Paper Number: 2004

Architecting Highly Available Networking Environments

Pamela Williams Dickerman

Advanced Technology Consultant

Michael Hayward

Hewlett-Packard Company 19111 Pruneridge Avenue Cupertino, CA 95014

Abstract

Networking has many hardware components; each could be a Single Point of Failure (SPOF). The key to eliminating failures within the network are understanding the topologies being used, understanding the failure points within those topologies, and removing these failure points from the network.

There are hardware and software products such as dual attached FDDI cards and HP's MC/ServiceGuard which provide increased network availability.

* What type of hardware redundancy is required to protect against SPOFs?
* What type of software is needed to work with the hardware?
* How do these products fit together?

This paper discusses how SPOFs in the network can be eliminated or how the fault can be segmented to such an extent that only a small percentage of users will be affected during an outage.

Single Points of Failure in Network Topologies

The key to eliminating communication downtime is to identify and eliminate the SPOFs within the network. There are many different types of network topologies, each with their own SPOFs. This paper will discuss three main topology types:

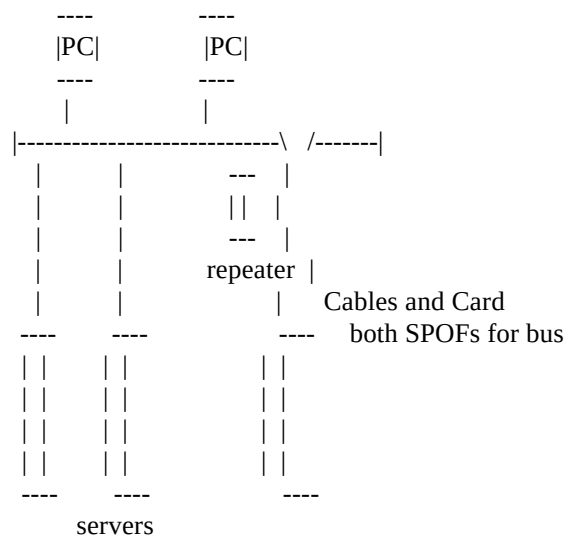
* Bus
* Star
* Ring

Bus Topologies

A bus type network is one where each system taps directly into the same

cable, has a matched impedance, has only two ends, and all signals are terminated at the cable ends. The most common bus topology uses ThickLAN and ThinLAN cabling. Both use coaxial type cable, commonly shortened to "coax". With bus topologies, a single segment can be extended by using a repeater. A repeater takes a signal from one segment and boosts the signal onto the next segment.

Figure 1: Bus Topology:



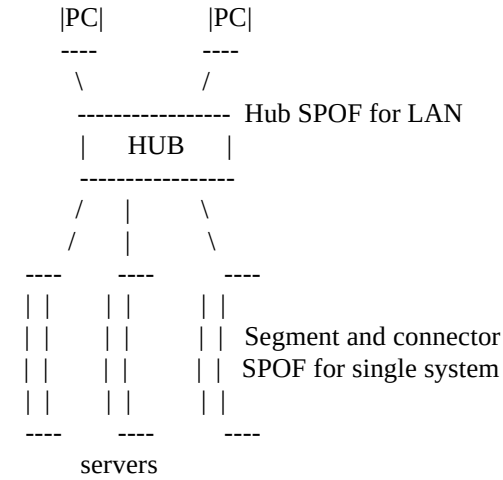
A SPOF with a bus topology is the cable. If the cable fails, no system connected to it can communicate on the network. With bus topologies, it can be difficult to isolate the exact location of the problem. A fault could be caused by a bad connector or cable break. Often with coax cables, finding the location of the fault is done with a Time Domain Reflectometer, indicator lights on the network interface cards, or tracing the cable segments.

The repeater in the diagram improves availability by segmenting cable, impedance and termination faults. If the repeater goes down, one LAN segment will be unable to communicate to other LAN segment. In addition, a LAN controller card failure isolates only one system from the LAN, unless the system is also a gateway to another LAN.

Star Topologies

A star topology is a point-to-point cabling scheme usually connecting a system to a network hub or concentrator with no connection in between. The two most common star topologies are twisted-pair cables connecting to hubs, and optical fiber connecting to optical concentrators.

