



PS Series Groups

Storage Array Network Performance Guidelines

Abstract

This Technical Report describes network configuration requirements and recommendations for high performance and reliability with a PS Series storage array.

Copyright © 2004—2006 EqualLogic, Inc.

September 2006

EqualLogic is a registered trademark of EqualLogic, Inc.

All trademarks and registered trademarks mentioned herein are the property of their respective owners.

Possession, use, or copying of the documentation or the software described in this publication is authorized only under the license agreement specified herein.

EqualLogic, Inc. will not be held liable for technical or editorial errors or omissions contained herein. The information in this document is subject to change. Your performance can vary.

PS Series Firmware Version 3.0

Table of Contents

| | |
|--|---|
| Introduction | 1 |
| Network Requirements and Recommendations | 1 |
| Spanning-Tree Protocol | 3 |
| Flow Control | 3 |
| Unicast Storm Control | 4 |
| Jumbo Frames | 4 |
| VLANs | 5 |
| Storage Array Network Configuration | 5 |
| Configuring a Network Interface | 7 |
| Documentation and Customer Support | 9 |

A PS Series group is an iSCSI SAN consisting of one or more EqualLogic® PS Series storage arrays (group members) connected to an IP network and managed as a single system. This Technical Report provides network requirements and performance recommendations for a PS Series group.

Note: Properly connected network cables are vital to array and group operation and performance. If you have a dual control module array, it is especially important that you use the correct cable configuration to take advantage of the array's failover capabilities. For the highest performance and availability, follow the network cable connection recommendations described in the *Hardware Maintenance* manual for your array model.

Detailed network configuration guidelines are beyond the scope of this document. All the usual rules for proper network configuration apply to the members (arrays) in a PS Series group.

Network Requirements and Recommendations

The networking requirements for group members are described in Table 1.

Table 1: Networking Requirements

| Requirement | Description |
|--|--|
| Appropriate network cable | For copper-based networks, use Category 5E or Category 6 cables with RJ45 connectors. You can also use Category 5 cables if they adhere to the TIA/EIA TSB95 standard. For fiber optic networks (used only with PS50E to PS2400E models), use cables with LC connectors. Fiber optic networks also require small form factor pluggable (SFP) connectors, which can be obtained from EqualLogic. |
| Network connection | Each array must have at least one active network connection and can have a maximum of three. |
| Network access to the group IP address | Each array must have at least one network interface configured on the same subnet as the group IP address. |
| Connectivity between switches | In a single-subnet group in which the arrays' configured network interfaces are connected to multiple switches, there must be network connectivity between the switches. |

For optimal performance and availability, follow the recommendations in Table 2:. The following sections provide background information on some of the recommendations.

Table 2: Networking Recommendations

| Recommendation | Description |
|---|--|
| Configure multiple network connections on an array. | Connecting multiple network interfaces on an array (assigning each interface an IP address and netmask) improves performance and availability. Connect the cables to different switches, if possible, for high availability. |
| In a multi-subnet group, ensure access to the group IP address. | In a multi-subnet group, we strongly recommend the arrays' configured network interfaces be able to access the subnet on which the group IP address resides. |

