200Gbps of Non-oversubscribed MPLS Access Capacity on a Single 40Gbps Wavelength Ring

Introduction

This white paper by Optimum Communications Services, Inc. (OCS) presents how OCS' self-optimizing Adaptive-Mesh network architecture can be utilized to provide twenty 10Gbps access points worth of non-oversubscribed, revenue-generating network access capacity over a single 40Gbps wavelength ring.

The Network Design Case

Fig. 1 below presents the network design case for this white paper. It is assumed for this case that:

- the network is to interconnect five sites, with two mutually protecting routers at each site;
- the most economic available high-capacity MPLS router network interface, for which the routers are able to do line-rate packet processing, is either a 10Gbps Ethernet (10GbE) or OC-192c card;
- the routers at each site are able to do load balancing among their 4 10Gbps links;
- the network shall provide site-to-site connectivity protection with no single-point-of-failures;
- the network shall support 40Gbps (worth 4 OC-192c:s) of throughput to/from each site;
- the network shall provide 20Gbps (worth 2 OC-192c:s) of guaranteed throughput between any pair of the sites under any traffic conditions (within the limits of the router interface capacity per each site), even in case of any single equipment or network link failure;
- given the above performance requirements, the objectives of the network design are reliability and manageability, i.e., the goal is uniformity and operational simplicity, rather than marginal optimization.

