

Workgroup: BESS WG  
Published: 21 March 2022  
Intended Status: Standards Track  
Expires: 22 September 2022  
Authors: Y. Wang  
ZTE Corporation

## **Centerlized EVPN Prefix Advertisement for Common Prefixes behind Different CEs**

### **Abstract**

In [Section 5.8](#) of [[I-D.wang-bess-evpn-arp-nd-synch-without-irb](#)], centerlized RT-5 advertisement are used for common prefixes behind different CEs, This draft describes the requirements for such scenarios. Then this draft reuse the procedures defined in [Section 6.2.2](#) of [[I-D.wz-bess-evpn-vpws-as-vrf-ac](#)] to support this scenario.

### **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 22 September 2022.

### **Copyright Notice**

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

## Table of Contents

- [1. Introduction](#)
  - [1.1. ECMP for Centerlized RT-5 Advertisement](#)
  - [1.2. Terminology](#)
- [2. Requirements](#)
- [3. Solution](#)
  - [3.1. Basic Control Plane Procedures](#)
    - [3.1.1. Centerlized CE-BGP](#)
    - [3.1.2. RT-2E Advertisement from PE1/PE2 to DGW1](#)
    - [3.1.3. RT-5G Advertisement from DGW1 to PE1/PE2/PE3/DGW1'](#)
    - [3.1.4. RT-2E Advertisement between PE1 and PE2](#)
- [4. Security Considerations](#)
- [5. IANA Considerations](#)
- [6. Normative References](#)
- [7. Informative References](#)
- [Author's Address](#)

## 1. Introduction

In [Figure 1](#), Both R1 and R2 advertise their prefixes to DGW1 according to [Section 5](#) of [[I-D.wang-bess-evpn-arp-nd-synch-without-irb](#)]. Subnet SN1 can only be reached through R1, so we say SN1 is R1's exclusive prefix. Then subnet SN2 is R2's exclusive prefix, because SN2 can only be reached throug R2. But subnet SN21 can be reached either through R1, or through R2, thus we say SN21 is a common prefix of R1 and R2.

When there are both exclusive prefixes and common prefixes behind some CEs, some special requirements should be considered, especially when some of these CEs will not aware which prefixes are the common prefixes. This draft describes the requirements and solutions related to these scenarios .

### 1.1. ECMP for Centerlized RT-5 Advertisement

R1 and R2 both establish a single CE-BGP session with DGW1. These CE-BGP session can be called the centerlized CE-BGP session.

SN21 is a common prefix of R1 and R2, when R4 send traffic to a host inside subnet SN21, DGW1' should load-balance that traffic between PE1, PE2 and PE3.

