INTER-AREA ROUTING, PATH SELECTION AND TRAFFIC ENGINEERING

How to meet Quality of Service requirements for traffic routed across MPLS and optical network boundaries

Version 1: November 2003

Ben Wright
Customer Services Consultant, Data Connection
Ben.Wright@dataconnection.com

Data Connection Limited
100 Church Street
Enfield
EN2 6BQ
United Kingdom
http://www.dataconnection.com

Table of Contents

1 Introduction

This paper explains and summarizes the various mechanisms that are being discussed by the IETF, OIF and ITU standards bodies for routing data across multiple areas. The focus of this paper is on methodologies that can select routes, spanning multiple areas, which can then be used for MPLS transport connections. However, many of the concepts discussed here are also applicable to other environments (including ATM networks).

For those new to the subject of Inter-area routing and path selection (but with some familiarity with Traffic Engineering and MPLS in general) the paper is intended to provide an introduction to the problems inherent in routing data across multiple networks. Readers familiar with these issues will also find this paper useful as it summarizes the ongoing standards body developments that relate to this topic and provides examples illustrating the ideas discussed there. These readers will probably find Chapters 4-9 most useful.

This document splits into three distinct sections.

- The first section, Chapters 2-3, introduces the concepts that are important to this paper, and illustrates why it is relatively hard to select a path for data that spans multiple networks, in comparison to selecting a path that is contained within a single area.
- The second section, Chapters 4-7, looks at one mechanism that carriers may be currently using to solve this problem and then introduces other methodologies under discussion within the standards bodies.
- The third section, Chapters 8-10, summarizes the solutions discussed above and the key issues facing carriers and equipment manufacturers looking to provide this function.

This paper assumes the reader is familiar with the basic principles of Traffic Engineering and MPLS. The Data Connection (DCL) white papers “MPLS Traffic Engineering: A choice of Signalling protocols” and “MPLS in Optical Networks” provide an introduction to these topics.

2 The Problem

The precise definition of an area is fundamental to understanding the issues we will be discussing, so we start by defining this concept.

2.1 What is an area?

An area, in the context of this paper, is a set of nodes and links that meet the both of the following requirements.

- The area has a single administrator.
- The area runs some kind of Interior Gateway Protocol (IGP). The scope of the IGP defines the area's boundary – full routing information about the area's nodes is only flooded to those nodes within the area.

This aim of this paper is to examine methodologies that allow nodes to select routes that meet a specified set of Traffic Engineering constraints. Therefore, we assume that the IGP is able to carry information that can be used to achieve this goal—for example, the IGP may disseminate information about the bandwidth available on each link throughout the area. OSPF-TE and IS-IS-TE, two of the most commonly used IGPs, are capable of carrying this type of information.

Section 2.5defines other terminology used in this paper.

2.2 Why divide a network into areas?

Networks are divided into different areas for a variety of reasons.

2.2.1 Inter-Carrier interworking

If multiple carriers operate within the network, a single organisation will not be responsible for administering the entire network. Therefore, by the above definition, the network must be split into multiple areas. As we shall see, setting up LSPs that span different carriers is likely to be harder than setting up connections that span areas within a single carrier.

However, even within a single carrier there are several reasons to divide the network into multiple areas.

2.2.2 Reasons of scale

If a carrier wanted to run a single IGP over their entire network, this would mean each node within the area would need to maintain information about all the links in their network. As the size of the network increases, the link state database gets larger and

- it is an increasingly onerous task for nodes to maintain the database
- it will take a node longer to sift through its database before a route can be calculated.

Even for carriers who presently maintain a single area, as more and more traffic traverses their network, the pressure to add new nodes and links to cope with this demand will also increase. Therefore these scalability concerns are likely to become more important in the future.

2.2.3 Administrative convenience

It may be sensible to allow different parts of a carrier's network to be administered by different parts of the organisation because the administrative burden may be too great for one group to administer the entire network. For example, splitting the network into areas based on geographical grounds, with a different group responsible for administering each area, may be a more efficient way of managing the network.

2.2.4 Business structure

The carrier network may be split into multiple areas to reflect the carrier's business structure. For example, if a carrier has acquired or merged with another carrier, it may choose to contain the newly acquired network in a separate area from the rest of the network.

2.2.5 Compatibility reasons

Network elements may support different routing protocols and it may be impossible to run a single IGP across the whole carrier network. Alternatively, a carrier may have deployed network elements from different vendors. By dividing the network into areas, a carrier can limit interoperability between different vendor devices to well-defined interfaces.

Given these reasons, it is highly likely that a given network will be split into multiple areas as shown in Figure 1 below.   A separate IGP will run in each area and a summary of each areas topology will be passed across the Inter-area link (the dotted line).

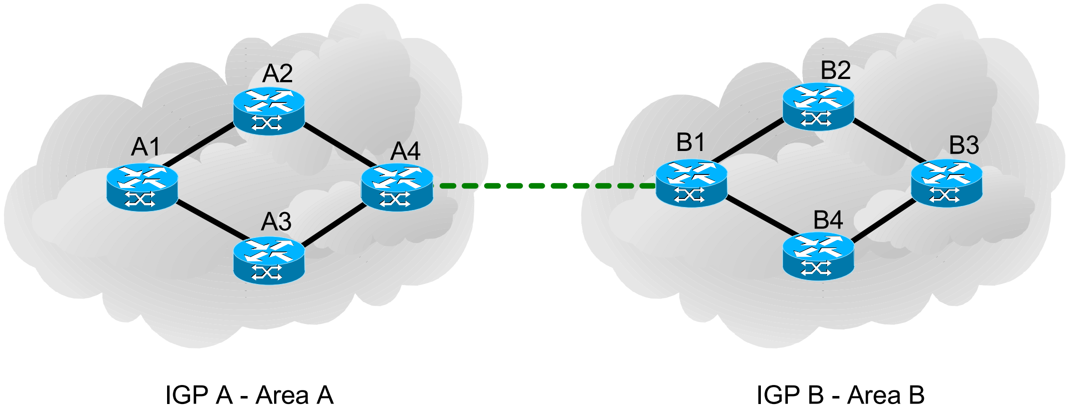

Figure 1 – A network split into 2 areas

2.3 The Problem

Increasingly, networks are being used to transport new types of data such as voice or video traffic and carriers are deploying MPLS networks to make this happen. Their aim is to meet the requirements of end-users, and guarantee that these new types of traffic can be transmitted

- reliably—for example, in order to meet end-user requirements for service availability, traffic may need to be protected by a diverse backup connection
- in accordance with the Quality of Service (QoS) associated with the type of data (for example, minimal and consistent delay for voice traffic).

Furthermore, carriers will require that, wherever possible, data is transmitted following a route which minimizes the impact on the performance of the carrier's overall network.

Increasingly, carriers are deploying MPLS networks to ensure that these requirements can be met. However, in order to set up an LSP that can meet these requirements, carriers must be able to calculate a suitable route (or sequence of routes) across the network. This paper focuses on how these routes can be calculated.

When the ingress point and egress point for an LSP are in the same area, there are well-understood methodologies for selecting such a route. However, as we shall see, extending this service across different areas or even different carrier networks is much more difficult.

Fundamentally, though, end-users do not recognize area boundaries. They still expect the same quality of service for data traffic routed across area borders as they do for local traffic. Therefore, carriers and equipment manufacturers are working together to produce devices that can make this happen.

2.4 Document structure

In Chapter 3, we explain why the methodologies used in Intra-area routing and path selection are not appropriate to networks divided into multiple areas, and consider the issues that make Inter-area and Intra-carrier routing intrinsically harder to accomplish.

In Chapter 4, we give an example of one way that carriers may be solving the problem today, and some of the limitations of that approach.

Chapter 5 introduces the concept of hierarchical routing, which illustrates another way in which the problem could be solved.

In Chapter 6, we look at Path Computation Servers—one way to solve this problem in Inter-area scenarios and Inter-carrier scenarios. We will focus on examples suggested in a recent IETF draft that indicates how this can be accomplished in Intra-carrier networks running IGPs such as OSPF or IS-IS.

In Chapter 7, we consider the recent developments in the OIF and ITU to allow Inter-area routing in an optical context. These developments build on the hierarchical routing concepts discussed in the Chapter 5.

Chapter 8 defines the requirements for Inter-carrier routing that have been laid out by the various standards bodies.

Chapter 9 summarizes the solutions discussed above and Chapter 10 provides references to the drafts and specifications about this subject.

2.5 Terminology

Before discussing the issues involved further, we first need to define the terms that we will use throughout this paper.

Path Vector Routing Protocol – A Path Vector routing protocol only distributes enough information throughout the network to allow a node to select a route to a particular destination based solely on the administrative weight of reaching that destination. Traffic Engineering information, such as available bandwidth or delay, cannot be disseminated by a Path Vector Routing Protocol. BGP and RIP are examples of Path Vector Routing Protocols.

Link State Routing Protocol – A Link State Routing Protocol distributes information about individual properties of the links in the area throughout the network. This allows nodes to build a database of link information—the Link State Database (LSDB). With this database, nodes can take into account the suitability of each link when calculating a route through the network. For example, a Link State Routing Protocol might distribute information about the bandwidth available on a particular link, allowing nodes to select a route to the destination that guarantees the bandwidth that a voice user requires. OSPF, ISIS and PNNI are examples of Link State Routing Protocols

Label Switched Path (LSP) - An LSP is an MPLS connection that allows data traffic to be forwarded through the network. For the purposes of this paper, TDM connections that can be programmed by GMPLS signaling protocols are classed as LSPs.

Traffic Engineering (TE) – Traffic Engineering is a branch of Network Engineering associated with optimising the performance of a network without altering its topology.

Open Shortest Path First (OSPF) – OSPF is a link state routing protocol that has standardized extensions which can be used to distribute Traffic Engineering information around the network.

Constrained Shortest Path First (CSPF) – CSPF is a class of algorithms that can be used to calculate a route that satisfies certain constraints. Typically these constraints are Traffic Engineering constraints such as bandwidth or delay.

Intermediate System to Intermediate System (IS-IS) – IS-IS is another example of a link state routing protocol that can also carry Traffic Engineering information.