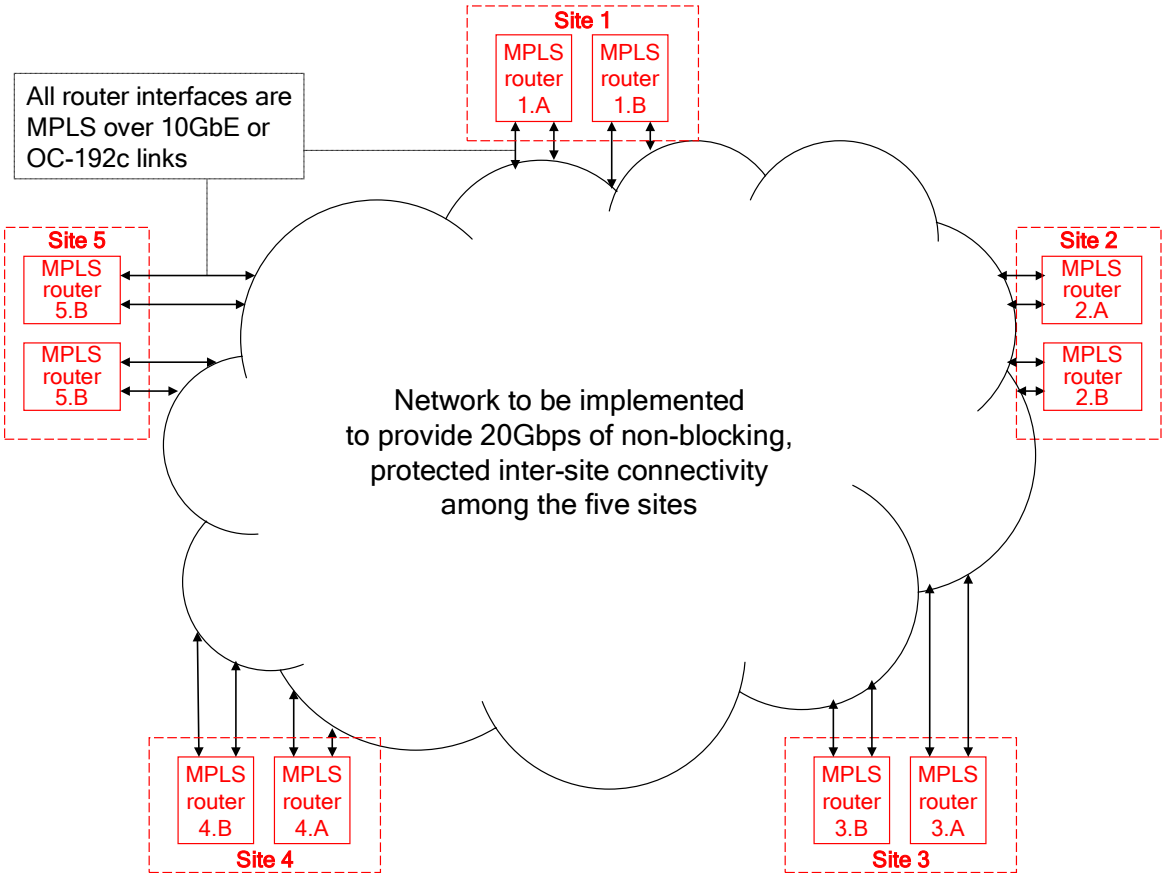

Figure 1. Network planning case diagram.

Regarding Fig. 1, it shall be understood that there are no restrictions to the geographic locations of the sites, and e.g. two of them can be located in the same physical location to provide doubled access capacity per such site. It is thus possible to provide asymmetric physical network designs even based on such a seemingly symmetrical case.

Packet-over-Lambda Architecture

Fig 2. below presents a network implementation using the packet-over-lambda architecture, based on OADMs and a doubled core MPLS switches.

