HIERARCHICAL INTER-DOMAIN MANAGEMENT FOR NETWORKS WITH CONDO-SWITCHES

Gregor v. Bochmann
School of Information Technology and Engineering (SITE)
University of Ottawa, Canada
bochmann@site.uottawa.ca

ABSTRACT
This paper presents a hierarchical approach to inter-domain routing and network management, especially intended for user-controlled lightpath provisioning (UCLP). The structuring of networks into subnetworks at several levels of the hierarchy provides an architecture for distributed processing of network management functions that is very scalable. Each network, as well as each subnetwork, represents an autonomous domain that communicates with its peer, child and parent networks through standard interfaces. A special feature of the architecture is the natural integration of condo-switches, that are switches with ports that belong to different networks, i.e. to different administrative domains. The paper gives the definition of the hierarchical inter-domain architecture with condo-switches and discusses procedures for routing and connection establishment within this structure. It is important to note that the internal structure of a given network (in terms of the interconnections between the internal subnetworks or the point-to-point links) remains hidden; only the list of subnetworks is normally available.

KEY WORDS
Inter-domain routing, user-controlled lightpath provisioning, optical networks, condo-switches

1. Introduction

There are basically two types of customer-owned and managed optical networks: metro dark fiber networks and long-haul wavelength networks. Schools, hospitals and government departments are acquiring their own dark fibers in metropolitan areas. They participate in so-called “condominium” dark fiber networks to better manage their connectivity and bandwidth. They light up the fibers with their own equipment and interconnect their fibers to either like-minded institutions or commercial service providers, Internet Exchanges as they so choose. In the long-haul area, many providers are selling or leasing point-to-point wavelength channels. Some providers are offering “condominium” wavelength solutions, where a number of customers share the capital costs of deploying long-haul optical networks. In return, each customer in the condominium consortium owns a set of wavelength channels. These purchased or leased wavelength channels can generally be treated as an asset rather than a telecom service. The institutions virtually extend their dark fiber networks many thousands of kilometers without having to purchase and maintain their own optical repeaters and associated equipment [1].

The switches within such a condominium network are connected through their ports to the fibers and/or wavelengths that belong to different owners. Therefore it is natural to apply the notion of "condominium" not only to the fibers and wavelengths of the network, but also to the switches: The different ports of a switch belong to different owners, the owners of the attached transmission lines [2]. We call such switches "condo-switches".

User-controlled lightpath provisioning (UCLP) is a traffic engineering mechanism in the context of customer-owned networks [3] (similar ideas have also been presented in [4]). Several UCLP systems have been developed and some have been used within the Canadian CA*net4 network and in international interconnection experiments. In this context, it is often necessary to build end-to-end lightpaths that are composed out of several concatenated path segments that are provided by separate condominium networks. These condominium networks are usually interconnected to one another through switches that allow the concatenation of a lightpath from one network to a lightpath continuing in the other network. If we consider a condominium network as an administrative domain, we therefore have to consider UCLP in an inter-domain context.

The UCLP system developed by CRC and the University of Ottawa [7] includes in its design already the notion of several independent administrative domains, called "federations". They were intended to represent the different provincial research networks within Canada that, together, make up the CA*net4 network. While access rights for administrators of these different federations are distinguished, the UCLP system provides full access for reading the lightpath configurations among all federations

in order to simplify end-to-end lightpath provisioning throughout the whole country.

In the context of UCLP, it has often been proposed that lightpaths that are not used by the owner for some extended period, could be advertised as free resources and leased by other users, possibly for a fee. This could possibly give rise to an open market of communication resources. One of the open questions is by which means the free resources could be advertised. Some form of (probably distributed) directory of available lightpaths would be required.

The UCLP systems existing today have not addressed the following two important questions:

- How could two independent UCLP systems communicate with one another in order to establish an end-to-end lightpath that traverses both networks ? – Some form of standard UCLP inter-domain protocol would be required.

- What kind of mechanism should be foreseen for advertising available lightpaths ? – Such a mechanism should be scalable to very large networks and to the inter-domain context.

