BESS WorkGroup
Internet-Draft
Intended status: Standards Track
Expires: 8 December 2025

W. Lin
Juniper
D. Rao
A. Sajassi
Cisco
L. Kreeger
Arrcus
J. Rabadan
Nokia

6 June 2025

EVPN Group Policy
draft-lrss-bess-evpn-group-policy-02

Abstract

Group Based Policy can be used to achieve micro or macro segmentation of user traffic. For Group Based Policy, a Group Policy ID, also known as Group Policy Tag, is used to represent a logical group that shares the same policy and access privilege. This document defines a backward compatible extension to Virtual eXtensible Local Area Network (VXLAN) that allows a Group Policy ID to be carried for the purposes of policy enforcement at the egress Network Virtualization Edge (NVE). It also defines a new BGP Extended Community that can be used to propagate Group Policy ID through a BGP route advertisement in the control plane. This is to facilitate policy enforcement at the ingress NVE when feasible.

Table of Contents

1. Introduction

The Group Based Policy (GROUPPOLICY, GROUPBASEDPOLICY) model defines an application-centric policy model where the application connectivity requirements are specified in a manner that is independent of the underlying network topology. In this model, Tenant System Interfaces (TSIs) are assigned to Tenant System Interface (TSI) Groups. Each TSI Group consists of TSIs that share the same network policies and requirements. Network policies are defined between the TSI Group of the traffic source and the TSI Group of the traffic destination. These policies are deployed when the TSI attaches to the network.

In certain deployment scenarios, the TSI to TSI Group mapping is only known at the Network Virtualization Edge (NVE) that the TSI is attached. This implies that the TSI Group of a packet destination may not be known until the packet reaches the egress NVE where the packet destination is attached. This could be a case where no distribution mechanism is used, or where the mapping cannot be installed in all NVEs. In such situations, it is required that the source TSI Group is carried with the packet so that policy can be applied at the egress NVE.

This document defines a backward compatible extension to VXLAN [RFC7348] that allows the source TSI Group ID to be carried so that policy can be applied when the destination TSI Group is determined at the egress NVE.

In scenarios where the TSI to TSI Group mapping decision can be made at the ingress NVEs, this document defines extensions to EVPN control plane for distribution of the destination TSI Group ID to ingress NVEs so that policy enforcement can be performed at the ingress NVEs. To accomplish this, this document defines a Group Policy ID BGP Extended Community to be sent along with EVPN routes to achieve the propagation of Group Policy ID from an egress node to an ingress node.

This document does not preclude the VXLAN Group Policy header extension from being used with other control plane mechanisms, or without any control plane.

Note: This document is a merged of two IETF documents - one that defines the VXLAN Group Policy header extensions (VXLAN-GRPPOLICY), and another one that deinfes the EVPN Extended Community extensions (EVPN-GPI-EC).

2. Requirements Language

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 BCP14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Terminology

* AC: Attachment Circuit

* DCI: Data Center Interconnect

* EVPN: Ethernet VPN, as per [RFC7432]

* GBP : Group Based Policy

* NVE: Network Virtualization Edge

* NVO: Network Virtualization Overlay

* PE: Provider Edge device

* TSI: Tenant System Interface - equivalent to an attachment circuit for a given tenant

* Tenant System Interface (TSI) Group: A TSI Group is a collection of TSIs that share the same network policies and requirements.

* VTEP: Virtual Tunnel Edge

* VXLAN: Virtual Extensible LAN

4. EVPN Control Plane Extensions

In an EVPN VXLAN overlay network, a policy group tag may be assigned based on the MAC, IP, port, VLAN, etc, or a combination of the above. Similar to the MAC/IP addresses in the EVPN network, once the Policy Group ID is known for a local host/server/VM attached to an EVPN network, its Group Policy ID can be advertised to other Network Virtualization Edge (NVE) devices in the control plane through the Group Policy ID Extended Community. The scheme used for classification and allocation of Policy Group IDs used for GBP in an EVPN overlay network with VXLAN encapsulation is outside the scope of this document.

Policy group tag propagation in the EVPN/BGP control plane can be applied to the EVPN route Type-1 (Ethernet A-D per EVI) [RFC7432], route Type-2 (MAC/IP Advertisement route) [RFC7432], route Type-3 (Inclusive Multicast Ethernet Tag route - IMET) [RFC7432], or EVPN route Type-5 (IP Prefix route) [RFC9136]. If Policy Group ID is allocated for a MAC address, IP host or prefix address through the GBP classification scheme, then its Group Policy ID can be encoded through the Group Policy ID Extended Community and be advertised alongside its corresponding EVPN Type-2 or Type-5 route. If Policy Group ID is allocated for an Attachment Circuit (AC), then its Group Policy ID can be encodoed through the Group Policy ID extended community and be advertised alongside its Ethernet A-D per EVI route.

