

Network Working Group
Internet-Draft
Intended status: Informational
Expires: September 19, 2016

Y. Luo
L. Ou
China Telcom Co., Ltd.
S. Lu
Tencent
S. Zhuang
N. Wu
Huawei
March 18, 2016

Use-cases for Traffic Steering in Operator Networks
draft-luo-grow-ts-use-cases-00

Abstract

Transporting data to their users through operators' networks is a fundamental service that can benefit both providers and consumers. Due to the dramatically increased popularity and desire of differentiated services and their constraints, it is essential for operators to provide the traffic steering service under limited network resources and maximize their benefits at the same time.

This document lists some typical use cases for traffic steering services. This document does not attempt to enumerate all kinds of scenarios, but rather it describes several key features of these scenarios from which solutions may be constructed.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 19, 2016.

Copyright Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Terminology	3
3. Use-cases for ISP	3
3.1. EoS-oriented Steering	3
3.1.1. An Example for Prioritized Users	3
3.1.2. Derived Requirements	4
3.2. Load Balancing Oriented Steering	4
3.2.1. An Example of Load Balancing between Aggregation and Core	5
3.2.2. An Example of Load Balancing in Core	5
3.2.3. An Example of Load Balancing among ISPs	6
3.2.4. An Example for Transit Selection	6
3.2.5. Derived Requirements	7
4. Use-cases for OTT	8
4.1. QoS-oriented Steering	8
4.2. Business-oriented Steering	9
4.3. Inbound Traffic Steering	9
5. IANA Considerations	9
6. Security Considerations	9
7. Acknowledgements	9
8. References	9
8.1. Normative References	9
8.2. Informative References	10
Authors' Addresses	10

1. Introduction

Transporting data to their users through operators' networks is a fundamental service that can benefit both providers and consumers.

Since data/information transport is playing a key role nowadays, operators have to face this increasing challenge through satisfying services with differentiated criterias, such as latency, throughput, reliability and even user-defined constraints. Kinds of steering techniques, either manually or automatically, have been introduced to achieve that goal. Some situations such as fault, utilization and also making profit have to be taken care at the same time.

This document lists some typical use cases for traffic steering services. This document does not attempt to enumerate all kinds of scenarios, but rather it describes several key features of these scenarios from which solutions may be constructed.

2. Terminology

- o QoS: Quality of Service
- o EoS: Experience of Service
- o ISP: Internet Service Provider
- o OTTSP: Over the Top Service Provider

3. Use-cases for ISP

3.1. EoS-oriented Steering

It is a reasonable commercial way to provide multiple paths to the same destination with differentiated experiences to prioritized users/services. This is an efficient approach to maximize providers' network resources and their profit and give choices to users/services.

3.1.1. An Example for Prioritized Users

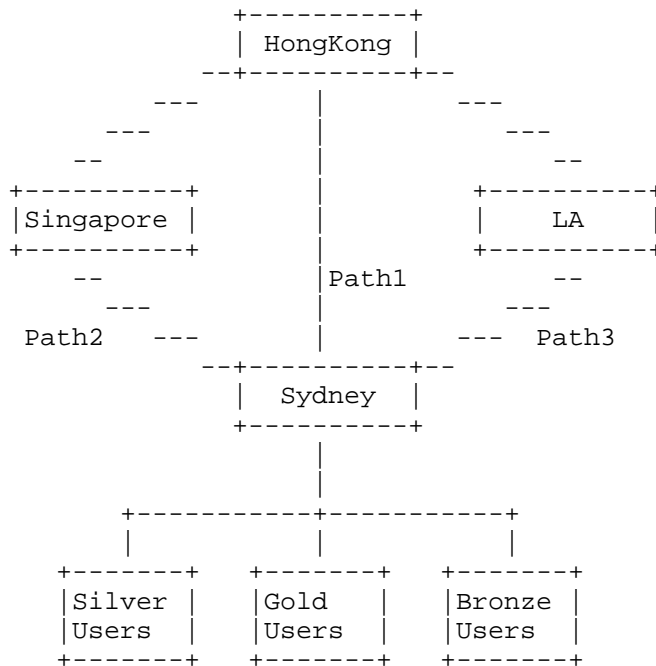


Figure 1 EoS-oriented path selection for ISP

As depicted above, there are three prioritized users in Sydney, saying Gold, Silver and Bronze, wish to visit website located in HongKong. An ISP provides three different paths with different experiences according to users' priority. The Gold Users may use Path1 with less latency and loss. The Silver Users may use the Path2 through Singapore with less latency but maybe some congestion there. The Bronze Users may use Path3 through LA with some latency and loss.

3.1.2. Derived Requirements

- o REQ01: A classification mechanism/system is REQUIRED to exist to identify users' traffic and the correspond priority respectively.
- o REQ02: A decision procedure/mechanism for path selection is REQUIRED to exist to decide traffic forwarding strategy based on the input from a classification mechanism.

3.2. Load Balancing Oriented Steering

It is a persistent goal for providers to increase the utilization ratio of their current network resources. Through transferring load to less preferred links, the throughput of the whole network may be increased.

