-<tbd>-gmpls-uni-routing-00.txt]

10.2. Informative References

[RFC4847] T. Takeda, "Framework and Requirements for Layer 1 VPNs", RFC 4847, April 2007.

11. Acknowledgments

TBD

Authors' Addresses

Vishnu Pavan Beeram
ADVA Optical Networking

Email: vbeeram@advaoptical.com

Igor Bryskin
ADVA Optical Networking

Email: ibryskin@advaoptical.com

Wes Doonan
ADVA Optical Networking

Email: wdoonan@advaoptical.com

John Drake
Juniper Networks

Email: jdrake@juniper.net

Gert Grammel
Juniper Networks

Email: ggrammel@juniper.net

Manuel Paul
Deutsche Telekom

Email: Manuel.Paul@telekom.de

Ruediger Kunze
Deutsche Telekom

Email: Ruediger.Kunze@telekom.de

Oscar González de Dios
Telefonica

Email: ogondio@tid.es

Cyril Margaria
Nokia Siemens Networks

Email: cyril.margaria@nsn.com

Friedrich Armbruster
Nokia Siemens Networks

Email: friedrich.armbruster@nsn.com