Interdomain Routing through QoS-Class Planes

David Griffin, Jason Spencer, and Jonas Griem, University College London

Mohamed Boucadair and Pierrick Morand, France Telecom R&D

Michael Howarth, Ning Wang, and George Pavlou, University of Surrey

Abolghasem (Hamid) Asgari, Thales Research & Technology Ltd.

Panos Georgatsos, Algonet S.A.

ABSTRACT

This article presents an approach to delivering qualitative end-to-end quality of service (QoS) guarantees across the multiprovider Internet. We propose that bilateral agreements between a number of autonomous systems (ASs) result in the establishment of QoS-class planes that potentially extend across the global Internet. The deployment of a QoS-enhanced Border Gateway Protocol (BGP) with different QoS-based route selection policies in each of the planes allows a range of interdomain QoS capabilities to coexist on the same network infrastructure. The article presents simulation results showing the benefits of the approach and discusses aspects of the performance of QoS-enhanced BGP.

INTRODUCTION

Various initiatives have attempted to add quality of service (QoS) to the Internet through resource reservation (integrated services or IntServ), differentiated forwarding (differentiated services or DiffServ), traffic engineering of routes (multiprotocol label switching traffic eengineering or MPLS-TE) or simply through overprovisioning. These approaches have focused mainly on the means of providing edge-to-edge performance guarantees within a single autonomous system (AS) or set of ASs under the control of one administrative entity. Internet service scope, however, cannot be constrained by the boundaries of a single administration and it is in the interests of all Internet network providers, service/content providers, and their customers for the scope of QoS-based services to be extended globally, across multiple domains.

The problem of how to extend QoS capabilities across multiple provider domains has not been solved satisfactorily to date. The source of the problem lies mainly with the autonomous nature of ISPs and their loose federation[1] that forms the global Internet. There is no central authority or regulation compared to networks such as the Public Switched Telephone Network where the ITU-T specifies reference models and performance targets so that end-to-end conversational voice quality is guaranteed.

Our solution to the interdomain QoS problem is tackled at two levels that mirror the business model of today's best-effort Internet. First of all, an offline traffic engineering (TE) process in each provider domain decides with which neighbouring ASs to establish QoS bindings on the basis of compatibility of intradomain QoS capabilities (similar to per domain behaviours [1]). This is analogous to the decision of today's ISPs with which ASs to establish peering or transit relationships for best-effort traffic. But in the case of QoS-based peering, additional agreements are required in the form of interprovider service-level specifications (SLSs) [2] that determine the technical aspects of the QoS-based connectivity services to be utilised, including the quantity, quality, and topological scope of the traffic to be exchanged.

The second level of our interdomain QoS solution is concerned with interdomain routing, and is focussed on QoS extensions to the Border Gateway Protocol (BGP [3]). QoS-enhanced BGP (q-BGP) [4] conveys additional information relating to the QoS of the paths it has selected to the destination prefixes that it announces. Enhanced route-selection processes are then able to establish appropriate interdomain routes so that traffic is forwarded according to the already established, additional QoS agreements

[1] *The term* loose federation *is used with regard to the end-to-end treatment of packets. Note that there are stricter constraints between any two adjacent ASs due to the customer-provider relationship between the ASs — especially for transit traffic.*

between ASs. Note that q-BGP sessions only exist between peers with which appropriate SLSs have been established. In this way, q-BGP routes are constrained to follow the SLSs that have been previously established by the offline TE algorithms. This also implies that those ASs who do not wish to participate in exchanging QoS-based traffic do not receive unwanted QoS information through BGP.

We believe that these two techniques — establishing interdomain QoS-class planes (later referred to as m-QCs, or meta-QoS-Class) through offline TE mechanisms, coupled with dynamic optimisation of the route selection process through q-BGP — provide a lightweight means of enabling end-to-end QoS across multiple domains.

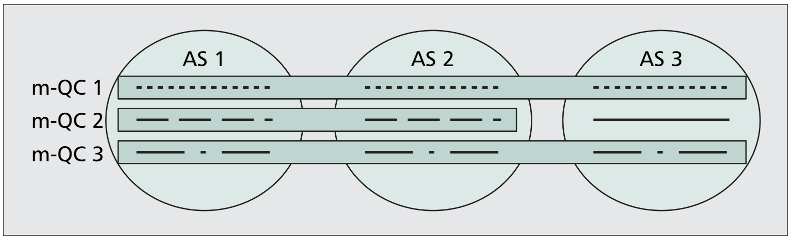
The following section describes the concept of m-QC planes and we describe enhancements to BGP for QoS-based interdomain routing. We present experimental results from simulations of q-BGP, demonstrating the effect of injecting QoS information into BGP and investigating the effect of different q-BGP route selection policies on end-to-end network performance. Finally, we present our conclusions and identifies aspects of future work.

