2011

Cost and Reliability Considerations in Designing the Next-Generation IP over WDM Backbone Networks

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

K. K. Ramakrishnan
AT&T Labs - Research, kkrama@research.att.com

Rakesh K. Sinha
AT&T Labs - Research, sinha@research.att.com

2011 Proceedings of 20th International Conference onComputer Communications and Networks (ICCCN)
Digital Object Identifier: 10.1109/ICCCN.2011.6006102

Cost and Reliability Considerations in Designing the Next-Generation IP over WDM Backbone Networks

(Invited Paper)

Byrav Ramamurthy*, K. K. Ramakrishnan†, Rakesh K. Sinha†
* University of Nebraska-Lincoln, Lincoln, NE, USA
† AT&T Labs - Research, New Jersey, USA
Email: byrav@cse.unl.edu,{kkrama,sinha}@research.att.com

***Abstract*—To accommodate the increasing demands for bandwidth, Internet Service Providers (ISPs) have deployed higher-speed links and reconfigurable optical add drop multiplexers (ROADMs) in their backbone networks. To address the reliability challenges due to failures and planned outages, ISPs typically use two backbone routers at each central office in a dual-home configuration. Thus at the IP layer, redundant backbone routers as well as redundant transport equipment to interconnect them are deployed to provide 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 which avoids such redundant resources by leveraging the capabilities of an intelligent optical transport network is a highly desirable objective. It is clear that such a redesign must lower costs without compromising on the reliability achieved by today's backbone networks. Modeling the costs and reliability of the network at all layers is an important step in achieving this objective. In this paper, we undertake an in-depth investigation of the cost and reliability considerations involved in designing the next-generation backbone network. Our work includes a detailed analysis of the operation, cost and reliability of the network at the IP layer and the multiple layers below it. We discuss alternative backbone network designs which use only a single router at each central office but use the optical transport layer to carry traffic to routers at other offices in order to survive failures or outages of the single local router. We discuss trade-offs involved in using these designs.**

I. Introduction

The demand for increased network capacity has been unrelenting. Growth in video over the network and increase in data use continue to contribute to the growth in demand. At the same time, society (businesses and consumers) expects a level of reliability of their IP service comparable to that provided by other utilities (electricity, etc.). Thus, higher bandwidths (and total aggregate network capacity) and reliability are critical requirements for the backbones of large IP service providers. The increase in capacity and reliability requirements have to be considered in the context of the enormous pressures for keeping costs down as margins for Internet service providers (ISPs) keep shrinking.

Reliability for IP networks is typically provided by redundancy. For example, when a link fails, the restoration mechanism must be able to find alternate routes for all affected flows. Outages and failures of routers are also handled by having redundant routers. However adding redundant reosurces also adds to the capital and operational cost of the network.

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. Furthermore, there are a variety of additional equipment and complex functionality in the ISP's backbone, beyond the routers and their line cards. However, cost reductions by simplifying the topology and reducing equipment costs 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.

In this paper, we undertake an in-depth investigation of the cost and reliability considerations involved in designing the next-generation backbone network. Section II includes a detailed description of the operation of the network at the IP layer and the multiple "optical" layers below it. In Section III, we propose alternative backbone network designs which use only a single router at each central office but use the optical transport layer to carry traffic to routers at other offices in order to survive failures or outages of the single local router. We provide a detailed analysis of the cost and reliability considerations involved and discuss trade-offs involved in using these designs. We discuss related work in Section IV and conclude the paper in Section V.

II. Background

A. Current state of backbone networks

The backbone network of a typical ISP can be quite complex, comprising multiple layers. Customer equipment homes on access routers (AR, also called a Provider Edge, PE, router), which in turn are connected to the core backbone routers (BR, also called Provider, P, router). An AR is often co-located with its BRs at a 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

978-1-4577-0638-7 /11/$26.00 ©2011 IEEE

use underlying ROADM network. While this type of redundant backbone router configuration in a PoP is typical of large ISPs, it can be expensive. If, for example, we wanted to reduce costs by keeping only one BR in a PoP, then each (non-remote) AR will be homed to one (co-located) BR. The resulting architecture will have unacceptable availability because any customers homed on this AR will lose their connectivity when this BR goes down. To overcome this, we further connect this AR to a second BR in a different PoP. This results in a different cost structure and a different network availability. Estimating the cost for this approach has to consider the reduction in the redundant router at ech PoP as well as the additional expense due to the need to transport the second AR to BR link. To understand this tradeoff requires a more careful understanding of the remaining lower layers below IP in terms of both cost and reliability impact.

