Bandwidth on Demand for Inter-Data Center Communication

Ajay Mahimkar, Angela Chiu, Robert Doverspike, Mark D. Feuer, Peter Magill, Emmanuil Mavrogiorgis, Jorge Pastor, Sheryl L. Woodward, Jennifer Yates

AT&T Labs – Research
{mahimkar,chiu,rdd,mdfeuer,pete,emaurog,jorel,sheri,jyates}@research.att.com

ABSTRACT

Cloud service providers use replication across geographically distributed data centers to improve end-to-end performance as well as to offer high reliability under failures. Content replication often involves the transfer of huge data sets over the wide area network and demands high backbone transport capacity. In this paper, we discuss how a **G**lobally **R**econfigurable **I**ntelligent **P**hotonic **N**etwork (GRIPhoN) between data centers could improve operational flexibility for cloud service providers. The proposed GRIPhoN architecture is an extension of earlier work [34] and can provide a bandwidth-on-demand service ranging from low data rates (e.g., 1 Gbps) to high data rates (e.g., 10-40 Gbps). The inter-data center communication network which is currently statically provisioned could be dynamically configured based on demand. Today's backbone optical networks can take several weeks to provision a customer's private line connection. GRIPhoN would enable cloud operators to dynamically set up and tear down their connections (sub-wavelength or wavelength rates) within a few minutes. GRIPhoN also offers cost-effective restoration capabilities at wavelength rates and automated bridge-and-roll of private line connections to minimize the impact of planned maintenance activities.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Network communications*

General Terms

Design, Performance, Reliability

Keywords

Inter-data center communication, ROADM, OTN

Hotnets '11, November 14–15, 2011, Cambridge, MA, USA.

1. INTRODUCTION

In the past few years, we have seen the rapid growth of cloud service offerings from companies such as Amazon [4], IBM [21], Yahoo [32], Apple [2], Microsoft [5], Google [15] and Facebook [10]. These cloud service providers (CSP) use multiple geographically distributed data centers to improve the end-to-end performance as well as to offer high availability under failures. Massive amounts of content are being collected by the data centers. The CSPs often replicate the content on a regular basis across multiple data centers. Inter-data center replication and redundancy impose high bandwidth requirements on the inter-data center wide area network.

Traditionally, a CSP leases or owns a dedicated line between its data centers. Greenberg *et al.* [16] reports that wide area transport is expensive and costs more than the internal network of a data center. This is also why some CSPs do not operate multiple geographically distributed data centers [20]. The peak traffic volumes between data centers are dominated by background, non-interactive, bulk data transfers (as also observed by Chen *et al.* [6]). The CSP runs backup and replication applications to transfer bulk data between its data centers. The scale of this data can range from several terabytes (*e.g.*, emerging scientific and industrial applications) to petabytes (*e.g.*, Google's Distributed Peta-Scale Data Transfer [36]). A recent survey conducted by Forrester, Inc. [14] further highlights that a majority of CSPs perform bulk data transfer among three or more data centers. They project that inter-data-center transport requirements will double or triple in the next two to four years.

There is a great deal of research literature on achieving full bisection bandwidth within a data center with improved network performance (*e.g.*, VL2 [17], DCell [19], BCube [18], MDCube [31], PortLand [25], c-Through [29], Helios [11], Proteus [28]). However, there are few recent studies on inter-data center bulk transfers [1, 6, 8, 23, 22]. Chen *et al.* [6] characterizes the inter-data center traffic characteristics using Yahoo! data-sets. NetStitcher [22] takes the interesting approach of stitching together unutilized bandwidth across different data centers by using multi-path and multi-hop store and forward scheduling. It effectively achieves inter-data center bulk transfers with existing capacity.

Our Approach. In this paper, we take a completely different approach to achieving dynamic inter-data center commu-

nication. We propose GRIPhoN - a **G**lobally **R**econfigurable **I**ntelligent **P**hotonic **N**etwork that would offer a Bandwidth on Demand (BoD) service in the core network for efficient inter-data center communication. We believe we are the first to address the inter-data center capacity issue from the carrier's perspective. The motivation behind BoD comes from the variability in traffic demands for communication across data centers. Non-interactive bulk data transfers between data centers are typically performed by the cloud operators and have different patterns than interactive end-user driven traffic. This gives us the opportunity to explore the use of different data rates at different times - for example, high data rate (10-40 Gbps) between data centers for non-interactive data transfers and low rate (1-10 Gbps) for supporting interactive sessions. GRIPhoN provides a platform for offering such dynamic connectivity. The inter-data center communication network which was previously statically provisioned can now be viewed as *adjustable*. GRIPhoN offers flexibility to the CSP in dynamically adjusting the bandwidth between its geographically distributed data centers based on the demand. The carrier also benefits from the intelligent re-use of the pool of resources across multiple customers.