Autonomous System (AS) – An AS is a collection of areas under the administrative control of a single operator.

Area Edge Point (AEP) – An AEP is a node in an area that is directly connected to a node in another area.

Area Border Router (ABR) – An ABR is a node in an OSPF network that is connected to multiple areas. An ABR is an example of an AEP.

Autonomous System Border Route (ASBRs) – A node in a network that is connected to multiple ASes.

Traffic Tunnel – A connection through a network (possibly spanning multiple areas) through which data or voice traffic can be passed. An MPLS LSP is an example of a traffic tunnel.

Head-End Area – The Head-End area contains the ingress point of a traffic tunnel or LSP.

Tail-End Area – The Tail-End Area contains the egress point of a traffic tunnel or LSP.

Multi-protocol Label Switching (MPLS) – MPLS allows packets to be forwarded opaquely, solely on the basis of a tag or label at the header of the packet. In an optical network, data is forwarded based on the wavelength of light used to carry it, or the timeslot in which it arrives. Signaling protocols such as RSVP are used to set up these connections.

Generalized MPLS (GMPLS) – GMPLS defines some extensions to the MPLS signaling protocols that can instruct nodes to program optical cross-connects for transmitting data based on, for example, the wavelength of light used, or the timeslot it is received in.

Quality of Service (QoS) – QoS refers to the capability of a network to provide a specified level of service appropriate to the type of data traffic that is being transmitted. For example, Voice calls may require different QoS to web browser traffic.

3 Inter-Carrier vs Inter-Area vs Intra-Area

In this section, we will consider will issues involved in selecting a suitable path for MPLS Traffic Engineering that is either

- Intra-area (the entire path is limited to a single area)
- Inter-area (the path spans multiple areas all owned by the same carrier)
- Inter-carrier (the path spans multiple carriers)

Examples of these types of path are shown in Figure 2. As we shall see, selecting paths that span multiple areas is inherently more difficult than the Intra-area case.

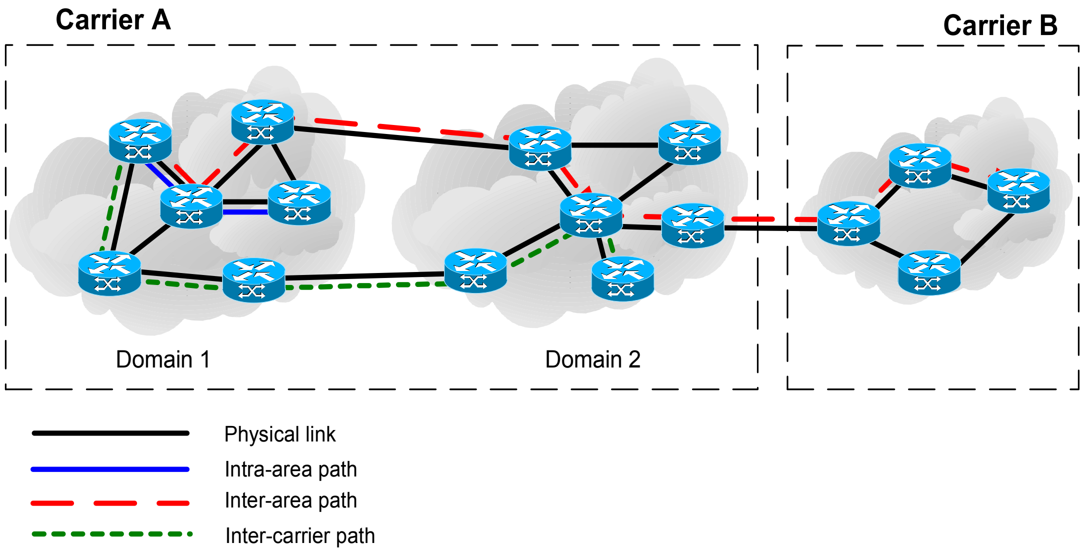

Figure 2 – Intra-area, Inter-area and Inter-carrier paths.

3.1 Intra-area path selection

It is relatively straightforward to select a path for an LSP if the ingress and egress point are within the same area. The IGP disseminates information about the area's nodes and links to all nodes within this area. For example, the IGP may disseminate the following information about each link.

- Available bandwidth.
- Administrative cost.
- Delay or latency.
- Any other characteristics of the link.

As LSPs are established and torn-down, Traffic Engineering characteristics associated with the link, such as the bandwidth available, will change. The IGP is therefore continuously updating the link state information. How this happens and how often it happens is beyond the scope of this paper.

The information disseminated by the IGP allows each node to build up a database of every link in its local area and the appropriate Traffic Engineering properties associated with them. If a node receives a request to set up an LSP that must meet a certain set of TE constraints, then this database can be used to calculate a suitable, complete, explicit least-cost route through the area to the specified egress point for the LSP. If required, the ingress point can calculate a diverse backup route (a route that does not use the same link or node) at the same time to provide additional reliability for the connection.

In the Intra-area case, the ingress point for the LSP can calculate the path the LSP should follow because it has full knowledge of all the nodes and links across the network that LSP will traverse. However, in Inter-area and Inter-carrier cases this is not true—the ingress point for the LSP will not know the details of every link and node that could be traversed.

3.2 Inter-area path selection

Unlike the Intra-area case, if the LSP must traverse multiple areas, there are some significant complications. Since the IGP only runs within a given area, it does not distribute topology information about all the links the LSP must traverse to the node initiating LSP set up. The information provided by the IGP can only be used to calculate a route to the Area Edge Point (AEP) and no further.

Without this knowledge a node cannot reliably choose the "best" route for an LSP to take (that is the least cost route that satisfies the constraints associated with the LSP). Instead, the node must try to approximate this function. The goal is to select a route that satisfies the specified constraints without incurring an unacceptable cost.

Broadly speaking, there are two ways in which a node can attempt to provide this function.

3.2.1 "Push" mechanisms

In these cases, an Inter-area routing protocol pushes down a summary of information about all other areas to the nodes in each area – how this summary is put together is discussed in Chapter 5. In this way, all nodes obtain a limited view of the network outside of their own area, and, therefore, they can select a sensible area edge-point that the LSP could traverse. Examples of this mechanism include PNNI and the OSPF-BGP model we will discuss in Chapter 4.

3.2.2 “Pull” mechanisms

When using pull mechanisms, nodes do not need to maintain a view of the rest of the network. Instead, when they receive an LSP set up request, they query a remote entity with a better or different view of the network. On the response to this query, the remote entity provides the node with a route to the destination and the node can the set up the LSP to follow this path. We shall discuss some examples of these mechanisms in Chapter 6.

3.2.3 Other Inter-area path selection considerations

There are further complications when setting up LSPs that span different areas. For example, in order to ensure that an LSP is protected against node or link failure by provisioning, carriers may want to provision a separate LSP to provide a backup connection. However, since it is possible that no node along the entire path has complete knowledge of the route an LSP should take, calculating routes that are diverse (do not use the same link or node) is much harder than in the Intra-area case.

3.3 Inter-carrier path selection

In order to meet customer requirements, carriers may need to provision resources across another carrier’s network and when doing this there are the following additional difficult issues to be resolved.

3.3.1 Issues of trust

When different carriers administer networks, it is important that a carrier does not disclose any sensitive internal information about the topology of its network to neighbouring carriers who may be competitors, as this could be used to gain competitive advantage.

For example, carriers may be reluctant to disclose details of the bandwidth available within their network or whether their topology allows them to protect a particular LSP. This makes it much harder to design an efficient path computation algorithm that can calculate routes spanning multiple carriers. Also, if a carrier discloses too little information, traffic may not be routed over their network—something that could have commercial implications.

Throughout this paper, we distinguish the Inter-area case from the Inter-carrier case, by saying that the issues of trust described above are applicable to the Inter-carrier case but not the Inter-area case. While this may not be true for all networks, in subsequent chapters we use this broad characterization as a useful tool for explaining the concepts important to Inter-area and Inter-carrier path selection.

3.3.2 Protection issues, standardization and SRLGs

Setting up a fully protected, end-to-end LSP across multiple carriers is particularly difficult. Ensuring that two LSPs use links in different areas or even different carrier networks does not necessarily guarantee that the LSP is satisfactorily protected. Different links may use the same physical conduit—for example different optical fibers may share the same cable. These links are likely to share the same fate—both links could be broken by a single event and therefore these links should not be used to protect one another.

To solve this, carriers may introduce the idea of Shared Risk Link Groups (SRLGs) within their networks. In this case, different links within the carrier network may be grouped together as belonging to a particular SRLG. Then, if an LSP is set up along a path using links belonging to a specific SRLG, it could set up a backup LSP to use a different SRLG as a form of protection.

However, links owned by different carriers can also share the same physical conduit. Therefore, the concept of SRLG must be extended to the Inter-carrier environment. For this to happen, nodes across all areas must be able to understand which SRLG is being referred to, which means that SRLG identifiers must be standardized across the network. This may be fairly straightforward in some Intra-carrier scenarios but is very difficult in the Inter-carrier environment.

As we shall see, no Inter-carrier routing strategy that meets these additional needs to the satisfaction of all concerned has been developed yet.

The key question is then; how do you compute paths across multiple areas and/or carriers' networks that satisfy the specified constraints?

4 Existing Inter-Area Strategies

Currently, some carriers use a combination of BGP and an IGP, such as OSPF-TE, to provide Inter-area routing functionality. With this model, BGP is an Inter-Autonomous System routing protocol and provides BGP capable nodes with information about the carrier's network as a whole. In the example below, we assume that OSPF-TE runs as an IGP in each area and disseminates Link State Traffic Engineering information about links in the area throughout the area.

In this chapter we examine the limitations associated with this model.

4.1 BGP-OSPF example

BGP is a Path Vector Routing Protocol and therefore cannot be used to disseminate Traffic Engineering information throughout the network. Instead, BGP can only provide reachability information—information about the set of addresses that are reachable by routes through a particular area. If a node must set up an LSP to a destination outside of its local area, this information can be used to determine the ABRs (Area Border Routers) that could be used to reach the destination, but it cannot be used to discover any information about the TE capabilities of the routes to the destination that go through each ABR.