Figure 2: Star Topology: Using Twisted-Pair



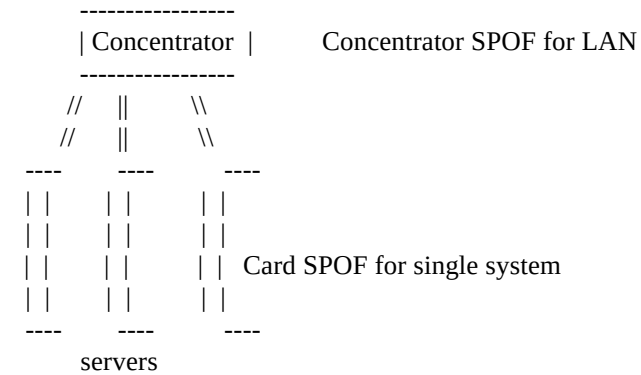
With Twisted-Pair, the SPOF is changed from the cable to the hub. If a single segment from a system to the hub or if the LAN card on that system fails, only that system is unable to use the network.

Optical fiber based networks use a concentrator, or a switch, rather than a hub. The two most common protocols running over optical fiber are Fiber Data Distributed Interchange (FDDI) and Fibre Channel (FC) with FDDI currently the largest percentage. This concentrator remains the SPOF for the LAN.

Hewlett-Packard (HP) makes two types of FDDI controller cards: Dual-Attach Stations (DAS) and Single-Attach Stations (SAS). SAS FDDI is a point-to-point connection between the system and the concentrator. Therefore, SAS configurations have the same SPOFs as the twisted-pair. DAS can be either star topology or ring topology. DAS ring topology is discussed in the next section.

DAS FDDI eliminates the segment as a SPOF. When there are no failures, only the first segment is being used. When a failure of a segment occurs, FDDI is able to transparently switch to the second segment and continue LAN operations.

Figure 3: Star Topology: Dual-Attached FDDI with a concentrator



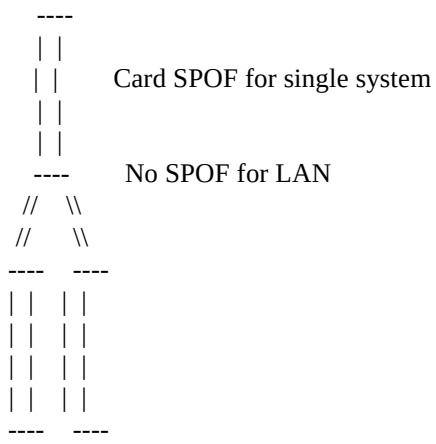
In Figure 3 we can see there are still two SPOFs with DAS FDDI in a ring topology.

* The concentrator.
* The DAS FDDI controller card on each system.

Ring Topologies

A ring topology is one where each system connects directly to the next system and the next system connects to the next and so on forming a ring.

Figure 4: Ring Topology: DAS FDDI without concentrators



In Figure 3 we can see there are still two SPOFs with DAS FDDI in a ring

The two most common ring topologies are Token Ring and DAS FDDI. (SAS can form a ring if only two system are used. For three or more systems, SAS requires a concentrator or gateway.) This paper will focus on DAS FDDI.

Unlike the SAS configuration, there is no need for a concentrator except to connect to SAS based systems. In addition, DAS FDDI has a self-healing capability. If there is a segment failure, DAS can wrap the failure and send the data in the other direction. Therefore, as shown in Figure 4, there is no SPOF that would bring down the entire LAN. However, like SAS, the DAS controller card on each system is a SPOF for that system to the LAN. DAS FDDI is most commonly used for server to server LAN connections.

Now that we have discussed some of the SPOFs in a network, the question is "how can we protect against these types of failures?". First let us list the two requirements we want to meet:

* The ability for the client and server to react then recover from a LAN fault.
* Provide redundant equipment and control software to manage the hardware to meet the first requirement.

The key is to ensure all SPOFs are protected and failover/recovery is automatic.

Client Recovery

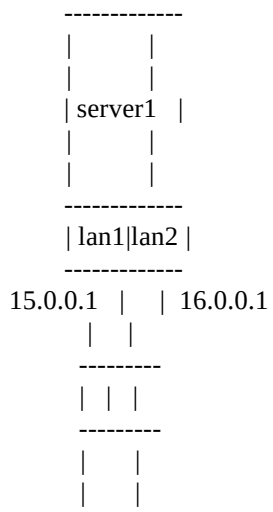
Client recovery during a network outage can involve a procedure as simple as the user reconnecting to a different address to something as advanced as automatic LAN recovery. First, manual and programmatic techniques will be discussed. Then, the benefits of automatic LAN recovery with MC/ServiceGuard will be discussed.

