IDR
Internet-Draft
Intended status: Informational
Expires: February 28, 2016

K. Patel
Cisco Systems
R. White
Ericsson
S. Hares
Huawei

August 27, 2015

Use Cases for an Interface to BGP Protocol
draft-keyupate-idr-i2rs-bgp-usecases-01.txt

Abstract

A network routing protocol like BGP is typically configured and analyzed through some form of Command Line Interface (CLI) or NETCONF. These interactions to control BGP and diagnose its operation encompass: configuration of protocol parameters, display of protocol data, setting of certain protocol state and debugging of the protocol.

Interface to the Routing System's (I2RS) Programmatic interfaces provides an alternate way to control and diagnose the operation of the BGP protocol. I2RS may be used for the configuration, manipulation, analyzing or collecting the protocol data. This document describes set of use cases for which I2RS can be used for BGP protocol. It is intended to provide a base for the solution draft describing a set of interfaces 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 February 28, 2016.

Table of Contents

1. Introduction

Typically, a network routing protocol like BGP is configured and results of its operation are analyzed through some form of Command Line Interface (CLI) or NETCONF. These interactions to control BGP and diagnose its operation encompass: configuration of protocol parameters, display of protocol data, setting of certain protocol state and debugging of the protocol.

The I2RS architecture document [I-D.ietf-i2rs-architecture] describes a mechanism to control network protocols like BGP using a set of programmatic interfaces. These programmatic interfaces allow one to control the BGP protocol by analyzing its operational state and routing protocol data, plus manipulating BGP's configuration to achieve various goals. The I2RS is not intended to replace any existing configuration mechanisms, (i.e.: Command Line Interface or NETCONF). Instead, I2RS is intended to augment those existing mechanisms by defining a standardized set of programmatic interfaces to enable easier configuration, interrogation and analysis of the BGP protocol.

This document describes set of use cases for which I2RS's programmatic interfaces can be used to control and analyze the operation of BGP. The use cases described in this document cover the following aspects of BGP: protocol parameter configuration, protocol route manipulation and tracking of protocol events. The goal is to inform the community's understanding of where the I2RS BGP extensions fit within the overall I2RS architecture. It is intended to provide a basis for the solutions draft describing the set of Interfaces to the BGP protocol.

1.1. 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].

1.2. Requirements for I2S

Each of the sections below (BGP protocol operation, BGP Route Manipulation, BGP Events, Central Membership for MPLS based VPNs, and Marking Overlapping BGP Routes) have specified a use case descriptions followed by a summary of I2RS requirements. Each requirement listed in these sections is given an number [REQnn] where nn is the unique BGP Requirement. Requirements duplicated from previous sections are repeated with the original requirements number.

2. Summary of Requirements for I2RS Module for BGP

This is a summary of the requirements for an IDR I2RS Yang Module listed in this document.

- BGP-REQ01: I2RS client/agent exchange SHOULD support the read, write and quick notification of status of the BGP peer operational state on each router within a given Autonomous System (AS). This operational status includes the quick notification of protocol events that proceed a destructive tear-down of BGP session

- BGP-REQ02: I2RS client SHOULD be able to push BGP routes with custom cost communities to specific I2RS agents on BGP routers for insertion in specific BGP Peer(s) to aid Traffic engineering of data paths. These routes SHOULD be tracked by the I2RS Agent as specific BGP routes with customer cost communities. These routes (will/will not) installed via the RIB-Info.

- BGP-REQ03: I2RS client SHOULD be able to track via read/ notifications all Traffic engineering changes applied via I2RS agents to BGP route processes in all routers in a network.

- BGP-REQ04: I2RS Agents SHOULD support identification of routers as BGP ASBRs, PE routers, and IBGP routers.

- BGP-REQ05: I2RS client-agent SHOULD support writing traffic flow specifications to I2RS Agents that will install them in associated BGP ASBRs and the PE routers.

- BGP-REQ06: I2RS Client SHOULD be able to track flow specifications installed within a IBGP Cloud within an AS via reads of BGP Flow

Specification information in I2RS Agent, or via notifications from I2RS agent

- BGP-REQ07: I2RS client-agent exchange SHOULD support the I2RS client being able to prioritize and control BGP's announcement of flow specifications after status information reading BGP ASBR and PE router's capacity. BGP ASBRs and PE routers functions within a router MAY forward traffic flow specifications received from EBGP speakers to I2RS agents, so the I2RS Agent SHOULD be able to send these flow specifications from EBGP sources to a client in response to a read or notification.

