idr
Internet-Draft
Intended status: Standards Track
Expires: March 7, 2020

J. Hu
Nokia

September 4, 2019

BGP Provisioned IPsec Tunnel Configuration
draft-hujun-idr-bgp-ipsec-01

Abstract

This document defines a method of using BGP to provide 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 March 7, 2020.

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 in this document means IPsec tunnel mode as defined in [RFC4301].

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 address range: NLRI and/or sub-TLV in tunnel encapsulation attribute

Other configurations are either derived or via tag 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/credentials, traffic selector protocol/port range, IKE SA and CHILD SA transform, lifetime, DPD timer, PFS ..etc: all these configurations are implicitly signaled via IPsec configuration tag 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
- A new sub-TLV for IPsec configuration tag

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

- Remote Endpoint: IPsec tunnel endpoint address
- 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:

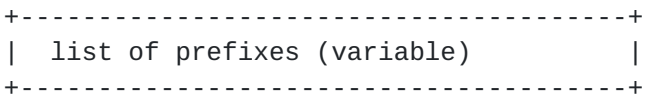

Figure 1: Source Prefix sub-TLV

Each prefix is encoded as following:

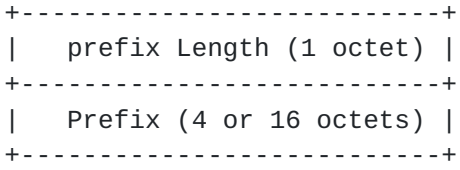

Figure 2: prefix

For a given IPsec tunnel TLV, local prefix sub-TLV MUST appear either zero or one time; remote prefix sub-TLV MUST appear only one time.

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].

For a given IPsec tunnel TLV, public routing instance sub-TLV MUST appear either zero or one time.

2.3. IPsec Configuration Tag sub-TLV

This sub-TLV represents the IPsec configurations (like IPsec transform) that are not explicit advertised by other sub-TLVs specified in this documentation; the meaning of this sub-TLV is local to the administrative domain. Follow are some examples:

- tag value T1 map to following configurations:
  - Certificate trust-anchor: CA-1
  - IKE_SA/CHILD_SA transform: AES-GCM-128
  - Diffie-Hellman Group: 15
  - Perfect Forward Secrecy: No
  - local/remote Traffic selector protocol: any
  - local/remote Traffic selector port range: any
  - IKE_SA lifetime: 24 hours
  - CHILD_SA lifetime: 1 hour
  - DPD interval: 30 seconds
  - ESP extended sequence number: no
- tag value T2 map to following configurations:
  - Certificate trust-anchor: CA-2
  - IKE_SA/CHILD_SA transform: AES-GCM-256
  - Diffie-Hellman Group: 20

* Perfect Forward Secrecy: Yes with group 20

* local/remote Traffic selector protocol: UDP

* local/remote Traffic selector port range: any

* IKE_SA lifetime: 48 hours

* CHILD_SA lifetime: 2 hours

* DPD interval: 10 seconds

* ESP extended sequence number: yes

The value field of this sub-TLV is 4 octets long. each IPsec tunnel TLV SHOULD only contain one IPsec configuration tag sub-TLV;

Figure 3: IPsec Configuration Tag

For a given IPsec tunnel TLV, IPsec configuration tag sub-TLV MUST appear only one time.

3. Operation

Following are the rules of operation:

1. All routers are in same administrative domain

2. All routers are pre-provisioned with Mapping between IPsec configuration tag value and IPsec configurations include authentication method/credentials

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 tunnel TLV, and router decides use IPsec based on local policy, then the router use first feasible CHILD_SA, a CHILD SA is considered as feasible 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 tag indicated in the IPsec configuration tag sub-TLV

5. If router can't find such CHILD SA, then it will use IKEv2 to create one; if there are multiple IPsec tunnel TLVs in U, then it need to select one from feasible TLVs, a IPsec tunnel 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. If there are multiple feasible IPsec tunnel TLV exists, then select the TLV using following rules in order:

   1. TLV with smallest local address range as indicated by Remote Prefix sub-TLV

   2. TLV with smallest remote address range as indicated by Local Prefix sub-TLV (NLRI prefix if local prefix sub-TLV is not included in TLV)

7. 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: tag mapped configuration

  + port range: tag mapped configuration

* 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: tag mapped configuration

  + port range: tag mapped configuration

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

Figure 4: Operation Example

There are following traffic protection requirements:

- subnetA - subnetB: ESP tunnel, CHACHA20_POLY1305 , mapping to tag Tag-1

- subnetA - subnetC: ESP tunnel, NULL-AES-GMAC-256 , mapping to tag Tag-2

- note: other IPsec configurations, like IKE_SA lifetime ..etc, are the same for both Tag-1 and Tag-2; not listed here for sake of

Both R1 and R2 are provisioned with IPsec authentication credentials and configurations corresponding to Tag-1 and Tag-2; both Tag-1 and Tag-2 map to traffic selector protocol any and port range any.

