

eVIP on LSB Internetworking

Evolved Virtual IP

INTERWORK DESCRIPTION

Copyright

© Ericsson AB 2015. All rights reserved. No part of this document may be reproduced in any form without the written permission of the copyright owner.

Disclaimer

The contents of this document are subject to revision without notice due to continued progress in methodology, design and manufacturing. Ericsson shall have no liability for any error or damage of any kind resulting from the use of this document.

Trademark List

All trademarks mentioned herein are the property of their respective owners. These are shown in the document Trademark Information.



Contents

1	Introduction	1
1.1	Prerequisites	1
2	Connect Cluster to External Data Communication Networks	3
2.1	Containment of VIP Addresses	3
2.2	Front End	4
2.3	Interlink Networks	6
3	Configure eVIP with OSPF	9
3.1	OSPF between Cluster and eVIP Gateway Router	9
3.2	OSPF Supervision	10
3.3	OSPF Areas to Cluster	11
3.4	FEE Interface	12
3.5	FEE and OSPF Router ID	13
4	Configure eVIP with Static Routing and BFD	15
5	Interworking Rules, Recommendations, and Limitations	17
5.1	Rules	17
5.2	Recommendations	18
5.3	Limitation	20
6	eVIP Gateway Router to DCN	21
7	Examples of Networks with L2 Switches	23





1 Introduction

Evolved Virtual IP (eVIP) is used to connect a cluster to one or more external Data Communication Networks (DCNs). A concept for collective addressing with VIP addresses is used. With eVIP, a shared IP address is used to address distributed functions in a multi-processing cluster. The shared IP addresses are called VIP addresses.

This document describes the interfaces and basic configuration concepts. It is intended to be used for network engineering and is to be read as a complement to *eVIP on LSB Management Guide*.

1.1 Prerequisites

Ensure that the manufacturer documents for the eVIP gateway routers are available.





2 Connect Cluster to External Data Communication Networks

A cluster is connected to DCNs through eVIP gateway routers, see Figure 1.

At least two eVIP gateway routers are used to connect the cluster to external DCNs to protect against a situation of router failure. The routers forward incoming traffic with destination IP addresses, corresponding to VIP addresses, to the cluster. The VIP addresses are in the cluster configured by eVIP. The interfaces from the cluster to the routers are in the cluster configured by eVIP Front-End Elements (FEEs).

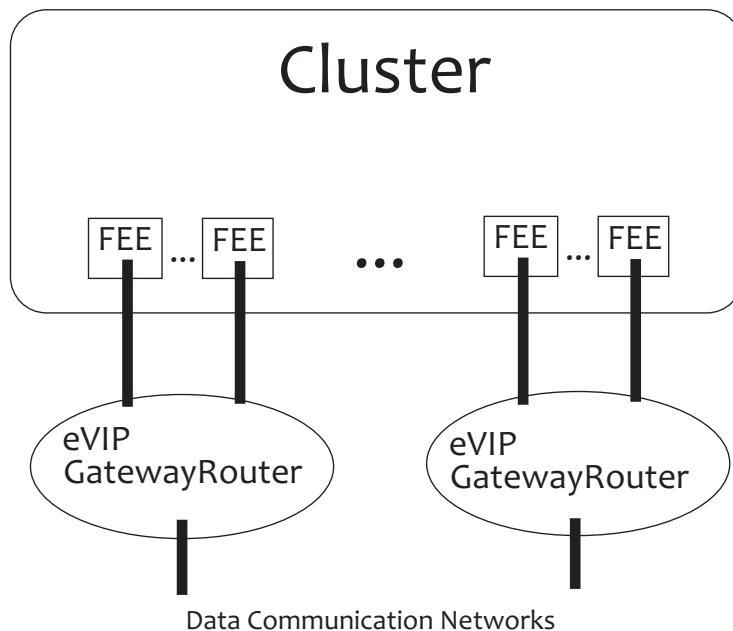


Figure 1 Cluster Connected to DCNs through eVIP Gateway Routers

2.1 Containment of VIP Addresses

VIP addresses are in eVIP configured to an Abstract Load Balancer (ALB), which is a technical term in eVIP used for configuring VIP addresses, external interfaces, and other eVIP resources. An ALB is a logical container used to scope configuration data and facilitate network separation. External network interfaces to a DCN are configured to an ALB.

Typically, separate DCNs are configured to separate ALBs, see Figure 2. An ALB is given a name when configured in runtime to identify the ALB, for example, O&M_Network_LB and Traffic_Network_LB.

Up to 8 ALBs can be configured in a cluster and each ALB must be given a unique name.