Interdomain QoS Delivery Through QoS-Class Planes

Achieving service differentiation across multiple ASs is hindered by the lack of co-ordination between operators of different ASs. In order to overcome this, we make use of the meta-QoS class (m-QC) concept defined in [5]. Each m-QC defines a qualitative QoS treatment across a domain, without requiring a particular means for engineering it. This leaves each operator free to choose its preferred method of engineering QoS, provided that m-QC requirements are fulfilled. Examples of m-QCs could be a one-way transit delay [6] of "very low" and packet loss rate [7] of "very low," or a delay-sensitive m-QC with delay value of "low" and loss value of "any." Unlike *hard guaranteed QoS classes* that are defined end-to-end with distinct performance characteristics, m-QCs do not imply a predefined/engineered end-to-end-QoS.

Interdomain m-QC planes are constructed by concatenating the appropriate local QoS classes across interdomain links, using well-known DiffServ Code Point (DSCP) values to distinguish the traffic belonging to different planes. Figure 1 illustrates this concept, where the dashed lines contained within each AS represent the local treatment that adheres to an m-QC. Where an AS has not created a local class conformant to an m-QC, it may not join the corresponding plane as is the case for m-QC2 in AS 3 in the figure, although it may still participate in other m-QC planes. Alternatively, it could participate in the m-QC plane with a local QoS class that exceeds the m-QC specification.

Since the traffic belonging to m-QC planes is differentiated by the DSCP value at interdomain links, each m-QC plane can route packets independently from one another with (an appropriately enhanced) BGP to optimise for each m-QCs engineering target performance. For instance, longer paths could be used for a non-delay-sensitive m-QC if this helps with load balancing and the use of low-utilisation links. This new ability to differentiate interdomain routing is the motivation for pursuing the QoS enhanced BGP that will be described in the next section. Note, however, that the mechanisms of multiple m-QC planes and differentiated interdomain routing can also support other TE goals such as resilience and load balancing.



■ **Figure 1.** *Parallel meta-QoS class planes (m-QCs) set up across multiple domains.*

In order to establish the interdomain m-QC planes the following steps need to take place:

- Establishment of interprovider SLSs, to agree the exchange of traffic belonging to a certain m-QC.
- Identification of the traffic flows and q-BGP announcements that fall into a particular m-QC. For traffic flows this is achieved at a packet level by using the DSCP field of the IP header and, in q-BGP, by means of a newly defined QoS attribute. The identifiers could either be globally known or agreed between adjacent AS peers at the time of SLS establishment.
- Announcement of the network prefixes that can be reached within each m-QC plane and associated optional QoS attributes. This is achieved dynamically through q-BGP, as described in the next section.
- DSCP swapping of data packets at each domain's ingress and egress points. M-QC traffic, when arriving at a domain, needs to be marked appropriately to receive the relevant local treatment (as per local QoS class engineered for the m-QC). When exiting from a domain, m-QC traffic needs also to be remarked to the value agreed in the SLS with the adjacent domain for that m-QC.

When traversing a set of ASs, the QoS treatment experienced by an IP datagram is now *consistent* in the sense that packet treatment received in each AS conforms to the corresponding m-QC.

QoS-Aware Interdomain Routing: Q-BGP

BGP is the most widely used interdomain routing protocol in the Internet for interconnecting adjacent ASs carrying best-effort traffic. Several proposals (e.g., [8] and [9]) have been made to extend the information conveyed by BGP. The enhancements specified below do not require any changes to the BGP state machine but allow

If several routes have the same value (or are within an acceptable margin to be considered equivalent), the second priority parameter is considered, and so on until a route is selected. If more than one route remains, then selection is based on AS_PATH length.

Route	One-way delay	One-way packet loss rate
R1	150 ms	5%
R2	120 ms	2%
R3	100 ms	3%
R4	200 ms	8%

■ **Table 1.** *Example q-BGP QoS attribute values.*

new features such as the different treatment of received announcements depending on the nature of the conveyed QoS-information.

Q-BGP [4] makes use of two new message attributes: QoS Service Capability and QoS_NLRI.

QoS Service Capability is used during capability negotiation when a q-BGP session is initiated between peers. It is an optional parameter of the OPEN message and it allows peering entities to learn about each other's QoS service capabilities, as previously agreed offline in the SLS. The parameter indicates the additional QoS information to be carried by q-BGP messages, depending on the category to which an offered interdomain QoS delivery solution belongs as follows:

Category one: An additional QoS Class identifier is exchanged in q-BGP messages, specifying the m-QC plane agreed in the SLS. QoS treatment characteristics need not be exchanged by q-BGP explicitly as they are standardised for the m-QC.

Category two: Both the m-QC plane identifier and the values of the associated QoS performance metrics are exchanged by q-BGP. For this category, the set of QoS attribute types is agreed during the SLS negotiation phase. The values of QoS performance metrics may be assigned either statically, by an administrator, or dynamically obtained through measurements. The way these values are set by the originator of the announcement is up to the domains' administrators and/or mutual agreement between AS peers.