BoD Service Vision and Today's Reality. We now outline the dynamic service vision of GRIPhoN and compare it to today's reality.

1. **Dynamic configurable-rate services.** The vision behind GRIPhoN is to offer dynamic multi-rate services for communication between geographically distributed data centers. Having a choice between multiple data rates offers flexibility to the CSPs in dynamically selecting the right bandwidth based on demand. Today carriers offer BoD private-line services in limited architectures and usually at rates $\leq$622 Mbps.
2. **Rapid establishment of new connections.** Dynamic bandwidth adjustments require rapid connection provisioning. This is achievable today at low data rates by re-configuring electronic circuit switches [9]. However, provisioning times for connections which require a full wavelength in the backbone are orders of magnitude slower than needed. This is primarily because there has been no call for faster times and hence neither the Element Management Systems (EMS) nor the optical hardware is optimized for fast speeds.
3. **Reduced outage time.** Following any network failure, it is important to quickly restore the service. For low-data-rate services, restoration times are on the order of milliseconds. However, no restoration is usually available today for full wavelength capabilities. There are two alternatives for private-line customers: either buy expensive 1+1 protection where if a primary connection fails, traffic is re-routed to a backup, or wait for the carrier to manually restore connections which means long outage times (4 to 12 hours typically).
4. **Minimal impact during maintenance.** Maintenance is a significant aspect of managing and operating large networks. Carriers would like to ensure minimal or no impact of maintenance on performance. Since the wavelength connection management is being manually handled today, there is a non-negligible impact on service.

GRIPhoN Contributions. GRIPhoN aims at bridging the gap between the dynamic service vision and today's reality as shown in Table 1. By offering dynamic configurable-rate services, GRIPhoN enables the CSP to actively adjust their inter-data center connections. Such a BoD service is not new to large carriers, at least for lower data rates. Such lower-data-rate services are already offered, for example the Optical Mesh Service (OMS) [9, 26, 27]. GRIPhoN scales these concepts to very high data rates and offers the first BoD service demonstration that can select data rates from sub-wavelength connections (*e.g.*, 1 Gbps) to full wavelength connections (*e.g.*, 10-40 Gbps). The sub-wavelength connections are provided by OTN (Optical Transport Network) switches in the network's OTN Layer. Full wavelength connections are established in the photonic layer by using colorless and non-directional reconfigurable optical add/drop multiplexers (ROADMs). A CSP leases dedicated optical access to the GRIPhoN core network at multiple data center locations and dynamically sets up optical connections between them. GRIPhoN enables dynamic and rapid connection management capabilities with the automated control of fiber cross-connects (FXC) to route signals to either the photonic or OTN layer. This enables a CSP to utilize wavelength and/or sub-wavelength resources.

GRIPhoN also offers cost-effective restoration capabilities at wavelength rates via automatic fault identification and dynamic re-establishment of connections. This reinstates customer connections far faster than repair of the underlying fault. Though not as fast as 1+1 protection, this would also be far less expensive. Finally, by using automated bridge-and-roll [34] of private line connections, GRIPhoN minimizes the impact during planned maintenance.

Comparison to prior work on dynamic optical networks. In contrast to CANARIE [3], CHEETAH [35], DRAGON [24], DWDM-RAM [13] and Lambda GRID [33] which are initiatives of research and education networks that serve universities and national laboratories, GRIPhoN is intended for the backbone network of a major carrier. Providing dynamic wavelength services on an inter-city commercial network presents challenges not only in the eventual scale that must be managed, but also in the transition phase from today's static network. Efficient network implementation across multiple layers and multiple customers, cost-effective service restoration and conformance with commercial operational practices have received less attention in the research and education initiatives, whereas these issues are the primary focus in GRIPhoN.