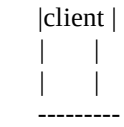
Multiple subnets and gateway protection

The client can react and recover from a network outage in a number of ways.

* Two separate subnets can be configured between the server and clients. Server and client applications can be programmed to switch communications to an alternate subnet IP address in the event the first IP address connection times out. * Another method is to configure the server and clients as gateways between the two subnets. This automatically protects against cable faults that occur between systems, but not against controller card faults.

Figure 5: Separate Subnet Example





In Figure 5, both systems have been configured with two LAN controller cards and with two separate cables between Server1 and Client1 to create two subnets. Each LAN card has a different IP address. If the cable connecting to lan1 fails, this would bring down access to the server's IP address 15.0.0.1. To recover, the client could connect to Server1 via the 16.0.0.1 address through lan2.

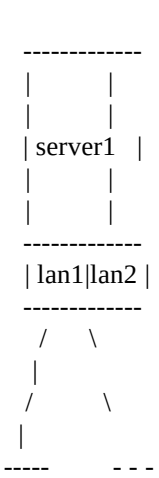
The client which normally does a telnet 15.0.0.1 would simply do a telnet 16.0.0.1 after a failure on the "15" network. A smart client application program could automatically try reconnecting via the second IP address.

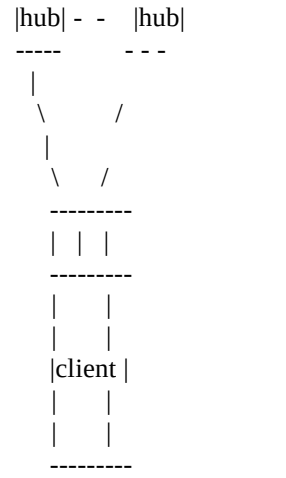
The tradeoff to the smart application approach is either making the subnet aware or using non-start applications where user connections are lost. Limitations to the gateway approach include no protection for LAN card failures and the configuration of multiple gateways can be extremely complex.

Local Failover Functionality of MC/ServiceGuard

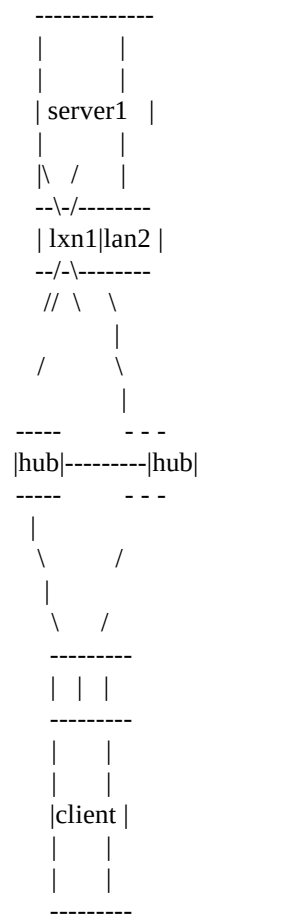
Products such as Hewlett-Packard's MC/ServiceGuard provide a different mechanism for automatically recovering from LAN failures. MC/ServiceGuard will transparently move the IP address from one LAN card on Server1 to the other standby LAN card. If the client is using a connection based protocol such as TCP, or even a connectionless protocol like UDP with a standard data transmission verification algorithm, the client applications and users will not know a LAN failure had occurred. MC/ServiceGuard can accomplish a local Ethernet recovery in less than 5 seconds and FDDI in less than 1 second.

