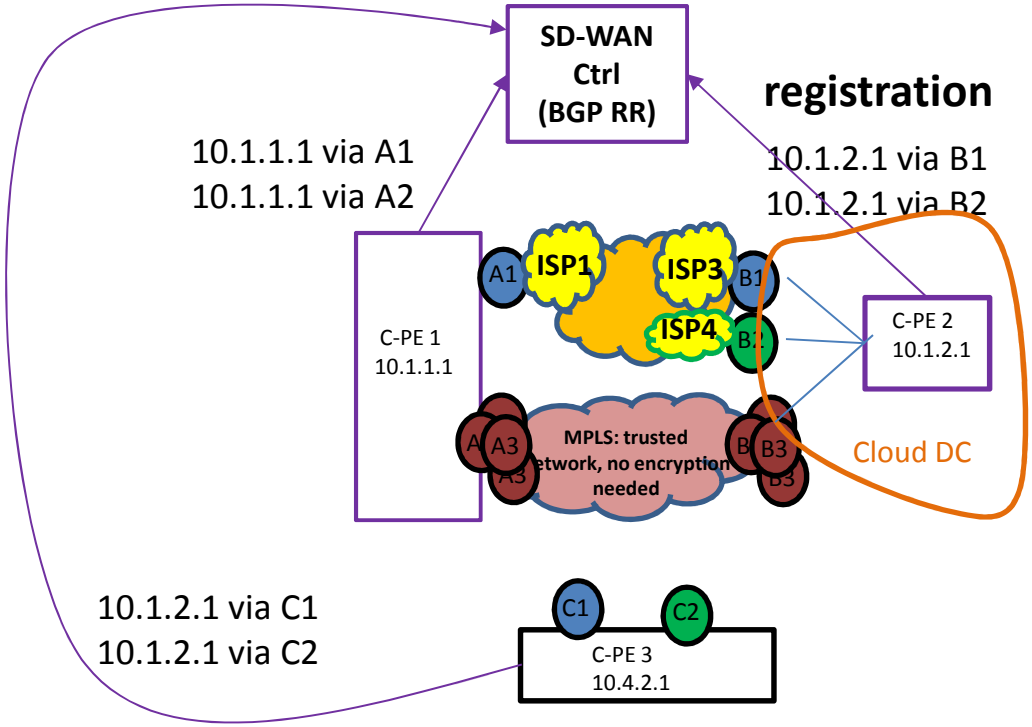
Analysis of various BGP UPDATE encoding options to achieve SDWAN WAN port properties propagation among SDWAN edges

Linda Dunbar

Aug 12, 2019

Goal: WAN Ports Property Propagation across SDWAN nodes in different domains

SD-WAN node's private address and WAN Ports/Addresses registration.



The constraints:

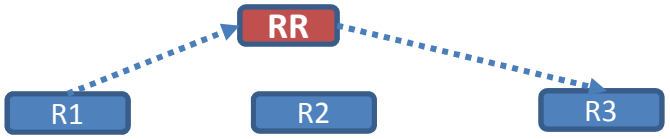
- C-PE1 & C-PE2 can spread across different geographical locations,
- their connection to RR can be over untrusted networks, and they might not know the reachable addresses for the peers they need to communicate (therefore needing RR to propagate)
- WAN ports can be from different network providers (A1/A2/A3/B1/B2/B3)
- Each PE advertise its WAN ports to Controller, which then propagate the advertisements to authorized peers.
- **PEs Loopback addresses & routes attached are not visible to some ISPs**

A1/A2/A3/B1/B2/B3 are logical address that can be applied to a set of ports

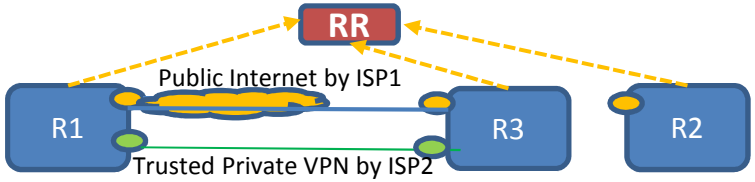
Security Sequence SDWAN Control Plane

Controller- Device Exchange

- Device (Router R1) sends routes to Controller (BGP RR) [Tunnel-Encap]
- Controller (RR) advertise tunnels for clients routes via Secure link (R1)
- The WAN ports properties are registered to Controller, and controller propagates.
- SDWAN edges pairwise secure channel requires Controller managed re-key schedule and distribution
- IPsec SA management independent from the attached client routes. i.e. IPsec parameters negotiation, public key exchange, and re-key schemes detached from client routes.