2. BANDWIDTH ON DEMAND SERVICE

In this section, we first present a simplified view of the services and network layers offered by the carrier. We then describe the design of the BoD service offered by GRIPhoN that can be utilized by the cloud service providers to dynamically adjust the bandwidth available between their data centers.

BoD service vision	Today's reality	GRIPhoN proposal
Dynamic configurable-rate	Maximum rate well below full wavelength rate	Rate configurable over wide range. Integrated services using OTN, FXC and wavelength switching
Rapid establishment of new connections	Takes several weeks for highest data rates	Automated Fiber Cross-connect (FXC) and ROADMs enable full wavelength connections in minutes
Reduced outage times	None (unless 1+1) for full wavelength rates	Automated outage detection and dynamic re-provisioning of impacted connections
Minimal impact during maintenance	Non-negligible impact on service	Automated bridge-and-roll [34]

Table 1: Bandwidth on Demand (BoD) service vision, today's reality and GRIPhoN proposal.

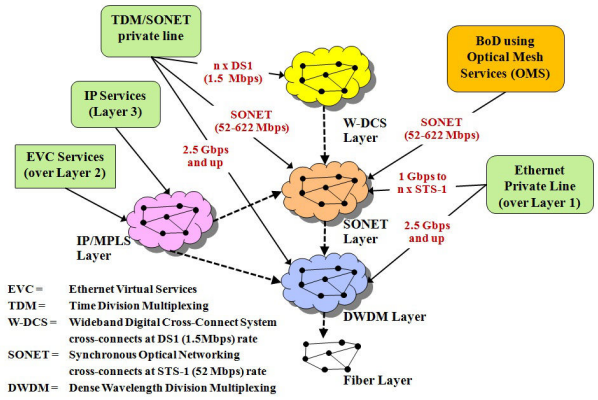

Figure 1: Carrier's view of current services & network layers.

2.1 Carrier's view of services & network layers

Fig. 1 provides a simplified representation of how today's technology layers are interrelated and how the service categories map to them. Most large carriers' current core transport networks consist of a Wideband Digital Cross-connect System (W-DCS) Layer, SONET Layer, DWDM Layer, and Fiber Layer.

Consider the network layers from the bottom up. At the very base is the fiber-optic layer. This layer consists of fiber-optic cables connecting the various nodes in the network. Laying these cables between cities is a huge capital investment, and hence this layer is very static. Built upon this fiber base is the transport layer. Dense-wavelength-division multiplexing (DWDM) is utilized in the core network because of it's huge capacity compared with all other technologies. A modern DWDM system utilizes anywhere from 40 to 100 wavelengths, each carrying signals at rates ranging from 10 to 100Gbps. Sub-wavelength channels at 2.5Gbps or 10Gbps can be provided via muxponders. These wavelength connections are bidirectional and multiplexed together onto a fiber-pair. Hence the transport layer is known as the DWDM layer. DWDM systems were initially point-to-point systems, with all traffic terminating at the two end nodes. If some connections were destined to travel further down the line, then they would be electronically regenerated before transmission on the next leg of their path. In recent years, ROADM technologies for DWDM transport networks have been deployed due to their capital and operational savings. A ROADM network typically includes a set of multi-degree ROADM nodes connected via fibers to form a mesh topology. Traffic may be added or dropped, regenerated, or expressed through at each ROADM. Optical transponders (OT) are connected to the ports of the ROADM to transmit and receive line-side optical signals and convert them to standard client-side optical signals. Optical-to-Electrical-to-Optical (OEO) regeneration is needed when the distance between terminating nodes exceeds a limit for adequate signal quality, known as the optical reach. When that happens, optical regenerators (REGENs) are used at one or more intermediate nodes. ROADM's are now being deployed with add/drop ports which are both "colorless" (so that any OT can be tuned to provide a signal at any wavelength) and "non-directional" (Any OT's signal can be used on any of the ROADM's inter-node fiber-pairs; this is also referred to as "steerable" or "directionless").

The SONET (Synchronous Optical Network) layer rides on top of DWDM layer with Broadband DCSs that cross-connect at STS-1 rate as its most common network element. The Add-Drop Multiplexer (ADM) is a special case of a DCS with 2 degrees to form SONET rings. It provides SONET connections at rates from STS-1 (52Mbps) to OC-192 (10Gbps). It carries both TDM and data traffic and provides an automatic protection/restoration mechanism to switch traffic from working circuits to backup circuits in less than a second. The Wide-band Digital Cross-connect System (W-DCS) is above the SONET layer and consists of DCS-3/1s and other DCS that cross-connect at greater than DS0 but below DS3 rates. It provides nxDS1 (1.5Mbps) TDM connections.

