Insertion of IPv6 Segment Routing Headers in a Controlled Domain
draft-voyer-6man-extension-header-insertion-04

Abstract

   The network operator and vendor community has clearly indicated that
   IPv6 header insertion is useful and required. This is notably the
   case when the entire journey of the packet remains in its source
   domain. In such a context, it does not matter where the extension
   header is inserted. The source domain may decide to place the IPv6
   extension header insertion where it suits its best: at the source of
   the packet or at any midpoint within the source domain.

Requirements Language

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

Status of This Memo

   This Internet-Draft will expire on December 23, 2018.

Copyright Notice

Table of Contents

1. Introduction

   We define the concept of "domain" as the set of nodes under the same
   administration. For example, a network operator infrastructure
   including routers and links grouped into BGP autonomous systems (ASs)
   and routing domains (running OSFP or IS-IS).

   We define "source domain" as the domain of the source of the packet.

2. **Source Domain and Packet Journey**

```
(-- Source Domain D --)
(                      )
( 1-----2-----3-----9 )
(       |     |       )
(       4-----5       )
(---------------------)
```
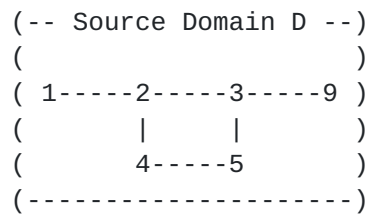
                        Figure 1: Source Domain

   In the previous diagram:

   o All the nodes 1 to 9 are in the same Source Domain D.

   o Node 1 originates a packet P1 destined to 9 (SA=1, DA=9).

   o Domain D runs a link-state routing protocols which implements the
      Fast Reroute (FRR) service through the Topology Independent Loop
      Free Alternates (TI-LFA,
      [I-D.bashandy-rtgwg-segment-routing-ti-lfa]).

   o All link metrics are set to 10.

   o Node 2's TI-LFA pre-computed backup path for the destination 9 is
      the Segment Routing Policy <5, 9> via outgoing interface (OIF) to
      node 4 according to [I-D.filsfils-spring-segment-routing-policy],
      [I-D.filsfils-spring-srv6-network-programming], and
      [I-D.ietf-6man-segment-routing-header]

   Within the 50 milliseconds of link 2-3 failure detection, node 2
   reroutes the traffic destined to 9 by inserting the pre-computed
   segment routing header (SRH) with SID list <5, 9> and forwards the
   packet to node 4. Node 4 forwards based on DA=5 to neighbor 5. Node
   5 updates the DA to 9 and removes the SRH. Node 9 receives the
   packet with (SA=1, DA=9).

   This FRR service is clearly beneficial for the operator of domain D:
   without this FRR operation, depending on the scale of the domain and
   the quality of the routing convergence implementation, traffic could
   be dropped for hundreds to thousands of milliseconds waiting for the
   routing plane to converge.

   This FRR service is largely deployed with MPLS.

   It is important to note that the operators industry is strongly
   requiring the same TI-LFA FRR service without the need to deploy or
   maintain the MPLS layer.

Obviously, this FRR service increases the size of the packet during
its journey within domain D. This is well-known to operators. Well-
known mitigation techniques have been deployed for more than 15 years
for the MPLS-based FRR service and the numerous VPN services. This
is often achieved by deploying a greater MTU value higher in the core
than at the ingress edge.

3. Transit Through a Source Domain

```
     (-- Source Domain D --)
     (                      )
  A---( 1-----2-----3-----9 )---B
     (        |     |       )
     (        4-----5       )
     (--------------------)
```
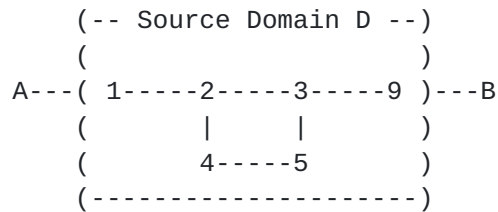
                   Figure 2: Transit Through a Source Domain

Consider a packet sent from A to B called P2 (A,B). A and B are
external nodes to the Source Domain D.

