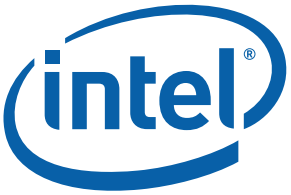


VXLAN Configuration Guide

Intel® Ethernet CNA X710 & XL710 on Red Hat* Enterprise Linux 7.1*

Technical Brief

Networking Division (ND)

October 2016

Revision 1.1
332671-002

Revision History

Revision	Date	Comments
1.1	October 6, 2016	Removed references to FCoE.
1.0	June 23, 2015	Initial release (Intel public).

Contents

NOTE: ***This page intentionally left blank.***

1.0 Introduction

In recent years, Cloud computing has emerged as one of the key usage models for data centers. Cloud computing framework allows data center providers to abstract physical resources from users, which allows sharing of physical resources. Intel® Virtualization technology is one of the key building block for Cloud computing. Data center providers are sharing processor, memory and I/O resources using Intel® Virtualization technology along with software from Microsoft*, VMware*, Citrix*, and Linux* distributors that provide mechanism to provision and manage virtualized hardware. The Intel® Ethernet Converged Network Adapters X710 & XL710 support industry-leading features such as VXLAN stateless offloads that are accelerating the performance and implementation of 10 and 40 Gigabit Ethernet in Cloud computing environments.

As Cloud computing matures, the need to virtualize networking resources has becomes vital. Cloud service providers expect network security and traffic segmentation in multi-tenant Cloud environments. Software Defined Networks (SDN) and Overlay Networks are industry standard techniques designed to achieve Network Virtualization. Network Overlays such as NVGRE (Network Virtualization using Generic Routing Encapsulation) and VXLAN (Virtual Extensible Local Area Network) achieve Network Virtualization by overlaying layer 2 network over physical layer 3 network. These techniques allow network scalability and efficient use of current network infrastructure. Cloud computing frameworks such as Amazon's EC2, VMware's Cloud Infrastructure, and Linux-based Open Stack use NVGRE or VXLAN to Virtualize, provision and manage network resources.

This document shows how to create Virtual Networks using VXLAN tunnels on Intel® Ethernet CNA X710 & XL710 in Red Hat* Enterprise Linux* version 7.

1.1 Intel® Ethernet Controller XL710

The 40 Gigabit XL710 Controller is designed for flexibility, with configurable port speeds of up to 2 x 40 GbE, or 4 x 10 GbE, ensuring a smooth transition to 40 GbE, It also provides a 222% increase in Gigabits per Watt in adapter power for approximately half the power cost when compared to using two previous generation dual-port adapters.

The XL710 offers the following features:

- 10/40 GbE Controller (Dual and Single 40 GbE, Quad and Dual 10 GbE configurations).
- PCI Express* (PCIe) 3.0, x8 including Direct I/O optimizations via TLP Processing Hints (TPH).
- Intelligent Off-load to enable high-performance with Intel® Xeon® servers.
- Network Virtualization off-loads including VXLAN and NVGRE.
- Industry-leading I/O virtualization innovations and performance with broad hypervisor and standards support.
- Intel® Ethernet Flow Director (for hardware application traffic steering).
- Excellent small packet performance for network appliances and NFV.
- Data Plane Developer Kit Optimize.
- Unified Networking providing a single wire for LAN and storage: NAS (SMB,NFS) and SAN (iSCSI).

The following are the Intel 40 Gigabit XL710 Controller-based Dual and Quad Adapter offerings:

Note: These boards do NOT ship with optics installed. Optics must be purchased separately.