- R1 advertise subnetA in BGP UPDATE, which has a tunnel encapsulation attribute that contains two IPsec tunnel TLVs:
  - TLV-1: endpoint R1TunnelAddr, tag sub-TLV Tag-1 and subnetB in Remote Prefix sub-TLV.
  - TLV-2: endpoint R1TunnelAddr, tag sub-TLV Tag-2 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, tag sub-TLV Tag-1 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, tag sub-TLV Tag-2 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: CHACHA20_POLY1305
  - 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-AES-GMAC-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 specific to IPsec tunnel:

- 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 tag sub-TLV is used, then both U1 and U2 MUST have matching tag sub-TLV, otherwise the VXLAN packet will not be sent through IPsec tunnels identified in U2

4.2. Other Operation Specifics

Following are some operation specific rules:

1. An IPsec dead peer detection mechanism, like IKEv2 DPD or BFD over IPsec, SHOULD be used to monitor liveness of IPsec tunnel;

2. If IPsec peer goes down, as described in section 5 of [I-D.ietf-idr-tunnel-encaps], packet forwarding router chooses another functional tunnel, specified by another tunnel TLV of same BGP route if there is any, to forward the packet; if there is no such tunnel, then router MAY drop the packet or MAY forward packet as it would had the Tunnel Encapsulation attribute not been present. this is matter of local policy.

3. After IPsec peer goes down, packet forwarding router SHOULD try to re-establish IPsec tunnel with certain hold-down timer and back-off mechanism. the detail is up to implementation. also IKEv2 session resumption [RFC5723] MAY be used to efficiently re-create tunnel;

4. When router receives a packet destined to a BGP route it advertised but does not have any of tunnel encapsulation in the BGP route, it MAY drop it or MAY accept it; this is matter of local policy. by default, the packet should be accepted.

5. As with all types of tunnel technology, IPsec tunnel adds overhead (crypto & encapsulation) to the packet, which often causes MTU issues, deployment SHOULD take tunnel overhead into MTU consideration.

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
- IPsec configuration tag

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 route filter include outbound route filter [RFC5291], Origin Validation [RFC6811] and BGPSec [RFC8205] could be used to further secure BGP UPDATE message.

IKEv2 cookie [RFC7296] and varies mechanisms defined including client puzzle defined in [RFC8019] could be used to protect IKEv2 from Distributed Denial-of-Service Attacks.

Follow latest IETF ESP/IKEv2 implementation requirement and guidance ([RFC8221] and [RFC8247] at time of writing) to make sure always using secure and up-to-date cryptographic algorithms;

7. Change Log

- v00 March 04, 2019: initial draft
- v01 Sep 04, 2019:
  - replaces color sub-TLV with a new IPsec configuration tag sub-TLV
  - add rule on selecting TLV when there multiple feasible TLVs in section Section 3
  - change crypto used in example of section Section 3
  - change title from "BGP Signaled IPsec Tunnel Configuration" to "BGP Provisioned IPsec Tunnel Configuration"
  - Add a section Section 4.2 on some operation specifics
  - add more content in Section 6
  - add specification of number of time each new sub-TLV allowed in a given tunnel TLV
  - add clarification in section Section 1 to clarify IPsec tunnel means IPsec tunnel mode
  - traffic selector protocol and port range now come from tag mapped configuration

8. References

8.1. Normative References

[I-D.ietf-idr-tunnel-encaps]
Patel, K., Velde, G., Ramachandra, S., and E. Rosen, "The BGP Tunnel Encapsulation Attribute", draft-ietf-idr-tunnel-encaps-13 (work in progress), July 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>.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, DOI 10.17487/RFC4301, December 2005, <https://www.rfc-editor.org/info/rfc4301>.

[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>.

8.2. Informative References

[RFC5291] Chen, E. and Y. Rekhter, "Outbound Route Filtering Capability for BGP-4", RFC 5291, DOI 10.17487/RFC5291, August 2008, <https://www.rfc-editor.org/info/rfc5291>.

[RFC5723] Sheffer, Y. and H. Tschofenig, "Internet Key Exchange Protocol Version 2 (IKEv2) Session Resumption", RFC 5723, DOI 10.17487/RFC5723, January 2010, <https://www.rfc-editor.org/info/rfc5723>.

[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>.

[RFC8019] Nir, Y. and V. Smyslov, "Protecting Internet Key Exchange Protocol Version 2 (IKEv2) Implementations from Distributed Denial-of-Service Attacks", RFC 8019, DOI 10.17487/RFC8019, November 2016, <https://www.rfc-editor.org/info/rfc8019>.

[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>.

[RFC8221] Wouters, P., Migault, D., Mattsson, J., Nir, Y., and T. Kivinen, "Cryptographic Algorithm Implementation Requirements and Usage Guidance for Encapsulating Security Payload (ESP) and Authentication Header (AH)", RFC 8221, DOI 10.17487/RFC8221, October 2017, <https://www.rfc-editor.org/info/rfc8221>.

[RFC8247] Nir, Y., Kivinen, T., Wouters, P., and D. Migault, "Algorithm Implementation Requirements and Usage Guidance for the Internet Key Exchange Protocol Version 2 (IKEv2)", RFC 8247, DOI 10.17487/RFC8247, September 2017, <https://www.rfc-editor.org/info/rfc8247>.

Author's Address

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

Email: jun.hu@nokia.com