

AN APPROACH TO INTER-DOMAIN TRAFFIC ENGINEERING

Geoffrey Cristallo	<u>Christian Jacquenet</u>
Alcatel	France Telecom R&D
Francis Wellespein, 1	42, rue des Coutures - BP 6243
2018 Antwerp	14066 Caen Cedex 4
Belgium	France
+32 (0)3 240 7890	+33 2 31 75 94 28/+33 2 31 73 56 26
geoffrey.cristallo@alcatel.be	christian.jacquenet@rd.francetelecom.com

Abstract

In this paper, we discuss the use of the Border Gateway Protocol, version 4 (BGP4), to exchange Quality of Service (QoS) information between domains. The propagation of this information across autonomous systems should contribute to the dynamic computation and the selection of routes that will participate in the enforcement of end-to-end QoS policies. This paper proposes the use of an additional BGP4 attribute, named the QOS_NLRI attribute, which will explicitly convey the QoS information related to the reachability of (a set of) destination IP prefixes, so that it may influence the BGP route selection process accordingly. We also present the preliminary results of a simulation work that aims at not only validating the technical feasibility of the approach, but also the scalability and the stability of an IP network, where the BGP route flapping conditions could easily be affected by the processing of such information.

Keywords: BGP4, inter-domain, traffic engineering, quality of service

1. Introduction

The Internet is becoming a privileged support for a wide range of IP service offerings, ranging from dial-up access to more sophisticated offers, such as Virtual Private Networks. The success of some of these services is now conditioned by the ability of the service provider to commit on the provisioning of a guaranteed level of quality, which will possibly be negotiated with the customer, depending on the network resource availability, among other considerations.

The capacity of a service provider to manage the network resources he's responsible of according to the QoS requirements of his customers has recently yielded some investigation in the field of intra-domain traffic engineering (see [1], for example), where most of the effort is currently focused on the exploitation of the implicit traffic engineering capabilities of Multi-Protocol Label Switching (MPLS, [2]).

Nevertheless, the concept of a multi-service Internet raises several issues as far as the provisioning of end-to-end quality of service over multiple domains is concerned. Indeed, it is not only a matter of exchanging QoS information between domains, but also a question of QoS policies, which may dramatically differ from one domain to another. Furthermore, the actual enforcement of a given QoS policy based upon an agreed (and ideally standardized) set of QoS parameters may differ from one domain to another.

On the other hand, the BGP protocol ([3]) has been used for the last decade or so to exchange reachability information between autonomous systems, based upon the manipulation of a set of attributes to describe the characteristics of a route that leads to a given destination. This inter-domain routing protocol has become a *de facto* standard for exchanging route information between the domains of the Internet, and we believe that BGP represents one of

the key components for the enforcement of end-to-end QoS policies, whereas QoS information could be propagated throughout the Internet by means of a specific attribute, hence possibly influencing the BGP route selection process for the computation of traffic-engineered paths.

This paper describes the contents and the use of the so-called "QOS_NLRI" attribute ([4]) that will be conveyed by BGP UPDATE messages. It also presents the preliminary results of a simulation work that will be pursued to better qualify the scalability of this approach.

The paper is organized as follows:

- Section 2 introduces some discussion on the motivation for exchanging and processing QoS information between domains,
- Section 3 depicts the overall approach that relies upon the use of the QOS_NLRI attribute and discusses basic operation,
- Section 4 presents the preliminary simulation results that aim at reflecting the technical feasibility of the approach,
- Finally a concluding section depicts the remaining issues, and how they should be addressed within the context of an ongoing simulation and development work.

2. Motivation for exchanging QoS information between domains

Value-added IP service offerings that are likely to be deployed over the Internet often imply the negotiation of QoS requirements between a customer and a provider, customers possibly being providers themselves. Such QoS information includes parameters like one-way transit delays, inter-packet delay variations, loss rates, *etc.* ([5]), and they can be inferred by a DiffServ Code Point (DSCP, [6]) marking indication, so that a given destination (an IP prefix or a host) could be reachable by different routes, depending on the level of quality both customer and provider have agreed upon.