3.2.1. An Example of Load Balancing between Aggregation and Core

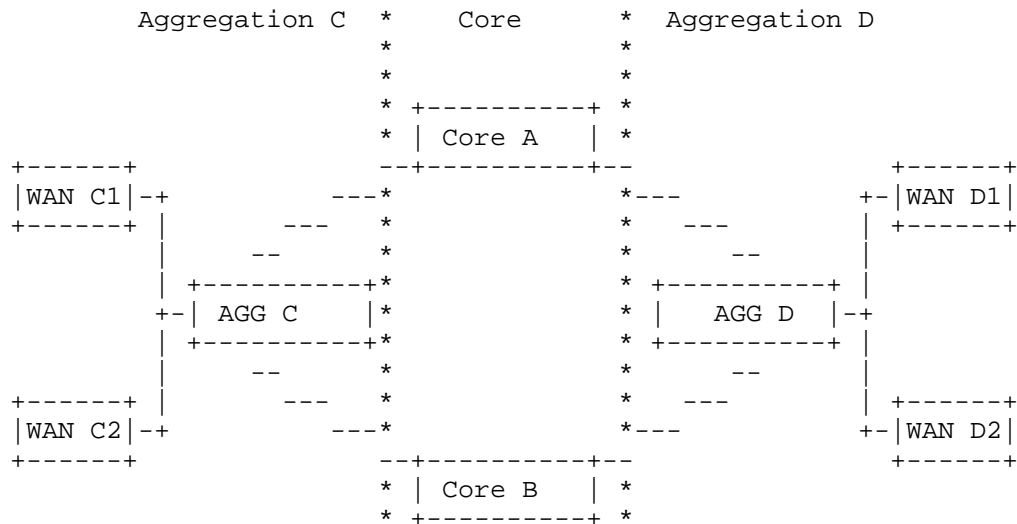


Figure 2 An Example of Load Balancing between Aggregation and Core

As depicted above, traffic from Aggregation C to Aggregation D follows the path AGG C->Core B->AGG D as the primary path. It is a reasonable practise to transfer some traffic load to less utilized path AGG C->Core A->AGG D when the primary path CBD has congestion.

3.2.2. An Example of Load Balancing in Core

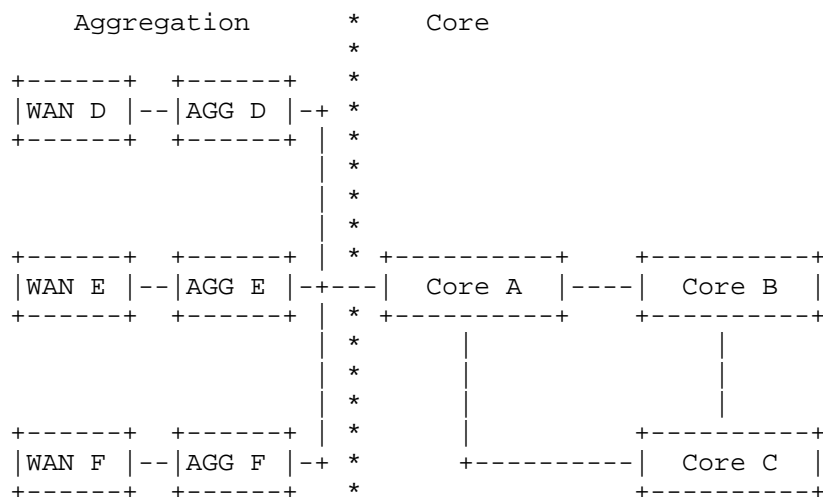


Figure 3 An Example of Load Balancing in Core

As depicted above, traffice from Core C to WAN area ususally passes through link CA in Core area. Part of traffice should be transferred to link CBA when link CA congested.

3.2.3. An Example of Load Balancing among ISPs

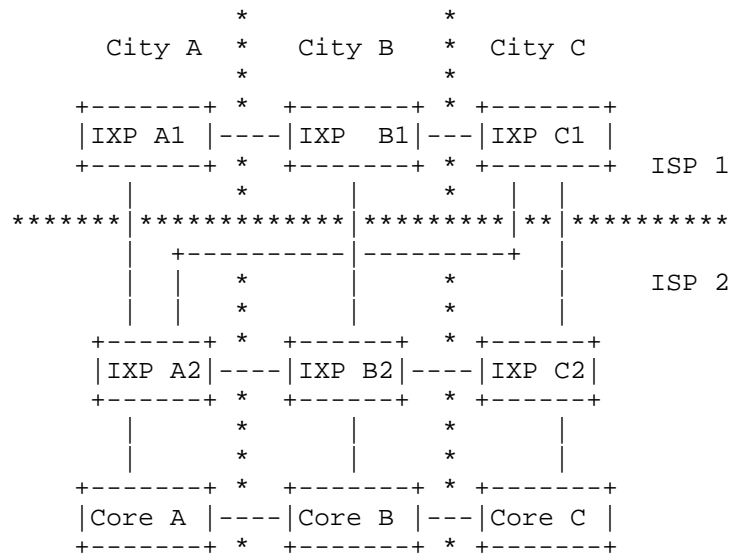


Figure 4 An Example of Load Balancing among ISPs

As depicted above, traffic from IXP C1 to A usually passes through link IXP C1->IXP A2->A. This is a long distant route, directly connect city C and city A. Part of traffic should be transferred to link IXP C1->IXP B1->IXP A1->IXP A2->A when primary link congested.

3.2.4. An Example for Transit Selection

It is quite common that multiple paths may exist for the same destination at the same time in an ISP network. Usually those paths with better QoS properties such as latency, loss, jitter and etc are often preferred. Since these properties keep changing from time to time, the decision of path selection has to be made dynamically.