This paper proposes a hierarchical structure of networks where sub-networks are interconnected by condo-switches and a hierarchical addressing scheme facilitates the routing problem. Within this hierarchical structure, the above two questions are answered in a natural manner.

The multi-level hierarchical structure of networks and subnetworks provides an architecture for distributed processing of management functions that is very scalable. Each network, as well as each subnetwork, represents an autonomous domain that communications with its peer, child and parent networks through standard interfaces. It is important to note that the internal structure of a given network (in terms of the interconnections between the internal subnetworks and other internal links) remains hidden; only the list of subnetworks is normally available.

After a short review in Section 2 of hierarchical routing in existing networks, Section 3 presents a simple example of a networking configuration with a hierarchical structure of domains (subnetworks), and gives a precise definition of the hierarchical domain architecture with condo-switches, which is proposed in this paper. Section 3.4 in particular, describes a standard set of functions that should be provided by each subnetwork in the hierarchy. Section 4, then, discusses procedures for finding routes and for setting up end-to-end connections through the hierarchical network architecture. Some issues of access rights are also discussed. Section 5 contains our conclusions.

2. Review of hierarchical routing in telephone and packet-switched networks

The telephony system uses a hierarchical numbering scheme that facilitates the routing of telephone calls through the public switched telephone network (PSTN). The following hierarchical levels can be identified: international prefix, regional prefix, local number (which usually is composed out of a number identifying the branch office and a number identifying the port number of the subscriber line).

In the Internet, the numbering scheme, that is, the IP address space, is structured at two levels: network prefix and host address suffix. Within a network representing a so-called autonomous domain (AD) of network management, the routing function is usually performed by the OSPF protocol which distributes the "link state", that is the configuration state of the whole network, to all nodes of the network. This routing approach is clearly not scalable for large networks. Therefore hierarchical routing has been introduced as an extension to OSPF, as shown in Figure 1.

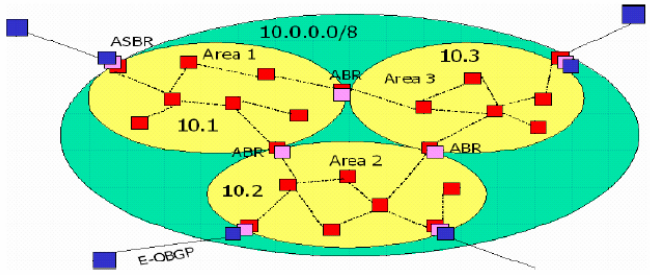

Figure 1: Hierachical routing with OSPF

A separate level of routing hierarchy in the Internet is inter-domain routing, which applies to routing between different ADs. The Border Gateway Protocol (BGP) is normally used for exchanging routing information between neighboring networks. The typical architecture, shown in Figure 2, foresees an inter-domain link between two routers belonging to the two respective networks. These routers are sometimes called "border gateway".

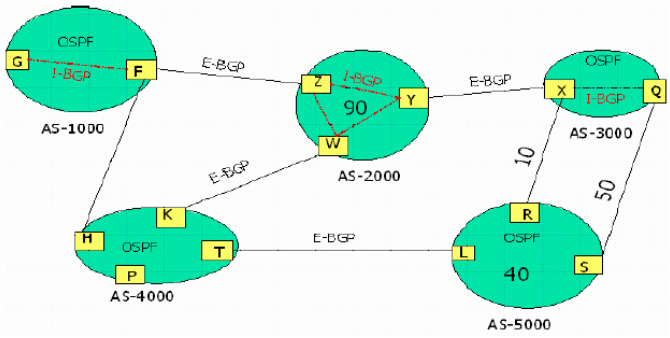

Figure 2: Internet inter-domain routing

We note that the routers labeled "ABR" in Figure 1 could be "condo-routers" where the different ports of the router belong to the different sub-networks that the router interconnects. In the architecture of Figure 2, on the other hand, there is no opportunity for introducing a "condo-router" that has different ports connected to different domains; there are only ports that are connected to an inter-domain link. We conclude that a hierarchical architecture including condo-switches should therefore resemble more Figure 1 than Figure 2.

