IDR Working Group   R. Raszuk
Internet-Draft   C. Cassar
Intended status: Standards Track   Cisco Systems
Expires: December 4, 2011   E. Aman
                                                        TeliaSonera
                                                        B. Decraene
                                                     France Telecom
                                                       June 2, 2011

BGP Optimal Route Reflection (BGP-ORR)
draft-ietf-idr-bgp-optimal-route-reflection-00

Abstract

   [RFC4456] asserts that, because the Interior Gateway Protocol (IGP)
   cost to a given point in the network will vary across routers, "the
   route reflection approach may not yield the same route selection
   result as that of the full IBGP mesh approach." One practical
   implication of this assertion is that the deployment of route
   reflection may thwart the ability to achieve hot potato routing. Hot
   potato routing attempts to direct traffic to the closest AS egress
   point in cases where no higher priority policy dictates otherwise.
   As a consequence of the route reflection method, the choice of exit
   point for a route reflector and its clients will be the egress point
   closest to the route reflector - and not necessarily closest to the
   RR clients.

   Section 11 of [RFC4456] describes a deployment approach and a set of
   constraints which, if satsified, would result in the deployment of
   route reflection yielding the same results as the iBGP full mesh
   approach. Such a deployment approach would make route reflection
   compatible with the application of hot potato routing policy.

   As networks evolved to accommodate architectural requirements of new
   services, tunneled (LSP/IP tunneling) networks with centralized route
   reflectors became commonplace. This is one type of common deployment
   where it would be impractical to satisfy the constraints described in
   Section 11 of [RFC4456]. Yet, in such an environment, hot potato
   routing policy remains desirable.

   This document proposes two new solutions which can be deployed to
   facilitate the application of closest exit point policy centralized
   route reflection deployments.

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 4, 2011.

Copyright Notice

Table of Contents

1. Introduction

There are three types of BGP deployments within Autonomous Systems
today: full mesh, confederations and route reflection.

BGP route reflection is the most popular way to distribute BGP routes
between BGP speakers belonging to the same administrative domain.
Traditionally route reflectors have been deployed in the forwarding
path and carefully placed on the POP to core boundaries. That model
of BGP route reflector placement has started to evolve. The
placement of route reflectors outside the forwarding path was
triggered by applications which required traffic to be tunneled from
AS ingress PE to egress PE: for example L3VPN.

This evolving model of intra-domain network design has enabled
deployments of centralized route reflectors. Initially this model
was only employed for new address families e.g. L3VPNs, L2VPNs etc

With edge to edge MPLS or IP encapsulation also being used to carry
internet traffic, this model has been gradually extended to other BGP
address families including IPv4 and IPv6 Internet routing. This is
also applicable to new services achieved with BGP as control plane
for example 6PE.

Such centralized route reflectors can be placed on the POP to core
boundaries, but they are often placed in arbitrary locations in the
core of large networks.

Such deployments suffer from a critical drawback in the context of
best path selection. A route reflector with knowledge of multiple
paths for a given prefix will pick the best path and only advertise
that best path to the the route reflector clients. If the best path
for a prefix is selected on the basis of an IGP tie break, the best
path advertised from the route reflector to its clients will be the
exit point closest to the route reflector. But route reflector
clients will be in a place in the network toplogy which is different
from the route reflector. In networks with centralized route
reflectors, this difference will be even more acute. It follows that
the best path chosen by the route reflector is not necessarily the
same as the path which would have been chosen by the client if the
client considered the same set of candidate paths as the route
reflector. Furthermore, the path chosen by the client might have
been a better path from that chosen by the route reflector for
traffic entering the network at the client. The path chosen by the
client would have guaranteed the lowest cost and delay trajectory
through the network.

Route reflector clients switch packets using routing information

learnt from route reflectors which are not on the forwarding path of
the packet through the network even in the absence of end-to-end
encapsulation. In those cases the path chosen as best and propagated
to the clients will often not be the optimal path chosen by the
client given all available paths.

Eliminating the IGP distance to the BGP nexthop as a tie breaker on
centralized route reflectors does not address the issue. Ignoring
IGP distance to the BGP next hop results in the tie breaking
procedure contributing the best path by differentiating between paths
using attributes otherwise considered less important than IGP cost to
the BGP nexthop.