To understand how this BGP-OSPF model could be implemented, consider the topology in Figure 3.

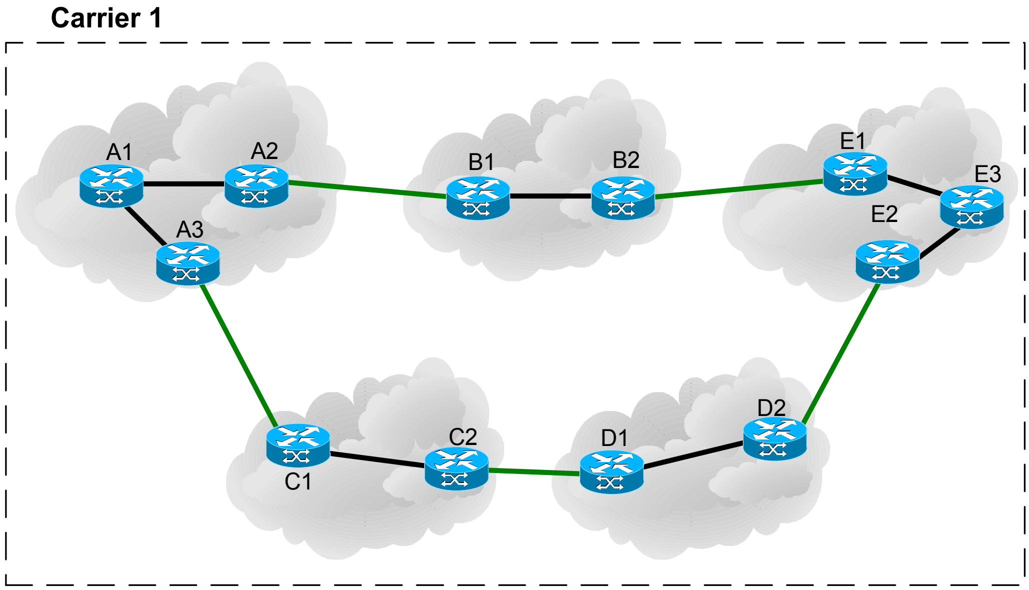

Figure 3 – Sample Inter-area scenario

Suppose node A1 in area A wants to set up an LSP to E3 that guarantees a particular level of bandwidth. If reachability information has been distributed across the area boundaries, A1 can select an Area Edge Point (AEP) through which E3 could be reached (in this example A2). Once this is done, A1 can then calculate a route to A2 through area A (the head-end area) that satisfies the constraints associated with the LSP. At this point, A1 then initiates LSP set up by sending a set up request through the head-end area to A2.

When the request reaches A2, it looks up the address of E3 in its routing table, realizes that B1 is the next hop, and forwards the request on to B1. B1 is then responsible for calculating a path to B2 (the IGP could be used to ensure that this path meets the end-user's service requirements, if possible). Similarly, on receipt of the LSP set up request, E1 calculates a route to the destination, E3.

However, there may not be any route through area B that meets the LSP constraints—and the attempt to set up the LSP would fail when the set up request (for example an RSVP Path message) reached B1. Similarly, it may be that there are no routes that pass through router E1 that could satisfy the Quality of Service Requirements for the LSP, for example if there is no available bandwidth on the link E1 – E3. In this case, the set up request would reach E1 before failing.

4.2 Crankback

To solve this, carriers can implement a crankback mechanism. If the LSP set up fails at E1, an error is passed back along the path that the LSP followed toward the ingress. On receipt of this error notification, other nodes on the path may attempt to recalculate a route to the destination area that takes a different route to E3. These nodes could then attempt to set up the LSP to follow this new route that might be able to support the LSPs requirements. In the above example, the error notification would have to be passed back all the way to the ingress point, A1, before the LSP could be re-routed.

In this way, the BGP-OSPF model can be seen as a trial and error approach for setting up the LSP.

4.3 Limitations

There are some significant limitations with this mechanism.

Firstly, if an LSP traverses a large number of areas, this approach could mean that it takes many attempts before an LSP can be successfully set up and many different routing calculations would need to be performed. This could result in the time taken to set up an LSP becoming unacceptably large. Therefore, to avoid this carriers may be required to over-provision their networks, to reduce the chance of LSP set up failing because resources for the connection were not available.

Furthermore, calculating diverse routes (for LSP protection) is extremely hard. Any node within the network will only know about the links within their local area, and the set of addresses that they can reach via certain ABRs. With this model, the easiest way to do this is to first set up an LSP; pass details of the route it took and the SRLGs it used back to the ingress point and then attempt to set up another LSP diverse from the returned SRLG. However, there is no standardized way to return SRLG information like this using the current MPLS signaling protocols. Further, this approach is not applicable to the Inter-carrier environment, since it mandates that low-level topology information about the route the first LSP takes through the entire (multiple carrier) network is returned to an ingress point. As discussed in the previous chapter, carriers will be unwilling to allow their low-level topology details to be disclosed in this way to other competing carriers.

If these limitations are acceptable to a carrier, this BGP model is a good way to solve the problems discussed in the previous chapters. However, if these restrictions are unacceptable, carriers have to provide ways to give the nodes in each area access to more information about the network than the reachability information that BGP provides. Hierarchical Link State Routing Protocols are one mechanism that can be used to try and provide this extra information.

5 Hierarchical Link State Routing

Hierarchical routing protocols provide a framework for aggregating and summarizing link state information relating to a particular area, and advertising it to nodes in other areas. The route is calculated in terms of the summarized topology advertised by the nodes running the hierarchical routing protocol.

We will consider in this chapter exactly how to represent the low level topology of a given area in a way that allows accurate Inter-area path computation based on this summarized topology information.

5.1 How hierarchical routing works

In a hierarchical system, we use the concept of levels to help aggregate properties of a particular area. Consider Figure 4, which is an example of a 2-level hierarchical network.

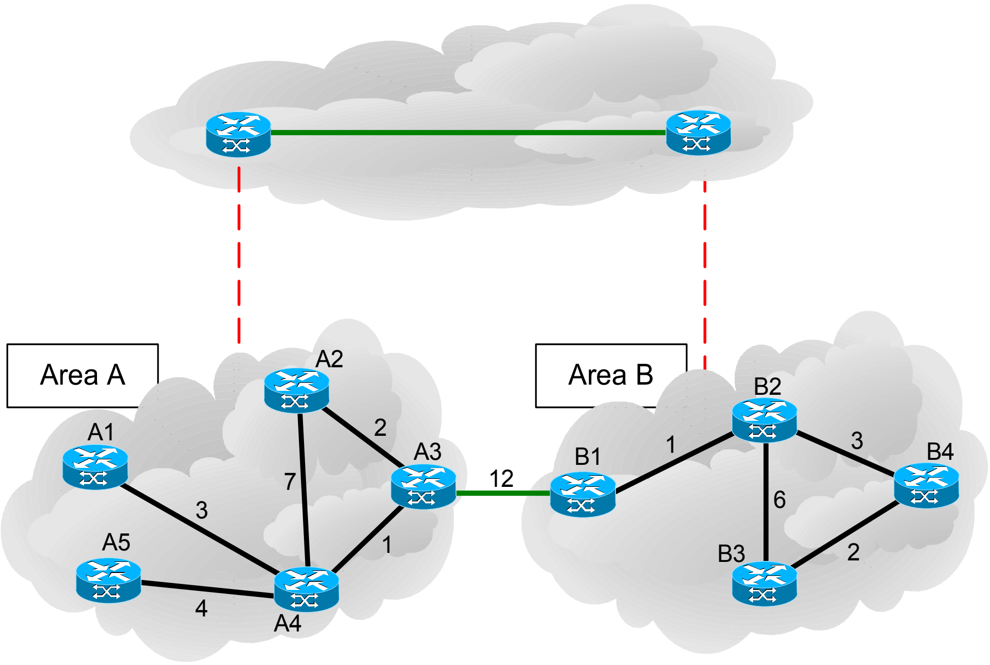

Figure 4 – Routing with two levels of hierarchy.

In this example, the physical nodes and links are at the lowest level of the hierarchy. The higher level consists of abstract nodes and links, which are a summarization of the lowest level topology. We will consider exactly how these areas are represented later in this chapter.

The numbers marked on the links in the diagram indicate the “cost” of traversing a particular link. In this example, cost is a single link metric incorporating traffic-engineering related features such as available bandwidth or delay. In a real network, representing a link by a single metric may be too simplistic, and more information about the characteristics of a link may need to be distributed. However, for simplicity, we consider a single metric-based system.

Every node in each area discovers Link State information about the links and nodes inside that area. These links are physical links, and so they should be viewed as being part of the lowest level.

At the higher level, each area is represented by an abstract node. If there is at least one physical link between two areas, then the corresponding abstract nodes have a virtual link between them. The set of abstract nodes at the higher level form a virtual area, and distribute information about each of the virtual links amongst themselves.

Each abstract node “feeds down” summarized information about these virtual links to the lower level nodes in the area it represents. From the perspective of the lower level nodes, there is now a link between node A3 and area B, which can be used to route traffic to that area.

5.2 Summarization

So what information should be included in the summary that is pushed down to the low-level nodes? On the one hand very little information could be included—in the example in the previous chapter, the nodes at lowest level of the hierarchy were only provided with reachability information by the BGP protocol. However, for some networks, this could cause significant problems when attempting to set up LSPs that guarantee bandwidth is available for traffic routed across multiple areas.

Alternatively, carriers can attempt to push a summary of the link state information relating to the other areas down to the lower level nodes. There is an important trade-off here—increasing the information that is included in the summary, will also increase the amount of information that the nodes at the lowest level of the topology must maintain.

As we shall see in the rest of this chapter, summarizing the Traffic Engineering capabilities of a whole area is a complex task.

5.3 Simple Node vs Complex Node Representation

Figure 4 provides an example of Simple Node Representation because each area is represented by a single node at the higher level of the hierarchy. However, this type of representation does not take into account differences in the Traffic Engineering constraints associated with different routes across an area. Consider the example in Figure 5.

