BGP remote Next-Hop attribute

draft-vandevelde-idr-remote-next-hop

Gunter Van de Velde
Sr Technical Leader
NOSTG, Cisco Systems

May 2013
BGP remote Next-Hop

- Keep in mind
  
  This technology is currently being worked upon at the IETF
  
  Currently no support yet by any vendor just yet
  
  Work in Progress
  
  Goal of presentation: make people think on the technology potential

- If you see a usage case or feedback around this technology:
  
  Contact
  
  - gunter@cisco.com
  
  - or any of the authors “draft-vandevelde-idr-remote-next-hop”
What is it?

• It is a “IP” network overlay technology

• Distributed transitive BGP based tunnel end-point awareness signalling

• Can be seen as alternative for MPLS tunneling within the Core ISP network resulting for no more need for full routing table on “P” routers (just like with MPLS)

• The overlay mechanism is tunnel technology agnostic (GRE, L2TP, IPinIP, VxLAN, etc)

• Transitive: The network overlay works Intra- and Inter-domain

• Expected convergence time: not faster or slower as traditional IP convergence
Motivation?

- Address family (IPv4, IPv6, VPNv4, VPNv6, IP+Label) agnostic
- Usage of proven and highly scalable Internet technologies (BGP, PIC, LFA, etc…)
- Cost optimization by getting rid of:
  - Core MPLS control plane
  - Internet and customer prefixes from core
  - Other technologies used to build a network overlay
- Usage of BGP technology:
  - Fast Convergence, High scalability, High availability, VPN Support
  - Highly secure by utilisation of BGP security technologies (RPKI Origin Authentication, TCP-AO, etc..)
  - BGP Remote-Next-Hop (http://datatracker.ietf.org/doc/draft-vandevelde-idr-remote-next-hop/)
- Incremental deployment supported
  - Due to the support of transitive distribution, it is possible to dynamic Internet wide overlay infrastructures
  - Existing BGP carries a distributed transitive global database of tunnel end-points
  - Can be deployed ‘RIGHT NOW’ assuming the BGP end-point support BGP rNH attribute
- Wide range of encapsulation protocols supported: VxLAN, GRE, IP-in-IP tunnels, etc…
  - Utilization of scalable and existing tunnel technology
  - Utilization of existing tunnel policy and RIB population mechanisms
  - Service differentiation: enable premium exit vs best-effort exit to Internet by Network Policy
- Backward compatible and support for gradual implementation
Toolset for BGP based Dynamic overlay tunnelling


• Other tunnel technologies: GRE, VxLAN, IP-in-IP, etc…

• BGP Route-Reflection (RFC4456)

• Prefix Independent Convergence

• BGP Diverse Path (RFC6774)


• BGP/MPLS VPN (RFC4364)
Address Distribution

IP Core
2.2.2.0/24
2001:db8::/64

Customer 1a
1.1.1.0/24

Customer 1b
1.1.2.0/24

Customer 1c
1.1.7.0/24

Next-Hop:
2.2.2.6/32
2001:db8::6/128

Next-Hop:
2.2.2.5/32
2001:db8::5/128

Next-Hop:
2.2.2.3/32
2001:db8::3/128

Next-Hop:
2.2.2.7/32
2001:db8::7/128

Next-Hop:
2.2.2.4/32
2001:db8::4/128

Customer 2a
1.1.3.0/24

Customer 2b
1.1.4.0/24

Customer 3a
1.1.5.0/24

Customer 3b
1.1.6.0/24

xTR1

xTR2

xTR3

xTR4

xTR5

xTR6

xTR7

Next-Hop:
2.2.2.2/32
2001:db8::2/128

Next-Hop:
2.2.2.1/32
2001:db8::1/128

Next-Hop:
2.2.2.0/32
2001:db8::/128
Address Distribution

• Core
  IGP: OSPF, EIGRP, ISIS
  MPLS Free Core
  BGP only is run only on the core edge and BGP RR
  support of IGP LFA

• Edge
  Location of the Tunnel in-/egress router
  BGP NLRI is used as remote network identifier and the attached BGP Remote-Next-Hop as Locator
  Forwarding in-/egress policy enforcement
  Multi-tunnel loadsharing

• Customer Networks
  Autonomous networks
    DC, finance, IT department, engineering, customers, etc…
    Independent address family agnostic address space
  Customer networks and services are network identifiers
BGP Remote-Next-Hop Attribute

- NLRI (Network Layer Reachability Information) is the customer network
- Next-hop is the traditional BGP Next-Hop used for traditional IP forwarding
- Remote-Next-Hop is the Tunnel End-Point used for dynamic tunnel based forwarding
- Multiple NLRI can point to identical Remote-Next-Hop

