VMware Network Throughput on AMD EPYC™ with Mellanox 100GbE NIC

Publication # **56354**	Revision: **1.00**
Issue Date: **June 2018**	

Advanced Micro Devices

Revision History

Date	Revision	Description
June 2018	1.00	Initial public release.

Software Requirements for Achieving 100Gbps using a Mellanox NIC

Some customers may find that when running VMware ESXi on AMD EPYC™ systems with Mellanox 100Gbps NICs, they are unable to achieve the network throughput that is expected. This can be addressed with the software deployment and optimizations detailed below.

AMD IOMMU Driver

To achieve the advertised throughput on a Mellanox ConnectX-4 or ConnectX-5 based Network Interface Card, the latest version of the AMD IOMMU driver released by VMware must be installed. The IOMMU driver is published and maintained by VMware; however, it is not released asynchronously for ESXi.

The table below contains the minimum versions of VMware ESXi for the updated AMD IOMMU driver:

ESXi Major Release	Minimum Level Required
ESXi 6.5	U2 or later
ESXi 6.7	TBD

Mellanox ConnectX-4 and ConnectX-5 100GbE NIC Driver

An additional requirement to achieve the advertised throughput in VMware is to install the latest Mellanox driver found on the VMware website. If version 4.16.13.5 or later is installed, the interrupts will properly pin to the NUMA node where the adapter is attached. This driver can be found on the VMware website.

Once the driver has been downloaded, extract the `offline_bundle` ZIP file, copy it the ESXi host through SCP, or the web interface, and install it.

Once the driver is installed, the host must be rebooted for changes to take effect.

VMware ESXi Modifications

The following settings are not strictly required but can help achieve a higher, sustained throughput in certain scenarios. The first group of settings only need to be applied on the ESXi host once. The second set of settings will need to be applied every time after the host boosts up.

Connect to the ESXi shell through the console or SSH, issue the following commands, and reboot the host:

```
esxcfg-advcfg -s 65535 /Net/VmxnetLROMaxLength
esxcli system settings kernel set -s netMaxPktsToProcess -v 128
esxcli system settings kernel set -s intrBalancingEnabled -v false
```

After the host has booted but before any virtual machines have powered on, issue the following commands (where vmnicX is the vmnic associated with the Mellanox adapter):

```
esxcli network nic ring current set -r 4096 -n vmnicX
esxcli network nic coalesce set -a false -n vmnicX
```

Test Results

Once the above changes are made, achieving line rate should be possible in a virtualized environment. In our test environment, two hosts were configured with Mellanox ConnectX-4 100Gbps NICs and connected back to back. We created eight virtual machines running Ubuntu 17.10, installed iperf version 2.xx, and ran iperf in server mode on the VMs receiving traffic, and ran iperf in client mode on the VMs sending traffic. Each VM was sending traffic in a single stream (-P 1) directly to a specific VM on the target host. When testing both Mellanox ConnectX-4 and ConnectX-5 in this manner, AMD measured ~80 Gbps on a single socket platform.