idr
Internet-Draft
Intended status: Standards Track
Expires: September 5, 2019

J. Hu
Nokia

March 4, 2019

BGP Signaled IPsec Tunnel Configuration
draft-hujun-idr-bgp-ipsec-00

Abstract

This document defines a method of using BGP to signal IPsec tunnel configuration along with NLRI, it uses and extends tunnel encapsulation attribute as specified in [I-D.ietf-idr-tunnel-encaps] for IPsec tunnel.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 5, 2019.

Table of Contents

1. Introduction

IPsec is the standard for IP layer traffic protection, however in a big network where mesh connections are needed, configuring large number of IPsec tunnels is error prone and not scalable. So instead of pre-provision IPsec tunnels on each router, this document defines a method to allow router to advertise the IPsec tunnel configurations it requires to reach a given NLRI via BGP. This document does not intend to be one solution for all cases, the main use case is to simplify IPsec tunnel provision in networks under single administrative domain; it uses standard based components (IPsec/IKEv2[RFC7296] and BGP) with limited changes. There is no change to IPsec/IKEv2, and only limited changes to BGP.

IPsec tunnel configurations typically include following parts:

- tunnel endpoint address (local and remote)
- public routing instance, routing instance where IPsec packet is forwarded in
- private routing instance, routing instance where payload packet is forwarded in
- tunnel authentication method and credentials
- IKE SA and CHILD SA transform (a.k.a crypto algorithms)
- CHILD SA traffic selector
- other: like lifetime, DPD timer, use of PFS ..etc

In order to minimize amount configurations signal via BGP, only following configurations are explicit advertised:

- local tunnel endpoint address: BGP tunnel encapsulation attribute
- public routing instance: sub-TLV in tunnel encapsulation attribute
- CHILD SA traffic selector: NLRI and/or sub-TLV in tunnel encapsulation attribute

Other configurations are either derived or via color mapping:

- remote tunnel endpoint address: dynamic learned when received IKEv2 IKE_SA_INIT request
- private routing instance: via route-target in same BGP UPDATE
- tunnel authentication and credentials: out of scope, could be PKI based authentication
- IKE SA and CHILD SA transform, lifetime, DPD timer, PFS ..etc: all these configurations are implicitly signaled via color sub-TLV in tunnel encapsulation attribute

[I-D.ietf-idr-tunnel-encaps] defines a generic tunnel encapsulation attribute for BGP, however it needs to be extended to support IPsec tunnel.

1.1. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Tunnel Encapsulation Attribute for IPsec

This document extends tunnel encapsulation attribute specified in [I-D.ietf-idr-tunnel-encaps] by introducing following changes:

- A tunnel type for IPsec tunnel: ESP tunnel mode (AH tunnel mode is not included in this document). Existing type 4 (IPsec in Tunnel-mode) in IANA "BGP Tunnel Encapsulation Attribute Tunnel Types" registry could be reused
- A new sub-TLV for public routing instance

- A new sub-TLV for remote address prefix
- A new sub-TLV for local address prefix

Following existing sub-TLVs apply to IPsec tunnel encapsulation attribute:

- Remote Endpoint: IPsec tunnel endpoint address
- Color: IPsec configuration attributes like ESP transform; the meaning of this sub-TLV is local to the administrative domain
- Embedded Label Handling: see Section 4 for detail

2.1. Local and Remote Prefix sub-TLV

Local prefix sub-TLV is an optional sub-TLV used to specify a list of address prefix that used as local traffic selector address ranges; if local prefix sub-TLV is not included, then prefixes in NLRI will be used; Remote prefix sub-TLV is a mandatory sub-TLV used to specify a list of address prefix that used as remote traffic selector address ranges; The IP version of local/remote prefix MUST be as same as IP version of prefix in NLRI. A single all zero prefix means any prefix is allowed. Local and remote prefix sub-TLV has same encoding as following:

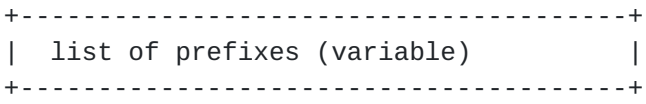

Figure 1: Source Prefix sub-TLV

Each prefix is encoded as following:

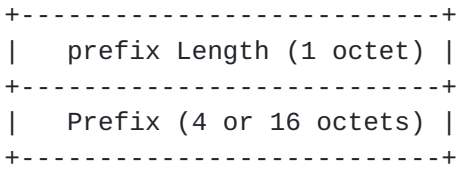

Figure 2: prefix

2.2. Public Routing Instance sub-TLV

Public routing instance sub-TLV is an optional sub-TLV used to specify the routing instance to which the remote point address belongs, if tunnel encapsulation attribute doesn't include this TLV, then the routing instance is the same to which BGP session belongs.

the value field of the sub-TLV consist a route target community as defined in [RFC4360].

