Troubleshoot ACI VMM Integration

Contents

Introduction

This document describes steps to understand and troubleshoot ACI Virtual Machine Manager Integration (VMM).

Background Information

The material from this document was extracted from the Troubleshooting Cisco Application Centric Infrastructure, Second Edition book, specifically the **VMM Integration - Overview, VMM Integration - vCenter Connectivity, VMM Integration - Host Dynamic Discovery** and **VMM Integration - Hypervisor Uplink Load Balancing** chapters.

Virtual Machine Manager Overview

ACI controllers have the capability to integrate with third-party virtual machine managers (VMMs).

This is one of the key features of ACI as it simplifies and automates operations for end-to-end networking configuration of the fabric and to workloads that connect to it. ACI offers a single overlay policy model that can be extended across multiple workload types, i.e. virtual machines, bare metal servers, and containers.

This chapter will specifically focus on some typical troubleshooting scenarios related to the VMware vCenter VMM integration.

The reader will walk through:

- Investigation on vCenter communication faults.
- Host and VM dynamic discovery process and failure scenarios.
- Hypervisor Load Balancing algorithms.

vCenter connectivity

Role-Based Access Control (RBAC)

The mechanisms by which APIC is able to interface with the vCenter Controller are dependent on the user account associated to a given VMM Domain. Specific requirements are outlined for the vCenter user associated with the VMM Domain to ensure that the APIC can successfully perform operations on the vCenter, whether it is pushing and retrieving inventory and configurations or monitoring and listening to managed inventory related events.

The easiest way to remove concern about such requirements is to use the administrator vCenter account that has full access; however, this kind of freedom is not always available to the ACI administrator.

The minimum privileges for a custom user account, as of ACI version 4.2, are as follows:

- **Alarms** APIC creates two alarms on the folder. One for DVS and another for port group. An alarm is raised when the EPG or VMM Domain policy is deleted on the APIC, however vCenter is unable to delete the corresponding Port Group or DVS due to having VMs attached to it.
- **Distributed Switch**
- **dvPort Group**
- **Folder**
- **Network** APIC manages the network settings such as add or delete port groups, setting

host/DVS MTU, LLDP/CDP, LACP etc.

- **Host** If using AVS in addition to above, the user needs the Host privilege on the datacenter where APIC will create DVS.**Host.Configuration.Advanced settingsHost.Local operations.Reconfigure virtual machineHost.Configuration.Network configuration**This is needed for AVS and the auto-placement feature for virtual Layer 4 to Layer 7 Service VMs. For AVS, APIC creates VMK interface and places it in VTEP port group which is used for OpFlex.
- **Virtual machine** If Service Graphs are in use, the Virtual machine privilege for the virtual appliances is also required.**Virtual machine.Configuration.Modify device settingsVirtual machine.Configuration.Settings**

Troubleshooting RBAC-related issues

RBAC issues are most often encountered during initial setup of a VMM Domain but could be encountered if a vCenter administrator were to modify permissions of the user account associated with the VMM Domain after initial setup has already taken place.

The symptom can present itself in the following ways:

- Partial or complete inability to deploy new services (DVS creation, port group creation, some objects are successfully deployed but not all).
- Operational inventory is incomplete or missing from ACI administrator views.
- Faults raised for unsupported vCenter operation, or for any of the scenarios above (e.g. port group deployment failure).
- vCenter controller is reported as offline and faults indicate that there is connectivity or credential related issues.

Solution for RBAC-related issues

Verify all the above permissions are granted to the vCenter user that is configured in the VMM Domain.

Another method is to login directly to the vCenter with the same credentials as defined in the VMM Domain configuration and attempt similar operations (port group creation, etc.). If the user is not able to perform these same operations while logged in directly to the vCenter, clearly the correct permissions are not granted to the user.

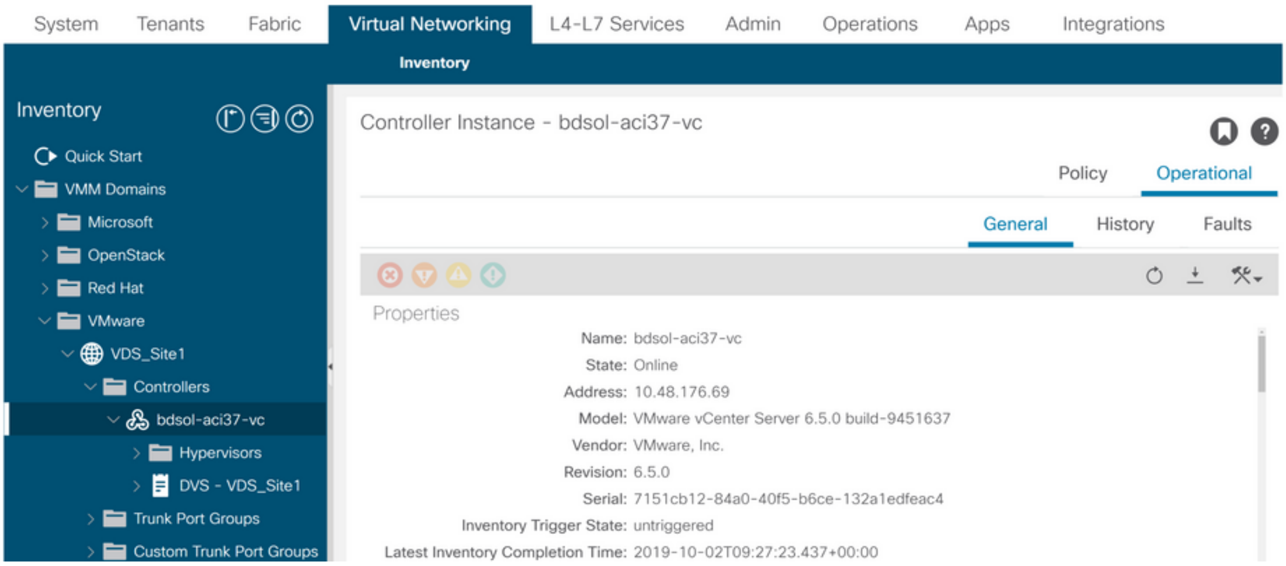
Connectivity Troubleshooting

When troubleshooting a VMM connectivity related issue, it is important to note some of the fundamental behaviors of how ACI communicates with vCenter.

The first and most pertinent behavior is that only one APIC in the cluster is sending configuration and collecting inventory at any given point. This APIC is referred to as the **shard leader** for this VMM Domain. However, multiple APICs are listening for **vCenter Events** in order to account for a scenario where the shard leader missed an event for any reason. Following the same distributed architecture of APICs, a given VMM Domain will have one APIC handling primary data and functionality (in this case, the shard leader), and two replicas (in the case of VMM they are referred to as **followers**). To distribute the handling of VMM communication and functionality across APICs, any two VMM Domains can either have the same or a different shard leaders.

The vCenter connectivity state can be found by navigating to the VMM controller of interest in the GUI or using the CLI command listed below.

VMWare VMM Domain - vCenter connectivity state



Name: bdsol-aci37-vc
State: Online
Address: 10.48.176.69
Model: VMware vCenter Server 6.5.0 build-9451637
Vendor: VMware, Inc.
Revision: 6.5.0
Serial: 7151cb12-84a0-40f5-b6ce-132a1edfeac4
Inventory Trigger State: untriggered
Latest Inventory Completion Time: 2019-10-02T09:27:23.437+00:00

```
apic2# show vmware domain name VDS_Site1 vcenter 10.48.176.69
Name                              : bdsol-aci37-vc
Type                              : vCenter
Hostname or IP                    : 10.48.176.69
Datacenter                        : Site1
DVS Version                       : 6.0
Status                            : online
Last Inventory Sync               : 2019-10-02 09:27:23
Last Event Seen                   : 1970-01-01 00:00:00
Username                          : administrator@vsphere.local
Number of ESX Servers             : 2
Number of VMs                     : 2
Faults by Severity                : 0, 0, 0, 0
Leader                            : bdsol-aci37-apic1

Managed Hosts:
ESX                VMs        Adjacency    Interfaces
---------------    --------   ----------   ------------------------------------------------
10.48.176.66       1          Direct       leaf-101 eth1/11, leaf-102 eth1/11
10.48.176.67       1          Direct       leaf-301 eth1/11, leaf-302 eth1/11
```

If a VMM controller is indicated to be offline, a fault will be thrown similar to below:

```
Fault fltCompCtrlrConnectFailed
Rule ID:130
Explanation:
This fault is raised when the VMM Controller is marked offline. Recovery is in process.
Code: F0130
Message: Connection to VMM controller: hostOrIp with name name in datacenter rootContName in
domain: domName is failing repeatedly with error: [remoteErrMsg]. Please verify network
connectivity of VMM controller hostOrIp and check VMM controller user credentials are valid.
```

The below steps can be used to troubleshoot connectivity issues between VC and the APICs.

1. Identifying the Shard Leader

The first step in troubleshooting a connectivity issue between the APIC and vCenter is understanding which APIC is the shard leader for the given VMM Domain. The easiest way to determine this information is to run the command 'show vmware domain name <domain>' on any APIC.

```
apic1# show vmware domain name VDS_Site1
Domain Name                        : VDS_Site1
Virtual Switch Mode                : VMware Distributed Switch
Vlan Domain                        : VDS_Site1 (1001-1100)
Physical Interfaces                : leaf-102 eth1/11, leaf-301 eth1/11, leaf-302 eth1/11,
                                  leaf-101 eth1/11
Number of EPGs                     : 2
Faults by Severity                 : 0, 0, 0, 0
LLDP override                      : RX: enabled, TX: enabled
CDP override                       : no
Channel Mode override              : mac-pinning
NetFlow Exporter Policy            : no
Health Monitoring                  : no

vCenters:
Faults: Grouped by severity (Critical, Major, Minor, Warning)
vCenter               Type      Datacenter            Status    ESXs   VMs    Faults
--------------------  --------  --------------------  --------  -----  -----  ---------------
10.48.176.69          vCenter   Site1                 online    2      2      0,0,0,0

APIC Owner:
Controller    APIC      Ownership
------------  --------  ---------------
bdsol-        apic1     Leader
aci37-vc
bdsol-        apic2     NonLeader
aci37-vc
bdsol-        apic3     NonLeader
aci37-vc
```

2. Verifying Connectivity to vCenter

After identifying the APIC which is actively communicating with the vCenter, verify IP connectivity with tools such as ping.

```
apic1# ping 10.48.176.69
PING 10.48.176.69 (10.48.176.69) 56(84) bytes of data.
64 bytes from 10.48.176.69: icmp_seq=1 ttl=64 time=0.217 ms
64 bytes from 10.48.176.69: icmp_seq=2 ttl=64 time=0.274 ms
64 bytes from 10.48.176.69: icmp_seq=3 ttl=64 time=0.346 ms
64 bytes from 10.48.176.69: icmp_seq=4 ttl=64 time=0.264 ms
64 bytes from 10.48.176.69: icmp_seq=5 ttl=64 time=0.350 ms
^C
--- 10.48.176.69 ping statistics ---
5 packets transmitted, 5 received, 0% packet loss, time 4084ms
rtt min/avg/max/mdev = 0.217/0.290/0.350/0.052 ms
```

