OpenFabrics Alliance
Interoperability Working Group (OFA-IWG)
April 2007 Interoperability Event Report

Amit Krig
Mellanox Technologies
2900 Stender Way
Santa Clara, CA 95054

June 27, 2007
Report Rev1.1

Enclosed are the results from OFA Interoperability testing performed on the following devices under test (DUTs):

Mellanox Technologies MHEA28-1TC HCA
Mellanox Technologies MHEA28-XT HCA
Mellanox Technologies MHGA28-1TC HCA
Mellanox Technologies MHGA28-XTC HCA
Mellanox Technologies MHES14-XT HCA
Mellanox Technologies MHES18-XTC HCA
Mellanox Technologies MHGS18-XTC HCA

The test suite referenced in this report is available at the OFA website, at test time release 1.10 (April 10, 2007) was used:

http://www.openfabrics.org/docs/InteropTestPlan.pdf   http://www.iol.unh.edu/services/testing/ofa/testplan.pdf

Test Procedure	IWG Test Status	Result/Notes
IB Link Initialize	**Mandatory**	Passed – no issues seen
IB Fabric Initialization	**Mandatory**	Passed – no issues seen
IB IPoIB	**Mandatory**	Passed – no issues seen
TI iSER	Beta	Refer to Comments
IB SRP	**Mandatory**	Passed – no issues seen
TI SDP	**Mandatory**	Passed with Comments
IB SM Failover	Beta	Partially tested due to time constraints of test period
TI MPI - OSU	Beta	Not tested due to time constraints of test period
TI MPI - Intel	Beta	Not tested due to time constraints of test period
TI uDAPL	Beta	Refer to Comments
FibreChannel Gateway(IB)	Beta	Not available for test
Ethernet Gateway(IB)	Beta	Not available for test

For specific details regarding issues please see the corresponding test result.

Testing Completed 4/27/2007

Bob Noseworthy
mailto:ren@iol.unh.edu

Review Completed 5/22/2007

Mikkel Hagen
mailto:mhagen@iol.unh.edu

Digital Signature Information

This document was created using an Adobe digital signature. A digital signature helps to ensure the authenticity of the document, but only in this digital format. For information on how to verify this document's integrity proceed to the following site:

http://www.iol.unh.edu/certifyDoc/

If the document status still indicates "Validity of author NOT confirmed", then please contact the UNH-IOL to confirm the document's authenticity. To further validate the certificate integrity, Adobe 6.0 should report the following fingerprint information:

MD5 Fingerprint: A303 D24B 3F7D 0E0D 27F2 B8BC 5FA0 1FC6

SHA-1 Fingerprint: 7BD1 A2EE 89DC AB98 2E32 F36A A9E6 E865 A0EE 88EE

Report Revision History

v1.0 Initial Release

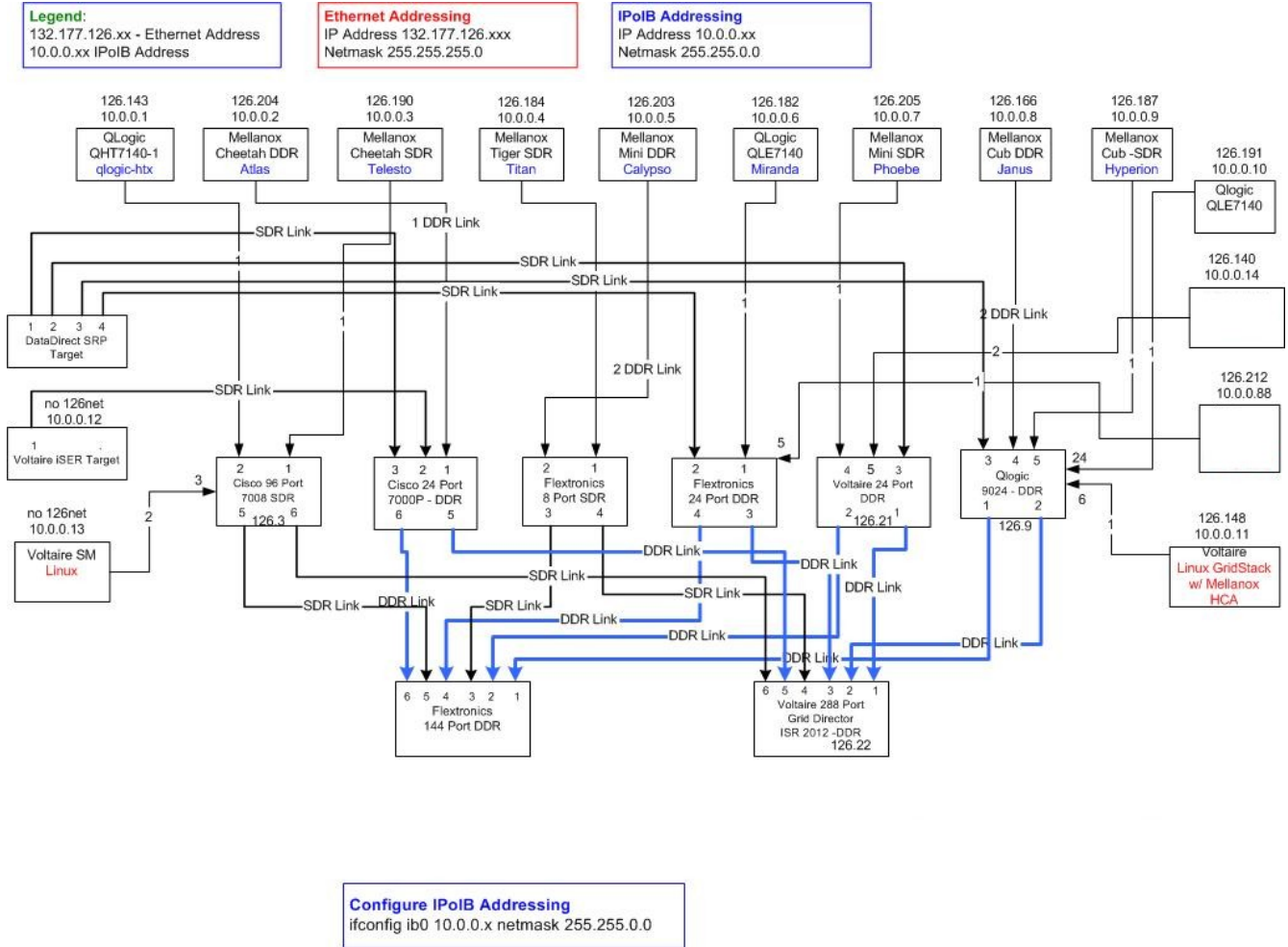v1.1 Changed results to Refer to Comments and Pass with Comments

Table 1: Result Key

The following table contains possible results and their meanings:

Result:	Description:
PASS	The Device Under Test (DUT) was observed to exhibit conformant behavior.
PASS with Comments	The DUT was observed to exhibit conformant behavior however an additional explanation of the situation is included, such as due to time limitations only a portion of the testing was performed.
FAIL	The DUT was observed to exhibit non-conformant behavior.
Warning	The DUT was observed to exhibit behavior that is not recommended.
Informative	Results are for informative purposes only and are not judged on a pass or fail basis.
Refer to Comments	From the observations, a valid pass or fail could not be determined. An additional explanation of the situation is included.
Not Applicable	The DUT does not support the technology required to perform this test.
Not Available	Due to testing station limitations or time limitations, the tests could not be performed.
Borderline	The observed values of the specified parameters are valid at one extreme and invalid at the other.
Not Tested	Not tested due to the time constraints of the test period.

Table 2: DUT and Test Setup Information

Figure 1: The IB fabric configuration utilized for any tests requiring a multi-switch configuration is shown below.

