

January 2001
147Q-0101A-WWEN

Prepared by High Availability

Compaq Computer Corporation

Contents

Increasing Availability of Cluster Communications in a Windows Cluster.....3

- Types of Cluster Communication.....3
- Significance of Cluster Communication Paths3
- Communication Points of Failure.....4
- Building Blocks of Communication Path Redundancy6
- Compaq Network Teaming and Configuration Fault Tolerant Features6
- Redundant Network Interface Controllers (NIC).....7
- Dual-Port Network Interface Controller (NIC).....7
- Interconnect Paths8
- PCI Hot Plug.....8

Elements of a successful Cluster failover in Network Communications:..... 11

- Compaq Hardware and Software..... 11
- Figure 6. Redundant NIC Configuration..... 12
- Microsoft Cluster Service..... 13

Other Information..... 14

- Compaq ServerNet II..... 14

Summary 15

Appendix A 16

- The Compaq Network Teaming and Configuration Feature..... 16
- General Configuration Notes ... 18

Increasing Network Availability in a Microsoft Windows Cluster

***Abstract:** This paper addresses the importance of network communication fault tolerance. Network communication mechanisms are defined. With the use of the Compaq Network Teaming and Configuration Utility, a redundant Network Interface Controller (NIC) pair can be created to provide network high availability. The load balancing feature of the Compaq Network Teaming and Configuration Utility in a clustered environment is beyond the scope of this paper.*

Notice

©2000 Compaq Computer Corporation

Aero, ActiveAnswers, Compaq, the Compaq logo, Compaq Insight Manager, Himalaya, NetFlex, NonStop, ProLiant, ROMPaq, SmartStart, StorageWorks, Tandem, BackPaq, CompaqCare (design), Contura, Deskpro, DirectPlus, LicensePaq, LTE, MiniStation, PageMarq, PaqFax, PaqRap Presario, ProLinea, QVision, QuickBack, QuickFind, RemotePaq, ServerNet, SilentCool, SLT, SmartStation, SpeedPaq, Systempro, Systempro/LT, TechPaq, and TwinTray are registered United States Patent and Trademark Office.

Armada, Cruiser, Concerto, EasyPoint, EZ Help, FirstPaq, Innovate logo,, LTE Elite, MaxLight, MultiLock, Net1, PageMate, QuickBlank, QuickChoice, QuickLock, ProSignia, SoftPaq, SolutionPaq, Systempro/XL, UltraView, Vocalyst, Wonder Tools logo in black/white and color, and Compaq PC Card Solution logo are trademarks and/or service marks of Compaq Computer Corporation.

Fastart, Netelligent, SANworks, and TaskSmart are trademarks and/or service marks of Compaq Information Technologies Group, L.P. in the U.S. and/or other countries.

Active Directory, Microsoft, Windows 95, Windows 98, Windows, Windows NT, Windows NT Server and Workstation, Windows NT Enterprise Edition, Microsoft SQL Server for Windows NT are trademarks and/or registered trademarks of Microsoft Corporation.

Pentium, Xeon, Pentium II Xeon, and Pentium III Xeon are registered trademarks of Intel Corporation.

UNIX is a registered trademark of The Open Group.

NetWare, GroupWise, ManageWise, Novell Storage Services, and Novell are registered trademarks and intraNetWare, Border Manager, Console One, Z.E.N.works, NDS, and Novell Directory Services are trademarks of Novell, Inc.

SCO, UnixWare, OpenServer 5, UnixWare 7, Project Monterrey, and Tarantella are registered trademarks of the Santa Cruz Operation.

Adobe, Acrobat, and the Acrobat logo are trademarks of Adobe Systems, Inc.

Other product names mentioned herein may be trademarks and/or registered trademarks of their respective companies.

The information in this publication is subject to change without notice and is provided "AS IS" WITHOUT WARRANTY OF ANY KIND. THE ENTIRE RISK ARISING OUT OF THE USE OF THIS INFORMATION REMAINS WITH RECIPIENT. IN NO EVENT SHALL COMPAQ BE LIABLE FOR ANY DIRECT, CONSEQUENTIAL, INCIDENTAL, SPECIAL, PUNITIVE OR OTHER DAMAGES WHATSOEVER (INCLUDING WITHOUT LIMITATION, DAMAGES FOR LOSS OF BUSINESS PROFITS, BUSINESS INTERRUPTION OR LOSS OF BUSINESS INFORMATION), EVEN IF COMPAQ HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

The limited warranties for Compaq products are exclusively set forth in the documentation accompanying such products. Nothing herein should be construed as constituting a further or additional warranty.

This publication does not constitute an endorsement of the product or products that were tested. The configuration or configurations tested or described may or may not be the only available solution. This test is not a determination of product quality or correctness, nor does it ensure compliance with any federal state or local requirements.

Increasing Network Availability in a Microsoft Windows Cluster
White Paper prepared by High Availability

First Edition (January 2001)
Document Number 147Q-0101A-WWEN

Increasing Availability of Cluster Communications in a Windows Cluster

Types of Cluster Communication

Two types of interconnects have an immediate and dramatic effect on the availability of cluster communication. These are intra-cluster communication, and cluster-to-LAN communication.