3. Operation

Following are the rules of operation:

1. All routers are in same administrative domain

2. All routers are pre-provisioned with following:

   * Authentication credential like PKI certificates and key

   * Mapping between color and IPsec configurations

3. If a given NLRI need IPsec protection, then advertising router need to include an IPsec tunnel encapsulation attribute, along with the NLRI in BGP UPDATE U;

4. When a router need to forward a packet along a path is determined by a BGP UPDATE which has a tunnel encapsulation attribute that contains one or more IPsec TLV, and router decides use IPsec based on local policy, then the router need to check if there is an existing CHILD SA could be used, a CHILD SA could be used when it meets all following conditions:

   * its private routing instance is same as routing instance to which the packet to be forwarded belongs

   * its public routing instance is same as indicated by the Public Routing Instance sub-TLV; if the sub-TLV doesn't exist, then it is same as routing instance to which BGP session belongs

   * its peer tunnel address is same as indicated by Remote Endpoint sub-TLV

   * the source and destination address of the packet to be forwarded falls in the range of CHILD SA's traffic selector

   * its transform and other configuration maps to the color indicated in the Color sub-TLV

5. If router can't find such CHILD SA, then it will use IKEv2 to create one; if there are multiple IPsec TLVs in U, then it need to select one from feasible TLVs, a IPsec TLV is considered as feasible when it meets all following requirements:

* the source address of the packet must fall in one of Remote Prefixes

* the destination address of the packet must fall one of Source Prefixes

* the Remote Endpoint, along with Public Routing Instance sub-TLV identifies an IP address that is reachable

6. After an IPsec TLV is selected, router uses IKEv2 to create the CHILD_SA:

   * public/private routing instance, peer's tunnel address are chosen based on above rules

   * Traffic Selector:

   * For each TS in TSi:

     + address range: the prefix specified in Remote Prefix sub-TLV

     + protocol: any

     + port range: any

   * for each TS in TSr:

     + address range: prefixes specified by Local Prefix sub-TLV if it exists; otherwise use the prefix specified by the NLRI

     + protocol: any

     + port range: any

The operation of BGP signaling IPsec configuration is illustrated with following example:

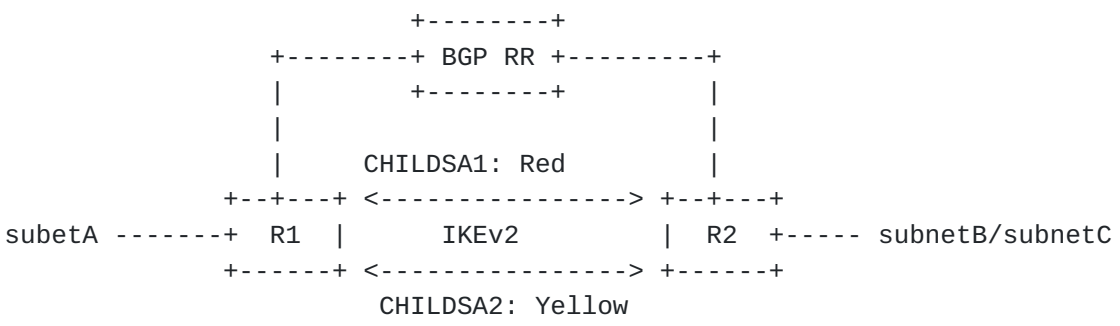

Figure 3: Operation Example

There are following traffic protection requirements:

- subnetA - subnetB: ESP tunnel, AES-CBC-256 with SHA-384, mapping to color red

- subnetA - subnetC: ESP tunnel, null encryption with only integrity protection, SHA-256, mapping to color yellow

Both R1 and R2 are provisioned with PKI key and certificate from same CA.

- R1 advertise subnetA in BGP UPDATE, which has a tunnel encapsulation attribute that contains two IPsec tunnel TLVs:

  * TLV-1: endpoint R1TunnelAddr, color sub-TLV red and subnetB in Remote Prefix sub-TLV.

  * TLV-2: endpoint R1TunnelAddr, color sub-TLV yellow and subnetC in Remote Prefix sub-TLV.

- R2 advertise subnetB in BGP UPDATE, which has a tunnel encapsulation attribute that contains one IPsec tunnel TLV: R2TunnelAddr, color sub-TLV red and subnetA in Remote Prefix sub-TLV.

- R2 advertise subnetC in BGP UPDATE, which has a tunnel encapsulation attribute that contains one IPsec tunnel TLV: R2TunnelAddr, color sub-TLV yellow and subnetA in Remote Prefix sub-TLV.

- R1 received a packet from subnetA destined to subnetB, since BGP UPDATE contain subnetB also contains an IPsec tunnel encapsulation attribute, there is no existing CHILD SA could be used, based on the rules described in this section, R1 select TLV-1 and uses IKEv2 to establish an IPsec tunnel to R2TunnelAddr, using certificate authentication, create 1st CHILD SA CHILDSA1:

    * ESP transform: AES-CBC-256 and SHA-384

    * Traffic Selector:

      + TSi: address subnetA, protocol any, port any

      + TSr: address subnetB, protocol any, port any

- after tunnel is created, R1 and R2 could forward traffic between subnetA and subnetB over CHILDSA1

- R1 received a packet from subnetA destined to subnetC, CHILDSA1 can't be used for this packet, R1 select TLV-2 to create 2nd CHILD SA, and given there is already an IKE SA between R1 and R2, R1 uses existing IKESA to create CHILDSA2:

    * ESP transform: Null encryption with SHA-256

    * Traffic Selector:

      + TSi: address subnetA, protocol any, port any

      + TSr: address subnetC, protocol any, port any

- R1 and R2 could forward traffic between subnetA and subnetC over CHILDSA2

4. Semantics and Usage of IPsec Tunnel Encapsulation attribute

IPsec tunnel encapsulation TLV has same usage and semantics as defined in [I-D.ietf-idr-tunnel-encaps] with following differences:

- Due to nature of IPsec, the payload packet could only be IPv4 or IPv6 packet, so it MAY be carried in any BGP UPDATE message whose AFI/SAFI is 1/1 (IPv4 Unicast), 2/1 (IPv6 Unicast).

- For 1/128 (VPN-IPv4 Labeled Unicast), 2/128 (VPN-IPv6 Labeled Unicast), these NLRI has embedded label, which cause the payload packet can't be encapsulated in ESP packet, however with IPsec tunnel encapsulation, the label could be ignored during encapsulation since CHILD SA itself could be used to identify the private routing instance; so an UPDATE that include IPsec tunnel encapsulation attribute, which contains value 2 of Embedded Label Handling Sub-TLV, could be used to signal this type of setup.

- For other types of AFI/SAFI, a nested tunnel setup could be used to get IPsec protection, for example, an 25/70 (EVPN) payload

packet could be encapsulated in VXLAN over IPsec tunnel. See Section 4.1 for further detail.

4.1. Nested Tunnel

A nested tunnel could be used for payload packet type that can't be encapsulated in IPsec tunnel directly, e.g. an Ethernet packet of EVPN service. Following is an example of using VXLAN over IPsec tunnel for EVPN service:

- R1 need to forward an Ethernet packet P
- the path along which P is to be forwarded is determined by BGP UPDATE U1, which has a VXLAN tunnel encapsulation attribute and the next-hop is router R2
- the best path to R2 is a BGP route that was advertised in BGP UPDATE U2, which has an IPsec tunnel encapsulation TLV.
- R1 will encapsulate P in a VXLAN tunnel as indicated in U1, then encapsulate VXLAN packet into IPsec tunnel as indicated in U2
- if color sub-TLV is used, then both U1 and U2 MUST have matching color sub-TLV, otherwise the VXLAN packet will not be sent through IPsec tunnels identified in U2

5. IANA Considerations

This document reuses "IPsec in Tunnel-mode"(4) as BGP Tunnel Encapsulation Attribute Tunnel Types.

This document will request new values in IANA "BGP Tunnel Encapsulation Attribute Sub-TLVs" registry for following sub-TLV:

- public routing instance
- remote address prefix
- local address prefix

6. Security Considerations

IKEv2 is used to create IPsec tunnel, which ensures following:

- Traffic protection keys are generated dynamically during IKEv2 negotiation, only known by participating peer of the IPsec tunnel; there is no central node to manage and distribute all keys.

- IKEv2 rekey mechanism refresh keys regularly; PFS(Perfect Forward Secrecy) provides additional protection;
- Secure authentication mechanism that only allow authenticated peer to create tunnel
- Traffic Selector guarantee that only agreed traffic is allowed to be forwarded within the IPsec tunnel;
- Using a separate, dedicate protocol(IKEv2) for key management/ authentication ensure they are not tied to BGP, all existing and future IKEv2 features could be used without changing BGP;

There is concern that malicious party might manipulate IPsec tunnel encapsulation attribute to divert traffic, however this risk could be mitigated by IKEv2 mutual authentication.

BGP Origin Validation [RFC6811] and BGPSec [RFC8205] could be used to further secure BGP UPDATE message.

7. References

7.1. Normative References

[I-D.ietf-idr-tunnel-encaps]
Rosen, E., Patel, K., and G. Velde, "The BGP Tunnel Encapsulation Attribute", draft-ietf-idr-tunnel-encaps-11 (work in progress), February 2019.

[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>.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, DOI 10.17487/RFC4360, February 2006, <https://www.rfc-editor.org/info/rfc4360>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

7.2. Informative References

[RFC6811] Mohapatra, P., Scudder, J., Ward, D., Bush, R., and R. Austein, "BGP Prefix Origin Validation", RFC 6811, DOI 10.17487/RFC6811, January 2013, <https://www.rfc-editor.org/info/rfc6811>.

[RFC7296] Kaufman, C., Hoffman, P., Nir, Y., Eronen, P., and T. Kivinen, "Internet Key Exchange Protocol Version 2 (IKEv2)", STD 79, RFC 7296, DOI 10.17487/RFC7296, October 2014, <https://www.rfc-editor.org/info/rfc7296>.

[RFC8205] Lepinski, M., Ed. and K. Sriram, Ed., "BGPsec Protocol Specification", RFC 8205, DOI 10.17487/RFC8205, September 2017, <https://www.rfc-editor.org/info/rfc8205>.

Author's Address

Hu Jun
Nokia
777 East Middlefield Road
Mountain View CA 95148
United States

Email: jun.hu@nokia.com