3. A hierarchical inter-domain model for networks with condo-switches

3.1. An example configuration

Figure 3 shows an example of a (physical) configuration consisting of a number of switches (larger round circles) and terminal devices, such as computers (smaller round circles) interconnected by communication links. In particular, this configuration allows an end-to-end connection between the terminals H1 and H2 through the switches S1, S2, S3, S4, S5, S6, and S7. This figure does not show any administrative domains, although the different communication links and switches may belong to different owners.

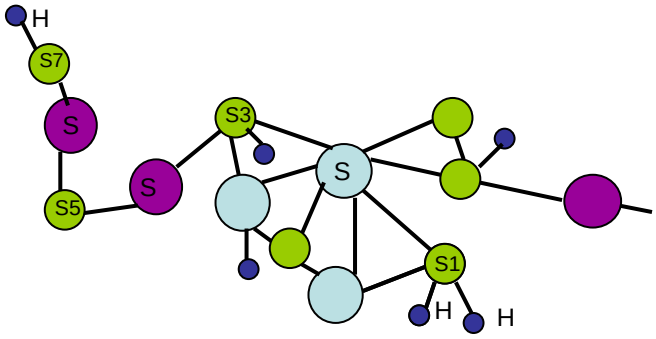

Figure 3: An example configuration of links, switching nodes, and terminal nodes

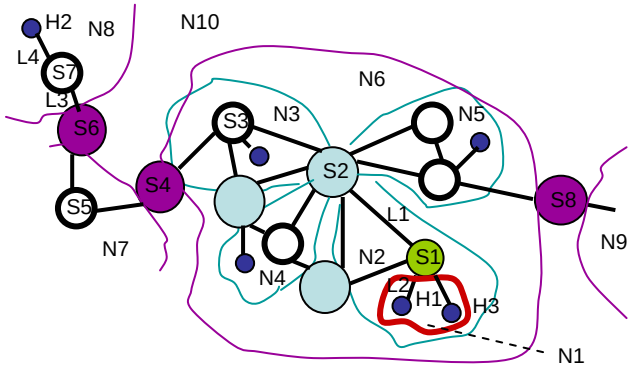

Figure 4: The configuration of Figure 3 with overlaid hierarchical inter-domain structure

Figure 4 shows the same configuration with the superposition of a hierarchical structure of administrative domains, in the following called "networks" or "sub-networks". For instance, network N6 consists of four sub-networks N2, N3, N4, and N5. These subnetworks do not contain any sub-subnetworks, except N2 which contains N1. The whole configuration shown in the figure is partitioned into four networks, N6, N7, N8 and N9. These networks are in fact subnetworks of the overall configuration, called network N10.

We note that the coloured switches in Figure 4 are condo-switches, that is, for each of these switches, different ports belong to different networks. For instance, switch S1 has two ports belonging to network N1 (connected to the two terminals H1 and H3), and two ports that are connected to links that belong to network N2. Similarly, switch S2 is a condo-switch belonging to networks N2, N3, N4 and N5, while switch S4 interconnects network N6 (and the subnetwork N3) with network N7.

3.2. Management functions provided by a network

The reason for introducing a hierarchical structure, as described for the example above, is to provide a framework for subdividing the management functionality between different autonomous domains. It is assumed that each (sub-) network is an autonomous management domain. The functions to be provided by a network (at any level of the hierarchy) are the following.

- To provide information for routing connections between the different sub-networks and/or the external condo-switches through which the network is connected to other peer networks.
- To accept requests for reserving bandwidth or lightpaths along the routes identified through the routing functionality of point 1.
- To provide a directory for available lightpaths for the following categories of connections: (a) between external condo-switches, (b) between external condo-switches of subnetworks, and (c) between an external condo-switch and an external condo-switch of a subnetwork.