The enforcement of end-to-end QoS policies is therefore conditioned by the provisioning of resources across domains. The characteristics of these resources will have to comply as much as possible with contractual QoS requirements, which means that the routers involved in the forwarding of the corresponding traffic should be aware of this QoS information, so that it might influence their routing decisions accordingly.

The following figure depicts an example where a standard BGP decision process, based on the contents of the AS_PATH attribute, would yield the selection of a route to Network A with transit delay characteristics that appear sub-optimal, when compared to other options.

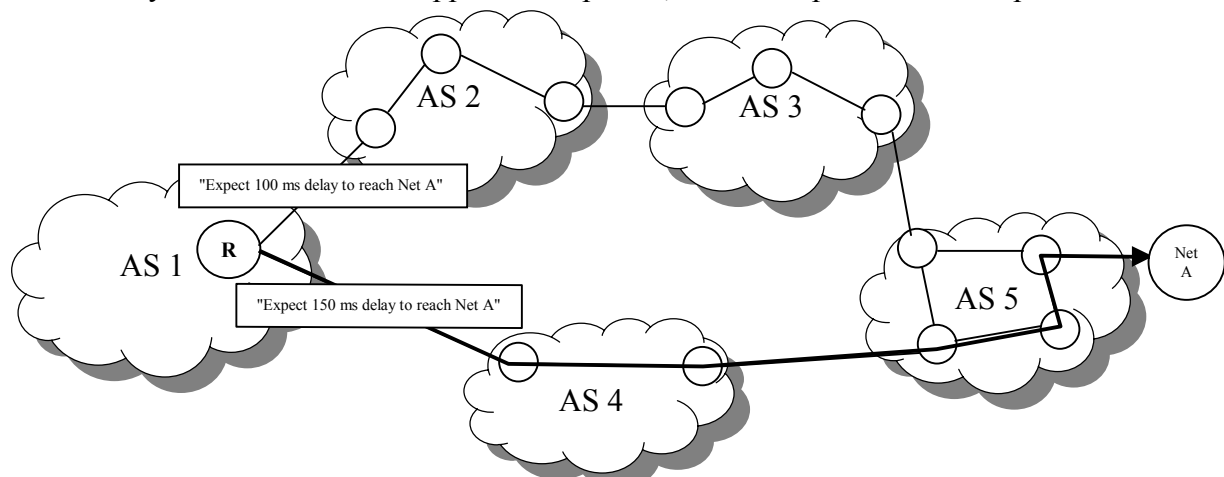


Figure 1: sub-optimal route selection to network A.

Indeed, from router R' perspective, the "route to network A" announcements sent by its two peering points (one BGP peer belonging to AS#2, the other one belonging to AS#4) would lead to the selection of the route (marked in bold type in the above figure) through AS#4, because of the contents of the respective AS_PATH attributes (the UPDATE message coming from AS#2 would indicate AS#2, AS#3 and AS#5 as the domains to cross to reach network A, while the UPDATE message coming from AS#4 would contain AS#4 and AS#5 only).

If we now assume that the selected route to network A experiences a 150 milliseconds transit delay, while a route that would cross AS#2, AS#3 and AS#5 experiences a 100 milliseconds transit delay, then the route that has been selected by router R according to the standard BGP decision process is a sub-optimal choice from this specific QoS parameter perspective.

Generally speaking, the BGP attributes that have been specified so far enable the enforcement of a high-level sort of inter-domain routing policy, where service providers can influence the selection of the adjacent domain to reach a given destination. Nevertheless, this is the kind of information that cannot be propagated across domains, and that currently lacks of "QoS-related" knowledge, based on the parameters that have been introduced in the beginning of this section.