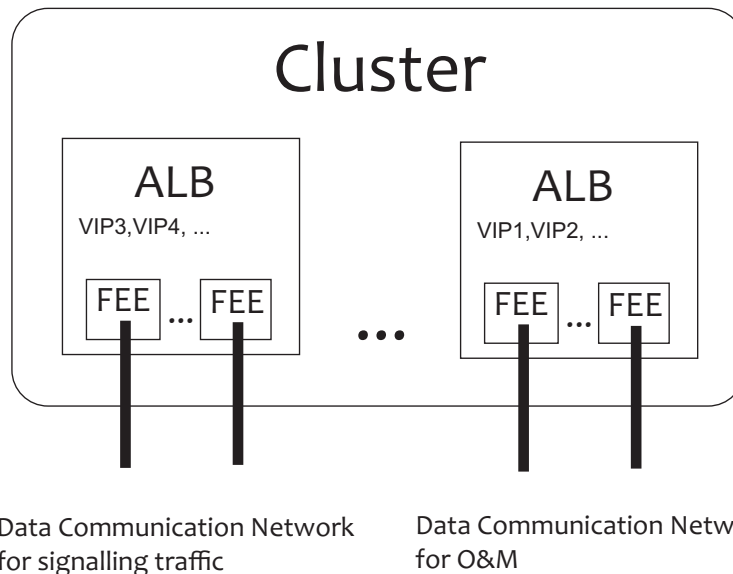


Figure 2 Cluster Connected to Two Separate DCNs

An ALB is configured with its own set of VIP addresses. The VIP addresses are virtual, that is, they are not configured to any particular physical interface or VLAN. The VIP addresses can be IPv4 or IPv6 addresses that are known to the external network, for example, public IP addresses. Each ALB can have a collection of IPv4 and IPv6 addresses configured as VIP addresses, for example, a set of non-contiguous addresses.

Note: VIP addresses from different ALBs must not overlap, that is, the same VIP addresses configured to one ALB must not be reused for another ALB.

2.2

Front End

The processing unit in the cluster with external eVIP interfaces is called “front-end processing unit”. The FEE instances are on the front-end blades. There is a one-to-one mapping between the external Layer 3 interfaces and the FEE instances. The external Layer 3 interfaces are used to interlink the FEEs with an eVIP gateway router. Each external Layer 3 interface of an FEE has an interface IP address, but these interface addresses are not VIP addresses. The external Layer 3 interfaces can reside in a VLAN or an untagged Ethernet network.

The FEEs are contained in an ALB. An ALB contains a set of FEE instances. Different FEE instances on the same blade can belong to different ALBs and then announce to the eVIP gateway routers the VIP addresses pertaining to their respective ALB, see Figure 3.

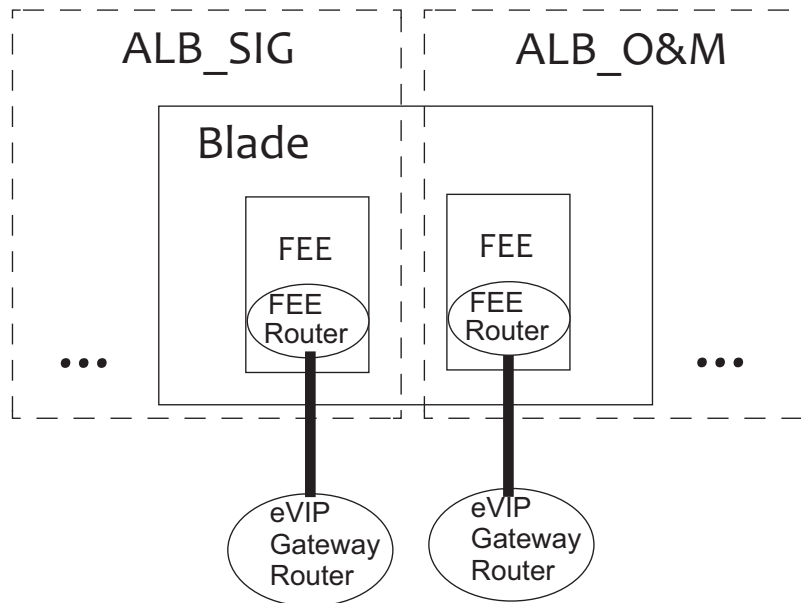


Figure 3 Blade with Two Interfaces Belonging to Different ALBs

At least two physical eVIP gateway routers are required to achieve redundancy. Dynamic routing is normally used in the internetworking between FEEs and eVIP gateway routers. The VIP addresses of an ALB are then announced to one or more eVIP gateway routers by a dynamic routing protocol OSPF. The OSPFv2 protocol is used for IPv4, but OSPFv3 is used for IPv6 VIP address announcements.

When FEEs are configured with router instances using OSPF, then the eVIP gateway routers regard the cluster as a collection of routers using the OSPF routing protocol. The eVIP gateway router therefore regards each connected FEE instance as a router.

