A New Perspective for a Global QoS-based Internet

Pierre Levis, Mohamed Boucadair, Pierrick Morand, Jason Spencer, David Griffin, George Pavlou, Panos Trimintzios

***Abstract*—Several Quality of Service (QoS) architectures and mechanisms have been proposed by standardisation bodies. However, those proposals were not always aligned with the spirit of IP networks. Furthermore, most research activities have focused to date on supporting QoS only within a single administrative domain. In this paper, we demonstrate that it is possible to provision QoS-enabled services that maintain the ability to easily connect any pair of users worldwide. For this purpose, we introduce the notion of Meta-QoS-Class and demonstrate its relevance for building QoS-enabled services across multiple domains. We then show how a QoS-enhanced Border Gateway Protocol (q-BGP) can be used together with Meta-QoS-Classes, for building a set of parallel Internet planes with different QoS capabilities. This concept opens up a new perspective for a global QoS-based Internet.**

***Index terms*—QoS, Traffic Engineering, Internet, Routing, inter-domain, Service Provider, SLS**

I. Introduction

Based on best current practice, we can hardly state that QoS has been deployed in Service Providers' networks for inter-domain and even intra-domain purposes. As such, the Internet remains an interconnection of best-effort networks, with the only global transport service usable throughout the Internet being the best-effort service. The reasons are manifold: premium QoS services are, at the moment, very difficult to realise [1]; most QoS research has an overly reactive nature [2]; and QoS deployment is definitely a complex process [3]. For example, today, there is no way for a video content provider to make it possible for their ready-to-pay customers to access the service, on a large scale, through guaranteed network performance.

We identify two types of customers that could be potentially interested in QoS over multi Service Provider domains. The first type includes corporate customers, who would use IP-based Virtual Private Network (VPN) services and request hard and permanent QoS guarantees for a limited set of well-known locations (VPN sites). Corporate behavior assumes that a number of Service Providers have mutual agreements to offer a QoS guaranteed transport service to connect the different sites of the given enterprise.

The second type of customers includes residential customers, who would request QoS in a more generic way, when they decide to use any service. This residential customer behavior assumes a large number of Service Providers have agreed on a global QoS infrastructure deployed all over the entire Internet.

The two scenarios described above are bound to differ in their respective solutions. QoS access across a limited set of domains is not the same type of problem as that of generic QoS access throughout the whole Internet; as such, there exist many QoS architectures [4]. So far, despite the fact that a lot of work has addressed QoS during the last decade, little has been undertaken in the way of addressing residential user needs. In this position paper we present a QoS solution exactly for this residential type of customers.

We do not consider all issues related to residential customers, we only address the problem of being able to reach any place, any time, through a QoS inter-domain routing infrastructure. This global QoS-based Internet solution should prevent QoS techniques and architectures from impairing the spirit on which the Internet has been devised [5]. The idea is, of course, not to exclude evolution just because the Internet should remain what it has been from the beginning, i.e. a best-effort only network - see [6] and [7] for reflections on design evolution. Our intention is simply to keep the features that form the fundamental basis on which to build QoS capabilities, open to the vast majority of citizens. These features should be broadly accepted and agreed on by all QoS actors. This community of actors is quite broad, encompassing: the user, the Service Provider, the vendor, the legislator, and so forth. It is important to understand that the priority is not necessarily on technical or financial considerations. We should preserve the facility to (1) spread Internet access, (2) welcome new services and applications and (3) communicate from any location to any other location.

Manuscript received May 28, 2004; revised May 09, 2005 and July 26, 2005. The paper was presented in part at the Conference on Software, Telecommunications and Computer Networks (SoftCOM) 2004.

P. Levis. M. Boucadair and P. Morand are with France Telecom R&D, Caen, France (e-mail: {pierre.levis, mohamed.boucadair, pierrick.morand}@francetelecom.com).

J. Spencer and D. Griffin are with University College London, London, UK (e-mail: {jsp,dgriffin}@ee.ucl.ac.uk).

G. Pavlou and P. Trimintzios are with University of Surrey, Guildford, UK (e-mail:{g.pavlou ptrimintzios}@eim.surrey.ac.uk).

On these grounds, the requirements are the following:

- IP networks should be ready to convey inter-domain QoS traffic before customers can initiate end-to-end QoS negotiations (just like inter-domain routing).
- A best-effort route must be available when no QoS route is known. Best-effort delivery must survive QoS and should remain the main Internet transport service.
- The inter-domain QoS delivery solution should not rely on the existence of a centralized entity that has the knowledge and the control of the whole Internet.
- It would be desirable for the potential solution to preserve the resilience feature of the current hop-by-hop IP routing approach.

A. Related work

EURESCOM P806 project [8] has proposed a QoS framework that takes into account inter-domain aspects. The proposal made in this paper and the P806 peoject share the idea that provider Service Level Specification (pSLS) should be negotiated only between adjacent providers. But differ in the way pSLS negotiations are activated. In P806, an end-user request sparks a chain of pSLSs to meet the QoS requirement. We deem this approach too dynamic to be scalable (same problem as the Resource ReSerVation Protocol (RSVP) approach). Our pSLSs are negotiated prior to end-users requests and can occur in any order.

IETF Differentiated Services (DiffServ) has stopped with Per Domain Behaviour (PDB) notion [9]. PDB is restricted to a single domain. There has been no inter-domain consideration in all the DiffServ work. Our work extends this effort and proposes a way to glue several DiffServ domains, and a way to circulate in this QoS infrastructure.

Bandwidth Brokers (BB) have been introduced in DiffServ architecture to manage bandwidth allocation within domains. The inter-domain aspect resides in relationships between adjacent BBs. BB could be deployed to manage the resources associated with Meta-QoS-Class planes which are introduced in this paper. Therefore, one BB could be placed on each domain in a given Meta-QoS-Class plane. Each BB would be responsible of allocating to end-users Meta-QoS-Class resources that have been negotiated in pSLSs. The Meta-QoS-Class plane would provide the QoS routing infrastructure.

Currently, there is no standardization effort for pomoting inter-domain related QoS archictetcures within the IETF. The recently charterd working group Path Computation Element (PCE) [10] focuses only on small chain of domains and does not enclose in its charter the QoS-specific issues but only traffic enginnering ones *a la* Multi-Protocol Label Switching (MPLS). In its side, Next Steps in Signalling (NSIS) working group treats signalling issues and does not investigate on specific inter-domain QoS problems [11].

B. Structure

This paper is an investigation towards a solution that meets the above requirements. We provide guidance for QoS services that are potentially accessible by the larger Internet community, regardless of the network access provider's location.