Figure 6: MC/ServiceGuard Local Failover Example





After



Key Points

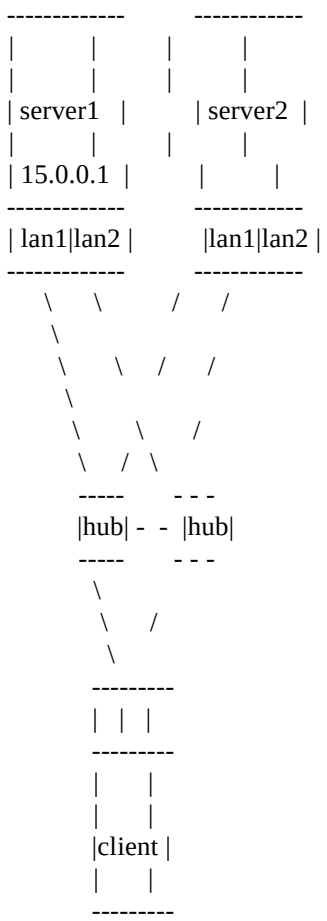
* Server and client are required to have MC/ServiceGuard installed to provide local switch functionality if both on same LAN. * Requires two HUBs cross connected for no SPOF
* Transparent to Applications

* Fast Failover * Requires idle LAN host adapter

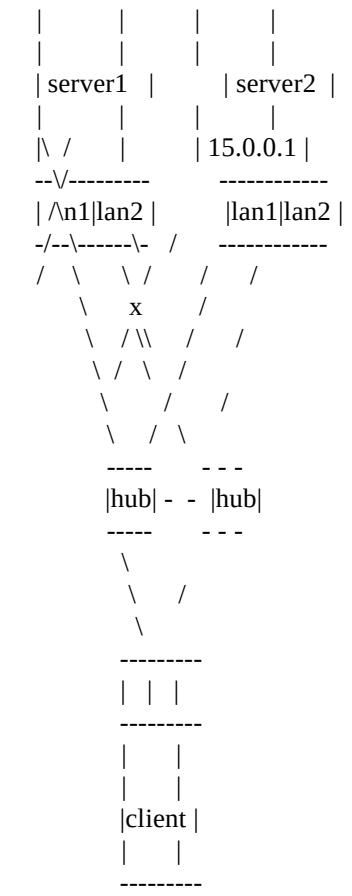
As will be discussed in the next section on LAN Topologies, the MC/ServiceGuard approach uses two hubs connected together to create one LAN with alternate paths. After a failure, the 15.0.0.1 address is transparently moved to the lan2 controller card. If the LAN is Ethernet, FDDI or TokenRing, MC/ServiceGuard issues a re-arp to broadcast the new IP to MAC level mapping out to all the clients. This causes the client to immediately know the new mapping and continue its connection. If the LAN is IEEE 802.3, this protocol uses proxy instead of arp protocol and has no method of distributing a change in IP to MAC level mapping; thus, requiring a 90 second timeout prior to the connection being disconnected. The client will then need to reconnect to the server.

Figure 7: MC/ServiceGuard Remote Failover Example

Before



After



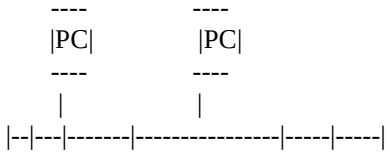
Another feature of MC/ServiceGuard is subnet failure detection with IP movement to an alternate system. This is useful if there are no alternate LAN cards available on the server, but there is an alternate server. As shown in Figure 7, the alternate server's LAN card becomes the redundant controller, but all connections are broken. Although the connections are broken, the clients continue to use the same IP address. They simply reconnect.

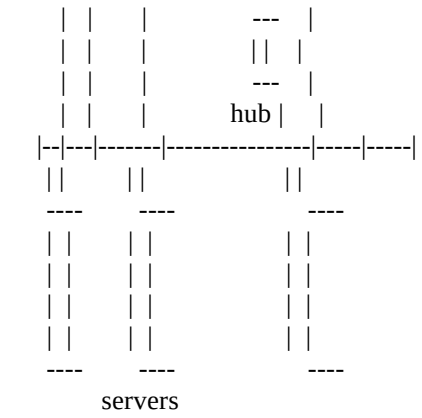
Eliminating SPOFs in LAN Topologies

In order for client reconnection and transparent failover to work, the client must have another hardware path to the server. This requires redundant equipment manage the hardware in the event of failure. This section will complete the picture by discussing the hardware necessary to create redundant LANs.

