Description

EVPN gateway support for all-active (A-A) multihoming adds a new redundancy model to our multi-domain EVPN solution introduced in [1]. This deployment model introduces the concept of a WAN Interconnect Ethernet Segment identifier (WAN I-ESI). The WAN I-ESI allows the gateway's EVPN neighbors to form L2 and L3 overlay ECMP on routes re-exported by the gateways. The identifier is shared by gateway nodes within the same domain (site) and set in MAC-IP routes that cross domain boundaries.

The operational semantics for the WAN I-ESI are the same as for a standard ethernet segment defined in [4]. In other words, no additional BGP/EVPN protocol extensions are required. The advertisement of Type-1 Auto Discovery and Type-4 Ethernet Segment routes is the same, the designated forwarder election process is the same, and the rules for local bias are the same. The only difference is that while traditional ethernet-segments represent the interface on which the route has been learned, this **virtual** ethernet segment represents the domain on which the route has been learned (i.e. local domain or remote domain).

Platform compatibility

- 7280R3 series
- 7500R3 series
- 7800R3 series

Feature History

Release	Update
4.31.0	Initial introduction
4.32.0F	• Support for domain identifier and loop detection with domain path • Support for directly attached hosts on the gateway (multi-homed and orphan)

Topology

The following diagram (Figure 1) provides a visual representation of three data centers (sites), where DC-A and DC-B provide redundancy via an all-active multihoming model, each using its own WAN I-ESI. Note that in other redundancy solutions, such as MLAG or the anycast gateway model, gateways within the same domain MUST share the same VTEP address. With A-A multihoming, gateways are configured with unique VTEP addresses.