In this case, determining the cost that should be associated with the virtual links such as the virtual link A-B is problematic. If we do not include the cost of the Intra-area links when calculating the cost of these virtual links then it appears that the cost of A-B is 16 units, A-D is 15 and B-C and D-C are both 10. Given this, if A1 wanted to calculate a route to C3 it would select the route A1-A2-D-C. At the lower level, this route is actually A1-A2-D1-D2-C2-C3 which has a cost of 32 units, which might be unacceptable. Also, in this example, the route A1-A2-B1-B2-C1-C3, which was overlooked, costs just 29 units.

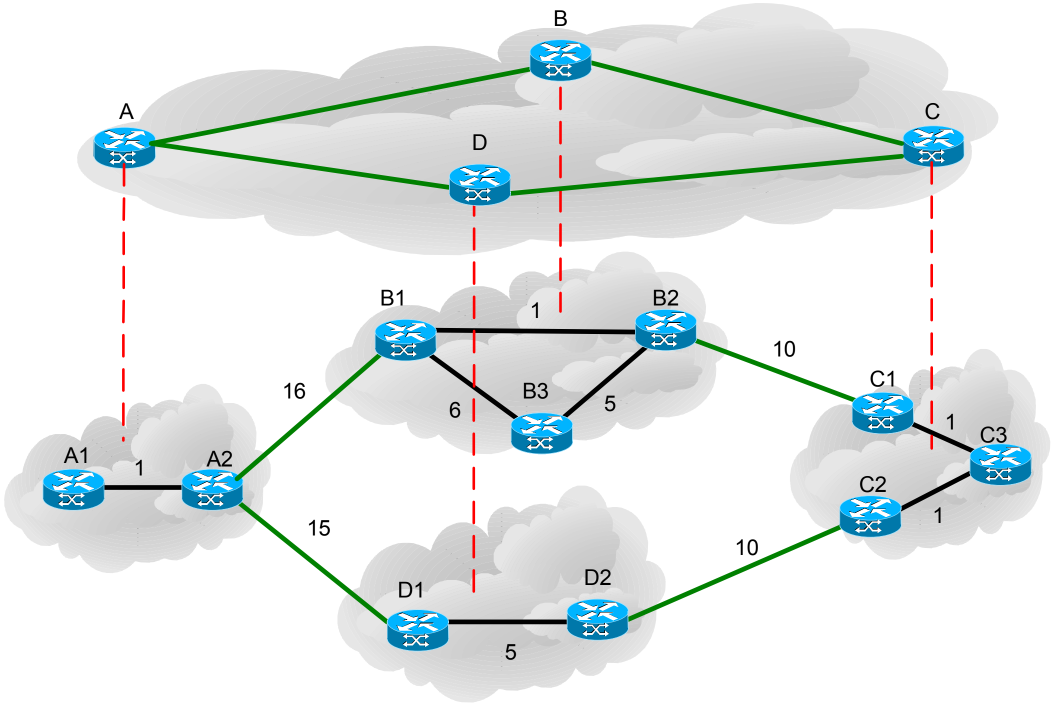

Figure 5 – Assigning Traffic Engineering parameters to virtual links

However, including the cost of the Intra-area links is not straightforward. A carrier could configure a fixed cost to traverse each abstract node in any direction but this may not represent the lowest level area topology. In the above example, if B1, B2 and B3 were all Area Edge Points, then the cost associated with traversing area B would be dependent on the direction in which the area is traversed. While B1 to B2 only has a cost of 1 unit, traversing the network from B1 to B3 costs at least 6 units.

Similarly, choosing a value to assign to a virtual link that indicates its bandwidth capabilities is problematic. There may be little bandwidth available for traffic traversing the area in one direction, but a large amount of bandwidth available for other directions. Representing an entire area as an abstract node, these Intra-area differences cannot be displayed in the routing protocol. To solve this problem, carriers may opt to use a different form of representation for each area – Complex Node Representation.

Using Complex Node Representation, the underlying area is not advertised merely as a single abstract node—but as a series of abstract nodes and links that illustrate, at the higher level, how the area can be traversed. Like any other links, these links can have properties such as available bandwidth, delay or cost associated with them. Exactly how a particular area is represented depends on the hierarchical routing protocol being used.

Associating multiple properties with these abstract links is not straightforward. Consider the network topology in Figure 6.

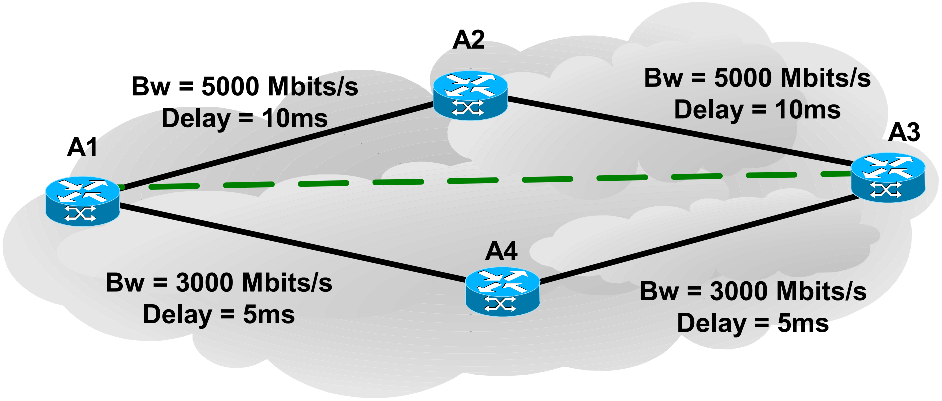

Figure 6 – Associating multiple properties to an abstract link

Suppose an abstract link is to be set up between A1 and A3 (the dotted line in Figure 6). There are two possible routes across the area A1-A2-A3, A1-A4-A3. The first route between A1 and A3 has a greater amount of bandwidth available, the latter minimizes delay. In networks such as this, associating bandwidth values with the abstract link becomes difficult. If the larger available bandwidth (5000 Mbits/s) and the smaller end-to-end delay (10ms) are associated to the link it would appear, at the higher level, that the area is capable of supporting an LSP that requires 4000 Mbits/s and an end-to-end delay of 10ms. This is not true.

However, if the carrier associates a lower bandwidth value with the abstract link (e.g. 3000 Mbit/s) it appears, at the higher level, that the area cannot support LSPs that require 4000 Mbits/s of bandwidth at all, which again is untrue.

To get round this problem, the area could be represented with multiple abstract links advertised between two nodes. However advertising more and more information into the higher level of hierarchy is not desirable—the whole point of hierarchical routing is to minimize the amount of link state information that nodes in other areas have to maintain.

As it happens, in many networks the differences in the cost incurred by traversing each area will be insignificant, especially in comparison with the costs associated with the Inter-area links. In this case, Complex Node Representation will offer little benefit, but would significantly increase both

- the amount of link state information each node must maintain
- the complexity of the technology running on each device, meaning a smaller investment on the part of the carrier.

Therefore, Simple Node Representation is likely to meet the requirements of most carriers.

5.4 Multiple levels of hierarchy

As the size of a carrier's networks gets larger, it is likely that more and more administrative areas will be added. Therefore, a virtual area consisting of one abstract node per area could potentially become too large. To get round this, additional levels can be added to the hierarchy. Different areas could be grouped together (on administrative or geographic grounds for example). That group of routers could then form a virtual area as in Figure 7.

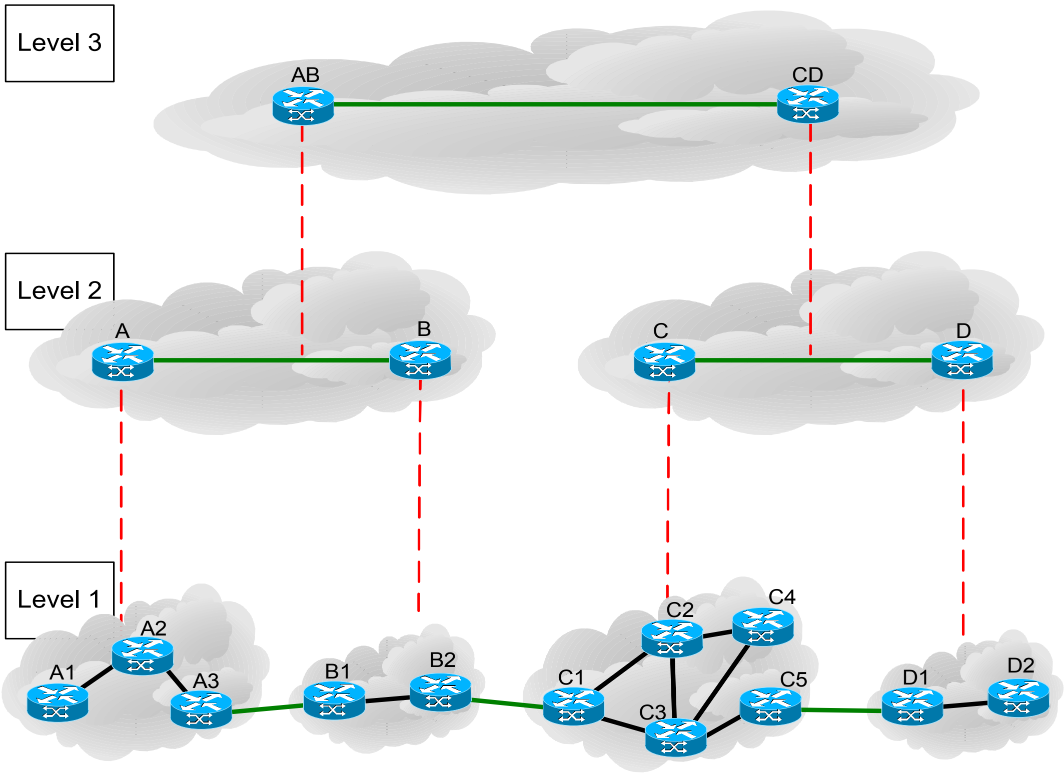

Figure 7 – Multiple levels of routing hierarchy

In this way, if A1 needs to set up an LSP to D2 say, A would select the route A1-A2-A3-B-CD-D2. Then, on receipt of the LSP setup request, nodes in areas B, C and D would expand this path in terms of the low-level topology and the LSP would be setup to follow this expanded route.

