A Modular RCP for Flexible Interdomain Route Control

Yi Wang
Princeton University
yiwang@cs.princeton.edu

Jennifer Rexford
Princeton University
jrex@cs.princeton.edu

ABSTRACT

This paper presents the MRCP (Modular Routing Control Platform), a routing control architecture that provides complete control and visibility of interdomain routing in a single AS. We propose a set of principles for making interdomain routing control more flexible and extensible, and show how the design of the MRCP adheres to the principles and enables new functionalities and services.

Categories and Subject Descriptors

C.2.3 [**Network Operations**]: Network Management

General Terms

Routing, Design, Control, Management, BGP

Keywords

Modularity, Flexibility, Extensibility

1. INTRODUCTION

Controlling interdomain routing is very complex for two reasons: first, operators often have many different goals in mind when they configure BGP; second, many of these goals are network-wide objectives (e.g., business policies, traffic engineering) that must be decomposed into router configurations for every router in the AS. Even worse, the decomposition of the operational goals into router configurations is still arguably a black art – largely a manual process with little systematic methodology and architectural support.

To address this issue, recent studies have shown that moving interdomain routing functionalities into a small collection of servers is a promising way to provide less convoluted and more flexible support for interdomain routing control [2, 1]. Particularly, the Routing Control Platform (RCP) has appealing properties such as having a network-wide view and making interdomain routing in an AS substantially easier to manage [2]. Given the fact that: a) different ISPs may have very different policies and management goals, and b) new extensions to BGP keep emerging [5, 7], it is also important for the control plane to *inherently* support for new BGP extensions and flexible expression of network-level control objectives. However, previous work on RCP discussed little about how the RCP can provide such flexibility and extensibility in an effctive way.

We address this issue in this paper. We argue that to best support control-plane flexibility and extensibility, the RCP should be built modularly. We first propose three principles for designing flexible interdomain routing control platforms, and then show how the modular design of the modular RCP (MRCP) adheres to the principles and provides architectural support for a wide range of new functionalities.

2. DESIGN PRINCIPLES

We propose the following three principles that we believe are essential to designing flexible interdomain routing control platforms.

- **Provide network-wide visibility**: Given network-level control and management objectives, the routing control platform must have a complete view of the network (an AS or institution) to convert these objectives into mechanisms and make decisions.
- **Decouple interdomain routing from intradomain routing**: The routing control platform should provide isolation between interdomain routing and intradomain routing, so that BGP routes chosen by the routing control platform are not affected by IGP routing changes.
- **Provide extensibility for new functionalities and services**: No matter how hard we try, we are always building tomorrow's legacy systems. Given the fact that new extensions to BGP keep emerging [5, 7], the routing control platform should provide flexible and extensible support to new functionalities and services, and make the transition as easy as possible.

3. MRCP ARCHITECTURE

In this section, we present the MRCP architecture. We make four major design decisions according to the above design principles.

3.1 Complete Control and Visibility

The MRCP controls the interdomain routing of an AS in

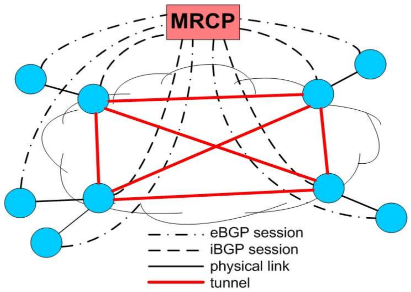

Figure 1: The modular routing control platform

a *logically* centralized manner, as shown in Figure 1[1]. The MRCP obtains a complete view of the network by learning all of the candidate routes from neighbor ASes through multi-hop external BGP (eBGP) sessions. It also directly controls the interdomain routing of the AS by assigning a BGP route per prefix to every router individually. Since MRCP talks with local routers through internal BGP (iBGP) sessions, this scheme also provides backwards compatibility.

3.2 Tunneling for a "BGP-free Core"

The coupling of interdomain and intradomain routing can cause a series of stability and performance problems [2, 6]. By using stateless tunnels between all the ingress-egress router pairs (encapsulating packets at ingress routers and decapsulating them at egress routers), the MRCP makes sure that its route assignments are both loop-free and isolated from IGP routing changes. This scheme also has great scalability benefits: all internal routers can be BGP ignorant (i.e., they don't need to store the huge BGP table in their memory or process BGP updates). Meanwhile, the MRCP only needs to talk to the edge routers of an AS. This tunneling can be supported by MPLS or IP-in-IP encapsulation. They are commonly used in ISPs and can be performed at line rate by high end routers [3].

