BGP UPDATE for SDWAN Edge Discovery
draft-dunbar-idr-sdwan-edge-discovery-00

Abstract

   The document describes encoding of BGP UPDATE messages for the SDWAN
   edge node discovery.

   In the context of this document, BGP Route Reflectors (RR) is the
   component of the SDWAN Controller that receives the BGP UPDATE from
   SDWAN edges and in turns propagates the information to the intended
   peers that are authorized to communicate via the SDWAN overlay
   network.

Status of this Memo

The list of Internet-Draft Shadow Directories can be accessed at
http://www.ietf.org/shadow.html

This Internet-Draft will expire on Dec 2, 2020.

Copyright Notice

Table of Contents

1. Introduction

[SDWAN-BGP-USAGE] illustrates how BGP is used as control plane for a
SDWAN network. SDWAN network is an overlay network with some special
properties.

The document describes a BGP UPDATE for SDWAN edge nodes to announce
its properties to its RR which then propagates to the authorized
peers.

2. Conventions used in this document

Cloud DC: Off-Premise Data Centers that usually host applications
             and workload owned by different organizations or
             tenants.

Controller: Used interchangeably with SDWAN controller to manage
             SDWAN overlay path creation/deletion and monitor the
             path conditions between sites.

CPE-Based VPN: Virtual Private Secure network formed among CPEs.
             This is to differentiate from most commonly used PE-
             based VPNs a la RFC 4364.

MP-NLRI: The MP_REACH_NLRI Path Attribute defined in RFC4760.

SDWAN End-point: can be the SDWAN edge node address, a WAN port
                  address (logical or physical) of a SDWAN edge node, or a
                  client port address.

OnPrem: On Premises data centers and branch offices

SDWAN: Software Defined Wide Area Network. In this document,
             "SDWAN" refers to the solutions of pooling WAN bandwidth
             from multiple underlay networks to get better WAN
             bandwidth management, visibility.control and application
             ID based forwarding for some applications.

3. Framework of SDWAN Edge Discovery

3.1. The Objectives of SDWAN Edge Discovery

The objectives of SDWAN edge discovery is for a SDWAN edge node to
discover its authorized peers to which its attached clients traffic
need to communicate. The attributes to be propagated includes the
SDWAN instances supported, the attached routes under specific SDWAN
instances, and the properties of the underlay networks over which
the client routes can be carried.

Some SDWAN peers are connected by both trusted VPNs and untrusted
public networks. Some SDWAN peers are connected only by untrusted
public networks. For the portion over untrusted networks, IPsec
secure tunnels have to be established. If an edge node has network
ports behind the NAT, the NAT properties needs to be discovered by
authorized SDWAN peers.

Just like any VPN networks, the attached client's routes belonging
to specific SDWAN instances have to be segmented and only exchanged
to the SDWAN peer nodes that are authorized to communicate.

3.2. Basic Schemes

There are two types of BGP UPDATE for SDWAN Edge Discovery:

  - UPDATE for the attached client routes:
     This is the traditional BGP UPDATEs, i.e. for a SDWAN node to
     advertise the attached client routes to remote peers. This
     UPDATE will continue using all the existing BGP attributes,

such as using the existing AFI/SAFI for IP or VPN. A new IPsec
Tunnel Type needs to be added to the Tunnel-Encap Path
Attribute [Tunnel-encap]. See Section 5 for the detailed
encoding.

- UPDATEs for underlay network properties:
  This UPDATE is for an edge node to advertise the properties of
  directly attached underlay networks, including the underlay
  network ISP information, NAT information, the supported IPsec
  keys, nonce, encryption algorithms, etc.

  This UPDATE is for peers to discover remote node's properties
  to establish IPsec tunnels and to traverse NAT. Each
  established IPsec tunnel has a unique identifier which can be
  referenced by the Tunnel-Encap Path Attribute to indicate the
  routes in the NLRI can be carried by the tunnel.

In the following figure: there are four types tunnels between C-PE1
and C-PE2:

a) MPLS-in-GRE path;

b) node-based IPsec tunnel [2.2.2.2<->1.1.1.1].

c) port-based IPsec tunnel [192.0.0.1 <-> 192.10.0.10]; and

d) port-based IPsec tunnel [172.0.0.1 <-> 160.0.0.1];

The regular BGP Route UPDATE messages indicate which tunnels the
client routes can be carried in the Tunnel-Encap Path Attribute. For
example, in the figure below, the client routes 10.1.1.1 and
20.1.1.1 can be carried by IPSec SA terminated at 192.0.0.1, IPSec
SA terminated at 170.0.0.1, or MPLS-in-GRE tunnel. Some routes can
be carried by the IPsec SA terminated at the node's Loopback
address.

```
                         +---+
             +--------------|RR |----------+
            /  Untrusted     +-+-+          \
           /                                 \
          /                                   \
      +---------+  MPLS Path               +-------+
 11.1.1.x| C-PE1   A1-----------------------------B1 C-PE2 |10.1.1.x
      |         |                           |       |
 21.1.1.x|        A2(192.10.0.10)------( 192.0.0.1)B2       |20.1.1.x
      |         |                           |       |
      | Addr    A3(160.0.0.1) --------(170.0.0.1)B3 Addr   |
      | 1.1.1.1 |                          |2.2.2.2|
      +---------+                          +-------+
```
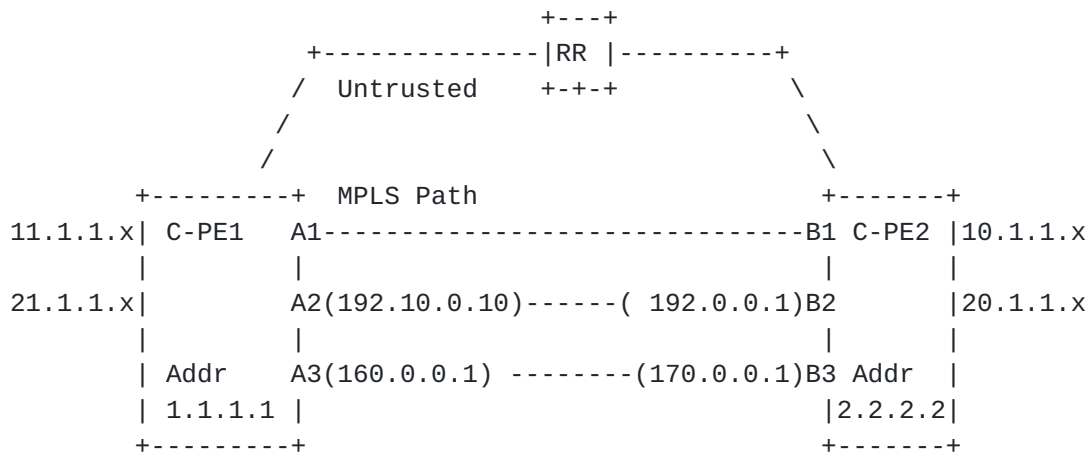
                    Figure 1: Hybrid SDWAN

   BGP UPDATE from C-PE2 for the attached client routes is like:

        Extended community: RT for SDWAN Instance 1
        Prefix: 10.1.1.x; 20.1.1.x
        NH: C-PE 2
        Tunnel Encap Attr [AFI/SAFI = 1/1)
          MPLS-in-GRE [tunnel type=11]
            Sub-TLV for MPLS-in-GRE [Section 3.2.6 of Tunnel-encap]
          IPSec SA for 192.0.0.1 [tunnel type=4]
            Tunnel-End-Point Sub-TLV [Section 3.1 of Tunnel-encap]
            IPsec SA sub-TLV [See the Section 5]
        Tunnel Encap Attr [AFI/SAFI = 1/1)
          IPSec SA for 170.0.0.1
            Tunnel-End-Point Sub-TLV /*the address*/
            IPsec SA sub-TLV

   Note: [Tunnel-Encap] Section 11 specifies that each Tunnel Encap
   Attribute can only have one Tunnel-End-Point sub-TLV. Therefore, two
   separate Tunnel Encap Attributes are needed to indicate that the
   prefix can be carried by either one.

