Run, Walk, Crawl: Towards Dynamic Link Capacities

Rachee Singh
rachee@cs.umass.edu
University of Massachusetts, Amherst

Monia Ghobadi
mgh@microsoft.com
Microsoft Research

Klaus-Tycho Foerster
ktfoerster@cs.aau.dk
Aalborg University

Mark Filer
mafiler@microsoft.com
Microsoft

Phillipa Gill
phillipa@cs.umass.edu
University of Massachusetts, Amherst

ABSTRACT

Fiber optic cables are the workhorses of today's Internet services. Operators spend millions of dollars to purchase, lease and maintain their optical backbone, making the efficiency of fiber essential to their business. In this work, we make a case for adapting the capacity of optical links based on their signal-to-noise ratio (SNR). We show two immediate benefits of this by analyzing the SNR of over 2000 links in an optical backbone over a period of 2.5 years. First, the capacity of 80% of IP links can be augmented by 75% or more, leading to an overall capacity gain of 145 Tbps in a large optical backbone in North America. Second, at least 25% of link failures are caused by SNR degradation, not complete loss-of-light, highlighting the opportunity to replace link failures by link flaps wherein the capacity is adjusted according to the new SNR. Given these benefits, we identify the disconnect between current optical and networking infrastructure which hinders the deployment of dynamic capacity links in wide area networks (WANs). To bridge this gap, we propose a graph abstraction that enables existing traffic engineering algorithms to benefit from dynamic link capacities. We evaluate the feasibility of dynamic link capacities using a small testbed and simulate the throughput gains from deploying our approach.

1 THE CASE FOR DYNAMIC WANS

Optical backbones are billion dollar assets, with fiber comprising their most expensive component. Today, optical backbones are used for all wide-area communications with service providers such as Google, Microsoft and Facebook, purchasing or leasing fiber from Tier 1 providers to operate their optical backbones. To cite an example, Level3 Communications [4] is one of the major Tier 1 providers with over 200,000 miles of fiber worth over $30 billion dollars [2, 3].

The research community is actively working on strategies to introduce programmability at higher layers of the networking stack. For instance, distributed traffic engineering in WANs has been replaced with programmable centralized controllers [17, 19], commodity hardware load-balancers have been replaced with software counterparts [10, 23], and switch management APIs have been replaced with fully programmable stacks [1, 7, 22, 25, 26]. With each change, providers have been able to improve the performance and lower the cost-of-ownership by introducing programmability into different layers of the networking infrastructure.

However, the physical layer is generally considered *static* and despite many experimental indications of the benefits of a programmable optical layer [8, 9, 11, 13, 16, 18, 20, 21, 24], it has not yet seen the light of the day. This impasse stems from the gap between optical component and network system designers. Building an operational network with a programmable physical layer requires operators to understand its impact on the IP layer and network management. On the other hand, designing a programmable optical device requires an understanding of the system requirements of operational networks.

This work highlights the need for dynamic link capacities in the WAN, and works to bridge the gap between network operators and hardware designers, taking a step towards the realization of a programmable optical layer. However, moving the needle in this space is tough. Providers need to quantitatively understand the impact of enabling dynamic capacities in the physical layer on the overall throughput and reliability of their networks. By studying the SNR of over 2,000 WAN links for 2.5 years we show that not only the throughput (as reported by prior work [20]) but also the availability of today's networks can be improved by adding programmability to the physical layer. Present day networks enforce a *binary* status (up/down) to fixed capacity links based on an SNR threshold. A dip in the SNR below the threshold results in the link being declared down when it may still be usable, but at a lower capacity. We find evidence of this by analyzing seven months of link failure tickets reported by WAN field operators. This means by enabling dynamic link capacities,

HotNets-XVI, November 30-December 1, 2017, Palo Alto, CA, USA

ACM ISBN 978-1-4503-5569-8/17/11…$15.00
https://doi.org/10.1145/3152434.3152451

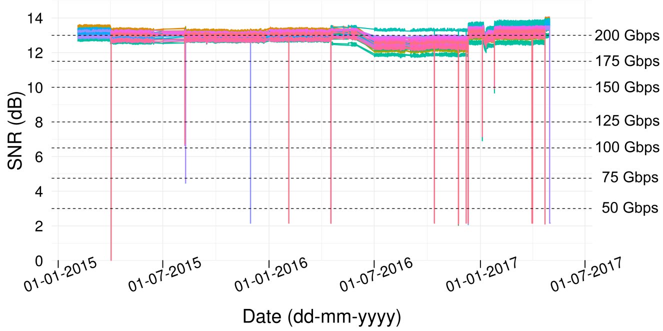


Figure 1: SNR changes of 40 optical wavelengths (i.e., IP links) on a wide area fiber cable. Dotted lines represent the feasible link capacity at and above a particular SNR.

not only can we drive the links faster, but also improve availability by driving them slower instead of failing them.