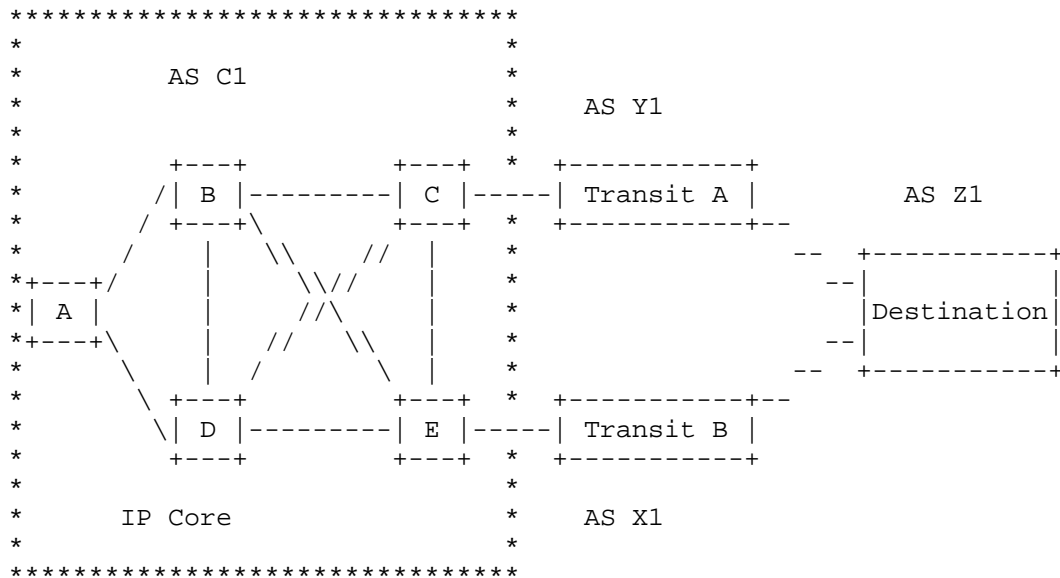


Figure 5 An Example for Transit Selection

As depicted above, the traffic to destination in AS Z1 from ISP IP core network (AS C1) has two choices on transit, saying Transit A and Transit B. Transit A will be preferred when the QoS of Transit B gets worse. As a result, the same traffic will go through Transit A instead.

3.2.5. Derived Requirements

- o REQ03: A resource monitoring mechanism/system is REQUIRED to exist for dynamically report the resource usage of target subnets.
- o REQ04: A decision procedure/mechanism for path selection is REQUIRED to exist to decide traffic forwarding strategy based on the input from a resource monitoring mechanism.
- o REQ05: A QoS monitoring mechanism/system is REQUIRED to exist for dynamically report the QoS conditions of those transits.
- o REQ06: A decision procedure/mechanism for path selection is REQUIRED to exist to decide traffic forwarding strategy based on the input from a QoS monitoring mechanism.
- o REQ07: A decision distribution mechanism/system is REQUIRED to exist to populate the adjustment behavior accordingly.

- o REQ08: The three mechanisms above are RECOMMENDED to be automatic ones.

4. Use-cases for OTT

4.1. QoS-oriented Steering

Actually similar situation exists in OTT service providers' networks too. An OTTSP may have multiple network exits separated in different sites. Depending on different conditions, it is optional for an OTTSP to choose one of those exits to connect with the same ISP.

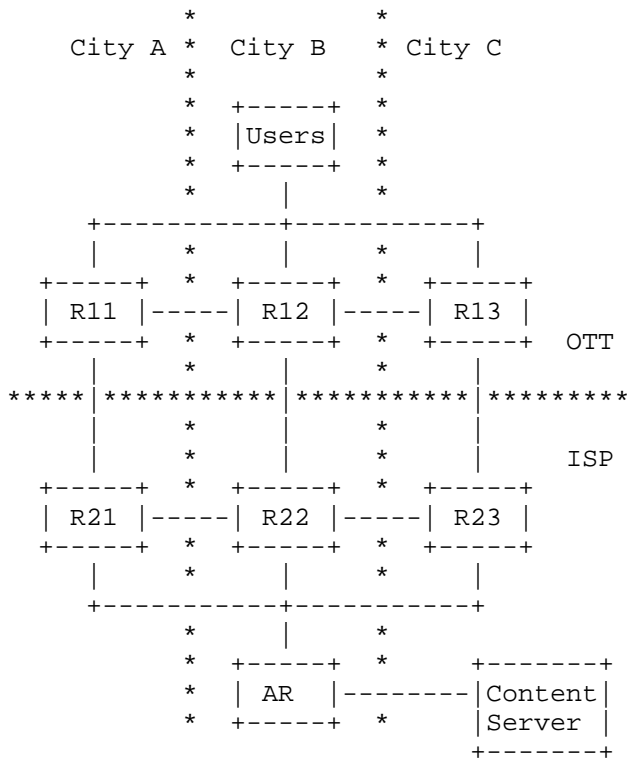


Figure 6 QoS-oriented Steering in OTT networks

As depicted above, the OTTSP has 3 exits with its ISP in City A, City B and City C. Based on network conditions, this OTTSP may choose different exits to steer its traffic into ISP's networks.

4.2. Business-oriented Steering

Besides network conditions, an OTTSP may make its steering strategy based on different business. For example, the OTTSP in the graph above may choose different exits on per-business base, which REQUIRES a mechanism/system exists to identify different businesses from traffic flow.

REQ09: A mechanism/system exists to identify different businesses from traffic flow.

4.3. Inbound Traffic Steering

Besides exits, an OTTSP may wish to have choices on entrances for inbound traffic. Because of internal policies, an ISP may choose to ignore or even prohibit an OTTSP's attempt to affect traffic paths. It will be beneficial for both providers if a perfect solution with detailed consideration may help to solve this problem, which is absolutely out of the scope of this document.

REQ10: An interactive mechanism/system is REQUIRED to exist for negotiation between OTT and ISP to solve the scenario of inbound traffic steering.

5. IANA Considerations

This document has no request to IANA.