QoS_NLRI is used to convey QoS-related information in UPDATE messages. Multiple QoS performance characteristics may be conveyed in a single QoS_NLRI attribute. It should be noted that both QoS Service Capability and QoS_NLRI attributes are optional, nontransitive.

Route Selection Process

The q-BGP route selection process should take into account the QoS-related information conveyed in the QoS_NLRI attribute of q-BGP UPDATE messages in order to select an appropriate route.[2] The process is a modified version of the classical BGP route selection process with UPDATE messages being filtered by m-QC plane identifier so that decisions are only made on announcements belonging to that m-QC plane. In addition to determining the AS's m-QC plane participation through prior SLS establishment, offline TE may further influence the q-BGP decisions through the optional configuration of LOCAL_PREF values to override dynamic q-BGP decisions based on QoS information. If more than one route has the highest LOCAL_PREF the subsequent selection process differs according to the category of interdomain QoS delivery solution, as explained above.

Category one (QoS-Class-identifier only): The q-BGP route selection process will choose the route that minimises the AS_PATH length for each m-QC plane since, in the absence of additional QoS attributes, this strategy has a good chance of maximizing end-to-end QoS when crossing ASes supporting the same m-QC.

Category two (QoS-Class-identifier and additional QoS information): since several QoS parameters may be advertised per destination prefix in each m-QC plane, the process prioritises the QoS parameters (the parameters and their priority depend on the m-QC specification). Thus, the route selection process chooses the best routes by initially examining the highest-priority QoS parameter. If several routes have the same value (or are within an acceptable margin to be considered equivalent), the second priority parameter is considered, and so on until a route is selected. If more than one route remains, then selection is based on AS_PATH length.

As an example, suppose that a q-BGP router has received the following routes from several peers for reaching the same prefix P1. Each of these routes is associated with a set of QoS performance values, as shown in Table 1.

If the q-BGP router is configured to prioritize the one-way delay metric, the selected route will be R3. But if the q-BGP router is configured to prioritise the one-way loss rate metric, the selected route will be R2.

Q-BGP Simulations and Experimental Results

In this section we demonstrate the performance gains achieved by q-BGP. Our simulation tools are event driven, using traffic models to approximate the QoS experienced by flows, including delivered bandwidth and delay due to congestion and propagation times. The first part of this section is a comparison of the impact of different q-BGP selection policies on delivered end-to-end traffic performance followed by an examination of the QoS improvement and scalability of q-BGP at bigger network sizes. Each simulation run considers a single m-QC plane and, therefore, the set of experiments models the case in which there is a hard reservation of interdomain link capacity for each SLS and each m-QC plane makes independent routing decisions through q-BGP.

Comparison of q-BGP Selection Policies

In this section we examine the relative performance of a range of different q-BGP selection policies in terms of their impact on the delivered delay and throughput on end-to-end flows. In

[2] *Our approach could also be extended to cover multiple BGP paths by adding a route index field in the QoS_NLRI attribute as proposed in draft-walton-bgp-add-paths-05.txt. However, we have chosen a single-path approach for simplicity at this stage.*

these experiments we concentrate on two QoS Attributes (QAs):

One-Way Delay (DELAYQA) QoS Attribute: The expected time for a packet to reach the prefix advertised. When traversing ASs and interdomain links, this value is formed through the concatenation of the various delay contributors:

```
Advertised DELAYQA = incoming advertisement DELAYQA + local QoS class delay + SLS queuing delay;
```

where local QoS class delay is the edge-to-edge delay across the current AS and SLS queuing delay is that introduced over the interdomain link with the AS from which the advertisement was received.

Bandwidth (BWQA) QoS Attribute: The bandwidth available to the prefix specified in the NLRI field. Local QoS classes are assumed to have sufficient bandwidth assigned[3] so the only restriction is the SLS capacity, thus the value advertised becomes:

```
Advertised BWQA = min (incoming advertisement BWQA, offered SLS capacity);
```

where offered SLS capacity is a portion of the bandwidth allocated to the m-QC on the interdomain link with the AS from which the advertisement was received.

Throughout the experiments we examine a number of route selection policies which make use of various combinations of QoS attributes. For added variability of policies we also use a QoS attribute equivalence margin. This margin effectively introduces a comparison granularity to QoS attributes.

We examine a number of route selection policies which make use of various combinations of QoS attributes. For added variability of policies we also use a QoS attribute equivalence margin. This margin effectively introduces a comparison granularity to QoS attributes.

In the following simulation runs, QA equivalence is calculated by:

```
if( floor( MessageA_QA/QAmargin ) = floor( MessageB_QA/QAmargin ) )
```

then the incoming message QAs (MessageA_QA and MessageB_QA) are equivalent and the decision must be performed on the next metric. For example, applying a one-way delay equivalence margin of 70 ms to the example in Table 1, R2 and R3 are equivalent with regards to one-way delay. With a margin of 90 ms, R1, R2, and R3 are equivalent.

Although results were obtained for many cases of equivalence margin value, for clarity the comparison graph concentrates on results from a more limited set of values, selected to highlight the major differences between the selection policies.