Controller manages the authorized recipients



All SDWAN nodes advertise their WAN port properties to Controller

Sequence similar to I2NSF Controller

There are many ways to skin the cat…

different encoding for BGP Update Messages

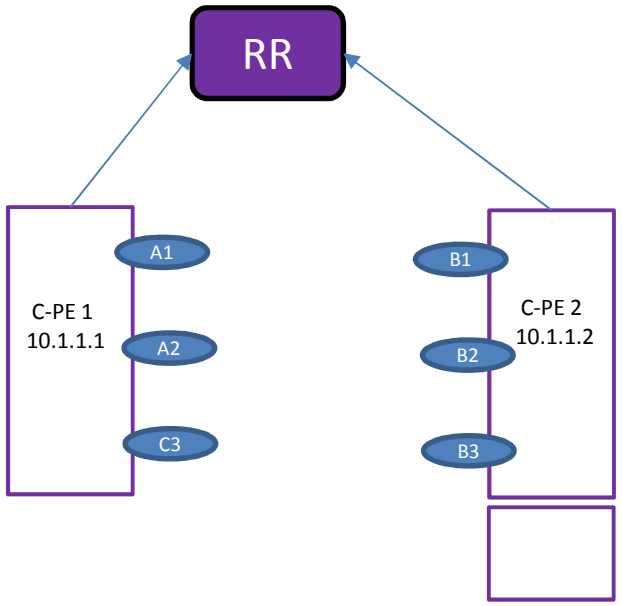
Option 1: Extending Tunnel-Encap with SAFI as IP to achieve WAN port registration?

C-PE1 needs to send out 3 separate Tunnel Update messages

Tunnel Update 1
- Tunnel-Type Public IPsec
- NLRI: 10.1.1.2
  1) subTLV: Remote endpoint: A1;
  2) SubTLV: Ipsec properties

Tunnel Update 2
- Tunnel-Type: Private network (no encap)
- NLRI: 10.1.1.2
  1) subTLV: Remote endpoint: A2;
  2) subTLV: properties associated with the private network;

Tunnel Update 3
- Tunnel-Type: Private Addr (no encap)
- NLRI: 10.1.1.2
  1) subTLV: Remote endpoint: A3;
  2) subTLV: properties associated with NAT;



C-PE2 needs to send out 3 separate Tunnel Update messages

Tunnel Update 1
- Tunnel-Type Public +IPsec
- NLRI: 10.1.1.2
  1) subTLV: Remote endpoint: B1;
  2) SubTLV: Ipsec properties

Tunnel Update 2
- Tunnel-Type: Private network (no encap)
- NLRI: 10.1.1.2
  1) subTLV: Remote endpoint: B2;
  2) subTLV: properties associated with the private network;

Tunnel Update 3
- Tunnel-Type: Private Addr (no encap)
- NLRI: 10.1.1.2
  1) subTLV: Remote endpoint: B3;
  2) subTLV: properties associated with NAT;

Need a new subTLV to carry the Site ID

Pros & Cons of the Option 1

- **Pros:**
  - no new SAFI introduced, the update messages can traverse existing routers
- **Cons:**
  - Same IPv4/IPv6 SAFI NLRI carries the WAN port information that is very different from clients' routes attached to the C-PEs.
  - The receivers (RR) has to do extra processing to differentiate the UPDATE messages from the attached routes UPDATE messages.

Option 2: Tunnel-Encap with SDWAN NLRI for SDWAN WAN Ports Prosperities & Policies

draft-dunbar-idr-sdwan-port-safi-03

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

SDWAN in cloud can have different TYPE

Can be identifiers to Port # or Cloud GW

The new SAFI=74 has been assigned by IANA for advertising properties of WAN ports that face untrusted networks

