LAYERED ARCHITECTURE FOR SCALABILITY IN CORE MESH OPTICAL NETWORKS

Jean-François Labourdette, Chris Olszewski, Sid Chaudhuri, and Eric Bouillet

Tellium, 2 Crescent Place, P.O. Box 901, Oceanport, NJ 07757
Corresponding author: jlabourdette@tellium.com

Abstract—In this paper, we examine the scalability of core optical mesh networks. As traffic demand grows and evolves, core network nodes need to switch ever-larger amounts of traffic at different rates (DS1, DS3, OC-3, OC-48, OC-192). In a flat or one-tier network architecture, each network node contains one or multiple identical switches that can switch at the lowest rate in the hierarchy, usually STS-1 level. As traffic grows, the switching capacity of a node can be scaled by interconnecting multiple such switches. However, scaling the network in this manner incurs a severe penalty in terms of interconnect capacity needed to interconnect switches in the same office. This penalty is dependent on the traffic forecast accuracy, increasing with the forecast uncertainty. We use theoretical and experimental results for scaling the switching capacity of a node by interconnecting multiple switches, and compare a flat or one-tier architecture with a layered (hierarchical) architecture where the network is scaled by organizing it in layers. There, layers are optimized to switch and groom at different rates. We show that there is a crossover point beyond which the layered architecture becomes more cost effective as the total traffic grows and as the traffic mix evolves toward higher rates.

A. INTRODUCTION

The phenomenal growth of the Internet has driven an exponential growth in data traffic over the past decade [1]. To accommodate this explosive traffic growth, core transport networks are being upgraded from legacy SONET/SDH rings to fast-restorable, capacity-efficient optical mesh networks [2-6]. The core optical network consists of backbone nodes interconnected by point-to-point WDM fiber links in a partial mesh interconnection pattern. Each WDM fiber link carries multiple wavelength channels (e.g. 160 OC-192 channels). Transmission rates of wavelength channels on long-haul WDM systems are currently evolving from OC-48 to OC-192, and are expected to evolve to OC-768 in the future. Multiple conduits (each containing multiple fibers) are usually incident at the backbone nodes from adjacent nodes. Figure 1 illustrates a core optical network. Diverse edge equipment, such as IP routers, FR/ATM switches, and Multi Service Provisioning Platforms (MSPP), are connected to an optical switch.

The deployment of an intelligent core optical network requires that the backbone nodes terminate wavelengths onto a scalable, strictly non-blocking core switch. An OEO-based core optical switch converts optical signals into the electrical domain at the ingress port, switches the electrical signals through an electrical switch matrix, and then converts signals into the optical domain at the egress port. Intelligence allows automatic topology discovery, rapid provisioning and fast restoration [7,8]. Furthermore, a scalable strictly non-blocking switch ensures that the expensive long haul wavelengths are not wasted due to blocking when traffic grows and traffic patterns change. Indeed, non-blocking switching in an office is required for flexible usage of WDM capacity. If there is uncertainty in traffic demand, penalty is paid either in switching interconnect or in WDM capacity. Intelligent optical switches enable re-configurable optical networking by effectively managing the network bandwidth, rapidly provisioning end-to-end circuits called lightpaths between the client devices, and supporting fast and capacity-efficient restoration of these lightpaths [4,6]. The grooming capability of a switch enables it to groom lower speed signals onto higher speed signals, and to bridge the mismatch between the wavelength transmission rates and the service rates. The granularity of the switch fabric thus drives the grooming granularity, i.e., the lowest rate at which the equipment can switch, multiplex, and demultiplex signals. The grooming granularity of the optical switch in turn determines the granularity at which network bandwidth is managed. The optimal grooming granularity of a network depends on the mix of traffic rates and on the traffic volume supported by the network.

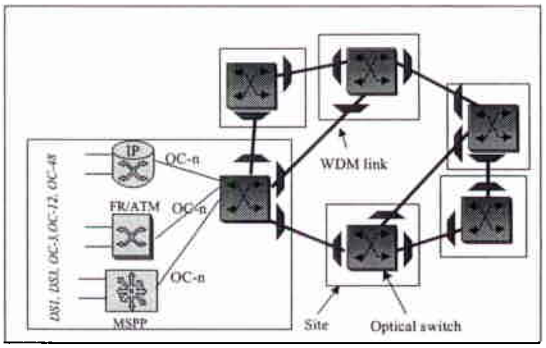

Figure 1: Core Optical Mesh Network

Traffic carried in the core optical network consists of data traffic that is packet-groomed by IP routers and FR/ATM-switches into OC-N (N=12, 48, 192) trunks. TDM switches aggregate traffic at lower rates such as STS-1 (52 Mbps) and VT1.5 (1.7 Mbps) and feed into the core at OC-N rates as well. Figure 2 illustrates a traffic bandwidth mix for a carrier backbone network. In this instance OC-48 (2.5 Gbps) and OC-192 (10 Gbps) services and trunks are a dominant (57%) and growing component of the core traffic mix.

