VMware multipathing failure due to timeout setting discrepancy with QLogic Host Bus Adapter (HBA)

Problem: VMware multipathing fails to kick in and link failures do not switch over paths. This results in disk I/O failures in place of switching over to functional link.

Description: This behavior happens because the default timeout setting for QLogic adapters called Keep Alive Timeout (KATO) is too high for statically discovered targets (Boot Disk for example). The multipath components should detect link failures quicker than its higher level (Disk level) components. By default for Linux 2.4 used by ESX, the disk timeout is 60 seconds and the underlying QLogic timeout turns out to be 65 seconds for the Boot disk with the default setting, So the Disk layer times out before the path fail over could happen. To fix this, you need to use QLogic configuration utility (iscli) and setup the default KATO settings.

Fix: This has to do with the QLogic HBA default settings for target Keep Alive Timeout (KATO).

Procedure for updating KATO Value in QLogic HBA

Make sure that you download and install **iscli-1.2.00-05_linux_i386.install.tar.gz** from the QLogic Website (www.qlogic.com).

Procedure:

Check the existing KATO values for the targets on Port 0 and Port 1

Display KATO value for Target 0 (Boot volume) on HBA Port 0:
 # iscli -t 0 0 | grep KeepAlive
TGT_KeepAliveTimeout: 30

Display KATO value for Target 0 (Boot volume) on HBA Port 1:
 # iscli -t 1 0 | grep KeepAlive
TGT_KeepAliveTimeout: 30

The above shows a boot volume with a KATO value of 30 which will lead to the problem. Now, you can set this value from the command line as well:

Change KATO value for Target 0 (Boot volume) on HBA Port 0 to the ESX Default (14):
 # iscli -tc 0 0 TGT_KeepAliveTimeout 14

Change KATO value for Target 0 (Boot volume) on HBA Port 1 to the ESX Default (14):
 # iscli -tc 1 0 TGT_KeepAliveTimeout 14

Set the port level KATO value for the BIOS which would apply the KATO value to any new statically bound target (keep in mind that this will not change the value for targets which are already bound, only NEW targets):

Change Port level KATO value for HBA Port 0:
 # iscli -n 0 KeepAliveTO 14

Change Port level KATO value for HBA Port 1:
 # iscli -n 1 KeepAliveTO 14

This action requires a reboot of the ESX server; at this point, reboot the ESX server to proceed further.

On reboot go to the QLogic BIOS and unbind and rebind the static targets (eg., ESX boot volume). This is required, otherwise the changed values will not come into effect.

On reboot of the ESX server edit the Task Management Timeout Value also as follows if it is not set to 10. This is needed as sometimes QLogic cards get shipped with a random value.

This value is also set at the port level, and at the target level, so you'll need to use the following commands to change it:

 # iscli -n 0 Task_Management_Timeout 10
 # iscli -n 1 Task_Management_Timeout 10
 # iscli -tc 0 0 TGT_TaskManagementTimeout 10
 # iscli -tc 0 0 TGT_TaskManagementTimeout 10

After setting these values you'll need to reboot and verify that the settings took with command similar to the following:

 # iscli -c 0 | egrep 'Keep|Task'
KeepAliveTO: 14
Task_Management_Timeout: 10
 # iscli -c 1 | egrep 'Keep|Task'
KeepAliveTO: 14
Task_Management_Timeout: 10
 # iscli -t 0 0 | egrep 'Keep|Task'
TGT_KeepAliveTimeout: 14
TGT_TaskManagementTimeout: 10
 # iscli -t 1 0 | egrep 'Keep|Task'
TGT_KeepAliveTimeout: 14
TGT_TaskManagementTimeout: 10

Again reboot the ESX server and then confirm using the get commands mentioned above is the value has been set for all the values that have been configured. If not reboot the ESX server and once again go to the QLogic BIOS and unbind and rebind the statically assigned targets and again reboot and check the same.

Problem: Restoration of the preferred link does not always restore the target access through this path.

Description: Registering a newly established path does not always complete even though the initiator-target session was re-established. In such situations StorTrends iTX stack has to force error recovery by issuing a session log-out as a part of recovery from such scenarios. The initiator would recognize this session killing and re-establish the session and complete the registering of the newly available path.

Fix: The iSCSI target module parameter for enabling this link recovery has to be set using the following commands:

tgtacl -t enable linkrecovery

Cli set –t tgtoptimizations –n -g 3 iscsi

Example:

Cli set –t tgtoptimizations –n c0v0 –g 3 iscsi
Cli set –t tgtoptimizations –n c0v0 –g 3 iscsi

Release Notes: This feature is needed for supporting SAN Boot and multi-pathing for VMware ESX servers. This requires a trivial out-of-band extension of the StorTrends iTX iSCSI target driver carried in StorTrends iTX Release 2.7 build 1030 (minor revisions 2.3 and below). This out –of-band extension to the iSCSI target driver is natively incorporated in minor revisions 2.4 and later.

Problem: The StorTrends iTX stack by default does not support multiple set of initiators to have exclusive target authorization settings.

Description: StorTrends iTX in its standard offering supports either a specific initiator to have Login access to a target or else any initiator could login to the target. Essentially it was one or all. Now we have implemented through CLI the feature where we allow a set of specified initiators to have Login access.

Fix: Use target authorization CLI commands to enable exclusive target access for a set of Initiators.

The method to update the initiator target settings is as follows.

tgtacl -t set -n -i -i < initiator_name>
tgtacl -t list -n
tgtacl -t delete -n

Example:

tgtacl -t set -n c0v0 -i iqn.1991-05.com.microsoft:VMware1 -i iqn.1991-05.com.microsoft:VMware2
tgtacl -t list -n c0v0
tgtacl -t delete -n c0v0

Note: This command allows a maximum set of four initiators for target access authorization.

Release Notes: This feature is needed for supporting SAN Boot and multi-pathing for VMware ESX servers. This requires a trivial out-of-band extension of the StorTrends iTX iSCSI target driver carried in StorTrends iTX Release 2.7 build 1030 (minor revisions 2.3 and below). This out –of-band extension to the iSCSI target driver is natively incorporated in minor revisions 2.4 and later.

Problem: Connectivity issue on different subnets dropping iSCSI sessions.

Description: StorTrends iTX in its standard offering supports responding to ARP requests on a single interface, even if the request source is through a different interface (standard Linux behavior). Also, during ARP announce, the standard behavior is to use any local address, irrespective of source IP address that came with the request. We needed to modify this setting so that we respond to ARP requests in the same interface in which the request was received, with the incoming IP address and our IP address properly filled in.

Fix: Update the /etc/sysctl.conf file with the following settings and reboot the StorTrends iTX.

net.ipv4.conf.default.arp_ignore = 2
net.ipv4.conf.default.arp_announce = 1
net.ipv4.conf.default.arp_filter = 1
net.ipv4.conf.all.arp_ignore = 2

net.ipv4.conf.all.arp_announce = 1
net.ipv4.conf.all.arp_filter = 1

Release Notes: This feature is needed for supporting SAN Boot and multi-pathing for VMware ESX servers. This requires only a configuration change and can be done when the stack will be used along VMware ESX servers.