The OSPF `hello` protocol takes care of the next hop supervision between FEEs and VIP gateway routers in both directions. The OSPF supervision can optionally (by configuration settings) be aided by the Bidirectional Forwarding Detection (BFD) protocol for more rapid link failure detection, see Section 3.2.1 OSPF Supervision with BFD on page 10.

The use of OSPF between FEEs and eVIP gateway routers is recommended, because the protocol prevents accidental configuration of routing loops and provides dynamic management of VIP addresses from the cluster.

2.2.1

Deployment of FEEs

In a physical, bare-metal environment there normally exists hardware constraints such as physical interfaces and cabling dictating where the FEEs can be deployed in the cluster.

In a virtualized environment, such as cloud, deployment constraints normally do not exist as the External Network interfaces are also virtualized. In the elastic

and scaling cloud environment, it is often desired to have a set of uniform processing units (Virtual Machines) making it possible to deploy FEEs on any of the processing units provided to the clustered application.

If the FEEs are on processing units that potentially can disappear during a scaling-in, there is a risk that eVIP can lose all external connectivity, causing a complete traffic outage.

To circumvent this, it is possible to configure FEEs as floating. This means that an FEE is automatically relocated to another processing unit if the current utilized processing unit disappears.

When relocating an FEE, the Layer 3 configuration is brought along. This means that each processing unit, that can accommodate a floating FEE, must have an External Network interface connected to the L3 network providing access to the eVIP Gateway Routers. This allows the FEE to re-establish connectivity with the eVIP Gateway Router when relocated.

During FEE relocation, the Layer 2 link breaks. If BFD link supervision is used, the disturbances are only minor as traffic is quickly relocated to other redundant links and FEEs. When the Layer 2 link is re-established, the BFD session is also re-established allowing the relocated FEE to carry traffic again.

2.3 Interlink Networks

The eVIP gateway routers are connected to the cluster through interlinking IP networks.

The interlinking networks can exist in directly connected cables or fiber between front-end interfaces and the eVIP gateway routers, that is, carried over dedicated physical Ethernet links, or the interlinking networks can be connected through an intermediary switch or switches.

An example of front-end blades interconnected with physical cables and one subnet per cable is shown in Figure 4.

For example configurations with switches, see Section 7 on page 23.

The interfaces on the cluster side front end, which are used for communication with the eVIP gateway routers, are called “external FEE interfaces”. These are so called Layer 3 interfaces and each external interface has an IP interface address.

The interlinking networks are configured in the FEEs and an eVIP gateway router. In an FEE this is done by configuring an IP subnet to a named bridged interface provided by Linux®, for example, as in a Linux Open Telecom Cluster (LOTIC) bundling.

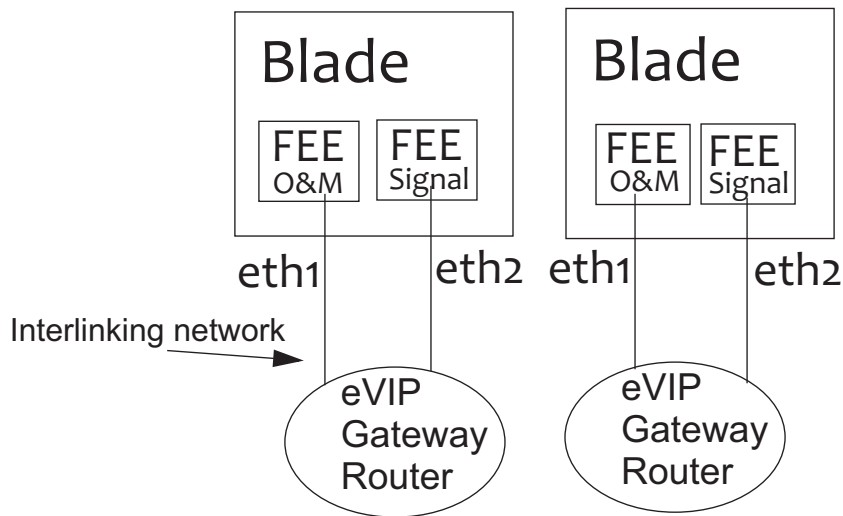


Figure 4 Front-End Blades Interconnected with Cables

An external eVIP interface can be part of a LAN or VLAN. The LANs or VLANs for the interlinking networks are mapped one-to-one to interlinking subnets. An external FEE interface is part of an interlinking subnet. There is a one-to-one mapping between external eVIP interfaces and FEEs, that is, for each FEE in eVIP there is only one external interface configured. A blade with four VLAN interfaces over two physical interfaces is shown in Figure 5.

Note: The VIP addresses configured to the ALBs are not part of the interlinking subnets. For example, the interlinking subnets are private IP addresses whereas the VIP addresses can be public IP addresses.