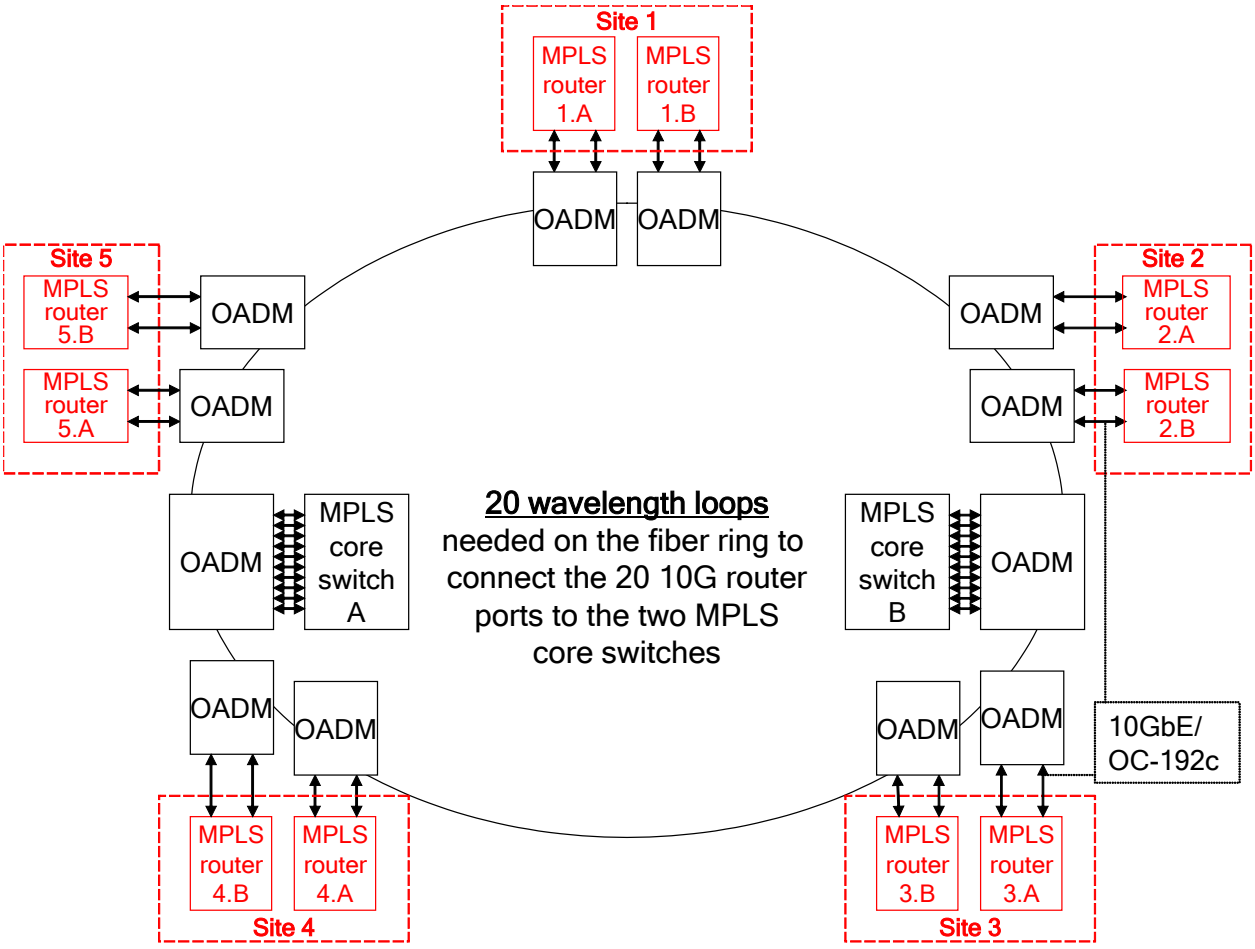

Figure 2. Conventional network architecture.

In the packet-over-lambda network of Fig. 2 the OADMs add-drop mux the 10Gbps ports of the MPLS routers to/from optical wavelengths (lambdas) on the fiber ring connecting the sites to the two MPLS core switches. Two of the four 10Gbps router ports per each of the five site are connected by the WDM network to the each of the two MPLS core switches, each thus requiring 2x5x10G=100Gbps of full-duplex packet switching capacity. Ten wavelengths on the fiber ring connecting the OADMs are needed for connecting the 2x5=10 router ports to each of the two MPLS core switches, and thus a total of 20 wavelength loops, are needed on the fiber ring.

Note that while it may be technically feasible, while staying with the basic packet-over-lambda architecture of Fig. 2, to implement the required inter-site network connectivity with fewer wavelengths, achieving such wavelength capacity requirement reduction would require complicating the network design, deviating from the goals of uniformity and operational simplicity, and thus defying the reliability and manageability objectives of the network design case. Therefore, such marginal optimizations at the cost of increasing complexity are not considered further in this white paper.

Adaptive-Mesh Architecture

Fig 3. below shows how the Adaptive-Mesh network architecture of OCS' Intelligent Transport Network™ (ITN) is able to provide the required 20x10Gbps of revenue-generating connectivity among the MPLS routers over a single 40G wavelength ring, and without any core MPLS switches.