However, as we have seen associating realistic traffic parameters with the abstract links is a difficult task, and will become more difficult as more levels are added to the hierarchy. For many carriers, the added complexity of implementing a hierarchical routing protocol with many levels will be unappealing.

5.5 Example of a hierarchical routing protocol - PNNI

PNNI is a hierarchical routing protocol that has been deployed in many ATM networks. The PNNI protocol works by electing Peer Group Leaders (PGLs). These PGLs take the form of an abstract node, and represent a particular Area at a higher level. PNNI supports complex node representation and can support up to 127 levels of hierarchy—which is likely to be many more than any carrier would choose to implement.

However, PNNI does assume that the PNNI routing protocol is being run at every level of hierarchy. Therefore using PNNI would mandate every Intra-area node runs PNNI as a routing protocol—a restriction that is likely to displease many carriers who have deployed networks that do not use this technology.

More information on PNNI can be found in the PNNI specification [11].

5.6 Limitations of hierarchical routing models

Setting up protected paths is still difficult using a hierarchical routing model. In the above example, A1 is responsible for selecting the path, but that node is unaware of the low-level topology information needed to allow it to calculate a diverse path that could protect the original LSP. SRLG information could be associated to the abstract links for this reason. However meaningfully associating a set of SRLGs to an abstract node representing multiple areas may be very difficult. In Chapter 7 we will discuss a different solution proposed by the OIF and ITU for providing this kind of protection.

In the examples discussed above, a summary of the information in the higher level is passed down to the lower level nodes. In large networks, this summary may be too large for some of the low-level nodes to maintain. In the next chapter, we discuss Path Computation Servers—a pull mechanism that can be used alongside the hierarchical routing protocol to reduce the amount of information that is pushed down to each node.

6 Path Computation Servers

A Path Computation Server (PCS) is a network node that can be utilized by other nodes in the network (Path Computation Clients—PCCs) to calculate a route for them. In this way, a summary of the network topology does not need to be passed down to every single low-level node. Instead, when these nodes need to set up an LSP, they query the PCS to work out the route the LSP should take. In this way, a PCS is a good example of the "pull mechanism" defined in the Chapter 3—Inter-area routing information is not pushed down to the Intra-area nodes, but those nodes pull the information from the PCS when the route is required.

Different areas could implement different mechanisms for querying the PCS. However, extensions to the RSVP signaling protocol to provide this querying mechanism have been proposed in another IETF draft [3].

The next sections discuss some examples of PCSs that are proposed in a recent IETF draft, draft-kompella-mpls-multiarea-te [1], which show how the PCS mechanism can be applied to the Inter-area environment. We will discuss how these concepts can be extended to address the issues of trust, inherent in Inter-carrier routing at the end of this chapter.

6.1 Background

There are a large number of IP networks already deployed, from the Internet to corporate LANs. A key focus of the IETF is to increase the efficiency of these networks as quickly as possible. Therefore, it would be particularly appealing to incumbent IP network carriers, if a solution to the Inter-area problem can be found that utilizes or develops existing technologies (rather than starting afresh). The draft [1] has suggested a series of ways to do just that, using PCSs.

The focus of the draft is on using protocols such as OSPF or IS-IS to solve these problems and in this section we use the example of a network running OSPF, which is a common scenario in IP networks. However, the principles discussed here are applicable to most link state IGPs.

Before discussing these examples, we need to introduce some concepts that relate to such networks.

6.1.1 Backbone areas and hierarchy

OSPF (like IS-IS for example) has two levels of hierarchy. The Area Border Routers operate at a higher level than the other devices in the network.

Each ABR is a member of the "backbone" area or "area 0" (there is a similar concept in IS-IS—the set of level 2 routers) and must be directly connected to another router in the backbone (assuming there are other such routers). This connection can be via a physical link or a virtual link (for example, an MPLS LSP may be provisioned between the two nodes, or the virtual link may just be a concatenation of multiple physical links). Like any other Intra-area links these links can have Traffic Engineering parameters such as delay or maximum bandwidth associated with them and this information is disseminated throughout area 0 (the way in which bandwidth and other properties are associated with these virtual links is beyond the scope of this paper).

Figure 8 shows an OSPF network. The dotted links are the backbone links and the solid links are Intra-area links.

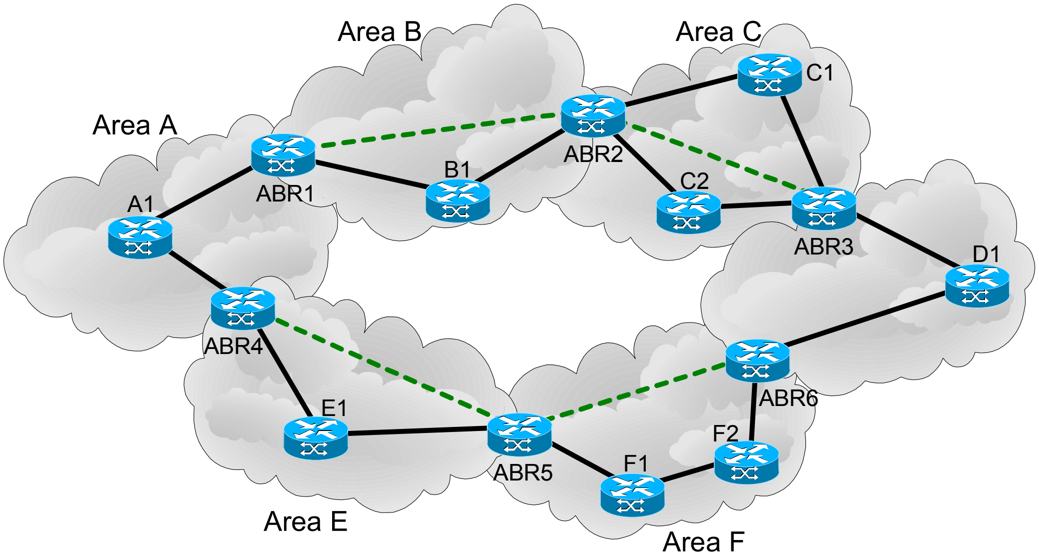

Figure 8 – An example of an OSPF network using multiple areas.

Therefore, in addition to the link state information for each area to which the ABR belongs, an ABR also includes link state information about area 0. Therefore, for example, any ABR can calculate the least-cost route from itself to any other ABR (even if it is not directly connected to it) and therefore to any other area. This idea is used in many of the suggestions below.

In OSPF, the ABR nodes have a better view of the topology of the whole network than the Intra-area nodes such as A1. This means that a non-ABR node can use an ABR node as the PCS and in sections 4.2 to 4.4 we discuss a series of mechanisms that use ABRs in this way.

Note, in OSPF, a single Area Border Router will actually be a member of every area it borders. This is not true for all IGPs—in IS-IS for example, nodes only belong to a single area. Therefore in a comparable example for IS-IS the areas would not overlap and, say, ABR 1 would be drawn as two nodes with a link between them, one in each area.

In the scenarios below, we consider a router (A1) in area A that needs to set up an LSP to D1 in area D. For carriers who have deployed IP networks the methods outlined in the following examples will be appealing—IETF drafts have defined standardized ways to implement this function that require only relatively simple changes to existing technologies to make them work. This means that they can be deployed very quickly.

6.2 Option 1—Omniscient Path Computation Server

This option requires at least one entity in the network to be aware of the low-level link state information from the entire network. When A1 receives a request to set up an LSP, it would query the entity acting as PCS to discover the route across the network to D1. The entity would then return the required route to the destination, which can be used by A1 to set up the LSP.

Further, since the PCS has complete knowledge of all the links in the network, the PCS could return the lowest level topology information back to the PCC (A1), which, since it is the ingress point, could use to set up the LSP.

Note that if there were only one such entity, the network would have a single point of failure for Inter-area TE route calculations. In order to make sure this important function was highly available, multiple omniscient entities may have to be deployed.

In the other options we shall discuss below, the PCS could only return the list of ABRs that the LSP must traverse.

6.2.1 Advantages

This is a very simple set up. Since the PCS is aware of the entire network, it is able to select least-cost routes for LSPs to take which will meet the required constraints for the LSP. In this way, the LSP can normally be able to set up first time.

Protecting LSPs is similarly easy, since the PCS can use its knowledge to calculate multiple diverse paths through the network.

6.2.2 Disadvantages

For large or medium-sized networks it is impractical to expect a single node to maintain link state information for every single link—the burden on this node would be immense.

6.3 Option 2—Local Path Computation Server

In OSPF, the ABRs have knowledge of the backbone area in addition to the link state information relating to the areas to which they belong. Therefore, in the above example, A1 may choose to query ABR 1 (acting as a PCS) to work out a route across the backbone to the tail-end area.

The PCS, ABR 1, would be free to select a route to the tail-end area that did not pass through itself if, for example, a route through itself were not available. In general, the PCS may not have knowledge of the link state of all the areas that the LSP must traverse; therefore it would just return a list of ABRs that the LSP must go through, rather than the full low-level details of the route.

This information could then be copied onto the LSP set up request by ingress. A1 could calculate a route to the first ABR in the route (in this case ABR1) using the Intra-area source-based calculation described in Chapter 3. Then, when the set up request reached the ABR1, ABR 1 would be responsible for calculating a route through its area to the next ABR in the list, ABR 2.

The Traffic Engineering characteristics associated with each of the high-level backbone links (the dotted lines in Figure 8) can be used to calculate a route to the final ABR, ABR 3, which meets the TE constraints for the LSP. However, since the ABR acting as a PCS, ABR 1, had no knowledge of the internal topology of the destination area, D, there may be no satisfactory route from ABR 3 to D1. In this case, the LSP set up would fail when it reached ABR3. A crankback mechanism could be employed and A1 could then retry the process again, instructing ABR 1 to select a route that avoided ABR 3.

It is harder to protect an LSP using this mechanism, in comparison to Option 1, since a single node does not have information about the entire network. Instead, carriers could implement protection by having the OSPF protocol associate SRLG information with each of the links across the backbone.