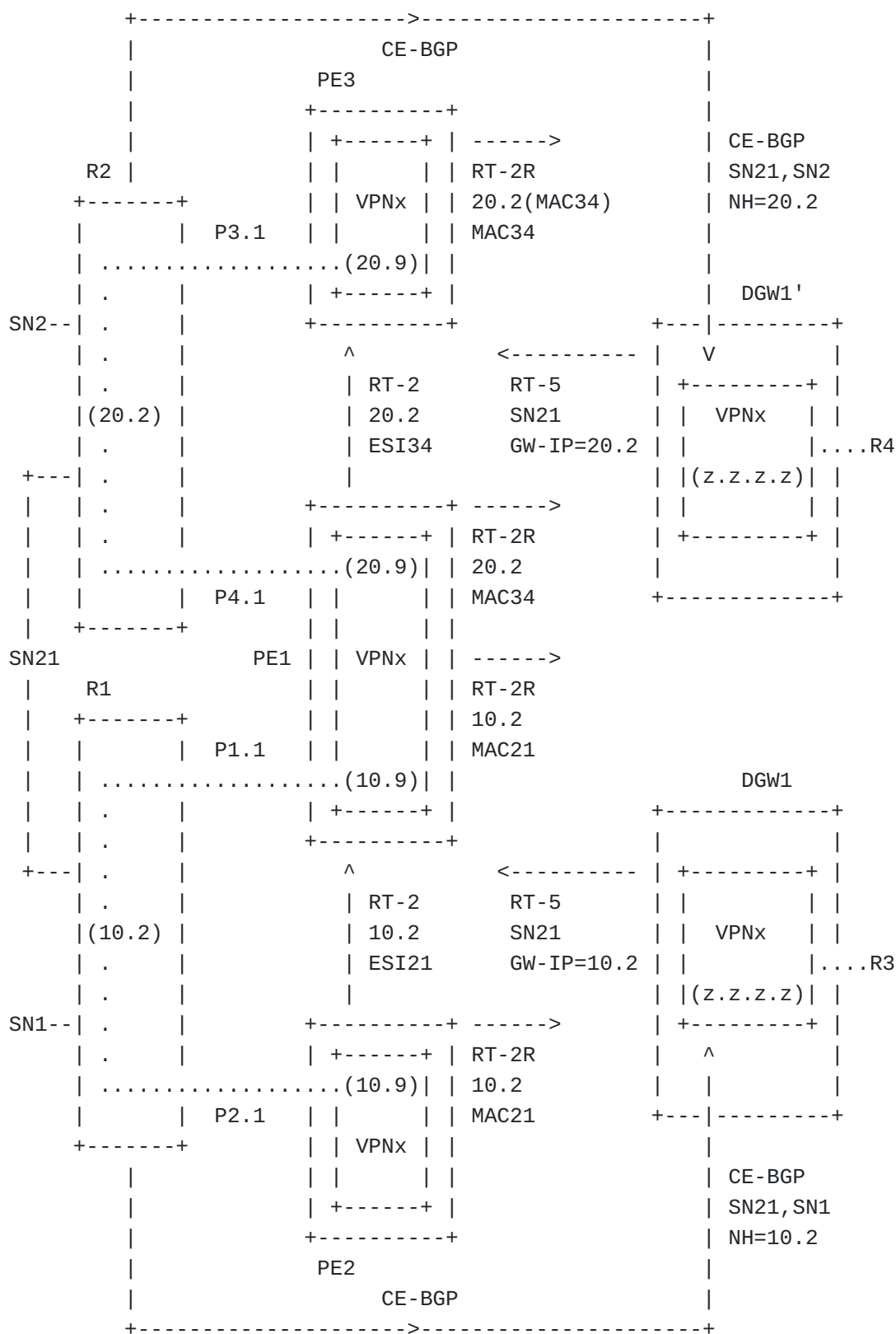


Figure 1: Centerlized RT-5 for common CE-Prefixes

Note that we just use centralized CE-BGP session to discover CE-prefixes, but we still expect a distributed Layer 3 forwarding framework.

## 1.2. Terminology

Most of the terminology used in this documents comes from [[RFC7432](#)] and [[RFC9136](#)] except for the following:

- \* **L3 EVI:** An EVPN instance spanning the Provider Edge (PE) devices participating in that EVPN which contains VRF ACs and maybe contains IRB interfaces or IRC interfaces.
- \* **CE-BGP:** The BGP session between PE and CE. Note that CE-BGP route doesn't have a RD or Route-Target.
- \* **RMAC:** Router's MAC, which is signaled in the Router's MAC extended community.
- \* **RT-2R:** When a MAC/IP Advertisement Route is used in the context of an IP-VRF, it is called as a RT-2R in this draft.
- \* **RT-5E:** An EVPN Prefix Advertisement Route with a non-reserved ESI.
- \* **RT-5G:** An EVPN Prefix Advertisement Route with a zero ESI and a non-zero GW-IP.
- \* **RT-5L:** An EVPN Prefix Advertisement Route with both zero ESI and zero GW-IP, but a valid MPLS label.
- \* **Internal Remote PE:** When PEx is called as an EVPN route ERy's internal remote PE, that is saying that, PEx is on the ES which is identified by ERy's ESI field. When ERy's SOI is not zero, that is aslo saying that PEx has been attached to the ethernet tag which is identified by the <ESI, SOI>.
- \* **External Remote PE:** When PEx is called as an EVPN route ERy's external remote PE, that is saying that, PEx is not on the ES which is identified by ERy's ESI field. When ERy's SOI is not zero, PEx may aslo be a PE which has not been attached to the ethernet tag which is identified by the <ESI, SOI>.
- \* **CE-Prefix:** When an IP prefix can be reached through CEx from PEy, that IP prefix is called as PEy's CE-prefix behind CEx in this

