Service-oriented Layer 1 Virtual Private Network for Grid Applications

Hanxi Zhang, Michel Savoie, Jing Wu,
Scott Campbell
Communications Research Centre
Canada, 3701 Carling Avenue,
Ottawa, Ontario, Canada K2H 8S2

Gregor v Bochmann
School of Information Technology &
Engineering, University of Ottawa
P.O. Box 450, Stn A,
Ottawa, Ontario, Canada, K1N 6N5

Bill St. Arnaud
CANARIE INC.
110 O'Connor St.
Ottawa, Ontario, Canada
K1P 5M9

Abstract: Emerging data-intensive Grid applications call for service-oriented layer 1 virtual private networks (VPNs) as their data plane. Layer 1 VPNs are created by dividing a physical network into web service enabled partitions. In this paper, we introduce the concept of fundamental lightpath, and propose that a fundamental lightpath be taken as the basic unit of optical network partitions. We then enlist the key web service operations a Lightpath Web Service (LP-WS) should support, such as concatenation, partitioning, etc. Furthermore, we discuss the LP-WS in the context of a business process, where institutions involved in a collaborative Grid project acquire a pool of LP-WSs from optical carrier networks, and then integrate these LP-WSs with discipline-specific web services into a workflow.

Keywords: end-to-end (e2e) connection, layer 1 virtual private network (VPN), lightpath web service (LP-WS)

I. INTRODUCTION

The emerging data-intensive Grid applications have stringent requirements on the underlying networks, in terms of throughput, latency and jitter. Many high-energy physics Grid applications desire high-speed networks capable of transferring bulk files in the order of terabytes at rates of 1Gbps or higher [1]. Some Grid applications, such as those featuring interactive and high-resolution object rendering, desire not only high bandwidth but also low latency and low jitter [1,2]. Best-effort IP networks such as the Internet cannot accommodate Grid applications exemplified above at a reasonable cost. Researchers are therefore looking at exploiting the vast bandwidth of fiber optical networks, especially those based on the Wavelength Division Multiplexing (WDM) technology, to serve data-intensive Grid applications.

Using dedicated, circuit-switched end-to-end (e2e) connections is widely recognized as a favorable choice (in many cases the only feasible choice) for data-intensive Grid applications, especially for long-lived Grid sessions that last for hours and days, if not weeks and months [3,4,5,6,7]. When setting up an e2e connection, cross-connections are performed at each optical switch along the path. In this paper the term 'optical switch' is used broadly to refer to electrical-optical switches based on Synchronous Optical Network (SONET) or Synchronous Digital Hierarchy (SDH), as well as all optical switches based on WDM.

In many distributed Grid application scenarios, resources such as computation and storage are at geographically dispersed locations. From the perspective of the networking community, the main challenge is to develop inter-domain optical routing and signaling protocols, so that e2e connections can be set up and torn down across multiple optical carrier networks. On the other hand, the Grid community takes a top-down perspective and sees two major requirements lying ahead.

First, the dynamic nature of Grid computing calls for application-driven provisioning of e2e connections. An e2e connection is traditionally provisioned manually by an administrator of the optical carrier network. However manual provisioning will not fit into the picture of Grid computing where e2e connections need to be set up and torn down on demand. The decisions as to when and where to set up e2e connections, how much bandwidth is needed for each e2e connection and when to tear down e2e connections are all parts of a Grid computing workflow. E2e connections are akin to dedicated physical wires that can be turned on and off by Grid applications. Grid applications shall be aware of optical network resources so that they can drive the layer 1 topology to fit their needs.

Second, resource allocation and management interfaces are needed between Grid applications and optical carrier networks. It is unlikely that an optical carrier network will dedicate all its optical network resources to one single Grid project. Instead, an optical carrier network shall divide its optical switches and optical links into partitions, and each partition is only visible to, and accessible, by the designated Grid project. That is what is defined by ITU-T Recommendation Y.1312 as 'layer 1 virtual private network (VPN)' [8].

In summary, setting up e2e connections for Grid applications is far more than a traffic engineering approach; it brings about the new paradigm of application-driven layer 1 VPN. Both requirements discussed above boil down to modeling optical networks and bandwidth as Grid resources on their own right. More specifically, the de facto Open Grid Services Infrastructure (OGSI) [9] is based on the Service-

Oriented Architecture (SOA) and mandates that all Grid resources be modeled as web services. A key step is therefore to define what is a partition of an optical network, and what is the web service representation of such a partition.