Similar to SR Policy NLRI/SAFI to specify specific polices

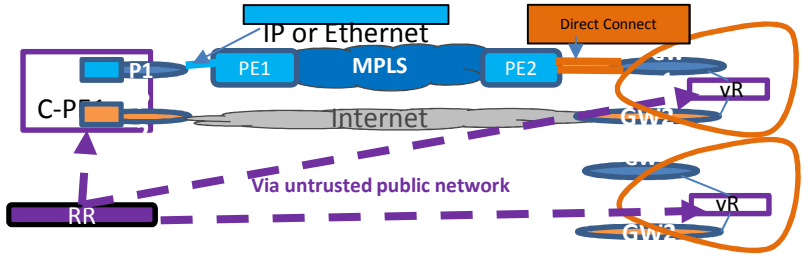
```
+-------------------+
|  NLRI Length      |
+-------------------+
|  Distinguisher    |
+-------------------+
|  Policy Color     |
+-------------------+
|  Endpoint         |
+-------------------+
```

- SDWAN NLRI Length: expressed in bits as defined in [RFC4760].
- SDWAN-Type: to define the encoding of the rest of the SDWAN NLRI.
- Port Distinguisher: SDWAN node Port identifier.
- SDWAN-Site-ID: used to identify a common property shared by a set of SDWAN nodes.
- SDWAN Node ID: the SDWAN node identifier, (e.g. system ID or the loopback address (IPv4 or IPv6)).

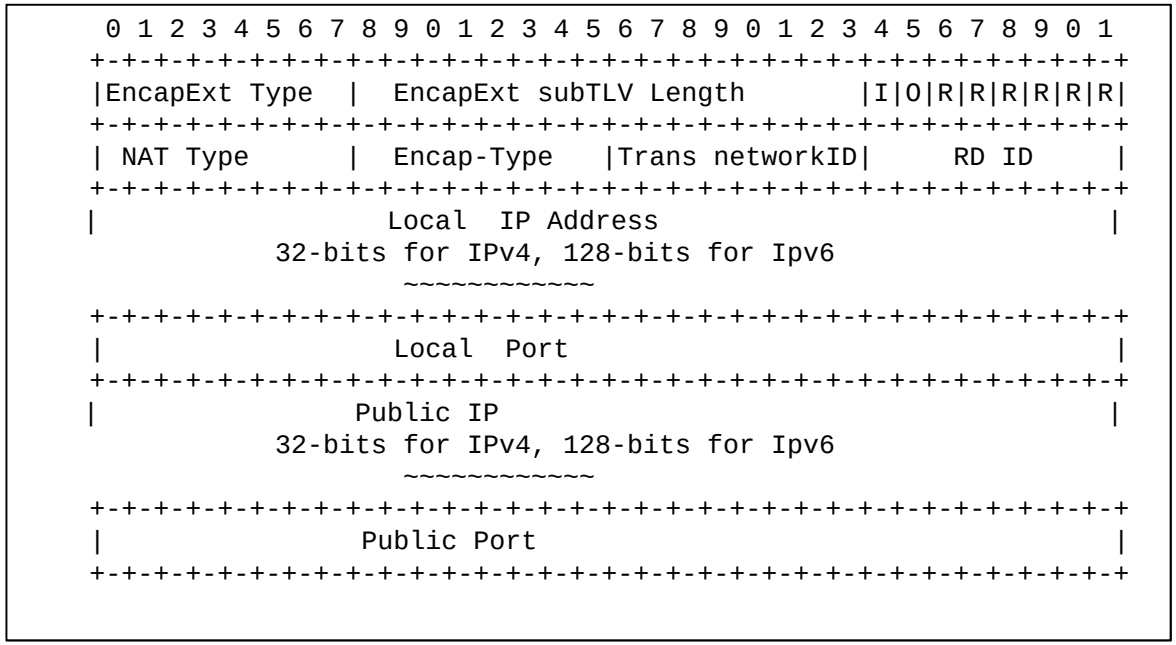
Advantage of new NLRI: to represent different address space: SDWAN WAN port, which can be port number & not IP address

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



Extended Port Property (Option 2 Continue)



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.

