End-to-End QoS in Interdomain Routing

Nishant Kumar and Girish P. Saraph
Indian Institute of Technology, Bombay
Mumbai, India
nishant@cse.iitb.ac.in, girishs@ee.iitb.ac.in

Abstract—Providing end-to-end Quality of Service (QoS) guarantees for interdomain routing remains a challenge for next generation Internet. As various real-time, mission-critical, and bandwidth-sensitive applications are migrating to the IP Networks the need for end-to-end QoS is becoming acute. Large existing base of BGP compliant networks rules out the acceptability of new routing protocol. However, inherent nature and functionality of BGP makes the QoS extensions rather difficult.

The paper presents a novel approach to achieve end-to-end QoS support by proposing a new Alliance Network model. The most important aspect of the proposed model is its compatibility with existing BGP infrastructure, such that the traditional BGP traffic continues in a normal fashion. The Alliance Network sets-up interdomain paths for premium traffic that require specific QoS guarantees using interdomain MPLS tunnels with resource reservation. The Alliance Network provides enhanced geographical reach and market penetration for the alliance partners through premium service offerings.

Index Terms—Border Gateway Protocol (BGP), Routing, Quality of Service (QoS), Networks, MPLS.

I. Introduction

Internet has grown tremendously in term of capacity, complexity, and the number of attached network domains. Simultaneously, there is migration of various services toward IP networks. Internet traffic is forwarded on a best effort basis without any guarantee on service and performance. This may result in poor service quality for applications requiring service differentiation. The need for timely delivery of real time application like telephony, video conferencing or guaranteed bandwidth for mission-critical applications has led to a high demand of end-to-end QoS guarantees. For such performance guarantees we need to switch from traditional datagram based interdomain routing to flow based or virtual connection based routing model.

QoS routing within an administrative domain or AS is addressed using virtual connections in ATM networks [1] or resource reserved tunnels in MPLS networks [2]. However, many connection requests span across multiple ASes e.g. a multi-national company having offices in many parts of the world and requiring them to be connected through a VPN with a specific level of performance. Hence, the QoS routing at the interdomain level is essential for end-to-end QoS guarantees. Such a task is significantly more complex and challenging than intradomain QoS routing. Internet has evolved with distributed path selection mechanisms, which maintain full independence of individual ASes. It only addresses the basic reachability needs without taking into account individual connection requirements. The internal network topology of an AS, its business relationships with other ASes, and its policy implementations are treated as proprietary information and interdomain routing must be done without detailed knowledge of the network topology, policies, and performance.

Border Gateway Protocol (BGP-4) is the current de-facto standard for interdomain routing protocol [3] [4]. It is essentially a path-vector protocol working under limitation of local AS policies. The policies of an AS are predominantly defined based on the types of business relationship with its neighboring ASes. A BGP router may receive multiple route advertisements (AS paths) for a particular network or subnet from different border routers from neighboring ASes. The router applies the network administrator-specified import policies to filter the acceptable routes and manipulate the attributes associated with them. The route advertisements consist of various attributes like local-pref, AS-Path, MED etc. The router then applies BGP decision process to find out the best path to reach a given network among the routes it has learnt. The decision is based on the policies and attributes attached to each route advertisement. Finally, the router applies the export policy to manipulate the attributes and to decide which routes are to be advertised to the neighboring AS routers.

Currently in Internet, there are two common types of business relationship between ASes: customer-provider and peer-peer [5]. In a customer-provider relation a customer AS pays for transit facility and access to other networks through the provider network AS. In a peer-to-peer relationship two ASes provide only access to each others customer networks in a non-transitive way. The third type of relationship, sibling-sibling, is not very common. The likely business relationships between the ASes can be deciphered from the entire set of AS paths based on the typical policy implementations [6].

Since BGP path selection is based on local AS policy implementation, it is locally optimal but globally sub-optimal. The local path selection does not take into account the QoS requirements of the end-user. It also applies local pruning at each intermediate AS, thereby only one locally selected path is advertised further downstream rather than multiple available options with different performance attributes. This approach has inherent limitations for the end-to-end QoS routing. It also has scalability issues due to the large routing table of sizes, which are in excess of 230,000 entries [7]. It leads to a number of problems, such as long convergence time after link failures, route flaps, forwarding loops, etc. [8] [9].

