Path Computation Element in SURFnet6

Is the Path Computation Element suitable for SARA's planning tool?

P. van Abswoude (Patrick.vanAbswoude@os3.nl)
R. Koning (Ralph.Koning@os3.nl)

System and Network Engineering

University of Amsterdam

Final Version

July 2, 2007

Abstract

This report describes the Path Computation Element (PCE) architecture, specified in RFC4655. It also motivates if the PCE is usable and applicable for SARA's planning tool. This planning tool is responsible for computing lightpaths in the SURFnet6 network. In our opinion the PCE architecture is applicable for SARA's planning tool although a lot of features like the Path Computation Client, the PCEP protocol and support for constraints and policies have yet to be implemented in their planning tool. Especially the inter-domain and inter-layer path computation properties are interesting. Unfortunately, an IGP protocol with Traffic Engineering extensions, like OSPF-TE or IS-IS-TE, has to be implemented in the SURFnet6 network before this can work.

Preface

As part of our study *System and Network Engineering* at the University of Amsterdam, we did a four week research project at SARA [1], an advanced computing and data centre in The Netherlands. We were located at the High Performance Networking department and did research on the planning tool currently used by the SURFnet[2] NOC[3]. SURFnet is the Dutch Internet service provider for educational and scientific institutes. The planning tool is used to plan lightpaths within the SURFnet6 network.

This document is the result of this research.

Acknowledgements

We would like to thank the people at SARA and especially our supervisors Ronald van der Pol and Andree Toonk for providing all the information and support we needed to conclude our research. We would like to express our gratitude to Lucy Yong for sharing her insights on future reservations. We would also like to thank our co-students at the University of Amsterdam, especially those who were also located at SARA for insight and input on our findings during the coffee breaks. We would also like to thank the people who reviewed this document: ing. G. v Malenstein, ing. C. Steenbeek, Erik Romijn and everyone else who helped us in some way.

[1] `http://www.sara.nl/`
[2] `http://www.surfnet.nl/`
[3] Network Operations Center

Contents

Chapter 1

Introduction

Since the introduction of fibre in 1840 [1], optical networks have been deployed all over the world. In the beginning a lot of these networks were focused on universities and research facilities because of the large quantities of data they had to transport, but nowadays optical networks are also used by Internet Service Providers (ISPs) and other large organisations.

SURFnet is an ISP for universities and research institutes in The Netherlands. These institutes are connected to the SURFnet6 network which has links to research networks all over the world.

The SURFnet6 network is a hybrid network consisting of a circuit switched optical part and a packet switched (traditional) routed IP part.

SURFnet6 offers lightpath services. To set up a lightpath in an optical network, a path has to be found from end-point to end-point and due to limited resources this also has to be done efficiently. Of course this can be done with Dijkstra's Shortest Path algorithm, but there were extra requirements. It is important to predict the properties of the circuit, so *constraints* are added to the algorithm to determine the QoS[1].

To calculate these kinds of paths in a network, the IETF has formed the PCE working group which is working on the PCE – Path Computation Element. This paper will discuss the usability and applicability of PCE for SARA's planning tool used to calculate paths in the SURFnet6 network.

[1] Quality of Service

Chapter 2

Problem description

2.1 The SURFnet6 network

The SURFnet6 network has been operational since the beginning of 2006 and contains over 6000 KM of dark fibre throughout the Netherlands. To implement this optical network, they use Nortel CPL (Common Photonic Layer) equipment for the DWDM signal, OM5200 devices for the conversion from the DWDM wavelength to Ethernet packets and OME6500 equipment for the SDH layer. Their clients are mostly universities and research institutes. Figure 2.1 is a global view of the SURFnet6 network.

Figure 2.1: The SURFnet6 network

2.1.1 Hybrid Networks

SURFnet6 is a hybrid network that provides two types of services: packet switched IP, and circuit switched lightpath services.

A lightpath is a path from one end-point to another inside an optical network with at least the following properties:

- Dedicated amount of bandwidth;
- Quality of Service;
- Deterministic;
- Predictable latency;
- Low jitter.[1]

These lightpaths are offered in various forms and will be explained in section 2.1.3.

2.1.2 The need for dedicated lightpaths

The reason why SURFnet provides these lightpaths is because they provide connectivity to universities and scientific institutions which require a lot of bandwidth. Cees de Laat of the University of Amsterdam has created a theory in 2002 that confirms this. He states that there are three types of users in a network [6]:

1. Lightweight users – do not require much bandwidth (browsing, mailing);
2. Business users – require a significant amount of bandwidth (multicast, streaming, VPNs);
3. Special (scientific) users – require a lot of bandwidth (data grids, virtualisation).

Figure 2.2 is a graphical representation of this theory.

It is true that nowadays the average lightweight user has the Internet use of a 'business user' because of the introduction of services like YouTube and IPTV, but business users also have expanded their Internet use by using more and more video conferencing.

2.1.3 Types of Lightpaths

Because of different demands, SURFnet distinguishes four kinds of lightpaths through the SURFnet6 network:

[1] Variation in the time between arriving packets.

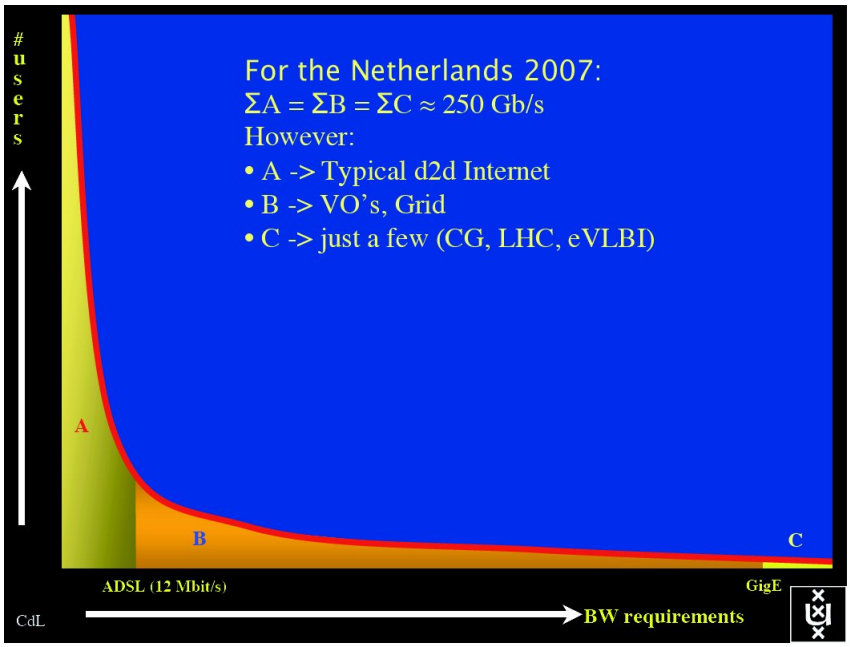


Figure 2.2: The different kinds of users (courtesy of Cees de Laat)

Unprotected lightpaths

Unprotected lightpaths are bi-directional lightpaths without any backup facility. If a network element or fibre fails, the lightpath is interrupted.

Redundant lightpaths

For a redundant lightpath a client needs two interfaces because it actually consists of two separate lightpaths. To eliminate the chance of link failure, the second lightpath goes through a different set of network elements, fibres and ducts.

Protected lightpaths

There are two types of protected paths, a loosely protected and a strictly protected path. The loosely protected paths are alternative paths using different port IDs and can go through the same devices; the strictly protected paths are built using completely different devices, fibres and ducts to provide an extra level of protection.

Optical Private Network

An optical private network is a network of protected or unprotected lightpaths comparable with multiple peer-to-peer connections.