For the flows that the ingress VXLAN Tunnel Endpoint (VTEP) has learned its destination group policy tag through EVPN/BGP control plane signaling, the policy enforcement can thus be carried out at the ingress node. Otherwise, policy enforcement can be carried out

at the egress node. If policy enforcement is carried out at the head-end VXLAN tunnel, the ingress node MUST set the GBP applied bit, the A-bit as it is specified in Figure 2, to 1 in the VXLAN header before forwarding the traffic to the VXLAN tunnel. Otherwise, the ingress node sets the A-bit to 0 in the VXLAN header.

4.1. Group Policy ID Extended Community

The Group Policy ID BGP Extended Community is a new transitive Opaque Extended Community with a Type value of 0x03. This Extended Community may be advertised along with an EVPN Type-1, Type-2, Type-3, or Type-5 route. This new Opaque Extended Community enables the EVPN route it is attached to, to propagate the Group Policy ID used for Group Based Policy in the control plane.

When the "Uniformed-Propagation-Mode" is used under the EVPN and IPVPN interworking use case, the Group Policy ID Extended Community is carried over by the GW PE when a route for a given IP or IPv6 prefix is propagated from one domain to another with a different address family.

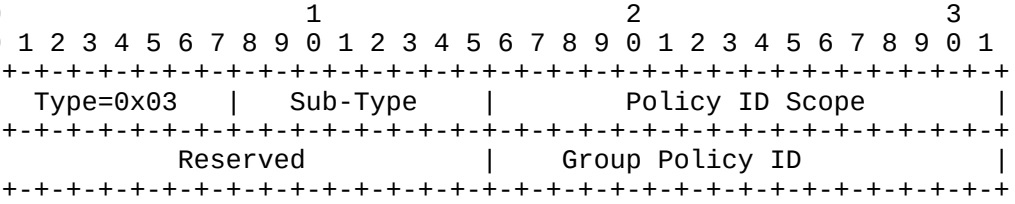

Figure 1: Group Policy ID Extended Community

Policy ID Scope: The Policy ID Scope field is 16-bit long, and is an optional field.

Group Policy ID (GPI): The GPI field is 16-bit long and it encodes the value of a Group Policy ID.

The reserved fields MUST be set to zero by the sender and ignored by the receiver.

If the Policy ID Scope is not set, any EVPN VXLAN NVE node that receives a route with a Group Policy ID may use the received value as is. If the Scope is set, a node that has the same locally configured Scope in the received route may use the received Policy ID value. A node that has a different local Scope than in the received route may need to translate the received Policy ID to a locally assigned value.

5. VXLAN Header Extension

This section defines the VxLAN header extension needed for Group Based Policy (VxLAN-GBP). Although in context of this document this VxLAN header extension is described with EVPN control plane, this header extension can be used with other control planes using VxLAN encapsulation such as [RFC9300] and [RFC9301].

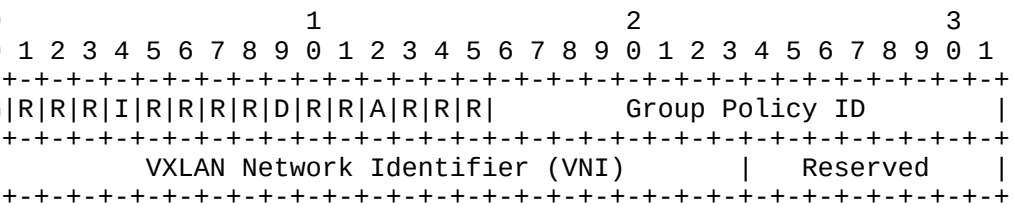

Figure 2: VXLAN Header Extension for GBP

The following bits are defined in addition to the existing VXLAN fields:

G Bit: Bit 0 of the initial word is defined as the G (Group Based Policy Extension) bit.

* G = 1 indicates that the source TSI Group membership is being carried within the Group Policy ID field as defined in this document.

* G = 0 indicates that the Group Policy ID is not being carried, and the G Bit MUST be set to 0 as specified in [RFC7348].

