End-to-End Quality of Service Engineering in Next Generation Heterogenous Networks
Edited by Abdelhamid Mellouk

Chapter 6

Pushing Quality of Service Across Inter-domain Boundaries

In this chapter, we introduce the current techniques and the remaining challenges for establishing inter-AS LSPs with QoS guarantees. We describe the workings of the inter-domain routing system. We discuss the consequences of path selection made by the current inter-domain routing system on the visibility of the paths. The limited visibility of path diversity does not actually prevent the establishment of inter-AS LSPs with QoS guarantees. Rather, the lack of QoS information requires clever heuristics to be designed in order to guide the search towards feasible QoS paths. We cover the existing signaling extensions to RSVP-TE that support the establishment of inter-AS LSPs, as well as the protection of those LSPs. The path computation techniques that have been proposed at the IETF are also detailed. Such computation techniques make it possible to find the LSP segments within each AS, in order to compose an end-to-end LSP with QoS guarantees when the sequence of ASs to be crossed is known. Finally, combine these three components, i.e. inter-domain routing, LSP signaling and path computation techniques. We show that inter-AS QoS is not beyond reach, but that more work needs to be done in specific areas, especially concerning heuristics to guide the search towards AS sequences across which feasible QoS paths can be found.

6.1. Introduction

Today's Internet essentially provides a best-effort service. More stringent services like virtual private networks (VPN) have recently been deployed [ROS 06], but

---

Chapter written by Bingjie FU, Cristel PELSSER and Steve UHLIG.

crossing autonomous system (AS) boundaries has proven to be difficult. An AS is a network under a single administrative authority. Typically, an autonomous system will correspond to the network of an Internet service provider (ISP), an enterprise or an academic network. The ability to provide QoS guarantees enables ISPs to propose new services and consequently leads to new sources of revenue for them. MultiProtocol Label Switching (MPLS) makes it possible to set up Label Switched Paths (LSPs) with QoS guarantees in a single AS. The traffic engineering (TE) extensions to the ISIS/OSPF routing protocol(s) [KAT 03, SMI 04] enable the intra-domain routing protocols to distribute the TE properties of the links. Once the topology and TE information is known, paths which meet the QoS guarantees can be easily computed with the Constrained Shortest Paths First (CSPF) algorithm or more elaborate algorithms like SAMCRA [VAN 01] and DAMOTE [BLA 03]. QoS may be provided to flows by establishing MPLS LSPs with RSVP-TE [FAR 07] along paths with certain properties. The QoS routing algorithm ensures that the selected path meets the QoS requirements. RSVP-TE makes it possible to reserve resources to guarantee the QoS that will be experienced through time. When the LSPs need to traverse multiple ASs administered by different ISPs, new path computation techniques, together with protocol extensions for establishing LSPs, need to be introduced.

In section 6.2 of this chapter, we first introduce the current routing system used for the Internet. We discuss its implications for the establishment of LSPs with QoS constraints across multiple domains. The methods for signaling the establishment of inter-domain LSPs are introduced in section 6.3. We review the work done in the Path Computation Element (PCE) working group (WG) of the Internet Engineering Task Force (IETF) and discuss the applicability of the paths proposed for computation architecture and techniques in section 6.4. In section 6.5, we discuss the solutions we envision for the problem of inter-domain LSP establishment and the issues that remain to be solved to make inter-domain QoS a reality. In section 6.6, we give some final views of our own and provide a conclusion to this study.

6.2. Background

The Internet, which initially connected only a few tens of routers, has grown exponentially in recent decades, resulting in a huge, distributed system composed of more than 25,000 autonomous systems (ASs) operated by different ISPs [HUS]. Originally, the Internet was a research network built to provide reachability and robustness. With its increasing use for commercial purposes, more stringent applications have been deployed, like voice-over IP (VoIP), video on demand (VOD) and virtual private networks (VPNs). These services require specific QoS demands to be guaranteed by the network. In order to provide such services, ISPs need to implement functions enabling QoS, such as MPLS, TE extensions to ISIS/OSPF and RSVP-TE. Some services may require multiple ASs to cooperate before providing

consistent guarantees across ASs. This introduces new operational and management challenges that do not forcibly fit with the original design of the best-effort Internet.

If several solutions to the QoS problem exist, ISPs would prefer to choose the one that is closest to their current network design. New functions, which can be realized incrementally based on the current system, are more likely to be deployed, especially in the short term. In this chapter, we build upon the existing features to show how we can evolve towards a QoS-capable Internet.

6.2.1. *The Internet as a distributed system*

The ideal situation would be for a certain router to know about the topology information of all ASs, to rely on CSPF or another QoS routing algorithm to compute the path based on the complete topological information, and to establish the LSP according to the path found by the QoS routing algorithm. It is not applicable in the general inter-domain framework, as ASs hide their internal structure from competitor ASs for security and business reasons [SPR 04, ZHA 05]; and at the same time, routing using a QoS routing algorithm on the complete topology and QoS information would not scale. As a consequence, a single node cannot compute an end-to-end path composed of individual routers for an LSP crossing multiple ASs. Instead, the computation of a path has to be distributed among multiple path computation elements nodes[1], where each node computes a segment of the path based on its knowledge of the local AS topology, the inter-domain reachability information provided by BGP and other information specifically required by the segment computation (like constraints on the IP-level path). The establishment of LSPs also needs to be distributed among the ASs crossed. Each AS is in charge of the establishment of LSP segments within its own network.

6.2.2. *Business relationships between ASs*

To obtain reachability across the whole Internet, ASs connect to each other through peering relationships. When two ASs are interested in connecting with each other, they negotiate and sign a contract that will rule the conditions under which the connectivity between them is to be used, such as how traffic will be exchanged, how the billing will be made, etc. A peering relationship between two ASs consists of a direct connection between them, either at a private location or a public exchange point. One or several physical links connect their border routers, and the border routers need to be configured consistently to the agreements of the peering contract.

Each inter-domain link, which connects two ASs, can be classified as one out of three types of business relationships, namely customer-to-provider,

1. An abstract node identifies a set of Label Switching Routers (LSRs).

provider-to-customer and peer-to-peer. Large ASs provide a transit service to their customer ASs. In addition, they may join with each other as peers to transit traffic between each other's customers, with shared costs and mutual benefits. Small ASs need to pay for the transit service provided by larger ASs to have Internet-wide reachability. Different techniques to deduce these business relationships have been proposed in the literature [GAO 00, XIA 04, DIM 07, BAT 07]. These techniques provide simplified business relationships [CAE 05], but modeling the reality of business relationships and routing policies based on the available data is very complex [MÜH 07]. Note that deduced business relationships are typically blind to multiple physical links between two ASs.

6.2.3. *Impact of inter-domain routing on path diversity*

The Internet uses the border gateway protocol (BGP) as the inter-domain routing protocol. The role of BGP is to propagate reachability information across the whole Internet. The way routing information is exchanged between BGP routers and the paths selected by BGP both depend on routing policies [GAO 00]. Routing policies [CAE 05] implement the business objectives (see section 6.2.2) of each ISP by modifying the flow of routing information and the path selection made by the BGP. As will be explained in this section, two factors influence the distinct routes known by a BGP router: business relationships and the best path selection of the BGP.

6.2.3.1. *BGP path selection*

Figure 6.1 illustrates how the BGP selects the best paths and propagates reachability information. For each destination prefix, i.e. a block of reachable IP addresses, a BGP router receives the best AS-paths from its neighboring BGP routers. An AS-path is a sequence of intermediate ASs which form a direct route for traffic from the source to reach the destination. For each prefix, a BGP router selects its best path in the following manner:

– it discards AS-paths having its own AS number to avoid loops;

– it filters AS-paths according to the import policies. That is, a router keeps the AS-paths which pass the import filters in the RIB-Ins, (see Figure 6.1, where *R1* only keeps the routes learned from *R10*, *R12* and *R13*);

– it selects its best AS-path from the AS-paths in the RIB-Ins (the route learned from *R12* in Figure 6.1) according to the BGP decision process.

RIB-In stands for routing information base-inbound, and contains all the valid routes that passed the import filters. The routes that are stored in the RIB-Ins then undergo the BGP decision process, which selects a single route towards any destination prefix. As shown in Figure 6.1, the BGP decision process is composed of a list of criteria (called "rules"). It aims at choosing a single route, called "best route", from a given set of routes tending towards a given destination. The criteria of

the decision process will reduce the set of candidate routes, by selecting only those that have the best values of the attribute considered. We present here a simplified version of the BGP decision process, with 7 rules. For details about the BGP decision process, see [ZHA 03].