Ethernet Virtual Circuits (EVCs) provide virtual links with guaranteed bandwidth. Ethernet private lines are links between customer routers or Ethernet switches, usually consisting of Gigabit Ethernet interfaces at customer ends and then encapsulated and rate-limited into pipes consisting of virtually concatenated SONET STS-1s. Circuit-based BoD services use virtual concatenation of channels fed from a dedicated access or metro pipe to the customer. With current services and network layers, the carrier offers BoD only at the SONET layer, not at the DWDM layer. With the GRIPhoN vision using future services & network layers, BoD at high data rates would be offered at the OTN layer as well as the DWDM layer.

Fig. 2 provides a view of such future services and network layers from the carrier's perspective. One of the key assumptions of this service evolution model is that the transport of Guaranteed Bandwidth connections can be categorized by bandwidth: below 1 Gbps is transported via the IP layer as EVCs; 1 Gbps up to the core wavelength rate is transported by the sub-wavelength layer as Ethernet Private Lines, most likely encapsulated into concatenated TDM pipes; high-rate private-line services (TDM connections of wavelength rate) are carried directly over the DWDM (Dense Wavelength Di-

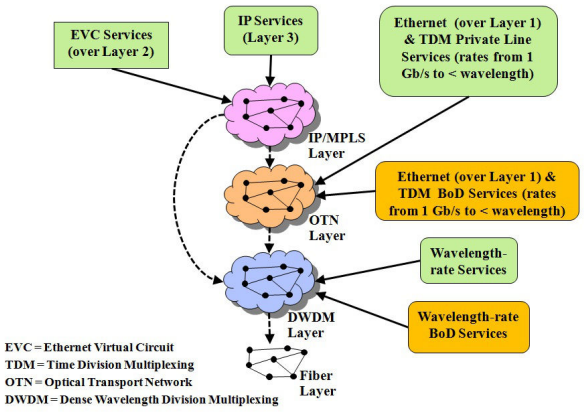

Figure 2: Carrier's view of future services & network layers.

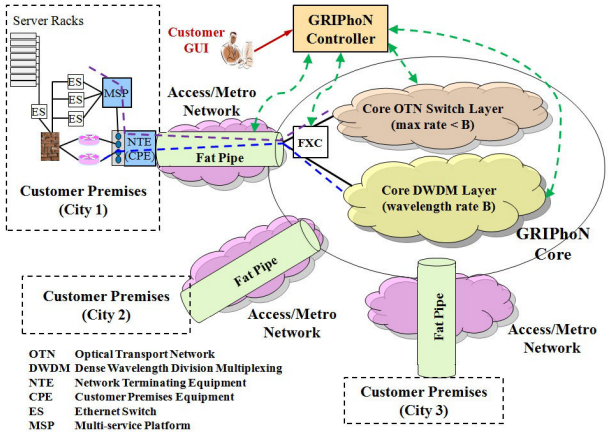

Figure 3: BoD for inter-data center communication using GRIPhoN.

vision Multiplexing) layer. The OTN layer is introduced as the sub-wavelength layer that provides higher switching capacity and better scalability than today's SONET/Broadband DCS layer. The OTN switches cross-connect at an ODU0 rate (1.25Gbps) and can support both TDM and Ethernet packet-based client signals. Using ITU-standardized digitally framed signals with digital overhead, the OTN layer supports connection management as well as Forward Error Correction for enhanced system performance. Compared to using muxponders in the DWDM layer to provide sub-wavelength connections, the OTN layer with its switching capability can achieve more efficient packing of wavelengths in the transport network. Moreover, it can provide automatic sub-second shared-mesh restoration similar to today's SONET layer.