DUT #1 Details			
Manufacturer	Mellanox Technologies	Firmware Rev	4.8.2
Model	MHEA28-1TC	Hardware Rev	A1
Speed	SDR	Driver Version	OFED 1.2 RC1
		IP Address in Fabric	132.177.126.187 / 10.0.0.9
Additional Comments/Notes			
AKA "Lion Cub HCA" - Dual-port, 10Gb/s w/ media adapter support, PCIe x8, 128MB memory, tall bracket, RoHS R5			

DUT #2 Details			
Manufacturer	Mellanox Technologies	Firmware Rev	5.2.0
Model	MHEA28-XT	Hardware Rev	A2
Speed	SDR	Driver Version	OFED 1.2 RC1
		IP Address in Fabric	132.177.126.205 / 10.0.0.7
Additional Comments/Notes			
AKA "Lion Mini HCA" - Dual-port, 10Gb/s w/ media adapter support, PCIe x8, mem-free, tall bracket			

DUT #3 Details			
Manufacturer	Mellanox Technologies	Firmware Rev	4.8.2
Model	MHGA28-1TC	Hardware Rev	A2
Speed	DDR	Driver Version	OFED 1.2 RC1
		IP Address in Fabric	132.177.126.166 / 10.0.0.8
Additional Comments/Notes			
AKA "Lion Cub DDR" - Dual-port, 20Gb/s w/ media adapter support, PCIe x8, 128MB memory, tall bracket, RoHS R5			

DUT #4 Details			
Manufacturer	Mellanox Technologies	Firmware Rev	5.2.0
Model	MHGA28-XTC	Hardware Rev	A2
Speed	DDR	Driver Version	OFED 1.2 RC1
		IP Address in Fabric	132.177.126.203 / 10.0.0.5
Additional Comments/Notes			
AKA "Lion Mini DDR" - Dual-port, 20Gb/s w/ media adapter support, PCIe x8, mem-free, tall bracket, RoHS R5			

DUT #5 Details			
Manufacturer	Mellanox Technologies	Firmware Rev	1.2.0
Model	MHES14-XT	Hardware Rev	A4
Speed	SDR	Driver Version	OFED 1.2 RC1
		IP Address in Fabric	132.177.126.184 / 10.0.0.4
Additional Comments/Notes			
AKA "Tiger HCA" - Single-port, 10Gb/s w/ media adapter support, PCIe x4, mem-free, tall bracket			

DUT #6 Details			
Manufacturer	Mellanox Technologies	Firmware Rev	1.2.0
Model	MHES18-XTC	Hardware Rev	A2
Speed	SDR	Driver Version	OFED 1.2 RC1
		IP Address in Fabric	132.177.126.190 / 10.0.0.3
Additional Comments/Notes			
AKA "Cheetah HCA" - Single-port, 10Gb/s w/ media adapter support, PCIe x8, mem-free, tall bracket, RoHS R5			

DUT #7 Details			
Manufacturer	Mellanox Technologies	Firmware Rev	1.2.0
Model	MHGS18-XTC	Hardware Rev	A3
Speed	DDR	Driver Version	OFED 1.2 RC1
		IP Address in Fabric	132.177.126.204 / 10.0.0.2
Additional Comments/Notes			
AKA "Cheetah DDR HCA" - Single-port, 20Gb/s w/ media adapter support, PCIe x8, mem-free, tall bracket, RoHS R5			

Mandatory Tests - IB Device Test Summary Results:

The following tables detail results for tests identified by the OFA-IWG as mandatory tests for the OFA Interoperability Logo Program (OFILP) per the OFA-IWG Interoperability Test Plan Release 1.10 (April 10, 2007)

Test Number and Name	Part(s)	Summary Note(s)	Result(s)
Group 1: IB Link Initialize	Test #1: Phy link up all ports	Link partners link as expected	PASS
	Test #2: Logical link up all ports switch SM	ibstatus reports active links	PASS
	Test #3: Logical link up all ports HCA SM	ibstatus reports active links	PASS
Discussion: Test #1: Phy link up all ports			
DDR cables were used for all link tests. Device LEDs validated visually. Link width and link speed validated via use of "ibdiagnet -lw 4x" and ibdiagnet -ls 10"			
Discussion: Test #2 & 3: Logical link up all ports switch SM / HCA SM			
The switch-under-test's SM, OFED's OpenSM, and the Voltaire SMHost (GVFM) SM were all used to validate that the link could be brought to the Active state as verified via the "ibstatus" command. All SMs were initially off as the switch was powered up. After checking for an initial physical link, the SM under test was started and the state of the link verified. Note, for HCA to HCA link checks, the nature of the direct cable connection prevented the validation of any SM but the OpenSM. Refer to the table below for specific link configurations tested.			

