Journal of Computing and Information Technology - CIT 19, 2011, 4, 223–229
doi:10.2498/cit.1002027

A Border-friendly, Non-overlay Mechanism for Inter-domain QoS Support in the Internet

Vitalian A. Danciu[1], Dieter Kranzlmüller[1,2], Martin G. Metzker[1] and Mark Yampolskiy[2,3]

[1]Ludwig-Maximilians-Universität München, Munich, Germany
[2]Leibniz Supercomputing Centre (LRZ), Garching, Germany
[3]German Research Network (DFN), Berlin, Germany

Many services provided over the Internet, like voice over IP and video on demand, increase the demand for assurances concerning the quality of the underlying network. A score of techniques for assurance of quality of service (QoS) have been devised for use within administrative domains. However, when paths cross the border of autonomous systems, assurance of end-to-end QoS remains an unsolved issue. Thereby the key challenge is the establishment of connection-oriented communication flows. We introduce a technique to establish ISO/OSI Layer 3 multi-domain communication paths. The proposed solution does not stress border-routers and is independent of domain-internal policies, while relying on the common forwarding mechanisms.

Keywords: routing, multi-domain switching, new generation networks, NGN, QoS

1. Introduction

The amount and diversity of user-faced applications depending on a good Internet connection quality is steadily growing. Examples are manifold and can be found in the areas of multimedia, like video-on-demand, telecommunication, like voice over IP (VoIP) or job-transfers in Clouds and Grids. In order to support the existing and upcoming services, a technique for quality assurance in Internet is needed.

With the exception of a few service-tailored solutions, today's effectiveness of mechanisms for Quality of Service (QoS) managements is limited by a provider's network border. Connections encountered in the Internet usually cross networks of multiple autonomous systems (AS), leaving true QoS assurance for multi-domain connections an unsolved issue.

Telephone and backbone network providers recognise that true quality assurance is only possible if resources are assigned to communication flows. The amount of these resources should be sufficient for the realisation of required properties and exclusive assignment prevents interferences with other communication flows. Experience made with IntServ supported by the RSVP protocol family further show that resource reservation along a communication path alone is not sufficient, as long as the enforcement of this communication path is not warrantied. Various extensions for RSVP show desperate attempts to cope with path changes in the Internet.

Discussion on New Generation Networks led to the conclusion that connection-oriented paths must become a cornerstone of QoS assurance in packed switched networks. Techniques like MPLS are limited to the administrative domain of a single network service provider. Extensions to MPLS like MPLS-TP or other approaches like the PBB-TE proposal for carrier grade Ethernet have not evolved beyond standardisation stage. Furthermore, introduction of such technologies to the Internet will require large-scale upgrades of network infrastructures, which is not likely to happen within a short time frame.

[0]Authors are listed in alphabetical order.

As the amount and the variety of end-user applications, relying on connections with guarantied quality parameters, are steadily growing, a solution based on existing technologies is needed.

The aspired solution was designed keeping in mind that border routers in the Internet are troubled by very large routing tables and, in general, are being expected to implement every conceivable inter-domain extension and enhancement. Hence, network operators' tolerance for extensions for whatever benefit is limited by the additional requirements imposed on their border routers.

Our contribution to enabling end-to-end QoS on inter-AS paths is an IP-based switching technique that allows providers to coordinate their efforts for providing QoS over the Internet. It is an *opt-in approach* in that it is applicable even if a limited number of AS operators choose to support it. The solution is being designed with acceptance by network operators in mind. The approach relies on regular IP routing at the AS borders and on locally available QoS mechanisms within the network.

In this paper, we focus on the core mechanism of our proposal. Aspects like data model for encoding QoS-relevant information and a protocol for interoperation between AS-providers will be addressed in other dedicated papers.

In the following Section 2, we discuss the requirements of quality-controlled paths in the Internet from both user and operator perspective. We survey existing approaches to inter-AS QoS in Section 3. The approach itself is detailed in Section 4, before we discuss limit cases in Section 5. Section 6 concludes the paper with a review of open questions and future research work.