D bit: Bit 9 of the initial word is defined as the Don’t Learn bit. When set, this bit indicates that the egress VTEP MUST NOT learn the source address of the encapsulated frame.

A Bit: Bit 12 of the initial word is defined as the A (Policy Applied) bit. This bit is only defined as the A bit when the G bit is set to 1.

* A = 1 indicates that the group policy has already been applied to this packet. Policies MUST NOT be applied by devices when the A bit is set.

* A = 0 indicates that the group policy has not been applied to this packet. Group policies MUST be applied by devices when the A bit is set to 0 and the destination Group has been determined. Devices that apply the Group policy MUST set the A bit to 1 after the policy has been applied.

Group Policy ID: 16 bit identifier that indicates the source TSI Group membership being encapsulated by VXLAN. The allocation of Group Policy ID values is outside the scope of this document.

6. Backward Compatibility

When considering backward compatibility, we need to consider a network that can have a mix of all the following NVEs:

* NVEs that don’t support GBP in either data plane or control plane

* NVEs that support GBP in data plane but not control plane

* NVEs that support GBP in both data plane and control plane

VXLAN [RFC7348] requires reserved fields to be set to zero on transmit and ignored on receive. This ensures that the G bit will never be set by VXLAN NVEs and therefore packets received from these NVEs can be assigned to a default Group Policy ID. It also ensures that VXLAN NVEs receiving packets with the G bit set will ignore the Group Policy ID. Due to this defined behavior by VXLAN NVEs, it allows the extensions described in this document to operate on the IANA assigned VXLAN UDP port (port 4789).

In some environments, there may be a mix of devices supporting the VXLAN Group Based Policy Extension and devices that do not. Devices supporting the VXLAN Group Based Policy Extension SHOULD assign traffic arriving without the G bit set to a default Group Policy ID for the purposes of policy enforcement. This guarantees backward compatibility on the egress NVEs supporting GBP functionality to do proper egress filtering.

When an ingress NVE receives an EVPN route with Group Policy ID extended community, if it understands this extended community, it processes it and adds the Group Policy ID along with the corresponding MAC/IP route in its forwarding table. If it doesn’t understand this extended community, it processes the EVPN route as before.

Now, when a packet from a host arrives on an ingress NVE that can perform GBP ingress policy enforcement, it looks up the destination MAC/IP address, retrieves the destination Group Policy ID and takes the appropriate action based on the pair of Group Policy IDs of the source and the destination. If the ingress NVE cannot perform the GBP ingress policy enforcement, it carries out the packet processing in accordance to the previous paragraph.

The ingress NVE may also use the presence of the Group Policy ID Extended Community to determine the capability of the advertising NVE to process the VXLAN Group Based Policy Extension and whether to include it in packets being sent towards the egress NVE.

7. Interconnecting multiple EVPN VXLAN domains

EVPN VXLAN based deployments may comprise of multiple EVPN networks, domains or sites.
In such cases, a VXLAN overlay may extend from an ingress node to an egress node across different domains; or it may be divided into multiple stitched overlay segments that are interconnected via DCI through gateway devices.

In this document, we simply refer to each EVPN network or site as a EVPN domain or domain for short unless it is explicitly specified otherwise.

From a control plane point of view, border GWs in each domain may learn routes of other domains either via direct peering sessions or via a set of external route reflectors.

In such deployments, the allocation and management of Group Policy IDs may be done independently in different domains, and consequently the allocated values scoped to each domain. Therefore, when a group policy tag is signaled with routes to a different domain, the tag needs to be translated to a value local to the receiving domain before it can be used in a group based policy at an ingress node in that domain.

A domain may receive routes from multiple sender domains. In order to facilitate simpler and flexible application of translation policies regardless of the deployed overlay design or control plane peering model, the advertised Policy ID may also carry with it a Policy ID Scope which identifies the allocation domain. Any suitable BGP node in the route distribution path can then consistently translate a received Policy ID based on the scope.

Scope assignment is done by the administrator or orchestration system managing the multi-domain deployment. The exact mechanism is out of the purview of this document.

8. EVPN Interworking with IPVPN

In the EVPN interworking use case as it is specified in the [I-D.ietf-bess-evpn-ipvpn-interworking], two or more EVPN networks/ domains are interconnected by a layer-3 IP-VPN network with VPN-IPv4/ VPN-IPv6 BGP address families. To support ingress policy enforcement, the Policy Group ID Extended Community needs to be propagated by the GW PEs sitting at the border of an EVPN domain and IP-VPN domain from one domain to another.