6. Security Considerations

This document has no security issue introduced.

7. Acknowledgements

TBD.

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.

8.2. Informative References

[I-D.chen-i2rs-ts-use-case]

Chen, M. and S. Hares, "I2RS Traffic Steering Use Case", draft-chen-i2rs-ts-use-case-01 (work in progress), July 2014.

[I-D.chen-idr-rr-based-traffic-steering-usecase]

Chen, M., Zhuang, S., Zhu, Y., and S. Wang, "Use Cases of Route Reflection based Traffic Steering", draft-chen-idr-rr-based-traffic-steering-usecase-00 (work in progress), July 2013.

[I-D.li-idr-flowspec-rpd]

Li, Z., Ou, L., Luo, Y., Lu, S., Zhuang, S., and N. Wu, "BGP FlowSpec Extensions for Routing Policy Distribution (RPD)", draft-li-idr-flowspec-rpd-01 (work in progress), October 2015.

Authors' Addresses

Yujia Luo
China Telcom Co., Ltd.
109 West Zhongshan Ave, Tianhe District
Guangzhou 510630
China

Email: luoyuj@gsta.com

Liang Ou
China Telcom Co., Ltd.
109 West Zhongshan Ave, Tianhe District
Guangzhou 510630
China

Email: oul@gsta.com

Sujian Lu
Tencent
Tengyun Building, Tower A, No. 397 Tianlin Road
Shanghai, Xuhui District 200233
China

Email: jasonlu@tencent.com

Shunwan Zhuang
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: zhuangshunwan@huawei.com

Nan Wu
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: eric.wu@huawei.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: December 31, 2016

Y. Luo
L. Ou
China Telcom Co., Ltd.
S. Lu
Tencent
S. Zhuang
N. Wu
Huawei
June 29, 2016

Use-cases for Traffic Steering in Operator Networks
draft-luo-grow-ts-use-cases-01

Abstract

Due to the dramatically increased network traffic and the desire of differentiated services, it is essential for operators to provide the traffic steering service under limited network resources and maximize their benefits at the same time.

This document lists some typical use cases for traffic steering services. This document does not attempt to enumerate all kinds of scenarios, but rather it describes several key features of these scenarios from which solutions may be constructed.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on December 31, 2016.

Copyright Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Terminology	3
3. Use cases and Requirements	3
3.1. Business-oriented Steering	3
3.1.1. An Example of Preferential Users	3
3.1.2. An Example of Preferential Services	4
3.1.3. Derived Requirements	5
3.2. Traffic Congestion Mitigation	6
3.2.1. An Example of Congestion Mitigation in Core	6
3.2.2. An Example of Congestion Mitigation among ISPs	7
3.2.3. An Example of Congestion Mitigation at International Edge	7
3.2.4. Derived Requirements	8
4. IANA Considerations	9
5. Security Considerations	9
6. Acknowledgements	9
7. References	9
7.1. Normative References	9
7.2. Informative References	9
Authors' Addresses	10

1. Introduction

Transporting data to their users through operators' networks is a fundamental service that can benefit both providers and consumers.

Since data/information transport is playing a key role nowadays, operators have to face this increasing challenge through satisfying services with differentiated criterias, such as latency, throughput, reliability and even user-defined constraints. Moreover, the internet traffic changes rapidly and is hard to be predicted, so

there is chance that operators' networks will be congested. However, the network capacity expansion takes time and could not meet the differentiated service requirement or solve the congestion problem in time. As a result, it's necessary to introduce traffic steering techniques into operators' networks.

This document lists some typical use cases for traffic steering in ISP networks and OTT networks. It does not attempt to enumerate all kinds of scenarios, but rather it describes several key features of these scenarios from which solutions may be constructed.

2. Terminology

- o QoS: Quality of Service
- o ISP: Internet Service Provider
- o MAN: Metropolitan Area Network
- o OTTSP: Over the Top Service Provider, or Content Operator
- o AR: Access Router

3. Use cases and Requirements

3.1. Business-oriented Steering

It is a reasonable commercial way to provide multiple paths to the same destination with differentiated experiences to preferential users/services. This is an efficient approach to maximize providers' network resources as well as their profit and offer more choices to network users.