To ease the impact of dynamic link capacities on higher layers of the networking stack, we propose an abstraction layer that enables existing IP traffic engineering (TE) schemes to adopt dynamic capacities without significant modifications. In doing so, we identify a problem in the design of Bandwidth Variable Transceivers (BVTs) — optical devices that enable dynamic capacity links. We discover that they are unnecessarily conservative with respect to the time taken to change a link's capacity. We propose a solution for reducing the reconfiguration time and evaluate it with a small testbed implementation. We hope that the benefits of easily adopting dynamic capacity links in WANs will motivate networking researchers and practitioners to direct their efforts towards the realization of a programmable physical layer.

2 QUANTIFYING THE OPPORTUNITY

In this section we quantify the benefits of designing networks with dynamic link capacities. In principle, our argument is based on a simple observation: SNR is a continuous variable, so should be the modulation (*i.e.,* capacity) of its corresponding signal. Below we summarize our main observations.

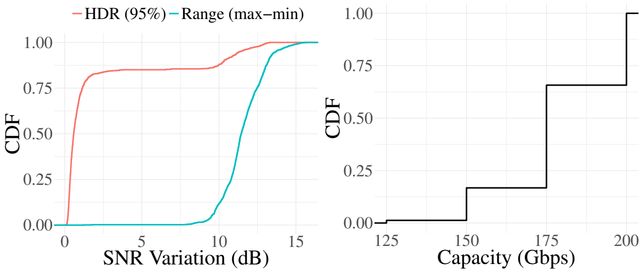
2.1 Improving Link Throughput

We study the SNR of over 2000 links[1] in a large company's WAN every fifteen minutes for a period of 2.5 years (Feb. 2015 – Jul. 2017) where each link has a fixed capacity of 100 Gbps. Figure 1 shows the distribution of SNR over time of 40 links that traverse the same WAN fiber cable. Each of these links corresponds to an optical wavelength. We observe that the SNR of these links is mostly stable but has occasional dips suggesting impairments in fiber or related optical hardware. Also, note that the SNR required for a fixed capacity link to carry 100 Gbps of traffic is 6.5 dB, however all optical wavelengths in this fiber have a much higher SNR.

Figure 2a (in blue) shows the distribution of the range (maximum - minimum) of all links' SNR over the period of our analysis. We find that the SNR range is wide, with the average SNR nearly 12 dB, suggesting that links observe dramatic drops in SNR over time. These drops have led operators to provision large margins between the actual SNR and the capacity threshold. To clarify the nature of variations in SNR, we compute the highest density region (HDR) of the SNR observed during this study (*i.e.,* the smallest interval in which 95% or more of the SNR values are concentrated) for each link in the WAN. In Figure 2a (red line), we plot the CDF of the width of the HDR of all links and find that HDR is less than 2 dB for 83% of them. This shows the dramatic changes in SNR are rare, and the SNR of WAN links is mostly stable. Based on this observation, one might be tempted to simply operate links closer to the actual SNR while still configuring the capacities statically. To quantify the gains in capacity from doing this, we compute the required SNR threshold of different modulation schemes available in our hardware: 100 Gbps, 125 Gbps, 150 Gbps, 175 Gbps, and 200 Gbps (shown as dashed horizontal lines in Figure 1). The computed thresholds are specific to our hardware, fiber length, fiber type, and wavelength. We calculate the feasible capacity for each link based on the lower SNR limit of its highest density region. We find a potential increase of 145 Tbps capacity can be had if we encode each link closer to its actual SNR. Figure 2b illustrates the CDF of the number of links which can be modulated with different capacities. The figure shows that the feasible capacity of 80% of our links is 175 Gbps or higher. This implies a possible gain of 75-100 Gbps over the default 100 Gbps capacity per link.