3.3 Ranking of Egress Routers

Internally, the MRCP associates each ingress-egress tunnel with a configurable rank. The combination of the tunneling and ranking provides a new tie-breaking scheme in the BGP route selection process. For example, if multiple edge routers that can reach a certain prefix, for each traffic ingress point the MRCP may choose to pick a potentially different egress point such that the rank of the ingress-egress router pair is minimum. It solves the over-sensitivity problem of the IGP-based tie-breaking scheme [6], and offers a flexible interface for better traffic engineering.

3.4 Route Compression

MRCP provides great opportunities to reduce resource consumptions at the routers – not only internal routers don't need to speak BGP at all, edge routers also need to know nothing but "prefix - next-hop" pairs. Therefore, the MRCP strips all the extra attributes before sending the best routes to local edge routers. This route compression can significantly reduce memory footprint of BGP at the routers.

[1]For simplicity, in this paper we only describe the case in which each AS has only one MRCP. In real deployment, robustness can be achieved by replicating MRCP servers.

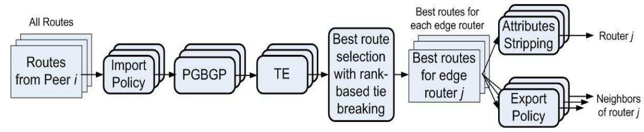

Figure 2: MRCP's modular pipeline architecture

4. MRCP IMPLEMENTATION

An MRCP is essentially an event (message) driven system – once a routing message arrives at the MRCP, it will trigger a sequence of processing, where each step is at a function module (policy enforcement, decision making, etc.). Therefore, we design MRCP as a pipeline of modules. Such modular architecture makes replacement of existing functions and introduction and sharing of new functions easier.

Figure 2 shows an example of MRCP's module-based pipeline architecture with two extension modules – a PGBGP module [5] and a traffic engineering module. By defining the right interfaces between modules, network operators can easily insert modules developed by third parties into their own MRCP.

We implement MRCP based on the XORP open routing platform [4]. Since XORP operates at single router level while MRCP operates at the AS level, we extend XORP in several major ways. Extentions include picking the best route per prefix for every edge router in a logically centralized manner, introducing the new rank-based tie-breaking scheme into the route selection process, making sure that the routes sent to a local edge router are consistent with the routes sent to the neighbor ASes of that edge router, route compression, etc. We choose to use IP-in-IP tunnels between ingress-egress routers since they are stateless and simple to use.

5. CONCLUSIONS

This paper presents the MRCP, a flexible interdomain routing control platform that featuring extensible support for new functionalities and services through it's modular architecture. We are testing its performance and also implementing new modules (such as PGBGP [5] and MIRO [7]) to evaluate its extensibility. We also plan to investigate how MRCP can support more expressive policy configuration.

6. REFERENCES

[1] O. Bonaventure, S. Uhlig, and B. Quoitin. The case for more versatile BGP route reflectors. Internet Draft draft-bonaventure-bgp-route-reflectors-00.txt, July 2004.

[2] N. Feamster, H. Balakrishnan, J. Rexford, A. Shaikh, and K. van der Merwe. The case for separating routing from routers. In *Proc. ACM SIGCOMM Workshop on Future Direction in Network Architecture*, August 2004.

[3] P. François and O. Bonaventure. An evaluation of IP fast reroute technique. In *Proc. CoNext*, October 2005.

[4] M. Handley, O. Hudson, and E. Kohler. XORP: An open platform for network research. In *Proc. ACM SIGCOMM Workshop on Hot Topics in Networking*, October 2002.

[5] J. Karlin, S. Forrest, and J. Rexford. Pretty good BGP: Improving BGP by cautiously adopting routes. In *Proc. International Conference on Network Protocols*, Nov 2006.

[6] R. Teixeira, A. Shaikh, T. Griffin, and J. Rexford. Dynamics of hot-potato routing in IP networks. In *Proc. ACM SIGMETRICS*, June 2004.

[7] W. Xu and J. Rexford. MIRO: Multi-path interdomain routing. In *Proc. ACM SIGCOMM*, September 2006.