2013

Multi-Layer Design of IP over WDM Backbone Networks: Impact on Cost and Survivability

Byrav Ramamurthy
University of Nebraska-Lincoln, bramamurthy2@unl.edu

Rakesh K. Sinha
University of Nebraska-Lincoln

K. K. Ramakrishnan
AT&T Labs - Research

Design of Reliable Communication Networks (DRCN), 2013 9th International Conference on the
Year: 2013
Pages: 60 - 70

Multi-Layer Design of IP over WDM Backbone Networks: Impact on Cost and Survivability

(Invited Paper)

Byrav Ramamurthy*, Rakesh K. Sinha† and K. K. Ramakrishnan†
* University of Nebraska-Lincoln, Lincoln, NE 68588, USA
† AT&T Labs - Research, Florham Park, NJ 07932, USA

***Abstract*—To address the reliability challenges due to failures and planned outages, Internet Service Providers (ISPs) typically use two backbone routers at each central office to which access routers connected in a dual-homed configuration. At the IP layer, redundant backbone routers and redundant transport equipment to interconnect them are deployed, providing reliability through node and path diversity. However, adding such redundant resources increases the overall cost of the network. Hence, a fundamental redesign of the backbone network avoiding such redundant resources, by leveraging the capabilities of an agile optical transport network, is highly desired.**

In this paper, we propose such a fundamental redesign of IP backbones. Our alternative design uses only a single router at each office but uses the agile optical transport layer to carry traffic to remote Backbone Routers (BRs) in order to survive failures or outages of the single local BR. Optimal mapping of local Access Routers (ARs) to remote BRs is determined by solving an Integer Linear Program (ILP). We describe how our proposed design can be realized using current optical transport technology. We evaluate network designs for *cost* and *performability*, the latter being a metric combining performance and availability. We show significant reduction in cost for approximately the same level of reliability as current designs.

I. Introduction

Traffic on the IP backbones of Internet Service Providers (ISPs) has been continually growing. However, ISP revenue has not kept pace and there is increasing pressure to reduce costs while maintaining high reliability. Reliability for IP networks is typically provided by edundancy to protect against failures, with restoration mechanisms finding alternate routes for affected flows. (Failures include unplanned outages as well as planned maintenance activities. In terms of impact on networks, the main difference is that the planned maintenance activities are typically scheduled during off-peak hours to minimize service impact. Moreover, before a planned maintenance, operators increase routing weight on all affected links to gracefully steer traffic away from them.) Failures of routers are typically handled by having redundant routers at each point-of-presence (POP). The typical deployment of dual homing access routers (AR) to a pair of core backbone routers (BR) to achieve a highly reliable IP backbone is a significant expense, that has been well recognized [9].

Cost reductions for an ISP's backbone are primarily achieved through reduction in the amount of equipment, both in terms of capital expenditure as well as operational costs. Currently, the dominant cost for an IP backbone is the cost of routers, particularly their line cards. But, there is a lot of additional equipment and complex functionality in an ISP's backbone, beyond just the routers and their line cards. Cost reductions by simplifying the network topology at different layers have to be carefully achieved, ensuring a proper tradeoff between cost and reliability. Reduction of equipment and costs at Layer 3 (router and line cards) should not result in significant additional deployment of components and capacity at a different layer. At the same time, moving to a simpler architecture to keep costs low, where for instance only a single backbone router exists at each POP, should not result in unacceptable availability.

The cost of transport equipment (transponders, ROADMs, regenerators, amplifiers and fiber) in an ISP's network is a significant contributor to the overall cost. We observe that there can be significant opportunities for sharing transport resources provisioned for restoration if the network primarily experiences a single failure at a time. (A single failure means planned maintenance or unplanned outage of a single network equipment. Notice that a single failure can bring down multiple links. E.g., failure of a single router fails all its adjacent links.) We recognize that there may be situations where multiple failures occur concurrently, but we consider these to be a lower probability event, and also more expensive to protect against. Therefore, we consider the appropriate cost-reliability tradeoff to be one where single failures are handled without impacting reliability adversely. Carriers generally build networks with headroom (overprovisioning) for both failure restoration as well as for future growth. This capacity can be shared across different possible failures.

In the approach we pursue in this paper, we envisage a network with only one backbone router (BR) at each POP. The access routers (ARs) homing on that primary BR under normal operation instead home on a remote BR when there is a failure of that primary BR. However, having the access routers home on the remote BRs require transport capacity to be provisioned for this purpose. The novelty in our design approach is to share the capacity in the network across different possible single failures without incurring protocol latencies at the IP layer to recover from a failure. We also propose that the capacity provisioned between the access routers and the remote BR under normal operation is minimal (and the links are assigned a high weight in the IGP protocol such as OSPF). Thus, the

access routers have an adjacency established both with the local primary BR as well as the remote backup BR. When the local primary BR fails, the transport resources are resized to have sufficient capacity to carry the traffic to and from the access routers homed on the corresponding remote BR. This design avoids the large IGP convergence latency that is often the case when a new adjacency is established, along with all the delays to establish the tranport circuit. We envisage an intelligent and agile Layer 1 network that can dynamically resize the transport circuit (we could certainly consider setting up a link-aggregation group that then has additional components added subsequent to detecting a failure).

In related work, Palkopoulou *et al.* [9], [8] performed a cost study of different architectural alternatives. They consider each access router connected to one or two backbone routers as well as having optical switches and/or a common pool of shared restoration resources. Huelsermann *et al.* [4] provide a detailed cost model for multi-layer optical networks, which we use in this paper. Chiu *et al.* [2] report a 22% savings for integrated IP/Optical layer restoration in a dual router architecture compared to pure IP based restoration. Their key idea is to move all inter-office links from a failed BR to another (surviving) BR in this POP using the optical layer. In an earlier paper [11], we describe the cost and reliability considerations involved in designing next-generation backbone networks.

Our proposal is to achieve a fundamental redesign of IP backbones that avoids redundant routers by exploiting the capabilities of agile optical transports. We evaluate the alternative backbone designs in terms of cost and performability (a metric combining performance and reliability). Section II includes a detailed description of the operation of the network at the IP and the transport layer In Section III, we propose alternative backbone network designs which use only a single router at each POP but use the agile optical transport layer to carry traffic to the remote BRs in order to survive failures or outages of the single local BR. Subsection III-A describes a possible realization of the proposed design using current optical transport technology. In Section IV, we describe our evaluation metrics, viz., cost and performability. In Section V, we describe the Integer Linear Program (ILP) formulation used to solve the problem of optimally mapping local Access Routers (ARs) to remote BRs in the new backbone network design. In Section VI, we describe the results comparing the cost and performability of our alternative design to that of the original design for a network modeled after a Tier-1 ISP backbone network. We then present our conclusions in Section VII.