2.2 GRIPhoN Design

Fig. 3 shows an overview of the GRIPhoN target service architecture that enables BoD service for dynamic inter-data center communication. The data center premises connect to the carrier's network via a fixed, dedicated access pipe. In order to allow for better grooming of the provided bandwidth, we partition the carrier's network into two separate layers that are (i) the Optical Transport Network (OTN) layer that provides low data rate connections (*e.g.*, 1 Gbps), and (ii) the Dense Wavelength Division Multiplexing (DWDM) layer that provides high data rate connections (*e.g.*, 40 Gbps). This allows a CSP to adjust the bandwidth according to their exact needs. For example, they can use lower-speed circuits to augment a high-speed circuit by using a combination of 2 x 1G OTN circuits and one 10G DWDM to achieve a total bandwidth of 12G instead of consuming a second 10G DWDM.

Reconfigurable Fiber Cross-Connect (FXC). In order to efficiently provide BoD services at wavelength rates, it is necessary to have a switch on the client-side of the OT [12, 30]. A client-side switch allows for dynamic sharing of transponders, which is useful in keeping costs low. While this switch could be electronic, the low cost, small footprint, and low-power consumption of a fiber-cross-connect (FXC) makes it an attractive technology. Unfortunately, an FXC is incapable of grooming traffic. Therefore, to provide BoD services at rates below the data rate of a single wavelength, electronic switching is necessary. This is provided by the OTN switch, a part of the OTN layer of the GRIPhoN network. This layer rides on top of the DWDM layer. When a connection is requested, the FXC, under the control of the GRIPhoN controller, directs the signal to either an OT, to be carried directly on the DWDM layer, or to a port on the OTN switch, where it can be combined with other OTN signals before transmission over the DWDM layer.

GRIPhoN controller. Connection establishment and release based on requests from the CSP are handled by the GRIPhoN controller. The GRIPhoN controller communicates with the network elements via the appropriate vendor-supplied EMS. The controller is responsible for keeping track of the available network resources in its database, communication with the network elements (FXC controllers, OTN switch EMS, ROADM EMS and NTE controllers) in order to create or tear down the connections ordered by the CSPs, capacity and resource management, inventory database management, failure detection, localization and automated restorations. To minimize service interruption during network reconfigurations due to restoration, the GRIPhoN controller executes a bridge-and-roll operation [7, 34] that first creates a full new wavelength path (the "bridge") while the original connection is still in use and then quickly "rolls" the traffic onto the new path when ready. The bridge-and-roll results in an almost hitless movement of traffic prior to scheduled maintenance or reversion following a failure restoration (moving traffic from backup paths to repaired primary). One constraint of the bridge-and-roll operation however, is that the new wavelength path has to be resource disjoint to the old path.

Customer Graphical User Interface (GUI). Each customer has a graphical user interface to GRIPhoN to visualize and manage his connections. The customer only visualizes the channelized or un-channelized interfaces (for sub-wavelength or wavelength connections, respectively) of the NTE on his

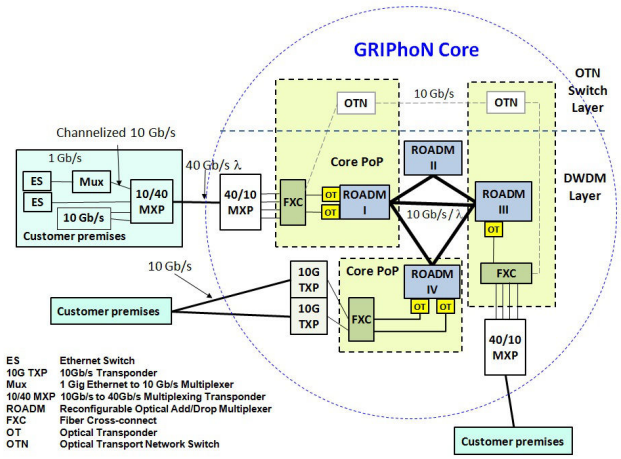

Figure 4: GRIPhoN Testbed.

premises. The GUI comprises capabilities for connection management (setting up or tearing down connections on demand) and simple fault management from the customer viewpoint, such as showing status of connections affected by outages, localizing the fault location, and indicating when restoration is performed. The complexity of the GRIPhoN network (access pipes, carrier equipments, network layers, GRIPhoN controller) is hidden from the customer.