draft. PEy's CE-prefix behind CEx is also called as PEy's CE-prefix for short in this draft.

- \* **Common CE-Prefix:** When an CE-Prefix can be reached through either CEy or CEz from PEy, in this draft, it is called as a common CE-Prefix of CEy and CEz, from the viewpoint of PEy.
- \* **Exclusive CE-Prefix:** When an CE-Prefix of PEy can be reached through CEy, and it can't be reached through other CEs of PEy, it is called as an exclusive CE-Prefix of CEy, from the viewpoint of PEy.
- \* **SNGW:** Sub-Net-specific Gate Way IP address, the SNGW of a subnet is an IP address which is used by the hosts of that subnet to be the nexthop of the default route of these host.
- \* **Overlay nexthop :** The CE-Prefix's nexthop IP address which is in the address-space of the L3 EVI.
- \* **Original Overlay nexthop :** The overlay nexthop which is advertised by the CE through a PE-CE route protocol.

## 2. Requirements

Before advertise SN1/SN2/SN21 to DGWs, R1 and R2 don't have to know which prefix is their common prefix, and which prefix is their exclusive prefix.

## 3. Solution

### 3.1. Basic Control Plane Procedures

#### 3.1.1. Centerlized CE-BGP

The CE-BGP session between R1 and DGW1 is established between 10.2 and z.z.z.z. The IP address 10.2 is called the uplink interface address of R1 in this document. The IP address z.z.z.z is called the centerlized loopback address of VPNx in this document. The IP address 10.9 is called the downlink VRF-interface address of PE1/PE2 in this document.

R1 advertises a BGP route for a prefix (say "SN21") behind it to DGW1 via that CE-BGP session. The nexthop for SN21 is R1's uplink interface address (say 10.2).

R2 advertises a BGP route for a prefix (say "SN21") behind it to DGW1' via that CE-BGP session. The nexthop for SN21 is R2's uplink interface address (say 20.2).

Note that the data packets from R1(R2) to the centralized loopback address may be routed following the default route on R1(R2). Thus DGW1 doesn't need to use the CE-BGP session to advertise prefixes of VPNx to R1(R2).

### **3.1.2. RT-2E Advertisement from PE1/PE2 to DGW1**

When PE1 and PE2 learns the ARP entry of 10.2, it advertises a RT-2R route to DGW1 (and DGW1').

When PE1 and PE3 learns the ARP entry of 20.2, it advertises a RT-2R route to DGW1 (and DGW1').

### **3.1.3. RT-5G Advertisement from DGW1 to PE1/PE2/PE3/DGW1'**

When DGW1 receives the SN21 from the CE-BGP session. The nexthop for SN21 is 10.2. So DGW1 advertises a RT-5G route to PE1/PE2/PE3 for SN21. The GW-IP value of the RT-5G route for SN21 is 10.2.

When DGW1' receives the SN21 from the CE-BGP session. The nexthop for SN21 is 20.2. So DGW1 advertises a RT-5G route to PE1/PE2/PE3 for SN21. The GW-IP value of the RT-5G route for SN21 is 20.2.

DGW1 and DGW1' may be the same device, in such case DGW1 should use the ADD-PATH of [[RFC7911](#)] to advertise two GW-IPs for the same prefix SN21.