II. ISP Backbone Architectures: Background

The backbone network of a typical ISP can be quite complex, comprising multiple layers. Customer equipment homes on access routers (AR), which in turn are connected to the core backbone routers (BR). BRs are located at point-of-presence (PoP, often called a central office). An ISP may have a large number of ARs that aggregate traffic into a BR.

Each PoP typically houses two BRs, with links between routers in the same PoP being typically Ethernet links over intra-office fiber. ARs are dual-homed to two BRs to provide the necessary level of service availability. ARs that are co-located (or close) to a PoP connect to the two BRs within the same PoP; ARs that are remotely located may be connected to two different PoPs. Figure 1 shows an AR connected to two BRs within the same PoP. The inter-office BR-BR links use underlying ROADM network. While this type of redundant backbone router configuration in a PoP is typical of large ISPs, it can be expensive. We can reduce cost by keeping only one BR in a PoP and then homing each AR to exactly one BR. However the resulting architecture will likely have unacceptable availability because any customers homed on this AR will lose their connectivity when this BR goes down.

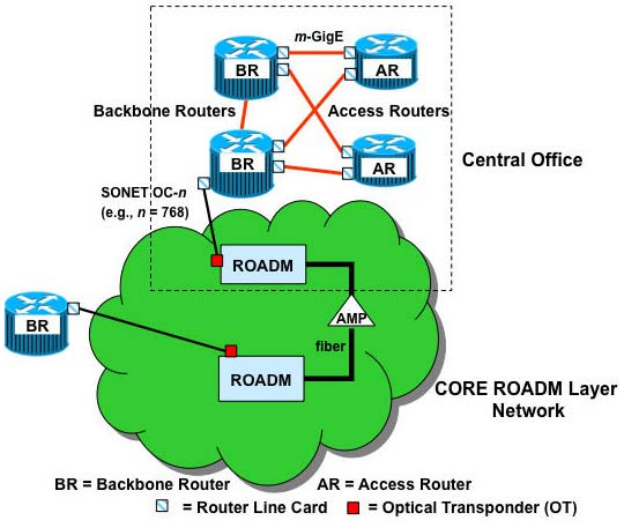

Fig. 1. Legacy backbone network.

Even though routers have become very reliable and (unplanned) complete router failures are rare, routers experience frequent outages because of planned software and hardware upgrades. A few router vendors support in-service software upgrades but as argued in [1], there is still a large base of deployed routers without such capability. Approaches for providing limited redundancy using mechanisms such as the Hot Standby Router Protocol (HSRP) etc., tend to be expensive. The overall effect is that upgrades still have a substantial impact [6] and 1:1 router redundancy remains a prevalent practice in carrier networks.

A. Physical Layer Restoration

When we attempt to provide resiliency to failures at the IP layer, we see the need to add redundant link(s) to the topology. The addition of a redundant link between two routers actually involves the set up of a multi-hop link over a complex topology at the lower layers. Furthermore, the creation of a backup link needs to ensure that it does not share components with the primary link (e.g., an amplifier, fiber or a ROADM). Moreover, components such as ROADMs themselves have

complex failure characteristics. Similarly, when seeking to reduce the cost of components in Layer 3 (router and its line cards) of the network, it is important to understand the impact of that reduction with the concomitant increase in cost and complexity at the lower layers, as well as the impact on availability. Thus, it is useful to examine the following questions: where is it most appropriate to provide restoration capabilities - at Layer 3 or should it be at a lower layer, or should it be a combination?

One of the arguments made against providing restoration exclusively at a lower layer (e.g., such as SONET) is that it is possibly inefficient because of the need to add substantial extra capacity for protection without the ability to take advantage of the statistical multiplexing that packet switching provides. Furthermore, one would still have to deal with failures of components at the higher layer (e.g., router line cards) [10]. One approach is to provide restoration at Layer 3. However, this comes at the cost of availability (including the time taken to restore from a failure), because the recovery from a failure is through complex distributed protocols that are dependent on timers that are set to large values. These considerations have led carriers to add protection at different layers on an ad-hoc basis to compensate for the different failure recovery capabilities at each layer and cost considerations. Thus, the overall system has evolved to be both expensive and complex. An additional observation is that carriers have to continually redo such evaluations and deployment of restoration mechanisms and capacity each time technology at a particular layer changes.

B. IP and MPLS Restoration

The traditional way of providing reliability in the IP network is to provide redundancy, especially at the router level in the IP backbone. IGP convergence tends to be slow. Production networks rarely shorten their timers to small enough values to allow for failure recovery in the sub-second range because of the potential of false alarms [3]. A common approach to providing fast recovery from single link failures is to use link-based FRR. While some level of shared redundancy is provided to protect against link failures, such as sharing of backup resources for mesh restoration (e.g., MPLS Fast-Reroute), the traditional means for providing protection against backbone router failures is to have a 1:1 redundant configuration of backbone routers at each PoP. So an AR in a non-zero OSPF area typically connects to two BRs in the backbone area (i.e., area 'zero'), to protect against single router failures. The 1:1 level redundancy is provided so that the traffic from the ARs feeding into each BR can be carried by the single BR, if second BR fails. Similarly, the link from the AR to the BR has to have sufficient capacity to carry all this traffic. Moreover from each BR, there has to be enough capacity to the other BRs at the different offices in the core backbone. As the capacity requirements go up, this approach of providing redundant BRs results in a dramatic increase in cost for the service provider to meet reliability requirements and allow for uninterrupted service even in the presence of a complete BR failure. We see a need therefore to modify the way service providers build their reliable IP backbone environments so that there is a reduction in cost while still providing the level of reliability expected from the backbone.

C. Optical Transport Layer Considerations

Inter-office links connecting backbone routers establish Layer 3 adjacencies used by protocols such as OSPF. In reality, a single inter-office link is a logical (or aggregate) link comprising, possibly, multiple physical links (such as SONET *circuits*) with potentially different capacities (e.g., OC-192 for 10G and OC-768 for 40G). In Fig. 2, for example, three 10G circuits between routers R1 and R2 form an aggregate link of capacity 30 Gbps. Aggregate links are used to reduce the number of OSPF adjacencies. A local hashing algorithm is used to decide which of the three circuits to use for any IP packet going over this aggregate link.

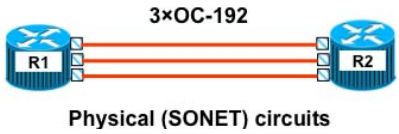

Fig. 2. Physical links that make up an aggregate L3 link.