If the vCenter was configured using the FQDN rather than IP address, the nslookup command can be used to verify name resolution.

```
apic1:~> nslookup bdsol-aci37-vc
Server: 10.48.37.150
Address: 10.48.37.150#53
Non-authoritative answer:
```

```
Name: bdsol-aci37-vc.cisco.com
Address: 10.48.176.69
```

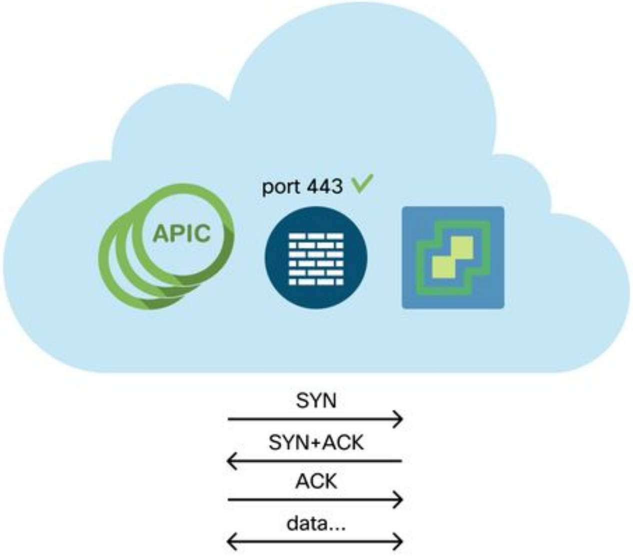
3. Check if OOB or INB is used

Check the APIC routing table to verify if out-of-band or in-band is preferred for connectivity and which gateway is used:

```
apic1# bash
admin@apic1:~> route
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
default         10.48.176.1     0.0.0.0         UG    16     0        0 oobmgmt
```

4. Ensure Port 443 is allowed between all APICs and the vCenter, including any firewalls in the path of communication.

vCenter <-> APIC - HTTPS (TCP port 443) - communication



General HTTPS reachability from the APICs to vCenter can be tested with a curl:

```
apic2# curl -v -k https://10.48.176.69
* Rebuilt URL to: https://10.48.176.69/*   Trying 10.48.176.69...
* TCP_NODELAY set
* Connected to 10.48.176.69 (10.48.176.69) port 443 (#0)
...
```

Verify that the shard leader has an established TCP connection on port 443 using the netstat command.

```
apic1:~> netstat -tulaen | grep 10.48.176.69
tcp 0 0 10.48.176.57:40806 10.48.176.69:443 ESTABLISHED 600 13062800
```

5. Perform a Packet capture

If possible, perform a packet capture along the path in between the shard leader and vCenter in an effort to identify if traffic is being sent and received by either device.

VMware Inventory

The following table shows a list of VMWare VDS parameters and specifies whether these are configurable by the APIC.

VMware VDS parameters managed by APIC

VMware VDS	Default Value	Configurable Using Cisco APIC Policy?
Name	VMM Domain name	Yes (Derived from Domain)
Description	'APIC Virtual Switch'	No
Folder Name	VMM Domain name	Yes (Derived from Domain)
Version	Highest supported by vCenter	Yes
Discovery Protocol	LLDP	Yes
Uplink Ports and Uplink Names	8	Yes (From Cisco APIC Release 4.2(1))
Uplink Name Prefix	uplink	Yes (From Cisco APIC Release 4.2(1))
Maximum MTU	9000	Yes
LACP policy	disabled	Yes
Port mirroring	0 sessions	Yes
Alarms	2 alarms added at the folder level	No

The following table shows a list of VMWare VDS Port Group parameters and specifies whether these are configurable by the APIC.

VMWare VDS Port Group parameters managed by APIC

VMware VDS Port Group	Default Value	Configurable using APIC Policy
Name	Tenant Name | Application Profile Name | EPG Name	Yes (Derived from EPG)
Port binding	Static binding	No
VLAN	Picked from VLAN pool	Yes
Load balancing algorithm	Derived based on port-channel policy on APIC	Yes
Promiscuous mode	Disabled	Yes
Forged transmit	Disabled	Yes
MAC change	Disabled	Yes
Block all ports	FALSE	No

VMware Inventory Troubleshooting

Inventory sync events occur to ensure that the APIC is aware of vCenter events that may require the APIC to dynamically update policy. There are two types of inventory sync events that can occur between vCenter and the APIC; a full inventory sync and an event-based inventory sync. The default schedule of a full inventory sync between the APIC and vCenter is every 24 hours, however these can also be manually triggered. Event-based inventory syncs are typically tied to triggered tasks, such as a vMotion. In this scenario, if a virtual machine moves from one host to another, and those hosts are connected to two different leaf switches, the APIC will listen for the VM migration event and, in the scenario of on-demand deployment immediacy, unprogram the EPG on the source leaf, and program the EPG on the destination leaf.

Depending on the deployment immediacy of EPGs associated with a VMM domain, failure to pull inventory from the vCenter could have undesirable consequences. In the scenario that inventory has failed to complete or is partial, there will always be a fault raised indicating the object or objects which have caused the failure.

Scenario 1 - Virtual machine with invalid backing:

If a virtual machine is moved from one vCenter to another, or the virtual machine is determined to have an invalid backing (e.g. a port group attachment to an old/deleted DVS), the vNIC will be reported to have an operational issues.

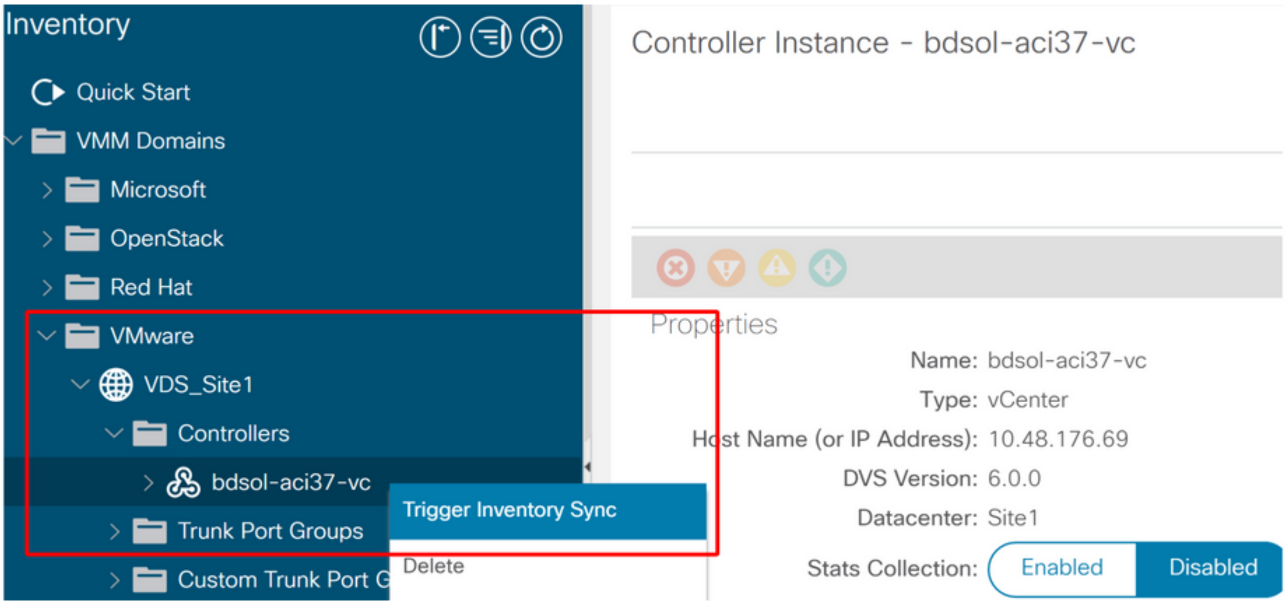
```
Fault fltCompVNicOperationalIssues
Rule ID:2842
Explanation:
This fault is raised when ACI controller failed to update the properties of a VNIC (e.g., it can
not find the EPG that the VNIC attached to).
Code: F2842
Message: Operational issues detected for VNic name on VM name in VMM controller: hostOrIp with
name name in datacenter rootContName in domain: domName due to error: issues.
Resolution:
Remediate the virtual machines indicated in the fault by assigning a valid port group on the
affected vNIC of the VM.
```

Scenario 2 — vCenter administrator modified a VMM managed object on the vCenter:

Modifying objects which are managed by the APIC from vCenter is not a supported operation. The following fault would be seen if an unsupported operation is performed on vCenter.

```
Fault fltCompCtrlrUnsupportedOperation
Rule ID:133
Explanation:
This fault is raised when deployment of given configuration fails for a Controller.
Code: F0133
Message: Unsupported remote operation on controller: hostOrIp with name name in datacenter
rootContName in domain domName detected, error: [deployIssues]
Resolution:
If this scenario is encountered, try to undo the unsupported change in vCenter and then trigger
an 'inventory sync' manually.
```

VMWare VMM Domain - vCenter controller - trigger inventory sync



Controller Instance - bdsol-aci37-vc

Properties

Name: bdsol-aci37-vc

Type: vCenter

Host Name (or IP Address): 10.48.176.69

DVS Version: 6.0.0

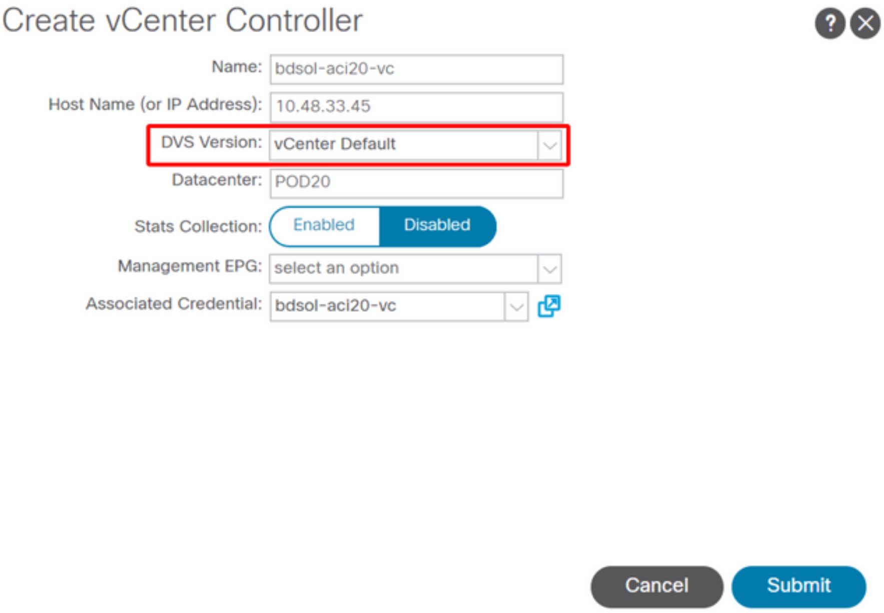
Datacenter: Site1

Stats Collection: Enabled [✓ Disabled]

VMware DVS version

When creating a new vCenter controller as part of a VMM Domain, the default setting for the DVS Version will be to use the 'vCenter Default'. When selecting this, the DVS version will be created with the version of the vCenter.