2. Assumptions and Requirements

AS operators are limited in the policy of their networks only by their own business obligations and by a small number of standards necessary for effective inter-networking. Their legal obligations are usually constrained within national borders that are exceeded by multi-national networks in the common case. It follows that for any new concept to enjoy significant adoption in the Internet, it needs to 1) have a strong financial incentive for operators to implement it, 2) not affect core operations adversely and 3) be inexpensively implementable, 4) scale, if it proves popular. If the popularity of such a concept depends on the acceptance of end-users, as with any end-to-end scheme, it should take their interests into account from its inception.

The Internet is a dynamic structure. At any time, links are added and removed, devices are introduced into or removed from the network, routes change and so on. An end-to-end approach in the network layer must be able to cope with these common events, and it must not rely on a "snapshot" view of the network.

Embracing these assumptions, we formulate the following design objectives that constitute requirements on our approach:

(1) Consideration for intra-AS interests.

- Do not presume to change or introspect intra-AS state, topology, or management information.
- Do not make assumptions regarding the QoS an AS can provide, or if the AS is willing to cooperate at all.
- Support paid-for transit; without it, there is no incentive for transit operators to consider the proposed extensions.

(2) Operations and limitations in the Internet.

- Do not strain routers at AS borders.
- Use the existing routing structure, and add only signalling. Do not transmit payload through an overlay, since this only makes the path transparent, not its quality attributes.
- Handle correctly the normal incidents in the Internet. The approach should be able to cope gracefully with route reconfiguration, inter-AS hops becoming unavailable and orphaned connections.

(3) Consideration for end-point interests.

- Make it easy to use in end-point client applications. Requiring vast changes to applications or operating systems may seriously hamper end-point-side acceptance of the approach.
- Do not make assumptions on what QoS properties an end-user will need for its connection. Such assumptions will limit the applicability of the approach.

3. Related Work

In this section we discuss three categories of related work: inter-AS path-switching using overlay networks, approaches using a specific technology stack and available technologies for implementing QoS in networks.

In [8, 12, 1] different overlay approaches for controlling packet transit over the Internet are introduced. QoS enforcement is implemented using available methods like DiffServ [3], IntServ [4], RSVP [5]. Following these approaches, our requirements for handling incidents and changes in the underlying network can only be achieved by including the providers. This will dictate the QoS technologies and capabilities providers have to employ and offer, contradicting our requirements for having the providers decide which services and quality their networks offer.

A method for establishing end-to-end paths for the specific stack of ATM/MPLS/IP has been elaborated in [15]. This approach cannot fulfill our requirements to leave intra-AS matters entirely to the providers, but clearly shows that establishing end-to-end paths requires a lot of communication and negotiation and that an explicit connection phase is a reasonable solution to this problem. The ongoing development of an MPLS multi-domain version MPLS-TP (formerly known as T-MPLS) is still far away from state of maturity for productive application. In [9, 2, 13, 6] the authors explore transporting IP packets between routers over optical links, with minimal highly use-case specific Layer 2 implementations. These articles show that QoS and resource allocation are achievable if the different approaches can be combined.

In [14, 11, 10] the authors show a measurable increase in QoS, when having control over the selected AS-path between two end-points. Clearly, fixed paths are a key to maintaining QoS and possibly as important as any other employed QoS technologies and protocols. An entire QoS description language is (for example) presented in [7]. An information model for the provisioning enabling end-to-end QoS is presented in [16]. Specifications like these may be used to describe QoS properties and requirements for a common basis in path negotiations and connection establishment.

4. Internet Inter-AS Path-switching

Our approach to provide inter-domain communication channels with QoS properties relies on introducing connection semantics to the traffic pertaining to those channels, while still relying on the standard IP forwarding techniques. We address inter-AS route selection, but leave intra-AS routing to the AS operators, whom we rely upon to ensure a declared level of quality for the traffic passing into/through their networks.

4.1. Conceptual Components

Each AS that wishes to support our QoS mechanism must provide three conceptual components detailed as follows.

1) A reserved, special purpose subnet, i.e. a network prefix dedicated to QoS channels. Its address space need not be the same in every AS.
2) An access function that manages connection end-points in the AS itself. We call a component implementing this function an *access gateway* (AG). AGs directly interact with end-points, to initiate and control end-to-end paths and multiplex the users' traffic onto Layer 3 segment addresses.
3) A forwarding function that differentiates all QoS channel traffic, including transit traffic. We call a component implementing this function a *forwarding gateway* (FG). The FG performs network address translation on incoming packets destination address and forwards them along a path specified for a given QoS-controlled channel.

