

## C o m m u n i t i e s

All Communities ▾ [Blogs ▾](#)

Communities &gt; Blogs

## VMware Storage Blog

[VMware vStorage reduces storage challenges](#) | [Main](#) |

October 19, 2009

## vStorage Multi Paths Options in vSphere

## Multi Path challenges

In most SAN deployments it is considered best practice to have redundant connections configured between the storage and the server. This often includes redundant host based adaptors, Fibre Channel switches, and controller ports to the storage array. This results in having four or more separate paths connecting the server to the same storage target. These multiple paths offers protection against a single point of failure and permit load balancing. In a VMware virtualization environment a single datastore can have active IO on only a single path to the underling storage target (LUN or NFS mountpoint) at a given time. The storage stack in an ESX server can be configured to use any of the available paths but only one at a time prior to vSphere. In the event of the active path failing, the ESX server will detect the failure and failover to an alternate path.

The two challenges that VMware Native Multi Path addresses has been the 1) presentation of a single path when several are available (aggregation) and 2) handling of failover and fallback when the active path is compromised (Failover). In addition the ability to alternate which path is active was supported through the introduction of round robin path selection in ESX release 3.5. Through the command line one can change the default parameters to have the path switched after a certain number of blocks or a set number of transactions going across an active path. These values are not currently available in the vCenter interface and can only be adjusted via the command line. Another issue which the ESX server storage stack has to address was how to treat a given storage device as some offered and active/active (A/A) controllers and others had active/passive controllers. A given LUN can be addressed by more than one storage controller at the same time in an active/active array. Where as the active/passive (A/P) array may access a LUN from one controller at a time and requires the transfer control to the other controller when access thru that controller is needed. To deal with this difference in array capabilities, VMware had to place logic in the code for each type of storage array supported. That logic defined what capabilities the array had with regards to active/active or active/passive as well as other attributes.

Path selection and fallback options included: Fixed, Most Recently Used and Round Robin. The Fixed path policy is recommended for A/A arrays. It allows the use to configure static load balancing of paths across ESX hosts. The MRU policy is recommended for A/P arrays. It prevents the possibility of path thrashing due to a partitioned SAN environment. Round robin may be configured for either A/A or A/P arrays, but care must be taken to insure that A/P arrays are not configured to automatically switch controllers. If A/P arrays are configured to automatically switch controllers (sometimes called pseudo A/A), performance may be degraded due to path ping ponging.

## vStorage API for MultiPathing

Pluggable storage architecture (PSA) was introduced in vSphere 4.0 to improve code modularisation, code encapsulation, and to enable 3<sup>rd</sup> party storage vendors multipath software to be leveraged in the ESX server. It provides support for the storage devices on the existing ESX storage HCL through the use of Storage Array Type Plugins (SATP) for the new Native Multipathing Plugins (NMP) module. It also introduces an API for 3<sup>rd</sup> party storage vendors to build and certify multipath modules to be plugged into the VMware storage stack.

In addition to the VMware NMP, 3<sup>rd</sup> party storage vendors can create there own pluggable storage modules to either replace the SATP that VMware offered. Or they could write a Path Selection Module (PSP) or MultiPath Plug (MPP) in that could leverage their own storage array specific intelligence for increased performance and availability.

As with earlier versions of ESX server, the VI admin can set certain path selection preferences via the vCenter path selection preferences. The path selection options of Fixed, MRU and Round Robin still exist as options in vSphere. A default value, set by the SATP that matched the storage array could be changed even if not advised as best practice. But with the new 3<sup>rd</sup> party SATP the vendor could set what defaults were optimized for their array. Through this interface, ALUA became an option that could be supported.

The VMware Native MultiPath module has sub-plugins for failover and load balancing.

- **Storage Array Type Plug-in (SATP)** to handle failover and the
- **Path Selection Plug-in (PSP)** to handle load-balancing.

- NMP "associates" a SATP with the set of paths from a given type of array.
- NMP "associates" a PSP with a logical device.
- NMP specifies a default PSP for every logical device based on the SATP associated with the physical paths for that device.
- NMP allows the default PSP for a device to be overridden.