Intra-cluster communication consists of information passed from one cluster node to another. The communication is performed over an interconnect. This consists of, at a minimum; two network Interface Cards (NICs) (one in each cluster node) and a crossover cable to connect the NICs.

Intra-cluster communication uses the interconnect data path to:

- Communicate individual resource and overall cluster status
- Send and receive cluster heartbeat signals
- Update the system registry information

Cluster-to-LAN communication consists of requests and responses to and from cluster nodes and network clients. This type of communication also exists in a non-clustered environment. As can happen with a stand-alone server, failure of a key network component results in downtime for network clients. Availability is of primary importance, especially when operating in a clustered environment. Ensuring network clients have access to their clustered applications and data depends on the availability of the cluster-to-LAN communication path.

Significance of Cluster Communication Paths

Since these communication mechanisms operate in a clustered environment, before discussing their significance, it is necessary to understand the terminology used to describe failures in a cluster. The Compaq ProLiant Cluster is highly available, rather than continuously available, so it is important to understand what parts of the system are vulnerable to faults. When a single hardware or software component fails and no component is available to take over, that component is identified as a single point of failure (SPOF). Due to the serious nature of single points of failure, a cluster should be designed to eliminate as many of them as possible.

Not all failures that interrupt cluster operations are single points of failure. As long as the cluster can recover from the failure, the data, applications, and network clients will return to normal operations as soon as the recovery process is complete. However, the period of time during the recovery process is considered unplanned or unscheduled downtime. While recovery is taking place, network clients will not have access to these cluster groups. Though not as catastrophic as a single point of failure, measures should also be taken to prevent these types of disruptions, and thereby reduce downtime.

As it pertains to cluster communications, there are two issues that adversely affect the operation of a cluster. The first issue momentarily disrupts operation by causing a failover event. The second issue is a single point of failure and disrupts operation until manual intervention by an administrator resolves the problem.

The first issue involves downtime associated with failover and failback events. When Microsoft Cluster Server detects an error that adversely affects the operation of a cluster group, it fails the cluster group from the one node to the other node. When Cluster Server detects an error that affects the operation of an entire cluster node, it fails all cluster groups running on that node to the other node. One such error occurs when communication between the cluster nodes is disrupted. If only one network connection exists in a cluster configuration, each time the network connection is disrupted for more than a few seconds Cluster Server will bring all cluster groups off-line on one of the nodes and fail them over to the other node. The process of failing over cluster groups takes time. The groups must be taken off-line on their primary node, the resources of each group (applications, drive volumes, IP addresses) must be transferred over to the other node, and the transferred data must be validated on the surviving node. While all of these operations occur, network clients are unable to access their cluster groups. Creating a redundant intra-cluster communication path easily and inexpensively minimizes the amount of downtime incurred due to this failure.

The second issue involves the network clients ability to access their clustered applications. Microsoft Cluster Server operates its failover and failback events at a cluster group level. A cluster group usually consists of an application, service, or file share, along with any dependent resources, such as drive volumes and IP addresses. Microsoft Cluster Server will likely be configured such that some cluster groups operate on cluster Node 1, and some on cluster Node 2. Each node is physically connected to the client LAN via a NIC, network cable, and a network hub.

In a Windows NT 4.0 Enterprise Edition environment a disruptive event will occur if the physical connection from an individual cluster node (ex. Node 1) to the client LAN is disrupted while the interconnect is still operational. For network clients whose cluster groups reside on Node 1, this event will prevent the clients from accessing their cluster groups (applications). Automatic failover of the cluster groups will not occur since Cluster Server, via the interconnect, believes both cluster nodes are operating normally. Until an administrator realizes the problem, discovers the root cause is a network error, and manually fails over all the cluster groups from Node 1 to Node 2, the clients cannot make use of Node 1's clustered applications. Creating a redundant cluster-to-LAN communication path easily and inexpensively minimizes the probability of this failure event.

Note: The above scenario does not apply to Windows 2000 Advanced Server or Windows 2000 Datacenter Server.

Communication Points of Failure

Several components make up the physical network of the cluster-to-LAN and intra-cluster communication paths. The failure of any one of these components renders the entire path inoperable, and results in the failure scenario previously described. Unless redundancy has been designed into the communication paths, a component failure will cause either a failover event or a complete disruption of access to certain cluster groups.

Understanding how each of these components plays a role in the interconnect and cluster-to-LAN data paths will help you comprehend the solutions discussed later in this paper. The following four hardware items are the primary points of failure.

- A port on a multi-port NIC (client or interconnect)
- A NIC (client or interconnect)

- A network cable
- A port on a network hub

Note: A fifth hardware item, a network hub, is also a single point of failure. However, the hub is viewed as a piece of the larger network, whose availability is a concern whether operating in a clustered environment or in a stand-alone server environment. In the “Examples” section, failure of a network hub is noted as a single point of failure when appropriate. Discussing how to resolve the effects of such a failure is beyond the scope of this document.

The diagram below depicts each of these failure points.