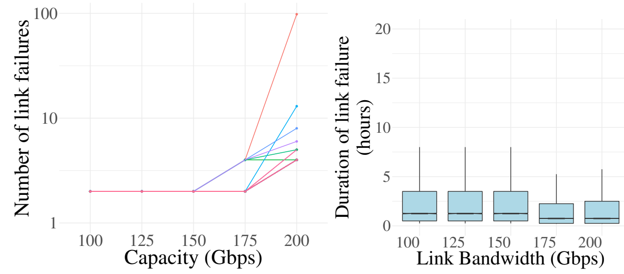
However, we find that the frequency of link failures increases if operators push the SNR threshold higher without enabling dynamic capacity links that adapt to the SNR. For example, we select a high quality WAN fiber where each link on this fiber has a high enough SNR to make all capacity denominations feasible over the period of 2.5 years. We then analyze the number of failures that these links would undergo if they were modulated with different capacities. Figure 3a shows that links on this fiber do not see any significant increase in the number of failures as the capacity is increased upto 175 Gbps. But some of the links would observe a large number of failures if we were to drive them at 200 Gbps. We characterize the duration of these link failures to evaluate the magnitude of disruption they could cause. Figure 3b plots the duration of link failures if all WAN links were modulated with different capacities (only if the capacity is feasible as per the link's SNR). We observe that such failure events last for several hours, making it unacceptable for operators to create more failures (as seen in Figure 3a) by tightening the SNR gap without having the ability to adapt the capacity if

[1]Without loss of generality, in this work, we assume a one-to-one mapping between an optical wavelength and an IP link and we use the word *link* to refer to an optical wavelength.



(a) (b)

Figure 2: (a) shows the CDF of two metrics of optical SNR variation: the size of high density region (95%) and the range of SNR. Observe that SNR stays within a narrow band of less than 2 dB, 83% of the time. But the range of SNR is much larger, suggesting dramatic but infrequent changes. (b) shows the capacities of WAN links if they were to be utilized according to their signal quality. Over 80% of links can gain 75 Gbps or more capacity over their existing static configuration of 100 Gbps.



(a) (b)

Figure 3: (a) the number of link failures for 40 different links (one color per link) for a given capacity. While increasing capacity up to 175 Gbps does not increase link failure events, achieving 200 Gbps capacity comes at the cost of increased link failures. (b) shows the duration of link failures if WAN links operate at a given capacity. Link failures for all capacities last for several hours on average.

SNR drops. We conclude that dynamic capacity links can enable operators to extract 75-100% capacity gains from WAN links without sacrificing link availability.

2.2 Improving Link Availability

Today, when the SNR of a link's optical signal drops below its pre-determined modulation threshold, the link is declared down. SNR drops may be caused by planned maintenance work (e.g., a line card replacement) or unplanned events (e.g., fiber cut, hardware failure, power failure). While some of these impairments make the link unusable (e.g. fiber cuts), others may simply lower the signal quality (e.g. failure of an amplifier) without completely shutting off the signal. Links undergoing failures due to lowered signal quality can still be used to send traffic at a reduced rate, highlighting an opportunity to improve link availability. We quantify this opportunity by manually analyzing seven months of unplanned failure tickets (250 events) reported by WAN field operators to understand their root causes. We identify three categories of root causes: (*i*) **Human:** In this category, an unplanned failure event happens while a scheduled maintenance is underway mostly due to human errors; (*ii*) **Fiber cuts:** This is an accidental break in an optical fiber; and (*iii*) **Hardware failure:** This refers to a failure of optical hardware, such as amplifiers, transponders, or optical cross connects.

We analyze the contribution of each root cause category to total outage duration (Figure 4a) and the percentage of events in each category (Figure 4b). We see that 20% of the total outage time (25% of events) can be traced to network events that occur while scheduled maintenance is underway. Similar observations have been made in a prior study [15]. We find that fiber cuts contribute to only 10% of the total outage durations (5% of events), highlighting that they are not the major cause of link failures in WANs. The remaining categories include hardware failure, human error and other undocumented failures. Failures with undocumented causes occur when technicians do not log the exact action taken, but we know they were not instances of fiber cuts. From this analysis, we conclude that over 90% of link failure events present an opportunity to harness the lowered capacity of such links. To narrow down the opportunity area identified from WAN link failure tickets, we analyze link SNR over time (as described in Section 2.1). For each link, we identify cases where the SNR falls below the 100 Gbps threshold (6.5 dB), causing the link to fail. Figure 4c shows the distribution of SNR values at link failures. We find that the lowest SNR in failure events is above 3.0 dB, nearly 25% of the time. We note that an SNR of 3.0 dB or above is sufficient to drive a link at 50 Gbps capacity. Therefore, 25% of link failures observed in our study could have been avoided by driving the impacted links at 50 Gbps, highlighting the improvement in availability offered by dynamic capacity links.