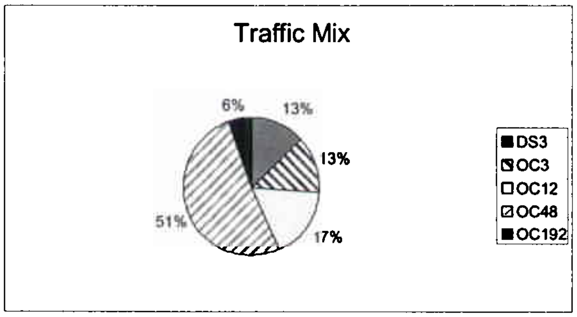

Figure 2: Traffic bandwidth mix in the core optical network, shifting toward OC-48 and above rates

As traffic demand grows and evolves, core network nodes need to switch ever-larger amounts of traffic at different rates (DS1, DS3, OC-3, OC-12, OC-48, OC-192). Furthermore, the traffic mix shifts towards higher rates over time. This evolution, if it outpaces the evolution of the optical switch size, requires that multiple switches be deployed at certain core network nodes. As described previously [9], interconnecting multiple switches within an office wastes ports, and this inefficiency has implications regarding the overall network architecture. For example, our analysis showed that if traffic forecast is only 70% accurate when deploying and connecting transmission and switching equipment in an office, as much as 30% of the switch ports may end up being used for connecting the switches together to allow non-blocking switching in the office for flexible usage of WDM capacity. While many other aspects impact the core mesh network architecture, we focus here on the effect of switch size.

In this paper we will consider two architectures for the core backbone network: (1) a flat architecture, (2) a layered architecture. In the flat architecture illustrated in Figure 3, the core optical switch can switch at the STS-1 (52 Mbps) granularity. The STS-1 switch handles all STS-N services (whose rates are multiples of the STS-1 rate) from client equipment. The STS-1 switch also terminates wavelengths (OC-48/OC-192) from the DWDM equipment. The flat architecture allows network bandwidth to be managed in increments of STS-1. In the layered architecture illustrated in Figure 4, the core optical switch operates at STS-48 (2.5 Gbps) granularity, and connected to the core optical switch are edge switches that can switch at the STS-1 granularity. The STS-48 switch directly handles OC-48 and OC-192 circuits. Edge STS-1 switches set-up, groom and switch at rates below STS-48. We term this the layered architecture because STS-48 switches perform "core-grooming" at STS-48 rate, and the STS-1 switches perform "edge-grooming" at STS-1 rate.

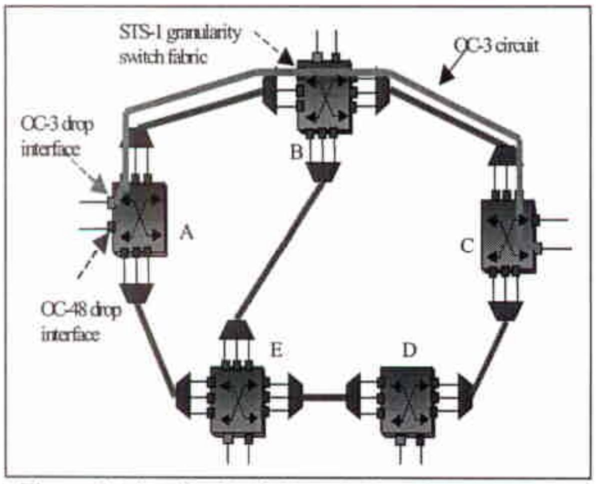

Figure 3: Flat (single tier) optical mesh network

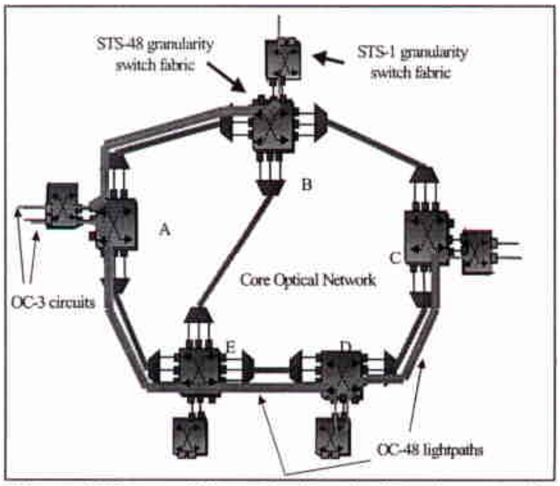
STS-48 granularity switch fabric
STS-1 granularity switch fabric
B
A
C
Core Optical Network
OC-3 circuits
E
D
OC-48 lightpaths

Figure 4: Layered (two-tier) optical mesh network with core grooming at STS-48 and edge grooming at STS-1