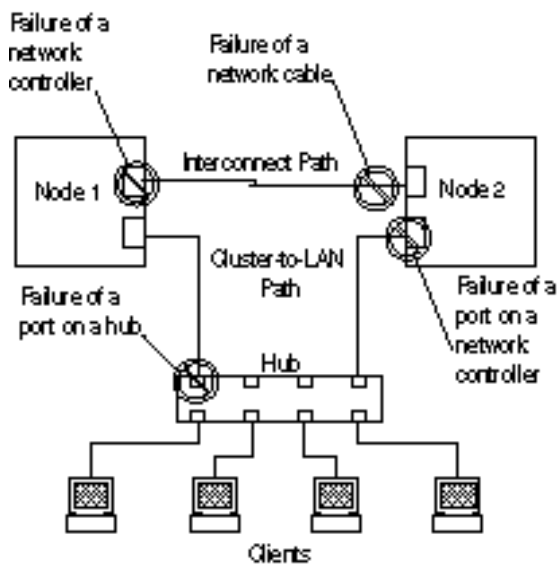


Figure 1: Communication Points of Failure

Building Blocks of Communication Path Redundancy

Now that you are able to identify the primary causes of failure in the communication paths, the next step is to understand what technologies are available to combat these points of failure. As you will see in the next section, an integration of hardware and software technologies provide the ability to create redundancy. This increases both the resiliency of cluster communications, and the overall availability of clustered applications and data.

Compaq Network Teaming and Configuration Fault Tolerant Features

The Compaq Network Teaming and Configuration Fault Tolerant feature consists of combining Compaq software with Compaq NICs. With this combination, two NICs, or a multi-port NIC, can be configured to be primary and backup paths for network communication; thus creating a redundant pair of network controllers. This feature is enabled with the Compaq Network Teaming and Configuration Utility, which can be found on the Compaq Support Software Diskette for Windows NT (NT SSD) or the Compaq Support Paq for Windows 2000 (NTCSP). Following is a sample screen from the utility.

Note: The Compaq Support Software for Windows NT (NT SSD) and the Compaq Support Paq for Windows 2000 (NTCSP) is located on the Compaq SmartStart CD. It can also be found at www.compaq.com.

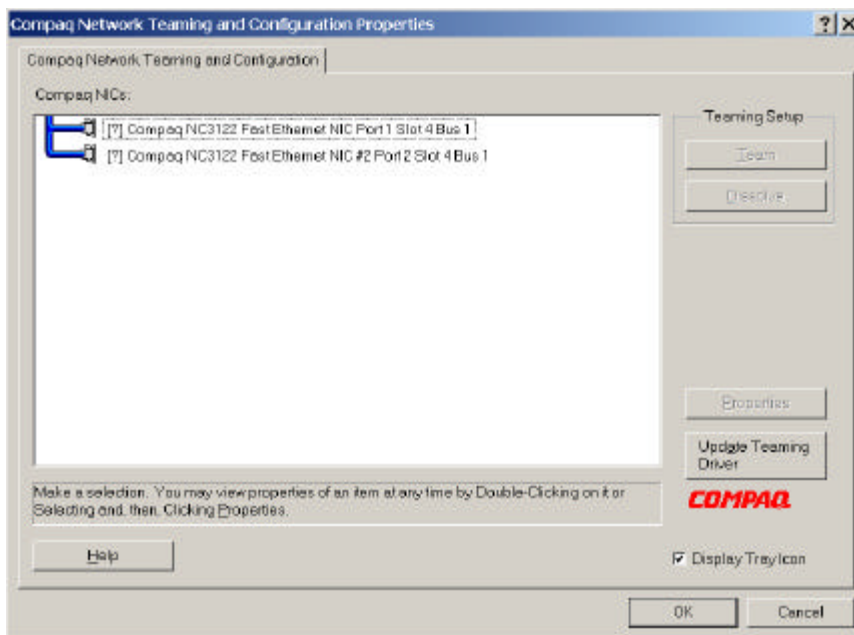


Figure 2: Illustration of a Dual NIC, no Teaming configured

Note: A quick summary of this feature can be found in Appendix A of this paper. For a complete description of the product, refer to the white paper entitled “*Compaq Advanced Network Error Correction Support in a Microsoft Windows NT Server Environment*”.

Redundant Network Interface Controllers (NIC)

To provide a maximum level of redundancy, customers can use Compaq NIC Teaming capabilities for selected Compaq network products to provide a redundant client network connection. This allows the use of two NICs in each server, one acting as an online spare for the other. If one of the NICs fails, the backup NIC takes over the IP address and functionality of the failed NIC. This feature is also tightly integrated with Compaq Insight Manager, providing proactive notification when the primary NIC fails. This configuration, when coupled with a dedicated interconnect for cluster communications, provides redundant paths for both client and cluster communications.

NIC redundancy is accomplished with the Compaq Network Teaming and Configuration Utility. This utility is available for use with the following Compaq NICs:

- Compaq 10/100 Fast Ethernet (Compaq NCxxxx)
- Compaq Netelligent

By using the NIC Teaming capabilities of Compaq NICs, an additional Compaq NIC can be added to a PCI slot to create a redundant pair. This redundant pair consists of the two NICs in the PCI slots, and is used for the client LAN connection. Both NICs in the pair must be connected to the same Ethernet hub. NIC Teaming redundant pairs should not be used for dedicated intra-cluster heartbeat connections.

Dual-Port Network Interface Controller (NIC)

Most NICs have a single-port with which a single network cable connects. Ordinarily, if two distinct network communication paths are needed from a single server, two NICs are placed in the server, and two expansion bus slots are used. A dual-port NIC, however, has two ports, each of which supports its own network connection. Only one expansion bus slot is used. Compaq offers a complete line of dual-ported NIC found at <http://www.compaq.com/products/servers/networking/index.html>.