Highly Available Bus Topologies

Figure 8: Highly Available Bus Topology





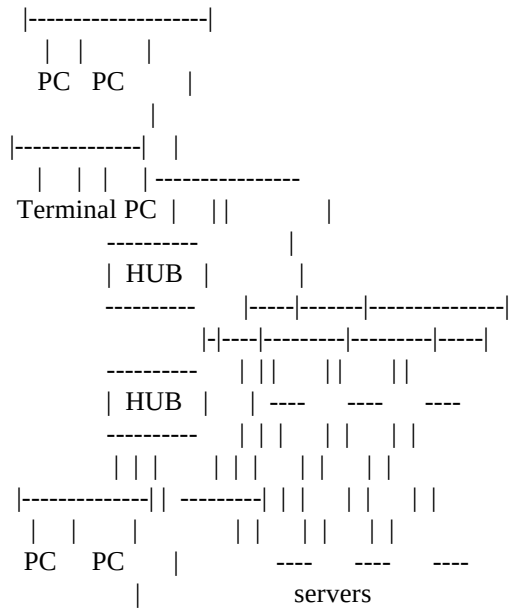
To eliminate the bus and interface cards as SPOFs, redundant busses are used. These two separate busses are linked together by a single hub.

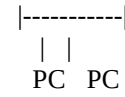
Since clients are usually a PC or workstation and MC/ServiceGuard does not run on these types of systems, we must use a segmentation scheme to isolate faults to the smallest possible area of impact.

In Figure 8 we can see the three servers have local LAN card switching functionality and will continue to communicate after a cable break. However, because each client is only connected to one cable, half the clients would lose communications with the servers with a cable break.

We can increase segmentation of the cables and decrease the area of impact caused by a cable break, by adding a second HUB and moving the clients onto multiple cable segments, as shown in Figure 9.

Figure 9: Highly Available Bus Topology using HUBS





In this example, there is still one virtual LAN with separate segments connected by hubs. A cable break from the PC to the LAN segment could cause a 25% loss of communications between the clients and the servers. If a HUB or the cable interconnecting one of the HUBs failed, 50% of the PCs would loss their connection. The highly user induced fault, client interconnect cable, has been decreased from causing 100% loss of communication with a single cable (as in Figure 1) to a 25% loss communication. The only way to remove the HUBs and interconnecting cable as a large percentage of communication failures, is to replace them with routers and use the "GateD" protocol. This will be discussed in the next section.

Figure 10: Highly Available Bus Topology with isolated clients using Routers

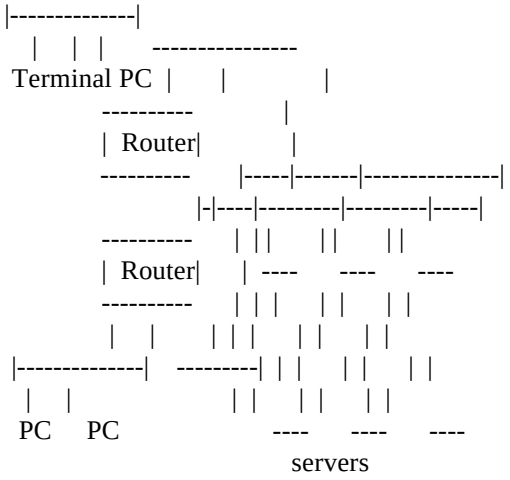
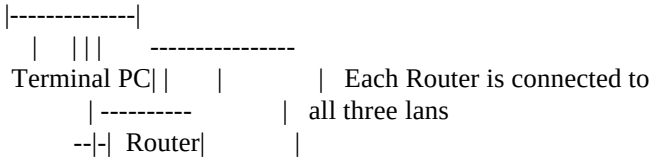
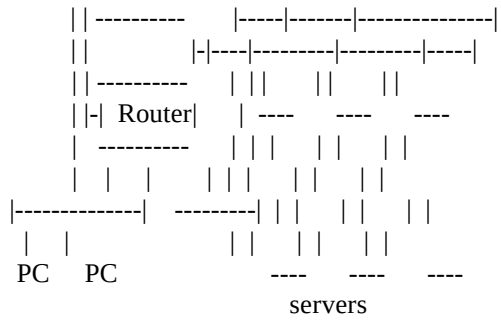


Figure 10 shows you how to isolate the clients on separate subnets via routers. Like the previous example, the local subnet for the servers is made highly available by using two cable segments and a hub. However, connection to the servers is made via the routers. The clients have a single point of failure at the router. If a router fails, the clients connected via that router will be unable to communicate with the server. The client's connections through the other router will still be up and communicating. By using routers, the clients have been isolated just as with the HUBs. A single router failure will still cause half of the clients to be disconnected from the servers.