- BGP-REQ08: I2RS Client SHOULD be able to read BGP route filter information from I2RS Agents associated with legacy BGP routers, and write filter information via the I2RS agent to be installed in BGP RR. The I2RS Agent SHOULD be able to install these routes in the BGP RR, and engage a BGP protocol action to push these routers to ASBR and PE routers.

- BGP-REQ09: I2RS client(s) SHOULD be able to request the I2RS agent to read BGP routes with all BGP parameters that influence BGP best path decision, and write appropriate changes to the BGP Routes to BGP and to the RIB-Info in order to manipulate BGP routes

- BGP-REQ10: I2RS client SHOULD be able instruct the I2RS agent(s) to notify the I2RS client when the BGP processes on an associated routing system observe a route change to a specific set of IP Prefixes and associated prefixes. Route changes include: 1) prefixes being announced or withdrawn, 2) prefixes being suppressed due to flap damping, or 3) prefixes using an alternate best-path for a given IP Prefix. The I2RS agent should be able to notify the client via publish or subscribe mechanism.

- BGP-REQ11: I2RS client SHOULD be able to read BGP route information from BGP routers on routes in received but rejected from ADJ-RIB-IN due to policy, on routes installed in ADJ-RIB-IN, but not selected as best path, and on route not sent to IBGP peers (due to non-selection).

- BGP-REQ12: I2RS client SHOULD be able to request the I2RS agent to read installed BGP Policies.

- BGP-REQ13: I2RS client SHOULD be able to instruct the I2RS Agent to write BGP Policies into the running BGP protocols and into the BGP configurations.

- BGP-REQ14: I2RS client-agent SHOULD be able to read BGP statistics associated with Peer, and to receive notifications when certain

statistics have exceeded limits. An example of one of these protocol statistics is the max-prefix limit.

- BGP-REQ15: The I2RS client via the I2RS agent MUST have the ability to read the LOC-RIB-In BGP table that gets all the routes that the CE has provided to a PE router.

- BGP-REQ16: The I2RS client via the I2RS agent MUST have the ability to install destination based routes in the local RIB of the PE devices. This must include the ability to supply the destination prefix (NLRI), a table identifier, a route preference, a route metric, a next-hop tunnel through which traffic would be carried

- BGP-REQ17: The I2RS client via the I2RS agent SHOULD have the the ability to read the loc-RIB-in BGP table to discover overlapping routes, and determine which may be safely marked for removal.

- BGP-REQ18: The I2RS client via the I2RS Agent SHOULD have the ability to modify filtering rules and initiate a re-computation of the local BGP table through those policies to cause specific routes to be marked for removal at the outbound eBGP edge.

This summary is also listed in the [I-D.ietf-i2rs-usecase-reqs-summary].

3. BGP Protocol Operation

It is increasingly common for services facilitated via BGP to be subject to severe, widespread disruptions (outages), primarily due to the destructive teardown of BGP sessions as a result of receiving malformed BGP attributes. Unfortunately, more fine-grained BGP error handling solutions, which would result in little to no impact on the operation of BGP protocol, remain elusive.

A planned Graceful must also carefully be handled to limit the amount of traffic loss during a the shutdown. While operational requirements for the BGP mechanism for graceful shutdown of a (set of) BGP sessions is described in [RFC6198], and the operational procedures are described in [I-D.ietf-grow-bgp-gshut], additional fine-grained BGP error handling could improve graceful shutdown of BGP sessions.

This section discussed how I2RS information could improve both the destructive teardown and the graceful teardown of sessions.

3.1. BGP Error Handling for Internal BGP Sessions

It is possible that I2RS could enable enhanced error handling techniques for Internal BGP sessions. At a minimum, I2RS-capable BGP routers could signal an event such as "Malformed Attribute Received" via an I2RS agent toward an I2RS client(s). I2RS client(s) may already have a real-time view of BGP routes, and corresponding BGP attributes, or may dynamically interrogate BGP routers in the network to identify the present propagation scope of the BGP route(s) that are affected. Finally, the I2RS client(s) could then signal back to I2RS agents on BGP routers to apply a filter that would block propagation of the BGP attribute or BGP route, as necessary, in order to temporarily aid in consistency of BGP routing information across the entire network until a permanent fix can be developed and deployed within BGP routers.

I2RS would enable the global visibility and global control over the operational state of BGP, within a given Autonomous System, that is necessary to facilitate the learning of, rapid response to and more fine-grained isolation/scoping of BGP protocol events that currently cause a destructive tear-down of BGP sessions that lead to widespread disruptions of services.