This paper makes the following contributions. Based on the theory of scaling the switching capacity of a node by interconnecting multiple switches [9], we analyze the flat and the layered network architectures, and identify conditions under which the layered network architecture is more cost-effective than the flat network architecture. This eventually happens when the traffic scales, and as the traffic mix evolves towards higher rates. This evolution, if it outpaces the evolution of STS-1 switch size, requires that multiple switches be deployed at certain core network nodes. In addition, the flat network architecture faces challenges to achieve fast and capacity efficient restoration of OC-N circuits. Shared mesh restoration of all individual OC-N circuits is difficult due to the complexity of handling so many circuits. For example, if a fiber carrying 160 OC-192 channels breaks, potentially, 160*192=30720 STS-1 circuits could be affected by the failure. This makes it very difficult to achieve shared mesh restoration times of the order of 100 msec while such times can be achieved with STS-48 switches operating at higher granularity [12,13]. See [12-16] for more information on restoration times in mesh networks. Fast restoration in the flat architecture is thus likely to require dedicated mesh (1+1) protection, thereby incurring the capacity penalty of dedicated mesh (1+1) protection compared to shared mesh restoration in the STS-48 layer of a two-tier architecture (two orders of magnitude less circuits need to be restored). In fact, because the number of switching involved, even dedicated mesh (1+1) protection may not guarantee fast restoration at STS-1 level. Indeed, large-scale networks have historically been organized hierarchically with multiple layers. It apparently seems ideal to have a single switch that is scalable, manageable, low-cost, and that can switch all rates and protocol formats. But practical considerations such as hardware and software scalability, manageability and reliability have always led to layered architectures, with each layer optimized independently. In the layered architecture, scalability and manageability are achieved by multiplexing traffic flows into larger streams as they traverse from the edge to the core, and demultiplexing them as they traverse from the core to the edge. Effectively, traffic flows are groomed and switched at a coarser granularity in the network core, and at a finer granularity at the edge.

The outline of the paper is as follows. Section B describes the architectural details of the flat and the layered networks along with procedures for the provisioning and restoration of lightpaths. Section C summarizes the theory of scaling the node switching capacity by interconnecting multiple switches together. Section D introduces metrics for comparing the two architectures, and presents an analysis for computing their relative costs. Section E describes the results of the model for a set of parameter values derived from realistic topology and traffic demand projections. Section F concludes this paper.

B. NODE & NETWORK ARCHITECTURE

Let us consider in more detail the two architectures for the core backbone network: (1) the flat architecture, and (2) the layered, or two-tier, architecture.

In the flat architecture (as shown in Figure 3), the core optical switch operates at STS-1 granularity. The STS-1 switch is connected with OC-N optical ports[1] to client equipment (such as IP routers, ATM/FR switches, Multi-Service Provisioning Platforms (MSPPs)) as shown in Figure 5. The STS-1 switch also terminates wavelengths (OC-48/OC-192) from the WDM systems connecting offices together. The flat architecture allows network bandwidth to be managed in STS-1 increments. As the traffic grows beyond the capacity of a single STS-1 switch, multiple STS-1 switches are interconnected to yield a larger STS-1 switching complex. This interconnection requires extra ports on the STS-1 switches to connect them together, and this is the key factor in our analysis.

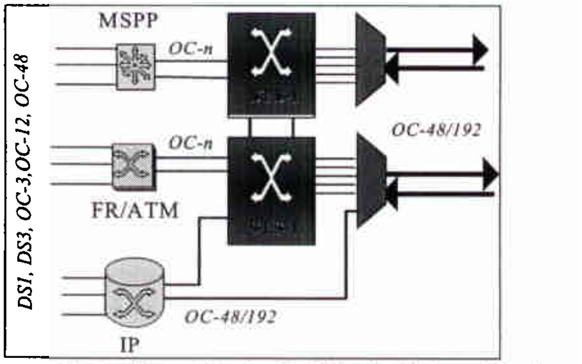

Figure 5: Node configuration in flat (single tier) architecture

The flat network architecture also faces challenges to achieve fast and capacity efficient restoration of OC-N circuits [12-16]. Shared mesh restoration of all individual OC-N circuits is difficult due to the complexity of handling and switching as many as tens of thousands of circuits following a fiber cut. This makes it very difficult to achieve restoration time of 100 msec. Fast restoration in the flat architecture is thus likely to require dedicated mesh (1+1) protection, thereby incurring the capacity penalty of dedicated mesh (1+1) protection compared to shared mesh restoration.

In the layered architecture (as shown in Figure 4), the core optical switch operates at STS-48 granularity. Connected to the core optical switch are switches that groom at STS-1 granularity as shown in Figure 6. The STS-48 switch directly handles OC-48 and OC-192 circuits. To groom traffic at a lower rate, edge STS-1 switches handle OC-N (N < 48) circuits. We term this a layered, or two-tier, architecture because there are STS-48 switches performing "core-grooming" at STS-48 rates, and STS-1 switches performing "edge-grooming" at STS-1 rates. In this layered architecture, wavelengths are managed in increments of OC-48.