For Mellanox Technologies HCAs	Switch SM	OpenSM	Voltaire SMHost (GVFM)
Cisco SFS 7000D 24-port DDR (Switch)	PASS	PASS	PASS
Cisco SFS 7008 96-port SDR (Switch)	PASS	PASS	PASS
Data Direct Networks S2A 9550 (SRP Target)	Not Applicable	PASS	Not Applicable
Flextronics 8 Modular SDR (Switch)	PASS	PASS	PASS
Flextronics 24 Modular DDR (Switch)	PASS	PASS	PASS
Flextronics 144 Modular DDR (Switch)	PASS	PASS	PASS
Qlogic SilverStorm 24-port DDR (Switch)	PASS	PASS	PASS
Qlogic QHT-7140 (HCA)	Not Applicable	PASS	Not Applicable
Qlogic QLE-7140 (HCA)	Not Applicable	PASS	Not Applicable
Voltaire ISR 9024D-M 24-port DDR (Switch)	PASS	PASS	PASS
Voltaire ISR 2012 288-port DDR (Switch)	PASS	PASS	PASS
Voltaire Gridstack (HCA)	Not Applicable	PASS	Not Applicable
Voltaire IPStor (iSER Target)	Not Applicable	PASS	Not Applicable

Test Number and Name	Part(s)	Summary Note(s)	Result(s)
Group 2: IB Fabric Initialization	Test #1: Verify all SMs configure fabric	Port is Active with all SMs	**PASS**

Discussion: Test #1: Verify all SMs configure fabric

The fabric configuration shown in Figure 1 was used for this test. 'ibdiagnet -c 1000' showed no Port errors counters increment. Only one SM is run at a time. All switches are power cycled between SM trials. All links are validated via use of 'ibdiagnet' and on each host 'ibstatus' to validate speed, width and link state. Refer to the table below for SM details.

For Mellanox Technologies HCAs	All ports Armed/Active	No Dup GUIDs	No Port errors
Cisco SFS 7000D 24-port DDR (Switch)	**PASS**	**PASS**	**PASS**
Cisco SFS 7008 96-port SDR (Switch)	**PASS**	**PASS**	**PASS**
OFED 1.2 RC1 OpenSM (SM)	**PASS**	**PASS**	**PASS**
Qlogic SilverStorm 24-port DDR (Switch)	**PASS**	**PASS**	**PASS**
Voltaire ISR 9024D-M 24-port DDR (Switch)	**PASS**	**PASS**	**PASS**
Voltaire ISR 2012 288-port DDR (Switch)	**PASS**	**PASS**	**PASS**
Voltaire SMHost (GVFM) (SM)	**PASS**	**PASS**	**PASS**

Test Number and Name	Part(s)	Summary Note(s)	Result(s)
Group 3: IPoIB	Test #1: Ping Test all to all	HCAs can ping all to all with all byte sizes	**PASS**
	Test #2: Connect disconnect host	Connectivity functions after topology change	**PASS**
	Test #3: File Transfer Procedure	4MB file SFTP transfers were validated	**PASS**

Discussion: Test #1: Ping all to all

IP connectivity to all HCAs using each class of SMs was validated. SMs include Cisco SM, OpenSM, Qlogic SM, Voltaire SM (9024), and Voltaire SMHost (GVFM) SM.

Note that to achieve ICMP echo requests (pings) with a packet size of 64, 256, 511, 512, 1024, 1025, 2044, 4096, 8192, 16384, 32768, 65507, ping -s # was utilized where # corresponds respectively in the following list: 36, 228, 483, 484, 996, 997, 2016, 4068, 8164, 16356, 32740, 65479

Discussion: Test #2: Disconnect and Reconnect HCA

Each class of SMs was tested (Cisco SM, OpenSM, Qlogic SM, Voltaire SM (9024), Voltaire SMHost (GVFM) SM).