We believe there is a need for a finer granularity, where service providers would have the ability to exchange information about the classes of service they currently support, the destination prefixes that can be reached by means of traffic-engineered routes that have been computed and selected within their domain, as well as any kind of QoS indication that may contribute to the enforcement of end-to-end QoS policies.

From this perspective, the BGP4 protocol appears to be a possible vector for conveying QoS information between domains.

3. An approach to inter-domain traffic engineering

3.1. Requirements

The choice of using the BGP4 protocol for exchanging QoS information between domains is not only motivated by the fact this is currently the only inter-domain (routing) protocol activated in the Internet, but also because the manipulation of attributes is a powerful means for service providers to announce QoS information with the desired level of precision. The approach presented in this paper relies upon the use of an additional attribute, and has identified the following requirements:

- *Keep the approach scalable.* The scalability of the approach can be defined in many ways that include the convergence time to reach a consistent view of the network connectivity, the number of route entries that will have to be maintained by a BGP peer, the dynamics of the route announcement mechanism (*e.g.*, how frequently and under which conditions should an UPDATE message containing QoS information be sent?), *etc.*
- *Keep the BGP4 protocol operation unchanged.* The introduction of a new attribute should not affect the way the protocol operates, but the information contained in this attribute may very well influence the BGP route selection process.
- *Allow for a smooth migration.* The use of a specific BGP attribute to convey QoS information should not constrain network operators to migrate the whole installed base at once, but rather help them in gradually deploying the QoS information processing capability.

3.2. The QOS_NLRI attribute

The proposed approach relies upon an additional BGP4 attribute, named the QOS_NLRI (QoS Network Layer Reachability Information) attribute. The QOS_NLRI attribute is an optional and transitive attribute, whose format is depicted in figure 2.

QoS Information Code (1 octet)
QoS Information Sub-code (1 octet)
QoS Information Value (2 octets)
QoS Information Origin (1 octet)
Address Family Identifier (2 octets)
Subsequent Address Family Identifier (1 octet)
Network Address of Next Hop (4 octets)
Network Layer Reachability Information (variable length)

Figure 2: the QOS_NLRI attribute.

The contents of the QOS_NLRI attribute ([4]) have been designed so that:

- There should be enough room to convey any kind of QoS information. The QoS information that has been identified so far includes transit delay information, loss rate, bandwidth information and Per Hop Behavior (PHB, [7]) identification,
- The QoS information is tightly related to the destination prefixes that are contained in the NLRI field of the attribute, so as to allow for a route-level granularity (instead of an AS-level granularity, as mentioned in section 2),
- The QoS information can be associated to other network protocols than IPv4, including its version 6 ([8]) whose header format includes implicit traffic engineering capabilities that should allow for an even finer level of granularity.

3.3. Basic operation

The QOS_NLRI attribute is conveyed in BGP UPDATE messages that are exchanged between peers of different domains. BGP routers that are capable of processing the information contained in the QOS_NLRI attribute can provide this indication to their peers by means of the Capabilities Advertisement mechanism, as defined in [9].

By processing the QoS information contained in the QOS_NLRI attribute, we basically mean that:

- The QoS information can be altered as long as it traverses the Internet. For example, one-way transit delay information can be modified by means of additive operation, whenever the information is propagated hop-by-hop between domains,
- The QoS information can actually influence the BGP route selection process.

Because it is an optional and transitive attribute, the QOS_NLRI attribute will be propagated across domains, even though there may be peers along the path that are unable to process the information conveyed in the attribute. In this case, the Partial bit of the attribute will be set by such peers, meaning that the information propagated by the attribute is incomplete.

How the QOS_NLRI attribute is actually fulfilled is clearly out of the scope of this paper, although one can think of many means, including:

- The use of a classical redistribution scheme, where traffic engineered routes that have been computed within a domain thanks to the use of an Interior Gateway Protocol (IGP)

with traffic engineering capabilities (like those defined in [10], for example) will be redistributed into BGP,

- The use of dynamic provisioning schemes, like the one depicted in [11],
- The use of protocols such as the One Way Delay measurement Protocol (OWDP, [12]) for the actual measurement of specific QoS indicators like transit delays,
- The use of specific counters, like the SNMP (Simple Network Management Protocol, [13]) interface counters that can provide statistical information on the perceived bandwidth of a given link, for example.