## VMware Blogs

- [The Console \(VMware Mgmt\)](#)
- [VMTN Blog \(Technical News\)](#)
- [Developer Center Blog](#)
- [ESXi Chronicles](#)
- [Team Fusion](#)
- [Uptime \(Business Continuity\)](#)
- [VI Team Blog](#)
- [Virtualization for SAP Solutions](#)
- [VIXAPI Blog](#)
- [VMware Code Central Blog](#)
- [VMware Communities Blog](#)
- [VMware Knowledge Base Blog](#)
- [VMware Knowledge Base Weekly Digest](#)
- [VMware Networking Blog](#)
- [VMware PhD \(Education\)](#)
- [VMware Security Blog](#)
- [VMware Storage Blog](#)
- [VMware ThinApp Blog](#)
- [VMware vApp Developer Blog](#)
- [VMware: Virtual Reality](#)
- [VMware View Blog](#)
- [VMware View-Point](#)
- [VMware Virtual Disk Development Kit Blog](#)
- [VMware vSphere Blog](#)
- [VMworld Team](#)
- [VROOM \(Performance\)](#)
- [vSphere PowerCLI Blog](#)
- [Workstation Zealot](#)

## Recent Comments

- [Backup on Introduction to vStorage APIs for](#)
- [Giri on Introduction to vStorage APIs for](#)
- [Storage on Introduction to vStorage APIs for](#)
- [Ito 5 tape on VMFS vs. NFS for VMware Infrastructure?](#)
- [seslicht on Introduction to vStorage APIs for](#)

## Categories

[Weblogs](#)

Rules can be associated with NMP as well as 3<sup>rd</sup> party provided MPP modules. Those rules stored in the /etc/vmware/esx.conf file.

Rules govern the operation of both the NMP and the MPP modules.

In ESX4.0 rules are configurable only thru the CLI and not through vCenter.

To see the list of rules defined use the following command:

```
# esxcli corestorage claimrule list
```

To see the list of rules associated with a certain SATP use:

```
# esxcli nmp satp listrules -s <specific SATP>
```

The primary functions of an SATP are to:

- Implements the **switching of physical paths** to the array when a path has failed.
- Determines when a **hardware component** of a physical path **has failed**.
- **Monitors the hardware state** of the physical paths to the storage array.

VMware provides a **default SATP** for each supported array as well as a **generic SATP** (an active/active version and an active/passive version) for non-specified storage arrays.

To see the complete list of defined **Storage Array Type Plugins (SATP)** for the **VMware Native Multipath Plugin (NMP)**, use the following commands:

```
# esxcli nmp satp list
```

Name	Default PSP	Description
VMW_SATP_ALUA_CX	VMW_PSP_FIXED	Supports EMC CX that use the ALUA protocol
VMW_SATP_SVC	VMW_PSP_FIXED	Supports IBM SVC
VMW_SATP_MSA	VMW_PSP_MRU	Supports HP MSA
VMW_SATP_EQL	VMW_PSP_FIXED	Supports EqualLogic arrays
VMW_SATP_INV	VMW_PSP_FIXED	Supports EMC Invista
VMW_SATP_SMM	VMW_PSP_FIXED	Supports EMC Symmetrix
VMW_SATP_LSI	VMW_PSP_MRU	Supports LSI and other arrays compatible with the SIS 6.10 in non-AVT mode
VMW_SATP_EVA	VMW_PSP_FIXED	Supports HP EVA
VMW_SATP_DEFAULT_AP	VMW_PSP_MRU	Supports non-specific active/passive arrays
VMW_SATP_CX	VMW_PSP_MRU	Supports EMC CX that do not use the ALUA protocol
VMW_SATP_ALUA	VMW_PSP_MRU	Supports non-specific arrays that use the ALUA protocol
VMW_SATP_DEFAULT_AA	VMW_PSP_FIXED	Supports non-specific active/active arrays
VMW_SATP_LOCAL	VMW_PSP_FIXED	Supports direct attached devices

To take advantage of certain storage specific characteristics of an array, one can install a 3<sup>rd</sup> party SATP provided by the vendor of the storage array, or by a software company specializing in optimizing the use of your storage array.

**Path Selection Plug-in (PSP).**

The primary function of the PSP module is to determine which physical path is to be used for an I/O request to a storage device. The PSP is a sub-plug-in to the NMP and handles load balancing. VMware offers three PSPs:

To see a complete list of Path Selection Plugins (PSP) use the following command:

```
# esxcli nmp psp list
```

Name	Description
VMW_PSP_MRU	Most Recently Used Path Selection
VMW_PSP_RR	Round Robin Path Selection
VMW_PSP_FIXED	Fixed Path Selection

**Fixed** — Uses the designated **preferred path**, if it has been configured. Otherwise, it uses the first working path discovered at system boot time. If the ESX host cannot use the preferred path, it selects a random alternative available path. The ESX host automatically reverts back to the preferred path as soon as the path becomes available.