In the previous section, it was noted that redundant NICs could be configured with two separate network controllers. An exciting feature of the Compaq Network Teaming and Configuration Utility is that it can be used to configure two ports of a dual-port NIC to be redundant. In this configuration, one of the NIC ports is configured as a hot backup for the other. The primary port will operate normally, sending and receiving data. Meanwhile, the second port remains in a standby state until the primary port encounters a failure. When the primary port encounters a failure, the standby port will take over. Therefore, no interruption of data flow is encountered.

Furthermore, the Compaq Network Teaming and Configuration and Correction feature can be employed with any two NICs, regardless of whether the NICs are single-port, dual-port, or a combination. For example, assume a dual-port NIC and a single-port NIC reside in Node 1. A redundant NIC configuration can be made using one of the ports on the dual-port NIC as the active port, and the port on the single-port NIC as the standby port. This applies only to NICs in the same family, Compaq 10/100 Fast Ethernet or Compaq Netelligent NICs. You cannot, for example, use a single port Netelligent NIC and team it with a Dual port Compaq 10/100 Fast Ethernet card.

Interconnect Paths

Building redundancy into your cluster communication paths requires knowledge of interconnect paths. Two types of interconnect paths exist. A *private interconnect* (also known as a dedicated interconnect) is used exclusively for intra-cluster communication. A *public interconnect* (also known as a shared interconnect) not only takes care of communication between the cluster nodes, it handles cluster-to-LAN communication.

Use of a private interconnect precludes heavy cluster-to-LAN network traffic from diminishing the flow of important intra-cluster communication. Additionally, a private interconnect is easy to set up, maintain, and monitor.

There are two methods of physically creating a private interconnect. The first directly connects the network controllers in each cluster node using a crossover cable. A network hub is not required since the crossover cable plugs directly into each controller and enables communication to occur between the NICs. The crossover cable appears to be a standard Ethernet cable, but it is not. The internal wiring of the cable differs from a standard Ethernet cable. You should consider labeling the crossover cable to distinguish it from a standard network cable. One crossover cable is included with each Compaq ProLiant Cluster kit.

The second physical interconnect utilizes a network hub, or even a series of hubs, repeaters, and switches. If the cluster nodes will be more than several meters apart, you will need to use this method. As long as both interconnect controllers reside on the same IP network, intra-cluster communication will occur over any combination of networking devices.

A public interconnect is not recommended as the primary path for intra-cluster communication, because cluster-to-LAN traffic can be heavy at times, and may interfere with node-to-node traffic. Still, it is recommended that a public interconnect be configured for intra-cluster communication (as a redundant path) while a dedicated interconnect is created to serve as the primary path.

Note: In the case of a cluster with greater than two nodes, a network hub is required for the intra-cluster communication.

PCI Hot Plug

Another important building block in communication path redundancy is PCI Hot Plug technology. PCI Hot Plug is an industry standard technology, which offers customers greater availability by eliminating both planned and unplanned downtime. This technology allows customers to remove and replace PCI expansion boards without having to power down the server or suspend any processes. PCI Hot Plug support allows users to replace a failed board with an identical board.

Compaq has implemented PCI Hot Plug so that each PCI bus slot can be controlled individually. This provides greater flexibility and availability, as service in adjacent slots is not effected if power is removed from one or more of the slots. In addition, greater availability is achieved through redundant failover capabilities available for certain Compaq PCI expansion boards. This function provides auto failover capabilities that allow a standby board to assume the workload, even while the failed board is being replaced.

PCI Hot Plug, when used in conjunction with redundant network interface controllers, brings even greater availability to your cluster communications. For example, assume you have two

Compaq 10/100 Fast Ethernet controllers placed in slots that support PCI Hot Plug. The two controllers have been configured with the Compaq Network Teaming and Configuration Utility to operate as redundant controllers. The primary controller encounters a failure, and network operation switches over to the standby (redundant) controller. At this point the primary controller can be physically removed from the cluster node without interrupting any operation of the cluster node. A new controller, the same model as the removed one, is placed in the open slot. Once the installation is complete, the newly installed controller becomes the standby for the currently active network controller. You are now back to a redundant configuration without having to power down the cluster node or disrupt client connections.

Microsoft Cluster Server allows the administrator to configure any certified network controller for intra-cluster communication, cluster-to-LAN communication, or both. Employing redundancy for the interconnect requires that at least two network controllers be configured, via Microsoft Cluster Server, for intra-cluster communication.

The following picture is of the main screen Cluster Administrator. Notice how the *New Cluster Network* item under the *Networks* folder is highlighted. The user has right clicked on the selection to bring up the small menu to the right of the selection.

Note: PCI Hot Plug is not supported on all Compaq servers; be sure to check the product specifications for your ProLiant server to see if PCI Hot Plug is available.