The external Layer 3 interfaces are typically, in the IPv4 case, part of small subnets used to interlink an eVIP gateway router with the corresponding FEE. In the IPv6 case, the interlinking is done at the link layer using IPv6, so called link local addresses. Conceptually, the OSPFv3 model (which is used for IPv6) is slightly different from OSPFv2. In the OSPFv3 case, the router is said to connect to the “link” and not to the “subnet”.

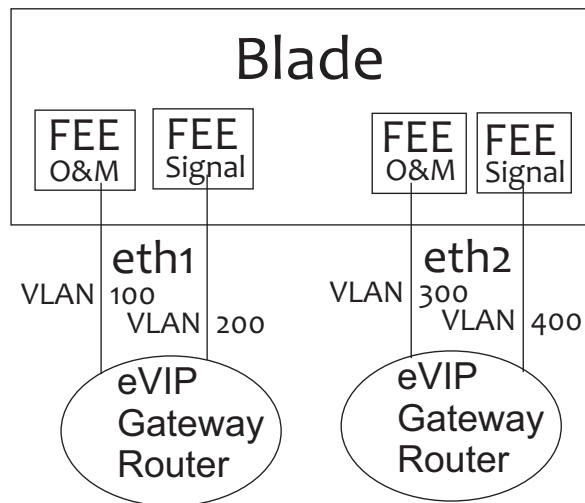


Figure 5 Blade with Four VLAN Interfaces over Two Physical Interfaces



3 Configure eVIP with OSPF

An eVIP gateway router is configured according to instructions provided by the manufacturer. When the eVIP gateway router is connected to the cluster, the eVIP function must be made aware of the gateway. This is done by configuring an FEE in eVIP for each Layer 3 interface on the cluster side that is used for interlinking the cluster to the gateway router, refer to *eVIP on LSB Management Guide*.

When configuring an FEE instance in eVIP, the following values must be known:

- Front-end node attached to the eVIP gateway router
- The name of the external eVIP interface on the front-end node
- The IP address of the eVIP gateway router interface
- OSPF parameters
- The ALB to which the FEE instance belongs

The front-end nodes and interfaces are identified by names provided by the used middleware or operating system, for example, LOTC.

The OSPF parameters include timer intervals, router ID, area ID, and a priority value. The OSPF parameters must correspond to the settings in the eVIP gateway router. Optionally, the BFD can be used with OSPF where the BFD parameters must correspond to the settings in the eVIP gateway router for stable behavior. For more information about the OSPF parameters, see Section 3.1 OSPF between Cluster and eVIP Gateway Router on page 9.

3.1 OSPF between Cluster and eVIP Gateway Router

The OSPF routing protocol is used between the cluster and the eVIP gateway router. For a description of the OSPF protocol, refer to [OSPF Version 2](#).

OSPF is used for the following tasks:

- Supervision of links and routers
- Redundancy switch-over for links and routers
- Establishment of equal cost paths to a VIP address destination in the cluster

The eVIP gateway router regards the cluster as a collection of OSPF neighbor routers. By eVIP the cluster starts a number of OSPF router instances, called FEE router instances, on the front-end nodes. The eVIP gateway router



automatically establishes OSPF adjacencies with the neighboring FEE router instances, which in this way become OSPF peers.

The FEE router instances use OSPF functionality for link supervision and to announce VIP addresses of an ALB to the eVIP gateway router. The announcement of VIP addresses to the eVIP gateway router attracts packet traffic to the announcing FEE router instances. In this way, packets with destination IP addresses matching the VIP addresses are attracted to interfaces of an ALB.

3.2 OSPF Supervision

For supervision, both the FEE router instances in the cluster and the eVIP gateway router periodically send out `hello` packets. Timer values are configurable. If there is a link failure, the eVIP gateway router “loses contact” with the corresponding FEE router inside the cluster and `hello` is not answered. OSPF updates the routing information so that incoming packets can be forwarded over other available links.

For outgoing packets, the FEE router instances in the cluster work in a similar way. The FEE router instances detect that contact is lost with the eVIP gateway router and eVIP ensures that the packets are forwarded over other available links belonging to the same ALB.

The timer values for the `hello` and `dead` intervals must be the same in the eVIP gateway router and the FEE router instance.

The default values for OSPF supervision are shown in Table 1.

Table 1 Default Values for OSPF Supervision

OSPF Parameter	OSPFv2 Default Value	OSPFv3 Default Value
<code>hello</code>	10 seconds	10 seconds
<code>dead</code>	40 seconds	40 seconds
<code>retransmit interval</code>	5 seconds	5 seconds
<code>transmit delay</code>	1 second	1 second
<code>router priority</code>	0 (zero)	0 (zero)

3.2.1 OSPF Supervision with BFD

Fast supervision of OSPF peers can be done by the BFD protocol. This requires that the router model used as eVIP gateway router at least supports BFD asynchronous mode with OSPF.