The remainder of this paper is organized as follows. Section 2 analyses the problem of end-to-end QoS based on agreements between Service Providers. Section 3 develops the concept of Meta-QoS-Class. Section 4 explains the use of Meta-QoS-Classes to build a QoS-enabled Internet. In this section, we also explain the design of the enhanced q-BGP protocol. Section 5 gives some experimental results and Section 6 some simulation results. Finally, After a conclusion in Section 7, Section 8 lists some possible new areas of research for inter-domain QoS facilitated by the introduction of the Meta-QoS-Class concept.

II. From a basic inter-domain QoS problem to the Meta-QoS-Class

A. Problem statement

We adopt a pragmatic view to tackle the problem of Internet QoS delivery. We consider the context where we have Service Providers (SPs) with QoS capabilities concatenated over the end-to-end path. These meshed QoS capabilities constitute the QoS infrastructure.

We start by closely examining the relevant requirements, opportunities and consequences for a given SP to integrate this QoS infrastructure. We focus mainly on SP-to-SP agreements rather than on SP-to-customer agreements.

Let's consider a given Service Provider that offers QoS-based services to its customers. The scope of these services is limited to its network domain boundaries. On the other hand, this Service Provider is aware that many other Service Providers, scattered over the Internet, also offer QoS-based services to their customers. This Service Provider is expected to want to benefit from the QoS infrastructure in order to expand its QoS-based service offerings to destinations outside its own administrative domain.

B. Reaching QoS agreements with neighbors

There are at least two main approaches for expanding one SP's QoS service to other SPs domains. In the first approach, the Service Provider negotiates agreements only with its immediate neighboring Service Providers [12]. This includes all the SPs that are directly accessible without the need to cross a third party SP. We call it the *cascaded* approach. In the second approach, the Service Provider negotiates directly with an appropriate number of downstream providers, one or more than one domain hop away. We call it the *centralized* approach [13].

There is a great deal of complexity and scalability issues related to the centralized approach, which represents a radical shift from current Internet practice. Therefore, we believe that the only realistic way forward is the cascaded approach. This is the approach we adopt in the rest of this paper.

C. Binding l-QCs

We assume that each SP domain implements QoS capabilities in order to provide QoS-based services. We use the term local-QoS-Class or l-QC to denote a basic QoS transfer capability within a SP domain. A l-QC is characterized by a set of attribute-value pairs, see Table 1 for an example, where the attributes express various packet transfer performance parameters such as (D, J, L): one-way transit delay (D), one-way transit variation delay (also known as jitter) (J) and packet loss rate (L). The provisioning of a l-QC solely relies upon engineering policies deployed within the domain. Typically, a combination of the elementary IP DiffServ QoS capabilities with traffic-engineering functions should ensure the performance of l-QCs. A l-QC is one occurrence of a PDB.

TABLE 1
EXAMPLE OF QOS-CLASS ATTRIBUTE-VALUE PAIRS

Attribute	Value
One-way Delay (D)	10 ms
One-way delay variation – Jitter (J)	0.1 ms
Packet Loss Rate (L)	10e-3

On a physical level, the QoS service extension to a domain owened by another SP signifies the l-QC extension outside the scope of a single domain. In particular, this means that packets from a flow originated in a domain, with a given DiffServ Code Point (DSCP) indicating a given l-QC, should experience a similar treatment when crossing the set of domains on the path towards its destination.

Two l-QCs from two neighboring SP domains are bound together when the two SPs have agreed to transfer traffic from one l-QC on the upstream domain to another l-QC on the downstream domain [14]. Then, if we assume that a Service Provider knows l-QCs capabilities advertised by its service peers, the basic technical question that this provider has to face is: on what basis shall I bind my l-QC to my neighbor SP l-QCs? Given one of my own l-QCs, which is the best match? Based on which criteria?

D. Limiting the scope of SP-to-SP agreements

In this section we look into the problems of SP-to-SP agreements that guarantee QoS over a chain of downstream domains.

Let’s assume that SPn knows from its neighbor SPn-1 a set of (Destination, D, J, L) where Destination is a group of IP addresses, and (D, J, L) is the QoS performance to get from SPn-1 to Destination. SPn uses this information to bind its own l-QCs with SPn-1 l-QCs. SPn knows the QoS performance of its own l-QCs and therefore, deduces the QoS performance it could guarantee to its customers in order to join Destination. If this is a viable service and business opportunity, SPn will buy from SPn-1 the (Destination, D, J, L) that best fits its operational objectives.

End-to-end QoS performance is guaranteed in a recursive manner: SP1 guarantees QoS performance for its own domain crossing; while for a given n, SPn guarantees SPn+1 QoS performance for the crossing of the whole chain of SPs (SPn, SPn-1, …, SP1).

In this model, when a Service Provider contracts an agreement with a neighbor SP, a large number of other SP-to-SP and SP-to-customer agreements are likely to rely on that single agreement if it happens to be part of the chain of Service Providers. Any modification in that agreement is likely to have an impact on numerous external agreements that use it. The problem that arises here, is that you are not free to reconsider your own agreements, because other Service Providers, that you may have not even heard of, include this agreement in their own agreements.

We call *SP chain trap* the fact that the degree of freedom to renegotiate, or terminate, one of your own agreements is restricted by the number of external (to your domain) agreements that depend on it. Within the scope of global Internet services, each Service Provider would find itself being part of a large number of SP chains.

This solution is not appropriate for global QoS coverage as it would lead to what we call *lake-freezing phenomenon*, ending up with a completely petrified QoS infrastructure, where nobody could renegotiate any agreement. We deem this lack of flexibility unacceptable for any Service Provider.

We do think that if a QoS-enabled Internet is desirable, with QoS services available potentially to and from any destination, as we are used to with the current Internet, any solution must resolve this problem and find other schemes for SP-to-SP agreements. For this purpose, we introduce the concept of Meta-QoS-Class.

E. The need for Meta-QoS-Classes

A Service Provider knows very little about agreements more than one SP domain hop away. These agreements can change and it is almost impossible to have an accurate visibility of their evolution.

Furthermore, a Service Provider cannot guarantee anything but its own l-QCs in order to avoid being trapped in SP chains. Therefore, a provider should take the decision to bind one of its l-QCs to one of its neighbor SP l-QCs based solely on:

- What it knows about its own l-QCs
- What it knows about its neighbor SP l-QCs

A Service Provider should not use any information related to what is happening more than one domain away. It should try to find the best match between its l-QCs and its neighbor

SP l-QCs. That is to say, it should bind one of its l-QCs with the neighbor l-QC that has the closest performance. Agreements are then based on guarantees covering a single SP domain.

For any n, SPn-1 guarantees SPn nothing but the crossing performance of SPn-1.

We are confronted, at this point, with a problem of QoS path consistency. If there is systematically a slight difference between the upstream l-QC and the downstream l-QC, we may end up with a significant slip between the first and the last l-QC. Therefore, we must have a means to ensure the consistence and the coherence of a QoS SP's domain path. The idea is to have a classification tool that defines two l-QCs as being able to be bound together if, and only if, they are classified in the same category. We call Meta-QoS-Class (MQC) each category of this l-QC taxonomy. From this viewpoint: *two l-QCs can be bound together if, and only if, they correspond to the same Meta-QoS-Class.*

III. The meta-QoS-class concept

A. Meta-QoS-Class based on application needs

The philosophy behind MQCs relies on a global common understanding of QoS application needs. Wherever end-users are connected, they more or less use the same kind of applications in quite similar business contexts. They also experience the same QoS difficulties and are likely to express very similar QoS requirements to their respective providers. Globally confronted with the same customers requirements, providers are likely to define and deploy similar l-QCs, each of them being particularly designed to support applications with the same type of QoS constraints. There are no particular objective reasons to consider that a Service Provider located in Japan would design a “Voice over IP” compliant l-QC with short delay, low loss and small jitter while another Service Provider located in the US would have an opposite view. Applications impose constraints on the network, independently of where the service is offered; see [15] for a survey on application needs. It should be understood that a MQC is actually an abstract concept. It is not a real l-QC provisioned in a real network.

B. Meta-QoS-Class definition

A MQCs is defined with the following attributes:

- *Name*: Name of the MQC.
- *Targeted use*: A list of potential supported services (e.g. VoIP) the MQC is particularly suited for. This list has to be extentible.
- *Performance*: Boundaries and limits for the values of the QoS performance attributes (e.g. D, J, L), whenever required. The performance can be expressed qualitatively or quantitatively.
- *Constraint on the flows*: Constraints on type of traffic to be put onto the MQC (e.g. only TCP-friendly).
- *Resources*: Constraints on the ratio: (resource for the class) to (overall traffic using this class).

Attributes could depend on the SP's domain diameter, for example a longer delay could be allowed for large domains. Performance attributes can be weighed in order to prioritize the ones the service is more sensitive to.

C. An example of a Meta-QoS-Class

To illustrate our concept and for the sake of clarity, we give an example of a MQC specification. This example is only considered as an illustration of the concept.

Name: Gold MQC: TCP-friendly and non TCP-friendly (actually two classes).

Targeted use: sensitive applications split into two different classes, one for TCP-friendly traffic and one for non TCP-friendly traffic. We differentiate between the two classes because, since we allow packet loss, a mix of TCP and non-TCP flows could put TCP flows at a disadvantage since the latter back-off in packet loss, especially with Random Early Detection (RED)-like mechanisms in routers.

Performance: low delay, low jitter, low loss.

Constraint on the flows: TCP friendly traffic for the TCP-friendly Class traffic.

Resources: on each output interface, the traffic for the class can be greater than the bandwidth reserved for the class (AF based), the difference between traffic and bandwidth has a direct impact on the loss rate.

D. Compliance of l-QCs to a Meta-QoS-Class

A Service Provider goes through several steps to expand its internal l-QCs. First, it classifies its own l-QCs based on MQCs. Second, it learns about available MQCs advertised by its neighbor. To advertise a MQC, a Service Provider must have at least one compliant l-QC and should be ready to reach agreements to let neighbor SP traffic benefits from it. Third, it contracts an agreement with its neighbor to send some traffic that will be handled accordingly to the agreed MQCs. The latter stage is the binding process. A l-QC can be bound only with a neighbor l-QC that is classified as belonging to the same MQC.

Note that when a Service Provider contracts an agreement with a neighbor it may well not know to what downstream l-QCs its own l-QCs are going to be bound. It only knows that when it sends a packet requesting a given MQC treatment (for example, owing to an agreed DSCP marking) the packet will be handled in the downstream SP domain by a l-QC compliant with the requested MQC.

E. What's in and out of a Meta-QoS-Class

A MQC typically bears properties relevant to the crossing of one and only one SP domain. However this notion can be extended, in a straightforward manner, to the crossing of several domains, as long as we consider the set of consecutive domains as a single virtual domain.

The MQC concept is very flexible with regard to new unanticipated applications. According to the end-to-end principle [16], a new unanticipated application should have little impact on existing l-QCs, because the l-QCs should have been designed, to the extent possible, to gracefully allow any new application to benefit from the existing QoS infrastructure they form. However, this issue does not concern the MQCs per se, because a MQC is an abstract concept that has no physical existence. It is solely the problem of l-QCs design and engineering. Therefore, a new unanticipated application could simply drive a new MQC and a new classification process for the l-QCs.

A hierarchy of MQCs can be defined for a given type of service (e.g. VoIP with different qualities). A given l-QC can be suitable for several MQCs (even outside the same hierarchy). In this case, several DSCPs are likely to be associated with a same l-QC in order to differentiate between traffic classes. Several l-QCs in a given SP domain can be classified as belonging to the same MQC.

The DiffServ concept of PDB should not be confused with the MQC concept. The two concepts share the common characteristic of specifying some QoS performance values. The two concepts differ in their purposes. The objective for the definition of a PDB is to help implementation of QoS capabilities within a single administrative network. A MQC does not describe the way to implement a l-QC or PDB. The objective for a MQC is to help agreement negotiation between Service Providers.

IV. THE FUNDAMENTAL USE CASE:THE QOS INTERNET AS A SET OF META-QOS-CLASS PLANES

A. MQC planes

We describe here the fundamental use case, for a QoS-enabled Internet, based on the MQC concept. Our purpose is to build a QoS-enabled Internet that keeps, as much as possible, the openness characteristics of the existing best-effort Internet, and more precisely conforms to the requirements expressed earlier in this paper.

The resulting QoS Internet appears as a set of parallel Internets or MQC planes. Each plane is devoted to serve a single MQC. Each plane consists of all the l-QCs bound accordingly to the same MQC. When a l-QC maps to several MQCs, it belongs to several planes. The end-users can select the MQC plane that is the closest to their needs as long as there is a path available for the destination.

Figure 1 depicts the physical layout of a fraction of the Internet, comprising four domains from four different SPs, with full-mesh connections.

Figure 2 depicts how these four SPs are involved in two different MQC planes.

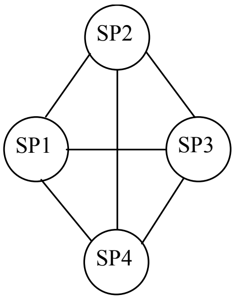


Fig. 1. A physical configuration

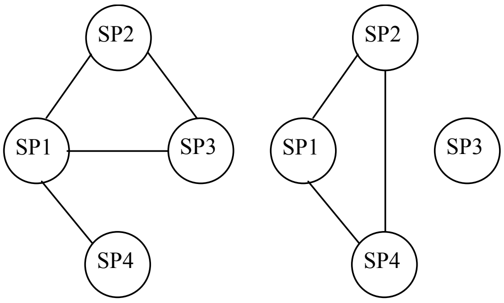


Fig. 2. Two Meta-QoS-Class planes

Considering the left-hand plane, we can deduce the following information: (1) each SP has at least one compliant l-QC for the given first MQC, and (2) a bi-directional agreement to exchange traffic for this class exists between SP1 and SP2, SP1 and SP3, SP1 and SP4, SP2 and SP3

Considering the right-hand plane, we can deduce the following information: (1) SP1, SP2 and SP4 have at least one compliant l-QC (SP3 maybe it has or not) for the given second MQC, and (2) a bi-directional agreement exists between SP1 and SP2, SP1 and SP4, SP2 and SP4.

We assume that in each MQC plane, because we want to stay close to the Internet paradigm, all paths are equal. Therefore, the problem of path selection amounts to: *do your best to find one path, which is as good as possible, within the selected MQC plane.* This is like the traditional routing system used by the Internet routers, applied to each of the MQC planes. Thus, we can rely on a Border Gateway Protocol (BGP)-like inter-domain routing protocol for the path selection process. We can call this protocol q-BGP. By destination, q-BGP selects and advertises one path for each MQC plane. From an abstract view, each MQC plane runs its own BGP protocol. When it comes to implementation, there can be only one q-BGP session between two SP domains,. We give some more details about the way we have designed our q-BGP proposal in sub-section C.

When, for a given MQC plane, there is no path available to a destination, the only way for a datagram to reach this destination is to use another MQC plane. The only MQC plane available for all destinations is the best-effort MQC plane (i.e. the current best-effort Internet). In case a best-effort path only is available, reachability is assured but not end-to-end QoS guarantees.

For a global QoS-based Internet, this solution stands only if MQC-based binding is largely accepted and becomes a current practice. This limitation is due to the nature of the service itself, and not to the use of MQCs. Insofar as we target global services we are bound to provide QoS in as many SP domains as possible. However, any MQC-enabled part of the Internet that forms a connected graph can be used for QoS communications, and be incrementally extended. Therefore, incremental deployment is possible, and does lead, to a certain extent, to incremental benefits. For example, in the Figure 2 right-hand plane, as soon as SP3 connects to the MQC plane it will be able to benefit from SP1, SP2 and SP4 QoS capabilities.

We can now elaborate a bit more on what it means for a Service Provider to contract an agreement with another Service Provider based on the use of MQCs. It simply means adding a link to the corresponding MQC plane, basically just what current traditional inter-domain agreement means for the existing Internet. As soon as a SP domain joins a MQC plane, it can reach all domains and networks within the plane.

This set of domains and networks is prone to evolve dynamically along with the appearance of new inter-domain agreements and the revocation of old inter-domain agreements. However, for a given SP-to-SP agreement, in a given MQC plane, any evolution elsewhere in this plane has no direct impact on this agreement. We are not, therefore, prone to the *lake-freezing phenomenon* and we can easily change our inter-domain agreements so far as our neighbor Service Providers agree.

We fully benefit from the resilience feature of the IP routing system: if a QoS path breaks somewhere, the q-BGP protocol will make it possible to compute another QoS path dynamically in the proper MQC plane.

Each Service Provider must have the same understanding of what a given MQC is about. A global agreement, on a set of standards, is needed. This agreement could be typically reached in an international standardization body. The number of MQCs defined, and consequently the number of MQC planes, must remain very small to avoid an overwhelming complexity. The need for standardization is evident as far as inter-domain QoS is concerned [17]. There must be also a means to certify that the l-QC classification made by a Service Provider conforms to the MQC standards. So the MQCs standardization effort should go along with some investigations on conformance testing requirements.

B. Levels of supported QoS guarantees

Any QoS inter-domain solution, either based on MQC or on a completely different approach, is valid as long as each Service Provider claiming to offer some QoS performance actually delivers the expected level of guarantee. In our MQC-based solution this is ensured by concatenation of local binding agreements, without any broad agreement covering the whole QoS path.

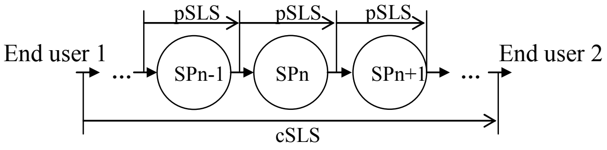


Fig. 3. pSLS and cSLS guarantee scope

It is very important to note, that here, we are only speaking of SP-to-SP agreements. Having SP-to-SP agreements limited to only one domain should not preclude having SP-to-customer agreements guaranteed edge-to-edge, from first domain ingress point to last domain egress point. Figure 3 shows the difference between the scope of SP-to-SP agreement, also known as pSLS and the scope of SP-to-Customer agreement, also known as customer SLS (cSLS).

There is often confusion about QoS as an underlying technology and QoS as a service offering [18].

If we target to offer harder administrative guarantees, for well-delimited services like VPN, we can, for example, use MQC information exchange by q-BGP to find a QoS path that fits the demand, and then reach an agreement for all SP actors of the selected QoS path, possibly enforcing the path by a MPLS tunnel [19].

More generally, MQC is a concept. MQC does not prohibit the use of any particular QoS mechanism or protocol at the data, control or management plane. For example, DiffServ, Traffic Shaping, Traffic Engineering, Admission Control, and so forth, are completely legitimate. MQC simply drives and federates the way QoS inter-domain relationships are built.

C. QoS-Enhanced Border Gateway Protocol (q-BGP)

In order to build inter-domain QoS delivery solutions, a SP needs to exchange QoS-related information with its neighbors to characterize, qualify and then classify the level of the QoS that will be experienced by its customer's traffic when transiting by the domain of one of its peer SPs. The exchange of QoS-related information and their corresponding characteristics can occur either at the Service Layer, especially during the pSLS negotiation phase: in this case, precise values of those QoS performance characteristics are agreed between two neighbours; or at the Control Layer, owing to the activation of appropriate protocols especially in the routing level. In both cases, SPs should activate appropriate mechanisms and implement adequate functionalities in order to ensure that inter-domain QoS service targets as agreed in pSLS or exchanged in the routing level are met.

Border Gateway Protocol (BGP, [20]) is the Internet inter-domain routing protocol for interconnecting adjacent autonomous systems. Within the context of the MESCAL project, we have investigated how BGP could be used as a means to convey QoS-related information between adjacent Autonomous Systems (AS) and proposed the QoS-Enhanced BGP (q-BGP) protocol. This enhanced version of BGP does not require any change of the BGP state machine but allows new features like the treatment differentiation of received announcements depending on the nature of QoS-information conveyed between two adjacent domains. From a q-BGP standpoint, two categories of QoS-delivery solutions have been identified: The first category that needs to exchange only an identifier of the MQC agreed during the pSLS negotiation phase. The QoS performance characteristics and their targets are negotiated and agreed in the pSLS; they are not exchanged in the routing level by q-BGP. The second category that requires that both the identifier of MQC plane and the QoS performance metrics values be exchanged by q-BGP. For the second category, the QoS performance metrics to be exchanged are agreed during the pSLS negotiation phase.

In order to implement the aforementioned features, q-BGP makes use of two new attributes listed below:

1. *QoS Service Capability:* This attribute is used during capability negotiation between two q-BGP peers when initiating the q-BGP session. It is included in the optional parameters of the OPEN message. This attribute allows peering entities to know about each other's QoS service capabilities, and indicates what information can potentially be carried by the q-BGP messages. A q-BGP speaker should use this capability attribute in order to indicate the group to which an offered inter-domain QoS delivery solution belongs to.

2. *QoS_NLRI:* Two flavours of this attribute are described in [21]. This attribute is used to convey QoS-related information. This attribute carries multiple QoS performance characteristics. The main important fields of this attribute are listed below:

- *QoS information Code*: this field identifies the type of QoS information (e.g. Packet rate, One-way delay metric, Inter-packet delay variation, etc.)
- *QoS information Sub-code*: this field carries the sub-type of the QoS information. Several sub-types have been identified like: Reserved rate, Available rate, Loss rate, Minimum one-way delay, Maximum one-way delay, etc.
- *QoS information value*: this field indicates the value of the QoS information. The corresponding units depend on the instantiation of the QoS information code. This could be either statically valid for a specified period, or dynamically, obtained through measurements. The way to set this value (for the originator of the announcement) is up to domains' administrators and/or mutual agreement between SP peers. When receiving a q-BGP route, the q-BGP receiver concatenates its local QoS values with the received ones in case this route may be re-advertised to other peers.
- *QoS Class Identifier*: This is used to distinguish the MQC plane in which belongs a given q-BGP announcement.

As far as QoS-related information is conveyed in q-BGP UPDATE messages, the route selection process should take into account this information in order to select one route from equal paths and determine the one to be actually used and stored in the Forwarding Information Base (FIB). This process could differ between inter-domain QoS delivery solutions that belong to the first category or the second one explained above. For the first category, the q-BGP route selection process is very similar to the classical BGP route selection one, i.e q-BGP route selection process will choose the route that minimises the AS_PATH hops for each MQC plane. Additional policies could be enforced according to the knowledge of the content of pSLSs and then drive the setting of some BGP metrics values like Local_Pref. For the second category, since several QoS parameters may be advertised for a given destination for each MQC plane, the process examines each QoS parameter in a prioritised order. Thus, the route selection process chooses the best routes by examining initially the highest priority QoS parameter. If several routes have the same weight for the highest priority parameter, the second priority parameter is considered, and this process is repeated as necessary until a route is selected. This process is achieved for each MQC plane [22].

D. Constructing inter domain Meta-QoS-Class planes

In order to construct QoS-enabled Internet planes driven by the MQC concept, each SP has to engineer its local QCs in order to comply with one or several MQCs. This engineering task is left solely to the SP. A SP might choose to engineer its entire network to support the highest quality traffic, complying with all MQCs at the same time. Once the engineering of local QCs and their classification according to MQCs have taken place, the steps listed below have to be followed:

- Establishment of bi-lateral pSLSs, to enable the exchange of traffic belonging to a MQC. These pSLSs activate q-BGP sessions per MQC plane;
- Identification of the traffic flows and q-BGP announcements that fall into a particular MQC. For traffic flows, this is achieved at a data level by using the DSCP and, in q-BGP, by means of the aforementionned QoS attributes, especillay the QC identifier field contained in QoS_NLRI attribute;
- Announcement of the network prefixes that can be reached within each MQC plane;
- DSCP swapping of data packets at each domain's ingress and egress points. MQC traffic when arriving at a domain needs to be marked accordingly, in order to receive the appropriate local treatment. When exiting a domain, MQC traffic needs to be remarked to the appropriate DSCP value, in order to be identified by the adjacent domain as to belonging to the particular MQC and receive the appropriate treatment.

When traversing an AS chain, the QoS treatment experienced by an IP datagram is consistent in all traversed ASs. The packet treatment received in each AS conforms to the corresponding MQC definition, through the engineering of suitable l-QCs. By using the MQC identifier included in the q-BGP UPDATE messages, each message can be processed within the context of the corresponding MQC plane.

V. Experiment results

Within the MESCAL project[1], a testbed has been set-up in order to test various functionalities including protocols and algorithms. Note that MQC concept cannot easily be tested because it is a means of classification of l-QCs and its added value could be validated when comparing between the results obtained if this concept is adopted and the ones obtained owing to the use of an alternative method. The conducted experiments within the testbed and simulations aim to validate the inter domain routing behaviours and performance within a MQC plane. Especially, The testbed is used to validate, at the data plane level, the DSCP marking/remarking between ASs in order to signal an inter-domain MQC, to validate the implementation of l-QCs in each domain using Linux traffic control features, to validate the Meta-QoS-Class concept and finally to validate the QoS-inferred inter-domain routing. Simulation has been used to study scalability issues.

The testbed is composed of eight ASs, q-BGP is activated at the boundaries of each domain, and a full mesh q-iBGP is activated within domains that are composed of more than one linux-based router. All routers have DiffServ capabilities for traffic classification, traffic conditioning and various scheduling disciplines. Hierarchical Token Bucket (HTB) scheduling discipline is used rather than Class Based Queuing (CBQ) [23] for implementing the classes of service. DSCP ingress re-marking is achieved by using IPFILTER and DSCP egress re-marking is achieved by using DSMARK queuing discipline [24]. Four l-QCs that belong to four distinct MQCs are configured in each AS.

Series of tests have been conducted within the testbed. DSCP swapping, QoS aggregation and route selection process were tested and conform to the specifications.

The figures below show the results of two tests conducted in the testbed. The DS field includes the DSCP value, plus two bits forced to zero.

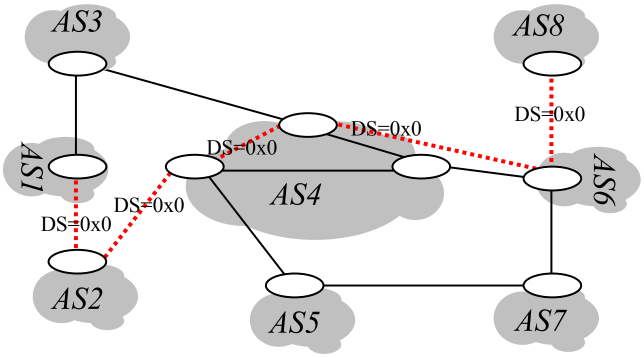


Fig. 4. Ping request following best-effort MQC

Border routers are represented in each AS by a single Linux-based router, except AS4 that is composed of three routers. Figure 4 is a snapshot of a ping tracker showing the path of a best effort ping request sent from a source located in AS1 towards a destination prefix located in AS8. This figure shows that q-BGP selected the AS2-AS4-AS6-AS8 path in the best-effort MQC plane (let's call it MQC0).

In the second scenario, depicted in Figure 5, all routers in the testbed are configured to prioritise the average one-way delay parameter for traffic belonging to MQC1. q-BGP will therefore select the path with the lowest average one-way delay, for any packet traveling inside MQC1.

MQC1 is signaled by distinct DS values between two BGP peers. The DS value used to signal MQC1 between AS1 and AS3 is 0x88, between AS3 and AS4 is 0xe8, between AS4 and AS6 is 0x48 and between AS6 and AS8 is 0xe8 (Note that distinct values are used to signal the other two MQCs deployed: MQC2 and MQC3). Within AS1, the l-QC identified by a DS value of 0x28 is classified to belong to MQC1.

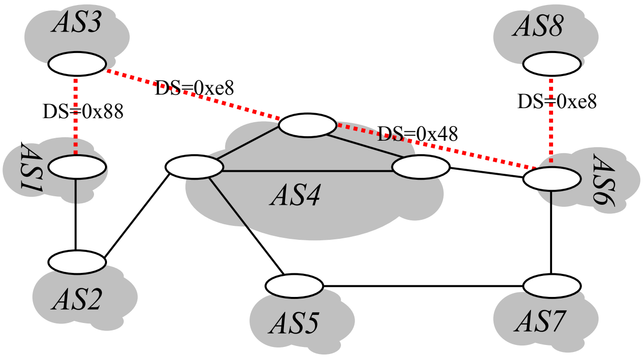


Fig. 5. Ping request following low-delay MQC

AS1 received q-BGP advertisements that gave an average one-way delay to AS8 prefixes of 350 ms via AS2 and 250 ms via AS3 for traffic marked with DS value of 0x88 (MQC1 DS value). Therefore, the traffic marked with a DS value of 0x28 (l-QC1 DS value), which is sent from AS1 will follow the path selected by q-BGP for MQC1 plane, i.e. AS3-AS4-AS6-AS8 as shown in Figure 5. q-BGP chose this path because it minimises the value of the average one-way delay parameter. The figure shows that traffic generated in AS1 with a DS of 0x28 is classified in MQC1 and the traffic is correctly marked/remarked when entering/exiting AS it encounters in its path.

VI. Simulation results

The testbed has demonstrated the feasibility and implementation of the q-BGP protocol and has concentrated on differentiated routing over multiple MQC planes. However, it would have been unfeasible to use it to examine large-scale behaviour. To this end, a simulator based on NS was created to examine the performance of q-BGP on much larger topologies and to extrapolate behaviour to Internet-scale topologies. The flexibility of a software-based simulator also eases further experimentation and prototyping on the

[1] www.mescal.org

protocol, such as an investigation into route selection processes, QoS_NLRI information type and calculation, and various other aspects of q-BGP policies. The simulator is epoch driven. Every AS performs its q-BGP processes at every epoch, and network monitoring is performed. Therefore a q-BGP message, if re-advertised through the network, will traverse the network at one hop per epoch.

When simulating an Internet-like network, a number of assumptions must be made. Firstly, topological traits such as the degree distribution and average connectivity must be assumed. While there are no truly representative Internet-like topology generators, the Barabasi-Albert (BA) model from the BRITE [25] topology generator was used. This creates power-law compliant topologies [26] when its preferential attachment option is used. It has been shown that the Internet is also a power-law compliant topology at the AS level. The second set of assumptions is on intra-AS traffic treatment: it was assumed that a premium, delay-optimised MQC across a single AS has a constant average delay between all border routers and the delay per AS has a uniformly random distribution (between 5 and 50 ms) over all ASs. The traffic matrix was assumed to be a full mesh, and therefore end-to-end one-way delay is always calculated for all pairs of source-destination ASs. In the simulation results presented below, it was assumed that no additional delay due to congestion was introduced as there was sufficient intra- and inter- domain capacity to prevent excessive queuing delays (premium service). Four different topologies were generated for each topology size (number of ASs) with other topological parameters, such as degree of connectivity, remaining constant. For each instance of the topology, twelve separate examples of intra-AS one-way delay allocations (l-QC delay values) were created. The results for each topology size are therefore the average over 48 simulation runs of different topologies and intra-AS delay distributions.

In the plots below each network instance was an AS topology of the specified size (number of ASs), with an average connectivity of four uni-directional inter-domain links per AS (the AS with the highest connectivity had 67 uni-directional inter-domain links and the top 10% ASs were connected to an average of 14 peer ASs). Local_pref was not set in any AS and therefore did not play a role in the route selection process. In the case of q-BGP category 1 (i.e. Only an identifier of MQC is used. This identifier is represented by mQCid), route selection was performed based on AS Path length and the AS number as tie-breaker, while in q-BGP category 2 (mQCid + QoS Info), route selection was performed on the One-Way-Delay (OWD) QoS attribute first, then AS Path length and then AS number as tie-breaker.

The improvement in delay can be seen in Figure 6 as a function of AS topology size. It can be seen that the benefit of additional QoS info in q-BGP messages is increasing with topology size. This is due to an increase in the number of alternative AS paths between a given source-destination pair (other than the default shortest AS-path length) as the topology size grows, and therefore the chances of finding an improved path on one-way delay grounds is increased.

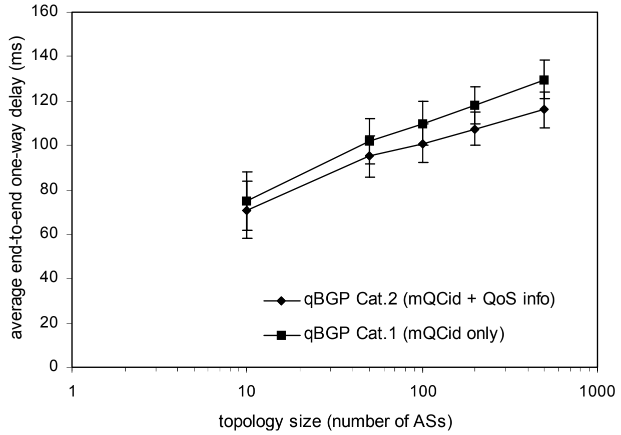


Fig. 6. The effect of topology size on the end-to-end delay experienced by demands

In Figure 7, we can see the total number of q-BGP messages sent from the first set of bootstrap messages, to a stable routing configuration. It should be noted that no message aggregation is performed in these simulations, either on network prefixes or QoS attributes.

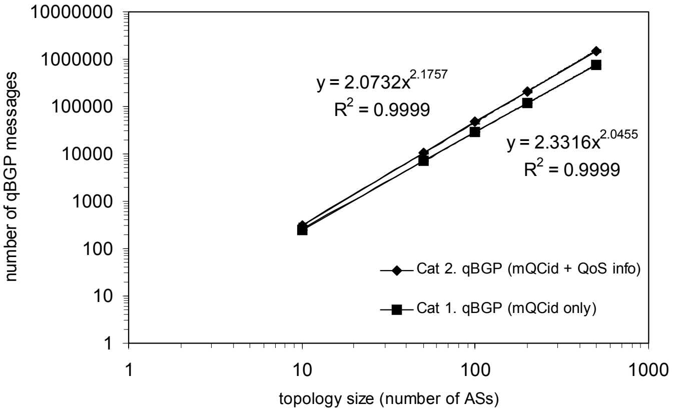


Fig. 7. The number of q-BGP messages sent from initialisation until it settles in a stable state with a full mesh of demands applied

The inclusion of QoS attributes in q-BGP therefore scales, in terms of number of q-BGP messages, in a similar way to q-BGP UPDATES and route selection based on mQCid only. Actually, the number of messages forms a power law with topology size, which is equivalent to the scaling of BGP today.

The main reason for the increased number of messages required for convergence is that, on QoS grounds, the preferred AS path may not always be the shortest one and therefore, UPDATE messages which are seen as better may arrive after messages following the shortest path, and as such, must be re-advertised after any UPDATE that arrived along the shortest path. This can be seen also in the amount of simulator epochs required before a stable and converged routing state is reached across the entire network. For a 100 nodes network the mQCid-only convergence time was 9.5 simulator epochs, while for the mQCid with OWD attribute this became 10.8 epochs.

In this section we have shown that the addition of QoS information as part of q-BGP messages does not scale significantly worse than shortest path based route selection policies, although more messages do have to be sent and the convergence time is minimally longer. This is caused by the increased propagation time as messages do not only follow the shortest path.

VII. Conclusion

In this is paper we have looked into the perspective for a global QoS-based Internet, based on existing mechanisms such as Differentiated Services. Towards this target we have introduced the new concept of Meta-QoS-Class. It significantly helps Service Providers to negotiate agreements. It avoids what we have called *Service Provider chain traps* leading to *lake-freezing phenomenon*. It Provides guidance for l-QC binding. It allows relevant l-QC binding with no knowledge of the following distant provider agreements. It enforces coherence in a QoS path without any knowledge of the complete end-to-end path.

The Meta-QoS-Class concept opens up an innovative way to achieve global QoS Internet connectivity that maintains the main features of the Internet. It could open a new path in the inter-domain QoS research area and enable new QoS models to be introduced. Along with this concept, we have also proposed a new q-BGP protocol that allows QoS route calculation and QoS route information dissemination. We have noticed satisfactory behavior on Internet-scale topologies via testbeds and simulations.

VIII. Future work

Some future work should be undertaken to refine the definition of a MQC. Each Service Provider must have the same understanding of what a given MQC is about. There must be also a means to certify that the l-QC classification made by a Service Provider conforms to the MQC standards. Security is a main concern in a QoS-enabled Internet. Flows entering a domain and requesting QoS are likely to arrive from any SP domain and to be destined to any SP domain. So, it is of primary importance for a Service Provider to be able to filter the flows whose requests are not legitimate. Some investigation must be conducted in this direction. Any research on QoS has particularly strong requirements in security [27]. The MQC concept opens the possibility of QoS services potentially reachable from anywhere on the Internet. Consequently, the menace of a spurious attack grows accordingly. Finally, the issue of how to provide the end user with QoS guarantees based on SP to SP QoS aggrements must be clarified and investigated.

Acknowledgement

We would like to thank the MESCAL partners for their inputs and fruitful discussions.

References

[1] B. Teitelbaum. (2001). Future Priorities for Internet2 QoS. Internet2 QoS Working Group document. [Online]. Available: http://qos.internet2.edu/wg/documents-informational/20011002-teitelbaum-qos-futures.pdf

[2] J. Crowcroft, S. Hand, R. Mortier, T. Roscoe, and A. Warfield: *QoS`s Downfall: At the bottom, or not at all!*, in Proc. ACM SIGCOMM Workshop on Revisiting IP QoS, pp. 109-114, Aug. 2003.

[3] G. Bell: *Failure to Thrive: QoS and the Culture of Operational Networking*, in Proc. ACM SIGCOMM Workshop on Revisiting IP QoS, pp. 115-120, Aug. 2003.

[4] G. Huston: *Next Steps for the IP QoS Architecture*, RFC 2990, Nov. 2000.

[5] D.D. Clark: *The Design Philosophy of the DARPA Internet Protocols*, in Proc. ACM SIGCOMM, pp. 106-114, Aug. 88.

[6] M. Blumenthal, D.D. Clark: *Rethinking the design of the Internet: The end to end arguments vs. the brave new world*, ACM Transactions on Internet Technology, Vol.1, No.1, pp. 70-109, Aug. 2001.

[7] J. Kempf, R. Austein: *The Rise of the Middle and the Future of End to End: Reflections on the Evolution of the Internet Architecture*, RFC 3724, Mar. 2004.

[8] U. Julita et al.: *A common Framework for QoS/Network Performance in a multi-Provider Environment*, EURESCOM Project P806, Del.1, Sep. 1999.

[9] K. Nichols, B. Carpenter: *Definition of Differentiated Services Per Domain Behaviors and Rules for their Specification*, RFC 3086, Apr. 2001.

[10] A. Farrel, J.P. Vasseur, J. Ash*: Path Computation Element (PCE) Architecture*, draft-ietf-pce-architecture-01.txt, Work in Progress, Jul. 2005.

[11] R. Hancock, G. Karagiannis, J. Loughney, and S. van den Bosch: *Next Steps in Signaling (NSIS): Framework*, RFC 4080, Jun. 2005.

[12] M.P. Howarth, P. Flegkas, G. Pavlou, N. Wang, P. Trimintzios, D. Griffin, J. Griem, M. Boucadair, P. Morand, H. Asgari and P. Georgatsos: *Provisioning for Inter-domain quality of service: the MESCAL app.roach*, IEEE Communications Magazine, Vol.43, No.6, pp. 129-137, Jun. 2005.

[13] H. Asgari, M. Boucadair, R. Egan, P. Morand, D. Griffin, J. Griem, P. Georgatsos, J. Spencer, G. Pavlou, M.P. Howarth: *Inter-Provider QoS Peering for IP Service Offering Across Multiple Domains*, in Proc. International Workshop on Next Generation Networking Middleware (NGNM '05), IFIP Networking Conference, May 2005.

[14] P. Levis, A. Asgari, P.Trimintzios: *Considerations on inter-domain QoS and Traffic Engineering issues through a utopian app.roach*, in Proc. International Conference on Telecommunications SAPIR Workshop, pp. 231-238, Aug. 2004, ©Springer-Verlag.

[15] D. Miras. (2002). A Survey on Network QoS Needs of Advanced Internet App.lications. Internet2 QoS Working Group working document. [Online]. Available: http://qos.internet2.edu/wg/app.s/fellowship/Docs/Internet2App.sQoSNeeds.html

[16] J.H. Saltzer, D.P. Reed, D.D. Clark: *End to End Arguments in System Design*, ACM Transactions on Computer Systems, Vol.2, No.4, pp. 277-288, Nov. 1984.

[17] M. Eder, H. Chaskar, S. Nag: *Considerations from the Service Management Research Group (SMRG) on Quality of Service (QoS) in the IP Network*, RFC 3387, Sep. 2002.

[18] B. Teitelbaum, S. Shalunov: *What QoS Research Hasn't Understood About Risk*, in Proc. ACM SIGCOMM Workshop on Revisiting IP QoS, pp. 148-150, Aug. 2003.

[19] D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, J. McManus: *Requirements for Traffic Engineering Over MPLS*, RFC 2702, Sep. 1999.

[20] Y. Rekhter, T. Li: *A Border Gateway Protocol 4 (BGP-4)*, RFC 1771, Mar. 1995.

[21] M. Boucadair: *QoS-Enhanced Border Gateway Protocol*, draft-boucadair-qos-bgp-spec-01.txt, Work in Progress, Jul. 2005.

[22] M.P. Howarth, M. Boucadair, P. Flegkas, N. Wang, G. Pavlou, P. Morand, T. Coadic, D. Griffin, A. Asgari and P. Georgatsos: *End-to-end quality of service provisioning through Inter-provider traffic engineering*, Computer Communications, Elsevier, to be published, end 2005.

[23] S. Floyd, V. Jacobson: *Link-sharing and Resource Management Models for Packet Networks*, IEEE/ACM Transactions on Networking, Vol.3, No.4, pp.. 365-386, Aug. 1995.

[24] W. Almesberger, J. Hadi Salim, A. Kuznetsov: *Differentiated Services on Linux*, in Proc. IEEE GLOBECOM, Vol.1B, pp.. 831-836, Dec. 1999.

[25] A. Medina, A. Lakhina, I. Matta, J. Byers: *BRITE: An App.roach to Universal Topology Generation*, In Proc. International Workshop on Modeling, Analysis and Simulation of Computer and Telecommunications Systems (MASCOTS '01), pp. 346-356, Aug. 2001.

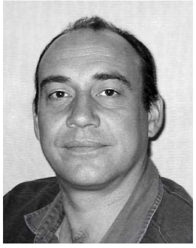
[26] T. Bu and D. Towsley: *On Distinguishing between Internet Power Law Topology Generators*, in Proc. IEEE INFOCOM, Vol.2, pp. 638-647, Jun. 2002.

[27] R. Atkinson, S. Floyd: *IAB Concerns & Recommendations Regarding Internet Research & Evolution*, RFC 3869, Aug. 2004.

Pierre Levis was an Assistant Professor at INT (Institut National des Telecommunications) Evry France, from 1990 to 1998. He was in charge of computer networks courses. His research interest was on Information Technology. Since 1998 he has been with France Telecom R&D Division. He has worked in the specification, the developpement, and the evaluation of IP service offerings. His research interests cover QoS, IPv6, network security, AAA, and mobile networks.

Mohamed Boucadair was an R&D engineer with France Telecom R&D in charge of dynamic provisioning, QoS, multicast, and intra/interdomain traffic engineering, and now works on VoIP services. He graduated from the Ecole Nationale Supérieure d'Ingénieur de Caen (Institut des Sciences de la Matière et des Rayonnements), a French school of engineers.

Pierrick Morand is the project coordinator of the EU IST-MESCAL project. He also leads a research group within France Telecom working to develop and enhance VoIP services for corporate business. He graduated from the Ecole Nationale Supérieure d'Ingénieur de Caen (Institut des Sciences de la Matière et des Rayonnements), a French school of engineers, and received his degree from the University of Caen, France.



David Griffin is a Senior Research Fellow in the Department of Electronic and Electrical Engineering, University College London (UCL), UK. He has a BSc in Electrical Engineering from Loughborough University, UK, and is currently completing a part-time PhD in Electrical Engineering from the University of London. Before joining UCL he was a Systems Design Engineer at GEC-Plessey Telecommunications, UK and then a Researcher in Telecommunications at the Foundation for Research and Technology - Hellas (FORTH), Institute of Computer Science, Crete, Greece.

Jason Spencer received his B.Eng. degree in Electronic Engineering and an M.Sc. degree in Telecommunications from University College London (UCL), London, U.K., in 1997 and 1998, respectively. He is currently working toward the Ph.D. degree at UCL on the interactions between network layers and their effects on network design. His research interests include network planning and management, decentralized network control, next-generation high-speed reconfigurable networks, complex systems, and large-scale system design.

Panos Trimintzios is a researcher at the Foundation for Research and Technology — Hellas (FORTH), Institute of Computer Science (ICS), Greece. Before that he was a research fellow at CCSR, University of Surrey. He received a B.Sc. in computer science and an M.Sc. in computer networks both from the University of Crete, Greece, and his Ph.D. from the University of Surrey. His research interests include network management and control, network security, network monitoring, policy-based networking, QoS service provisioning, and GRIDs.

George Pavlou is professor of communication and information systems at CCSR, University of Surrey, where he leads the activities of the Networks Research Group. He holds a Diploma in engineering from the National Technical University of Athens, Greece, and M.Sc. and Ph.D. degrees from University College London (UCL).