6.3.1 Advantages

With this option, no single node is required to maintain information about all the links in the network. Therefore, this mechanism will scale significantly better than Option 1.

This option also means that LSP set up is not initiated until the ingress point knows which Area Border Router to use. By doing this, it is much more likely that the LSP can be set up first time, contrasting with the BGP and OSPF example discussed in Chapter 4. This means that it is less likely that resources at the intervening nodes will be unnecessarily allocated and then deallocated if the original route selected by A1 turns out to be invalid and the LSP set up time is likely to be reduced.

6.3.2 Disadvantages

With this model, there is a burden placed on the ABRs who must calculate routes for each of the non-ABRs in their areas. If an ABR is acting as a PCS for a large number of clients, then it could receive a potentially large number of Route Queries.

In addition, these routes are often calculated using a Constrained Shortest Path First algorithm (CSPF), which is likely to be a relatively CPU-intensive operation. In this set up, the ABRs are responsible for calculating all the routes; therefore the burden placed on the ABRs may be significant. If an ABR is unable to deal with all of these requests promptly, it could cause significant delays to LSP set up.

In summary, using this mechanism, there could still be a lengthy delay before the LSP can be set up because

- initial attempts to do this could still fail
- the initiating node must still wait for a response from the PCS, which could take a significant amount of time.

The next option extends this mechanism to solve the first of the two problems.

6.4 Option 3—Local and remote Path Computation Servers

This mechanism develops the ideals utilized by Option 2, but instead of returning a route to just the tail-end area, the PCS also returns a route, including a route through the tail-end area, to the destination router.

In this case, on receipt of the Route Query request from A1, ABR 1 would, as before, calculate a preferred route from A1 to the tail-end area. It would then forward a Route Query request to a PCS in the tail-end area, say ABR 3, asking for a route to the destination that includes the pre-calculated route, in this case specifically including ABR 1 and ABR 3. If a suitable route existed, ABR 3 would respond with the a list of ABRs that the LSP should traverse, in the example above the list would contain ABR 1, ABR 2 and ABR 3.

If no satisfactory route existed from ABR 3 to the destination D1, (for example if there were too little bandwidth available on the link ABR 3 – D1) then ABR3 would inform ABR 1 on the response. ABR 1 could then send other Route Queries to ABR 3, for each ABR in area A that could be used to reach area D. For example, it could query ABR 3 to discover whether a route existed from ABR 4 to the destination. If a route exists, ABR 3 would provide the list of ABRs that the LSP should traverse (ABR 4, ABR 5, ABR 6).

6.4.1 Advantages

This mechanism calculates a route through the entire network that will support the QoS requirements of the LSP. Therefore, after using this method normally an LSP would be able to be set up first time.

6.4.2 Disadvantages

The route determined by this mechanism may still not be optimal in terms of overall network performance or cost. The second PCS can only calculate the optimal route from Area A to the destination, and is unaware of the internal topology of area A. Therefore, the route it selects may select an ABR in area A that it would be relatively expensive to route the call through.

This option relies uses two CSPF route calculations before LSP set up is initiated. As discussed above, these calculations are CPU-intensive, and therefore this could significantly degrade the performance of the network. The burden placed on ABRs would be even greater than the previous method and therefore carriers may need to deploy relatively powerful ABRs to cope with this.

6.5 Applications to Inter-carrier scenarios

In many ways, it would seem that the Path Computation Server mechanism is well placed to overcome the issues of trust associated with an Inter-carrier environment. If carrier A needs to set up an LSP across carrier B, it could query a node in carrier B's network for a route to the destination. Then, just as ABR 1 in the above example did not return to A1 any low-level topology information about the areas that it traversed, the PCS in carrier B's network could hide any sensitive topological information from the route that it returned to carrier A. Indeed if it returned a route consisting of only the destination hop, this would still allow the LSP to be set up, but no confidential information would be disclosed.

However, as with the BGP-OSPF model discussed in Chapter 4, guaranteeing end-to-end protection is virtually impossible. A PCS in carrier B may be able to calculate diverse routes for LSPs to take across its network, however it has no knowledge of the internals of carrier A's network. With the options discussed above, protection was achieved because ABR 1 had knowledge of the backbone as well as the head-end area, while ABR 3 had knowledge of the backbone and the tail-end area.

6.6 Fully Hierarchical Link State Routing vs PCS

As we have seen, fully hierarchical routing involves a summarized version of the network topology being pushed down to the nodes at the lowest level, which can use this information to calculate routes without the aid of any other network element. If the LSP ingress can use its own Link State Database (including summarized information about the other areas) rather than querying a PCS, it allow quicker LSP set up times. Further, performing the route calculation at these nodes, rather than requiring a PCS, reduces the burden placed on the nodes with a wider perspective of the network.

However, as we saw in the previous chapter, it is very difficult to meaningfully summarize the topology information of an area or collection of areas—this added complexity might dissuade carriers from using a fully hierarchical solution, where information is passed down to the nodes at the lowest-level. Furthermore, if a carrier wishes to minimize the control plane information nodes at the lowest level of the hierarchy must store, then the PCS-based method is very appealing.

For most carriers the solution probably lies in integrating a combined approach. Hierarchical routing protocols can be used to disseminate information amongst those nodes that can also serve as a PCS, for some of the lower level, less powerful devices. This type of model was suggested in recent developments in the optical networking world.

7 The Optical Solution

The ITU and OIF bodies have also made significant progress in the Intra and Inter-carrier path selection scenarios. The OIF have produced specifications in this area [4], [5] and at a recent demonstration at the March 2003 OFC Conference, they showcased optical networking topology working in a multiple area (Intra-carrier) environment.

7.1 UNI, I-NNI or E-NNI?

The ITU and OIF distinguish between different types of links between nodes. Links are categorized as follows, as shown in Figure 9.

- UNI (User to Network Interface). This interface is assumed to be an interface between a carrier network and an end-user device. The end user device, the UNI client (UNI-C), requests a connection across the carrier network from the UNI Network device (UNI-N), part of the carrier's network. This is not a trusted interface, and no Traffic Engineering or topology information is passed across the UNI.

- I-NNI (Interior – Network Node Interface). I-NNI links are between two nodes within a single administrative area. This is a trusted interface; full topology information can (usually) be exchanged across this.

- E-NNI (Exterior – Network Node Interface). E-NNI links are between two nodes existing in different administrative areas. The level of information disseminated across the interface is subject to an administrative policy. Inter-carrier links could be an E-NNI link where the administrative policy is used to ensure sensitive information is not disclosed to competing neighboring carriers.

7.2 How the OIF / ITU proposal works

Figure 9 shows a sample Network topology.

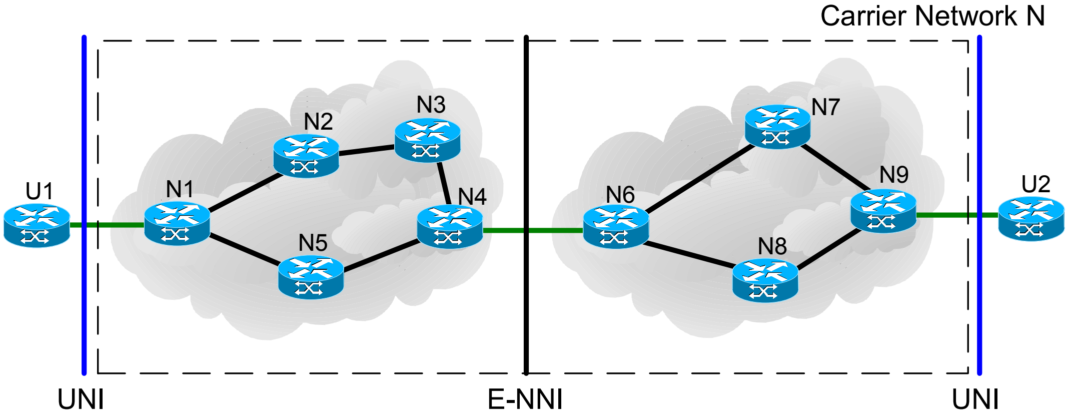

Figure 9 – UNI and E-NNI interfaces

The UNI-C, U1 requests a connection to U2, another UNI-C (it does this using a variant of the GMPLS signaling protocol). Once N1 receives this request, it can

- grant the request immediately (if there is a pre-existing connection able to service this request across the carrier network N), and instruct N9 to attempt to set up the connection between N9 and U2
- provision resources across the network to support this connection (if a suitable connection does not already exist)
- reject the request.

In the second case, resources may be provisioned using GMPLS signaling or some alternative mechanism. In all cases though, these connections may span multiple administrative areas within the carrier's network.

The end-user may require a protected connection. In this case, the user would request that a second connection (diverse from the first) is set up by the UNI-N. This way, if a node or link goes down across the network, there is a pre-provisioned connection across the Network that can still be used for sending data.

7.3 NNI routing and signaling

The NNI routing function was defined in a recent specification [13]. This specification defines an Inter-area routing protocol called DDRP (Domain to Domain Routing Protocol). This protocol is based on OSPF although an equivalent version could be developed based on other Intra-area TE capable routing protocols such as IS-IS.

DDRP is a hierarchical routing protocol, although for the demonstration only two levels of hierarchy were utilized. Within each area, nodes can distribute information about the links and nodes within it (level 1 information). However, the level 2 routing capable devices can distribute information about the following.

- The Inter-area links and their Traffic Engineering capabilities.
- Reachability information about the addresses for clients connected to the area.
- Information about "abstract links" and other client addresses across other areas. Within a particular area an "abstract link" may be defined between two border nodes. The Traffic Engineering capabilities associated with this link are derived by aggregating the capabilities of the corresponding physical links.

Using this information, routers could build up a database that would allow them to select paths across the entire network that would meet QoS requirements. In the demonstration, DDRP did not push down summarized topology information to all the low-level nodes. Instead, some low-level nodes used a PCS-based method to calculate routes spanning multiple areas. However, the specification does not describe certain key issues that carriers must consider before implementing a DDRP based solution. In particular, it does not address

- what type of information is included in the summary that is fed-down to the nodes at the lowest level, which they can then use for path selection
- how Traffic Engineering properties should be associated to any abstract links
- how a policy should be implemented to prevent confidential topology information from being leaked over the area boundary to a competing carrier.