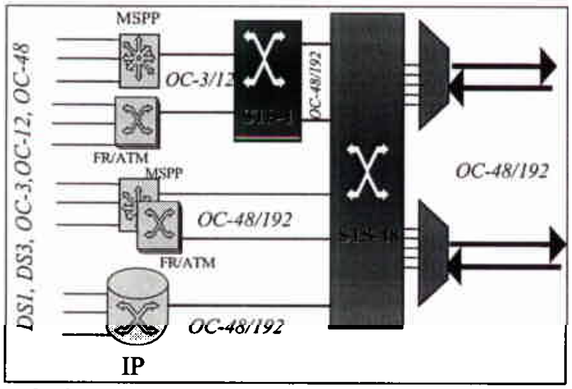

Figure 6: Node configuration in layered (two-tier) architecture

The STS-48 switch also terminates wavelengths from WDM systems. Connected to the STS-48 switch are one or more STS-1 switches. The STS-1 switch handles services below STS-48 rates (e.g., DS3, OC-3, OC-12), and aggregates them into OC-48 pipes. OC-48 or OC-192 circuits between backbone nodes are setup as lightpaths by

[1] Some OC-48 or OC-192 trunks may be directly connected to the WDM systems since sub-rate grooming is not required. This would be done at the expense of provisioning flexibility.

finding a route in the core network, and configuring the STS-48 switches along the route. Services below OC-48 rates (called subrate services) between backbone nodes are setup as follows: A set of OC-48 or OC-192 lightpaths between the core switches serve as trunks (or an *overlay* topology) for purposes of routing subrate services. For example, the overlay topology may be identical to the physical topology, with a direct lightpath between all neighbors. If there is a direct lightpath between a pair of backbone nodes, and there is enough capacity on that lightpath, a subrate service between the node-pair can use that lightpath. If there is no direct lightpath with enough capacity, then the subrate service has to traverse multiple lightpath hops, and at each intermediate node gets re-groomed at the STS-1 switch and switched onto the lightpath towards the next hop. Intermediate grooming is a natural aspect of routing on an overlay topology [10,11]. Because of the possibly less efficient packing of sub-OC-48 connections into OC-48 trunks, the resulting OC-48 wavelength utilization would in general be lower than in the flat network architecture. However, the effect is limited as follows. The amount of intermediately groomed traffic between a node-pair is bounded and less than a fraction of an OC-48, because, as soon as such traffic between two nodes exceeds a threshold, a more economical solution is to set up a direct lightpath between those nodes. The layered network has a fast, efficient and scalable restoration architecture. OC-48 and OC-192 lightpaths between STS-48 switches are restored using shared mesh restoration, achieving in the order of 100 msec restoration time, and allowing maximum sharing efficiency. Sub-rate circuits that ride on lightpaths are automatically restored thanks to the restoration of the lightpath. Connections between the STS-1 switch and STS-48 switch may be protected by a 1:N protection scheme.

C. Node Scaling

As traffic through a node (both add-drop and passthrough) grows, the switching capacity of that node needs to scale as well. Otherwise, either network capacity will be used inefficiently because circuits are routed on longer paths, or the network will block circuits. Typically, the switching capacity at a node is scaled in the following fashion. First, the switch is scaled in-service up-to the maximum allowable limit that is imposed by the architectural design of the switch fabric. Then, multiple such switches are interconnected with inter-machine ties to yield a larger switching capacity. The focus of this section is an analysis of the second case above, i.e., the scaling of the switching capacity of a node by interconnecting multiple smaller switches. For such a switching system, we derive the interconnect capacity needed to maintain a certain blocking probability under arbitrary traffic conditions.

Each node has links to neighboring nodes in the network. Each link contains multiple fibers, and each fiber carries multiple WDM channels. Each WDM channel terminates at a port on a switch at the node. Each node also has client devices that terminate traffic. The client devices are connected to a switch at the node. Each switch is strictly non-blocking of size M x M, and has M ports (each switch port has 1 input channel and 1 output channel). As traffic grows, switches are incrementally added as illustrated in Figure 7, and inter-connected with existing switches with inter-machine ties. There are N switches interconnected using a full-mesh interconnection topology. Each interconnection tie between switches has C bidirectional channels, using up C ports at each of the switches. The useful ports in the system are those ports that are not used for interconnections between switches. A traffic request is a bidirectional connection request from an ingress port to an egress port. A connection request is setup as follows. If the ingress and egress ports belong to the same switch then the connection is established within the switch. If the ports belong to different switches, then, the connection is established with a path between the ports, and cross-connections are made at each switch on the path.

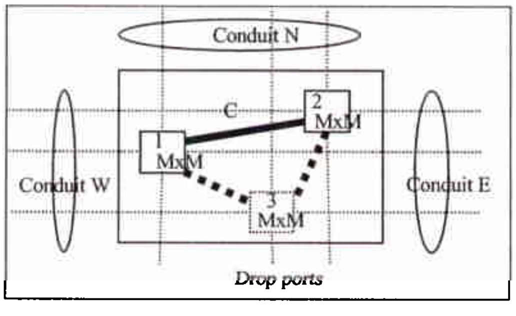

Figure 7: Model of a node. Multiple switches are interconnected in a full-mesh interconnection pattern. Switch 3 is incrementally added by interconnecting to switches 1 and 2, and by connecting its ports to the DWDM systems on the line side, and client devices on the drop side

The offered traffic, L, is the set of bidirectional connection requests, which can be any subset of the permutations of the useful ports of the system. The system may block at most a fraction B of the offered traffic. The carried traffic, S, is the set of connection requests that are set up and not blocked, $S = L(1\text{-}B)$. The internal traffic T (driving the amount of interconnection capacity used) is the total number of bidirectional interconnection channels (on inter-tie links) used to carry S. The number of useful ports that connect to the DWDM system and client equipment is P. A fraction Q of the traffic that is not blocked has ingress and egress ports on the same switch. The amount of interconnection capacity X used for internal traffic is the total number of interconnect ports used to carry $T{=}S(1\text{-}Q)$ connections. Given such a model of the switching system, we are interested in obtaining its switching efficiency, i.e., the relationship between the useful ports, and the interconnection ports. The set of parameters used to model a switching system with complete and uniform mesh connectivity between the switches is summarized in Table 1.