Traditionally BGP was designed to exchange only the reachability information. It is being extended to address the issues like long convergence [8], scalability, QoS, TE [10]. Also, [11] [12] describes the current approaches for end-to-end QoS support. Reference [11] approach is based on
extending BGP to advertise interdomain QoS routing information. Since, the requirements for QoS at interdomain level puts a large volume of information to be processed at routing domain [14] [15], it is not scalable for the entire global Internet. Currently, it is possible to set up inter-provider, interdomain MPLS tunnels [12] [13] with specific resource reservations and QoS guarantees. However, in the absence of any accepted framework it requires protracted discussions and agreements between two service providers.

This paper presents a novel approach to achieve end-to-end QoS support by proposing new Alliance Network model for a set of interconnected ASes. Alliance Network uses an overlay model to interoperate with the BGP-based Internet. It also addresses the implementation scheme, business perspective, and advantages of the proposed alliance model. This paper is organized as following: Section II presents an evolutionary path for the BGP based Internet to support additional QoS routing scheme. The alliance model is presented in section III along with the implementation issues. Section IV discusses the business perspective for the alliance network and gives various advantages of the proposed model. The paper is concluded with a brief summary.

II. Evolutionary Path for BGP-based Internet

Since BGP is a de-facto standard for interdomain routing, it is the main protocol supported at the core of the Internet. All the core router infrastructure and communication systems are BGP compliant. It is not possible to change the existing infrastructure to support new protocol or system model. Thus, it is important to make sure that any new model is compatible with the BGP and coexists with it. Also, it is preferable from the cost point of view that the new model is only a software based upgrade to the existing infrastructure rather than it requiring new hardware. Another important factor is that the regular Internet traffic must continue to flow in a traditional manner without any interruption and only premium traffic that requires the end-to-end QoS guarantees should use the new facility by paying premium rates.

In order to ensure the evolutionary path it is important to revisit the BGP approach. BGP routers in an AS make TCP based peering sessions with their counterparts in other ASes through external BGP (eBGP) sessions and exchange network reachability information. These routers use internal BGP (iBGP) to distribute the routes learnt through eBGP to all the BGP routers within the AS. These routers are in a full logical mesh of iBGP sessions with each other. The \(N^2\) scaling problem of iBGP sessions, in a large AS can be addressed by using a route reflector [3]. A route reflector coordinates the iBGP information exchange. For a larger AS multiple route reflectors may be used. However, it is important to ensure that they are in consistent state of routing information, otherwise it could lead to poor path selection and forwarding loops.

Routing Control Platform, RCP [16] is a promising solution for an evolutionary upgrade of the BGP based Internet. It separates the interdomain routing control plane from the forwarding plane. RCP server may be viewed as an enhanced version (i.e. software upgrade) of a common route reflector. It centralizes the BGP import and export policy implementation for the AS administrator and frees up other routers for the forwarding tasks. The RCP in an AS domain may be deployed in three stages discussed below [16]:

1) First stage of the RCP deployment enhances the route reflector functionality by coordinating the filter policy implementation in the AS. RCP learns the BGP advertised paths from the border routers through iBGP sessions. It selects the best path among the paths received and sends it to all routers in the AS via iBGP.

2) In the second stage eBGP sessions from BGP routers of other ASes are extended all the way to RCP rather than the border routers of that AS. If neighboring AS also has an RCP then the eBGP extends from one RCP to the other as shown below. This enables the RCP to learn multiple routes through eBGP, since the filter is not applied at the border router.

3) In the third stage RCP functionality is further enhanced by supporting new protocol between the RCPs via TCP connection (indicated by green line). At the same time the RCP continues with its BGP communication.

This new routing protocol provides a basis for end-to-end QoS support in interdomain routing. At the same time BGP based network operation remains uninterrupted.