3 DEPLOYMENT CONSTRAINTS

Previous work has made the community aware that over-provisioning in the physical layer is wasteful, but modulating optical links closer to their observed quality increases the likelihood of link failures. Therefore, to gain throughput without compromising reliability of links, we have to adapt link capacities based on the SNR. Engineering a system that adapts to signal quality in fiber optics presents practical challenges to both optical hardware and WAN traffic engineering. In this section, we analyze the challenges that have kept the physical layer from being programmable so far. We propose preliminary solutions in an attempt to bridge the gap between the need for programmability in the fiber and the existing optical and networking infrastructure.

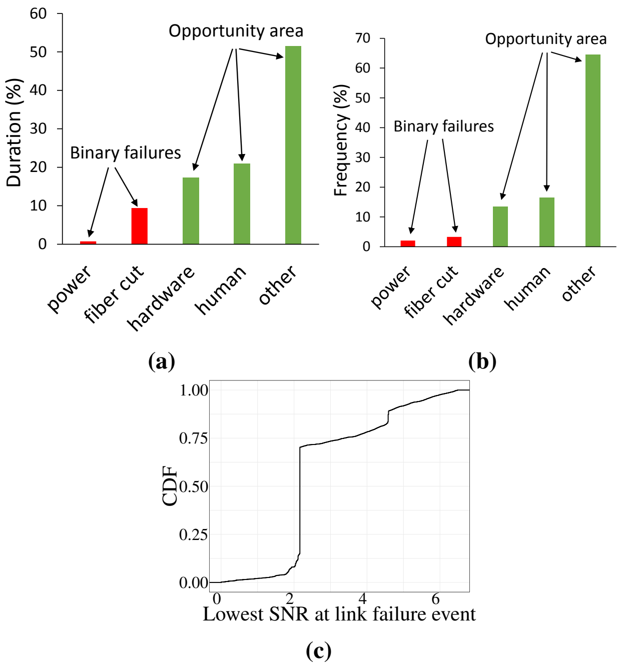


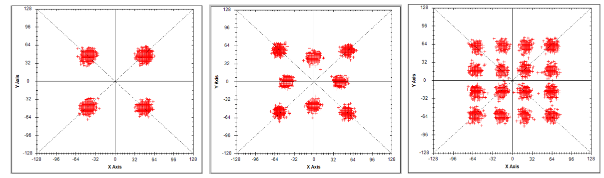
Figure 4: Categorization of major failure root causes in terms of duration of events (a) and frequency of events (b). Contrary to common belief, fiber cuts are not the major root cause of failures in WANs. Unplanned events during planned maintenance or hardware failures are more probable; they contribute to more outage duration than fiber cuts. Figure (c) shows the distribution of the SNR values at link failure events.

3.1 Towards Hitless Capacity Change

Bandwidth variable transceivers (BVTs) are not yet optimized for the latency of a modulation change. State-of-the-art BVTs can only change the link modulation after bringing it to a lower power state. This translates to a link failure for higher layer protocols. The duration of such link failures is a major challenge in the deployment of dynamic capacity links in production networks. To quantify this, we build a testbed consisting of one fiber link connected to a BVT from Acacia [5]. We program changes to the link modulation using the transceiver's MDIO interface. Figure 5 shows the constellation diagrams—i.e., the representation of a signal modulated by a digital modulation scheme—of QPSK, 8QAM and 16QAM modulations in our testbed. We change the link's modulation 200 times and analyze the time taken.

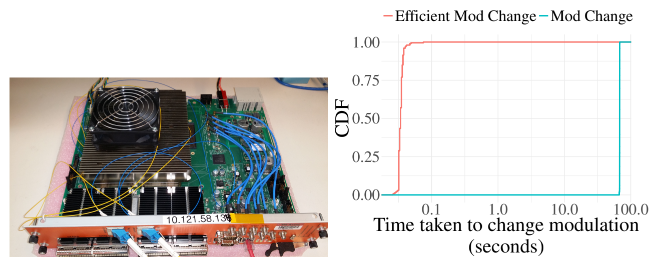
Figure 6b shows the distribution of time taken for modulation changes. We observe that the average downtime of the link undergoing capacity change is 68 seconds, a cause for concern in WANs. This also highlights that today's optical hardware caters to a very cautious network operator who over-provisions fixed capacity fiber to sidestep the high latency associated with modifying link capacities.