- Intel® Ethernet Converged Network Adapter X710-DA4
  - X710DA4FH, XL710DA4FHBLK (Retail, Quad Port FH)
  - X710DA4FHG1P5 (OEM Gen, Quad Port FH)
  - X710DA4G1P5 (OEM Gen, Quad Port LP)
- Intel® Ethernet Converged Network Adapter X710-DA2
  - X710DA2, XL710DA2BLK (Retail, Dual Port)
  - X710DA2G1P5 (OEM Gen, Dual Port)
- Intel® Ethernet Converged Network Adapter XL710-QDA2
  - XL710QDA2, XL710QDA2BLK (Retail, Dual Port)
  - XL710QDA2G1P5 (OEM Gen, Dual Port)
- Intel® Ethernet Converged Network Adapter XL710-QDA1
  - XL710QDA1, XL710QDA1BLK (Retail, Single Port)
  - XL710QDA1G1P5 (OEM Gen, Single Port)

Power efficiency is critical to IT specialists as energy consumption is a real concern in data center operations. The Intel Ethernet Controller provides a low-power interface to eliminate the need for additional power. It also offers the manageability IT personnel require for remote control and alerting.

This controller provides multiple interface options, a smaller footprint for reduced infrastructure and cabling costs, lower power consumption, and intelligent off-loads that do not require disabling key features and flow direction to balance high volume traffic flows.

1.2 Requirements

1.2.1 Hardware Requirements

- An Intel® Ethernet Converged Network Adapter X710 or XL710.
- A server platform that supports Intel® Virtualization Technology for Directed I/O (Intel® VT-d).
- A server platform with an available PCI Express*: x8 5.0 Gb/s (Gen2) or x8 8.0 Gb/s (Gen3) slot.

1.2.2 Software Requirements

- Red Hat Enterprise Linux Version 7.1.
- Intel® Ethernet Converged Network Adapter X710 or XL710 Linux Drivers, available at:

  http://sourceforge.net/projects/e1000/files/
- Linux Bridge Utilities. Typically included with Linux distribution. Red Hat Enterprise Linux version 7.1 includes Linux Bridge Utility version 1.5.

2.0 Installation and Configuration

2.1 VXLAN Network Setup

Figure 1 shows a typical VXLAN tunnel setup. VM-A and VM-B are part of tenant network for ABC Corporation. The Cloud Service Provider has assigned a unique Virtual Network Identifier (VNI) 5000 to ABC Corporation's network. A unique VNI is required for each tenant utilizing the physical network to maintain network security and the tenant's network traffic segregation. Network traffic flows from VM-A on "Host1" to VM-B on "Host2" as follows:

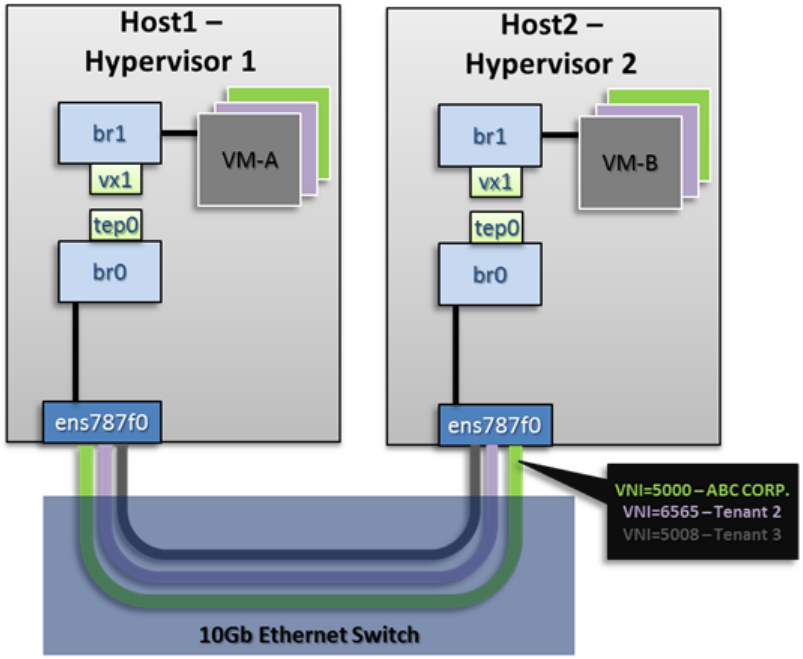


Figure 1. Typical VXLAN Tunnel Setup

Process Flow

1. Traffic from VM-A flows via bridge "br1" using egress internal port "vx1".
2. A VXLAN header is added to the outgoing Ethernet frames and handed to bridge "br0" via ingress internal port "tep0".
3. Ethernet frames with VXLAN header are now placed on the wire using eth2 device destined for VM-B on "Host2".
4. The process is reversed on "Host2," and the Ethernet frame is handed to VM-B

Providing network access to additional tenants is as simple as creating an additional bridge for each tenant using a unique VNI, and creating a corresponding Tunnel Endpoint (TEP) port.