M	Number of ports per switch
N	Number of switches at site
B	Blocking probability of site; given any set of connection requests, at most a fraction B of the connection requests is blocked
L	Offered traffic, set of bidirectional connection requests
S	Carried traffic, $S{=}L(1\text{-}B)$
P	Useful ports: number of ports connected to DWDM systems and client equipment.
Q	Fraction of traffic for which ingress and egress ports are on the same switch[2]
T	Multi-hop carried traffic: traffic that is carried in one or more hops over the interconnect capacity between multiple switches $T = S(1\text{-}Q) = L(1\text{-}B)(1\text{-}Q)$
X	Interconnection capacity (ports) for full mesh connectivity between switches; $P + X \leq NM$.

Table 1: Parameter set used to model a switching system

We can express the ratio of interconnection ports to the total ports X/MN as [9]:

$$\frac{X}{MN} \geq \frac{2(1-B)(1-Q)\left(\frac{N-1}{N}\right)}{1+2(1-B)(1-Q)\left(\frac{N-1}{N}\right)}$$

Figure 8 illustrates the interconnection ports required (as a fraction of the total ports at the site) as a function of Q, the traffic prediction accuracy for different values of the number N of switches per site. Here we assume that traffic/connections are never blocked.

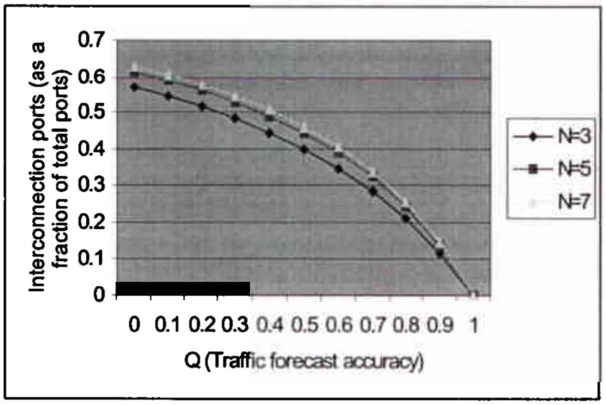

Figure 8: Interconnection ports (as a fraction of the total ports) as a function of Q

[2] Also a measure of forecast accuracy.

In the extreme case when Q=0, i.e., for completely arbitrary traffic patterns, and no blocking, the above result indicates that approximately 2/3 of the ports at a switch must be dedicated for interconnections with other switches. When about Q=0.7, i.e., 70% of traffic does not need interconnection hops, about 30% of the ports need to be interconnection ports. The number of useful ports is at most (about) a third of the total ports at the site. We define k to represent the number of interconnect ports needed as a ratio of the number of useful ports.

D. Network Architecture Analysis

In this section we analyze the flat and layered network architectures and derive simple formulas for the total network cost. For the sake of simplicity we assume that the traffic mix consists of two service rates: OC-48 and OC-3. We also assume that all wavelengths are at OC-48 rate.

Topology and Routing	N	Number of nodes
	d	Average node degree
	H	Average hop distance on fiber topology; approximated as $\log N / \log d$
	P	Bandwidth overhead resulting from packing of OC-3 into OC-48 connection (a packing utilization of 70% results in a 30% overhead, which is representative of real networks)
		Average hop distance for OC-3 in two-tier architecture due to re-grooming at intermediate nodes[3] (if no intermediate re-grooming takes place, β=1)
Traffic	T	Total traffic in Gbps
	X	Total OC-3 demand in Gbps, X=(1-α)T, $0 \leq \alpha \leq 1$
	Y	Total OC-48 demand in Gbps, Y=αT
Switch size	K_1	Maximum size of STS-1 switch
	k	Ratio of interconnect ports to useful ports when multiple STS-1 switches are interconnected (if number of OC-48 equivalent ports per site exceeds K_1)
	K_2	Maximum size of STS-48 switch
	r	Ratio of protection ports to useful ports to protect the intra-office interconnection between STS-1 and STS-48 switches. r is equal to 1 if 1+1 protection is used, or 1/N is 1:N protection is used.
Bandwidth	B_{1T}	Network bandwidth required to route demand in one-tier network
	B_{2T}	Network bandwidth required to route demand in two-tier network
Restoration	R_s	Ratio of protection capacity to working capacity for shared mesh restoration
	R_d	Ratio of protection capacity to working capacity for dedicated mesh (1+1) restoration

Table 2: network and traffic parameters

In general, the restoration ratios defined in Table 2 verify the following inequalities $R_s \leq 1 \leq R_d$. Practical values for R_s and R_d can be found in [4,6,17,18].