It is important to note that a given network must provide these functions without the knowledge of the internal structure of its subnetworks. It may, however, use the functions provided by its subnetworks (which are of the same nature as those described above).

Given this hierarchical structure, the establishment of an end-to-end connection between two terminals will in general involve several networks. For the example of Figure 4, for instance, the establishment of an end-to-end connection between the terminals H1 and H2 may proceed as follows: Since the two terminals are located within the networks N6 and N8, respectively, the parent network of these two networks, network N10, will

determine that the route should include a segment from switch S4 through network N7 to switch S6. The network S8 will determine the route from S6 to the terminal H2 which resides directly in this network, and network N6 would be responsible for finding a route from the terminal H1 to the switch S4. This will be done in two steps: first a segment from S2 (the chosen external condo-switch of the subnetwork N2 which contains H1) to S4, and then a segment from H1 to S2. The former segment can be obtained from the subnetwork N3 by requesting a route between its external switches S2 and S4; and the latter segment will become the responsibility of the subnetwork N2.

To facilitate this kind of hierarchical routing, we suggest to introduce a hierarchical naming scheme for networks, switches and terminals based on the hierarchical structure of the networks within which they reside. For instance, the addresses of the terminals H1 and H2 would be the following, assuming that N10 is at the highest hierarchical level: address of H1 = root/N10/N6/N2/N1/H1; address of H2 = root/N10/N8/H2.

3.3. Definition of hierarchical networks with condo-switches

We give in the following more precise definitions of the concepts (written in **bold**) that were informally introduced above.

Switch: A switch has several **ports** (input, output or both-way). Each port (assuming that it is used in the given configuration) is connected to a network. A switch can establish **cross-connections** between different ports.

Network: A network has a number of **external switches**, **internal switches** and **subnetworks**. A switch S is an external switch of network N if at least one port of S is connected to N or a subnetwork of N, and at least one other port of S is connected to the parent network or another network N' or a subnetwork of N', where N' is a peer of network N in the network hierarchy. A switch S is an internal switch of network N if it is not an external switch of N, but it is an external switch of at least one of its subnetworks. A **connection** can be established within a network. We distinguish the following kinds of connections:

- **External connection**: a connection between two external switches of the network.
- **Internal connection**: a connection between two internal switches of the network.
- **Semi-external connection**: a connection between an internal switch and an external switch of the network.

Special cases of networks: We say that a network is a normal network if it contains at least one subnetwork. There are also "primitive networks" that do not contain subnetworks (and therefore do not need the management functionality discussed above); examples are point-to-point **links**, or broadcast (multi-point) network, such as wireless Ethernet. All the links shown as black lines in Figures 3 and 4 are such "primitive networks".

Special cases of switches: We may distinguish the following types of switches: (a) normal cross-connects (often assumed to be non-blocking), (b) add-drop multiplexers, (c) **terminal devices** that normally have only a single port (although multi-homed hosts may be considered for reliability or performance reasons), and (d) distributed virtual switches that may be implemented in the form of several (distributed) physical switches that are interconnected by some network (for which these switches are external switches).

End-to-end connection: An end-to-end connection is a sequence of network connections (external, internal and/or semi-external) at different levels of the network hierarchy that are interconnected by cross-connections provided by intermediate switches. For instance, the end-to-end connection from terminal H1 to terminal H2 in Figure 4 consists of the following connection segments: link from H1 to S1 (semi-external connection in network N1), link from S1 to S2 (semi-external connection in N2), connection from S2 to S4 (external connection of N3), connection from S4 to S6 (internal connection in N10), link from S6 to S7 (semi-external connection in N8), and link from S7 to H2 (internal connection in N8). These segments are interconnected by cross-connections in the switches S1, S2, S4, S6 and S7.

The special case of a distributed virtual switch, mentioned under point (d) above, appears to be quite similar to a network. However, there are some important differences between a network and a switch at the conceptual level, as indicated in Figure 5. Externally, a network is characterized by its external switches and the external connections it can establish between them. A switch is characterized by its ports and the cross-connections it can establish between them.