In this paper, we introduce the concept of fundamental lightpath, and propose that a fundamental lightpath be taken as the basic unit of optical network partitions. We then enlist the key web service operations a Lightpath Web Service (LP-WS) should support, such as concatenation, partitioning, etc. Furthermore, we discuss the LP-WS in the context of a business process, where institutions involved in a collaborative Grid project acquire a pool of LP-WSs from optical carrier networks, and then integrate these LP-WSs with discipline-specific web services into a workflow. We also address how inter-domain optical routing and signaling issues are transformed into operations on LP-WSs.

This paper is organized as follows. In Section II we introduce the concept and web service representation of a fundamental lightpath. In Section III we discuss about the business process associated with LP-WSs, including the planning of LP-WSs, workflow composition, error handling and accounting and administrative aspects. In Section IV we conclude the paper.

II FUNDAMENTAL LIGHTPATH AND LP-WS

II.1 Fundamental Lightpath

Optical switches support two types of line cards: add-drop cards and pass-through cards. Add-drop cards are typically gigabit Ethernet cards. The host where a data-intensive Grid service is run gets connected to an add-drop card of the local optical switch via a dedicated local loop. Whether the Grid service is a proxy for sensor networks or instruments, a storage service or a visualization service, it is just a data source and/or a data sink from an optical network's perspective. Pass-through cards are used to multiplex traffic from add-drop cards, and relay multiplexed traffic between adjacent switches.

An e2e connection consists of a series of dedicated channels between a pair of add-drop ports. In the case of WDM networks the constituent channels are wavelengths, whereas in the case of SONET networks the constituent channels are bandwidth-contiguous Synchronous Transport Signal (STS) channels. In many cases, e2e connections need to be set up across heterogeneous domains, such as one based on SONET and the other based on WDM. A common approach to bridge between two heterogeneous domains is to drop the connection off an egress switch of one domain, and then add the connection back to an ingress switch of the other domain via a dedicated Ethernet segment or a Multi-protocol Label Switching (MPLS) tunnel, as shown in Fig. 1 below.

——————	STS Channel
················	Wavelength
– – – – –	Local Loop
▬▬▬▬▬	Dedicated Ethernet Segment or MPLS Tunnel
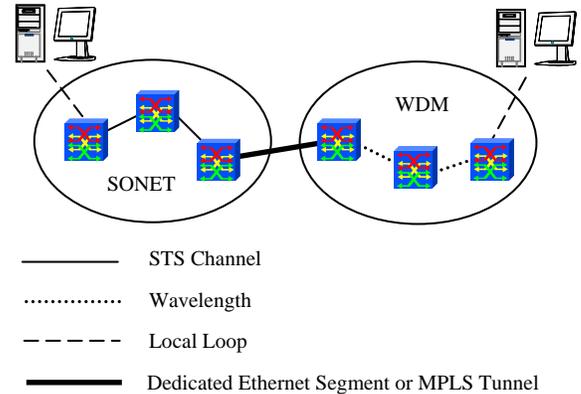
Fig.1   An E2e Connection Across Heterogeneous Domains

In Fig. 1, STS channels, wavelengths and dedicated Ethernet segments are logically identical in that they are all dedicated channels between two adjacent optical switches. Regardless of the specific technology being used, such dedicated channels can either be terminated on Grid services, or be concatenated to make longer dedicated channels that can potentially be terminated on Grid services.