Pros & Cons of the Option 2

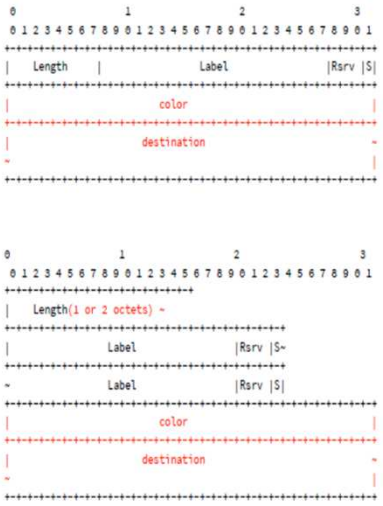
- **Pros:**
  - **Clean design and processing on the receivers (RRs). Simpler** processing to differentiate the UPDATE messages from the attached routes UPDATE messages.
- **Cons:**
  - New NLRI is introduced, the update messages can't traverse existing routers
    - Since the the Tunnel UPDATE message with the new SDWAN NLRI/SAFI is strictly between SDWAN edge nodes and their respective RR(s) via a secure tunnel, the SDWAN UPDATE messages are not going to traverse existing routers. Therefore, it doesn't cause any issues.

Option 3:
Using the new SAFI introduced for BGP labeled Colored Unicast

draft-szarecki-idr-bgp-lcu-traffic-steering

- The BGP-LCU SAFI is for carry the traffic property across network domains,
  - similar to carry the WAN port properties across different domains/locations
- Foundation:
  - Traffic treatment encoded as 32b integer – COLOR
    - Agreed among all domains
  - Intra-domain tunnels marked w/ COLOR is satisfies desired treatment

LCU NLRI



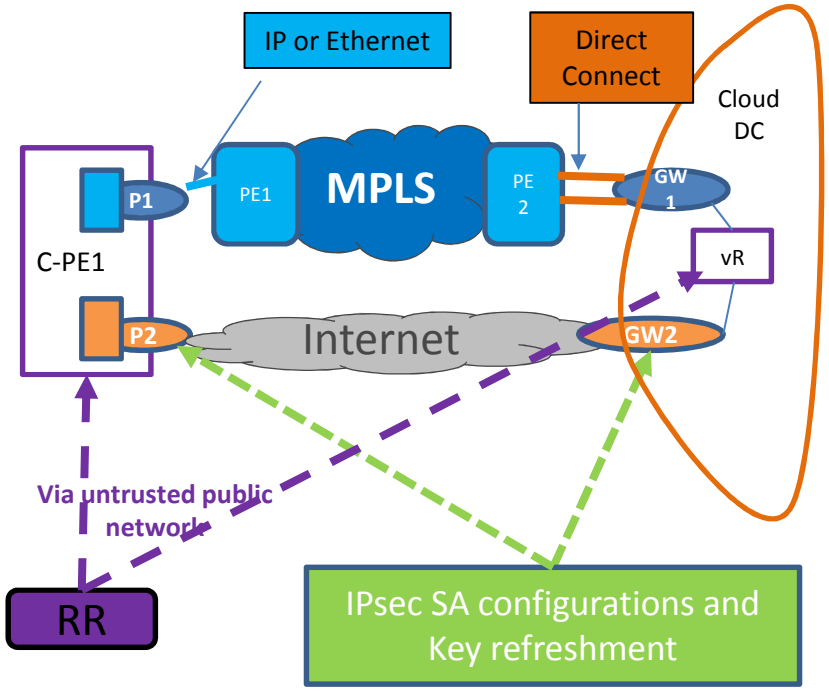
- Follows RFC8277
  - Prefix := <COLOR,DESTINATION>
    - COLOR:= integer (32b)
    - DESTINATION:= IPv4/IPv6 subnet address
  - Length 1 or 2 octets (511B)
    - COLOR, IPv6/128 DESTINATION w/ a lot of labels (160+)

Pros & Cons of the Option 3

- **Pros:**
  - leverage the newly proposed NLRI for carrying Traffic Color across domains
  - Similar goal as SDWAN needing to propagating WAN port properties across domain/geolocations
- **Cons:**
  - Need to attach the attributes which haven't been specified by the draft yet.
  - Need to ask merging the content with draft-dunbar-idr-sdwan-port-safi-03.