**Most Recently Used (MRU)** — Uses the first working path discovered at system boot time. If this path becomes unavailable, the ESX host switches to an alternative path and continues to use the new path while it is available.

**Round Robin (RR)** — Uses an automatic path selection rotating through all available paths and enabling load balancing across the paths. It only uses active paths and is of most use on Active/Active arrays. However in Active/Passive arrays, it will load balance between on ports to the same Storage Processor. However, Round Robin policy is not supported in MSCS environments.

3<sup>rd</sup> party vendors can provide their own PSP to take advantage of more complex I/O load balancing algorithms that their storage array might offer. However, 3<sup>rd</sup> party storage vendors are not required to create plug-ins. All the storage array specific code supported in ESX release 3.x has been ported to ESX release 4 as a SATP plug-in.

**MultiPath Plugins (MPP) or Multipath Enhancement Module (MEM)**

Multipath Plugin (also referred to as Multipath Enhancement Modules (MEM)) enable storage partners to leverage intelligence within their array to present one connection that is backed by an aggregation of several connections. Using the intelligence in the array with their own vendor module in the ESX server, these modules can increase the performance with load balancing and increase availability of the multiple connections. These modules offer the most resilient and highly available multipathing for VMware ESX servers by providing coordinated path management between both the ESX server and the storage.

A MPP “claims” a physical path and “manages” or “exports” a logical device. The MPP is the only code that can associate a physical path with a logical device. So that path can not be managed by both a NMP and MPP at the same time.

The PSA introduces support for two more features of the SCSI 3 (SPC-3) protocol: TPGS and ALUA.

However, although SCSI-3 is supported for some VMFS functions (LSI emulation for MSCS environments) the volume manager functions within VMFS are still SCSI-2 compliant. As such, the 2TB per LUN limit still applies.

**Target Port Group Support (TPGS)** - Is a mechanism that gives storage devices the ability to specify path performance, and other characteristics to a host, like an ESX server. A Target Port Group (TPG) is defined as a set of target ports that are in the same target port Asymmetric Access State (AAS) at all times with respect to a given LUN.

Hosts can use AAS of a TPG to prioritize paths and make failover and load balancing decisions.

Control and management of ALUA AASs can operate in 1 of 4 defined modes:

1. **Not Supported** (Report and Set TPGs commands invalid).
2. **Implicit** (TPG AASs are set and managed by the array only, and reported with the Report TPGs

## Archives

[February 2010](#)

[October 2009](#)

[September 2009](#)

[June 2009](#)

[September 2008](#)

[July 2008](#)

[Subscribe to this blog's feed](#)

command)

3. **Explicit** (TPG AASs are set and managed by the host only with the Set TPGs command, and reported with the Report TPGs command)
4. **Both** (TPG AASs can be set and managed by either the array or the host)

**Asymmetric Logical Unit Access (ALUA)** is a standard-based method for discovering and managing multiple paths to a SCSI LUN. It is described in the T10 SCSI-3 specification SPC-3, section 5.8. It provides a standard way to allow devices to report the states of their respective target ports to hosts. Hosts can prioritize paths and make failover/load balancing decisions.

Since target ports could be on different physical units, ALUA allows different levels of access for target ports to each LUN. ALUA will route I/O to a particular port to achieve best performance.

**ALUA Follow-over Feature:**

A Follow-over scheme implemented to minimize path thrashing in cluster environments with shared LUNs. Follow-over supported for explicit ALUA only and is not part of the ALUA standard.

When an ESX host detects a TPG AAS change that it did not cause:

- o It will not try to revert this change even if it only has access to non-optimized paths
- o Thus, it follows the TPG of the array

---

Posted by Paul Manning on October 19, 2009 | [Permalink](#) |



## TrackBack

TrackBack URL for this entry:

<http://www.typepad.com/services/trackback/6a00d8341c328153ef0120a5f6c428970b>

Listed below are links to weblogs that reference [vStorage Multi Paths Options in vSphere](#):

## Comments

You can follow this conversation by subscribing to the [comment feed](#) for this post.

## Post a comment

If you have a TypeKey or TypePad account, please [Sign In](#).

Name:

Email Address: (Not displayed with comment.)

URL:

☐ Remember personal info?

Comments: