**MESCAL**

*Management of End-to-end Quality of Service*
*Across the Internet at Large*

**IST-2001-37961**

# D1.4: Issues in MESCAL Inter-Domain QoS Delivery: Technologies, Bi-directionality, Inter-operability, and Financial Settlements

| **Document Identifier:** MESCAL/WP1/TRT/D1.4/ | |
|---|---|
| **Deliverable Type:** Report<br>**Deliverable Nature:** Public | **Contractual Date:** - None<br>**Actual Date:** 30 January 2004 |

**Editor:** Abolghasem (Hamid) Asgari, Thales Research & Technology (TRT) UK Limited

**Authors:**
*FTR&D:* Pierrick Morand, Mohamed Boucadair
*TRT:* Hamid Asgari, Richard Egan, Mark Irons
*UCL:* Jonas Griem, David Griffin, Jason Spencer
*UniS:* P. Flegkas, P. Trimintzios
*Algo:* Takis Damilatis, Panagiotis Georgatsos

**Abstract:** This deliverable comprises of six chapters addressing various issues in QoS-based service delivery. The first chapter describes the Inter-domain QoS peering approaches. Second chapter explains the Internet Exchange Points and their implications on the three MESCAL solution options presented in D1.1. Chapter 3 describes various optical network architectures, control planes and protocols, as well as current optical technology, the current state of dynamic optical networks and their impact on MESCAL. Chapter 4 discuses how to provide bi-directional services in MESCAL environment considering all three solution options. Chapter 5 discusses inter-operability of the three MESCAL solution options from a service and technical point of view examining the co-existence of distinct MESCAL solution options deployed within the same AS and inter-working of a MESCAL solution option through a domain that supports different solution option(s). Finally, Chapter 6 looks at issues related to the financial settlements for a bi-directional service including who should pay for the service.

**Keywords:** Peering, Cascade, IXP, AS, GMPLS, *cSLS*, *pSLS*, QC, QoS

| | | | |
|---|---|---|---|
| France Telecom Research and Development | FTR&D | Co-ordinator | France |
| Thales Research and Technology | TRT | Principal Contractor | UK |
| University College London | UCL | Principal Contractor | UK |
| The University of Surrey | UniS | Principal Contractor | UK |
| Algonet SA | Algo | Principal Contractor | Greece |

Project funded by the European Community under the
"Information Society Technology" Programme (1998-2002)

# Executive Summary

This deliverable comprises of six chapters addressing various issues in QoS-based service delivery. This deliverable is additional to those specified in the MESCAL Technical Annex. It complements the Specification of Business Models and a Functional Architecture for Inter-domain QoS Delivery (Deliverable D1.1) by considering some further aspects of Inter-domain QoS delivery.

The first chapter describes the Inter-domain QoS peering approaches. There can be many approaches for the interconnection of providers' networks for offering QoS services across multiple domains. In this chapter we focus on two approaches, namely the Cascaded and the Centralised, for constructing QoS-based service classes. Then, we compare and analyse the strengths and limitations of these two approaches in implementation and offering end-to-end QoS-based services.

Internet Exchange Points (IXPs) are public peering points that play important part in the overall infrastructure of the Internet. Chapter 2 surveys IXPs networking infrastructure and the services they provide. The IXP participation in the inter-domain connectivity chain and its implication on MESCAL solutions are investigated. Some solutions are provided to overcome the problems that arise by the involvement of IXPs in the QoS delivery chain with regards to the MESCAL solution options proposed in D1.1.

Chapter 3 presents various optical network architectures, control planes and protocols, as well as current optical technologies, the current state of dynamic optical networks and their impact on MESCAL. It describes the standards and technologies available to configure and manage an optical network. The integration of MESCAL solution options with optical control planes and possible future optical network developments are discussed.

In chapter 4, bi-directionality of services using the cascaded approach is considered for all three solution options proposed in D1.1. The primary challenge is in constructing the QoS-enabled reverse path for return traffic. This chapter identifies the issues, presents a detailed discussion of the resulting implications and provides methods for resolving them.

Chapter 5 discusses inter-operability of the three MESCAL solution options from a service and a technical point of view. Two main scenarios are examined: the first scenario examines the co-existence of three MESCAL solution options deployed within the same domain; the second scenario focuses on the extension of the scope of a given MESCAL solution option through a domain that supports different solution option(s). Issues related to these scenarios are highlighted and solutions are proposed.

Chapter 6 looks at business relationships and the issues related to the financial settlement for Inter-domain QoS services. New business relationships and arrangements for financial settlement are proposed for the MESCAL-enabled QoS-aware Internet, which can co-exist with current business practices.

The key achievements of this deliverable is the detailed investigation on the issues related to interconnection peering and technologies, inter-operability, and bi-directionality for the MESCAL solution in order to offer end-to-end QoS-enabled services. This detailed study will enable better dissemination, standardisation, realisation, and experimentation processes to take place in the remainder of the project.

# Table of Contents

# List of Figures

# List of Tables

# CHAPTER 1: INTER-DOMAIN QoS PEERING

## 1.1 Introduction

In order to provide access to the global Internet, ISPs must interact with each other; there cannot be a single ISP offering global Internet coverage.

The MESCAL solution to the problem of QoS-based service delivery in the Internet, across different Network Provider (NP) [1] domains, adopts a *hop-by-hop, cascaded model* for the interactions between NPs both at the service and network (IP) layers. Service layer interactions result in the establishment of service agreements between NPs, *pSLSs* in MESCAL terminology, aggregating customer service traffic, which need to be supported by appropriate service management and traffic engineering capabilities per provider domain as well as by BGP-based interactions at the IP layer for QoS inter-domain routing purposes.

The theme of this chapter is to evaluate the essence of the MESCAL approach regarding the particular model of interactions between providers adopted.

Currently, in the best-effort Internet, there exist two forms of distinct relationships between ISPs for traffic exchange, underlined by respective business agreements: *peering* and *transit*. Peering is termed as the business relationship, whereby ISPs reciprocally provide only access to each other's customers. Peering is a non-transitive relationship. Peering is a mutual agreement between ISPs to exchange data between themselves, normally for no fee or charge. Transit is the business relationship, whereby one transit provider provides access to all destinations in its routing table (could be global Internet) to another ISP for a charge. The business relationships and financial settlements between providers in the current and in a MESCAL-enabled QoS-aware Internet are discussed in detail in Chapter 6.

It is clarified that the terms 'peering', 'peering approaches' and similar, are used throughout this document (and in other MESCAL documents) to denote that two providers interact with each other for the purpose of expanding the topological scope of their offered services, under any business relationship which may govern this interaction; it should be taken that this implies a peering or a peer-to-peer business relationship as previously introduced. Obviously, when these terms are used in the context of business relationships, they will imply the specific business relationship.

There are many models for the interconnection and service-layer interactions between providers' networks for offering QoS services across multiple domains. Eurescom specified organisational models for the support of inter-operator IP-based services [P1008-D2]. An organisational model is required not only to establish a complete end-to-end customer service, but also the hierarchy of management information flows required to provide and maintain such a service. These Eurescom organisational models are specified for the interconnection of operators' networks and associated management systems that may be grouped into three configurations known as the *cascade model* the *hub model*, and the *mixture model*, which is a combination of the cascade and hub models, to cope with certain management requirements in complex interconnection scenarios. A variation on the theme of the hub model is denoted as the *star model*. Further insight into these forms of management organisational models, can be found in [ETSI 300820] -which is about ATM network interconnection management- and the EURESCOM Project P813-PF [P813-D1] -which is about technical development and support for European ATM service. These organisational models are strongly influenced by experience in the telecommunications industry of provision of international telephony and other services for which network interconnection is a requirement, both in commercial and regulatory terms.

The chapter is organised as follows. First, we build on organisational models similar to the above, the concepts to establish a set of inter-domain QoS peering approaches in order to construct end-to-end

[1] The terms NP (Network Provider), ISP (Internet Service Provider) as well as, simply, the term provider, are used interchangeably throughout the document to denote a business entity owning a network and being responsible for its operation and the provision of Internet connectivity aspects.

QoS-based services across the Internet at large scale. The challenge is to adapt established knowledge and methodologies to the requirements of/approaches for new IP QoS-based services, which will be attractive both to customers and providers. Then, we compare and analyse the strengths and limitations of the major approaches; namely the cascaded (adopted by MESCAL) and the centralised approaches, in offering end-to-end performance efficiently. Finally, the main points of the analysis are summarised and conclusions are drawn.

## 1.2 Inter-domain QoS Peering approaches

The type of inter-domain peering impacts the service negotiation procedures, the required signalling protocols, the QoS binding, and path selection. The following peering approaches are considered:

- The *centralised* approach where a Network Provider negotiates *pSLSs* directly with an appropriate number of downstream providers to construct an end-to-end QoS service. With this approach, service peers are not necessarily BGP peers.
- The *cascaded* approach where a NP only negotiates *pSLS*s with its immediate neighbouring provider/s to construct an end-to-end QoS service. With this approach, service peers are also BGP peers.
- The *hub* approach, which is similar to the centralised approach, where the Service Provider (SP), as a distinct entity from NP, is the central point that negotiates and establishes *pSLSs*.
- The *hybrid* approach, which is the mixture of centralised and cascaded approaches.

Within the MESCAL project, the first two major approaches have been considered for further study in order to construct end-to-end QoS-based services across the Internet at large scale. The following sections provide a description of all approaches for inter-domain peering concentrating on the two major ones.

### 1.2.1 Centralised Approach

The centralised approach disassociates *pSLS* negotiations from the existing BGP peering arrangements. The originating domain knows the end-to-end topology of the Internet and establishes *pSLS*s with a set of potential domains (neighbours, transit, and distant ASs) in order to reach a set of destinations, to offer end-to-end QoS-based services.

As shown in Figure 1, the SLS Ordering at the central point (AS1) has the responsibility for managing the overall requested QoS service/connection. To manage customer requests, the provider (AS1) directly requests peering agreement (*pSLS1 & pSLS2*) with providers AS2 and AS3 and with any other network provider involved in order to create an end-to-end QC (from AS1 to AS3). Each AS is responsible for the connection inside its domain and its inter-domain link interfaces.

In general, responsibility for management of the inter-domain physical links between ASs may be owned and managed on a bi-lateral basis. We assume that management of the egress interfaces of the border routers is the responsibility of that particular domain.

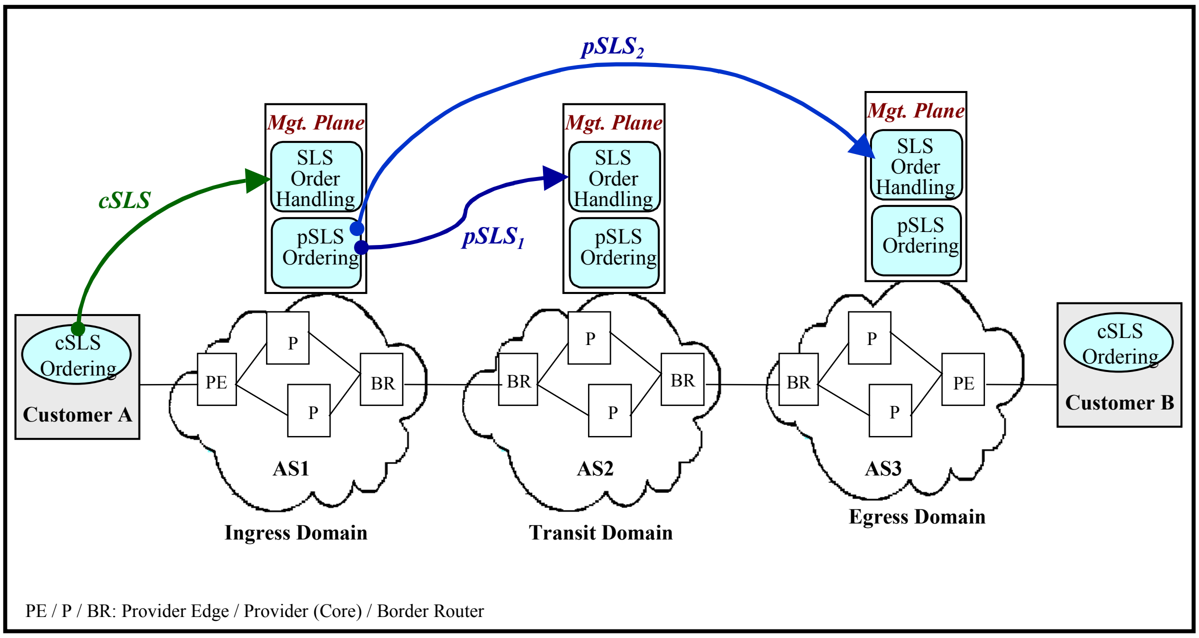


**Figure 1: Centralised Approach.**

## 1.2.2 Cascaded Approach

In the cascaded approach, each NP makes *pSLS* contracts with the immediately adjacent interconnected NPs. Thus, the QoS peering agreements are between BGP peers, but not between providers more than "one hop away". This type of peering agreement is used to provision the QoS connectivity from a customer to reachable destinations that may be several domains away.

Figure 2 gives an overview of the operations in this approach. The domain AS3 supports an intra-domain QoS capability (l-QC1). AS2 supports an intra-domain QoS capability (l-QC2) and is a BGP peer of AS3. AS2 and AS3 negotiate a contract (*pSLS2*) that enables customers of AS2 to reach destinations in AS3 with a QoS (e-QC1). This process can be repeated recursively to enable AS1 to also reach destinations in AS2 and AS3, but at no point do AS1 and AS3 negotiate directly. In each step of the cascade, the upstream provider (AS) acts in the consumer role to the provider immediately downstream, which is acting in the provider role.

It is each provider's responsibility to make appropriate *pSLS*s with the immediate downstream provider making it possible for individual customer IP QoS services to be created and managed along the entire route.

Within the context of MESCAL project, the solution options explained in [D1.1] are based on the cascaded approach.

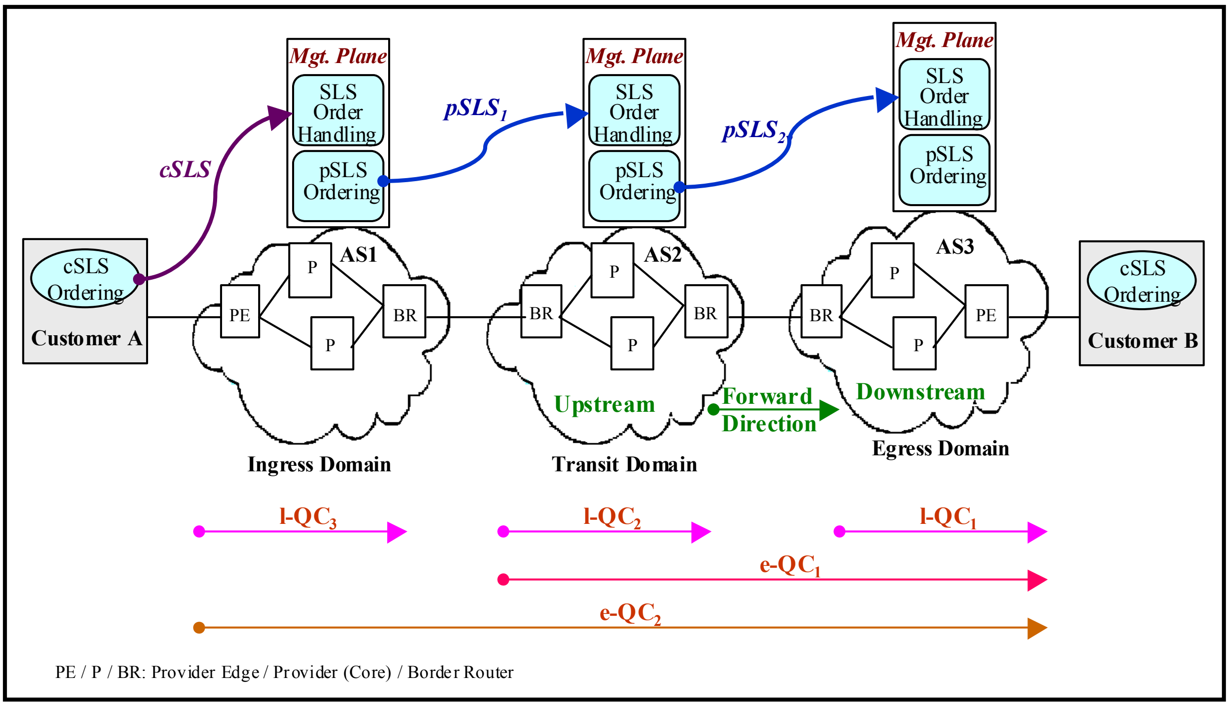


**Figure 2: Cascaded Approach.**

### 1.2.3 Hub Approach

In the hub approach, depicted in Figure 3, the role of Service Provider (SP) is separated from the NP. Here, the SP takes the responsibility for the overall service management of any given customer IP QoS service instance. This is achieved by making *pSLS* contracts with a chain of NPs so as to create an end-to-end service. The SP has the necessary facilities to support IP QoS services and all necessary management and control functions to control and to interact with the network providers. An end-to-end customer service instance is built using *pSLS* agreements between the SLS Ordering of SP and the SLS Order Handling entities of the interconnected networks. The hub approach is “Service Provider centric” as the SP has customers in different domains while the cascaded approach is “Network Provider centric”.

It should be noted that the MESCAL business model [D1.1, Chapter 2] precludes the Service Provider as a separate entity from its focus, thus MESCAL does not support the hub approach.

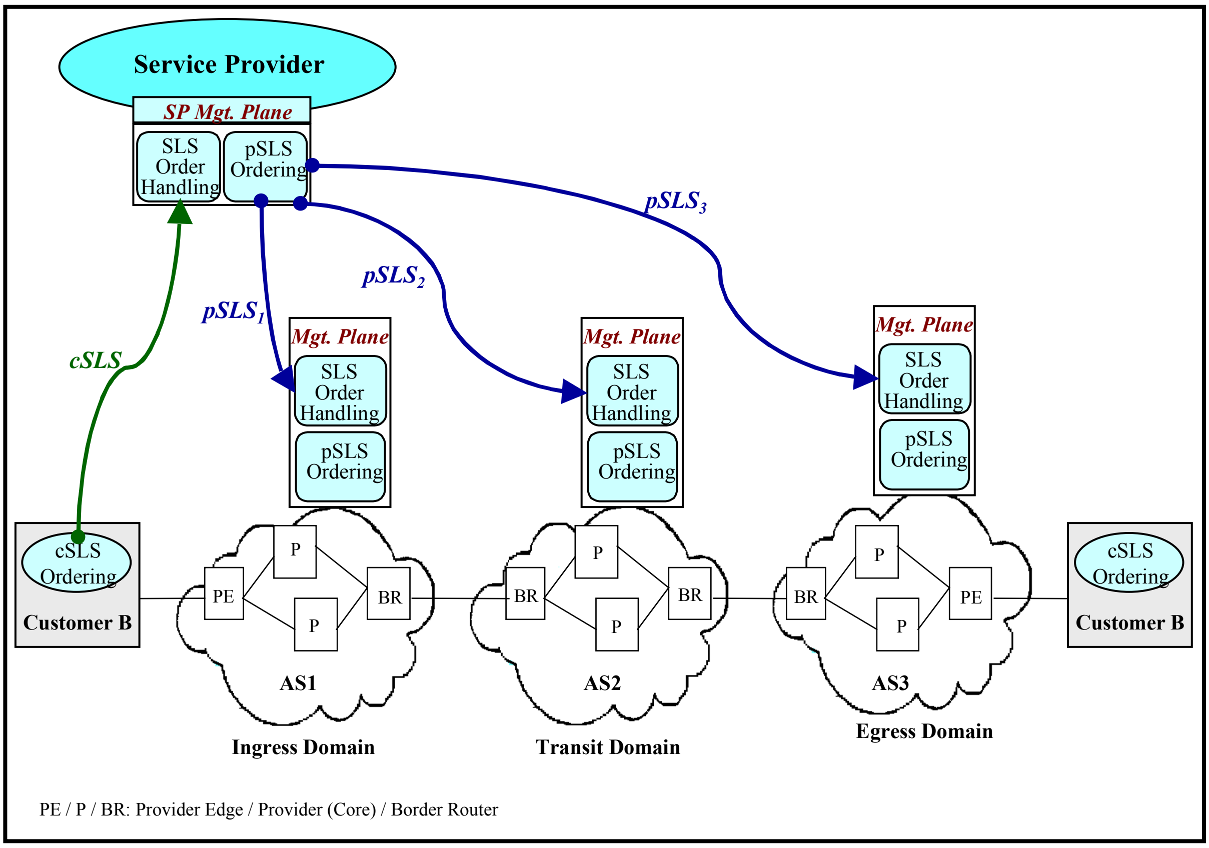


**Figure 3: Hub Approach.**

## 1.2.4 Hybrid Approach

The hybrid approach is combination of the centralised and the cascaded approaches, in which provider's roles are adjusted according to the requirements for providing end-to-end services. An example of using the hybrid approach is shown in Figure 4 where there is not a *pSLS* agreement in place between AS3 and AS4 to provide a QoS service between customer A and customer B using the cascaded approach. Thus, SLS Ordering of AS1 may request a *pSLS* with AS4 in order to complete the end-to-end peering connection and provide the service to its customer.

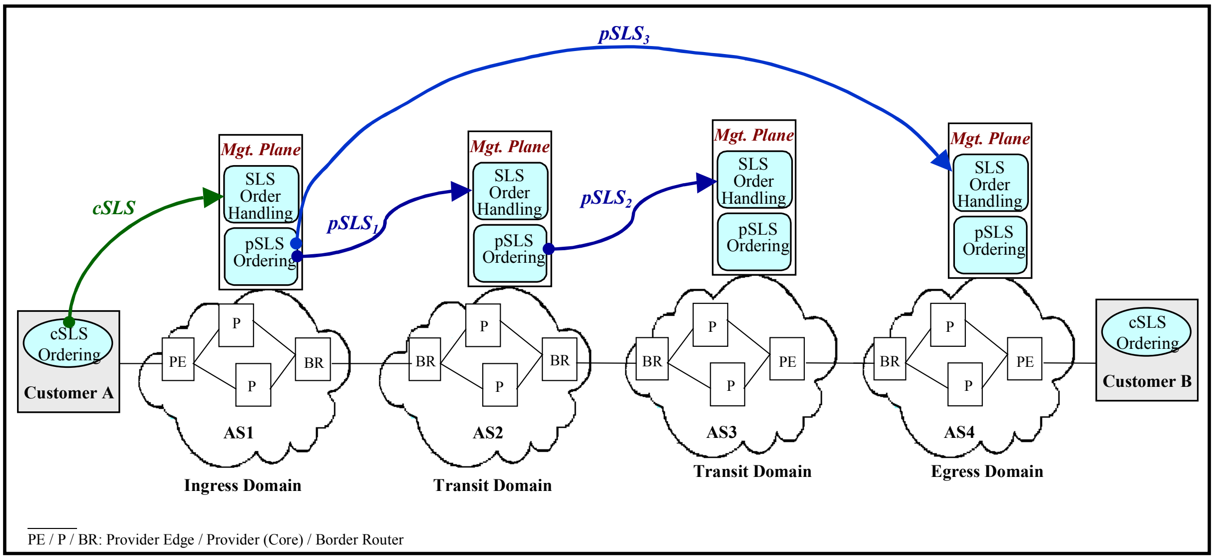


**Figure 4: Hybrid Approach.**

To provide the service, AS1 communicates with AS2 and AS4 directly and AS2 communicates with AS3 in order to establish *pSLS1, pSLS3,* and *pSLS2* respectively. That means there is no *pSLS* between the AS1 and AS3 while there is a *pSLS* agreement between AS2 and AS3. Towards AS1, AS2 manages and is responsible for the AS2 network and the link to AS3. Towards AS1, AS4 is responsible for the AS4 network. Towards AS2, AS3 is responsible for the AS3 network and the link to AS4.

The value of the hybrid approach is that it combines the benefits of the Cascaded and Centralised approaches. However, it also suffers from the limitations of both approaches.

The recursive implementation or "nesting" of multiple instances of the centralised approach is also possible, but this is almost indistinguishable from the hybrid approach. However, if a nested centralised approach were used, it would mean that any NP responsible for a portion (where this could be either the initiator, transit or destination ASs) of the end-to-end connection could delegate the implementation of this connection to several other NPs.

## 1.3 Comparison of Peering Approaches

In the following sections we focus on the two major approaches, the centralised and the cascaded for comparison.

### 1.3.1 Strengths and Limitations of the Centralised Approach

**Topology information:** the central point (initiator) in the centralised approach requires an up-to-date topology of Internet including the existence and operational status of every physical link between ASs for selecting the appropriate ASs in order to perform mapping & binding of QCs as well as negotiating and establishing *pSLS* agreements.

**QC mapping and binding**: Each domain must advertise its capabilities (i.e., o-QCs) to the outside world. Here, there is a one-to-one map between an l-QC and an o-QC. The central point needs to know all domains' advertised o-QCs for QC mapping and bindings to form e-QCs.

**Inter-domain routing:** Since the central point has access to the overall topology information and the interconnected links, it is possible to find and set-up the optimal routes to the destinations. The centralised approach establishes an explicit inter-domain route from source to destination, which is *pSLS* constrained and may not be the BGP route.

**Load balancing**: The initiator, who does have all topology data, has the possibility to set up protection routes via different networks for re-routing or load balancing purposes.

**Peering points**: the central point tendency is towards peering points between different NPs that satisfy its own requirements. Thus, the transit domain operator does not have control over the selection of egress points.

**Flexibility**: The centralised approach provides a high-degree of flexibility at the *pSLS* negotiation level and connectivity in selecting the chain of ASs from the global domain in order to establish end-to-end service connections.

***pSLS***: *pSLS* agreements are tailored to the initiator requirements. This type of *pSLSs* may cost more to establish as they are adapted to the initiator requirements. Aggregation of traffic demands can only happen at the traffic forecast process according to the *pSLSs*.

**Information exchange:** The initiator is directly involved with the customer and with each individual NP in the chain in providing the peering connection. Any *pSLS* related message must directly be sent from any NP in the chain to the initiator.

***pSLS* assurance**: The initiator can itself deal with *pSLS* violations if it has the opportunity to directly verify the performance of *pSLSs* established with each network provider involved. However, it must obtain related monitoring information (delay, loss, and throughput) taken from all involved ASs in order to perform *pSLS* violation detection. The greater the number of *pSLSs*, the more information to manage.

***cSLS* assurance**: The initiator can detect any *cSLS* violations. Active measurement sessions for one-way delay/loss measurements are set-up between the initiator and the destination AS. Thus, the initiator can obtain related monitoring information (delay, loss, and throughput) from the destination AS in order to verify the *cSLS* performance.

**Traffic Conditioning and QC Enforcement**: *cSLS* related information or aggregated QoS class information specific to the initiator domain has to be communicated to all domains in the chain for performing traffic classification and conditioning.

**Scalability**: There are the following major concerns:

- A drawback of the centralised approach is the need for topology related information of the Internet. For a relatively small number of networks this may be feasible, but it raises scalability concern when large number of networks are involved. Generally, there is a scope associated with every *cSLS* in order to specify the boundaries (in terms of IP addresses) of the *cSLS* requested for. With this information the central point realise with whom (ASs) to negotiate *pSLSs* for reaching to requested destinations.
- The central point may end up with many *pSLSs* to manage.

  To analyse this and for simplicity, we assume the number of l-QCs is equal across all domains (e.g., to support EF, AF, and BE traffic). We constrain the possible combinations between the QCs. Thus, we assume l-QC for providing EF is only mapped to the similar l-QC in the following AS. We also assume a single peering point between two ASs.

  $N_{lqc}$ = Number of l-QCs in each domain which is a constant value.

  $N_s$ = Number of *pSLS* agreements required from a central point to reach an AS for a single e-QC. By e-QC we mean the end-to-end QC constructed by the central point for offering o-QC to the central point customers. Central points negotiates l-QCs and construct e-QC that an e-QC = l-QC1 + l-QC2 + …+ l-QCn in terms of delay from the central point of view.

  $i$ = Number of transit hops (ASs) plus the egress hop in order to construct an end-to-end path from source to destination.

  $N_d$ = Number of AS domains in the Internet.

  $N_p$ = Number of *pSLSs* from a central point to reach all ASs for all e-QCs.

  $N_{pt}$ = Number of total *pSLSs* required to offer QoS-based services across Internet.

  $$N_s = i$$

  $$N_p = N_{lqc} * \sum_{j=1}^{N_d-1}(N_d - j) = N_{lqc} * \left[\frac{N_d * (N_d - 1)}{2}\right]$$ in the worst case as shown in Figure 5 (a).

  $$N_p = N_{lqc} * (N_d - 1)$$ in the simple case as shown in Figure 5 (b).

  $$N_{pt} = N_d * N_p$$

  The worst case is that where the central point is the furthest away from the destination AS. By assuming the number of l-QCs is constant, the order is $O(N_d^2)$ in the worst case and the order is $O(N_d)$ for the simple case.

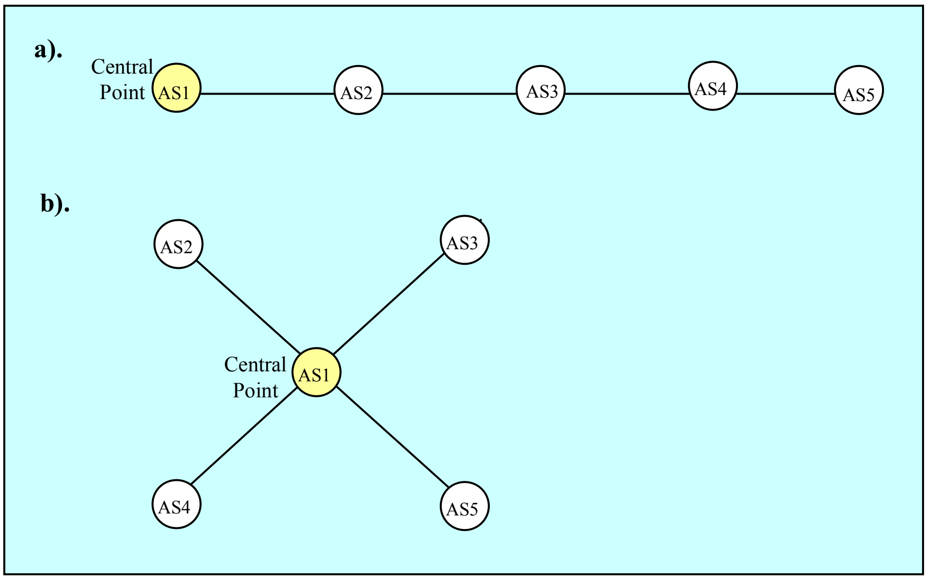


**Figure 5: Location of initiator (central point) in the network.**

Thus, an order of $O\left(N_d^2\right)$ of *pSLSs* may need to be established by a central point making the scalability of centralised approach a cause for concern.

- The processing time of any SLS Monitoring for SLS assurance may grow proportionately to the number of *pSLSs* being monitored by the central point.

**Implementation Complexity**: The centralised approach requires topological information and knowledge of l-QCs of all domains. That increases the complexity for the establishment of *pSLSs*, and therefore can be difficult to implement.

**Service offering**: In general and with regard to offering unidirectional services to the customers, the Pipe connections can be implemented with both centralised and cascaded approaches. The centralised approach may not be well suited for implementing the Hose and Funnel models. For the Pipe (1, 1) or Hose (1, N) *cSLSs,* there is only one ingress node. But in the case of hose, as there may be N distinct destinations (or egress points) to be reached in separate NPs the initiator NP may need to establish at least a large number of *pSLSs* to ensure packet delivery. For example, in the case b of Figure 5 where the destinations are only one hop away from the central point (AS1), the number of *pSLSs* to establish is equal to N. But using similar example as in the case b of Figure 5, if each destination is M hop away from AS1, the number of *pSLSs* to establish is equal to N*M.

For the Funnel (M, 1) similar to Hose, the number of *pSLS* contracts is increased proportionally with the number of destinations or egress points (N).

For bi-directional service offering, the establishment of bi-directional *pSLSs* is straightforward using the centralised approach because the initiator can request for two unidirectional *pSLSs* in opposite directions at the same time.

## 1.3.2 Strengths and Limitations of the Cascaded Approach

**Topology information:** Each NP in the chain only needs to know its adjacent neighbours and the status of related interconnection links.

**QC mapping and binding**: Each NP only needs to know its own l-QCs and the e-QCs advertised by its neighbouring domains in order to construct end-to-end QCs (e-QCs). This is true for every NP involved in the chain in order to construct its e-QCs.

**Inter-domain routing:** In this approach, inter-domain routing is also *pSLS* constrained. By *pSLS* constrained, we mean that traffic will only pass through the ASs where there are *pSLS* agreements already in place. If there is no *pSLS* agreement there is no way for an AS to transport the respected traffic. Each AS participating in the chain does not have all topology data and there are less possibility to set-up an optimal route to the destination. Each NP is constrained to select downstream neighbour in order to reach to the destination. Thus, the route selection for establishing *pSLS* agreement need a selection mechanism as it is possible for each AS to receive advertisements for similar o-QCs (provided by different downstream domain and using different routes to destinations).

**Load balancing**: The initiator NP, who does not have all topology data, has less possibility to set-up the protecting route via different networks.

**Peering points**: Each AS's tendency is towards peering points which to satisfy its own requirements rather than any specific requirements of initiators. Thus, each AS could have a limited number of peering points to manage.

**Flexibility**: A limitation of the cascaded approach is that it gives the service initiator NP less flexibility and control of the whole IP service path. The initiator NP is obliged to use the e-QCs constructed by the downstream domain.

***pSLS***: A provider is able to aggregate traffic demands for establishing a single *pSLS* with its adjacent provider's domain if that traffic enters the provider's domain from the same ingress point, has the same QoS requirements, and destined for the same destination point in spite of the fact that traffic are originated from different sources. Therefore, *pSLS* agreements can be tailored at nearly optimum level by aggregating the customers' traffic demands.

The e-QCs are constructed recursively, therefore, the traffic from different sources to the same destination having the same QoS-class can be aggregated as their path merge, i.e., they use the same o-QC from current to the destination domain. This provides the opportunity for constructing *pSLSs* downstream based on the aggregated traffic demands as shown in Figure 6. This figure shows that *pSLS5* agreement between AS3 and AS4 for transporting AS3 traffic is established for the aggregate traffic coming from AS3 but with different original sources and destined for the same destination domain utilising the same o-QC.

This aggregation cannot occur in the centralised approach, as traffic demands cannot be merged.

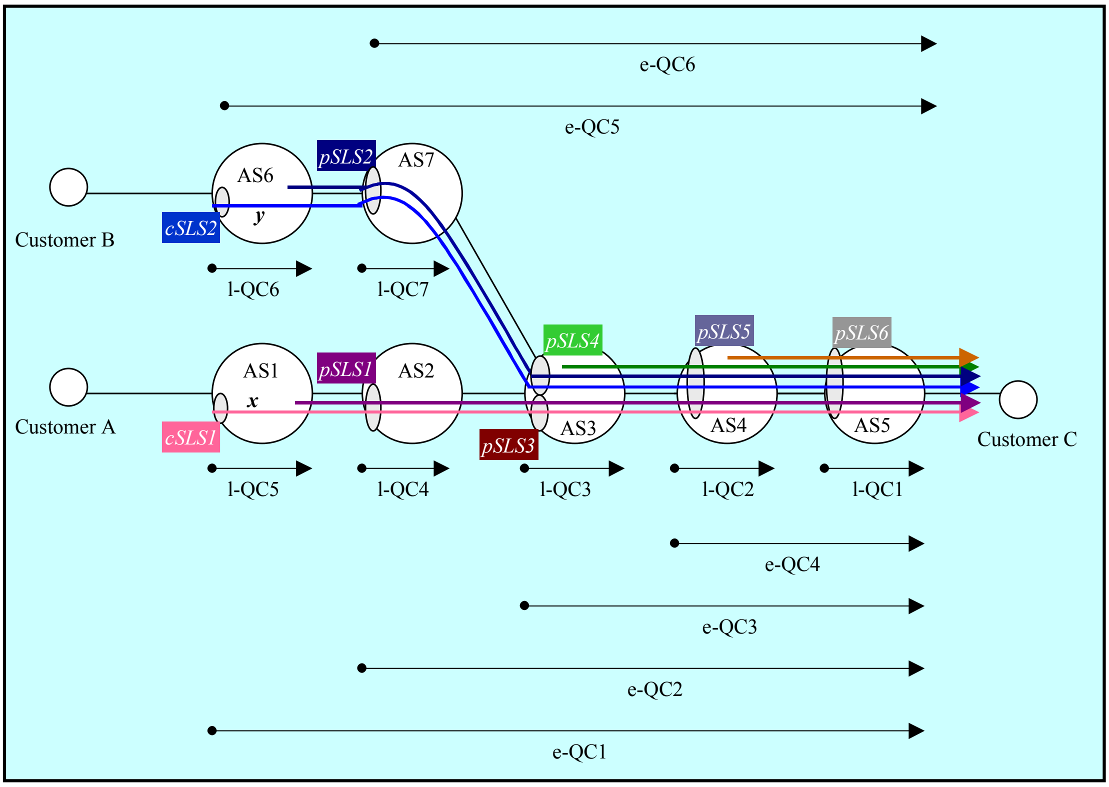


**Figure 6: Traffic demand aggregation for establishing *pSLSs*.**

**Information exchange:** The initiator NP is not directly involved with each individual network provider. Any SLS-related messages can not be directly sent to the initiator other than by propagating back through the adjacent upstream providers. In general, it takes longer for these messages to get to the initiator.

***pSLS* assurance**: Each domain can only directly detect the violations of *pSLSs* established with its neighbouring domains. The initiator NP or any upstream provider does not have the possibility to directly verify the performance of each network provider involved in the end-to-end chain if a *cSLS/pSLS* violation is detected.

***cSLS* assurance**: The initiator can detect *cSLS* violations. Similar to the centralised approach the initiator must obtain related monitoring information (delay, loss, and throughput) from the destination AS in order to verify the *cSLS* performance. As the destination domain does not have direct SLS relationship with initiator, the information related to cSLS performance has to be propagated back to the initiator by the adjacent upstream domains, otherwise an association is established.

**Traffic Conditioning and QC Enforcement**: *cSLS* related information might not be communicated to all domains in the chain. Information exchange at the aggregated QoS class level could be enough.

**Scalability:** There is no major scalability concern:

- In centralised approach the initiator has to know the Internet wide network topology while in the cascaded approach the initiator only know the links to neighbouring provider's networks. Therefore, it is expected that this model will be scaleable. Topology related messages such as messages regarding links only need to be sent to the "adjacent" providers. The parameters that will impact scalability are: the number of *pSLS* to establish and the qBGP traffic.
- As traffic demand aggregation can happen at *pSLS* level, each NP may only have a limited number of *pSLSs* to manage.

  To analyse this and for simplicity, similar assumptions as in the centralised approach are used here.

  $N_{lqc}$ = Number of l-QCs in each domain which is a constant value.

  $N_{oqc}$ = Number of o-QCs offered to each destination that is a constant value. (This could be equal to the number of l-QCs.)

  $N_d$ = Number of AS domains in the Internet.

  $N_s$ = Number of *pSLS* agreements required between two adjacent domains to reach an AS for a single e-QC.

  $N_{req}$ = Number of *pSLSs* required to be in place in the network to reach from a source to specific destination for a single e-QC.

  $i$ = Number of transit hops (ASs) plus the egress hop in order to construct an end-to-end path from source to destination.

  $N_p$ = Number of *pSLSs* from an AS to reach all destination ASs for all e-QCs.

  $N_{pt}$ = Number of total *pSLSs* required to offer QoS-based services across Internet.

  $$N_s = 1$$

  $$N_{req} = i$$

  $$N_p = N_{oqc} * (N_d - 1)$$

  $$N_{pt} = N_p * N_d$$

  Thus, an order of $O(N_d)$ of *pSLSs* needs to be established by an NP making the cascaded approach more scalable.

**Implementation Complexity**: This approach minimises the complexity of establishing pSLSs, less topological information is required and may be easier to implement.

**Service offering**: the Pipe model, Funnel model, and Hose model connections can be implemented for unidirectional service offering with cascaded approach. Using the same example as in case b of Figure 5 regardless of whether the destinations are one or M hops away from the central point (AS1), the number of *pSLSs* to establish is equal to N.

For bi-directional service offering, the establishment of *pSLSs* in reverse direction can be difficult using the cascaded approach depending on the MESCAL solution option. This is thoroughly discussed in Chapter 4.

### 1.3.3 Strengths and Limitations of the Hybrid Approach

The combined strengths and limitations of the centralised and cascaded approaches are applicable to the hybrid peering approach.

## 1.4 Conclusion

A single point of control for the service instances is the compelling feature of the centralised approach. The use of the centralised approach for more than a few interconnected NPs would be increasingly difficult to manage. Providers would prefer to offer services which reflect current Internet structure and for whom the use of the centralised approach would be inappropriate in many instances. Such providers would probably consider using the cascaded approach, which reflects the loosely coupled structure of Internet. Within the context of MESCAL, we focus on and provide solutions using the cascaded approach.

The cascaded approach makes it possible to build IP QoS services on a global basis while only maintaining contractual relationships with adjacent operators. Hence, this approach is more scalable than the centralised approach. This also reflects the current behaviour of BGP. A limitation of the cascaded approach is that it gives the service initiator less control of the whole IP service path.

With regard to service offering, Pipe VPNs can be implemented with both the centralised and cascaded approaches, but the centralised approach may not be suitable for implementing Hose and Funnel model connections.

# CHAPTER 2: IXPS AND THEIR IMPLICATIONS ON MESCAL SOLUTION OPTIONS

## 2.1 Interconnection Methods

In order to provide access to the Internet, an ISP must have connectivity to the global Internet. There are two options to interconnect ISP domains: Direct circuit interconnection and Internet Exchange based interconnection. The direct circuit interconnection option requires the lease of point-to-point circuits between ISPs. This scales linearly with the number of autonomous domains. The Internet Exchange Points (IXPs) interconnection option provides rich inter-ISP connectivity within the exchange. While the role of IXPs is often invisible to the end user, they form a very important part of the overall infrastructure of the Internet. An IXP is a physical network infrastructure operated by an entity with the purpose of facilitating the exchange of Internet traffic between ISP domains. Any ISP that is connected to an IXP can exchange traffic with any other ISPs connected to the same IXP, using a single physical connection to the IXP, thus overcoming the scalability problem of individual interconnection links. Also, by enabling traffic to take a more direct route between many ISP networks, an IXP can improve the efficiency of the Internet.

IXPs are also called 'public peering points'. IXPs may also provide private peering that is similar to providing a direct interconnection to peering ISPs. IXPs are used for peering. Most IXPs include clauses in the terms and conditions of the IXP that forbid transit across the IXP medium and most IXPs monitor traffic to verify this [Raveedran00]. In order to provide transit relationships happening over an IXP, the provider who wishes to sell transit, can co-locate a router at IXP and any ISP who wants to buy transit, should either directly connect to the provider’s router or co-locate its own router at the IXP and add a back-to-back connection to the provider’s router co-located at the IXP.

## 2.2 IXP Models

There are different types of IXP models: commercial, non-commercial, and government/educational.

- Commercial IXPs are built and operated by a Telco or co-location provider. Profit is not from the IXP services, but from the services that support the IXP, such as co-location spaces, telecom services, etc. Examples of commercial IXPs are MAE, PacBell NAP, Equinix, etc.
- Non-commercial IXPs are co-operatives funded by membership fees paid by the connected ISPs, and are operated for the benefit of the member ISPs and the Internet community at large. These are neutral IXPs for Internet peering. Most European IXPs are non-commercial, for example LINX, AMX-IX.
- Government and Educational institutions build IXPs to enhance their own connectivity. StarTap is an IXP for R&D networks.

## 2.3 IXP’s Networking Infrastructure

Some IXPs limit their scope of activities to purely providing a switched infrastructure, whilst others offer extra technical services. The vast majority of IXPs have adopted a layer 2 switched Ethernet architecture. Gigabit Ethernet public peering has proven to be successful and is the current practise. Amsterdam IX (AMX-IX) and LINX are examples that use Gigabit Ethernet switches at multiple locations. There are examples of other architectures such as ATM and FDDI however these are not common and they have been deprecated. PAC-Bell NSP is a L2 ATM IXP that uses PVCs between ISPs. A L3 IXP is an old Internet exchange technique and is not the best option for today’s Internet. In L3 IXPs, all traffic is exchanged inside a router.

The switch equipment is essential to the IXP. A simplified view of an Ethernet-based IXP infrastructure is given in Figure 7. ISPs can co-locate their border router (e.g., BR1) at the IXP and provide the connections (e.g., P1 to BR1) to the IXP. These ISPs are responsible for configuration of their routers co-located at the IXP. The IXP is responsible for configuration and operation of interconnection medium. It should be noted that it is quite possible to operate the IXP using a block of addresses from a member ISP, or even a third party, but for orderly management and administration of the IXP it is preferable for it to have its own address allocation.

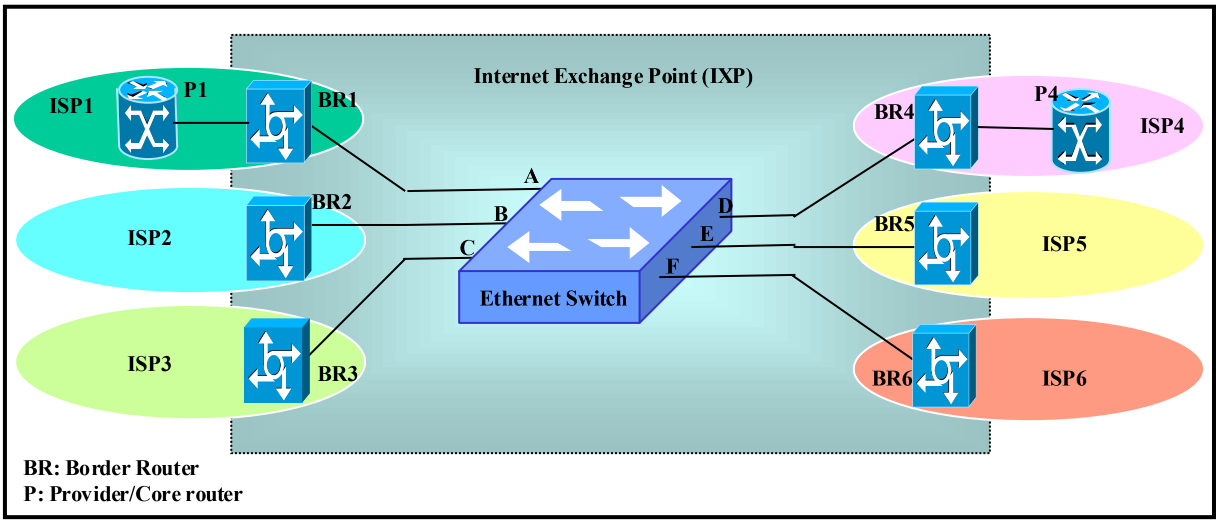


**Figure 7: A simplified view of an IXP infrastructure.**

IXPs normally have multiple switches for redundancy and providing continuity of service during routine maintenance or upgrade of any switch device. Many European IXPs have expanded to provide access to their switched infrastructure on multiple sites in a given metropolitan area. Most switches are currently Gigabit Ethernet switches offering 10/100Mbps as the basic interface standard. A number of IXPs have extended their service to provide access at Gigabit speeds (AMX-IX).

As an IXP example, the LINX network infrastructure operates a dual switch vendor architecture that consists of a number of high performance layer 2 switches from Extreme & Foundry. The switches are placed in LINX sites at managed facilities in London. A private dark fibre ring running Gigabit connects the sites together.

The concentration of ISP connections at an IXP can make it a very convenient place for one ISP to have a direct physical connection to another ISP with whom they exchange significant traffic. Also, if two ISPs have a substantial amount of traffic on the IXP medium, it is possible to create a private interconnect within the IXP to offload this traffic through a private connection. This helps to minimise traffic overload on the IXP’s medium. ISPs can use the same leased line into the IXP and could decide to easily change their connectivity to other ISP/NSPs as long as they provide IP connection at the IXP.

### 2.3.1 IXP Services and Ancillary Equipment

In addition to the main switch infrastructure, there is a range of ancillary equipment that is not essential to the core function of the IXP such as Router Server, Route Reflector and Multicast Server. IXPs also offer a range of services: looking glass, IXP’s web pages, web caching and replication services, content co-location, traffic analysis and other tools.

**Collector router:** To assist the IXP and members in troubleshooting, some IXPs provide a router with which all members peer and announce their routes. The router listens to, or 'collects' these announcements, but does not announce any routes itself; hence some IXPs use the term 'collector' router for this equipment. This provides a central 'view' of the IXP.

**Looking glass:** Looking glass is a management tool. Using looking glass the user is able to see permitted information on others machine i.e., BGP route table, BGP summary, route-reflector, trace, ping, whois, etc. For a list of most of the looking glass sites see http://neptune.dti.ad.jp/. Looking glass in an IXP allows some group (ranging from participants to the general public) to see layer 3 adjacencies and some statistics for the exchange; this can be derived from data seen by a participant at an exchange or from a route collector. The route collector listens to the announcements, but does not announce any routes itself. This enables to have a central 'view' of the IXP.

**Transit router:** Where an IXP has server equipment hosting, for example, their web site and email, and possibly some staff requiring Internet access, a router with full Internet connectivity is obviously required. Connectivity for this router, and therefore for the IXP's own network infrastructure, is often provided by member ISPs.

**Web and email servers:** An IXP will, of course, require equipment to host their web site and email. IXP web pages contain information on the IXP, the status, IXP statistics, etc.

**Content co-location:** This is to co-locate strategic content at IXP.

## 2.4 BGP Peering Modes at IXP

There are three modes of BGP peering in Ethernet based IXPs: Meshed Peering, Route Reflector Peering, and Router Server Peering. Router Server and Meshed peering are the common modes used on the IXPs around the world.

### 2.4.1 Meshed Peering

At the IXP, each ISP's Border Router (BR) establishes a peering session with every other BR if so desired for peering. This creates a full-mesh BGP peering among all the ISPs on the same IXP. Each BR is an eBGP speaker that peers with every other BRs of other ISPs. This is to facilitate the full routing exchange among the ISPs attached to the IXP. The number of peering sessions each ISP router needs to process is *O(N)*. That is, if there were *N* ISPs present at an IXP, each would have *N-1* peering sessions. When *N* is a large number, a sizeable load could be placed on each router in order to maintain the required peering sessions and process the required routing information. Figure 8 shows meshed BGP peering

In meshed peering, each ISP BR router would need to perform two major functions: route processing and packet forwarding. A heavy traffic load could put a substantial extra burden on the routers. The load would be particularly heavy if the number of peering sessions was not small, the number of destination routes was large, and the policy was complicated. It would be ideal to have the routers concentrate on forwarding packets, and have another system handle routing.

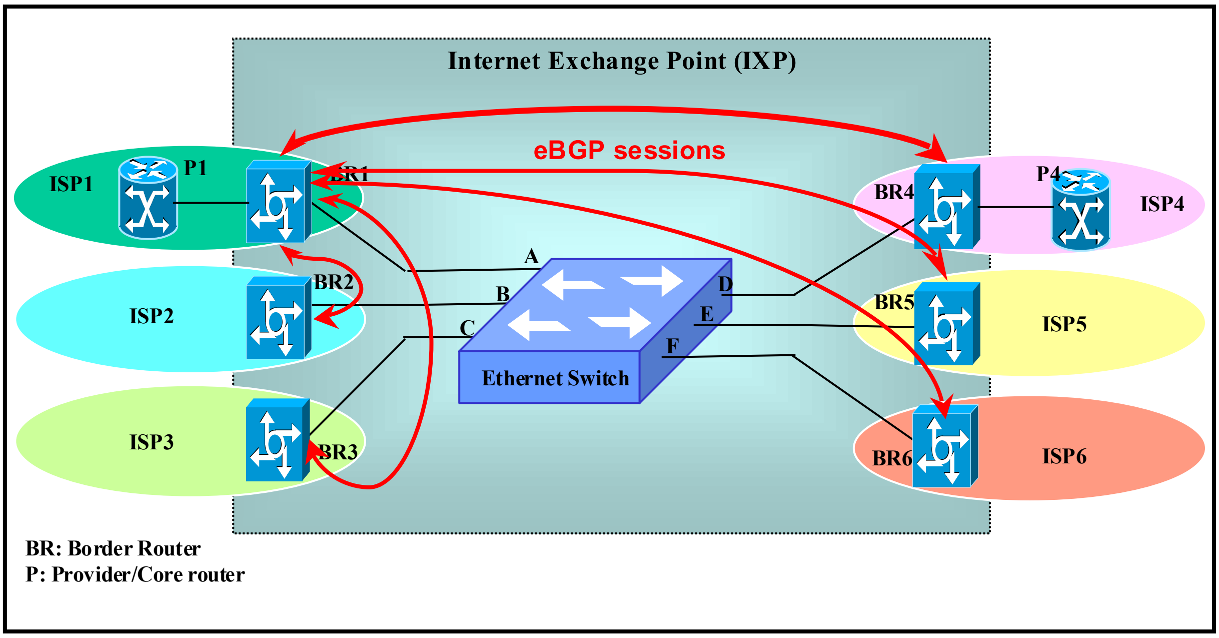


**Figure 8: Meshed ISP peering from the perspective of BR1.**

### 2.4.2 Route Reflector Peering

This mode relies on the BGP Route Reflector (RR) technology to enable peering between ISPs [Raveedran00]. BGP route reflector was originally designed to allow iBGP meshes to scale by added hierarchy. RFC 1966 provides the details specification of how BGP route reflector work. The L2 route reflector IXP uses routers as dedicated “BGP Route Reflector” to minimise the number of peering sessions a member router has to configure. Since the number of BGP peering sessions between ISPs is reduced, smaller routers can be used on the L2 RR IXP, reducing the cost of entry to the IXP. In a basic L2 RR IXP, the IXP provides the medium, the route reflector router and another router for the IXP services. The ISP members provide their own routers and connections to the IXP. The IXP management is responsible for the configuration and operations of the interconnection medium, the router reflector, and services on the IXP. The ISPs are responsible for the configuration of their routers. The L2 RR IXP needs its own BGP AS number that can add an AS number to the BGP path. Since it is discouraged to have transit service on the IXP, private AS numbers may be used on the IXP.

The ISP peering using a route reflector is shown in Figure 9. Each ISP’s border router (e.g. BR1) creates one iBGP session with the IXP’s RR. The ISP’s border router (e.g., BR1) will either originate or redistribute their routing information into IXP’s AS. The IXP’s RR will reflect these advertisements to the other routers (e.g., BR2, BR3, BR4, and BR5) on the IXP. While BGP routing information flows between the ISP’s border router and the RR, traffic between the ISP routers will not flow through the RR. When the BGP route is advertised by an ISP into the IXP’s AS, the RR will preserve the BGP Next Hop of the prefix. Hence, when 191.16.1.0/24 is advertised into the IXP’s AS, router BR4 will see a next hop of BR1. All traffic from P4 destined to ISP1 will go directly from BR4 to BR1.

Each ISP is responsible for its own router at the IXP (e.g., ISP1 for BR1). They will need to configure the router to advertise their IPv4 address block to the IXP while taking the advertisements from all other ISPs and communicate it through out their network. The advertisements happen on the router inside the ISP’s own AS number, e.g., ISP1 would advertise an aggregate of their CIDR block from router P1 via an eBGP session to router BR1. The link between router P1 and BR1 is provisioned, managed, and controlled by ISP1. Router P1 have IGP configured, but router BR1 would not have an IGP configured. Router BR1 would only have static routes and BGP running to pass information into its forwarding table. This would insure no leakage of ISP1’s IGP to any other ISP on the IXP. Router

P1 and BR1 would have eBGP configured between them. Each router would use the link IP address for their peering, not the loopback interface. Router BR1 would not have a default route configured for preventing one ISP hijacks bandwidth from another ISP.

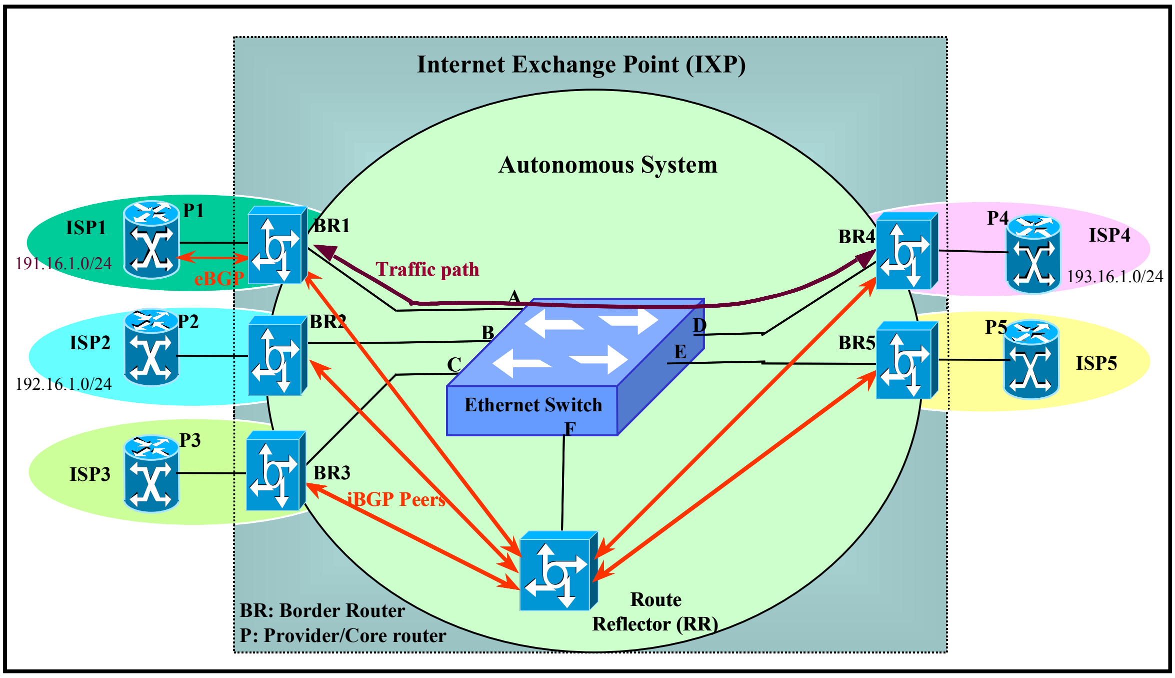


**Figure 9: ISP peering using route reflector.**

### 2.4.3 Route Server Peering

The Route Server (RS) peering is a way to scale the BGP session on an IXP. Similar to the RR, the Route Server separates routing from packet forwarding. It processes routing information for each ISP's router, thus enabling the ISP routers to concentrate on packet switching. The big difference between the RR and RS is that RS uses eBGP for each of the peers while the RR uses iBGP. Figure 10 shows the BGP peering using route server.

At the IXPs, tens of BRs are expected to attach. In RS peering BRs at an exchange point peer only with an RS. The RS thus reduces the number of peering sessions each ISP router needs to process from *O(N)* to *O(1)*. The Route Server facilitates and simplifies inter-domain routing among providers' routers at the Internet interconnection points by gathering routing information from ISP routers, processing the information based on the ISP's routing policy requirements, and passing the processed routing information to each ISP router. In order for the RS to tailor its route processing to meet the policy requirements of an ISP, the ISP must register its inter-domain routing policy information in the Internet Routing Registry. The Route Server will derive a given ISP's routing policy based on the information registered in the IRR. As a BGP peer, RS obviously needs to have an AS of its own. The RS is configured for propagation of routing information, from a policy database, with selection and export criteria of each BR at the IXP. RIPE maintains a policy database in Europe for enabling easier debugging of inter-domain routing, determination of Internet connectivity, etc. RS interrogates Internet Routing Registries, builds a database of the entries in the registries for the member networks, and provides a routing table based on this information. An IXP member's router may then build its routing table with just one peering session with the RS rather than taking many routing tables from all its peers. The principal aim is to reduce the processing power required in the member router connected to the IXP. Each BR advertises its selected routes to the RS and RS performs the inter-domain route computation on behalf of BRs.

All three peering modes are functionally equivalent as in either arrangement the same routes are selected at the IXP. RS and RR peering have some operational advantages over meshed: they simplify the BR administration and reduce route processing at BRs. In meshed, when a new BR attaches to the IXP, every other BR has to peer with this new BR. In RR/RS, the new BR only peers with RR/RS. There are two drawbacks: RS/RR reduces BR processing and storage at the cost of increasing their own. RS and RR are single point of failures.

The Route Server does not forward packets among the ISP routers attached to a connection point. Instead, it uses BGP's third-party routing information capabilities to pass routing information from one ISP to another, with the next hop pointing to the ISP router that advertises the route to the RS. Traffic is therefore exchanged directly among the ISP routers on the IXP, even though the Route Server provides the routing information.

The Route Server has the ability to create a Routing Information Base (RIB), known as a 'View,' for each ISP peer router. The view created for a given ISP maintains routing information, which meets the policy requirements of that particular ISP. The view makes it possible for an ISP peering with the Route Server to obtain the same routing information from the RS that it would if it peered with every other ISP on the IXP. That is, the Route Server could give a different path towards a given destination to different ISPs, if such paths were available and if such policy were required by the ISPs. The RS will not distribute routes learned from one ISP to another ISP without the permission of both. This permission will be expressed in terms of the ISP's routing policy registered in the IRR.

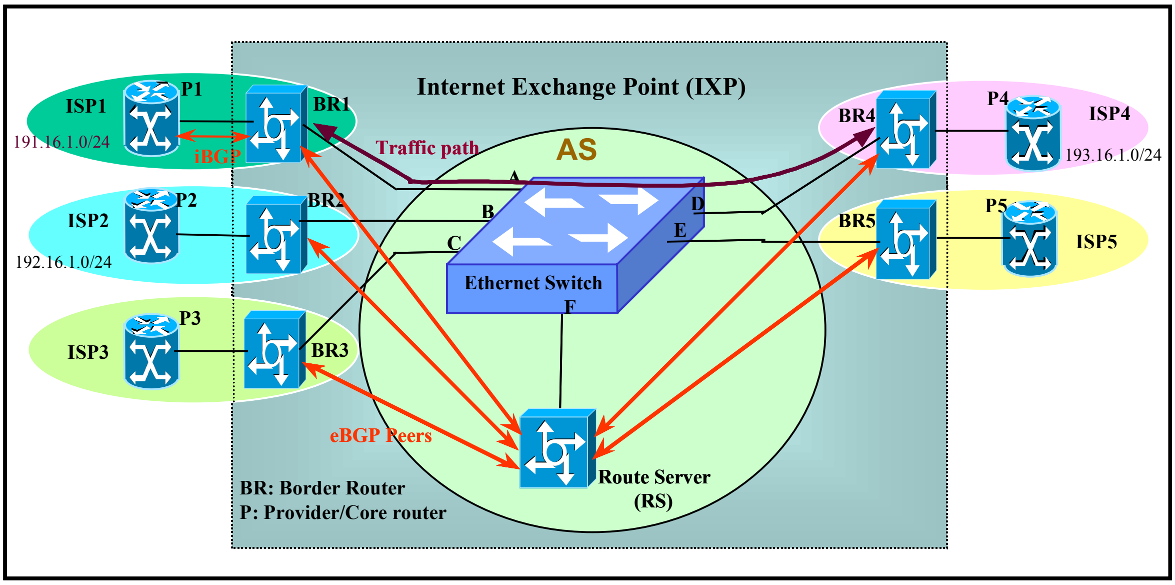


**Figure 10: ISP peering using route server.**

## 2.5 IXP Implication on MESCAL

### 2.5.1 IXP's AS Number in the AS Path

The RS can be configured to insert or suppress its own AS number in the AS path when passing routes from one ISP to another via BGP. This option is configurable on a peer-by-peer basis, and is configured according to the wishes of each ISP. This means that the Route Server can be viewed transparently when passing routing information. This can be done for RR peering as well.

### 2.5.2 BGP Peering Mode used at the IXP

Depending on the BGP peering mode used in the IXP, both RR and RS routers must run qBGP as it is used in the loose and the hard guarantee solution options. This is for RR in order to establish internal qBGP between itself and border routers and for RS to establish external qBGP between itself and border routers.

### 2.5.3 Performance Issues related to IXP's Network

Ports on a typical Ethernet hub all connect to a common backplane within the hub, and the bandwidth of the network is shared by all nodes that are attached to the hub. If two nodes establish a session that uses a significant amount of bandwidth, the network performance of all other nodes that are attached to the hub is degraded. To reduce degradation, the switch treats each port as an individual segment. When the stations on different ports need to communicate, the switch forwards the frames from one port to the other port at wire speed to ensure that each session receives full bandwidth. Ethernet switches support simultaneous, parallel connections between Ethernet segments. Normally, Ethernet operates in half-duplex mode. Gigabit Ethernet and 10-Gigabit Ethernet ports are normally operate in full duplex (e.g., Cisco's Catalyst 6500 series switches). In full-duplex mode, nodes can transmit and receive at the same time. To switch frames between ports efficiently, the switch maintains an address table. When a frame enters the switch, it associates the MAC address of the sending station with the port on which it was received. Switches build the address table by using the source address of the received frames.

The Ethernet services provided by the IXP's interconnect medium make use of protection techniques. These switches are all interconnected together using multiple diverse paths over dark fibre/WDM network. IXP implementation of Gigabit Ethernet even support "Jumbo frames" that can be as large as SDH frames, so there are no issues with datagram fragmentation in using GE to carry the traffic. "Jumbo frames" extends the Ethernet frames to 9000 bytes to support 8K application datagrams (Cisco's Catalyst support 9216 bytes).

Congestion in the switch may occur if incoming traffic from two/more ports compete for an outgoing switch port. Figure 7 shows a simplified view of an IXP infrastructure. We assume that ISP1, ISP2, and ISP3 all have peering agreements with ISP4. If traffic from A, B, and C ports are destined for D simultaneously, the sum of traffic volume from these three ports must be less that the capacity that port D can handle. Otherwise packets are dropped at the switch. This has to be taken into account by IXP.

Congestion also occurs at the BR if the volume of traffic exiting from the switch does not match the capacity of the link between BR and the ISP core. This is regarded as an internal ISP planning/engineering issue and the ISP should resolve it.

#### *2.5.3.1 An IXP Example (AMX-IX) in terms of Networking and Traffic*

Amsterdam IX (AMS-IX) is a non-profit, neutral and independent exchange. AMS-IX uses Gigabit Ethernet switch supporting 10/100Mbps connections. It dropped FDDI several years ago. AMS-IX is a distributed exchange, currently present at four independent locations in Amsterdam, as shown in Figure 11. The AMS-IX Infrastructure consists of one Ethernet switch at each location, interconnected by 10-Gigabit Ethernet over dark fibre. ISPs can connect to the AMS-IX infrastructure, using half-/full-duplex for 10BaseT or 100BaseTX, or full-duplex for 1000BASsX. Each switch is equipped with multiple interface boards with 10BaseT/100BaseTX and 1000BASsX ports.

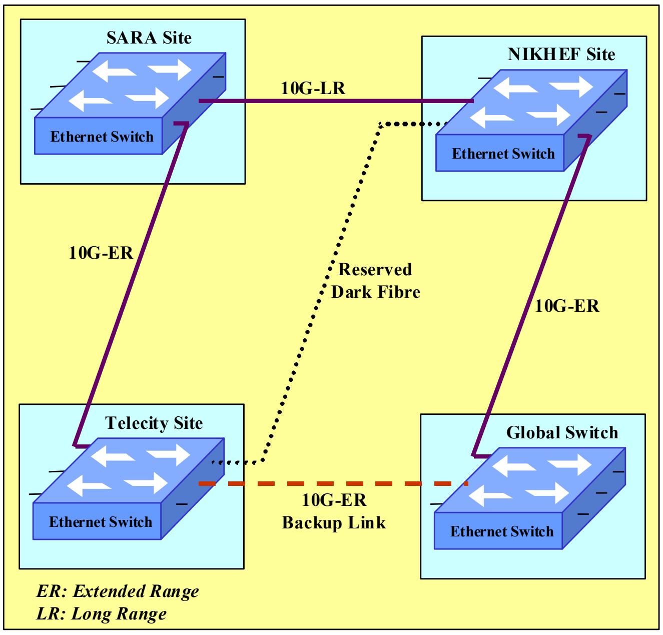


**Figure 11: Current AMS-IX Topology (source AMX-IX website).**

More than 160 ISPs are connected via 10-, 100-, 1000 Mbps Ethernet interfaces to the AMS-IX. These ISPs include all major Dutch ISPs and many international ISPs (especially from the United States, the United Kingdom, Germany, Belgium and the Nordic countries). Graphs in Figure 12 are taken from AMS-IX site and they show the hourly traffic on a normal day and monthly traffic in AMS-IX respectively. The aggregate incoming peak traffic is about 18 Gbps. Graph for Monthly aggregate traffic shows a steady increase of traffic in AMS-IX. In general, it is expected the IXPs provide non-congested environment for ISPs' traffic. For example, the load between two of AMS-IX sites, SARA and NIKHEF, was in the order of 2.5 to 3 gigabit/s. In May 2002, AMS-IX upgraded the existing link from 4 to 10 Gigabit/s.

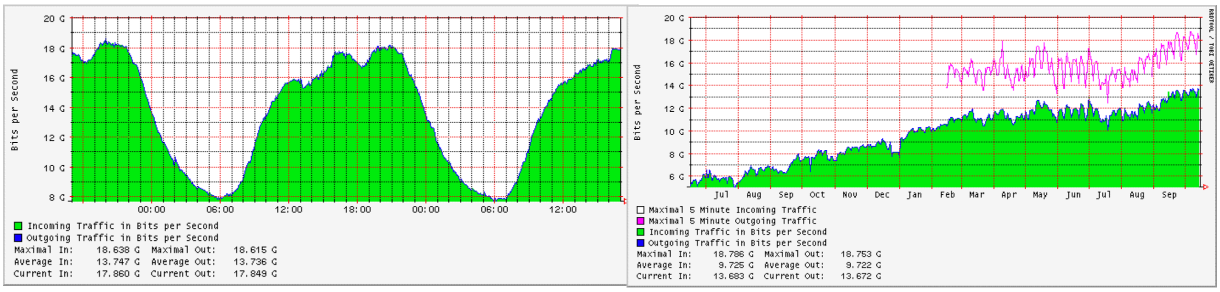


**Figure 12: Hourly and monthly (2002-03) aggregate traffic in AMS-IX (source AMX-IX website).**

## *2.5.3.2 Congestion Avoidance at the IXP*

Over provisioning is the first solution to avoid congestion at the IXP networking infrastructure. This is the view stated in the above example.

Generally, Ethernet lacks traffic engineering capabilities to ensure that a given amount of bandwidth on an Ethernet network is provisioned for a given ISP's service. There is normally no traffic differentiation in the IXP switch infrastructure. But it is possible to use prioritised services using 802.1p to have traffic differentiation at the IXP medium.

### 2.5.4 Peering Agreements

IXPs are not, generally, involved in the peering agreements between connected ISPs; this is a matter for the two ISPs involved to agree on peering. IXPs do however have requirements that an ISP must meet to connect to the IXP. The requirements that set by IXPs for public peering may have effects on the inter-domain peering between ISPs. For example, use of the AMS-IX IP unicast peering is limited to the following:

- The AMS-IX infrastructure is based on the IEEE 802.3 standard. This means that LLC encapsulation (802.2) is not permitted.
- 100base and 10base Ethernet interfaces attached to AMS-IX ports must be explicitly configured i.e. they should not be auto-sensing.
- Frames forwarded to AMS-IX ports shall have one of the following ether types: IPv4, ARP, and IPv6.
- Frames forwarded to an individual AMS-IX port shall all have the same source MAC address.
- Use of proxy ARP on the router's interface to the Exchange is not allowed.
- Frames forwarded to AMS-IX ports shall not be addressed to a multicast or broadcast MAC destination address.
- etc.

### 2.5.5 Configuration of Border Routers at the IXP

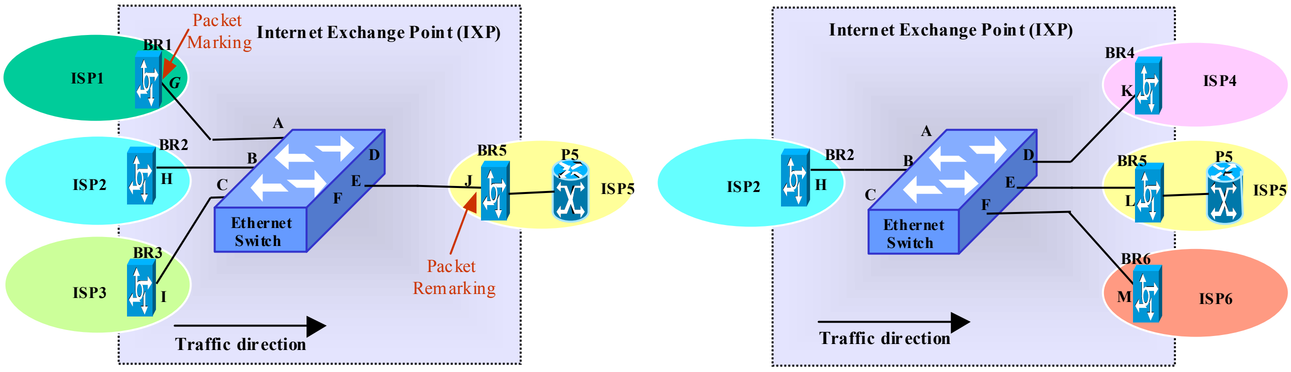
Each ISP is responsible for its own router at the IXP. The link between ISP’s core router and ISP’s border router co-located at the IXP is provisioned, managed, and controlled by ISP. Therefore, configuration of border router interface is the responsibility of the ISP without any involvement from the IXP. There is an issue in which the BR interface does not serve only a single ISP but a number of ISPs who use the IXP and have peering agreements with the aforementioned ISP. This issue is investigated in detail in the following section.

The use of RR/RS in the IXP minimises the number of peering sessions a member border router has to configure. Since the number of BGP peering sessions in the border routers are less, the processing requirements required in these routers is reduced and thus smaller routers are used. Care should be taken in not to overwhelm these border routers with high processing requests.

### 2.5.6 Implication of Interconnection States

There are two interconnection states (Fan-in & Fan-out) that arise from IXP involvement in the inter-domain chain. As shown in Figure 13a, ISP1, ISP2, and ISP3 each have a peering agreement with ISP5. The link connecting port E of switch to interface J of BR5 is shared by traffic from ISP1, ISP2, and ISP3 whose next AS hop is ISP5. This is called Fan-in state.

In the Fan-out state shown in Figure 13b, ISP2 has peering agreements with ISP4, ISP5, and ISP6. One interconnection link from ISP2 and consequently one interface (BR2-H) is used to connect to the IXP switch where the other three downstream ISPs are connected individually. Thus, the link connecting BR2-H interface to port B of switch is shared by traffic from ISP2 whose next AS hop is ISP4, ISP5, or ISP6.



a): Fan-in state

b): Fan-out state

**Figure 13: Interconnection states.**

With regard to interconnection states, the issue is how to identify the packets and classify them for remarking at the ingress points of border routers (e.g., BR5-J interface in Figure 13a), as packet's DSCP value may not be adequate information for packet remarking process. The same argument is correct in identifying the traffic for metering and policing purposes. The following sections discuss the DSCP marking/remarking issue related to each solution option and the traffic metering and policing issue.

## *2.5.6.1 Packet Marking/Remarking at Border Routers*

### *2.5.6.1.1 Loose Guarantees Solution Option*

In this solution option, the *Meta-QoS-classes* are used to indicate the requested QoS across the Internet. A *Meta-QoS-class* indicator is used both for both intra-domain and inter-domain QoS purposes. In intra-domain, the end-user submits a datagram with an indication of the requested *Meta-QoS-class* . Each provider chooses an appropriate l-QC for treating this datagram within its domain. The *Meta-QoS-class* indicator is kept in the datagram. The DSCP can be used as this indicator [D1.1-03].. When the datagram reaches a domain's boundary, the *Meta-QoS-class* indicator is used to indicate the QoS class between domains. This could be a global value agreed by all providers or a local value understandable by two adjacent eBGP peers.

If the *Meta-QoS-class* concept is globally known at the boundaries between the domains with respect to the use of DSCP (i.e., the same DSCP is used for identifying a *Meta-QoS-class* ), there will not be any problem concerning either "Fan-in state" or "Fan-out state". The other solution is as follows. The same DSCP value identifying each *Meta-QoS-class* must be used by the border routers (i.e., BR1, BR2, BR3 in Figure 13a) for traffic that is routed to BR5 as the next hop (see Figure 13a). This is also true for "Fan-out sate" because the DSCP values assigned by BR2 must be recognised in the same fashion by the three BRs (i.e., BR4, BR5, and BR6 in Figure 13b).

Therefore, it is preferable to use globally known values for *Meta-QoS-classes* or at a minimum to use well-known values for *Meta-QoS-classes* within the boundaries of IXP infrastructure. Otherwise, the solutions that are presented in section 2.5.7 might be considered.

### *2.5.6.1.2 Statistical Guarantees Solution Option*

In this solution option, the QC mappings happen at two levels. The first level is mapping within the AS, between the local l-QCs/e-QCs and the o-QC. The second mapping is an external mapping between the l-QCs and/or e-QCs of one AS with o-QCs of the adjacent ASs. DSCP is used to signal QC mapping. Thus, there are DSCP setting at both ingress point (DSCP remarking) and egress point (DSCP marking) of ASs according to the requirements set at *pSLS* agreement between two ASs.

When there are private links between the ISPs, DSCP marking/remarking at the AS border routers are straightforward. In the case shown in Figure 14, there are three private links between ISP1-ISP5, ISP1-

ISP5, and ISP3-ISP5 connected to three individual interfaces at BR5. Table 1 shows that packets are marked at the egress interfaces of BRs (A, B, C) and are remarked to appropriate DSCPs (for l-QC mapping) at the BR5 interfaces (D, E, F) based on their current DSCP and the interface they arrived at.

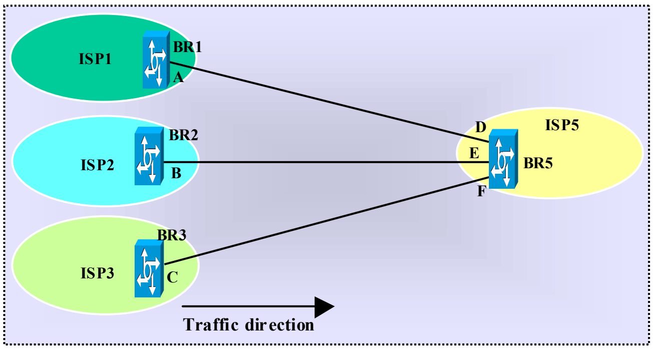


**Figure 14: ISP private interconnection.**

| Packet Marking | | | |
|---|---|---|---|
| *AS* | *Interface* | *Current DSCP* | *Mark DSCP to* |
| ISP1 | BR1-A | DSCP11 | DSCP21 |
| ISP2 | BR2-B | DSCP11 | DSCP31 |
| ISP3 | BR3-C | DSCP11 | DSCP41 |

| Packet Remarking | | | |
|---|---|---|---|
| *AS* | *Interface* | *Current DSCP* | *Remark DSCP to* |
| ISP5 | BR5-D | DSCP21 | DSCP11 |
| ISP5 | BR5-E | DSCP31 | DSCP11 |
| ISP5 | BR5-F | DSCP41 | DSCP11 |

**Table 1 : Packet's DSCP marking/remarking at BRs's interfaces for private interconnection links.**

In the IXP case (Figure 13a), if packet marking is performed at the egress points of ASs/ISPs (i.e., BR1-G, BR2-H, BR3-I) based on the bilateral peering agreements between any two ISPs, the BR5 cannot rely solely on the current packet's DSCP for remarking and it needs more information in order to identify the packet and apply the proper remarking. This is because the BR5-J interface serves traffic coming from three different BRs and the peeing agreements might established for the use of different DSCP for similar QoS classes. A solution may seem for BR5 to have a common understanding of QoS classes for packets that have the same DSCP but coming from different ISPs (ISP1, ISP2, ISP3). This common understanding must be achieved within the IXP. This approach may not be suitable for solution option 2 because it severely limits the number of QoS classes available for use.

The packet marking/remarking processes are shown in the following two tables. The DSCP marking at the BR1-G, BR2-H, and BR3-I interfaces (Figure 13a) are performed as shown in Table 2.

| Packet Marking | | | | |
|---|---|---|---|---|
| *AS* | *Interface* | *Current DSCP* | *Next Hop* | *Mark DSCP to* |
| ISP1 | BR1-G | DSCP11 | BR5-J | DSCP21 |
| ISP2 | BR2-H | DSCP12 | BR5-J | DSCP21 |
| ISP3 | BR3-I | DSCP13 | BR5-J | DSCP21 |

**Table 2: DSCP marking for the IXP case at ISPs' BR egress interfaces.**

At the BR5 interface (Figure 13a), packets must be remarked somehow to convey the appropriate l-QCs as shown in Table 3. A number of solutions to this problem are given in section 2.5.7.

| Packet Remarking | | | | |
|---|---|---|---|---|
| ***AS*** | ***Interface*** | ***Current DSCP*** | ***Other information*** *(see section 2.5.7)* | ***Remark DSCP to*** |
| ISP5 | BR5-J | DSCP21 | e.g., MAC address /VLAN Tag/GRE header | DSCP11 |
| ISP5 | BR5-J | DSCP21 | " | DSCP12 |
| ISP5 | BR5-J | DSCP21 | " | DSCP13 |

**Table 3: DSCP remarking for the IXP case at ISPs' BR ingress interfaces.**

In the Fan-out state, packet marking can be performed similar to the process shown in Table 2. But as the BRs (BR4, BR5, or BR6 in Figure 14b) receive packet from different upstream ISPs, the same problem stated for Fan-in state occurs for packet remarking.

#### *2.5.6.1.3 Hard Guarantees Solution Option using MPLS*

In this solution option, inter-AS QoS LSPs are constructed between edges of networks for transporting customer traffic. The LSPs could be either multicoloured or mono-coloured. In any case a DSCP to MPLS-EXP mapping has to be performed at the first LSR before injecting customer traffic to the correct LSP. For getting the correct PHB treatment at the AS, EXP to EXP mapping has to be performed at the border routers where the traffic enters another domain. As this option is also based on the *Meta-QoS-class* concept, the same recommendation given in section 2.5.6.1.1 is applied here.

It should be mentioned that in this option, MPLS LSPs are constructed edge-to-edge to provide QoS-enabled services. Thus, the border routers co-located at the IXP must be able to provide MPLS over Ethernet service. In addition for providing inter-AS MPLS tunnels, the IXP must allow and provide any necessary functionality for MPLS over Ethernet transport.

### *2.5.6.2 Traffic metering and policing*

At the border routers, traffic metering is used to measure temporal properties of aggregated QoS flows selected by the classifier against a traffic profile specified in the *pSLS*. Policing aims at controlling these traffic flows based on the information provided by meters. With regard to Fan-in state, the same issue of identifying packets for metering/policing at the ingress points of border routers (e.g., BR5-J interface in Figure 13a) applies here. Therefore, metering and policing has to be performed before remarking the packets to their new DSCP values at the border routers. To perform metering/policing according to the *pSLS* agreements, the border router need to inspect more information to identify packets as it has been discussed in section 2.5.7.

## 2.5.7 Potential Solutions: Traffic Classification using Layer 2/3 Information

Traffic classification is required when implementing marking, metering, policing, shaping, and queuing functions. There can be solutions to consider for identifying/classifying packets for DSCP remarking if bilateral agreements set between any two ISPs. Below, briefly explains three solutions that can be also used for metering and policing purposes.

### *2.5.7.1 MAC-based classification*

The simplest solution is to use MAC-based classifiers. In the Fan-in state and in BR5-J, source MAC address identifies the border router interface (e.g., BR1-G in Figure 13a) that packet arrived from. In the Fan-out state, the destination MAC address identifies for the border router (BR2-H interface in Figure 13b) packet's next hop address. It is possible to use MAC address information to classify packets (using access lists) at ingress point of BRs (BR4, BR5, BR6) and then remark the DSCP at BR5-J interface appropriately. This approach can be used for the three solution options. Cisco IOS provides this feature. Linux also support MAC address classifiers.

#### *2.5.7.2 Setting up VLANs*

The other alternative option is the use of VLANs [IEEE98]. This is to set up a VLAN between any two BRs whose ISPs have a peering agreements. The 802.1Q specification establishes a standard method for inserting VLAN membership information into Ethernet frames. A packet identification (packet tagging) process is normally used for layer 2 VLAN segmentation. Packets are assigned a unique packet identifier within each header. This header information designates the VLAN membership of each packet. This provides a virtual interface between two BR routers (e.g., BR1 and BR5 in Figure 13a). The virtual interface identification can be used for packet's DSCP remarking. IEEE 80.1Q allows up to 4095 VLANs. The VLAN approach may not be scalable if there is a need for fully meshed or a large amount of VLANs to construct.

#### *2.5.7.3 Using GRE tunnels*

Generic Routing Encapsulation (GRE) is a standard-based tunnelling protocol that can encapsulate a wide variety of protocols (packet types) inside IP tunnels, creating a virtual point-to-point link between points over an IP network. GRE is an encapsulation protocol defined in RFC-1072 (supported by Cisco IOS). GRE encapsulation consists of a packet header with components that allow it to identify data for processing when it arrives at the tunnel end. These components include an IPv4 tunnelling header; a GRE header with optional fields that include tunnel key, checksum, and sequencing fields; and the payload (i.e., tunnelled Layer 3 packet). Traffic from BR1 destined for BR5 is sent through a GRE tunnel. GRE tunnelling allows ISP1 and ISP5 to appear to be directly connected. The way to set up the virtual link between ISP1 and ISP5 is to encapsulate traffic from ISP1 in a GRE IP packet with a source and destination addresses. If the source address of the packet is set to BR1-G interface address, and the destination address is set to the BR5-J interface address (Figure 13a), all traffic sent from network ISP1 to ISP5 is transmitted across IXP network, through GRE tunnel. This GRE source/destination IP address should allow to identify the packets for DSCP remarking.

GRE is a point-to-point tunnelling protocol, so it would be operationally overwhelming to deploy GRE tunnelling in a full-mesh configuration, because of the difficulty in managing a large number of tunnels in a border router. Therefore, GRE is ideal in limited deployments.

## 2.6 Summary

In this document, the interconnection methods for inter-domain peering between providers are described. IXPs as the public peering points play an important part of overall Internet infrastructure. We explained the IXPs' networking infrastructure and the services they provide. The BGP peering modes at IXPs are explained. Then, the IXP participation in the inter-domain connectivity chain and its implication on MESCAL solutions are investigated in detail. Some solutions are provided to overcome the problems risen from the involvement of IXP in the QoS delivery chain with regards to the MESCAL solution options.

# CHAPTER 3: OPTICAL NETWORK TECHNOLOGIES & THEIR IMPLICATION ON MESCAL

## 3.1 Introduction

This chapter considers the underlying transport network provided by Physical Connectivity Providers (as per the MESCAL business model described in D1.1) and its usefulness and interfacing to IP Network Providers offering one or more MESCAL service options. The transport provided will be used by the IP Network Provider either internally to an AS or will provide connectivity to Internet Exchange Points or between Network Providers and their customers and peers. The provisioning can occur at a range of time scales. On a monthly scale new IP peering agreements will cause the network planner to request new or additional physical connectivity between IP Network peers. On a short time scale the Intra- and Inter- Domain provisioning cycles could cause the creation of new links and/or the modification of existing links' capacities.

To provide the large capacity required, these links are predominantly made available in the optical domain. This chapter describes the standards and technologies available to configure and manage an optical network. The integration of MESCAL Solution Options into optical control planes and possible future optical network developments are discussed.

## 3.2 Approaches to Optical Networking

The current most widespread approach to optical networking is the provisioning of static wavelengths within fibres in WDM (Wavelength Division Multiplexing) systems. Newer DWDM (Dense Wavelength Division Multiplexing) can provide of the order of 64 wavelengths per fibre over distances of 1000km with 40 Gigabits/sec per wavelength (Lucent LambdaXtreme transport [LCNT]). Current deployed WDM technologies usually require manual configuration of physical equipment rather than a highly automated configuration scheme. For this reason reconfiguration can take months, hence the evolution towards manageable intelligent optical networks (such as the Lucent system stated above).

### 3.2.1 Static Point-to-Point Links

Static point-to-point links provide fixed paths for wavelengths between two geographic locations. Network configuration is performed through the use of electrical switching or physical reconfiguration by an engineer. Electrical domain switching is not however fast enough for new applications and emerging line speeds and therefore new all-optical approaches to wavelength switching are being developed.

Static point-to-point links rely on Layer 2 switching to provide less restrictive transport, through the use of technologies such as SDH (Synchronous Digital Hierarchy), FDDI (Fibre Distributed Data Interface) or Ethernet (such as RPR (Resilient Packet Rings)). These technologies provide features like capacity management, resilience and routing but their bandwidth scalability is limited by electrical switching speeds.

As there are no management or control plane means to configure the wavelengths this option is not considered further as a suitable solution for dynamic network provisioning for MESCAL.

### 3.2.2 Intelligent Dynamic Optical Networks

To support the physical connectivity demands of MESCAL solution options at the fastest possible provisioning speeds with the least restrictions (capacity granularity, enforced topology, hierarchy etc.), it is envisioned that intelligent dynamic optical network would be required. While SDH and many other Layer 2 protocols could support the capacity requirement, their switching and transmission bandwidth limits are being approached. The emerging technologies considered here are GMPLS (Generalised Multi-Protocol Label Switching) and ASON (Automatically Switched Optical Networks). These technologies provide an overlapping set of features that could be used in future networks to provide all optical dynamically re-configurable capacity.

### *3.2.2.1 GMPLS*

GMPLS (Generalised Multi-Protocol Label Switching) is the union of existing MPLS solutions, MPLambdaS (MPλS) and label switching through TDM networks. MPLambdaS provides for the configuration of optical forwarding as well as features associated with MPLS such as label nesting and link bundling. While MPLS was predominantly label switching through IP, ATM and FR clouds, GMPLS is now a unified control plane for label switching through packet, TDM, wavelength and spatial switched environments.

#### *3.2.2.1.1 MPLambdaS and Lambda Switching*

MPLambdaS (MPλS) uses optical cross-connects to space and wavelength switch between interfaces. The MPLambdaS equivalent of labels is now wavelengths (it does not however switch optically on label headers within the data being transmitted) and fibres, conceptually therefore a typical MPLambdaS node may appear as shown in Figure 15. This figure shows an optical cross-connect (OXC) performing lambda switching (itf1, itf2,… = interface 1, 2, …) according to a label enabled control plane. The equipment is really an optical cross connect (OXC) switch in the data plane, coupled to an LSR (Label Switched Router) type control device in the control plane which does not deal with the actual forwarding but only controls the configuration of the OXC.

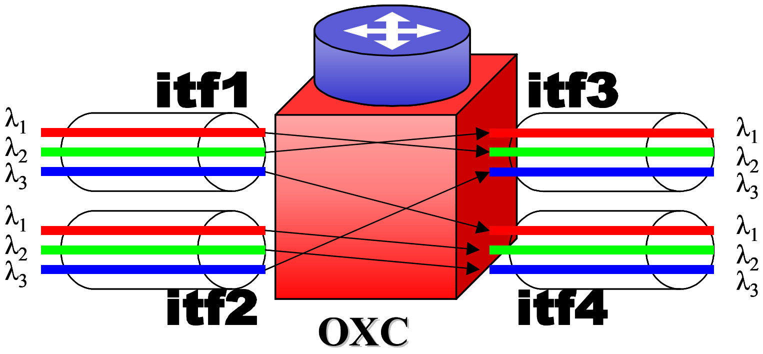


**Figure 15: An OXC performing lambda switching according to a label enabled control plane.**

A typical configuration may look like:

```
<itf1, λ₁> → <itf3, λ₂>
<itf1, λ₂> → <itf3, λ₁>
<itf1, λ₃> → <itf4, λ₁>
…
<itf2, λ₃> → <itf3, λ₃>
…
```

Given a fibre topology and wavelength switching, an example MPLambdaS network is shown in Figure 16.

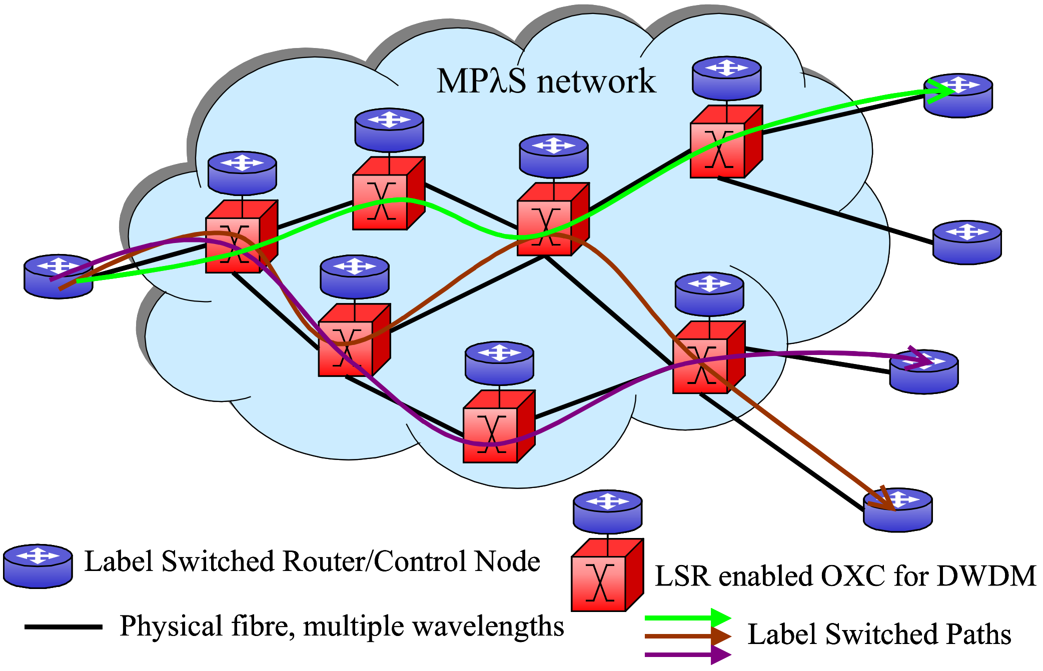

MPλS network
Label Switched Router/Control Node
LSR enabled OXC for DWDM
Physical fibre, multiple wavelengths
Label Switched Paths

**Figure 16: An MPLambdaS network where the end-to-end light paths are decided by the GMPLS control plane.**

The MPLambdaS control nodes are connected via a control plane (control plane links are not shown in the figure) network that does not necessarily follow the data plane topology and could be in or out of band and is carried over IP. A separate control plane is required in TDM and optical networks since they cannot inspect packet headers and therefore cannot switch depending on labels in headers.

#### *3.2.2.1.2 GMPLS Protocols*

As the concepts in GMPLS are very similar to those of MPLS, existing protocols are used with minor modifications to deal with the concept of wavelengths or TDM streams.

##### *3.2.2.1.2.1 Signalling*

GMPLS supports signalling protocols such as RSVP-TE (Resource reSerVation Protocol - Traffic Engineering) [RSVPTE] and CR-LDP (Constraint-based Routing Label Distribution Protocol) [CRLDP]. These protocols are used to initiate the creation of LSPs (Label Switched Paths), the distribution of network information (label exchange) and support various features such as bi-directional LSPs (a single LSP is usually uni-directional) and protection LSPs. The protection LSPs are pre-calculated and can provide span or path protection using shared (M:N) or dedicated (1+1) protection capacity. RSVP-TE also has support for wavebands (the aggregation of contiguous wavelengths in a single fibre which will be switched together) thereby providing the ability to quickly switch very large amounts of bandwidth in optical networks.

These protocols are carried by IP over the control plane.

##### *3.2.2.1.2.2 Routing*

Aside from the off-line traffic engineering for path creation in MPLS-TE, there are also dynamic routing protocols available such as OSPF-TE and IS-IS-TE. These provide link advertisement and route discovery and distribution. Extensions for GMPLS include fibre identification methods and the consideration of wavelength bandwidth.

One of the unresolved issues in GMPLS dynamic routing is layer inter-operability and the effect that the re-configuration of large amounts of capacity (LSPs) may have on adjacent layer link metrics.

These protocols are carried by IP over the control plane.

#### *3.2.2.1.2.3 Link Management*

GMPLS uses LMP (the Link Management Protocol) for OAM (operations, administration and maintenance), link verification (sending keep-alive messages and PING/Hello style packets), and fault isolation.

LMP uses both the control plane (to share link information such as physical interface connectivity) and the data plane (to perform keep-alive messages and link testing).

#### *3.2.2.1.3 GMPLS functionality*

In GMPLS a number of MPLS concepts remain, such as LSP nesting and link bundling.

#### *3.2.2.1.3.1 LSP nesting*

In LSP nesting one or more LSPs can be encapsulated with a new label and therefore multiple clients (with different labels) can be multiplexed into an encapsulating LSP to create LSP hierarchies. This allows for the protection and restoration of entire groups of LSPs and can provide a transparent route through transit networks for multiple incoming LSPs. In MPLambdaS this LSP nesting is implicit since a wavelength (equivalent to a label) must always by encapsulated by a fibre (another label).

Note that LSPs must terminate in the same type of network, so a PSC (Packet Switch Capable) LSP cannot terminate at the edge of the optical network (since the labels are meaningless unless the terminating equipment is capable of understanding the network type).

#### *3.2.2.1.3.2 Link and LSP bundling*

To provide more scalable routing and capacity allocation GMPLS supports link and LSP bundling. While LSP bundling was possible in MPLS, GMPLS adds the ability to bundle multiple physical optical links (negotiated over LMP) between physically adjacent nodes to be presented as a single link. This can decrease the LSA (Link State Advertisement) size if an IGP (Interior Gateway Protocol) is used and also allows more flexible capacity allocation and a finer capacity granularity.

#### *3.2.2.1.4 GMPLS Deployment*

The deployment of GMPLS can follow two models: GMPLS Peer Deployment Model and GMPLS Overlay Deployment Model.

GMPLS Peer Deployment Model is used where optical and existing electrical LSR domains are isolated. This model would suit the situation where the ISP connectivity (electrical LSR based) provider and inter-AS connectivity (point-to-point optical) provider are different entities or isolated for some administrative purpose.

GMPLS Overlay Deployment Model is used where there is a single control plane for the optical and electrical LSR domains. This can lead to better network resource usage (a single control plane would mean more complete knowledge of the network architecture and better routing of LSPs) at the expense of more complicated business and technical interactions between physical connectivity providers. It can however be shown [SALSA] that for a time-varying traffic demand that is known *a priori,* a unified optimisation of both (G)MPLS and DWDM networks can perform nearly as well as a dynamic adaptable configuration of separate layers.

### *3.2.2.2 ASON*

Automatically Switched Optical Networks are a set of control plane components that are used to manipulate transport network resources in order to provide the functionality of setting up, maintaining and releasing end-to-end connections. The recommendations cover the control of not only purely optical networks such as OTN (Optical Transport Network, ITU-T Rec. G.872) but also legacy SDH (ITU-T Rec. G.803) networks.

ASON's main purpose is to facilitate the fast and efficient configuration of connections within a transport layer network to support both switched (by user request, or in this case the IP Network Provider) and soft permanent (by a management request, for network management purposes) connections. ASON also allows for the reconfiguration or modification of existing connections. There is also support for link monitoring, topology discovery and propagation, connection restoration (re-routing of failed connections over spare resources) and protection (pre-calculated and pre-allocated standby capacity). ASON also supports the aggregation of links (and not only between physically adjacent nodes) to provide greater control over the granularity of connections.

The architecture of ASON is such that the transport and control plane network is divided into domains, which are divided into routing areas. The division into domains allows for the creation of administrative domains for reasons such as geography or domains of different types of equipment. The routing areas, which can be further divided into routing sub-areas, are created for routing scalability reasons. These areas can be designed to create a hierarchy of routing areas where the internal configuration of which is transparent to outside the domain or routing area.

The current ASON ITU-T draft recommendation only specifies reference points in the interfacing of domains, routing areas and other networks, as well as a functional architecture for the control components required to support the ASON architecture. The recommendation states that the control plane must be control protocol neutral and as such does not provide control plane protocol specifics.

## 3.2.3 Related Technologies and Protocols

While not directly related to a specific optical control plane, a number of technologies are often linked to next generation optical networks. GFP (Generic Framing Procedure) is used for framing, LCAS (Link Capacity Adjustment Scheme) for bandwidth negotiation and VCat (Virtual Concatenation) for the aggregation of capacity channels [GLV]. They could be used alongside existing optical network architectures such as OTN (ITU-T G.709) to provide additional functionality that may be required to interface to an inter-domain QoS delivery solution like MESCAL.

### *3.2.3.1 GFP*

GFP, the Generic Framing Procedure [GLV], is a lightweight encapsulation framework for the encapsulation of packet switched and TDM based services. It provides low overhead (small header) and low frame delimiter complexity and is used to wrap services to be transported through electrical or optical networks. It supports client multiplexing so that multiple client layers can be wrapped and transported over the same transport network.

### *3.2.3.2 LCAS*

LCAS, the Link Capacity Adjustment Scheme, is a two-way handshaking and signalling protocol [GLV] and provides for the negotiation of increasing capacity in channels while leaving existing traffic undisturbed. LCAS is being proposed in ITU-T draft G.7042/Y.1305.

### *3.2.3.3 VCat*

VCat, Virtual Concatenation, is a scheme [GLV] for the concatenation of capacity channels such as those provided by SDH or OTN. This concatenation is transparent to the entire transport network except for the end nodes and therefore the channels can be routed diversely and can have their own protection schemes. Its use with SDH is described in ITU-T draft G.707/Y.1332. It is usually used to provide better link utilisation through the matching of LAN speeds (such as 100Mbit/s Ethernet) with SDH capacity (matching to two virtually concatenated VC3 (~100Mbit/s)).

In the absence of a GMPLS or ASON system the above three technologies used together could form the basis for a lightweight abstraction layer to optical networks with the ability to negotiate and re-negotiate connections of any capacity or granularity. Alongside GMPLS and ASON they can be a complete solution that will provide flexible bandwidth over legacy networks with GMPLS providing forwarding and ASON providing connection management and interfacing to the customer.

## 3.2.4 Optical Technologies and Equipment

While these control plane technologies are being proposed the technology does not necessarily exist to support pure optical switching at high speeds and as such service provisioning and capabilities are restricted more by available equipment than network architecture. A general view of current optical technologies can be seen in Figure 17. The diagram shows the number of available ports versus the speed at which they can be switched for various OOO (Optical-Optical-Optical, i.e. pure optical) and OEO (Optical-Electrical-Optical, i.e. optical with a conversion to the electrical domain) technologies. OEO technologies can reconfigure fastest (yellow boxes) but still would not be fast enough to switch wavelengths on a per packet basis. OOO technologies (darker grey boxes) such as MEMS (Micro-optical Electro-Mechanical Systems) are slower to switch but would be fast enough to reconfigure for a new GMLambdaS configuration (since it is simply a redirected wavelength and not a per packet forwarding decision). It can therefore be seen that the technology currently exists to support the time-scales of the demands a MESCAL system may make, but not future per-packet optical switching.

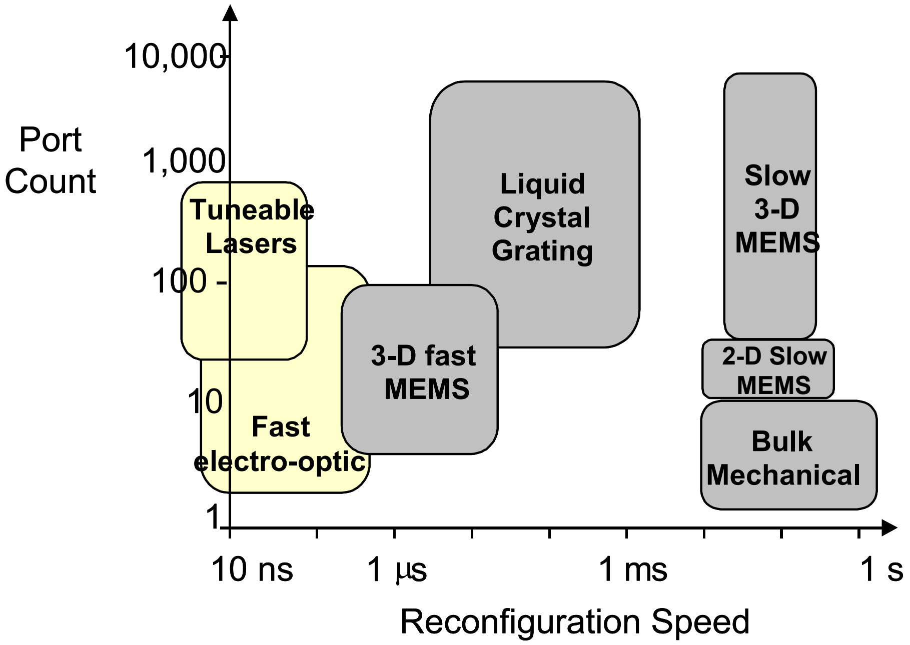


**Figure 17: Current optical technologies, comparing their reconfiguration speeds to the current available port count.**

While the standards, architectures and protocols described above are still in draft form, products are already available which MESCAL Solution Options could interface to. The equipment ranges in size/capacity, capability and technology, from long-haul all optical switches to metro-edge switches and routers.

### *3.2.4.1 Ciena CoreDirector*

The Ciena CoreDirector Intelligent Optical Core Switch [CIENA], already available and currently deployed in test optical networks is capable of switching up to 640 Gbps and supports up to 64 ports of 10 Gigabit of SDH/SONET switching and is capable of supporting ASON or GMPLS control planes. It can support packet (Ethernet, POS), optical and TDM (SONET/SDH) switching. The optical switching technology is OEO using fast tuneable lasers. Provisioning times of approximately 30 minutes are claimed.

#### *3.2.4.2 Lucent LambdaXtreme*

For long haul applications Lucent have the LambdaXtreme Transport device [LCNT] which is a purely optical (OOO) OADM (Optical Add-Drop-Multiplexer). Lucent's "Navis" management software, which controls the device, can be interfaced to GMPLS and ASON networks. The device supports up to 64 wavelengths over 1000km at 40Gb/s per wavelength or 128 wavelengths at 10 Gbps per wavelength over 4000km.

#### *3.2.4.3 Cisco ONS 15454*

For metro edge applications the Cisco ONS 15454 [CISCO], under the UCP (Unified Control Plane) implementation supports GMPLS as well as associated protocols such as RSVP-TE and LMP. UCP also supports communication with ASON networks.

## 3.3 MESCAL and Optical Networks

If the technology exists to provide the services of a physical connectivity provider, we must also consider how the MESCAL Solution Options may take advantage of their features. The business model in MESCAL makes a separation between the business entities controlling the IP network and the physical connectivity, which is similar to the way current optical networks operate. Since optical networks currently carry traffic for a wide range of networking technologies it is common for IP, SDH and DWDM networks to be operated independently. As they are operated independently the different network layers are not usually planned and optimised by a single process either; rather these networks are independently planned and then provide a demand matrix to the network planner of the bearer network. It is also possible a single IP network provider may use multiple physical connectivity providers (which may not all be optical networks) and this therefore further separates IP and optical networks. For this reason it is believed that interoperability and integration with optical control networks is beyond the scope of MESCAL. However, of the three different solution options, the third solution option could be interfaced more closely to GMPLS enabled optical networks than current IP networks.

### 3.3.1 MESCAL Solution Options 1 and 2

In MESCAL Solution Options 1 and 2 all network intelligence remains in the IP network and therefore after planning the IP layer the demands are sent to one or more optical network operators (as physical connectivity providers). Interaction between IP and optical networks will be for the creation or reconfiguration of connectivity during the resource provisioning cycle at a management or business agreement level. The interaction can be with any type of optical control plane, either ASON or GMPLS.

### 3.3.2 MESCAL Solution Option 3

The use of MPLS in MESCAL Solution Option 3 allows for the partial integration between the MESCAL architecture and an underlying optical network controlled by GMPLS. By interfacing the GMPLS control planes between IP network providers and the optical network providers a more optimal route for LSPs can be computed. The IP network PCSs can then request light-paths through optical networks or, depending on required bandwidth, negotiate the stacking of labels and therefore the multiplexing of multiple LSPs over existing light-paths, allowing for fine control of capacity.

## 3.4 Conclusions

There are a number of optical control plane technologies currently available and in development. The design of these technologies does not impose any immediate limitations at the time scales in question and provides control interfaces to current and future optical networks. The only possible interoperability issue is that of capacity granularity and the ability to multiplex clients and sub-divide the bandwidth of each wavelength. In the electrical domain technologies such as GFP, Ethernet or existing SDH technology do however provide this multiplexing and therefore can alleviate this problem, but may not be scalable to future line speeds. Used together with link/LSP bundling in GMPLS or link aggregation in ASON, or VCat it would be possible to efficiently allocate fine-grained capacity up to very high speeds (multiple wavelengths). A further important feature common to GMPLS, ASON and LCAS is that existing connections don't have to be torn-down before increasing their capacity, a feature that would be useful for transparent provisioning cycles.

The organisational separation of IP and optical networks would mean that there is no direct link between MESCAL Solution Options 1 and 2 and DWDM networks, and therefore any of the technologies listed here would be suitable for the dynamic provisioning of bandwidth. MESCAL Solution Option 3's use of MPLS however would allow for a closer integration as the IP network providers PCSs could now directly interface to the optical network's PCSs for faster more efficient provisioning.

# CHAPTER 4: BI-DIRECTIONALITY OF SERVICES

## 4.1 Introduction

EURESCOM P1008 Project [P1008-01] considered all streams as uni-directional streams but its P1103 Project [P1103-02] acknowledged that bi-directional traffic handling and the general requirements of providing broadcast and multicast services with assured QoS should be considered by the project. There is no indication in [P1103-02] of how to implement bi-directional services at the network layer using the cascaded model.

The MESCAL solutions allow QoS-based IP delivery service between two end-points (sender and receiver) spanning a substantial number of domains, with loose/statistical/hard guarantees. The general requirements of providing bi-directional services with some assured QoS and handling their respective traffic appropriately should be considered by MESCAL. It may be feasible for MESCAL to make it possible for two separate *SLSs* to cover each directional scope of a service.

In the cascaded approach discussed in chapter 1, each Network Provider (NP) or ISP forms *pSLS* contracts with the immediately adjacent interconnected NPs. Thus, the QoS peering agreements are only between BGP peers. This process is repeated recursively to provision the QoS connectivity from a customer to reachable destinations that may be several domains away. Figure 18 shows an example for end-to-end uni-directional QoS service implementation using the cascaded approach. Each NP/ISP administers its own domain and the inter-connection links that it is responsible for. For example in Figure 18, ISP1 is responsible for the network provisioning and resource allocation in AS1 including the configuration of both *"**a**"* and *"**b**"* interfaces. Based on the forecasts and/or c/pSLS subscription requests, ISPs provision their network and allocate their resources for offering the QoS-based service.

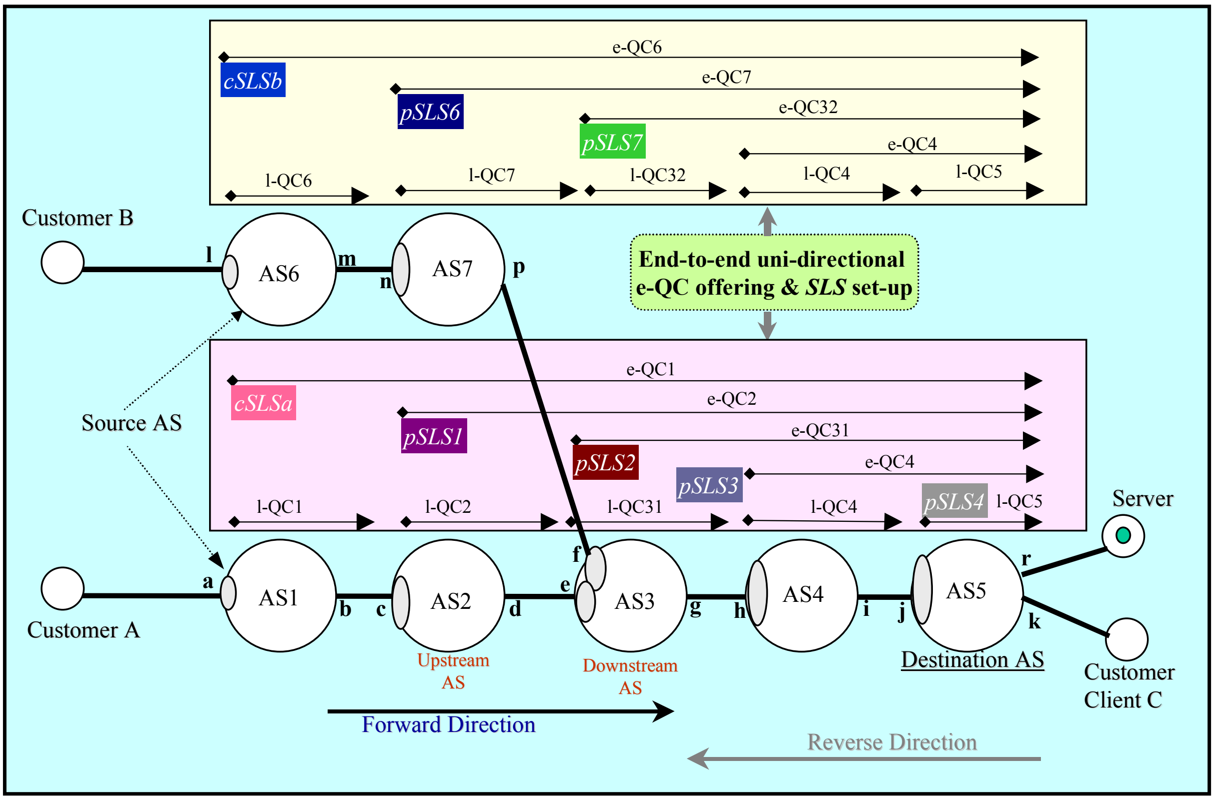


**Figure 18: End-to-end uni-directional QoS service implementation.**

In this chapter, we will consider the bi-directionality using the cascaded approach, starting from solution option 2 proposed in D1.1 where most problems arise, because of the end-to-end e-QC ideas applied in this option. We investigate two methods for supporting bi-directional services. We also discuss bi-directionality for MESCAL solution options 1 and 3 proposed in D1.1 although these are less exposed to the problems, through the lack an e-QC concept.

## 4.2 Bi-directionality in Statistical Guarantees Solution Option (2)

There are some fundamental problems to be solved in order to provide bi-directional services with solution option 2. This section identifies these problems, presents a detailed discussion on the resulting implications and provides two methods of providing bi-directionality in solution option 2.

### 4.2.1 Problem Space: The scope and QC for the reverse path

In the cascaded model, the scope (the source and the reachable destinations) of the desired e-QCs for the forward direction is part of the c/pSLS during the negotiation phase. Taking an example in Figure 18, during *cSLSa* negotiation phase the tuple (Source Customer, Destination Customer, e-QC) is known as (A, C, e-QC1). However, when the reverse direction is considered from the destination AS's point of view there are the following apparent problems in constructing the bi-directional services using the cascaded approach:

**I**) From the destination AS's point of view (AS5 in Figure 18), the destination for the traffic in the reverse direction (sources of traffic in forward direction) are not known. This is due to the fact that the cSLSa is between Customer A and AS1 and AS1 knows the above tuple whereas AS5 is unaware of it. As the destinations for return traffic are unknown, AS5 cannot verify/find whether there are e-QCs formed through QC binding operations to reach the desired destination (e.g., Customer A) or not. Thus, how does AS5 find the scope for the reverse direction (e.g., AS1) in order to see any e-QC is formed and offered to reach AS1 customers?

**II**) Every time another upstream AS forms an e-QC that utilises the l-QC of the destination AS (AS5), the scope of the return paths for AS5 extends and AS5 does not know this.

**III**) Which QoS class (l-QC) at each AS (e.g., AS5) should be used for return traffic?

**IV**) How should this l-QC be mapped to an e-QC offered by the upstream AS (e.g., AS4)?

**V**) Should the customer at the e-QC source (Customer A at AS1) pay for return traffic?

The destination AS (AS5) has no explicit information to answer the above questions.

Regarding problem **II**, changing/extending/amending the scope of e-QCs is a delicate business as an e-QC defines a service with strict properties. It should be noted that *Meta classes* used in MESCAL solution option 1, do not imply a predefined/engineered end-to-end-QoS. Unlike *Meta classes* with general concept that define "the best I can do at this price"*,* e-QCs are defined end-to-end with distinct performance characteristics. This means that if the scope of the e-QC broadens, it can only do so within its end-to-end performance boundaries and constraints. An e-QC defining a service from London to Athens is not likely to be extensible to include a location in Sydney. Whereas, this is possible with *Meta class* concept employed in the MESCAL solution option 1. Therefore, extending the scope of an e-QC is not a straightforward task.

There are two methods to tackle the problem of providing QoS enabled path in reverse direction. The first method extends the single cascade with bi-directional capabilities. The second method employs unidirectional cascades in forward and reverse directions to build bi-directional services.

### 4.2.2 Method 1: Single Cascade with Bi-directional Capabilities

#### *4.2.2.1 Method Outline*

As it is shown in Figure 18, the cascaded approach is used for constructing e-QCs and setting up *pSLSs* for enabling Customer A traffic to reach Customer Client C. One possible solution for setting up a reverse path is to negotiate *pSLSs* in the reverse direction between peer ASs with an open destination scope (*). An open scope is necessary when considering that as the e-QC is sold on, it can become part of a new e-QC, the scope and QoS parameters of which cannot be known by the Destination AS (i.e., AS5 in Figure 18). To allow the upstream AS (e.g., AS4) to offer the e-QC to further upstream ASs (e.g., AS3) without the need for amending the scope of pre-existing downstream pSLSs every time the scope changes, the (*) is required. This potentially solves the bi-directionality problem at the pSLS level, but it raises some issues in implementing the e-QCs and invoking the service:

1. As the Destination AS is not aware of the forward path e-QC used by the Source AS, which l-QC/e-QC should it use for the reverse path (problem III)? The problem is yet more apparent when the QoS path is asymmetric and the forward and reverse paths are not supposed to use the same QC. Unless explicitly signalled, the destination AS cannot deduce information to chose a reverse path QC.
2. How is the admission control applied for reverse traffic so that no false/unwanted (unpaid-for) traffic is injected?
3. There is also a return path implication on the functional model. BGP updates are required to take place in the direction towards the Destination AS to enable return path reachability (*pSLS* ordering in the upstream AS needs to configure its local qBGP). However, this has the scope problem (problem I): each time a service is sold onto an upstream peer, it requires BGP updates for the new return path reachability to be propagated down through the cascade. This is exactly the sort of scalability problems that the cascade model is supposed to prevent.

Both issues 1 and 2 call for the use of a signalling mechanism for communicating between the Source and Destination ASs in order to inform the Destination AS the sink for return traffic (e.g., Customer A). While this does not seem to be in line with the cascade approach, which implies relationships only between cascade neighbours, it also raises the need for the Source AS to specify:

- Who the destination AS is in order to identify for the destination AS the scope of return traffic?
- What is the desired QC level for return traffic?
- How to contact and build trust relationships between the Source and Destination ASs? This is problematic, because there is no direct business relationship between the two ASs.

The identified technical problems will be further investigated in the next section, however, some fundamental problems of the reverse path through a single cascade have been identified. Back propagation of BGP and possibly SLS information is required to create a reverse path, thereby adding considerable complexity to the solution. The end-to-end signalling requirement violates a basic concept of the cascade model, where direct business relationships only exist between peers.

#### *4.2.2.2 An example implementation - Single cascade by using e-QC enabled c/pSLSs in forward direction & l-QC enabled pSLSs with no explicit e-QC binding in reverse direction*

This is to implement bi-directional QoS-enabled services by employing e-QC enabled *c/pSLSs* in forward direction and l-QC enabled *pSLSs* with no explicit e-QC binding in reverse direction. While each ISP provides the QoS environment by provisioning its network and allocating resources in the forward direction, it can provide a similar environment for return traffic in the reverse direction within its own domain.

This implementation is based on the following facts and concepts:

- The inter-domain routing is *pSLS* constrained. By *pSLS* constrained, we mean that traffic (in either direction) will only pass through the ASs where there are *pSLS* peering agreements already in place. If there is no *pSLS* agreement there is no way for an AS to transport the respected QoS traffic. Generally, the peering agreement between two neighbouring domains should serve traffic in both directions. Even the availability of best-effort route as the minimum requirement for the return traffic may be considered through normal peering.
- As each ISP can provide the foundation for handling traffic in both directions through *pSLS* agreement/s, it is viewed that the inter-domain route between given ASs are set-up in a symmetric way but the intra-domain routes can be asymmetric.
- There will be only a single *pSLS* in place for similar service (with a given performance target) provided by an AS to its upstream domains in order to get to specific destinations irrespective of the sources of traffic.
- While there is a *pSLS* negotiation for forward direction between two neighbouring domains, there can also be *$p_r$SLS* negotiation for return path between the same two domains at the same time. A service is established if both *SLSs* are agreed. The performance requirements for these two types of *SLSs* within the domain are the same but the end-to-end scopes are different.
- As an alternative option, there could be no *$p_r$SLS* negotiation for the return path between neighbouring ASs. The requirements for the reverse direction can be part of the *pSLS* negotiation for the forward path.
- *pSLSs* are uni-directional. *pSLSs* are established for transporting traffic in forward direction whereas *$p_r$SLS* are established for transporting traffic in reverse direction. The scopes for handling QoS of these two *pSLSs* are different. The first one is normally targeted for e-QCs and the second one is targeted for l-QCs within the domain.
- While each ISP (AS) binds its own l-QCs with the e-QCs offered by the downstream domains and offer new e-QCs to upstream domains, they are also able to offer the same or similar l-QC (as in forward direction) for reverse direction within its domain.
- Each AS is only concerned with l-QCs for the reverse direction and does not consider/construct e-QCs for the return path. If each AS configures the reserve direction performance target the same as to the forward direction, the mathematical operation of l-QCs performance values results in similar values for the end-to-end path in the reverse direction to the e-QCs in forward direction. Although there is no e-QC set-up in the reverse direction, the return traffic receives similar treatment to traffic in forward direction.
- Obviously, traffic in reverse direction can have different bandwidth requirements, which may be specified as part of *$p_r$SLS* negotiation. The QoS performance requirements (delay, loss, jitter) in both directions are the same (or very close) as ISPs can normally offer l-QCs between any given ingress-egress pair within their domain.
- While admission control and Traffic Conditioning parameters are set and configured for the forward direction in the domain, it is possible to set them for reverse direction appropriately (see next section).

As shown in Figure 19, AS5 provisions and configures its networks to provide l-QC5 service to upstream domains. It also provisions and configures its networks to provide *l-QC5* service for return traffic generated by Customer C. The performance characteristic of l-QC5 and *l-QC5* are the same. It advertises both to the upstream domains (e.g., AS4). This procedure occurs recursively in terms of e-QCs for forward direction traffic handling and *l-QCs* for return traffic handling. There is no e-QC binding operation for *l-QCs*. The service advertised by AS2 to AS1 is o-QC2 for e-QC2. AS1 provisions and configures its networks to provide e-QC1 service to Customer A. It also provisions and configures its networks to provide *l-QC1* service for return traffic originated by Customer C towards Customer A. The forward (and may be the return) capabilities are advertised by AS1 to its customers.

Customer A is now able to negotiate and establish *cSLSa* with AS1 to send QoS traffic to customer C in forward direction. The reserve direction provides the best possible service for symmetric return traffic but the performance targets for return traffic may not be quantified precisely. Table 4 shows the scope of c/pSLSs and interfaces that are involved for both directions for the cases shown in Figure 19.

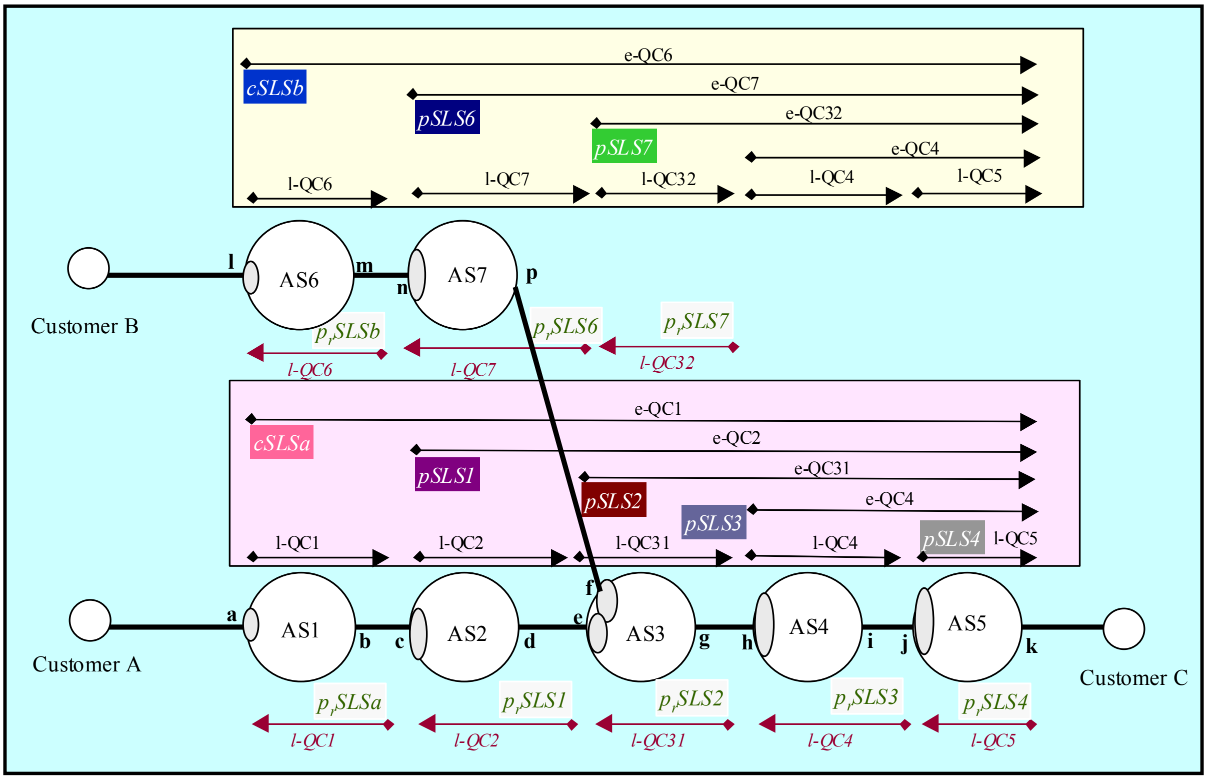


**Figure 19: ISP peering for bi-directional services using method 3.**

| ***Forward Direction*** | | | | ***Return Direction*** | | | |
|---|---|---|---|---|---|---|---|
| ***c/pSLS*** | **Scope (interface)** | **o-QC** | **Interfaces to configure (inclusive)** | ***c/pSLS*** | **Scope (interface)** | **o-QC** | **Interfaces to configure (inclusive)** |
| *cSLSa* | AS1-AS5 (**a** to **k**) | e-QC1 | **a** to **b** | $p_r$*SLSa* | AS1 (**b** to **a**) | *l-QC1* | **b** to **a** |
| *pSLS1* | AS2-AS5 (**c** to **k**) | e-QC2 | **c** to **d** | $p_r$*SLS1* | AS2 (**d** to **c**) | *l-QC2* | **d** to **c** |
| *pSLS2* | AS3-AS5 (**e** to **k**) | e-QC31 | **e** to **g** | $p_r$*SLS2* | AS3 (**g** to **e**) | *l-QC31* | **g** to **e** |
| *pSLS3* | AS4-AS5 (**h** to **k**) | e-QC4 | **h** to **i** | $p_r$*SLS3* | AS4 (**i** to **h**) | *l-QC4* | **i** to **h** |
| *pSLS4* | AS5 (**j** to **k**) | l-QC5 | **j** to **k** | $p_r$*SLS4* | AS5 (**k** to **j**) | *l-QC5* | **k** to **j** |
| *cSLSb* | AS6-AS5 (**l** to **k**) | e-QC6 | **l** to **m** | $p_r$*SLSb* | AS6 (**m** to **l**) | *l-QC6* | **m** to **l** |
| *pSLS6* | AS7-AS5 (**n** to **k**) | e-QC7 | **n** to **p** | $p_r$*SLS6* | AS7 (**p** to **n**) | *l-QC7* | **p** to **n** |
| *pSLS7* | AS3-AS5 (**f** to **k**) | e-QC32 | **f** to **g** | $p_r$*SLS7* | AS3 (**g** to **f**) | *l-QC32* | **g** to **f** |

**Table 4: The c/pSLS agreements and their scopes in both directions.**

The admission control and traffic conditioning parameters are set and configured for the forward direction in the domain. Traffic classification will be based on the DSCPs for forward direction. The scope of *cSLS1* is from ***a*** to ***k*** and customer A should assign DSCP value to represent l-QC1. The assignment of l-QCs (DSCP values) is carried out at the BR routers (e.g., ingress interfaces) for packets in the forward direction according to the procedures specified in D1.1 for different solution options. An example procedure is shown in Table 1.

| Packet Marking | | | |
|---|---|---|---|
| ***AS*** | ***Interface*** | ***Current QC*** | ***Mark to*** |
| AS1 | a | l-QC1 | l-QC1 |
| AS2 | c | l-QC1 | l-QC2 |
| AS3 | e | l-QC2 | l-QC31 |
| AS3 | f | l-QC7 | l-QC32 |
| AS4 | h | l-QC31, l-QC32 | l-QC4 |
| AS5 | j | l-QC4 | l-QC5 |
| AS6 | l | l-QC6 | l-QC6 |
| AS7 | n | l-QC6 | l-QC7 |

**Table 5 : Packet marking at ingress interfaces of border routers.**

When *pSLS* and *$p_r$SLS* are agreed, the admission control and Traffic Conditioning parameters must also be to set and configured for the reverse direction at each domain. It is possible to set them for reverse direction appropriately. For the reverse direction and at the boundary routers (**k**, **i**, **g**, **d**, **b**, **p**, and **m** interfaces in Figure 19), three pieces of information are required for traffic conditioning. The scope of *pSLS* specifies the destination prefix of forward traffic. This is going to be the source of return traffic. As long as the boundary router knows the traffic source prefix and the next AS hop in the reverse direction (which is the AS who requested the *pSLS*) in addition to the packet's DSCP, it is able to correctly map the packet's DSCP to the new *l-QCs* that is to be used within the domain. Once the packet is marked correctly, it will take the appropriate intra-domain route towards the specified interface to the next AS hop. This procedure is repeated until return traffic arrives at its appropriate end-point.

It is possible that two different streams of return traffic originated from a destination to a source may require use of the same l-QC in one of the transit domains. This creates a splitting problem at the egress point of the domain. In order to avoid the splitting problem, v-QCs can be used at the ingress point of that domain to differentiate the two streams at the egress point of the domain and assign the correct DSCP. This implies that a different v-QC is needed for each *$p_r$SLS,* otherwise more state information is required for inspecting and classifying packets that may create major scalability problem.

This implementation has the following pros and cons:

- It follows the cascaded approach and its scalability follows the cascaded approach.
- It offers bi-directional services at the network layer, with symmetric performance values. For asymmetric performance targets, the other explained and complex methods can be used.
- This method not only provides e-QCs in the forward direction but also to the same effect, it does the same in reverse direction.
- The issue on the scope of return traffic is solved as *$p_r$SLS* extend hop-by-hop rather than end-to-end and it resolves the QC operation at every hop through the information available in *pSLSs* and *$p_r$SLSs*.
- There is no issue related to change/extend/amend the scope of e-QCs for return traffic, every time another upstream AS forms an e-QC that utilises the l-QC as the scope of the return path stays within the domain.
- Similar l-QCs used in forward direction are used in the reverse direction for return traffic.

- The charging/billing is only done between the two neighbouring ISPs who peer and the customer at the end of e-QC pays for the service (forward and return traffic).
- It is as simple as the cascaded approach to implement and it is deployable with no extra burden.
- This method is based on the ***symmetric*** path for both directions.
- To avoid splitting problem, more state information is required for packet classification that raises the ***scalability*** problem.

## 4.2.3 Method 2: Multiple Uni-directional Cascades

### *4.2.3.1 Method Outline*

This method allows the establishment of uni-directional *SLSs* for sending traffic only. The bi-directionality of services is left to be initiated by the application layer. The suitable e-QCs have to be set-up separately by the source and destination ASs. There is no guarantee that a suitable e-QC for the return path will exist for any given forward e-QC, except by virtue of a "customer God". The "customer God" ensures that suitable reverse path e-QCs exist in the destination AS, based on application requirements. This method would potentially provide the environment for having bi-directional services using the cascaded approach in both directions.

We assume the topology shown in Figure 18 where Customer A (client) is connected to a server through 5 interconnected ASs, "***Client-AS1-AS2-AS3-AS4-AS5-Server***". For the client to send traffic to the server with a given QoS (e.g. upload a file using e-QC1) the client needs to establish *cSLSa* with the AS1 for this QoS. AS1 has *pSLS1* that supports forwarding of this QoS traffic with the AS2, AS2 in the same way has *pSLS2* for forwarding of this QoS traffic with the AS3, AS3 has *pSLS3* with AS4 and AS4 has *pSLS4* with AS5. So the traffic reaches the server with the desired quality. The correct DSCP, the conditioning and the billing of the client's traffic are based on the *cSLSa* established with AS1. Because there is no return QoS path established for any necessary return traffic such as ACKs, etc the best-effort route is used. This is in line with the MESCAL assumption that whenever a QoS route to destinations is not available, the best effort route may be used as an alternative [D1.1 Section 3.1]. Having dissimilar QoS classes in the forward and reverse directions may cause problems with flow control protocols such as TCP. Any reverse path with a lower quality than its corresponding forward path could cause TCP to degrade the forward path quality. This is not a desirable effect and measures should be taken to avoid the use of flow control enabled protocols on asymmetric paths.

In the case where the client wants to receive traffic from the server with a given QoS (i.e., to download a file), the client must contact the server at the application layer with a request to send traffic to the client. The QoS requirements of the sending traffic as well as the billing details are also agreed between the two. The client may contact the server through a normal best-effort route and the client-server contact at the application layer may require a modification to the application. Alternatively, the client might use another application to arrange for setting up the client-server connection and then pass it onto the client application once the connection is set-up. The application layer communication between customers or client/server will need a way to describe and agree on the QoS levels to be used in each direction. This could be done by exchanging details of the specific e-QCs they have subscribed to in their respective *cSLSs*, or it could be done at a more abstract level in a customer language without exposing exactly how this is mapped to the e-QCs/*cSLSs*/QoS parameters they have with their respective ISPs. Thus, there is a need to know on how the customers agree on the QoS levels they will use.

If the necessary uni-directional *pSLSs* (from server to client) are already in place, the *cSLS* can be established. Otherwise, the appropriate actions need to be taken to establish *pSLSs* dynamically through the cascaded approach for fulfilling this request. Upon the establishment of *cSLS* agreement, the server can send traffic, which is forwarded to the client, through established appropriate *pSLSs* between AS5 and AS4, AS4 and AS3, AS3 and AS2 and between AS2 and AS1. In the case that such a *cSLS* cannot be established the server rejects the clients request. The correct admission control parameters and traffic conditioning of the server's outgoing traffic is based on the *cSLS* established between the server and AS5. The billing of the client for the received traffic is based on his agreement with the server.

The next section explains how to implement uni-directional SLS-based QoS paths in order to achieve bi-directionality.

#### *4.2.3.2 Building multiple cascades to implement bi-directional services*

Figure 20 shows the cascaded implementation for forward direction in which the sources from AS1 (e.g., Customer A), AS2, AS3, AS4 and AS5 can reach Customer C in AS5 with the specified desired e-QC quality. Figure 20 shows the establishment of cascade for reverse direction. However, there are a few implications/issues though, arising from implementation.

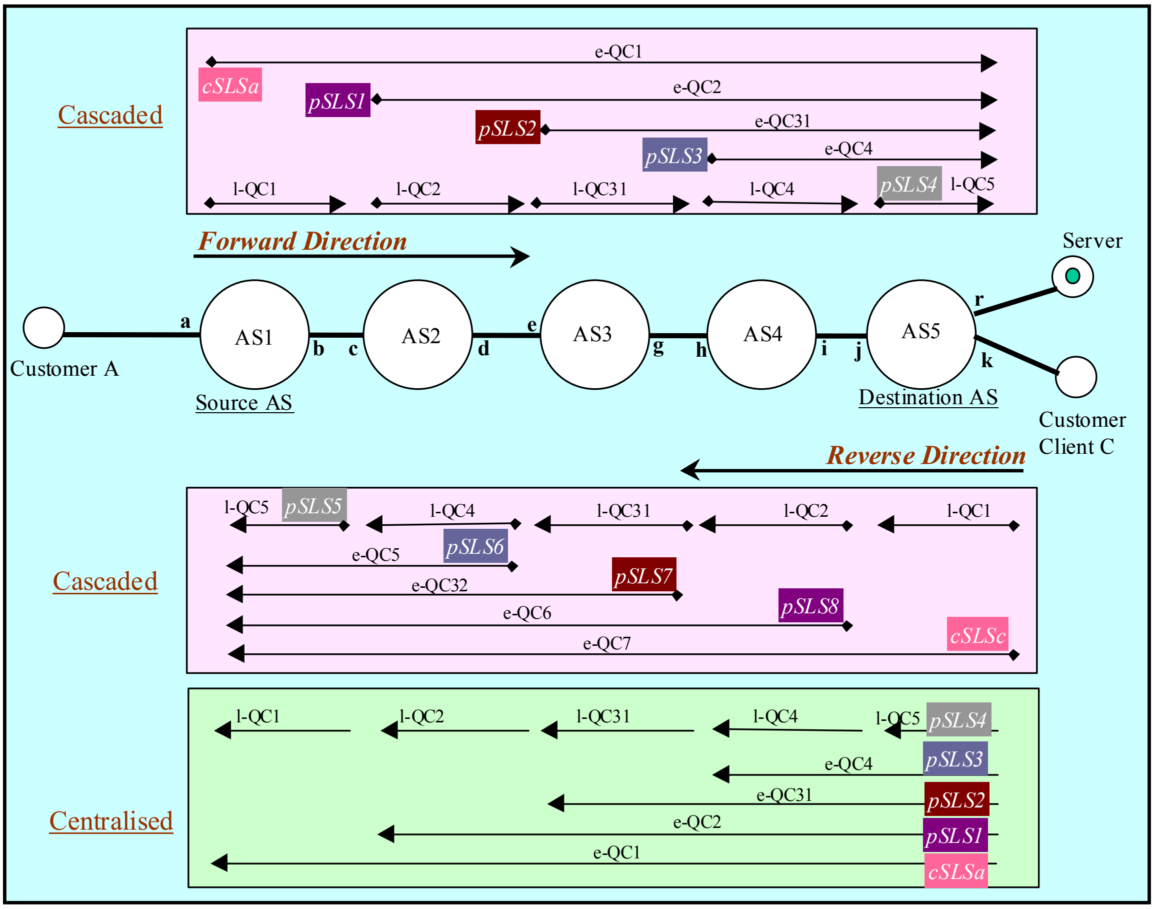


**Figure 20: Cascaded e-QC/SLS set-up in both directions.**

1. The most obvious problem is the requirement to have a suitable reverse path e-QC. Since the existence of this path depends on the willingness of the Destination AS to set it up, its existence and availability can be determined either during the invocation phase of the service or via some other pre-set-up means of communication between the involved ASs (as explained in method 1), customers or third party services such as a VoIP gatekeeper. This implies that there is a negotiation between Customers/ASs/third parties before invocation has taken place in order to agree the forward and reverse QoS levels. These are then mapped to e-QCs at each end according to the inter-domain capabilities (and pre-existing *cSLSs*) at each Source and Destination ASs.

2. From the service set-up point of view, it is now imperative for communication to take place between the two involved ASs (or Customers) in order to do accounting and billing. Presumably only one communicating party should be charged for the whole bi-directional service. This should not be a problem for most types of service. For example, most

asymmetric services like VoD require a business relationship between the client and the server anyway, which will now simply include extra charge for the return path QoS-enabled traffic. For services such as VoIP it is more tricky to see how to mirror the business model of the current telephone network where you do not usually pay to receive calls. Perhaps, one solution is to assume that there is a 3rd party VoIP service provider who has business relationships with the ASs at each end: a customer signals to the VoIP SP to initiate a call; the VoIP SP invokes cSLSs in each direction on behalf of the calling and called parties; the VoIP SP pays the ASs at each end of the call; the calling party pays the VoIP SP for the call.

3. From the network point of view, consider the case of a new ISP joining the MESCAL Internet who is willing to offer QoS-enabled services. This ISP is required to purchase some e-QCs from its peering ISPs in order to send traffic. However, these peering ISPs need to sell the QCs offered by the new ISP to the MESCAL Internet community (potentially a very large number of ASs) before this new ISP can receive any QoS-enabled traffic.

    a. If this new ISP is a small entity, then its peering partners may have little interest in setting up e-QCs based on the new ISPs l-QCs. Even if there were a will to do this, it would likely take some time to propagate through the Internet. There is no guarantee that upstream ASs will bind with an AS's l-QCs/e-QCs, hence customers may not be able to receive any QoS traffic. If the upstream ASs do bind then there is an unknown propagation delay before the chain of bindings take place through to the destination ASs. This could be an issue for new ISPs or existing ones offering new o-QCs: they can immediately offer QoS services to their customers for sending traffic (as soon as they have established *pSLSs* with their peers and received the corresponding qBGP updates) but they do not know when they will be in a position to receive QoS traffic. A partial solution to this problem is to assume that small regional/local ASs "pay" the tier 1 ASs they peer with to bind with their o-QCs and promote the resulting e-QCs to other ASs (tier 1s as well as their customer regional/local ASs). This could speed up the process, as the business model is usually that smaller ISPs are customers of larger ISPs, rather than true peers, and that the net flow of cash is from the local/regional ISPs towards the tier 1 providers.

    b. If the new ISPs service offerings (o-QCs) could somehow be included in the scope of existing *pSLSs* that upstream peering ASs have, the problem could be solved. When an AS (e.g. ASx, a regional ISP) offers new o-QCs, which are then adopted by a tier 1 provider as a constituent of the o-QCs it offers (perhaps just extending/amending the scope of the o-QCs it already offers to include the new destination prefixes in ASx), then the *pSLSs* of all customers of that tier 1 provider need to be modified if they want to sent to ASx's prefixes. Unless the scope of those pSLSs was already destination "*".

Now we explain additional complexity arising from multiple reverse direction cascades.

As it is shown in Figure 20, in the sources from AS1 (e.g., Customer A), AS2, AS3, AS4 and AS5 can reach Customer C in AS5 with the specified desired e-QC quality. For reverse direction, can there be a reverse path e-QC for every forward path e-QC using the cascaded approach in order to allow bi-directional QoS offering? As it is depicted in Figure 20, it may not possible to achieve this using the cascaded approach. This is only possible by constructing e-QCs in reverse direction by using centralised approach. As an example (bottom of Figure 20), e-QC4 and l-QC31 need to be combined to form e-QC31 in reverse direction and this can only be achieved using the centralised approach. Therefore, the source AS is starting point for the cascade in reverse direction and not destination AS.

In general, the multiple cascaded implementation requires to build multiple reverse cascades in reverse direction to allow transporting return traffic.

In the forward path, the e-QC paths merge as they get towards the Destination AS. This provides the opportunity for constructing *pSLSs* downstream based on the aggregated traffic demands. However, the paths in reverse direction de-merge as they depart from a Destination AS towards Source ASs. Consequently, for a single Destination AS to provide return service to its upstream domains, multiple cascades of e-QCs and *pSLSs* set-ups are required. None of these e-QCs and *pSLSs* can be combined and merged. Thus, for a single forward direction cascaded, there must be multiple cascades in reverse direction, depending on the number of Source ASs that are served by a Destination AS (Figure 21). Figure 21 shows the implementation of multiple cascades in which not only they serve to reach from customer C to sources in AS5 (with l-QC1 quality), AS4 (with e-QC45), AS3 (with e-QC35), AS2 (with e-QC25), and AS1 (with e-QC15) but also sources with desired quality from AS4 to AS3 (with e-QC34), AS4 to AS2 (with e-QC24), AS4 to AS1 (with e-QC14), AS3 to AS2 (with e-QC23), AS3 to AS1 (with e-QC13), and AS2 to AS1 (with e-QC12).

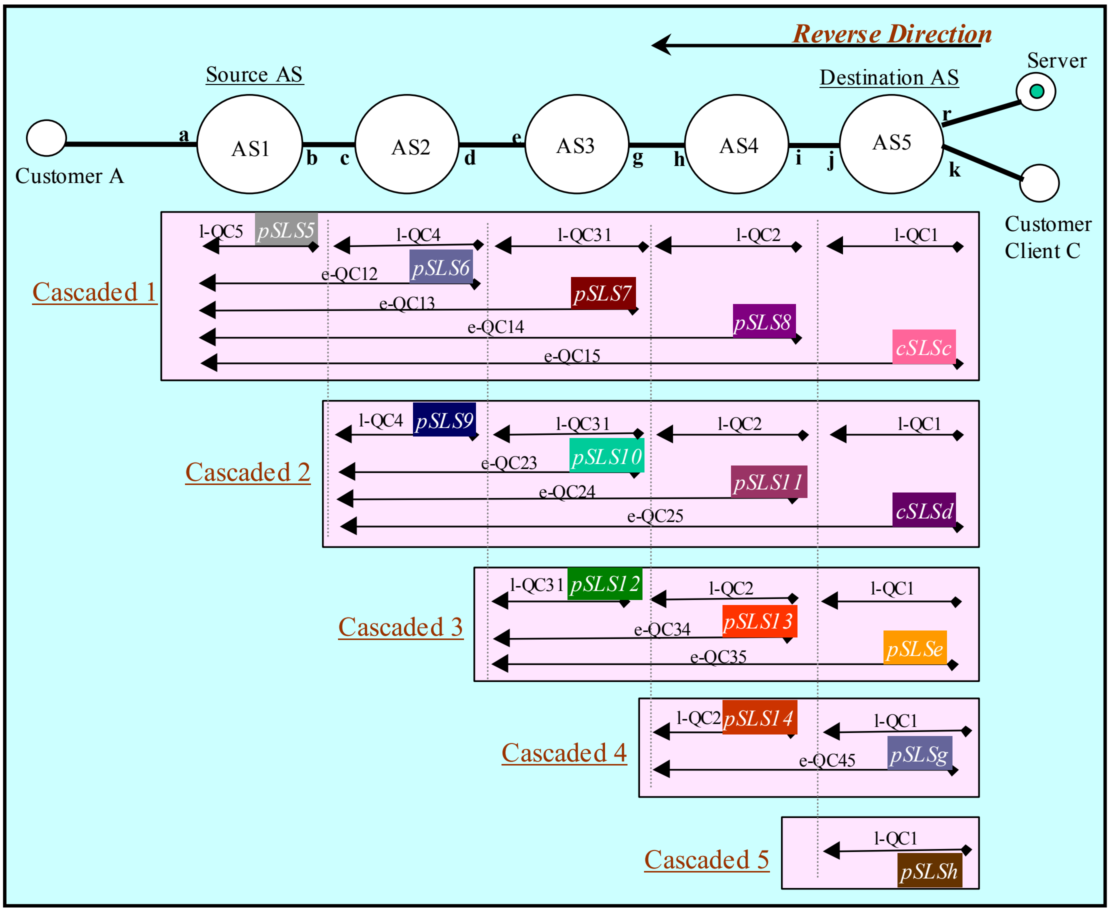


**Figure 21: Implementation of multiple cascades.**

## 4.3 Bi-directionality in Loose Guarantees Solution Option (1)

In this solution option, an AS advertises the Meta-QoS-classes that it supports within its administrative domain. Other domains can make *pSLS* arrangement with this domain to make use of offered Meta-QoS-classes. Although each domain can find out whether it can reach certain destinations in a Meta-QoS-class plane through qBGP updates it receives, there is no need for strict cascade approach to build e-QCs as *pSLSs* are established with open destination scope (*). There is also no end-to-end QoS guarantee defined/agreed, which is the fundamental difference between the loose guarantee solution option and statistical guarantee solution option.

Thus, in order to provide bi-directionality in solution option 1, the following procedures can be carried out:

- *pSLSs* agreed between two domains are not tied with certain destinations as in solution option 2. Hence, as *pSLSs* are uni-directional and they are established for transporting traffic in forward direction, *$p_r$SLS* can be established for transporting traffic in reverse direction. The boundary for handling QoS of these two *pSLSs* are the same i.e., Meta-QoS-classes within the domain.

- As stated above, while there is a *pSLS* negotiation for forward direction between two neighbouring domains, there can also be *$p_r$SLS* negotiation for return path between the same two domains at the same time. The performance targets for these two types of *SLSs* within the domain are the same, i.e., Meta-QoS-class x. As an alternative option, there could be no *$p_r$SLS* negotiation for return path between neighbouring ASs. The requirement direction can be part of the *pSLS* negotiation for the forward path.
- The reverse direction *$p_r$SLS* can have different bandwidth requirement, which may be specified as part of *$p_r$SLS* negotiation.
- The path for forward traffic and return traffic may be different depending on the q-BGP updates but the *SLS* agreements between all involved ASs are in place to handle the traffic in both directions irrespective of the paths traffic may take in forward and reverse directions.

There might be a different Meta-QoS-class requirement in the reverse direction than the forward direction. To address this, there can be application level communication between the two parties (customers) involved in order to specify the QoS requirements in either direction. This may also require having a trust relationship between the two involved ASs. Problem **V** stated in section 4.2.1 is also applicable here. There can be a scenario in which an entity (e.g., a Gatekeeper) is assigned for billing. Customers pay that entity and it pays the Source and Destination ASs (ISPs).

## 4.4 Bi-directionality in Hard Guarantees Solution Option (3)

The *pSLS* set-up in solution option 3 is the same as solution option 1 as it provides more flexibility for LSP set-up. If solution option 2 approach is used for establishing the pSLSs and consequently LSP set-up, the definition of e-QC and predetermined scope of *pSLS* limit the extent of LSP set-up to only the destination/s specified by the scope of *pSLSs*. The open destination scope (*) of *pSLSs* in solution option 1 & 3 makes it possible to establish LSPs between any two points in the network as far as these points are covered by the established *pSLSs*.

In solution option 3, neighbouring domains also establish pSLSs between themselves. q-BGP runs between the domains, which already have established pSLSs. Solution option 3 uses q-BGP to announce PCS unique identifiers (PCSID) across the Internet in order for "option-3" ASs to be able to discover a path towards every AS having a PCS.

When an LSP is required to set-up between 2 addresses, at the service level, the service provider communicates with management entities' of source and destination ASs. It provides the management entity's source AS with head-end of LSP and possibly PCSID of that domain, and the management entity's destination AS with tail-end of LSP and possibly PCSID of that domain. It also provides necessary information for destination AS in order to verify/authenticate the source AS's request for LSP establishment as well information required for charging purposes. There is no direct communication between two ASs.

Following the service level communication, the source AS's PCS calculates a PCS-path towards the destination AS, and it's up to each AS in the PCS-path to establish the LSP. The LSP creation request is propagated downstream to appropriate PCSs. The requests include the AS's ASBR, PCSID, and the tail-end address of LSP. This procedure is repeated until the request reaches the destination PCS. After authenticating the identity of LSP requester (source AS's PCS), the destination PCS send a reply message back to the downstream domain's PCS accepting the request and include the LSP loose path (destination, ASBR) addresses in the message. The next downstream domain's PCS does the same adding its own relevant ASBR addresses to the LSP loose path. The originating PCS does the same and it is in a position to request for a RSVP reservation for LSP establishment using the LSP loose path.

In order to have bi-directional communication, *pSLS* and *$p_r$SLS* can be set-up in the same fashion as solution option 1. Thus, based on these *SLSs*, LSPs can be created in forward and reverse directions as described above in order to build bi-directional services.

## 4.5 Conclusions

This chapter discussed the complexity added to the three solution options for creating support for bi-directional services. The cascaded approach for QoS peering is considered. The main issue is how to construct the QoS-enabled reverse path for return traffic. We identified and discussed some fundamental issues with solution option 2 in order to provide bi-directional services and provided detailed discussions on the resulting implications. These are to do with finding the source/s for return traffic, changing/extending of the scope of the return path, the selection and mapping of l-QC/e-QC for the return direction, etc. We provided two methods to establish *SLSs* for enabling providers to offer bi-directional services. The pro and cons of these methods are discussed in detail. The first method does not have the problem of specifying the QoS class for reverse direction as the method is defined for symmetric paths. We believe the most feasible solution for providing bi-directionality is the use of multiple cascades as discussed in method 2.

Providing for bi-directional services in solution option 1 and 3 causes less complication, because *pSLSs* are based on the Meta-QoS-classes and q-BGP is used to get the reachability information within a Meta-QoS-class plane. In solution option 1, there is no strict end-to-end QoS guarantee in forward direction hence there is no need to create a reverse path with a strict end-to-end QoS guarantee.

In solution option 3 and based on *pSLSs* and *$p_r$SLS,* LSPs can be created in forward and reverse directions in order to build bi-directional services. These SLSs allow establishing LSPs between any two points in the network as far as these points are covered by the established *pSLSs* and *$p_r$SLS*.

General conclusion for all solution option 1 and 2 is the requirement for service/application level signalling between the two parties involved. This is to find-out about the Meta-QoS-class plane for reverse direction, information for billing and admission control in solution option 1, to specify the desired sink for return traffic for the Destination AS and the l-QC/e-QC for return traffic, information for billing and admission control in solution option 2. In solution option 3, service level communication is also required to pass to source AS head-end of LSP and possibly PCSID of that domain and destination AS with tail-end of LSP and possibly PCSID of that domain and necessary information for authentication and billing purposes.

# CHAPTER 5: INTER-OPERATABILITY OF MESCAL SOLUTION OPTIONS

## 5.1 Introduction

In this chapter, we discuss inter-operability issues between the solution options, which have been described in [D1.1]. Inter-operability will be discussed for the following two cases:

- Two adjacent peering ASs have implemented and deployed different solution options.
- A single AS has deployed more than one solution option.

Concerning the hard guarantee solution option, it should be noted that this solution option has been defined based upon the loose guarantee solution option. But in this chapter, for the purpose of the discussion, we consider that a provider can decide to sell hard service option only, though its network is technically able to offer the loose service option.

## 5.2 Comparing the Solution Options

Table 6 below summarises the functionalities and concepts that need to be taken into account when deploying the solution options described in [D1.1]. Implementation of these functionalities may differ depending on the solution option.

"M" = mandatory, "C" = conditional, "O" = optional, "-"= indifferent

| | **Loose Solution Option** | **Statistical Solution Option** | **Hard Solution Option** |
|---|---|---|---|
| **e-QC relies on *Meta-QoS-class* concept** | M | O | M |
| **e-QC supports bandwidth constraints** | - | M | M |
| **q-BGP deployment** | M | O | M |
| ***Meta-QoS-class* signaling** | M | O | M |
| **v-QC processing** | - | M | - |
| **Inter-domain bandwidth control** | M | M | M |
| **o-QC advertisement (as defined in the context of the functional model)** | O | O | O |
| **MPLS deployment** | -<br>Could be locally (Intra-domain) deployed | O | M |
| **Intra-domain TE deployment** | M | M | M |
| **RSVP-TE** | - | - | M |
| **PCS deployment** | - | - | M |

**Table 6: Comparing the three solution options.**

## 5.3 Inter-working of Solution Options: Service Considerations

### 5.3.1 Services resulting from inter-working

Table 7 indicates the equivalent service option resulting from the interconnection of two ASs operating different service options. Technical issues related to inter-working are not taken into account in this section. Only, the resulting service is considered.

This table must be read from the point of view of an upstream provider (e.g., AS1) inter working with a downstream provider (e.g., AS2). The service resulting from the interconnection of the two providers' ASs is expressed as a service option

| | | AS 2 | | |
|---|---|---|---|---|
| | | **Loose solution option** | **Statistical solution option** | **Hard solution option** |
| **AS 1** | **Loose solution option** | Loose service option | Loose service option | Loose service option |
| | **Statistical solution option** | Loose service option | Statistical service option | Statistical service option |
| | **Hard solution option** | Loose service option | Statistical service option | Hard service option |

**Table 7: Service resulting when inter-working two options.**

Shaded cells represent inter-working cases whose resulting QoS characteristics guarantees are lower than those provided by the upstream service option, from AS1 point of view. Note that, the compatible service offered by both ASs (diagonal of table) will not be considered hereafter.

As a result, it can be concluded that the transition in service offering would be valid -from a service guarantee perspective- only in the three following cases:

- The loose service option ➔ the statistical service option (i.e., the loose service can be offered (pass through) an environment that offers statistical service).
- The loose service option ➔ the hard service option
- The statistical service option ➔ the hard service option

*However, the above transition service logic won't be strictly respected as discussed below in order to be able to address transit scenarios and traffic bi-directionality issues.*

### 5.3.2 Transit scenario and translation scenario

The basic motivation for a network provider to establish a *pSLS* with a potential service peer is to extend the scope of its supported service options even thanks to the establishment of "*heterogeneous*" agreements. *Heterogeneous* refers to contracts established between ASs that each offers distinct service option).

Two main scenarios can be considered. In the example shown in Figure 22, the solution option x compliant domain offers weaker QoS guarantees than the solution option y compliant domains:

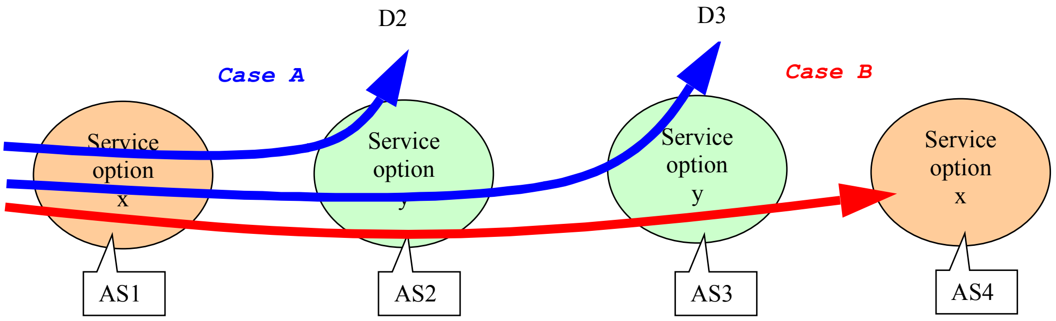


**Figure 22: The transit scenario.**

- Case A: AS1 wants to reach destinations that are "located" in AS2 or AS3 that support service option y. To do so, it is necessary to establish a heterogeneous *pSLS* between AS1 and AS2. The approach is valid since the service guarantees offered by AS2 are better than those provided by AS1.
- Case B: AS1 wants to cross a set of ASs that support service option y to reach another domain that offer service option x. To do so, it is required to establish two *pSLS*s:
  - One between AS1 and AS2: this case is valid since service option y provides better service guarantees than service option x.
  - One between AS3 and AS4: from AS3 point of view this case is difficult to justify since AS4 offers a weaker service (service option x) than the service guarantees offered by AS3.

*Thus, if transit scenarios are considered to be pertinent we need to relax the constraints that "prevent" (at least doesn't encourage) a solution option y compliant domain to establish a pSLS with a solution option x compliant domain.*

### 5.3.3 The Bi-directionality Issue

The MESCAL project doesn't make any assumptions about the services that could be built upon the solution options it provides. Then, and from a pure service perspective, MESCAL should offer a means to provide bi-directional IP services. The bi-directionality should be understood in terms of connectivity and QoS guarantees so that an end-user could have consistent (both directions) IP communications since most of existing applications rely on round-trip exchanges.

If only one *pSLS* (that is supposed to be unidirectional) is established between AS1 and AS2, the downstream traffic could only be returned using best-effort grade of service (see Figure 23). The resulting communication could be a best-effort communication although the end-user accepted to pay for sending its upstream IP traffic with QoS guarantees. This might lead to commercial misunderstandings and could invalidate the added value of MESCAL solutions if the bi-directionality issue is left open.

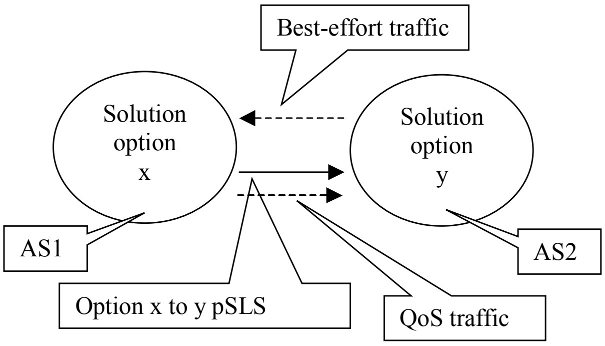


**Figure 23: The bi-directionality scenario.**

It should be noted that in a bi-directional QoS communication that involves two users: user1 and user2, even if the traffic follows a different AS path for the two directions (from user1 to user2 and from user2 to user1), as far as a solution option x to solution option y translation occurs somewhere from user1 to user2 the reverse translation y to x must exist further on towards user2 or from user2 to user1.

It is probably out-of the scope of MESCAL to study bi-directional *cSLS* and *pSLS* assurance services but MESCAL should at least ensure that the solution options could inter-work with *pSLS*s that have been excluded above.

*Thus, the establishment of a reverse pSLS will be examined, when necessary, within the following sections.*

## 5.4 Bandwidth Consideration in the Three Solution Options

End-to-end bandwidth consideration is a critical issue that needs further studies and specific discussions. The aim of this section is just to briefly remind the way each solution option handles it.

- Within the context of the loose solution option, no end-to-end bandwidth guarantees are provided to customers. Only a maximum amount of bandwidth is negotiated per *Meta-QoS-class* for both *pSLS* and *cSLS*. We will call **`LooseBwMgt`** as the bandwidth management for the loose solution option. Bandwidth Management function includes all bandwidth related tasks such as: reservation, computation of available bandwidth, bandwidth consideration per *pSLS*
- For the statistical solution option, end-to-end bandwidth guarantees are associated to each contracted o-QC within the scope of a *cSLS* or a *pSLS*. This requires sophisticated configuration and admission control to be implemented. We will call **`StatistBwMgt`** as the bandwidth management for the statistical solution option.
- Within the context of the hard solution option, strict end-to-end bandwidth guarantees need to be fulfilled. Sophisticated resource reservation mechanisms are necessary in order to deploy this option. We will call **`HardBwMgt`** as the bandwidth management for the hard solution option.

## 5.5 Solution Options Co-existence in the same AS

In this section, we focus on the co-existence of several service options in the same autonomous system. For this purpose, we examine the impact of the deployment of each service option on the network infrastructure and will qualify the compatibility of these solution options.

Note that the basis of comparison relies upon the technical description of the three solution options provided in [D1.1] and summarised in the Table 6

### 5.5.1 Case 1: The loose and the statistical solution options

#### *5.5.1.1 Main differences between solution options*

In order to determine whether the two solution options; the loose and the statistical solution options, could exists together in the same AS, we briefly highlight below their main technical differences:

- The use of the *Meta-QoS-class* concept: the loose solution option relies on the *Meta-QoS-class*es concept and makes it mandatory, which strongly constraints and standardises negotiations between two ASs. *Meta-QoS-class*es guarantee that a flow will receive a consistent treatment all along the AS path. The statistical solution option doesn’t make it mandatory but can sometime (on a per agreement basis) use this notion as a negotiation means to negotiate o-QCs.
- The use of q-BGP: The QoS-inferred BGP is introduced in order to convey QoS information that will help the BGP route selection process. The choice of the best route is no more based only on the optimisation of the number of the AS crossed, but also on the optimisation of its end-to-end QoS parameters. In the loose solution option, an AS learns Internet routes (each route made up

with l-QCs belonging to the same *Meta-QoS-class*) with associated QoS performance characteristics thanks to the activation of q-BGP. The choice of the routes to propagate or to activate relies on an ad-hoc processing of this information and is achieved on a per *Meta-QoS-class* basis. From a statistical solution option standpoint, the choice of the next hop AS is constrained by the o-QC and is particularly obtained from the management plane deduced from the established *pSLS*s. The use of q-BGP is not mandatory, and thus less freedom is left to the q-BGP route selection process (except in case that it is activated between two adjacent domains and used for load balancing purposes).

- The *pSLS*s or *cSLS*s negotiated in the case of the loose solution option can be seen as an authorisation to send traffic in a given *Meta-QoS-class* plane. Note that no end-to-end guarantees are negotiated between two service peers or between a customer and a provider. In the statistical solution option, the *pSLS* or the *cSLS* determines the QoS level that the peer guarantees for the traffic sent up to its final destination(s). This means that strict end-to-end QoS guarantees are to be fulfilled.
- The **`LooseBwMgt`** and **`StatistBwMgt`** are greatly distinct since the loose solution option considers all/any destinations within a *Meta-QoS-class* and achieves shaping operations on this basis, while the statistical solution option considers each individual destination (or set of destinations) associated with each *pSLS*. In the statistical solution option, more complex policies are needed and strict constraints are required for the admission control.

Based on the technical differences listed above, two scenarios are examined in order to evaluate a possible co-existence of the aforementioned solution options in a given autonomous system. The relevant technical issues and problems are discussed in each section and recommendations are given at the end of each section to deal with the problems.

### *5.5.1.2 Discussion*

***Impacts and problems***: In this coexistence case, a given destination (located outside the domain implementing the two solution options) can be made reachable with the loose and/or the statistical solution option. In terms of implementation, each local-QoS-class deployed in the AS can be potentially shared by the two solution options. As a consequence, a datagram requesting a particular l-QC (to signal a *Meta-QoS-class* plane or a specific e-QC) is forwarded along different inter-domain AS paths depending on the service option which were requested by the end-user. In other words the egress point could be different for a given destination depending on the solution option. If the same intra-domain signaling code (e.g., DSCP) was used it would be impossible to differentiate the two egress points.

Thus, in order to be able to handle this case, each solution option should assign distinct and non-overlapping ranges of DSCP values to signal the available l-QCs within a domain. This will allows applying the right routing policies depending on the service option.

At the boundaries of its domain, a provider will have to police the incoming traffic and to shape the outgoing traffic. Both policing and shaping *should* differentiate the traffic between the two service options in order to apply terms of the established *pSLS*s. Separate ranges of DSCP *would* greatly help.

Intra-domain traffic engineering will have to be compliant with the two solution options.

The management plane has to manage the two solution options including *cSLS*s, *pSLS*s, QC operations, traffic engineering policies, and so on. So that it meets the objectives of each solution option when sharing common network resources. This management will have to be intelligent enough so that cross-interactions on resources between the solution options can be controlled.

This later problem occurs several times in the remaining sections. We will refer to it as the ***"Multiple Solution Option Management problem"*** *(MSOM* problem).

***Recommendations:***

- *Use different ranges of DSCP values for the two solution options.*
- *Build a management system able to handle simultaneously the two solution options on top of a common and shared network infrastructure*

### *5.5.1.3 Scenario 1.1*

***Hypothesis***: The deployed statistical solution option doesn't make use of the q-BGP protocol.

***Impacts and problems***: The QoS constrained route selection process depends on the solution option that is considered. Routes that are learned, thanks to q-BGP, will be used only by the loose solution option. The statistical solution option will construct its own routes using established *pSLS*s (e.g. fixed path and injection of ad-hoc route in the IGP). From an inter-domain routing perspective, there is no coexistence conflict between the two service options in the same AS, provided that the MSOMP is solved and external routes needed for the statistical solution option are injected in dedicated range of DSCP routing plane.

***Recommendations:***

- *None*

### *5.5.1.4 Scenario 1.2*

***Hypothesis***: the deployed statistical solution option makes use of the q-BGP protocol.

***Impacts and problems***: In this scenario, we assume that the two solutions options use q-BGP to exchange network accessibility together with QoS performance information and also to select the optimal routes.

Use of q-BGP by each solution option is different since:

- The loose solution option uses *Meta-QoS-class*es. Each q-BGP message ***must*** indicate the *Meta-QoS-class* plane and contain an announcement. The q-BGP process ***may*** use end-to-end QoS performance parameters contained in each announcement in order to select a route for a given *Meta-QoS-class* plane. When receiving an announcement, the learned prefix is propagated in the appropriate intra-domain *Meta-QoS-class* plane (an l-QC which implement the corresponding *Meta-QoS-class*) thanks to the binding, which has been selected, by the management plane. All destinations learned for a given *Meta-QoS-class* will always feed the same intra-domain *Meta-QoS-class* plan(s).
- Within the context of the statistical solution option, the use of q-BGP is optional. The choice of the next hop (and then the path to a given destination) is based on the contracted *pSLS*s. Then, *routing decisions can be completely management-based* since the *pSLS*s include -implicitly- some routing-related information. Routing can be fully based on a fixed path. Nevertheless, in the statistical solution option, q-BGP could be used as a means to enhance the effectiveness of this solution option by:
  - Providing alternative routes in case of failure: of course this requires the establishment of *pSLS*s that enable to reach the same destination and for the same o-QC.
  - Achieve load balancing between paths that serve the same destination(s) and in which traffic will experience the same o-QC.

  The statistic solution option ***must*** use the end-to-end QoS parameters in order to choose a route. When receiving an announcement for a remote prefix, each individual binding depends on the o-QC offered for this destination prefix belongs to. Several remote o-QC signalled with the same inter-domain DSCP code point could be bound on different l-QC to form new o-QCs. There are no systematic mechanisms to feed an intra-domain DSCP routing plan.

***Recommendations:***

- *The q-BGP process must have a means to distinguish the announcements made per solution option so that it can process each announcement according to the service option it belongs to.*

### 5.5.2 Case 2: The Loose and the Hard Solution Options

As described in [D1.1], the loose and the hard solution options are designed and have been specified to work together. Even if the q-BGP protocol is used by the two solution options to learn the QoS capabilities of the Internet, the co-existence of these two solution options could encounter some problems related to the common use of q-BGP.

In order to illustrate this problem, let consider two ASs: AS1 and AS2, which have agreed a *pSLS* allowing them to extend the hard service option. AS1 offers the loose and the hard service option but AS2 offers only the hard solution option. In this example, when receiving routes from AS2, AS1 could use this information within the scope of its loose service option in order to reach learned destinations in AS2. This occurs because the two service options can share the same announcements.

***Recommendations:***

- *Differentiate q-BGP updates per service option.*
- *Or, recommend that the hard service option cannot be offered in a domain if the loose service option is not offered in that domain.*

### 5.5.3 Case 3: The Statistical and the Hard Solution Options

The implementation of the hard solution option is technically based on the loose solution option. Within this context no "pure" IP traffic (but MPLS labelled traffic) would be allowed but *Meta-QoS-class*es and q-BGP should be supported. Moreover, if it were decided that the hard solution options could not be deployed independently of the loose service option (see case 2's recommendations) this would make this case more similar to case 1.

The MSOM problem occurs as it occurred between the loose and the statistical solution options. The management system will have to handle three different service views on top of a shared network infrastructure.

Inter-domain QoS signaling has to provide mechanisms to separate the announcements of the two service options.

***Recommendations:***

- *Use different ranges of DSCP values for the two solution options.*
- *Build a management system able to handle simultaneously the two solution options on top of a common and shared network infrastructure.*
- *Differentiate q-BGP updates per service option.*

***Note:*** *An alternative to deploy both solution options would be to establish LSPs for hard service option purposes using the statistical service option capabilities. This alternative isn't included in [D1.1]. PCS behaviours* ***may*** *be different from what is described in [D1.1] since the inter-AS path would be directly deduced from the pSLS. Thus, LSPs destinations would be constrained by the o-QC definition. The bandwidth to be reserved for the LSPs is taken from the one negotiated for o-QCs. The use of MPLS techniques on top of this statistical solution option should be strongly motivated since (if we except tunnelling aspects) the grade of service they offer would be very close in terms of QoS* *guarantees. It could be argued that more constrained "a la statistical" o-QC is built that would provide the same level of QoS service than LSPs. A provider that would have deployed such an enhanced statistical service option would not consider necessary to deploy a hard solution option.*

### 5.5.4 Case 4: Deploying All Solution Options

This case is the same as the above case but IP traffic is accepted within the scope of the loose service option.

***Recommendations:***

- *Same recommendations as above but they are extended to the three service options.*

## 5.6 Solution options Compatibility in an Inter-domain Scenario

### 5.6.1 The Loose and Statistical Solution Options

In this section we discuss the transition scenarios between the loose and the statistical solution options when they are deployed in two adjacent ASs.

The discussions focuses mainly on the transition from an AS implementing the loose solution option to another AS implementing the statistical solution option. The reverse direction will be briefly examined.

Two main scenarios are studied, that should find ad-hoc clauses this means that the way the adaptation have to be done should be agreed between two service peers and thus be present in an SLA or SLS in *pSLS*:

- The downstream AS adapts its announcements.
- The upstream AS adapts itself.

Figure 24 is used to illustrate the scenarios discussed hereafter.

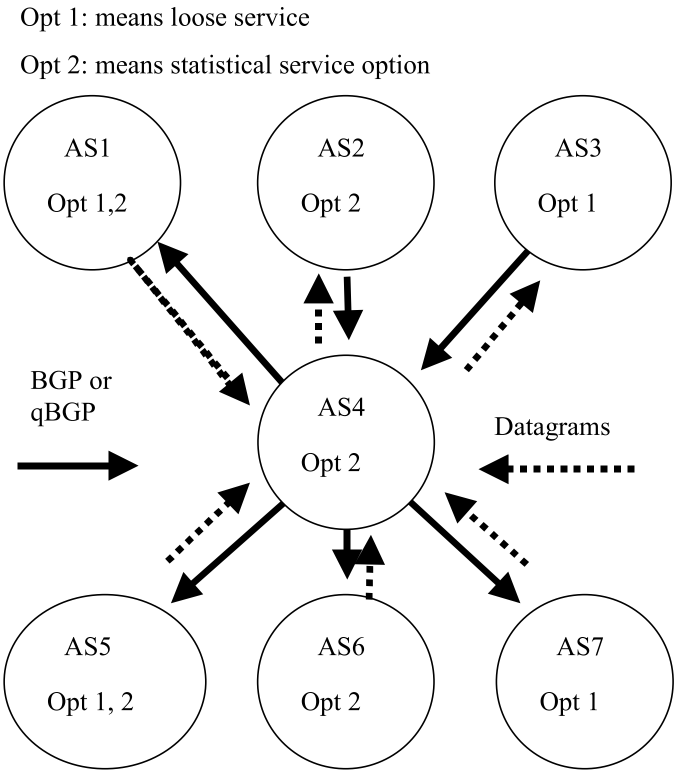


**Figure 24: Examples of interaction between loose and statistical service options.**

The *Meta-QoS-class* paradigm is one of the fundamental keys of the loose solution option. Within this service option, a service peer is informed about the destination an AS could reach per *Meta-QoS-class* plane.

Referring to the operating mode of the statistical solution option, and within this scenario, AS4 doesn't send q-BGP information on this basis. This leads to a blocking state since an AS implementing the loose solution option won't be able to use correctly received q-BGP information. If the loose solution option persisted in using this information this would certainly generate routing and service inconsistencies from the loose service option viewpoint.

Two solutions could be adopted:

- AS4 could adapt its announcements so that they conform to the *Meta-QoS-class* paradigm: this scheme could be handled easily if AS4 would support *Meta-QoS-class*es. In other words, AS4 should determine the *Meta-QoS-class* similar to the case it would announce information related to o-QC. We refer to this case as scenario 1.
- AS5 and AS7 could translate q-BGP announcements received from AS4 into the *Meta-QoS-class* paradigm. This means that AS5 and AS7 would have to extract/deduce/compute the *Meta-QoS-class* information that would be appropriate to reach each learned destination. In other words, AS5 and AS7 would consider each o-QC learned from AS4, as being a virtual combination of a set of compatible *Meta-QoS-class*es enforced by each ASs identified in the AS_PATH attribute of the q-BGP messages. We refer to this case as scenario 2.

### *5.6.1.1 Scenario 1*

***Hypothesis***: AS4 adapts its announcements to the *Meta-QoS-class* paradigm.

***Impacts and problems***: This hypothesis implies that AS4 will send QoS information on a *Meta-QoS-class* basis. In doing this, each o-QC's information has to be injected into a *Meta-QoS-class*. The immediate following questions are raised:

- *What methodology to follow in order to identify the Meta-QoS-class plane in which an o-QC's information must be injected?*
- *Who applies this methodology?*

D1.1 [D1.1] doesn't give any answer to this problem. The technical description of the statistical option should be revisited in order to clarify how this solution option could handle this issue. Proposals for handling this issue are proposed below:

- ***What Methodology to be used?***

1. The methodology could rely on an arbitrary choice. This would work but this is not a valid option since no end-to-end service consistency would be ensured. This means that every service peer could execute its proprietary methodology and no agreement has been made between him and its neighbour.
2. The AS responsible for deducing an equivalent *Meta-QoS-class* from an o-QC could use the following information reported by q-BGP:
    a. The AS_PATH, which gives the number of AS hops.
    b. The aggregated QoS performance characteristics associated with the destinations related to the o-QC.

    This information ***should*** allow to compute average values for l-QC QoS parameters by dividing each individual e-QC's QoS performance values by the number of its associated AS hops. The average values for l-QC QoS parameters would then be classified with regard to *Meta-QoS-class*. The destinations associated with the o-QC would then be announced in the deduced *Meta-QoS-class*. Note that with this approach, we lose semantic aspects of the *Meta-QoS-class* definition.
3. An alternative would be to force the statistical solution option to construct the o-QC on a set of cascaded l-QCs that would conform to the same *Meta-QoS-class*. No computation would be necessary anymore at the border router.

    At this stage, the second proposal is preferred since it is aligned with the current description of the statistical solution option, but the third one could significantly help.

***Who applies this methodology?***

1. The AS implementing the statistical solution option could decide (thanks to a methodology processing) the set of o-QC it intends to export to the loose service option. The q-BGP listener would learn available routes as it learns them when peering with another loose solution compliant domain.
2. The AS implementing the loose solution option could select (via an ad-hoc o-QC advertisement and discovery service) the set of o-QC he wishes to import and would apply itself the above methodology. The AS supporting the statistical solution option would announce the corresponding destination (s) in the related plane as specified by its service peering.

In this scenario, the QoS information carried in q-BGP messages are made compliant with those of the loose service option.

In the example shown in Figure 25, AS5 and AS7 can now extend the scope of their loose service option when co-operating with AS4. Furthermore, AS4 is considered as a loose solution option compliant AS since it meets all service peering requirements of the loose service option.

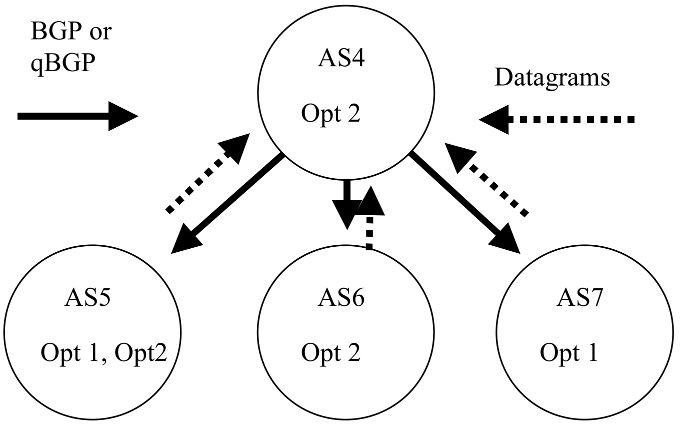


**Figure 25: Scenario 1 example.**

Within this scenario, three additional problems need to be resolved as explained below.

- ***The signalling problem***

Inter-domain QoS signalling is a critical problem within the context of this scenario when a given AS offers the two solution options or both ASs offer the two solution options. To illustrate this problem (see Figure 25), let's consider the case where AS5 and AS4 have established a *pSLS* allowing AS5 to extend the statistical service option with AS4 destinations. AS4 sends q-BGP UPDATE messages to inform AS5 about its intra and/or inter-domain routes and QoS capabilities. From an intra-domain perspective, AS5 needs a means to exploit those announcements in the appropriate service option and more particularly in the right option dependent routing space. This is due to the conclusion drawn in section 5.5.1.2 that separate ranges of DSCP (and routing spaces) should be used when the two solution options have to coexist.

In this example, at least q-iBGP should be able to handle such a case but q-eBGP should also do the same when the two peering ASs would support the two service options.

This issue could be tackled by ***indicating in each q-BGP message, which service option it serves***. A technical solution could rely on the use of the community attribute so as to be able to differentiate q-BGP announcements per service option.

- ***The MSOM problem***

The management plane of the AS supporting the statistical option is affected since it must be able to support "loose *pSLS*s" together with the related implementation policies. The two solutions options must be supported; maintaining full support for the statistical option and partial (inter-working mode) support for the loose solution option.

- ***The bandwidth management problem***

In the example shown in Figure 26, AS7 has established *pSLS*s with AS4. AS7 will shape the outgoing traffic as all ASs supporting the loose service option are supposed to do it as per *Meta-QoS-class*. Since AS4 implements adaptation functions to support inter-working with the loose service option it will police the overall exchanged traffic according to a loose service option. The QoS characteristics (e.g., signalling code, etc.) of the received traffic will then be mapped by AS4 onto the appropriate o-QCs. But for each supported o-QC and related *pSLS*, there is an admission control as well as a second policing step that will be applied by AS4.

Since bandwidth management differs between the two solution options (different level of aggregation) there is a high probability that the traffic sent by AS7 doesn't conform to the individual and dedicated bandwidth of each o-QC (*pSLS*) for a given *Meta-QoS-class*.

As a result, even if AS7 sends a traffic conforming to the agreed *pSLS*, the dispatching of this traffic between the o-QCs might exceed the individual dedicated bandwidth of some of them, leading AS4 to reject part of this traffic.

In order to illustrate this problem, let's consider the example below (see Figure 26).

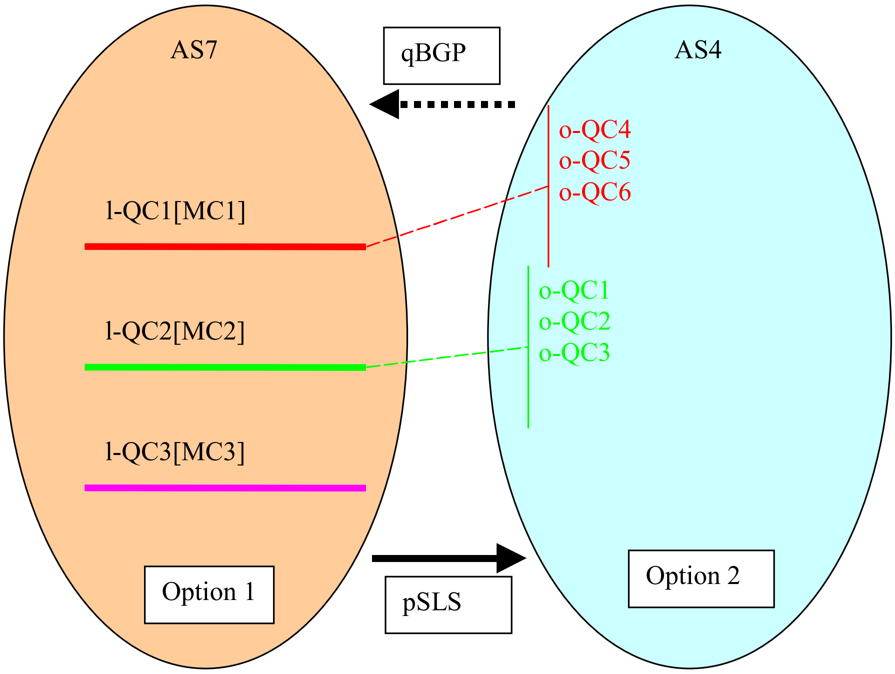


**Figure 26: Bandwidth management problem.**

- AS7 will shape the traffic using the following rules:
  - Bandwidth between AS7 and AS4 for MC1 **MUST** not exceed the Sum(Bandwidth (o-QC4(*pSLS*)), Bandwidth (o-QC5(*pSLS*)), Bandwidth (o-QC6(*pSLS*)))
  - Bandwidth between AS7 and AS4 for MC2 **MUST** not exceed the Sum (Bandwidth (o-QC1(*pSLS*)), Bandwidth (o-QC2(*pSLS*)), Bandwidth (o-QC3(*pSLS*)))
- AS4 will configure the following rules:
  - Bandwidth between AS7 and AS4 for o-QC1 for a precise *pSLS* **MUST** not exceed the Bandwidth (o-QC1)
  - Bandwidth between AS7 and AS4 for o-QC1 for a precise *pSLS* **MUST** not exceed the Bandwidth (o-QC2)
  - Bandwidth between AS7 and AS4 for o-QC1 for a precise *pSLS* **MUST** not exceed the Bandwidth (o-QC3)
  - Bandwidth between AS7 and AS4 for o-QC1 for a precise *pSLS* **MUST** not exceed the Bandwidth (o-QC4)
  - Bandwidth between AS7 and AS4 for o-QC1 for a precise *pSLS* **MUST** not exceed the Bandwidth (o-QC5)
  - Bandwidth between AS7 and AS4 for o-QC1 for a precise *pSLS* **MUST** not exceed the Bandwidth (o-QC6)

Since AS7 implements a loose inter-domain shaping it is not able to detect if a particular o-QC(negotiated within a *pSLS*) bandwidth threshold has been reached. Then AS7 will continue to send traffic until the Sum (Bandwidth (o-QC4), Bandwidth (o-QC5), Bandwidth (o-QC6)) or Sum (Bandwidth (o-QC1), Bandwidth (o-QC2), Bandwidth (o-QC3)) is reached. Then AS4 will reject the traffic exceeding the individual contracted bandwidth of each o-QC (within a dedicated *pSLS*).

***As a result, a problem of optimisation of the inter-domain bandwidth will always occur.***

***Recommendations:***

- *Specify a methodology the statistical solution option should follow in order to adapt o-QCs to the Meta-QoS-class concept of the loose solution option.*
- *Solve the bandwidth management problem.*
- *When the two solution options need to coexist in the same AS:*
  - *Differentiate q-BGP announcements per solution option*
  - *Use different range of DSCP per option for a given PDB.*

### *5.6.1.2 Scenario 2*

***Hypothesis***: In this scenario, we consider that the AS supporting the loose solution option will adapt BGP messages received from the statistical solution option to the *Meta-QoS-class* paradigm. In addition, it will shape the traffic, as a statistical solution option would do it.

***Impacts and problems***:

In order to study this scenario, let's consider the following example (see Figure 27):

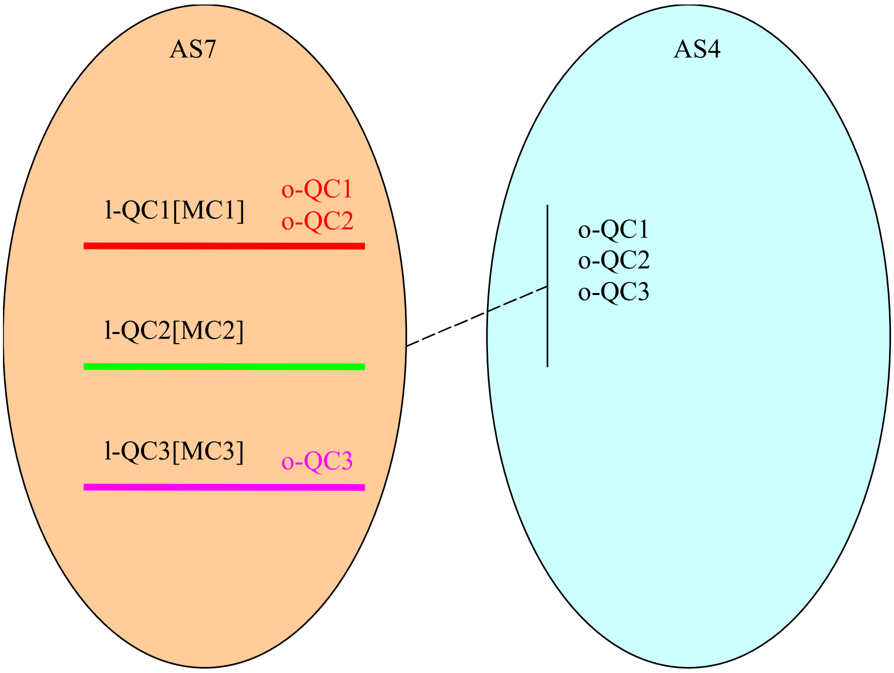


**Figure 27: Scenario 2.**

AS7 has established a *pSLS* with AS4. It can benefit from o-QC1, o-QC2 and o-QC3. Consequently, AS4 sends q-BGP messages to AS7, which has to inject routes associated to these o-QCs in the appropriate *Meta-QoS-class*es in order to extend its loose solution option.

AS7 should:

- Understand q-BGP announcement sent by a statistical service option speaker.
- Inject each learned o-QC in the appropriate *Meta-QoS-class* plane.
- Operate the loose q-BGP route selection process.

The problem of the bandwidth management doesn’t occur here because AS7 achieves a shaping that conforms to the statistical service option requirements.

- *The MSOM problem*

The management plane of the AS supporting the loose option is impacted since it must be able to support "statistical *pSLS*s" together with the related implementation policies. The two solutions options must be supported: maintaining full support for the loose option and partial support (inter-working mode) for the statistical solution option.

> ***Recommendations:***
>
> - *Specify a methodology the loose solution option should follow in order to adapt o-QCs to the Meta-QoS-class concept of the loose solution option*

### *5.6.1.3 Bi-directional aspects in the context of inter-working of loose and statistical solution options*

In order to comply with one of the assumptions of the MESCAL project that states to "*provide an interface to the service level allowing the introduction of more sophisticated services",* and considering that bi-directional services are part of those sophisticated services, the reverse path set-up, from the statistical solution option to the loose solution option, should be studied (see Figure 28).

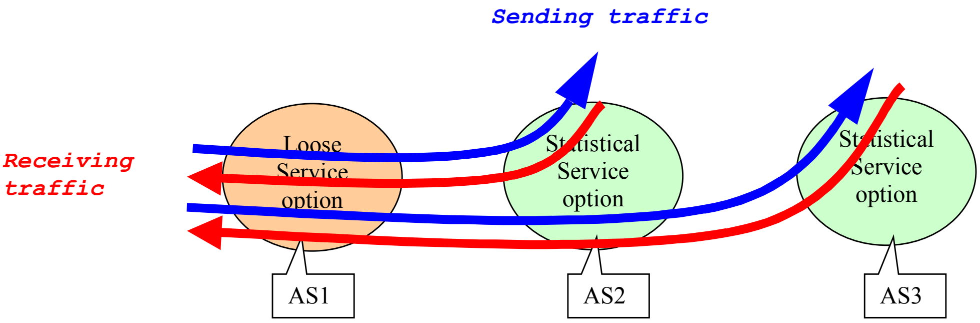


**Figure 28: Bi-directionality problem.**

As discussed in section 5.3, *pSLS*s that would be established from a statistical to a loose service option (red flows in Figure 28) would not bring any added-value in terms of QoS guarantees and would only be motivated by bi-directionality (or inter-domain transit) considerations.

In the previous sections, the main problem was in injecting an o-QC in a *Meta-Class-plane*. In this section it would be the other way around: ***how to transform the QoS requirements of a destination learned from a loose service option into an o-QC that remains compatible with the statistical solution option?***

- The simplest way to solve this inter-working issue would be to deploy the loose solution option and to make it mandatory each time a reverse path needs to be created.
- The second way would assume that this reverse *pSLS* would be only asked by the AS implementing the loose service option, for serving very specific and particular business objectives. In this context, there would be a limited number of o-QC to build by AS2. Each *cSLS* sold to AS2 customers and using these o-QCs could only be loose *cSLS* since no end-to-end QoS guarantees could be provided. Mapping and binding could be achieved as the loose option does. The AS supporting the loose option could choose the l-QC with is the closest from the *Meta-QoS-class* it selected.
  - Additionally, routing inconsistencies should be avoided (see Figure 29).

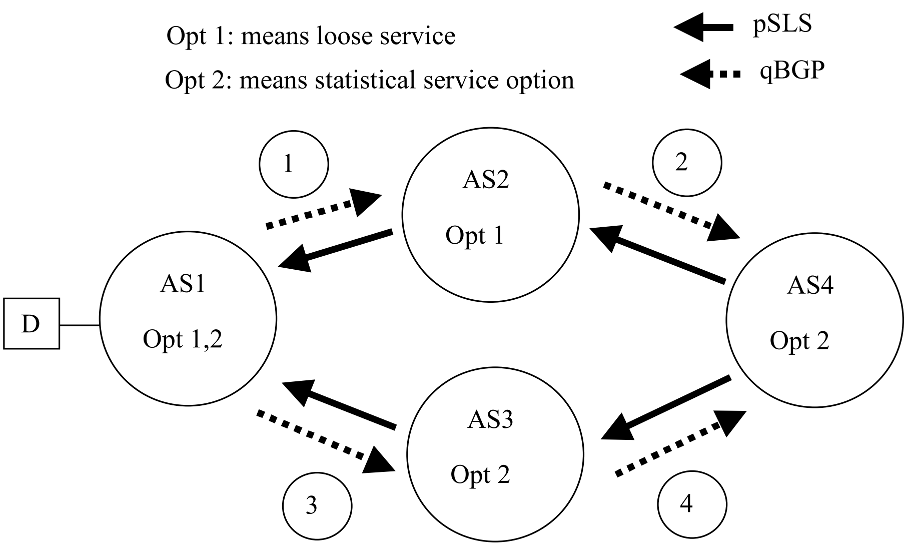

Opt 1: means loose service
Opt 2: means statistical service option
pSLS
qBGP
1
AS2
Opt 1
2
D
AS1
Opt 1,2
AS4
Opt 2
AS3
Opt 2
3
4

**Figure 29: potential routing inconsistencies.**

As illustrated above, a destination "D" could be announced within the scope of the loose option between AS1 and AS2 and then announced by AS2 to AS4 to become and o-QC within the scope of an inter-working peering agreement. In parallel, this same destination could be known of AS4 via AS3 thanks to a pure statistical *pSLS*. If this destination were injected in the same DSCP plane, routing problems would appear and a loose o-QC (*"loose QC" means an o-QC that is built thanks to a pSLS bought from a loose solution option enabled domain*) could be sold to another AS.

In order to solve this potential problem:

  - We could signal loose o-QCs differently from the real statistical o-QCs using a specific range of DSCP.
  - Track and prevent, at the management plane, all potential routing collisions in taking care of never binding a destination learned from such an inter-working peering onto a l-QC already bound onto this same destination so that q-BGP cannot never select this route, or load balancing can never use this route, for pure statistical services. This means that: because of the resulting o-QC is considered as loose, this shouldn't be announced to other peers as pure statistical solution option.

- Egress shaping:
  - Should be achieved correctly if the statistical AS would behave as a loose service option at the egress ASBR and would achieve a *Meta-QoS-class* based shaping. It would be easier if all destination learned from a given *Meta-QoS-class* would have been bound onto the same l-QC otherwise it would be complicated

## 5.6.2 The loose and the hard Solution Options

This section discusses technical inter-working aspects between solution option 1 and 3.

The Cases studied are listed in Table 8 depending on the service options offered by each provider:

| Case | Provider 1 ($AS_x$) | Provider 2 ( $AS_y$) |
|---|---|---|
| **A** | 1 | 3 |
| **B** | 1 & 3 | 1 |
| **C** | 1 & 3 | 3 |
| **D** | 1 & 3 | 1 & 3 |

**Table 8: Loose and the hard solution options related scenarios.**

Those Cases are discussed in both directions i.e., $\mathbf{AS_x}$ to $\mathbf{AS_y}$ and $\mathbf{AS_y}$ to $\mathbf{AS_x}$.

We examine what would happen if a provider decided to establish a *pSLS* with another provider supporting a different set of service options. In particular, we assume here that q-BGP mechanism is activated and the exchanged information are shared and used by both service options 1 and 3.

### 5.6.2.1 Cases A and C

Despite option 3 has been elaborated above option 1, the issue of considering a provider offering only service option 3 has nevertheless been raised during the elaboration phase of this document.

By supporting option 3 only, we consider the case of a service provider that would decide to not sell any option 1 *pSLS* whilst it could have the technical capabilities for doing it, since solution option 3 relies on top of solution option 1. This decision would apply to existing and future peering contracts. Remember that supporting option 3 means that qBGP must be supported, activated and must be *Meta-QoS-class* aware.

Figure 30 is used to discuss Case A and Case C.

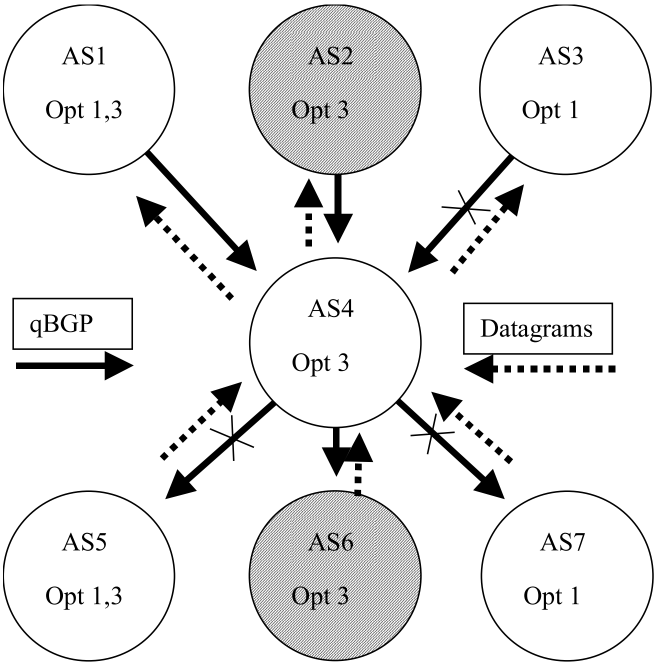


**Figure 30: Case A and C example.**

Firstly, the discussion is based on the AS4 viewpoint that only offers service option 3 to its customers. We also examine *pSLS*s that it potentially is able to buy.

- AS4 can establish a *pSLS* with AS3. In this case, it will receive q-BGP information. Since AS4 decided not to activate service option 1, it will never forward any IP traffic to AS3. If it tries to establish LSPs towards AS3 this will fail since AS3 is not an option 3 compliant. **Case A is an invalid case.** It is quite possible to carry traffic (coming from AS3 towards AS5) belonging to e.g., a *Meta-QoS-class* passing through an LSP in AS4. This can be an e2e service provided for option 1 customers and not option 3. Thus, the case A is invalid if AS4 wishes to provide hard guarantee e2e to its customers. These are hybrid scenarios and for implementing them we have to combine *pSLS* from different options, use very particular engineering and adaptation techniques which are out the scope of the pure option 1 and 3. It is decided to remove those discussions from the document since their added value is not significant. If AS4 establishes a *pSLS* with AS1, it will then receive q-BGP advertisements. Since both ASs are option 3 compliant, AS4 can establish end to end LSPs toward AS1.

We can now examine what could AS4 sell:

- If AS4 accepts to sell a *pSLS* to AS7, AS7 will then receive q-BGP advertisements. These advertisements will be computed and possibly propagated to others ASs. But, since AS4 decided to support only option 3, all incoming IP traffic from AS7 will be dropped. This confirms again that **case A is an invalid case.**

If AS4 sells a *pSLS* to AS5, AS5 will also receive q-BGP announcements. These advertisements will then be used by the solution option 1 and the same drawback as above will appear. **Case C is an invalid case** too. "Thus, even if it is valid to establish an LSP from an option 3 compliant AS toward an option 1 and 3 compliant AS (as shown by arrow between AS4 and AS1), no return LSP can be established as just clarified in the bullet point and shown by arrow between AS4 and AS5)". In this scenario, AS4 would be able to establish an LSP towards AS1 but the reverse would not be correct. The only way for returning the traffic would be to use a best-effort level of guarantees.

We can conclude that an AS that is only compliant with option 3, would only be able to peer with option 3 compliant ASs for providing bi-directional services.

### *5.6.2.2 Case D*

In this scenario (see Figure 30) we suppose that AS1 and AS2 have both commercially activated options 1 and 3. Thus, any option 3 compliant domain, can use each route reported by q-BGP to establish inter-domain LSPs.

It should also be considered that:

- Generally speaking, a provider can decide to apply filters in order to propagate a subset of its own networks or a subset of networks it learned with q-BGP from remote ASs. Some filtering capabilities (network and AS-PATH filtering) must be made available to the network service provider to achieve this. Additionally, this filtering must be achieved on a per *Meta-QoS-class* basis.
- In order to reflect commercial practises, the case of a provider that would decide to apply different inter-domain routing policy for service options 1 and 3 should also be considered. For instance, we could imagine that a given AS accepts terminating traffic for option 1 but not for option 3. This policy could be applied to all *pSLS*s or decided/negotiated case by case, each time a *pSLS* needs to be established.

There are several scenarios, which are shown in Table 9 and discussed below, depending on the commercial policy the provider decides to apply at the peering point and should be realised by the *pSLS* it sells.

| | Option 1 | Option 3 |
|---|---|---|
| Scenario 1 | N | N |
| Scenario 2 | Y | N |
| Scenario 3 | N | Y |
| Scenario 4 | Y | Y |

**Table 9: Scenarios for Case D.**

In Table 9, we indicate, per service option, whether a given network prefix known by AS2 is made known to the peering AS1 partner (Y) or not (N). This network prefix can be a network prefix of the provider selling the *pSLS* (AS2) or a network prefix it learned thanks to *pSLS*s already established with other providers (AS4 or AS5 for instance).

In the following sections we discuss these different scenarios according to topology shown in Figure 31 where AS1 and AS2 have both activated the two service options. Scenarios listed in the Table 9 are successively applied to the peering relationship between AS1 and AS2. We consider that AS1 requests a *pSLS* from AS2 whose peering policy conforms to the above scenarios for a given network. We also consider that a *pSLS* exists between these two ASs, which allows AS1 to exchange some traffic within the scope of service options 1 & 3.