We also define a set of loaded cost parameters for interface cards and network bandwidth, as shown in Table 3. The loaded interface cost captures the start-up equipment cost, the switch fabric cost, and other components of the switch in addition to the interfaces. Thus the per Gbps costs reflect the costs of a fully loaded switch.

[3] Note that P decreases as β increases since more intermediate re-grooming increases the packing.

Interface	C_1	Loaded cost per Gbps of OC-3 interface on STS-1 switch
	C_2	Loaded cost per Gbps of OC-48 interface on STS-1 switch
	C_2	Loaded cost per Gbps of OC-48 interface on STS-48 switch
Bandwidth	C_λ	Loaded cost per Gbps per hop (300 miles) of OC-48 wavelength (includes WDM transponder and OA costs)

Table 3: loaded costs

In a flat network, the total network cost is comprised of (a) the interface card costs: the cost of OC-3 drop interfaces, $2XC_1$, the cost of the OC-48 drop interfaces, $2YC_2$, the cost of the network-side interfaces, $2B_{1T}C_2$, and if multiple STS-1 switches need to be interconnected at a site the cost of the interconnect ports $2k(X+Y+B_{1T})C_2$, and (b) the cost of network bandwidth, $B_{1T}C_\lambda$. The total network cost is thus:

$$C_{1T} = 2XC_1 + 2YC_2 + 2B_{1T}C_2 + 2k(X + Y + B_{1T})C_2 + B_{1T}C_\lambda$$

If the traffic was unprotected, the total network bandwidth would be $B_{1T\,unprotect.} = H\,(X+Y)$. We assume that the traffic is protected with shared-mesh restoration, and that shared mesh restoration is performed on OC-48 native traffic and on OC-48 resulting from packing the native OC-3 traffic[4]. The network bandwidth required is thus:

$$B_{1T\,shared\,mesh} = H\,(1+Rs)[Y+(1+P)X]$$

On the other hand, if we assume that, in order to achieve fast restoration (of the order of 100 msec), the single-tier network implements dedicated mesh restoration[5] for all circuits[6], we have:

$$B_{1T\,ded.\,mesh} = H\,(1+Rd)[Y+X]$$

We can then plug the appropriate formula for B_{1T} into the total network cost formula.

In a layered network, the total cost of the network is comprised of (a) the interface card costs: cost of OC-3 drop side interfaces (on STS-1 switch), $2XC_1$, cost of OC-48 drop side interfaces (on STS-48 switch), $2Y\underline{C}_2$, cost of network-side interfaces on the STS-48 switch, $2B_{2T}\underline{C}_2$, and cost of OC-48 interfaces for interconnecting STS-1 switch and STS-48 switch, $2(1+P)X\beta(1+r)(C_2+\underline{C}_2)$[7], and (b) the cost of network bandwidth, $B_{2T}C_\lambda$. The total network cost is thus:

$$C_{2T} = 2XC_1 + 2Y\underline{C}_2 + 2B_{2T}\underline{C}_2 + 2(1+P)X\beta)1+r)(C_2+\underline{C}_2) + B_{2T}C_\lambda$$

If the traffic was unprotected, the total network bandwidth would be $B_{2T\,unprotect.} = H\,(X+Y)$. We assume that the traffic is protected with shared-mesh restoration, and that restoration is performed on OC-48 native traffic and on OC-48 resulting from packing the native OC-3 traffic. The network bandwidth required is thus:

$$B_{2T\,shared\,mesh} = H(1+Rs)[Y+(1+P)X]$$

We can then plug the appropriate formula for B_{2T} into the total network cost formula.

E. Results and Interpretation

We assume values for the different parameters that are derived from realistic topologies, traffic patterns, and cost estimates. We assume a uniform traffic pattern and we vary the traffic mix from 0% OC-48 (100% OC-3) to 100% OC-48 (0% OC-3). The maximum switch sizes are $K_2 > K_1$, and the total traffic T is chosen such that the average traffic[8] per node is larger than K_1. In such a case, multiple STS-1 switches need to be interconnected together in the flat network architecture, and we assume that 30% additional ports are needed for interconnecting the STS-1 switches ($k = 0.3$). We further assume that the switch sizes are not exceeded in the two-tier network architecture. Finally, the intra-office interconnection between the STS-1 and STS-48 switches is 1:8 protected ($r = 1/8$). The topology of the network is chosen so that $N = 50$ and $d = 2.5$, resulting in $H = 4.2$. We also assume that $\beta = 1.5$ and

[4] In a similar way to what one would do in a two-tier architecture.

[5] End-to-end, or path-based, 1+1 protection.

[6] That is at sub-rates.

[7] Include 1+1 or 1:N protection of links between the STS-1 switch and the STS-48.

[8] Including transit traffic.

a packing overhead of P = 0.3. The restoration parameters, ratios of protection to working capacity, are taken to be typical $R_s = 0.7$ and $R_d = 1.3$ [4, 6, 17, 18].

