Configuring SMB Direct with RoCEv2 in Windows

Guidelines for Using SMB Direct support on Windows using RDMA over converged Ethernet (RoCE) v2

General Guidelines and Limitations:

- Cisco UCS Manager release 4.1.x and later releases support Microsoft SMB Direct with RoCEv2 on Microsoft Windows Server 2019 and later. Cisco recommends that you have all KB updates from Microsoft for your Windows Server release.

Note RoCEv2 is not supported on Microsoft Windows Server 2016.

- Cisco recommends you check UCS Hardware and Software Compatibility specific to your UCS Manager release to determine support for Microsoft SMB Direct with RoCEv2 on Microsoft Windows.

- Microsoft SMB Direct with RoCEv2 is supported only with Cisco UCS VIC 1400 Seriesand 15000 Series adapters. It is not supported with UCS VIC 1200 Series and 1300 Series adapters. SMB Direct with RoCEv2 is supported on all UCS Fabric Interconnects.

Note RoCEv1 is not supported with Cisco UCS VIC 1400 Series and Cisco UCS VIC 15000 Series.

- RoCEv2 configuration is supported only between Cisco adapters. Interoperability between Cisco adapters and third party adapters is not supported.

- RoCEv2 supports two RoCEv2 enabled vNIC per adapter and four virtual ports per adapter interface, independent of SET switch configuration.

- RoCEv2 cannot be used on the same vNIC interface as NVGRE, NetFlow, and VMQ features.

- RoCEv2 cannot be used with usNIC.

- RoCEv2-enabled vNIC interfaces must have the no-drop QoS system class enabled in UCS Manager.

- The RoCE Properties queue pairs setting must for be a minimum of 4 queue pairs.

- Maximum number of queue pairs per adapter is 2048.

- The QoS No Drop class configuration must be properly configured on upstream switches such as Cisco Nexus 9000 series switches. QoS configurations will vary between different upstream switches.

- The maximum number of memory regions per rNIC interface is 131072.

- UCS Manager does not support fabric failover for vNICs with RoCEv2 enabled.

- SMB Direct with RoCEv2 is supported on both IPv4 and IPv6.

- RoCEv2 cannot be used with GENEVE offload.

MTU Properties:

- In older versions of the VIC driver, the MTU was derived from either a UCS Manager service profile or from the Cisco IMC vNIC MTU setting in standalone mode. This behavior changes on Cisco UCS VIC 1400 Series and later adapters, where MTU is controlled from the Windows OS Jumbo Packet advanced property. A value configured from UCS Manager or Cisco IMC has no effect.

- The RoCEv2 MTU value is always power-of-two and its maximum limit is 4096.

- RoCEv2 MTU is derived from the Ethernet MTU.

- RoCEv2 MTU is the highest power-of-two that is less than the Ethernet MTU. For example:

    - if the Ethernet value is 1500, then the RoCEv2 MTU value is 1024

    - if the Ethernet value is 4096, then the RoCEv2 MTU value is 4096

    - if the Ethernet value is 9000, then the RoCEv2 MTU value is 4096

Windows NDPKI Modes of Operation:

- Cisco's implementation of Network Direct Kernel Provider Interface (NDPKI) supports two modes of operation: Mode 1 and Mode 2. Mode 1 and Mode 2 relate to the implementation of Network Direct Kernel Provider Interface (NDKPI): Mode 1 is native RDMA, and Mode 2 involves configuration for the virtual port with RDMA. Cisco does not support NDPKI Mode 3 operation.

- The recommended default adapter policy for RoCEv2 Mode 1 is Win-HPN-SMBd .

- The recommended default adapter policy for RoCEv2 Mode 2 is MQ-SMBd.

- RoCEv2 enabled vNICs for Mode2 operation require the QoS host control policy set to full.

- Mode 2 is inclusive of Mode 1: Mode 1 must be enabled to operate Mode 2.

- On Windows, the RoCEv2 interface supports MSI & MSIx interrupt modes. By default, it is in MSIx interrupt mode. Cisco recommends you avoid changing interrupt mode when the interface is configured with RoCEv2 properties.

Downgrade Limitations: Cisco recommends you remove the RoCEv2 configuration before downgrading to any non-supported RoCEv2 release. If the configuration is not removed or disabled, downgrade will fail.

Overview of Configuring RoCEv2 Modes 1 and 2 in Windows

Configuration of RoCEv2 on the Windows platform requires first configuring RoCEv2 Mode 1, then configuring RoCEv2 Mode 2. Modes 1 and 2 relate to the implementation of Network Direct Kernel Provider Interface (NDKPI): Mode 1 is native RDMA, and Mode 2 involves configuration for the virtual port with RDMA.

To configure RoCEv2 mode 1, you will:

- Configure a no-drop class in CoS System Class. By default, Platinum with CoS 5 is a default in UCS Manager.

- Configure an Ethernet adapter policy for Mode 1 in UCS Manager.

- Configure Mode 1 on the host system.

RoCEv2 Mode 1 must be configured before configuring Mode 2.