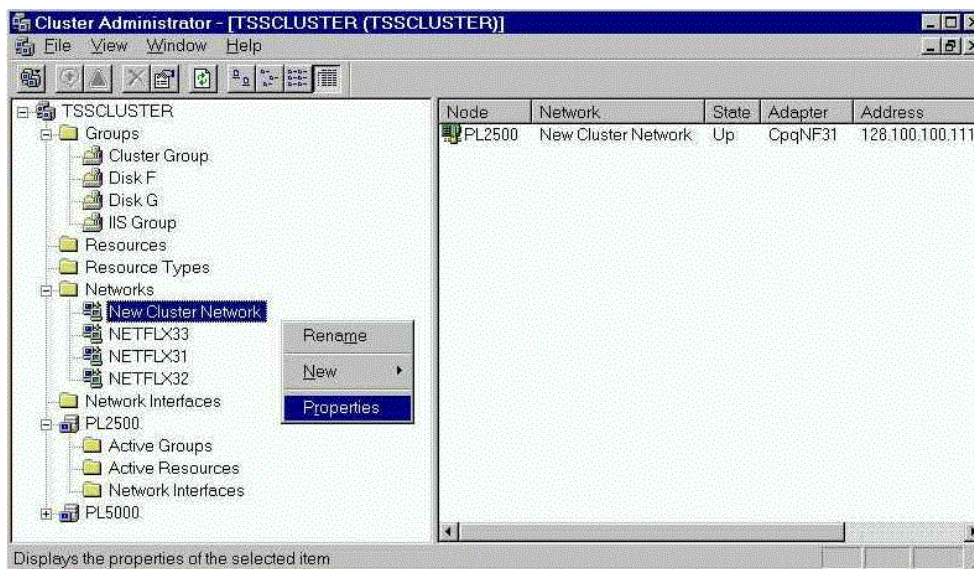


Figure 3: Screen Shot of The Cluster Administrator's Network Controller Configuration Screen

Clicking on *Properties* brings up the following picture of Cluster Administrator's Properties screen for network controllers. This is where controllers are configured for cluster-to-LAN use, intra-cluster communication, or for both.

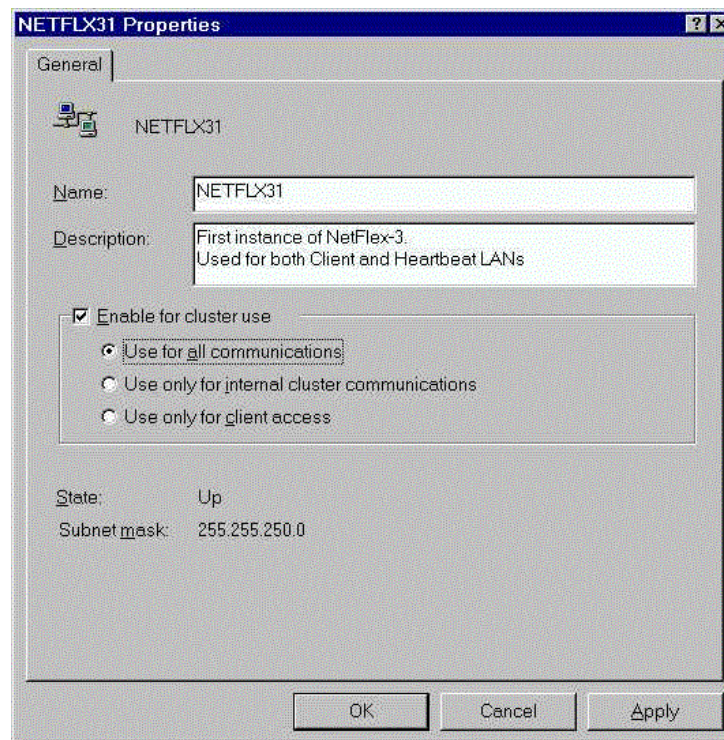


Figure 4: The Cluster Administrator Properties Screen for Network Controllers

Elements of a successful Cluster failover in Network Communications:

Compaq Hardware and Software

Figure 5 is a basic cluster configuration using single port NICs for the public and private network connections without redundancy.

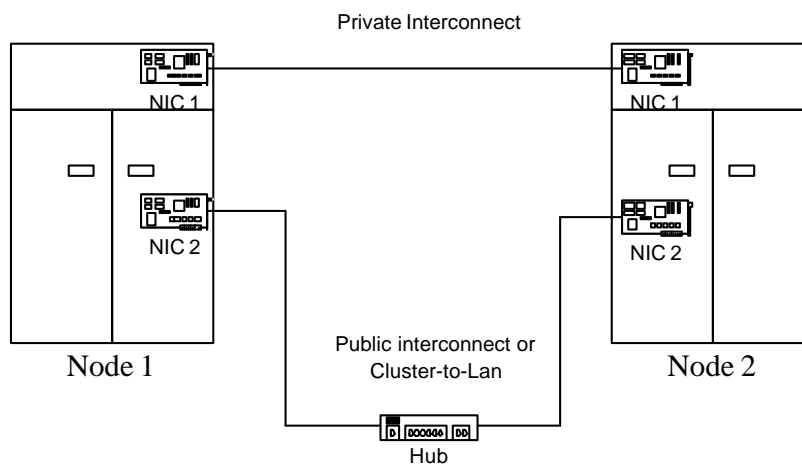


Figure 5. Basic Cluster Configuration

It is possible to replace the standard Compaq NIC, which has a single Ethernet port, with a dual port Compaq Ethernet NIC. In figure 6, the intra-cluster heartbeat is moved from the single-port NIC to one of the ports on the dual port NIC. Then a NIC Teaming redundant pair is created using the other available port on the dual port NIC and the single port NIC. This redundant pair is used for both the client LAN connection and the intra-cluster heartbeat. By configuring the heartbeat communication to use both the dedicated interconnect and the client network, a backup path is provided for intra-cluster communications. This method provides redundancy for the client LAN and intra-cluster connections. Using redundant Compaq NICs for the client network also protects against NIC failure. In either case, cluster communications continue to operate uninterrupted if a NIC fails.