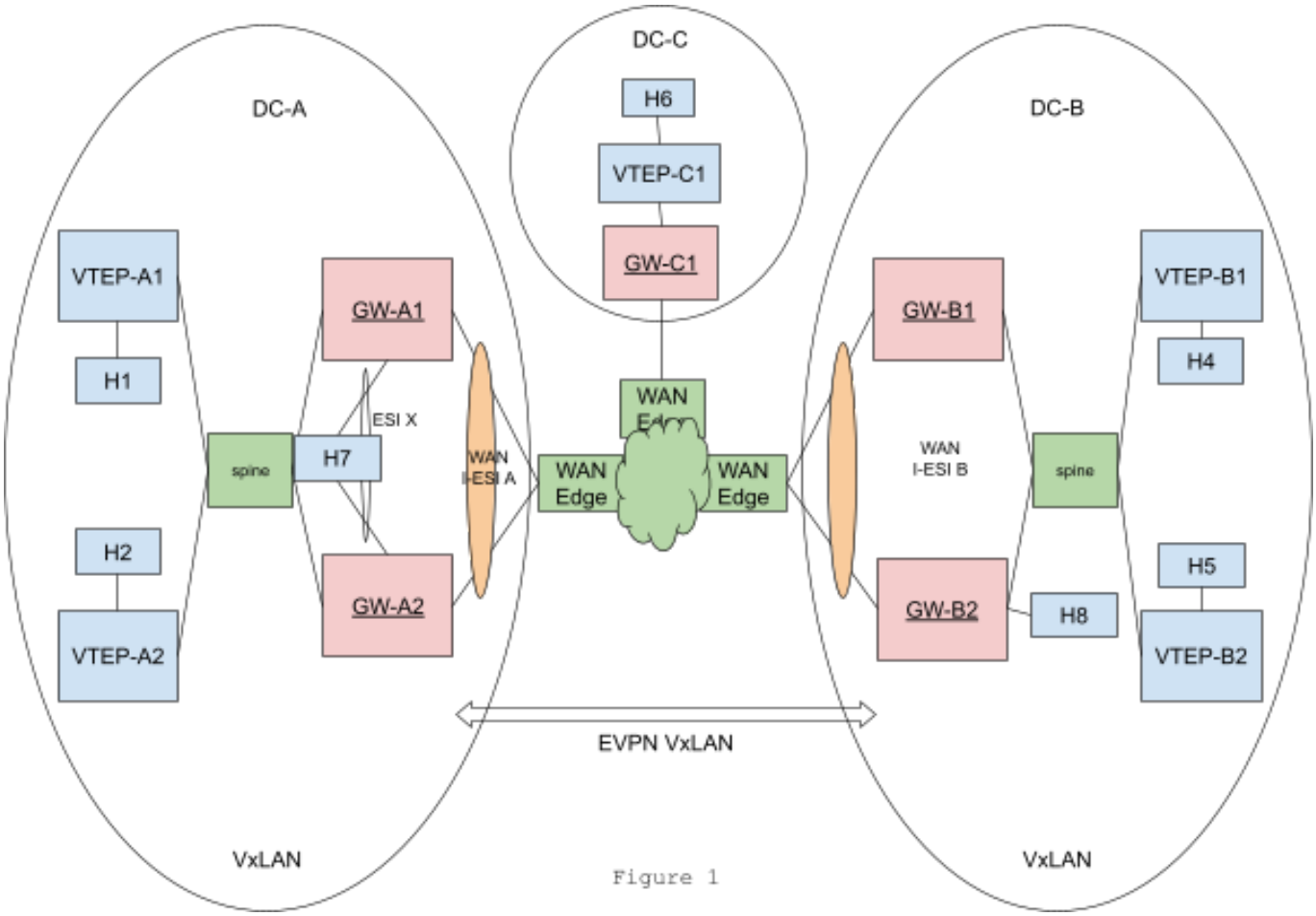

Figure 1

From the perspective of a local domain VTEP in DC-A (e.g. VTEP-A1), hosts in DC-B and DC-C will appear to be multihomed behind an Ethernet Segment (I-ESI A) attached to GW-A1 and GW-A2. Figure 2 is a representation of how local domain VTEPs see remote hosts, with the dotted lines indicating how, on the local domain VTEP, the remote hosts appear behind both GW-A1 and GW-A2.

In release 4.32.0, support for directly connected hosts attached to the gateway (H7, H8) was added. Directly connected multi-homed hosts on the gateway do not use the I-ESI and instead are configured as a native ethernet-segment under the interface configuration.

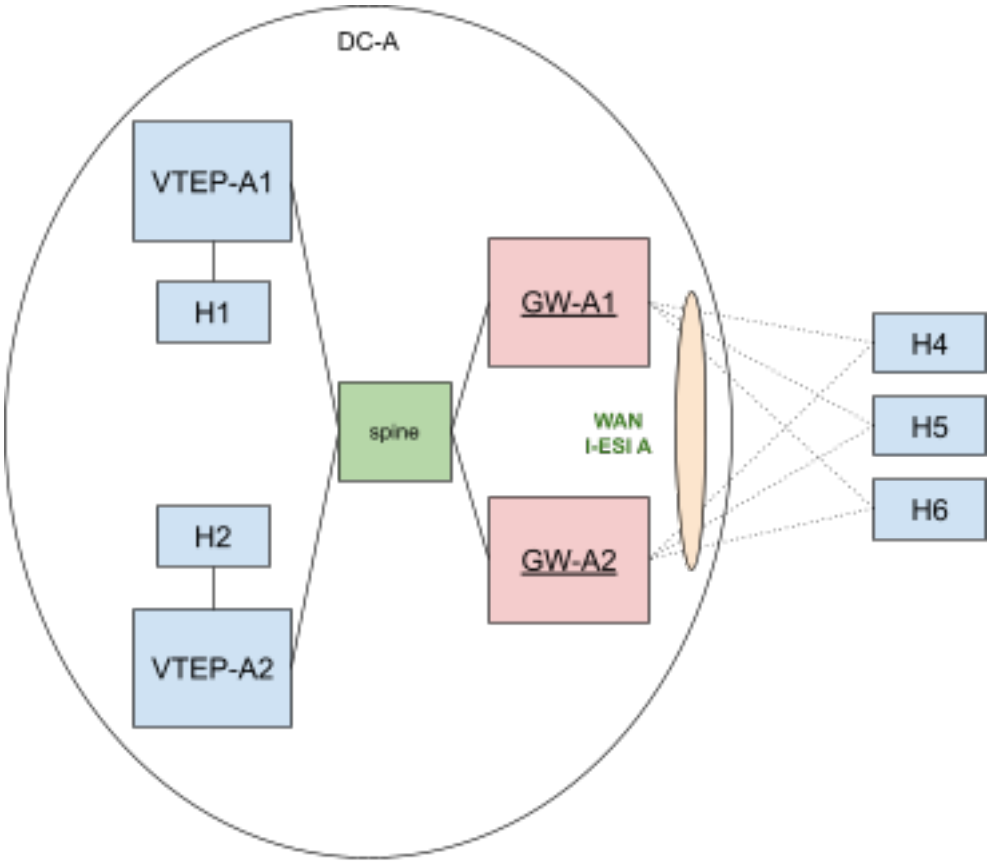

Figure 2

EVPN Peering Models

Within a site, the devices can have EVPN sessions using iBGP full-mesh, iBGP with a route reflector, or eBGP. Similarly, the gateways can have EVPN sessions using the same configurations.