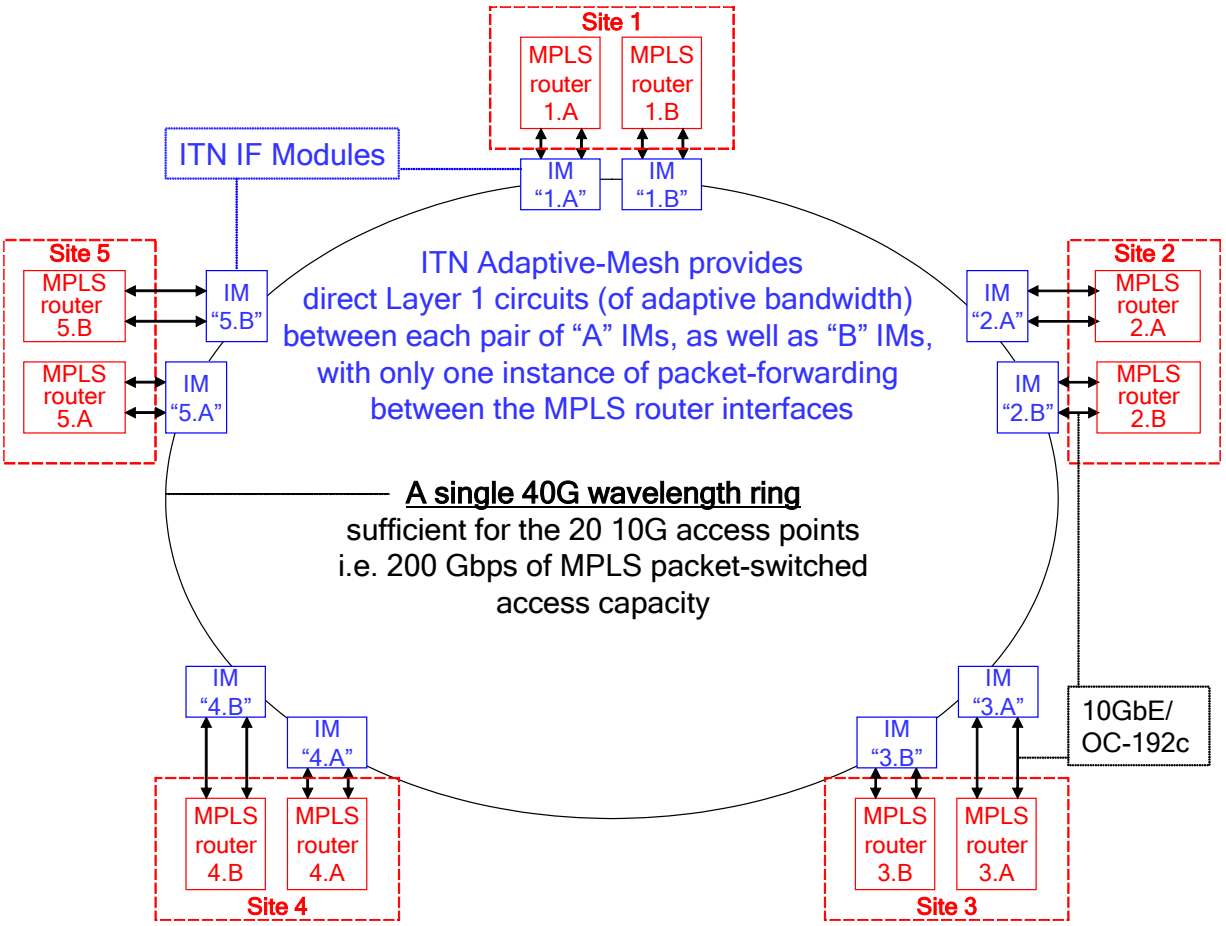

Figure 3. Adaptive-Mesh network diagram.

In Fig. 3 there is a direct Adaptive-Mesh between each of the "A" instances of the ITN Interface Modules (IMs), as well as between each of the "B" instances of IMs.

Each Adaptive-Mesh is non-blocking, since within Adaptive-Mesh, per each 10G egress link to the MPLS routers, there is a dedicated 10Gbps pool of network capacity for transporting data to that site from the other sites. Such 10Gbps pools, each comprising 192 STS-1 timeslots, thus form dynamically channelized multi-source, single-destination packet transport buses, providing adaptive-bandwidth STS-X (X=0..192) circuits between the IMs. Without blocking, these buses can be routed on an OC-768 ring among the IMs to each of the twenty 10G access points[1]. Within each packet transport bus in Adaptive-Mesh, the allocation of 192 STS-1 timeslots among adaptive-bandwidth STS-X circuits is continuously optimized, at rate of 72000 optimization cycles per second, based on data load variations between the MPLS routers.

As a result, Adaptive-Mesh delivers traffic between the MPLS routers it interconnects at a maximum rate achievable given the capacities of the routers interfaces (e.g. 10Gbps), while requiring the minimum amount of network bandwidth sufficient to prevent blocking. Adaptive-Mesh thus maximizes the network cost-efficiency, due to its patent-pending real-time self-optimization capability.

[1]For derivation of Adaptive-Mesh ring capacity requirements, please refer to Appendix A of this document.