OSPF with BFD supervision works as follows:

- When an OSPF adjacency is established with a neighbor, OSPF activates the BFD software in the router, which starts to check reachability to the neighbor addresses.
- A BFD session is initiated and the sending of `hello` packets starts.
- When the BFD failure detection interval expires without receiving a `hello`, the OSPF software is informed so that the corresponding OSPF adjacency can be taken down and routes recalculated.

With the BFD, the failover time can be shortened compared to OSPF without the BFD.

A feature of the BFD is called “echo”. The BFD echo can optionally be configured in the FEE router. For an eVIP gateway router with many BFD sessions to answer, the processing load can be overwhelming for some router architectures. In this situation, the use of the echo function can reduce the processing load on the eVIP gateway router. When the echo function is enabled, the normal BFD `hello` packets are sent at a slower rate. Fast detection is still achieved by echo packets, which generally cost less to process but this depends on the router architecture in the eVIP gateway router. The BFD echo packets are transmitted so that they are echoed back from the interface and data plane of the opposite side, for example, the Network Interface Card (NIC).

Note: The timer values for the BFD and FEE router instance must be coordinated with settings in the eVIP gateway router.

For the BFD parameters and recommended settings, see Section 5.2 Recommendations on page 18.

3.2.2

SPF Exponential Hold Time Backoff

SPF runs when there is a topology change. The SPF algorithm is a very CPU intensive process. To prevent frequent SPF calculation the system waits for a period of time, called the hold time, and delay SPF calculations during network instability. SPF exponential backoff makes it possible to make the hold time variable change based on the frequency of topology changes. The hold time minimum and maximum periods can be configured using the `spf_delay` and `spf_interval` parameters. The hold time changes exponentially when a topology change occurs.

3.3

OSPF Areas to Cluster

OSPF has a concept of routing areas. The type of routing area used between the cluster and the gateway router must be a stub area. Stub area is a technical term used in OSPF.

Note: In a stub area, all routers must be configured to belong to a stub area, this includes FEE routers and eVIP gateway routers.

A single stub area is sufficient for the eVIP gateway purpose. However, it is possible to use several stub areas, see Figure 6. For example, separate stub areas can be used for the ALB serving separate DCNs.

An OSPF area has an identity that is given in the decimal dot notation format of an IP address. The area identity can correspond to an existing IP address or can be fictitious. However, within the stub area the router ID must be unique.

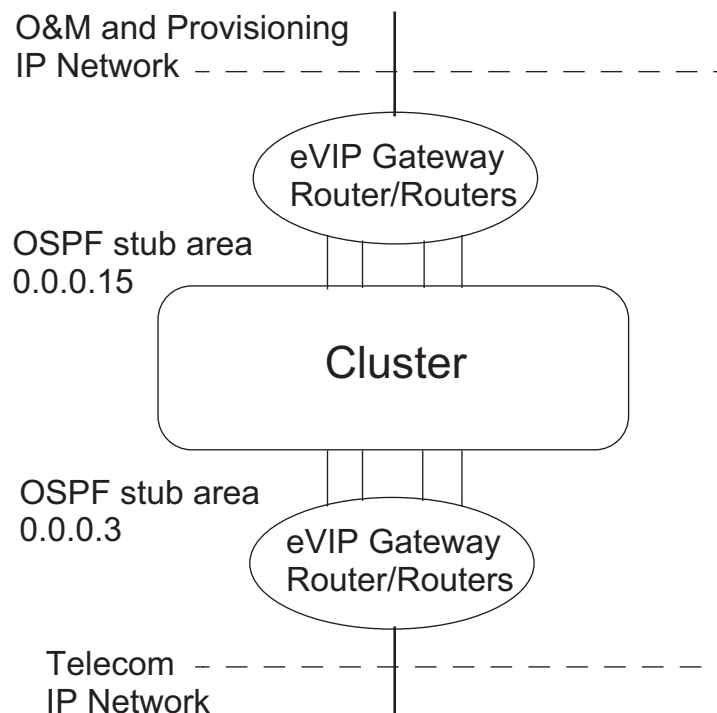


Figure 6 Cluster Connected to Different Networks with Different OSPF Stub Areas

3.4 FEE Interface

External FEE Layer 3 Interfaces have a local interface IP address on networks interlinking FEEs to an eVIP gateway router. Only one such Layer 3 interface can belong to an FEE instance. For example, a blade with two external Layer 3 interfaces connected to an eVIP gateway router requires two FEEs, one for each Layer 3 interface. The Layer 3 interfaces can in this example either be two VLANs or two (untagged) physically separate Ethernet interfaces.