Configuration

The following sample shows the configuration on GW-A1:

```
GW-A1(config-router-bgp)# show active
router bgp 64512
   ...
   maximum-paths 4 ecmp 4
   bgp bestpath d-path
   ...
   !
   vlan 10
      rd evpn domain all 10.255.1.1:10
      route-target import export 64500:10
```

```
   route-target import export evpn domain remote 64501:10
   redistribute learned
!
vrf red
   rd 10.255.1.1:0
   route-target import evpn 64500:20000
   route-target export evpn 64500:20000
   router-id 10.255.1.1
...
address-family evpn
   neighbor WAN-RR activate
   neighbor WAN-RR domain remote
   neighbor SPINE-RR activate
   domain identifier 10:1
   domain identifier 10:2 remote
   ...
   !
   evpn ethernet-segment domain all
      identifier 0011:1111:1111:1111:1111
      route-target import 00:01:00:01:00:01
   !
   layer-2 fec in-place update
   !
!
```

In the sample configuration, GW-A1 peers with a route reflector SPINE-RR in the local domain and peers with a route reflector WAN-RR in the remote domain.

As previously stated, the WAN I-ESI represents the domain on which we've learned the host bindings. For example, type-2 routes learned in the local domain, when re-exported, will be advertised into the remote domain using the WAN I-ESI. Similarly, routes in the remote domain, when re-exported, will be advertised into the local domain using the WAN I-ESI.

In addition to the new ESI configuration, it is still necessary to still configure a remote domain route-target under the mac-vrf stanza, a remote domain route-distinguisher under the mac-vrf stanza, and to configure the neighbor as remote domain as indicated in [1].

As of 4.32.0, for route loop detection, domain identifiers along with the domain bestpath step must be configured for each domain [6]. The same identifier must be used for the same domain on all the gateways in one site. This configuration will prevent re-exported routes by one gateway from being installed on the other gateway. For releases prior to 4.32.0, it is required to configure an RFC rule to reject mac-ip routes from the peer, the RFC rule described should be augmented to also reject mac-ip routes.

If the gateways of one site exchange routes in both domains, the following RCF rule must be configured. The purpose of this RCF rule is to reject IMET routes received by the gateway peer in the remote domain, as it is required that the peer gateway exists in the local domain. For this configuration, the gateways in one DC should use the same community, but the community used should be different from the other DC's communities.

```
GW-A1(config-router-bgp)# show active
router bgp 64512
   ...
   address-family evpn
      ...
      neighbor WAN-RR rcf in evpnDciMhBlockGwRx()
      neighbor WAN-RR rcf out evpnDciMhBlockGwTx()
      ...
      !
   !
router general
   control-functions
      code
       function evpnDciMhBlockGwTx() {
          if evpn.route_type is EVPN_IMET
          {
              community add ;
           }
           return true;
        }

        function evpnDciMhBlockGwRx() {
           return community has_none ;
        }
      EOF
```

For L3 ECMP to form, the ***maximum-paths*** CLI must also be configured on all the VTEPs in the deployment.

Platform TCAM Profile

In addition to the TCAM rule of ***vxlan routing*** required for Multi-Domain EVPN VXLAN, additional user defined TCAM profile needs to be configured in CLI for this feature specifically on 7280R3/7500R3/7800R3 platforms.