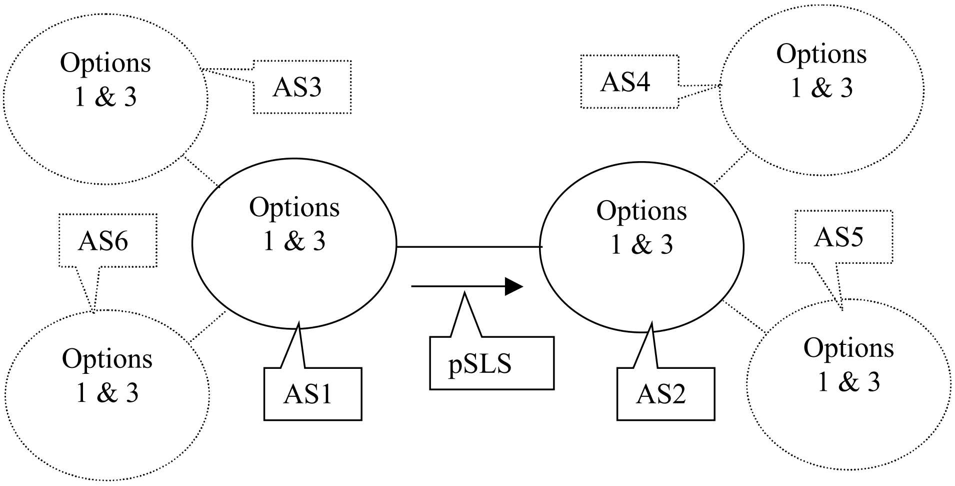


**Figure 31: Example topology for Case D.**

### *5.6.2.2.1 Scenario 1*

The network prefix (it does not matter whether this prefix belong to AS2 or has been learned via already existing peering) should not be announced at the peering point. Since q-BGP won't propagate this network it will not be reachable within both service options 1 & 3.

### *5.6.2.2.2 Scenario 2*

The network prefix is announced by q-BGP. Thus, it is aware (AS1 is now aware of the existence of this network prefix) of AS1 and other remote ASs for which a cascade of *pSLS*s exists. AS1, or a remote AS (AS3 and AS6 for instance) can potentially ask for LSPs toward this destination while it should not.

A simple way to handle that case is to let the PCS achieve the "filtering" at the application level. Each time a neighbour AS requests an LSP with a prohibited target tail-end destination, the PCS rejects the request. Implementation is "easy" and impacts only the PCSs but:

- It triggers unnecessarily a number of requests between PCS, thus potentially overloading the PCSs.

- It gives an unrealistic view of the QoS Internet capabilities within the scope of option 3 since a requesting provider believe a destination is accessible within the scope of option 3 while it is not.

#### *5.6.2.2.3 Scenario 3*

Since option 1 and option 3 share the same q-BGP information, the network is nevertheless announced by q-BGP. Thus, it is aware of AS1 and other remote ASs for which a cascade of *pSLS*s exists. AS1, or a remote AS, can potentially send QoS IP traffic toward this destination within the context of option 1 while it should not.

AS1 can install and propagate a route toward this network while AS2 can decide to discard any datagrams toward this destination.

This is a dangerous case leading to major inter-domain routing inconsistencies.

#### *5.6.2.2.4 Scenario 4*

No particular problem is detected for this scenario.

#### *5.6.2.2.5 PCS and pSLS interactions*

When the PCS in AS2 receives a request from a remote PCS, it must verify that a service option 3 *pSLS* has been established with the requesting remote domain. AS2 PCS must consequently be aware of all service option 3 *pSLS*s, which have been established by AS2 and will only consider requests from the corresponding ASs.

#### *5.6.2.2.6 Conclusion for case D scenario*

Table 10 summarises issues raised in the previous above bullets.

| | Option 1 | Option 3 | Conclusion |
|---|---|---|---|
| Scenario 1 | N | N | No q-BGP announcement needed for the network |
| Scenario 2 | Y | N | Option 3 path computation inconsistencies. Inefficiency should be avoided. |
| Scenario 3 | N | Y | Option 1 routing inconsistencies. Major drawback. |
| Scenario 4 | Y | Y | No problem |

**Table 10: Conclusion for case D scenario.**

It can be concluded that managing different routing policy between the two service options:

1-implies PCSs to be aware of the *pSLS* established at the peering points of the domain in order to be able to identify remote PCS it will accept requests from.

2-generates routing and path computation inconsistencies, thus generating an obvious risk of inefficiency for option 3 and some major failures for option 1.

### *5.6.2.3 Case B*

In the following section we consider that AS1 and AS2 have both activated service option 1 and AS1 has also activated service option 3 (see Figure 32).

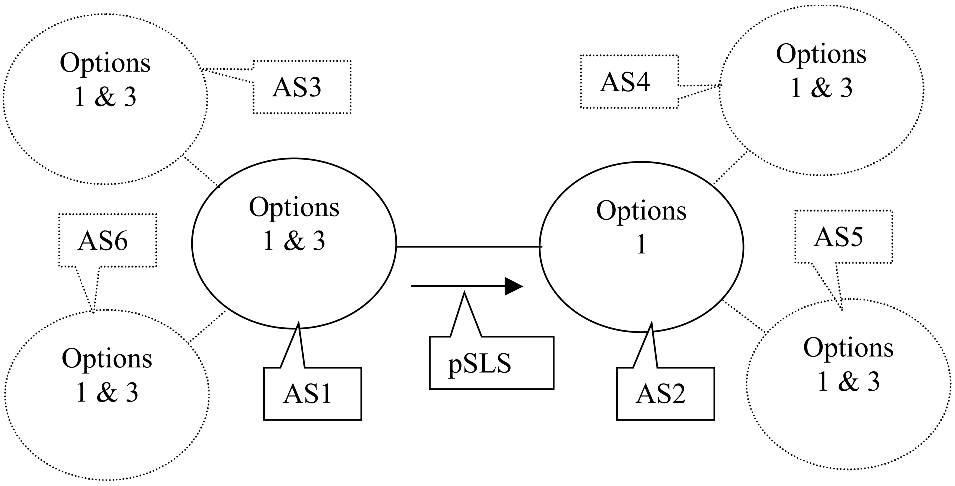


**Figure 32: Case B.**

Scenarios listed in the Table 11 are successively applied to the peering relationship between AS1 and AS2. AS1 requesting a *pSLS* from AS2 whose peering policy conforms to the scenarios listed below.

| | Option 1 | Option 3 |
|---|---|---|
| Scenario 1 | N | N |
| Scenario 2 | Y | N |

**Table 11: Case B.**

There are no particular problems since the inter-working is only achieved on an option 1 basis.

### *5.6.2.4 Concluding remarks on inter-working of option 1 & option 3*

The major drawbacks resulting from the above discussions are as follows:

- It is not possible to deploy only service option 3, excepted if peering occurs with ASs having the same policy.
- It is not possible to deploy different routing policies for solutions options 1 and 3. Corresponding filtering actions can lead to an inconsistent view of the QoS Internet and involve a risk of inefficiency.

The simplest way to handle these interoperability issues is to conform to the initial spirit that lead to the specification of option 1 and 3 and conclude:

- A MESCAL compliant AS must deploy and activate service option 1 each time a *pSLS* is established. In other words ASs should be option 1 or options 1 & 3 compliant.
- Filtering is done at service option 1 level and applies to service option 3, without any modification.

### *5.6.2.5 Possible signalling solutions*

#### *5.6.2.5.1 A Single signalling Channel*

Figure 33 shows an issue that occurs when no distinction is made between "option-1" prefixes and "option-3" prefixes.

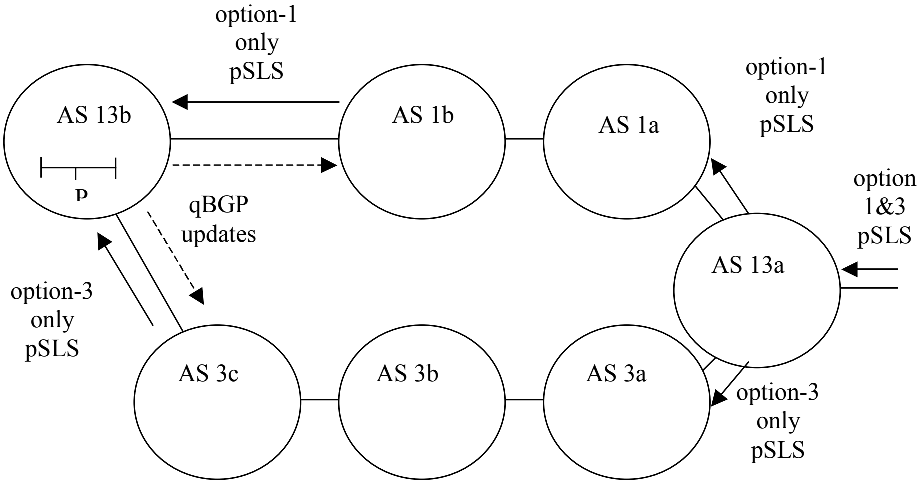


**Figure 33: Single signalling.**

In this example, AS 13b originates the prefix P. According to its established *pSLS*s, it advertises this prefix to AS 1b according to existing "option-1 only" *pSLS*, whereas it advertises this prefix to AS 3c according to existing "option-3 only" *pSLS*. This prefix is forwarded as is to AS 13a. AS 13a learns prefix P thanks to establishment of two distinct *pSLS*s: "option-1 only" and "option-3 only", and has to forward this prefix to its external peers. In this example, AS 13a has an "option 1&3" *pSLS* with an external peer. The question is: how should it announce prefix P to this external peer?

It can choose between one of the 2 paths, and forward it. In this case, the path and the associated QoS attributes of the update are not relevant to the option in the path not been chosen. This is not a good solution.

A possible solution would be to make mandatory the deployment of the loose service option (option-1) when deploying the hard service option (option-3). In other words, there would be no "option-3 only" AS as already stated in §5.6.2.4.

Within the context of this solution, all q-BGP messages are intended for the loose and the hard service options purposes. This solution is fully aligned with the [D1.1] specifications. One of the advantages of this solution is that it doesn't increase the size of the routing tables and doesn't degrade the performances of the switching/routing devices (more than what is needed to include *Meta-QoS-class*).

Nevertheless, this solution doesn't provide any information that will help to optimise the PCS communications since there is no indication on the deployment of the hard solution option within a given domain. In addition, the path computed by q-BGP doesn't take into account the deployment of the hard service option. As a conclusion, the PCS communications will sometimes fail as there may be "option-3" holes in the path selected by the PCS.

To illustrate this problem, let's consider Figure 34. The AS13a receives q-BGP UPDATE messages that announce the reachable destinations per *Meta-QoS-class*. The PCS located in the AS13a will use the collected information thanks to q-BGP in order to compute an AS path that will be used to reach the destinations located in the AS13d and deduce the next PCS to be contacted. In this example q-BGP will report two possible paths, but one of the two AS13a possible requests will fail since one of this paths contains an AS deploying the loose service option only. This could be avoided if there were a way to inform about the existence of *the hard service option holes* in the AS path computed via q-BGP and then predict the failure of the PCS requests.

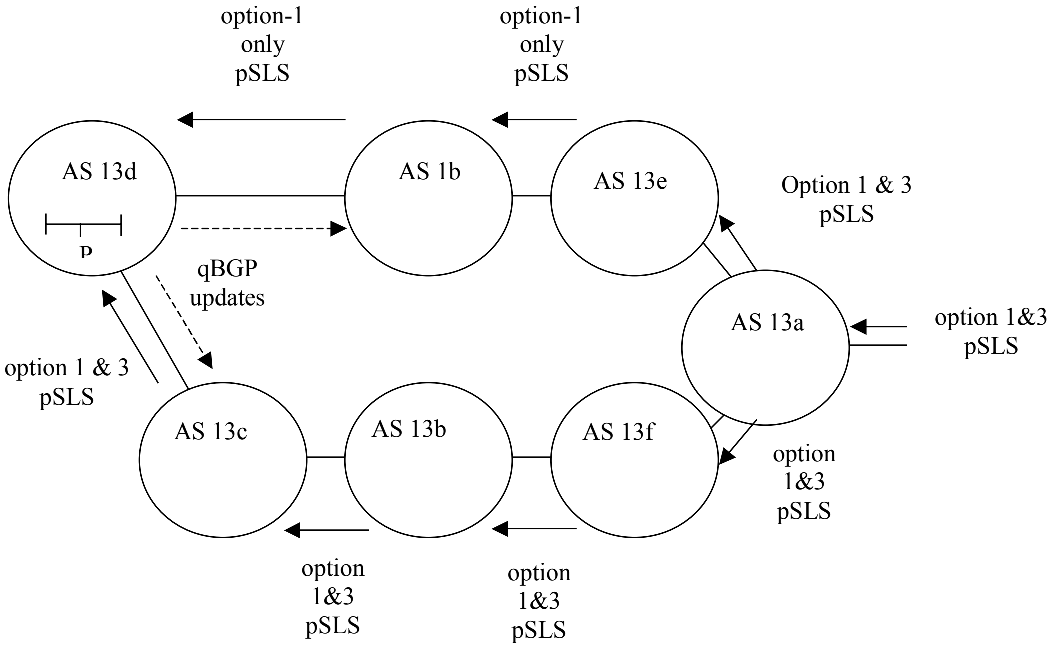


**Figure 34: Hard guarantee hole.**

The proposed solution is to introduce a dedicated flag in q-BGP messages that is removed when an AS offering only the loose service option is crossed (This can be done using an OPTIONAL NON TRANSITIVE community attribute). This modification would help PCSs to filter paths containing "option-3" holes, thus enabling to choose pertinently the next AS to be contacted and then get a better vision hard service option deployment scope.

Unfortunately this solution is not optimal. If AS13a learns 2 possible routes and selects the one that contains the option 3 hole, this latter will be announced to downstream service peers. As a result of this selection process none of the downstream AS will be able to establish LSPs towards AS13d prefixes through AS13a whilst we know a valid path exits.

We could modify the behaviour of the AS13a as follows. If AS13a learns 2 (or more) routes for a same destination, one of them having no option-3 hole, AS13a will announce this destination with the best QoS performance characteristics of learned routes and will indicate that this route doesn't contain any option 3 hole.

This improvement doesn't provide service option-3 with an accurate view of the real QoS performance characteristics of the route since these characteristics can sometimes be those of the path selected for solution option-1.

### *5.6.2.5.2 A Double signalling Channel*

Another way to deal with the aforementioned issue is to be able to differentiate updates for "option-1" related prefixes and "option-3" related prefixes. This can be done using the community attribute of q-BGP. Let's consider the configuration shown in Figure 35.

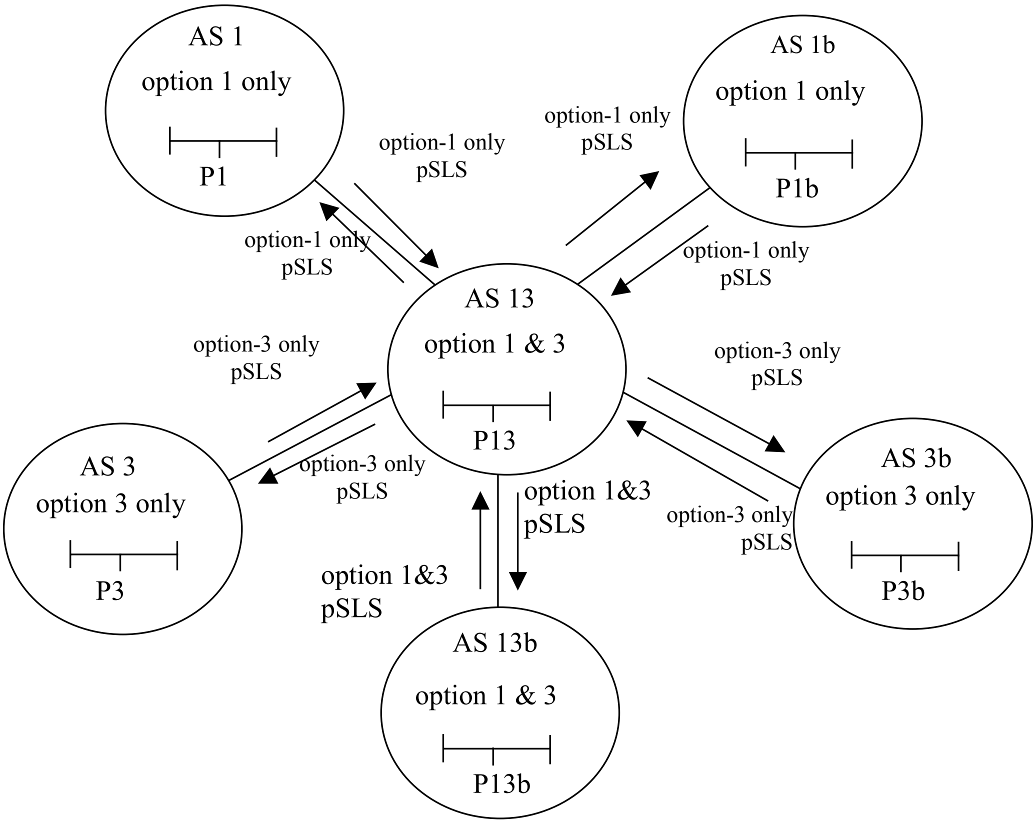


**Figure 35: Double signaling.**

An "option 1&3" AS (AS 13) has *pSLS*s with two "option-1 only" ASs (AS 1 and AS 1b), two "option-3 only" ASs (AS 3 and AS 3b) and with another "option 1&3" AS (AS 13b). Each AS advertises a prefix of its own.

The syntax is as follows:

- AS x is "option-x only" and originates prefix Px
- AS xb is "option-x only" and originates prefix Pxb
- AS xy is "option x&y" and originates prefix Pxy
- AS xyb "option x&y" and originates prefix Pxyb

In the simplest case, a *pSLS* between two ASs involves only the common option(s) offered by the two ASs. We assume that there exist some well-know community attribute values related to options available within ASs. In this document, those values are "1", "2" and "13". We assume that the value "13" is known and understood by "option-1 only" and "option-3 only" ASs.

Originating a prefix within BGP:

Each AS originating a prefix must include the community attribute in its BGP updates as specified below:

- If the advertised prefix is announced in the scope of an "option-1 only" *pSLS*, the community attribute must be set to "1";
- If the advertised prefix is announced in the scope of an "option-3 only" *pSLS*, the community attribute must be set to "3";
- If the advertised prefix is announced in the scope of an "option 1&3" *pSLS*, the community attribute can be set to "1", "3" or "13".

Forwarding BGP updates to external peers:

Before forwarding a BGP UPDATE message to an external peer, a BGP speaker must filter the BGP update messages according to the community attribute values and the *pSLS* it agreed on with its external peer. It must proceed as below:

- If the *pSLS* with the external peer is "option-1 only" and the BGP update has value "1" in its community attribute, the BGP speaker forwards the update as it is.
- If the *pSLS* with the external peer is "option-1 only" and the BGP update has value "3" in its community attribute, the BGP speaker does not forward the update.
- If the *pSLS* with the external peer is "option-1 only" and the BGP update has value "13" in its community attribute, the BGP speaker replaces the value of the community attribute with "1" and forwards the update.
- If the *pSLS* with the external peer is "option-3 only" and the BGP update has value "3" in its community attribute, the BGP speaker forwards the update as it is.
- If the *pSLS* with the external peer is "option-3 only" and the BGP update has value "1" in its community attribute, the BGP speaker does not forward the update.
- If the *pSLS* with the external peer is "option-3 only" and the BGP update has value "13" in its community attribute, the BGP speaker replaces the value of the community attribute with "3" and forwards the update.
- If the *pSLS* with the external peer is "option 1&3" and the BGP update has value "1" in its community attribute, the BGP speaker forwards the update as it is.
- If the *pSLS* with the external peer is "option 1&3" and the BGP update has value "3" in its community attribute, the BGP speaker forwards the update as it is.
- If the *pSLS* with the external peer is "option 1&3" and the BGP update has value "13" in its community attribute, the BGP speaker can forward the update as is or replace the value of the community attribute with "1" or "3" and forward the update.

With the above algorithm:

- AS 1 and AS 1b will never receive prefixes P3 and P3b, but will be able to join P13 and P13b using option-1 only.
- AS 3 and AS 3b will never receive prefixes P1 and P1b, but will be able to join P13 and P13b using option-1 only.
- AS 13b will be able to join prefixes P1 and P1b using option-1 only, and P3 and P3b using option-3 only.

Receiving eBGP updates:

When a BGP speaker receives an eBGP update, it must forward it to all its internal peers (prior to this that, it can check that community attribute is related to the corresponding *pSLS*. If not, it can re-tag the update or discard it). When creating a FIB, the BGP speaker must use the updates which community attribute matches the corresponding option of the created FIB.

This solution does not follow the assumption made in D1.1 about option-3 that "*q-BGP is running between domains, which have agreed to establish a pSLS. Each domain receives, per Meta-QoS-class plane, the set of destinations that can be reached within each Meta-QoS-class plane it supports, together with some aggregated QoS performance information.*" Indeed, if a prefix is forwarded several times according to its related service options, it means that each domain should receive per *Meta-QoS-class* **and per option** the set of destinations. This is an issue that to a great extent increases the memory space needed for RIBs and FIBs, as well as the number of q-BGP updates.

### *5.6.2.5.2.1 PCS based routing for option 3*

A solution to this issue would be to avoid creating RIBs and FIBs for option-3 prefixes. Indeed, when a LSP is created, IP routing is no more needed for IP forwarding of user traffic. Calculating an AS path can be sufficient for PCSs to create the LSP. Figure 36 illustrates this solution.

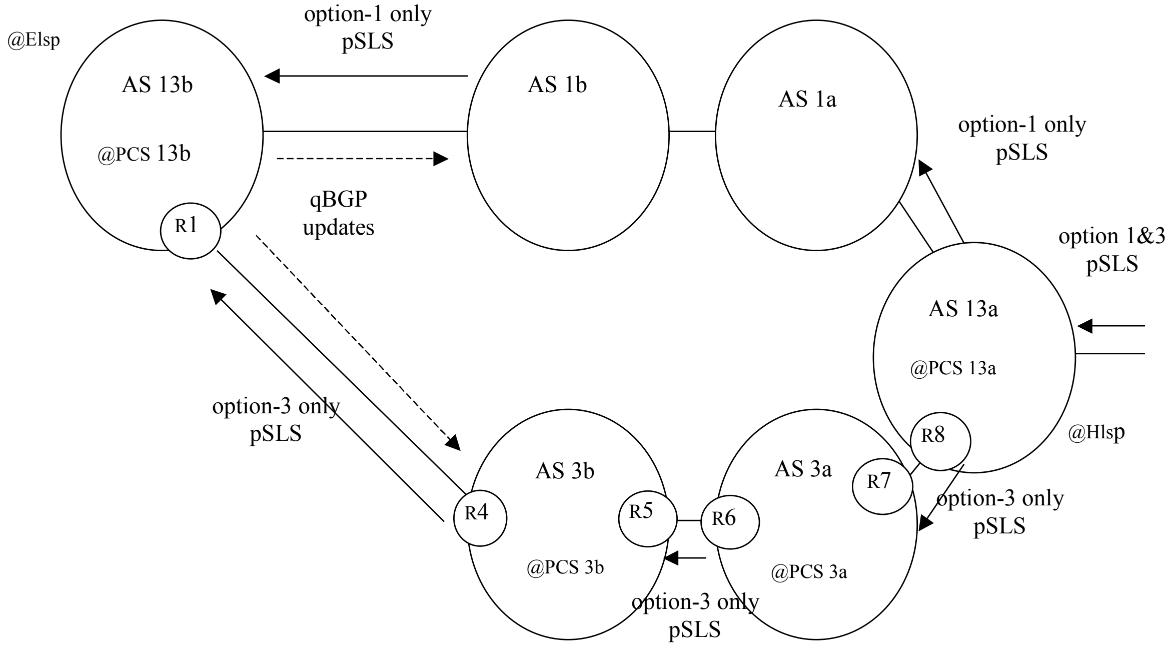


**Figure 36: PCS based routing for option 3.**

In Figure 36 routers "Rx" are ASBR.

This solution assumes the following:

- Each AS implementing option 3 has a PCS. This PCS is associated with an identifier that is unique in the whole Internet. This identifier can be the IP address of the PCS (@PCS x). This PCS Id is considered as identifying the AS itself as well.
- At the service level, when an AS wants to establish an LSP with another AS, there must be an agreement between these two ASs. This agreement specifies both the tail-end address of the LSP and the PCS identifier of the terminating AS.

The principle of this solution is to use q-BGP to announce PCS identifiers across the Internet in order for "option-3 only" and "option 1&3" ASs to be able to discover a path towards a set of AS supporting solution option-3. q-BGP updates including a PCS identifier use the community attribute with a specific value called PCSID. An ASBR must filter these updates and forward them only to external peers with which an "option-3 only" or "option 1&3" *pSLS* has been established. "Option-1 only" ASs must never receive an update with PCSID community attribute. It could be considered as a mean of PCS discovery to announce PCSID updates to "option-1 only" ASs. But "option-1 only" ASs must not forward these updates to any other external peers. Such announcement is relevant for a given *Meta-QoS-class* and contains QoS performance characteristics computed in same way as for solution option 1.

Following is an example using Figure 36. According to pre-established *pSLS*s, AS 13b advertises its PCS Id to AS 3b using q-BGP. Then, the @PCS 13b is forwarded to AS13a via AS3b and AS3a.

A PCS is an internal q-BGP peer. It receives all q-BGP updates, but it has to deals only with updates that include the PCSID community attribute.

In this example, AS 13a wants to establish an LSP between one of its address @Hlsp and an address belonging to AS 13b, @Elsp. An agreement is made between AS 13b and AS 13a. They agree on the LSP tail-end addresses and the PCS identifiers (@PCS 13a and @PCS 13b). The process is as follows:

- The PCS of AS 13a calculates the best path towards @PCS 13b using the internal q-BGP updates that include the PCSID community attribute. The result of its calculation for the best route is to use the BGP Next-Hop R7. Thanks to its IGP-TE, the PCS can deduce that R8 is the corresponding ASBR. So, it sends a request to PCS of AS 3a to build a LSP on its behalf from R7 to @Elsp of @PCS 13b.
- The PCS of AS 3a calculates the best path towards @PCS 13b using the internal q-BGP updates that include the PCSID community attribute. The result of its calculation for the best route is to use the BGP Next-Hop R5. If the IGP-TE of AS 3a can create an LSP section from R7 to R5 (via R6), it sends a request to PCS of AS 3b to build a LSP on behalf of @PCS 13a from R5 to @Elsp of @PCS 13b.
- The PCS of AS 3b calculates the best path towards @PCS 13b using the internal q-BGP updates that include the PCSID community attribute. The result of its calculation for the best route is to use the BGP Next-Hop R1. If the IGP-TE of AS 3b can create an LSP section from R5 to R1 (via R4), it sends a request to PCS of AS 13b to build a LSP on behalf of @PCS 13a from R1 to @Elsp of @PCS 13b.
- The PCS of AS 13b is the terminating AS for the requested LSP. After the authentication of the requester @PCS 13a, and if the IGP-TE of AS 13b can create an LSP section from R1 to @Elsp, the PCS of AS 13b replies affirmatively to PCS of AS 3b and include the path [@Eslp,R1] in the reply message.
- Then, PCS from AS 3b replies affirmatively to PCS of AS 3a and include the path [@Eslp, R1, R4, R5] in the reply message.
- Then, PCS from AS 3a replies affirmatively to PCS of AS 13a and include the path [@Eslp, R1, R4, R5, R6, R7] in the reply message.
- Then, the PCS of AS 13a adds at the end of the path [R8, @Hlsp]. AS 13a can now ask for a RSVP reservation using the reverse path [@Hlsp, R8, R7, R6, R5, R4, R1, @Elsp].

Note that the amount of bandwidth to be reserved is also signalled.

If, during this process, a PCS can't find a path to the @PCS destination, it replies negatively to the requesting PCS. Then, the requesting PCS must calculate another path towards the destination, and try another PCS. This means that the PCS must keep track of all internal q-BGP updates it receives in order to be able to calculate alternative paths.

With this solution, LSP tail-end addresses @Hlsp and @Elsp don't need to be advertised by the q-BGP. The LSP source AS doesn't need to know how to route to @Elsp from an IP point of view. It just needs to be able to calculate an AS-path towards the PCS Id of the destination AS. This is done by using the q-BGP updates including the PCSID community attribute. Therefore, even though PCS Ids are IP addresses, which included in q-BGP updates, there are of no interest to IP routing. This means that PCSID updates are processed by BGP in the same way as other updates, except that these are not included in routers' RIBs. Those updates are used only by PCSs.

As far as IGP routing is concerned, it is important to emphasise that the IGP must know how to reach the BGP Next-Hop (the external BGP peer) in order to calculate its internal TE path.

### 5.6.3 The statistical and the hard solution options

This inter-working scenario could be seen as an extension of the statistical service option over a hard solution option.

Since the hard solution option is technically based on the loose solution option, the problems to solve are the same as those identified in the previous section. In particular, the signalling and the multiple service options management problems need to be taken into account.

## 5.7 Conclusions

This chapter discussed the inter-working of the three MESCAL solution options from a service and a technical point of view. Two main scenarios have been successively examined: the first scenario has examined the co-existence of distinct MESCAL solution options deployed within the same AS, while the second one focused on the extension of the scope of a given MESCAL solution option through a domain that supports different solution option(s). Issues encountered in the aforementioned scenarios have been highlighted and solutions proposed.

The discussions led us to identify major issues/problems raised by the inter-working of different solution options when one/more provider/s attempt to deploy more than one service option in a given AS. These issues/problems are listed as below:

- The signalling issue/problem
- The MSOM issue/problem
- The bandwidth management issue/problem.

Coexistence and inter-working considerations are complex and, in order to solve the aforementioned issues, we needed to introduce the following set of features hereafter reminded:

1. Disjoint inter-domain announcement features (for both q-iBGP and q-eBGP).
2. Disjoint intra-domain routing spaces, per service option.
3. Multiple inter-domain bandwidth management "*philosophies*" (*Meta-QoS-class* based, o-QC based, LSP based).
4. Enhanced management systems that would be able to handle simultaneously several service options within a domain.
5. Appropriate (but approximate) methodologies in order to "inject" an o-QC into an *Meta-QoS-class*, and vice-versa
6. Specific adaptation functions that an ASBR will have to support when involved in an inter-working peering context.

MESCAL service and protocol specifications will take into account these requirements. Some of these requirements will have a direct impact on protocol and algorithm design and implementation while others will constraint the activation and the configuration of some network functions (Bandwidth management for instance).

# CHAPTER 6: BUSINESS RELATIONSHIPS & FINANCIAL SETTLEMENTS

## 6.1 Introduction

The MESCAL solution to the problem of QoS-based service delivery in the Internet, across different network provider (NP) domains, adopts a hop-by-hop, cascaded model for the interactions between NPs, which are seen both at the service and network (IP) layers. Interactions at the service layer aim at the establishment of agreements for QoS traffic exchange, pSLSs in MESCAL terminology, to allow NPs to expand the reach of their offered QoS-based services beyond the boundaries of their domains. Interactions at the IP layer are required to enable NPs to find, determine and maintain suitable QoS routes for forwarding traffic in the Internet, beyond their boundaries. In addition to appropriate protocols for supporting these interactions, the MESCAL solution prescribes the required service management and traffic engineering functionality per NP domain, to gracefully and effectively meet the requirements emerging from these interactions, while optimising the utilisation of the network resources. Chapter 1 discussed and analysed the strengths and limitation of the two major approaches for service interactions between NPs (referred to as peering approaches), cascaded versus centralised.

Driven by the different levels of QoS guarantees on packet transfer performance targets and bandwidth that could be provided to services - loose, statistical and hard QoS guarantees - three corresponding technical options of the general MESCAL solution have been specified. As such, each solution option suits the needs of different service types, therefore targeting different customer/user segments. Furthermore, each option approaches the inter-domain QoS problem from a different angle, and its required functionality pertains to different levels of operations complexity and scalability. Solution option 3, which could also be viewed as an add-on feature to the other two solution options, is suitable for services requiring hard QoS guarantees but with the inherent limitation that cannot scale to the size of the Internet users/destinations. Following the aggregate philosophy of Diffserv networks, solution option 1 has been designed to provide for loose (qualitative) QoS guarantees across the Internet, while solution option 2 delivers statistical guarantees (i.e. not per flow but per flow aggregates) for quantitative QoS targets, in addition to qualitative QoS guarantees. The technical targets, aspects and constraints of the three MESCAL solution options have been presented in [D1.1], while suitable protocols and algorithms are described in [D1.2].

This chapter explores the viability of the proposed MESCAL solution and the associated options in terms of their applicability from business (not technical) perspectives. Specifically, as service accounting, billing and marketing aspects are outside the scope of MESCAL, viability from business perspectives is addressed at the level of business relationships between NPs and related financial settlements for exchanging QoS traffic; accounting and data collection methods, charging, rating and pricing models are not addressed.

Such an analysis is useful for validating the MESCAL approach and the proposed solutions from yet another angle, that of business viability. Furthermore, it is a prerequisite for the deployment of the MESCAL solutions. The analysis is based on evident assumptions and requirements as well as on paradigms from current practice.

The chapter is organised as follows. First, the business relationships and financial settlements in today's best-effort Internet are presented. Then, these aspects are discussed from the perspectives of the MESCAL solution, drawing also implications for the MESCAL pSLSs.

## 6.2 Current Business Practices in the Internet

### 6.2.1 Internet Connectivity and Business Relationships

The global Internet is a collection of independently operated networks that have been organised into what is considered to be a *three-tiered* hierarchy [HUST]. The connectivity and position in the tier model is dependent on the size of the ISP (Internet Service Provider)[2], it's geographic reach, capacity - in terms of link speeds and routing capability - and the available reachable prefixes. This three-tier model is shown in Figure 37.

Tier 3 ISPs typically cover cities or regions of a country and as they cannot afford their own national links, they must interconnect with Tier 2 or other Tier 3 ISPs. Tier 2 ISPs usually cover entire countries but do not cross international boundaries. Tier 2 ISPs provide transit for traffic between their customers, i.e., Tier 3 ISPs. For international connectivity, Tier 2 ISPs normally connect to Tier 1 ISPs that span more than one geographical zone (e.g., country). Therefore, Tier 3 ISPs are typically the ones with end users and the larger ISPs only have other ISPs as their customers, however very large organisations could also be the customers of Tier 2 ISPs (not depicted in the Figure 37).