One possible valid solution or workaround to this problem requires
sending all domain external paths from the RR to all its clients.
This approach suffers the significant drawback of pushing a large
amount of BGP state to all the edge routers. In many networks, the
number of EBGP peers over which full Internet routing information is
received would correlate directly to the number of paths present in
each ASBR. This could easily result in tens of paths for each
prefix.

Notwithstanding this drawback, there are a number of reasons for
sending more than just the single best path to the clients. Improved
path diversity at the edge is a requirement for fast connectivity
restoration, and a requirement for effective BGP level load
balancing. Protocol extensions like add-paths
[I-D.ietf-idr-add-paths] or diverse-path
[I-D.ietf-grow-diverse-bgp-path-dist] allow for such improved path
diversity and can be used to address the same problems addressed by
the mechanisms proposed in this draft. In practical terms, add/
diverse path deployments are expected to result in the distribution
of 2, 3 or n (where n is a small number) 'good' paths rather than all
domain external paths. While the route reflector chooses one set of
n paths and distributes those same n paths to all its route reflector
clients, those n paths may not be the right n paths for all clients.
In the context of the problem described above, those n paths will not
necessarily include the closest egress point out of the network for
each route reflector client. The mechanisms proposed in this
document are likely to be complementary to mechanisms aimed at
improving path diversity.

2. Proposed solutions

This document proposes two simple solutions to the problem described
above. Both of these solutions make it possible for route reflector
clients to direct traffic to their closest exit point in hot potato

routing deployments, without requiring further state to be pushed out
to the edge. These solutions are primarily applicable in deployments
using centralized route reflectors, which are typically implemented
in devices without a capable forwarding plane.

The two alternatives are:

   "Best path selection for BGP hot potato routing from client's IGP
   network position"

   "Angular distance approximation for BGP warm potato routing"

Both solutions rely upon all route reflectors learning all paths
which are eligible for consideration for hot potato routing. In
order to satisfy this requirement, path diversity enhancing
mechanisms such as add paths/diverse paths may need to be deployed
between route reflectors.

In both of these solutions the route reflector selects and
distributes a route to each client based on what would be optimal
from the client's perspective. By optimal we refer in this document
to the decision made during best path selection at the IGP metric to
BGP next hop comparison step. Clearly the overall path selection
preference may be chosen based other policy step and provisions as
defined in this document would not apply.

In the respective solutions the choice is made either factoring in
IGP costs or the configured angular distance to the next hop. The
route reflector makes different decisions for different clients only
in the case where the tie breaker for path selection would have been
the IGP distance to the BGP nexthop (as in hot potato routing).

A signficant advantage of this approach is that the RR clients do not
need to run new software or hardware.

[3](#). **Best path selection for BGP hot potato routing** from customized IGP
    network position

This section describes a method for calculating the order of
preference of BGP paths from the point of view of each separate route
reflector client. More specifically, the route relflector will
compute the IGP metric to the BGP nexthop from the position of the
client to which the resulting path will be distributed, if the IGP
metric is the tie breaker applied to a set of possible paths. In the
subsequent model authors will propose virtual reflector placement at
operator's selected IGP location.

In the case of a hierarchical IGP deployment where the client is in a
different level in the hierarchy to the route reflector, the route
reflector will compute IGP distance to the BGP nexthop from the Area
Border Routers (ABR) leading to the client in lieu of the route
reflector client itself, and use the shortest distance from these
ABRs to the nexthop. This provides an approximation to the desired
functionality. Rather than a client picking the closest path, the
client would be picking the exit point closest to the client region
as defined by area or level. In cases where one or more nexthops are
in the same region as the client, one of those nexthops would be
preferred, with tie breaking within those nexthops performed from the
route reflector's position in the network.

It is assumed that reachability through a set of ABRs is always
advertised through identical prefixes from those ABRs. If a nexthop
is reachable through multiple ABRs but the ABRs advertise
reachability through prefixes of different length, then only the ABR
advertising the longest prefix will be considered as a viable path to
the nexthop.

BGP best path selection and its distribution has a natural
consequence of limiting the amount of state in the network. That is
not in itself a drawback. BGP speakers will rarely need to receive
all available BGP paths. In network deployments with multiple
upstream peerings or with very dense peering schemes, the number of
available BGP paths for a given BGP prefix can be high. Real network
deployments with the number of paths for a prefix ranging from 10s to
100s have been observed. It would be wasteful to propagate all of
those paths to all clients, such that each client can select paths
according to the position of the nexthop relative to the client.

Whenever a BGP route reflector would need to decide what path or
paths need to be selected for advertisement to one of its clients,
the route reflector would need to virtually position itself in its
client IGP network location in order to choose the right set of paths
based on the IGP metric to the next hops from the client's
perspective.

This technique applies in deployments with or without diverse paths
or the various path selection modes contemplated in add-paths.

In the network architectures consisting of more then single pair of
route reflectors it is required that all reflectors are fully meshed
and have ability to learn and maintain all external BGP paths. In
the event of constructing a hierarchy of reflectors to relax the full
RR mesh requirements ORR should not be run between such route
reflectors.

3.1. Client's perspective best path selection algorithm

   For each centralized route reflector the proposal assumes that the
   route reflector participates in a common IGP with its clients. There
   are two scenarios to consider - flat versus hierarchical IGP network.

3.1.1. Flat IGP network

      Reflectors run SPF from the client IGP node point of view such
      that the cost of BGP nexthops from the client can be determined if
      necessary. For the purpose of BGP path selection the interesting
      product of this calculation is the ability to determine the IGP
      distance from a client to a BGP next hop. This distance to a
      nexthop would be interesting in cases where that next hop is for a
      path which is contending with otherwise equally preferred paths.
      This approach works in tunneled as well as conventional hop-by-hop
      IP forwarding cores.

      When the path selection tie breaker for a prefix is the IGP metric
      to the BGP nexthops of the contending paths, then the route
      reflector will determine the order of preference of the contending
      paths by considering the distance from the client to the path
      nexthops in order to decide what path/s to advertise to a client
      (or group of clients where feasible). It should be noted that an
      operator may wish to provide a distance tolerance value, such that
      beyond a certain granularity, differences between IGP metric are
      invisible to the path selection algorithm. This will allow a
      route reflector some leeway in selecting between paths such that
      rather than pick one path over another on the basis of a
      difference in distance which is operationally irrelevant, the
      route reflector can choose to optimize for update generation
      grouping. Furthermore, this tolerance will reduce the likelihood
      of generation of BGP updates when the IGP topology changes in a
      way which is not operationally relevant. In the case that a path
      is selected from a set for a given prefix while ignoring
      differences in distance within the tolerance figure, then that
      same path must always be preferred for all clients where the paths
      are within the tolerance figure

3.1.2. Hierarchical IGP network

   Hierarchy introduces two challenges:

      The first challenge is that the RR IGP view may differ from a
      client IGP view by virtue of one or the other having a summarised
      view versus the other. Summarisation, by its nature, loses
      information. Consider the example where a client within a PoP
      sees two prefixes with two metrics for two egress points within

the PoP, but where the RR only sees a single summary covering
reachability to both nexthops as injected by the ABR. For
clarification purposes in the case of ISIS by ABR we refer to
L1/L2 node. However it needs to be observed that inter area
networks running LDP are required to disable summarisation of all
FEC advertised in LDP (typically all loopbacks) unless [RFC5283]
is deployed. Such deployments are not likely to suffer
summarisation difficulties.

The second challenge is that in cases where the client is in a
different level of hierarchy from the RR, the RR can not build a
Shortest Path First (SPF) tree with the client node as root,
simply because the topology derived by the IGP will not include
the client node. It will instead only include reachability to the
client from one or more ABRs. In order to overcome this problem,
the RR could compute an SPF tree from the ABRs in the area. The
RR would then determine the shortest distance from a client which
lives behind the ABRs, to a nexthop, by adding the advertised
distances from an ABR to the client and the distance from the ABR
to a nexthop, for each ABR, and picking the minimum. This assumes
that IGP metrics on links are symmetric; i.e. that the distance
from the ABR to the client or nexthop is equal to the distance
from the client or nexthop to the ABR.

There are cases where the above approach does not help. If RR is
trying to arbitrate amongst a set of paths for a client which is
in the same hierarchy as some of those paths, and in a different
hierarchy to the RR, the opaqueness of the region containing the
client at the RR defeats the selection process. It is impossible
to determine the relative position of the RR client and the paths
within the client region.

