



Sun[™] Solaris[™]

Implementing Highly Available Network Access

Abstract

This Technical Report describes how to implement highly-available network access from Sun Solaris 8 and Solaris 9 servers to a PS Series group.

Copyright © 2004, 2005 EqualLogic, Inc.

January 2005

EqualLogic is a registered trademark of EqualLogic, Inc.

Sun, Sun Microsystems, the Sun logo, and Solaris are trademarks, registered trademarks, or service marks of Sun Microsystems, Inc. in the U.S. and other countries

All other 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.

PS Series Firmware Version 2.0 or later.

Table of Contents

Implementing Network-Based Multipathing on Solaris.....	1
PS Series Group Requirements	2
Configuring the Active-Active Scenario	3
Preparing the Solaris Server (Active-Active)	3
Setting Up For High Availability (Active-Active).....	3
Configuring the Active-Passive Scenario.....	5
Preparing the Solaris System (Active-Passive).....	5
Setting Up For High Availability (Active-Passive)	5
Documentation and Customer Support	6

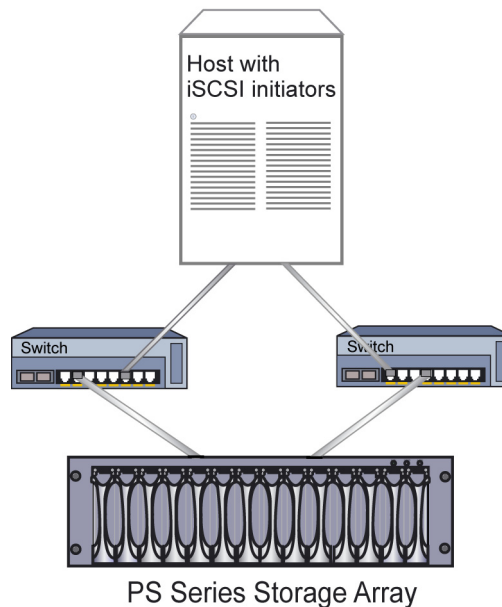
Implementing Network-Based Multipathing on Solaris

You can implement highly-available network access from Sun Solaris 8 and Solaris 9 servers to a reliable and scalable PS Series group. The foundation of a group is the PS Series storage array, which provides multiple network connections, in addition to no-single-point-of-failure hardware.

To use multipathing, you must configure the IP Network Multipathing component in your Solaris operating system environment. Multipathing utilizes the operating system's IP failover capability and assumes no special Ethernet hardware, software, or switches. See the Solaris multipathing documentation for detailed information in installation and configuration.

Multipathing requires multiple physical paths to a storage device (for example, to each array in a PS Series group). In the case of a failure on one of the paths (Ethernet adapter, cable, or switch failure) an alternate path can be used without disruption. The following figure shows a configuration in which there are multiple network paths between a host and an array:

Multiple Network Paths to a PS Series Storage Array



You can configure network multipathing in the following ways:

- An active-active configuration where all the paths can carry data simultaneously to the device. If a path fails, the alternate path automatically handles all I/O.
- An active-passive configuration where only one path is active at any given time. If that path fails, the alternate path automatically assumes the role of the active Ethernet connection.

Note: To configure network multipathing, you need root access to your Solaris server.

The following sections describe both configurations.

PS Series Group Requirements

Properly connecting network cables to a PS Series storage array is vital to array operation and performance. With the dual control module array, it is especially important that you use the correct configuration to take advantage of the array's failover capabilities. A thorough discussion of this topic is beyond the scope of this document. For more information, consult the Technical Report, *Network Connection and Performance Guidelines*.

Briefly, in a dual control module array, each control module has three network interface ports, labeled `eth0`, `eth1`, and `eth2`. This provides three pairs of network interface ports although only one port in a pair can be used for I/O at one time.

For multipathing access to a PS Series group, each array in the group must have multiple network interfaces connected to a network (usually through a network switch), each assigned a unique IP address. In a dual control module array, both ports in a pair share the same IP address. To improve availability, it is recommended that you connect and configure all three multiple network interfaces.

After creating the PS Series group and connecting cables, you can create volumes. See the *QuickStart* or the *Group Administration* manual for information about setting up a group and creating volumes. You can then connect to the volumes using the iSCSI initiators installed on the Solaris servers.

Access control records are used to restrict server access to data in a PS Series group. A volume and its snapshots share a list of access control records (sometimes called the access control list). You can configure a record to apply to the volume, its snapshots, or both, as needed.

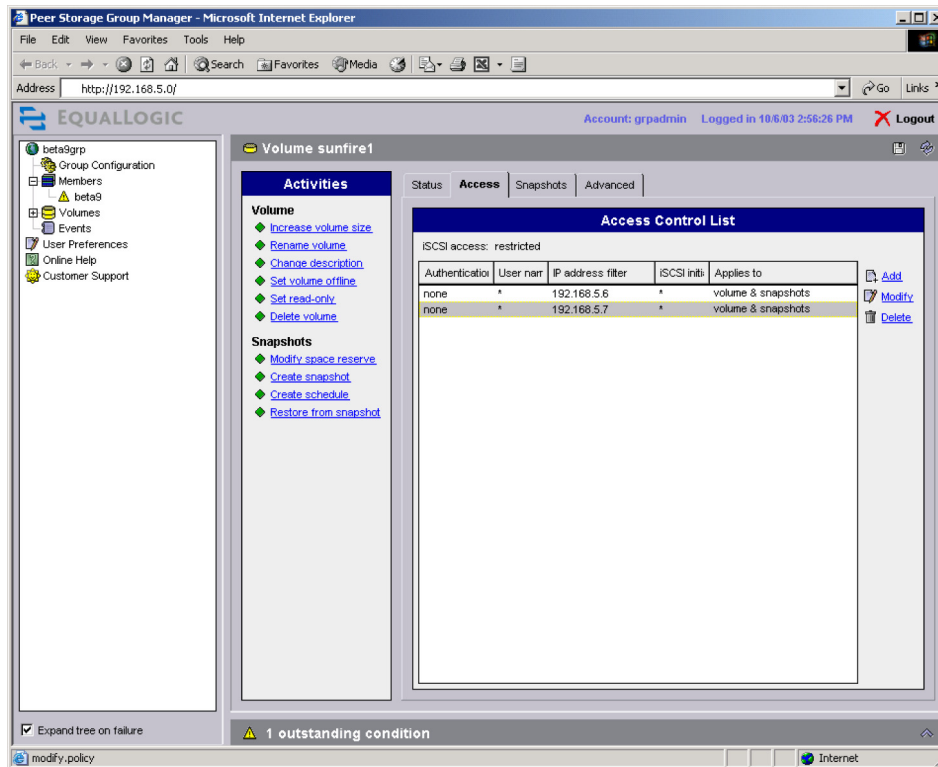
In each access control record, you can specify an IP address, iSCSI initiator name, or CHAP (Challenge Handshake Authentication Protocol) user name (or any combination). A server must match *all* the requirements in *one* record in order to access the volume or snapshot. For example, if a record includes both an IP address and a CHAP user name, a server must present the IP address *and* supply the CHAP user name and its associated password (using the iSCSI initiator) in order to match the record.

Note: If you use IP addresses or iSCSI initiator names to restrict access, create an access control record for each IP address or initiator name presented by the server.

If you use a CHAP user name to restrict access, it is recommended that you also specify an IP address in the access control record. If you only use CHAP, initiators that support discovery will attempt to log in to the target, even if they do not have the right access credentials, resulting in a large number of events logged in the group and an inefficient use of resources.

If you are limiting access by IP address, you need to configure access control records to allow access from all possible IP addresses associated with a Solaris server. The following figure shows two access control records – one for each of the possible IP addresses used in the active-active scenario (described in the following section):

PS Series Group Showing Access Control Records



Configuring the Active-Active Scenario

Preparing the Solaris Server (Active-Active)

You must have at least two Ethernet adapters installed on the server on the same subnet and connected to the network. Both adapters are in use simultaneously. If a failure occurs in one adapter, the server directs all network access automatically from the failed adapter to the other adapter, ensuring uninterrupted access to the network.

It is also recommended that you plug each of these adapters into different switches for redundancy.

Each adapter requires two IP addresses. For each adapter, one address is designated as highly available or “active” (used by applications) and the other address is designated as a test address. If you plan to create more than two paths to your storage, you need more than four addresses.

Setting Up For High Availability (Active-Active)

The following example shows how to make the IP addresses associated with the host names `sunfire1` and `sunfire2` highly available.

Note: If you do not know the adapter names, the output from the `ifconfig -a` command provides you with this information. In the following example, the adapter names are `bge1` and `bge2`.

1. From a root login session, create or modify your existing `/etc/hostname.<adapter>` files. Replace the **bolded** entries with your own site-specific information.

```

/etc/hostname.bge1
sunfire1 netmask + broadcast + group iscsigroup up \
addif testaddr1 deprecated -failover netmask + broadcast + up

/etc/hostname.bge2
sunfire2 netmask + broadcast + group iscsigroup up\
addif testaddr2 deprecated -failover netmask + broadcast + up

/etc/hosts
192.168.5.6      sunfire1      loghost
192.168.5.7      sunfire2
192.168.5.8      testaddr1
192.168.5.9      testaddr2

/etc/netmask
192.168.5.0      255.255.255.0

```

2. Reboot the system and verify the network configuration by running `ifconfig -a`. Note that `bge1` and `bge2` each have two IP addresses bound to them: one active address and one test address.

```

[root@sunfire1 /]# ifconfig -a
lo0: flags=1000849<UP,LOOPBACK,RUNNING,MULTICAST,IPv4> mtu 8232 index 1
    inet 127.0.0.1 netmask ff000000
bge1: flags=1000843<UP,BROADCAST,RUNNING,MULTICAST,IPv4> mtu 1500 index 2
    inet 192.168.5.6 netmask ffffffff0 broadcast 192.168.5.255
    groupname iscsigroup
    ether 0:3:ba:3f:51:cc
bge1:1: flags=9040843<UP,BROADCAST,RUNNING,MULTICAST,DEPRECATED,IPv4,NOFAILOVER> mtu
1500 index 2
    inet 192.168.5.8 netmask ffffffff0 broadcast 192.168.5.255
bge2: flags=1000843<UP,BROADCAST,RUNNING,MULTICAST,IPv4> mtu 1500 index 3
    inet 192.168.5.7 netmask ffffffff0 broadcast 192.168.5.255
    groupname iscsigroup
    ether 0:3:ba:3f:51:cd
bge2:1: flags=9040843<UP,BROADCAST,RUNNING,MULTICAST,DEPRECATED,IPv4,NOFAILOVER> mtu
1500 index 3
    inet 192.168.5.9 netmask ffffffff0 broadcast 192.168.5.255

```

3. Test that you have configured the system correctly by introducing a network failure into the system. In the following example, the failover was caused by pulling the `bge1` network cable:

```

[root@sunfire1 /]# Oct  3 12:55:03 sunfire1 bge: NOTICE: bge1: link down
Oct  3 12:55:05 sunfire1 in.mpathd[80]: NIC failure detected on bge1 of group
iscsigroup
Oct  3 12:55:05 sunfire1 in.mpathd[80]: Successfully failed over from NIC bge1 to NIC
bge2

```

The following output shows the interface state after the failover:

```

[root@sunfire1 /]# ifconfig -a
lo0: flags=1000849<UP,LOOPBACK,RUNNING,MULTICAST,IPv4> mtu 8232 index 1
    inet 127.0.0.1 netmask ff000000
bge1: flags=19000842<BROADCAST,RUNNING,MULTICAST,IPv4,NOFAILOVER,FAILED> mtu 0 index 2
    inet 0.0.0.0 netmask 0
    groupname iscsigroup
    ether 0:3:ba:3f:51:cc
bge1:1:
flags=19040843<UP,BROADCAST,RUNNING,MULTICAST,DEPRECATED,IPv4,NOFAILOVER,FAILED> mtu
1500 index 2
    inet 192.168.5.8 netmask ffffffff0 broadcast 192.168.5.255

```

```

bge2: flags=1000843<UP,BROADCAST,RUNNING,MULTICAST,IPv4> mtu 1500 index 3
    inet 192.168.5.7 netmask ffffffff0 broadcast 192.168.5.255
    groupname iscsigroup
    ether 0:3:ba:3f:51:cd
bge2:1: flags=9040843<UP,BROADCAST,RUNNING,MULTICAST,DEPRECATED,IPv4,NOFAILOVER> mtu
1500 index 3
    inet 192.168.5.9 netmask ffffffff0 broadcast 192.168.5.255
bge2:2: flags=1000843<UP,BROADCAST,RUNNING,MULTICAST,IPv4> mtu 1500 index 3
    inet 192.168.5.6 netmask ffffffff0 broadcast 192.168.5.255

```

In the following output, notice is given that the link is up on `bge1` and the interface has been restored to its initial state:

```

[root@sunfire1 /]# Oct  3 12:55:40 sunfire1 bge: NOTICE: bge1: link up 1000Mbps Full-
Duplex
Oct  3 12:55:45 sunfire1 in.mpathd[80]: Successfully failed back to NIC bge1
Oct  3 12:55:45 sunfire1 in.mpathd[80]: NIC repair detected on bge1 of group
iscsigroup

```

Configuring the Active-Passive Scenario

Preparing the Solaris System (Active-Passive)

You must have at least two Ethernet adapters installed on the server on the same subnet and connected to the network. Only one adapter is in use at a time. If a failure occurs in one adapter, the server directs all network access automatically from the failed adapter to the other adapter, ensuring uninterrupted access to the network.

It is also recommended that you plug each of these adapters into different switches for redundancy.

In an active-passive configuration with two adapters, you need at least three IP addresses. One address is designated as highly-available or “active” (to be used by applications). The system uses the other two addresses as test addresses – one for each of the two interfaces. If you plan to create more than two paths to your storage, you need additional addresses.

Setting Up For High Availability (Active-Passive)

The following example shows how to make the IP address associated with the host name `sunfire` highly available.

Note: If you do not know the adapter names, the output from the `ifconfig -a` command provides you with this information. In the following example, the adapter names are `bge1` and `bge2`.

1. From a `root` login session, create or modify your existing `/etc/hostname.<adapter>` files. Replace the **bolded** entries with your own site-specific information.

```

/etc/hostname.bge1
sunfire netmask + broadcast + group iscsigroup up \
addif testaddr1 deprecated -failover netmask + broadcast + up

/etc/hostname.bge2
testaddr2 netmask + broadcast + deprecated group iscsigroup up \
-failover standby up

```

```

/etc/hosts
127.0.0.1      localhost
192.168.5.5   sunfire      loghost
192.168.5.6   testaddr1
192.168.5.7   testaddr2

```

```

/etc/netmask
192.168.5.0   255.255.255.0

```

2. Reboot the system and verify the network configuration. The active interface (bge1) has two IP addresses bound to it (192.168.5.5 and 192.168.5.6) . The passive interface, bge2, has one IP address bound to it (192.168.5.7). It is in a down state until a failover occurs.
3. You can test that you have configured the system correctly by introducing a network failure into the system. The following example shows a NIC failure detected on bge1, caused by pulling the bge1 network cable. The example also show the successful failover to bge2:

```

[root@sunfire etc]# Oct  6 09:42:13 sunfire bge: NOTICE: bge1: link down
Oct  6 09:42:15 sunfire in.mpathd[80]: NIC failure detected on bge1 of group
iscsigroup
Oct  6 09:42:15 sunfire in.mpathd[80]: Successfully failed over from NIC bge1 to NIC
bge2

```

Run `ifconfig -a` to verify your results as shown in the following sample output:

```

[root@sunfire etc]# ifconfig -a
lo0: flags=1000849<UP,LOOPBACK,RUNNING,MULTICAST,IPv4> mtu 8232 index 1
    inet 127.0.0.1 netmask ff000000
bge1: flags=19000842<BROADCAST,RUNNING,MULTICAST,IPv4,NOFAILOVER,FAILED> mtu 0 index 2
    inet 0.0.0.0 netmask 0
    groupname iscsigroup
    ether 0:3:ba:3f:51:cc
bge1:1:
flags=19040843<UP,BROADCAST,RUNNING,MULTICAST,DEPRECATED,IPv4,NOFAILOVER,FAILED> mtu
1500 index 2
    inet 192.168.5.6 netmask ffffffff broadcast 192.168.5.255
bge2:
flags=29040843<UP,BROADCAST,RUNNING,MULTICAST,DEPRECATED,IPv4,NOFAILOVER,STANDBY> mtu
1500 index 3
    inet 192.168.5.7 netmask ffffffff broadcast 192.168.5.255
    groupname iscsigroup
    ether 0:3:ba:3f:51:cd
bge2:1: flags=21000843<UP,BROADCAST,RUNNING,MULTICAST,IPv4,STANDBY> mtu 1500 index 3
    inet 192.168.5.5 netmask ffffffff broadcast 192.168.5.255

```

The following output shows the successful fail back to bge1:

```

[root@sunfire etc]# Oct  6 09:43:05 sunfire bge: NOTICE: bge1: link up 1000Mbps Full-
Duplex
Oct  6 09:43:10 sunfire in.mpathd[80]: Successfully failed back to NIC bge1
Oct  6 09:43:10 sunfire in.mpathd[80]: NIC repair detected on bge1 of group iscsigroup

```

Documentation and Customer Support

Visit the EqualLogic Customer Support website, where you can download the latest documentation and firmware. You can also view FAQs, the Knowledge Base, and Tech Reports and submit a service request.

EqualLogic PS Series storage array documentation includes the following:

- *Release Notes*. Provides the latest information about PS Series storage arrays and groups.
- *QuickStart*. Describes how to set up the hardware and start using a PS Series storage array.
- *Group Administration*. Describes how to use the Group Manager GUI to manage a PS Series group. This manual provides comprehensive information about product 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 PS Series storage array hardware.

To access the Customer Support website, from the EqualLogic website (www.equallogic.com), click Support and log in to a support account. If you do not have an account, create one by clicking the link under the login prompt.

To contact customer support, send e-mail to supportnp@equallogic.com. If the issue is urgent, call 1-877-887-7337 to speak with a member of the customer support team.