A fundamental lightpath represents a dedicated channel between two adjacent optical switches. A fundamental lightpath is essentially a partition of an optical network, not only because multiple dedicated channels can be multiplexed into a fiber link between two adjacent switches, but also in the sense that a fundamental lightpath is tightly coupled with two ports, one on each end switch. An end-port of a fundamental lightpath can be either a physical port or one of the virtual ports sharing a physical port, depending on the bandwidth and the underlying technology of the fundamental lightpath. Fig. 2 below illustrates an optical network that is partitioned into fundamental lightpaths.
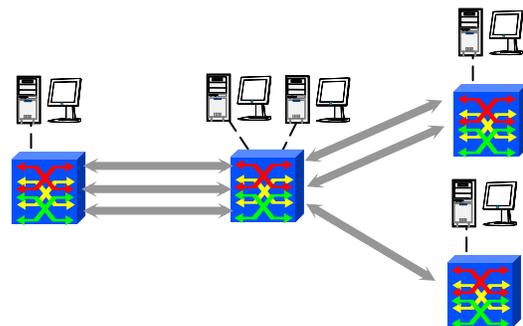
Fig. 2   Fundamental Lightpaths As Optical Network Partitions

II.2   Lightpath Web Service (LP-WS)
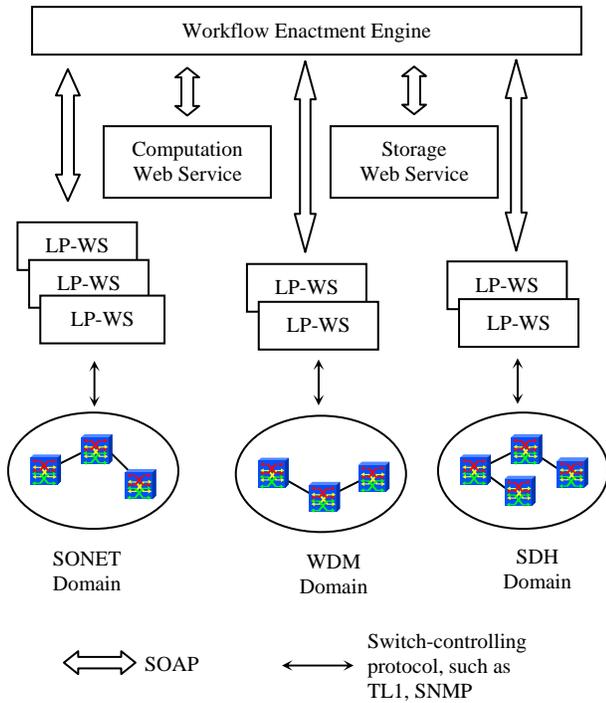
Fig. 3   Layered Control Hierarchy

Operation	Description
Concatenate*	Stitch two or more LP-WSs with the same bandwidth into a longer LP-WS
Unlink*	Chop an LP-WS into two or more shorter LP-WSs with the same bandwidth; the reverse operation of Concatenate
AddDrop*	Cross-connect an LP-WS with a data source or a data sink, used when creating an e2e connection
UnAddDrop*	Release the cross-connection between an LP-WS and a data source or a data sink, used when tearing down an e2e connection; the reverse operation of AddDrop
Partition	Divide an LP-WS into two or more children LP-WSs that go between the same pair of switches but with lower bandwidth; only applies to SONET/SDH channels of certain sizes
Bond	Merge two or more LP-WSs between a pair of switches into an LP-WS with higher bandwidth; only applies to SONET/SDH channels of certain sizes; reverse operation of Partition
Query	Return the Universal Resource Identifiers (URIs) of children LP-WSs (i.e., partitions); only applies to an LP-WS currently partitioned
SetState*	Set status based on the results of an operation
GetState*	Retrieve both static and dynamic service data
Diagnose*	For error handling; more details will be given in Section III

Table 1   LP-WS Operations

Note: an operation marked with an '*' is mandatory

Fundamental lightpath based optical network partitioning fits naturally into the picture where heterogeneous dedicated channels need to be concatenated to set up an e2e connection. The web service representation of a fundamental lightpath, subsequently referred to as a Lightpath Web Service (LP-WS), shall be based on a generic web service interface such that Grid applications can manipulate LP-WSs from heterogeneous domains in a logically consistent fashion, whereas the domain and technology specific details are left to the implementation of each LP-WS instance.

In Fig. 3, we illustrate that LP-WSs made available by different domains appear homogeneous in the sense that they all implement the same LP-WS web service interface. A distributed Grid workflow is launched from a workflow enactment engine, which invokes LP-WSs and non-lightpath web services (such as computation and storage) both using the Simple Object Access Protocol (SOAP), and interoperability is thus achieved at the application level. When high-level operations such as 'concatenate with another LP-WS' or 'unlink from another LP-WS' (the reverse operation) are called upon the LP-WS web service interface, an LP-WS instance will delegate the job to corresponding network elements, using a switching-controlling protocol such as Transaction Language 1 (TL1), Simple Network Management Protocol (SNMP), etc.

An LP-WS should support the following static attributes: the switch/slot/port IDs associated with the two end points of the fundamental lightpath; channel type, e.g., wavelength, STS channel, etc; and finally detailed information about the channel, e.g., the wavelength value of a WDM channel, the start position and bandwidth of a STS channel, etc.

In Table 1 above, we list the basic operations that shall be supported by the LP-WS web service interface, which is to be specified by a Web Service Definition Language (WSDL) file. LP-WS implementations must be smart enough to validate whether it is physically feasible to perform an operation on the fundamental lightpath(s) involved.

An LP-WS is a stateful and secure web service. An LP-WS is stateful in that its status as to whether it is concatenated with other LP-WSs, or whether it is partitioned into children LP-WSs, etc., keeps changing over time. The dynamic states of an LP-WS are reflected by its service data, i.e., dynamic service attributes. In order to facilitate Authentication, Authorization and Accounting (AAA), an LP-WS shall be created as a secure web service, and such a secure web service is only accessible by a client that presents an X.509 digital certificate endorsed by the LP-WS's creator, e.g., an optical domain administrator. By associating a digital identity with a network partition, we can gracefully control who can have access to which ports and channels in an optical domain.

Depending on the bandwidth and start position in the SONET or SDH frame, some SONET/SDH based fundamental lightpaths can be partitioned or bonded. Consider that in a certain stage of a distributed Grid workflow it is desirable to have multiple low bandwidth e2e connections running over a single optical link to support parallel data sessions, and later on a single high bandwidth e2e connection is needed over the same optical link.

Partitioning and bonding operations of the LP-WS make it possible to realize such smart workflows.

III THE BUSINESS PROCESS

III.1 Planning of Initial Set of LP-WSs

The business process associated with LP-WSs consists of two stages. First, upon research institutions' requests, an optical domain makes some network partitions available in the form of LP-WSs. Second, the institutions integrate those LP-WSs with discipline-specific web services into a workflow, execute the workflow and analyze results.

A distributed Grid project usually involves a small number of data-intensive Grid services. These data source and data sink Grid services may be connected to different optical domains. Research institutions planning to run a distributed Grid project should first decide upon the set of LP-WSs to be acquired from optical domains. Instead of considering all intra-domain and inter-domain links, institutions will focus their choice on fundamental lightpaths that go between 'relevant switches', i.e., those switches where data source/sink Grid services are added/dropped, plus a few transit switches when necessary.

The sufficiency of a set of LP-WSs depends on a few factors, including the distribution of data source and data sink Grid services, the topologies of optical domains, the interconnections between domains and the requirements of the Grid workflow, etc. The bottom line is that by starting with an initial set of LP-WSs, a Grid project can always have e2e connections of desired bandwidth between those meaningful data source and data sink pairs, by concatenating some of the LP-WSs assigned to the Grid project. It will be a plus if some fault-tolerance and rerouting capabilities can be achieved with the set of LP-WSs. A sufficient and economical set of initial LP-WSs shall therefore be worked out between research institutions and optical domain administrators.