Note: A NIC Teaming redundant pair should not be created across the two ports of a dual port NIC, as this will not provide port redundancy in the case of an entire NIC failure.

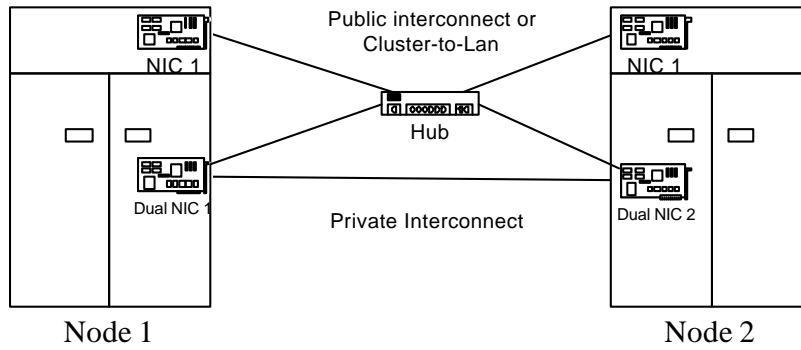


Figure 6. Redundant NIC Configuration

Figure 7 shows the elimination of any single point of failure. In this configuration, NIC 2 and NIC 3 are connected to separate network hubs, in order to eliminate the network hub as a single point of failure.

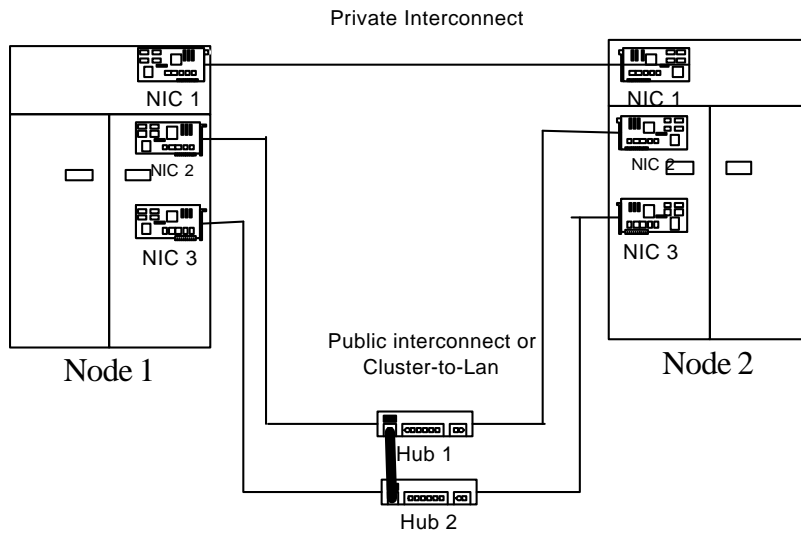


Figure 7. No Single Point of Failure Configuration

Microsoft Cluster Service

Before you begin the Configuration of Microsoft Cluster Service (MSCS) you need to know the dos and don'ts. It is highly recommended that you first read the Microsoft Cluster Administrator Guide. You should also consult the readme files included with MSCS and consult the Web for known configuration problems and special situations.

An important consideration for redundant NIC teams, is the Microsoft operating system that is being used. Microsoft Windows 2000 Advanced Server and Windows 2000 Datacenter Server provides a NIC monitoring resource as part of the Cluster Service. This means that if the client LAN NIC fails, the Cluster Service will recognize this and fail over the applications and services to the other node. Microsoft Windows NT Server 4.0, Enterprise Edition does not have this NIC monitoring feature as part of the cluster software. In the case of a failure, the cluster software will not fail over any applications or services, and clients will not be able to access them. Under Windows NT Server 4.0, Enterprise Edition, it is highly recommended that a redundant NIC team be used for the client LAN connection. Although Windows 2000 Advanced Server provides a NIC monitoring feature, it will still cause a disruption of client services when the failover occurs. For maximum availability, a redundant NIC team is recommended.

The following is a summary of do's and don'ts extracted from three Knowledge Base articles:

Articles

- Q254101 Network Adapter Teaming and Server Clustering
- Q258750 Recommended Private "Heartbeat" Configuration on a Cluster Server
- Q259267 Microsoft Cluster Service Installation Resources

Do's

- Form a second private interconnect if you can.
- Use teaming on the public network for maximum failover protection
- Set your adapter to a specific speed 10 MB/Sec or 100 MB/Sec for the private interconnect
- Set your Duplex mode to Half Duplex for the private interconnect
- Remove all unnecessary network traffic from the network adapter that is set to Internal Cluster communications only (this adapter is also known as the heartbeat or private network adapter)
- Remove NetBIOS from the private interconnect
- Set the proper Cluster communication priority order.
- Set the proper adapter binding order.
- Define the proper network adapter speed and mode.
- Configure TCP/IP correctly.