Moreover, the network design of Fig. 3 is completely uniform, as it is made simply by repeating the same site-interface design of two IMs per site for each of the five sites, without requiring any other equipment. Furthermore, since ITN is able to forward MPLS packets directly based on their MPLS Labels, without requiring any packet-layer forwarding or routing tables, ITN is extremely straightforward in operation.

It is also seen that in Fig. 3 any of the MPLS router sites (e.g. site 1) could be equally well be another ITN Adaptive-Mesh network. With such extensible Adaptive-Mesh architecture, e.g. four regional inter-router networks per Fig. 3 with site 1 at each replaced by an interface to an inter-regional Adaptive-Mesh network can be interconnected to form a (national) network with 4x40G=160G of MPLS access capacity per each of the four regions, for a total (national) network access capacity of 4x160G=640Gbps.

Thus, the ITN Adaptive-Mesh architecture can be extended to arbitrary number of network access points, without a single packet layer forwarding or routing table between the MPLS routers interconnected by ITN, and while maintaining the uniformity and operational simplicity of the basic network of Fig. 3. A reduction in the order of 20X in the required fiber-optic wavelength capacity, along with proportional reduction in equipment capacity, is achieved as well, compared to equal-perform performance network alternatives that are not able to utilize adaptive-bandwidth Layer 1 as the Adaptive-Mesh does.

Accordingly, the characteristics of the dynamic bandwidth allocation optimization of ITN Adaptive-Mesh, resulting in maximization of revenue-generating data throughput, can be summarized as follows:

- fully automatic -- equipment other than the ITN nodes (IMs) do not need to be aware of the dynamic timeslot allocation within Adaptive-Mesh;
- fully transparent -- at MPLS layer, the MPLS routers see each others (instead of ITN) directly through Adaptive-Mesh networks;
- standards compatible and inter-operable -- the external network interfaces of ITN (e.g. MPLS over PPP over OC-192c/10GbE) are directly inter-operable with standard router network interfaces.

Summary

By comparing the network designs of Fig. 2 and Fig. 3 it is observed that the Adaptive-Mesh architecture of ITN is able to achieve with a single OC-768 ring, i.e., a single wavelength, the required inter-site connectivity that would require 20 wavelengths on the ring and proportionately more equipment capacity if implemented using the packet-over-lambda architectures. Accordingly, for network upgrades similar to the case in this white paper, the Adaptive-Mesh architecture of OCS' ITN can achieve in the order of 20 times higher cost-efficiency. Considering the additional savings resulting from the operational simplicity of ITN, the real cost-efficiency difference in favor of ITN would be even higher.

Finally, the philosophy of this white paper is to present generic, multi-use network architectural diagrams that can be applied as basis for multiple types of specific network design cases. Rather than examining detailed optimization scenarios, the main characteristics of two main architectural alternatives, one based on non-adaptive Layer 1 and another based on adaptive Layer 1, that meet the common network design objectives have been compared to find aspects of the architectural alternatives resulting in cost-efficiency differences of macroscopic scope (e.g. 10X or 20X) rather of marginal scope (e.g. 10% or 20%). It is considered that the cost-efficiency differences between architectures found in generic comparisons should be measurable in multiples rather than percentages, for the cost-efficiency difference results to be of general merit and applicable as a general rule for choice of preferable architecture in network planning cases within the general scope of this paper, i.e., packet-switched backbone network upgrades.

Obviously, there are different architectures that could be considered other than the two main architectures presented in the foregoing, and there is room for additional considerations regarding various traffic types (e.g. multicast). Please refer to the Appendices for further reading regarding these topics.

For more info about OCS and the ITN solution, please visit www.ocsipholding.com.

Appendix A: Derivation of Adaptive-Mesh Requirements for Network Bandwidth

Fig. A.1 below illustrates the routing of the packet transport buses, called Adaptive Concatenation Multiplexer Buses (AMBs) that consist of M (an integer) STS-1 timeslots each, between the Intelligent Transport Network™(ITM) Interface Modules (IMs).