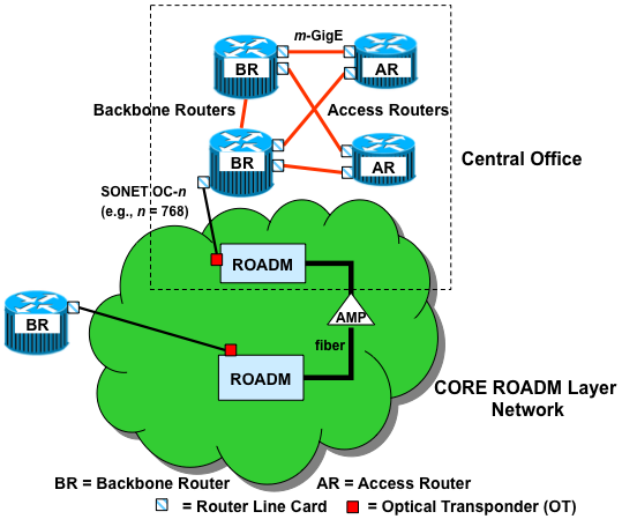


Fig. 1. Legacy backbone network.

B. Motivation for our work

Even though routers have become very reliable and (unplanned) complete router failures are rare events, routers experience frequent outages because of planned software and hardware upgrades. A few router vendors support in-service software upgrades but as argued in [2], there is still a large base of deployed routers without such capability. Approaches for providing limited redundancy through Hot Standby Router Protocol (HSRP) etc. tend to be very 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.

When we attempt to provide resiliency to failures at the IP layer, we encounter the need to add a redundant link to the topology. The addition of such a 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 the network, even though the components at Layer 3 (router, its line cards) are often a greater fraction of the cost, 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 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 somewhat 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]. So, 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 excessively 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.

C. IP layer

The traditional way of providing reliability in the IP network is to provide redundancy and duplication, 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 [4]. 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 sources feeding into each BR can be carried by the single BR if the other 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 other BRs in 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 the 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.

D. Optical transport layer

Inter-office links connecting backbone routers establish the Layer 3 adjacencies used by protocols such as OSPF. In reality, a single inter-office link is a logical (or aggregate) link comprising of possibly, multiple physical links (such as SONET *circuits*) with perhaps, different capacities (e.g. OC-768 and OC-192). In Fig. 2, for example, three OC-192 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.

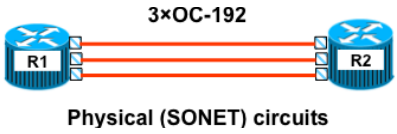


Fig. 2. Physical links that make up an aggregate L3 link.

Each physical link occupies either a complete wavelength (e.g. OC-768 circuit in a 40Gbps wavelength system) or a sub-wavelength (e.g. OC-192 circuit in a 40Gbps wavelength system). To carry such a physical circuit (either OC-768 or OC-192), an *optical transponder* (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. Depending on the capacity of the circuit that needs to be carried over a wavelength, the type of transponder is chosen (e.g. OC-768 transponder or OC-192 transponder). 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 functions of a multiplexer and transponder. With a muxponder at each end of a 40Gbps wavelength path, upto 4 OC-192 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 4 OC-192 transponders tend to be higher than the cost of a 4 x OC-192 muxponder. So, in practice, multiple OC-192s are carried over a single wavelength using muxponder and OC-768 transport equipment. However in rare cases, where we have only a single OC-192 circuit (and no anticipated capacity growth), it may be cheaper to use OC-192 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 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 drop and add 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). A regen can be constructed by using two transponders placed back-to-back, but often it can be constructed in a simpler manner and at lower cost.

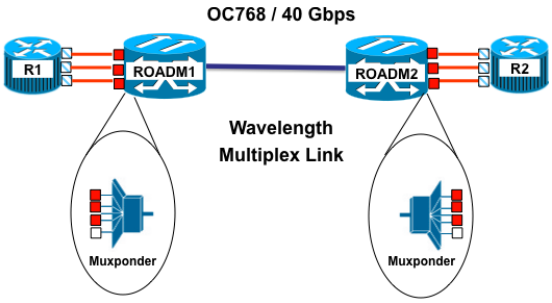


Fig. 3. Routing of the physical links over a multiplex link.