- Disable the Media Sense feature (in Windows 2000 only).

Don'ts

- Use teaming on the private interconnect of a server cluster.
- Use the “Auto-Detect” setting on your Network Adapter for the private interconnect.

Other Information

All Compaq ProLiant Servers ship with a NIC. In some servers, the NIC is integrated onto the motherboard of the computer. In other cases, the NIC is a separate controller placed in one of the expansion bus slots.

When configuring your cluster's network communications, it is recommended that, whenever feasible, you utilize the NIC shipped with the server. The reason for this recommendation is simply that you use one less expansion bus slots than if you placed another NIC in the server.

If you will be using a configuration that requires the Compaq Advanced Fault Detection and Correction feature, check to see if your integrated NIC supports it. If it does you can use the integrated NIC as part of a redundant pair. If it does not, you should consider using the NIC as part of a dedicated interconnect or dedicated cluster-to-LAN network path.

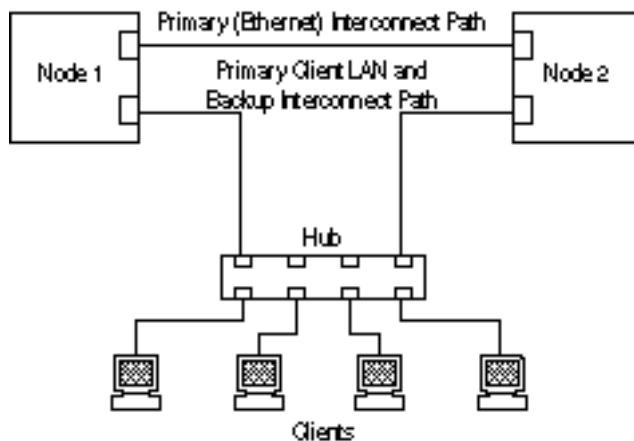


Figure 8: Dedicated Ethernet Interconnect

Compaq ServerNet II

The Compaq ServerNet II controller is a bi-directional, high-bandwidth, low-latency interconnect. A Compaq ServerNet II controller can function as a gigabit Ethernet controller, which can be used for intra-cluster communication. Due to Compaq ServerNet II cable specifications, the ServerNet II controller cannot be physically connected to an Ethernet hub. The

ServerNet II controller is not supported by the Compaq Network Teaming and Configuration utility.

Note: For additional Compaq ServerNet II information, please refer to www.compaq.com/highavailability

Summary

This paper has addressed the importance of network communication fault tolerance. Network communication mechanisms have been defined. The Compaq Network Teaming and Configuration Utility was used to create a redundant Network Interface Controller (NIC) pair, providing network high availability.