3. TESTBED

In this section, we describe our laboratory prototype implementation of GRIPhoN and present preliminary results on wavelength connection management. Fig. 4 shows our GRIPhoN testbed with three customer premises, and the core GRIPhoN network with DWDM and OTN layer. The DWDM layer consists of Reconfigurable Optical Add/Drop Multiplexers (ROADMs) to provide wavelength switching (currently at 10 Gbps, with plans to go to 40 Gbps). In our prototype, we use two 3-degree ROADMs and two 2-degree ROADMs. Wavelength-tunable optical transponders (OT) are installed at the ROADM add/drop ports and are used to setup end-to-end wavelength connections. Client-side FXCs allow for dynamic sharing of OT's and REGENs. The OTN layer is in the process of installation. Each of the three customer premises sites that could host a data center facility includes servers, Ethernet switches, low-speed multiplexers (1Gbps/10Gbps), and a 10Gbps/40Gbps Muxponder (10/40 MXP). The servers provide video-on-demand (VoD) content across multiple facilities. The 1/10Gbps multiplexer aggregates from multiple Ethernet switches and transmits over a high-speed (10Gbps) channelized line. The 10/40Gbps Muxponder emulates Network Terminating equipment (NTE) and has four 10Gbps ports on the client side and a 40Gbps transmission rate on the line side (towards the carrier). The line-side is the "fat pipe" shown in Fig. 3, and it emulates a metro network which brings customer traffic to the core network. Central Office terminals (COT) would receive the customer data, in our prototype this is emulated by another 10/40MXP.

Path length (hops)	1 (I-IV)	2 (I-III-IV)	3 (I-II-III-IV)
Connection establishment time (seconds)	62.48	65.67	70.94

Table 2: Dependence of wavelength connection establishment times and the path length in the ROADM layer.

We have constructed a customer GUI that has capabilities for dynamically setting up and tearing down connections at chosen rates. It shows four 10Gbps ports at each customer premises. In this paper, we focus on DWDM layer experiments. The 10Gbps connection is established from the customer to the Core PoP (Core Point-of-Presence) over the customer's fat pipe controlled through the EMS of the 40Gbps link. The wavelength connection that will be used to traverse the backbone network is set up between a pair of OTs installed at the source and destination ROADM nodes (in this case, in their respective core PoPs). The establishment of a wavelength connection ranges from 60 to 70 seconds. There are two contributions to this time: (i) ROADM Element Management System (EMS) configuration steps, and (ii) optical tasks, such as ROADM reconfiguration, laser tuning, power balancing and link equalization. The times associated with both components at present are not constrained by any fundamental limitations; rather, they represent a lack of current carrier requirements for speed. We are now working with equipment suppliers to further understand the setup times and ways to reduce them. The 60-70 seconds for wavelength connection establishment is orders of magnitude better than today's provisioning time in the DWDM layer. This was primarily achievable due to the automated reconfiguration of fiber cross-connects and ROADMs using the GRIPhoN controller. Tearing down a wavelength connection takes around 10 seconds. We also performed preliminary analysis on the dependence of the connection provisioning times on the path lengths (number of hops) in the ROADM (or, DWDM) layer. Table 2 summarizes the results over ten iterations. As the path length increases, the connection provisioning increases.

4. RESEARCH CHALLENGES

The BoD services offered by GRIPhoN introduce an entirely new set of research and operational challenges. An effective, integrated network design or restoration process across IP, OTN and DWDM layers necessitates cross-layer management. The dynamic services, the intelligent and autonomous network, and the integration of multiple network layers together present several challenges:

Network resource planning. Ensuring adequate network resources to support anticipated demand from the CSPs is made more difficult by the existence of dynamic services. In order to support rapid connection provisioning and faster restorations, the carrier must plan ahead, where and when to deploy the spare resources (especially OTs). Obviously, it would be very expensive for the carrier to provision in lieu of all possible usage scenarios. Thus, they need to forecast demand and carefully manage the pool of GRIPhoN resources. The carrier should also ensure isolation of services across different CSPs. This resource planning at first glance may seem similar to the planning that was performed in providing plain old telephony services (POTS) with resources (phone

circuits) statistically shared by multiple users. However, in this network the number of users is smaller and the cost of a line is far greater, making accurate planning far more critical.

Network re-grooming. One attractive application of GRIPhoN that is tolerant of the connection times demonstrated in this work is network grooming. As the GRIPhoN network grows, additional routes between nodes will be added. This will make paths that were previously unavailable more appropriate for some connections than the originally established paths. The carrier may then want to re-provision the inter-data center communication network with better paths (reducing latency and/or off-loading the original paths). The process of re-provisioning connections to achieve an improved network configuration is called re-grooming. In order to perform re-grooming with minimal impact to the CSP, the GRIPhoN bridge-and-roll can be used to migrate the wavelength connections [34].