After selecting the best route, the BGP router attaches the AS number of its AS to the AS-path and propagates this AS-path to its neighboring BGP routers according to the export policies (outbound filters). In Figure 6.1, the route is not advertised to *R31*. Although the BGP router may learn many AS-paths from its peers, *only one* is selected as the best AS-path and used to forward traffic towards each destination prefix. Moreover, only one is advertised to the neighbors.

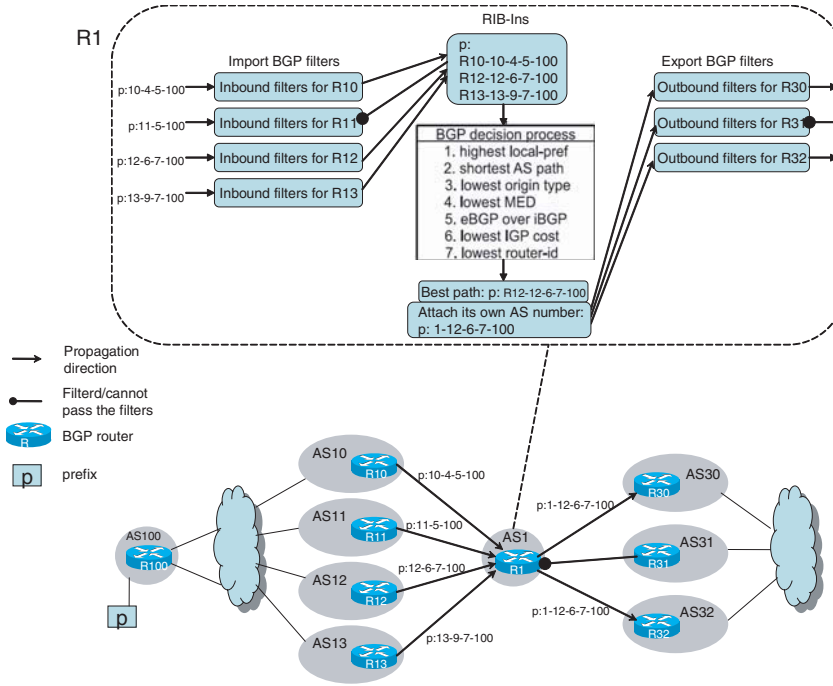

Figure 6.1. *BGP path selection and propagation*

6.2.3.2. *Routing policies*

Routes learned from different types of peers tend to be considered differently. For example, a route learned from a customer will be preferred to a route received from a peer, with a route learned from a provider being the least preferred. The reason for doing this is the cost associated with the use of a neighbor to forward traffic; that is, an AS receives money to provide connectivity to its customers, shares the cost of the traffic exchanged with its peers, and pays its providers to send traffic

through them. This preference among differently learned routes can be transferred into a value of the `local-pref` attribute of the BGP routes. As shown in Figure 6.1, the `local-pref` attribute is used in the first step of the BGP decision process. The value of the `local-pref` attribute will typically be assigned to a route upon reception of a route from a peer.

Business relationships also introduce constraints on BGP route propagation, e.g., the so called valley-free route propagation property [GAO 00]. This property is verified [MÜH 07] because ASs typically want to avoid propagating routes which result in using themselves to transit traffic between two larger ASs [FEA 04]. An AS propagates routes to its providers, but only those learned from its customers, not the routes learned from its other providers or peers. An AS propagates routes to its peers, but only the routes learned from its customers, not the routes learned from its providers and other peers. An AS propagates all its routes to its customers, that is, the routes learned from its customers, providers and peers.

Other routing policies also introduce constraints on the paths known by a router. For instance, ASs may announce only subsets of their prefixes to their neighbors in order to control the amount of incoming traffic they receive from them [WAN 03].

The AS-paths received by a BGP router from its neighbors may only be a small subset of all the AS-paths its neighbors know, because each BGP router advertises only its best paths. Moreover, because of the redundant connectivity in the intra-domain and/or the inter-domain topology, the IP-level path to which an AS-path corresponds is one out of possibly many IP-level paths consistent with this AS-level path.

Thus, the AS-paths seen by a BGP router (in the RIB-Ins) towards a prefix, represent only a fraction of all the usable connectivity. By usable connectivity, we mean the IP-level paths that are consistent with routing policies. In the next section, we show the routing diversity obtained from simulations on a topology with routing policies extracted from observed data.

6.2.3.3. *Observed Internet routing diversity*

In this section, we present a lower bound on the available routing diversity in the Internet. We borrow from Mühlbauer *et al*. [MÜH 07] the inter-domain connectivity, extracted using BGP data from more than 1,300 BGP observation points, including those provided by RIPE NCC [ROU a], Routeviews [ROU b], GEANT [INT], and Abilene [ABI]. The AS-level model from [MÜH 07] allows each AS to be composed of multiple entities, in order to capture the intra-domain routing diversity. The AS-level connectivity is minimal while enabling the propagation of all AS-paths observed in the BGP data. This minimal connectivity might lead to an underestimation of the actual number of AS-paths learned by an AS. Consequently, the results shown in this section only provide a lower bound on path diversity in the Internet.

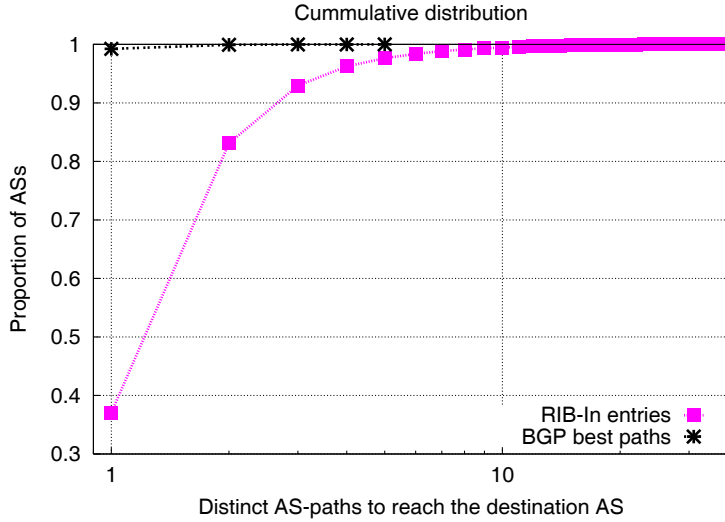

Figure 6.2. *Number of AS-paths learned/selected to reach a certain destination AS*

The BGP path selections at the routers are simulated on the above connectivity. The only policies configured are those preventing valley-free paths to propagate. Business relationships are inferred using the CSP algorithm [HAO 06] on the data used in [MÜH 07]. Because it has been observed in [MÜH 07] that trying to implement route preferences consistent with the inferred business relationships leads to path choices in the simulations that are often inconsistent with the paths observed in the BGP data, we only prevent non-valley-free paths from propagating and leave the shortest AS paths to propagate.

We rely on C-BGP [QUO 05] to calculate the outcome of the BGP decision process and the set of learned routes at every router of the AS-level connectivity. We present the results for a single randomly chosen AS destination, even though we tried different AS destinations and obtained similar distributions for all of them. Simulation results for the selected AS destination are shown in Figure 6.2. On the x-axis of Figure 6.2 we show the number of distinct AS paths that are observed in all routers of an AS towards a given destination AS. On the y-axis, we show the cumulative fraction of ASs that locate x or less distinct AS paths towards the destination AS. There are two curves in Figure 6.2: one for the best paths selected by the BGP routers in the simulation, and another for the routes in RIB-Ins of the BGP routers of the simulation. Observations from BGP data only provide information concerning best routes selected by BGP routers, not RIB-Ins. The upper curve of Figure 6.2 shows that more than 99% of the ASs only have a single AS path to reach the given AS. If the RIB-Ins of the routers are considered, on the other hand, more path diversity is available for large ASs. For the

AS destination considered, about 36% of all the ASs learn only 1 AS-path. These ASs are probably stub AS, that have a single upstream provider, hence they cannot learn more than a single AS path towards any destination. 64% of all ASs learn multiple AS-paths towards the destination AS, with some of them knowing up to 38 distinct AS-paths. This is a lower bound on the actual number of AS paths that are known by ASs compared to reality. Thus, we expect that even greater diversity is, in practice, hidden in the BGP routers.

The previous results indicate that best path selection, and the fact that the BGP routers only advertise their best route, significantly reduces the path diversity visible in the routers. The actual path diversity that might be usable in practice (RIB-Ins) is highly underestimated because it is not completely visible from BGP data. Because BGP data only gives a limited view of the Internet, using the BGP data to determine the RIB-Ins will inevitably give only a lower bound indication. The poor visibility of path diversity does not mean that good QoS paths cannot be found if the path diversity actually present in current BGP routers can be exploited.