**NLRI (EID)**

**Traditional BGP Attributes**

- Next-Hop
- Remote-Next-Hop #1
- Remote-Next-Hop #2
- Remote-Next-Hop #3

**Supported Tunnel Encapsulation**

- Addresses-used for Tunnel end-point
- Remote customer Network or service

Next-hop is used for traditional IP forwarding
Address Distribution: BGP Table at xTR1

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-Hop</th>
<th>Remote-Next-Hop</th>
<th>Tunnel Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.2.0/24</td>
<td>2.2.2.2</td>
<td>2.2.2.2</td>
<td>GRE</td>
</tr>
<tr>
<td>1.1.3.0/24</td>
<td>2.2.2.3</td>
<td>1.10.3.3</td>
<td>GRE</td>
</tr>
<tr>
<td>1.10.3.3/32</td>
<td>2.2.2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.4.0/24</td>
<td>2.2.2.4</td>
<td>2.2.2.4</td>
<td>VxLAN</td>
</tr>
<tr>
<td>1.1.5.0/24</td>
<td>2.2.2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.7</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.5</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.7.0/24</td>
<td>2.2.2.6</td>
<td>2001:db8::6</td>
<td>GRE</td>
</tr>
</tbody>
</table>
Traditional BGP Forwarding

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-Hop</th>
<th>Remote-Next-Hop</th>
<th>Tunnel Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.2.0/24</td>
<td>2.2.2.2</td>
<td>2.2.2.2</td>
<td>GRE</td>
</tr>
<tr>
<td>1.1.3.0/24</td>
<td>2.2.2.3</td>
<td>1.10.3.3</td>
<td>GRE</td>
</tr>
<tr>
<td>1.1.4.0/24</td>
<td>2.2.2.4</td>
<td>2.2.2.4</td>
<td>VxLAN</td>
</tr>
<tr>
<td>1.1.5.0/24</td>
<td>2.2.2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.7</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.5</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.7.0/24</td>
<td>2.2.2.6</td>
<td>2001:db8::6</td>
<td>GRE</td>
</tr>
</tbody>
</table>
Traditional BGP Forwarding

Customer 1a
1.1.1.0/24

Customer 1b
1.1.2.0/24

Customer 1c
1.1.7.0/24

Customer 2a
1.1.3.0/24

Next-Hop:
2.2.2.1/32
2001:db8::1/128

Customer 2b
1.1.4.0/24

Next-Hop:
2.2.2.4/32
2001:db8::4/128

Customer 3a
1.1.5.0/24

Customer 3b
1.1.6.0/24

Normal IP Packet

Router with Full BGP Table
Tunnel Based Forwarding: Case 1
Reduction of Core state and size of routing table

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-Hop</th>
<th>Remote-Next-Hop</th>
<th>Tunnel Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.2.0/24</td>
<td>2.2.2.2</td>
<td>2.2.2.2</td>
<td>GRE</td>
</tr>
<tr>
<td>1.1.3.0/24</td>
<td>2.2.2.3</td>
<td>1.10.3.3</td>
<td>GRE</td>
</tr>
<tr>
<td>1.1.4.0/24</td>
<td>2.2.2.4</td>
<td>2.2.2.4</td>
<td>VxLAN</td>
</tr>
<tr>
<td>1.1.5.0/24</td>
<td>2.2.2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.7</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.5</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.7.0/24</td>
<td>2.2.2.6</td>
<td>2001:db8::6</td>
<td>GRE</td>
</tr>
</tbody>
</table>

© 2012 Cisco and/or its affiliates. All rights reserved.
Tunnel Based Forwarding: Case 1

Reduction of Core state and size of routing table

Core only needs routing information for tunnel endpoints
**Tunnel Based Forwarding: Case 2**

Reduction of Core state and size of routing table with virtualization

---

**Table:**

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-Hop</th>
<th>Remote-Next-Hop</th>
<th>Tunnel Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.2.0/24</td>
<td>2.2.2.2</td>
<td>2.2.2.2</td>
<td>GRE</td>
</tr>
<tr>
<td>1.1.3.0/24</td>
<td>2.2.2.3</td>
<td>1.10.3.3</td>
<td>GRE</td>
</tr>
<tr>
<td>1.10.3.3/32</td>
<td>2.2.2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.4.0/24</td>
<td>2.2.2.4</td>
<td>2.2.2.4</td>
<td>VxLAN</td>
</tr>
<tr>
<td>1.1.5.0/24</td>
<td>2.2.2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.7</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.5</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.7.0/24</td>
<td>2.2.2.6</td>
<td>2001:db8::6</td>
<td>GRE</td>
</tr>
</tbody>
</table>
Tunnel Based Forwarding: Case 2
Reduction of Core state and size of routing table with virtualization

Core only needs routing information for PE and Customer Tunnel endpoints

Tunnelled IP Packet – IPv4 GRE Header

Next-Hop: 2.2.2.1/32 2001:db8::1/128
Tunnel Based Forwarding: Case 3
Reduction of Core state and size of routing table

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-Hop</th>
<th>Remote-Next-Hop</th>
<th>Tunnel Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.2.0/24</td>
<td>2.2.2.2</td>
<td>2.2.2.2</td>
<td>GRE</td>
</tr>
<tr>
<td>1.1.3.0/24</td>
<td>2.2.2.3</td>
<td>1.10.3.3</td>
<td>GRE</td>
</tr>
<tr>
<td>1.10.3.3/32</td>
<td>2.2.2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.4.0/24</td>
<td>2.2.2.4</td>
<td>2.2.2.4</td>
<td>VxLAN</td>
</tr>
<tr>
<td>1.1.5.0/24</td>
<td>2.2.2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.7</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.5</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.7.0/24</td>
<td>2.2.2.6</td>
<td>2001:db8::6</td>
<td>GRE</td>
</tr>
</tbody>
</table>
Tunnel Based Forwarding: Case 3
Reduction of Core state and size of routing table

Core only needs routing information for tunnel endpoints
Tunnel Based Forwarding: Case 4
IPv4 over IPv6 enabled core

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-Hop</th>
<th>Remote-Next-Hop</th>
<th>Tunnel Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.2.0/24</td>
<td>2.2.2.2</td>
<td>2.2.2.2</td>
<td>GRE</td>
</tr>
<tr>
<td>1.1.3.0/24</td>
<td>2.2.2.3</td>
<td>1.10.3.3</td>
<td>GRE</td>
</tr>
<tr>
<td>1.1.4.0/24</td>
<td>2.2.2.4</td>
<td>2.2.2.4</td>
<td>VxLAN</td>
</tr>
<tr>
<td>1.1.5.0/24</td>
<td>2.2.2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.7</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.5</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.7.0/24</td>
<td>2.2.2.6</td>
<td>2001:db8::6</td>
<td>GRE</td>
</tr>
</tbody>
</table>

© 2012 Cisco and/or its affiliates. All rights reserved.
Tunnel Based Forwarding: Case 4
IPv4 over IPv6 enabled core

Core only needs routing information for IPv6 tunnel endpoints and IPv4 BGP endpoints

Tunneled IP Packet – IP-in-IPv6 Header

Next-Hop: 2.2.2.1/32 2001:db8::1/128

Customer 1a 1.1.1.0/24
Customer 1b 1.1.2.0/24
Customer 1c 1.1.3.0/24
Customer 2a 1.1.4.0/24
Customer 2b 1.1.5.0/24
Customer 3a 1.1.6.0/24
Customer 3b 1.1.7.0/24
Tunnel Based Forwarding: Case 5
IPv4 over IPv6 enabled core

<table>
<thead>
<tr>
<th>Destination</th>
<th>Next-Hop</th>
<th>Remote-Next-Hop</th>
<th>Tunnel Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.2.0/24</td>
<td>2.2.2.2</td>
<td>2.2.2.2</td>
<td>GRE</td>
</tr>
<tr>
<td>1.1.3.0/24</td>
<td>2.2.2.3</td>
<td>1.10.3.3</td>
<td>GRE</td>
</tr>
<tr>
<td>1.1.3.3/32</td>
<td>2.2.2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.4.0/24</td>
<td>2.2.2.4</td>
<td>2.2.2.4</td>
<td>VxLAN</td>
</tr>
<tr>
<td>1.1.5.0/24</td>
<td>2.2.2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.7</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.6.0/24</td>
<td>2.2.2.3</td>
<td>2001:db8::7</td>
<td>IP-in-IPv6</td>
</tr>
<tr>
<td>1.1.7.0/24</td>
<td>2.2.2.6</td>
<td>2001:db8::6</td>
<td>GRE</td>
</tr>
<tr>
<td>1.1.7.0/24</td>
<td>2.2.2.6</td>
<td>2001:db8::6</td>
<td>GRE</td>
</tr>
</tbody>
</table>

IPv4 over IPv6 enabled core
Core only needs routing information for IPv6 tunnel endpoints and IPv4 BGP endpoints.
Conclusion

- BGP based Dynamic Tunnelling is allows a single IP based control base
- High scalability due to proven BGP technology
- Fast Convergence due to proven BGP and IGP tuning technology
- Network core devices enjoy reduction in the size of the BGP table
- BGP based Dynamic Tunnelling allows virtualisation based upon IP technology
- IPv4 and IPv6 agnostic
- Incremental Global implementation is supported
- BGP based Security is supported and scalable
Thank You