   BGP UPDATE from C-PE1 for the attached client routes is like:

        Prefix 11.1.1.x
        NH: C-PE 1

```
      Tunnel Encap Attr [AFI/SAFI = 1/1)
       MPLS-in-GRE /*11.1.1.x can only be reached by MPLS network*/
           Sub-TLV for MPLS-in-GRE [Section 3.2.6 of Tunnel-encap]


      Extended community: RT for SDWAN Instance 1, 2
      Prefix 21.1.1.x
      NH: C-PE 1
      Tunnel Encap Attr (AFI/SAFI = 1/1)
       MPLS-in-GRE
           Sub-TLV for MPLS-in-GRE
       IPSec SA for 192.10.0.10
           Tunnel-End-Point Sub-TLV
           IPsec SA sub-TLV
      Tunnel Encap Attr (AFI/SAFI = 1/1)
       IPSec SA for 160.0.0.1
           Tunnel-End-Point Sub-TLV
           IPsec SA sub-TLV
```

3.3. Edge Node Discovery using BGP

   The basic scheme of SDWAN Edge node discovery using BGP consists of:

   - Upon powering up, a SDWAN edge node establish a secure tunnel
     (such as TLS, SSL) with the SDWAN central controller whose
     address is preconfigured on the edge node. The central
     controller will inform the edge node of its local RR. The edge
     node will establish a transport layer secure session with the
     RR (such as TLS, SSL).

   - The Edge node will advertise its own properties to its
     designated RR via the secure transport layer tunnel. This is
     different from traditional BGP, where each node sends its
     properties (BGP UPDATE) to its neighbors, which in turn
     propagate to all the nodes in the network.

   - The RR propagates the received information to the authorized
     peers.

       - The authorized peers can establish the secure data channels
          (IPsec) and exchange more information among each other.
   For a SDWAN deployment with multiple RRs, it is assumed that there
   are secure connections among those RRs. How secure connections being
   established among those RRs is the out of the scope of the current
   draft. The existing BGP UPDATE propagation mechanisms control the
   edge properties propagation among the RRs.

   For some special environment where the RR and communication to RR
   are highly secured, [SDN-IPsec] IKE-less can be deployed to simplify
   IPsec tunnel establishment among edge nodes.

4. BGP UPDATE to Support SDWAN Segmentation

   One SDWAN network can be segmented to multiple instances. Each SDWAN
   edge node may need to support multiple SDWAN instances. One client's
   traffic may need to be mapped to different SDWAN segmentations based
   on client's policy. Therefore, we need encoding to differentiate
   SDWAN instances. For example, in the picture below, the
   "PaymentFlow" (payment applications) can only be propagated to
   "Payment GW". Other flows can be propagated to all other nodes. This
   is very similar to VPNs. But need to differentiate from traditional
   MPLS VPNs because a SDWAN edge may also support traditional MPLS
   VPNs.
   [Note: SDWAN Instance ID is configured the same way as VRF, or EVI
   as in EVPN. For node with both MPLS and IPsec ports, the label for
   MPLS can be used for SDWAN Instance ID]

```
                              +---+
                  +--------------|RR |----------+
                 /  Untrusted    +-+-+           \
                /                                 \
               /                                   \
         +---------+  MPLS Path                  +-------+
   11.1.1.x| C-PE1   A1-------------------------------B1 C-PE2 |10.1.1.x
         |         |                              |       |
   21.1.1.x|       A2(192.10.0.10)------( 192.0.0.1)B2       |20.1.1.x
         |         |                              |       |
         | Addr    A3(160.0.0.1) --------(170.0.0.1)B3 Addr  |11.2.2.x
         | 1.1.1.1 |                         /  |2.2.2.2|
         +---------+                        /    +-------+
                                           /
                                          /PaymentFlow
                                         /
                              +----+----+
                              | payment |
                              | Gateway |
                              +---------+
```