VMWare VMM Domain - vCenter controller creation



Create vCenter Controller

Name: bdsol-aci20-vc

Host Name (or IP Address): 10.48.33.45

DVS Version: vCenter Default

Datacenter: POD20

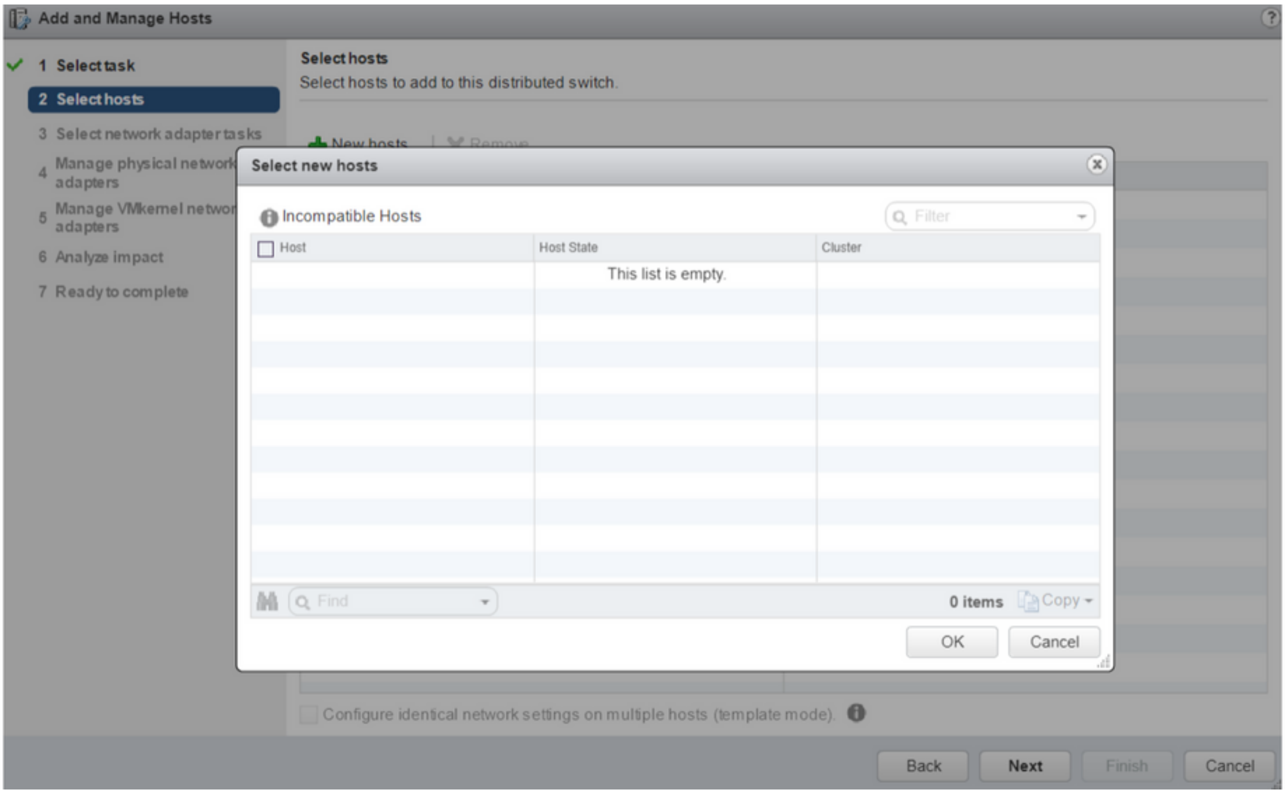
Stats Collection: Enabled [✓ Disabled]

Management EPG: select an option

Associated Credential: bdsol-aci20-vc

This means that in the example of a vCenter running 6.5 and ESXi servers running 6.0, the APIC will create a DVS with version 6.5 and hence the vCenter administrator will be unable to add the ESXi servers running 6.0 into the ACI DVS.

APIC managed DVS - vCenter host addition - empty list



Select hosts

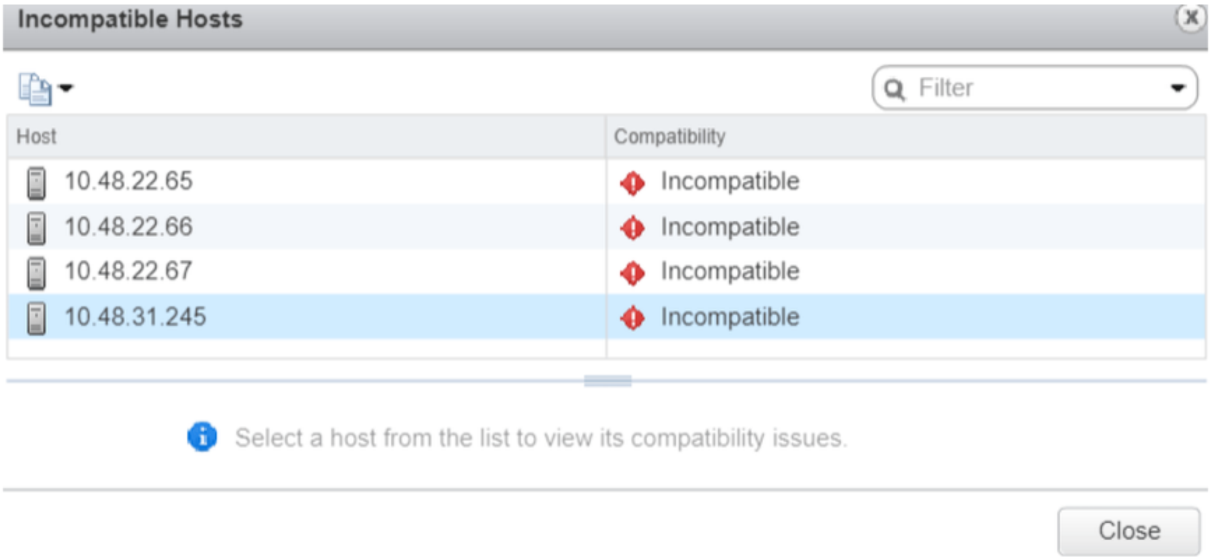
Select hosts to add to this distributed switch.

Host	Host State	Cluster
	This list is empty.	

0 items

□ Configure identical network settings on multiple hosts (template mode).

APIC managed DVS - vCenter host addition - incompatible hosts



Host	Compatibility
10.48.22.65	Incompatible
10.48.22.66	Incompatible
10.48.22.67	Incompatible
10.48.31.245	Incompatible

Select a host from the list to view its compatibility issues.

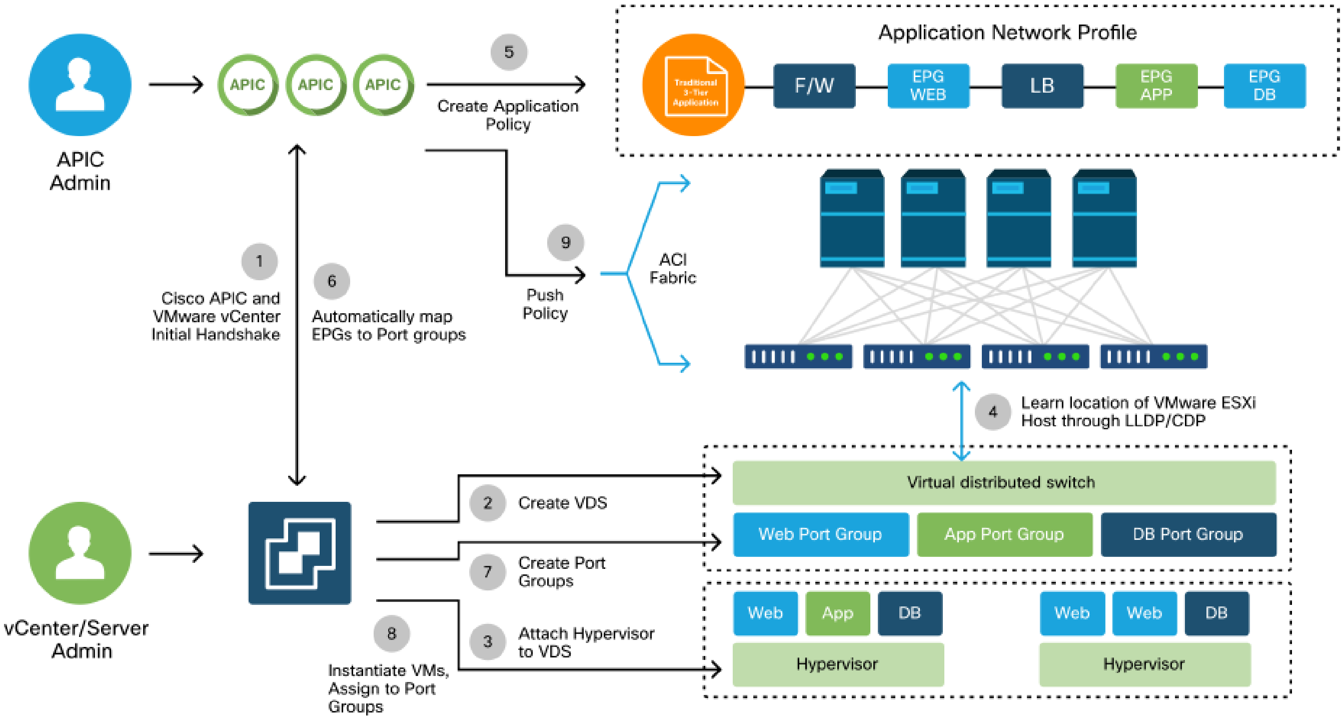
So, when creating a VMM Domain make sure to select the correct 'DVS Version' such that the necessary ESXi servers can be added to the DVS.