Refer to the default fabric configuration. All HCAs were connected except the MiniDDR HCA (10.0.0.5).

All SMs were disabled, all switches were then power cycled.

The SM under test was enabled, and all HCAs were confirmed to be reachable except for the station at 10.0.0.5.

The QLE7140 HCA (10.0.0.6) was removed from the fabric. All HCAs were confirmed to be reachable except 5 and 6.

The MiniDDR HCA (5) was then connected to the fabric. All HCAs are now reachable except 6.

The Qlogic QLE7140 HCA is now connected to the Flextronics 8 port SDR switch (at port 7) rather than the Flextronics 24 port switch (at port 1). All HCAs are now reachable.

Using host system 10.0.0.5, "ifconfig ib1 down" results in a loss of IP connectivity, "ifconfig ib1 up" results in a restoration of IP connectivity

Discussion:. Test #3: File transfer procedure

For expediency, SFTP was utilized as FTP servers were not configured at test time.

File transfer was validated from the Tiger SDR host (10.0.0.4) to each other station by sending a 4MB file to the remote station and then copying it back and comparing the file.

Tests were repeated 4 times. All file transfers completed successfully and the file was identical in all cases.

Test Number and Name	Part(s)	Summary Note(s)	Result(s)
Group 5: IB SRP	Test #1: Basic dd application	dd read completes successfully	PASS
	Test #2: Kill IB SM	dd read completes successfully	PASS
	Test #3: Disconnect Host	dd works after link restored	PASS
	Test #4: Disconnect Target	works after link restored in 5sec	PASS

Discussion: Test #1: Basic dd application

Note, the test plan v1.10 was modified from requiring only "2 HCAs" to include representative hosts from all vendors – the specific devices used for validation of the DataDirect Network's SRP target included the Mellanox MHGS18-XTC (Cheetah DDR), MHES14-XT (Tiger SDR), and the MHGA28-1TC (Cub DDR). For this test, 6 hosts logged into the DataDirect Network's SRP targets and performed simultaneous reads via the command 'dd if=/dev/sdb of=/dev/null count=600 bs=10M'. On the hosts, the ib_srp module is loaded and 'ibsrpdm -c' and 'lsscsi' is used to find the correct info to add the targets. An example of the add target command used on each host follows: 'echo id_ext=23000001ff0401d4,ioc_guid=23000001ff0401d4,dgid=fe80000000000000023000001ff0401d4,pkey=ffff,service_id=d40104ff01000023 > /sys/class/infiniband_srp/srp-mthca0-1/add_target'

Discussion: Test #2: Kill IB SM

Note, during testing the SM should not be stopped until login and data transfer on all connections is confirmed, thus ensuring all devices have properly logged in and commenced transfers before terminating the current SM.

In all cases for all SMs, each host was observed to complete the current in-progress transfer when the SM is killed.

SMs tested included: Cisco SFS 7000D 24-port DDR (Switch), OFED 1.2 RC1 OpenSM, Qlogic SilverStorm 24-port DDR (Switch), Voltaire SMHost (GVFM) SM , Voltaire ISR 9024D-M 24-port DDR (Switch)

Discussion: Test #3: Disconnect Host

Note, per test plan v1.10, one host was used for this "Step B" of IB SRP testing. The host (Voltaire Gridstack HCA) was logged into lun2 of the DataDirect Networks SRP target and could perform dd operation without issue. The link to the host was then disconnected and then reconnected. The same dd operation could still be performed without additional user intervention/commands.

Discussion: Test #4: Disconnect Target

After consultation, it was determined that a target should drop login information if a link is removed for an extended period of time, however it was deemed reasonable for a target to maintain login information if the link is removed for less than 30seconds. For this test, the host used (Mellanox MHGA28-1TC) was used to login to lun2 and dd was performed. The link to interface 2 (lun2) of the DataDirect Networks SRP target was then removed and restored 5 seconds later. Within 10seconds the logical link was up for the link and the DDN SRP target was observed to still maintain the login information, which was confirmed when the host could again perform the same dd operation without additional user intervention/commands.