The solution for hierarchical IGP networks also assumes that if
RRs are present and are responsible for calculation of BGP best
path to clients they are either placed in each local area
coinciding with area containing clients or they are placed in the
core (area 0/level 2) of the network.

3.2. Aside: Configuration-based flexible route reflector placement

The ability to exploit topology information available in the IGP in
ways described above can also be used to virtually place the RR at
different points in the network for purposes other than hot potato
routing.

A route reflector can be globally configured to "pretend" its logical
location is one of any of the other nodes within a given IGP area/
level flooding scope regardless of its physical connectivity.

Such flexibility provides a useful tool for reflector virtualization, and supports moving or replacing physical route reflectors without any effect on routing. Such a change can be permanent or it could be performed during network maintenance in order to minimize network impact.

A possible variation would allow the virtual placement of RR to be effected on a per-AF or AF plus update/peer group granularity. It should be noted that this approach provides for splitting one centralized route reflector such that it is virtually positioned at various network locations, with the network location depending upon of address family or address family plus update/peer group.

Virtual slicing of a centralized route reflector relaxes the need to propagate all BGP paths between RRs in a alternative conventional distributed RR deployment. It is expected that such RRs would be deployed in redundant sets, and that those RRs would not need to be physically colocated, while still benefiting from the possibility of being logically colocated, and therefore not compromising any of the best path selection symmetry.

3.3. Route reflector client grouping

It may be appropriate to allow the operator, or the route reflector itself, to group clients together using IGP distance between clients to determine grouping. All the operation discussed above which relied upon computing best path for each client, and measuring distances from each client to different nexthops, would instead be performed for each group of clients. Configurable thresholds can be used to determine which IGP metric changes should be visible to BGP, and trigger best paths recomputation. The latter would be beneficial in existing BGP RR code too.

Alternatively route reflector client grouping could be accomplished statically by the operator by coloring clients belonging to a common group (for example being part of the same POP). In order to accomplish such marking it is proposed that BGP OPEN message be augmented with an optional paramiter indicating the Group ID given peer belongs to.

3.3.1. Route Reflector Client Group ID

This is an Optional Parameter in BGP OPEN message that is used by a BGP speaker to convey to its route reflectors the Group ID value. Such value will allow automatic and predictable peer grouping on the route reflectors as deemed necessary from operator's network architecture.

The parameter contains precisely one set of [Group_ID Code, Group_ID
Length, Group_ID Value] encoded as shown below:


```
+----------------------------+
| Group ID Code (1 octet)    |
+----------------------------+
| Group ID Length (1 octet)  |
+----------------------------+
| Group ID Value (4 octets)  |
+----------------------------+
```

The use and meaning of these fields are as follows:

Group ID Code:

    Group ID Code is a one octet field that identifies Group ID
    optional parameter of BGP OPEN message. Value TBD by IANA
    Recommended value: 3.

Group ID Length:

    Group ID Length is a one octet field that contains the length
    of the Group ID Value field in octets. It is fixed and equals
    to 4.

Group ID Value:

    Group ID Value is a fixed length field of size equal to
    four octets that contains the numerical value of group given
    BGP speaker should be part of on the route reflector.

    Two special values are reserved:

        0x00000000 - No grouping preference
        0xFFFFFFFF - Do not group this BGP speaker

    An implementation may allow automatic population of
    GROUP_ID value using IGP area identifier.

Route reflectors or EBGP speakers receiving such Group IDs from their
respective BGP peers as part of the BGP OPEN procedure MAY use them
when constructing update or peer groups in addition to any of the
existing grouping mechanism already available. An implementation may

allow operator to explicitly allow or disallow honoring such grouping
or provide means for manual overwrite via explicit configuration.

[3.4](). Discussion

This is not the first instance where a router participating in an IGP
is required to build the SPF tree using a root other than itself.
Determination of loop free alternate paths as described in [[RFC5714]()]
is one such example.