Host dynamic discovery

Host / VM discovery process

VMM integration in ACI differentiates itself from manual provisioning in that the fabric can dynamically discover where hosts and applicable virtual machines are connected to efficiently deploy policy. Through this dynamic process, ACI can optimize utilization of hardware resources on the leaf switches as VLANs, SVIs, zoning rules, etc. are deployed on nodes only when there is an endpoint connected which requires the policy. The benefit to the network administrator, from an ease of use perspective, is that ACI will provision VLAN/policy where VMs connect in an automated way. To determine where policy must be deployed, the APIC will use information from multiple sources. The following diagram outlines the basic steps of the host discovery process when using a DVS-based VMM Domain.

VMWare VMM Domain — Deployment workflow



APIC Admin
APIC
APIC
APIC
5
Create Application Policy
Application Network Profile
Traditional 3-Tier Application
F/W
EPG WEB
LB
EPG APP
EPG DB
1
Cisco APIC and VMware vCenter Initial Handshake
6
Automatically map EPGs to Port groups
9
Push Policy
ACI Fabric
4
Learn location of VMware ESXi Host through LLDP/CDP
vCenter/Server Admin
2
Create VDS
7
Create Port Groups
8
Instantiate VMs, Assign to Port Groups
3
Attach Hypervisor to VDS
Virtual distributed switch
Web Port Group
App Port Group
DB Port Group
Web
App
DB
Hypervisor
Web
Web
DB
Hypervisor

In short the following key steps are happening when:

- LLDP or CDP is exchanged between hypervisor and leaf switches.
- Hosts report adjacency info to vCenter.
- vCenter notifies APIC of adjacency info: APIC knows about host via inventory sync.
- APIC pushes policy to the leaf port: please review "Resolution Immediacy" sub-section within this section to further understand these conditions.
- If vCenter adjacency info is lost, APIC can remove policy.

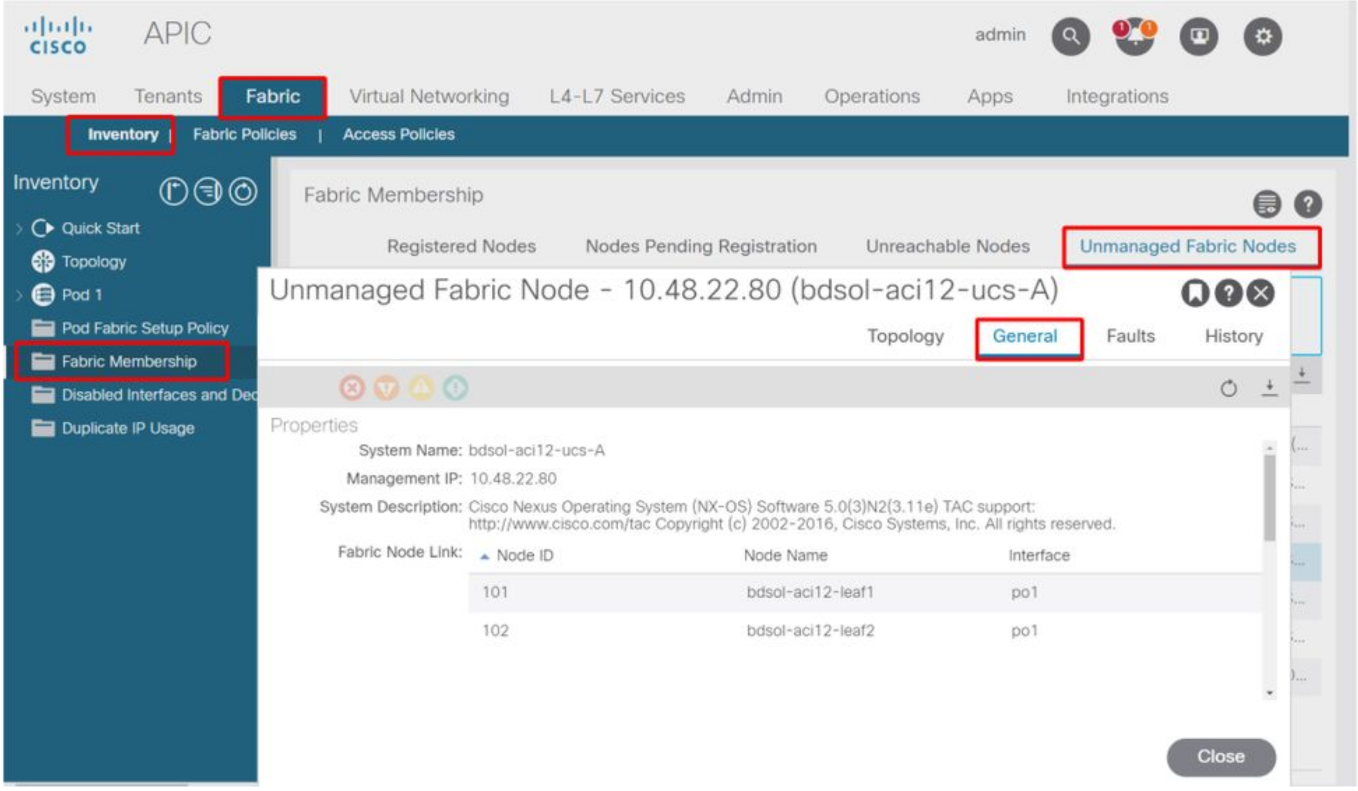
As can be seen, CDP/LLDP plays a key role in the discovery process and it is important to make sure this is properly configured and both sides are using the same protocol.

Fabric LooseNode / intermediate switch - use case

In a deployment using a blade chassis with an intermediate switch between the leaf switches and the hypervisor, the APIC needs to 'stitch' the adjacency together. In this scenario, multiple discovery protocols could be used as the intermediate switch may have different protocol requirements than the host.

In a setup with a blade server and an intermediate switch (i.e. blade chassis switch), ACI should detect the intermediate switch and map the hypervisors behind it. The intermediate switch is referred to in ACI as a LooseNode or an 'Unmanaged Fabric Node'. The detected LooseNodes can be viewed under 'Fabric > Inventory > Fabric Membership > Unmanaged Fabric Nodes'. By navigating to one of these types of servers in the GUI, the user can view the path from leaf to intermediate switch to host.

APIC UI — Unmanaged fabric nodes (LooseNodes)



Unmanaged Fabric Node - 10.48.22.80 (bdsol-aci12-ucs-A)

Properties

System Name: bdsol-aci12-ucs-A

Management IP: 10.48.22.80

System Description: Cisco Nexus Operating System (NX-OS) Software 5.0(3)N2(3.11e) TAC support: http://www.cisco.com/tac Copyright (c) 2002-2016, Cisco Systems, Inc. All rights reserved.

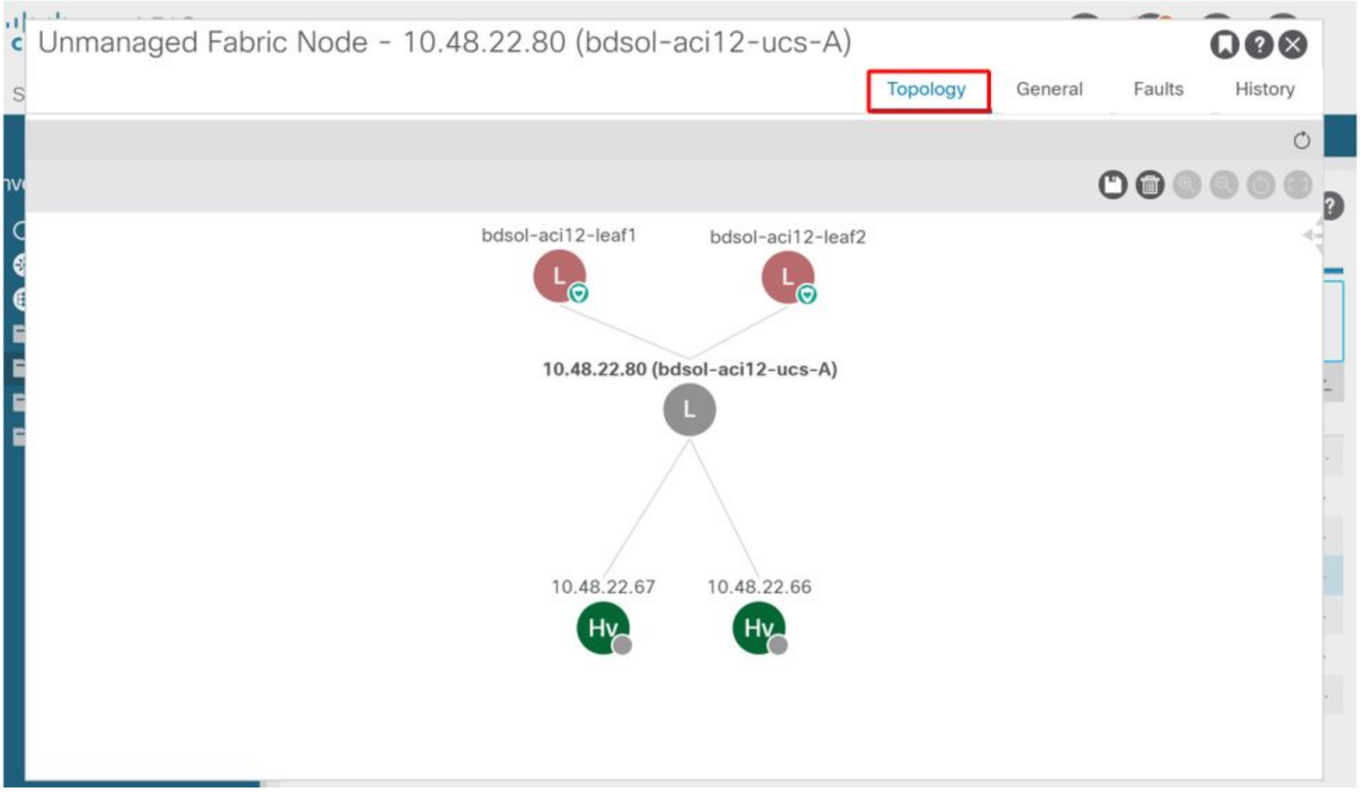
Fabric Node Link:

Node ID	Node Name	Interface
101	bdsol-aci12-leaf1	po1
102	bdsol-aci12-leaf2	po1

With LLDP or CDP discovery in place, ACI can determine the topology for such LooseNodes, given that the hypervisor downstream of the intermediate switch is managed through VMM integration, and the leaf itself has an adjacency to the intermediate switch from downstream.

This concept is illustrated by the image below.