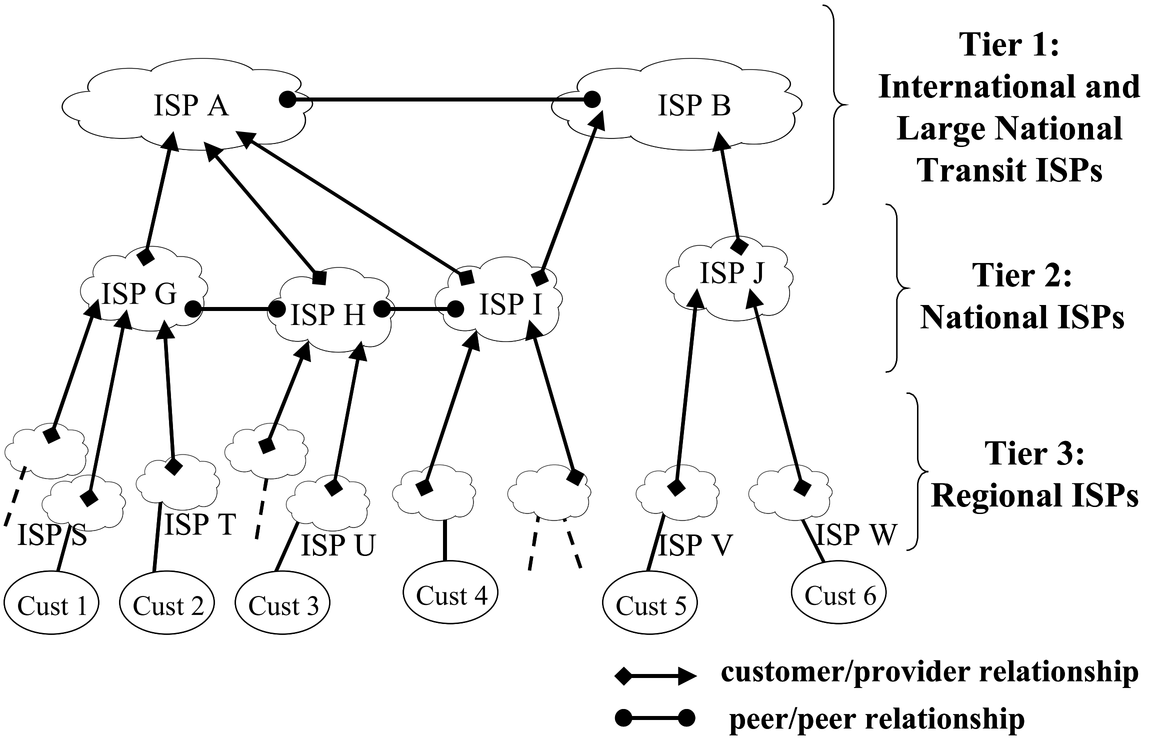


**Figure 37: Three-Tier Internet model with peering/transit agreements.**

Business relationships between ISPs generally fall into two categories:

- *customer-provider*, and
- *peer-to-peer*.

In the customer-provider relationship, a provider ISP provides the Internet connectivity service to its customer ISP. Usually, this type of business relationship is between ISPs belonging to different levels of the three-tier Internet model, with the ISP in the lower tier being a customer of the ISP in the upper tier.

[2] The term ISP, used heavily today, is used throughout in this section in an equivalent meaning to the term NP (network provider) used in other sections of this chapter and the document as a whole.

The peer-to-peer relationship is a kind of 'short-cut' to prevent traffic flowing into the upper tiers and allows for the direct flow of traffic between the peer-to-peer ISPs. Usually, this type of business relationship is between ISPs of similar size (belonging to the same tier). Predominantly, a peer-to-peer relationship can be taken as a non-transitive relationship. Peer-to-peer ISPs reciprocally provide only access to each other's customers, i.e. peer-to-peer ISPs mutually agree to exchange traffic between themselves, not transiting traffic to their providers or to other peer-to-peer ISPs.

Today's Internet follows a three-tier hierarchical business model, with the business relationships between ISPs being completely determined by their relative positioning in the hierarchy. Tier 1 ISPs are purely transit networks that do not source or sink their own traffic and only transport traffic for customer ISPs while having a peer-to-peer relationship with other Tier 1 ISPs. Tier 2 ISPs are customers to Tier 1 ISPs, while they could also have peer-to-peer relationships with adjacent Tier 2 ISPs. Last, Tier 1 ISPs can only be customers to Tier 2 ISPs.

The above business relationships occur between adjacent (directly connected) ISPs. The decision to interconnect two ISPs and the type of interconnection is dependant on geography, the available physical layer, the ISP's demand matrix and the size of the two ISPs (level of the three-tier architecture they can be positioned in), which in turn affected by and various other reasons, such as economical/political conditions, business viability and market penetration/potential.

## 6.2.2 Financial Settlements

The financial settlements between ISPs primarily depend on their business relationship, which in turn depends on their positioning in the three-tier Internet model as outlined in the previous section. They are varied and often considered confidential, but they can fall into three main categories as shown in Table 12 and described in the following sections:

| Type of business relationship | Type of financial settlement |
|---|---|
| customer-provider | service-provider settlement (cf. section 6.2.2.1) |
| peer-to-peer | negotiated-financial settlement (cf. section 6.2.2.2)<br>SKA (sender-keeps-all) settlement (cf. section 6.2.2.3) |

**Table 12: Financial settlements currently in the Internet.**

### *6.2.2.1 Service-Provider Settlement*

Under the service-provider settlement, a customer (end-customer or ISP) pays a flat rate or a usage-based amount to the provider ISP for reachability to networks, which the provider ISP can reach through its peers, customers or through its own provider ISPs (although if the customer is an ISP then it is rare for an ISP to transit traffic from a peer ISP into a provider ISP). Usage-based payment is typical for ISP to ISP agreements and sometimes for ISP to end-customer agreements, where the amount paid to the provider ISP is a function of the amount of traffic sent or received, the time of day and possibly other traffic attributes. The customer will always pay whether the traffic is being sent or received (i.e. the customer must still pay the provider for totally symmetric traffic). This type of settlement usually applies in customer-provider business relationships.

### *6.2.2.2 Negotiated-Financial Settlement*

Under the negotiated-financial settlement, the traffic volume in each direction is monitored and then payment must be made on the net flow of traffic. This settlement is usually used in peer-to-peer business relationships, where the ISPs are of a similar size. It is generally thought that even if the traffic is not balanced in each direction, it could still be cheaper than a service provider settlement with the ISP in the above tier.

#### *6.2.2.3 Sender Keeps All Settlement*

Under the SKA (Sender-Keeps-All) settlement both ISPs do not pay each other for traffic exchange, and usually split the physical layer costs between them. Such a settlement is usually used in peer-to-peer business relationships, especially between tier 2 ISPs whose customers send traffic destined for the peer ISP and therefore there is a cost saving because this traffic is not sent into an upper tier ISP that must be paid based on a service provider settlement. This settlement is really a special case of the negotiated financial settlement, where the net traffic flow is considered to be zero, because either the traffic is symmetric or because the perceived gain to each party is considered worth the agreement.

### 6.2.3 Summary – Flow of Traffic and Money

Due to the policies of ISPs and the financial settlement agreements they use, there are a number of restrictions on the flow of traffic between networks. Traffic entering an ISP destined for a network that is not directly connected to it, must be forwarded to the ISP in the tier above, unless the destination network is a customer of an ISP that has a peer-to-peer business relationship with that ISP. Therefore, if a customer, say customer 1 in Figure 37, wants to reach another customer, say customer 3 in Figure 37, ISP G may decide to forward the traffic to ISP H rather than ISP A. However, when there is an SKA agreement between ISP H and ISP G, then ISP H cannot act as a transit network for ISP G to reach Customer 4 from Customer 1. Instead, ISP G must forward the traffic to ISP A, which will then forward to ISP I. This is a policy decision based purely on financial compensation considerations, as transiting traffic under an SKA agreement will probably end up costing more, as it will finally reach a link with a non-SKA agreement.

An important aspect to note about the three-tier model of the Internet, is that the customer-provider business relationship applies generally along the vertical axis and, as such, there is a net flow of money from the bottom up i.e. customers always pay their provider in tier above and once traffic reaches the highest required tier 1 ISP, then the traffic will go against the flow of money.

## 6.3 MESCAL Business Relationships and Settlements

### 6.3.1 MESCAL QoS Business Cases and Relationships

First, it should be made clear that the purpose of the business cases, relationships and settlements emerging from the MESCAL solution is for QoS provisioning in the Internet. By viewing QoS as an add-on feature to plain connectivity services, the MESCAL implied business model, relationships and settlements should only be seen in addition to those holding today in the best-effort Internet, as already outlined in the previous section.

Based on its hop-by-hop, cascaded approach for interactions between providers to the end of QoS-based service delivery in the Internet, the MESCAL solution advocates two basic business cases. The business cases are differentiated by the type of QoS-based services that could be offered in the Internet. Specifically:

(A) a business case for the provisioning of QoS-based services relying only on loose QoS guarantees - qualitatively expressed performance targets, and no bandwidth guarantees.

(B) a business case for the provisioning of QoS-based services relying on statistical guarantees for quantitative performance targets and bandwidth, in addition to qualitative QoS guarantees.

It should be noted that in either of the above business cases, services relying on hard QoS guarantees could also be provided.

The two MESCAL business cases bear their own business interactions and financial settlements. These are described in the following sections. Table 13 depicts the applicability of the proposed MESCAL solution options per case.

| MESCAL business case | MESCAL solution option |
|---|---|
| A -Provisioning Internet QoS-based services with loose QoS guarantees | primarily 1[3], but also 3 for services with hard QoS guarantees |
| B - Provisioning Internet QoS-based services with statistical QoS guarantees | primarily 2, but also 3 for services with hard QoS guarantees |

**Table 13: MESCAL business cases and solution options.**

### *6.3.1.1 Business Case A – Provisioning of QoS-based Services with Loose QoS Guarantees*

This business case directly corresponds to the hierarchical, three-tier business model of the Internet as it stands today, with the business relationships and related financial settlements being as described in the previous section. Obviously, in this case the business agreements between the providers should be underlined by appropriate MESCAL pSLSs for allowing the exchange of QoS traffic. Figure 38 (compare with current practice, shown in Figure 37) depicts this business case of a MESCAL-enabled QoS Internet.

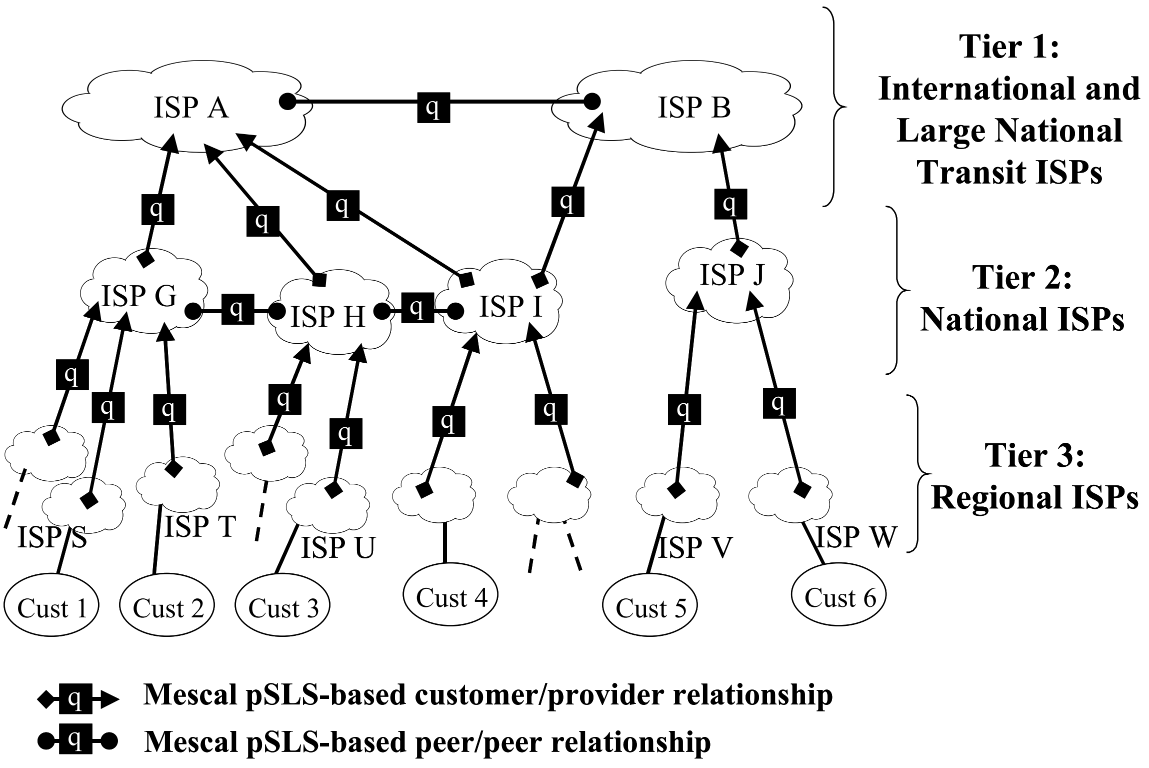

ISP A
ISP B
q
Tier 1: International and Large National Transit ISPs
ISP G
ISP H
ISP I
ISP J
Tier 2: National ISPs
Tier 3: Regional ISPs
ISP S
ISP T
ISP U
ISP V
ISP W
Cust 1
Cust 2
Cust 3
Cust 4
Cust 5
Cust 6
Mescal pSLS-based customer/provider relationship
Mescal pSLS-based peer/peer relationship

**Figure 38: MESCAL business case A – provisioning of services with loose QoS guarantees.**

### *6.3.1.2 Business Case B - Provisioning of QoS-based Services with Statistical QoS Guarantees*

In this business case, the Internet is considered *flat* as opposed to being hierarchical, as in the previous case. The term flat is meant twofold: no hierarchy and common (not distinct) business relationships. Specifically, although ISPs can still be categorised into a three-tier hierarchy, in the flat Internet, this hierarchy remains at a logical level and it does not influence or dictate the nature of the business relationships between ISPs. ISPs can have business relationships with any ISP (provided they can be

[3] A suitable version of solution option 2 (for qualitative QoS-based services) could also apply in the first business case; similarly, solution option 2 could also apply in the second business case, only for qualitative QoS-based services.

physically interconnected) they may deem appropriate to interact with for expanding the reach of their QoS-based services, despite the tier they may happen to reside. And, the business relationships between ISPs are all the same, following the so-called *upstream-QoS-proxy* or just *QoS-proxy* business relationship.

In the upstream-QoS-proxy relationship, either of the ISPs may agree with the other ISP to provide a transit QoS-based connectivity service to (a subset of) anywhere it can reach in the Internet with this QoS. The ISP offering the transit QoS service would have built its QoS reach capabilities based on similar agreements with (some of) its directly attached ISPs, which in turn would have built their own QoS reach capabilities based on similar agreements with (some of) their own adjacencies and so on. Therefore, each ISP in a chain of QoS-proxy relationships established in the same direction appears as kind of a 'proxy' of the ISPs further along this direction.

Figure 39 (compare with previous business case, shown in Figure 38, and with current practice, shown in Figure 37) depicts this business case of a MESCAL-enabled QoS Internet.

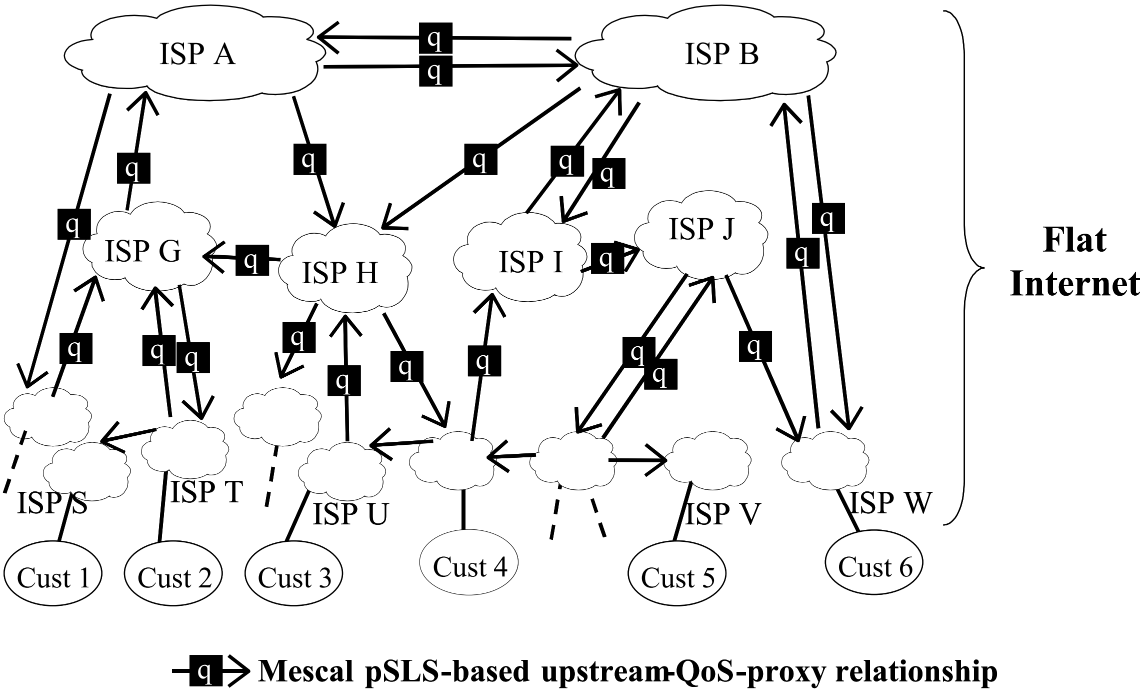


**Figure 39: MESCAL business case B –provisioning of services with statistical QoS guarantees.**

Two things are worth noting about this business relationship. First is its liberal, 'reflective' nature. As an ISP could use the other ISP as a QoS transit (proxy) to the Internet, so the other ISP could use this ISP as a QoS transit. Furthermore, a number of agreements could be established between ISPs under such a business relationship. Second, its strong collaborative and transitive nature, which is built in a cascaded fashion. ISPs can transit QoS traffic anywhere they can get in the Internet under QoS constraints, by alternately combining similar capabilities down the direction of the QoS traffic. Of course, QoS reachability is interrupted when suitable QoS-proxy agreements cannot longer be established e.g. an ISP does not offer QoS connectivity in its domain or even if it does, it is not willing to participate in a chain of QoS delivery.

The QoS-proxy business relationship differs from the business relationships holding in the current best-effort Internet in the connotation of the established agreements on traffic exchange and subsequently in the directionality of the traffic flows. In customer-provider relationships, agreements are established for transporting traffic from/to the customer ISPs to/from the provider ISPs and in the peer-to-peer relationships, agreements are established for the ISPs to exchange traffic on a mutual basis; in QoS-proxy relationships, agreements may be established independently in either way, as each ISPs wishes, and so does traffic flow.

All together, the flat Internet business model as described above is quite different from its three-tier hierarchical counterpart. This is because of the relaxation of the hierarchy. In the flat Internet, traffic can flow freely, under QoS constraints, between providers bound to QoS-proxy relationships. In the hierarchical Internet, the flow of traffic is constrained by the different types of business relationships, which are dictated by the relative positioning of the ISPs in the hierarchy. For instance, in the hierarchical Internet, transit services are mainly provided through customer-provider relationships (going one tier up) and not through peer-to-peer relationships (through ISPs in the same tier), whereas, in the flat Internet, any ISP may provide transit services, by definition. It should be noted that the type of business relationships between ISPs and associated peering policies have an impact on inter-domain routing e.g. route stability, router memory and processing capabilities and protocol overhead. Therefore, such issues need to be studied carefully, before moving to a new Internet business model (such studies have been undertaken by MESCAL).

The differences between the flat and the hierarchical Internet business models, as discussed above, are attributed to the diverse types of services each model is set-up to provide. In a best-effort or loose-QoS-based connectivity service Internet, geographical coverage is clearly the strongest selling point; hence, the hierarchical business model and the customer-provider and peer-to-peer relationships. In a statistical-QoS-based service Internet, provided that there is a global demand for related services, the ability to provide such QoS becomes clearly an asset, even prevailing to geographical coverage. As such, what is required, is to seek for suitable QoS peers to deliver such QoS everywhere in the Internet is needed e.g. at home; at the desktop; hence, the MESCAL QoS-proxy relationship, which equally applies to all ISPs - regardless of their size.

Finally, the following observation is worth making. The flat QoS-aware Internet business model resembles the current practice in the traditional telecom service world (PSTN) and VoIP. The QoS-proxy business agreements could be seen as corresponding to call-termination agreements established between telephony operators/providers. Furthermore, in the telco's world, synergies between operators are built mainly on grounds of reliability and competitiveness, much like as the flat Internet business model implies. This resemblance is not surprising; telephony calls and services based on statistically guaranteed quantitative QoS metrics, provided that there is a global demand for them, are very similar in that they are both commodities, which need to be widely offered at a certain quality.

### *6.3.1.3 Mixture of Business Cases*

The MESCAL business cases presented in the previous sections are not exclusive. As they are both built in addition to the current model of the best-effort Internet, they could co-exist, forming in the Internet segments (islands), each providing for different types of QoS; from best-effort and qualitative to statistically guaranteed QoS. Figure 40 depicts such a mixed business case. Note that the co-existence of the MESCAL business models from a technical perspective is analysed in Chapter 5.

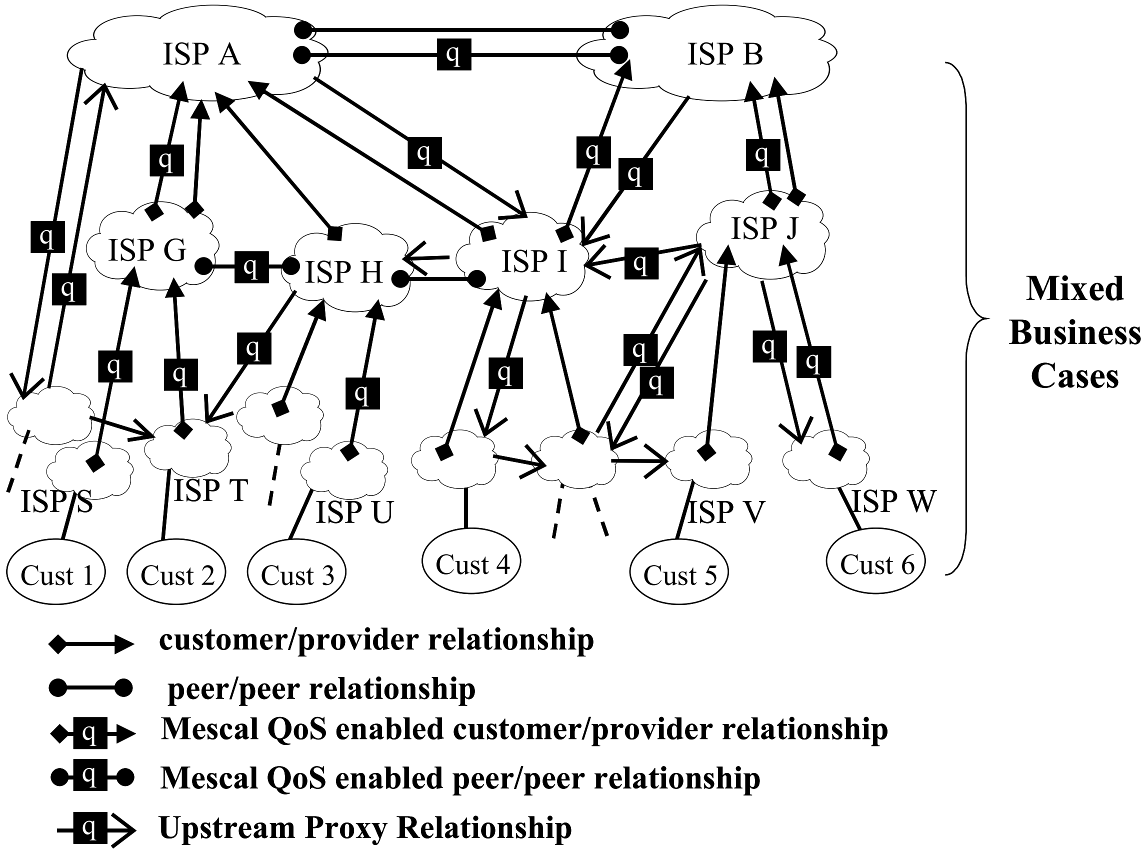

ISP A
ISP B
ISP G
ISP H
ISP I
ISP J
ISP S
ISP T
ISP U
ISP V
ISP W
Cust 1
Cust 2
Cust 3
Cust 4
Cust 5
Cust 6
q
Mixed Business Cases
customer/provider relationship
peer/peer relationship
Mescal QoS enabled customer/provider relationship
Mescal QoS enabled peer/peer relationship
Upstream Proxy Relationship

**Figure 40: Business models towards a QoS-aware Internet.**

### 6.3.2 Implications on pSLSs

The business agreements between ISPs in the previously identified MESCAL business cases for a QoS-aware Internet should be underlined by appropriate MESCAL pSLSs for allowing the exchange of QoS traffic. The type of business relationships affects certain characteristics of the pSLSs, which are discussed below.

In the customer-provider business relationship, pSLSs are always requested by the customer ISP. They have the connotation of agreements for the customer ISP to 'join in' (send and receive traffic) the QoS-aware Internet as seen by the provider ISP. Such pSLSs are bi-directional and can only offer qualitative QoS guarantees for all destinations that can be reached by the provider ISP as well as hard QoS guarantees.

pSLSs between peer-to-peer ISPs may be requested by either of the ISPs. They have the connotation of mutual agreements for the exchange of QoS traffic from one ISP domain to the other ISP domain. Such pSLSs are bi-directional and, according to the QoS context they apply, offer qualitative or hard QoS guarantees within the scope of the ISP domains.

In the upstream-QoS-proxy business relationship, pSLSs may be requested by either ISP. In this respect, we distinguish between the client ISP requesting the pSLS, and the server ISP offering and agreeing to provide the pSLS. Such pSLSs have the connotation of agreements for the server ISP to deliver QoS traffic from the client ISP to (a subset of) the destinations that can be reached from the server ISP with this QoS. They are unidirectional, in the direction from the client to the server ISP and may offer qualitative and/or statistical or hard quantitative QoS guarantees to certain destinations in the Internet (those reachable by the server ISP). According to the nature of this relationship, an ISP can be a client of the other ISP, while at the same time is a server for the other ISP.

### 6.3.3 Financial Settlements

First, it should be made clear that the financial settlements in a MESCAL-enabled QoS-aware Internet, which are discussed in this section, are considered in addition to the settlements made for best-effort connectivity.

In a MESCAL QoS-aware Internet, the exchange of QoS traffic between ISPs is financially settled on the basis of pSLSs that the ISPs have established. Broadly speaking, the following two principles govern these settlements:

- The ISP who requests the pSLS pays. Considering that QoS is a commodity, this has the intuitive connotation of 'customer always pays'.
- In the cases where either of the ISPs may request pSLSs from the other ISP, or the pSLSs have the connotation of mutual agreements, in addition to the case of each ISP paying the other ISP, against pSLSs, payment reconciliation may take place a priori, as part of a business agreement.

Based on the above, the financial settlements in the current best effort Internet apply also in the MESCAL QoS-aware Internet. Table 14 depicts the financial settlements per MESCAL business relationship.

| Type of business relationship | Type of financial settlement |
|---|---|
| MESCAL pSLS-based customer-provider | service-provider settlement (cf. section 6.2.2.1) |
| MESCAL pSLS-based peer-to-peer | negotiated-financial settlement (cf. section 6.2.2.2)<br>SKA (sender-keeps-all) settlement (cf. section 6.2.2.3) |
| MESCAL pSLS-based QoS-proxy | service-provider settlement (cf. section 6.2.2.1), if only one ISP requests pSLSs<br>negotiated-financial settlement (cf. section 6.2.2.2), if ISPs request pSLSs from each other<br>SKA (sender-keeps-all) settlement (cf. section 6.2.2.3), if ISPs request pSLSs from each other |

**Table 14: Financial settlements in the MESCAL QoS-aware Internet.**

### 6.3.4 Summary – Flow of Traffic and Money

Two business cases have been identified for a MESCAL-enabled QoS-aware Internet; one for providing services based only on qualitative QoS guarantees and one for additionally providing services based on statically guaranteed quantitative QoS metrics. In both cases, services relying on hard QoS guarantees could also be provided, however not for the mass market because of scalability limitations inherent in the technical solution.

The qualitative-QoS Internet business case directly corresponds to the three-tier, hierarchical model currently in place, whereas the statistical-QoS Internet business case advocates a flat Internet, where the business relationships between ISPs are of the same type, which is not affected nor dictated by the tier levels the ISPs may reside. This type of business relationship is of a strong transitive nature and could be thought as being the QoS Internet counterpart of a call-termination agreement in the PSTN and VoIP business world. The differences between the hierarchical and the flat Internet business models have been outlined and discussed from QoS delivery perspectives in section 6.3.1.

In the flat Internet, the net flow of money always follows the flow of traffic. In the hierarchical Internet, assuming that a tier 1 ISP must always be involved, the net flow of money follows the flow of traffic until a tier 1 ISP is reached, at which point on, the net flow of money goes against the traffic.

Whether the MESCAL QoS business models will be deployed in the Internet, and which ones will prevail, primarily depends on the current and emerging conditions and norms of the global service market e.g. on whether there will be demand for QoS-based services and for which type of QoS guarantees.

Finally, the following point is worth noting. The identified, MESCAL pSLS-based, business relationships and related financial settlements are in addition to those currently required for best-effort

Internet connectivity. As such, both MESCAL business cases can be intermixed, while they can also coexist with the current business practices in the best-effort Internet.

In conclusion, MESCAL does not distort the current business practices in the Internet and advocates a safe, incremental and best-effort compatible migration path towards a QoS-aware Internet.

# REFERENCES

[ASON] ITU-T Draft Recommendation G.8080/Y.1304

[CISCO] "Cisco ONS 15454 SONET and Cisco ONS 15454 SDH multiservice provisioning platforms" http://cisco.com/en/US/products/hw/optical/ps2006/index.html

[CIENA] Ciena CoreDirector Intelligent Optical Core Switches: "Intelligent, standards–based (GMPLS, G.ASON) control plane, dynamic point-and-click provisioning, per-circuit priority, arbitrary concatenation, automated grooming, SONET/SDH Gateway", http://www.ciena.com/products/coredirector/coredirector_303.htm

[CRLDP] P. Ashwood-Smith and L. Berger, ed., IETF RFC 3472 "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Constraint-based Routed Label Distribution Protocol (CR-LDP) Extensions"

[D1.1] Paris Flegkas, et al., "D1.1: Specification of Business Models and a Functional Architecture for Inter-domain QoS Delivery", http://www.mescal.org/, June 2003.

[D1.2] Michael Howarth, et al., "D1.2: Initial specification of protocols and algorithms for inter-domain SLS management and traffic engineering for QoS-based IP service delivery and their test requirements", http://www.mescal.org/, January 2004.

[ETSI-300820] ETSI EN 300820-1, "Telecommunications management Network (TMN); ATM Management information model for the X-interface between Operation Systems (OS) of a Virtual Path (VP)/Virtual Channel (VC) cross connected network; Part 1: Configuration management", (Enhanced standard text submitted to ETSI for public enquiry, October 1999).

[GLV] IEEE Communications Magazine May 2002 Vol. 40 No. 5[GMPLS] E. Mannie, ed., "Generalized Multi-Protocol Label Switching Architecture" IETF Internet draft, draft-ietf-ccamp-gmpls-architecture-07.txt.

[HUST] G. Huston, "ISP Survival Guide", John Wiley and Sons 1998. ISBN 201-3-45567-9

[IEEE-VLANs] IEEE standard for Virtual Local Area Networks, IEEE 802.1Q, 1998.

[IXP] World-wide Exchange Point Information: http://www.ep.net/ep-main.html

[Govinden] Ramesh Govindan, Cengiz Alaettinog-lu, Kannan Varadhan, and Deborah Estrin, "Route servers for inter-domain routing", Computer Networks and ISDN Systems, 30(12), pp.1157-1174, 1998.

[LCNT] Lucent LambdaXtreme Transport http://www.lucent.com/

[P813-D1] P813-PF**,** "Technical Development and Support for European ATM Service Introduction", Deliverable 1, Guidelines to ETSI TS 101 674-2 "Technical Framework for the Provision of Interoperable ATM Services; Network Management (X-interface) Specification for Phase 1 Implementation", February 2000.

[P1008-D2] Project P1008, "Inter-operator interfaces for ensuring end-to-end IP QoS", Deliverable 2, Selected Scenarios and requirements for end-to-end IP QoS management, January 2001.

[Raveendran] Barry Raveendran Greene, "L2 Internet eXchane Point (IXP) using a BGP Route Reflector, Technical Design, Configuration, and General Advice about IXPs", Draft version 4, Dec., 2000, http://macross.dynodns.net/idr/L2_Route_Reflector_IXP_v0.4.pdf.

[RFC 1966] T. Bates, R. Chandra, "BGP Route Reflection An alternative to full mesh IBGP", Experimental RFC, June 1996.

[RFC 1863] D. Haskin, "A BGP/IDRP route server alternative to a full mesh routing", RFC 1863, October 1995.

[RS03] Route Server Technical Overview, http://www.rsng.net/overview.html.

[RSVPTE] L. Berger, ed., IETF RFC 3473 "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions".

[SALSA] S. Salsano, F. Ricciato, M. Listanti and A. Belmonte "Off-line Configuration of a MPLS over WDM Network under Time-Varying Offered Traffic", IEEE Infocom 2002, New York June 2002.