Figure 2 shows the various elements of an Ethernet frame with VXLAN header.

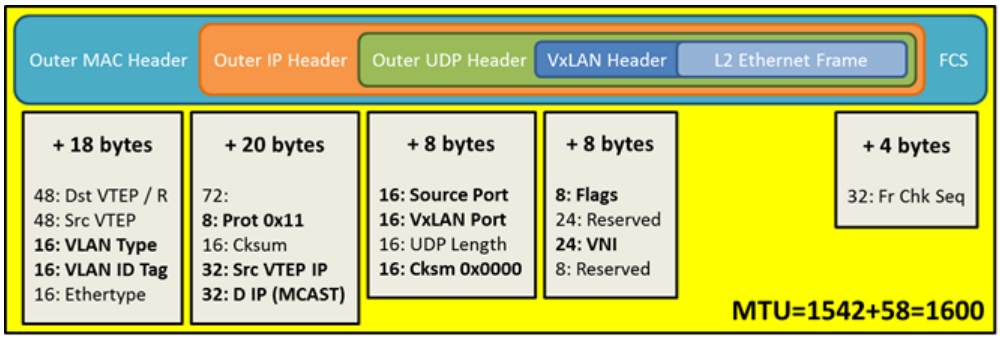


Figure 2. Ethernet Frame with VXLAN Header

Figure 3 shows the corresponding network packet.

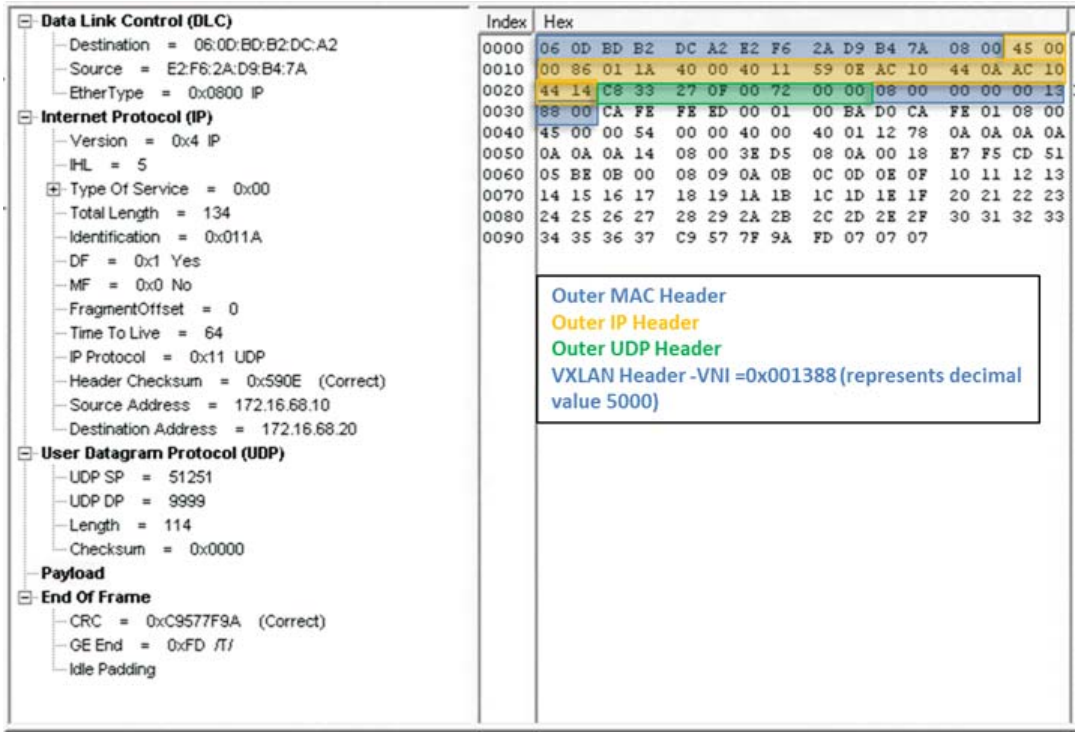


Figure 3. Corresponding Network Packet

2.2 Server Setup

This section shows various setup and configuration steps for enabling SR-IOV on Intel® Ethernet CNA X710 or XL710 server adapters.