Each physical link occupies either a complete wavelength (e.g., 40G circuit in a 40Gbps wavelength system) or a sub-wavelength (e.g., 10G circuit in a 40Gbps wavelength system). To carry such a physical circuit (either 10G or 40G), an *optical transponder* (OT, or simply, a transponder) should be installed at either end of a wavelength path. An optical transponder is a device which enables the transmission and reception of a client signal over a wavelength in the WDM transport layer using optical-to-electronic-to-optical (O/E/O) conversion. The type of transponder (e.g., 10G transponder or 40G transponder) is chosen depending on the capacity of the circuit that needs to be carried over a wavelength. Usually it is cheaper to carry multiple sub-wavelength circuits over a single wavelength than to carry them separately. This requires the use of a special device known as a *muxponder*. The muxponder combines the functionality of a multiplexer and a transponder. With a muxponder at each end of a 40Gbps wavelength path, upto four 10G circuits can be carried across it. Such a wavelength path which has been partitioned to carry sub-wavelength circuits is called a *multiplex* link (see Fig. 3).

The combined cost of four 10G transponders tend to be higher than the cost of a 4 x 10G muxponder. So, in practice, multiple 10Gs are carried over a single wavelength using a muxponder and 40G transport equipment. However in rare cases, where we have only a single 10G circuit (and no anticipated capacity growth), it may be cheaper to use 10G transport equipment.

The wavelength paths mentioned above originate and terminate at ROADM nodes in the optical transport layer of the network. Usually a ROADM node is located at each office adjacent to a backbone router. The router port is connected to

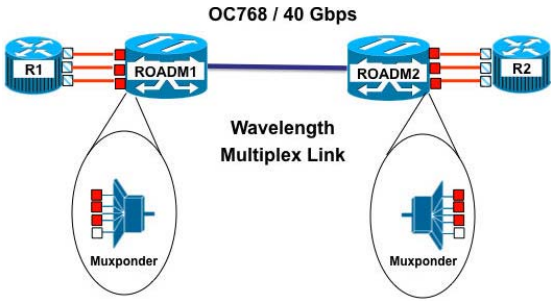

Fig. 3. Routing of the physical links over a multiplex link. Transponders in use are represented by filled squares and those not in use by empty squares.

the ROADM using a short-reach wavelength (termed λ_0) transported over a fiber-optic cable with its own pair of transponders. The ROADM is a device which allows optical signals (wavelengths) to be added, dropped or bypassed (switched) in a reconfigurable manner. A fully flexible ROADM is colorless, directionless and non-blocking. This means that any subset of wavelengths can be switched from any input fiber to any output fiber. The ROADM through its add and drop capability allows for a wavelength to be regenerated using an O/E/O method. Regeneration is essential to clean up the wavelength signal to overcome bit-error rate (BER) degradation due to noise and crosstalk. Regneration is performed on each individual wavelength as needed and involves the use of a special device known as a *regenerator* (or simply, a regen). Although a regen can be constructed using two transponders placed back-to-back, it can often be constructed in a simpler manner and at lower cost.

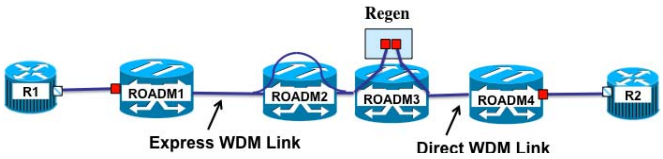

Fig. 4. An Express link can bypass regeneration at intermediate ROADM nodes. The express link from ROADM 1 to ROADM 3 bypasses regeneration at ROADM 2. Regenerators on the circuit are represented by two adjacent filled squares.

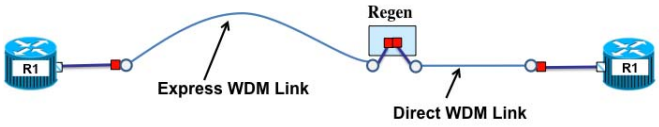

Fig. 5. A physical link can span multiple Direct WDM links and Express WDM links.

Neighboring ROADM nodes are connected using an optical path consisting of one or more fiber segments separated by optical amplifiers. Note that an amplifier is a purely optical device which is used to combat signal attenuation by boosting the power of all the wavelengths carried by the optical fiber. Such an optical path connecting adjacent ROADM nodes is termed as a *Direct WDM* link (see Fig. 4). A regenerator-free path can span multiple fiber segments and multiple ROADM nodes depending on the *optical reach* of the signal. The optical reach is a vendor-specific metric that is dependent on various physical parameters of the components. Thus a regenerator-free optical fiber path traversing multiple fiber segments and multiple ROADM nodes can be used to connect two distant nodes and such a link is termed as an *Express WDM* link A physical link (e.g., a 10G circuit) between two routers can span multiple Direct WDM links and multiple Express WDM links (see Fig. 5). In addition each of these WDM links can be multiplexed to carry sub-wavelength circuits (e.g., 4x10G circuits over a 40 Gbps wavelength).

III. Architecture Alternatives

In this section we describe the different architecture alternatives that use a single backbone router (BR) at each POP as a means of reducing cost along with the transport alternatives to carry the traffic to a remote backbone router.

The first option (Option-1) for reducing the cost of a backbone is to eliminate one BR from each POP, thus avoiding the cost of the additional BR. While this may be a simple approach, we still need to ensure that this is done in a manner that the reliability of the service provided by the backbone network is not adversely impacted. Of the two BR_1 and BR_2 in each office, we eliminate BR_2 and move all of its links to BR_1. The cost reduction comes from eliminating roughly half of AR-BR links and all of $BR_1 - BR_2$ intra-office links. However this design cannot protect against any BR outage and our performability evaluation in Section VI shows an unacceptable drop in performability. Option-1 is thus referred to as UR, 'Unreliable design', in Section VI.

To improve performability, our second option (Option-2) improves on Option-1 by adding a link from each AR to an additional BR, located in a remote POP (called *remote BR*, in the rest of the paper). While this does save the cost of the $BR_1 - BR_2$ intra-office links, it results in increased transport cost for connecting the ARs to the remote BRs. It also saves the chassis cost of the eliminated BR_2s but may require extra line cards (with the expectation that this does not result in an additional chassis) on the remote BRs as we have to add more links to the remote BRs. The number of inter-office links, which tends to dominate the layer-3 cost, does not change substantially as we are effectively replacing each "AR - (local, second) BR" link with an "AR - (remote) BR" link.

The final option (Option-3) improves on Option-2 dynamically by setting up an "AR - remote BR" link upon failure of the local BR. We first eliminate the BR_2 router from each office and identify a remote BR for each AR. However, instead of setting up permanent full capacity "AR-remote BR" links (as in Option-2), we size these links dynamically upon failure of the local BR, taking advantage of the agility available in newer optical transport equipment. Since we design for a single BR failure at a time, we need at most one "AR - remote BR" link at any given time. The cost advantage over Option-2 comes from multiplexing gains achieved by sharing the backup capacity, as we may be able to share transport resources as

well as ports on ARs and remote BRs. We illustrate the source of savings in router ports with the following example. Suppose three ARs connect to the same remote BR and require (respectively) 8, 9, and 7 10G-connections upon failure. In Option-2, this will require $8+9+7=24$ 10G ports. However with Option-3, we will only need $\max(8,9,7)=9$ 10G ports. We also get multiplexing gains from the sharing of transport resources among AR-(remote) BR connections. Moreover in Option-2, each AR will need enough ports to connect to its local BR as well as to its remote BR. However, in Option-3, we can reuse the same AR ports to connect to either the local or the remote BR with the use of flexible fiber crossconnects.