3.1.1. An Example of Preferential Users

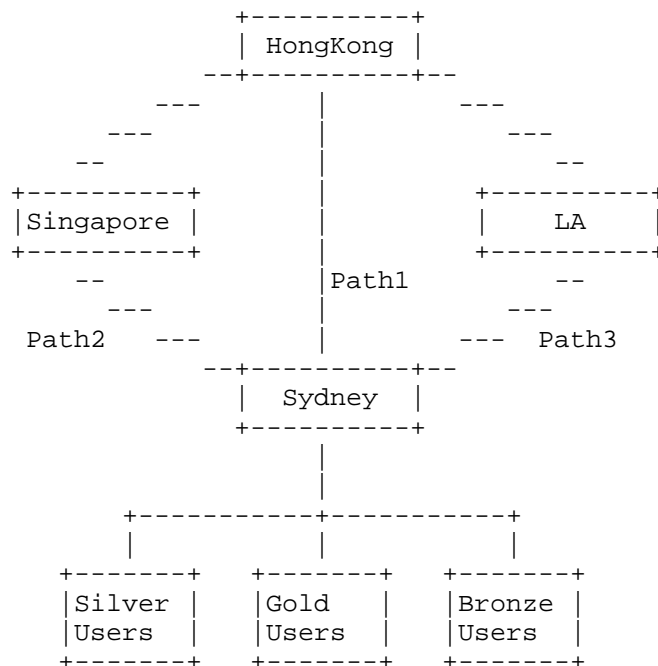
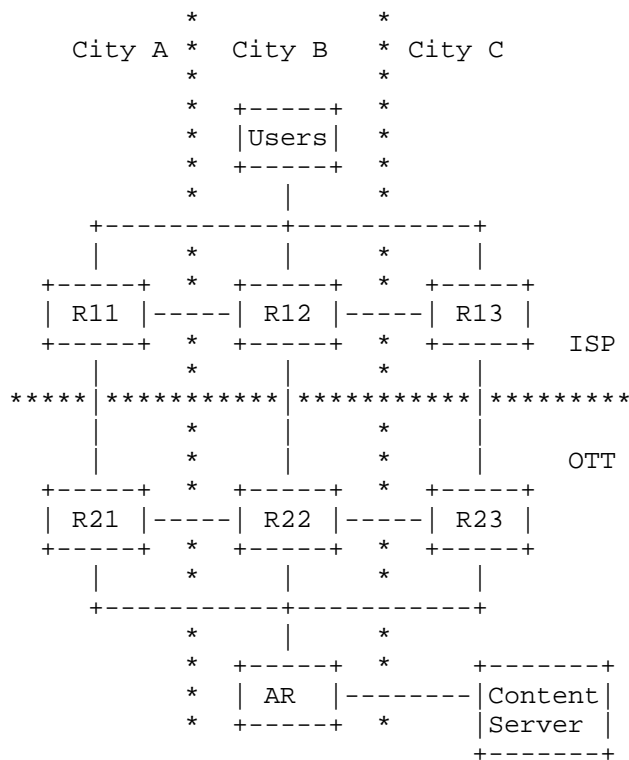


Figure 1 Differentiated Path Selection for Different User

In the above ISP network, there are three kinds of users in Sydney, saying Gold, Silver and Bronze, and they wish to visit website located in HongKong. The ISP provides three different paths with different experiences according to users' priority. The Gold Users may use Path1 with less latency and loss. The Silver Users may use the Path2 through Singapore with less latency but maybe some congestion there. The Bronze Users may use Path3 through LA with some latency and loss.

3.1.2. An Example of Preferential Services



As depicted above, the OTTSP has 3 exits with one ISP, which are located in City A, City B and City C. The content is obtained from Content Server and send to the exits through AR. an OTTSP may make its steering strategy based on different services. For example, the OTTSP in the graph above may choose exit R21 for video service and exit R22 for web service, which REQUIRES a mechanism/system exists to identify different services from traffic flow.

- o REQ01: A classification mechanism/system is REQUIRED to identify users' priority from user traffic or to identify different services from traffic flow.
- o REQ02: A decision procedure/mechanism for path selection is REQUIRED to exist to decide traffic forwarding strategy based on the input from a classification mechanism.

3.2. Traffic Congestion Mitigation

It is a persistent goal for providers to increase the utilization ratio of their current network resources, and to mitigate the traffic congestion. Traffic congestion is possible to happen anywhere in the ISP network (MAN, IDC, core and the links between them), because internet traffic is hard to predict. For example, there might be some local online events that the network operators didn't know beforehand, or some sudden attack just happened. Even for the big events that can be predicted, such as annual online discount of e-commerce company, or IOS update of Apple Inc, we could not guarantee there is no congestion. What is more, the network capacity expansion is usually an annual operation, which could be delayed by any links of the engineering. As a result, the temporary traffic steering is always needed. The same thing happens to the OTT networks as well.

It should be noted that, the traffic steering is absolutely not a global behavior. It just acts on part of the network, and it's temporary.

3.2.1. An Example of Congestion Mitigation in Core

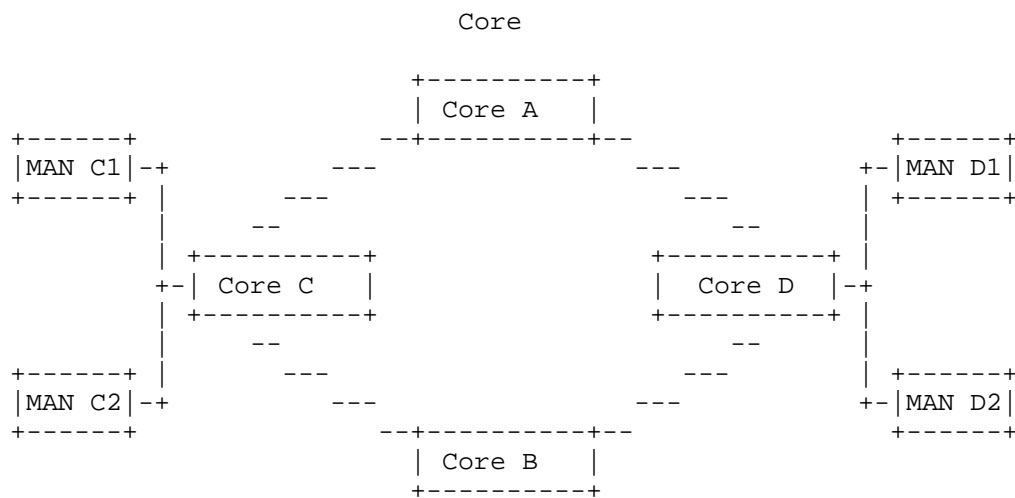


Figure 3 An Example of Congestion Mitigation in Core

As depicted above, traffic from MAN C1 to MAN D2 follows the path Core C->Core B->Core D as the primary path, but somehow the load ratio becomes too much. It is reasonable to transfer some traffic load to less utilized path Core C->Core A->Core D when the primary path has congestion.