2.2 Current Software used in SURFnet6

Ronald van der Pol and Andree Toonk from SARA have written the current software for calculating end-to-end lightpaths within the SURFnet6 network [27]. It's called the 'Planning Tool'; it consists of a Perl script that calculates the shortest path using Dijkstra's constraint based shortest path algorithm using topology data from NDL – Network Description Language[2] –, and network state data from a MySQL[3] database.

SARA is also planning to implement the ability to account for SRLGs – Shared Risk Link Groups –, which is currently done by hand. This adds the ability to calculate paths that not only uses alternative network elements, but also adds the guarantee the path does not go through the same fibres and ducts, reducing the risk of connection failures even more.

2.2.1 Technical explanation of the planning tool

The current planning tool consists of a web interface where a user is able to select two end-points (figure: 2.3).

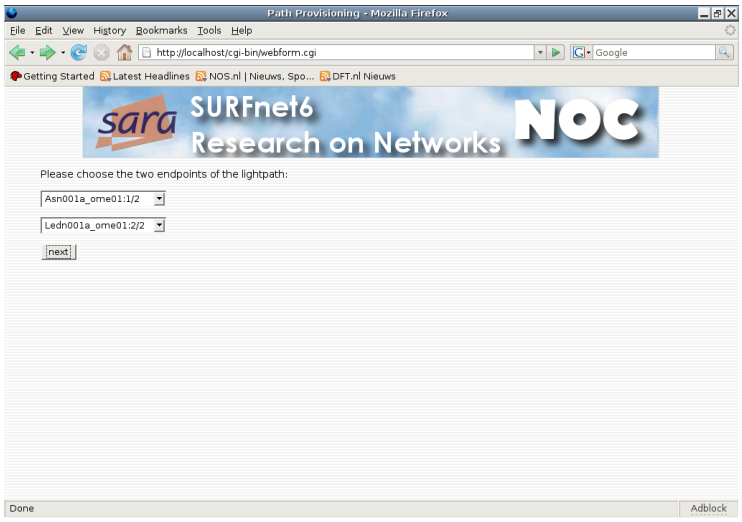


Figure 2.3: Select end-points

Then the user selects the kind of lightpath and the amount of bandwidth required (figure: 2.4).

When all the data is collected, the planning tool runs a CSPF (Constraint based Shortest Path First) algorithm, which will be explained later in this section. The output contains the network elements which the network administrator has to configure to set-up the lightpath (figure: 2.5).

[2]NDL is a format for specifying network topology's in XML[26]
[3]`http://www.mysql.com`

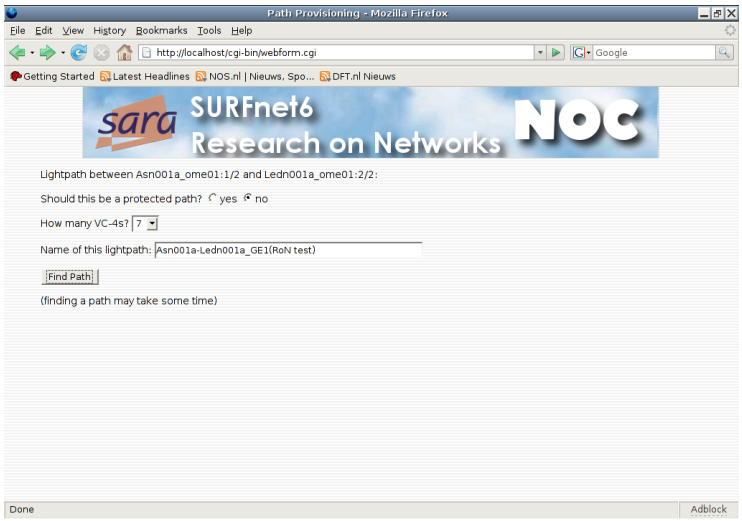


Figure 2.4: Select type, name and bandwidth

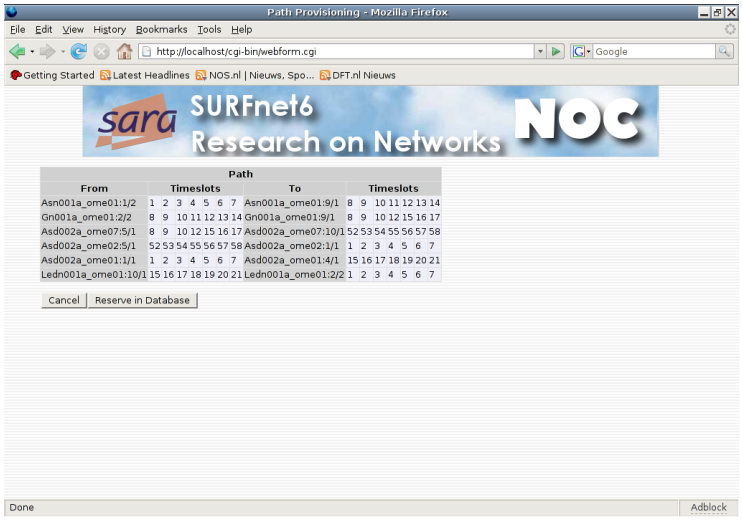


Figure 2.5: The output

SARA uses a very simple protocol to communicate with the planning tool. All information is sent and received in plain text strings and can be easily interpreted. In this way an administrator can connect to it with telnet, and do path computation requests.

The algorithm

The following is an example of the CSPF algorithm used in SARA's planning tool. In this example, we want a lightpath between `university 1` and `university 2` with the following constraints:

- Bandwidth – 2.5 Gbps available.
- Type of lightpath – Protected.
- Type of protection – Strict (the protected lightpath can't go though the same network elements).

First, the topology is loaded and the two end-points are selected (figure 2.6).

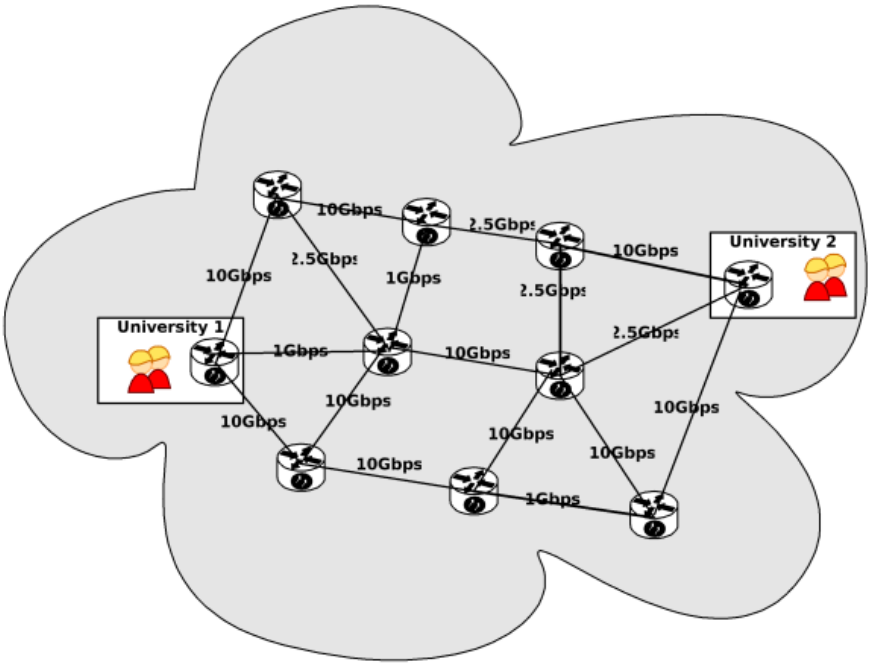


Figure 2.6: The example network topology

Then, the constraints are applied. First, all the links that do not comply with the bandwidth constraint are removed (figure 2.7).

Then Dijkstra is applied for the first time (figure 2.8).