6.2.4. *Inter-AS LSP requirements*

As shown in the previous section (section 6.2.3), the selection of the best BGP route does not depend on the quality of the path in terms of delay, bandwidth, etc. This is verified by the literature [HUF 02, BOL 93, SAV 99, ZHA 01]. Thus, the best BGP route may not be suitable for a particular type of traffic with given QoS requirements. To provide specific QoS guarantees, it might be necessary to use paths different from those chosen by BGP. One solution for this is to use MPLS LSPs with RSVP-TE. In this section, we discuss the requirements for inter-AS LSPs.

In [ZHA 05], ISPs expressed several requirements for MPLS inter-AS traffic engineering (TE). Among these requirements is the ISP's desire to keep internal AS resources and the set of hops followed by the TE-LSP confidential. This confidentiality requirement entails the constraint of only partly specifying the hops that the TE-LSP must traverse, since global topology information is not available. Moreover, it must be possible to perform path optimization inside each transited AS, where the required information is available. In addition, end-to-end optimization of inter-AS LSPs is also required by ISPs.

A second requirement, the protection requirement, concerns the restoration capabilities of inter-AS LSPs. The proposed solution has to be able to provide rapid local protection against link, shared risk link group (SRLG) and node failures. An SRLG is a group of links that may fail at the same time. It is a set of links that share a common physical resource such as an Ethernet switch, a fiber, an optical cross-connect, etc. Additionally, the proposed solution should support the establishment of multiple link/SRLG/node diversely routed inter-AS TE LSPs between a pair of LSRs.

A last requirement, the scalability requirement, is that the proposed solution should be scalable in terms of the amount of IGP flooding, the additional information carried by BGP, the amount of RSVP-TE signaling messages exchanged and state to retain.

6.3. RSVP-TE extensions to support inter-domain LSPs

In this section, we discuss extensions to enable the establishment of traffic engineered inter-domain LSPs towards a prefix destination. The local or global protection of these inter-domain LSPs is discussed afterwards.

We present the extensions to RSVP-TE proposed in [PEL 03] that fulfill both the confidentiality and the protection requirements concurrently, while trying to keep the solution scalable. This solution also attempts to impact only the head-end LSR, the intermediate AS border routers (ASBRs) on the path of the inter-AS LSP and the tail-end LSR of the LSP therefore allowing a smooth migration towards the support of inter-AS LSPs. It does not impact the current BGP and MPLS traffic engineering techniques. Moreover, it does not require additional IGP flooding.

6.3.1. *Explicit routing of an LSP*

The explicit route object (ERO) is well-suited to the establishment of inter-AS LSPs, in that it enables the head-end of the LSP to partially specify the path to be followed by the LSP. Following nodes crossed by the `Path` message enable us to achieve this objective as the `Path` message goes along. More precisely, the head-end LSR of an inter-AS LSP is only able to fill the ERO with nodes that belong to the same AS and eventually with the list of ASs that will be crossed by the `Path` message. At the entrance of each AS, the ASBR computes the path of the LSP towards the downstream AS and completes the ERO accordingly. This process is illustrated in Figure 6.3. We see that *R0* computes the path towards *AS1* and sets the ERO accordingly. Inside *AS1*, *R3* completes the ERO towards the next AS, *AS2* and so on. These paths are computed based on the LSP's destination address. The ERO specifies only a set of hops on the path of the inter-AS LSP and it leaves the responsibility of the local path optimization to each crossed AS, according to a set of constraints also carried inside the `Path` message of the LSP. This fulfills the local path optimization requirement from the second paragraph of section 6.2.4. A mechanism for the reoptimization of loosely routed LSPs signaled with RSVP-TE is defined in [VAS 06].

The ERO object may be constructed at the head-end LSR and is either based on a manual configuration that specifies the ASs and/or the ASBRs to be crossed by the LSP, based on the BGP routing table, or based on the path computation result of a path computation entity such as the Path Computation Element (PCE) introduced in section 6.4.

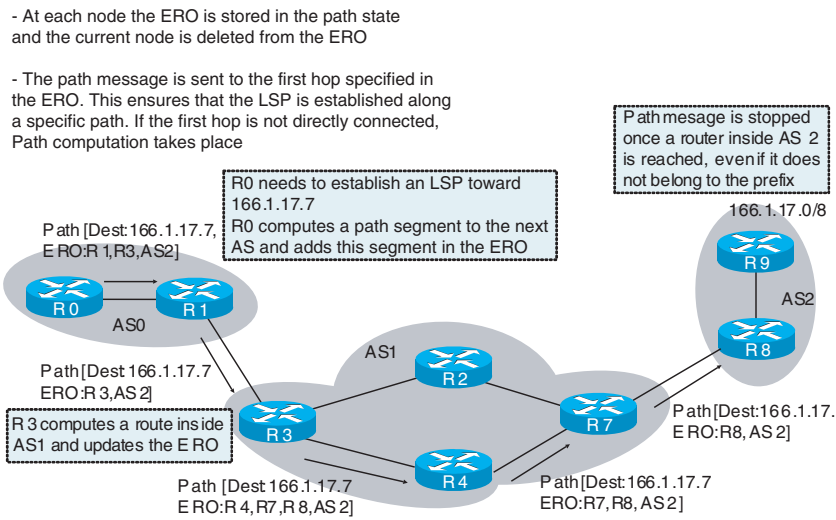

Figure 6.3. *Establishment of an inter-AS LSP*

The inter-domain path selection could be performed by relying on QoS information distributed by extensions to BGP. Such QoS enabled advertisements were proposed in [XIA 02] and are still being researched. Later in this chapter, we look more deeply into inter-domain path selection techniques.

6.3.2. *RRO aggregation and the path key*

The Record Route Object (RRO) enables us to obtain the path followed by an LSP. It is used to detect loops within the LSP's path, in order to pin the LSP onto its path and to compute the LSP disjoint from this LSP for global or local protection.

We note that recording the path of an inter-AS LSP may be in contradiction to the desire of the ISPs to hide the internal topology of ASs. Therefore, we proposed in [PEL 03] to modify the processing of this object at the ASBRs so as to withhold the complete path followed by the LSP inside the current AS from neighboring ASs. We call this process "RRO aggregation".

RRO aggregation consists of marking the sub-object added by the ingress ASBR inside the AS. Thus, at the last router of the AS, i.e. the egress ASBR, the list of nodes in the AS are removed from the RRO. These sub-objects are replaced by the address of the ingress ASBR, the AS number and the address of the egress ASBR in order to retain enough information to perform loop detection, disjoint path computation and route pinning of the inter-AS LSP. We use the example topology in Figure 6.3 to illustrate RRO aggregation. In *AS1*, the ingress ASBR *R3* adds its address inside the

RRO and marks it. The following LSRs (*R4* and *R7*) add their addresses inside the RRO. The egress ASBR, *R7* in Figure 6.3, removes all addresses starting from the marked sub-object, representing the address of *R3*. It replaces these sub-objects with the address of the ingress ASBR (*R3*), its AS number (*AS1*) and its own address (*R7*).

[BRA 07] proposes using path keys to fulfill the confidentiality requirement of ISPs. The path key contains the identifier of the node that knows the list of nodes composing the confidential path segment. Path key sub-objects (PKS) can be stored inside the ERO of the RSVP path messages. Such sub-objects must follow the node responsible for expanding the path key, that is, the first node of the confidential path segment. This node sends the path key to the node with an identifier contained in the PKS for expansion of the path key into a sequence of nodes.

6.3.3. *Protection of inter-AS LSPs*

Restoration capabilities need to be provided to inter-AS LSPs against link, node and SRLG failures. In this section, we introduce the local protection of inter-AS LSPs [PEL 03]. The possibility of establishing complete link or node disjoint LSPs can be useful to balance traffic on these disjoint LSPs, or provide reliability against failures. We suggest means of addressing this problem in section 6.3.4.

6.3.3.1. *Local protection of inter-AS LSPs*

Techniques to protect AS core nodes and links joining these nodes are described in [PAN 05]. The protection of links connecting distinct ASs, called “inter-domain” links, is discussed in [PEL 02]. These techniques can be combined with those described in [PAN 05] to protect inter-AS LSPs all the way along their path.

Here, we use an example (Figure 6.4) to illustrate how to locally protect inter-AS LSPs against the failure of the egress ASBR (*R13*) and of the SRLGs of the link preceding this ASBR (*R11 - R13*). This problem is best solved by using two detour LSPs at the node *R11* on the path of the working LSP, preceding the egress ASBR. A detour LSP protects against the SRLGs of the intra-AS link *R11 - R13*. A second detour LSP protects against the egress ASBR failure. The detour protecting against the SRLGs has to merge in the same AS as the link it is required to protect, i.e. it has to merge with the working LSP at the egress ASBR *R13*. This is because other ASs neither know this intra-AS link nor its SRLGs. The detour protecting against the egress ASBR needs to exclude *R13* and merge with the working LSP in the next AS, *AS2*.