3.2.2. An Example of Congestion Mitigation among ISPs

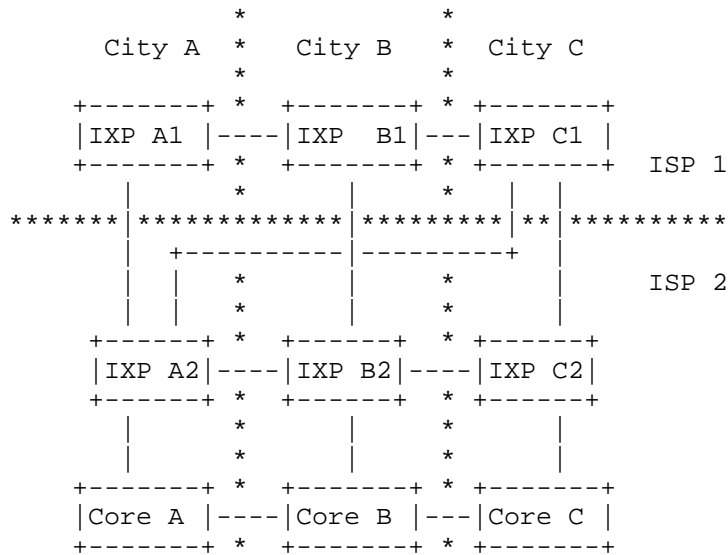
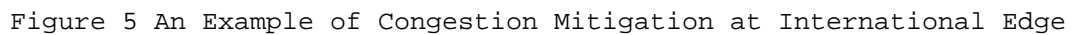


Figure 4 An Example of Congestion Mitigation among ISPs

As depicted above, ISP1 and ISP2 are interconnect by 3 exits which are located in 3 cities respectively. The links between ISP1 and ISP2 in the same city are called local links, and the rest are long distance links. Traffic from IXP C1 to Core A in ISP 2 usually passes through link IXP C1->IXP A2->Core A. This is a long distant route, directly connecting city C and city A. Part of traffic could be transferred to link IXP C1->IXP B1->IXP A1->IXP A2->Core A when the primary route congest.

3.2.3. An Example of Congestion Mitigation at International Edge

An ISP usually interconnects with more than 2 transit networks at the international edge, so it is quite common that multiple paths may exist for the same foreign destination. Usually those paths with better QoS properties such as latency, loss, jitter and etc are often preferred. Since these properties keep changing from time to time, the decision of path selection has to be made dynamically.



3.2.4. Derived Requirements

- [Page 8]

- o REQ07: A decision distribution mechanism/system is REQUIRED to exist to populate the adjustment behavior accordingly.
- o REQ08: The seven mechanisms above are RECOMMENDED to be automatic ones.

4. IANA Considerations

This document has no request to IANA.

5. Security Considerations

This document has no security issue introduced.

6. Acknowledgements

TBD.

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.

7.2. Informative References

[I-D.chen-i2rs-ts-use-case]
Chen, M. and S. Hares, "I2RS Traffic Steering Use Case", draft-chen-i2rs-ts-use-case-01 (work in progress), July 2014.

[I-D.chen-idr-rr-based-traffic-steering-usecase]
Chen, M., Zhuang, S., Zhu, Y., and S. Wang, "Use Cases of Route Reflection based Traffic Steering", draft-chen-idr-rr-based-traffic-steering-usecase-00 (work in progress), July 2013.

[I-D.li-idr-flowspec-rpd]
Li, Z., Ou, L., Luo, Y., Lu, S., Zhuang, S., and N. Wu, "BGP FlowSpec Extensions for Routing Policy Distribution (RPD)", draft-li-idr-flowspec-rpd-02 (work in progress), June 2016.

Authors' Addresses

Yujia Luo
China Telcom Co., Ltd.
109 West Zhongshan Ave,Tianhe District
Guangzhou 510630
China

Email: luoyuj@gsta.com

Liang Ou
China Telcom Co., Ltd.
109 West Zhongshan Ave,Tianhe District
Guangzhou 510630
China

Email: oul@gsta.com

Sujian Lu
Tencent
Tengyun Building,Tower A ,No. 397 Tianlin Road
Shanghai, Xuhui District 200233
China

Email: jasonlu@tencent.com

Shunwan Zhuang
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: zhuangshunwan@huawei.com

Nan Wu
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

Email: eric.wu@huawei.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: October 5, 2016

R. White
Linkedin
A. Retana
Cisco Systems, Inc.
S. Hares
Huawei
April 4, 2016

Filtering of Overlapping Routes
draft-white-grow-overlapping-routes-04

Abstract

This document proposes an optional mechanism to remove a prefix when it overlaps with a functionally equivalent shorter prefix. The proposed mechanism does not require any changes to the BGP protocol.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on October 5, 2016.

Copyright Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of

the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Requirements Language	3
3. Overlapping Route Filtering Mechanism	3
3.1. Marking Overlapping Routes	4
3.2. Preferring Marked Routes	4
3.2.1. Using a Cost Community	4
3.2.2. Using the Local Preference	4
3.3. Handling Marked Routes Within the AS	5
3.4. Handling Marked Routes at the Outbound Edge	5
4. Examples of Filtering Overlapping Routes	5
4.1. IPv4 Example	5
4.2. IPv6 Example	6
5. Operational Considerations	6
5.1. Advantages to the Service Provider	7
5.2. Implications for Router processing	7
5.3. Implications for Convergence Time	7
6. Security Considerations	7
7. IANA Considerations	8
8. Acknowledgements	8
9. References	8
9.1. Normative References	8
9.2. Informative References	8
Appendix A. Change Log	8
A.1. Changes between the -00 and -01 versions.	8
A.2. Changes between the -01 and -02 versions	9
A.3. Changes between the -02 and -03 versions	9
Authors' Addresses	9

1. Introduction

One cause of the growth of the global Internet's default free zone table size is overlapping routes injected into the routing system to steer traffic among various entry points into a network. Because padding AS Path lengths can only steer inbound traffic in a very small set of cases, and other mechanisms used to steer traffic to a particular inbound point are ineffective when multiple upstream providers are in use, advertising longer prefixes is often the only possible way for an AS to steer traffic into specific entry points along its edge.