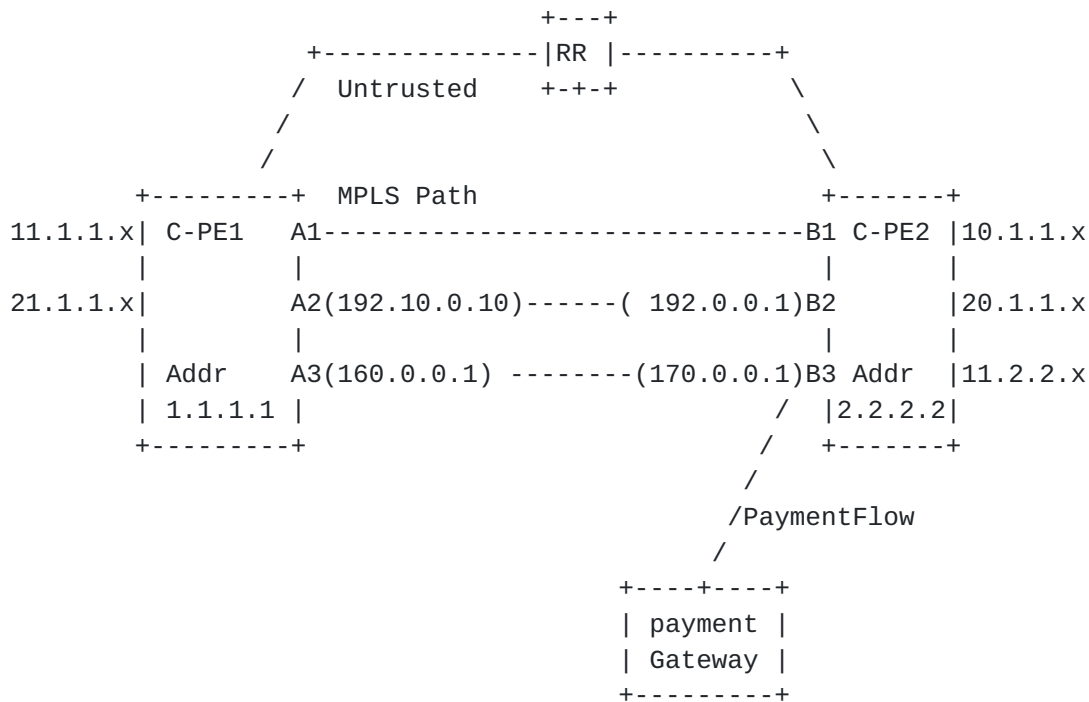   Figure 2: SDWAN Segmentation

4.1. Constrained Propagation of Edge Capability

   BGP has built-in mechanism to dynamically achieve the constrained
   distribution of edge information. RFC4684 describes the BGP RT
   constrained distribution. In a nutshell, a SDWAN edge sends RT
   Constraint (RTC) NLRI to the RR for the RR to install an outbound
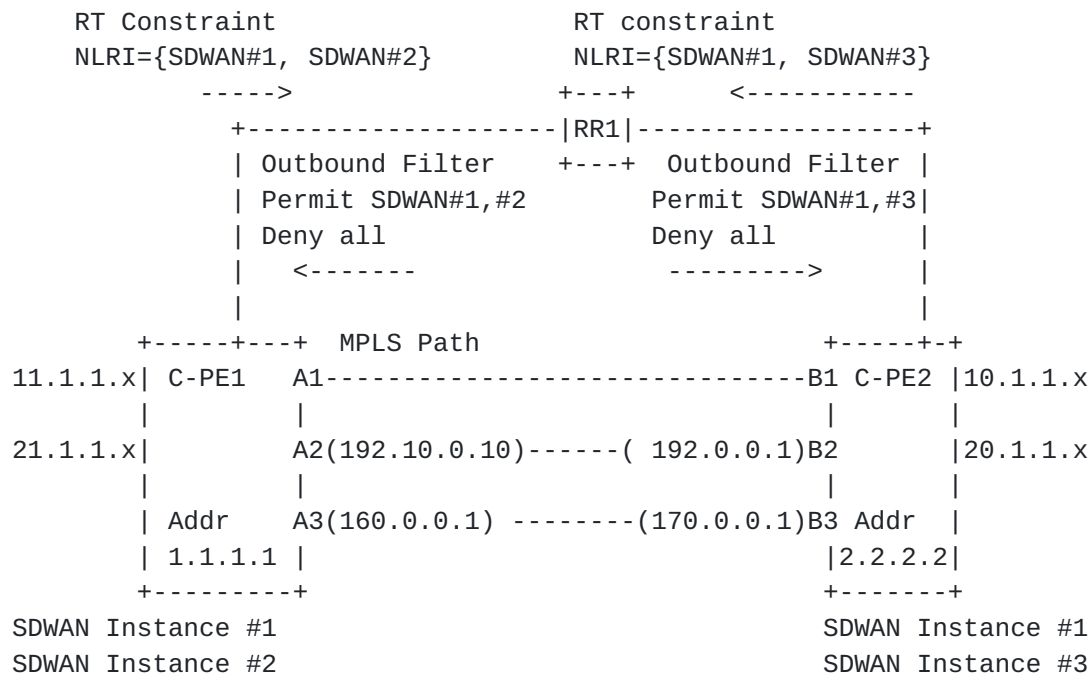   route filter, as shown in the figure below:

```
        RT Constraint                      RT constraint
        NLRI={SDWAN#1, SDWAN#2}            NLRI={SDWAN#1, SDWAN#3}
               ----->               +---+      <-----------
               +--------------------|RR1|------------------+
               | Outbound Filter   +---+  Outbound Filter |
               | Permit SDWAN#1,#2       Permit SDWAN#1,#3|
               | Deny all                Deny all         |
               |    <-------             --------->       |
               |                                          |
         +-----+---+   MPLS Path                   +-----+-+
    11.1.1.x| C-PE1   A1-------------------------------B1 C-PE2 |10.1.1.x
         |         |                             |       |
    21.1.1.x|         A2(192.10.0.10)------( 192.0.0.1)B2       |20.1.1.x
         |         |                             |       |
         | Addr    A3(160.0.0.1) --------(170.0.0.1)B3 Addr  |
         | 1.1.1.1 |                             |2.2.2.2|
         +---------+                             +-------+
    SDWAN Instance #1                            SDWAN Instance #1
    SDWAN Instance #2                            SDWAN Instance #3
         Figure 3: Constraint propagation of Edge Property
```

However, as SDWAN overlay network span across untrusted networks,
RR can't trust the RT Constraint (RTC) NLRI BGP UPDATE from any
nodes. Polices must be configured on RR to filter out unauthorized
nodes to be registered as interested in certain SDWAN instances.
RR can only process the RTC NLRI from authorized peers for a SDWAN
instance.

It is out of the scope of this document on how RR is configured
with the policies to filter out unauthorized nodes for specific
SDWAN instances. The policy configuration could be by manual
commands or by management systems.

When the RR receives edge node property BGP UPDATE from an edge
node, it propagates the received UPDATE message to the nodes that
are in the Outbound Route filter for the specific SDWAN instance.

4.2. SDWAN Segmentation for Control Plane

SDWAN Instances is represented by the SDWAN Target ID in the BGP
Extended Community.

Same as Route Target for VPN, a different Type is used to
differentiate SDWAN Instances from MPLS VPN instances. This is
especially useful when a CPE supports both MPLS VPN and SDWAN
Segmentation (instances).

Encoding:

 RFC4360: Extended Community for SDWAN Route Target
   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
  +-+
  | Type high | Type low(*) | |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+   Value |
  | |
  +-+

  0 1 2 3 4 5 6 7
  +-+-+-+-+-+-+-+-+
  |I|T| 6-bit val |
  +-+-+-+-+-+-+-+-+

The high-order octet of the Type Field
T bit =0 (transitive) when SDWAN edge sends to its RR which then
propagates to remote peers based on outbound filters.

RFC4760 states that Route Target community is transitive
For SDWAN, an edge receiving the SDWAN Update shouldn't forward it
to other nodes.
T bit =1 (non-transitive) when RR propagates the UPDATE to SDWAN
EDGE

[IANA Consideration:
   Following the encoding scheme specified by RFC7153, need IANA to
   assign the following values for the "Type High" Octet:
   - Transitive (when edge announce the advertisement to its RR):
     Ox0A, which is the number after 0x08 for Flow Spec Redirect.
   - Non Transitive (when RR send to remote edges): Ox4A

      Request a new value of the low- order octet of the Type field for
      this community (different from the VPN Route Target 0x02)?
   ]

4.3. SDWAN Instance Identifier in Data Plane

   From data plane perspective, packets from different SDWAN network
   instances (or segmentations) need to have their corresponding SDWAN
   instance identifier encoded in the header.

   For a SDWAN edge node which can be reached by both MPLS and IPsec
   path, the client packets reached by MPLS network will be encoded
   with the MPLS Labels based on the scheme specified by RFC8277.

   For GRE Encapsulation within IPsec tunnel, the GRE key field can be
   used to carry the SDWAN Instance ID. For NVO (VxLAN, GENEVE, etc.)
   encapsulation within the IPsec tunnel, Virtual Network Identifier
   (VNI) field is used to carry the SDWAN Instance ID.

   [Note: the SDWAN Instance ID is same as EVI in EVPN, or VNI if VxLAN
   is used].

5. IPsec Tunnel Type for client's routes in Tunnel-Encap

5.1. Encoding

   This document introduces the following extension to the tunnel
   encapsulation attribute specified in [Tunnel-Encaps]:

      . Support for the following IPsec tunnel types: IPsec in Tunnel
        [value 4] using two new SUB-TLVs: IPsec-SA-ID, and IPsec-SA-
        Group.

   Editor's note: The IPSEC-SA-Group was designed to provide better
   scaling for multiple SA terminated at one endpoint. One end point
   can have multiple SAs, one SA can encrypt client data to or from
   CPE1 and another one for CPE2.

IPsec-SA-ID Sub-TLV

```
 0                   1
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     IPsec SA Identifier       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The IPsec SA identifier (2 Octet) is for cross reference the IPsec
SA attributes being advertised by the Underlay Network Properties
Advertisement UPDATE [Section 6].

If the client traffic needs to be encapsulated in a specific type
within the IPsec ESP Tunnel, such as GRE or VxLAN, etc., the
corresponding Tunnel-Encap Sub-TLV needs to be appended right after
the IPsec-SA-ID Sub-TLV.

Editor Note: 4 octets can be considered as well for IPsec-SA-ID.

IPsec-SA-Group Sub-TLV:

```
 0                   1                   2                   3
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        InnerEncapType         |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  IPsec SA Identifier #1       |   IPsec SA Identifier #2      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                               |IPsec SA Identifier #n         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                IPsec-SA-Group Sub-TLV
```

IPsec-SA-Group Sub-TLV is for the scenario that the client prefix
can be carried by multiple IPsec SAs with the same inner
encapsulation. Multiple IPsec SA IDs are included in the IPsec-ID-
Group Sub-TLV. If different inner encapsulation is desired within
IPsec tunnels, then multiple IPsec-SA-Group Sub-TLVs can be included
within one Tunnel Encap Path Attribute.

InnerEncapType (2 octet) indicates the encapsulation type for the
payload within the IPsec ESP Tunnel. The Inner Encap Type value will
take the value specified by the IANA Consideration Section (12.5) of
[Tunnel-Encap]:

   - types 8 (VXLAN), 9 (NVGRE), 11 (MPLS-in-GRE), and 12 (VXLAN-
     GPE) in the "BGP Tunnel Encapsulation Tunnel Types" registry.
   - types 1 (L2TPv3), 2 (GRE), and 7 (IP in IP) in the "BGP Tunnel
     Encapsulation Tunnel Types" registry.

   For each of the Tunnel Types specified, the detailed encapsulation
   value field as specified by Section 3.2 of [Tunnel-Encap] is
   appended right after the IPsec Sub-TLV.

   Since IPsec SA has a lot of attributes, such as public keys, nonce,
   encryption algorithms etc., the IPsec Tunnel Identifier [ID] is used
   instead of listing all those attributes in the client routes update.
   Using IPsec Tunnel ID greatly reduces the size of BGP client UPDATE
   messages, especially when the client routes can be carried by
   multiple IPsec tunnels. Another added benefit of using IPsec Tunnel
   ID is that the client routes can be advertised independently from
   the IPsec SA rekeying process.

   The Tunnel Ending Point Sub-TLV specified by the Section 3.1 of
   [Tunnel-Encap] has to be attached to identify the IPsec Tunnel
   terminating address.
   There can be multiple IPsec tunnels terminating at one WAN port or
   at one node, e.g. one tunnel for going to destination "A", another
   one for going to destination "B". Use SDWAN for retail industry as
   an example, it is necessary for all shops at any location to only
   exchange Payment System traffic with the Payment Gateway, while
   other traffic can be exchanged with any nodes.
   Therefore, there could be multiple IPsec Sub-TLVs bound with one
   Tunnel Ending Point Sub-TLV.

   However, it is quite common in SDWAN deployment that all IPsec
   attributes from one node or one port are the same for all
   destinations. In that case, IPsec SA ID is set to 0 and the
   terminating address can be used to cross reference the IPsec SA
   attributes which are advertised by the Underlay Network Property
   advertisement UPDATE.

5.2. Encoding Example

5.2.1. Multiple IPsec SAs Sharing One Tunnel End Point

   The encoding example is for the following scenario:

      - there are three IPsec SAs terminating at the same WAN Port
         address (or the same node address)
      - Two of the IPsec SAs use GRE (value =2) as Inner Encapsulation
         within the IPsec Tunnel
      - One of the IPsec SA uses VxLAN (value = 8) as the Inner
         Encapsulation within its IPsec Tunnel.

   Here is the encoding for the scenario:

```
    0                   1                   2                   3
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Tunnel-Type =4 (IPsec)        |       Length =                |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |               Tunnel-end-Point Sub-TLV                        |
   ~                                                              ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |IPsec-SA-group:InnerEncapType=2|       Length = 4             |
   +------------------------------+------------------------------+
   |   IPsec SA Identifier = 1     |    IPsec SA Identifier = 2    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    GRE-KEY (4 Octets)                         |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |IPsec-SA-ID:  =3               |         Reserved             |
   +------------------------------+------------------------------+
   ~                       VxLAN Sub-TLV                          ~
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   The Length of the Tunnel-Type = 4 (IPsec) is the sum of the
   following:
   - Tunnel-end-point sub-TLV total length
   - the IPsec-SA-Group Sub-TLV length + 4 (the two octets for
      InnerEncapType + the two octets for the Length field)
   - GRE-Key Length (4)
   - The IPsec-SA-ID Sub-TLV length: 4
   -          The VxLAN sub-TLV total length

5.2.2. Multiple IPsec SAs Having different Tunnel End Points

If IPsec SAs are terminating at different addresses, which can be different WAN ports or Node lookback address, then multiple Tunnel Encap Attributes have to be included in the client UPDATE.

The encoding example for the Figure 1:

- there is one IPsec SA terminating at the WAN Port address 192.0.0.1; and another IPsec SA terminating at WAN Port 170.0.0.1;
- Both IPsec SAs use GRE (value =2) as Inner Encapsulation within the IPsec Tunnel

```
 0                   1                   2                   3
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Tunnel-Type =4 (IPsec)       |        Length =               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Tunnel-end-Point Sub-TLV                           |
~            for  192.0.0.1                                     ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  IPsec SA Identifier = 1     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                  GRE Sub-TLV                                  ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Tunnel-Type =4 (IPsec)       |        Length =               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Tunnel-end-Point Sub-TLV                           |
~            for  170.0.0.1                                     ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  IPsec SA Identifier = 1     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                  GRE sub-TLV                                  ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

6. Underlay Network Properties Advertisement using MP-NLRI

For the Figure 1 in Section 3, C-PE2 needs to advertise its IPsec SA associated attributes, such as the public keys, the nonce, the

supported encryption algorithms for the IPsec tunnels terminated at
192.0.0.1, 170.1.1.1 and 2.2.2.2 respectively.

Using the IPsec Tunnel [ISP4: 160.0.0.1 <-> ISP2:170.0.0.1] as an
example: C-PE1 needs to advertise the following attributes for
establishing the IPsec SA:
  NH: 160.0.0.1
  SDWAN Node ID
  SDWAN-Site-ID
  Tunnel Encap Attr (Type=SDWAN)
        ISP Sub-TLV for information about the ISP4
        IPsec SA Nonce Sub-TLV,
        IPsec SA Public Key Sub-TLV,
        IPsec SA Sub-TLV for the supported transforms
              {Transforms Sub-TLV - Trans 2,
              Transforms Sub-TLV - Trans 3}

C-PE2 needs to advertise the following attributes for establishing
IPsec SA:
  NH: 170.1.1.1
  SDWAN Node ID
  SDWAN-Site-ID
  Tunnel Encap Attr (Type=SDWAN)
        ISP Sub-TLV for information about the ISP2
        IPsec SA Nonce Sub-TLV,
        IPsec SA Public Key Sub-TLV,
        IPsec SA Sub-TLV for the supported transforms
              {Transforms Sub-TLV - Trans 2,
              Transforms Sub-TLV - Trans 4}

As both end points support Transform #2, the Transform #2 will be
used for the IPsec Tunnel [ISP4: 160.0.0.1 <-> ISP2:170.0.0.1].

6.1. Controller Facilitated IPsec Tunnels for SDWAN Networks

IPsec is a common technique used to encrypt traffic traversing
untrusted networks. IPSec operation between two peer nodes need to
perform Internet Key Exchange (IKEv2), which can be broken down into
the following steps:

- IKE_SA_INIT exchanges: This pair of messages negotiate
  cryptographic algorithms, exchange nonces, and do a Diffie-
  Hellman exchange.

- IKE_AUTH: this pair of messages authenticate the previous
  messages, exchange identities and certificates, and establish
  the first Child SA. Based on the authentication used: Pre-
  Shared Key, RSA certificates or EAP the number of messages
  exchanged in IKE_AUTH can grow.

- CREATE_CHILD_SA - This is simply used to create additional
  CHILD-SAs as needed

- INFORMATIONAL- During an IKEv2 SA lifetime, peers may desire to
  exchange some control messages related to errors or have
  notifications of certain events. This function is accomplished
  by INFORMATIONAL exchange.

In SDWAN environment, each SDWAN edge node might need to establish
IPsec tunnels to multiple peers, and there can be multiple IPsec
tunnels for different client traffic between any two peers. In
addition, SDWAN edge nodes can be far apart. Upon powering up, a
SDWAN edge might not know their authorized communication peers and
might not have the policies configured for aligning traffic with
their specific IPsec Tunnels. Therefore, the IPsec operation in
SDWAN environment are usually facilitated by its SDWAN Controller.

[SDN-IPsec] describes two different mechanisms to achieve controller
facilitated IPsec configuration: IKE case vs. IKE-less case. Under
the IKE case, the Controller is in charge of provisioning the
required information to IKE, the Security Policy Database (SPD) and
the Security Association Database (PAD). The SDWAN peers exchange
the Internet Key Exchange (IKE) protocol and manage SPD and SAD.
Under the IKE-less case, the Controller will provide the required
parameters to create valid entries in the SPD and the SAD into the
edge nodes. Therefore, the edge node will only need to
implementation IPsec encryption while automated key management
functionality is moved to the Controller.

For BGP controlled SDWAN networks, there is already a secure
management tunnel established between RR and the edge nodes, all the
negotiations exchanged in IKEv2 can be carried by BGP UPDATE
messages to/from the Route Reflector (RR), which will propagate the
information to the intended destinations. More importantly, when an
edge node needs to establish multiple IPsec tunnels to many
different SDWAN edge nodes, all the management information can be
multiplexed into the secure management tunnel between RR and the

edge node. Therefore, there is reduced amount of work on authentication in processing in BGP Controlled SDWAN networks.

In addition, BGP has a built-in mechanism to constrain SDWAN UPDATE messages only to the authorized peers that an SDWAN edge node can communicate [RFC4364].

6.2. NLRI encoding for Underlay Network Properties

For the MPLS VPN, the underlay network is controlled by the VPN service provider, therefore, there is no need for nodes to advertise any underlay properties to remote peers.

For the untrusted underlay network to which a SDWAN edge is connected, many attributes need to be advertised to remote nodes, such as:

- ISP information of the underlay network,
- NAT property
- the geolocation of the SDWAN edge
- IPsec SA attributes, such as public keys, nonce, supported
   encryption algorithms, etc.
- the IPsec tunnel termination address

Here is the encoding for those attributes in the NLRI field within the MP_REACH_NLRI Path Attribute of RFC4760, under a SDWAN SAFI (code = 74):

```
+-----------------+
|   NLRI Length   | 1 octet
+-----------------+
|   Site-Type     | 1 Octet
+-----------------+
|   IPSec-SA-Type | 1 Octet
+-----------------+
|   Port-Local-ID | 4 octets
+-----------------+
|  SDWAN-Site-ID  | 4 octets
+-----------------+
|  SDWAN-Node-ID  | 4 or 16 octets
+-----------------+
```

where:

- NLRI Length: 1 octet of length expressed in bits as defined in
  [RFC4760].
- Site Type: 1 octet value. The SDWAN Site Type defines the
  different types of Site IDs to be used in the deployment. The
  draft defines the following types:
     Site-Type = 1: For simple deployment, such as all edge nodes
     under one SDWAN management system, a simple identifier is
     enough for the SDWAN management to map the site to its
     precise geolocation.
     Site-Type = 2: to indicate that the value in the site-ID is
     locally significant, therefore, need a Geo-Loc Sub-TLV to
     fully describe the accurate location of the node. This is for
     large SDWAN heterogeneous deployment where Site IDs has to be
     described by proper Geo-location of the Edge Nodes [LISP-
     GEOLoc].
- IPSec-SA-Type: 1 octet value. The IPSec SA Type represents two
  different types of IPSec SA Sub-TLV encoding to be carried with
  the NLRI.
     IPSec-SA-Type = 1 : Simple IPSec Security Association
     properties defined in IPSec SA Sub-TLV.

            IPSec-SA-Type = 2: The full set of IPSec Sub-TLVs including
            Nonce, Public Key, Proposal and Transform Sub-TLVs.
      -            Port local ID: SDWAN edge node Port identifier, which can be
        locally significant. The detailed properties about the network
        connected to the port are further encoded in the Tunnel Path
        Attribute.
      - SDWAN-Site-ID: used to identify a common property shared by a
        set of SDWAN edge nodes, such as the property of a specific
        geographic location shared by a group of SDWAN edge nodes. The
        property is used to steer an overlay route to traverse specific
        geographic locations for various reasons, such as to comply
        regulatory rules, to utilize specific value-added services, or
        others.
      - SDWAN Edge Node ID: a routable address (IPv4 or IPv6) within the
        WAN to reach this node or port.

      [Editor's note on using SDWAN SAFI for the underlay network
      property advertisement:
            SDWAN SAFI [IANA assigned =74] is used instead of IP SAFI in
            the MP-NLRI [RFC4760] Path Attribute to advertise the
            underlay network properties to emphasize that the address in
            the NLRI is NOT client addresses.
            If the same IP SAFI used, receiver needs to add extra logic
            to differentiate regular BGP MP-NLRI client routes
            advertisement from the SDWAN underlay network properties
            advertisement. The benefit of using the same IP SAFI is that
            the UPDATE can traverse existing routers without being
            dropped. Since the SDWAN underlay network UPDATE is only
            between SDWAN edge and its corresponding RR, there won't be
            any intermediated routers. Therefore, this benefit becomes
            not applicable.
      ]

    The underlay network property encoding structure is as follows:

      SDWAN SAFI NLRI: <Site-Type, IPSec-SA-Type, Port-Local-ID, SDWAN-
      Site-ID, SDWAN-Node-ID>

Encoding for Site-Type = 1, IPSec-SA-Type = 1 is defined below:

Attributes:

 Tunnel Encaps Attribute (23)

        Tunnel Type: SDWAN-Underlay (to be assigned by IANA)

                NAT Sub-TLV (Optional)

                IPsec-SA Attribute Sub-TLV (Mandatory Base Sub-TLV)

                ISP of the Underlay network Sub-TLV (Optional)

Encoding for Site-Type = 2, IPSec-SA-Type = 1 is defined below:

Geo-Prefix and Geo-Point Sub-TLV (Mandatory)

Attributes:

 Tunnel Encaps Attribute (23)

        Tunnel Type: SDWAN-Underlay (to be assigned by IANA)

                NAT Sub-TLV (Optional)

                IPsec-SA Attribute Sub-TLV (Mandatory Base Sub-TLV)

                ISP of the Underlay network Sub-TLV (Optional)

The Geo-Prefix and Geo-Point Sub-TLV is defined in [LISP-GEOLOC].

6.3. Underlay Properties encoding in the Tunnel Path Attribute

The underlay properties are encoded in the Tunnel Encapsulation
Attribute defined in [Tunnel-Encap] using a new Tunnel-Type TLV
(code point to be assigned by IANA).

The Tunnel Encaps Attribute are defined as follows:

```
 0                   1                   2                   3
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Tunnel-Type(=SDWAN-Underlay )| Length (2 Octets)            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             Value                            |
|                                                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
         SDWAN Underlay network Sub-TLV Value Field
```

   Where:
      Tunnel Type is SDWAN-Underlay (to be assigned by IANA).

6.4. Extended Sub-TLV for NAT

   When a SDWAN edge node is connected to an underlay network via a
   port behind NAT devices, traditional IPsec uses IKE for NAT
   negotiation. The location of a NAT device can be such that:
     - Only the initiator is behind a NAT device. Multiple initiators
       can be behind separate NAT devices. Initiators can also connect
       to the responder through multiple NAT devices.
     - Only the responder is behind a NAT device.
     - Both the initiator and the responder are behind a NAT device.

   The initiator's address and/or responder's address can be
   dynamically assigned by an ISP or when their connection crosses a
   dynamic NAT device that allocates addresses from a dynamic address
   pool.

   Because one SDWAN edge can connect to multiple peers via one
   underlay network, the pair-wise NAT exchange as IPsec's IKE is not
   efficient. In BGP Controlled SDWAN, NAT information of a WAN port is
   advertised to its RR in the BGP UPDATE message. It is encoded as an
   Extended sub-TLV that describes the NAT property if the port is
   behind a NAT device.

   A SDWAN edge node can inquire STUN (Session Traversal of UDP Through
   Network Address Translation RFC 3489) Server to get the NAT
   property, the public IP address and the Public Port number to pass
   to peers.

```
         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
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
        |Port Ext 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                                 |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   Where:

     o Port Ext Type: indicate it is the Port Ext SubTLV.
     o PortExt subTLV Length: the length of the subTLV.
     o Flags:
         - I bit (CPE port address or Inner address scheme)
            If set to 0, indicate the inner (private) address is IPv4.
            If set to 1, it indicates the inner address is IPv6.

         - O bit (Outer address scheme):
            If set to 0, indicate the public (outer) address is IPv4.
            If set to 1, it indicates the public (outer) address is
            IPv6.

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

     o 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).

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

6.5. IPsec Security Association Property Sub-TLVs

Editor's Note:

RFC7296 specifies the IPsec SA attributes exchange among two
peers to establish peer-wise IPsec SA. [Controller-IKE]
specifies the structure for a controller to facilitate the
exchange of the RFC7296 specified IPsec SA attributes among
many nodes.

[CONTROLLER-IKE] specifies the Device Information Message
(DIM) for the edge node to advertise to its controller, which
includes DH public number, nonce, peer identity, an indication
whether this is the initial distributed policy, and rekey
counter. The originating edge node distributes the DH public
value along with the other values in the DIM (using IPsec
Tunnel TLV in Tunnel Encapsulation Attribute) to other remote
C-PEs via the RR. Each receiving C-PE uses this DH public
number and the corresponding nonce in creation of IPsec SA
pair to the originating C-PE - i.e., an outbound SA and an
inbound SA. The detail procedures are described in section 5.2
of [CONTROLLER-IKE].

[SECURE-VPN] proposes the BGP UPDATE Sub-TLV structure to
carry the information specified by [Controller-IKE] to be
propagated among peers via BGP.

To expedite the standardization process, this draft aligns
with the IPsec Sub-TLVs described in the Section 6.1, 6.2 and
6.3 of [SECURE-EVPN], with some optimization.

For scalability reason, this draft advertises the IPSec SA
related attributes at a different pace than client routes
UPDATEs. Client Routes UPDATE only references the identifier
for the prior established IPsec SAs.

The optimized IPsec SA attributes are represented by a set of Sub-
TLVs:

  - IPsec SA Nonce Sub-TLV
  - IPsec SA Public Key Sub- TLV
  - IPsec SA Proposal Sub-TLV: to indicate the number of
     Transform Sub-TLVs
         o Transforms Substructure Sub-TLV

For BGP controlled SDWAN network, very often an edge node doesn't
know its peer identity. Then the peer identity field can be null.

6.6. IPsec SA Nonce Sub-TLV

The Nonce Sub-TLV is based on the Base DIM sub-TLV as described the
Section 6.1 of [SECURE-EVPN]. IPsec SA ID is added to the sub-TLV,
which is to be referenced by the client route NLRI Tunnel Encap Path
Attribute for the IPsec SA. The following fields are removed
because:

  - the Originator ID is carried by the NLRI,
  - the Tenant ID is represented by the Route Target Extended
     Community, and
  - the Subnet ID are carried by the BGP route UPDATE.