For local protection against other resources such as the ingress ASBR, see [DE 04].

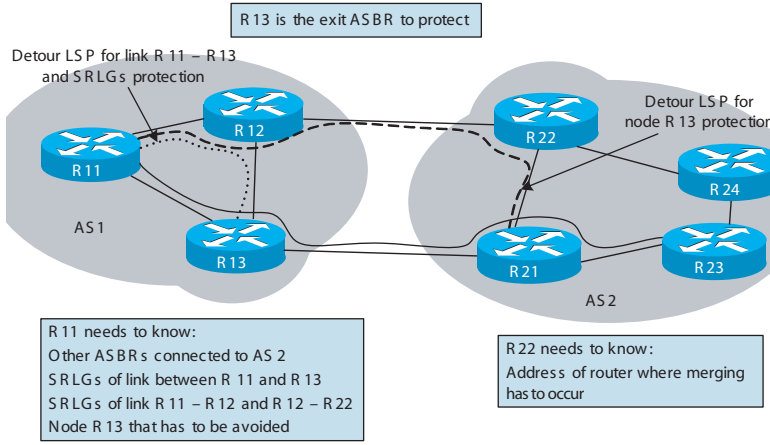

Figure 6.4. *Local protection against egress ASBR and against the SRLGs of the link preceding this ASBR*

An LSP that is used to protect a set of LSPs crossing common resources is called a bypass tunnel. The establishment of bypass tunnels for the protection of inter-AS LSPs is analogous to the establishment of detour LSPs, shown previously.

6.3.4. *End-to-end disjoint LSPs*

Sprintson *et al.* [SPR 07] propose a distributed routing algorithm for finding two disjoint inter-AS QoS paths. They rely on a link-state aware topology abstracted from the multi-domain network. For an analysis of different schemes to establish end-to-end disjoint LSPs, we refer the reader to [TAK 07]

6.4. State of the art in inter-domain PCE

As shown in section 6.3, with the RSVP-TE extensions, inter-domain LSPs can be established if some guiding information (e.g., intermediate ASBRs on the path) is provided. In this section, we summarize the work done in the PCE WG of the IETF on the problem of finding an inter-domain path that respects certain QoS constraints. We focus on the proposed path computation techniques which enable us to determine the guiding information. Finally, we discuss the limitations and/or applicability of these path computation techniques.

6.4.1. *PCE-based architecture*

The PCE WG of the IETF is working on an architecture [FAR 06a] for the computation of paths to support MPLS TE LSPs. This architecture aims to be applied within a single domain or within a small group of domains (where a domain is a layer, IGP area or AS with limited visibility).

A PCE is an entity that can collect QoS, topology and reachability information. It can carry out path computation on behalf of a set of routers (path computation clients (PCCs)). PCEs, in the same AS or in different ASs, can be configured to communicate and cooperate [ASH 06] with each other in order to find feasible end-to-end paths. A traffic engineering database (TED) is used to store the information the PCE is interested in. It is used by the PCE to carry out path computations.

The generic requirement to allow the communications within and between PCEs was covered in [ASH 06].

An additional mechanism, called PCE discovery, is required for the PCCs to learn the list of PCEs that are available in their domain and in neighboring domains. The requirements for such a protocol are expressed in [LE 06, OKI 07].

6.4.2. *Path computation methods*

As can be seen from the name, PCE-based architecture focuses on how to compute paths. In this section, we introduce the path computation methods proposed. First we explain the function of the TED, which is used by PCEs to store traffic engineering related information, then we present the path computation techniques. These techniques work under the assumption that the AS-path is known *a priori*.

6.4.2.1. *TED*

The PCE computes path segments related to given QoS and diversity constraints based on a TED. The content of the TED for inter-domain TE has been discussed at the IETF [FAR 06b]. It depends on the domain of the PCE. The TED contains at least the topology of the domain and the TE attributes of the links belonging to the domain. In addition, it may contain the TE attributes of the links at the border of the domain, for example the inter-AS links. This information is distributed by the TE extensions to the Interior Gateway Protocols (IGP) [KAT 03, SMI 04, CHE 07]. Moreover, the TED must also contain reachability information for destinations outside the domain of the PCE. This information is currently distributed by BGP for destinations outside the AS.

6.4.2.2. *Per-domain path computation*

This technique relies on establishing and computing the inter-domain MPLS LSPs simultaneously. It makes use of RSVP-TE's ability to establish inter-domain MPLS LSPs, and to crankback [FAR 07]. This is the capability (1) to stop the establishment of an LSP at a node when it cannot compute a path that respects the constraints of the LSP and (2) to establish the LSP along a different path.

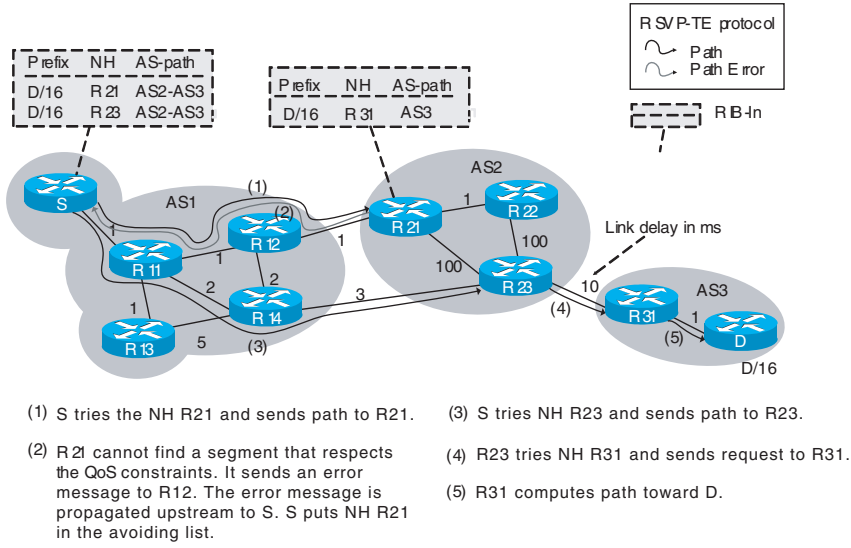

Figure 6.5. *Per-domain path computation*

As we mentioned in section 6.3.1, inside RSVP-TE, it is possible to indicate inside the ERO the path or a portion of the path to be followed by an LSP. The per-domain path computation technique, described in [VAS 07a], relies on this object. It consists of completing at the ingress router of a domain, the ingress ASBR, the path computation up to the BGP next-hop (NH), i.e. last reachable hop towards the destination. This node is either the first hop inside the downstream domain or the last hop inside the current domain. The computed path segment is then stored inside the ERO of the RSVP-TE `Path` message. This message is forwarded along the path specified inside the ERO and requests the establishment of the LSP along the path.

Either a dedicated PCE or the ingress routers in an AS should be responsible for the computation of the path segments. In Figure 6.5, the ingress ASBRs compute the paths. Upon reception of a RSVP `Path` message requesting the establishment of an LSP, a ASBR computes the paths. The ASBR tries usable NHs to reach the destination and consistent with the given AS-path. The means of determining the usable NHs is still an issue for debate. Here, we consider as usable NHs those known from the BGP propagation. The ASBR stores the list of NHs that have already been tried for a given LSP and that lead to an impracticable path with regard to the constraints. When the ASBR is not able to complete the path with a segment respecting the QoS properties of the LSP, crankback is performed [FAR 07], that is, the ASBR generates an RSVP `Path Error` message and sends it upstream. The upstream ASBR computes a new segment avoiding the NHs that have already been tried.

Figure 6.5 illustrates the per-domain path computation technique with a path computation that takes place at the ingress ASBRs. In this example, an LSP with the delay constraint of 100 ms has to be established from *S* to *D* using a given AS-path *AS2 - AS3*. The source of the LSP *S* first tries to use NH *R21*. It computes a path segment towards *R21* respecting the constraints based on its knowledge of the internal topology of *AS1*. *S* generates an RSVP `Path` message with an ERO object that contains the computed path segment. Then the `Path` message is sent along the segment. This leads to the establishment of the LSP along the path segment.

At the ingress ASBR inside *AS2*, *R21*, the process described in the previous paragraph is repeated, that is, *R21* computes the path segment towards NH *R31* in order to reach the tail-end of the LSP. However, *R21* is not able to provide a path segment that respects the constraints. Consequently, crankback occurs at *R21*.

R21 sends a `Path Error` message upstream. When the `Path Error` message arrives at *S*, *S* tries to compute a path segment avoiding NH *R21*, which leads to an infeasible path. *S* computes a path segment that ends at NH *R23*. It sends a `Path` message along the path segment towards *R23*. *R23*, in the downstream AS, *AS2*, carries out the computation of a path segment starting at *R23* and ending at the entrance *R31* inside the downstream AS *AS3*. This path segment is inserted inside the ERO of the `Path` message and the establishment of the LSP continues until the LSP's tail-end is reached.