After the first run, all the nodes that are part of the calculated shortest path (and are already in use) are removed from the topology (the line in dashed red). Because a protected path is desired, Dijkstra is applied again on the remaining topology which leaves us with a second shortest path (figure 2.9).

Now there is a lightpath (black – the solid line) and a secondary 'protection' lightpath (blue – the dashed line), that complies with the given constraints (figure 2.10).

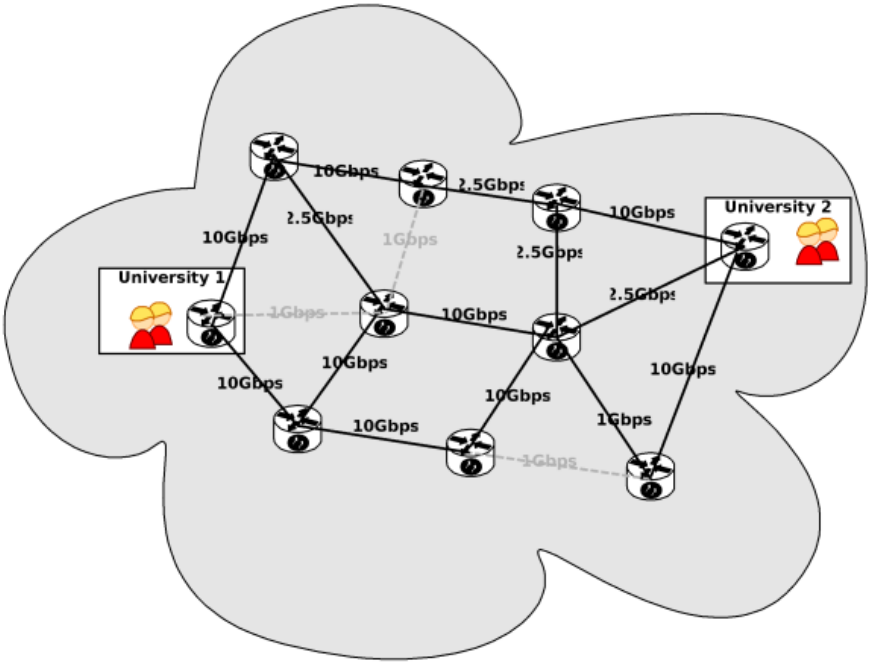


Figure 2.7: The topology with the bandwidth constraint applied

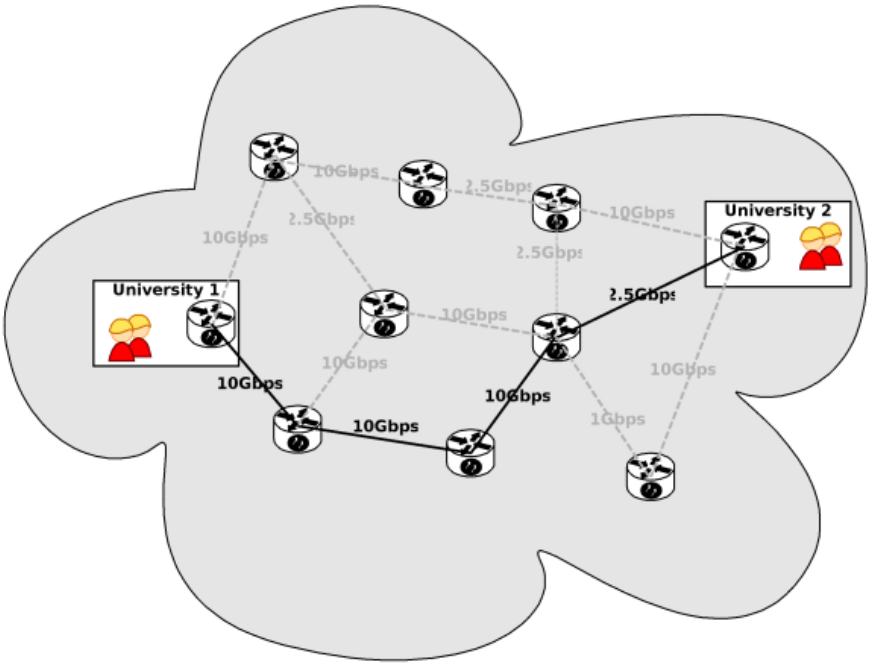


Figure 2.8: The topology with Dijkstra first

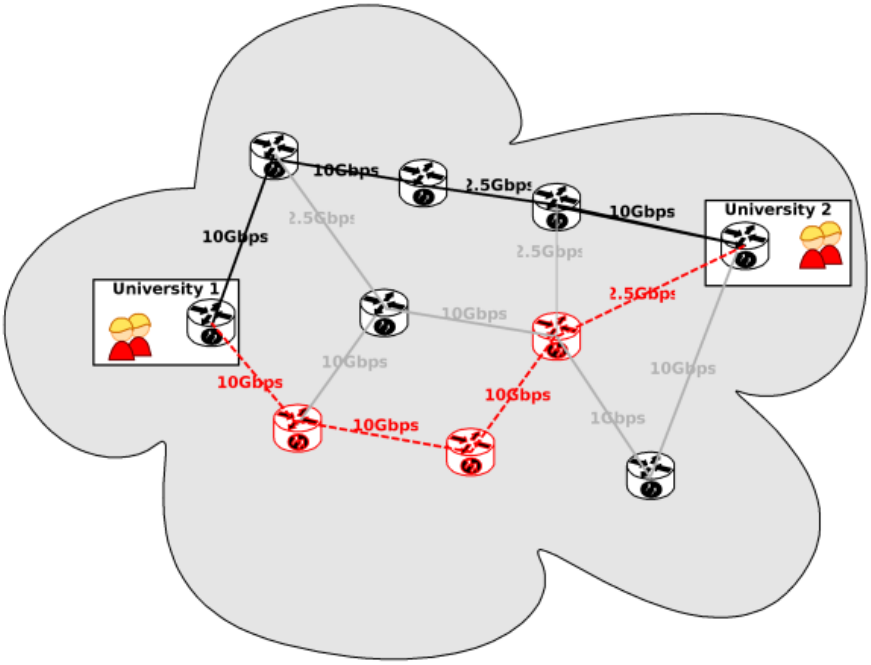


Figure 2.9: The topology with Dijkstra second

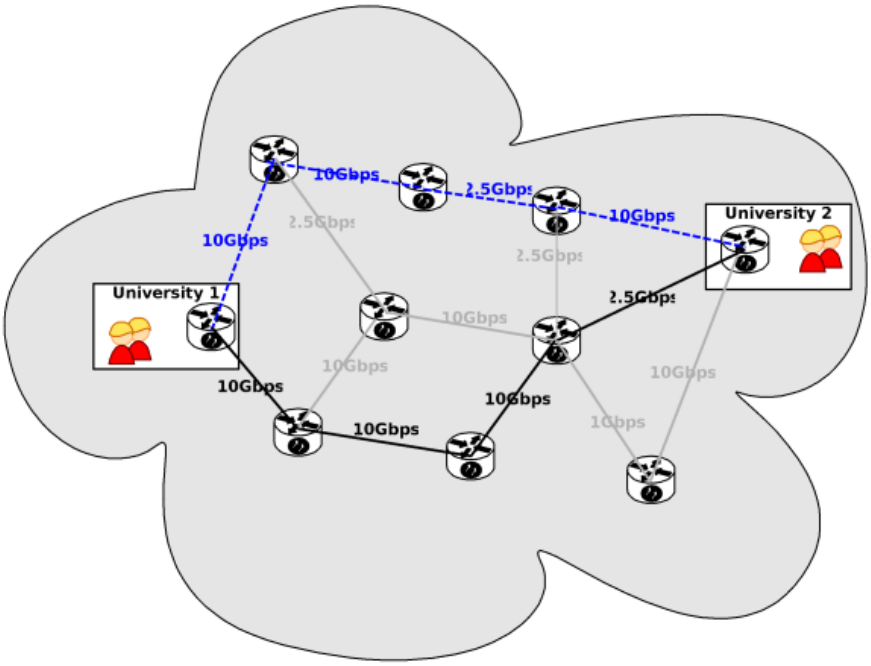


Figure 2.10: The calculated lightpaths final

2.3 Assignment

The PCE working group of the IETF is currently working on a description of the PCE (Path Computation Element) and some guidelines on the usage of this element in intra- and inter-domain networks. This PCE resembles the planning tool described earlier in this document.