For the Uniform-Propagation-Mode defined in the [I-D.ietf-bess-evpn-ipvpn-interworking], when propagating an EVPN IP prefix route across the domain boundary to IP-VPN network, the Gateway PE SHOULD propagate communities, extended communities and large communities except for all the EVPN extended communities. The Policy Group ID Extended Community defined in this document is a new transitive Opaque Extend Community. It is not subject to stripping at the GW PE when the Uniform-Propagation-Mode is used, and SHOULD be propagated.

9. Acknowledgements

The authors would like to thank Ayan Banerjee, Tom Edsall, Thomas Graf, Jeffrey Zhang, and Jeff Haas for their careful review and valuable feedbacks. The authors also would like to thank Prasad Miriyala and Selvakumar Sivaraj for their contributions.

10. Security Considerations

This document describes an extension to VXLAN to carry the Group Policy Identifier of the source endpoint. These identifiers must be distributed to participating VTEPs that are encapsulating traffic from the endpoints sourcing traffic. The control plane protocol should ensure that these identifiers are securely distributed to the network elements participating in the policy enforcement domain.

Additionally, the Group Policy Identifier field being carried in the packet directly impacts the network policy applied to the traffic.

There is a risk that these identifiers may be spoofed and proper integrity protection should be put in place to ensure that these fields can only be populated by trusted entities. Due to the importance of these fields, confidentiality may also be required to ensure that traffic cannot be targeted for attack based on the policy identifiers.
In some environments, these attacks are mitigated through physical security. In other environments, traditional security mechanisms like IPsec that authenticate and optionally encrypt VXLAN traffic including the bits and fields described in this document.

Furthermore, when EVPN control plane is used, then all the security considerations in [RFC7432] and [RFC7623] apply directly to this document because it leverages the control and data plane procedures described in those documents.

11. IANA Considerations

For the Group Policy ID Extended Community defined in this document, IANA has allocated the following codepoint in the Sub-type registry of Type 0x03 Transitive Opaque Extended Community.

Sub-Type	Name	Reference
0x17	Group Policy ID Extended Community	[this document]

This document requests IANA for the following allocations from the two-octet registry called "VxLAN Flags" in the VxLAN header.

Field	Name	Reference
Bit 0	G bit	[this document]
Bit 9	D bit	[this document]
Bit 12	A bit	[this document]

Furthermore, this document requests IANA for the allocation of the 16-bit reserved field in the first word of the VxLAN header as Group Policy ID.

Field	Name	Reference
Bits 16-31	Group Policy ID	[this document]

12. References

12.1. Normative References

[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>.

[RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <https://www.rfc-editor.org/info/rfc7432>.

[RFC7623] Sajassi, A., Ed., Salam, S., Bitar, N., Isaac, A., and W. Henderickx, "Provider Backbone Bridging Combined with Ethernet VPN (PBB-EVPN)", RFC 7623, DOI 10.17487/RFC7623, September 2015, <https://www.rfc-editor.org/info/rfc7623>.

[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>.

[RFC9136] Rabadan, J., Ed., Henderickx, W., Drake, J., Lin, W., and A. Sajassi, "IP Prefix Advertisement in Ethernet VPN (EVPN)", RFC 9136, DOI 10.17487/RFC9136, October 2021, <https://www.rfc-editor.org/info/rfc9136>.

12.2. Informative References

[RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014, <https://www.rfc-editor.org/info/rfc7348>.

[RFC9300] Farinacci, D., Fuller, V., Meyer, D., Lewis, D., and A. Cabellos, Ed., "The Locator/ID Separation Protocol (LISP)", RFC 9300, DOI 10.17487/RFC9300, October 2022, <https://www.rfc-editor.org/info/rfc9300>.

[RFC9301] Farinacci, D., Maino, F., Fuller, V., and A. Cabellos, Ed., "Locator/ID Separation Protocol (LISP) Control Plane", RFC 9301, DOI 10.17487/RFC9301, October 2022, <https://www.rfc-editor.org/info/rfc9301>.

Authors' Addresses

Wen Lin
Juniper
Email: wlin@juniper.net

Dhananjaya Rao
Cisco
Email: dhrao@cisco.com

Ali Sajassi
Cisco
Email: sajassi@cisco.com

Lawrence Kreeger
Arrcus
Email: lkreeger@gmail.com

Jorge Rabadan
Nokia
Email: jorge.rabadan@nokia.com