To configure RoCEv2 mode 2, you will:

- Either create an Ethernet VMQ connection policy for RoCEv2 or use the UCS Manager MQ-SMBd policy.

Windows Requirements

Configuration and use of RDMA over Converged Ethernet for RoCEv2 in Windows Server requires the following:

- Windows 2019 or Windows Server 2022 with latest Microsoft updates

- UCS Manager release 4.1.1 or later

- VIC Driver version 5.4.0.x or later

- UCS M5 B-Series or C-Series servers with VIC 1400 Series adapters: only Cisco UCS VIC 1400 Series or VIC 15000 series adapters are supported.

Note   All Powershell commands or advanced property configurations are common across Windows 2019 unless explicitly mentioned.

All Powershell commands or advanced property configurations are common across Windows 2019 and 2022 unless explicitly mentioned.

Configuring SMB Direct Mode 1 on UCS Manager

To avoid possible RDMA packet drops, make sure same no-drop COS is configured across the network.

Before you begin

Configure a no-drop class in UCSM QoS Policies and use it for RDMA supported interfaces. Go to **LAN** > **LAN Cloud** > **QoS System Class** and enable **Priority** Platinum with CoS 5.

Procedure

Step 1	In the **Navigation** pane, click **Servers**.
Step 2	Expand **Servers** > **Policies**.
Step 3	Expand the node for the organization where you want to create the policy.
	If the system does not include multitenancy, expand the **root** node.
Step 4	Expand **Adapter Policies** and choose the existing adapter policy for Win-HPN-SMBd.

If using a user-defined adapter policy, use the configuration steps below.

a) On the **General** tab, scroll down to **RoCE** and click the **Enabled** radio button.
b) In the **RoCE Properties** field, under **Version 1**, click the **Disabled** radio button. For **Version 2**, click the **Enabled** radio button.

c) For **Queue Pairs**, enter **256**.

d) For **Memory Regions**, enter **131072**.

e) For **Resource Groups**, enter 2.

f) For **Priority**, choose **Platinum No-Drop COS.** from the dropdown.

This setting assumes you are using the default No-Drop policy.

g) Click **Save Change**s.

Step 5   Next, create an Ethernet Adapter Policy. In the Navigation pane, click **LAN**.

Step 6   Expand **LAN** > **Policies**.

Step 7   Right-click the **vNIC Templates** node and choose **Create vNIC Template**.

Step 8   Go to **vNIC Properties** under the General tab and modify the vNIC policy settings as follows:

a) Set **MTU** to **1500** or **4096**.

b) For the Adapter Policy, select **Win-HPN-SMBd**

c) For the **QoS policy**, specify **Platinum**.

Step 9   Click **Save Changes**.

Step 10   After you save the changes, UCS Manager will prompt you to reboot. Reboot the system.

What to do next

When the server comes back up, configure RoCEv2 mode 1 on the Host.

# Configuring SMB Direct Mode 1 on the Host System

You will configure connection between smb-client and smb-server on two host interfaces. For each of these servers, smb-client and smb-server, configure the RoCEv2-enabled vNIC as described below.

Before you begin

Configure RoCEv2 for Mode 1 in UCS Manager.

Procedure

Step 1   In the Windows host, go to the Device Manager and select the appropriate Cisco VIC Internet Interface.
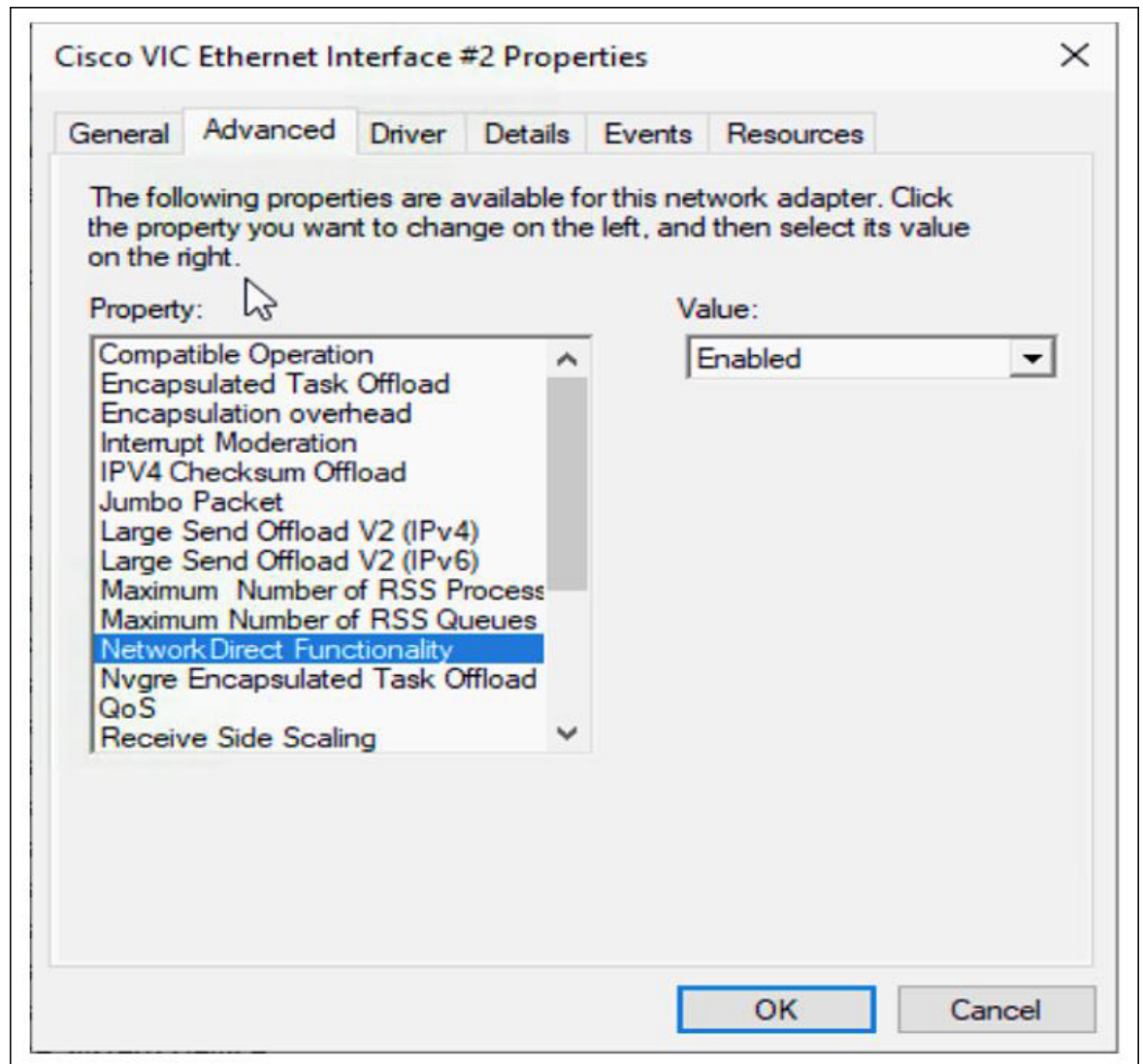
Step 2   Go to **Tools** > **Computer Management** > **Device Manager** > **Network Adapter** > click on **VIC Network Adapter** > **Properties** > **Advanced** > **Network Direct Functionality**. Perform this operation for both the smb-server and smb-client vNICs.
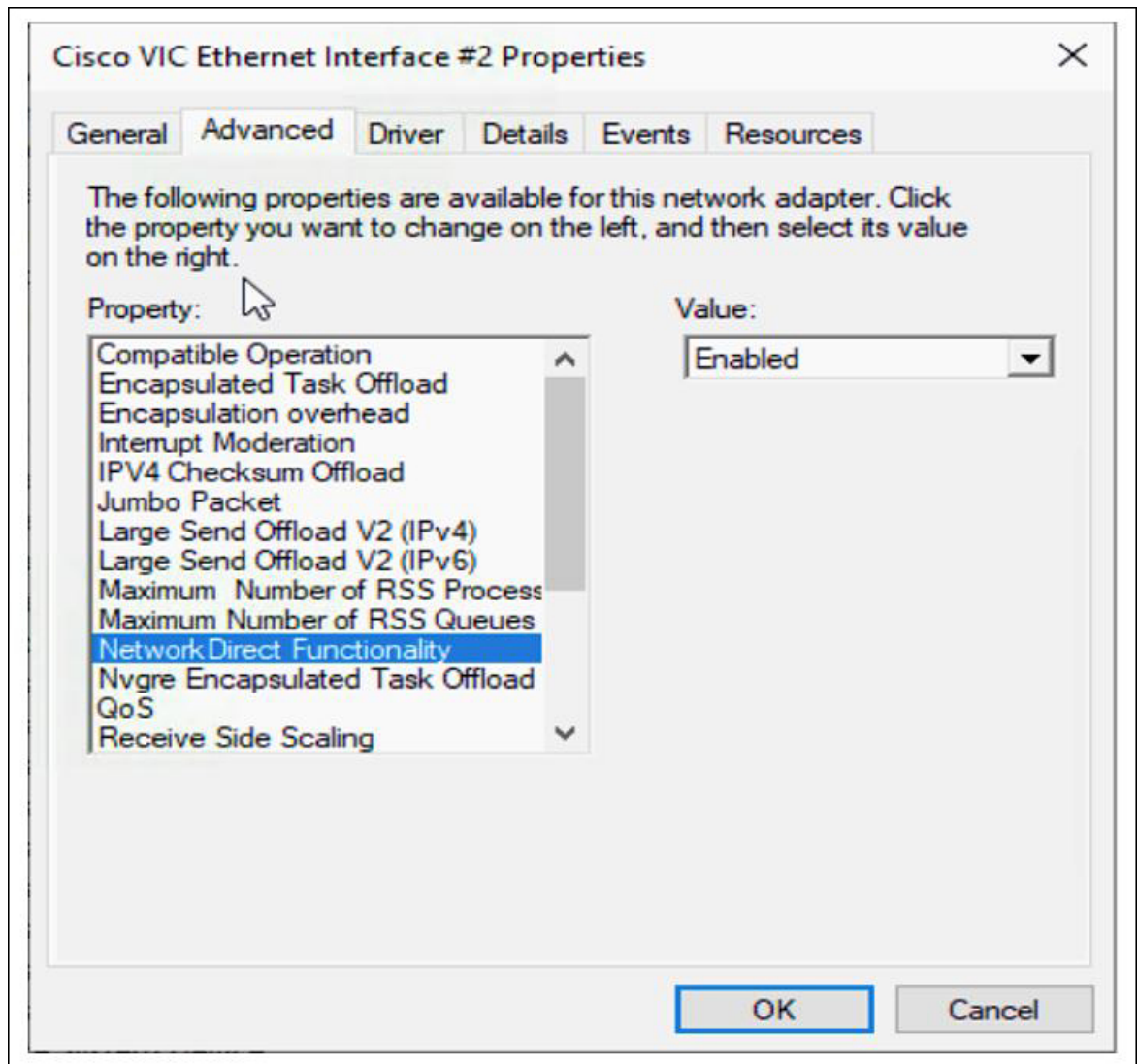
Step 3 Verify that RoCE is enabled on the host operating system using PowerShell.

