draft-dunbar-idr-sdwan-port-safi-05

Linda Dunbar

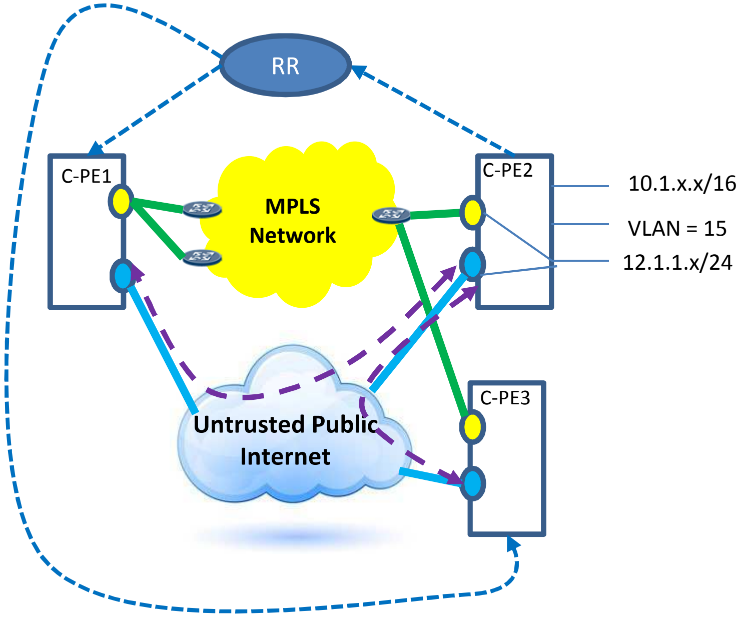
Sue Hares

IETF 106 Nov 2019

Intention of the draft

- Informational draft
- Purpose:
  - We applied a Port-SAFI in the first come first serve (FCFS) category (SAFI =74)
  - Intended to inform how the Port-SAFI is used for SDWAN overlay network
- We would like to hear your feedback.

Port Based IPSec Tunnel
Confederated IPsec via RR



Regular MPLS BGP Routes Update

BGP UPDATE Messages from C-PE2 to announce all the routes attached

- MP-NLRI Path Attribute
  - Nexthop (C-PE2)
  - NLRI
    - 10.1.x.x.
    - VLAN 15
    - 12.1.1x

BGP UPDATE Messages from C-PE2 to RR for WAN port properties:

- MP-NLRI Path Attribute:
    - Port Identifier encoding
- Tunnel-Encap Path Attribute:
    - NAT for the WAN Port
    - IPsec SA-12 (C-PE1->C-PE2 via the specific port)
    - IPsec SA-32 (C-PE3->C-PE2 via the specific port)

New NLRI for the WAN Port

N subTLVs in the Tunnel Encap Path Attribute

Attributes for End Point Identity

```
+------------------+
|   NLRI Length    | 1 octet
+------------------+
|  Network-Type    | 2 Octets
+------------------+
|Port-Distinguisher| 4 octets
+------------------+
|  SDWAN-Site-ID   | 4 octets
+------------------+
|  SDWAN-Node-ID   | 4 or 16 octets
+------------------+
```

SDWAN
Can have different TYPE

Locally significant
within the node

routable address across WAN

- NLRI Length: expressed in bits as defined in [RFC4760].
- Network-Type: SDWAN
- Port Distinguisher: Locally significant Port identifier.
- SDWAN-Site-ID: Globally unique site identifier.
- SDWAN Node ID: Locally significant node identifier (system ID or the loopback address (IPv4 or IPv6)).

Advantage of new NLRI: to represent different address space than client routes: SDWAN WAN port; similarr approach as the new NLRI used for SR Policy
Disadvantage of new NLRI: intermediate Routers can drop the UPDATE due to not recognizing the new NLRI.
Not applicable to SDWAN overlay, as the UPDATE to RR is simple IP forwarding, not terminated by any routers/switches in between

SubTLV for the NAT Property of the WAN Port

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|EncapExt Type  |  EncapExt subTLV Length       |I|O|R|R|R|R|R|R|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| NAT Type      |  Encap-Type   |Trans networkID|     RD ID     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Local  IP Address                            |
           32-bits for IPv4, 128-bits for Ipv6
                   ~~~~~~~~~~~~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Local  Port                                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Public IP                                      |
           32-bits for IPv4, 128-bits for Ipv6
                   ~~~~~~~~~~~~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Public Port                                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Flags:

-I bit (CPE port address or Inner address scheme)

  If =0 → inner addr is IPv4.
  If =1 → inner address is IPv6.

-O bit (Outer address scheme):

 If =0 → the public (outer) address is IPv4.
 If =1 → the public (outer) address is IPv6.

-R bits: reserved for future use. Must be set to 0 now.

NAT Type: without NAT; 1:1 static NAT; Full Cone; Restricted Cone; Port Restricted Cone; Symmetric; or Unknown (i.e. no response from the STUN server).