 The format of this Sub-TLV is as follows:
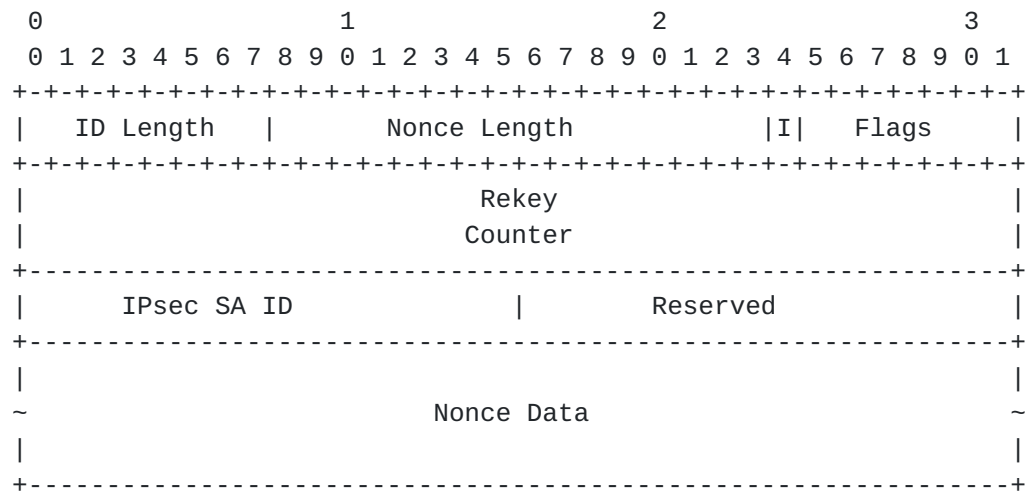
```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   ID Length   |        Nonce Length       |I|   Flags       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Rekey                             |
|                            Counter                          |
+-------------------------------------------------------------+
|       IPsec SA ID         |          Reserved               |
+-------------------------------------------------------------+
|                                                             |
~                        Nonce Data                           ~
|                                                             |
+-------------------------------------------------------------+
```

   IPsec SA ID - The 2 bytes IPSec SA ID could 0 or non-zero values. It
   is cross referenced by client route's IPSec Tunnel Encap IPSec-SA-ID
   or IPSec-SA-Group Sub-TLV in Section 5. When there are multiple
   IPsec SAs terminated at one address, such as WAN port address or the
   node address, they are differentiated by the different IPsec SA IDs.

6.7. IPsec Public Key Sub-TLV

   The IPsec Public Key Sub-TLV is derived from the Key Exchange Sub-
   TLV described in [SECURE-EVPN] with an addition of Duration filed to
   define the IPSec SA life span. The edge nodes would pick the
   shortest duration value between the SDWAN SAFI pairs.

   The format of this Sub-TLV is as follows:

```
 0                   1                   2                   3
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Diffie-Hellman Group Num    |          Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                             |
~                     Key Exchange Data                        ~
|                                                             |
+-------------------------------------------------------------+
|                          Duration                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

6.8. IPsec SA Proposal Sub-TLV

The IPsec SA Proposal Sub-TLV is to indicate the number of Transform
Sub-TLVs. This Sub-TLV aligns with the sub-TLV structure from
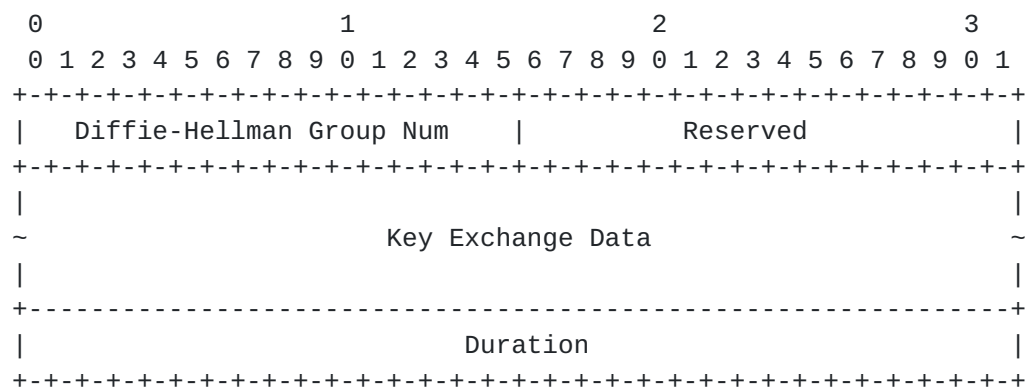[SECURE-VPN]

The Transform Sub-sub-TLV will following the section 3.3.2 of
 RFC7296.

6.9. ISP of the Underlay network Sub-TLV

The purpose of the Underlay network Sub-TLV is to carry the ISP WAN
port properties with SDWAN SAFI NLRI. It would be treated as
optional Sub-TLV. The BGP originator decides whether to include this
Sub-TLV along with the SDWAN NLRI. If this Sub-TLV is present, it
would be processed by the BGP receiver and to determine what local
policies to apply for the remote end point of the Underlay tunnel.

The format of this Sub-TLV is as follows:

```
     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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |     Type      |     Length    |      Flag     |   Reserved    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |Connection Type|   Port Type   |        Port Speed             |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

Type: To be assigned by IANA

Length: 6 bytes.

Flag: a 1 octet value.

Reserved: 1 octet of reserved bits. It SHOULD be set to zero on
transmission and MUST be ignored on receipt.

Connection Type: There are two different types of WAN
Connectivity. They are listed below as:

Wired - 1
WIFI - 2
LTE - 3

       5G - 4

       Port Type: There are different types of ports. They are listed
       Below as:

       Ethernet - 1
       Fiber Cable - 2
       Coax Cable - 3
       Cellular - 4

       Port Speed: The port seed is defined as 2 octet value. The values
       are defined as Gigabit speed.