For this feature to work, a TCAM feature "***tunnel vxlan multi-homing domain remote***" is required in

the TCAM profile. However, there are not enough TCAM resources to fit this feature when MLAG egress filtering is also enabled. So it is also required to disable MLAG egress filtering by configuring the TCAM feature "***mlag egress filter disabled***" in the TCAM profile.

vxlan-multihoming-dci

Designated Forwarder and Local Bias Semantics

The designated forwarder election process for the WAN I-ESI is the same as for a traditional Ethernet segment. The local bias rules are also the same, though they have been extended to support virtual interfaces (VXLAN tunnels). For BUM traffic originating from the local domain, all gateways will receive a copy, but only the designated forwarder will bridge the traffic to remote domain VXLAN tunnels. For traffic originating from the remote domain, all gateways will receive a copy, but only the designated forwarder will bridge in the local domain. Gateway VXLAN tunnels in the local domain are identified as terminating at VTEPs that share the WAN I-ESI with our gateway node.

Failover and Redundancy

For this feature, *full box gateway failures*, which refer to situations where the entire gateway device reloads or becomes non-operational, are natively supported. From Figure 1, if GW-A1 goes down or loses connectivity to all other devices, all cross-DC traffic will now be sent through GW-A2 only.

However, *partial gateway failures*, which refer to situations where the gateway loses connectivity to a single domain (either remote or local), require additional configuration. An event handler may be used to disable the interconnect ESI when all the links to either the local domain or the remote domain go down. Below is an example of the event handler that accomplishes such task:

```
event-handler spineandwan
   trigger on-intf Ethernet1/1,2/1,9/1,10/1 operstatus
   trigger on-config disabled
   action bash
      if [ $OPERSTATE_0 == "linkdown" ] && [ $OPERSTATE_1 == "linkdown" ]
      then
         python3 /mnt/flash/esishut.py
      elif [ $OPERSTATE_2 == "linkdown" ] && [ $OPERSTATE_3 == "linkdown" ]
      then
         python3 /mnt/flash/esishut.py
      else
         python3 /mnt/flash/esiunshut.py
      fi
   EOF
   delay 1
```

In this example ***$OPERSTATE_0*** and ***$OPERSTATE_1*** refer to the link status of the spine connection while ***$OPERSTATE_2*** and ***$OPERSTATE_3*** refer to the link status of the WAN connections. When all the spine links go down, the event manager will call esishut.py, which will change the interconnect ethernet-segment to a shutdown state:

```
evpn ethernet-segment domain all
   identifier 0011:1111:1111:1111:1111
   route-target import 00:01:00:01:00:01
   disabled
!
```

The python script to disable the I-ES (***esishut.py***) when a link is up is defined here:

```
#!/usr/bin/env python3
# Copyright (c) 2024 Arista Networks, Inc.  All rights reserved.
# Arista Networks, Inc. Confidential and Proprietary.

import time
import Cell
import PyClient
import Tac

Tac.run( [ "FastCli", "-A", "-p", "15", "-c", " enable \n config \n router bgp 212 \n
 address-family evpn \n  evpn ethernet-segment domain all  \n disabled " ] )
```

The python script to enable the I-ES (***esiunshut.py***) is defined here:

```
#!/usr/bin/env python3
# Copyright (c) 2024 Arista Networks, Inc.  All rights reserved.
# Arista Networks, Inc. Confidential and Proprietary.

import time
import Cell
import PyClient
import Tac

Tac.run( [ "FastCli", "-A", "-p", "15", "-c", " enable \n config \n router bgp 212 \n
 address-family evpn \n  evpn ethernet-segment domain all  \n no disabled " ] )
```

More information about event handler can be found in the eos user manual [5].

Note that this handler should be used in the event of losing underlay connectivity to either the local domain or the remote domain. If the topology can prevent partial failures, through an underlay backup path or some other means, then that is preferred.

Show Commands

The following command shows the designated forwarder elected for the WAN I-ESI for VLAN 10 on GW-A1.

```
GW-A1#show bgp evpn instance vlan 10
EVPN instance: VLAN 10
  Route distinguisher: 10.255.1.1:10
  Route distinguisher remote: 10.255.1.1:10
  Route target import: Route-Target-AS:64500:10
  Route target export: Route-Target-AS:64500:10
  Route target import remote: Route-Target-AS:64501:10
  Route target export remote: Route-Target-AS:64501:10
  Service interface: VLAN-based
  Local VXLAN IP address: 10.255.1.1
  VXLAN: enabled
  MPLS: disabled
  Local ethernet segment:
    ESI: 0011:1111:1111:1111:1111
      Interface: Vxlan1
      Mode: all-active
      State: up
      ES-Import RT: 00:01:00:01:00:01
      DF election algorithm: modulus
      Designated forwarder: 10.255.1.1
      Non-Designated forwarder: 10.255.2.1
```