Figure 11: Highly Available Bus Topology with Cross-Connected Routers





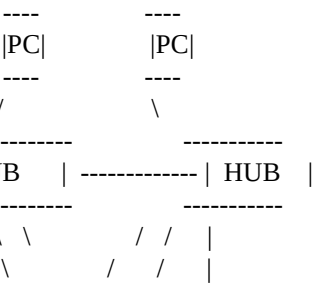
In Figure 11, each router is connected to each LAN. The client LANs have access to the servers through both routers. The advantage of using routers over HUBs becomes clear when we cross connect the routers and start using gated protocols to provide current subnet status. The routers will poll each other for current subnet access maps and broadcast this information to anyone listening. The clients and servers need to be configured in listen only mode to receive current route information from the routers. The clients and servers should not be using the IP broadcast address (for example, by configuring the default route as the local system) to route traffic to other subnets. Instead they should be using specific router IP addresses and gated-configured primary and alternate route paths. In this way the maximum throughput of the LAN is preserved (for example, no broadcast storms are being created) and errors such as duplicate IP addresses are avoided.

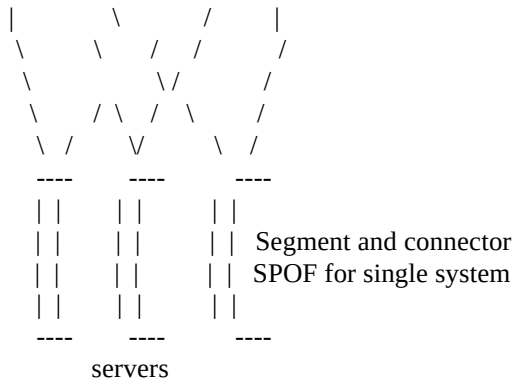
For customers who are not concerned about maximizing network throughput and do not want to take on the task of configuring gated on all their systems, the default route set to local will work as an alternative. Configure each local node as the default router and it will broadcast all IP packages destined for the other subnets.

Highly Available Star Topologies

In the Star topology section at the beginning of this paper, we learned the hub or the concentrator is a SPOF for the LAN. Implementing duplicate stars removes this SPOF. Just as discussed in the previous section on increased bus availability, a hub or concentrator failure will cause the biggest percentage of communication outage. The cable has been changed from a high impact point to a low impact point, because all cables are point to point instead of having multiple taps as in the bus configurations.

Figure 12: Highly Available Star Topology





In Figure 12, there are two ethernet cards on each system using twisted-pair cabling. The first ethernet card is connected to the one hub, and the standby ethernet card is connected to the alternate hub. The hubs are interconnected with one cable to make them one subnet. If a LAN card on a server fails, the second LAN card will take over the communications.

Routers can also be easily added to the star topology.
The same rules apply for this configuration as discussed in "Figure 11: Highly Available Bus Topology with Cross-Connected Routers". The cables can be replaced with token ring or optical fiber and the hubs replaced with concentrators. Everything would work the same as discussed for the twisted-pair bus topology.

Highly Available Ring Topologies

As indicated in the first section, DAS FDDI has a self-healing feature which added with the fact that there is no concentrator in a ring topology means there is no SPOF for the LAN itself. However, the DAS FDDI card on each system is a SPOF for that server.

One way to react to a failure of a DAS FDDI card is to implement the remote failover feature of MC/ServiceGuard. If a DAS card fails, the server application and its IP address are moved to the second server. The client must reconnect, but the outage duration will be minimal. For higher levels of availability, dual DAS FDDI rings could be implemented. However, the overhead and cost of this solution compared with the small increase in availability is a difficult tradeoff. Finally, a combination of a SAS star topology as a primary LAN with a DAS ring as a alternate LAN is another solution.

Further Considerations of Highly Available Networks

Eliminating the Single Points of Failure in a network and having client recovery mechanisms in place are only the first steps in creating highly available networks. Ongoing planning, clear documentation of the current networking environment, and written procedures for replacement of failed parts must be well understood.

It takes careful planning and investigation, but highly available networks can be setup and maintained.