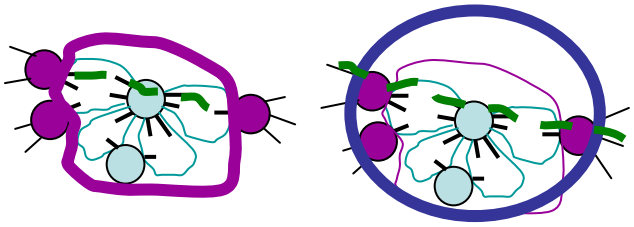

Figure 5: Left: a network with three external switches and one external connection between two external nodes. **Right:** a switch with seven ports and a cross-connection between two of its ports

3.4. Service interface provided by a network

In order to allow for end-to-end connection establishment throughout a configuration with hierarchical inter-domain structure, each network has to provide certain functions through one or several service access points (SAP). For simplicity, we assume in the following that each network has a single SAP which is identified by a URL, such as "uclp.uottawa.ca" for a hypothetical SAP of the UCLP network administered by the University of Ottawa. The following functions should be provided:

- Find a route between two given subnetworks: The resulting route is given in the form of an alternating sequence of switches and subnetworks. For instance, the result of a request for a route between networks N8 and N9, given to the SAP of network N10 (see Figure 4), would be the sequence "S8, N6, S4, N7, S6". This function includes the selection of the end-points of the route; in our example, the external switches S8 and S6 of the source and destination networks, respectively.
- Find a route from a given external switch to a given subnetwork (similarly).
- Find a route between two given external switches (similarly). For example, request the SAP of network N6 to find a route from S8 to S4, may result in "S8, N5, S2, N3, S4".
- Find and reserve a lightpath along a given route within the network.
- Advertise a given lightpath as available.
- Find an available lightpath along a given route within the network.
- Administrator functions for managing the configuration, access rights and accounting.

The switches are in some sense considered as independent entities, since many of them are condo-switches and therefore belong to more than a single domain. We therefore assume that each switch also has a SAP identified by a URL. Besides administrator functions, the switch SAP only provides the following function:

- Set or unset a cross-connection between two given ports.

4. Management issues

4.1. Connection establishment procedure

Given the services described in Section 3.4, the following procedure may be executed by a user who wants to establish an end-to-end connection between two given terminals. In the following description, we use the network example of Figure 4. The procedure consists of two phases: (a) Finding a route, and (b) reserving a lightpath on that route and setting up the corresponding cross-connections in the intermediate switches. The end-to-end route will be represented by a sequence of partial routes such that each partial route corresponds to exactly one network. For the example of an end-to-end connection from H1 to H2 we expect to obtain an end-to-end route which is composed out of the segments described in the paragraph entitled "End-to-end connection" in Section 3.3.

The end-to-end route can be found by following the following steps:

1. Determining the highest-level network involved: We assume here that a hierarchical naming scheme for switches (and in particular for terminal devices) as explained in Section 3.2 is used. The highest-level network is identified by the common prefix of the names of the two end-points. For our example, these two names are "root/N10/N6/N2/N1/H1" and "root/N10/N8/H2", and the common prefix is "root/N10" which identifies network N10.
2. Finding the route segment in the highest-level network: This route segment is obtained by invoking the "find route between subnetworks" function on the SAP of the highest-level network. This function will also provide external switches at the next-lower subnetworks which will play the role of source and destination for this route segment. In our example, the names of the two end-points indicate that the subnetworks of interest are N6 and N8. The obtained route segment will be "S4, N7, S6".
3. For each of the two sub-networks identified in Step 2, recursively go down the hierarchy of subnetworks until the end-point is reached and establish a route segment for each subnetwork which represents a semi-external route for that subnetwork. Starting with subnetwork N6 in our example, this leads to the following operations and resulting route segments: (1) request the SAP of N6 to "find route between external switch S4 and subnetwork N2" resulting in the segment "S4, N3, S2" and the selection of the switch S2; (2) request the SAP of N2 to "find route between external switch S2 and subnetwork N1" resulting in the segment "S2, L2, S1" and the selection of switch S1; (3) request the SAP of N1 to "find route between external switch S1 and switch (terminal) H1" resulting in the route segment "S1, L2, H1".