DWDM layer management. The connection establishment times we have demonstrated are far slower than any fundamental limitations on the DWDM layer. To reduce the connection establishment time will place additional requirements on both the physical hardware and software control used in the DWDM layer. The optical transport system must be able to turn wavelengths on/off and route them appropriately without affecting other connections. This has implications for the entire DWDM layer, from how quickly a new wavelength is turned on, to the power transient tolerance of the optical line (including both amplifiers and receivers). The latter requirement is already being addressed by carriers requiring that a cable cut in one part of a mesh network will not affect traffic in another part of the network. Achieving a DWDM layer with dramatically faster end-to-end connection times in a *cost-effective* manner requires that the entire system's dynamics be considered.

5. SUMMARY

In this paper, we presented the design of Globally Reconfigurable Intelligent Photonic Network (GRIPhoN) between data centers that can provide BoD service ranging from low data rates (*e.g.*, 1 Gbps) to wavelength rates (*e.g.*, 40 Gbps). GRIPhoN provides flexibility to the cloud service providers to dynamically set up and take down their wavelength connections between their geographically distributed data centers when performing tasks like content replication or non-interactive bulk data transfers.

Acknowledgement

We thank Adel Saleh, the DARPA Program Manager of the CORONET Program, for his inception of the program and for his guidance. We appreciate the support of the DARPA CORONET Program, Contract N00173-08-C-2011 and the U. S. Army RDE Contracting Center, Adelphi Contracting Division, 2800 Powder Mill Rd., Adelphi, MD under contract W911QX-10-C00094. We thank Amin Vahdat (our shepherd), Rakesh Sinha and the HotNets anonymous reviewers for their insightful feedback. We also thank Fujitsu and Ciena for their equipment and technical support.