3.2. Summary of I2RS Capabilities and Interactions

- BGP-REQ01: I2RS client/agent exchange SHOULD support the read, write and quick notification of status of the BGP peer operational state on each router within a given Autonomous System (AS). This operational status includes the quick notification of protocol events that proceed a destructive tear-down of BGP session

4. BGP Route Manipulation

Multiprotocol BGP [RFC4760] provides support to carry routing information for different BGP address families. Route manipulation is heavily done across these different address families for different reasons. BGP IPv4 and IPv6 address families use BGP Communities [RFC1997] and other IBGP and EBGP attributes to manipulate BGP routes for Traffic Engineering purpose. BGP VPN address families use Extended Communities [RFC4360] to filter unwanted BGP routes. BGP Flowspec address family [RFC5575] is used to install Flow based filters to filter unwanted data traffic. The following sub-sections describe the use of IRS towards BGP Route Manipulation for different BGP address families.

4.1. Customized Best Path Selection Criteria

The BGP customized Bestpath facilitates custom bestpath computations within a BGP speaking network. It is usually used within an IBGP network. Customized bestpaths use special extended communities known as cost communities. Cost communities carry enough information; Point of Insertion (POI) and the cost value to signal where in BGP bestpath the customize checks need to be done. Both, the traffic engineering as well as backdoor (SHAM) links use customized bestpath computation.

With I2RS, it would be possible for an I2RS client to push routes with custom cost communities on the BGP routers for Traffic Engineering purpose. I2RS client now can act as a central entity keeping track of all Traffic engineering data that get applied to BGP routes within an IBGP network.

4.2. Flowspec Routes

The BGP flowspec address family is used to disseminate the traffic flow specification to the BGP Autonomous System Border Routers (ASBRs) and Provider Edge (PE) routers. Both, the BGP ASBRs and the PEs would translate the received BGP traffic flow specification into an Access Control List (ACL) and install it in router’s forwarding path. Using such ACLs routers can now classify, shape, rate limit, filter, or redirect traffic flows.

With I2RS, it would be possible for an I2RS client to push traffic flow specifications to the BGP ASBRs and the PE routers. I2RS client can act as a central entity tracking all the traffic flow specifications that are installed within an IBGP network. I2RS client could also prioritize and control the announcement of traffic flow specifications according to various ASRBs and PE router’s capacity. BGP ASBRs and PE routers MAY forward traffic flow specifications received from EBGP speakers to I2RS Agents. This would allow I2RS agents to centrally manage and track any externally received traffic flow specifications.

4.3. Route Filter Routes for Legacy Routers

The BGP Route Filter address family is used to disseminate the Route Target filter information between VPN BGP speakers. This information is then used to build a route distribution graph that helps in limiting the propagation of VPN NLRI (network Layer Reachabilty Information) within a VPN network. However, it requires that all the BGP VPN routers are upgraded to support this functionality. Otherwise, the graph information is incomplete when a VPN network

consists of legacy routers that participates in VPN but does not implement the BGP route filter address family.

With I2RS, it would be possible for an I2RS client to push router filter information to BGP RR routers on behalf of all legacy routers that participates in VPN but does not support or implement the BGP route filter address family. I2RS client can act as a central entity tracking all the configured Route Filters for legacy routers and push them on appropriate RRs who in turn would push it to ASBRs and PE routers. In this way, I2RS agents help build an optimal route distribution graph that would assist in filtering of VPN NLRIs in a VPN network.

4.4. Optimized Exit Control

Optimized Exit Control is used to provide route optimization and load distribution for multiple network connections between networks. Network operators can monitor IP traffic flows and then could define policies and rules based on traffic class performance, link bandwidth monetary costs, link load distribution, traffic types, link failures, etc.

With I2RS, it would be possible for an I2RS client to manipulate BGP routes and its parameters that influence BGP bestpath decisions. I2RS client could act as a central entity that would monitor and manipulate BGP routes based on central network based policies. Such routes would then be injected by a I2RS client into the network so as to get the load distribution for multiple network connections.

4.5. Summary of I2RS Capabilities and Interactions

- BGP-REQ02: I2RS client SHOULD be able to push BGP routes with custom cost communities to specific I2RS agents on BGP routers for insertion in specific BGP Peer(s) to aid Traffic engineering of data paths. These routes SHOULD be tracked by the I2RS Agent as specific BGP routes with customer cost communities. These routes (will/will not) installed via I2RS RIB.

- BGP-REQ03: I2RS client SHOULD be able to track via read/ notifications all Traffic engineering changes applied via I2RS agents to BGP route processes in all routers in a network.