We investigate the cause of latency in capacity reconfiguration and find that majority of this latency is



(a) 100 Gbps **(b) 150 Gbps** **(c) 200 Gbps**

Figure 5: Small testbed evaluation of dynamic capacity adjustment of two optical links.



(a) Evaluation board **(b)**

Figure 6: (a) shows the testbed we build for evaluating the feasibility of capacity variable links. (b) is a CDF of the time taken to change modulation (capacity) of a fiber link in our testbed. Link capacity changes take 68 seconds on an average. But we demonstrate ways to change the modulation efficiently such that it takes only 35 milliseconds on average.

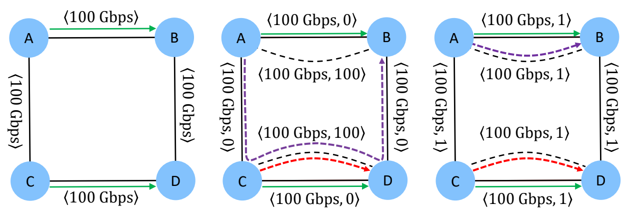
associated with turning the laser back on after reprogramming the transceiver module. We plot the distribution of time taken to change modulations without turning off the laser and find that it only takes approximately 35 ms on average. This suggests an opportunity to strive towards hitless capacity changes in the fiber.

3.2 Separating Layer1 from TE

Introducing complexity in widely deployed WAN traffic engineering controllers like those of SWAN [17] or MPLS-TE is a cause of concern for network operators. Informing TE programs of physical layer phenomenon (*e.g.*, signal quality) complicates them, making them bug prone. This is a reason for operators' reluctance to deploy dynamic capacity links in the wide area. To address this concern, we propose an abstraction that incorporates SNR to make decisions about the capacity of the underlying physical links. The abstraction then provides, as input, an augmentation of the network's topology and demands to traffic engineering algorithms, keeping the capacity variability aspect opaque. The next section provides the graph construction algorithm and correctness theorem.

4 GRAPH ABSTRACTION

In this section, we describe how we use a graph abstraction to enable existing traffic engineering (TE) algorithms such as



(a) Current flows **(b) Few increases** **(c) Short paths**

Figure 7: Intuition for graph abstraction: In 7a, we have the current network, with 100 Gbps flows (in green) between A,B and C,D, respectively. In 7b, we annotate each link with $<$ *capacity*,*cost* $>$, where the cost reflects the penalty of using the (updated) link. When both A and C want to send an additional 25 Gbps flow (in dotted red and purple), updating one link's capacity suffices. If short paths are favored, all costs are set equal, as in 7c.

SWAN [17] and B4 [19] to integrate dynamic capacity links. We argue that TE algorithms should integrate dynamic capacity links in their formulation via this abstraction to maintain the benefits of OSI layered design. As discussed in Section 3, updating a link's capacity is not a hitless operation with state-of-the-art hardware. Hence, operators ought to look for a balance between the traffic churn caused by the modification of a link's capacity and its potential benefit. We allow network operators to take advantage of programmability in the physical layer while keeping the IP layer algorithms unchanged.

We draw inspiration from the concept of *Fibbing* by Vissicchio et al. [27][2] and introduce an augmented view of the IP layer topology into the TE algorithm. This augmented view is designed such that the TE algorithm's flow assignment instructs changes in the capacity of physical links in the network topology. The next section describes this procedure in detail and shows that augmentation provides an optimal cross-layer abstraction.

4.1 Augmenting the IP Layer Topology

We modify the input topology of the TE algorithm such that at each stage of the TE recomputation, we add *fake* links corresponding to each physical link capable of increasing its current capacity. This is illustrated in Figure 7. A sample physical topology is shown in Figure 7a, with four vertices connected via bi-directional links of equal capacity. Suppose that the SNR of the links between nodes (A,B) and (C,D) is high enough to allow doubling the links' capacities. Figure 7b shows fake links (black dashed lines) added between nodes (A,B) and nodes (C,D). Each of the fake links has the same capacity as the real ones. However, the activation of a fake link is associated with a cost which is a function of the amount of traffic disrupted when the link switches to a higher bandwidth.

[2] To obtain central control over distributed route computation, a fake topology is injected into the routing protocol in [27], with the process denoted as fibbing.

Algorithm 1: Graph augmentation procedure

Input : WAN topology $G < V,E,U,P >$
Output : Augmented topology $G' < V,E',P' >$
foreach $e = (v,w) \in E$ **do**
  $P'(e) = 0$ `\\ can be adapted for other penalty functions, e.g., Fig. 7c`
  if $U[v,w] > 0$ **then**
    $E' = E' \cup \{(v,w,U[v,w],P[v,w])\}$
return $G' < V,E' \leftarrow E' \cup E,P' >$

This allows our abstraction to balance between the benefits of changing a link's capacity and the relative cost of the service interruption caused by it. The TE operators are free to set these costs to be as conservative or aggressive as they desire.

Example. Consider a network with the initial stage illustrated in Figure 7a where all links are configured with 100 Gbps capacity, and the traffic matrix is $A \rightarrow B = C \rightarrow D = 100$ Gbps. The green arrows illustrate the assignment of traffic flows. Assume that in the next round of TE computation, both traffic demands have increased to 125 Gbps, and the links (A,B) and (C,D) have the SNR to carry an extra 100 Gbps, with the cost of changing the modulation set at 100. In this new setting, the penalty-minimizing solution of the flow problem will route the additional traffic such that the capacity of only one link is increased, *e.g.,* (C,D).

Algorithm 1 formalizes the procedure of augmenting a topology to allow dynamic link capacities. The input is the topology G with vertices V, links E, and matrices U,P representing the increased capacity $U(e)$ and the cost $P(e)$ of updating each link's capacity e. The output is an augmented graph G' with the same set of vertices V, but the set of links E' is augmented with fake links and a new cost matrix P' indicating the penalty associated to each (real and fake) link.

Optimality. We show that an augmented topology is the right abstraction to incorporate dynamic capacity links into TE formulations.

THEOREM 1. *Let G be a topology consisting of links with variable capacities, with penalty function P. There is an augmented topology G' such that solving the min-cost max-flow problem on G' is equivalent to solving max-flow on G.*

The proof of Theorem 1 is not fully reported due to space constraints. Our construction relies on a three-step procedure:

(1) Augment the inputs of the TE algorithm by inserting a combination of fake network components and demands.
(2) Run the unmodified TE algorithm on the fake input.
(3) Directly translate the output of Step (2) into:
  (a) decisions about which link capacities should be modified; and
  (b) the flow-paths of the current traffic demands.

As such, the TE algorithm is unaware of the links with variable capacities, yet we obtain cross-optimization for both the

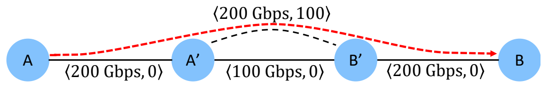


Figure 8: Intuition for an unsplittable flow of 200 Gbps.

optical and the IP layer. The key idea is "compiling" optical components into a language the TE algorithm understands.

4.2 Our Abstraction is Expressive

We suggest using the current link traffic as a penalty function, but the TE operator can set the penalty values arbitrarily: *e.g.,* factoring in the disruption duration or adjusting the penalty according to the traffic priority class. The TE operator can also steer the traffic disturbances in a more finely-grained manner: (i) if a flow is not allowed to be disturbed at all, (i-a) links on its path are not allowed to change their capacity, and (i-b) the flow along with the capacity it uses is hidden from the TE optimization; (ii) if the flow can be temporarily rerouted, but will not suffer from disruption, we can make use of the consistent network updates [17] toolkit by introducing an intermediate network state. As such, after identifying the links to be updated E_U, we remove E_U from the topology and invoke the TE controller again. More intricate solutions [12] can easily be adapted.

When balancing increasing link capacities versus latencies (hop count), link weights can be added in parallel to penalties. To enforce short paths at all costs, *e.g.,* for interactive traffic between data centers [17], we set every link weight to unit size, resulting in the solution of Figure 7c where all flows take only one hop. The network operator is free to adopt a middle-ground or to choose different weight policies for different parts of the network.

Furthermore, observe that in Figure 7b, an unsplittable flow of 200 Gbps cannot be routed from A to B, even after the link capacity is increased. We can adapt our abstraction with intermediate vertices A', B' on (A, B), ensuring that the (abstracted) link remains at a capacity of 200 Gbps, but allowing for a single path of 200 Gbps as illustrated in Figure 8.

Reductions in link capacities can be due to low SNR. Our proposed abstraction handles such events by removing the corresponding fake edges from the augmented topology. Since TE controllers will treat all edges (real and fake) in the same way, removal of a fake edge from the topology will trigger the same set of operations as a real edge removal.

