Guide to NetworkDirect Logo Testing

Author: Eric Lantz (elantz@microsoft.com)

Date: August 8, 2008

Version: 3 (Release)

References

- Windows HPC Server 2008 SDK on MS Connect (ND interface definition, test executables and test sources, HPCS2008 software (including WinSrv2008 OS):
  https://connect.microsoft.com/site/sitehome.aspx?SiteID=12
- Windows Logo Kit (DTM 1.2: LAN tests only):
  http://www.microsoft.com/whdc/winlogo/WLK/default.mspx
- NetworkDirect Logo Requirements on LogoPoint:
  (winqual account required to use LogoPoint but not for these requirements links)
  - http://www.microsoft.com/whdc/winlogo/hwrequirements.mspx
  - http://download.microsoft.com/download/d/e/1/de1e0c8f-a222-47bc-b78b-1656d4cf3cf7/WLP-Reqs-DEVICE_03-21-08.pdf

- Post Your Drivers on Windows Update (make your customer's life easier!):
  http://www.microsoft.com/whdc/winlogo/drvsign/WinUp.mspx#
- MPICH Test Suite at Argonne National Lab:
  http://www-unix.mcs.anl.gov/mpi/mpi-test/tsuite.html
- Windows stack in OpenFabrics Alliance:
  https://wiki.openfabrics.org/tiki-index.php?page=OpenIB+Windows

Scope of this Document

This document assumes the reader is familiar with the NetworkDirect Service Provider Interface (ND SPI) which is used for RDMA-style networking in Windows HPC Server 2008 and describes the process for logo testing and digitally signing that provider. Testing with both the current version of the Windows Logo Kit Device Test Manager (version 1.2) and the upcoming DTM release (version 1.3) are discussed.

NetworkDirect Logo Testing – The Big Picture

The Microsoft High Performance Computing (HPC) team has created a new, high-performance, RDMA networking interface for Windows named NetworkDirect. A portion of this project has involved working with networking hardware vendors to make drivers (called a NetworkDirect Provider) for this new interface. To enable computer OEMs to offer these high-performance drivers bundled in their Windows-logo'd systems, the providers must be Microsoft-certified via the NetworkDirect and LAN logo programs.

The NetworkDirect logo program is comprised of 2 sets of logo tests which are run in a series of Device Test Manager (DTM) "jobs" for a given driver and device:

1. NetworkDirect Logo Tests
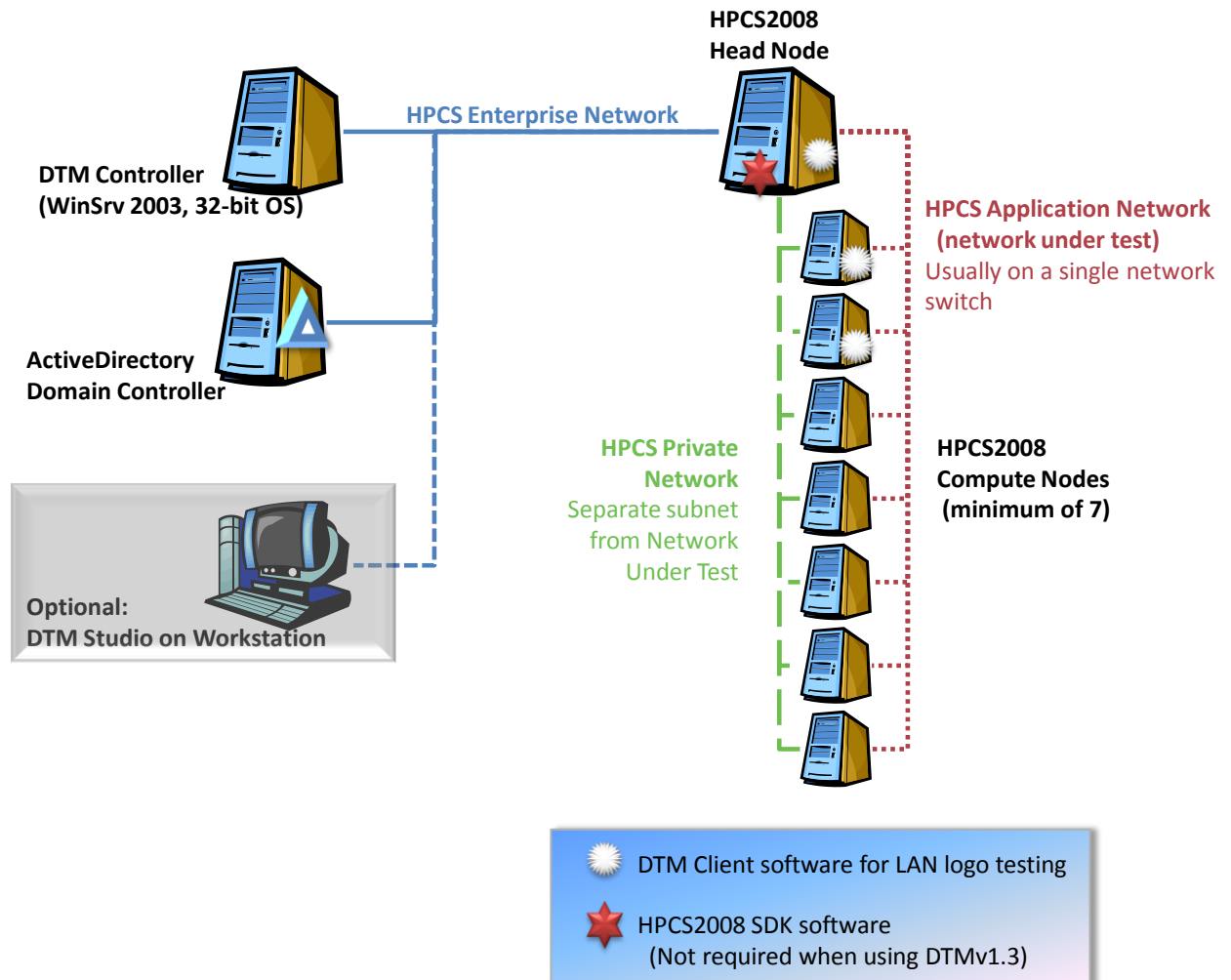   The NetworkDirect logo tests are run on an 8-node (minimum) cluster and include both:
   - Functional tests of the ND SPI implementation
   - Message Passing Interface (MPI) tests which verify the provider in key usage scenarios for MPI applications

2. LAN Logo Tests
   The NetworkDirect interface relies on an NDIS mini-port driver for both: a) IP address translation (to native address of the network fabric); and b) communications fail-over in the event of a failure to connect via NetworkDirect. Thus, the NDIS driver must be verified with the LAN logo test for Windows Server 2008 which runs between 2 nodes using the Windows Logo Kit. The LAN logo tests cover dynamic partitioning and a small number of PCI tests in addition to the core LAN tests.

What You'll Need to Run the NetworkDirect Logo Tests

The hardware/software setup for ND logo testing is diagrammed below:

Hardware

1. Eight (8) server-class computers from which to form a Windows HPC Server 2008 cluster.

    - 64-bit architecture (x64)
    - AMD Opteron, AMD Athlon 64, or Intel EM64T
    - 6 GB of RAM (minimum)
    - 50 GB of Disk (minimum)

- One (1) of the cluster computers must have a minimum of 3 network cards. The remaining seven (7) cluster computers must have a minimum of 2 network cards (one of which is the device under test)

2. One (1) server-class computer upon which the Windows Logo Test DTM Controller will be installed
3. Management network connecting all servers (network cards, switching)
4. Network under test (network cards, switching)

Software
1. Window HPC Server 2008 (pre-release version on MS Connect)
   With which you'll construct an 8-node cluster
2. Windows HPC Server 2008 Software Development Kit (SDK) (pre-release version on MS Connect)
   Includes the ND Logo tests and their source code
3. Windows Logo Kit which includes the Device Test Manager (DTM)
   Used to run the LAN logo test
4. Windows Server 2003 SP1 or SP2, 32-bit Operating System
   Used for the DTM Controller
5. Winqual vendor account
   Needed to submit your LAN logo results and ND+LAN drivers for signature.
   Also used for paying your test fees (about $250 per submission)