- BGP-REQ04: I2RS Agents SHOULD support identification of routers as BGP ASBRs, PE routers, IBGP routers.

- BGP-REQ05: I2RS client-agent SHOULD support writing traffic flow specifications to I2RS Agents that will install them in associated BGP ASBRs and the PE routers.

- BGP-REQ06: I2RS Client SHOULD be able to track flow specifications installed within a IBGP Cloud within an AS via reads of BGP Flow Specification information in I2RS Agent, or via notifications from I2RS agent.

- BGP-REQ07: I2RS client-agent exchange SHOULD support the I2RS client being able to prioritize and control BGP's announcement of flow specifications after the reading of status information on capacity of BGP routers (ASBR and PE). BGP ASBRs and PE routers functions within a router MAY forward traffic flow specifications received from EBGP speakers to I2RS agents, so the I2RS Agent SHOULD be able to send these flow specifications from EBGP sources to a client in response to a client query or as part of pub/sub event notification.

- BGP-REQ08: I2RS Client SHOULD be able to read BGP route filter information from I2RS Agents associated with legacy BGP routers, and write filter information via the I2RS agent to be installed in BGP RR. The I2RS Agent SHOULD be able to install these routes in the BGP RR, and engage a BGP protocol action to push these routers to ASBR and PE routers.

- BGP-REQ09: I2RS client(s) SHOULD be able to request the I2RS agent to read BGP routes with all BGP parameters that influence BGP best path decision, and write appropriate changes to the BGP Routes to BGP and to the RIB-Info in order to manipulate BGP routes

5. BGP Events

Given the extremely large number of BGP Routes in networks, it is critical to have scalable mechanisms that can be used to monitor for events affecting routing state and, consequently, reachability. In addition, similar tools are needed in order to monitor BGP protocol statistics, which help operators and developers better understand scalability of software and hardware that BGP utilizes.

I2RS could provide a publish-subscribe capability to applications to:

- request monitoring of BGP routes and related events; and,

- subscribe to the I2RS client to receive events related to BGP routes or other protocol-related events of interest.

5.1. Notification of Routing Events

There are certain IP prefixes, for example those that are arbitrarily classified by a given network operator as "high visibility" by its end-users, for which immediate notification of changes in their state

are extremely useful to know about. Upon notification of such events, a Network Operations Center (NOC) could respond to customer inquiries in a more timely fashion; alternatively, the NOC may decide to perform Traffic Engineering to restore service, etc.

Currently, the only way to learn of such events is for a BGP monitoring system to establish a BGP session with a multitude of BGP routers in an AS. Then, the BGP monitoring system needs to look through all BGP UPDATE's in order to identify those events that are of interest to it. Note, this doesn't account for the fact that there are several applications that might be simultaneously interested in learning of events to a given IP prefix nor the fact that some applications may want to dynamically insert or remove "IP prefixes of interest", depending on the needs of their constituent applications.

With I2RS, it is conceivable that applications could tell an I2RS client, through a North-Bound API, their "IP prefixes" (or, AS_PATH's, BGP communities, etc.) that are of interest. For example, a NOC application may be interested in changes to high visibility content or service-provider Web sites; alternatively, a security application may be interested in events associated with a different set of IP prefixes. The I2RS client would then consolidate the list of IP prefixes, and associated characteristics, to be monitored and program BGP routers in an AS to observe this subset of routes for changes. Some examples of changes in routing state might include:

- an IP prefix being announced or withdrawn
- an IP prefix being suppressed, due to route flap dampening
- an alternative best-path being chosen for a given IP prefix

When the requisite events for a BGP Route are observed by a BGP router, it would notify I2RS agents.

The I2RS agents would have a publish/subscribe mechanism whereby various sets of applications may subscribe to events of interest. The I2RS client would then publish these events so applications would immediately receive them and take the appropriate domain-specific action necessary.

5.2. Tracing Dropped BGP Routes

It is extremely useful to operators to be able to rapidly identify instances where a BGP route is not being propagated within an Autonomous System. At a minimum, this could result in sub-optimal performance when attempting to reach such destinations.

There are two instances when this scenario will occur. First, when a Service Provider is using "Soft Reconfiguration Inbound", it allows their ASBR routers to receive a copy of a BGP route, but show that route was not permitted into the Adj-RIB-In most likely as a result of the inbound BGP policy not permitting that IP prefix. Thus, this BGP route is not even eligible for BGP Path Selection. The second instance is where the BGP route is permitted by the inbound BGP policy into the Adj-RIB-In, but due to BGP Path Selection (i.e.: lower LOCAL_PREF, longer AS_PATH length, etc.) was not chosen as the best path and, subsequently, this particular BGP route is not forwarded on to other internal BGP speakers in the AS. In both instances, the BGP route is only visible within the ASBR on which that BGP route was first learned. Needless to say, in large Service Provider networks with a numerous interconnects to a single customer it can be very time-consuming to discover where such a BGP route is learned before ultimately determining why the route was blocked or not preferred.

With I2RS, it would be possible for an I2RS client to rapidly gather information from across a large set of BGP routers in the network to determine at what ASBR’s the BGP route is being learned. Next, the I2RS client could interrogate those routers BGP policies to determine the root cause of why the route was either not learned or not preferred in BGP. Finally, if necessary, the I2RS client(s) could amend BGP policies and push them out to BGP routers to permit the BGP route or make it a preferred route according to the BGP path selection algorithm.

5.3. BGP Protocol Statistics

There are a variety of statistics related to the operation of BGP that are invaluable to network operators. These statistics generally help operators, and developers, understand the present state and future scalability of BGP.

One statistic that is invaluable to operators is the current number of BGP routes learned through an eBGP session. Operators then apply a command against each eBGP session to limit the maximum number of BGP routes that may be learned through that eBGP session before a warning message is triggered and/or the eBGP session is torn down completely. This configuration capability is often referred to as a "max-prefix limit". This command must be routinely audited and, if necessary, adjusted in order to not trigger a false warning or teardown due to the natural organic growth in BGP routes learned from a given BGP neighbor.

I2RS agents could provide an invaluable capability to help audit and re-program the "max-prefix limit" on a periodic basis, which is

generally once per day. Specifically, the first task would be for an I2RS client to validate that there is a "max-prefix limit" applied to every eBGP session. (If there is not, that should either trigger a red alarm to the NOC to manually fix this condition or for the I2RS client to automatically apply a "max-prefix limit" that would alleviate this hazardous condition). Assuming there is a "max-prefix limit" already in place, the I2RS client would simultaneously retrieve, from each BGP router, the current number of BGP routes learned through a BGP session and value used for the "max-prefix limit" on that same BGP session. These two values could then be handed off to an application that determines if adjustments in the "max-prefix limit" value are required for each BGP session. The application would then notify the I2RS client of the subset of eBGP sessions and their associated change in "max-prefix limit" value, whereby the I2RS client would then adjust the BGP protocol configuration on each requisite BGP router in the network. Finally, it should be noted that the above is just one method whereby "max-prefix limit" values are adjusted. It's similarly possible that the BGP routers may, through the I2RS, pull the "max-prefix limit" values for each eBGP neighbor they have on-board on a periodic basis and validate their accuracy.

The above is just one use case related to BGP protocol statistics. There are wealth of other BGP protocol statistics or state information that would be invaluable to have programmatic visibility into that operators do not have today.

5.4. Summary of I2RS Capabilities and Interactions for Event statistics

I2RS SHOULD have the ability to:

- BGP-REQ10: I2RS client SHOULD be able instruct the I2RS agent(s) to notify the I2RS client when the BGP processes on an associated routing system observe a route change to a specific set of IP Prefixes and associated prefixes. Route changes include: 1) prefixes being announced or withdrawn, 2) prefixes being suppressed due to flap damping, or 3) prefixes using an alternate best-path for a given IP Prefix. The I2RS agent should be able to notify the client via publish or subscribe mechanism.

- BGP-REQ11: I2RS client SHOULD be able to read BGP route information from BGP routers on routes in received but rejected from ADJ-RIB-IN due to policy, on routes installed in ADJ-RIB-IN, but not selected as best path, and on route not sent to IBGP peers (due to non-selection).

- BGP-REQ12: I2RS client SHOULD be able to request the I2RS agent to read installed BGP Policies

- o BGP-REQ13: I2RS client SHOULD be able to instruct the I2RS Agent to write BGP Policies into the running BGP protocols and into the BGP configurations.

- o BGP-REQ14: I2RS client-agent SHOULD be able to read BGP statistics associated with Peer, and to receive notifications when certain statistics have exceeded limits. An example of one of these protocol statistics is the max-prefix limit.

6. Central membership computation for MPLS based VPNs

MPLS based VPNs use route target extended communities to express membership information. Every PE router holds incoming BGP NLRI and processes them to determine membership and then import the NLRI into the appropriate MPLS/VPN routing tables. This consumes resources, both memory and compute on each of the PE devices.

An alternative approach is to monitor routing updates on every PE from the attached CEs and then compute membership in a central manner. Once computed the routes are pushed to the VPN RIBs of the participating PEs.

This centralization of membership control has a few advantages.

- o The membership mechanism (route-targets) need not be configured in each of the PEs and can be expressed once centrally.

- o No resources in the PEs need to be spent to categorize routes into the VRF tables that they belong and to filter out unwanted state.

- o Doing it centrally means the availability of almost unlimited compute capacity to compute membership and hence can be done in a scaleable manner.

- o More sophisticated routing policies and filters can be applied during the central import/export process than can be expressed and performed using the traditional route target mechanism.

- o Routes can be selectively pushed only to the participating PE’s further reducing the memory load on the individual routers in the network. This further obviates for a distributed mechanisms such as rt constraints to reduce unnecessary path state in the routers.

Note that centrally computation of membership can be applied to other scenarios as well such as VPLS, MVPNs, MAC VPNs and others. Depending on the scenario, what gets monitored from the CE might vary. Central computation will especially help VPLS where multi-

homing and load balancing using distributed techniques has particularly been a challenge.

Also note that one of the biggest promises of central route computation is simplification and reduction of computation and memory load on all devices in the network. This use case is just one example that illustrates these benefits of central computation very well.

Summary of I2RS Capabilities and Interactions:

- BGP-REQ15:The I2RS client via the I2RS agent MUST have the ability to read the loc-RIB-In BGP table that gets all the routes that the CE has provided to a PE router.

- BGP-REQ16:The I2RS client via the I2RS agent MUST have the ability to install destination based routes in the local RIB of the PE devices. This must include the ability to supply the destination prefix (NLRI), a table identifier, a route preference, a route metric, a next-hop tunnel through which traffic would be carried

7. Marking Overlapping Traffic Engineering Routes for Removal

It is often the case that routes are advertised not to provide reachability (in the strict sense), but rather to provide optimal reachability, or to engineer the path traffic takes to a particular destination. While this can improve the efficiency of a network's operation, it can also increase the amount of state carried in the control plane beyond the point where the additional state has any real effect on traffic flow. Removing Overlapping Routes [I-D.white-grow-overlapping-routes] provides a mechanism designed to remove these traffic engineering routes once they are beyond the point of actually impacting traffic flows in the network.

Summary of I2RS Capabilities and Interactions:

- BGP-REQ17: The I2RS client via the I2RS agent SHOULD have the the ability to read the loc-RIB-in BGP table to discover overlapping routes, and determine which may be safely marked for removal.

- BGP-REQ18: The I2RS client via the I2RS Agent SHOULD have the ability to modify filtering rules and initiate a re-computation of the local BGP table through those policies to cause specific routes to be marked for removal at the outbound eBGP edge.

8. IANA Considerations

This document makes no request of IANA.

9. Security Considerations

The BGP use cases described in this document assumes use of I2RS programmatic interfaces described in the I2RS framework mentioned in [I-D.ietf-i2rs-architecture]. This document does not change the underlying security issues inherent in the existing in [I-D.ietf-i2rs-architecture].

10. Acknowledgements

The authors would like to thank Ed Crabbe, Joel Halpern, Wes George, Carlos Pignataro, Jon Mitchell, Rex Fernando, Hannes Gredler, Shane Amante, Bill Atwood for their comments and suggestions.

11. References

11.1. Normative References

[I-D.ietf-i2rs-architecture]
Atlas, A., Halpern, J., Hares, S., Ward, D., and T. Nadeau, "An Architecture for the Interface to the Routing System", draft-ietf-i2rs-architecture-09 (work in progress), March 2015.

[I-D.ietf-i2rs-usecase-reqs-summary]
Hares, S. and M. Chen, "Summary of I2RS Use Case Requirements", draft-ietf-i2rs-usecase-reqs-summary-01 (work in progress), May 2015.

[RFC1997] Chandra, R., Traina, P., and T. Li, "BGP Communities Attribute", RFC 1997, DOI 10.17487/RFC1997, August 1996, <http://www.rfc-editor.org/info/rfc1997>.

[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>.

[RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, DOI 10.17487/RFC2629, June 1999, <http://www.rfc-editor.org/info/rfc2629>.

[RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", BCP 72, RFC 3552, DOI 10.17487/RFC3552, July 2003, <http://www.rfc-editor.org/info/rfc3552>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <http://www.rfc-editor.org/info/rfc4271>.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, DOI 10.17487/RFC4360, February 2006, <http://www.rfc-editor.org/info/rfc4360>.