Cost	Interface	C_2	1
		C_2	1
		C_1	2.5
	Bandwidth	C_λ	5

Table 4: Parameter values assumed for the analysis, costs are normalized

In the first part of the analysis, we assume that both network scenarios implement shared mesh restoration of the end-to-end OC-48 circuits[9]. The cost difference between the two-tier and the one-tier architecture, $\Delta C = B_{2T\ shared\ mesh} - B_{1T\ shared\ mesh}$, is obtained from the respective cost formulas for both networks.

$$\Delta C \ = \ 2(1+P)X\beta)1+r)(C_2+\underline{C}_2)-2k(X+Y+B_{1T})C_2$$

Let's first consider the extreme case where all the traffic is OC-48 (X=0). The cost difference ΔC is negative if the total traffic is such that the maximum STS-1 switch size K_1 is exceeded. If that happens, multiple STS-1 switches need to be deployed and interconnected together, with ports wasted for intra-office interconnect. In the two-tier architecture, no lower-tier STS-1 switches need to be deployed, and one STS-48 switch per site is sufficient to handle all the traffic.

$$\Delta C \ = \ -2k(Y+B_{1T})C_2$$

Let's now consider the other extreme case where all the traffic is OC-3 (Y=0).

$$\Delta C \ = \ 2(1+P)X\beta)1+r)(C_2+\underline{C}_2)-2k(X+B_{1T})C_2$$

From this equation, the cost difference ΔC is positive when all the traffic is OC-3, that is the two-tier architecture is more expensive than the one-tier architecture, if the interconnect penalty k satisfies the condition:

$$k \le \frac{2(1+P)\beta)1+r)}{1+H(1+Rs)(1+P)} \approx .43$$

For most set of parameters, the cost difference will be positive, unless the interconnect penalty becomes extremely severe. As a result of the relative cost when all the traffic is OC-48 (two-tier is cheaper) and when all the traffic is OC-3 (one tier is cheaper), there will be a traffic mix for which the two cost curves intersect, if the total traffic per site exceeds the STS-1 switch size.

For the parameter values chosen, curves (b) and (c) in Figure 9 illustrate the cost comparison of the two architectures when all traffic is shared-mesh restored[10]. When the proportion of OC-48 in the traffic mix exceeds 50%, the layered architecture (curve (c)) becomes cheaper than the flat architecture (curve (b)). As the proportion of OC-48 in the traffic mix increases, the cost of the layered network declines sharply as (1) the OC-3 into OC-48 packing bandwidth penalty disappears, (2) fewer interconnections between the STS-48 and STS-1 switch are required[11], and (3) the cost of the OC-48 is less per Gbps than the cost of the OC-3 interface. The cost of the flat network declines less sharply, the decline being due uniquely to (1) the OC-3 into OC-48 packing bandwidth penalty disappearing, and (2) the cost of the OC-48 being less per Gbps than the cost of the OC-3 interface. A similar cost crossover also occurs when all traffic is unprotected.

In the second part of the analysis, we assume that dedicated mesh restoration is required in the flat network to achieve fast restoration of the end-to-end sub-rate circuits. The corresponding cost curves (a) and (c) are shown in Figure 9. The bandwidth overhead of dedicated mesh (1+1) restoration over shared-mesh restoration results in more network bandwidth and more network-side interface capacity. The two-tier architecture is thus cheaper than the one-tier architecture for the entire traffic mix. The cost of the two-tier network declines sharply as the traffic mix moves towards OC-48 for the same three reasons mentioned earlier. The price of the one-tier network declines

[9] Recall though that shared-mesh restoration in the single-tier architecture, even by managing logical OC-48 equivalent circuits, will not achieved the 100 msec restoration time of the two-tier architecture.

[10] On an end-to-end, or path, basis.

[11] Since OC-48 traffic terminates directly on the STS-48 switch.

slightly, due entirely to the fact that the cost of the OC-48 is less per Gbps than the cost of the OC-3 interface. Since there is no OC-3 into OC-48 packing bandwidth penalty for dedicated mesh protection, the bandwidth cost component in the one-tier architecture remains constant.

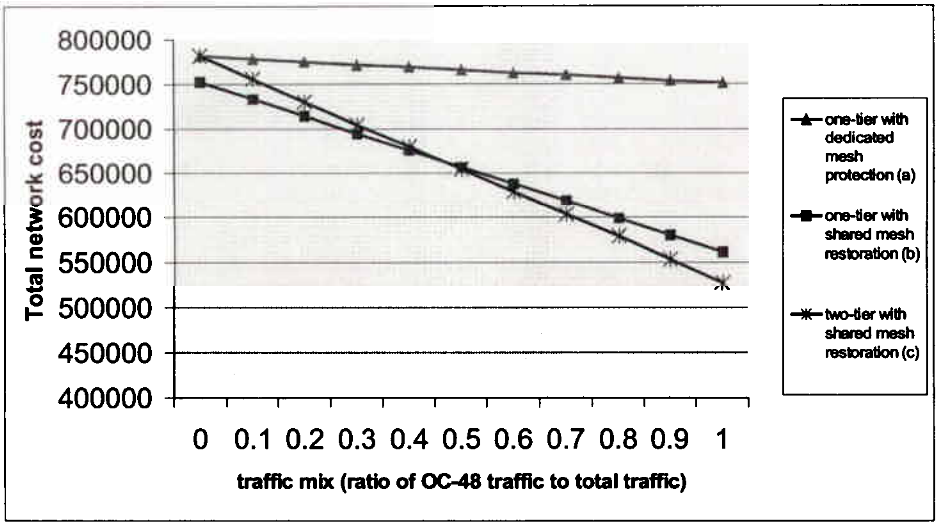

Figure 9 Price comparison when one-tier is capable of dedicated mesh (1+1) protection (a) or shared mesh restoration (b); the layered network performs shared mesh restoration (c)