We refer to Option-3 as our "proposed architecture" and denote it as SR-100 in Section VI. Option-2 is not discussed further in this paper.

A. Realization of the proposed architecture

Note that the creation of a dynamic link would likely result in a new router adjacency. The introduction of a new router adjacency dynamically would typically causes a large latency for the IGP protocol to converge, thus increasing the outage restoration time. Our solution (first proposed in [11]) is instead to set up a permanent AR-remote BR link at a *low rate* to maintain protocol state (e.g., using keep-alive messages). Upon a failure, we dynamically resize this link to the required full rate. This avoids bringing up new router adjacencies as well as propagation of failure information through Link State Advertisements (LSA). The AR whose local BR has failed can recover connectivity to the rest of the network, through its remote BR adjacency, without the need for the entire network to converge. We recognize the possibility of short-term congestion, while the network is converging, but overall the complete reroute process would be transparent to the routing control plane. It is therefore similar to the case of having two BRs in each POP, but at a significantly reduced cost.

We propose to use a service platform similar to that utilized by AT&T's GRIPhoN project [5]. A simplified diagram of a POP is shown in Figure 6. For simplicity, we show only one AR located in the POP even though in reality we have several ARs homing on this BR. In some cases, the ARs may be 100s of miles away from this BR. The BR, AR, ROADM, and OTN equipment (not shown) are interconnected by an FXC (fiber crossconnect) switch. Under failure free operation, an AR has several 10G connections to the local BR. In our design, it also has a low-rate connection to a pre-determined remote BR. Upon failure of the local BR, we resize the AR - remote BR connection. One way of achieving this is to set up a Link Aggregation Group (LAG) between the AR and the remote BR and add additional individual circuits to it as needed. We exploit the OTN layer for sub wavelength circuits (e.g., for setting up the initial low rate 1G, ODU-0, connection, as in [5]), and the DWDM layer is used for adding wavelength connections, e.g., multiples of 40G. We use an FXC to reuse the ports that are on the AR to the local BR. As shown in Figure 6, upon failure of the backbone router (BR) at POP A, all the ports on the access router at POP A are connected to the BR at the remote POP B.

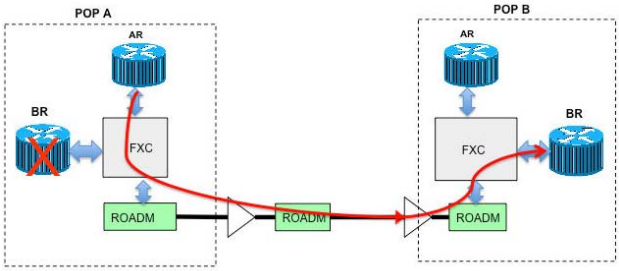

Fig. 6. Re-homing upon BR failure.

IV. Evaluation of network designs

We evaluate network designs for *cost* and *performability*. The overall cost of the backbone network includes the cost of the routers (chassis, line cards) as well as all of the transport equipment used for the interconnection of the routers.

A. Router cost

Router equipment includes router ports, line-cards, and chassis. Given the set of circuits in a design, we compute the required number of 10G Ethernet ports as well as 10G and 40G transport ports. Then we estimate the number of line-cards (and chassis) based on the number of required ports. The access topology in production networks tends to be extremely complicated, with a mixture of low-rate and high-rate connections. There are various aggregator switches or routers arranged in hierarchical pattern for multiplexing low-rate connections so as not to exhaust ports on BRs. We considered a simplified access model and assumed that (a) each AR is located in a POP and (b) AR-BR links are 10G Ethernet.

B. Transport layer cost

Transport equipment includes optical transponders, regenerators, and muxponders. Transponders and regenerators are used on a per circuit basis so the cost of a circuit is the cost of the two optical transponders (one on each end) and the cost of the regenerator(s). Muxponders are shared over multiple sub-wavelength circuits (in our case, 4 10G circuits) so we charge each circuit one-fourth the cost of a muxponder. In addition, we have pre-deployed amplifiers, ROADMs, and fibers. Because this last set of transport equipments is *common* to multiple circuits (e.g., one amplifier is used in all wavelengths), we amortize the 'common' cost contribution to any circuit on a per wavelength-km basis.

For a 40G circuit, the cost consists of the 40G transponders and 40G regenerators used on each WDM link of the end-to-end path. Note that such WDM links can be either a Direct WDM link or an Express WDM link. For example, in Fig. 7, a new 40G circuit is carried over 2 Express WDM links (curved lines) followed by a Direct WDM link (straight line). Hence

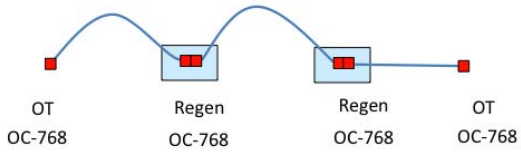

Fig. 7. Cost of a 40G circuit.

the cost for the circuit is 2 x (Cost of 40G transponder) + 2 x (Cost of 40G regen).

For a 10G circuit, the cost computation is a bit more complicated. This is due to the fact that a 10G circuit is often carried over a multiplex link (see Section II). Deploying a pair of muxponders to create the first sub-wavelength circuit on a 40G WDM link ensures that additional sub-wavelength circuits do not incur this cost again. Note that both 10G regenerators and 40G regenerators may appear in the end-to-end path carrying a 10G circuit.

Thus the cost for a new 10G circuit depends on whether a new multiplex link needs to be created in the network or not. If a new multiplex link is created, it may possibly use a sequence of Express WDM links and Direct WDM links. For illustration purposes, we describe four options for carrying a 10G circuit across the transport network. In all of these figures, new components are shown in dark shaded portions, straight lines represent Direct WDM links, curved lines represent Express WDM links and wavy lines represent multiplex links.

In Case 1 (see Fig. 8), the new 10G circuit uses a new multiplex link carried over two Express WDM links (curved lines) and a Direct WDM link (straight line) using an unused wavelength on each link. The wavelength is operated at 40 Gbps and muxponders are used at both ends to carry the new 10G circuit. The total cost for carrying the 10G circuit is 2 x (Cost of 10G transponder) + 2 x (Cost of 40G regen) + 2 x (Cost of muxponder). Here for comparing different options, we ignore the common cost (ROADMS, fiber, etc.). Note that three additional 10G circuits may be carried over this end-to-end multiplex link in the future thanks to the muxponders.