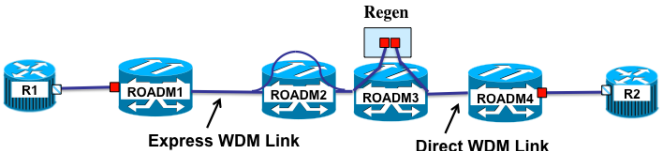


Fig. 4. An Express link can bypass ROADM nodes. The express link from ROADM 1 to ROADM 3 bypasses ROADM 2 (without any regeneration at the intermediate node).

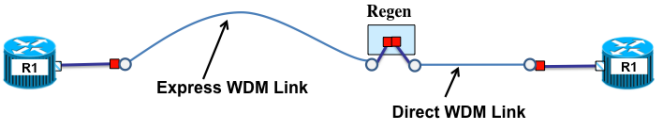


Fig. 5. A physical link can span multiple Direct WDM links and Express WDM links.

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. OC-192 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. 4xOC-192 circuits over a 40 Gbps wavelength).

III. Alternative Designs for Next-Generation Backbone Networks

We propose an IP backbone design that takes advantage of the increasing flexibility and agility being provided by the underlying lower layer technologies such as Layer 2 switching and Layer 1 optical networks to build a lower cost IP backbone that includes only a single BR at each office, while still being robust against a single router failure. Our intent is to lower cost while keeping availability at acceptable levels, by carrying the traffic to a remote BR instead of having a redundant local BR.

There are several different options for connecting to a remote BR that we are currently investigating. We summarize them below.

1) Each AR is connected to a local BR and a remote BR with full capacity link. We save on the cost of a BR but we pay more in terms of transportation cost and (because the link to remote BR is less reliable than the link to local BR) decreased performability.
2) Each AR is connected to a local BR only. When failure happens, we dynamically connect this AR to a pre-determined remote BR. This works well for scheduled maintenance events but for unscheduled failures, setting up a link and protocol convergence time (with conservative values for the timers preferred by operators) may take several 10s of seconds, resulting in decreased performability. The main advantage over the previous option is cost reduction due to multiplexing. A reasonable network design goal is to protect against a *single* BR failure. When a BR fails, all ARs homed to the failed BR need to be connected to a remote BR. However ARs homed to other (non-failed) BRs do *not* need a link to a remote BRs. The overall effect is that unlike in the previous option where all AR-remote BR links are present, we only need a subset of AR-remote BR links at any given time and thus the resulting design requires less resources.
3) Each AR is connected to a local BR at full rate and to a remote BR at a low rate. The link to remote BR has just enough capacity to maintain protocol state (e.g, keep-alive messages) so that we do not lose time in setting up a new link. When a failure happens, we dynamically resize this link to full rate. This combines the benefits of the previous two options: the cost of the second option and restoration of the first option but requires necessary protocol design that we are working on currently.

A. Baseline network modeling

As a first step in the process, we evaluate the cost and reliability of a baseline model of a Tier-1 ISP backbone network. This backbone has a dual-router architecture and all logical links are aggregate links, consisting of a mixture of OC-192s and OC-768s. We use it as a baseline and then check how our suggested architectural changes affect the cost and availability. However, because of the long ordering cycles for additional capacity, production networks always have excess capacity for anticipated traffic growth. To have a fair baseline, we start with a network of appropriate cost and capacity to make our cost savings realistic. So, we made the following changes to production network capacity. We simulated all single failures, including routers, router line-cards, router ports, transponders, regens, ROADMs, optical amplifiers, and fiber cuts. For each failure, we simulated the ciruits that go down and how those affected flows get rerouted by OSPF. For each logical link, we obtained the highest utilization across all failures. If this utilization was more than 100%, we reduced the number of circuits. However removal of circuits also resulted in removal of corresponding network elements – regenerators, OTs, router ports – thus changing the set of single failures and therefore 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.

B. Router cost

A router itself has several components.

1) The intra-office links between the two routers in the same office are 10G so each link contributes two 10G ports.
2) Each inter-office link contributes two OC-192 or OC-768 ports.
3) 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 hierachical pattern for multiplexing low-rate connections so as not to exhaust ports on BRs. We considered a simplified model and assumed that all access facing ports on BRs are 10G and estimated their numbers from a demand matrix and by assuming an average utilization of 60%.
4) Where only the costs for line cards, but not of individual ports, are known, we translated the number of ports into the number of line cards, with the added twist that the GigE cards can be over-provisioned.
5) We derive the number of chassis based on the number of line cards.

C. Transport layer cost

At the transport layer, there are common costs which we compute based on the number of miles traversed by each circuit. The common cost includes the cost of all pre-deployed components such as fiber, amplifiers, ROADMs and other equipments in the network. Apart from the common cost, the total costs include the costs incurred by each circuit (OC-768 or OC-192) carried over the network.