SARA asked us to analyse the documents of the PCE working group and tell whether or not this PCE could be a useful extension or replacement of the current planning tool. In addition to this, SARA asked us to look if the PCE-based architecture also covers some features they are planning to add in the next version of their planning tool, such as Shared Risk Link Groups and 'future reservations'.

Based on this information we formulated the following research question:

> *Is the PCE-Based architecture described in RFC4655 usable for SARA's planning tool?*

We also formulated six sub-questions:

- How does the IETF's PCE retrieve its knowledge of the network?
- How does IETF's PCE retrieve the status of the network elements?
- How does IETF's PCE find a path?
- How does IETF's PCE return a found path?
- Does IETF's PCE look at Shared Risk Link Groups (SRLG)?
- Which additional advantages can be gained by using the IETF PCE architecture?

In the following chapters, our findings about PCE are documented, and we will answer the research question.

Chapter 3

Path Computation Element

3.1 About PCE

RFC4655[8] describes a PCE- based architecture, it also defines a PCE as follows:

> *"A Path Computation Element (PCE) is an entity that is capable of computing a network path or route based on a network graph, and of applying computational constraints during the computation."*

The architecture described is heavily tied with Traffic Engineering routing protocols and the Generalized Multiprotocol Label Switching (GMPLS) but it should function within other environments.

The main motivation for a PCE-based architecture is the offloading of path computation which can be heavy on large networks due to the extra constraints, from the nodes itself to a separate entity. This entity can run centralised on a dedicated node, or distributed on several nodes which have a complete view of the network. Another motivation is to add the possibility of finding paths across multiple layers (e.g. an IP layer and an optical layer), or multiple domains and adding policies to these computations for enhancing security and performance.

3.2 Current status of PCE

The current status of PCE is a work in progress. There are currently three informational RFCs [8] [11] [2] and about 17 Internet-Drafts, all being worked on by the PCE working group.

3.3 Path Computation Element Components

This section will explain the PCE architecture as explained in RFC4655[8].

The main function a PCE is calculating paths between two end-points with given constrains. The idea is that a Path Computation Client (section 3.3.1) will give two end-points and constraints as input. The PCE will use the Traffic Engineering Database (TED; section 3.3.2) to calculate the path and will return a message with the calculated path.

3.3.1 Path Computation Client

The Path Computation Client (PCC) is the client of a PCE. Its main function is to request a path from a PCE by it giving two end-points and constraints.

The PCC could be a standalone client or can be implemented on the network elements.

If the PCC is implemented on the network elements (nodes), the ingress node sends a path computation request to the PCE. When the ingress node receives a reply from the PCE, the path will be set up. In order to do this the nodes in the network must run an Interior Gateway Protocol (IGP) protocol with Traffic Engineering support. One of the reasons the IGP protocol with TE extensions is needed, is because the nodes will learn the location of the PCE and it's capabilities through this protocol.

To set restrictions on the PCC, a network administrator could use policies. Policies will be explained in section 3.4.2

3.3.2 Traffic Engineering Database

To gain knowledge about the current network state and the topology the PCE consults a Traffic Engineering Database (TED). The TED contains the topology of the network. The information the TED must contain are:

- The various network elements;
- Which interfaces a network element has;
- The capacity of each interface;
- To which network element(s)/interfaces the interfaces are connected;
- How much capacity is used at the moment;
- Type of interface (SDH, DWDM, Ethernet etc.).

A TED could be filled by routing protocols like OSPF-TE[12] or IS-IS-TE[23] or could be constructed by hand as long as it contains enough information to find paths. The only demand with a self-created TED is that the TED must be updated as network resources are used or released. To calculate a path, constraints are applied to this information and an algorithm should be run to calculate the path.

The RFCs do not specify any specific way to get information from the TED or how a path should be found. It only states that a path should be returned in a proper manner, *if* a path could be found.

3.3.3 Constraints

Instead of calculating the shortest path – something that is normally done in an IP-based network – the PCE is able to calculate the shortest path with constraints. The constraints are based on the content of the TED. RFC4657[2] states that the following constrains must be supported:

- MPLS-TE and GMPLS generic constraints:
  - The amount of bandwidth needed.
  - Affinities inclusion/exclusion[1]
  - Link, Node, Shared Risk Link Group (SRLG) inclusion/exclusion
  - The ability to use OSPF or ISIS metrics as a constraint.
  - The ability to use OSPF or ISIS Traffic Engineering metrics as a constraint (e.g. delay)[9].
  - Restrict the maximum amount of nodes a path must traverse.
- MPLS-TE specific constraints
  - MPLS Class-type, a method for labelling types of traffic.
  - Local Node and Bandwidth protection. Provides the ability to protect the link with the MPLS Fast ReRoute [18] method defined in RFC4090[19]
  - Node protection (Provides routing around a failed node using fast re-route[18])
- GMPLS specific constraints
  - Switching type, encoding type (SONET, SDH)
  - Link protection type (Link, Node, SRLG)

Although all the constraints are currently based on properties used in (G)MPLS-TE, these properties are mostly general and therefore usable by other techniques.

3.3.4 Inter-PCE path computation

Not all PCEs require a complete view of the network. A PCE can consult another PCE to assist it in calculating a path through a different part of the network. In this way one can choose to use a PCE for different areas in a network. This can be usable if one wants to distribute processing power or

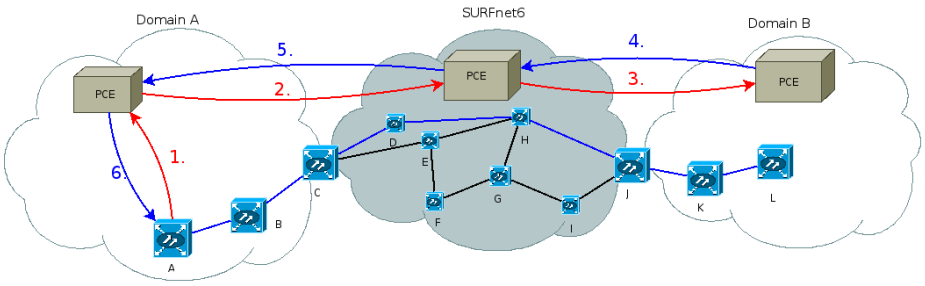
[1] A PCC may request the PCE to exclude points of failure.

wants to enhance security. It also adds the ability for inter-domain and inter-layer communication.

Inter-domain

Inter-domain and inter-layer path computation is still open for further research because the above mentioned papers are not going into detail about the underlying technique – they simply state that it is within the scope of PCE. We simply state it here because it could be an interesting option for the SURFnet network.

Inter-domain path computation is the calculation of a path through multiple domains. The IETF has created an Internet-Draft [20] about inter-domain communication between PCEs. Explaining inter-domain path computation is best with an example.



1. Path request from A to L.
2. Path request from C to L.
3. Path request from J to L.
4. The path is: "J – K – L".
5. The path is: "C – D – H – J – K – L".
6. The path is: "B – C – D – H – J – K – L".