Any packet transiting through source domain D must be unchanged when
it exits domain D.

Therefore, node 1 encapsulates the packet P2 in an outer IPv6 header
with SA=1 and DA=9. Resulting in packet P3 (1,9)(A,B).

From the viewpoint of domain D, packet P3 is the same as packet P1 of
the previous use-case. Indeed, domain D only considers the outer
header when forwarding P3 and the outer header is: (SA=1, DA=9). As
with packet P1, the entire journey of packet P3 is contained within
source domain D.

Node 2 may thus rightfully insert an SRH on packet P3 to implement a
sub-50 milliseconds FRR operation upon the loss of the link 2-to-3
and node 5 can remove this SRH.

The transparency of the service is guaranteed: the insertion and
removal of the SRH on packet P3 has no impact on packet P2. P2 at
the exit of the domain D is the same as at the entrance of the domain
D.

Customers of the transit service offered by source domain D do demand
FRR services. The 50 millisecond FRR operation provides a much
better service availability than 100's to 1000's of milliseconds of
loss for each routing transition.

4. Impact of SRH Insertion Within a Source Domain

   This section discusses the impact of SRH insertion within a source
   domain for traffic transiting the source domain, or traffic generated
   within the source domain.

   Any SRH inserted on a packet within a source domain MUST be removed
   before delivery to destination. This requirement ensures the
   destination node will not receive a packet with an SRH not inserted
   by the source SR Node. Therefore there is no impact of an
   inadvertent SRH being received at a destination node.

   There are however two points of impact associated with ICMP error
   generation back to the source:

      Path MTU discovery [RFC8201] may generate ICMP error messages to
      the packet source.

      Hop Limit may be exhausted and generate ICMP error messages to the
      packet source.

4.1. ICMP Error message processing

   Using the example packet P1 from Section 3. If Hop Limit decrements
   to 0 or a Packet Too Big (PTB) error is generated at node 4, after
   the SRH is inserted, the destination address in P1 is 5.

   This results in an ICMP error message generated to node 1, as per
   [RFC4443] but with an unfamiliar destination address.

4.1.1. ICMP Error message processing with routing header

   During parsing of the ICMP error message at node 1, the invoking
   packet's protocol receives the error. In the case of UDP and TCP,
   the invoking packet four-tuple (source address, source port,
   destination address, destination port) identifies a UDP or TCP
   session.

   Since the original destination (node 9) is not the current
   destination of the invoking packet, the lookup cannot succeed in
   current implementations, and the error is not delivered to the source
   UDP or TCP session.

   This is common for any use of routing headers regardless of whether a
   routing header is inserted at source or by an intermediate node.

4.2. Destination outside the Source Domain

Since the SRH inserted within an intermediate node MUST be removed
when all segments within the SRH have been visited, it is not
possible to leak the SRH to the destination outside the source
domain.

5. Security Considerations

This document proposes to insert an SRH to a transit packet if such
packet is originated and destined within a controlled/trusted domain.

Insertion of SRH is safe when confined within a source domain.

In such conditions, the security of the original packet is not
compromised by header insertion. The packet reaches the destination
or leaves the source domain without any inserted header.

A source domain can operate SRv6-based services for internal traffic
while preventing any external traffic from accessing these internal
SRv6-based services. Several mechanisms exists and are currently
used today, for example:

o Access-lists (ACL) on the each externally facing interface in
   order to drop any incoming traffic with SA or DA belonging to the
   internal SID space.

o ACL to prevent access to local SIDs from outside the operator's
   infrastructure.

o Support Unicast-RPF on source address on external interface.

6. IANA Considerations

This document doesn't introduce any IANA request.