The `Get-NetOffloadGlobalSetting` command shows NetworkDirect is enabled.

```
PS C:\Users\Administrator> Get-NetOffloadGlobalSetting

ReceiveSideScaling          : Enabled
ReceiveSegmentCoalescing    : Enabled
Chimney                     : Disabled
TaskOffload                 : Enabled
NetworkDirect               : Enabled
NetworkDirectAcrossIPSubnets : Blocked
PacketCoalescingFilter      : Disabled
```

Note If the NetworkDirect setting is showing as disabled, enable it using the command:

```
Set-NetOffloadGlobalSetting -NetworkDirect enabled
```

Step 4 Bring up Powershell and enter the command:

```
get-SmbClientNetworkInterface
```

```
PS C:\Users\Administrator>
PS C:\Users\Administrator> Get-SmbClientNetworkInterface

Interface Index RSS Capable RDMA Capable Speed    IpAddresses        Friendly Name
--------------- ----------- ------------ -----    -----------        -------------
14              True        False        40 Gbps  {10.37.60.162}     vEthernet (vswitch)
26              True        True         40 Gbps  {10.37.60.158}     vEthernet (vp1)
9               True        True         40 Gbps  {50.37.61.23}      Ethernet 2
5               False       False        40 Gbps  {169.254.10.5}     Ethernet (Kernel Debugger)
8               True        False        40 Gbps  {169.254.4.26}     Ethernet 3


PS C:\Users\Administrator>
```

Step 5   Enter **enable - netadapterrdma [-name] ["Ethernetname"]**

**Step 6**   Verify the overall RoCEv2 Mode 1 configuration at the Host as follows:

    a) Use the Powershell command **netstat -xan** to verify the listeners in both the smb-client and smb-server Windows host; listeners will be shown in the command output.

```
PS C:\Users\Administrator>
PS C:\Users\Administrator> netstat -xan

Active NetworkDirect Connections, Listeners, SharedEndpoints

  Mode    IfIndex Type        Local Address          Foreign Address        PID

  Kernel        9 Listener    50.37.61.23:445        NA                     0
  Kernel       26 Listener    10.37.60.158:445       NA                     0
PS C:\Users\Administrator>
```

    b) Go to the smb-client server fileshare and start an I/O operation.

    c) Go to the performance monitor and check that it displays the RDMA activity.
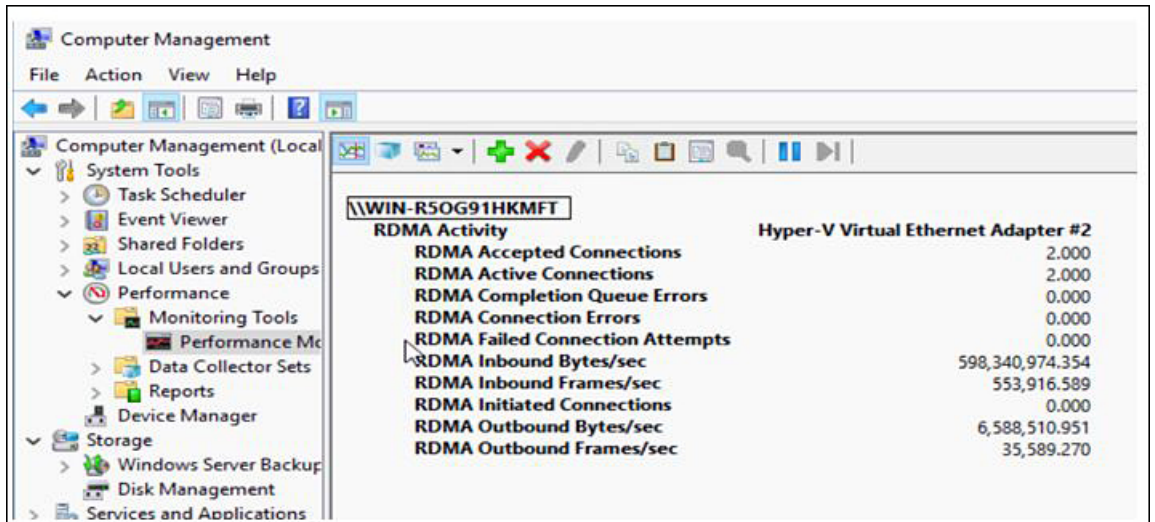
Step 7   In the Powershell command window, check the connection entries with the **netstat -xan** output command to make sure they are displayed. You can also run **netstat -xan** from the command prompt. If the connection entry shows up in netstat-xan output, the RoCEv2 mode1 connections are correctly established between client and server.

```
PS C:\Users\Administrator> netstat -xan

Active NetworkDirect Connections, Listeners, SharedEndpoints

Mode     IfIndex Type        Local Address       Foreign Address      PID

Kernel        4 Connection   50.37.61.22:445     50.37.61.71:2240     0
Kernel        4 Connection   50.37.61.22:445     50.37.61.71:2496     0
Kernel       11 Connection   50.37.61.122:445    50.37.61.71:2752     0
Kernel       11 Connection   50.37.61.122:445    50.37.61.71:3008     0
Kernel       32 Connection   10.37.60.155:445    50.37.60.61:49092    0
Kernel       32 Connection   10.37.60.155:445    50.37.60.61:49348    0
Kernel       26 Connection   50.37.60.32:445     50.37.60.61:48580    0
Kernel       26 Connection   50.37.60.32:445     50.37.60.61:48836    0
Kernel        4 Listener     50.37.61.22:445     NA                   0
Kernel       11 Listener     50.37.61.122:445    NA                   0
Kernel       32 Listener     10.37.60.155:445    NA                   0
Kernel       26 Listener     50.37.60.32:445     NA                   0
```

Note   IP values are representative only.

Step 8   By default, Microsoft's SMB Direct establishes two RDMA connections per RDMA interface. You can change the number of RDMA connections per RDMA interface to one or any number of connections.

For example, to increase the number of RDMA connections to 4, type the following command in PowerShell:

```
PS C:\Users\Administrator> Set-ItemProperty -Path `
"HKLM:\SYSTEM\CurrentControlSet\Services\LanmanWorkstation\Parameters"
ConnectionCountPerRdmaNetworkInterface -Type DWORD -Value 4 -Force
```
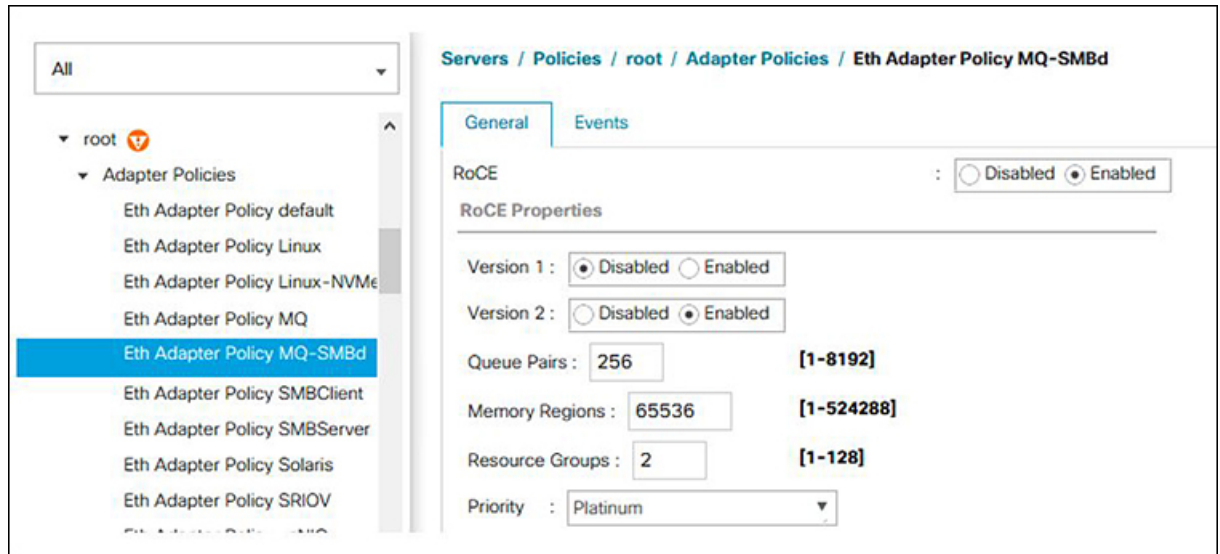
Configuring Mode 2 on UCS Manager

You will apply the VMQ Connection Policy as vmmq.

