CPE Based VPN + SD-WAN

https://datatracker.ietf.org/doc/draft-dm-net2cloud-gap-analysis/

Linda.Dunbar@Huawei.com
Andy Mails (agmalis@gmail.com)
Christianjacquenet@orange.com
Mehmet.toy@verizon.com
Use Case 1: Classic SD-WAN
CPE based VPN: Integrating SR Routes & Internet Routes

L2 or L3 VPNs over IP WAN

Tunnels via Internet

Tunnels via SR domain

SD-WAN Controller (RR)
Use Cases 2: SD-WAN end points are far apart, Different apps need different paths

For communication between “A1” <-> “Z1”:
Optimal path: “A1” <-> E1 <-> C1<->C4 <-> “E2”<->Z1 (at Site-3)

Problems:
- It is very difficult, if even possible, for PEs to determine which egress PEs is optimal for flows between “E1” <-> “E2” (as multiple PEs can reach E2 via SD-WAN paths).
- Steer the SD-WAN path over the Enterprise VPN as much as possible for better quality & control (cost, traffic management, delay, etc)

SD-WAN paths over public internet can have unpredictable performance, especially over long distances and cross state/country boundaries.
Gap: RFC5512 & draft-ietf-idr-tunnel-encaps-09

• Tunnel-Encap removed SAFI =7 for distributing encapsulation tunnel information.
  □ Tunnels are associated with routes. The SD-WAN paths need to be established before data arrival. There is no Routes to associate with the Tunnel
  □ Using a “Fake Route” for tunnel end point: create deployment complexity.
    ➢ Each CPE has many tunnels to all their peers: therefore need many “fake address”.
    ➢ 10000 CPEs will need 10’s thousands of “fake address” → difficult to manage

• The BGP Route Update doesn’t have enough fields to carry detailed information of the remote CPEs: such as
  □ Site-ID, System-ID, Port-ID
  □ IPsec configuration information sent by the “Controller (RR)” to the CPEs.
  □ for two peer CPEs to negotiate IPsec keys, based on the configuration sent from the Controller.
  □ UDP NAT private address <-> public address mapping
  □ CPEs tend to communicate with a few other CPEs, not all the CPEs need to form mesh connections. Using BGP, CPEs can easily get dumped with too much information of other CPEs that they never need to communicate with.
    ➢ NHRP only sends the relevant information for the interested end points for establishing tunnels. Therefore, there is a need for some form of “Registration” methods.
Gap of draft-rosen-bess-secure-l3vpn-01

- [Secure-l3vpn] is for limited number of remote C-PEs, whereas SD-WAN overlay deployment may have over 1000s of nodes
- [Secure-l3vpn] needs heavy configuration, whereas SD-WAN needs Zero touch provisioning, i.e. auto synchronize Ipsec config,
- For RR communication with CPE, this draft only mentioned IPSEC. Needs lightweight secure connection for Zero touch provisioning.
- Multiple WAN port per CPE, need way to distinguish different Tunnels from one CPE. Just Red & Black is not enough
- The draft assumes that C-PE “register” with the RR. But it doesn’t say how
- IPsec requires periodic refreshment of the keys. How to synchronize the refreshment among multiple nodes?
- IPsec usually only send configuration parameters to two end points and let the two end points to negotiate the KEY. Large scale SD-WAN needs Controller to authenticate Peers.
To stimulate the discussions.....

PROTOCOLS WORK FOR IETF?
Managed Overlay WAN Services: 100’s or 1000’s CPEs

**Goals:**
- **Zero Touch Provisioning for CPEs**
  - Upon powered up, CPE sends request to its factory default Central Controller address to retrieve its local RR address.
- **Large scale number of CPEs overlay existing networks**
  - Allow controller to change routes to traverse specific sites (instead of through specific CPEs)
  - E.g. instead of Site 2 -> Site 1, use Site2 -> Site 3 -> Site 1 (for performance, cost, or temporary detour).
  - To simplify complex full mesh of too many CPEs

**How:** Hierarchical management
- Partition CPEs into Sites, each Site is a logical entity for remote sites
- Why:
  - Enable detour based on sites, instead of CPEs,
  - Avoid complexity of managing full mesh of all CPEs.
  - Hide CPE identity from others (some deployment needs this feature)
More details:

- Each CPE registers with LC (RR) to establish secure connection for BGP over TLS/DTLS
- CPE WAN Ports information are advertised via LC to all other CPEs
- Each CPEs need to inform its LC on its targeted Sites to establish Tunnels
- CPE receives Tunnel information for the targeted CPEs to establish tunnel
  - CPE Tunnel Key: LinkId、SystemIP、SiteId
- Route Advertisement will carry CPE Tunnel Key, Tunnels are aggregated per Tunnel Key
- Vrf is represented by using VN ID to replace MPLS Label
Desired BGP Extension