[RFC5492] Scudder, J. and R. Chandra, "Capabilities Advertisement with BGP-4", RFC 5492, DOI 10.17487/RFC5492, February 2009, <http://www.rfc-editor.org/info/rfc5492>.

11.2. Informative References

[I-D.ietf-grow-bgp-gshut]
Francois, P., Decraene, B., Pelsser, C., Patel, K., and C. Filsfils, "Graceful BGP session shutdown", draft-ietf-grow-bgp-gshut-06 (work in progress), August 2014.

[I-D.mcpherson-irr-routing-policy-considerations]
McPherson, D., Amante, S., Osterweil, E., and L. Blunk, "IRR & Routing Policy Configuration Considerations", draft-mcpherson-irr-routing-policy-considerations-01 (work in progress), September 2012.

[I-D.white-grow-overlapping-routes]
White, R., Retana, A., and S. Hares, "Filtering of Overlapping Routes", draft-white-grow-overlapping-routes-03 (work in progress), October 2014.

[RFC2622] Alaettinoglu, C., Villamizar, C., Gerich, E., Kessens, D., Meyer, D., Bates, T., Karrenberg, D., and M. Terpstra, "Routing Policy Specification Language (RPSL)", RFC 2622, DOI 10.17487/RFC2622, June 1999, <http://www.rfc-editor.org/info/rfc2622>.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, DOI 10.17487/RFC4760, January 2007, <http://www.rfc-editor.org/info/rfc4760>.

[RFC5575] Marques, P., Sheth, N., Raszuk, R., Greene, B., Mauch, J., and D. McPherson, "Dissemination of Flow Specification Rules", RFC 5575, DOI 10.17487/RFC5575, August 2009, <http://www.rfc-editor.org/info/rfc5575>.

[RFC6198] Decraene, B., Francois, P., Pelsser, C., Ahmad, Z., Elizondo Armengol, A., and T. Takeda, "Requirements for the Graceful Shutdown of BGP Sessions", RFC 6198, DOI 10.17487/RFC6198, April 2011, <http://www.rfc-editor.org/info/rfc6198>.

[RFC6890] Cotton, M., Vegoda, L., Bonica, R., Ed., and B. Haberman, "Special-Purpose IP Address Registries", BCP 153, RFC 6890, DOI 10.17487/RFC6890, April 2013, <http://www.rfc-editor.org/info/rfc6890>.

Appendix A. BGP Configuration

The configuration of BGP is arduous to establish and maintain, particularly on networks whose services have a requirement for complex routing policies. This need is magnified by the need to routinely perform changes to large numbers of BGP routers to, for example: add or remove customer's BGP sessions, announce or withdraw (customer) IP prefixes in BGP, modify BGP policies to effect changes in Traffic Engineering, audit BGP routers to ensure they have consistent and appropriate BGP policies, and others.

There are three categories of BGP configuration:

1. Local BGP routing protocol configuration: local Autonomous System Number (ASN), BGP path selection properties of the router, injection of (aggregate) routes into BGP, etc.

2. Local BGP policies: policies designed to filter and/or manipulate BGP attributes associated with BGP routes learned through BGP sessions. These policies typically live in the global configuration of a BGP router, but are applied on a per-BGP neighbor basis (or, group of BGP neighbors); and,

3. BGP neighbor sessions: remote ASN, remote IP address, address families, BGP policies to applied to routes, max-prefix limits, etc.

The sum total of BGP configuration on a BGP router is typically the largest quantify of configuration on Service Provider's BGP routers, by a fairly large margin. When that is combined with the large set of routine configuration changes, mentioned above, it should be fairly clear that systematic reading, configuration and control of

BGP routers through a mechanism like I2RS would greatly benefit all operators of BGP routers.

While it may not be possible to provide programmatic APIs for esoteric vendor-specific policy configuration, it is possible to provide such API's for BGP protocol specific configuration and the more commonly used BGP routing policies.

A.1. BGP Protocol Configuration

Ability to enable and disable new address families within a BGP protocol for a network of BGP speaking routers is a challenge. The challenge is mainly in keeping track of BGP speaker's feature capabilities and then configuration of new address families on a multiple BGP speakers within a given network. With the necessary information, I2RS agents allow a network operator to push configuration information for enabling and disabling of new address families on a partial or entire set of BGP speakers within a given network. This would assist in building BGP overlay networks as needed.