These components are illustrated in Figure 1, that shows an established path through three AS, where AS 65229 is a transit AS, while AS 65016 and AS 65815 are ISPs of end-points A and B, respectively.

Note that AG and FG functions are shown as implemented on the same physical component. Applications run on the *initiating end-point* (initiator, A) and the *target end-point* (target, B) signal information and requests concerning channel management.

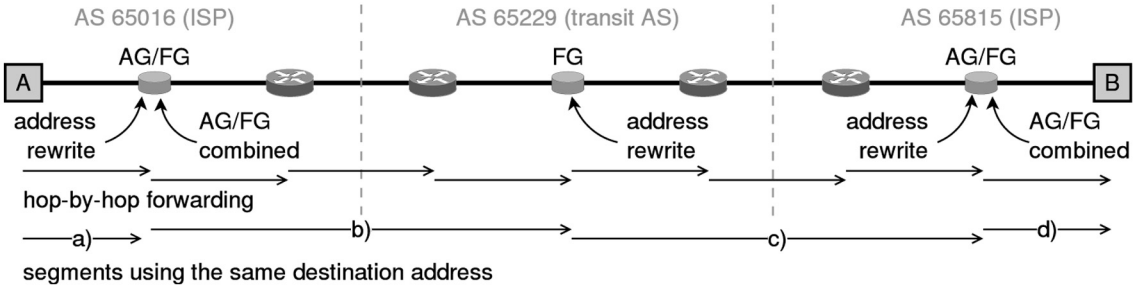

Figure 1. An end-to-end path across three AS. Segments a–d) are defined by FGs

4.2. Protocol Outline

The core of our approach is handing packets from AS to AS along a predetermined path through the Internet.

We propose recursive path establishment as each AS may know only about and be able to negotiate with its direct neighbors.

As the first step, the initiator requests a connection to the target from an AS-local AG, along with requirements for specific QoS. Using regular intra-AS routing, an AG determines the next FG (which is trivial if AG and FG are the same machine) and passes it the path request. If a FG can fulfill a path request, i.e. if the AS it serves can fulfill the requirements regarding QoS, it passes the (modified) request on to the next FG, determined using Internet routing information. This step is repeated by FGs until the target AS, i.e. AS containing the target end-point, is reached.

If a FG is unable to meet a path request, it signals this back to the instance it received the request from. Having received a fail-signal, a component may attempt to find a path via another AS (back-tracking). If a suitable end-to-end path is found, the end-points transmit via their AGs. If no path could be found, the setup of the QoS-controlled channel has failed.

a) Forwarding example

There are three routers functioning as FGs, each rewriting the destination address of incoming packets. This divides the path between A and B into four segments *a*) through *d*). Rewriting destination addresses of incoming packets to segment addresses, announced by FGs, ensures the packets are routed from FG to FG. Combined with Inter-AS (BGP) routing information, it can be ensured that packets are handed from one AS to the next.

In Figure 1 an application on A sends user-data addressed to B. Having crossed segment *a*) the traffic is received by the AG (co-located with the FG) in AS 65016. Recognising the incoming packets as belonging to an established path, their destination addresses are rewritten to the segment address allocated by the FG in AS 65229. Through regular Internet routing the packets are forwarded to AS 65229, segment *b*), where the destination addresses are rewritten again and the packets forwarded to the FG in AS 65815, segment *c*). Here the addresses are rewritten to the original destination address and forwarded to B, segment *d*).

AS operators can control the ingress and egress points of path segments by refining their routing (BGP) configuration and may use the per segment address to implement special QoS transit through their networks.

b) Resource allocation

Allocating resources for path segments shows many characteristics of a Layer 3 end-to-end connection. To coordinate resource allocation we distinguish between four states of resources during the connection phase, illustrated in Figure 2. The states are grouped to identify a connection either as *idle* or *busy*.