Figure 3.1: An example of inter-domain path computation

In figure 3.1 is an example of inter-domain path computation. A PCC of domain A requests a path from node `A` to node `L` in domain `B`. The idea is that the PCE from domain `A` sends a path computation request to the PCE from SURFnet requesting the path from node `C` to `L`. Then SURFnet's PCE contacts the PCE in domain `B` to request for its portion of the path. Considering everything goes well, the PCE of domain `A` can concatenate the path from `A` to `C` with the path returned from the SURFnet PCE and return a full path from `A` to `L`. In this case, the PCEs which send a path computation request to another domain are also PCCs.

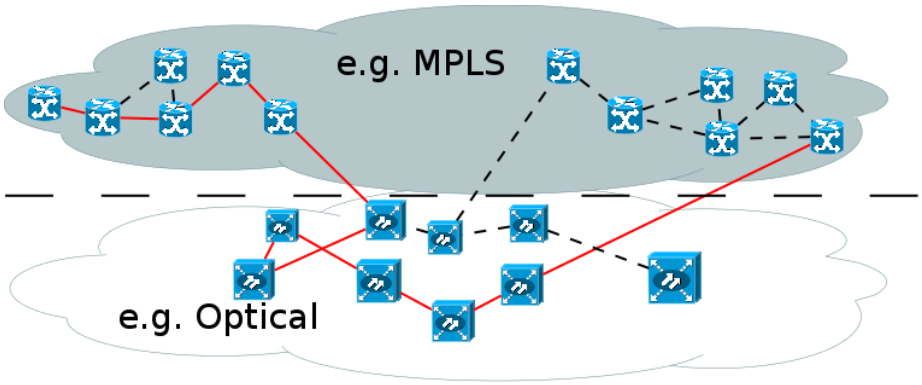
Inter-domain path computation brings some security complications. The security complications are discussed in section 3.4.

Inter-layer

Inter-layer communication is applied when the objective is to perform path computation though multiple network layers. This means that the PCE is able to calculate a path that is going through multiple layers – which could be valuable for the hybrid network of SURFnet6. The reason why a network administrator wants inter-layer path computation is stated in an Internet-Draft containing the framework for PCE-based inter-layer MPLS and GMPLS Traffic Engineering[16]. The main reason is, that it is important to optimize network resource utilization globally in all layers rather than optimizing only one layer at the time. This is also called inter-layer traffic engineering (or inter-layer TE). This is important because a higher layer hop could have the value of 1 (1 hop) but the underlying circuit could really exist of multiple hops.

The framework Internet-Draft[17] defines two models for inter-layer path computation. The first model is the single PCE inter-layer path computation. In this scenario, inter-layer path computation is performed by a single PCE that has knowledge of all layers in the whole domain – a so called multi-layer PCE. The second model is the multiple PCE inter-layer path computation. In this scenario, there is at least one PCE per layer. If the PCE on the 'higher' layer gets a path request which he cannot compute, it will "consult" the PCEs on lower layers to compute a path on the lower layers. Then a PCE in a lower layer will return a path and the higher layer PCE is able to return a complete path through multiple layers.

In figures 3.2 and 3.3 is an example of path computation through multiple layers.

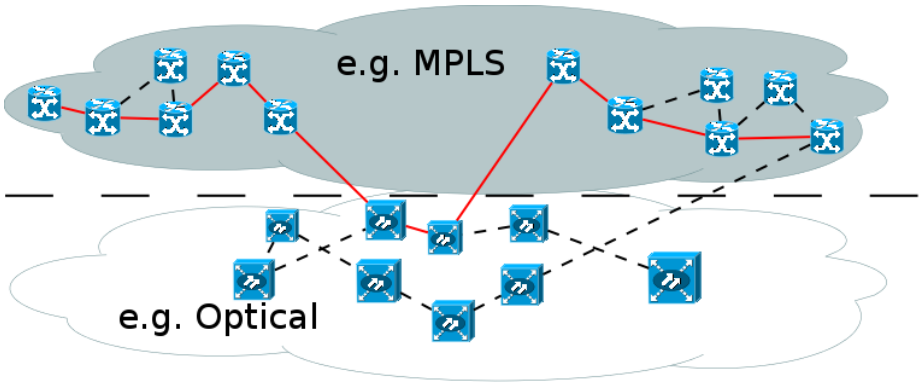


Without multi-layer PCE, SPF is done on each layer separately. The MPLS considers the optical layer as a single hop and counts 6 hops. Overall this results in a suboptimal path of 11 hops.

Figure 3.2: An example of inter-layer path computation

In the first figure, there is only a PCE on the MPLS[2] layer. This is an example of a path computation on e.g. a MPLS layer that has the most optimal path on the MPLS layer but a suboptimal path on the underlying optical layer.

[2]MPLS is a way to create a circuit switched network upon a packet switched network



With multi-layer PCE, SPF is done on all layers together. The PCE can now also see the optical layer. It results in an optimal overall path of 10 hops.

Figure 3.3: An example of inter-layer path computation

In the second figure, the PCE considers multiple layers making sure the path is optimal on both layers (inter-layer TE).

3.3.5 Communication

Requirements for communication between PCC and PCE and communication between PCE and PCE are described in RFC4657[2]. It states that there must be one protocol covering PCC-PCE and PCE-PCE communication because they practically do the same requests. This protocol is called the Path Computation Element Communication Protocol (PCECP).

The communication is client-server based, the PCC will send a message containing one or more path requests to the PCE. If the PCC did not cancel the request, the PCE returns a positive or negative response message. The request must contain a source and destination and one or more path constraints, e.g. minimum bandwidth. A positive response consists of one or more paths, if no path could be found the PCE gives a negative response. The paths returned by the PCE must be easily convertible to Explicit Route Objects (EROs; section 3.3.5) and acceptable for use in (G)MPLS enabled networks.

There are some additional requirements for availability, security, extensibility, and scalability. These requirements are described in detail in RFC4657.

At the time of writing the IETF is working on a protocol specification, the PCEP protocol[28], this is an implementation of the requirements written in RFC4657 and it is currently in draft form.

IETF is planning to use TCP as transport protocol mainly because of reliability and flow control. A session has to be established before a PCC can send requests to the PCE and receive a reply. Within these sessions keep-alive messages will be send to check for connection failure. Notifications are used to give information about events like PCE congestion. If a PCC wants to cancel a request, it also sends a notification.

A priority could be given to a PCE request. The standard value is 0 but can be from 1 to 7, where 7 is the highest priority.

Request Format

The format for a path computation request exists of a `RP` (Request Parameters) object, an `END_POINT` object containing two IPv4 or IPv6 endpoints and some optional objects. The optional objects are used for specifying some constraints like bandwidth, an `IRO` (Include Route Object) which holds the nodes the path (must) contain, whether the path may be load balanced[3] and if two computation requests should be synchronised with added constraints whether the second path should not traverse the same elements (Link, Node or SRLG) of the first path.

The `RP` object consists of some crucial information like a random `Request-ID` number and `flags` to define the characteristics of the request like, priority, whether its about a strict or a loose path, whether re-optimization is required and whether the path is uni- or bi-directional.

Re-optimization is the ability to change certain characteristics of an existing path; this e.g. could be changing bandwidth requirements or even re-routing the existing path through other nodes. If re-optimization is required the request must also contain a `bandwidth` object and a `RRO` (Record Route Object) containing the path it followed, the syntax of this object is the same as the `IRO` and the `ERO` described later on, however some sub-objects may be different.

Response Format

The response of a PCE to a PCC can be either positive or negative.

A response consists of a `RP` object, a `NO_PATH` object for paths that failed to compute and `ERO` objects for each successfully returned path.

In a negative response the `NO_PATH` object is accompanied with a `IRO`. An `IRO` contains the elements that could not comply with the requested requirements. This could be caused by for example policy restrictions or a failed network element.

In positive response the `ERO` represents a path accompanied with some optional objects, for representation of bandwidth and additional metrics.