| Recommendation | Description |
|--|---|
| Use a switched, Gigabit Ethernet network. | For the best performance, you should connect arrays and hosts to a switched network and make sure that all network connections between hosts and arrays are Gigabit Ethernet. An array can operate at 10 and 100 Megabits, but performance will be significantly degraded. |
| Use redundant network paths between hosts and arrays. | A multipathing solution helps to ensure that no single point of failure exists between hosts and arrays. |
| For replication, use a reliable, adequately sized network link. | <p>For effective and predictable replication, be sure that the network link between the partners is reliable and provides sufficient bandwidth for copying data.</p> <p>See the <i>Group Administration</i> manual for more information on replication and network links.</p> |
| Do not use STP on switch ports that connect end nodes. | <p>It is recommended that you do not use Spanning-Tree Protocol (STP) on switch ports that connect end nodes (iSCSI initiators or storage array network interfaces). However, if you want to use STP or RSTP (preferable to STP), you should enable the port settings—available on some switches—that let the port immediately transition into STP forwarding state upon linkup (for example, Cisco’s PortFast option). This functionality can reduce network interruptions that occur when devices restart, and should only be enabled on switch ports that connect end nodes.</p> <p>Note: The use of Spanning-Tree for a single-cable connection between switches is encouraged, as is the use of trunking for multi-cable connections between switches.</p> <p>See <i>Spanning-Tree Protocol</i> on page 3 for more information.</p> |
| Use Flow Control on switches and NICs. | <p>It is recommended that you enable Flow Control on each switch port that handles iSCSI traffic. If your server is using a software iSCSI initiator and NIC combination to handle iSCSI traffic, you must also enable Flow Control on the NICs to obtain any performance benefit. PS Series storage arrays will correctly respond to Flow Control.</p> <p>See <i>Flow Control</i> on page 3 for more information.</p> |
| Disable unicast storm control on switches that provide this functionality. | <p>It is recommended that you disable unicast storm control on each switch that handles iSCSI traffic. However, the use of broadcast and multicast storm control is encouraged on switches.</p> <p>Not all switches provide unicast storm control functionality.</p> <p>See <i>Unicast Storm Control</i> on page 4 for more information.</p> |
| Use Jumbo Frames on switches and NICs. | <p>Enable Jumbo Frames on each switch that handles iSCSI traffic. If your server is using a software iSCSI initiator and NIC combination, you must also enable Jumbo Frames on the NICs that handle iSCSI traffic to obtain any performance benefit and ensure consistent behavior.</p> <p>See <i>Jumbo Frames</i> on page 4 for more information.</p> |
| Use VLANs. | <p>Configure VLANs on switches to separate iSCSI traffic from other network traffic.</p> <p>See <i>VLANs</i> on page 5 for more information.</p> |

Spanning-Tree Protocol

Spanning-Tree Protocol (STP, defined in the IEEE 802.1d specification), is a link management protocol that prevents loops in an Ethernet network by ensuring that only one active path exists between switches. Upon linkup, a switch performs a 30 to 50 second STP calculation to transition ports into forwarding or blocking state. During this time, no user data passes through the ports, resulting in an increase in the time it takes for links to come online.

Because STP can increase the time it takes to recover from a PS Series array control module failover or a network switch failure, it is recommended that you do not use STP on switch ports that connect end nodes (iSCSI initiators or storage array network interfaces).

Note: The use of STP for a single-cable connection between switches is encouraged, as is the use of trunking for multi-cable connections between switches.

If you must use STP, it is recommended that you enable the switch port settings that allow the immediate transition of the port into STP forwarding state upon linkup. This functionality—not available on all switches—can reduce network interruptions that occur when devices restart. For example, Cisco enhanced the original STP 802.1d specification with a proprietary feature called PortFast, which immediately transitions a port into STP forwarding state upon linkup.

The settings should be enabled only on switch ports that connect end nodes (iSCSI initiators and storage array network interfaces). Do not connect other hubs, switches, concentrators, or bridges to these ports, because this may cause infinite packet loops.

It is also preferable to use Rapid Spanning Tree Protocol (RSTP, defined in the IEEE 802.1w specification) instead of STP. RSTP allows a switch port to bypass the Spanning-Tree listening and learning states and quickly enter the STP forwarding state. To achieve fast convergence on a port, RSTP relies upon two variables: edge ports and link type. Switch ports connecting end nodes (iSCSI initiators and storage array network interfaces) should be configured as edge ports to take advantage of RSTP features.

See your switch documentation for more information on STP.

Flow Control

On many networks, there can be an imbalance in the network traffic between the devices that send traffic and the devices that receive the traffic. This is often the case in SAN configurations in which many servers (initiators) are communicating with storage devices (targets). If senders transmit data simultaneously, they may exceed the throughput capacity of the receiver. When this occurs, the receiver may drop packets, forcing senders to retransmit the data after a delay. Although this will not result in any loss of data, latency will increase because of the retransmissions, and I/O performance will degrade.