Determining the shortest path and associated cost between any two
arbitrary points in a network based on the IGP topology learned by a
router is expected to add some extra cost in terms of CPU resource.
However SPF tree generation code is now implemented efficiently in a
number of implementations, and therefor this is not expected to be a
major drawback. The number of SPTs computed in the general non-
hierarchical case is expected to be of the order of the number of
clients of an RR whenever a topology change is detected. Advanced
optimizations like partial and incremental SPF may also be exploited.
By the nature of route reflection, the number of clients can be split
arbitrarily by the deployment of more route reflectors for a given
number of clients. While this is not expected to be necessary in
existing networks with best in class route reflectors available
today, this avenue to scaling up the route reflection infrastructure
would be available. If we consider the overall network wide cost/
benefit factor, the only alternative to achieve the same level of
optimality would require significantly increasing state on the edges
of the network, which, in turn, will consume CPU and memory resources
on all BGP speakers in the network. Building this client perspective
into the route reflectors seems appropriate.

[3.5](). Advantages

The solution described provides a model for integrating the client
perspective into the best path computation for RRs. More
specifically, the choice or BGP path factors in the IGP metric
between the client and the nexthop, rather than the distance from the
RR to the nexthop. The documented method does not require any BGP or
IGP protocol changes as required changes are contained within the RR
implementation.

This solution can be deployed in traditional hop-by-hop forwarding
networks as well as in end-to-end tunneled environments. In the
networks where there are multiple route reflectors and unencapsulated
hop-by-hop forwarding, such optimizations should be enabled on all
route reflectors. Otherwise clients may receive an inconsistent view
of the network and in turn lead to intra-domain forwarding loops.

With this approach, an ISP can effect a hot potato routing policy
even if route reflection has been moved from the forwarding plane to
the core and hop-by-hop switching has been replaced by end to end
MPLS or IP encapsulation.

As per above, the approach reduces the amount of state which needs to
be pushed to the edge in order to perform hot potato routing. The
memory and CPU resource required at the edge to provide hot potato
routing using this approach is lower than what would be required in
order to achieve the same level of optimality by pushing and
retaining all available paths (potentially 10s) per each prefix at
the edge.

The proposal allows for a fast and safe transition to BGP control
plane route reflection without compromising an operator's closest
exit operational principle. Hot potato routing is important to most
ISPs. The inability to perform hot potato routing effectively stops
migrations to centralized route reflection and edge-to-edge LSP/IP
encapsulation for traffic to IPv4 and IPv6 prefixes.

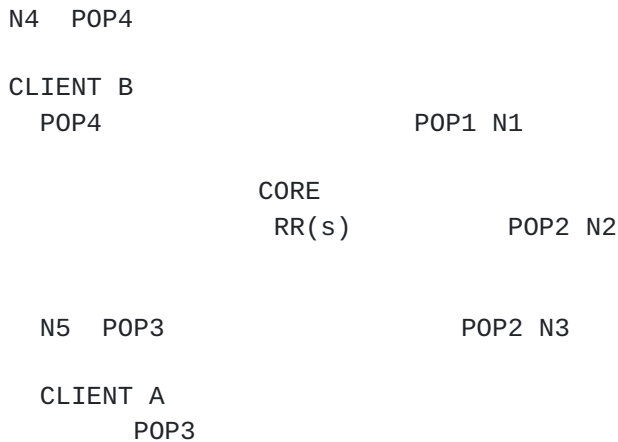## 4. Angular distance approximation for BGP warm potato routing

This section describes an alternative solution to the use of IGP
topology information to virtually position the RR at the client
location in the network. This solution involves modelling the
network topology as a set of elements (regions, PoPs or routers)
arranged in a circle. Route reflector clients and inter-domain exit
points would then be statically assigned to those elements such that
one can compute the angular distance between route-reflector clients
and the various exit points in order to infer the distance between
any two elements. This measure of distance can be used as an
effective alternative to the IGP distance as a tie breaker in the
path selection algorithm if necessary.

4.1. Problem statement

This solution addresses the problem described in earlier sections,
while attempting to minimize computational overhead. The aim of the
proposed solution is to enable a route reflector to provide a route
reflector client with an exit point for a prefix which is 'closest'
to the client rather than the route-reflector, without having to
distribute all paths to that client, or having to derive each
client's view of the network topology. The measure of closest is
based on a simplistic description of network topology provided by the
operator.