These longer prefix routes, called overlapping routes in this document, are often advertised along with a shorter prefix route, called a covering route, in order to ensure connectivity in the case

of link or device failures. Overlapping routes not only add to the load on routers in the Internet core by simply expanding the table size; these routes may be less stable than the covering routes they are paired with.

Given the importance of an autonomous system's ability to steer traffic into specific entry points, simply removing the longer prefixes in a longer prefix (overlapping)/shorter prefix (covering) pair of routes isn't a viable solution.

This document proposes an optional mechanism to remove overlapping routes that are no longer useful for steering traffic towards a specific entry point in a particular AS. Removing these routes would reduce the global table in size, and reduce its instability, while removing no capabilities, nor increasing the average path length.

The mechanism proposed is simple to implement, requiring no changes to BGP [RFC4271] either in packet format or in the decision process. The removal described in this document is akin to filtering, not to route aggregation.

The intent of the mechanism is for it to be used based on local decisions and policies, not on an Internet-wide fashion. It is assumed that network operators using this mechanism have an incentive to do so.

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. Overlapping Route Filtering Mechanism

The handling of overlapping prefixes received from an external peer can be broken down into four parts: marking overlapping routes, preferring marked routes, handling marked routes within the AS, and handling marked routes at the AS exit point.

The initial step in successfully filtering overlapping routes is to identify and mark them. This document proposes the use of a BGP community called BOUNDED for that purpose. Because the operation suggested takes place inside an Autonomous System (AS), then any locally assigned community can be used.

The term BOUNDED is used to refer to a locally assigned community used to mark overlapping routes, and to these marked routes as well.

3.1. Marking Overlapping Routes

As each prefix is received by a BGP speaker from an external peer, it is evaluated in the light of other prefixes already received. If two prefixes overlap in space (such as 192.0.2.0/24 and 192.0.2.128/25, or 2001:DB8::/32 and 2001:DB8:1:/48), the longer prefix SHOULD be BOUNDED if it fully overlaps the covering prefix and it is the best path to the destination.

An overlapping prefix is said to fully overlap the corresponding covering prefix if both have identical AS_PATH attributes (both in length and contents) and the same NEXT_HOP.

3.2. Preferring Marked Routes

Since the same overlapping route may be received at several peering points along the edge of the AS, and the covering route may not be present at each of these points, BOUNDED routes SHOULD be preferred over unmarked routes for overlapping routes to be properly handled. A router which marks an overlapping route should also use one of the two mechanisms described here to insure the marked route is preferred throughout the AS.

Only one method described in this section SHOULD be deployed in any given AS.

3.2.1. Using a Cost Community

The recommended method for preferring BOUNDED routes is to use a Cost Community [I-D.ietf-idr-custom-decision] with the Point of Insertion set to ABSOLUTE_VALUE. This mechanism leaves all existing local policy controls in place within the AS.

If this method is used, only the BOUNDED routes need to be tagged using a lower than default Cost, as routes without a Cost Community are considered to have the default value.

3.2.2. Using the Local Preference

An alternate mechanism which may be used to prefer BOUNDED routes is to set their Local Preference to some number higher than the normal standard policy settings for a particular prefix. It's not important that any particular BOUNDED route win over any other one; so simply adding a small amount to the normal Local Preference, as dictated by local policy, will ensure a BOUNDED route will always win over an unmarked route, so only these routes reach the outbound edge of the AS.

3.3. Handling Marked Routes Within the AS

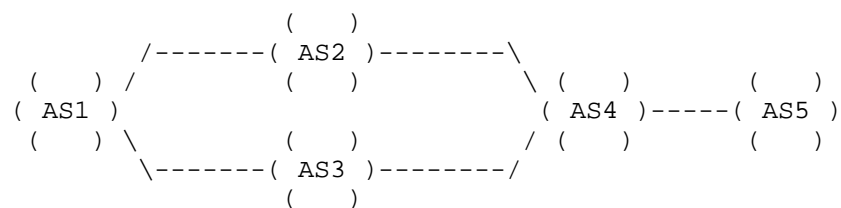
Routes marked with the BOUNDED community MAY not be installed in the local RIB of routers within the AS. This optional step will reduce local RIB and forwarding table usage and volatility within the AS.

3.4. Handling Marked Routes at the Outbound Edge

If local policy dictates, routes marked with the BOUNDED community SHOULD NOT be advertised to external peers. If they are advertised, they MAY then be marked with the NO_EXPORT community.

4. Examples of Filtering Overlapping Routes

Assume the following configuration of autonomous systems:



This network is used in both of the following examples.

4.1. IPv4 Example

- o AS1 is advertising 192.0.2.128/25 to both AS2 and AS3.
- o AS2 is advertising both 192.0.2.128/25 and 192.0.2.0/24 into AS4.
- o AS3 is advertising 192.0.2.128/25 into AS4
- o Each BGP connection (session) is handled by a separate router within each AS (for instance, AS4 peers with AS2 and AS3 on separate routers).