**Before you begin**

Configure RoCEv2 Policies in Mode 1.

Use the pre-defined default adapter policy "MQ-SMBd", or configure a user-defined Ethernet adapter policy with the following recommended RoCE-specific parameters:

- RoCE: Enabled

- Version 1: disabled

- Version 2: enabled

- Queue Pairs: 256

- Memory Regions: 65536

- Resource Groups: 2

- Priority: Platinum

Create a VMQ connection policy with the following values:

- Multi queue : Enabled

- Number of sub-vNIC: 16

- VMMQ adapter policy: MQ-SMBd
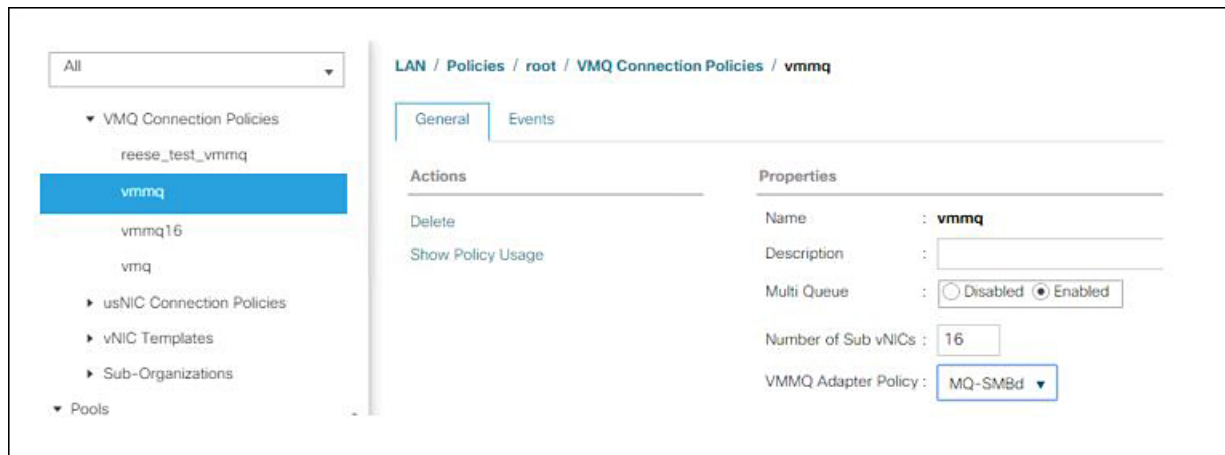
Procedure

Step 1   In the **Navigation** pane, click **Servers**.

Step 2   Expand **Servers** > **Service Profiles**.

Step 3   Expand **Service Profiles** > **vNICs** and choose the VMQ Connection policy profile to configure.

Step 4   Go to **vNIC Properties** under the General tab and scroll down to the Policies area. Modify the vNIC policy settings as follows:

a)   For the Adapter Policy, make sure it uses `Win-HPN-SMBd` or the adapter policy configured earlier for Mode 1.

b)   For the **QoS policy**, select `best-effort`.

Step 5   Click **Save Changes**.

Step 6   In the Navigation pane, click **LAN**.

Step 7   Expand **LAN** > **Policies** > **QoS Policy Best Effort**.

Step 8   Set **Host Control** to `Full`.
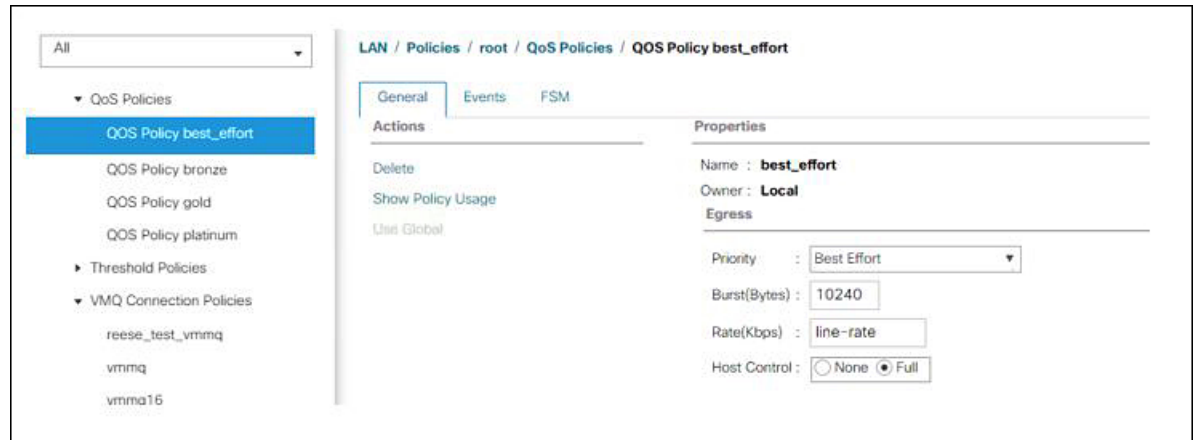
Step 9   Click **Save Changes**.