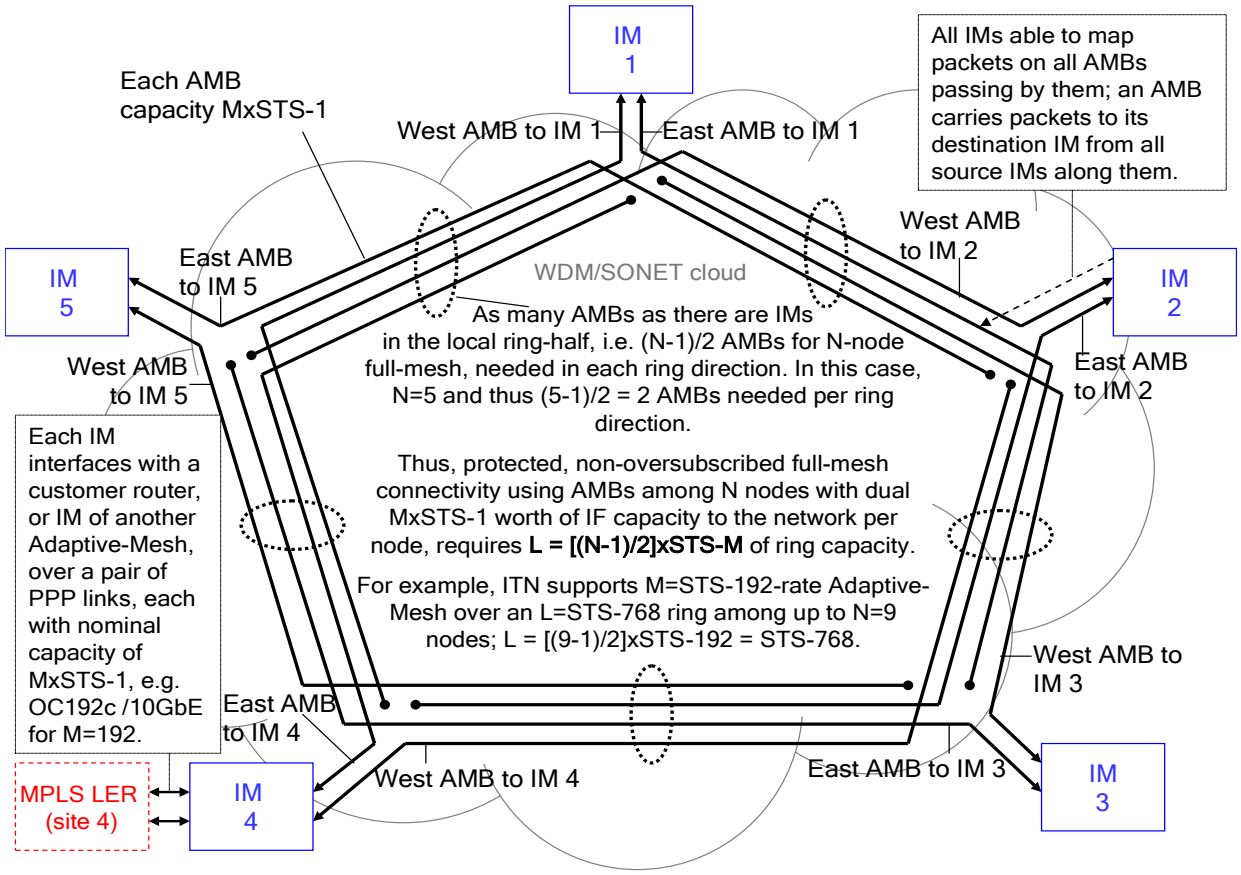

Figure A.1. Routing of the AMBs that form an Adaptive-Mesh. Even though for simplicity only two source nodes (IMs) per AMB are shown, the architecture supports any number of source IMs per AMB. Moreover, it is not required that the AMBs form a full-mesh, or that they are routed over a fiber ring.