4. Simulation work

4.1. Principles

The simulation work aims at validating the technical feasibility of the approach (hence qualifying the added value introduced by the use of the QOS_NLRI attribute). It is also intended to focus on the scalability of the approach, within the context of the enforcement of inter-domain traffic engineering policies.

The preliminary results of this simulation are presented in this section of the paper, and they have been obtained thanks to the following two-step procedure:

- Model a BGP peer that has the ability to process the information conveyed in the QOS_NLRI attribute, so as to influence the route selection process.
- Model an IP network composed of several autonomous systems, each AS being composed of several BGP peers. Some of them have the ability to process the information conveyed in the QOS_NLRI attribute (as per section 3.3), the others don't.

From a simulation perspective, the aforementioned QoS-based route selection process is currently based upon the manipulation of the Partial bit of the attribute, together with a specific kind of QoS information, which is the one-way transit delay.

Upon receipt of an UPDATE message that contains the QOS_NLRI attribute, BGP peers will first check the contents of the Loc_RIB (Local Routing Information Base) to see if a route already exists towards the destination described in the NLRI field of the attribute. If not, the route described in the QOS_NLRI attribute will be installed in the peer's routing table. If yes, the route selection process has been designed as follows:

- If both routes are said to be either incomplete (Partial bit has been set) or complete (Partial bit is unset), the route with the lowest delay will be selected,
- Otherwise, a route with the Partial bit unset is always preferred over any other route, even if this (complete) route reflects a higher transit delay.

If ever both Partial bit and transit delay information are not sufficient to make a decision, the standard BGP decision process (according to the breaking ties mechanism depicted in [3]) is performed.

4.2. Preliminary results

The following table reflects the first results obtained from a simulation network composed of 9 autonomous systems and 20 BGP peers. Three parameters have been taken into account:

- The percentage of BGP peers that have the ability to process the information conveyed in the QOS_NLRI attribute,

- The transit delays "observed" (and artificially simulated) on each transmission link: the higher the delays, the lower the percentage of serviced QoS requirements,
- The QoS requirements themselves, expressed in terms of delay: as such, the strongest requirements (*i.e.* the lowest delays) have less chance to be satisfied.

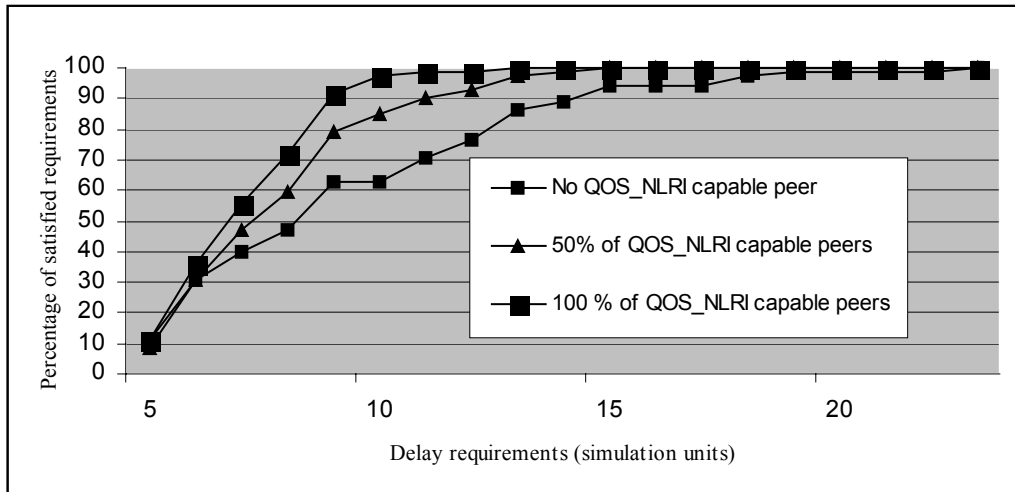


Figure 3: preliminary results of QOS_NLRI-capable peer simulation.