Flow Control eliminates this problem. When enabled, Flow Control allows the receiver to instruct the sender to “throttle back” when the receiver senses that it is being overwhelmed by packets. The receiver does this by sending “pause frames” to the sender, which causes the sender to slow packet transmission for a short period of time. This lets the receiver process its backlog so it can later resume accepting input. The amount of delay introduced by this action is dramatically less than the overhead caused by TCP/IP packet retransmission.

It is recommended that you enable Flow Control on the switch ports that handle iSCSI traffic. If a server is using a software iSCSI initiator and NIC combination to handle iSCSI traffic, you must also enable Flow

Control on the NICs to obtain any performance benefit. PS Series storage arrays will automatically correctly respond to Flow Control.

See your switch and NIC documentation for information about enabling Flow Control.

Unicast Storm Control

A traffic “storm” occurs when a large outpouring of packets creates excessive network traffic that degrades network performance. Many switches have traffic storm control features that prevent ports from being disrupted by broadcast, multicast, or unicast traffic storms on physical interfaces. These features typically work by discarding network packets when the traffic on an interface reaches a percentage of the overall load (usually 80 percent, by default).

Because iSCSI traffic is unicast traffic and can typically use the entire link, it is recommended that you disable unicast storm control on switches that handle iSCSI traffic. However, you *can* use broadcast and multicast storm control on switches.

See your switch documentation for information on disabling unicast storm control.

Note: Not all switches provide unicast storm control functionality.

Jumbo Frames

Ethernet traffic is transported in *frames*. A standard Ethernet frame allows up to 1500 bytes of data to be transferred in a single Ethernet transaction. Jumbo Frames extend Ethernet frames to 9000 bytes and allow more data to be transferred with each Ethernet transaction. Jumbo Frames help reduce the interrupt overhead on the server.

PS Series storage arrays support standard Ethernet frames and Jumbo Frames up to 9000 bytes (9018 including the TCP header itself). This is sometimes referred to as the “Alteon standard.” PS Series storage arrays also support path MTU, which is the ability to automatically detect the maximum frame size between TCP/IP end points.

It is recommended that you enable Jumbo Frames on each switch that handles iSCSI traffic. If your server is using a software iSCSI initiator and NIC combination to handle iSCSI traffic, you must also enable Jumbo Frames on the NICs to obtain any performance benefit and ensure consistent behavior.

To simplify troubleshooting initial deployments, make sure that your iSCSI initiator, switches, and storage arrays are fully operational *before* enabling Jumbo Frames on the NICs.

Note: To obtain a performance benefit with Jumbo Frames and ensure consistent behavior, *all* devices in the network path between servers and the PS Series group—including the switches and the NICs used for iSCSI traffic—must have Jumbo Frames enabled.

Switches configured for Jumbo Frames will support both standard Ethernet frames and Jumbo Frames. However, if a NIC is configured for Jumbo Frames, but the switch is not, you may experience inconsistent behavior. For example, the switch will function properly if the frames are small, but once the NIC attempts to send frames larger than 1500 bytes, the switch will not be able to handle the frames and will drop them. Also, if some switches are configured for Jumbo Frames, but others are not, you may experience inconsistent behavior if routing changes occur after the connection has been established.

Many low and mid-range network switches do not support Jumbo Frames. On switches that do support them, the feature is usually disabled by default, so you need to manually enable it.

Usually, VLANs must be configured in order to enable Jumbo Frames on switches. See your switch documentation for details.

Every initiator connection has a corresponding event message. The message indicates whether the connection is using standard or Jumbo Frames. Use the Group Manager GUI to look for connections with standard frames.

VLANs

VLANs provide performance benefits similar to those of separate network switches, but with fewer physical switches and less management overhead. Because VLANs use logical rather than physical connections, they are flexible to operate.

Using VLANs to separate iSCSI SAN traffic from LAN traffic also allows you to enable switch-based security enhancements, including port blocking and address filtering.

It is recommended that you configure virtual LANs (VLANs) on your switches to separate iSCSI SAN traffic from other network traffic.