Research institutions should be able to submit their request for LP-WSs by electronic means, preferably graphical user interfaces showing the network topologies of optical domains. Upon receiving the request for an LP-WS, an optical domain will validate if the requested port and channel resources are available. If the answer is positive, an LP-WS instance will be created. An optical domain will either advertise the newly created LP-WS to an agreed-upon UDDI service, or inform the requesting institutions directly about the URI of the LP-WS. In either case, all research institutions involved in a Grid project will have the pointers to all LP-WSs assigned to the project. In this sense, all LP-WSs can be thought of as stored in a logically centralized registry. In Fig. 4 below, we show that domain boundaries dissolve within a Grid project's layer 1 VPN, which consists of the initial set of LP-WSs.
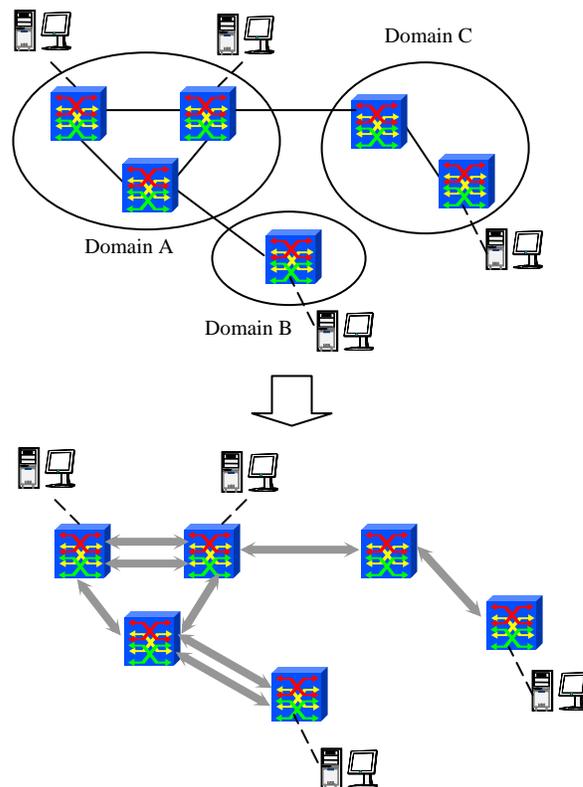
Fig. 4   A Layer 1 VPN Constructed With LP-WSs Acquired From Multiple Domains

Within a service-oriented layer 1 VPN, both optical routing and signaling are application-driven. Routing is a matter of querying LP-WSs to find out a combination that goes from an ingress port to an egress port. Signaling is a matter of bundling the 'concatenate' or 'unlink' calls upon all LP-WSs along a route into one transaction, when creating or deleting an e2e connection.

III.2 Workflow Composition

Researchers can compose Grid workflows using a graphical web service workflow composer, such as Taverna [10], as illustrated in Fig. 5 below. The 'available web services' panel will be populated with available LP-WSs as well as non-lightpath web services. In order to compose a workflow diagram, a user drags web service icons from the 'available web services' panel and imports flow control icons from the graphical composer. The workflow diagram, i.e., the graphical representation of a workflow, can be saved as an XML file, which can then be executed, edited and reused. In a workflow XML file, flow control logics such as conditional execution, loops, parallelism and error handling, etc., are all defined with XML tags. As of this writing the most commonly accepted workflow description schema is Business Process Execution Language (BPEL) [11].
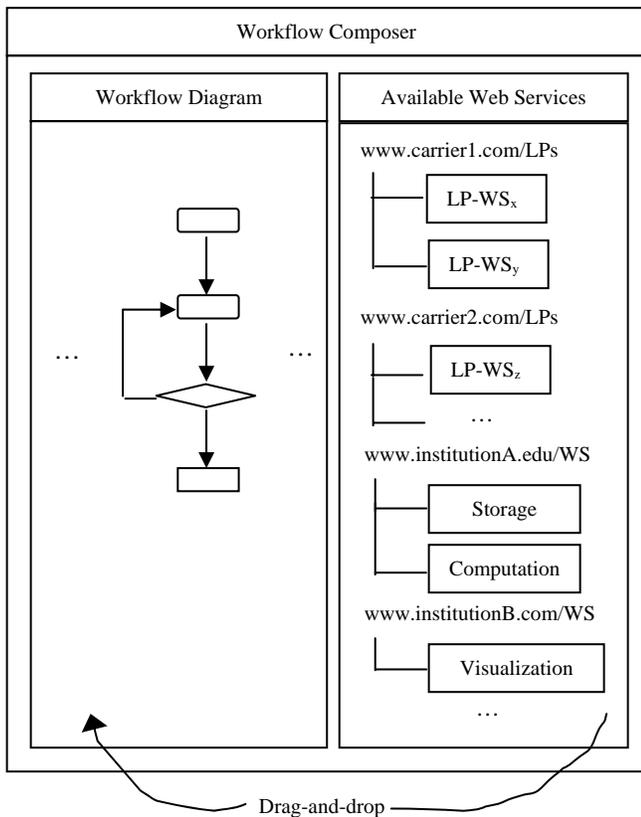
Fig. 5 Graphical Workflow Composer

With the recent advances in semantic web services and semantic Grid [12,13], it is possible for a user to have a workflow automatically composed by specifying some key words and/or with descriptive statements, such as 'find the telescope instrument connected to the WDM switch in Ottawa and the storage service connected to the SONET switch in Toronto; create an e2e connection between the instrument and the storage facility; turn on the telescope and start retrieving data'. That would require ontology and semantic definitions for both LP-WSs and non-lightpath Grid services.