Consider the following example of an ISP network topology drawn to

reflect the location of the nodes and POPs:

          N4   POP4

          CLIENT B
            POP4   POP1 N1

                          CORE
                           RR(s)   POP2 N2

            N5   POP3   POP2 N3

            CLIENT A
                  POP3

   N - represents the different exit points for a given prefix. POP2 is
   a geographically large PoP with two paths; N2 and N3.

   In a deployment where the centralized RRs tie break on the basis of
   their IGP-based view of the network, N1 above would be advertised to
   all clients on the basis that it is closest to the RR. Path N4 would
   be a more appropriate choice for client B. Similarly, N5 would be
   more appropriate for client A since path N5 is closer to client A
   then path N1.

4.2. Proposed solution

   The proposed solution revolves around the operator establishing the
   angular position of the route-reflector clients and inter-domain exit
   points in the network. The route reflector then picks the path to
   advertise to a client based on the client's angular position versus
   the angular position of the inter-domain exit points originating the
   paths. The operator can choose the granularity of angular position
   appropriate to the desired goals. On one hand, the coarseness of the
   angular position will effect the operator overhead; versus the
   optimality of routing on the other. The finest granularity possible
   will be the relative position of originating clients.

   Note that this solution has nothing to do with actual IGP link
   metrics and resulting topology in the network.

   It can be shown that for each network topology, elements such as AS

exit points can be mapped on to a circle. By putting POPs, Regions
or individual clients onto the hypothetical circle we can identify an
angular location for each element relative to some fixed direction;
for example defining the angular north of the circle at 0 degress.

The angular position of elements in the network can be conveyed to a
route reflector in a number of ways:

   Assignment of angular position of each RR client through
   configuration on the route reflector itself; per client
   configuration on RR

   Assignment of angular position of an RR client at each client,
   then propagating it to RRs.

The proposed angular distance approximation is compatible with both
flat and hierarchical IGP deployments.

In the example illustrated above the route reflector might learn or
be configured with the following set of paths and corresponding
angular positions:

Prefix X/Y   N1   N2   N3   N4   N5

Location
in degrees   60   85   120   290   260

If the absolute angular position of clients A and B were as follows:

   Client A: 260 degrees

   Client B: 290 degrees

Then the corresponding angular distances for those clients versus the
exit points can be calculated as follows:

 Prefix X/Y   N1   N2   N3   N4   N5

 Client A   200   175   140   30   0

 Client B   230   205   170   0   30

With an RR running the BGP best path algorithm modified to use the
angular distance from the client to the nexthops, rather than its IGP
distance to the nexthops as tie breaker, each client is provided with
its closest path with the measure of closeness reflecting the angular

position as configured by the operator.

The model used by the operator in order to determine the angular
position of a client or exit point, might involve grouping elements
together by region or PoP, or might involve no grouping at all.
Implementations should allow the operator to pick the appropriate
granularity.

4.3. Centralized vs distributed route reflectors

In an environment where the RR clusters are distributed (yet
centralized enough to make hot potato routing hard), and each RR
cluster serves a subset of clients, it becomes necessary to propagate
the angular position of the clients between route reflectors. This
can be achieved as follows:

    Deploy add-paths between route reflectors in order to maximize
    path diversity within the cluster.

    A non AS transitive BGP community of type (TBA by IANA) can be
    used to encode and propagate angular position between 0 and 359 of
    a client. This community is only relevant to the route reflectors
    of a given BGP domain and should be stripped either at the ASBR
    boundary or when propagating updates to BGP peers which are not
    route reflectors.

The angular position marking could also be added by clients and
advertised to the route reflector. This would require some
configuration effort.

5. Deployment considerations

The solutions are primarily intended for end-to-end tunneled
environments, i.e. where traffic is label switched or IP tunneled
across the core. If unencapsulated hop-by-hop forwarding is used,
either misconfigurations or conflicts between these optimizations and
classical BGP path selection rules could lead to intra-domain
forwarding loops. Under certain circumstances the solutions can also
be deployable without end-to-end tunneling. In particular the best
path selection based on the client's IGP best-path selection is
guaranteed not to cause any forwarding loops (other than micro loops
associated with reconvergence) when deployed in a flat IGP area
provided that no distance tolerance value is used so that the path
choice is truly made on a per-client basis.

It should be self evident that this solution does not interfere with
policies enforced above IGP tie breaking in the BGP best path

algorithm.