1. Install Intel® Ethernet CNA X710 or XL710 server adapter in an available PCI-Express x8 slot. (Ensure that the x8 slot is electrically connected as x8, some slots are physically x8 but electrically support only x4. Verify this with your server manufacturer or system documentation.)
2. Power up the server.
3. Enter the server's BIOS setup and make sure the virtualization technology and Intel® VT-d features are enabled.
4. Install Red Hat Enterprise Linux 7.1 on the server.
5. Make sure all Linux KVM modules, libraries, user tools, and utilities have been installed during the operation system installation.
6. The Red Hat Enterprise Linux installation process may require a server reboot upon successful operating system install.
7. Log in to the newly-installed Red Hat Enterprise Linux operating system using the "root" user account and password.
8. Download and install the latest Intel® Ethernet CNA X710/XL710 Linux driver, available at:

   http://sourceforge.net/projects/e1000/files/

2.3 VXLAN Tunnel Using Open vSwitch

Download, configure, compile, and install Open vSwitch. Follow the instructions provided within the Open vSwitch package.

Figure 4 shows the configuration for Host1, while Figure 5 shows the configuration for Host2.

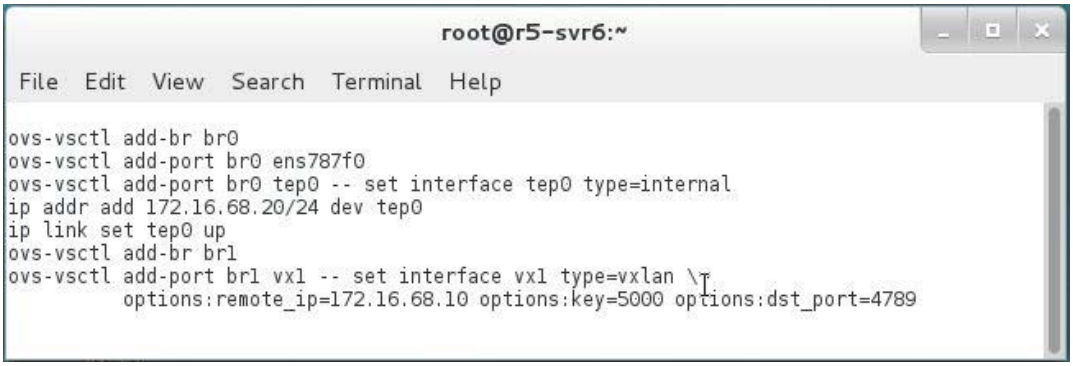


Figure 4. Host1 Configuration

Figure 5. Host2 Configuration

The following command shows Open vSwitch configuration:

```
# ovs-vsctl show
```

Figure 6 shows output of the Open vSwitch configuration for Host1.

Figure 6. Host1 Open vSwitch Configuration

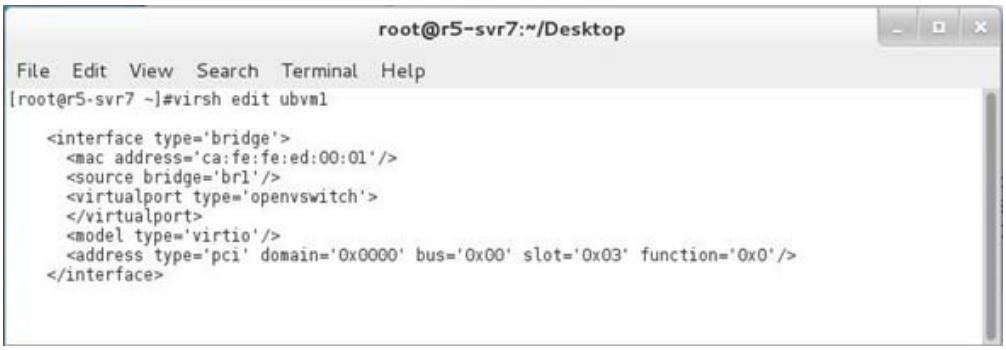
2.3.1 Network Connectivity to Guest Virtual Machines