Once the route has been calculated, the NNI signaling specification [5] defines a mechanism, based on GMPLS, to provision resources to set up the Inter-area LSP.

As with the UNI, the signaling specification indicates that when a node requests that an LSP is set up across the E-NNI it can indicate that the LSP should be diverse from another connection, or a SRLG. In this way, LSPs set up across the E-NNI can be protected.

7.4 Possible applications to Inter-carrier Traffic Engineering

There are two possible ways in which the framework above could be used to traffic engineer data between networks.

7.4.1 Inter-carrier using UNI

It is possible to implement an Inter-carrier interface using the UNI but, as we shall see in this section, providing protection for LSPs in this way is difficult.

If one carrier wished to request a connection across another carrier's network, they could use the UNI interface in order to do this, emulating an end-user. Since no Traffic Engineering information flows across the interface the supporting carrier will not disclose confidential topological information to the requester.

However, in the case where the UNI-C is part of another carrier's network rather than an end-user, there is a fundamental technical issue that is not covered by the NNI or UNI specification. While an end-user is likely to only be connected to a single UNI-N, a carrier could potentially have access to many different UNI-N nodes, in different carriers' networks. Consider the example topology shown in Figure 10.

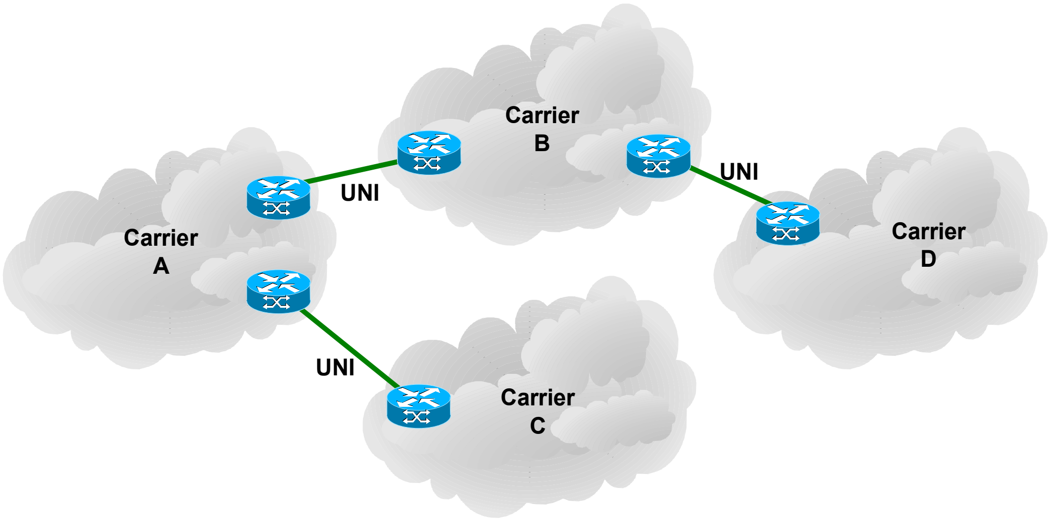

Figure 10 – Inter-carrier routing using UNI interfaces

In this case, when trying to set up an LSP from carrier A to carrier D, carrier A must decide whether the connection should be provisioned across carrier B (which would work) or carrier C (where there is no valid route to the destination). However, reachability information is not exchanged across the UNI, therefore it is not clear how the carrier can decide this.

Furthermore, even if carrier A could choose carrier B somehow, they would have to know address information relating to the UNI link between carriers B and D to ensure that carrier B set up the connection correctly. Again, it is not clear how carrier A would learn this information.

7.4.2 Inter-carrier using E-NNI

The E-NNI could be used to support Inter-carrier interfaces. This would be a more symmetric relationship—by advertising an E-NNI into two carrier's networks either carrier could (potentially) use that link for sending data traffic across the other carrier's network.

However, although the signaling across this interface has been defined, there is no such precise specification for the routing capabilities (as discussed, the current specification only deals with the Intra-carrier case), and the demonstration gives few clues about this. However, as we shall see in the next chapter, although a precise specification for this is lacking, various standards bodies have outlined stringent requirements that any possible solution to this problem must meet.

7.5 Summary

The OFC demonstration provides an important proof of concept, showing the way that inter-area optical networks could work together, and the specifications [4], [5] develop this to provide a clear plan for Inter-area and Inter-carrier signaling.

However, since a consensus on exactly what routing information should be passed across the E-NNI is yet to be reached, a crucial piece of the inter-carrier puzzle is still missing.

Furthermore, there are also some questions that need answering about DDRP.

- Are two levels of hierarchy enough, and will this technology scale to a large number of areas?
- How can the technology be used to calculate routes that are diverse from both an Inter and Intra-area level?

While the demonstration and specification provide an important step towards developing a coherent strategy for tackling Inter-carrier routing, we are still some way from having a precise specification that could be implemented to achieve that goal. However, those operating in both the packet switching and optical networking spaces have suggested a set of requirements that any candidate technology must implement if it is to be seen to solve this problem.

8 Inter-Carrier Routing Requirements

The packet and optical networking worlds have, through the IETF and ITU, laid out a series of stringent requirements that any Inter-carrier routing function must meet.

Predictably, the focus of the IP networking world (described in some IETF drafts) is on developing a routing strategy that draws heavily on existing technology and is primarily aimed at allowing MPLS Inter-carrier Traffic Engineering. While the format of this routing protocol is not strictly defined, a draft [6] hints that the Path Computation Server method outlined in the chapter 4 might be extended to meet the Inter-carrier requirement.

However, the optical world has a very different set of requirements that do not presuppose the signaling or routing protocols that should be used. Standards bodies such as the ITU or OIF are developing specifications to help carriers meet these requirements. For more information on the role of the standards bodies please see the Data Connection Whitepaper – "ASON and GMPLS a comparison".

8.1 Packet-switching MPLS Traffic Engineering requirements

As far as the IETF are concerned, there are three key requirements. [6]

- Bandwidth and QoS guarantees across other carrier's networks. A global Service Provider or carrier may want to extend their Point of Presence (PoP) to a region where the local Service Provider already has a network present. To do this, the global carrier must be able to guarantee network resources, such as bandwidth, across the local carrier's network.

- Optimization. When an Inter-carrier LSP is set up, it should follow the route that incurs the lowest cost and can still support the QoS requirements of an LSP.

- Fast Recovery. Service Providers should be able to offer the same protection services to users across multiple Areas that they expect when setting up an Intra-area connection.

The draft [6] further suggests that key features of Intra-area MPLS Traffic Engineering should be extended to Inter-carrier Traffic Engineering.   Therefore, for example

- Inter-carrier LSPs should carry DiffServ information to guarantee QoS
- in the event of an Inter-carrier link failing, a rapid protection mechanism should be able to route around this and this mechanism should interoperate with Intra-area policies such as RSVP Fast Reroute [12]
- if a new or better route becomes available, upstream LSRs should be notified so they can re-signal the LSP.

Further, when setting up an LSP between areas owned by different Service Providers, it should be possible to

- "hide" hop information relating to a particular area from other areas on the path for the LSP
- reject, on the basis of a pre-configured policy, either the Inter-carrier LSP set up or the modification of properties concerning an existing Inter-carrier LSP.

But how can an optimal Inter-carrier route be calculated? The IETF [6] suggest that the Path Computation Server method discussed in chapter 6 could be used for this purpose.

8.2 Optical Inter-Carrier Routing requirements

The ITU have defined a series of stringent generic requirements on the routing architecture, protocol and the inputs into path selection [7]. These requirements are intended to ensure the Inter-carrier strategy takes into account both the TE requirements of packet-switching networks, as well as those of optical networks as well.

The IPO IETF working group have developed these standards into specific requirements from optical carriers [8].

- No topological information should be exchanged across Inter-carrier interfaces.
- Information exchanged across all interfaces (UNI, I-NNI or E-NNI) should be configurable by the specific carrier.
- Routing protocols used to disseminate Inter-area or Inter-carrier information should be orthogonal to any Intra-area routing protocols (since there are many legacy networks, running different routing protocols inside their areas, they should not be required to change this). Note that PNNI, a hierarchical routing protocol we discussed previously, fails this requirement—PNNI requires that the PNNI routing protocol be run at every level of the hierarchy.

The first of these requirements indicates that a hierarchical link state routing mechanism that discloses link state information (even in summarized form) about the carrier's network is not applicable. It is possible that an Inter-carrier PCS mechanism (discussed in Section 6.5) could solve this problem. However, as we saw in Chapter 4, if no link state information is provided across the interface, protecting LSPs by setting up a backup for each connection becomes very difficult.

Given these contrasting requirements, the standards bodies face a stern challenge to try and develop a strategy for Inter-carrier Routing that satisfies both the optical and packet switching camps.

9 Summary

While Intra-area path selection and Traffic Engineering is relatively well defined, providing this function in an Inter-area or even Inter-carrier scenario is significantly more challenging. However, this is likely to become an increasingly important issue as carrier networks grow and transport more and more types of data traffic, each with their own QoS requirements.

As we have seen, there are several mechanisms to select routes that span multiple areas and meet the Traffic Engineering constraints of an LSP. The mechanisms we have discussed are outlined in the Table of solutions on page 36.

However, Service Providers or carriers must consider some important questions when deciding which to use.

9.1 Key issues

9.1.1 What type of information can be leaked across area boundaries?

There is a sliding scale of information that could be passed across the network boundary. At one extreme, carriers could just provide reachability information. At the other end of the scale, detailed link state information could be passed over the area boundary. There is a trade-off here—by limiting this information, carriers are decreasing the amount of information a node needs to maintain, but also making it harder for nodes to calculate routes that traverse area boundaries. Furthermore, for competitive reasons, carriers are likely to want to carefully control the link state information that is leaked to other carriers.

9.1.2 Push or pull mechanism?

As discussed in chapter 3, there are two general types of path selection mechanism; push or pull. The former push down routing information to the lowest level nodes and therefore those nodes must store more information. With pull mechanisms, the required information is pulled from nodes operating at the higher level of hierarchy (with a wider view of the network) at LSP set up time.