Typically, VLANs are separate subnets in TCP/IP networks, so connectivity between VLANs usually requires L3 routing, which is not supported by all switches. VLANs also support different network settings, including Jumbo Frames.

Note: PS Series storage arrays do not support VLAN tagging.

See your switch documentation for information on configuring VLANs.

Storage Array Network Configuration

A PS Series storage array must have at least one and can have up to three active network connections at one time. Multiple network connections are *strongly* recommended for performance and availability.

General recommendations for connecting network cables are as follows:

- Connect network interfaces in the following order: Ethernet Port 0, Ethernet Port 1, and Ethernet Port 2. Connecting and configuring all three network interfaces provides the maximum network bandwidth.
- Connect network interfaces to different network switches to increase availability.
- In a dual control module array, to ensure network connectivity if a control module fails, connect cables to the network interface ports on both control modules. For example, connect a cable to Ethernet Port 0 on CM0 and Ethernet Port 0 on CM1.

See the *Hardware Maintenance* manual for your array model for detailed information about connecting network cables.

As examples, the following figures show the recommended network configuration for single control module PS400E and PS3000 arrays. If one interface fails, initiators can be redirected to the other interfaces.

Figure 1: PS400E Array Single Control Module Configuration

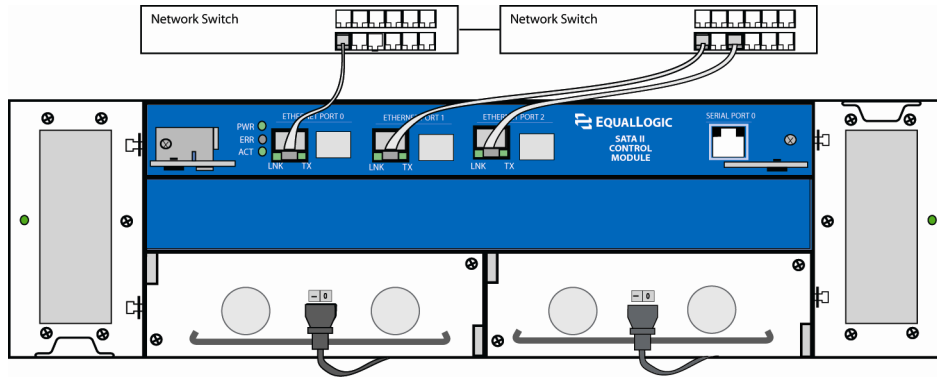
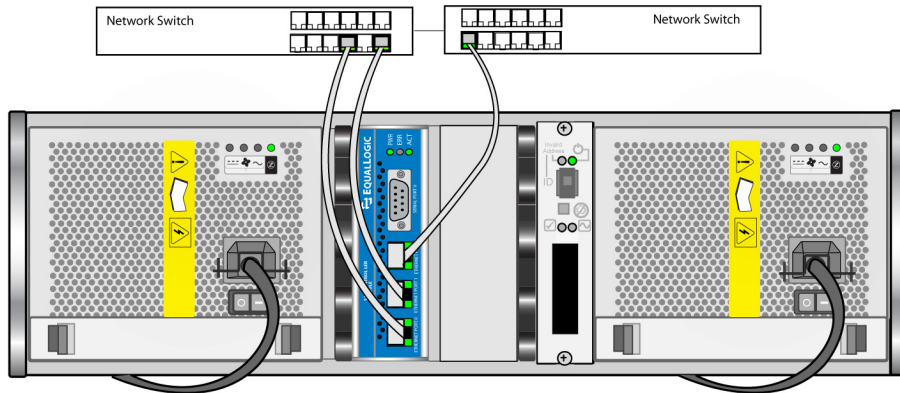


Figure 2: PS3000 Array Single Control Module Configuration



The following figures show the recommended network configuration for dual control module PS400E and PS3000 arrays. This configuration provides the highest network availability and the maximum network bandwidth.