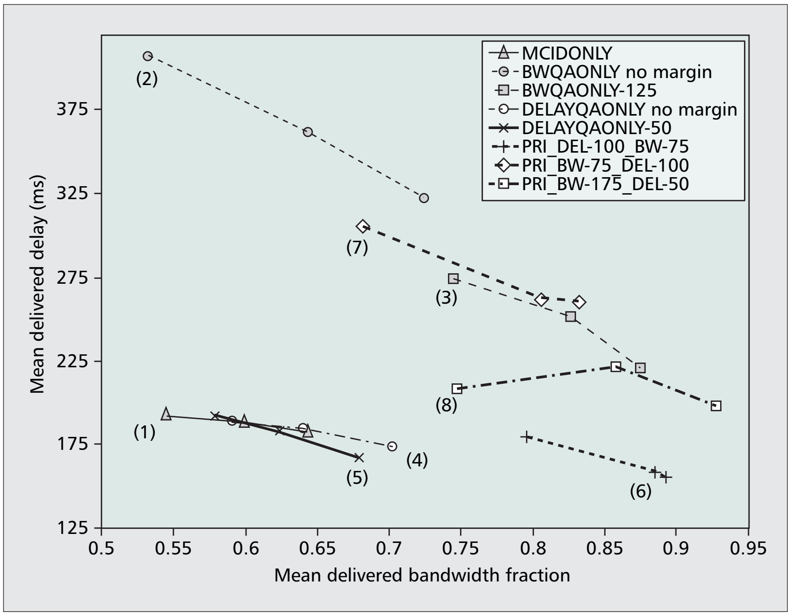
The route selection processes examined here are explained below. Note that we are investigating the performance of the dynamic q-BGP route selection process here rather than investigating an administratively set routing policy; therefore, LOCAL_PREF values are assumed to be the same for all routes.

- M-QC identifier only: routing decisions within the m-QC plane are based first on AS Path length, and use ASN (AS number) as a tie breaker. Given that the simulations focus on a single m-QC plane these tests are effectively without any additional QoS information injected into q-BGP and are therefore equivalent to classical BGP. This policy is denoted as MCIDONLY.
- Single QoS attributes of either DELAYQA-only or BWQA-only: the routing decision is performed based on the QoS attribute first (within the boundaries of the QAmargin), and then on AS path length and ASN. These policies are denoted as DELAYQAONLY and BWQAONLY.
- A two-level priority scheme where, depending on the priorities specified in the policy, either one of DELAYQA or BWQA is checked first, and then if found equivalent (depending on the bandwidth (BW) and one-way-delay (DEL) parameters and margins) the other QA is checked. If that, too, is equivalent the decision is the based on AS path length and the ASN. These policies are denoted as PRI_BW-*s*_DEL-*t* for when BWQA has a higher priority than DELAYQA, and PRI_DEL-*t*_BW-*s* when DELAYQA has a higher priority than BWQA. The *s* denotes BWQA equivalence margin parameter and *t* the DELAYQA equivalence margin.

To provide realistic and useful SLS capacities for simulation purposes, a SLS capacity generator was used in these experiments to provide a base-line SLS capacity matrix in which one routing configuration is capable of satisfying the input traffic demand exactly. To avoid favouring shortest-path routing strategies, the SLS capacity generator allocates capacity to interdomain links on routes that do not always follow the shortest path. It would be very difficult for any routing heuristic to find the precise (non-shortest-path) routing configuration that was used to generate the SLS capacity matrix. An SLS scaling factor is then used to scale the capacities. Scaling the capacities is equivalent to overprovisioning above the original base-line capacity matrix. The number of paths with available resources increases with the scaling factor and it becomes easier for the routing policy to find a suitable routing configuration. The performance of different q-BGP selection policies can be compared by examining the SLS scaling factor required to achieve equivalent end-to-end QoS.

While there are no truly representative Internet-like topology generators, the BRITE Barabasi–Albert (BA) model (http://www.cs.bu.edu/brite/) was used to generate the test topologies. This creates power-law compliant topologies when its preferential attachment option is used, and it has been shown that the Internet is also a power-law compliant topology at the AS level [10]. Network sizes (number of ASs) of 100 were used in the following simulation experiments. Since we wanted to see the effect of q-BGP policies on end-to-end performance, a simple M/M/1 queuing model was used to approximate the effect on delivered QoS of queuing delays at interdomain links. It was also assumed that all demands were inelastic. The measurements taken are the delivered end-to-end one-way delay (the sum of local QoS class delays across each AS, interdomain link queuing and propagation delays) and the mean delivered

[3] *SLS agreements between ASs imply that each AS undertakes to engineer its local resources to honor the QoS and traffic quantity clauses for the m-QC.*