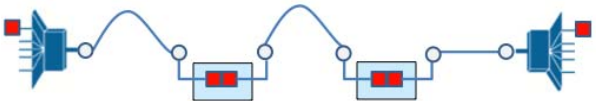

Fig. 8. Case 1: Using a new multiplex link routed over two Express WDM links and a Direct WDM link.

In Case 2 (see Fig. 9), the new 10G circuit is carried over such an existing multiplex link. The total cost for carrying the 10G circuit is 2 x (Cost of 10G transponder).

Fig. 9. Case 2: Using an existing multiplex link.

In Case 3 (see Fig. 10), the new 10G circuit is carried over an existing multiplex link and a new multiplex link that spans two Express WDM links and a Direct WDM link. An unused wavelength is operated at 40Gbps on each WDM link and muxponders are used at both ends to carry the new 10G circuit (similar to Case 1). The total cost for carrying the 10G circuit is 2 x (Cost of 10G transponder) + 2 x (Cost of 40G regen) + 2 x (Cost of muxponder) + (Cost of 10G regen). The additional cost incurred here (compared to Case 1) is due to a 10G regen which is required to transport the 10G circuit across the old and the new multiplex links.

Fig. 10. Case 3: Using an existing multiplex link and by creating an additional, new, multiplex link.

Finally, in Case 4 (see Fig. 11), the new 10G circuit is carried over two existing multiplex links. The total cost for carrying the 10G circuit is 2 x (Cost of 10G transponder) + (Cost of 10G regen).

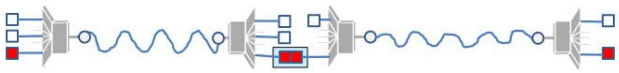

Fig. 11. Case 4: Using a sequence of two existing multiplex links.

As noted earlier, different from these four cases, a new multiplex link may use other sequences of Express WDM links and Direct WDM links.

C. Network cost

For computing network cost, we used normalized equipment prices reported in [4], which are based on data from IST's Next Generation Optical Networks for Broadband European Leadership (NOBEL) project phase 2. Notice that these are different from the equipment price numbers used in our previous paper [11]. These relative costs should be treated as examples only, for illustrating the efficacy of our approach, across a wide range of variation of the relative costs of the various components used in a typical IP backbone network. Equipment prices tend to vary over time and so in Section VI, we include a sensitivity analysis of how our estimated savings change with equipment prices.

D. Performability

For evaluating *performability*, we used the *nperf* tool [7]. The tool considers a set of *failure scenarios* representing failures of one or more network components. For each failure scenario, we first determine the set of failed circuits. A single component failure can bring down multiple circuits. E.g., when a router fails, all its incident circuits also fail; an amplifier failure or fiber cut fails all circuits routed over those components. The set of failed circuits in a scenario is the union of failed circuits from the failure of the individual network components.

Next we determine the effect of these failed circuits on *logical* links. Recall that a logical link may be an aggregate of multiple circuits that gives the appearance of a single link to the routers. If only a subset of the member circuits fail, then the net effect is a reduction in this (aggregated) logical link's capacity but the link does not fail. In this paper, we assume that the network uses OSPF routing. If none of the links fail, then the flows stay on their original routes but may experience packet loss due to congestion as some of the links in the route may have reduced capacity. If some the links fail then OSPF routing recomputes a new set of routes (based on routing weights assigned to each link) and reroutes some of the flows. There are two possible sources of packet loss: (a) it is possible that a failure scenario may disconnect the network graph and thus a flow may not have any possible route. Even if a flow, with failed links in its current route, does have an alternate route, it takes several seconds to detect the failure and reroute this flow, during which time some packets get lost. We broadly categorize this type of packet loss as resulting from *unavailability of routes* (b) the amount of flow sent on a link may exceed its capacity. This may happen either because a link lost a subset of its member circuits (and thus has reduced capacity) or because many different flows got rerouted to this link. We categorize this packet loss as resulting from *link congestion.*

For each failure scenario, we determine amount of traffic loss due to *unavailability of routes* and *link congestion.* In addition to the loss computation the tool also computes the probability of this failure scenario, based on vendor and proprietary field tested mean time between failures (MTBF) and mean time to repair (MTTR) of different components. In our evaluations, we consider $10,000$ of the most likely failure scenarios. The end results are two probability distributions of traffic losses: (a) due to unavailability of routes, and (b) due to link congestion. While comparing different designs in Section VI, we report '$1-$expectations of these distributions' and (respectively) call them *No route* and *Congestion* performability. E.g., a 'no route' performability of 0.999 means that over a long period of time, we expect $1-0.999=0.0001$ fraction of traffic to have no route.

E. Cost-Performability trade-off

There is an obvious trade-off between cost and performability. E.g., increasing link capacities improves congestion performability but also increases the cost. So the cost and performability of a design should always be considered together and not in isolation. In our evaluations, we considered a design goal of surviving all single failures (router ports, complete router, amplifier, optical transponder etc.) to determine the appropriate capacities on links. Then we ran *nperf* tool on 10,000 most probable single and multiple failure scenarios to evaluate the performability. Considering single failures is a standard practice because these failures cover a large fraction of failure probability space. However we want to emphasize that this design heuristic is one of several possible reasonable choices. We could reduce the capacities a bit to reduce cost at lower performability or increase capacities to increase cost and performability. Ultimately the real merit of our results is that we show substantial cost savings while offering acceptable performability. The exact trade-off between cost and performability in our designs can be somewhat adjusted depending on the requirement of the ISP.

F. Baseline network design

We used the following method to compute link capacities that are barely sufficient to survive any single failure. We started with a model where each logical link was an aggregate collection of 10G and 40G circuits. We allow for the possibility that a link may have a *single* circuit.) We simulated all single failures using the *nperf* tool. For each failure, we computed the circuits that go down and how those affected flows get rerouted. Then for each logical link, we obtained the highest utilization across all single failures. If this utilization was less than 100%, we reduced the number of circuits in that (aggregate) logical link. However removal of circuits also resulted in removal of the corresponding network elements – regenerators, OTs, router ports – thus changing the set of single failures and therefore the highest utilization. If this utilization was more than 100%, we added circuits on that (aggregate) logical link which, similar to the previous case, changed the set of single failures and thus the highest utilization.

So we iterate over this process, each time adding or removing circuits, until we had a network where the maximum utilization for single failure was close to 100% for all links. Finally if a logical link contained four or more 10G circuits, we replaced them with a 40G circuit. This network design is referred to as BL, Baseline design, in Section VI.