ERO objects

An `ERO` (Explicit Route Object) is a RSVP (ReSerVation Protocol)[4] property and is defined in RFC3209[3], RFC3473[4] and RFC3477[13]. The `ERO` object consists of multiple sub-objects also defined in these RFCs. These sub-objects

[3]Load balanced in this context is that the PCE may return two paths of 5Gb when a path for 10Gb is requested.

[4]A resource reservation set-up protocol designed for an integrated services Internet[5]

have a bit to specify if it is about a loose or a strict[5] hop, and `type`, `length` and `value` fields. The following sub-objects are specified:

IPv4 Prefix (type 1): This contains an IPv4 address and a subnet mask to identify a network or host.

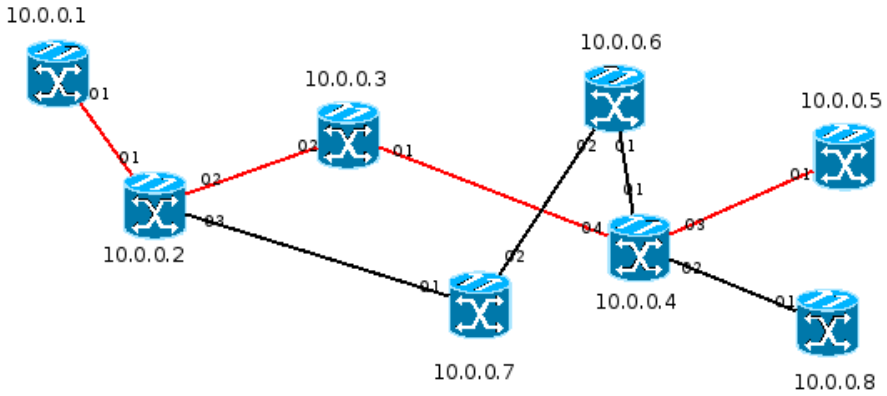
IPv6 Prefix (type 2): This contains an IPv6 address and a subnet mask to identify a network or host.

Label (type 3): This contains a label used in (G)MPLS context.

Unnumbered Interface ID (type 4): This contains a `ROUTER-ID`, commonly an IP address and the `interface ID` of the router.

Type 32: Autonomous System Number This sub-object contains an AS number.

Figure 3.3.5 shows a part of a network containing a path. The path in this network consists only of strict hops. The hops are identified by using the IP address of the router in combination with the interface number, that's why each hop is represented by a `subobject` of `type` 4 (Unnumbered interface ID).



The path goes from 10.0.0.1 to 10.0.0.5, therefore the `ERO` for this path looks as follows:

Strict/Loose	Type	ROUTER-ID	interface ID
Strict	4	10.0.0.1	1
Strict	4	10.0.0.2	2
Strict	4	10.0.0.3	1
Strict	4	10.0.0.4	3

Figure 3.4: ERO example

An `IRO` and `ERO` are practically the same because the protocol draft defines an `IRO` as an object which contains the same sub-objects as an `ERO`. The only difference is the functionality.

3.3.6 Discovery

PCE discovery can be implemented in two ways namely static or dynamic.

[5]Strict hops represent exactly one node; loose hops can represent multiple nodes.

With static configuration, the client has the PCE's location (IP-address) and capabilities statically configured. Another option is that only the location is statically configured and the capabilities are dynamically configured.

With dynamic configuration, the PCCs should be able to dynamically discover the location of PCEs in its domain and in case of inter-domain computation also PCEs in other domains. RFC4674 [11] describes the requirements for the PCE discovery protocol. The requirements are elaborated in [22] for OSPF and [21] for ISIS.

It turns out that the dynamic discovery of PCEs is handled by an IGP protocol, like OSPF-TE or IS-IS-TE, running on the control plane. PCEs, its capabilities and their status are announced inside the TLV (Type, Length, Value) messages carried in the IGP protocol e.g. Router Information LSAs [15] if the IGP protocol used is OSPF-TE.

There are two main types of information that can be carried in a Router Information LSA. `PCE Discovery Information` is used for announcing the location, the visibility, the scope and the neighbours of the PCE. `PCE Congestion Information` is optional and is used to report congestion and the estimated duration of the congestion.

RFC4657 [2] states that identification of PCC-PCE or PCE-PCE is based on IP addresses.

3.3.7 Manageability

There has to be a way to manage the PCEs or PCCs in a network. An administrator must be able to monitor the PCEs, to turn off some or all functionality and to change the application of policies. A part of this will be configurable though a SNMP MIB because this is a widely supported standardised interface.

As described in [25] the MIB is divided in three modules, the standard MIB, the PCEP protocol MIB, and the PCE discovery MIB. The function of the standard MIB is to create a root to hold the PCEP, the discovery MIB and to define some common objects.

An overview of functions in the PCEP MIB [14] are:

- Client configuration and status information.
- Peer configuration and information.
- Session configuration and information.
- Notifications to indicate session changes.

An overview of functions in the PCE discovery MIB [24]:

- The ability to turn off PCE discovery.
- How many PCEs are discovered and how.
- Information about of known PCEs and their ability.

- Congestion information of discovered PCEs.

It is very likely these MIBs are going to be extended as the PCE framework develops. More detailed information about the MIB contents is described in the above mentioned drafts.

3.4 Security Aspects of PCE

Security measures can be taken on two levels. One is the protocol level, where some initial measures can be taken to provide integrity and authentication. Second, policies can be applied to the PCE or PCC. This chapter provides information about some measures that can be taken to secure the PCE and PCC against different kind of attacks.

3.4.1 Protocol Security

The use of an external PCE, a PCE not running on a router itself, brings some security issues such as:

- The possibility to intercept PCE requests or responses.
- False impersonation of PCE or PCC.
- Falsification of PCE discovery, policy information or PCE capabilities.
- Information disclosure to non-authorised PCCs.
- Denial of Service attacks on the PCE.

These issues bring about the same risks as running an IGP protocol on the network and in an intra-domain network this could be controlled. However in an inter-domain network the security implications must be considered, because of the communication with a domain under control of another entity.

This leads to the following demands for the implementation.

- There must be a mechanism to authenticate discovery info
- There must be a method to verify discovery info
- There must be a method to encrypt discovery info
- There must be a method to restrict scope of discovery to a set of authorised PCCs and a filter at domain boundaries.

The PCEP protocol draft specifies how the IETF is planning to implement these details. To provide authentication and integrity of the messages and the information inside they are planning to use TCP-MD5 signature option like BGP does (RFC2385 [10]). IPSec tunnels can be used between PCEs to provide privacy and protection against sniffing. To protect against DoS attacks the IETF advises to use input shaping like throttling incoming PCEP messages and to use mechanisms as access-lists to only allow connections from authorised hosts.

3.4.2 Policies

Another important feature is the ability to apply policies on the requested data. In this way one is able to restrict the information given to another PCE or PCC. This is particularly useful for inter-domain communication and can prevent exposure of the network topology to a fellow domain.

RFC4655, describes two ways for managing policies. An external policy component can be set up to facilitate multiple PCEs or a PCE can keep its local policy database. Which policies one has to use is out of the scope but each PCE should be able to have is own policy information.

RFC4655 also separates three different types of policies:

User-specific: These policies look at the information of the user or service initiating the request, such as user-id or a VPN-ID. This could be implemented on either a PCE or PCC, but in order to implement this on the PCE the PCC should be able to provide the right information.

Request-specific: These policies look at the information in the request itself. This could be useful for adding extra constraints or diversities at the PCE side.

Domain-specific: These policies look at the domain of the requesting PCC and the domains involved in the resulting paths. In this way one can choose to restrict some functionality for a specific domain.