■ **Figure 2.** *The effect of q-BGP selection policy on delivered delay and delivered bandwidth fraction (the portion of offered traffic that arrives at the destination). MCIDONLY (1) is a policy based purely on m-QC id (effectively a single instance of classical BGP, like the current Internet), BWQAONLY (2, 3) is a policy based only on a bandwidth QoS attribute, DELAYQAONLY (4, 5) is a policy based purely on a delay QoS attribute, and PRI_DEL-*_BW-* (6) is a policy where route selection is based first on delay, and then bandwidth, and PRI_BW-*_DEL-* (7,8) is where selection is first on bandwidth, then delay. The numbers in the policy names denote the equivalence margin size. Each point in every three-point trace is for 50t, 100, and 150 percent SLS overprovisioning (from left to right) over the baseline SLS capacities.*

bandwidth fraction: the portion of offered traffic that is delivered to the destination. Throughout these experiments the values of QoS attributes propagated were based on static predefined values, and not on dynamic or measured values.

Figure 2 shows a scatter plot of mean delivered delay against mean delivered bandwidth fraction for a range of q-BGP route selection policies. The results are shown for three SLS scaling factors of 1.5, 2.0, and 2.5 — the three points that run from left to right on each of the curves.

Single QoS Attribute q-BGP Policy: Bandwidth

The policy of selection based on BWQA-only with no equivalence margin (2) delivers higher bandwidth fractions than MCID-only for higher SLS scaling factors, but performs worse than MCID-only in congested networks (i.e., low SLS scaling factor). The reason for the latter is due to a phenomenon that we call QA-rush: where resources (ASs, interdomain links) with the highest advertised QoS (e.g., high capacity, or low delay) are selected more often than those with inferior QoS attribute values, resulting in a greater load on those resources, potentially saturating them and increasing delivered queuing delays and packet loss.

In all cases the adoption of the BWQA-only policy shows worse delivered delay than MCID-only, since it selects the largest capacity route at any cost, causing increased congestion and greater queuing delays. As the QA values are static and administratively set, they do not change to reflect this saturation, and consequently the overall performance suffers.

By adding a margin of equivalence (e.g., of 125 bandwidth units as shown for the BWQAONLY-125 (3) curve), the performance is improved in terms of both delivered delay and bandwidth when compared to selection based on the absolute widest path (2). This also out-performs MCID-only (1) in terms of delivered bandwidth fraction but not delay. The policy of using an equivalence margin improves performance by increasing the number of equivalent bandwidth paths and allowing route selection within the set of best bandwidth paths to be done on the basis of the AS path length, thereby adding diversity to the overall routing behaviour.

Single QoS Attribute q-BGP Policy: Delay

The policy of selection based on DELAYQA-only (4) shows negligible improvement over selection based on shortest AS path (MCID-only) in terms of delay and only marginal improvement in terms of delivered bandwidth. One of the reasons for this is that in the simulation scenarios — as in the real world — the shortest AS path is often the one with shortest delay. If the simulated inter-AS topology were selected carefully so that the ASs along shortest-path routes had large local QoS-class delays then a more marked improvement in performance of the DELAYQA-only selection policy might be observed. This is also why there is little difference in the results of DELAYQA-only with and without an equivalence margin: (4) and (5), respectively.

Two Level Priority q-BGP Policy

The best performing route selection policies are those that select paths according to both advertised delay and bandwidth. PRI_DEL-100_BW-75 (6) is first of all selecting paths on the grounds of smallest advertised delay, with a margin of equivalence of 100 ms, and subsequently selecting between these on the basis of widest advertised bandwidth with a margin of equivalence of 75 bandwidth units, falling back on AS path length and finally AS number if a tie breaker is required. This policy delivers the best overall performance in terms of bandwidth and delay at all three SLS scaling factors.

It is interesting to compare PRI_DEL-100_BW-75 (6) to PRI_BW-75_DEL-100 (7) — that is, the same bandwidth and delay margins, but with the priority reversed. In the latter case (7), both delivered bandwidth and delay are worse than the former (6) and worse than selection based on BWQA-only with a wider margin of equivalence (3). This is again caused by the QA-rush, and the rush is alleviated by increasing the BWQA equivalence margin (8).

It can be seen that with different margins of equivalence, a selection policy with the same priority order of QoS attributes can deliver significantly improved delay/bandwidth performance. This can be seen by comparing PRI_BW-75_DEL-100 with PRI_BW-175_DEL-50. It appears, therefore, that it is better for the path-selection process not to be too narrow in its choice of the set of best paths on the highest-priority QoS attribute so that more potential paths are passed to the selection step based on the second priority attribute and therefore a greater chance of finding a good path according to the second priority QoS attribute.

It is therefore important that the policies for selecting QoS-aware paths are carefully considered. We have shown that selecting paths based on bandwidth can deliver higher bandwidth but also significantly higher delay than the best-effort (BE) Internet; and that selecting paths based on delay provides only a small improvement on the BE Internet, since the shortest AS path is effectively the one with the lowest delay. Further, if QoS parameters are compared too strictly, use of QoS information can in fact hinder delivered QoS. However, significant QoS improvements over the BE Internet are obtained if two QoS parameters (delay and bandwidth) are used to select paths, with better overall QoS obtained by prioritizing delay over bandwidth, rather than bandwidth over delay.