- **New SD-WAN Path SAFI and NLRI**
  - Need a new SAFI to be defined to represent SA-WAN paths (IANA). The SD-WAN SAFI use a new NLRI defined here
- **SD-WAN Path Attributes**
  - IPsec Attributes:
    - configuration parameters from Controller to CPEs
    - IPsec Key exchange between CPEs
  - NAT Private address <-> Public address mapping for remote CPEs
- **SD-WAN Remote Endpoint: Site ID (+optional CPE ID and Port ID)**
- **SD-WAN Path Policy Sub-TLVs**

```plaintext
+---------------------+   
| NLRI Length        | 1 octet   
|---------------------|
| Tunnel Type         | 2 octets  
|---------------------|
| Path Distinguisher  | 4 octets  
|---------------------|
| Side ID (Color)     | 4 octets  
|---------------------|
| Endpoint            | 4 or 16 octets  

where:
New Tunnel Type: SD-WAN

The Remote Endpoint and Color sub-TLVs, as defined in [I-D.ietf-idr-tunnel-encaps] are used to represent Site ID and CPE System ID.

Path Distinguisher is to differentiate multiple SD-WAN paths terminated at one CPE
```
### BGP extension to distribute IPsec attributes

**Detailed Tunnel information advertisement**

01234567890123456789012345678901
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
<table>
<thead>
<tr>
<th>NAT Type</th>
<th>Encap Type</th>
<th>Trans network ID</th>
<th>RD ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private IP Address</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private Port</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public IP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Port</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NAT Type Include:**
- without NAT
- 1:1 Static NAT
- Full Cone
- Restricted Cone
- Port Restricted Cone
- Symmetric

**Encap Type include:**
- SD-WAN IPSEC,GRE

**Transport Network ID:** Global Unique value for each CPE

**RD ID:** Routing Domain ID. global unique

**Private IP:** WAN private IP;
**Private Port:** The destination Port number for Remote CPE to establish IPsec to the local CPE;
**Public IP:** The IP after NAT
**Public Port:** The Port after NAT

<table>
<thead>
<tr>
<th>01234567890123456789012345678901</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transform</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>SPI</td>
</tr>
<tr>
<td>key1 length</td>
</tr>
<tr>
<td>key2 length</td>
</tr>
<tr>
<td>key3 length</td>
</tr>
<tr>
<td>Duration</td>
</tr>
</tbody>
</table>

**Transform:** 1 byte
**AH | ESP | AH+ESP**
**Transport mode | transport mode**

**AH:** 1 byte
- Authentication algorithm:
  - md5 | sha1 | sha2-256 | sha2-384 | sha2-512 | sm3
  - Local can have multiple Authentication methods. Local & Remote CPE negotiate for the strongest one.

**ESP:** 2 bytes
- Authentication method (High 4bits):
  - md5 | sha1 | sha2-256 | sha2-384 | sha2-512 | sm3
  - Local can have multiple Authentication methods. Local & Remote CPE negotiate for the strongest one.
- Default, ESP uses SHA2-256 method

**ESP Encryption (High 4bits):**
- 3des | des | aes-128 | aes-192 | aes-256 | sm1 | sm4
- Default, ESP uses AES-256 encryption

**Reserve:** 1 byte
**SPI:** 4 bytes
**Key1:** AH authentication key
**Key2:** ESP authentication key
**Key3:** ESP encryption key
**Duration:** SA lifespan
What is the difference between Transport network ID and RD ID?
Linda Dunbar, 6/22/2018
### Tunnel Encapsulation

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sd-wan</td>
<td>Used when Tunnel is established</td>
</tr>
</tbody>
</table>

```plaintext
+------------------+
| 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 |
| +---------------------------------------------------------------+
| | Tunnel Type (2 Octets) | Length (2 Octets) |
| +------------------+
| 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 |
| +---------------------------------------------------------------+
| | Value |
```
Tunnel Information Advertisement Method

- Tenant Separation Method:

  - For Tenant Separation: CPEs belonging to same Tenant are added to a Peer Group peer group1 route-policy tenant1-in import peer group1 route-policy tenant1-out export route-policy tenant1-in permit node 10 apply community 100:1 additive route-policy tenant1-out permit node 10 if-match community-filter 1 ip community-filter 1 permit 100:1 Others are configured in similar way

  - CPE1:
    - Receiving SD-WAN IPSEC information, report WAN ports information to Controller via SD-WAN SAFI
    - RR send to CPE2, CPE3 (using Policy Filtering to only send to Peers belong to same tenant)
    - CPE2, CPE3 upon receiving SD-WAN IPSEC config information, start to negotiate with peers on IPsec tunnel establishing and establish the key.
    - SD-WAN IPSEC tunnel is added to the Service Tunnel (IDR-Tunnel-encap) to be used for Route Advertisement
Thank you

www.huawei.com