III.3 Error Handling

Critical network outages do occur sometimes. Instead of having each LP-WS periodically check the availability of its physical resources, a simpler and less costly approach is to embed error handing inside a workflow, using redundant LP-WSs.

Suppose that somewhere in a Grid workflow we need to make an e2e connection. We have two candidate routes between the source and the destination: $LP\text{-}WS_1$ -> $LP\text{-}WS_2$ and $LP\text{-}WS_3$ -> $LP\text{-}WS_4$, and these two routes are node-disjoint. For brevity, we use pseudo-codes instead of a workflow language in Fig. 6 below to illustrate how error handling can be embedded inside the workflow.

```
try
{
   // equivalent to LP-WS2.concatenate(LP-WS1);
   LP-WS1.concatenate(LP-WS2);
}
catch (Exception e) //something is wrong with at least one LP-WS
{
    //check resource availability of both LP-WSs
   if(LP-WS1.diagnose( ) == RESOURCE_NOT_AVAILABLE )
        LP-WS1.setState(LPState.UNAVAILABLE);
   if (LP-WS2.diagnose( ) == RESOURCE_NOT_AVAILABLE )
        LP-WS2.setState(LPState.UNAVAILABLE);
   try
   {
        // try a node-disjoint route
        LP-WS3.concatenate(LP-WS4);
   }
   catch(Exception e)
   {
        // nothing can be done now, abort
        exit("both routes failed!");
   }
}
```

Fig. 6 Error-handling Embedded Inside A Workflow

The 'diagnose' operation of an LP-WS checks the availability of the LP-WS's physical resources. By setting the state of an LP-WS to unavailable, the icon corresponding to the LP-WS will become stale on the workflow composer.

III.4 Accounting and Administration

We now come to the accounting and administrative aspects of LP-WSs. Basically research institutions pay on a per LP-WS basis, depending on the bandwidth of an LP-WS, the web service operations supported, and the desired service duration, etc. For an LP-WS straddling two optical domains, the two domains can make arrangements as to how the payment should be split. Research institutions should be given the choice to renew LP-WSs when they are about to expire, and an optical domain should have mechanisms for revoking those LP-WSs whose service period have expired. An optical domain is not only responsible for making sure that the physical resources associated with an LP-WS are available, but also responsible for making sure that an LP-WS is up and running during the requested service period. If an LP-WS is not accessible for some reason, the effect to end-users is equivalent to physical resources being unavailable. Therefore in a production environment, an optical domain should create LP-WSs in a web service application server and run them in a persistent fashion, so that LP-WSs would not lose their states as a result of a system crash or restart.

The accounting and administration complexities discussed above require that on top of the traditional network element oriented management, optical domains should have management tools showing the association between physical resources and LP-WSs.
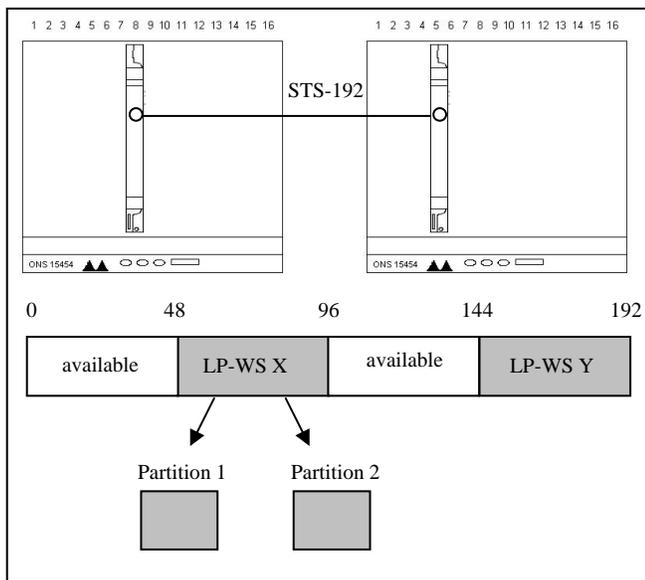
Fig. 7 LP-WS Oriented Management Tool