APIC UI — Unmanaged Fabric Node path

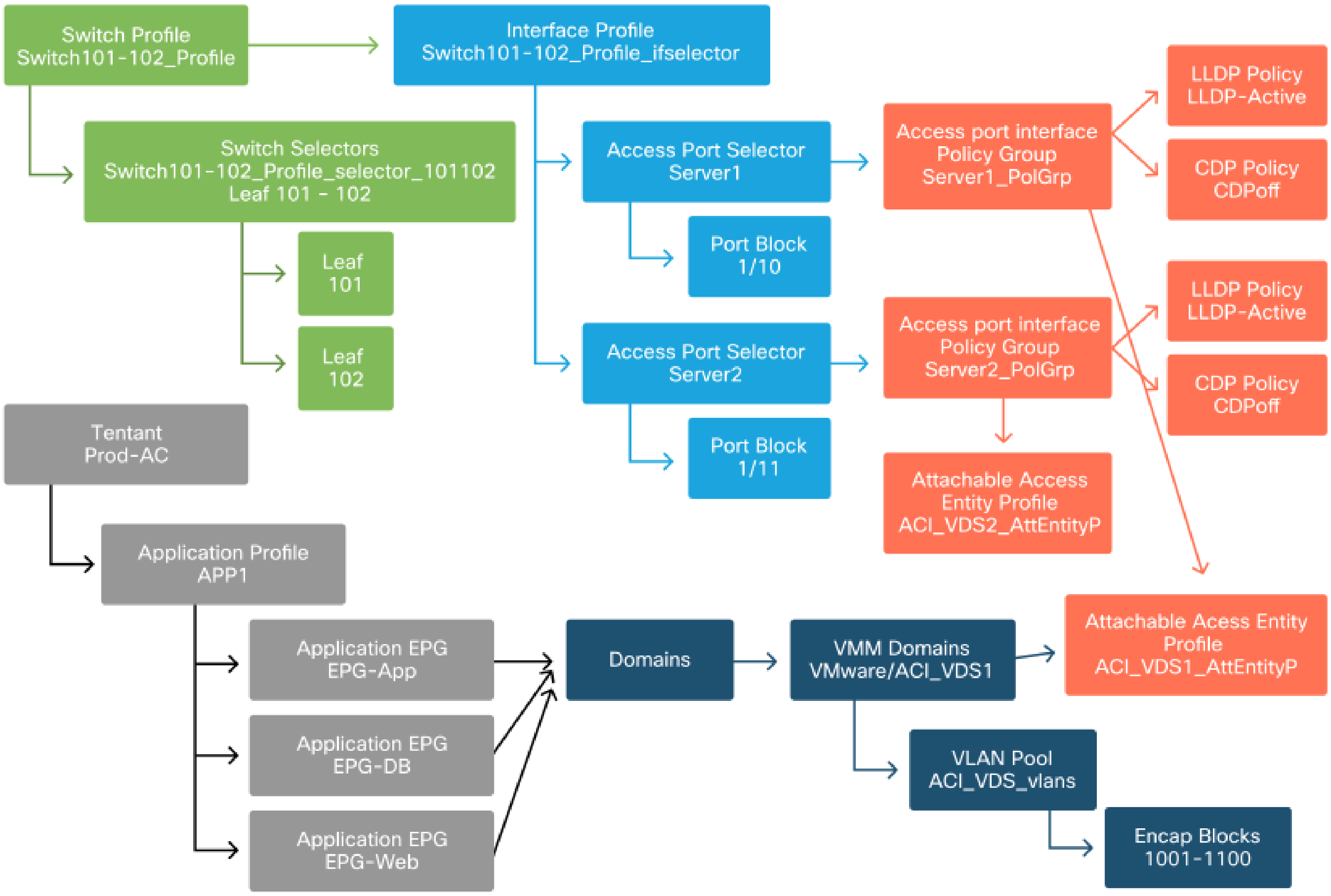


Resolution Immediacy

In scenarios where critical services utilize the VMM-integrated DVS such as management connectivity to vCenter/ESXi, it is prudent to use the Pre-provision Resolution Immediacy. With this setting, the mechanism of dynamic host discovery is removed and instead policy / VLANs are statically programmed on the host facing interfaces. In this configuration, the VMM VLANs will always be deployed to all interfaces tied to the AEP referenced by the VMM Domain. This removes the possibility that a critical VLAN (such as management) is removed from a port due to a discovery protocol-related adjacency event.

Refer to the below diagram:

Pre-provision deployment example



If Pre-provision was set for an EPG in the ACI_VDS1 VMM Domain, then VLANs would be deployed on links for Server1 but not Server2 as Server2's AEP does not include the ACI_VDS1 VMM Domain.

To summarize the resolution immediacy settings:

- On-Demand - Policy is deployed when adjacency is established between leaf and host and a VM attached to the port group.
- Immediate - Policy is deployed when adjacency is established between leaf and host.
- Pre-provision - Policy is deployed to all ports using an AEP with the VMM Domain contained, no adjacency is required.

Troubleshooting scenarios

VM cannot resolve ARP for its default gateway

In this scenario, VMM integration has been configured and the DVS has been added to the hypervisor but the VM cannot resolve ARP for its gateway in ACI. For the VM to have network connectivity, verify that adjacency has established and VLANs are deployed.

First, the user can check the leaf has detected the host by using 'show lldp neighbors' or 'show cdp neighbors' on the leaf depending on the protocol selected.

```
Leaf101# show lldp neighbors
Capability codes:
 (R) Router, (B) Bridge, (T) Telephone, (C) DOCSIS Cable Device
 (W) WLAN Access Point, (P) Repeater, (S) Station, (O) Other
Device ID            Local Intf      Hold-time  Capability  Port ID
bdsol-aci37-apic1     Eth1/1          120                    eth2-1
bdsol-aci37-apic2     Eth1/2          120                    eth2-1
bdsol-aci37-os1       Eth1/11         180        B           0050.565a.55a7
S1P1-Spine201         Eth1/49         120        BR          Eth1/1
S1P1-Spine202         Eth1/50         120        BR          Eth1/1
Total entries displayed: 5
```

If needed from a troubleshooting perspective, this can be validated from the ESXi side both on the CLI and GUI:

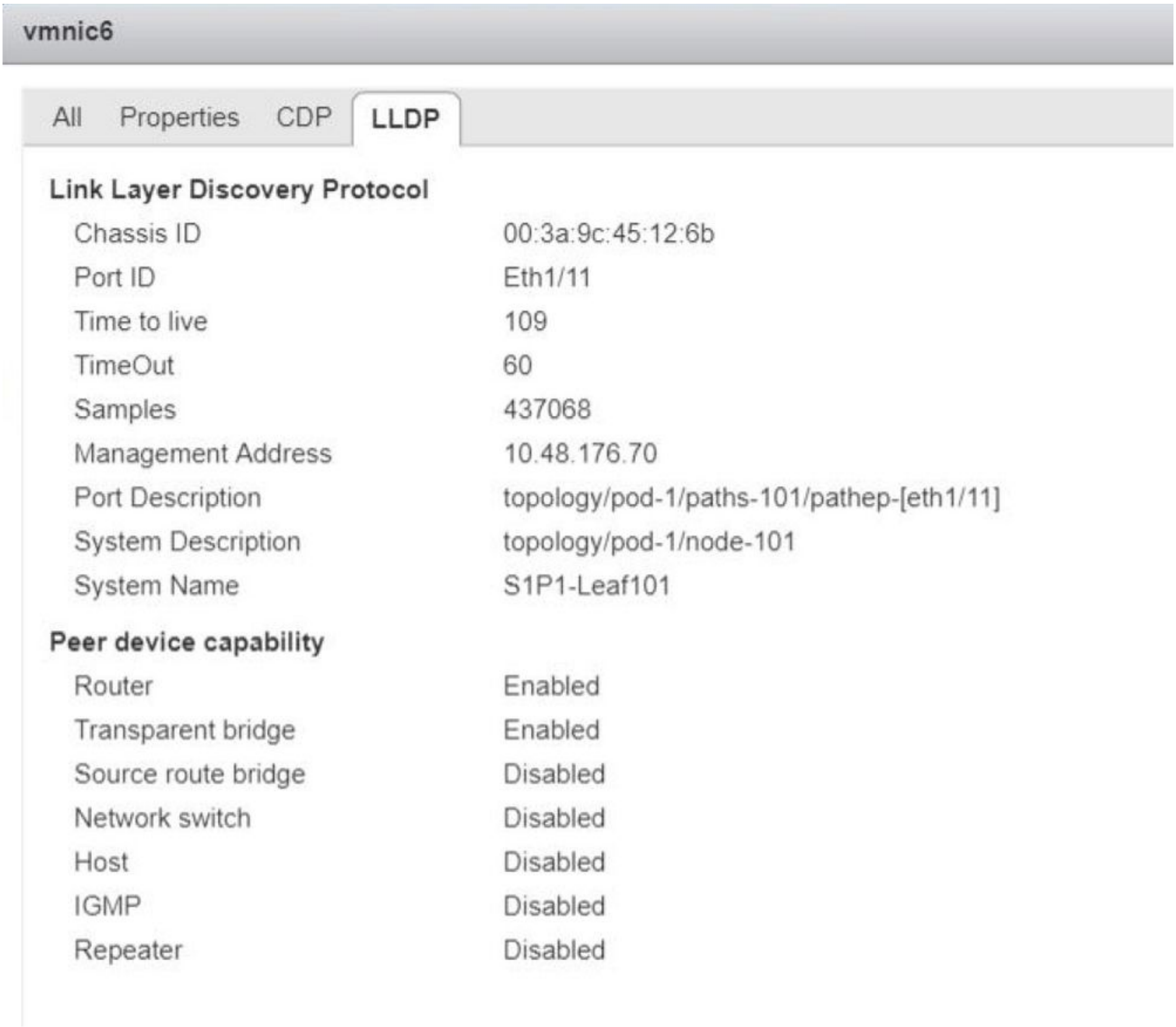
```
[root@host:~] esxcli network vswitch dvs vmware list
VDS_Site1
  Name: VDS_Site1
 ...
  Uplinks: vmnic7, vmnic6
  VMware Branded: true
  DVPort:
        Client: vmnic6
        DVPortgroup ID: dvportgroup-122
        In Use: true
        Port ID: 0

        Client: vmnic7
        DVPortgroup ID: dvportgroup-122
        In Use: true
        Port ID: 1

[root@host:~] esxcfg-nics  -l
Name    PCI          Driver      Link Speed     Duplex MAC Address       MTU    Description
vmnic6  0000:09:00.0 enic        Up   10000Mbps Full   4c:77:6d:49:cf:30 9000   Cisco Systems
Inc Cisco VIC Ethernet NIC
vmnic7  0000:0a:00.0 enic        Up   10000Mbps Full   4c:77:6d:49:cf:31 9000   Cisco Systems
Inc Cisco VIC Ethernet NIC

[root@host:~] vim-cmd hostsvc/net/query_networkhint --pnic-name=vmnic6 | grep -A2 "System Name"
              key = "System Name",
              value = "Leaf101"
           }
```

vCenter Web Client - host - vmnic LLDP/CDP adjacency details



vmnic6

Link Layer Discovery Protocol

Chassis ID	00:3a:9c:45:12:6b
Port ID	Eth1/11
Time to live	109
TimeOut	60
Samples	437068
Management Address	10.48.176.70
Port Description	topology/pod-1/paths-101/pathep-[eth1/11]
System Description	topology/pod-1/node-101
System Name	S1P1-Leaf101