For an OC-768 circuit, the cost consists of the OC-768 transponders and OC-768 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. 6, a new OC-768 circuit is carried over 2 Express WDM links (curved lines) followed by a Direct WDM link (straight line). Hence the cost for the circuit is 2 x (Cost of OC-768 transponder) + 2 x (Cost of OC-768 regen).

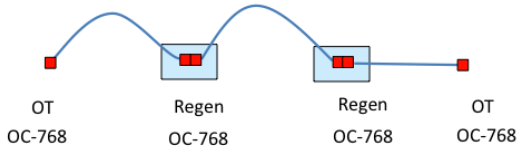
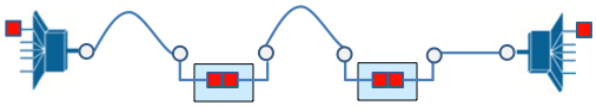


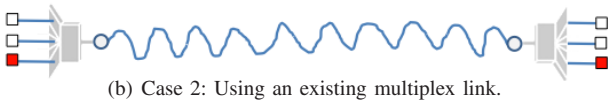
Fig. 6. Cost of a OC-768 circuit.

For an OC-192 circuit, the cost computation is a bit more complicated. This is due to the fact that an OC-192 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 OC-768 WDM link ensures that additional sub-wavelength circuits do not incur this cost again. Note that both OC-192 regenerators and OC-768 regenerators may appear in the end-to-end path carrying an OC-192 circuit.

Thus the cost for a new OC-192 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. Hence, we identify four different options for carrying an OC-192 circuit across the transport network. In Case 1 (see Fig. 7(a)), the new OC-192 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 OC-192 circuit. The total cost for carrying the OC-192 circuit is 2 x (Cost of OC-192 transponder) + 2 x (Cost of OC-768 regen) + 2 x (Cost of muxponder). Here for comparing different options, we ignore the common cost (ROADMS, fiber, etc.). Note that three additional OC-192 circuits may be carried over this end-to-end multiplex link in the future thanks to the muxponders. In Case 2 (see Fig. 7(b)), the new OC-192 circuit is carried over such an existing multiplex link. The total cost for carrying the OC-192 circuit is 2 x (Cost of OC-192 transponder). In Case 3 (see Fig. 7(c)), the new OC-192 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 OC-192 circuit (similar to Case 1). The total cost for carrying the OC-192 circuit is 2 x (Cost of OC-192 transponder) + 2 x (Cost of OC-768 regen) + 2 x (Cost of muxponder) + (Cost of OC-192 regen). The additional cost incurred here (compared to Case 1) is due to a OC-192 regen which is required to transport the OC-192 circuit across the old and the new multiplex links. Finally, in Case 4 (see Fig. 7(d)), the new OC-192 circuit is carried over two existing multiplex links. The total cost for carrying the OC-192 circuit is 2 x (Cost of OC-192 transponder) + (Cost of OC-192 regen).

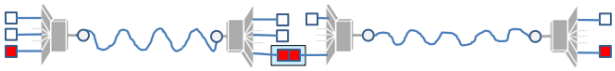


(a) Case 1: Using a new multiplex link routed over two Express WDM links and a Direct WDM link.



(b) Case 2: Using an existing multiplex link.

(c) Case 3: Using an existing multiplex link and by creating a new multiplex link.



(d) Case 4: Using a sequence of two existing multiplex links.

Fig. 7. Four options for carrying an OC-192 circuit. New components are shown in dark shaded portions. Straight lines represent Direct WDM links. Curved lines represent Express WDM links. Wavy lines represent multiplex links.

D. Cost breakdown for the baseline network

For the baseline network, the cost breakdown into individual components is given in Table I. The table does not include the cost of the access network. As mentioned previously, the access topology tends to be extremely complicated. There can be many aggregating switches or routers between the customer router and the BR. However our suggsted changes only affect the last link to BR (which we capture by counting the number of access facing ports on BRs) and not the rest of the access topology.

As mentioned earlier, the router port costs dominate the overall Layer 3 cost. For our baseline, the router chassis contributes 13.9% of the cost but this percentage will go down as traffic grows and the chassis gets fully utilized.

Equipment	Percentage of network cost
Router chassis	13.9
Router ports	44.0
Transport	42.1

TABLE I
COST BREAKDOWN DUE TO VARIOUS EQUIPMENT CLASSES.

E. Network availability

The failure impact of commonly used components is as follows.