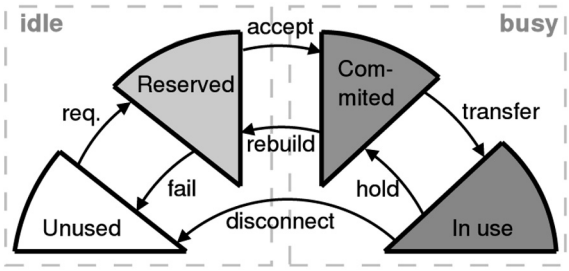

Figure 2. We distinguish four states of connection segment resource allocation

A segment is *unused* when it has been identified on a viable path through the Internet. During allocation, an AS may be *requested* to allocate resources for the connection from the unused segment. At this point resources are *reserved*, but it remains unclear whether the creation of the end-to-end channel will succeed: a later consulted AS may obviate successful establishment of this particular path, thus its establishment will *fail*, causing non-viable reservations to be released during back-tracking.

Once it is certain that resources could be allocated in every AS, the initiator's AG *accepts* the proposed path, resources are *committed* and the requested connection is ready for use. The connection can be considered established now, but user-data is not yet transmitted. Once a connection is *in use*, user-data is *transferred* via the established path.

c) De-allocation

When the end-points determine a connection is no longer needed, they ask the AG to *disconnect* and resources are freed similarly to allocation. When a connection is not in use, FGs *hold* a connection so that the end-points may resume sending data easily, but the AS are given a chance to alter the path. When a connection (or its quality) can no longer be maintained, it has to be *rebuilt* using an alternative (sub-)path, possibly containing other networks. If no suitable path could be found or resource reservation *failed*, the requested connection cannot be established.

5. Discussion

Our idea describes how service providers may employ capabilities of common Layer 3 protocols and routing procedures to enable end-to-end paths through the Internet.

a) Application layer multiplex

The most apparent challenge in Layer 3 channels for Layer 7 applications is multiplexing. A channel meeting QoS requirements must be exclusive for specific Layer 7 flows and provisioned separately from every other Layer 3 traffic. This multiplex is handled by access nodes, but not elaborated in this work. Implementing access nodes as separate components allows pushing the multiplex towards the initiator, maybe even to customer sites. Special treatment of traffic can be better accomplished once packet destinations have been rewritten.

b) Scalability

Initially, a transit AS provider needs to allocate a subnet and a router acting as FG to be able to realise the Layer 3 paths. As destination addresses are first changed by AGs, they must be placed along the "normal" path between two end-points where such connections will be established.

More AGs may be put up as needed and to provide more sophisticated QoS. AGs should be placed near the customers' uplinks when offering this service to initiating customers like end-users. To offer this service to special targets, e.g., a VoIP gateway, AGs should be placed to catch VoIP traffic, but hardly any other traffic. The FG may be expanded to be an entire Layer 2 subnet which is accessed exclusively by traffic of known (and paid for) characteristics.

c) Channel state model

The resource reservation states depicted in Figure 2 are grouped to identify a connection either as *idle* or *busy*. Looking at QoS, the transition from idle to busy is a crucial juncture for providers, because once resources are allocated they are blocked by the connection and may not be used for other traffic, no matter whether the data is actually transmitted or not. If blocked resources were used for other traffic, QoS cannot be guaranteed. On the other hand, today's accounting is based on packets or bytes passing through routers. Our distinction between idle and busy allows to have both: 1) a tentative request/reservation in order to determine paths and 2) definitely committed resources providers can charge customers for.

d) Management integration

Our approach is non-invasive, rather than requiring changes to present components and protocols. We make no assumptions on the QoS that providers can or must provide. Clients are free to request arbitrary attributes from their providers. Of course, paths cannot be established when requirements can't or won't be met by providers. Using Layer 3 segment addresses allows for easy per connection accounting and the explicit connection phase leaves room for quality negotiations between providers. Thus we allow paid-for services and transit. Through embedding segment address subnets into the

normal addressing and thus forwarding behaviour of Internet protocols, we don't add to the tasks of border routers. All functionality is provided by the ISP, thus client applications need little knowledge to use this Layer 3 service. The only hooks client applications need is for requesting and terminating paths. Resource reservation has been divided into four phases to have paths on top of the Internet topology, which always has to be considered subject to change.