For VPN address families, the main challenge lies in the complex VPN configuration required to setup the control plane for Customer VPNs. The configuration involves creating a Virtual Routing and Forwarding instance (VRF), a Route Distinguisher (RD) that ensures each customer prefixes remains unique across VPNs, and Route Targets (RT) that help ensure that the Customer prefixes are segregated appropriately so that they do not cross the VPN boundaries. I2RS would allow a network operator to push such configuration from a central location where a global VPN provisioning information could be stored. This helps avoid manual configuration of a VPN on multiple routers. Instead the configuration is controlled and pushed though a central I2RS client using a programmatic set of APIs on targeted set of BGP speakers.

Use of I2RS agents to announce protocol configuration information would simplify and automate configuration of BGP protocol in IBGP deployments where the protocol based policies are seldom used. To facilitate such a centralized configuration model, BGP speakers could be extended to use programmatic APIs to announce their feature capabilities as part of protocol initialization to the centralize I2RS agents. This would assist I2RS agents to auto-discover BGP protocol capabilities of various BGP speakers in a given network. I2RS agents in turn would use the information towards enabling/ disabling of BGP specific features on BGP speakers.

A.2. BGP Policy Configuration

Filtering of BGP routes is strongly recommended to control the announcements of BGP prefixes across the internet. Most providers make extensive use of BGP prefix filtering policies at the edge of their networks. The reasons for filtering BGP prefixes are:

- Avoid Unwanted Route Announcements. Filter prefixes that MUST NOT be routed [RFC6890]. Filter prefixes that are not allocated by Internet Routing Registries.

- Facilitate Route Summarization. Filter prefixes beyond certain agreed prefix mask length between providers. Route Summarization helps control BGP RIB and FIB table size.

- Defensive Security. Filter prefixes from Stub customer ASes that are not owned by the customers. Filter customer prefixes announced by other providers. This helps avoid prefix hijacking.

A set of standards-based schemas to enable configuration of Local BGP policies and BGP neighbor sessions was realized through the Routing Policy Specification Language (RSPL) [RFC2622]. The RPSL defined a standards-based schemas, or 'objects' as it called them, that defined:

- binding of IP prefixes to (one or more) Origin AS, (route objects);

- collections of routes (route-set objects);

- collections of Autonomous Systems (as-set objects); and,

- routing policy of an Autonomous System to/from its adjacent neighbor AS'es, (aut-num objects)

Each ASN is responsible for creation, modification and deletion of its RPSL objects in an Internet Routing Registry (IRR). IRR's are typically operated by Regional Internet Registries (RIR's) and a few dozen larger ISP's and independent organizations. The IRR's provide a well-known location for all organizations attached to the Internet to retrieve or update RPSL objects.

While still widely and actively used by Internet Service Providers, the prevailing belief is that the data contained in the IRR's is inaccurate, primarily due to a lack of deployed authorization method with respect to the creation of modification of RPSL objects. It should be noted that this criticism is not directed at the previously defined RPSL schemas, but rather at the data contained in RPSL

schemas by end-users of the IRR system. Please refer to the IRR And Routing Policy Configuration Considerations [I-D.mcpherson-irr-routing-policy-considerations] document for a more thorough discussion of the history and present state of the IRR's.

Currently, RPSL schemas are exchanged between non-routing systems (servers) used within the IRR system. In addition, open-source and proprietary applications create or modify RPSL schemas, as necessary, to signal the announcement (or, withdrawal) of an IP prefix from an ASN or the creation (or, teardown) of a neighbor relationship between two adjacent ASN's. Most importantly, these RPSL schemas are consumed by similar applications to automatically build routing policies, (i.e.: lists of IP prefixes, corresponding Origin ASN's and/or AS_PATH's), that then get translated to device-specific syntax (i.e.: CLI) before being pushed into individual BGP routers to effect routing policy on the network. It is common for Internet Service Providers to perform updates to these routing policies across their entire network on a daily basis.

With I2RS it would be desirable to change the last step in the above process so that BGP policies derived from RPSL schemas, and other information sources, are translated into standards-based schemas that are then pushed, or pulled, into individual BGP routers. More generally, I2RS agents could use API's to gather information required to build various types of BGP routing policies plus the corresponding set of Autonomous System Border Routers (ASBR's) where such policies need to be applied in the network and, finally, making those changes to individual network elements so those BGP policies take effect in the network. In doing so, a network operator now has a centralized way of building and making these policies take effect across the network in a coordinated manner.

Authors' Addresses

Keyur Patel
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
USA

Email: keyupate@cisco.com

Russ White
Ericsson

Email: russw@riw.us

Susan Hares
Huawei

Email: shares@ndzh.com