A software router inside FEE is configured to each FEE instance for interworking with the eVIP gateway router. For each eVIP gateway router, there is at least one FEE router instance configured in eVIP. Each FEE router instance has its own router ID, see Figure 7.

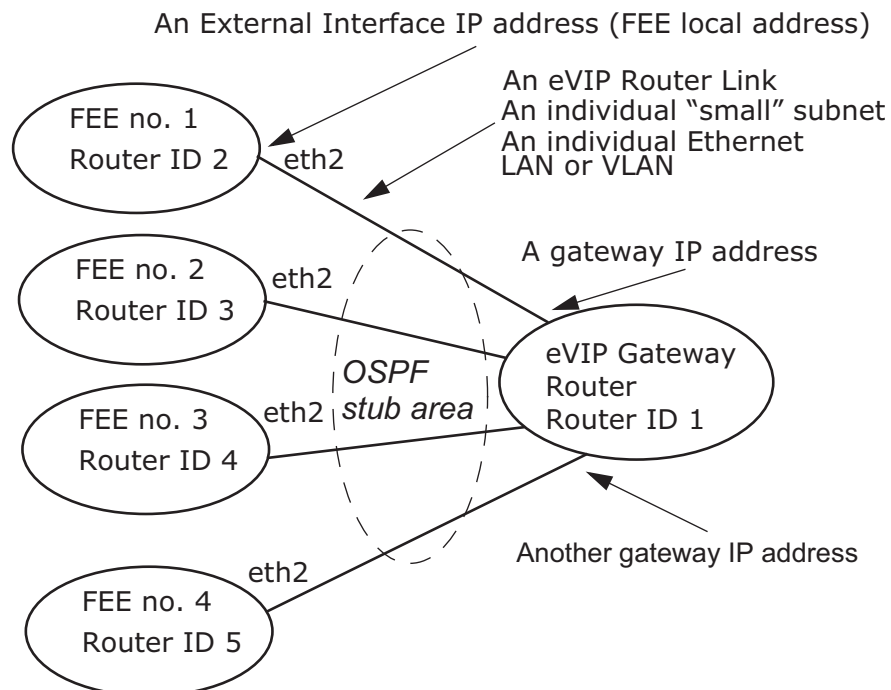


Figure 7 Typical Configuration with One Gateway Interface per eVIP Router Link

The VIP addresses are configured to an ALB, which is associated to one or more FEE instances. The VIP addresses are announced to the eVIP gateway router by OSPF. Packets arriving to the eVIP gateway router with destination IP addresses corresponding to one of these VIP addresses are forwarded by the eVIP gateway router to the cluster.

Usually, an ALB is configured with several FEE instances, which form a pool of alternative paths from the eVIP gateway router; this corresponds to so called Equal-Cost Multipaths (ECMPs) routing.

A default gateway is automatically injected by an OSPF Area Border Router (ABR), typically the eVIP gateway routers, into the stub area that is selected as gateway for the outgoing traffic.

3.5 FEE and OSPF Router ID

By convention, the router ID for an FEE instance in eVIP is specified to be identical to the IP address of the external eVIP interface connected to the eVIP gateway router. However, this is not strictly necessary. The router ID for a FEE instance in eVIP can, for example, be a fictitious IP address given in the decimal dot notation format.





4 Configure eVIP with Static Routing and BFD

Static routing with BFD supervision can be used between FEEs and the eVIP gateway router.

When static routing is used, static routes in the eVIP gateway router must be configured with VIP addresses of an ALB as route destination and the “interface addresses” on the FEE side as next-hop. If there is a connection failure between a FEE and an eVIP gateway, it is detected by BFD and then another available FEE is automatically selected.

When static routing is used, eVIP can select any eVIP gateway router with working connectivity to FEEs in the outgoing traffic case.

When static routing is used, as opposed to the case with the dynamic routing protocol OSPF, the selection of eVIP gateway router has no means to consider the network connectivity situation beyond the first-hop, that is, from the FEE to the eVIP gateway router.





5 Interworking Rules, Recommendations, and Limitations

This section specifies the interworking rules, recommendations, and limitations.

5.1 Rules

This section specifies the interworking rules for eVIP on LSB.