Step 10   After you save the changes, UCS Manager will prompt you to reboot. Reboot the interface.

What to do next

When the server comes back up, configure Mode 2 on the Host.

Configuring Mode 2 on the Host System

This task uses Hyper-V virtualization software that is compatible with Windows Server 2019 and later.

Before you begin

- Configure and confirm the connection for Mode 1 for both the UCS Manager and Host.

- Configure Mode 2 in UCS Manager.

Procedure

Step 1   Go the the Hyper-V switch manager.

Step 2   Create a new Virtual Network Switch (vswitch) for theRoCEv2-enabled Ethernet interface.

a) Choose **External Network** and select **VIC Ethernet Interface 2** and **Allow management operating system to share this network adapter**.

b) Click **OK** to create the create the virtual switch.
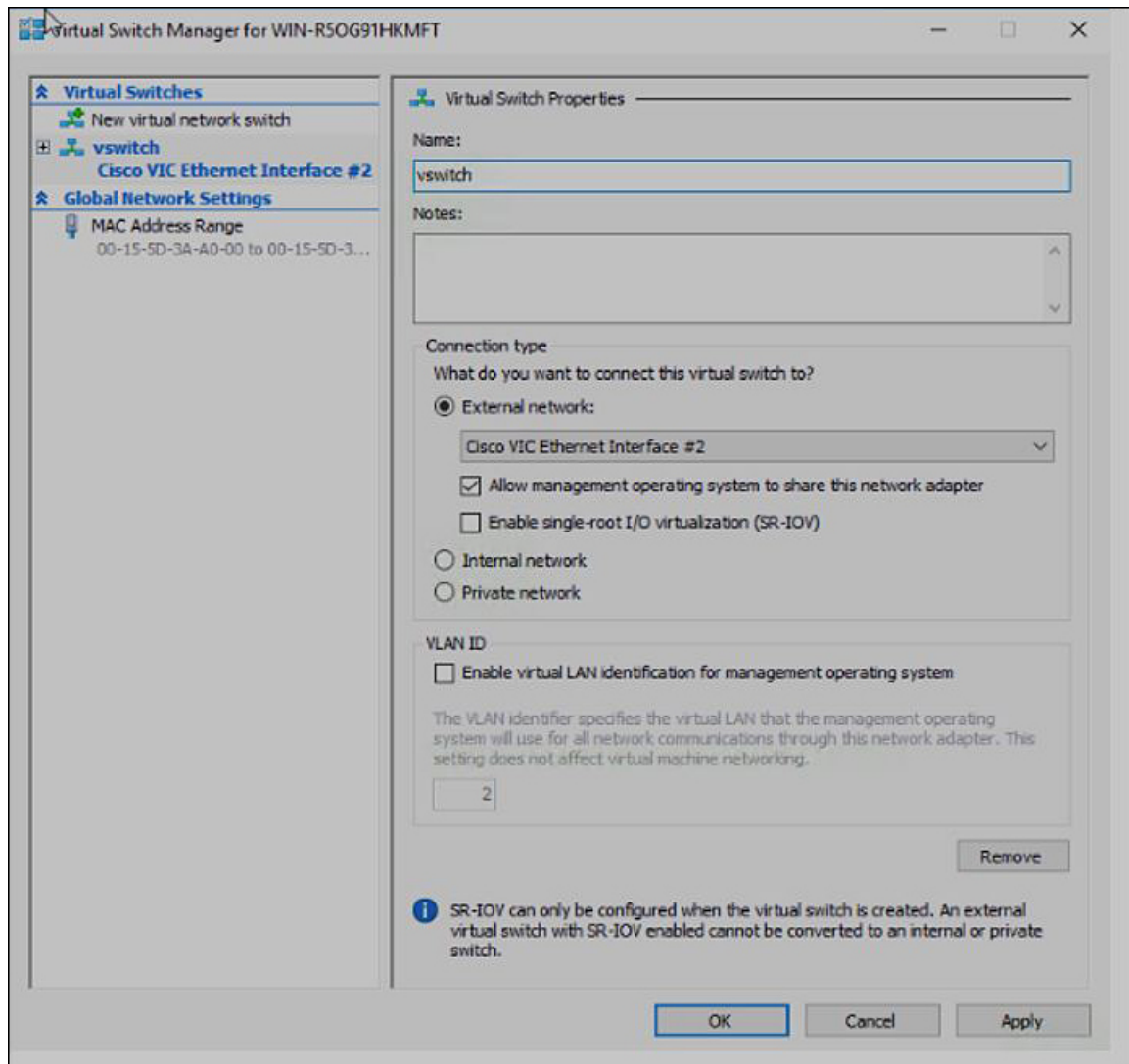
Bring up the Powershell interface.