6. REFERENCES

[1] S. Agarwal, J. Dunagan, N. Jain, S. Saroiu, A. Wolman, and H. Bhogan. Volley: automated data placement for geo-distributed cloud services. In *NSDI*, 2010.
[2] Apple icloud. `http://www.apple.com/icloud/`.
[3] B. S. Arnaud, J. Wu, and B. Kalali. Customer-controlled and -managed optical networks. In *Journal of Lightwave Technology*, 2003.
[4] Amazon Simple Storage Service. `aws.amazon.com/s3/`.
[5] Windows azure. `http://www.microsoft.com/windowsazure/`.
[6] Y. Chen, S. Jain, V. K. Adhikari, Z.-L. Zhang, and K. Xu. A first look at inter-data center traffic characteristics via yahoo! datasets. In *IEEE INFOCOM*, 2011.
[7] A. L. Chiu, G. Choudhury, G. Clapp, R. Doverspike, J. W. Gannett, J. G. Klincewicz, G. Li, R. A. Skoog, J. Strand, A. von Lehmen, and D. Xu. Network design and architectures for highly dynamic next-generation ip-over-optical long distance networks. In *Journal of Lightwave Technology*, 2009.
[8] M. Chowdhury, M. Zaharia, J. Ma, M. I. Jordan, and I. Stoica. Managing data transfers in computer clusters with Orchestra. In *ACM SIGCOMM*, 2011.
[9] R. Doverspike. Practical aspects of bandwidth-on-demand in optical networks. In *Panel on Emerging Networks, Service Provider Summit, OFC*, 2007.
[10] Facebook Statistics. `www.facebook.com/press/info.php?statistics`.
[11] N. Farrington, G. Porter, S. Radhakrishnan, H. H. Bazzaz, V. Subramanya, Y. Fainman, G. Papen, and A. Vahdat. Helios: a hybrid electrical/optical switch architecture for modular data centers. In *ACM SIGCOMM*, 2010.
[12] M. D. Feuer, D. C. Kilper, and S. L. Woodward. ROADMS and their system applications. In *Optical Fiber Telecommunications VB*. New York: Academic Press, 2008.
[13] S. Figueira, S. Naiksata, H. Cohen, D. Cutrell, P. Daspit, D. Gutierrez, and D. B. Hoang. DWDM-RAM: Enabling grid services with dynamic optical networks. In *IEEE International Symposium on Cluster Computing and the Grid*, 2004.
[14] Forrester research. `http://info.infineta.com/l/5622/2011-01-27/Y26`.
[15] Google. `http://www.google.com/corporate/datacenter/index.html`.
[16] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel. The cost of a cloud: research problems in data center networks. In *ACM SIGCOMM CCR*, 2009.
[17] A. Greenberg, J. R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D. A. Maltz, P. Patel, and S. Sengupta. VL2: a scalable and flexible data center network. In *ACM SIGCOMM*, 2009.
[18] C. Guo, G. Lu, D. Li, H. Wu, X. Zhang, Y. Shi, C. Tian, Y. Zhang, and S. Lu. BCube: a high performance, server-centric network architecture for modular data centers. In *ACM SIGCOMM*, 2009.
[19] C. Guo, H. Wu, K. Tan, L. Shi, Y. Zhang, and S. Lu. DCell: a scalable and fault-tolerant network structure for data centers. In *ACM SIGCOMM*, 2008.
[20] Perspectives - James Hamilton's Blog, Inter-Datacenter replication & geo-redundancy. `http://perspectives.mvdirona.com/2010/05/10/InterDatacenterReplicationGeoRedundancy.aspx`.
[21] Ibm smart cloud. `http://www.ibm.com/cloud-computing/us/en/`.
[22] N. Laoutaris, M. Sirivianos, X. Yang, and P. Rodriguez. Inter-datacenter bulk transfers with NetStitcher. In *ACM SIGCOMM*, 2011.
[23] N. Laoutaris, G. Smaragdakis, P. Rodriguez, and R. Sundaram. Delay tolerant bulk data transfers on the internet. In *ACM SIGMETRICS*, 2009.
[24] T. Lehman, J. Sobieski, and B. Jabbari. DRAGON: a framework for service provisioning in heterogeneous grid networks. In *IEEE Communications Magazine*, 2006.
[25] R. Niranjan Mysore, A. Pamboris, N. Farrington, N. Huang, P. Miri, S. Radhakrishnan, V. Subramanya, and A. Vahdat. Portland: a scalable fault-tolerant layer 2 data center network fabric. In *ACM SIGCOMM*, 2009.
[26] K. Oikonomou and R. Sinha. Network design and cost analysis of optical vpns. In *OFC*, 2006.
[27] Optical mesh service (OMS). `http://http://www.business.att.com/wholesale/Service/data-networking-wholesale/long-haul-access-wholesale/optical-mesh-service-wholesale/`.
[28] A. Singla, A. Singh, K. Ramachandran, L. Xu, and Y. Zhang. Proteus: A topology malleable data center network. In *ACM HotNets*, 2010.
[29] G. Wang, D. G. Andersen, M. Kaminsky, M. Kozuch, T. S. E. Ng, K. Papagiannaki, and M. Ryan. c-Through: Part-time optics in data centers. In *ACM SIGCOMM*, 2010.
[30] S. L. Woodward, M. D. Feuer, J. L. Jackel, and A. Agarwal. Massively-scaleable highly-dynamic optical node design. In *OFC/NFOEC*, 2010.
[31] H. Wu, G. Lu, D. Li, C. Guo, and Y. Zhang. MDCube: a high performance network structure for modular data center interconnection. In *ACM CoNEXT*, 2009.
[32] Yahoo! `http://www.yahoo.com/`.
[33] O. Yu, A. Li, Y. Cao, L. Yin, M. Liao, and H. Xu. Multi-domain lambda grid data portal for collaborative grid applications. *Future Gener. Comput. Syst.*, 2006.
[34] X. J. Zhang, M. Birk, A. Chiu, R. Doverspike, M. D. Feuer, P. Magill, E. Mavrogiorgis, J. Pastor, S. L. Woodward, and J. Yates. Bridge-and-roll demonstration in griphon (globally reconfigurable intelligent photonic network). In *OFC*, 2010.
[35] X. Zheng, M. Veeraraghavan, N. S. V. Rao, Q. Wu, and M. Zhu. CHEETAH: Circuit-switched high-speed end-to-end transport architecture testbed. In *IEEE Communications Magazine*, 2005.
[36] D. Ziegler. Distributed peta-scale data transfer. `http://www.cs.huji.ac.il/~dhay/IND2011.html`.