Push mechanisms, such as hierarchical routing protocols, are also more complex to implement and carriers must pay careful attention to exactly what information is summarized. However, there are deployed networks using these protocols (such as PNNI) and therefore these mechanisms are proven to scale.

Carriers are likely to implement a combined strategy, hierarchical routing protocols will push down summarized topology information to a set of nodes that will act as PCS for the remaining nodes in the network. If a carrier does this, then working out which points of the network will use which path selection mechanism will be an important issue to resolve.

9.1.3 For push mechanisms, how much information should be summarized?

The hierarchical routing mechanisms require that the link state information for a particular area be summarized. Putting the issues of trust discussed previously to one side, the less information about the internal topology of an area that is advertised across area boundary, the smaller the associated maintenance overhead for the entire network. However, restricting the amount of information that is summarized makes it more likely that sub-optimal paths will be selected, reducing the overall efficiency of the network.

9.2 Table of solutions

The following table provides a concise summary of the mechanisms discussed in this paper, although it is not intended to provide an exhaustive guide to every available solution to this problem. For more details on the individual mechanism specified below, see the appropriate section or references.

Description	Refs	Push or Pull	Does it allow TE across area boundaries	Allows information hiding (Inter-carrier)	How well does it scale?	Can Protection be easily implemented?
PCS 1 – Use omniscient entity (Section 6.2)	[1], [2], [3]	Pull	Yes	No	Major concerns for large networks.	Yes, the entity knows the full route.
PCS 2 – Use a local PCS to calculate a route to the tail-end area (section 6.3).	[1], [2], [3].	Pull	Yes	No	Possible burden on ABRs. Only two levels of hierarchy discussed in [1]	Yes. Full route is passed back to ingress.
PCS 3 – Use local and remote PCS to calculate route to destination (section 6.4)	[1], [2], [3]	Pull	Yes	No	Possible burden on ABRs (greater burden than PCS 2)	Yes. Full route is passed back to ingress.
PCS 4 – Use PCS without passing back topology information. (Section 6.5)	None as yet.	Pull	Yes	Yes	Some concerns as above.	May not be easy, not defined in a specification as yet.
Distribute just reachability information across area boundaries (Section 4.1)	Numerous, depending on protocol used.	Push	No	No	Should scale well.	Not without distributing topology information across the network boundary.
PNNI as routing protocol (section 5.5)	[11]	Push	Yes	Privacy issues are not explicitly considered	Very well. PNNI has proven scalability.	With difficulty. PNNI could disseminate SRLG information can be used to calculate diverse route.
OIF-NNI using DDRP as the routing protocol (section 7)	[4], [5], [13]	Push	Yes	Yes	Should scale well.	It is not clear exactly how diversity support can be provided.

9.3 Conclusion

As we have seen, selecting a path for an LSP that satisfies the LSP constraints is particularly difficult when the LSP must span multiple areas. Different standards bodies are developing different strategies to tackle these issues—the IETF is actively working on drafts ([1], [2], [3], [6]) that show how existing technologies can be leveraged to provide a solution relevant to many packet switched networks. However, the OIF and ITU are currently developing a different DDRP-based approach for optical networks ([4],[5],[7]).

This difference is motivated by the significant differences in the requirements of optical carriers in comparison with say, incumbent carriers owning IP networks. Given this, some of these solutions are likely to co-exist in the future.

Within each of the rough positions assumed by the IETF or OIF and ITU, there are many solutions being discussed. Each solution has its own drawbacks; for example some solutions may not scale well but implementing other more scalable approaches is a significantly more complex issue. However, carriers will not decide which solution to implement on the technical issues we have discussed alone. They will carefully weigh up the investment needed to implement each solution against the savings it could provide to the running costs of their network. It is these factors together that will determine which solutions are accepted by the industry and deployed in tomorrow's networks.

10 References

The following documents are referenced within this white paper. All RFCs and Internet drafts are available from www.ietf.org URLs are provided for other references.

Note that all Internet drafts are "work in progress" and may be subject to change, or withdrawn without notice.

10.1 References

1	draft-kompella-mpls-multiarea-te	This draft defines the set of mechanisms suitable for Inter-area (not Inter-carrier) path selection.
2	draft-kompella-mpls-rsvp-constraints	This draft extends the RSVP signaling protocol to pass constraints relating to the path selection process along the route of the LSP on the Path message. This way, if another node is asked to compute (part of) the Path for that LSP, it can ensure the path it computes meets the original constraints. This is necessary to perform constraint-based routing when non using source-based path selection.
3	draft-vasseur-mpls-computation-rsvp	This draft extends RSVP to provide a "Path Computation" message needed to query the PCS.
4	oif-UNI-01.0	This document defines the UNI 1.0 specification.
5	Oif2003.179.03	This is the NNI Signaling specification that defines I-NNI and E-NNI interfaces.
6	draft-ietf-tewg-interas-mpls-te-req	This draft defines the requirements for Inter-carrier (or Inter-AS) path selection and routing from an IETF perspective
7	G7715	This document defines the ITU requirements on routing protocols, including those that will be used to allow Inter-area and Inter-carrier routing.
8	draft-ietf-ipo-carrier-requirements	This draft defines the optical network carriers' control plane requirements.
9	"Interdomain Optical Routing", (Greg M. Bernstein, Vishal Sharma, Lyndon Ong, Journal of Optical	This article discusses the issues involved with Inter-area routing (both Intra-carrier and Inter-carrier cases).

	Networking, February 02)	
10	draft-ietf-mpls-lsp-hierarchy	This draft indicates how hierarchical LSP setup can be signaled.
11	Private Network-Network Interface specification, PNNI 1.0 af-pnni-0055.000	This is the PNNI specification, defined by the ATM forum.
12	draft-ietf-mpls-rsvp-lsp-fastreroute	This draft defines the RSVP fast re-route function that can be used to provide protection for LSPs.
13	oif2003.259.00	This is the E-NNI routing specification for DDRP (an OSPF-based routing protocol).

10.2 Other useful documentation

10.2.1 RFCs

13	RFC2205	Resource ReSerVation Protocol (RSVP)— Version 1 Functional Specification
14	RFC3209	RSVP-TE: Extensions to RSVP for LSP Tunnels
15	RFC2328	OSPF Version 2
16	RFC1771	A Border Gateway Protocol 4 (BGP-4)
17	RFC1142	OSI IS-IS Intra-area Routing Protocol
18	RFC3270	Multi-Protocol Label Switching (MPLS)—Support of Differentiated Services
19	RFC3272	Overview and Principles of Internet Traffic Engineering

10.2.2 Internet drafts

20	draft-katz-yeung-ospf-traffic	Defines extensions to OSPF to support Traffic Engineering
21	draft-ietf-ccamp-gmpls-routing	Defines routing extensions in support of Generalized MPLS
22	draft-ietf-ccamp-ospf-gmpls-extensions	Traffic Engineering extensions to OSPF in support of GMPLS
23	draft-boyle-tewg-interarea-reqts-00	Requirements for support of Inter-area and Inter-AS MPLS Traffic Engineering

10.2.3 Other ITU documents

24	G.8080	Architecture for Automatic Switched Optical Networks (ASON)
25	G.807	Requirements for Automatically Switched Networks.

10.3 Other Data Connection white papers

Data Connection has published a series of white papers. Other white papers include "MPLS Virtual Private Networks", which looks at the use of MPLS as a tunneling protocol for VPNs, and "VPN Technologies—a comparison", which examines the VPN technologies that are currently under discussion.

These white papers can be downloaded at http://www.dataconnection.com.

11 About Data Connection

Data Connection Limited (DCL) is the leading independent developer and supplier of MPLS, IP Routing, LMP, ATM, SS7, MGCP/Megaco, SCTP, VoIP Conferencing, Messaging, Directory and SNA portable software products. Customers include Alcatel, Cabletron, Cisco, Fujitsu, Hewlett-Packard, Hitachi, IBM Corp., Microsoft, Nortel, SGI and Sun. Data Connection is headquartered in London UK, with US offices in Reston, VA and Alameda, CA. It was founded in 1981 and is privately held. During each of the past 20 years its profits have exceeded 20% of revenue. Last year sales exceeded $35 million, of which 90% were outside the UK, mostly in the US.

The DC-MPLS product family provides OEMs with a flexible source code solution with the same high quality architecture and support for which Data Connection's other communications software products are renowned. It runs within Data Connection's existing high performance portable execution environment (the N-BASE). This provides extensive scalability and flexibility by enabling distribution of protocol components across a wide range of hardware configurations from DSPs to line cards to specialized signaling processors. It has fault tolerance designed in from the start, providing hot swap on failure or upgrade of hardware or software.

DC-MPLS is suitable for use in a wide range of IP switching and routing devices including Label Switch Routers (LSRs) and Label Edge Routers (LERs). Support is provided for a range of label distribution methods including Resource ReSerVation Protocol (RSVP), Constraint-based Routed Label Distribution Protocol (CR-LDP) and Label Distribution Protocol (LDP). The rich feature set gives DC-MPLS the performance, scalability and reliability required for the most demanding MPLS applications, including VPN solutions for massively scalable access devices.

DC-MPLS integrates seamlessly with Data Connection's IP Routing, LMP, ATM, SS7, MGCP and other protocol products, and uses the same proven N-BASE communications execution environment. The N-BASE has been ported to a large number of operating systems including VxWorks, Linux, OSE, pSOS, Chorus, Nucleus, Solaris and Windows NT, and has been used on many processors including x86, i960, Motorola 860, Sparc, IDT and MIPS.

DC-Directory is a complete directory solution, which combines the best elements of X.500 and Internet standards (LDAP, HTTP) with comprehensive management and administration applications to provide an open, scalable directory service for enterprises and Service Providers. Directories are increasingly being considered to present a unified management system for network hardware, MPLS systems, network authentication and VPN security.

Ben Wright is a Customer Services consultant for the DC-MPLS and DC-Routing product families.

Data Connection is a trademark of Data Connection Limited and Data Connection Corporation. All other trademarks and registered trademarks are the property of their respective owners.