- | | |
|---------------|--|
| Rule 1 | eVIP gateway routers connected to traffic networks must be duplicated for redundancy. |
| Rule 2 | Each ALB must be configured with an individual name. |
| Rule 3 | If VLANs are used in networks interlinking an eVIP gateway router and FEEs, there must be a one-to-one correspondence between the VLANs and subnets. |
| Rule 4 | IPv6 routing must be configured between FEE and eVIP Gateway router, when IPv6 VIP addresses are used to external networks. |
| Rule 5 | <p>VIP addresses from different ALBs must not overlap, that is, the same VIP addresses configured to one ALB must not be reused for another ALB.</p> <p>Note: For outgoing traffic leaving the ALB. That is with source addresses being VIP addresses, the destination addresses can overlap in traffic from different ALBs. For example, destination addresses to remote parties in different VPNs can overlap. In such case, sockets must be bound to the corresponding ALB. This is done by arrangements in the application and then it follows that any default route must not be configured on the PNs where this application software is deployed.</p> |
| Rule 6 | The VIP addresses configured to the ALBs must not be part of the interlinking subnets configured in the FEEs. |
| Rule 7 | eVIP gateway routers must not use OSPF authentication towards the cluster. Specify no authentication in the eVIP gateway routers. |
| Rule 8 | Priority 0 (zero) must be configured to the OSPF instances of the FEE routers of the cluster. |



- | | |
|----------------|--|
| Rule 9 | eVIP gateway routers must not use a priority value equal to 0 (zero). Specify a positive integer greater than 0 (zero). |
| Rule 10 | The OSPF router ID in an eVIP gateway router must not be equal to the IP address of any internal interface in the cluster. |
| Rule 11 | The timer values for the <code>hello</code> and <code>dead</code> intervals must be the same in the eVIP gateway router and the FEE router instance. |
| Rule 12 | The OSPF <code>hello</code> interval must be set to a value greater than or equal to 3 seconds. If the interval is set to less than 3 seconds, false alarms can occur. Default is 10 seconds. |
| Rule 13 | The router <code>dead</code> interval must be set to a value greater than twice the <code>hello</code> interval. Recommended value for <code>dead</code> interval is three times the <code>hello</code> interval. Default is 40 seconds. |
| Rule 14 | The type of routing area used between the cluster and the gateway router must be a stub area. |
| Rule 15 | In a stub area, all routers must be configured to belong to a stub area, this includes FEE routers and eVIP gateway routers. |
| Rule 16 | Within a stub area, the router ID must be unique. Use unique router IDs all across a cluster. |
| Rule 17 | Static routes can only be configured with BFD; static routes without BFD are not supported. |

5.2 Recommendations

This section specifies the interworking recommendations for eVIP.



Recommendation 1

In general, use OSPF between FEEs and eVIP gateway routers, because of the following:

- The protocol prevents accidental configuration of routing loops.
- The protocol provides dynamic management of VIP addresses.
- The protocol is inherently a stable robust protocol.

BFD can further be used with OSPF for rapid link failure detection.

Static routing can only be used together with BFD.

However, static routing with BFD (in both directions between FEEs and an eVIP gateway router, without OSPF) is a permissible configuration in cases where requirements on network design do not require dynamic routing end-to-end. In these scenarios, it is assumed that the network engineering of the network beyond the eVIP gateway router is robust enough such that it can avoid situations of undesirable blackholing of packets.

Recommendation 2

The BFD parameters and the recommended settings are shown in Table 2.

Recommendation 3

Utilizing the FEE floating scheme, see Section 2.2.1 Deployment of FEEs on page 5, can disrupt dynamic routing protocols such as OSPF. The relocation of an FEE is seen as a routing instance disappearing for a short period and thus triggering routing updates to be sent by the peers. Such updates can spread across the entire External Network causing route-flapping.

Therefore, in network scenarios where a set of FEEs are required to use OSPF, this particular set of FEEs is configured as “fixed element” FEEs.

Table 2 BFD Parameters and Recommended Settings

BFD Parameter	Recommended BFD Setting
echo	off
bfd interval	200
minrx	200
multiplier	5



5.3 Limitation

The interworking limitation for eVIP on LSB is as follows:

- A system limit of maximum 8 ALBs can be configured in a cluster.
- The number of internetworking links an eVIP system can support is limited by the number of FEEs that can be provisioned. For information on the number of supported FEEs, refer to Section *Constraints* in *eVIP on LSB System Architecture Description*.



6 eVIP Gateway Router to DCN

From the eVIP gateway router to the DCNs, the interworking options are many and depend on the network design and the capabilities of the router used as eVIP gateway router. It is a wide field beyond the scope of this document to be covered in detail. Typically the following methods are used:

- Border Gateway Protocol (BGP) routing.

BGP routing requires that eVIP addresses (routes) are redistributed from the OSPF stub area into the BGP. Access Lists (ACLs), or so called route maps, are used to filter out routes that are not to be seen in the BGP autonomous system, for example, the interlinking networks between an eVIP gateway router and the cluster.

- OSPF routing.

OSPF routing in the external network can be done either over the area 0, the backbone area, or, if the router used as eVIP gateway router is capable thereof, redistribute eVIP addresses from the stub area to a non-backbone area of a separate OSPF Autonomous System. Redistribution must only be done in one direction, that is, from the stub area to the external Autonomous System.

- Static routing.

Static routing with alternative routes configured with different metrics or ECMPs can be used.