V. ILP formulation for optimal AR-remote BR mapping

We start with a design where each POP has only one BR that all ARs in this POP are connected to. As outlined earlier, when the local BR fails, the traffic from that office moves to a pre-determined remote BR. We need to find the mapping from ARs to remote BRs that minimizes the additional network cost while ensuring that all flows originating at this POP have a route with sufficient link capacities. We consider a generalized version of the problem where each flow is classified either as *priority* or as *best-effort* and we only need to worry about restoring priority traffic. If there is no class-of-service, all traffic is treated as being restorable – hence the same as 'priority' traffic.

We find this optimal mapping using an integer linear program (ILP). The ILP formulation assumes that the routing of a circuit only depends on the two end-points of a circuit, e.g, along shortest path on the L1/L2 network. A more ambitious goal, not undertaken here, would have been to consider a joint optimization of AR to remote BR mapping as well as routing of the needed circuits.

All ARs connected to a given BR are mapped to the same remote BR. So for modeling purpose we collapse all these ARs into a single (super) AR. Given n POPs, we number ARs

and BRs from $1 \ldots n$ and without loss of generality assume that the (super) AR and BR in the i-th POP ($1 \leq i \leq n$) are both numbered i. Thus the i-th AR is connected to its (local) i-th BR and can be remotely connected to any of the remaining $n-1$ remote BRs. So altogether there are n^{n-1} possible connections.

The additional cost of protecting BR failures has four different components.

A. Cost components for AR-remote BR mapping

1) At each AR, we need additional (10G) OTs to set up links to remote BRs. We do not need any additional (10G) router ports because we can reuse the router ports used to connect to the local BR. The number of additional OTs is equal to the number of router ports used by priority traffic and can be precomputed as m_i (corresponding to the number of 10G links) for the i-th AR.
2) At each BR, we need additional (10G) OTs and (10G) router ports to accommodate the link from remote ARs. The number M_j of OTs/ports at the j-th BR is maximum m_i across all the ARs mapped to this BR. Notice that we also have a (permanent) low rate connection between ARs and remote BRs that we are not accounting for in the above statement. E.g., if two ARs map to the same remote BR and they each require 50Gbps of uplink then we will need two 1Gbps (ODU-0) connection permanently and will have to resize one of the 1Gbps connection to 50Gbps connection upon failure of the local BR. So the total additional capacity needed on the remote BR will be 51 Gbps and not 50Gbps. A similar statement applies to additional OTs on ARs. Because the cost of these permanent connections are small compared to the the rest of the cost and to keep the ILP formulation simple, we ignore them in the remainder of this section. They can be added to the final cost once we have determined the AR to remote BR mappings.
3) If the i-th AR is mapped to jth-remote BR, then we need transport capacity (equal to the amount of priority traffic from i-th AR) to set up this link upon failure of the local BR. Transport cost includes cost of regenerators, fiber, ROADM, and amplifiers.
4) We may also need *additional* capacity at certain inter-office BR-BR links. Consider the following scenario: Suppose we decide to map the AR at location L_1 to the remote BR at location L_2 and let L_3 be a different location. When the local BR at L_1 fails, all the traffic that was flowing between L_1 and L_3 now shifts and is carried between L_2 and L_3. It is possible that the links on L_2 do not have enough capacity to carry all this traffic and some of the links in the $L_2 - L_3$ path require additional capacity.

B. Parameters and variables

- Parameter m_i is the number of 10G links needed at the i-th AR to carry all its priority traffic. This can be determined from the number of 10G ports on the i-th AR needed to carry all its priority traffic and is an input to the ILP.
- Variable M_j is the number of additional 10G OTs/ports required at the j-th BR for setting up the AR-remote BR links in the final mapping.
- Variable r_{ij} is 1, if i-th AR is connected to j-th BR upon failure of its local BR; and is 0 otherwise.
- We index network resources from 1 to N. Then let parameter s_k^{ij} be the number of units of the k-th resource we need for cost items (3) and (4) above upon "remapping" of the i-th AR to the j-th BR, when the i-th BR fails.
- Parameter c_k is the unit cost of the k-th resource.
- Variable S_k is the number of units of the k-th resource we need in the final mapping.

We can precompute s_k^{ij} as following:

1) Add a link from the i-th AR to the j-th BR in the $nperf$ model. This link should have capacity $m_i \times 10G$ and very high OSPF weight.
2) Use $nperf$ to simulate the failure of the (local) i-th BR and compute utilization of each edge needed to restore all priority traffic from this AR.
3) For any edge with utilization above 100%, determine the extra capacity needed to keep utilization under 100%. E.g., if the utilization is 200%, we need to double the capacity.
4) Route all additional inter-office links as well as $m_i \times 10G$ capacity from i-th AR to the j-th BR link. The number of k-th resource needed for this set of circuits is s_k^{ij}.

Tables I and II summarize the parameters and variables.

	Number	Description
n	1	Number of ARs/BRs
N	1	Number of transport resources
m_i	n	Number of 10G OTs at the i-th AR for priority traffic; can be precomputed from the number of ports on the i-th AR
c_k	N	Unit cost of the k-th transport resource
s_k^{ij}	Nn^2	Number of units of the k-th transport resource needed upon remapping of the i-th AR to the j-th remote BR, when the i-th BR fails.

TABLE I
PARAMETERS IN THE ILP FORMULATION

	Number	Description
M_j	n	Number of 10G OTs/ports at the j-th BR
r_{ij}	n^2	Boolean variable: 1 iff i-th AR gets mapped to j-th remote BR
S_k	N	Number of units of the k-th transport resource needed

TABLE II
VARIABLES IN THE ILP FORMULATION

C. Objective function

Our goal is to minimize the total cost of the network. Thus, the objective is:

$$\min[\sum_{1\le j\le n} M_j*(\text{cost of 10G OTand10G port})+\sum_{1\le k\le N} c_k*S_k]$$

To compute the additional network cost, we need to add, to the ILP solution, $m_i *$ (cost of 10G OTs) and the cost of maintaining the AR to remote BR *permanent* connections.

D. Constraints

1) Each AR is connected to exactly one BR:

$$\forall i, \quad \sum_{1\le j\le n} r_{ij} = 1.$$

2) Each AR is connected to a remote BR (to remove the degenerate case of AR having two connections to its local BR):

$$\forall i, \quad r_{ii} = 0.$$

3) Each BR needs ports/OTs equal to the maximum number of ports on one of its connected ARs:

$$\forall j, \quad M_j = \max_i\{m_i | r_{ij} = 1\}.$$

Because m_i's are input to the ILP (not variables), the above constraint can be equivalently expressed as n^2 linear constraints:

$$\forall i, j, \quad M_j \ge m_i * r_{ij}.$$

(When $r_{ij} = 0$, the inequality is vacuously true. When $r_{ij} = 1$, the inequality becomes $M_j \ge m_i$ and because we are minimizing M_j in our objective function, we know that one of these inequalities will be tight and we will get $M_j = \max\{m_i | r_{ij} = 1\}$.