Step 3 Configure the non-default vport and enable RDMA with the following Powershell commands:

```
add-vmNetworkAdapter -switchname vswitch -name vp1 -managementOS

enable-netAdapterRdma -name "vEthernet (vp1"
```


a) Configure set-switch using the following Powershell command.

```
new-vmswitch -name setswitch -netAdapterName "Ethernet x" -enableEmbeddedTeam $true
```

This creates the switch. Use the following to display the interfaces:

```
get-netadapterrdma

add-vmNetworkAdapter -switchname setswtch -name svp1
```

You will see the new vport when you again enter

```
get-netadapterrdma
```

b) Add a vport.

```
add-vmNetworkAdapter -switchname setswtch -name svp1
```

You will see the new vport when you again enter

```
get-netadapterrdma
```

c) Enable the RDMA on the vport:

```
enable-netAdapterRdma -name "vEthernet (svp1)"
```

Step 4 Configure the IPV4 addresses on the RDMA enabled vport in both servers.

Step 5 Create a share in smb-server and map the share in the smb-client.

a) For smb-client and smb-server in the host system, configure the RoCEv2-enabled vNIC as described above.

b) Configure the IPV4 addresses of the primary fabric and sub-vNICs in both servers, using the same IP subnet and same unique vlan for both.

c) Create a share in smb-server and map the share in the smb-client.

Step 6 Finally, verify the Mode 2 configuration.

a) Use the Powershell command **netstat -xan** to display listeners and their associated IP addresses.

```
PS C:\Users\Administrator>
PS C:\Users\Administrator> netstat -xan

Active NetworkDirect Connections, Listeners, SharedEndpoints

  Mode    IfIndex Type           Local Address           Foreign Address         PID

  Kernel        9 Listener       50.37.61.23:445         NA                      0
  Kernel       26 Listener       10.37.60.158:445        NA                      0
PS C:\Users\Administrator>
```

b) Start any RDMA I/O in the file share in smb-client.
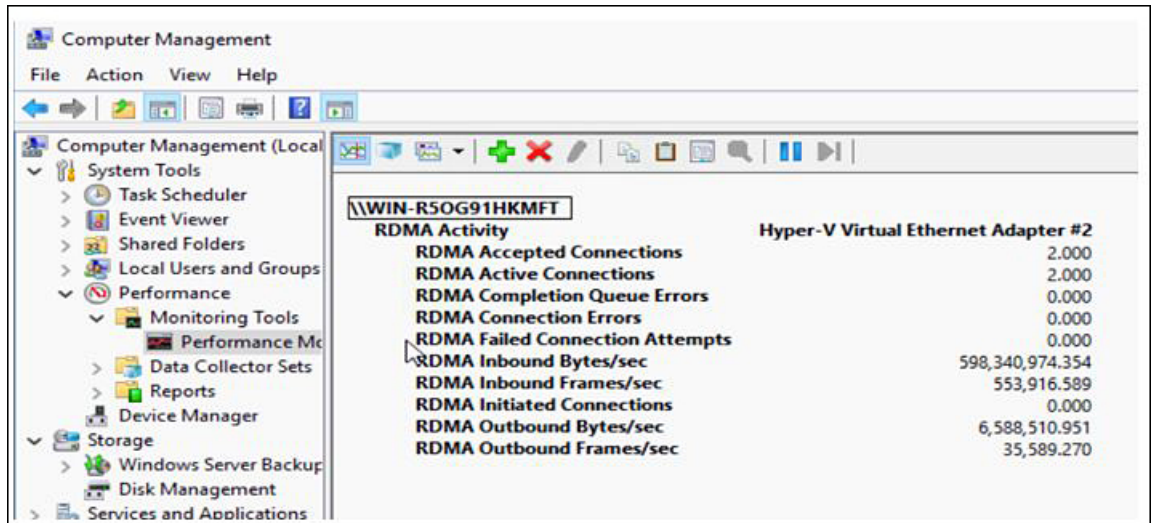
c) Issue the **netstat -xan** command again and check for the connection entries to verify they are displayed.

```
PS C:\Users\Administrator>
PS C:\Users\Administrator> netstat -xan

Active NetworkDirect Connections, Listeners, SharedEndpoints

  Mode   IfIndex Type           Local Address            Foreign Address           PID

  Kernel       9 Connection     50.37.61.23:192          50.37.61.184:445          0
  Kernel       9 Connection     50.37.61.23:448          50.37.61.184:445          0
  Kernel       9 Connection     50.37.61.23:704          50.37.61.214:445          0
  Kernel       9 Connection     50.37.61.23:960          50.37.61.214:445          0
  Kernel       9 Connection     50.37.61.23:1216         50.37.61.224:445          0
  Kernel       9 Connection     50.37.61.23:1472         50.37.61.224:445          0
  Kernel       9 Connection     50.37.61.23:1728         50.37.61.234:445          0
  Kernel       9 Connection     50.37.61.23:1984         50.37.61.234:445          0
  Kernel       9 Listener       50.37.61.23:445          NA                        0
  Kernel      26 Listener       10.37.60.158:445         NA                        0
PS C:\Users\Administrator>
```

What to do next

Troubleshoot any items if necessary.