Figure 3: PS400E Array Dual Control Module Configuration

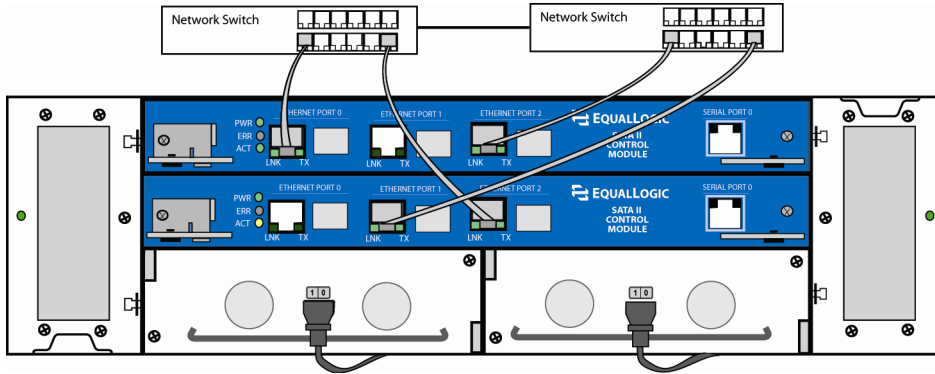
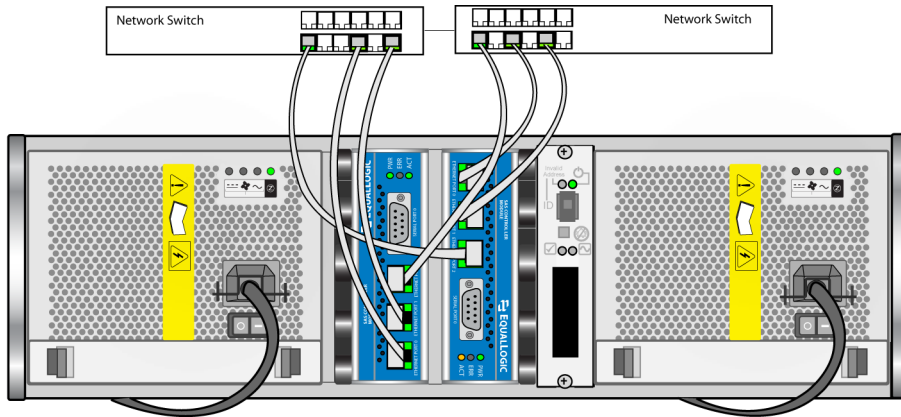


Figure 4: PS3000 Array Dual Control Module Configuration



Configuring a Network Interface

After connecting a network interface to a network, as described in the *Hardware Maintenance* manual for your PS Series array model, you can configure the interface by assigning it an IP address and subnet mask. You can also modify an existing network interface configuration.

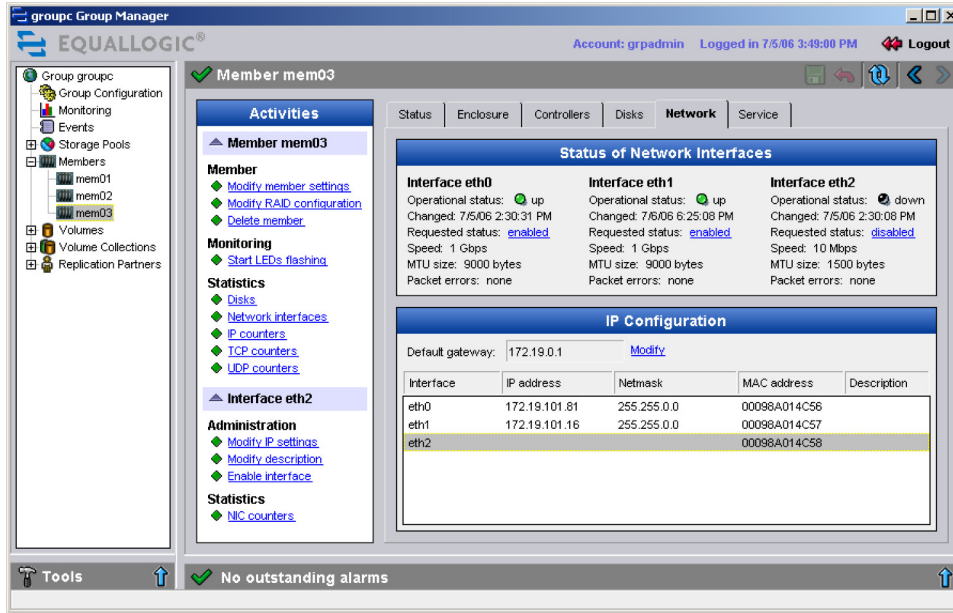
Note: Changing the IP address for a functioning network interface will disconnect all iSCSI initiators that are connected to that interface. However, all initiators should reconnect automatically.