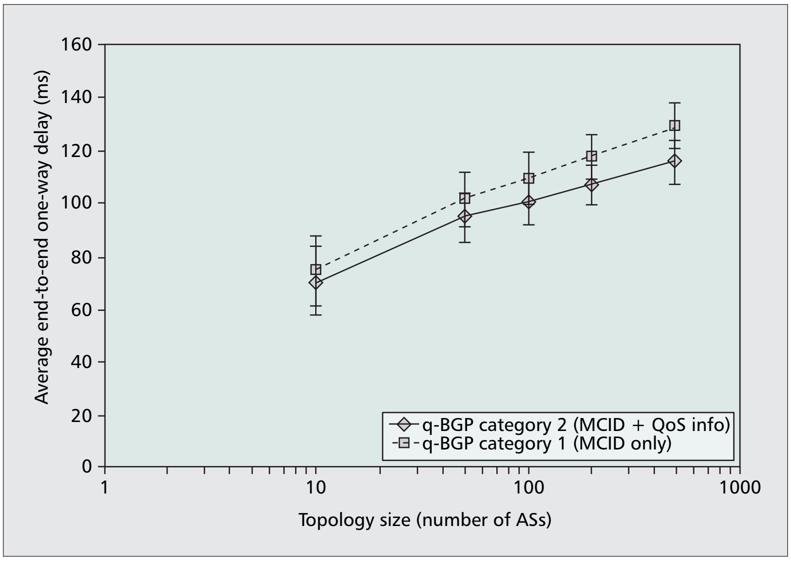
Q-BGP AND QOS AT LARGER SCALES

In this second set of simulation runs we examined the behaviour of q-BGP in larger networks. Detailed modelling of internal AS topologies was avoided by assuming that intra-AS traffic treatment caused a constant average one-way delay between all border routers. The average delay was constant for a given AS but varied between ASs with a uniformly random distribution (between 5 and 50 ms). In the set of simulation results presented below it was assumed that no additional delay due to congestion was introduced at either interdomain or intradomain links, which is equivalent to the case of overprovisioned networks — a simplification introduced to reduce computation requirements for large-scale simulation runs. Four different topologies were generated for each topology size (defined by the 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 were created. The results for each topology size are therefore the average over 48 simulation runs of different topologies and intra-AS delay distributions.

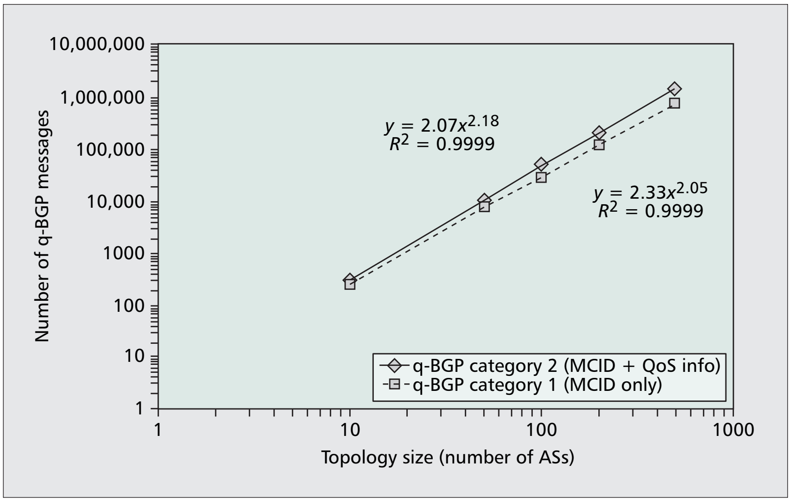
The improvement in delay can be seen in Fig. 3 as a function of AS topology size. It can be seen that the benefit of additional QoS information in q-BGP messages increases with topology size. This is due to a greater number of alternative AS paths between a given source-destination pair in a larger topology. Therefore, the chances of finding an improved path based on one-way delay are increased.

The use of additional QoS information in q-BGP brings an additional overhead in terms of an increased number of q-BGP UPDATE messages. Figure 4 shows the total number of q-BGP messages sent from the first set of bootstrap messages through to a stable routing configuration. It should be noted that there is no message aggregation in these simulations, either on network prefixes or QoS attributes. When the two plots are extrapolated to a topology size of 18,000 ASs the q-BGP category two routing scheme produces approximately three times as many messages as category one q-BGP. The inclusion of additional QoS information 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 m-QC id only. By this we mean that 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 the preferred AS path, on QoS grounds, may not always be the shortest one. Imagine, from the perspective of the AS receiving q-BGP UPDATEs, that the shortest AS path to a particular destination prefix has three AS hops, but the total one-way packet delay (in the data plane) as reported in q-BGP is significantly greater than an alternative five-hop AS path. According to the q-BGP route selection priority rules, the longer path with a smaller delay should be preferred. The q-BGP message received via