Note that when other PEs receive these RT-5 route for SN21, the ECMP behavior is already defined in [Section 4.1](#) of [[RFC9136](#)] as the following:

- \* Based on the BD-10 Route Target in DGW1 and DGW2, the IP Prefix route is also imported, and SN1/24 is added to the IP-VRF with Overlay Index IP2 pointing at the local BD-10. In this example, it is assumed that the RT-5 from NVE2 is preferred over the RT-5 from NVE3. If both routes were equally preferable and ECMP enabled, SN1/24 would also be added to the routing table with Overlay Index IP3.

Figure 2: ECMP for GW-IP based RT-5

### **3.1.4. RT-2E Advertisement between PE1 and PE2**

The RT-2R routes advertisement between PE1 and PE2 is used to sync subnet 10.0's ARP entries to each other in order to avoid ARP missing. The ESI Value of these two RT-2R routes is ESI21.

The RT-2R routes advertisement between PE1 and PE3 is used to sync subnet 20.0's ARP entries to each other in order to avoid ARP missing. The ESI Value of these two RT-2R routes is ESI34.

#### 4. Security Considerations

TBD.

#### 5. IANA Considerations

There is no IANA consideration needed.

#### 6. Normative References

[I-D.wang-bess-evpn-arp-nd-synch-without-irb] Wang, Y. and Z. Zhang, "ARP/ND Synching And IP Aliasing without IRB", Work in Progress, Internet-Draft, draft-wang-bess-evpn-arp-nd-synch-without-irb-08, 1 September 2021, <<https://datatracker.ietf.org/doc/html/draft-wang-bess-evpn-arp-nd-synch-without-irb-08>>.

[I-D.wz-bess-evpn-vpws-as-vrf-ac] Wang, Y. and Z. Zhang, "EVPN VPWS as VRF Attachment Circuit", Work in Progress, Internet-Draft, draft-wz-bess-evpn-vpws-as-vrf-ac-02, 28 August 2021, <<https://datatracker.ietf.org/doc/html/draft-wz-bess-evpn-vpws-as-vrf-ac-02>>.

[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>>.

[RFC7911] Walton, D., Retana, A., Chen, E., and J. Scudder, "Advertisement of Multiple Paths in BGP", RFC 7911, DOI 10.17487/RFC7911, July 2016, <<https://www.rfc-editor.org/info/rfc7911>>.

[RFC8214] Boutros, S., Sajassi, A., Salam, S., Drake, J., and J. Rabadan, "Virtual Private Wire Service Support in Ethernet VPN", RFC 8214, DOI 10.17487/RFC8214, August 2017, <<https://www.rfc-editor.org/info/rfc8214>>.

[RFC8365] Sajassi, A., Ed., Drake, J., Ed., Bitar, N., Shekhar, R., Uttaro, J., and W. Henderickx, "A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)", RFC 8365, DOI 10.17487/RFC8365, March 2018, <<https://www.rfc-editor.org/info/rfc8365>>.

[RFC9135] Sajassi, A., Salam, S., Thoria, S., Drake, J., and J. Rabadan, "Integrated Routing and Bridging in Ethernet VPN

(EVPN)", RFC 9135, DOI 10.17487/RFC9135, October 2021,  
<<https://www.rfc-editor.org/info/rfc9135>>.

[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>>.

## 7. Informative References

### [I-D.sajassi-bess-evpn-ip-aliasing]

Sajassi, A., Badoni, G., Warade, P., Pasupula, S.,  
Krattiger, L., Drake, J., and J. Rabadan, "EVPN Support  
for L3 Fast Convergence and Aliasing/Backup Path", Work  
in Progress, Internet-Draft, draft-sajassi-bess-evpn-ip-  
aliasing-04, 7 March 2022, <[https://datatracker.ietf.org/  
doc/html/draft-sajassi-bess-evpn-ip-aliasing-04](https://datatracker.ietf.org/doc/html/draft-sajassi-bess-evpn-ip-aliasing-04)>.

### Author's Address

Yubao Wang  
ZTE Corporation  
No. 68 of Zijinghua Road, Yuhuatai District  
Nanjing  
China

Email: [wang.yubao2@zte.com.cn](mailto:wang.yubao2@zte.com.cn)