This technique may or may not rely on PCEs. Moreover, the path computation and the LSP establishment take place at the same time. The computation ends once a path respecting the constraints is found even if it is not the shortest path. With the per-domain path computation technique, if PCEs are used in the computation, they do not communicate among themselves. Thus, discovery of PCEs in neighboring domains is not required.

6.4.2.3. *Backward recursive path computation*

The Backward Recursive PCE-based Computation (BRPC) technique is described in [VAS 07c]. It has been designed to find the shortest path for a constrained inter-AS LSP request. It makes the assumption that the list of domains to be crossed by the LSP is known prior to the computation. Thus, the computed path is the best path that can be obtained along this inter-domain path.

A PCE, which uses the BRPC technique to compute the path, communicates with other PCEs in order to request the computation of path segments contained in regions for which it does not possess enough topological information. Cooperative PCEs can communicate with each other using the protocols specified in [VAS 07b].

In this technique, the LSP's head-end sends a PCReq message specifying the constraints for the LSP to the PCE of its domain. Then, a PCReq message is sent from the PCE of one domain to the PCE of the downstream domain. Upon reception, the PCEs in the downstream domain, of multiple path segments starting at the entrances of the downstream domain and ending at the LSP's tail-end together with their QoS properties, a PCE is capable of computing the best segments starting at the entrances of its domain and ending at the tail-end of the LSP, with regard to the constraints. These segments are sent to the upstream PCE inside a PCRep message.

Figure 6.6 illustrates the computation of inter-AS constrained paths by means of BRPC. The LSP to establish is subject to a maximum delay constraint of 100 ms. The head-end of this LSP is router *S* in *AS1*. The tail-end of the LSP, node *D*, belongs to *AS3*. The longest matching prefix advertised for *D* is *D*/16. First, the AS-path is determined. The AS-path which needs to be followed by the LSP is *AS2 - AS3*. The central part of the figure shows the physical topology of the ASs and their interconnections. In the top part of the figure, we see the PCEs of each ASs, labels for the messages exchanged between PCEs and the BGP routes known by the PCEs. The content of the messages exchanged between PCEs is shown at the bottom of the figure.

The LSP's head-end sends a PCReq message to the PCE of its AS, *PCE1*. *PCE1* sends a PCReq message to *PCE2*, the PCE inside *AS2*. The PCReq message contains the address of the LSP tail-end and the constraints for the LSP. The delay constraint is not necessary because the output of the computation technique is the shortest delay path following the given AS-path. If the path delay returned to the LSP's head-end is above the delay constraint, there is no suitable path for the LSP respecting the given AS-path. The LSP establishment fails.

PCE2 sends a PCReq to *PCE3* because *AS3* is the downstream AS to *AS2* in the given AS-path. *PCE3* computes a path segment from *R31* to *D*, the LSP's tail-end. It then sends the segment with its delay in PCRep message (*3*) upstream to *PCE2*.

When *PCE2* receives PCRep (*3*), it computes path segments from the entrances inside its domain to the destination of the LSP. For this purpose, the PCE performs a shortest path first (SPF) computation on the graph composed of the local topology, the inter-AS links and the segments received from the downstream PCE. This results in two path segments starting at node *R21* and node *R23* respectively and ending at *D*.

Next, *PCE2* sends the resulting segments and their delays inside PCRep (*4*) to *PCE1*. After receiving the reply from *PCE2*, *PCE1* computes the end-to-end path based on the local topology, the inter-AS links connected to *AS1* and the received segments. The resulting path is *S - R11 - R14 - R23 - R31 - D* with a delay of 17 ms.

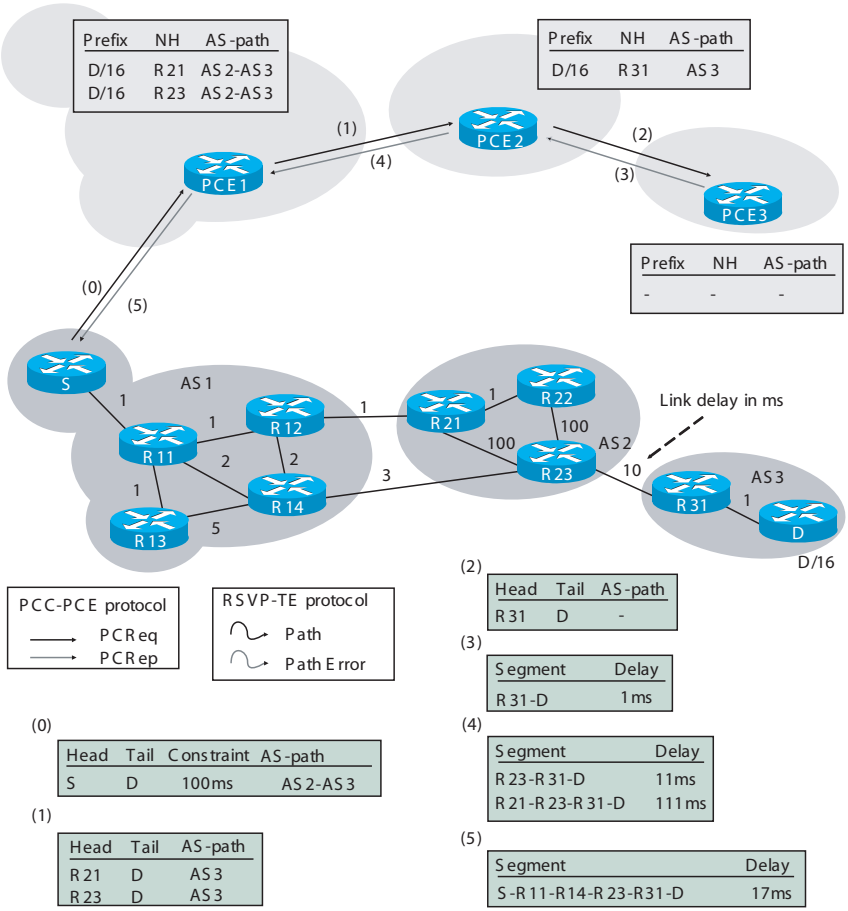

Figure 6.6. *Backward recursive PCE-based path computation*

This path is sent in PCRep (*5*) to the head-end of the LSP, *S*. Finally, *S* initiates the establishment of the LSP along this path. For this purpose it stores the path returned by *PCE1* inside the ERO. Thus, the RSVP `Path` message follows the computed path and the LSP is established along this path.

In order to respect the confidentiality requirement of ISPs (see section 6.2.1), PCEs may return an aggregated RRO or path keys [BRA 07] inside PCRep messages, instead of returning path segments that reveal sequences of hops inside their domains.

BRPC relies on PCEs that communicate and cooperate in order to find the shortest path respecting the given QoS constraints along a given AS-path. Here, PCEs have to discover the PCEs in neighboring domains [LE 06]. With BRPC, contrary to the

per-domain path computation technique, it is possible to simultaneously compute a pair of disjoint LSPs, as described in [VAS 07c], when the AS-path for the pair of LSPs is given.

6.4.3. *Applicability of the path computation techniques*

The PCE-based architecture makes it possible to set up inter-AS LSPs with end-to-end QoS constraints. However, the following aspects still need to be addressed before we obtain a solution that is workable in practice:

– all the proposed path computation methods assume an awareness of the AS-path. If such an AS-path is not specified, a solution has to be found that will compute this AS-path;

– as the QoS properties of the AS-paths are not known by the source PCE, using a randomly picked AS-path gives little confidence in finding an end-to-end path respecting the QoS constraints. Hence, the QoS properties of AS-paths must be obtained or determined in one way or another.

Recall the lower bound on routing diversity provided in section 6.2.3.3. Using the BGP best paths, we may obtain a poor sample of the available end-to-end QoS and might fail in finding a feasible path. If the AS-path is not given to the PCE for path computation, the PCE may need to try all known AS-paths. The number of AS-paths known by the PCE might be several times that of the BGP best AS-paths for the AS.

6.5. Towards inter-AS QoS

Section 6.3 discussed the techniques available to support inter-domain LSPs. Section 6.4 introduced the path computation architecture and techniques. These two components are able to work together and establish inter-AS LSPs with end-to-end QoS constraints. However, before inter-AS LSPs with QoS guarantees are a reality, enough information needs to be known by the entities that will compute the end-to-end path in each AS. The techniques presented in section 6.4 assume that the AS sequence of a feasible QoS path is known. Given the lack of QoS information currently available for the Internet, this seems to be quite an assumption.