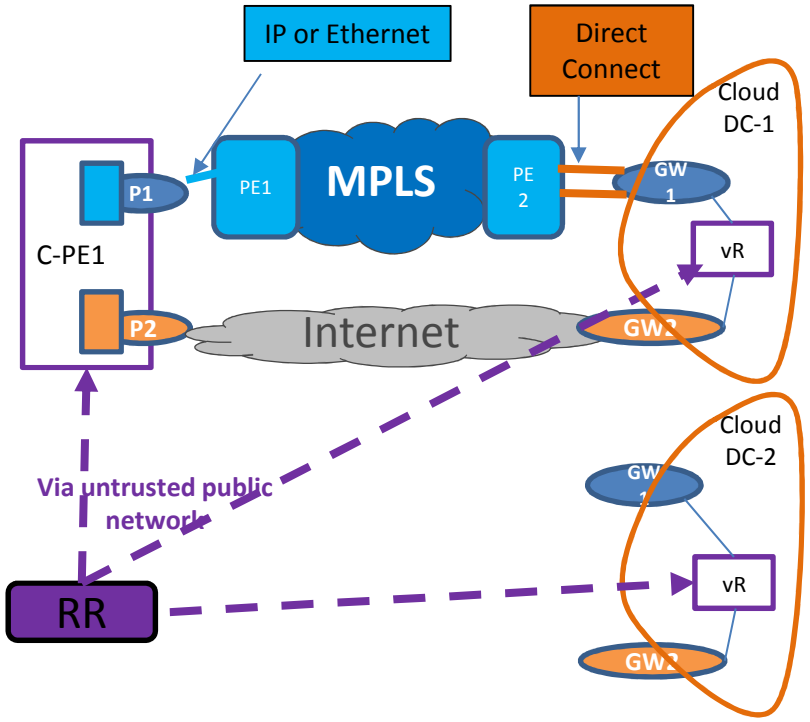
BACKUP SLIDES

Key Characteristics of Hybrid SDWAN: Multi-players



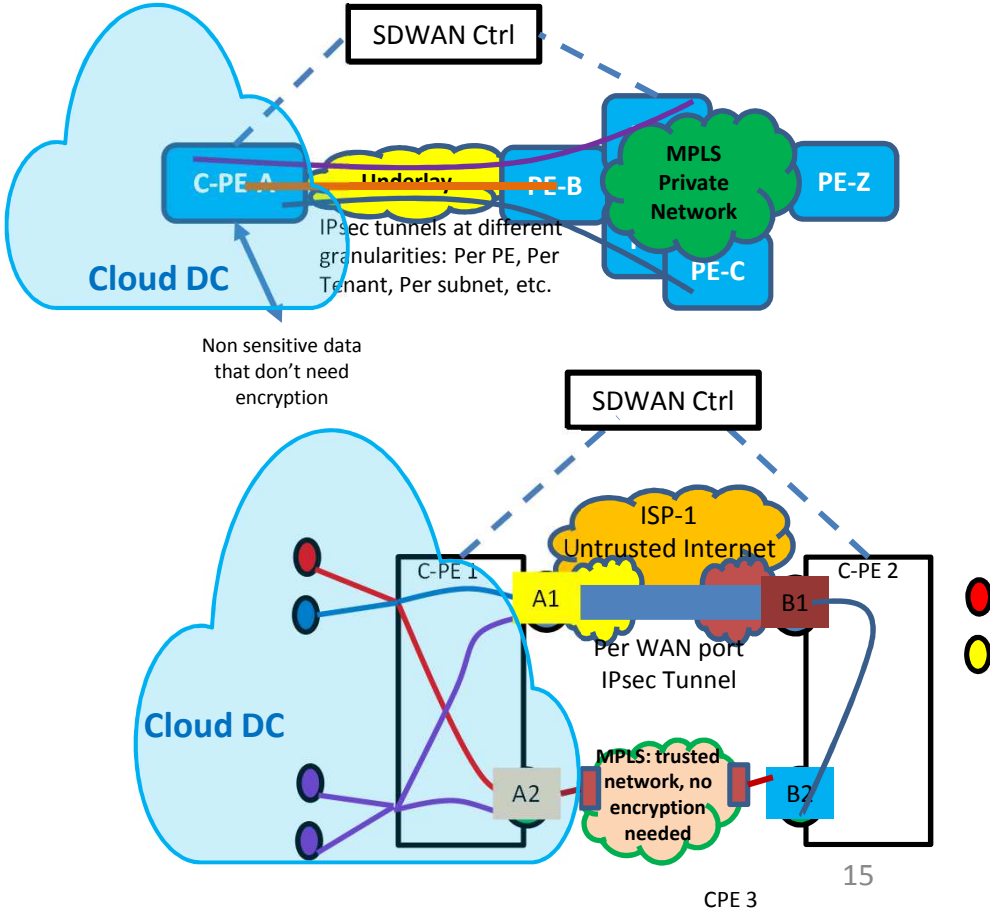
- Multiple parties: Enterprises, VPN providers and Cloud Providers
- Multiple paths with different performance & authentication parameters.
- the connections between BGP RR & CPEs are over public network, therefore requires TLS, DTLS, or IPsec.