Test Number and Name	Part(s)	Summary Note(s)	Result(s)
Group 6: TI SDP	Test #1: netperf procedure	All tests completed successfully	**PASS**
	Test #2: FTP procedure	4MB file FTP transfers were validated	**PASS**
	Test #3: IB SCP Procedure	SCP file transfer validated	**PASS**
Discussion: Test #1: netperf procedure			
Note, test plan v1.10 calls for export SIMPLE_LIBSDP=1, however this was known not to work in OFED1.2 RC1. Module ib_sdp is loaded. Each host machine's netperf server is configured with 'export LD_PRELOAD=/usr/lib64/libsdp.so; export LIBSDP_CONFIG_FILE=/etc/libsdp.conf; netserver -p 13009'. Note the port number (13009) is varied per machine, 13000 + the last IP address octet. Netperf is then run to and from all hosts in the fabric. An example netperf command follows: 'netperf -p $13003 -H 10.0.0.3 -l 1 -t TCP_STREAM -- -m 10 -s 1024' Note message sizes 10, 100, 1000, 10000 and buffer sizes of 1024 and 6000 were validated on all connections. All netperf operations were observed to complete successfully.			
Discussion: Test #2: FTP procedure			
FTP servers were configured as specified in the test plan and a 4MB file was transferred to and from each host and a binary comparison performed to validate the transfer. 'lsmod	grep sdp' was used to validate that ib_sdp was loaded with >0 dependencies.			
Discussion:. Test #3: IB SCP Procedure			
scp was used successfully between all hosts to transfer files while ib_sdp was loaded.			

Beta Tests - IB Device Test Results:

The following table details results for tests identified by the OFA-IWG as beta tests for the OFA Interoperability Logo Program (OFILP) per the OFA-IWG Interoperability Test Plan Release 1.10 (April 10, 2007)

Test Number and Name	Part(s)	Summary Note(s)	Result(s)
Group 4: TI iSER	Test #1-5	Multiple issues observed	**Refer to Comments**
Discussion: Test Results			

All iSER testing occurred between the Voltaire iSER target and 9 HCA systems within the topology outlined in Figure 1. Though successful iSER dd exchanges were observed, of 9 systems tested, 1 system failed to list the iSER target device in the results of sg_map,   2 systems locked up during testing, and 1 system was observed to crash. The cause of these incidents is unknown, but was investigated throughout a day of testing at the interoperability event by attending vendor staff. Note, this testing was driven, with thanks, by attending vendor staff and not overseen by UNH-IOL.   Retesting occurred the following week in an effort to provide additional details of the failure conditions. In the time available, two systems were re-investigated, one that had locked up, and one that had worked properly. In this retesting, inconsistent results from previous testing was observed, suggesting one of several issues: A) Test process error: The test procedure outlined in test plan v1.10 lacked sufficient detail to govern the test process, through the efforts of attending vendor staff, a more detailed process was developed during the week of testing; however, it is possible that some of this process was implemented in error. B) OFED 1.2 RC1 iSER support issues. C) Vendor device iSER issues (either target or initiator). There is insufficient observational evidence to support either of the last two items at this time, further testing and improvements to the test process are required to this Beta test process.

Two sets of system output are included in the Appendix at the end of this report. The first is captured from a system that was previously observed to lockup during iSER testing. The second from a system that was not observed to experience any issue.

Test Number and Name	Part(s)	Summary Note(s)	Result(s)
Group 5: IB SM Failover/Handover	Test #1-4	No issues observed	**Refer to Comments**
Discussion: Test Results			

Testing time prevented complete execution of all switch permutations. Modifications to the v1.10 test plan were developed during the testing event to further clarify the process, and additional clarifications are sought to better test true handover from on SM to another from the same vendor and allow for failover testing for SMs from different vendors. Due to these pending modifications, and the incomplete execution of the current test plan procedure, further testing is required for this Beta test process. Note – of the same-vendor SMs tested, no issue was observed for:
- standby SMs becoming master SMs in the event that the current master SM is removed, or
- the standby SM's priority is adjusted to a priority higher than the current master SM, or
- the standby SM's priority is adjusted to the same priority of the current master SM but the GUID of the standby SM is lower than the current master.