The Process

Process for ND Logo Testing BEFORE Windows 7 Beta [targeted for Nov2008] (using Manual Process and DTMv1.2)

The NetworkDirect logo tests will be integrated with version 1.3 of the Device Test Manager framework which is expected to ship in November 2008. Prior to November, hardware vendors can logo test their NetworkDirect providers via the following "manual" process:

1. Create an 8-node cluster (see instructions and diagram above) using Windows HPC Server 2008 (HPCS2008) which includes a Private network (usually a simple 1-Gigabit Ethernet subnet) and an Application network composed of the networking hardware under test.
2. Download and install the Windows HPC Server 2008 SDK onto the head node of your HPCS2008 cluster. The HPCS2008 SDK which includes the ND tests and a script to run them named *run-NDLogoSuite.cmd*.
3. Download and install the Windows Logo Kit v1.2 which is a testing framework for the LAN tests.

   a) Install the DTM controller on a Windows Server 2003 SP1 or SP2, 32-bit Operating System (you may install a 32-bit OS on 64-bit hardware if you wish)
   b) Install the DTM Client software on 2 of the compute nodes in your HPCS2008 cluster for the LAN testing. One of these machines must have 6GB RAM.

c) Install the DTM Studio software, which is used to start and monitor your LAN logo tests, on either the DTM controller computer or another computer on the management network (such as your personal workstation).

4. Run the ND logo tests

   a) Use the HPCS2008 SDK to run the ND tests

      1. Log into the cluster head node as an administrator, and navigate to the ND scripts folder: [HPCS2008 SDK root]\NetworkDirect\scripts

      2. Right-click on the ND logo testing script, *run-ndLogoSuite.cmd*, and choose "Run as Administrator".
         This will run all functional and MPI logo tests and collect the results a zip file which is placed in this shared folder on the cluster head node:
         [HPCS2008 SDK root]\NetworkDirect\scripts\results\

      3. Iterate until the ND provider passes all the ND tests from the HPCS2008 SDK or you have secured a waiver for failing tests from NDLogos@microsoft.com. Note that all tests must be rerun after any change in the driver.

   b) Use the DTM Device Explorer to run the LAN tests

      1. Instructions for using the DTM are available in the DTM Studio user interface and in this Windows Logo Kit whitepaper:
         http://www.microsoft.com/whdc/DevTools/WDK/DTM/DTMhowto.mspx

      2. Iterate until the NDIS driver passes all the LAN tests in DTMv1.2 or you have secured a waiver for failing tests from NDLogos@microsoft.com. Note that all tests must be rerun after any change in the driver.

5. Submit your passing ND test results to Microsoft

   a) Email the following to NDLogos@microsoft.com

      1. ND test result cab file from the following directory on the head node:
          [HPCS2008 SDK root]\NetworkDirect\scripts\results\

      2. ND provider under test (binaries with INF installer)

   b) The Microsoft HPC team will review your test results and perform additional verification tests on your ND provider at scale. You will be notified of any failures or discrepancies. You will receive a digitally signed email from Eric Lantz (MS HPC team) including the checksum of your provider/installer when the ND provider has passed inspection. SAVE THIS SIGNED EMAIL.

6. Submit your passing LAN test results to Microsoft

   a) Create a CPK package of your LAN test results and your ND/LAN driver and installer per the instructions in the DTM.

      • CRITICAL STEP:
        Include the certification of your ND results (the signed email from step 5c above) in your

LAN CPK. You can include this email by [TBD steps in the DTM].

Also, please add this note in the notes section of the submission: "This submission is for both the NetworkDirect and LAN logo programs. Winqual team please contact "Darin Zhen" if any manual review is required."

    b) Refer to the [Winqual Web site](#) and then navigate to under "Windows Logo Program" select hardware and then create logo submission using your CPK results package.

    c) WinQual will digitally sign your ND/LAN driver and return it to you.

7. Thank You and Congratulations! You're done.

Process for ND Logo Testing AFTER Windows 7 Beta [targeted for Nov2008] (using DTMv1.3)

When the Windows Logo Kit v1.3 ships (coincident with Windows 7 Beta) the ND logo program will be fully integrated into the WLK and thus all ND testing and results collection will be driven through the DTM just like the LAN logo program today. ***There will be no need to run the ND tests manually and submit logs to NDsubmission.*** Instead two logo submissions will be done through the WinQual website- one for ND and one for LAN.