Application components can be available from multiple Cloud DCs.



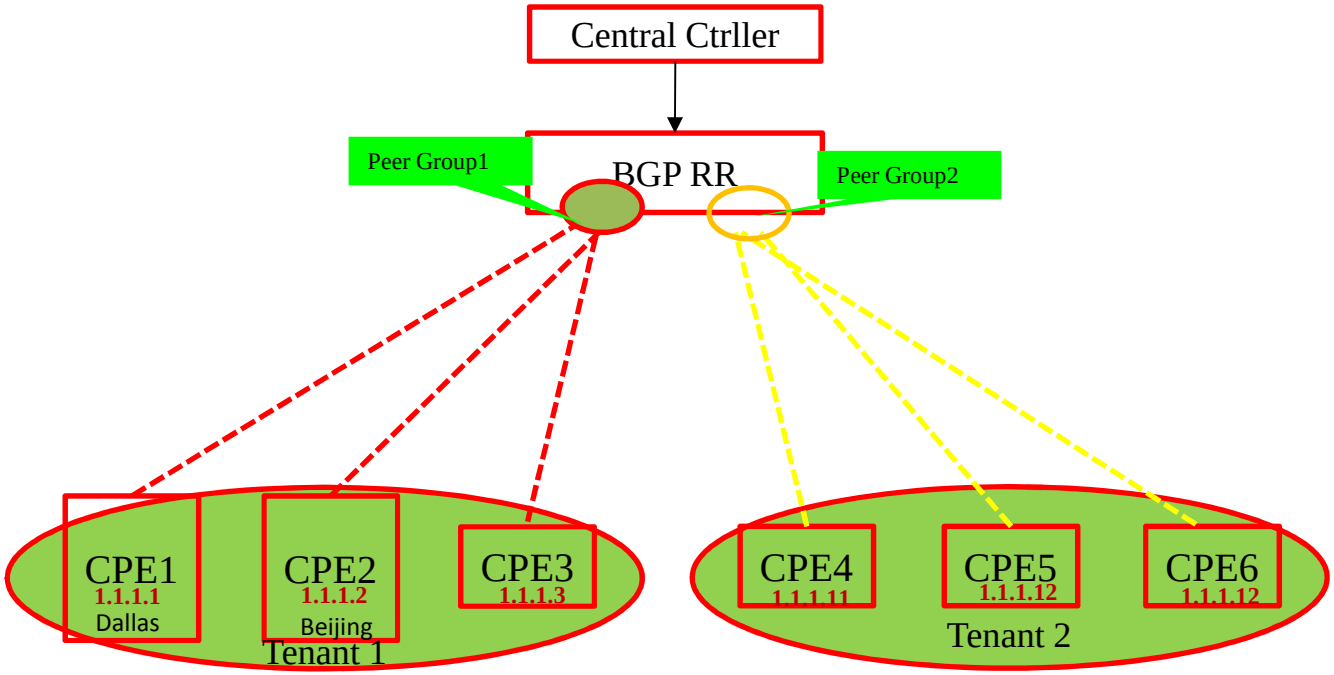
SD-WAN Scenarios

- **Homogeneous SD-WAN:**
  - edge encrypting all sensitive traffic to other edge nodes, regardless if the underlay is private or public.
- **SD-WAN over Hybrid Networks**
  - Traffic over Private VPN networks (e.g. MPLS) can go natively without encryption to achieve better performance, and
  - traffic over internet are carried by IPsec tunnels.
  - User specified policy governs a flow
- Already in practice in Networks



SDWAN Ports scoped Advertisement

- Tenant Separation Method：



Why BGP as Control Plane for SDWAN WAN Ports Registration

- Compelling reasons of using BGP:
  - BGP already widely deployed as sole protocol (see RFC 7938)
  - Robust and simple implementation
  - Wide acceptance – minimal learning
  - 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.

- Alternative: NHRP, DSVPN/DMVPN
  - In addition to more proposal changes needed, it doesn't scale well
- Prior extension of BGP for non client routes reachability:
  - Flowspec, BGP LS, Segment routing policies, etc