The Type-1 and Type-4 routes advertised by each gateway VTEP can be displayed with:

```
GW-A1#show bgp evpn route-type auto-discovery detail
...
GW-A1#show bgp evpn route-type ethernet-segment detail
...
```

The following command shows the MAC-IP advertised to the remote domain with ESI for the

WAN I-ESI:

```
GW-A1(config)#show bgp evpn route-type mac-ip 10.10.0.3 domain remote detail
BGP routing table information for VRF default
Router identifier 0.0.0.1, local AS number 300
BGP routing table entry for mac-
ip 2ae2.e1f7.b57c 10.10.0.3 remote, Route Distinguisher: 10.255.1.1:10
 Paths: 1 available
  Local
    - from - (0.0.0.0)
      Origin IGP, metric -, localpref 100, weight 0, valid, internal, best
      Extended Community: Route-Target-AS:64500:20000 Route-Target-
AS:64501:10 TunnelEncap:tunnelTypeVxlan
      VNI: 10010 L3 VNI: 20000 ESI: 0011:1111:1111:1111:1111
      D-PATH: 10:1:EVPN
```

Note that when GW-A1 advertises routes from the remote domain to the local domain, the ESI for the WAN I-ESI is also serialized:

```
GW-A1(config)#show bgp evpn route-type mac-
ip esi 0011:1111:1111:1111:1111 domain local detail
BGP routing table entry for mac-
ip 3eb6.d466.75b9 10.10.0.4, Route Distinguisher: 10.255.2.1:10
 Paths: 1 available
  400
    10.255.2.1 from 10.101.0.1 (0.0.3.1)
      Origin IGP, metric -, localpref 100, weight 0, valid, internal, best
      Extended Community: Route-Target-AS:64500:10 Route-Target-
AS:64500:20000 TunnelEncap:tunnelTypeVxlan EvpnRouterMac:00:00:78:05:00:00
      VNI: 10010 L3 VNI: 20000 ESI: 0011:1111:1111:1111:1111
      D-PATH: 10:2:EVPN
```

To confirm that L2 ECMP is formed over the VXLAN tunnel, run the following command:

```
GW-A1#show vxlan address-table address b6bd.9600.81d0
          Vxlan Mac Address Table
----------------------------------------------------------------------

VLAN  Mac Address     Type      Prt  VTEP             Moves   Last Move
----  -----------     ----      ---  ----             -----   ---------
  10  b6bd.9600.81d0  EVPN      Vx1  10.255.3.1       1       1:25:05 ago
```

```
                                    10.255.4.1
```

To confirm that L3 ECMP is formed over the VXLAN tunnel, the following command can be run.

```
GW-A1#show ip route vrf red 10.10.0.4
...
 B E      10.10.0.4/32 [200/0] via VTEP 10.255.4.1 VNI 20000 router-mac
                                    00:00:78:06:00:00 local-interface Vxlan1
                                via VTEP 10.255.3.1 VNI 20000 router-mac
                                    00:00:78:02:00:00 local-interface Vxlan1
```

Troubleshooting

- Verify all troubleshooting steps mentioned in [1]
- Verify that the configured ESI for the WAN I-ESI is consistent on the gateways and unique from any other ESI in the network
- Verify that the same designated forwarder is elected for the WAN I-ESI on the gateways in the same site
- Verify that L2 ECMP and L3 ECMP are formed as expected via VXLAN tunnels
- Verify that the TCAM profile has the following features configured:
    - ***feature tunnel vxlan multi-homing domain remote***
    - ***feature mlag egress filter disabled***

Limitations

- EVPN single-active multihoming is not supported for WAN I-ESI.
- Centralized routing model is not supported.
- V6 overlay and V6 underlay are not supported.
- Domain identifiers for directly attached hosts are not supported.

Resources

1. Multi-domain EVPN VXLAN TOI
2. EVPN VXLAN All-Active Multihoming TOI
3. RFC 9014: Interconnect Solution for Ethernet VPN (EVPN) Overlay Networks
4. RFC7432: BGP MPLS-Based Ethernet VPN
5. EOS User Manual
6. EVPN DCI GW Domain Path TOI