Before considering a new approach to interdomain QoS routing, it is important to consider the tiered structure of the BGP based Internet. Studies from [6] show that business relationships between the ASes play important role in shaping the BGP import and export policies of the AS. For example, a provider AS advertises its customer routes to all the ASes
connected, but it does not advertise the routes learnt from its provider or peer. This is termed as policy opaqueness [14] which leads to the outcome of a path that is locally optimal but globally suboptimal for the end-to-end route. The business relations between ASes lead to a tiered structure of Internet as shown in the figure below.

Fig. 2 BGP-based tier structure

Here, each node is an AS and the arrow indicates a customer to provider link and the dashed line is a peer link. Any valid BGP path has multiple uphill segments, followed by up to one transverse (peer link) segment, and then down-hill segments. Thus, in figure 2 when AS-13 is learning path to reach AS-3 only two blue dashed paths are advertised through BGP. This makes sure that policies are satisfied over the BGP path. However, if the types of business relationships can be ignored then many more paths are possible through simple connectivity as shown by two red dashed paths. If these paths are to be enabled for end-to-end QoS routing along with resource reservation then a commercial model has to be evolved such that each transit AS in the path gets paid for the services offered by it. The alliance model mechanism proposed in the next section provides such a model.

Since the QoS routing requires path selection based on different QoS parameters specified by end-user, it is important to learn about multiple paths and have capability to establish reserved tunnels or connections along the best suited path. The opaque nature of BGP policy based routing prunes many such paths. It is better to have a tier-less mesh structure between the cooperating ASes to enable multiple path options as shown in the figure below. It is the same connection topology as the BGP-based structure. Under tier-less structure both blue and red paths are enabled and the best suited path is chosen.

Fig. 3 Tier-less Mesh Structure

A graphical representation of how multiple AS paths may be useful is shown in the figure below. It shows that there is a regular BGP-based best effort path to go from ISP-1 to ISP-6. However, there is another path with maximum bandwidth and one more path with low latency. Depending upon the QoS requirements a suitable path can be chosen.

Fig. 4 Multiple paths in interdomain routing

It is important to first define the type of business model for such a tier-less network. It has to make commercial sense to all the ASes involved in it. The next section addresses this question.

III. Alliance Network Model

We start with a definition of an Alliance Network. It consists of a set of interconnected ASes from multiple tiers:
1. To provide end-to-end QoS support for the end users
2. Enabling optimal multi-parameter based inter-domain path selection
3. By sharing QoS parametric information throughout the network
4. Along with resource reservation for MPLS tunnels
5. Based on revenue sharing mechanisms between the providers
6. That provides cost-performance based choices to the end user
7. Through a single partner provider for billing and support.

The implementation of Alliance Network is based on the extension of the RCP approach presented in section II. The RCP enables coexistence with BGP and support for new protocol. The new protocol is used to support overlay model for interdomain QoS routing. We further extend the functionality of the RCP server to include connection management services and call it RCMP. Since the model assumes prior agreement on the revenue sharing mechanism and exchange of QoS information, it is not scalable to global Internet. The business model is similar to alliance model in airline industry where players collaborate as well as compete on different segments and provide common ticketing and sharing of connections on certain routes. A pictorial representation of an Alliance Network model is shown in the figure 5.
In Alliance Network the task of interdomain routing is divided into two planes: Routing Control and Management Plane (RCMP) and Forwarding Plane (FP). The RCMP is responsible for following jobs:

1. Establishing and maintaining connections with other RCMPs in the network
2. Handling individual premium connection requests with specific QoS requirements
3. Computing multiple paths to a destination network
4. Signaling the control and management information for path exploration and set-up
5. Path selection based on multiple parameters
6. Setting up LSPs with specific QoS guarantees.
7. Using aggregation mechanisms in setting up the MPLS tunnels
8. Admission control and book keeping of existing tunnels
9. Service management and billing for the end user.

A. Routing Control and Management Plane: Information and Components

Each Autonomous System has a single, logically centralized RCMP entity, which could be a dedicated server with a fast computation and networking capability. It need not participate in the forwarding plane operations. These entities have following information and components that help them in doing their job:

1. Complete network topology of that AS.
2. Service Management component containing path characteristics (delay, jitter, reserved and available bandwidth, etc.) for paths within the AS, transit paths, and connections to neighboring AS
3. Multiple Path Computation Element (MPCE)
4. Network reachability information learnt from other RCMP entities.