Per Fig. A.1, the ring capacity requirement of an Adaptive-Mesh (A-M) between N=5 IMs at rates M=STS-192 is L=(5-1)/2xSTS-192=2xSTS-192. Thus, on an OC-768 ring, supporting 4 STS-192 rings, two such 5-IM A-Ms, each requiring 2 STS-192 rings, can thus be arranged, as is done in the network of Fig. 3.

Note also that per Fig. A.1, the A-M diagram of Fig 3. provides up to full 40Gbps of inter-site connectivity between any pair of the sites without blocking. Since the network of Fig. 3 is made of two A-Ms per Fig. A.1, one A-M between the “A” and another A-M between the “B” IM instances of Fig. 3, it is directly visible that 10Gbps of throughput is available to each of the MPLS routers per each of its two 10G access links, i.e. 20Gbps in total, from any one of the other MPLS routers on the local ring side of the access link. For instance, it can be seen that the full 2x10Gbps of traffic from e.g. site 4 routers can be connected to site 1 routers, using the 10Gbps STS-192 AMBs on the clock-wise ring direction on both of the two A-Ms.

However more study is appropriate for the case of more than 20Gbps of traffic load from site 4 to site 1, even up to full 40Gbps i.e. worth the full capacity of both sites. In that case, if there is no traffic load to site 1 from any other site, the network capacity to site 1 should be fully allocated to traffic from site 4. In such a case, the IMs 4.A and 4.B will forward packets they receive over the right-hand-side 10G access links from the routers 4.A and 4.B destined to site 1, via 3.x and 2.x IMs, which will re-forward the packets, based on their destination identifier, over the two counter-clock-wise direction STS-192 AMBs of both A-Ms through IMs 1.x to routers at site 1. Thus, ITN Adaptive-Mesh network of Fig. 3 is able to provide full 2x2x10Gbps = 40Gbps of traffic throughput between any of the sites interconnected.

Appendix B: Comparison to RPR, SONET ADM and Ethernet architectures

A summary of how the RPR, SDH/SONET ADM and Ethernet alternatives would compare with the packet-over-lambda and Adaptive-Mesh architectures of Fig. 2 and Fig. 3:

RPR:

- Requires 10 times more packet processing interface capacity per each node:
  - With RPR, each node along the route of a packet has to process at packet-level each packet entering and leaving the node, even if the packet is simply passing between the ring interfaces of the node.
  - With Adaptive-Mesh, packet-forwarding is only needed for the lower capacity access interfaces (10Gbps), but not the high capacity ring interfaces (40Gbps). Furthermore, with Adaptive-Mesh, no packet-switches are needed due to distributed forwarding and full mesh of (adaptive-bandwidth) Layer 1 circuits.
  - Thus, while with Adaptive-Mesh an IM node with 2x10G of access capacity can do with 20Gbps of packet forwarding capacity, an RPR node requires 20G(access)+40Gx2(both ring IFs) = 100Gbps of packet forwarding capacity, and 100Gbps of packet-switching capacity i.e. 100G(IF)+100G(switch)=200Gbps worth of packet processing capacity; 10 times more per each node than Adaptive-Mesh.
  - RPR packet processing is also considerably more complicated (both the hardware logic as well management software and service provisioning) than the look-up-table-free MPLS forwarding of Adaptive-Mesh.
- No bandwidth efficiency gain achievable over Adaptive-Mesh, as any less network bandwidth than the 40Gbps ring required by Adaptive-Mesh would result in the network becoming blocking, even if the traffic flows through it were fully optimized;
- Loss of predictability of network performance:
  - In an RPR implementation of the network of Fig. 1, a packet may have to processed at up to five intermediate RPR nodes between two MPLS routers.
  - Each instance of packet processing along the path of packet is a potential point of congestion.
  - Consequently, the delay, jitter and probability of packet loss are increased per each RPR node between routers to be interconnected.
  - Note that in Adaptive-Mesh, packets travel on Layer 1 circuits (of adaptive bandwidth) between router access interfaces of the IMs, with minimum delay jitter, and no chance of packet loss, ensuring that network bandwidth is used only for packets that will get through the their destination.
  - As RPR lacks a way to globally allocate network capacity based on the traffic loads across the whole network, its throughput performance is non-deterministic.

