January 2022

OPTIMIZED MULTI-HOMING IN INTER-AS OPTION B FOR ETHERNET VIRTUAL PRIVATE NETWORKS

Satya Mohanty

Mankamana Mishra

Ali Sajassi

OPTIMIZED MULTI-HOMING IN INTER-AS OPTION B FOR ETHERNET VIRTUAL PRIVATE NETWORKS

AUTHORS:
Satya Mohanty
Mankamana Mishra
Ali Sajassi

ABSTRACT

It is becoming increasingly common to stretch Ethernet Virtual Private Network (EVPN) services across Inter-Autonomous System (Inter-AS) boundaries. In such Inter-AS deployments, providers do not like to expose loopbacks of a provider edge (PE) router's loopbacks in one AS to the other AS, since this typically ensures less maintenance overhead and also allays security concerns. Inter-AS Option B for EVPN is ideally suited for such deployments as the next-hops of the EVPN routes for Inter-AS Option B are reset at the AS Border Router (ASBR), so the receiving AS does not know the loopback addresses of an advertising AS. This proposal provides a new technique to enable Inter-AS Option B for EVPN implementations that involves minimal configuration at ASBRs, which, in some instances, may enable multi-path and/or Prefix-Independent Convergence (PIC) -type load balancing at ASBRs.

DETAILED DESCRIPTION

Consider an illustrative EVPN example, as shown in Figure 1, below, in which a Customer edge (CE) device is reachable via provider edge routers, PE1 and PE2.

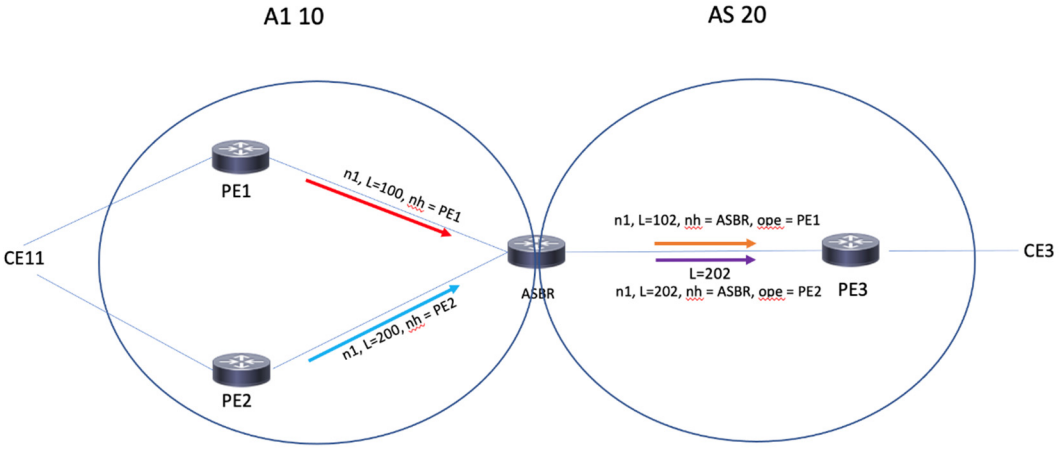

Figure 1: Example EVPN Architecture

Published by Technical Disclosure Commons, 2022

Due to local learning on an Ethernet Segment (ES) from Address Resolution Protocol (ARP) requests for the architecture of Figure 1, a corresponding type-2 EVPN route (IP1, MAC1) in the context of an EVPN instance (EVI), say EVI1, is created at both PE1 and PE2 with key ‘n1’and advertised to the ASBR from PE1 and PE2 respectively.

The Route Distinguishers (RD) of the type-2 routes are 1:1 and 2:2 respectively. The route keys are distinct, so they are treated as separate nets at the ASBR and then advertised to PE3 where they are brought together into the interested EVI. Simultaneously, there will be corresponding type-1 per Ethernet Segment-Auto-Discovery (ES-AD) (mass withdraw route) and per-EVI AD routes that are advertised. As the next-hops are reset at the ASBR, PE3 will not be able to determine from the Border Gateway Protocol (BGP) paths of these nets from which PE the type-1 or the type-2 routes originated. This means that since the Media Access Control (MAC) addresses and ES-AD routes obtained at the receiving AS cannot be correlated, the procedures of the mass withdraw as described in Request For Comments (RFC) 7432 can no longer be relied upon.

One potential technique through which to solve this problem is described in https://datatracker.ietf.org/doc/html/draft-heitz-bess-evpn-option-b-01 via the Originator PE (OPE) sub-Type Length Value (sub-TLV) in which the OPE-essentially encodes the identity of the Originating Router (IP Address) in the tunnel encapsulation attribute. In that case, it becomes very easy for the receiver PE, PE3 to unambiguously identify from which PE the EVPN advertisement was sent.