Computing an end-to-end QoS path requires us to find a trade-off between the amount of QoS information to be distributed across the Internet and the complexity of finding QoS paths. We expect that a solution to the inter-AS QoS problem will address the following two aspects:

– what information should be distributed across the Internet to enable the QoS path computation and how should this information be distributed?

– how should the propagated QoS information be used by an AS willing to establish a QoS path?

In the subsequent sections, we discuss each of the above two questions separately. The first question, which consists of two inter-related aspects, the content and the distribution of QoS information, is discussed in section 6.5.1. Section 6.5.2 then presents a possible scenario of end-to-end LSP computation with QoS constraints.

6.5.1. *Distributing QoS Information for inter-AS LSPs*

Several types of information have to be available to make the end-to-end LSP computation and establishment possible. First, loose or strict path information has to be disseminated. By loose topological information, we mean AS sequences that provide a given QoS. Less loose topological information would be IP-level paths. In practice, we do not expect ISPs to be willing to reveal IP-level information, both for confidentiality reasons, and also because it would make the computation framework impossible to scale.

Second, QoS information about the link state has to be available, at least within ISP networks. For this, traffic engineering extensions to IGP protocols have to be used, and measurements have also to be carried out by the ISPs inside their AS. Traffic engineering extensions to IGP protocols [KAT 03, SMI 04] can provide information about the maximum bandwith and maximum reservable bandwidth on a link. For delay-related information, measurements have to be used [CHO 07]. Depending on how the path computation will be carried out, this information may have to be propagated to special nodes (e.g. PCEs) and kept up to date.

Finally, restrictions are likely to apply on the use of the resources, as ISPs want to retain control of their own resources. Today, routing policies are not explicitly revealed beyond AS boundaries, because they are bound to contractual agreements. We see no reason why the same rules would not apply to QoS paths. ISPs will decide how much resources can be allocated to QoS paths, and under which conditions. This will probably lead to distinct business relationships for QoS traffic. If ISPs are bound by contractual agreements for QoS traffic, then it is unlikely that a centralized framework for distributing and computing QoS paths will be used. As the current routing architecture is highly distributed, it is unlikely that a centralized solution would be adopted in the short term, i.e. within the next few years.

A general solution for providing QoS information for NHs and routes is for ASs to summarize their QoS states towards a specific destination and propagate this information to neighboring ASs according to their QoS-related business agreements. For the summary of an AS's QoS states towards the destination, work needs to be done in order to allow the QoS summaries from each AS to be combined and propagated along the AS sequence. Moreover, the QoS state should be summarized in such a way that the summary is not very sensitive to the variations of the real QoS state, such that the QoS summary does not need to be updated very frequently, as proposed in [XIA 02].

6.5.1.1. *RIB-Ins*

One type of information that can be used to extend the space of forwarding routes from using only the BGP best paths, is the RIB-Ins of all border routers in the AS. We have shown in section 6.2.3.3 that much more path diversity can be exploited if all the routes known to ASs can be used to establish inter-AS LSPs. As routes present in the RIB-Ins of BGP routers are consistent with routing policies, the sole drawback of using RIB-In entries is that they are currently hidden. The scalability of BGP routing lies partly in this hiding of the full path diversity. If the RIB-Ins of all routers of an AS together with the export policies of the AS have to be known to an entity like a PCE, scalability can become an issue.

6.5.1.2. *Content of TED*

The PCE of an AS may be responsible for computing the QoS status for border routers inside the AS. It can be configured to compute and assign QoS states for border routers in the AS's desired way, e.g. billing as a big drive for the routing policies might be taken into account in the form of preferences to next-hops/routes. ASs would configure their PCEs to assign higher preferences to cheap next-hops/routes which can still provide certain QoS guarantees. The PCE can also selectively store the QoS states for next-hops/routes in the TED. Only a few next-hops/routes with "good" QoS states will contribute to the establishment of inter-AS LSPs with QoS constraints, and whether the QoS states of the rest of NHs/routes are known in the TED will not affect the path computation outcome. Scalability in the size of the TED can be obtained by caching only the most promising entries (not the full RIB-Ins) from a QoS perspective.

6.5.1.3. *Centralized vs. distributed information dissemination*

Distributing the QoS information across the Internet can be achieved in two ways. First, each AS can push their QoS information to a centralized system that will handle the requests for inter-AS LSP establishment. Requests for Inter-AS LSPs with QoS guarantees will be sent to the centralized system, which will try to find an end-to-end AS-path that meets the QoS guarantees. Such a solution is currently being investigated by the IPsphere Forum [IPS]. The IPsphere Forum is developing a solution to determine the ASs to cross for an LSP with given QoS requirements and taking into account the business relationships of the ASs. Participation in this IPsphere Forum is based on membership, and their work is not publicly available. We will not discuss further the centralized solution to the QoS problem, but rather focus on distributed solutions based on the current routing architecture.

This second method relies on the current distributed model of routing in the Internet. In this model, each AS is responsible for propagating reachability information about every destination to its neighbors. The same principle could apply to the propagation of QoS-related information. Some works [BON 01, XIA 02] have proposed QoS extensions to BGP. The idea is to use BGP to piggyback QoS

information, since BGP already propagates reachability information across the Internet. [BON 01] proposed to add a new type of attribute to BGP messages. This new type of BGP attribute gives freedom to ASs to specify whether the QoS information is transportable or if only their neighbors should know about it. Several attributes are proposed in [BON 01], like maximum bandwidth, available capacity, minimum and maximum transit delay. The QoS attributes are associated with the forwarding path towards the destination prefix of the BGP route. This QoS information thus only concerns the best paths chosen by BGP, which we know represent only a small fraction of the whole path diversity known to ASs on the Internet. [XIA 02] proposed four statistical QoS metrics, to make the QoS extensions to inter-domain routing scalable while ensuring the optimum QoS routing. [XIA 02] proposes to represent QoS information by intervals of the metric values, $[l, u]$, together with a probability p that the instantaneous metric value belongs to this interval. As in [BON 01], the QoS information in [XIA 02] is related to the best routes selected by BGP.

6.5.2. *Computing inter-AS LSPs with end-to-end QoS constraints*

We expect that providing ASs with appropriate AS-paths will be one of the challenges involved in solving the inter-AS QoS problem. Given that obtaining a complete view of the topology and TE information is hardly possible, a PCE that computes a portion of a constrained inter-domain LSP must rely on heuristics to choose an appropriate AS-path and BGP next-hop among those given for the destination. If a bad choice is made by the heuristic at a PCE, a downstream PCE may not be able to complete the computation of the path. In that case, the solution is to rely on crankback to try alternative AS-paths and next-hops. Even though we do not expect that this will be necessary, in the worst case the PCE still needs to search through the whole space of possible AS-paths and next-hops.

In [PEL 06], Pelsser *et al.* propose two heuristics, namely "nearest NH" and "Vivaldi" [DAB 04], for the selection of the ingress node (i.e. the next-hop) in the downstream domains and combine them with the per-domain path computation technique. The "nearest NH" selects the next-hop based on the delay information locally available to the domain, and the "Vivaldi" next-hop selection is based on an estimation of the delay of the path that transits through the candidate next-hop. These heuristics select next-hops for inter-AS LSPs with the end-to-end delay as the QoS constraint. It is shown from simulations that these heuristics can limit the computational requirements for finding feasible end-to-end QoS paths.

If, as discussed in section 6.5.1, the QoS information is available, then it can be directly used as a heuristic for NH/AS-path selection. It should be designed specifically for path computation for inter-AS LSPs with QoS constraints. It should thus maximize the likelihood that a feasible path will be found.

Once the AS-level path on which to establish the LSP is chosen, it is up to the path computation techniques to compute intra-AS segments thanks to the TED (see section 6.4.2). The content of the TED is used to find out the best next-hop that satisfies the QoS constraints on the AS-path. Once the next-hop is found, the segment is found by the path computation techniques. We show in Figure 6.7 how QoS descriptions can be computed and recorded in TEDs, and propagated with BGP. In Figure 6.8, we show how the per-domain path computation technique uses this information on QoS descriptions to compute the path.