The policies are handled by a separate policy component. RFC4655 states that there are multiple options for how policy information is coordinated namely:

- Policy decisions may be made by PCCs before consulting PCEs. This type of decision includes selection of PCE, application of constraints, and interpretation of service requests.
- Policy decisions may be made independently at a PCE, or at each cooperating PCE. That is, the PCE(s) may make policy decisions independent of other policy decisions made at PCCs or other PCEs.
- There may also be explicit communication of policy information between PCC and PCE, or between PCEs to achieve some level of coordination of policy between entities. The type of information supports policies, has important implications on what policies may be applied on each PCE, and the requirements for the exchange of policy information inform the choice or implementation of communication protocols including PCC-PCE, PCE-PCE, and discovery protocols.

Chapter 4

Usability of PCE within SURFnet6

This section will explain the possible problem areas, the usability of the PCE for SURFnet6. It will also give an explanation of the changes SARA needs to make to their current planning tool to implement the PCE as designed by the IETF.

4.1 Possible problem areas

4.1.1 Reservations

In section 2.2 is an explanation of the current lightpath planning software. As explained in the mentioned section, the current software works with two elements to compute and reserve a path namely:

- Network Description Language.
- Network State Database.

In its current form, the network state database contains all the reserved network elements and the NDL contains information about the network topology. The combination of these two elements make sure the current software has the ability to make 'normal' reservations.

SARA makes a distinction between two kinds of reservation namely 'normal' reservations and future reservations.

Normal reservations

A normal reservation is the most expensive reservation in a network. If, for example, today is Monday and someone requests a lightpath from this Friday

until Friday next week. In its current form, the path is being calculated, reserved and set up instantly by the network administrators– four days early. The lightpath also needs to be broken-down by hand on the end-date. This means that if someone else wants a path over the same network elements from Monday until Thursday, something that is theoretically possible, the planning tool will return that it cannot be done because the path is already reserved – the one from Friday until Friday next week. In this situation, network elements are unnecessarily occupied.

Future reservations

Future reservations are slightly different but have another problem. Take the same example as in the previous section. With future reservations, the reservation will become active (either by hand or automated) on the day it needs to be active. The problem is that, in order to support future reservations, the PCE must have the ability to look in the future – it needs to see on a certain date which network elements are in use at that time.

PCE however, does not support future reservations of any kind. There have been discussions on the IETF's PCE mailing list[1] in May 2006 between Lucy Yong and the members of the PCE working group at the IETF, Dimitri Papadimitriou, J.P. Vasseur and Adrian Farrel. In one of the last posts Dimitri Papadimitriou states:

> *"[...] i can understand that some NRENs, research nets, etc. are looking at LSP / resource scheduling functionality (... the mythical resource broker) for single domain application specific networks but PCE is not meant to incorporate such functionality [...]"*

We contacted Lucy Yong, because she was also interested in a 'reservation' ability of the PCE. She replied:

> *"As I understand, that PCE currently does not consider reservation capability and people still think that function belongs to NMS.*
> *...*
> *I thing as PCE development moves forward, at some point, this issue has to be addressed."*

This confirms the need for such a feature in PCE.

The subject has been addressed once more in a discussion of the CCAMP working group which also consists of members of the PCE working group [2] in March 2007 but the outcome is that PCE, in its current form, will not support any kind of future reservation. However it does state that the members of the CCAMP working group are considering to implement 'future reservations' if there is enough interest for it.

The limitation here exists mainly in the protocol, because, as told in section 3.3.2, the TED can be self-constructed in such a way that it can support future

[1] `http://www1.IETF.org/mail-archive/web/PCE/current/msg00749.html`
[2] minutes-68 – `http://tools.IETF.org/wg/ccamp/minutes?item=minutes68.html`

scenarios. But if the PCEP protocol does not support sending time constraints in a request, one is not able to calculate a path in the future.

Lucy Yong, active member of the CCAMP working group, has written a draft [29] that describes how future reservations can be implemented as a separate entity a `Reservation System`. This RS is responsible for sending a path request to the GMPLS control plane at the appropriate time. However, it looks like the RS does not guarantee if the requested amount of bandwidth is available at the time. This means it is possible to use this approach in a network which just uses solely a RS for making path requests, losing the ability to request a path on the fly. Also GMPLS has to be implemented in order for it to work.

SARA and SURFnet could point out to the CCAMP working group that they need 'future reservations', so that the IETF is aware of the demand for this feature.

4.2 Overall benefits

If SARA decides to implement the PCE in their planning tool, it would mean that SARA is using an IETF standard.

Thereby, if other (research) networks implement the PCE in their network with the same guidelines, they can have the ability to use inter-domain path calculation to find a path through the SURFnet6 to other networks.

Inter-layer path computation could also be useful for the SURFnet6 network. If a PCE is used on multiple network layers, traffic engineering becomes more effective because the network traffic can now be engineered over all layers.

4.3 Implementation considerations

This section will be a reflection on the PCE and will explain if the PCE is usable for SARA's planning tool. We will discuss the implementation considerations SARA has to take into account when they want to implement the PCE architecture.

4.3.1 Mandatory implementations

Implementation of a PCC

There has to be some implementation of a PCC in order to communicate with the PCE and to be able to send path computations requests. The current planning tool already has something similar to a PCC (the web interface).

Implementation of the PCEP protocol

In its current form, SARA's planning tool has a simple communication protocol between the PCC and the PCE. The client currently exists of a web interface or a telnet client and the requests and replies consist of a set of human-readable strings. There has to be emphasised that the status of the PCEP protocol is still an Internet-Draft and it could be that changes occur before it becomes an Internet standard. Another thing is that, with the implementation of the PCEP protocol, SARA is unable to plan future reservations.

Implement the constraints

The current planning tool has to be adapted to support the constraints described in section 3.3.3. The support SARA has to add includes SRLG, the ability to restrict hop count, and the handling of priorities.

Implement policies

To comply with the IETF's PCE standard, policies have to be implemented in the planning tool to add a level of security and restrict the ability to request information about the network structure.

4.3.2 Optional implementations

Implementation of an IGP-TE protocol

If SARA wants to enable the dynamic discovery of PCEs or wants to use the inter-PCE functionality (section 3.3.4), SARA needs to implement an IGP-TE protocol such as OSPF-TE or IS-IS-TE in the SURFnet6 network. This is because the PCE will announce itself through this protocol (section 3.3.6) which makes sure the PCC can find the PCE and also gets the knowledge on which layer the announced PCE(s) can compute paths (section 3.3.4). If other domains run the same IGP-TE protocol, they will be able to use the SURFnet PCE to compute paths through their network.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

To conclude this document we will give the answer to the sub-questions mentions in section 2.3.

How does the IETF's PCE retrieve its knowledge of the network?
The PCE receives its knowledge of the network from the Traffic Engineering Database (or TED). The TED can be constructed by an IGP-TE protocol such as OSPF-TE or IS-IS-TE or can be constructed manually (see section 3.3.2).

How does IETF's PCE retrieve the status of the network elements?
The PCE receives the status of the network elements from the TED. If an IGP-TE protocol is implemented in the network the TED is automatically updated and if the TED in constructed manually, the TED has to be updated manually (see section 3.3.2).

How does IETF's PCE find a path?
To find a path, the PCE utilises a Constraint based Shortest Path First (CSPF) algorithm. When a path request is committed, the constraints are applied to the network topology. Then Dijkstra's shortest path first algorithm is applied to find the shortest path (see section 2.2.1).

How does IETF's PCE return a found path?
The PCE returns a path in `ERO` objects. Multiple `ERO` objects can be attached to a response. If the answer is negative, it returns a `NO_PATH` object together with an `IRO` containing the elements that failed (see section 3.3.5).

Does IETF's PCE look at Shared Risk Link Groups (SRLG)?
The PCE is designed to look at SRLGs. In case of a protected path, the SRLG option is responsible that the lightpaths don't go through the same fibres or fibreducts (see sections 2.2 and 3.3.5).