The solution applies to NLRIs of all address families which can be
route reflected and which can be tie broken by IGP distance to the
nexthop.

It should be noted that customized per-client or group of clients
best path selection is already in use today in the context of
Internet Exchange Point (IXP) route servers. In an IXP route server
the client best path is selected as a result of different policies
rather than IGP metric distance to BGP next hop.

A possible scalability impact of optimizing path selection to take
account of the RR client position is that different RR clients
receive different paths, and therefore update/peer group efficiency
diminishes. This cost is imposed by the requirement given the
requirement is to optimize the egress path from the client's
perspective. It is also not unlikely that groups of clients will end
up receiving the same best path/s, in which case, inefficiency of
update generation will be minimized. It should be noted that in the
cases described under flexible router placement where placement is
determined on a per update/peer group basis or per route reflector,
the scale benefits of peer groupings are retained.

6. **Security considerations**

No new security issues are introduced to the BGP protocol by this
specification.

7. **IANA Considerations**

IANA is requested to allocate a type code for the Standard BGP
Community to be used for inter cluster propagation of angular
position of the clients.

IANA is requested to allocate a new type code from BGP OPEN Optional
Parameter Types registry to be used for Group_ID propagation.

8. **Acknowledgments**

Authors would like to thank Eric Rosen, Clarence Filsfils, Uli
Bornhauser Russ White, Jakob Heitz and Mike Shand for their valuable
input.

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.

[RFC4271]   Rekhter, Y., Li, T., and S. Hares, "A Border Gateway
            Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RFC4360]   Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended
            Communities Attribute", RFC 4360, February 2006.

[RFC5492]   Scudder, J. and R. Chandra, "Capabilities Advertisement
            with BGP-4", RFC 5492, February 2009.

9.2. Informative References

[I-D.ietf-grow-diverse-bgp-path-dist]
            Raszuk, R., Fernando, R., Patel, K., McPherson, D., and K.
            Kumaki, "Distribution of diverse BGP paths.",
            draft-ietf-grow-diverse-bgp-path-dist-03 (work in
            progress), January 2011.

[I-D.ietf-idr-add-paths]
            Walton, D., Retana, A., Chen, E., and J. Scudder,
            "Advertisement of Multiple Paths in BGP",
            draft-ietf-idr-add-paths-04 (work in progress),
            August 2010.

[RFC1997]   Chandrasekeran, R., Traina, P., and T. Li, "BGP
            Communities Attribute", RFC 1997, August 1996.

[RFC1998]   Chen, E. and T. Bates, "An Application of the BGP
            Community Attribute in Multi-home Routing", RFC 1998,
            August 1996.

[RFC4384]   Meyer, D., "BGP Communities for Data Collection", BCP 114,
            RFC 4384, February 2006.

[RFC4456]   Bates, T., Chen, E., and R. Chandra, "BGP Route
            Reflection: An Alternative to Full Mesh Internal BGP
            (IBGP)", RFC 4456, April 2006.

[RFC4893]   Vohra, Q. and E. Chen, "BGP Support for Four-octet AS
            Number Space", RFC 4893, May 2007.

[RFC5283]   Decraene, B., Le Roux, JL., and I. Minei, "LDP Extension

                for Inter-Area Label Switched Paths (LSPs)", RFC 5283,
                July 2008.

   [RFC5668]   Rekhter, Y., Sangli, S., and D. Tappan, "4-Octet AS
                Specific BGP Extended Community", RFC 5668, October 2009.

   [RFC5714]   Shand, M. and S. Bryant, "IP Fast Reroute Framework",
                RFC 5714, January 2010.

Authors' Addresses

   Robert Raszuk
   Cisco Systems
   170 West Tasman Drive
   San Jose, CA 95134
   US

   Email: raszuk@cisco.com

   Christian Cassar
   Cisco Systems
   10 New Square Park
   Bedfont Lakes, FELTHAM TW14 8HA
   UK

   Email: ccassar@cisco.com

   Erik Aman
   TeliaSonera
   Marbackagatan 11
   Farsta,   SE-123 86
   Sweden

   Email: erik.aman@teliasonera.com

   Bruno Decraene
   France Telecom
   38-40 rue du General Leclerc
   Issi Moulineaux cedex 9,   92794
   France

   Email: bruno.decraene@orange-ftgroup.com