It can be represented using the following block diagram:

B. Routing Control and Management Plane: Processing

The RCMP entities have a complete topographic view of an AS. It learns about the various networks that are reachable within an AS and through its customers. It broadcasts this information to all other RCMP entities in the alliance network. The RCMP entities exchange information with neighboring RCMPs over secure TCP connections and always remain in session through a keepalive/refresh message. The update message is sent in case of any change in status.

Each RCMP in the network learns a set of aggregated networks reachable from other RCMP and connectivity between them. Once the information about different networks reachable through the Alliance Network is available to all other RCMPs, they can use this information for calculating multiple paths to reach a given network destination. The Multi-path computation element (MPCE) in RCMP does the computation of different paths from the source to destination network. An efficient algorithm for the selection of multiple paths is presented in Ref. [17]. It enables optimization of multiple QoS parameters. MPCE path selection mechanism based on VS algorithm [17] is discussed in later part of this section.

The sequence of steps followed for setting up of an end-to-end interdomain QoS path is detailed here:

1. A path set-up request from end-user with specific QoS needs is received by an RCMP in the Alliance Network
2. The path-originating RCMP uses network reachability information and MPCE to calculate multiple paths.
3. Each path-vector is expressed in terms of constituent RCMP entities along the path.
4. A probe message is sent along each possible path. It carries parametric information for the desired QoS requirements and also has fields for updating QoS path parameters.
5. If an intermediate AS agrees to provide the resources, (typically by accommodating the request in the existing transit LSP tunnel or by setting up new LSP) it updates the path based QoS parameters in the probe message and also adds the costs involved. It temporarily reserves the requested resources.
6. In the first round, the probe messages are tabulated at the destination RCMP and sent back to the source RCMP.
7. The source RCMP makes the decision about which path to select using path based QoS parameters. If multiple paths match the QoS requirements, then the lowest cost factor would decide the best path.
8. Now specific resource reservations are done along the selected path. The reservations done in the first round make sure those resources are not lost during the probing
phase. Other RCMP paths time-out on the reservations and automatically release the resources.

9. The path set-up and resource reservation message is sent downstream along the selected path. The reservation confirmation and label information comes back in the upstream direction and full label-switched path is set-up.

C. QoS parameters update

The QoS path parameters can be additive such as number of hops, service charges or path delay. Where as, in case of maximum available bandwidth along the path, it is a minimum value of available bandwidths for all links along the path. In case of delay jitter we treat this quantity as an RMS value and use square of that as an additive parameter.

D. MPCE: Path Selection

The RCMP learns about different interdomain routing connection within the Alliance Network using the information exchange mechanism discussed earlier. The problem of finding an optimal path subject to multiple QoS parameters or constraints is NP-Complete problem and is not easily tractable. The constraints can be based on number of hops, service charges, delay, delay jitter, bandwidth, path protection etc. These constraints may be additive or non-additive in nature.

BGP suffers from lack of QoS support and currently researchers are looking to address this issue by proposing extensions to BGP. Its ubiquitous nature restricts a new protocol to replace it. Therefore our approach is concentrating on overlay model which co-exist with BGP. Different approaches to interdomain QoS routing under multiple constraints are presented in [18] [19] [20]. However they are heuristic algorithms based on source-oriented routing. This requires each AS (where the source is located) to have QoS parametric information of all interdomain paths for path computation. This approach is not scalable and suffers from long convergence under dynamic conditions. It also does not take into account the independence and privacy of individual ASes along the path. Therefore we are looking at an innovative routing scheme that has distributed path computation and the local decision-making for path-setup is effective even under limited information visibility.

We approach this problem using a heuristic algorithm, VS-routing scheme. The detail and the simulation results of this scheme is discussed in ref. [17]. VS routing scheme is based on a very concise representation of the network topology information using multi-dimensional Virtual Space configuration. The VS embedding involves evolution of equivalent multi-body representation of network topology using an iterative scheme. The logical development of the evolution process is consistent with many natural systems that evolve using multi-body interaction and energy minimization. The resultant VS configuration is highly directional to enable simple geometric routing. VS routing scheme has limited and fast information exchange, since the dynamic information is exchanged locally. Thus it adapts quickly to dynamic situations such as link failures or loading changes. It also has a very attractive scaling property and exhibits inherent stability and fast adaptability to dynamic conditions. The VS routing also enables multi-path routing without requiring additional resources or information exchange.