Peer device capability

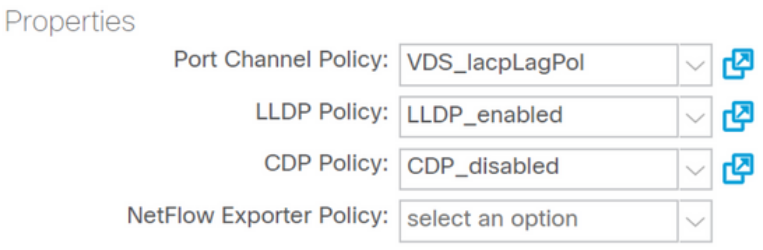
Router	Enabled
Transparent bridge	Enabled
Source route bridge	Disabled
Network switch	Disabled
Host	Disabled
IGMP	Disabled
Repeater	Disabled

If the leaf LLDP adjacency cannot be seen from the ESXi host, this is often caused by using a network adapter which is configured to generate LLDPDUs instead of the ESXi OS. Make sure to validate if the network adapter has LLDP enabled and hence is consuming all LLDP information. If this is the case, be sure to disable LLDP on the adapter itself so it is controlled through the vSwitch policy.

Another cause might be that there is a mis-alignment between the discovery protocols used between leaf and ESXi Hypervisor. Make sure on both ends to use the same discovery protocol.

To check if the CDP/LLDP settings are aligned between ACI and the DVS in the APIC UI, navigate to 'Virtual Networking > VMM Domains > VMWare > Policy > vSwitch Policy'. Make sure to only enable either LLDP or CDP policy as they are mutually exclusive.

APIC UI - VMWare VMM Domain - vSwitch policy



Properties

Port Channel Policy: VDS_lacpLagPol

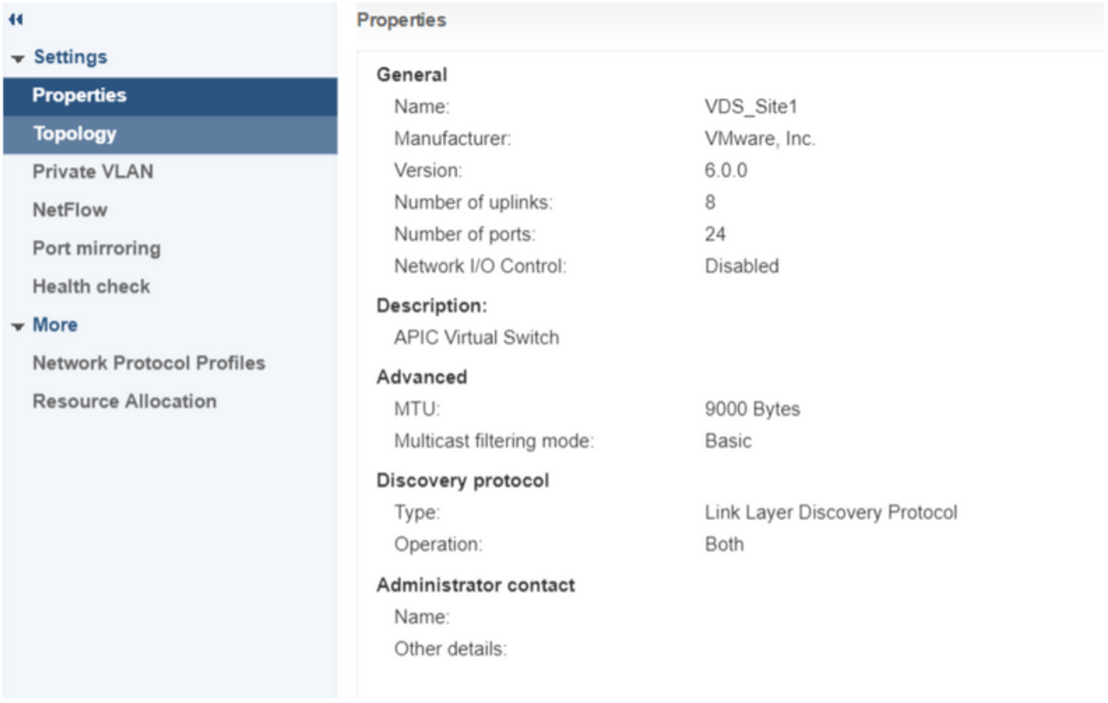
LLDP Policy: LLDP_enabled

CDP Policy: CDP_disabled

NetFlow Exporter Policy: select an option

In vCenter go to: 'Networking > VDS > Configure'.

vCenter Web Client UI - VDS properties



Properties

General

Name: VDS_Site1

Manufacturer: VMware, Inc.

Version: 6.0.0

Number of uplinks: 8

Number of ports: 24

Network I/O Control: Disabled

Description:

APIC Virtual Switch

Advanced

MTU: 9000 Bytes

Multicast filtering mode: Basic

Discovery protocol

Type: Link Layer Discovery Protocol

Operation: Both

Administrator contact

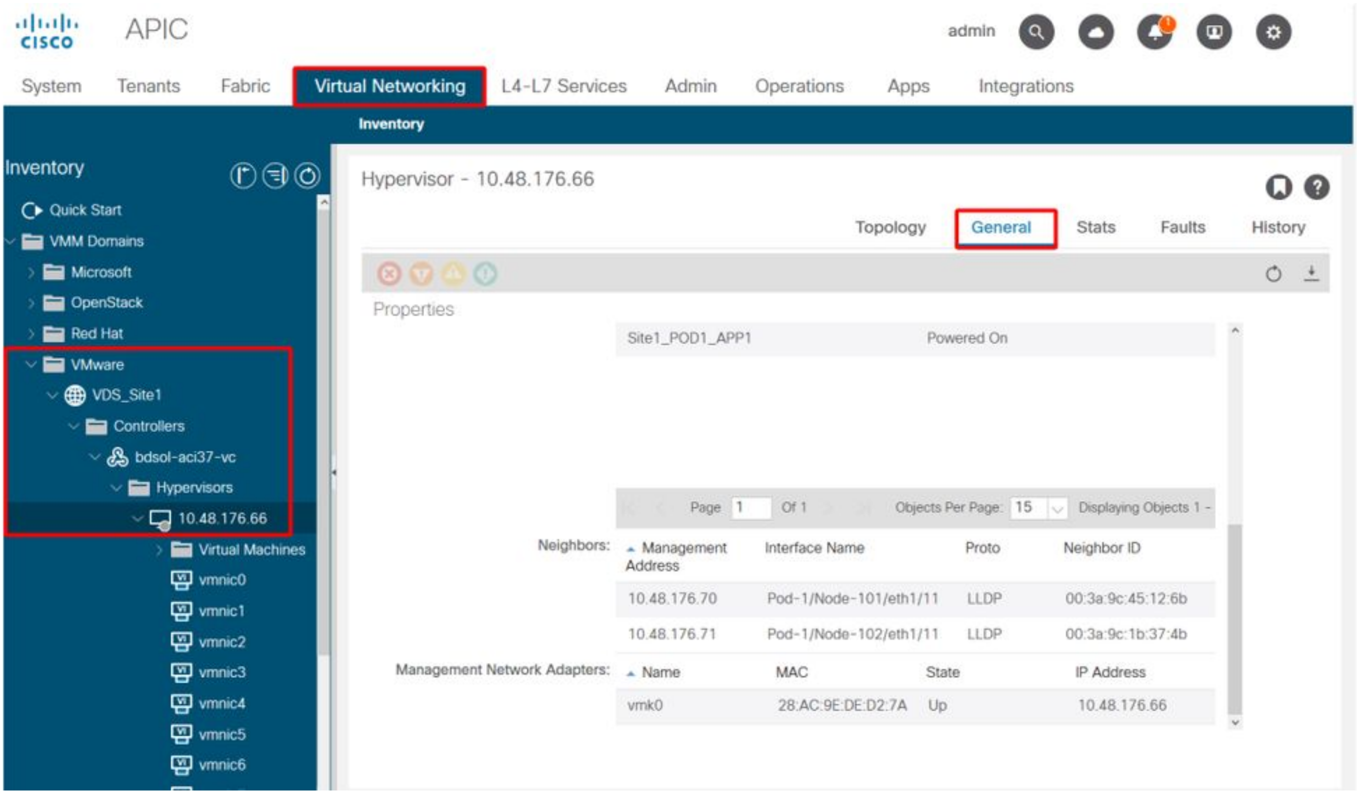
Name:

Other details:

Correct the LLDP/CDP settings if needed.

Then validate the APIC observes the ESXi host's LLDP/CDP neighborship against the leaf switch in the UI under 'Virtual Networking > VMM Domains > VMWare > Policy > Controller > Hypervisor > General'.

APIC UI - VMWare VMM Domain - Hypervisor details



Hypervisor - 10.48.176.66

Site1_POD1_APP1 Powered On

Page 1 Of 1 Objects Per Page: 15 Displaying Objects 1 -

Neighbors:

Management Address	Interface Name	Proto	Neighbor ID
10.48.176.70	Pod-1/Node-101/eth1/11	LLDP	00:3a:9c:45:12:6b
10.48.176.71	Pod-1/Node-102/eth1/11	LLDP	00:3a:9c:1b:37:4b

Management Network Adapters:

Name	MAC	State	IP Address
vmk0	28:AC:9E:DE:D2:7A	Up	10.48.176.66

If this is showing expected values, then the user can validate that the VLAN is present on the port toward the host.

```
S1P1-Leaf101# show vlan encap-id 1035

 VLAN Name                             Status    Ports
 ---- -------------------------------- --------- -------------------------------
 12   Ecommerce:Electronics:APP        active    Eth1/11

 VLAN Type  Vlan-mode
 ---- ----- ----------
 12   enet  CE
```

vCenter/ESXi management VMK attached to APIC-pushed DVS

In a scenario where vCenter or ESXi management traffic needs to utilize the VMM integrated DVS, it is important to take some extra care to avoid a stalemate in activating the dynamic adjacencies and activate the required VLANs.

For vCenter, which is typically built before VMM integration is configured, it is important to use a physical domain and static path to assure the vCenter VM's encap VLAN is always programmed on the leaf switches so that it can be used before VMM integration is fully set up. Even after setting up the VMM integration, it is advised to leave this static path in place to always assure availability of this EPG.