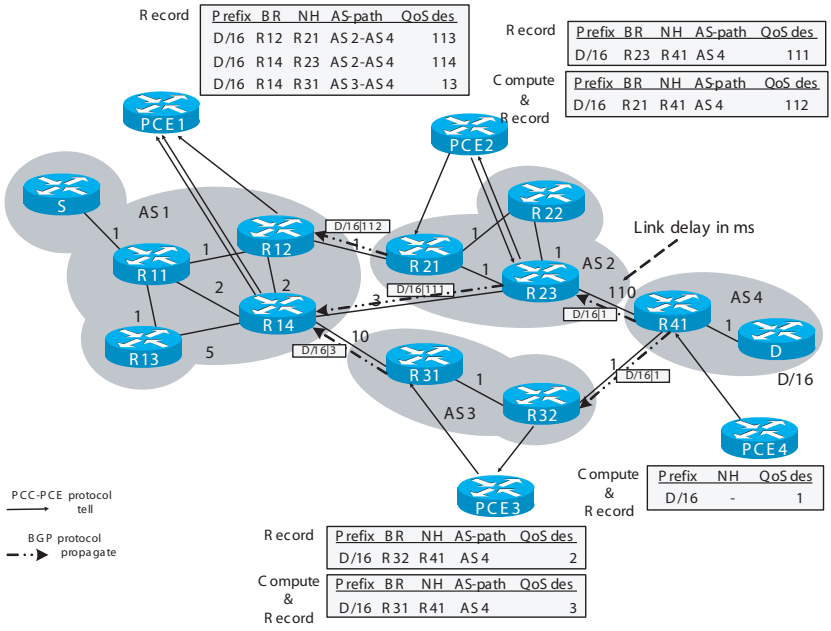

Figure 6.7. *QoS summary propagation*

In Figures 6.7 and 6.8, we use delay as the QoS metric. BGP routers propagate their QoS states to all peering BGP routers, while in reality, they may propagate to a selected subset of peering BGP routers according to policies. Each BGP router (BR in the figures), upon receiving BGP QoS updates from border routers towards a destination, will inform the PCE in this AS of this updated QoS information towards the destination, as well as the location from which the updates were received (i.e. the next-hop). The PCE will compute and record the QoS state of all the BGP routers in this AS towards the destination. Thus, the PCE knows the AS-wide QoS state towards all reachable destinations and via which NH border router this QoS state might be achieved.

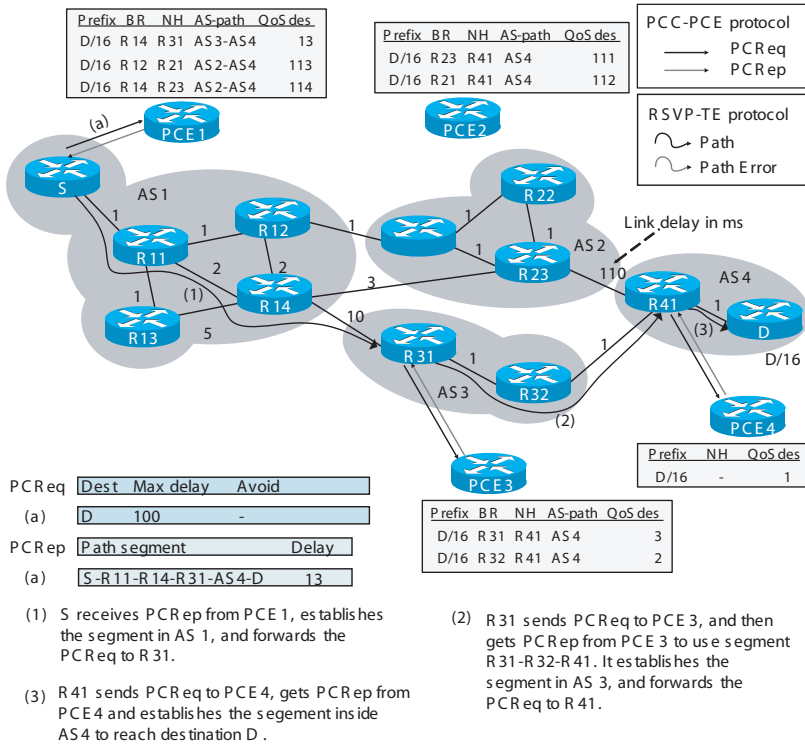

Figure 6.8. *Per-domain path computation using QoS information*

Upon receiving a path computation request (PCReq), the PCE will compute the path based on the information recorded in the TED including the intra-domain topology information. Different path computation techniques (per-domain or backward recursive) can be used, as our framework tells the PCE where to forward the PCReq but does not stick to a single path computation technique. In the example in Figure 6.8, an LSP with a delay constraint of 100 ms needs to be established from *S* to *D*. The head-end LSR *S* sends PCReq *(a)* to the PCE in charge of the path computation in *AS1*. *PCE1* knows from its TED that next-hop *R31* has the smallest delay towards *D/16*, and computes the path segment in *AS1* towards *R14*/*R31*. PCReq is forwarded to *R31* which sends PCReq to *PCE3*. *PCE3* selects next-hop *R41* and computes the path segment inside *AS3* towards *R32*/*R41*. *R41* receives the PCReq and asks *PCE4* to compute the path segment towards *D/16*. The LSP can then be established with the segments computed in each AS.

6.6. Conclusion and perspectives

In this chapter, we explained the complexity of establishing inter-AS LSPs with QoS guarantees. We described the working of the inter-domain routing system, and the assumptions on which it relies. We discussed the consequences of the path selection by the current inter-domain routing system on the visibility of the inter-domain paths. The limited visibility of the available paths does not actually prevent us from establishing inter-AS LSPs with QoS guarantees. Rather, this limited visibility requires clever heuristics to be designed in order to find the most feasible QoS paths. We covered the existing signaling extensions to RSVP-TE that support the establishment of inter-AS LSPs, as well as the protection of those LSPs. The path computation techniques that have been proposed at the IETF were also detailed. These computation techniques make it possible to find the LSP segments within each AS, so as to compose an end-to-end LSP with QoS guarantees. Finally, we amalgamate these three components, i.e. inter-domain routing, signaling, and path computation techniques, and show that an inter-AS QoS is not beyond reach, whilst recognizing that more work needs to be done in specific areas.

Even though the content of this chapter is intended to be as factual as possible, we believe that it argues in favor of the feasibility of providing end-to-end QoS in the Internet. The challenges that have to be faced in order to push QoS beyond AS boundaries are by no means insurmountable. Dissemination of QoS information using the current inter-domain routing system has already been proposed several times. Due to concerns about the scalability of inter-domain routing, however these proposals have not received much interest from the networking community. We believe that the current Internet evolution towards more stringent services, demands end-to-end QoS capabilities. Before the deployment of MPLS by ISPs, pushing QoS on the Internet as proposed through Differentiated Services (DiffServ) [BLA 98] or Integrated Services (IntServ) [SHE 97] was very difficult, as it involved significant changes to the core of the Internet. With the wider deployment of BGP/MPLS VPNs in large ISP networks, the situation of today's Internet core makes QoS more and more relevant. We have shown in this chapter that most of the building blocks to make QoS path computation possible have already been defined, albeit not completely standardized or tested as yet. The work carried out at the IETF within the Path Computation Element (PCE) working group is a proof of the effort expended by the community to make the computation of end-to-end QoS paths possible. This working group has however decided so far to limit its work to path computation within a single domain, given the difficulties of crossing AS boundaries when it involves different companies. We believe that the community should investigate further into inter-AS QoS to make the Internet a better place for QoS-demanding applications.

6.7. Acknowledgments

Cristel Pelsser acknowledges the numerous discussions and support of Olivier Bonaventure and Stefaan De Cnodder. Bingjie Fu is supported by STW project DTC. 6421.

6.8. References

[ABI] ABILENE OBSERVATORY, http://abilene.internet2.edu/observatory/.

[ASH 06] ASH J. and LE ROUX J.-L., "A path computation element (PCE) communication protocol generic requirements", *Request for Comments 4657, IETF*, September 2006.

[BAT 07] BATTISTA G.D., ERLEBACH T., HALL A., PATRIGNANI M., PIZZONIA M. and SCHANK T., "Computing the types of the relationships between autonomous systems", *IEEE/ACM Trans. Networking*, vol. 15, no. 2, p. 267–280, April 2007.

[BLA 98] BLAKE S., BLACK D., CARLSON M., DAVIES E., WANG Z. and WEISS W., "An architecture for differentiated services", *Request for Comments 2475, IETF*, December 1998.

[BLA 03] BLANCHY F., MÉLON L. and LEDUC G., "An efficient decentralized on-line traffic engineering algorithm for MPLS networks", in *Proc. of 18th International TELETRAFFIC CONGRESS - Providing Quality of Service in Heterogeneous Environments*, 2003.

[BOL 93] BOLOT J.-C., "End-to-end packet delay and loss behavior in the Internet", *Proc. of ACM SIGCOMM*, p. 289–298, 1993.

[BON 01] BONAVENTURE O., "Using BGP to distribute flexible QoS information", Internet draft, draft-bonaventure-bgp-qos-00.

[BRA 07] BRADFORD R., VASSEUR J.-P. and FARREL A., "Preserving topology confidentiality in inter-domain path computation using a key-based mechanism", Internet draft, draft-ietf-pce-path-key-01.txt.