Most Linux distributions ship with the Kernel Based Virtual Machine (KVM) solution, command line KVM management tools such as **virsh**, and a Graphical User Interface (GUI)-based **virt-manager**. Cloud Infrastructure typically creates Virtual Machines as per tenant's specifications. These VMs communicate over a segregated network using Network Overlays such as NVGRE and/or VXLAN. Cloud Infrastructure Service Providers may choose to implement either NVGRE, VXLAN, or both for providing network connectivity to their tenants.

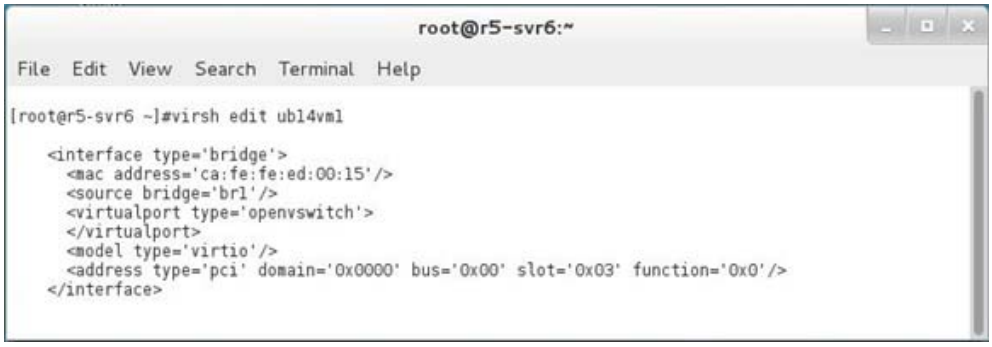
The following sections shows how to configure VMs to communicate over VXLAN. Unfortunately, the GUI-based VM management tool **virt-manager** does not allow VM network interface assignment to Open vSwitch at the time this document was authored. The **virsh** tool is required to edit a VM's configuration. Use the following command on Linux console to edit the VM settings:

```
virsh edit <virtual_machine_name>
```

Add the following section to device section of the file:



In the following example, the VM is configured to use VXLAN network configuration from Section 2.3. Each VM must be configured for unique MAC address.



2.4 VXLAN Tunnel Using Linux Bridge

The Linux Bridge utility package is installed as a part of Virtualization software group during Red Hat Enterprise Linux installation.

Figure 7 shows typical virtual network setup.

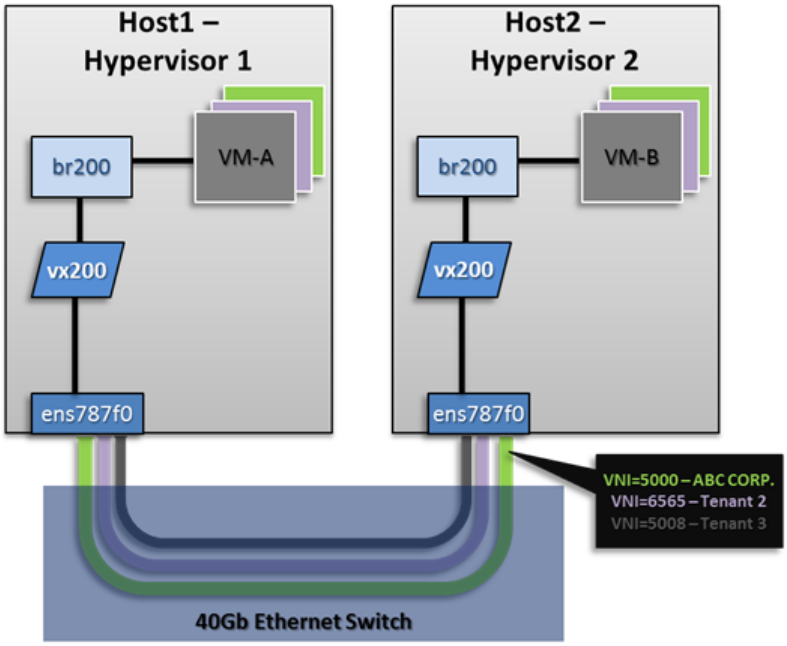


Figure 7. Typical Virtual Network Setup

Section 2.4.1 shows how to VXLAN virtual network using Linux Bridge Utilities.