However, there is one scenario in which neither RFC 7432 nor the above Internet Engineering Task Force (IETF) draft involving the ability to correlate MAC addresses and ES-AD routes at the receiving AS will help. For example, considering the architecture of Figure 1, assume that the OPE is present so that PE3 can unambiguously infer whether the given net, 'n1', originated from PE1 or PE2. However, in case of a node failure of PE1, PE3 can deduce that PE1 went down only after PE3 receives the corresponding withdraw of the ES-AD route of PE1. PE3 can then switch the traffic to be re-directed to PE2 by imposing the out-going label '202' to the traffic. Because the loopbacks of either AS are not redistributed into the other, the convergence is subject to BGP convergence, and any ingress protection schemes will not be optimal.

This proposal seeks to addresses this problem by providing a solution that achieves traffic re-direction with minimal delay. Figure 2, below, illustrates example details associated with this solution, which involves bringing two paths of a network together.

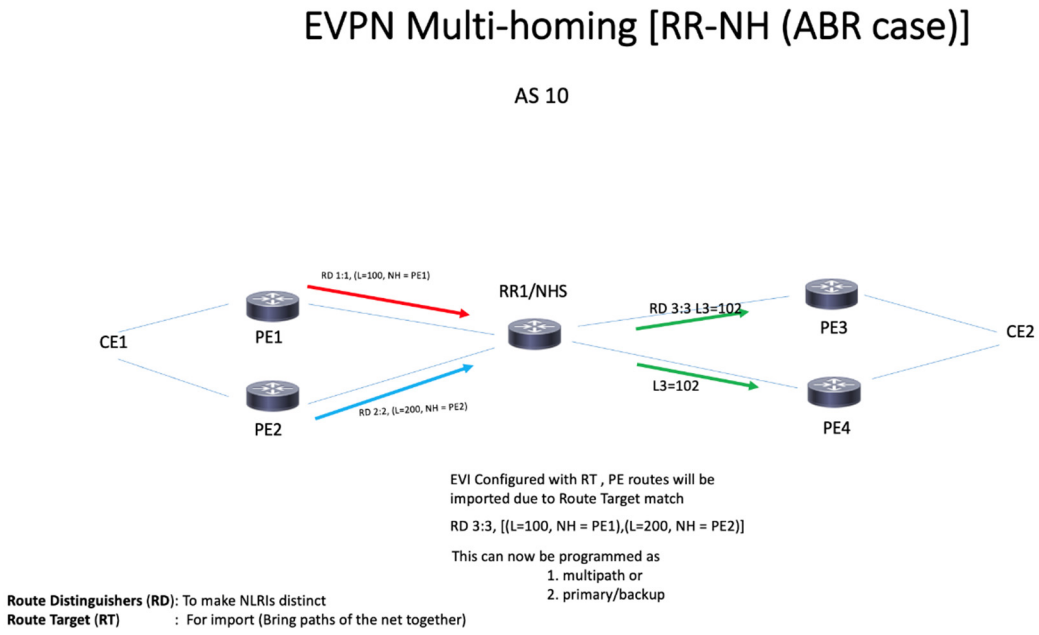

Figure 2: Solution Architecture

To achieve the solution, a basic EVI configuration is utilized, which is configured with import Route-Targets that are the same as those carried in the type-2 routes (This just means that the EVI from where these routes are sourced is configured with the same Route-Target). The multicast EVPN optimization procedures provided in draft-ietf-bess-evpn-bum-procedure-updates-08 are modelled on Virtual Private LAN Service (VPLS) and Multicast VPN (MVPN) Inter-AS operations and Section 9.2 of RFC 6514 already requires configuration of Route-Targets. Thus, the requirement to provide EVI and Route-target configuration at the ASBR is not in conflict with existing procedures.

Once the two paths of the network are brought together, the individual next-hops and received labels as well as the EVI-AD and ES-AD routes are processed by the EVPN module, which already has the logic to determine whether to program these received next-hops and labels as either Equal-Cost Multipath (ECMP) routes (to facilitate Active/Active load balancing) or Active/Backup routes (depending on the mode on which the Ethernet Segment is operating).

As per the mode, a decision is made and a local label is allocated by label logic for the ASBR and the hardware (e.g., FIB) is programmed appropriately. The best-path of the imported network is advertised in the control plane to the other AS with the local label. To facilitate forwarding when packets arrive with the local label, corresponding flows will be load balanced (ECMP) to PE1 or sent only to the Active next-hop in case of Single-Active mode of operation.

If PE1 goes down, the next-hop becomes inaccessible at the ASBR following an Interior BGP (IBGP) notification. In the Active/Active case, the surviving member of the ECMP Forwarding Equivalence Class (FEC) points to PE2. In the Active/Standby case, the standby is immediately promoted to Active. Thus, in either case, the traffic path can be repaired fairly quickly. Contrasting this solution with instances in which ingress protection may be provided at PE3, such ingress protection will rely on the mass-withdraw route. Further, if instead of the remote PE not being in the adjacent AS, but rather several ASes away, the superiority of the solution provided by this proposal over the proposed IETF draft solution becomes more apparent. Accordingly, this proposal provides a new technique to enable Inter-AS Option B for EVPN implementations that involves minimal configuration at ASBRs, which, in some instances, may enable multi-path and/or PIC-type load balancing at ASBRs.