FGs may search for alternative paths for local repair or fail-over of connection segments. Thus, planned changes as well as changes in response to faults can be prepared and executed without interfering with our mechanism.

6. Conclusion

Assurance of end-to-end inter-domain network QoS requires cooperation of all network operators between end-points. To secure this cooperation, a QoS management scheme must not infringe on operators' management policies or operations.

The approach we have proposed is designed to work in the Internet, as it externalises the decisions pertaining to the additional function (QoS management) into a special gateway. It requires only minimal configuration changes to the traditional routing infrastructure, and it can be refined and extended independently. It refrains from introducing tunneling and overlay techniques in order to accommodate any QoS technologies that may have been deployed in operators' networks.

We envision this mechanism to provide a generic base for the management of connections in the Internet, extensible with functions different from QoS management. While this paper has addressed the fundamental mechanism itself, practical applicability additionally requires a protocol specification for the exchanges between FG as well as an information/data model serving as a common base for QoS negotiation.

7. Acknowledgment

The authors wish to thank the members of the Munich Network Management (MNM) Team for helpful discussions and valuable comments on previous versions of this paper. The MNM Team directed by Prof. Dr. Dieter Kranzlmüller and Prof. Dr. Heinz-Gerd Hegering is a group of researchers at Ludwig-Maximilians-Universität München (LMU), Technische Universität München (TUM), the University of the Federal Armed Forces and the Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Science. `http://www.mnm-team.org`

References

[1] I. F. AKYILDIZ, T. ANJALI, L. CHEN, J. C. DE OLIVEIRA, C. SCOGLIO, A. SCIUTO, J. A. SMITH, AND G. UHL. A new traffic engineering manager for DiffServ/MPLS networks: design and implementation on an IP QoS testbed. *Computer Communications*, 26(4):388–403, 2003. 2

[2] D. AWDUCHE AND Y. REKHTER. Multiprotocol lambda switching: combining MPLS traffic engineering control with optical crossconnects. *Communications Magazine*, IEEE, 39(3):111–116, 2002. 3

[3] S. BLAKE, D. BLACK, M. CARLSON, E. DAVIES, Z. WANG, AND W. WEISS. An Architecture for Differentiated Service. RFC 2475 (Informational), December 1998. Updated by RFC 3260. 3

[4] R. BRADEN, D. CLARK, AND S. SHENKER. Integrated Services in the Internet Architecture: an Overview. RFC 1633 (Informational), June 1994. 3

[5] R. BRADEN, L. ZHANG, S. BERSON, S. HERZOG, AND S. JAMIN. Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification. RFC 2205 (Proposed Standard), September 1997. Updated by RFCs 2750, 3936, 4495. 3

[6] W. COLITTI, K. STEENHAUT, D. COLLE, M. PICKAVET, J. LEMEIRE, AND A. NOWÉ. Integrated routing in GMPLS-based IP/WDM networks. *Photonic Network Communications*, 31(1):1–15, 2008. 3

[7] SVEND FROLUND AND JARI KOISTINEN. QML: a language for quality of service specification. Technical Report HPL-98-10, HP Labs, February 1998. 3

[8] JIAYUE HE, RUI ZHANG-SHEN, YING LI, CHENG-YEN LEE, JENNIFER REXFORD, AND MUNG CHIANG. DaVinci: dynamically adaptive virtual networks for a customized internet. In *CoNEXT ò08: 2008 ACM CoNEXT Conference*, pages 1–12, New York, NY, USA, 2008. ACM. 2

[9] D.K. HUNTER AND I. ANDONOVIC. Approaches to optical Internet packet switching. *Communications Magazine, IEEE*, 38(9):116–122, 2002. 3

[10] SHU TAO KUAI, SHU TAO, KUAI XU, ANTONIO ESTEPA, TENG FEI, LIXIN GAO, ROCH GUÉRIN, JIM KUROSE, DON TOWSLEY, AND ZHI LI ZHANG. Improving voip quality through path switching. In *Proceedings of IEEE INFOCOM*, pages 2268–2278, 2005. 3