5 RELATED WORK

A recent study from Jin *et al.* [20] on cross-layer optimization between IP and optical layers is similar in spirit to our motivation of bridging the gap between optical and IP layer phenomenon. In their work, Jin *et al.* show optical reconfiguration can provide latency gains for deadline driven bulk transfers. In contrast, our work focuses on the reconfiguration of the capacity of a network topology, without the migration of wavelengths across links. Moreover, we provide measurements from an operational backbone and argue for a programmable optical layer with a focus on throughput and reliability.

Govindan *et al.* [15] studied 100 failure events across two WANs and data center networks, offering insights on challenges of maintaining high levels of availability for content providers. Although they do not isolate optical layer failures, they report on root causes of failures, including optical transmitters. We complement their work by focusing on optical layer failures.

Ghobadi *et al.* studied Q-factor data from Microsoft's optical backbone [13, 14]. Our work complements their analysis by providing insights on temporal changes of SNR as well as root causes of failures. Further, we evaluate the repercussions increasing link capacities by quantify the link failures caused at static but higher capacities.

Similarly, Filer *et al.* [11] studied the deployed optical infrastructure of Microsoft's backbone and discussed the benefits of optical elasticity. They expressed a long-term goal of unifying the optical control plane with routers under a single Software Defined Network controller and recognize YANG [6] and SNMP as potential starting points for a standard data model and control interface between the optical layer and the WAN traffic controller. In this work, we explore how programmability in the optical layer can bring throughput gains or cost savings. We also provide abstractions for making optical elasticity feasible.

6 CONCLUSION

In this work, we quantify the benefits of dynamic capacity links on network throughput and reliability. We highlight the practical constraints preventing dynamic capacity links from being a reality in wide area networks. The first challenge is posed by the hardware delay in modifying a link's capacity. We analyze the cause of this delay in current optical transceivers and propose a potential solution to reduce this delay from over a minute to a few milliseconds. We address the second major challenge, stemming from operator reluctance to introduce complexity in TE controllers, by proposing a graph abstraction to incorporate dynamic capacity links in network topologies. Our graph abstraction augments the network topology in such a manner that existing TE controllers can operate unmodified, while taking advantage of dynamic capacity links in the WAN.

ACKNOWLEDGEMENTS

We would like to thank Victor Bahl, Jamie Gaudette, Jeff Cox, Liban Buni, and Kelly Becker for enabling this study. We also thank Mike Pan and Ratul Mahajan for helpful discussions. Klaus-Tycho Foerster is supported by the Danish VILLUM FONDEN Reliable Computer Networks (ReNet) project.

REFERENCES

[1] DPDK: Data plane development kit. http://dpdk.org/.

[2] Level 3. http://arstechnica.com/information-technology/2016/10/centurylink-to-buy-level-3-get-another-200000-miles-of-fiber/.

[3] Level3 and global crossing fiber map. http://www.telecompetitor.com/level-3-to-acquire-global-crossing/.

[4] The Level3 network. http://www.level3.com/~/media/files/maps/en-network-services-level-3-network-map.ashx.

[5] Acacia Communications. Acacia Communications Announces the Industry's First Coherent Flex-rate 400G 5x7 Transceiver Module. http://ir.acacia-inc.com/phoenix.zhtml?c=254242&p=irol-newsArticle&ID=2103147.

[6] M. Bjorklund. YANG - A Data Modeling Language for the Network Configuration Protocol (NETCONF). RFC 6020, Oct. 2010.

[7] P. Bosshart, D. Daly, G. Gibb, M. Izzard, N. McKeown, J. Rexford, C. Schlesinger, D. Talayco, A. Vahdat, G. Varghese, and D. Walker. P4: Programming protocol-independent packet processors. *SIGCOMM Comput. Commun. Rev.*, 44(3):87–95, July 2014.

[8] M. Channegowda, R. Nejabati, and D. Simeonidou. Software-defined optical networks technology and infrastructure: Enabling software-defined optical network operations (invited). *J. Opt. Commun. Netw.*, 5(10):A274–A282, Oct 2013.

[9] N. Cvijetic, A. Tanaka, P. N. Ji, K. Sethuraman, S. Murakami, and T. Wang. SDN and OpenFlow for dynamic flex-grid optical access and aggregation networks. *Journal of Lightwave Technology*, 32(4):864–870, Feb 2014.