F. Conclusion

Large-scale transport networks have always been organized hierarchically into multiple layers for scalability and manageability. In this paper we compare two architectures for a core optical mesh network: a flat or single-tier architecture and a layered, or two-tier, architecture. The results are based on the theory of scaling the switching capacity of a node by interconnecting multiple switches. Assuming that multiple switches are interconnected together, we model the interconnection capacity required to operate the resulting switching complex in a non-blocking way in face of inaccurate traffic forecast. Using these results, we show that the layered network is cheaper than the flat network when the traffic scales beyond the capacity of the STS-1 switch, and the proportion of OC-48 demand in the traffic mix exceeds some threshold. While we carried out this simple quantitative comparison based mostly on switch sizes, there are many other aspects that come into play [19] and would impact the cost of single-tier and two-tier core optical mesh network architectures and change the comparison. On the other hand, the argumentation presented here reflects a fundamental behavior and applies to networks in general.

ACKNOWLEDGMENTS

The authors would like to thank Ramu Ramamurthy for his contribution to this work.

REFERENCES

[1] K.G. Coffman and A.M. Odlyzko, "Growth of the Internet", *Optical Fiber Telecommunications IV B: Systems and Impairments*, I. P. Kaminow and T. Li, eds. Academic Press, 2002, pp. 17-56.

[2] T.E. Stern and K. Bala, "Multi-wavelength Optical Networks: A Layered Approach", Reading, MA: Addison Wesley, 1999.

[3] C. Olszewski et al., "Network Migration: Evolution from Ring to Mesh", NFOEC 2003, Orlando, Florida, September 2003.

[4] G. Ellinas et al., "Routing and Restoration in Mesh Optical Networks", Optical Network Magazine, Jan-Feb 2003.

[5] B. Doshi et al., "Optical Network Design and Restoration", Bell Labs Technical Journal, Optical Networking Special issue, Jan-March 1999.

[6] J. Labourdette et al., “Routing Strategies for Capacity-Efficient and Fast-restorable Mesh Optical Networks”, Photonic Network Communications, June-Dec 2002.

[7] Special Issue of Optical Network Magazine on “Standards Activities: Addressing the Challenges of Next-Generation Optical Networks”, Vol. 4, Issue 1, Jan/Feb 2003

[8] G. Bernstein, J. Yates, and D. Saha, “IP-centric Control and Management of Optical Transport Networks”, *IEEE Communications Magazine*, pp.161-167, October 2000

[9] R. Ramamurthy, J-F. Labourdette, S. Chaudhuri, “Scaling Switching Capacity by Interconnecting Multiple Switches”, to be submitted for publication.

[10] Special Issue of Optical Network Magazine on “Telecommunications Grooming”, Vol. 2, Issue 3, May/June 2001.

[11] K. Zhu and B. Mukherjee, “A Review of Traffic Grooming in WDM Optical Networks: Architectures and Challenges”, Optical Network Magazine, Vol. 4, Issue 2, March/April 2003.

[12] A. Akyamac, et al, “Ring Speed Restoration and Optical Core Mesh Networks”, Proc. of 7th European Conference on Networks and Optical Communications (NOC), Darmstadt, Germany, June 2002.

[13] A. Akyamac et al., “Optical Mesh Network Modeling: Simulation and Analysis of Restoration Performance”, NFOEC 2002, Dallas, TX, Sept 2002

[14] M. Goyal, G. Li, J. Yates, “Shared Mesh Restoration: A Simulation Study”, in *OFC '02*, Anaheim, CA, March 2002.

[15] S. Koo, S. Subramaniam, “Trade-Offs Between Speed, Capacity and Restorability in Optical Mesh Network Restoration”, in *OFC '02*, Anaheim, CA, March 2002.

[16] C. Janczewski et al., "Restoration Strategies in Mesh Optical Networks: Cost, Performance and Service Availability", NFOEC 2002, Sept 2002, Dallas, TX

[17] R. R. Iraschko, M. H. MacGregor, and W. D. Grover, “Optimal Capacity Placement for Path Restoration in STM or ATM Mesh-Survivable Networks”, IEEE/ACM Transactions on Networking, vol. 6, no. 3, pp. 325-336, June 1998.

[18] J. Doucette and W. Grover, “Comparison of Mesh Protection and Restoration Schemes and the Dependency on Graph Connectivity”, 3rd international workshop on the Design of Reliable Communication networks (DRCN 2001), Budapest, Hungary, October 2001.

[19] Chris Olszewski et al., “Network Economics”, NFOEC 2002, Sept. 2002, Dallas, TX.