Figure 8 shows the configuration for Host1.

```
root@r5-svr7:/
File  Edit  View  Search  Terminal  Help

[root@r5-svr7 /]#
[root@r5-svr7 /]# ip link set ens787f0 mtu 1600
[root@r5-svr7 /]# ip link add br200 type bridge
[root@r5-svr7 /]# ip link add vx200 type vxlan id 5000 group 239.1.1.1 dstport 4789 dev ens787f0
[root@r5-svr7 /]# brctl addif br200 vx200
[root@r5-svr7 /]# ip link set br200 up
[root@r5-svr7 /]# ip link set vx200 up
[root@r5-svr7 /]# ifconfig ens787f0 172.16.10.7/24
[root@r5-svr7 /]# ip link set ens787f0 up
[root@r5-svr7 /]#
[root@r5-svr7 /]#
```

Figure 8. Host1 Configuration

Table 1 contains descriptions of the commands used in the example in Figure 8.

Table 1. Command Descriptions

Command	Description
`ip link set ens787f0 mtu 1600`	Configures MTU size of 1600 for port 1 of the Intel® Ethernet CNA X710 or XL710 server adapter.
`ip link add br200 type bridge`	Creates Linux bridge named "br200".
`ip link add type vxlan id 5000 group 239.1.1.1 dstport dev ens787f0`	Creates VXLAN interface named "vx200" and assigns VNI value of 5000. Sets up Multicast group 239.1.1.1 registration for resolving ARP requests. Assigns standard UDP Port 4789 for VXLAN traffic.
`ip link set br200 up`	Changes the state of br200 bridge to UP/ACTIVE.
`ip link set vx200 up`	Changes the state of vx200 interface to UP/ACTIVE.
`ifconfig ens787f0 172.16.10.7/24`	Assigns IP address to the physical port, also known as Tunnel End Point (TEP).
`ip link set ens787f0 up`	Changes the state of physical interface (TEP) interface to UP/ACTIVE.

Figure 9 shows the configuration for Host2.

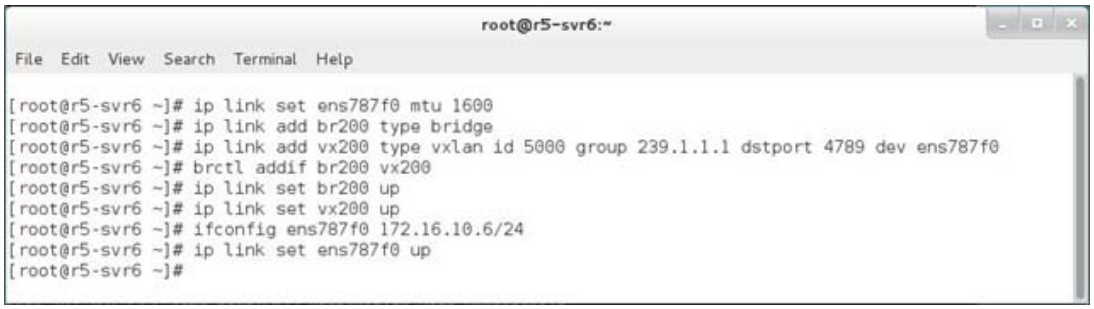


Figure 9. Host2 Configuration

2.4.1 Network Connectivity to Guest Virtual Machines

This section shows how to configure VMs to communicate via VXLAN. The GUI-based VM management tool **virt-manager** is used to assign a VM network interface to Linux Bridge "br200".

Figure 10 shows VM setup screen.