For the ESXi hypervisors, as per the "Cisco ACI Virtualization Guide" on Cisco.com, when migrating onto the vDS it is important to make sure that the EPG where the VMK interface will be connected is deployed with the resolution immediacy set to Pre-provision. This will make sure the VLAN is always programed on the leaf switches without relying on LLDP/CDP discovery of the ESXi hosts.

Host adjacencies not discovered behind LooseNode

Typical causes of LooseNode discovery issues are:

- CDP/LLDP is not enabled CDP/LLDP must be exchanged between the intermediate switch, the leaf switches and ESXi hostsFor Cisco UCS, this is accomplished via a network control policy on the vNIC
- A change in the management IP of the LLDP/CDP neighbor breaks connectivity The vCenter will see the new management IP in the LLDP/CDP adjacency, but will not update APICTrigger a manual inventory sync to fix
- VMM VLANs are not added to the intermediate switch The APIC doesn't program third party blade/intermediate switches.Cisco UCSM integration app (ExternalSwitch) available in 4.1(1) release.VLANs must be configured and trunked to uplinks connected to ACI leaf nodes and downlinks connected to hosts

F606391 - Missing adjacencies for the physical adapter on the host

When seeing the fault below:

```
Affected Object: comp/prov-VMware/ctrlr-[DVS-DC1-ACI-LAB]-DVS1/hv-host-104
Fault delegate: [FSM:FAILED]: Get LLDP/CDP adjacency information for the physical adapters on
the host: bdsol-aci20-os3 (TASK:ifc:vmmmgr:CompHvGetHpNicAdj)
```

Please review the workflow in section "VM cannot resolve ARP for its default gateway" as this means there are missing CDP/LLDP adjacencies. These adjacencies should be verified end-to-end.

Hypervisor uplink load balancing

When connecting hypervisors such as ESXi to an ACI fabric, they will typically be connected with multiple uplinks. In fact, it is recommended to have an ESXi host connected to at least two leaf switches. This will minimize the impact of failure scenarios or upgrades.

In order to optimize how uplinks are used by the workloads running on a hypervisor, VMware vCenter configurations allow configuring multiple load balancing algorithms for VM-generated traffic towards the hypervisor's uplinks.

It is crucial to have all hypervisors and the ACI fabric aligned with the same load balancing algorithm configuration to ensure correct connectivity is in place. Failure to do so may result in intermittent traffic flow drops and endpoint moves in the ACI fabric.

This can be seen in an ACI fabric by excessive alerts such as:

```
F3083 fault
ACI has detected multiple MACs using the same IP address 172.16.202.237.
MACs: Context: 2981888. fvCEps:
uni/tn-BSE_PROD/ap-202_Voice/epg-VLAN202_Voice/cep-00:50:56:9D:55:B2;
uni/tn-BSE_PROD/ap-202_Voice/epg-VLAN202_Voice/cep-00:50:56:9D:B7:01;
or
[F1197][raised][bd-limits-exceeded][major][sys/ctx-[vxlan-2818048]/bd-[vxlan-16252885]/fault-
F1197]
Learning is disabled on BD Ecommerce:BD01
```

This chapter will cover VMWare ESXi host connectivity into ACI but is applicable for most hypervisors.

Rack server

When looking at the various ways an ESXi host can connect to an ACI fabric, they are divided in 2 groups, switch dependent and switch independent load balancing algorithms.

Switch independent load balancing algorithms are ways to connect where no specific switch configuration is needed. For switch dependent load balancing, switch-specific configurations are required.

Make sure to validate if vSwitch Policy is in line with the 'ACI Access Policy Group' requirements as per the table below.

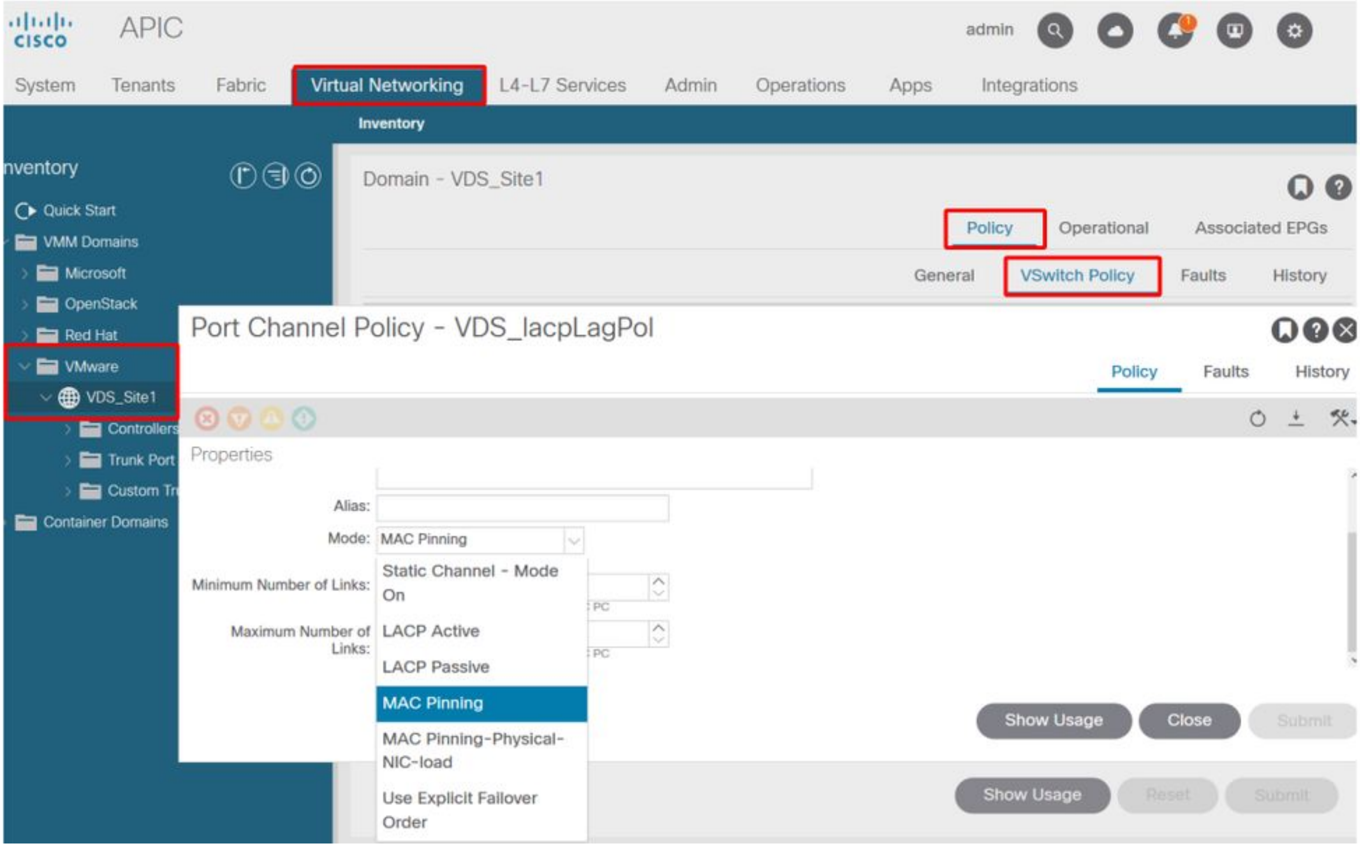
Teaming and ACI vSwitch policy

VMware Teaming and Failover Mode	ACI vSwitch Policy	Description	ACI Access Policy Group - Port Channel Required
Route based on the originating virtual port	MAC Pinning	Select an uplink based on the virtual port IDs on the switch. After the virtual switch selects an uplink for a virtual machine or a VMKernel adapter, it always forwards traffic through the same uplink for this virtual machine or VMKernel adapter.	No
Route based on Source MAC hash	NA	Select an uplink based on a hash of the source MAC address	NA
Explicit Failover Order	Use Explicit Failover Mode	From the list of active adapters, always use the highest order uplink that passes failover detection criteria. No actual load balancing is performed with this option.	No
Link Aggregation(LAG) - IP Hash Based	Static Channel - Mode On	Select an uplink based on a hash of the source and destination IP addresses of each packet. For non-IP packets, the switch uses the data at those fields to compute the hash. IP-based teaming requires that on the ACI side a port-channel / VPC is configured with 'mode on'.	Yes (channel mode set to 'on')
Link Aggregation(LAG) - LACP	LACP Active / Passive	Select an uplink based on a selected hash (20 different hash options available). LACP based teaming requires that on the ACI side a port-channel / VPC is configured with LACP enabled. Make sure to create an Enhanced Lag Policy in ACI and apply it to the VSwitch Policy.	Yes (channel mode set to 'LACP Active/Passive')
Route based	MAC Pinning -	Available for distributed port groups or	No

on Physical NIC Load (LBT)	Physical-NIC-load	distributed ports. Select an uplink based on the current load of the physical network adapters connected to the port group or port. If an uplink remains busy at 75 percent or higher for 30 seconds, the host's vSwitch moves a part of the virtual machine traffic to a physical adapter that has free capacity.

See the screenshot below on how to validate Port-Channel Policy as part of the vSwitch Policy in place.

ACI vSwitch Policy — Port Channel Policy


Port Channel Policy - VDS_lacpLagPol

Alias:

Mode: MAC Pinning

Minimum Number of Links:

Maximum Number of Links:

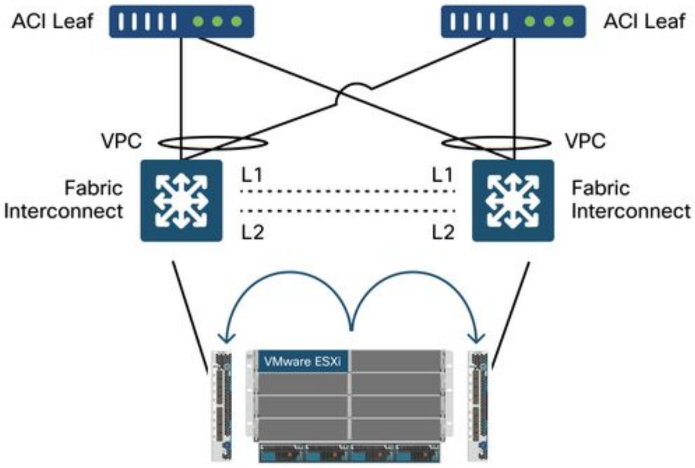
Note : For a more in-depth description of VMware networking features, please review vSphere Networking at **https://docs.vmware.com/en/VMware-vSphere/6.5/com.vmware.vsphere.networking.doc/GUID-D34B1ADD-B8A7-43CD-AA7E-2832A0F7EE76.html**

Cisco UCS B-Series use case

When using Cisco UCS B-Series servers, it is important to note they connect within their chassis to UCS Fabric Interconnects (FIs) that do not have a unified dataplane. This use case equally applies to other vendors which employ a similar topology. Because of this there can be a difference between the load-balancing method used from an ACI leaf switch side and the vSwitch side.