To configure a network interface using the GUI, click:

Members → *member_name* → Network tab

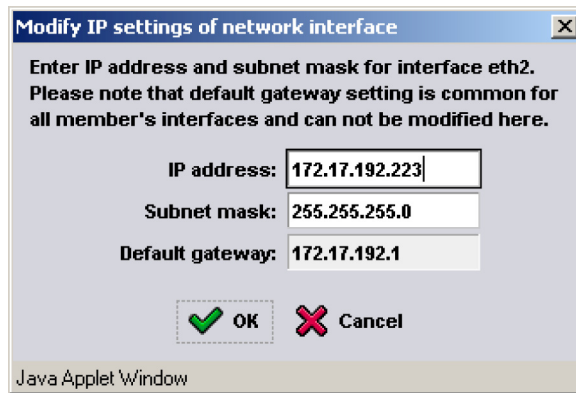
The Member Network window appears, displaying the current network configuration for each network interface on the member. As shown in the Figure 5:, eth2 (Ethernet Port 2) does not have an IP address assigned.

Figure 5: Member Network



To configure (or modify) a network interface, double-click the network interface in the IP Configuration panel or select the network interface and click Modify IP settings in the Activities panel. The Modify IP Settings dialog box appears.

Figure 6: Modify IP Settings



Specify the following information about the network interface, then click OK:

- **IP address.** You must specify an IP address. In a multi-subnet group, each member must have at least one network interface that is on the same subnet as the group IP address.
- **Subnet mask.** You must specify a subnet mask (netmask) for the IP address.

Note: The default gateway is the same for all network interfaces on an array. You cannot modify the default gateway using the Modify IP Settings dialog box.

You can also use the CLI to configure a network interface and set its status to up or down:

```
member select member_name eth select port ipaddress ip_address netmask mask_ip  
member select member_name eth select port up
```

The port variable can be 0, 1, or 2 for eth0, eth1, and eth2, respectively.

The following examples assign an IP address and netmask to eth2, and enable it for use:

```
> member select lab23 eth select 2 ipaddress 152.17.2.144 netmask 255.255.255.0  
> member select lab23 eth select 2 up
```

Documentation and Customer Support

EqualLogic PS Series storage array documentation includes the following:

- *Release Notes*. Provides the latest information about PS Series arrays.
- *QuickStart*. Describes how to set up the hardware and start using PS Series arrays.
- *Group Administration*. Describes how to use the Group Manager GUI to manage a PS Series group. This manual provides comprehensive information about concepts and procedures.
- *CLI Reference*. Describes how to use the Group Manager command line interface to manage a group and individual arrays.
- *Hardware Maintenance*. Provides information on maintaining the storage array hardware.

In addition, visit the EqualLogic Customer Support website, where you can download documentation and firmware and view FAQs, Solutions, and Technical Reports, if you have a valid service contract or an array that is still under the first year of warranty. You can also submit a support request from the Customer Support website.

- From the EqualLogic website (www.equallogic.com), click Login and log in to your support account.
- If you do not have an account, create one by clicking the link under the login prompt.

To contact EqualLogic Customer Support, send e-mail to supportnp@equallogic.com. If the issue is urgent call to speak with a member of the Customer Support team:

- In the United States: 877-887-7337
- From outside the United States: +00 1 919-767-5729