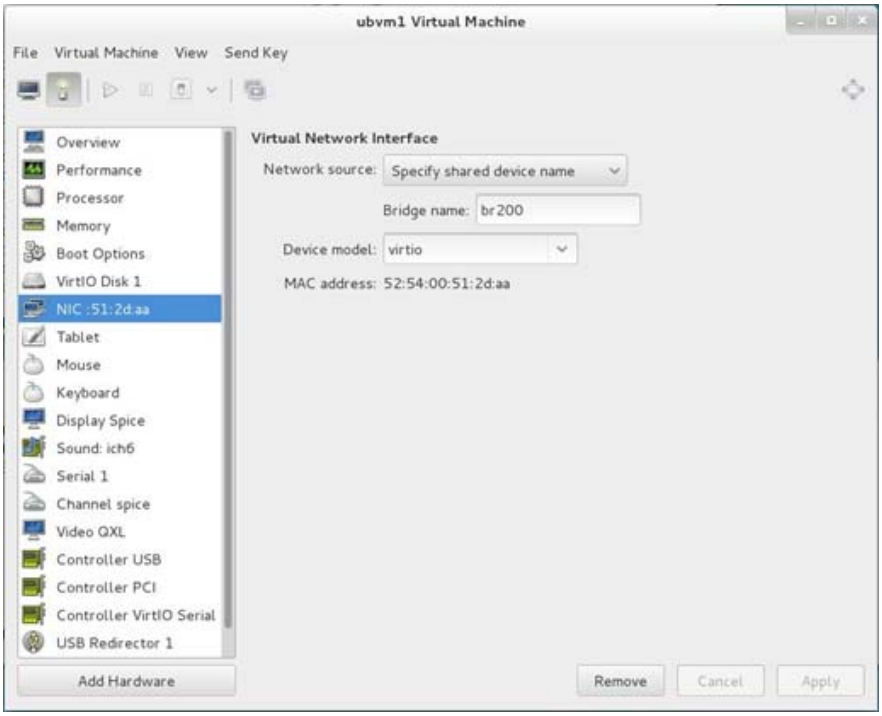


Figure 10. VM Setup Screen

2.5 Setup Validation

Intel® Ethernet CNA X710 & XL710 products supports Network Virtualization offloads including Geneve, VXLAN, and NVGRE. Support for Network Virtualization Offloads is enabled by default on Linux driver (i40e) for Intel® Ethernet CNA X710 & XL710 products. The following command allows users to confirm Network Virtualization offloads support status.

```
ethtool -k <ethernet_interface_name>
```

Figure 11 shows the Network Virtualization offload feature named `tx-udp_tnl_segmentation` by i40e Linux driver status.