- A transponder, regenerator, or a router port fails a single circuit.
- An amplifier or a fiber cut fails multiple circuits. The fiber cut probability is roughly proportional to its length.
- ROADMs are engineered so that a complete ROADM almost never fails but only a part of the ROADM carrying links in a given direction fails.

- A complete router or a router line card can go down either because of scheduled maintenance or due to unplanned failure. Scheduled maintenance typically is performed during off-peak hours and great care is taken to avoid any service impact. For example, in a dual router architecture, the assumption is that while a BR is being upgraded, any AR homed to this BR still has reachability to the rest of the network using a second BR it is homed to. If that second BR were to go down before the scheduled maintenance time of this BR, the maintenance activity will be postponed. If the second BR failure were to occur during the maintenance of the first BR, the maintenance event is halted.

A network responds differently to different types of failures. If the failure happens in Layer 2, one of two things happens. If the network has restoration available at this layer, the IP layer may not even find out about this failure. This is achieved by configuring the timers in such a way that the lower layer completes its restoration within L3's 'detection windows.' If the network does not offer L2 restoration or if the L2 restoration is not successful, then the failure information is propagated to Layer 3. The L3 response depends on the type of routing and restoration protocol used. For example, if only a subset of circuits in a link aggregate fails, then the (logical) link's capacity changes but it remains up. A protocol such as OSPF that is insensitive to link capacities will not reroute any flows in this case. On the other hand, a different protocol may conclude that there is not enough capacity for all the flows and may try to reroute some of them.

We evaluate a network's performability (a metric combining performance and reliability) using the *nperf* tool [7]. A network state is defined by a set of failed components. The tool assigns a probability to each state based on the MTTR (mean time to repair) and MTBF (mean time between failures) of components. It also simulates the failure and the restoration and estimates losses due to (a) the restoration protocol being unable to find a route for flows (b) the restoration protocol rerouting flows over links with insufficient capacity, and (c) the disconnects during the restoration protocol convergence. By multiplying the losses in any network state with its probability and then summing up over all states, we get a measure that is averaged over the space of time and failures. For example, an expected loss of 0.00001 means that over a very long duration, an average of 0.00001 fraction of traffic will be lost or alternatively the network has five 9's of reliability ($1 - 0.00001 = 0.99999$). We are currently investigating the relative performability of various options described in Section III.

IV. Related work

Palkopoulou *et al.* [9] 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 common pool of shared restoration resources. The cost estimates are done on two reference networks, a 17 node German network and a 14 node US network, and by assuming uniformly distributed traffic. Huelsermann *et al.* [5] provide a detailed cost model for multi-layer optical networks. Chiu *et al.* [3] 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 the other (surviving) BR in this office using the optical layer. They achieve a savings in BR ports by a clever reuse: when a BR fails, all the ports on the surviving BR in this office used for intra-office links are unused and can be reused for the new aforementioned inter-office links. Instead of removing one BR from a dual BR design, the study by Oikonomou *et al.* [8] focuses on reducing port count from both BRs. The key observation is that planned router maintenance happens far more frequently than (unplanned) router failures and planned maintenance is scheduled when the traffic load is a fraction of the peak traffic. Finally, the RouterFarm architecture proposed by Agrawal *et al.* [1] minimizes impact of router upgrades by rehoming the customers to an alternate router during the maintenance window. This is done by reconfiguring optical layer and copying customer router configuration to the alternate router.

V. Conclusion

Increasing costs in operating today's ISP backbone networks have resulted in carriers looking for alternative designs which leverage the strengths of an increasingly agile optical transport layer. Improvements achievable by simplifying the Layer 3 (router) topology and reducing equipment costs at this layer have to be carefully evaluated, ensuring a proper tradeoff between cost and reliability. The modeling of cost and reliability of a baseline network topology described in this paper is an important first step towards evaluating the feasibility of alternative designs for the backbone network, a goal for our future work.

References

[1] M. Agrawal, S.R. Bailey, A. Greenberg, J. Pastor, P. Sebos, S. Seshan, K. van der Merwe, and J. Yates. RouterFarm: Towards a dynamic, manageable network edge. In *INM*, 2006.

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

[3] 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.

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

[5] 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.

[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] K.N. Oikonomou, R.K. Sinha, and R.D. Doverspike. A network design technique for selective restoration. In *OFC*, 2011.

[9] E. Palkopoulou, D.A. Schupke, and T. Bauschert. Qualitative evaluation of homing architectures in multi-layer networks and availability analysis. In *ONDM*, 2009.

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