Figure 3 shows the impact of the introduction of QOS_NLRI-capable peers in the network, where the enhanced route selection process results in an increase of the percentage of satisfied QoS requirements, especially in the region where these requirements are stronger.

While this figure demonstrates the technical feasibility of the approach, the results presented here remain obviously preliminary, as discussed in the next section.

5. Conclusion

Although the results presented in the previous section are encouraging, this BGP4-based approach to inter-domain traffic engineering raises issues that remain un-addressed, like the scalability of the approach and the QoS route aggregation capabilities of enhanced BGP peers, so that its introduction in operational networks remain compliant with the requirements and directions that have been identified in [14].

The simulation work is currently ongoing with the manipulation of other QoS information (like the PHB identification), as well as a strong focus on the qualification of the scalability. This paper has presented a possible approach to inter-domain traffic engineering, which has proven its feasibility without modifying the BGP state machine as described in [3]. It allows smooth migration strategies because of the transitive nature of the QOS_NLRI attribute as well as the preservation of a BGP backward compatibility.

6. Acknowledgments

Part of this work has been funded by the European Commission, within the context of the TEQUILA (Traffic Engineering for QUALity of service in the Internet, at LARge scale, [15]) project, which is itself part of the IST (Information Society Technologies) research program.

The authors of this paper would like to thank all the partners of the TEQUILA project for the fruitful discussions that have been conducted within the context of the traffic engineering specification effort of the project.

7. References

- [1] Fortz, B., Thorup, M., "Internet Traffic Engineering by Optimising OSPF Weights", IEEE INFOCOMM 2000, March 2000.
- [2] Rosen, E., Wiswanathan, A., Callon, R., "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.
- [3] Rekhter, Y., Li, T., "A Border Gateway Protocol 4 (BGP-4)", RFC 1771, March 1995.
- [4] Cristallo, G., Jacquenet, C., "Providing Quality of Service Indication by the BGP-4 Protocol: the QOS_NLRI Attribute", draft-jacquenet-qos-nlri-03.txt, Work in Progress, July 2001.
- [5] Demichelis, C., Chimento, P., "IP Packet Delay Variation Metric for IPPM", draft-ietf-ippm-ipdv-08.txt, Work in Progress, November 2001.
- [6] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.
- [7] Black, D., Brim, S., Carpenter, B., Le Faucheur, F., "Per Hop Behavior Identification Codes", RFC 3140, June 2001.
- [8] Deering, S., Hinden, R., "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.
- [9] Chandra, R., Scudder, J., "Capabilities Advertisement with BGP-4", RFC 2842, May 2000.
- [10] D. Katz, D. Yeung, K. Kompella, "Traffic Engineering Extensions to OSPF", draft-katz-yeung-ospf-traffic-06.txt, Work in Progress, October 2001.
- [11] Jacquenet, C., "A COPS Client-Type for IP Traffic Engineering", draft-jacquenet-ip-te-cops-02.txt, Work in Progress, June 2001.
- [12] Shalunov, S., et al., "A One-way Delay Measurement Protocol", draft-ietf-ippm-owdp-03.txt, Work in Progress, February 2001.
- [13] Case, J., Fedor, M., Schoffstall, M., Davin, J., "A Simple Network Management Protocol", RFC 1157, May 1990.
- [14] Labovitz, C., Ahuja, A., Wattenhofer, R., Venkatachary, S., "The Impact of Internet Policy and Topology on Delayed Routing Convergence ", IEEE INFOCOMM 2001, April 2001.
- [15] <http://www.ist-tequila.org>.