When the router in AS4 peering with AS2 receives both the 192.0.2.128/25 and the 192.0.2.0/24 prefixes, it will mark 192.0.2.128/25 as BOUNDED, and set a Cost Community (as described in Section 3.2.1) so the marked overlapping route is preferred over unmarked routes within AS4.

The border router between AS4 and AS3 will receive the longer prefix from AS3, and the preferred BOUNDED overlapping route through iBGP. It will prefer the marked route, so the unmarked route towards 192.0.2.128/25 will not be advertised throughout AS4.

If the link between AS1 and AS2 fails, the longer length prefix will be withdrawn from AS2, and thus the peering point between AS2 and AS4 will no longer have an overlapping set of prefixes. Within AS4, the border router which peers with AS2 will cease advertising the 192.0.2.128/25 prefix, which allows the AS3/AS4 border router to begin advertising it into AS4, and through AS4 into AS5, restoring connectivity to AS1.

4.2. IPv6 Example

- o AS1 is advertising 2001:DB8:1:/48 to both AS2 and AS3.
- o AS2 is advertising both 2001:DB8:1:/48 and 2001:DB8::/32 into AS4.
- o AS3 is advertising 2001:DB8:1:/48 into AS4
- o Each BGP connection (session) is handled by a separate router within each AS (for instance, AS4 peers with AS2 and AS3 on separate routers).

When the router in AS4 peering with AS2 receives both the 2001:DB8:1:/48 and 2001:DB8::/32 prefixes, it will mark 2001:DB8:1:/48 as BOUNDED, and set a Cost Community (as described in Section 3.2.1) so the marked overlapping route is preferred over unmarked routes within AS4.

The border router between AS4 and AS3 will receive the longer prefix from AS3, and the preferred BOUNDED overlapping route through iBGP. It will prefer the marked route, so the unmarked route towards 2001:DB8:1:/48 will not be advertised throughout AS4.

If the link between AS1 and AS2 fails, the longer length prefix will be withdrawn from AS2, and thus the peering point between AS2 and AS4 will no longer have an overlapping set of prefixes. Within AS4, the border router which peers with AS2 will cease advertising the 2001:DB8:1:/48 prefix, which allows the AS3/AS4 border router to begin advertising it into AS4, and through AS4 into AS5, restoring connectivity to AS1.

5. Operational Considerations

The intent of the mechanism described in this document is for it to be used based on local policies, not on an Internet-wide fashion. It is assumed that network operators using this mechanism have an incentive to do so.

The practice of filtering exists today on the Internet. While there may be local benefits to applying manual filters and/or the mechanism

specified in this document, the operator should be aware of the impact it may have on neighboring autonomous systems' policies [I-D.cardona-filtering-threats].

The benefits and implications associated with this proposal are discussed in the sections below. The text references the sample network in Section 4.

5.1. Advantages to the Service Provider

AS4, in each of the situations, reduces the number of prefixes advertised to transit peering autonomous systems by the number of longer prefixes that overlap with aggregates of those prefixes, so that AS5 receives fewer total routes, and a more stable routing table. While one copy of the prefix continues to be carried through the autonomous system, this entry can be removed from the local forwarding table.

5.2. Implications for Router processing

This proposal requires a BGP speaker to perform an additional check on receiving a route, checking the route against existing routes for overlapping coverage of a set of reachable destinations. This additional work, in terms of processing requirements, should be easily offset by the overall savings in processing through the reduction of the forwarding table size, and the additional stability in the routing table due to the removal of longer length prefixes.

5.3. Implications for Convergence Time

If the route to the AS providing the route to the covering route should be lost, the overlapping route must now propagate into the autonomous systems which had formerly received only the covering route. This behavior increases convergence time and may create situations in which reachability is temporarily compromised. Unlike the case where manual filters are used, normal BGP behavior should restore reachability without changes to the router configuration.

6. Security Considerations

This document presents a mechanism for an autonomous system to mark and filter overlapping prefixes. Note that the result of this operation is akin to the implementation of local route filtering at an AS boundary. As such, this document doesn't introduce any new security risks.

7. IANA Considerations

This document has no IANA actions.

8. Acknowledgements

Cengiz Alaentinoğlu, Daniel Walton, David Ball, Ted Hardie, Jeff Hass, Barry Greene, Bill Herrin and Robert Raszuk gave valuable comments on this document.

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

9.2. Informative References

- [I-D.cardona-filtering-threats]
Cardona, C. and P. Francois, "Making BGP filtering a habit: Impact on policies", draft-cardona-filtering-threats-02 (work in progress), July 2013.
- [I-D.ietf-idr-custom-decision]
Retana, A. and R. White, "BGP Custom Decision Process", draft-ietf-idr-custom-decision-04 (work in progress), November 2013.
- [RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

Appendix A. Change Log

A.1. Changes between the -00 and -01 versions.

- o Updated authors' contact information.
- o Changed intended status to Informational.
- o General editorial changes.
- o Clarified the intent of the draft in several places.
- o Clarified when a route should be marked (3.1).
- o Edited the operational considerations section.

- o Updated ACKs.
- A.2. Changes between the -01 and -02 versions
- o Updated authors' contact information.
 - o General editorial changes.
 - o Refined the text about marking routes.
- A.3. Changes between the -02 and -03 versions
- o Updated authors' contact information.
 - o Added IPv6 examples.
 - o Minor editorial changes.
- A.4. Changes between the -03 and -04 versions
- o Updated authors' contact information.

Authors' Addresses

Russ White
Linkedin

Email: russ@riw.us

Alvaro Retana
Cisco Systems, Inc.
7025 Kit Creek Rd.
Research Triangle Park, NC 27709
USA

Email: aretana@cisco.com

Susan Hares
Huawei

Email: shares@ndzh.com