Encap Type： the supported encap types for the port facing public network, such as IPsec+GRE, IPsec+VxLAN, IPsec without GRE, GRE (when packets don't need encryption)
Transport Network ID: Central Controller assign a global unique ID to each transport network；
RD ID: Routing Domain ID， Need to be global unique.
Local IP: The local (or private) IP address of the port；
Local Port: used by Remote SDWAN node for establishing IPsec to this specific port.
Public IP: The IP address after the NAT. If NAT is not used, this field is set to NULL.
Public Port: The Port after the NAT. If NAT is not used, this field is set to NULL.

SubTLV for the Port Based IPsec

```
The IPsecSA sub-TLV is for the SDWAN edge node to establish IPsec security
association with their peers via the port that face untrusted network:
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |IPsec-SA Type  |IPsecSA Length                 | Flag          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |  Transform    | Transport    | AH             |   ESP         |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |         SPI (rekey counter                                    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | key1 length   |   DH Public Key                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | key2 length   |       Nonce                                   |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | key3 length   |         key3                                  |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |        Duration                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

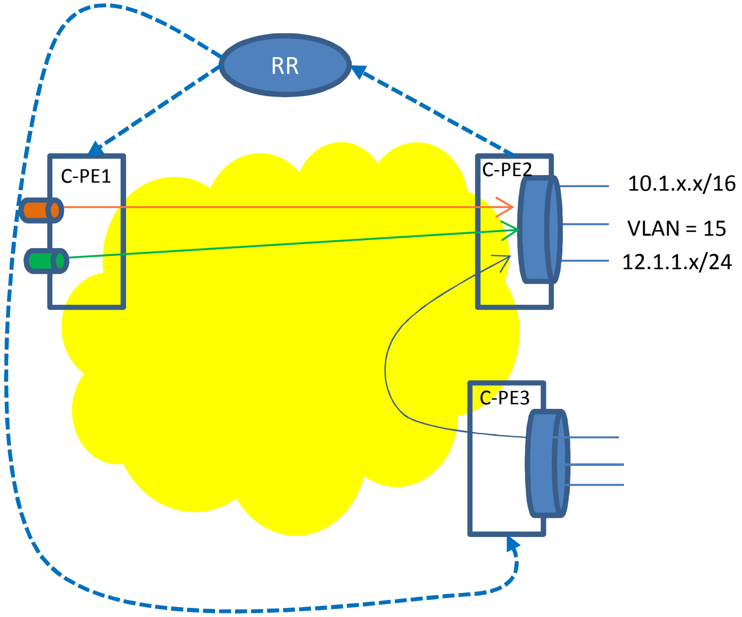
Device Information Message (DIM) are derived from draft-carrel-ipsecme-controller-ike-01

Next Step

- Call for WG adoption
- Why?
  - Demonstrate how BGP is used in Port Based IPsec to scale SDWAN overlay

BACKUP SLIDES

Recap of BESS' presentation on BGP for Homogeneous SDWAN



One BGP UPDATE Message from C-PE2 to RR:

- multiple routes encoded in the MP-NLRI Path Attribute
  - 10.1.x.x/16
  - VLAN #15
  - 12.1.1.x/24
- IPsec attributes are encoded in the Tunnel-Encap Path Attribute
  - IPsec attributes for all possible remote nodes, or
  - IPsec attributes for specific remote nodes, or
  - IPsec attributes for specific remote subnets

  ….

WHY BGP

- here are some of the Compelling reasons of using BGP to distribute SDWAN edge properties among peers that might be spread across the globe:
- (note: the BGP for SDWAN Edges is running at different layers than the BGP for underlay networks, i.e. not “FLAT” BGP. They are among SDWAN edges, not for exposing to underlay provider as you stated EBGP. When the underlay network service providers use SDWAN to temporarily expand bandwidth in some segments, they have more reason to use BGP to minimize amount of learning & configuration of introducing new protocols in their environment)

- – BGP already widely deployed as sole protocol (see RFC 7938). Even if not for this purpose of propagating SDWAN WAN port properties, the BGP base protocol implementation is supported by virtually all switches/routers (virtual & physical). Even AWS VPC export the BGP routes.
- – Wide acceptance – minimal learning (which is very important requirement for operations)
- – Robust and simple implementation,
- – Reliable transport
- – Guaranteed in-order delivery
- – Incremental updates
- – No flooding and selective filtering
- – RR already has the capability to apply policies to communications among peers.
- Bottom line: It is much easier to add one function than adding a brand-new protocol stack.

- Alternative: extending LISP, NHRP, DSVPN/DMVPN
- – In addition to more proposal changes needed, NHRP/DSVPN/DMVPN don’t scale well.
- – More learning, more barrier to be deployed, just think how many decades of painful journey deploying IPv6.
- –
- Prior extension of BGP for non-client routes reachability: Flowspec, BGP LS, Segment routing policies, etc