■ **Figure 3.** *The effect of topology size on the end-to-end delay experienced by demands.*



■ **Figure 4.** *The number of q-BGP messages sent from initialization until the network settles in a stable state with a full mesh of demands applied.*

the neighbouring AS announcing the three-hop path is likely to arrive earlier than the one from the other neighbouring AS announcing the five-hop path, due to the accumulation of processing time and propagation delay of the q-BGP route selection process at each intermediate AS. In the absence of the five-hop shorter-delay announcement, q-BGP will select the first route and announce this to its peers. On receipt of the subsequent announcement of the shorter-delay path, q-BGP will select the latter route and propagate it to its peers: thereby increasing the total number of q-BGP messages and introducing a transient routing instability. One way of mitigating this would be for ASs to wait for some period to be sure they have received all likely updates, rather than make immediate decisions. This would improve the transient stability of the solution, but at the cost of longer convergence times.

Significant QoS improvements over the BE Internet are obtained if two QoS parameters (delay and bandwidth) are used to select paths, with better overall QoS obtained by prioritizing delay over bandwidth, rather than bandwidth over delay.

Conclusions and Future Directions

This article has proposed a simple scheme for providing qualitative QoS guarantees across multiple domains. The solution has two aspects. Firstly, offline TE processes drive the establishment of QoS-based SLSs with neighbouring ASs for exchanging traffic belonging to one of a limited set of well-known meta-QoS-classes. The establishment of similar bilateral agreements between other ASs will result in the creation of meta-QoS-class planes which extend across the Internet. Secondly, a QoS-enhanced BGP protocol and route selection process ensures that interdomain routing topologies are dynamically constructed for each plane. By itself this yields performance improvements over the standard best-effort Internet by virtue of packets receiving similar differentiated treatment within each AS along the path to its destination.

Simulation has suggested that performance can be enhanced further by including additional QoS information into q-BGP messages and by utilising an enhanced route selection process to select routes based on this QoS information. The results show that delivered end-to-end delay or bandwidth is improved when q-BGP selection policies select paths based on related QoS attributes. However, if the equivalence margin of QoS attributes on competing paths is set too narrow, then a degradation of performance compared to that offered by classical BGP selection policies may be observed due to the phenomenon of "*QA rush*," where the common sections of the paths are overloaded. This can be mitigated by increasing the margin of equivalence when comparing QoS attributes, so that selection is also performed on metrics of a lower priority, such as a second QoS attribute or AS path length. Now, while sufficient quantities of better paths are retained, there are more choices of paths resulting in sufficient routing diversity which alleviates congestion.

It has been demonstrated that different route selection policies result in different delivered performance. It is important to state that this is in addition to any service differentiation implemented by utilising different PHBs/packet-forwarding priorities within the routers of each AS. On the other hand, this result indicates that end-to-end QoS differentiation is achievable even with homogenous forwarding behaviour.

While the performance benefits of QoS-based path selection have been demonstrated, it has also been shown that the cost of the solution is not prohibitive in terms of the number of additional q-BGP UPDATE messages. Simulation results of the DELAYQA-only q-BGP policy with no equivalence margin (which is the worst-case scenario in terms of number of messages) show that the number of q-BGP messages required for stable interdomain routing scales with AS-topology size in a similar way to classical BGP. With larger equivalence margins, the total number of messages is reduced.

Ongoing work includes the investigation of soft partitioning of m-QC planes and interdomain SLS capacities, studying the interaction of multiple routing policies over several m-QC planes, and the impact of injecting dynamically measured QoS metrics into q-BGP while avoiding potential instabilities caused by route-flapping.

Acknowledgments

This work was undertaken under the Information Society Technologies (IST) MESCAL and AGAVE projects, which are partially funded by the Commission of the European Union. We would also like to thank our project partners who have implicitly contributed to the ideas presented here.

References

[1] K. Nichols and B. Carpenter, "Definition of Differentiated Services Per Domain Behaviors and Rules for Their Specification," RFC 3086, Apr. 2001.

[2] P. Georgatsos *et al.*, "Provider-level Service Agreements for Inter-domain QoS delivery," *Proc. 4th Int'l. Wksp. Advanced Internet Charging and QoS Technologies*, Springer, Sept. 2004.

[3] Y. Rekhter and T. Li, Eds., "A Border Gateway Protocol 4 (BGP-4)," IETF RFC 1771, Mar. 1995.

[4] M. Boucadair, "QoS-Enhanced Border Gateway Protocol," IETF Internet draft, July 2005.

[5] P. Levis *et al.*, "The Meta-QoS-Class Concept: a Step Towards Global QoS Inter-Domain Services," *Proc. IEEE Int'l. Conf. Software, Telecommun. and Comp. Networks*, Oct. 2004.

[6] G. Almes, S. Kalidindi, and M. Zekauskas, "A One-Way Delay Metric for IPPM," RFC 2679, Sept. 1999.