Appendix A

The Compaq Network Teaming and Configuration Feature

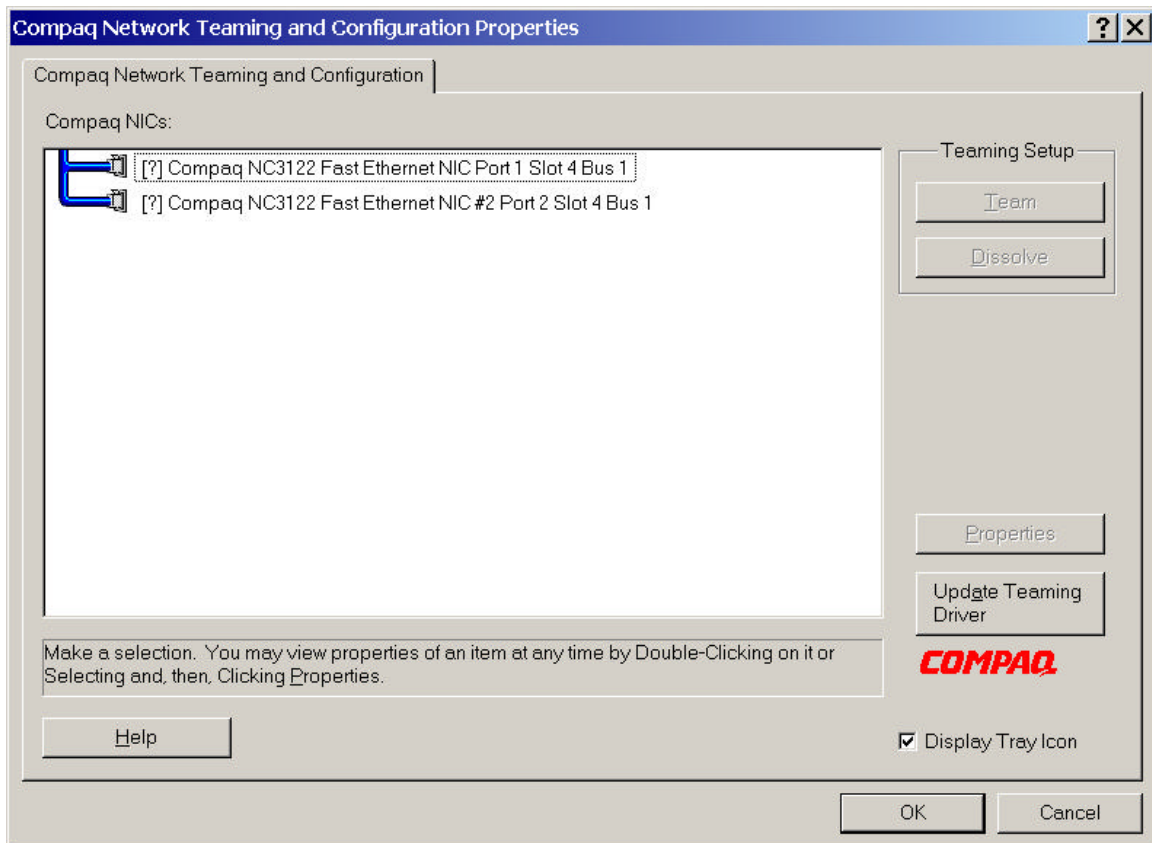


Figure 9: Compaq Network Teaming and Configuration main page

Team Configuration Screen

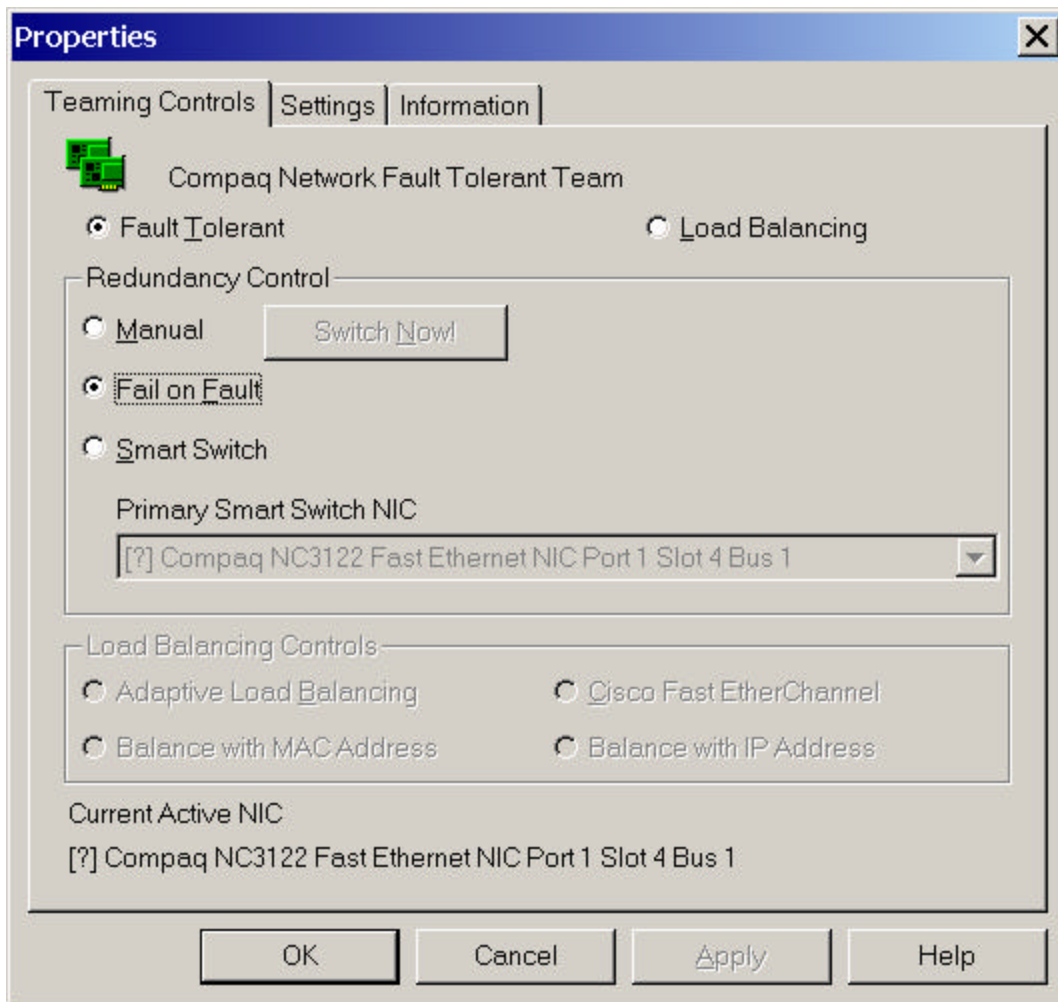


Figure 10: Compaq Network Teaming and Configuration Properties page

The Compaq Fast Ethernet or Gigabit Server NICs provide several options for increasing throughput and fault tolerance when running Windows NT 4.0, Windows 2000, or NetWare 4.1x or newer:

- Fault Tolerant - provides automatic redundancy for your NIC. If the primary NIC fails, the secondary takes over.
- Load Balancing - creates a team of NICs to increase transmission throughput. Also includes NFT. Works with any 100Base-TX or Gigabit switch. ALB works with IP only.

Note: The load balancing feature is not currently recommended for use in a clustered environment.

General Configuration Notes

- Install Windows NT 4.0 Service Pack 4 or later prior to configuring NIC Teaming.
- Windows NT versions prior to 4.0 do not support NIC Teaming options.
- The Compaq Network Teaming and Configuration Utility is supported on Windows 2000; Service Pack 1 for Windows 2000 is not required.
- NICs that are teamed must reside on the same IP network.