- Generally (beyond the scope of eVIP), regardless of the routing protocols or configuration, use BFD to supervise connections between nodes in the network end-to-end. This increases the robustness of the overall network and provide faster fault detection.





7

Examples of Networks with L2 Switches

This section shows the following example configurations with switches:

- Configuration with L2 switches, four VLANs, and two subnets, see Figure 8.
- Configuration with L2 switches, two VLANs, and two subnets, see Figure 9.
- Configuration with bonded Ethernet external interfaces, see Figure 10.

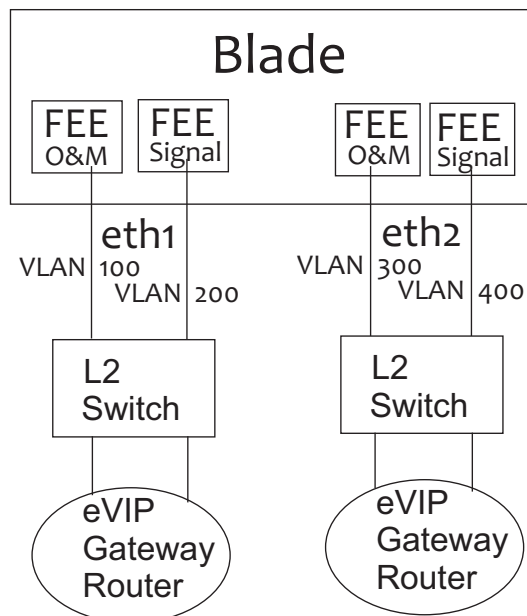


Figure 8 Configuration with L2 Switches, Four VLANs, and Two Subnets

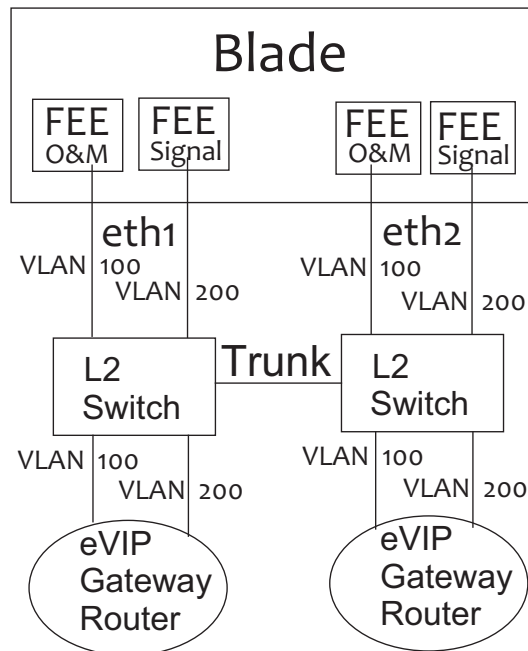


Figure 9 Configuration with L2 Switches, Two VLANs, and Two Subnets

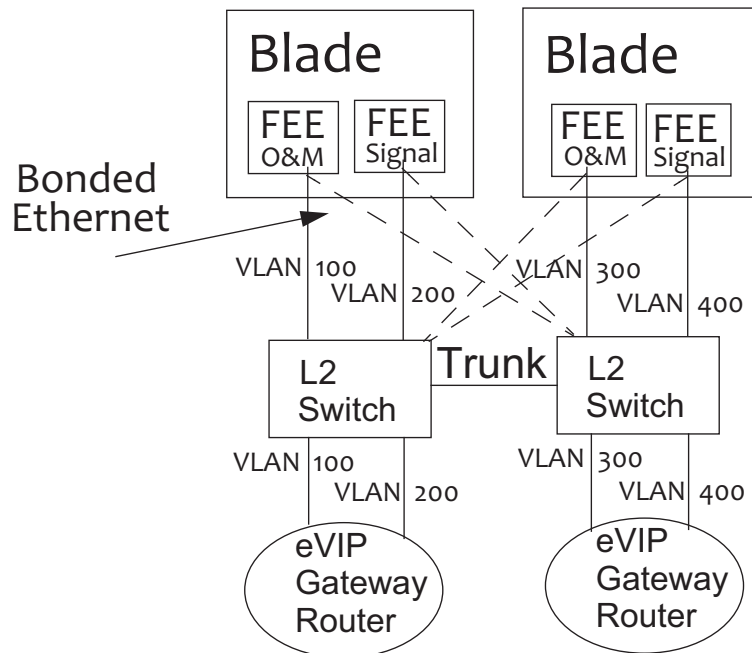


Figure 10 Configuration with Bonded Ethernet External Interfaces

Note: An FEE can only have a single external Layer 3 interface. However, a Layer 3 interface can have more than one underlying Layer 2 interface, for example, bonded Layer 2 interfaces.