4) Because we consider at most one BR failure at a time, the additional units of k-th transport resource is maximum across all AR to remote BR mappings:

$$\forall k, \quad S_k = \max_{i,j}\{s_k^{ij} | r_{ij} = 1\}.$$

As with the previous constraint, this is equivalent to Nn^2 linear constraints:

$$\forall i, j, k, \quad S_k \ge s_k^{ij} * r_{ij}.$$

For a fixed i and k, the n constraints are $\forall j, S_k \ge s_k^{ij} * r_{ij}$. However we have a separate constraint stating that for a fixed i, exactly one r_{ij} is one and the rest are zero. So we can rewrite these n constraints as a single constraint (albeit of n terms) $S_k \ge \sum_j s_k^{ij} * r_{ij}$. Thus the above set of constraints can be rewritten as Nn constraints:

$$\forall i, k, \quad S_k \ge \sum_j s_k^{ij} * r_{ij}.$$

VI. Results

We started with the topology and traffic matrix modeling a Tier-1 ISP backbone network. This is a *baseline* design to estimate changes in cost and performance of our proposed designs. This model has POPs in major US cities, where each POP houses two BRs connected by a set of 10G Ethernet links. Each AR is located in a POP and is connected to two BRs in its POP by a set of 10G Ethernet links. Each inter-city link connecting BRs is an aggregate of 10Gs and 40Gs. As explained in Section IV, we sized each logical link to survive all single failures. Because of the long ordering cycles for additional capacity, production networks always have excess capacity for anticipated traffic growth. This would have inflated our cost savings as we would be starting with a network of higher cost than necessary. So to create a fair baseline, we resized the capacities on each link to barely survive all single failures of router ports, complete routers, amplifiers, fiber cuts, and optical transponders. This design is referred to as BL in Table III.

Design name	Design description
BL	Baseline design. Each AR is dual homed to two local BRs. Restoration design to protect all traffic upon any single failure
UR	Unreliable design. Each AR homes to a single local BR. Restoration design to protect all traffic upon any single failure except complete router outage. Drop all traffic from ARs when their local BR fails.
SR-100	Each AR homes to a single local BR. Assume 100% of the traffic is priority. Restoration design to protect *all* traffic upon any single failure. Rehome ARs to a remote BR when their local BR fails.
SR-75	Assume 75% of the traffic is priority. Each AR homes to a single local BR. Restoration design to protect *priority* traffic upon any single failure. Rehome ARs to a remote BR when their local BR fails.
SR-50	Assume 50% of the traffic is priority. Each AR homes to a single local BR. Restoration design to protect *priority* traffic upon any single failure. Rehome ARs to a remote BR when their local BR fails.
SR-25	Assume 25% of the traffic is priority. Each AR homes to a single local BR. Restoration design to protect *priority* traffic upon any single failure. Rehome ARs to a remote BR when their local BR fails.

TABLE III
DESIGN NAMES AND DESCRIPTIONS

A simplistic option to reducing the cost of a backbone is to eliminate one BR from each POP and then moving all of the links from the removed BR to the surviving BR. For inter-city BR-BR links, we sized their capacities to survive all single failures except complete router outages. This is referred to as UR in Table III and Table IV shows its cost and performability. (This design is called Option-1 in Section III.) The cost reduction comes from eliminating roughly half of AR-BR links, all of $BR_1 - BR_2$ intra-office links, and chassis related to removed BRs. We also save on the inter-city BR-BR links because with all links concentrated on a single BR

(instead of being spread out over a pair of BRs), we get better capacity multiplexing. However this design cannot protect against any BR outage and has less than three 9's of 'no route' performability which is an unacceptable threshold in carrier grade networks.

Design	% Savings from BL	Performability	
		No route	Congestion
BL	0	0.999986	0.999965
UR	35.12	0.998957	0.999936
SR-100	30.72	0.999979	0.999978

TABLE IV
COST AND PERFORMABILITY

The last row of Table IV shows our proposed design (referred to as SR-100 in Table III) where any AR, upon failure of its local BR, homes to a remote BR. The rehoming as well as the additional capacity is computed by ILP described in Section V starting from UR. For performability evaluation, we assume that when the local BR fails, traffic originating at that AR *after a brief interruption (for 1 minute, in our experiments) then gets rehomed to the remote BR*. As we can see, rehoming adds very little to the overall cost (cost savings from BL reduce from 35.12% to 30.72%) but matches the performability of the baseline design. The reason for such a small additional cost is because by setting up these remote connections dynamically (instead of permanent connections), we are exploiting statistical multiplexing in use of transport resources. The minor improvement in congestion performability in SR-100 over BL is incidental (some of the capacity we added for remote homing happened to help with congestion in multiple failures).

A. Designing for restoration of priority traffic only

In networks supporting class of service (CoS), priority and best effort traffic has different SLAs. We consider network designs where we provide restoration capacity for priority traffic only. Notice that just because we do not consider best effort traffic in our restoration design, it does not mean that all best effort traffic gets dropped upon a failure. E.g., say link L, upon failure F_1, needs 10 units of additional capacity for priority traffic and, upon a different failure F_2, needs 20 units of additional capacity for priority traffic. Further, suppose that we add $\max(10, 20) = 20$ units of additional capacity on link L. Upon failure F_2 indeed all the additional capacity will be taken by priority traffic and all affected best-effort traffic will be dropped. However upon failure F_1, priority traffic only needs 10 units of capacity and the remaining 10 units will be used to restore best-effort traffic.

Table V lists the performability of designs assuming (respectively) 75%, 50%, and 25% of the traffic is classified as priority. The first two rows repeat the results from Table IV and (because they do not consider CoS) have the same performability for all traffic. For designs, SR-75,SR-50, and SR-25, we first size their link capacities so that all *priority* traffic survives any single failure except complete router outage and then we find the remote BR mapping and additional capacities using the ILP in Section V.

Design	% Savings from BL	Performability			
		Priority		Best effort	
		No route	Cong	No route	Cong
BL	0	0.99998	0.99996	0.99998	0.99996
UR	35.12	0.99896	0.99994	0.99896	0.99994
SR-100	30.72	0.99998	0.99998	0.99998	0.99998
SR-75	40.59	0.99998	0.99982	0.99998	0.99617
SR-50	48.9	0.99993	0.99997	0.99997	0.99572
SR-25	55.94	0.99997	0.99998	0.99997	0.99510

TABLE V
COST AND PERFORMABILITY WITH CLASS OF SERVICE

We see substantial improvement in cost savings as the fraction of priority traffic goes down. E.g, if half of traffic is best-effort (SR-50), we are getting a savings of nearly 50% where performability of priority traffic nearly matches those in BL. The only drop is in the congestion performability of best effort traffic where the application layer may be able to deal with a small amount of lost packets. The minor differences in no route performability (in the 5th decimal place) is because our design heuristic of getting all link utilizations near 100% is somewhat coarse. As argued at the end of Section IV, with a proper network design tool, we can tweak the performability and costs of these designs.