1. The current plan is for no changes to the hardware/software configuration described above.

    a) As with the DTMv1.2, create an 8-node cluster using Windows HPC Server 2008 (HPCS2008) which includes a management network (usually a simple 1-Gigabit Ethernet subnet) and a network composed of the networking hardware under test.

2. Download and install the Windows Logo Kit v1.3 which is a testing framework for the logo tests.

    a) Install the DTM controller on a Windows Server 2003, 32-bit Operating System (you may install a 32-bit OS on 64-bit hardware if you wish)

    b) Install the DTM Client software on:
- The head node of your HPCS2008 cluster
- Two (2) of the compute nodes in your HPCS2008 cluster

    c) Install the DTM Studio software, which is used to start and monitor your logo tests, on either the DTM controller computer or another computer on the management network (such as your personal workstation).

3. Run the ND and LAN logo tests with the DTM

    a) Use the DTM Device Explorer to create a NetworkDirect logo submission and run the tests

    b) Iterate until the ND provider and NDIS driver pass all the tests in DTMv1.3 or you have secured a waiver for failing tests from the WinQual team.

4. Submit your passing test results to Microsoft

a) Create a CPK package of your ND and LAN test results and your ND/LAN driver and installer per the instructions in the DTM.
b) Refer to the Winqual Web site and then navigate to under "Windows Logo Program" select hardware and then create logo submission using your CPK results package.
c) WinQual will digitally sign your ND/LAN driver and return it to you.

5. Thank You and Congratulations! You're done.

Background Information

Current NetworkDirect Logo Requirements

Requirement ID	Title	Group Name	Form Factor
NETWORK-0229	NetworkDirect devices must adhere to the NetworkDirect SPI specification	Network Devices	Server2008 - Server Device
NETWORK-0230	NetworkDirect devices must be installable with WIM driver injection	Network Devices	Server2008 - Server Device
NETWORK-0231	NetworkDirect devices must implement an NDIS interface	Network Devices	Server2008 - Server Device
NETWORK-0235	NetworkDirect devices must properly services key MPI usage patterns	Network Devices	Server2008 - Server Device
NETWORK-0236	NetworkDirect devices must meet the specified minimum for data transfer rate, latency, and connection creation rate	Network Devices	Server2008 - Server Device
NETWORK-0239	NetworkDirect devices must register memory to be used for communication at a rate and scale that meets the specified minimum	Network Devices	Server2008 - Server Device
NETWORK-0240	NetworkDirect devices must exhibit fewer than specified maximum number of data errors during data integrity tests	Network Devices	Server2008 - Server Device
NETWORK-0241	NetworkDirect devices must access remote memory solely within prescribed byte-level	Network Devices	Server2008 - Server

	boundaries			Device

NetworkDirect Logo Tests – Unit Tests

Test Name	Test Description/ ND-Requirement	# Nodes	#Proc/ Node	Command
1. ndping.exe	Send/recv bandwidth, polling and blocking mode. NETWORK-0236	2	1	Polling mode: Client: ndping c <server IP> <server port> p1 Server: ndping s <server IP> <server port> p1 Blocking mode: Client: ndping c <server IP> <server port> b1 Server: ndping s <server IP> <server port> b1
2. ndrping.exe	RDMA write/read bandwidth, polling and blocking mode. NETWORK-0236	2	1	Polling mode: Client: ndrping c <server IP> <server port> p1 Server: ndrping s <server IP> <server port> p1 Blocking mode: Client: ndrping c <server IP> <server port> b1 Server: ndrping s <server IP> <server port> b1
3. ndpingpong.exe	Send/recv latency, polling and blocking mode. NETWORK-0236	2	1	Polling mode: Client: ndpingpong c <server IP> <server port> p1 Server: ndpingpong s <server IP> <server port> p1 Blocking mode: Client: ndpingpong c <server IP> <server port> b1 Server: ndpingpong s <server IP> <server port> b1
4. ndrpingpong.exe	RDMA write latency, polling only (receiver must poll on memory). NETWORK-0236	2	1	Polling mode: Client: ndrpingpong c <server IP> <server port> p1 Server: ndrpingpong s <server IP> <server port> p1 Blocking mode: Client: ndrpingpong c <server IP> <server port> b1 Server: ndrpingpong s <server IP> <server port> b1