The VS configuration is constructed for each additive QoS path parameter and different path options are explored using the VS configuration. The criticality of one constraint over the other could determine which configuration is used for primary path selection. Path selection based on multiple QoS constraint in an interdomain scenario using VS routing with simulation results have been reported in detail in Ref. [21]. It shows that over 80% feasible path requests can be successfully satisfied under various dynamic scenarios.

A table below gives an example of the tabulated QoS path parameters learnt through probe messages along the multiple paths. The table corresponds to four separate AS-paths from the Alliance Network shown in Fig. 3.

<table>
<thead>
<tr>
<th>No.</th>
<th>AS-Path</th>
<th>BGP based</th>
<th>Available Bandwidth (MB)</th>
<th>Total Delay (ms)</th>
<th>Delay Jitter (rms)</th>
<th>Cumulative Service Charge units(MB-hr)</th>
<th>Reservation Possible</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>3-1-2-6-13</td>
<td>Yes</td>
<td>7.5</td>
<td>100</td>
<td>&lt;1.00</td>
<td>22</td>
<td>Yes</td>
</tr>
<tr>
<td>P2</td>
<td>3-1-2-7-13</td>
<td>Yes</td>
<td>4.5</td>
<td>145</td>
<td>&lt;1.00</td>
<td>25</td>
<td>No</td>
</tr>
<tr>
<td>P3</td>
<td>3-1-5-6-13</td>
<td>No</td>
<td>40</td>
<td>145</td>
<td>&lt;1.00</td>
<td>25</td>
<td>Yes</td>
</tr>
<tr>
<td>P4</td>
<td>3-10-4-7-13</td>
<td>No</td>
<td>12</td>
<td>25</td>
<td>&lt;0.50</td>
<td>32</td>
<td>Yes</td>
</tr>
</tbody>
</table>

It is seen from the table that paths P1 and P2 are the only BGP advertised paths and can continue to support best-effort traffic. Whereas, the Alliance Network can support path P3 with high available bandwidth for certain broadband tunnels or path P4 with low latency and jitter for certain real-time applications. Since P4 has highest cost / MB-hr of reservation it is useful for only premium traffic that can afford it. Since P1 and P4 or P3 and P4 do not have any path overlap, they are suitable for providing protection.

IV. Business Perspective and Advantages of Alliance Network

The main advantage of the Alliance Network model is that it provides mechanism for end-to-end QoS guarantees in interdomain routing while coexisting with BGP based Internet. It provides evolutionary path to the existing infrastructure to support this additional functionality. Certain new types of premium applications with tight QoS specifications can be supported through this mechanism.

For a service provider it enables enhanced geographical reach and market penetration through the alliance partners. When multinationals are expanding in new regions or aligning with new global partners the existing service provider may not have any market presence to serve the network requirements. Whereas, the Alliance Network may be able to set-up the desired connections quickly.

At the resource management level, a network operator of an individual AS in the network may set-up fixed set of transit tunnels between the border routers. Any new request for the transit facility is accommodated by adding the new flow over existing LSP tunnel using label-stacking in MPLS. Thus, each connection request need not be handled individually and the QoS performance parameters for the tunnel are already known and can be easily updated on the probe messages. These
tunnels could be set-up to use the underutilized links within the AS. Thus, unused network resources can be used to generate additional revenues through transit traffic.

One major concern of the service providers is the privacy of their internal network topology, implementation policies, and business relationships with others. In the present model none of this information is revealed to the other ASes.

While setting-up the path reservation the source RCMP agrees to the cumulative cost of the QoS path that is selected. This allows the settlement of financial transactions within the alliance partners. The customer deals with only the source RCMP for single billing and support. The LSP is set-up only for the duration for which the services are paid for, enabling bandwidth on demand.

In conclusion, Alliance Network model presents a very practical scheme for implementing end-to-end QoS routing in the interdomain, inter-provider environment. It will enable new services over the existing infrastructure.

REFERENCES