In Fig. 7 above, we illustrate what an LP-WS oriented management tool could look like. By double-clicking a link on a network topology graph, an optical domain administrator is presented with a window like shown in Fig. 7. An administrator should be able to navigate comfortably between the view in Fig. 7 and the URI-oriented view as follows:

http://www.carrier.com/lightpaths/LP-WS?X
http://www.carrier.com/lightpaths/LP-WS?X/Partition?1
http://www.carrier.com/lightpaths/LP-WS?X/Partition?2
http://www.carrier.com/lightpaths/LP-WS?Y

IV CONCLUSIONS

In this paper, we presented a service-oriented framework that allows distributed Grid applications to control their private and dedicated transport networks, and the Grid 'virtual organization' paradigm can therefore be achieved at layer 1. The proposed framework is based on the LP-WS, which is the web service representation of an optical network partition. It is important that heterogeneous optical network partitions be abstracted using the same LP-WS web service interface. The proposed solution also involves an interesting new business model, where optical carrier networks lease their network partitions in the form of LP-WSs to Grid project participants, who will then integrate LP-WSs with discipline-specific web services into a workflow. The boundary between applications and networks becomes blurry, because now they are all web services.

Detailed design and prototyping of the proposed approach are underway at the Communications Research Centre. Demonstrations of prototype implementations will be reported when available.

REFERENCES

[1] GGF Grid High-Performance Networking Research Group, "Networking Issues for Grid Infrastructure", August 2004: https://forge.gridforum.org/docman2/ViewCategory.php?group_id=53&category_id=750

[2] GGF Grid High-Performance Networking Research Group, "Grid Network Services Use Cases", August 2004: https://forge.gridforum.org/docman2/ViewCategory.php?group_id=53&category_id=750

[3] R. Boutaba, W. Golab, Y. Iraqi and B. St. Arnaud, "Lightpaths on Demand: A Web Services Based Management System", IEEE Communications Magazine, Special Issue on XML-based Management, July 2004

[4] R. Boutaba, W. Golab, Y. Iraqi, T. Li and B. St. Arnaud, "Grid-Controlled Lightpaths for High Performance Grid Applications", Journal of Grid Computing, Special Issue on High Performance Networking, Vol. 1, No. 4, pp. 387-394, 2003

[5] B. St. Arnaud, A. Bjerring, O. Cherkaoui, R. Boutaba, M. Potts and W. Hong, "Web Services Architecture for User Control and Management of Optical Internet Networks", Proceedings of the IEEE, Vol. 92 No. 9, pp. 1490-1500, September 2004

[6] S. Oudenaarde, Z. W. Hendrikse, F. Dijkstra, L. Gommans, C. Laat and R. Meijer, "An Open Grid Services Architecture Based Prototype for Managing End-to-End Fiber Optic Connections in a Multi-Domain Network", accepted for publication in the FGCS special issue on DATATAG: http://staff.science.uva.nl/~delaat/articles/2004-3-aaa.pdf

[7] J. Wu, S. Campbell, J. M. Savoie, H. Zhang, G. v. Bochmann and B. St. Arnaud, "User-managed end-to-end lightpath provisioning over CA*net 4", Proc. National Fiber Optic Engineers Conference (NFOEC), Orlando, FL, USA, Sept 7-11, 2003, pp. 275-282

[8] ITU-T Rec. Y.1312, "Layer 1 Virtual Private Network Generic Requirements and Architectures", Sept. 2003: http://www.itu.int/itudoc/itu-t/aap/sg13aap/history/y1312

[9] Open Grid Services Infrastructure (OGSI) version 1.0: http://www.ggf.org/documents/GWD-R/GFD-R.015.pdf

[10] Taverna Project: http://taverna.sourceforge.net/

[11] Business Process Execution Language (BPEL) specification: http://www-128.ibm.com/developerworks/library/ws-bpel/

[12] W3C Semantic Web: http://www.w3.org/2001/sw/

[13] Semantic Grid Community Portal: http://www.semanticgrid.org/