[7] G. Almes, S. Kalidindi, and M. Zekauskas, "A One-Way Packet Loss Metric for IPPM," RFC 2680, Sept. 1999.

[8] R. Srihari, D. Tappan, and Y. Rekhter, "BGP Extended Communities Attribute," Feb. 2005, draft-ietf-idr-bgp-ext-communities-08.txt

[9] K. Patel and S. Hares, "AS Path Based Outbound Route Filter for BGP-4," Dec. 2004, draft-ietf-idr-aspath-orf-07.txt

[10] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-Law Relationships of the Internet Topology," *SIGCOMM '99*, 1999.

Biographies

DAVID GRIFFIN (dgriffin@ee.ucl.ac.uk) is a senior research fellow in the Department of Electronic and Electrical Engineering, University College London (UCL), United Kingdom. He has a B.Sc. degree in electrical engineering from Loughborough University, United Kingdom, and is currently completing a Ph.D. degree in electrical engineering part-time from the University of London. Before joining UCL he was a systems design engineer at GEC-Plessey Telecommunica-

tions, United Kingdom, and then a researcher in telecommunications at the Foundation for Research and Technology — Hellas (FORTH), Crete, Greece.

JASON SPENCER (jsp@ee.ucl.ac.uk) received a B.Eng. degree in electronic engineering and an M.Sc. degree in telecommunications from UCL in 1997 and 1998, respectively. He is currently working toward a 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.

JONAS GRIEM (jgriem@ee.ucl.ac.uk) is completing a Ph.D. degree at UCL. His work involves intradomain as well as interdomain traffic engineering on IP networks. He has been involved in several European and U.K. research projects on IP network management and peer-to-peer networking. Prior to his research work, he completed an M.Eng. degree in electronic and electrical engineering at UCL.

MOHAMED BOUCADAIR (mohamed.boucadair@orange-ftgroup.com) is an R&D engineer within France Telecom Group. He has been involved in IST research projects (TEQUILA, MESCAL, AGAVE) working on dynamic provisioning, interdomain traffic engineering, and voice over IP issues. He is an author and co-author of several Internet drafts in the field of VPN management, dynamic routing protocols, and network configuration.

PIERRICK MORAND (pierrick.morand@orange-ftgroup.com) graduated in 1985 from ENSI Caen and received a degree from the University of Caen. In 1989 he joined the research laboratories of France Telecom where he participated and led various projects (X400, agent-based service platforms, VoIP, VPN, COPS). He was the project coordinator of the IST-MESCAL project which ended in August 2005, and is now the leader of a research group working to develop VoIP services for corporate business.

MICHAEL HOWARTH (M.Howarth@surrey.ac.uk) is a lecturer in networking at the University of Surrey, UK. He holds a Bachelor's degree in engineering science and a D.Phil. in electrical engineering, both from Oxford University, and an M.Sc. in telecommunications from the University of Surrey. Prior to joining Surrey he worked for several networking and IT consultancies. He is a Chartered Electrical Engineer.

NING WANG (N.Wang@surrey.ac.uk) is a post-doctoral research fellow at the Center for Communication Systems Research, University of Surrey. He holds a B.Eng. degree in computing from Changchun University of Science and Technology, China, an M.Eng. degree in electronic engineering from Nanyang Technological University, Singapore, and a PhD degree in electronic engineering from University of Surrey, UK. His major research interests include Internet QoS provisioning, traffic engineering and network management.

GEORGE PAVLOU (G.PAVLOU@SURREY.AC.UK) is Professor of Communication and Information Systems at the Centre for Communication Systems Research, University of Surrey, United Kingdom, where he leads the activities of the Networks Research Group. He holds an M.Eng. in electrical and mechanical engineering from the National Technical University of Athens, Greece, and M.Sc. and Ph.D. degrees in computer science from UCL. His research interests include network dimensioning, traffic engineering and management, quality of service, multimedia service control, mobile ad hoc networks, programmable networks, and communications middleware.

ABOLGHASEM (HAMID) ASGARI (Hamid.Asgari@thalesgroup.com) is an assistant chief engineer with Thales Research and Technology, United Kingdom, specializing in data communication systems/networks. He holds a B.Sc. degree from Beheshti University, Tehran, Iran, and an M.Sc. degree from the University of Auckland, New Zealand, both in computer science, and a Ph.D. degree in the design and analysis of ATM networks from the University of Wales, Swansea.

PANOS GEORGATSOS (pgeorgat@algonet.gr) received a B.Sc. degree in mathematics from the National University of Athens in 1985 and a Ph.D. degree in computer science from Bradford University, United Kingdom, in 1989. He is currently working for Algonet S.A., Athens, Greece, as head of the R&D Group in telecommunications. His research interests are in the areas of service management, network planning, resource dimensioning, dynamic routing, analytical modeling, and architectures for distributed systems.

While the performance benefits of QoS-based path selection have been demonstrated, it has also been shown that the cost of the solution is not prohibitive in terms of the number of additional q-BGP UPDATE messages.