→ RPR does not meet the manageability and reliability objectives of the network design case, and while increasing complexity and equipment cost, provides no benefits over Adaptive-Mesh.

SONET ADMs to add-drop-mux OC-192s in to OC-768 ring:

- Increases the upfront equipment cost of the network.
- Requires the ADM OC-N interfaces and cross-connects to be configured.
- Requires an additional type of network elements to be managed.
- Adds more potential point of failures to the network, via increasing the number of nodes and network links between the nodes.

→ Compromises the objectives of manageability and reliability.

Ethernet:

- 10GbE cannot be add-drop-muxed into OC-768 ring, forcing to use 4 wavelengths on the ring for 4 10GbE ports of an OADM.
- Assuming Ethernet-switching is applied for full interface capacity at each site, including network ring interfaces as well as site access (router) interfaces, the network *suffers from all the complexity disadvantages of RPR*. Also, like is the case with RPR or any other approach without Layer 1 circuit based transport capability across intermediate nodes between the routers, Ethernet is inherently non-deterministic in performance compared to networks with L1 circuit switching and transport.

→ Ethernet does not meet the manageability and reliability objectives of the network design case, and while increasing the complexity, provides no benefits over Adaptive-Mesh.

Appendix C: Multi-cast scenarios

Regarding multi- and broad cast traffic types:

- Note that regardless of how the network (of Fig. 1) is implemented, it is not possible for the routers to receive more traffic than is the fixed capacity of their network access links (e.g. 40Gbps per site). Thus, it is not possible for e.g. two access sites to broadcast at 40Gbps traffic to all other sites, as the combined volume of the two broadcast streams of 2x40Gbps=80Gbps would exceed the 40Gbsp receive capacity limit of each access site. Since the broadcast throughput constraint is dictated by the 40Gbps receive access capacity limit of each access site, the inter-site network needs to support only 40Gbps of combined traffic throughput (broad-, multi- or uni-cast) to each of the access sites from the other sites -- which requirement is met by the Adaptive-Mesh.
- While it might initially appear that it would be more bandwidth-efficient to send a multicast packet over just a single L1 path from which all the nodes in the network are able to drop their copy of the multi-cast packet, compared to sending a multi-cast packet over multiple AMBs to the multiple destinations of the multi-cast packet, such bandwidth efficiency gains are only artificial. This is due to that in order for the network to support any mix/match of (uni-, multi- and broad-cast) packets in a non-blocking manner, each access site (of e.g. 40Gbps) will require 40Gbps of dedicated network bandwidth for receiving traffic from the other access sites. It is observed that in fact the highest network capacity requirement occurs in case of uni-cast (instead multi-cast) packets, when each access site is to receive their unique set of packets from the other sites, in which case a network that does not provide a dedicated 40Gbps worth of capacity for transport packets to each of the access sites is a blocking network with reduced revenue-generating traffic delivery capability. Since a dedicated 40Gbps of network transport capacity to each access site is thus anyway required, any bandwidth savings achievable for multi-casting packets via the use of shared transport path (instead of different AMBs) are only artificial, at least as long as the network needs to support also uni-cast packets.
- Note furthermore that, from a multi-cast stream reachable via a common intermediate node, in Adaptive-Mesh only one instance of packets need to be sent to such intermediate node, which will then do the multi-cast forwarding as necessary at that stage in the network.
- For remote access networks, ITN provides a Packet Demux Bus architecture for traffic flowing toward the remote access sites, wherein several e.g. GbE access sites share the same e.g. 20 STS-1s worth of network capacity to/from the CO/POP.