B. Cost sensitivity with respect to router and transport equipment costs

Our proposed designs have lower transport as well as router related costs compared to the baseline but the percentage savings are lower for transport cost compared to the router related cost. The projected cost savings are dependent on unit equipment costs and if router equipment costs were to go down (compared to transport equipment costs) then our projected savings will also go down. We estimated our cost savings based on equipment prices reported in [4] but recent trend towards cheaper Ethernet based switching have pushed the router costs down so Table VI show a sensitivity analysis of our estimated cost savings. Each design has three rows. The top row lists the savings with the equipment cost reported in [4]. If the transport equipment prices go up (relative to router equipment prices), our savings will improve and we do not show them in the table. However the next two rows shows how our savings go down if router equipment became twice (router cost multiplier is 0.5) or 10 times cheaper (router cost multiplier is 0.1). We see that even in the case of a 10x reduction in router prices, our cost savings remain nearly 18% for SR-100 to nearly 50% for SR-25.

C. Cost sensitivity with respect to traffic scaling

Finally we examine how our savings vary with traffic matrix scaling by increasing traffic 10 fold. This has a major impact on the design as the increased traffic nearly eliminates the need for sub-wavelength 10G circuits. As shown in Table VII, our

Design	Router equipment cost multiplier	% Savings from BL
UR	1.0	35.12
	0.5	33.47
	0.1	28.38
SR-100	1.0	30.72
	0.5	27.58
	0.1	17.92
SR-75	1.0	40.59
	0.5	36.67
	0.1	24.58
SR-50	1.0	48.90
	0.5	46.93
	0.1	40.84
SR-25	1.0	55.94
	0.5	54.37
	0.1	49.56

TABLE VI
COST SENSITIVITY RELATIVE TO ROUTER AND TRANSPORT EQUIPMENT COSTS.

cost savings are not sensitive to traffic scaling. In fact, they improve slightly with the higher traffic.

Design	Router (unit) cost multiplier	% Savings from BL
SR-100	1.0	30.72
	0.5	27.58
	0.1	17.92
SR-100 with 10x original traffic	1.0	34.94
	0.5	33.1
	0.1	27.09

TABLE VII
COST SENSITIVITY RELATIVE TO TRAFFIC SCALING.

VII. CONCLUSION

Network service providers continue to see increased pressures to reduce the cost of their IP backbones. A significant cost is incurred by the core backbone routers, and the redundancy of dual routers at each point of presence (POP). The increasing reliability for the core IP routers enables ISPs to exploit an elegant design that leverage the strengths of an increasingly agile optical transport to avoid the high cost of redundant core routers while achieving the same level of availability and performance. However, operational aspects in a network still impact router availability, especially with the inability to seamlessly upgrade the hardware and software of these routers.

Our design carefully ensures that connectivity is maintained upon single failures, including that of a complete core router, and also seeks to avoid congestion and packet loss under such failure conditions. We proposed an architecture that dynamically sizes the capacity of the links between the access-routers and a remote backbone router. We achieve almost the same level of performability as the baseline dual router design, while achieving a cost savings of approximately 30%.

We recognize the current trend among ISPs to provide higher availability to certain classes of traffic (e.g., VPN traffic), rather than all the traffic flowing over their backbone. When protection and restoration is provided only to high priority traffic, we see a substantial cost reduction.

We also recognize that almost all cost based design decisions are highly dependent on the unit costs of routers and optical network components at any given time. To understand this, based on the near term trends of which components are experiencing cost reductions as technology evolves, we evaluate the sensitivity of our design to the relative costs of the different components. We examine a range of reductions in the cost of backbone routers (all the way down to 10% of current costs) and show that we are still able to achieve worthwhile cost reductions while achieving acceptable performability. Finally, our results are robust to the traffic matrix scaling up. Our results for the cost reduction for the IP backbone makes a compelling case for our architecture.

Our overall approach should point to a new trend in how backbone networks are architected, achieving a suitable trade-off between cost and reliability while at the same time ensuring that fast restoration is achieved when a backbone router fails.

ACKNOWLEDGMENT

Byrav Ramamurthy was supported by an NSF FIA grant (CNS-1040765) and the AT&T VURI program. This work has greatly benefited from our discussions with Jorge Pastor, Martin Birk, Bob Doverspike, Guangzhi Li, Pete Magill and Kostas Oikonomou (all at AT&T Labs-Research).

REFERENCES

[1] S.R. Bailey, V. Gopalakrishnan, E. Mavrogiorgis, J. Pastor, and J. Yates. Seamless access router upgrades through IP/Optical integration. In *NFOEC*, 2011.

[2] A. L. Chiu, G. Choudhury, M. D. Feuer, J. L. Strand, and S. L. Woodward. Integrated restoration for next-generation IP-over-Optical networks. *J. of Lightwave Technology*, 29(6):916–924, 2011.

[3] M. Goyal, K.K. Ramakrishnan, and W. Feng. Achieving faster failure detection in OSPF networks. In *IEEE ICC*, 2003.

[4] R. Huelsermann, M. Gunkel, C. Meusburger, and D. A. Schupke. Cost modeling and evaluation of capital expenditures in optical multilayer networks. *J. of Optical Networking*, 7(9):814–833, 2008.

[5] A. Mahimkar, A. Chiu, R.D. Doverspike, M. Feuer, P. Magill, and Mavrogiorgis E. Bandwidth on demand for inter-data center communication. In *ACM Workshop on Hot Topics in Networks (HotNets-X)*, 2011.

[6] A. Mahimkar, H. Song, Z. Ge, A. Shaikh, J. Wang, J. Yates, Y. Zhang, and J. Emmons. Detecting the performance impact of upgrades in large operational networks. In *SIGCOMM*, 2010.

[7] K.N. Oikonomou, R.K. Sinha, and R.D. Doverspike. Multi-layer network performance and reliability analysis. *International J. of Interdisciplinary Telecommunications and Networking*, 1(3):1–30, 2009.

[8] E. Palkopoulou. *Homing Architectures in Multi-Layer Networks: Cost Optimization and Performance Analysis*. PhD thesis, TU Chemnitz, Germany, 2012.

[9] E. Palkopoulou, D. Schupke, and T. Bauschert. Quantifying CAPEX savings of homing architectures enabled by future optical network equipment. *Telecommunication Systems*, pages 1–7, August 2011.

[10] S. Phillips, N. Reingold, and R.D. Doverspike. Network studies in IP/Optical layer restoration. In *OFC*, 2002.

[11] B. Ramamurthy, K. K. Ramakrishnan, and R. K. Sinha. Cost and reliability considerations in designing the next-generation IP over WDM backbone networks. In *Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN 2011)*, pages 1–6, August 2011.