Which additional advantages can be gained by using the IETF PCE architecture?
If SARA wants to implement PCE in the SURFnet6 network, an additional advantage is the inter-domain and inter-layer path computation options (see section 3.3.4). There has to be emphasised that an IGP-TE protocol has to be implemented before the inter-PCE functionality is usable.

PCE can certainly be an interesting extension for SARA's planning tool, but still a lot of work is required to fit the IETF's requirements. The current planning tool is used in a more static context and IETF's PCE is designed to be very dynamic and optimised for use with (G)MPLS and routing protocols with Traffic Engineering extensions, like OSPF-TE or IS-IS-TE – something that yet has to be implemented on the SURFnet6 infrastructure.

A disadvantage is that the dynamic approach of the PCE adds significant complexity to the protocol, while the protocol used by SARA's planning tool is much easier to understand and implement and has all the functionality that is currently needed.

However, because of SURFnet's partners and its use of multiple network layers, inter-domain and inter-layer path computation is a very interesting subject. The standardised PCE protocol can facilitate easy communication with SURFnet's partners and will cause better interoperability between the networks – something that will be addressed more in the future. The properties of inter-layer computation can add to the efficiency of the currently running platforms. We have to emphasise that additional research to this subject is required because of the security implications with the implementation of an IGP-TE protocol.

On the long term, we think the dynamic properties of a Traffic Engineering routing protocol combined with PCE can certainly be of interest for the SURFnet6 network. This also requires the implementation of a TE routing protocol.

On the short term, the PCE can be used in a static context and it could make the transition to a dynamic environment more convenient.

Considering the added value of PCE is desirable, we think that SARA should starts with the adaptation of their planning tool to meet the requirements of the Path Computation Element.

5.2 Future Work

5.2.1 Inter-layer and Inter-domain

Inter-layer and inter-domain path computation is very interesting for the SURFnet6 network but this part of PCE is still open for research. How does inter-domain and inter-domain path computation work in a non-GMPLS enabled network? How can this technique be implemented in SURFnet6? There has been done research on inter-layer path computation by the University of Amsterdam[7] based on ITU-T G.805. Maybe this in interesting for inter-layer path computation based on PCE?

5.2.2 Algorithms

The choice for CSPF is the most obvious choice because it works with the current description of the PCE. It could be that it is feasible to implement different CSPFs for different path computations (i.e. a CSPF per layer). Maybe there is a way to speed up the current implementations of CSPF. The algorithm could, for example, be faster by changing the order of the constraints. Some research in this area is suggested.

5.2.3 Policies

The PCE is able to work with policies for path computation. These policies can for example exist of priority policies (which request goes first), bandwidth policies (who gets more or less bandwidth) or inter-domain policies (how much information is given with a path computation request). It could be feasible to research which policies are needed in SURFnet6 with priority, usability and feasibility in mind.

Bibliography

[1] Optical fiber - wikipedia, the free encyclopedia, June 2007.

[2] J. Ash and J. L. Le Roux. RFC4657 - path computation element (pce) communication protocol generic requirements, September 2006.

[3] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, and G. Swallow. RFC3209 - rsvp-te: Extensions to rsvp for lsp tunnels, December 2001.

[4] L. Berger. RFC3473 - generalized multi-protocol label switching (gmpls) signaling resource reservation protocol-traffic engineering (rsvp-te) extensions, 2003.

[5] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin. RFC2205 - resource reservation protocol (rsvp), September 1997.

[6] Cees de Laat. Why is optical networking interesting?, 2002.

[7] Freek Dijkstra, Bert Andree, Karst Koymans, Jeroen van der Ham, and Cees de Laat. A multi-layer network model based on itu-t g.805, May 2007.

[8] A. Farrel, J. P. Vasseur, and J. Ash. RFC4655 - a path computation element (pce)-based architecture, August 2006.

[9] F. Le Faucheur, R. Uppili, A. Vedrenne, P. Merckx, and T. Telkamp. RFC3785 - use of interior gateway protocol (igp) metric as a second mpls traffic engineering (te) metric, May 2004.

[10] A. Heffernan. RFC2385 - protection of bgp sessions via the tcp md5 signature option, August 1998.

[11] Ed. J.L. Le Roux. RFC4674 - requirements for path computation element (pce) discovery, October 2006.

[12] D. Katz, K. Kompella, and D. Yeung. RFC3630 - traffic engineering (te) extensions to ospf version 2, September 2003.

[13] K. Kompella and Y. Rekhter. RFC3477 - signalling unnumbered links in resource reservation protocol - traffic engineering (rsvp-te), 2003.

[14] A. S. Kiran Koushik and E. Stephan. Internet-draft - pce communication protocol(pcep) management information base. draft-kkoushik-pce-pcep-mib-00 (work in progress).

[15] A. Lindem, N. Shen, J. P. Vasseur, R. Aggarwal, and S. Shaffer. Internet-draft - extensions to ospf for advertising optional router capabilities, May 2007. draft-ietf-ospf-cap-11 (work in progress).

[16] Eiji Oki. Internet-draft - pcc-pce communication requirements for inter-layer traffic engineering, March 2007. draft-ietf-pce-inter-layer-req-04 (work in progress).

[17] Eiji Oki, J. L. Le Roux, and A. Farrel. Internet-draft - framework for pce-based inter-layer mpls and gmpls traffic engineering, March 2007. draft-ietf-pce-inter-layer-frwk-03 (work in progress).

[18] P. Pan, D. Gan, G. Swallow, J. P. Vasseur, Dave Cooper, A. Atlas, and M. Jork. Internet-draft - fast reroute extensions to rsvp-te for lsp tunnels, August 2003. draft-ietf-mpls-rsvp-lsp-fastreroute-02 (work in progess).

[19] P. Pan, G. Swallow, and A. Atlas. RFC4090 - fast reroute extensions to rsvp-te for lsp tunnels, May 2005.

[20] J. L. Le Roux. Internet-draft - pce communication protocol (pcecp) specific requirements for inter-area multi protocol label switching (mpls) and generalized mpls (gmpls) traffic engineering, December 2006. draft-ietf-pce-pcecp-interarea-reqs-05 (work in progress).

[21] J. L. Le Roux, J. P. Vasseur, Yuichi Ikejiri, and Raymond Zhang. Internet-draft - is-is protocol extensions for path computation element (pce) discovery, May 2007. draft-ietf-pce-disco-proto-isis-05 (work in progress).

[22] J. L. Le Roux, J. P. Vasseur, Yuichi Ikejiri, and Raymond Zhang. Internet-draft - ospf protocol extensions for path computation element (pce) discovery, May 2007. draft-ietf-pce-disco-proto-ospf-05 (work in progress).

[23] H. Smit and T. Li. RFC3784 - intermediate system to intermediate system (is-is) extensions for traffic engineering (te), June 2004.

[24] E. Stephan. Internet-draft - definitions of managed objects for path computation element discovery, March 2007. draft-ietf-pce-disc-mib-01 (work in progress).

[25] E. Stephan. Internet-draft - definitions of textual conventions for path computation element, March 2007. draft-ietf-pce-tc-mib-01 (work in progress).

[26] Jeroen van der Ham, Paola Grosso, Ronald van der Pol, Andree Toonk, and Cees de Laat. Using the network description language in optical networks, 2007.

[27] Ronald van der Pol and Andree Toonk. Light path planning and monitoring in surfnet6 and netherlight, 2007.

[28] JP. Vasseur and JL. Le Roux. Internet-draft - path computation element (pce) communication protocol (pcep), March 2007. draft-ietf-pce-pcep-07 (work in progress).

[29] L. Yong and Y. Lee. Internet-draft - ason/gmpls extension for reservation and time based automatic bandwidth, October 2006. draft-yong-ccamp-ason-gmpls-autobw-service-00 (work in progess).