[CAE 05] CAESAR M. and REXFORD J., "BGP routing policies in ISP networks", *IEEE Network Magazine*, 2005.

[CHE 07] CHEN M., ZHANG R. and DUAN X., "OSPF extensions in support of inter-AS multiprotocol label switching (MPLS) and generalized MPLS (GMPLS) traffic engineering", Internet draft, draft-ietf-ccamp-ospf-interas-te-extension-02.txt.

[CHO 07] CHOI B., MOON S., ZHANG Z., PAPAGIANNAKI K. and DIOT C., "Analysis of point-to-point packet delay in an operational network", *Computer Networks*, vol. 51, no. 13, p. 3812–3827, Elsevier North-Holland, Inc., 2007.

[DAB 04] DABEK F., COX R., KAASHOEK F. and MORRIS R., "Vivaldi: a decentralized network coordinate system", in *Proceedings of the ACM SIGCOMM Conference*, 2004.

[DE 04] DE CNODDER S. and PELSSER C., "Protection for inter-AS MPLS tunnels", Internet draft, draft-decnodder-ccamp-interas-protection-00.txt.

[DIM 07] DIMITROPOULOS X., KRIOUKOV D., FOMENKOV M., HUFFAKER B., HYUN Y., K CLAFFY and RILEY G., "AS relationships: inference and validation", *ACM Comput. Commun. Rev.*, vol. 37, no. 1, 2007.

[FAR 06a] FARREL A., VASSEUR J.-P. and ASH J., "A path computation element (PCE) based architecture", *Request for Comments 4655, IETF*, August 2006.

[FAR 06b] FARREL A., VASSEUR J.-P. and AYYANGAR A., "A Framework for inter-domain multiprotocol label switching traffic engineering", *Request for Comments 4726, IETF*, November 2006.

[FAR 07] FARREL A., SATYANARAYANA A., IWATA A., FUJITA N. and ASH G., "Crankback signaling extensions for MPLS and GMPLS RSVP-TE", *Request for Comments 4920, IETF*, July 2007.

[FEA 04] FEAMSTER N., BALAKRISHNAN H. and REXFORD J., "Some foundational problems in interdomain routing", in *ACM SIGCOMM Workshop on Hot Topics in Networks (HotNets)*, 2004.

[GAO 00] GAO L., "On inferring autonomous system relationships in the Internet", *Proc. IEEE Global Internet*, 2000.

[HAO 06] HAO J., MEULLE M. and NGUYEN Q., "Formulation CSP et approches heuristiques pour l'inférence des accords d'interconnexion dans l'Internet", in *ROADEF'06*, 2006.

[HUF 02] HUFFAKER B., FOMENKOV M., PLUMMER D., MOORE D. and K CLAFFY, "Distance metrics in the Internet", *IEEE International Telecommunications Symposium*, 2002.

[HUS] HUSTON G., http://www.potaroo.net/.

[INT] INTEL-DANTE MONITORING PROJECT, http://rtmon.gen.ch.geant2.net/.

[IPS] IPSPHERE FORUM - THE BUSINESS OF IP, http://www.ipsphereforum.org/home.

[KAT 03] KATZ D., KOMPELLA K. and YEUNG D., "Traffic engineering (TE) extensions to OSPF version 2", *Request for Comments 3630, IETF*, September 2003.

[LE 06] LE ROUX J.-L., "Requirements for path computation element (PCE) discovery", *Request for Comments 4674, IETF*, October 2006.

[MÜH 07] MÜHLBAUER W., UHLIG S., FU B., MEULLE M. and MAENNEL O., "In search for an appropriate granularity to model routing policies", in *Proceedings of the ACM SIGCOMM Conference*, 2007.

[OKI 07] OKI E., "PCC-PCE communication and PCE discovery requirements for inter-layer traffic engineering", Internet draft, draft-ietf-pce-inter-layer-req-06.txt, work in progress, November 2007.

[PAN 05] PAN P., SWALLOW G. and ATLAS A., "Fast reroute extensions to RSVP-TE for LSP tunnels", *Request for Comments 4090, IETF*, May 2005.

[PEL 02] PELSSER C. and BONAVENTURE O., RSVP-TE extensions for interdomain LSPs, Technical Report no. 2002-09, University of Namur, October 2002.

[PEL 03] PELSSER C. and BONAVENTURE O., "Extending RSVP-TE to support inter-AS LSPs", *2003 Workshop on High Performance Switching and Routing (HPSR 2003)*, 2003.

[PEL 06] PELSSER C. and BONAVENTURE O., "Path selection techniques to establish constrained interdomain MPLS LSPs", in *Proc. of IFIP International Networking Conference*, 2006.

[QUO 05] QUOITIN B. and UHLIG S., "Modeling the routing of an autonomous system with C-BGP", *IEEE Network Magazine*, 2005.

[ROS 06] ROSEN E. and REKHTER Y., "BGP/MPLS IP virtual private networks (VPN)s", *Request for Comments 4346, IETF*, February 2006.

[ROU a] RIPE'S ROUTING INFORMATION SERVICE, http://www.ripe.net/ris/.

[ROU b] ROUTEVIEWS PROJECT, http://www.routeviews.org/.

[SAV 99] SAVAGE S., COLLINS A., HOFFMAN E., SNELL J. and ANDERSON T., "The end-to-end effects of Internet path selection", *Proc. of ACM SIGCOMM*, p. 289–299, 1999.

[SHE 97] SHENKER S., PARTRIDGE C. and GUERIN R., "Specification of guaranteed quality of service", *Request for Comments 2212, IETF*, September 1997.

[SMI 04] SMIT H. and LI T., "Intermediate system to intermediate system (IS-IS) extensions for traffic engineering (TE)", *Request for Comments 3748, IETF*, June 2004.

[SPR 04] SPRING N., MAHAJAN R., WETHERALL D. and ANDERSON T., "Measuring ISP topologies with rocket fuel", *IEEE/ACM Transactions on Networking*, vol. 12, no. 1, p. 2–16, February 2004.

[SPR 07] SPRINTSON A., YANNUZZI M., ORDA A. and MASIP-BRUIN X., "Reliable routing with QoS guarantees for multi-domain IP/MPLS networks", in *Proceedings of IEEE Infocom*, 2007.

[TAK 07] TAKEDA T., IKEJIRI Y. and VASSEUR J.-P., "Analysis of inter-domain label switched path (LSP) recovery", Internet draft, draft-ietf-ccamp-inter-domain-recovery-analysis-02.txt, work in progress, September 2007.

[VAN 01] VAN MIEGHEM P., NEVE H.D. and KUIPERS F., "Hop-by-hop quality of service routing", *Computer Networks*, vol. 37, no. 3-4, p. 407–423, 2001.

[VAS 06] VASSEUR J.-P., IKEJIRI Y. and ZHANG R., "Reoptimization of multiprotocol label switching (MPLS) traffic engineering (TE) loosely routed label switched path (LSP)", *Request for Comments 4736, IETF*, November 2006.

[VAS 07a] VASSEUR J.-P., AYYANGAR A. and ZHANG R., "A per-domain path computation method for establishing inter-domain traffic engineering (TE) label switched paths (LSPs)", Internet draft, draft-ietf-ccamp-inter-domain-pd-path-comp-05, work in progress, April 2007.

[VAS 07b] VASSEUR J.-P. and LE ROUX J.-L., "Path computation element (PCE) communication protocol (PCEP)", Internet draft, draft-ietf-pce-pcep-08.txt, work in progress, July 2007.

[VAS 07c] VASSEUR J.-P., ZHANG R., BITAR N. and LE ROUX J.-L., "A backward recursive PCE-based computation (BRPC) procedure to compute shortest interdomain traffic engineering label switched paths", Internet draft, draft-ietf-pce-brpc-06.txt, work in progress, September 2007.

[WAN 03] WANG F. and GAO L., "Inferring and characterizing Internet routing policies", in *ACM SIGCOMM Internet Measurement Workshop*, 2003.

[XIA 02] XIAO L., LUI K.-S., WANG J. and NAHRSTEDT K., "QoS extension to BGP", *Proc. of the 10th IEEE International Conference on Network Protocols*, 2002.

[XIA 04] XIA J. and GAO L., "On the evaluation of AS relationship inferences", *IEEE Global Communications Conference (GLOBECOM)*, Dallas, TX, November 2004.

[ZHA 01] ZHANG Y. and DUFFIELD N., "On the constancy of Internet path properties", *Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*, p. 197–211, 2001.

[ZHA 03] ZHANG R. and BARTELL M., *BGP Design and Implementation*, CISCO Press, 2003.

[ZHA 05] ZHANG R. and VASSEUR J.-P., "MPLS inter-autonomous system (AS) traffic engineering (TE) requirements", *Request for Comments 4216, IETF*, 2005.