[11] SHU TAO KUAI, SHU TAO, KUAI XU, YING XU, TENG FEI, LIXIN GAO, ROCH GUÉRIN, JIM KUROSE, DON TOWSLEY, AND ZHI LI ZHANG. Exploring the performance benefits of end-to-end path switching. In *Proceedings of IEEE ICNP*, pages 418–419, 2003. 3

[12] ZHI LI AND PRASANT MOHAPATRA. QRON: QoS-aware routing in overlay networks. *IEEE Journal on Selected Areas in Communications*, 22(1):29–40, 2004. 2

[13] G. MOHAN AND E. CHENG TIEN. QoS routing in GMPLS-capable integrated IP/WDM networks with router cost constraints. *Computer communications*, 31(1):19–34, 2008. 3

[14] SHANSI REN, LEI GUO, AND XIAODONG ZHANG. ASAP: an AS-aware peer-relay protocol for high quality VoIP. *International Conference on Distributed Computing Systems*, 0:70, 2006. 3

[15] S. SÁNCHEZ-LOPEZ, X. MASIP-BRUIN, J. SOLE-PARETA, AND J. DOMINGO-PASCUAL. Fast setup of end-to-end paths for bandwidth constrained applications in an IP/MPLS-ATM integrated environment. *Computer Networks*, 51(3):835–852, 2007. 3

[16] M. YAMPOLSKIY, W. HOMMEL, P. MARCU, AND M. K. HAMM. An information model for the provisioning of network connections enabling customer-specific End-to-End QoS guarantees. *In Proceedings of the IEEE SCC 2010 International Conference on Services Computing*, Miami, USA, July 2010. IEEE Computer Society. 3

Received: June, 2011
Accepted: November, 2011

Contact addresses:

Vitalian A. Danciu
Ludwig-Maximilians-Universität München
Oettingenstr. 67
80538 Munich, Germany
e-mail: danciu@mnm-team.org

Dieter Kranzlmüller
Leibniz Supercomputing Centre (LRZ)
Boltzmannstr. 1
85748 Garching, Germany
e-mail: kranzlmueller@mnm-team.org

Martin G. Metzker
Ludwig-Maximilians-Universität München
Oettingenstr. 67
80538 Munich, Germany
e-mail: metzker@mnm-team.org

Mark Yampolskiy
Vanderbilt University
1025 16th Ave S
Nashville, TN 37212, USA
e-mail: yampol@mnm-team.org

VITALIAN A. DANCIU received his diploma and doctoral degrees from the Ludwig-Maximilians-Universität München, Germany. He has published refereed papers on a range of management topics including networks and services, virtualized infrastructure, IT management processes and policy-based management. A member of the Munich Network Management Team since 2003, he continues to teach and research at his alma mater.

UNIV.-PROF. DR. DIETER KRANZLMÜLLER is full professor of computer science at the Ludwig-Maximilians-Universität München (LMU), member of the board of the Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities, and Scientific Director of the Center for Digital Technology & Management (CDTM). He is member of the Executive Board of the EGI.eu Organisation and German representative on the European Grid Initiative (EGI) Council. He chairs the MNM-Team (Munich Network Management Team), which is engaged in networks and distributed systems in general, and networks, grids, clouds and HPC in particular.

MARTIN G. METZKER graduated from the Technische Universität München with a diploma in computer science. In June 2009 he joined the Munich Network Management Team and became a researcher at Ludwig-Maximilians-Universität (LMU) in Munich, Germany. His research interests are network management, virtualization and quality of service in virtual infrastructures.

MARK YAMPOLSKIY first studied applied mathematics at one of the technical universities in Moscow, before studying computer science at the Technical University Munich (TUM). He successfully defended his Ph.D. dissertation in the area of computer networks and network management. The focus of his research is dedicated to quality assurance in multi-domain network connections. Situated at the Leibniz Supercomputing Centre (LRZ), he is working as member of the Géant research collaboration. Within this collaboration, he is involved in numerous research and service activities tackling network management issues in multi-domain environments. Among other, he is in charge of the design and development of the monitoring system for multi-domain backbone connection Géant E2E Links. Mark Yampolskiy is a member of Munich Network Management (MNM) Team, a team of researchers focusing on various aspects of networking, network management, and collaborative distributed environments.