During the second phase, the subnetworks involved in the route segments obtained during the first phase will be requested to find and reserve a lightpath along their corresponding route segment. In the case of wavelength division multiplexing (WDM) without wavelength conversion, the constraint of a uniform wavelength along

the whole lightpath must be enforced during this phase (see for instance [5] and [6]). The last operations during this phase are the set-up of the appropriate cross-connections on all the intermediate switches.

4.2. Finding the service access points of networks

The procedure described in Section 4.1 requires accessing the SAPs of various networks and a user owning a given terminal may only be registered to the local subnetwork and may not know the URLs of the other networks and switches involved in the establishment of an end-to-end connection from his terminal. The following approach may be used to solve this problem.

Let us assume that the SAP of each network provides the following additional functions:

1. Global network identification: this function returns the global hierarchical name of the network (e.g. "root/N10/N6/N2" for network N2) and the URL's of all the higher-level networks that appear in the global network name. Optionally, it may also return a chain of authentication certificates which authenticate the network.
2. Identification of subnetworks: this function returns a list containing the name and URL of all subnetworks of the network.
3. Identification of external and internal switches: this function returns a list of switches visible in the scope of this network and the URL of the SAP of these switches. Note: The URL of the switches may also be provided as an attribute of the switches that appear in route segments that are returned by the "find route" functions described in Section 3.4.

For configuration management purposes, we also assume that each network provides a function for registering a new subnetwork. Through this function, the subnetwork will learn about its own global network identification. A special case of this function is the establishment of a new link between switches of the given network; in this case, no SAP will be created for the link, since it does not represent a functional subnetwork in the network hierarchy. Similarly, a function for registering a new switch within the network must be provided.

4.3. Access control

We do not pretend to cover the topic of access control in this paper in detail. The purpose of this section is simply to highlight the issues and point to some possible directions for finding solutions.

In our existing UCLP system [3] with a two-level hierarchy of "federations" within a "network", we have identified two user roles: (a) administrators, and (b) normal users. Normal user can find routes, reserve end-to-end connections and advertise as available partial connection segments, called Lightpath Object (LPO). Administrative users can also perform configuration management functions. Further studies have shown that a pure role-based access control model appears to be too limiting for managing access rights in the context of multi-domain UCLP. A discretionary access model based on certificates has therefore been proposed [8].

It is not clear what kind of access control model would be appropriate for a hierarchical multi-domain system as proposed in this paper. For finding an appropriate model, one may also consider other distributed applications, such as peer-to-peer computing in order to establish an access control framework that is not only suitable for UCLP, but also for other kinds of distributed applications [9]. In this context, the model proposed with the Astrolab system [10] appears to be interesting. It suggests that (a) each network would have a certification authority (CA) that issues signed certificates for user authentication and for assigning roles and/or rights to certain users; (b) each network will also have its own access rights policies that determine how to interpret the certificates provided by users; and (c) the policies of a subnetwork may override the policies of its parent networks. Secure authentication of networks, switches and users could be provided through a public-private key infrastructure (PKI) which may have a hierarchical structure similar to the network hierarchy.

5. Conclusions

This paper presents a general hierarchical approach to inter-domain routing and network management, especially intended for user-controlled lightpath provisioning (UCLP). The structuring of networks into subnetworks at several levels of the domain hierarchy provides an architecture for distributed processing of network management functions that is very scalable. Each network, as well as each subnetwork, represents an autonomous domain that communications with its peer, child and parent networks through standard interfaces provided at the service access points of these networks. A special feature of this architecture is the natural support for condo-switches, that are switches with ports that belong to different networks, i.e. different administrative domains.

Procedures for finding routes in this hierarchical network configuration are described in detail, as well as procedures for establishing lightpaths along such routes.