7. Manageability Considerations

TBD

8. Contributors

The authors would like to thank the following for their
contributions: Stefano Salsano, Antonio Cianfrani, David Lebrun,
Olivier Bonaventure, Prem Jonnalagadda, Milad Sharif, Hani Elmalky,
Ahmed Abdelsalam, Robert Raszuk, Arthi Ayyangar, Dirk Steinberg, Wim
Henderickx.

9. Acknowledgements

   TBD

10. References

10.1. Normative References

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

   [RFC4443] Conta, A., Deering, S., and M. Gupta, Ed., "Internet
              Control Message Protocol (ICMPv6) for the Internet
              Protocol Version 6 (IPv6) Specification", STD 89,
              RFC 4443, DOI 10.17487/RFC4443, March 2006,
              <https://www.rfc-editor.org/info/rfc4443>.

   [RFC8201] McCann, J., Deering, S., Mogul, J., and R. Hinden, Ed.,
              "Path MTU Discovery for IP version 6", STD 87, RFC 8201,
              DOI 10.17487/RFC8201, July 2017,
              <https://www.rfc-editor.org/info/rfc8201>.

10.2. Informative References

   [I-D.bashandy-rtgwg-segment-routing-ti-lfa]
              Bashandy, A., Filsfils, C., Decraene, B., Litkowski, S.,
              Francois, P., and d. daniel.voyer@bell.ca, "Topology
              Independent Fast Reroute using Segment Routing", draft-
              bashandy-rtgwg-segment-routing-ti-lfa-04 (work in
              progress), April 2018.

   [I-D.filsfils-spring-segment-routing-policy]
              Filsfils, C., Sivabalan, S., Hegde, S.,
              daniel.voyer@bell.ca, d., Lin, S., bogdanov@google.com,
              b., Krol, P., Horneffer, M., Steinberg, D., Decraene, B.,
              Litkowski, S., Mattes, P., Ali, Z., Talaulikar, K., Liste,
              J., Clad, F., and K. Raza, "Segment Routing Policy
              Architecture", draft-filsfils-spring-segment-routing-
              policy-06 (work in progress), May 2018.

   [I-D.filsfils-spring-srv6-network-programming]
             Filsfils, C., Li, Z., Leddy, J., daniel.voyer@bell.ca, d.,
             daniel.bernier@bell.ca, d., Steinberg, D., Raszuk, R.,
             Matsushima, S., Lebrun, D., Decraene, B., Peirens, B.,
             Salsano, S., Naik, G., Elmalky, H., Jonnalagadda, P., and
             M. Sharif, "SRv6 Network Programming", draft-filsfils-
             spring-srv6-network-programming-04 (work in progress),
             March 2018.

   [I-D.ietf-6man-segment-routing-header]
             Previdi, S., Filsfils, C., Leddy, J., Matsushima, S., and
             d. daniel.voyer@bell.ca, "IPv6 Segment Routing Header
             (SRH)", draft-ietf-6man-segment-routing-header-13 (work in
             progress), May 2018.

   [RFC4798] De Clercq, J., Ooms, D., Prevost, S., and F. Le Faucheur,
             "Connecting IPv6 Islands over IPv4 MPLS Using IPv6
             Provider Edge Routers (6PE)", RFC 4798,
             DOI 10.17487/RFC4798, February 2007,
             <https://www.rfc-editor.org/info/rfc4798>.

Authors' Addresses

   Daniel Voyer (editor)
   Bell Canada

   Email: daniel.voyer@bell.ca

   John Leddy
   Comcast
   4100 East Dry Creek Road
   Centennial, CO 80122
   US

   Email: John_Leddy@comcast.com

   Clarence Filsfils
   Cisco Systems, Inc.
   Brussels
   BE

   Email: cfilsfil@cisco.com

   Darren Dukes (editor)
   Cisco Systems, Inc.
   Ottawa
   Canada

   Email: ddukes@cisco.com

   Stefano Previdi
   Individual Contributor
   Italy

   Email: stefano@previdi.net

   Satoru Matsushima
   Softbank

   Email: satoru.matsushima@g.softbank.co.jp