[10] D. E. Eisenbud, C. Yi, C. Contavalli, C. Smith, R. Kononov, E. Mann-Hielscher, A. Cilingiroglu, B. Cheyney, W. Shang, and J. D. Hosein. Maglev: A fast and reliable software network load balancer. *NSDI*, 2016.

[11] M. Filer, J. Gaudette, M. Ghobadi, R. Mahajan, T. Issenhuth, B. Klinkers, and J. Cox. Elastic optical networking in the microsoft cloud. *Journal of Optical Communications and Networking*, 8(7):A45–A54, Jul 2016.

[12] K.-T. Foerster, S. Schmid, and S. Vissicchio. Survey of consistent network updates. *CoRR*, abs/1609.02305, 2016.

[13] M. Ghobadi, J. Gaudette, R. Mahajan, A. Phanishayee, B. Klinkers, and D. Kilper. Evaluation of elastic modulation gains in Microsoft's optical backbone in North America. *Optical Fiber Communication Conference*, page M2J.2, 2016.

[14] M. Ghobadi and R. Mahajan. Optical layer failures in a large backbone. *IMC*, 2016.

[15] R. Govindan, I. Minei, M. Kallahalla, B. Koley, and A. Vahdat. Evolve or die: High-availability design principles drawn from googles network infrastructure. *SIGCOMM*, 2016.

[16] M. Gunkel, A. Mattheus, F. Wissel, A. Napoli, J. Pedro, N. Costa, T. Rahman, G. Meloni, F. Fresi, F. Cugini, N. Sambo, and M. Bohn. Vendor-interoperable elastic optical interfaces: Standards, experiments, and challenges [invited]. *IEEE/OSA Journal of Optical Communications and Networking*, 7(12):B184–B193, Dec 2015.

[17] C.-Y. Hong, S. Kandula, R. Mahajan, M. Zhang, V. Gill, M. Nanduri, and R. Wattenhofer. Achieving high utilization with Software-driven WAN. *SIGCOMM*, 2013.

[18] C. R. Jackson, R. Nejabati, F. Agraz, A. Pagès, M. Galili, S. Spadaro, and D. E. Simeonidou. Demonstration of the benefits of SDN technology for all-optical data centre virtualisation. *Optical Fiber Communication Conference*, page Tu3L.3, 2017.

[19] S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu, J. Zolla, U. Hölzle, S. Stuart, and A. Vahdat. B4: Experience with a globally-deployed software defined wan. *SIGCOMM*, 2013.

[20] X. Jin, Y. Li, D. Wei, S. Li, J. Gao, L. Xu, G. Li, W. Xu, and J. Rexford. Optimizing bulk transfers with software-defined optical WAN. *SIGCOMM*, 2016.

[21] L. Liu, D. Zhang, T. Tsuritani, R. Vilalta, R. Casellas, L. Hong, I. Morita, H. Guo, J. Wu, R. Martinez, and R. M. noz. First field trial of an OpenFlow-based unified control plane for multi-layer multi-granularity optical networks. *Optical Fiber Communication Conference*, page PDP5D.2, 2012.

[22] D. Loher. SONiC: Software for Open Networking in the Cloud. https://azure.github.io/SONiC/.

[23] P. Patel, D. Bansal, L. Yuan, A. Murthy, A. Greenberg, D. A. Maltz, R. Kern, H. Kumar, M. Zikos, H. Wu, C. Kim, and N. Karri. Ananta: Cloud scale load balancing. *SIGCOMM*, 2013.

[24] D. Simeonidou, R. Nejabati, and S. Azodolmolky. Enabling the future optical Internet with OpenFlow: A paradigm shift in providing intelligent optical network services. *International Conference on Transparent Optical Networks*, pages 1–4, June 2011.

[25] A. Sivaraman, A. Cheung, M. Budiu, C. Kim, M. Alizadeh, H. Balakrishnan, G. Varghese, N. McKeown, and S. Licking. Packet transactions: High-level programming for line-rate switches. *SIGCOMM*, 2016.

[26] A. Sivaraman, S. Subramanian, M. Alizadeh, S. Chole, S.-T. Chuang, A. Agrawal, H. Balakrishnan, T. Edsall, S. Katti, and N. McKeown. Programmable packet scheduling at line rate. *SIGCOMM*, 2016.

[27] S. Vissicchio, L. Vanbever, and J. Rexford. Sweet little lies: Fake topologies for flexible routing. *SIGCOMM*, 2015.