Although the context of this discussion are condominium networks with condo-switches providing lightpaths to end-users, the proposed architecture appears also to be quite suited for other kinds of networks. Different types of "lightpaths" may be considered. While originally the term "lightpath" was developed in the context of WDM and

wavelength routing in optical networks, the term may also be used to denote a label-switched path in MPLS. And the management of MPLS flows, at the inter-domain level, may very well use the hierarchical structure proposed in this paper.

We note that one aim of the introduction of our hierarchical management framework was the definition of a standard protocol by which the management systems of different autonomous domains could communicate with one another in order to provide some global functions, such as end-to-end connection establishment through heterogeneous domains. The network services defined in Section 3.4 represent an **abstract** definition of such a protocol. This definition is given in terms of a set of remote procedure calls (RPC) that should be supported by each autonomous domain, that is, each network. While the definition given in this paper only defines the semantics of these standard services, a second step towards the establishment of a standard protocol is the selection of the communication syntax to be used.

Based on our experience with our existing UCLP system [3], two approaches to defining the syntax for such a communication standard come immediately to mind: (a) defining a Java interface corresponding to the services defined in Section 3.4 and use Java RMI to access the SAPs of the networks, possibly using Jini for finding these SAPs, and (b) defining a Web Service [11] corresponding to the services defined in Section 3.4 , describing it in WSDL and using the XML-encoded SOAP protocol to access the SAPs of the networks.

The following issues require further study: (1) defining an appropriate framework for the management and control of access rights and (2) dealing with network configurations that are not hierarchical.

Acknowledgements

I would like to thank the people of the CRC-UofO UCLP team, Scott Campbell, Jun Chen, Michel Savoie, Jing Wu, Hanxi Zhang, and particularly Mathieu Lemay for many interesting discussion in connection with UCLP and hierarchical management structures.

References:

[1] "Frequently Asked Questions about Customer Owned Dark Fiber, Condominimum Fiber, Community and Municipal Fiber Networks", http://www.canarie.ca/ canet4/library/customer.html, March 2002.

[2] T. Anderson and J. Buerkle. “Requirements for the Dynamic Partitioning of Switching Elements”, IETF RFC 3532, May 2003.

[3] 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.

[4] R. Boutaba and A. Polyrakis. “Projecting Advanced Enterprise Network and Service Management to Active Networks.” IEEE Network, Vol.16, No.1, Jan./Feb. 2002, pp.28-33.

[5] G. M. Bernstein, V. Sharma and L. Ong. “Interdomain Optical Routing”, Journal of Optical Networking, Vol.1, No.2, Feb. 2002, pp.80-92.

[6] M. G. Khair, An Implementation Approach for an Inter-Domain Routing Protocol for DWDM, Master Thesis, SITE, University of Ottawa, March 2004.

[7] J. Wu, H. Zhang, S. Campbell, M. Savoie, G. v. Bochmann and B. St.Arnaud, A Grid oriented lightpath provisioning system, Proc. Globecom Workshop on "High Performance Global Grid Networks", 2004.

[8] J. Chen, A Distributed Networtk Management System for User-Controlled Lightpath Provisioning and its Security Requirements, Master Thesis, SITE, University of Ottawa, December 2004.

[9] I. Foster, C. Kesselman, S. Tuecke, The anatomy of the Grid: Enabling scalable virtual organizations, Intern. Journal on Supercomputer Applications, 15 (3), 2001.

[10] R. VAN RENESSE, K. P. BIRMAN, and W. VOGELS, Astrolabe: A Robust and Scalable Technology for Distributed System Monitoring, Management, and Data Mining, ACM Transactions on Computer Systems, Vol. 21, No. 2, May 2003, Pages 164–206.

[11] B. St.Arnaud, A. Bjerring, O. Cherkaoui, R. Boutaba, M. Potts, W. Hong, Web Services architecture for user control and management of optical Internet networks, Proc. Of the IEEE, Vo. 92, No. 9, pp. 1490-1500, August 2004.