5. ndconn.exe	Client connection establishment scalability. NETWORK-0236	2	1	Client: ndconn c <server IP> <server port> t1 Server : ndconn s <server IP> <server port> t4
6. ndconn.exe	Server connection establishment scalability. NETWORK-0236	2	1	Client: ndconn c <server IP> <server port> t4 Server : ndconn s <server IP> <server port> t1
7. ndmrrate.exe	Memory registration scalability (pipelined registrations via overlapped. NETWORK-0239	2	1	ndmrrate <local IP>
8. ndmrlat.exe	Memory registration latency (single registration). NETWORK-0239	2	1	ndmrlat <local IP>
9. ndmpic	Simulates MS-MPI's connection establishment logic, establishes multiple connections between two nodes. NETWORK-0235	2	1	Client: ndmpic 0 <remote IP> Server: ndmpic 1 <remote IP>
10. ndmw	Simulates MS-MPI's memory window usage model between two nodes. NETWORK-0235	2	1	Client: ndmw c <server IP> <server port> Server : ndmw s <server IP> <server port>

NOTE: All of the unit tests accept a "l<logfile name>" parameter (that's a lower case 'L', followed by the logfile name) which will redirect STDOUT and STDERR to the log file.

NetworkDirect Logo Tests – MPI Tests

All MPI tests verify the NetworkDirect Logo requirement: NETWORK-0235.

All MPI test command are of the form:

```
Job submit
    /projectname:ND_Logo_Tests
    /jobname:[TEST NAME]
    /workdir:%%CCP_CLUSTERNAME%%\ND_Tests
    /STDOUT: [TEST NAME]\%%CCP_JOBID%%_STDOUT.txt
    /STDERR: [TEST NAME]\%%CCP_JOBID%%_STDERR.txt
    /numnodes:[# NODES]
    mpiexec  -cores [# CORES/NODE]  [TEST NAME]
```

Test Name	Test Description	# Nodes	#Cores/ Node
1. mpi_allreduce.exe	Applies a reduction operation and places the result in all tasks in the group. This is equivalent to an MPI_Reduce followed by an MPI_Bcast.	8	1
2. mpi_alltoall.exe	Each rank performs a scatter operation, sending a message to all the ranks in the group in order by index.	8	1
3. mpi_alltoallv.exe	All-to-all communication with variable length messages and offsets.	8	1
4. mpi_bcast.exe	Sends messages from the one rank to all other ranks.	8	4
5. mpi_intercomm _create1.exe	Combines two communicators into one and ensures message traffic is sorted by communicator.	8	1
6. mpi_isend_perf.exe	Latency tester for non-blocking send around a ring of ranks.	8	1
7. mpi_reduce.exe	Applies a reduction operation on all ranks in the communicator and places the result in one rank. In addition to the normal loops provided by the MPITEST environment (message length, communicator size and type, and data type) this test also loops over the identity of the root process and over the operation to be used in the reduce.	8	1
8. mpi_send _flood.exe	Basic blocking send operation, with one rank flooding another with messages before they are received.	8	1
9. accfence1 .exe	Accumulate and replace then synchronize RMA calls with an MPI RMA Window Fence.	4	1
10. async.exe	Uses asynchronous I/O. Each process writes to separate files and then reads them back.	8	1
11. coll2.exe	Test MPI_Gather	8	1
12. coll13.exe	Tests MPI_AlltoAll followed by an MPI_AllReduce	8	4
13. putpscw1.exe	Test RMA put (MPI_Put) with Post/Start/Complete /Wait	4	1
14. scattern.exe	Sends a vector and receives individual elements using MPI_Scatter	8	4
15. sendMany.exe	Each rank posts a non-blocking receive from the others and then MPI_Send's to each.	8	1
16. sendrecv1.exe	Send-receive various data types between 2 ranks with data checking.	2	1
17. transpose1.exe	Transposes a matrix using put, fence, and derived datatypes using MPI_vector and MPI_hvector to contain the data values.	2	1