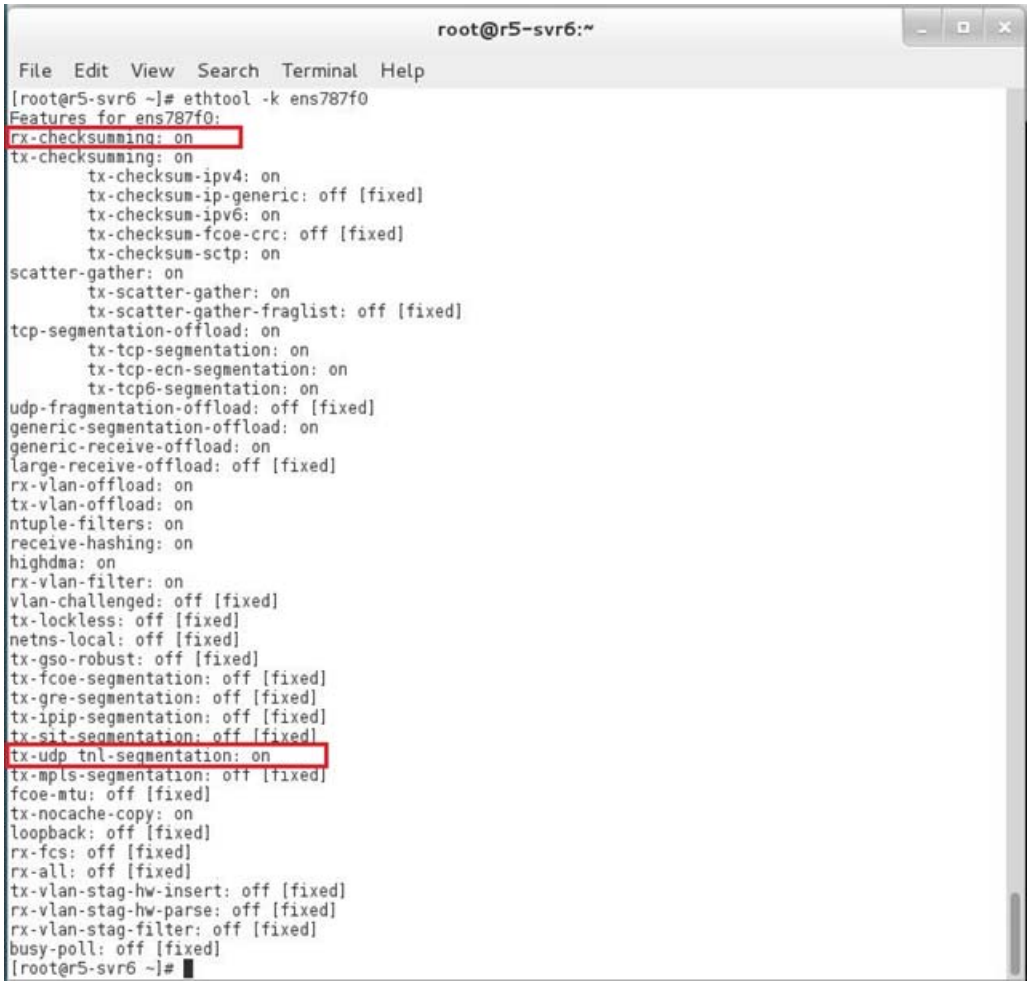


Figure 11. Network Virtualization Offload Feature

Network Virtualization offloads provide improved network throughput performance. The user can experiment with effects on Network throughput performance by disabling and enabling the RX-Checksum feature while running **iperf** or **netperf** (common network performance measurement tools).

The following command allows the user to disable or enable the RX-Checksum feature:

```
ethtool -K <ethernet_interface_name> rx-checksum off
```

2.6 Recommendations/Suggestions

1. To take advantage of new features that are being added to Linux, consider updating your Linux Kernel as often as possible using your Red Hat Subscription. Updated Linux Kernel may provide additional performance benefits as well.
2. Use the **set_irq_affinity** script for optimal interrupt affinitization across available processor cores. This script is included in the Intel® Ethernet CNA X710/XL710 Linux i40e driver package. Linux distributions use a generic **irqbalance** service to balance interrupts across available processor cores. The **irqbalance** service must be disabled before executing the **set_irq_affinity** script.

   Use the following commands to manage services:

   `systemctl status irqbalance` (shows service status)

   `systemctl stop irqbalance` (stops the service)

   `systemctl disable irqbalance` (disables the service)
3. VXLAN packets are larger than the standard Ethernet packets. Maximum Transmission Unit (MTU) size must be increased to accommodate VXLAN packets. Ensure that all physical switch ports participating in VXLAN network are configured to MTU size of 1600.

3.0 Summary

Intel's best-of-breed 10 and 40 GbE solutions are now available with Network Virtualization capabilities. Customers get world-class Ethernet support along with Network virtualization support in mainstream Linux distributions in a single adapter. Intel® Ethernet CNA X710 and XL710 server adapters provide hardware offloads that deliver significant performance improvements for Network virtualization deployments.

4.0 Customer Support

Intel® Customer Support Services offers a broad selection of programs, including phone support and warranty service. For more information, contact us at:

support.intel.com/support/go/network/ adapter/home.htm

Service and availability may vary by country.

5.0 Product Information

To see the full line of Intel Network Adapters for PCI Express, visit www.intel.com/go/ethernet.

To speak to a customer service representative regarding Intel products, please call 1-800-538-3373 (U.S. and Canada) or visit support.intel.com/support/go/network/contact.htm for the telephone number in your area.

NOTE: ***This page intentionally left blank.***

LEGAL