6.10. Simplified IPsec Security Association sub-TLV

       For a simple SDWAN network with edge nodes supporting only a few
       pre-defined encryption algorithms, a simple IPsec sub-TLV can be
       used to encode the pre-defined algorithms, as below:

```
     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-simType   |IPsecSA Length               | Flag          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |Protocol Type  | IPsec Mode   | AH algorithms  |ESP algorithms |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |          ReKey Counter                                       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | key1 length   |         Public Key                         ~
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | key2 length   |         Nonce                              ~
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |    IPsec  SA ID           |        Reserved                 |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |        Duration                                             |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```


   Where:

o IPsec-SimType: The type value has to be between 128~255 because
  IPsec-SA subTLV needs 2 bytes for length to carry the needed
  information.
o IPsec-SA subTLV Length (2 Byte): 25 (or more)
o Flags: 1 octet of flags. None are defined at this stage. Flags
  SHOULD be set to zero on transmission and MUST be ignored on
  receipt.
o IPsec Protocol Types (1 Byte): the value can be AH, ESP, or
  AH+ESP.
o IPsec Mode (1 byte): the value can be Tunnel Mode or Transport
  mode
o AH algorithms (1 byte): AH authentication algorithms supported,
  which can be md5 | sha1 | sha2-256 | sha2-384 | sha2-512 | sm3.
  Each SDWAN edge node can have multiple authentication.
  algorithms; send to its peers to negotiate the strongest one.
o ESP (1 byte): ESP authentication algorithms supported, which
  can be md5 | sha1 | sha2-256 | sha2-384 | sha2-512 | sm3. Each
  SDWAN edge node can have multiple authentication algorithms;
  send to its peers to negotiate the strongest one. Default
  algorithm is AES-256.
    o When node supports multiple authentication algorithms, the
      initial UPDATE needs to include the "Transform Sub-TLV"
      described by [SECURE-EVPN] to describe all of the
      algorithms supported by the node.

o Rekey Counter: 4 bytes
o Public Key: IPsec public key
o Nonce.IPsec Nonce
o IPsec SA ID: The 2 bytes IPSec SA ID could 0 or non-zero
  values. It is cross referenced by client route's IPSec Tunnel
  Encap IPSec-SA-ID or IPSec-SA-Group Sub-TLV in Section 5. When
  there are multiple IPsec SAs terminated at one address, such as
  WAN port address or the node address, they are differentiated
  by the different IPsec SA IDs.
o Duration: SA life span.

6.11. Remote Endpoint

The Remote Endpoint sub-TLV is not used for SDWAN NLRI because

o Tunnel end point address is already included in the Tunnel-end-
  point Sub-TLV.
o The underlay networks to which an SDWAN edge node are connected
  might have different identifiers than their corresponding AS
  numbers on the SDWAN controller. The SDWAN controller might use
  its own specific identifiers for the underlay networks.
o The Transport-Network-ID in the EncapExt sub-TLV represents the
  SDWAN unique network identifier.

If the Remote Endpoint Sub-TLV is present, it is ignored by the RR
and other SDWAN edge nodes.

7. Error & Mismatch Handling

Each C-PE device advertises SDWAN SAFI Underlay NLRI to the other C-
PE devices via BGP Route Reflector to establish pairwise SAs between
itself and every other remote C-PEs. During the SAFI NLRI
advertisement, the BGP originator would include either simple IPSec
Security Association properties defined in IPSec SA Sub-TLV based on
IPSec-SA-Type = 1 or full-set of IPSec Sub-TLVs including Nonce,
Public Key, Proposal and number of Transform Sub-TLVs based on
IPSec-SA-Type = 2.

The C-PE devices would compare the IPSec SA attributes between the
local and remote WAN ports. If there is a match on the SA Attributes
between the two ports, the IPSec Tunnel would be established.

The C-PE devices would not try to negotiate the base IPSec-SA
parameters between the local and the remote ports in the case of
simple IPSec SA exchange or the Transform sets between local and
remote ports if there is a mismatch on the Transform sets in the
case of full-set of IPSec SA Sub-TLVs.

As an example, using the Figure 1 in Section 3, to establish IPsec
Tunnel between C-PE1 and C-PE2 WAN Ports A2 and B2 [A2: 192.10.0.10
<-> B2:192.0.0.1]:

C-PE1 needs to advertise the following attributes for establishing
the IPsec SA:
  NH: 192.10.0.10
  SDWAN Node ID

```
     SDWAN-Site-ID
     Tunnel Encap Attr (Type=SDWAN)
         ISP Sub-TLV for information about the ISP3
         IPsec SA Nonce Sub-TLV,
         IPsec SA Public Key Sub-TLV,
         Proposal Sub-TLV with Num Transforms = 1
              {Transforms Sub-TLV - Trans 1}
```

   C-PE2 needs to advertise the following attributes for establishing
   IPsec SA:

```
     NH: 192.0.0.1
     SDWAN Node ID
     SDWAN-Site-ID
     Tunnel Encap Attr (Type=SDWAN)
         ISP Sub-TLV for information about the ISP1
         IPsec SA Nonce Sub-TLV,
         IPsec SA Public Key Sub-TLV,
         Proposal Sub-TLV with Num Transforms = 1
              {Transforms Sub-TLV - Trans 2}
```

   As there is no matching transform between the WAN ports A2 and B2 in
   C-PE1 and C-PE2 respectively, there will be no IPsec Tunnel be
   established.

8. Manageability Considerations

      TBD - this needs to be filled out before publishing

9. Security Considerations

   The document describes the encoding for SDWAN edge nodes to
   advertise its properties to their peers to its RR, which
   propagates to the intended peers via untrusted networks.

   The secure propagation is achieved by secure channels, such as
   TLS, SSL, or IPsec, between the SDWAN edge nodes and the local
   controller RR.

   [More details need to be filled in here]

10. IANA Considerations

This document requires the following IANA actions.

    o SDWAN Overlay SAFI = 74 assigned by IANA
    o SDWAN Route Type

11. References

11.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
          Requirement Levels", BCP 14, RFC 2119, March 1997.

11.2. Informative References

[RFC8192] S. Hares, et al, "Interface to Network Security Functions
          (I2NSF) Problem Statement and Use Cases", July 2017

[RFC5521] P. Mohapatra, E. Rosen, "The BGP Encapsulation Subsequent
          Address Family Identifier (SAFI) and the BGP Tunnel
          Encapsulation Attribute", April 2009.

[CONTROLLER-IKE] D. Carrel, et al, "IPsec Key Exchange using a
          Controller", draft-carrel-ipsecme-controller-ike-01, work-
          in-progress.

[LISP-GEOLOC] D. Farinacci, "LISP Geo-Coordinate Use-Case", draft-
          farinacci-lisp-geo-09, April 2020.

[SDN-IPSEC] R. Lopez, G. Millan, "SDN-based IPsec Flow Protection",
          draft-ietf-i2nsf-sdn-ipsec-flow-protection-07, Aug 2019.

[SECURE-EVPN] A. Sajassi, et al, "Secure EVPN", draft-sajassi-bess-
          secure-evpn-02, July 2019.

[Tunnel-Encap]E. Rosen, et al, "The BGP Tunnel Encapsulation
          Attribute", draft-ietf-idr-tunnel-encaps-09, Feb 2018.

[VPN-over-Internet] E. Rosen, "Provide Secure Layer L3VPNs over
            Public Infrastructure", draft-rosen-bess-secure-l3vpn-00,
            work-in-progress, July 2018

[DMVPN] Dynamic Multi-point VPN:
            https://www.cisco.com/c/en/us/products/security/dynamic-
            multipoint-vpn-dmvpn/index.html

[DSVPN] Dynamic Smart VPN:
            http://forum.huawei.com/enterprise/en/thread-390771-1-
            1.html

[ITU-T-X1036] ITU-T Recommendation X.1036, "Framework for creation,
            storage, distribution and enforcement of policies for
            network security", Nov 2007.

[Net2Cloud-Problem] L. Dunbar and A. Malis, "Seamless Interconnect
            Underlay to Cloud Overlay Problem Statement", draft-dm-
            net2cloud-problem-statement-02, June 2018

[Net2Cloud-gap] L. Dunbar, A. Malis, and C. Jacquenet, "Gap Analysis
            of Interconnecting Underlay with Cloud Overlay", draft-dm-
            net2cloud-gap-analysis-02, work-in-progress, Aug 2018.

[Tunnel-Encap] E. Rosen, et al "The BGP Tunnel Encapsulation
            Attribute", draft-ietf-idr-tunnel-encaps-10, Aug 2018.

12. Acknowledgments

Authors' Addresses

    Linda Dunbar
    Futurewei
    Email: ldunbar@futurewei.com

    Sue Hares
    Hickory Hill Consulting
    Email: shares@ndzh.com

    Robert Raszuk
    Email: robert@raszuk.net

    Kausik Majumdar
    CommScope
    Email: Kausik.Majumdar@commscope.com