Below is a UCS FI topology with ACI:

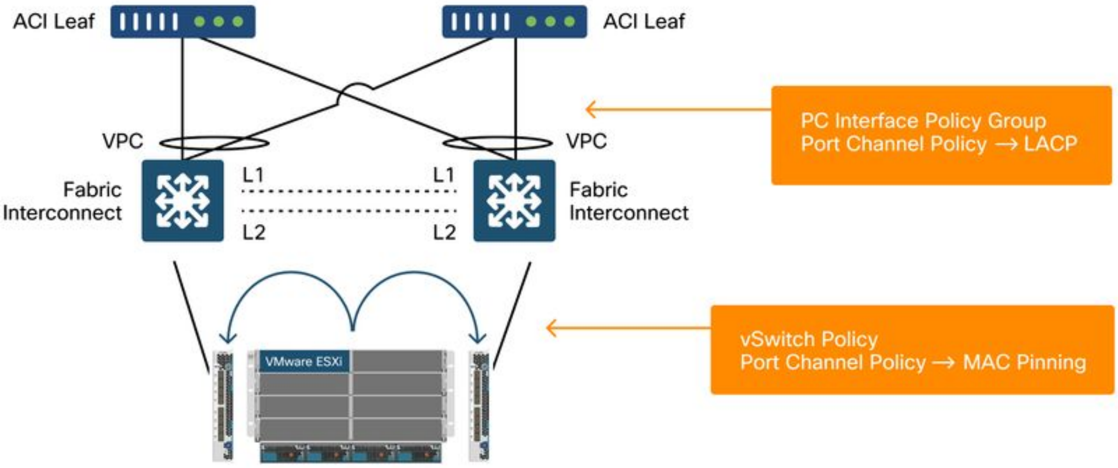
Cisco UCS FI with ACI leaf switches - topology



Key things to notice:

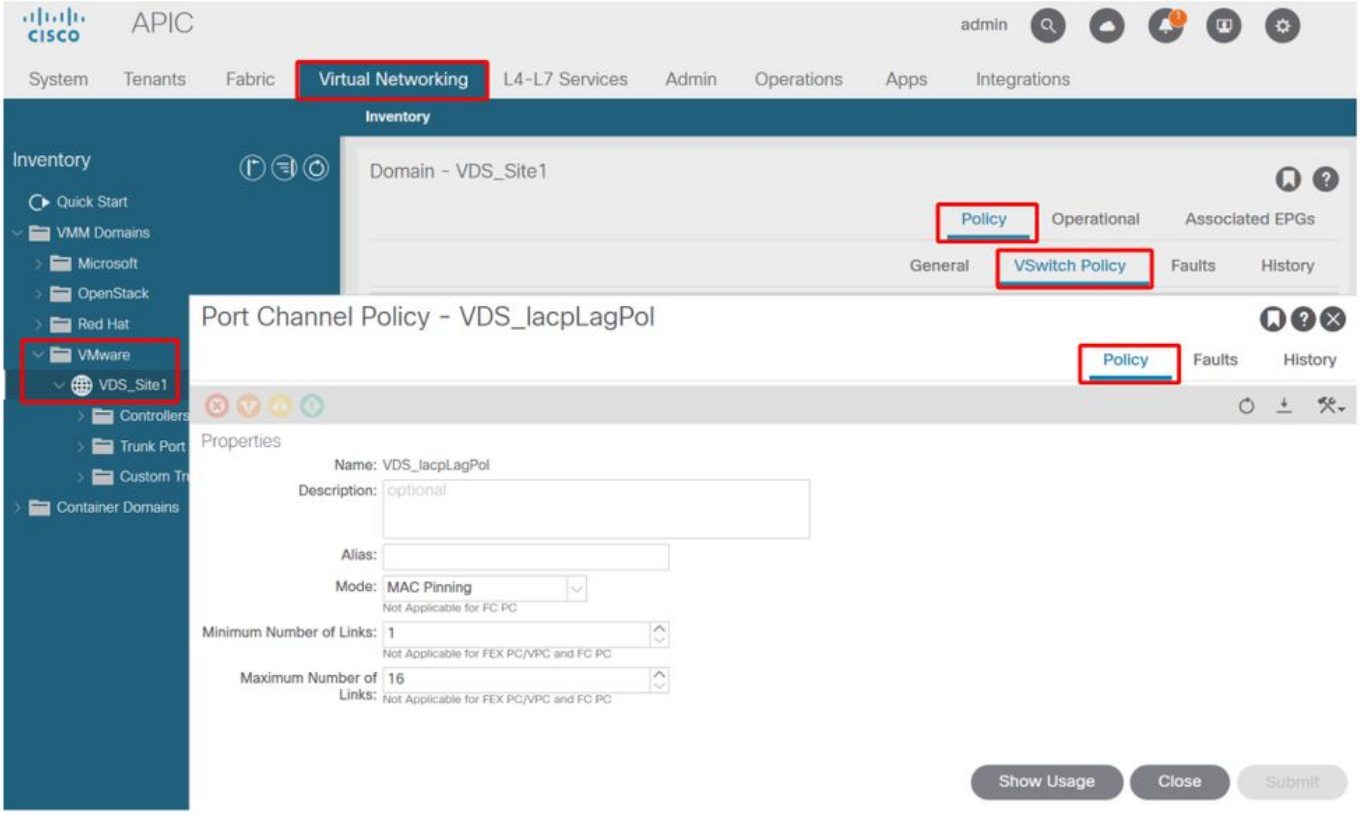
- Each Cisco UCS FI has a port-channel towards the ACI leaf switches.
- The UCS FIs are directly interconnected for heartbeat purposes only (not used for dataplane).
- Each blade server's vNIC is pinned to a specific UCS FI or uses a path toward one of FIs by using UCS Fabric Failover (Active-Standby).
- Using IP-hash algorithms on the ESXi host's vSwitch will cause MAC flaps on the UCS FI’s.

In order to correctly configure this, do the following:



When MAC Pinning is configured on the Port-Channel Policy as part of the vSwitch Policy in ACI, this will show as 'Route based on the originating virtual port' teaming configuration of the port groups on the VDS.

ACI — Port Channel Policy as part of vSwitch Policy



Port Channel Policy - VDS_lacpLagPol

Properties

Name: VDS_lacpLagPol

Description: optional

Alias:

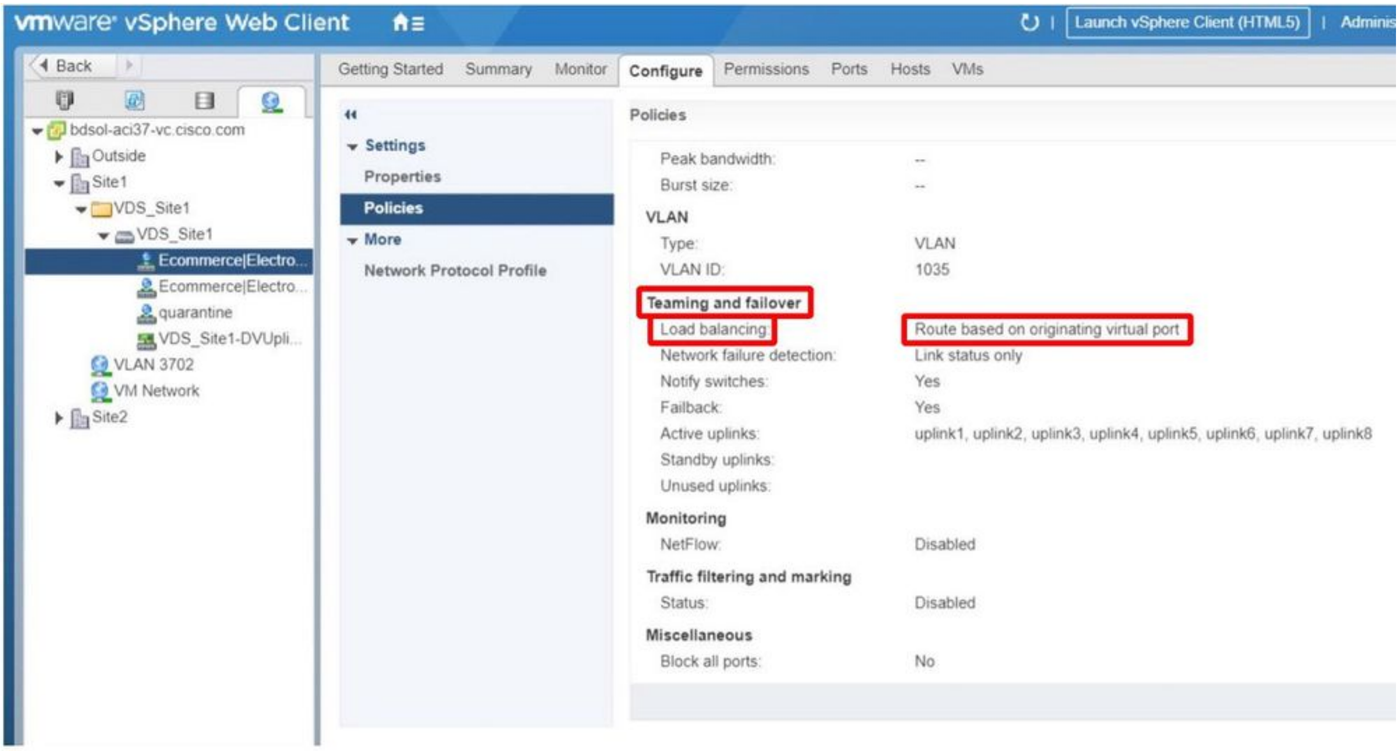
Mode: MAC Pinning
Not Applicable for FC PC

Minimum Number of Links: 1
Not Applicable for FEX PC/VPC and FC PC

Maximum Number of Links: 16
Not Applicable for FEX PC/VPC and FC PC

The Port Channel Policy used in the above example is auto-named by the wizard hence it's called "CDS_lacpLagPol" although we use Mode "MAC Pinning".

VMWare vCenter — ACI VDS — Port Group — Load Balancing setting



Policies

Peak bandwidth: --
Burst size: --

VLAN
Type: VLAN
VLAN ID: 1035

Teaming and failover
Load balancing: Route based on originating virtual port
Network failure detection: Link status only
Notify switches: Yes
Failback: Yes
Active uplinks: uplink1, uplink2, uplink3, uplink4, uplink5, uplink6, uplink7, uplink8
Standby uplinks:
Unused uplinks:

Monitoring
NetFlow: Disabled

Traffic filtering and marking
Status: Disabled

Miscellaneous
Block all ports: No