Test Number and Name	Part(s)	Summary Note(s)	Result(s)
Group 8: TI MPI – Ohio State University	Test #1-14	Not tested due to lack of time	**Not Available**
Group 9: MPI – Intel MPI	Test #1-21	Not tested due to lack of time	**Not Available**
Discussion: Test Results			

Testing time prevented execution of these tests. Further automation and early execution of IB link and fabric initialization tests should allow for greater test time within the testing week of future events.

Test Number and Name	Part(s)	Summary Note(s)	Result(s)
Group 10: TI uDAPLTEST Commands	Test #1-10	Issues observed	**Refer to Comments**

Discussion: Test Results

Time was taken to test and during the week of April 23-27 retest this portion of the test plan to investigate issues seen during testing. Initial tests during the week of April 16 were run such that all HCAs were running the dapltest server and the dapltest clients simultaneously, this was later modified to perform only on concurrent server-client test pair at a time. Despite repeated efforts to vary the timing of the scripts and the order in which operations were, inconsistent results were observed. At least three failure modes could be seen:

1)
Most commonly seen, the client would fail to connect to the server, eventually timeout, and the following error would be displayed:
Warning: conn_event_wait DAT_CONNECTION_EVENT_UNREACHABLE
or
Warning: conn_event_wait DAT_CONNECTION_EVENT_NON_PEER_REJECTED

2)
On at least one occasion, the client was seen to hang indefinitely, presumably in a similar state to that observed above, but in this one case, no timeout occurred within >15min and the client was aborted.

3)
Finally, the server was seen to seg fault on rare occasion – initially this is believed to only occur if more than one client was attempting to connect to the same server, but requires further investigation. This was directly observed during the first week of testing when manually running the tests in a many-to-many scenario - during pair-wise testing the following week, the server output was not observed/captured, so it is conceivable that some of the DAT_CONNECTION_EVENT messages were due to a crashed server (but in at least one case, a DAT_CONNECTION_EVENT_UNREACHABLE was returned for test plan test 2.1, 2.2 and 2.3, but 2.4 completed successfully - indicating that the server was still running on the target machine)

These events did not occur often - however the script performing the pair-wise Group2 tests between all HCAs could not be run successfully 3-times in a row without at least one DAT_CONNECTION_EVENT_UNREACHABLE. At times, one complete run through the script could be completed without error.
In at least one case, there was higher occurrence of the connection unreachable events with one particular machine - following a reboot of this machine the occurrence of the DAT_CONNECTION_EVENT_UNREACHABLE from the host seemed to cease. As the machines were not frequently rebooted, it is possible, however unlikely, that there is some interaction between previous test efforts (iser, sdp, srp, etc) leaving the machines in an 'unclean' state - or it is just as possible that repeated dapltest executions is a factor -- but additional testing, and thus time, would be required to further understand this.

Due to these pending issues, further testing is required for this Beta test process.

Test Number and Name	Part(s)	Summary Note(s)	Result(s)
Group 12: Fibre Channel Gateway (IB)	Test #1-10	No supporting devices available	**Not Available**
Group 13: Ethernet Gateway (IB)	Test #1-7	No supporting devices available	**Not Available**
Discussion: Test Results			

These beta tests could not be performed as no such gateway devices were available during the testing period.

Appendix A: Example Testing Scripts

Examples of some testing scripts employed during the testing period are embedded below for reference.

Appendix B: iSER Retest Output

Two sets of system output are embedded below for reference.

The first is captured from a system that was previously observed to lockup during iSER testing, upon retest this system was able to perform dd operations with user intervention (restarting iscsi).

The second capture is from a system that was not observed to experience any issue during initial testing, but was not observed to complete dd operations during retesting.

Retesting occurred several days after initial testing, however system changes should include nothing greater than a reboot.