WaveSwitch 100
Ethernet Switch
Configuration Guide

PLAIN TREE
SYSTEMS
INC

Part number: 11618

Document issue: Rev A

Date of issue: March 1995

Technical support and product information

Product support staff are available to answer technical questions between 8:00 a.m. and 8:00 p.m. EST, Monday to Friday.

Telephone:	1 800 831 1095 (toll free) 1 613 831 8883
Fax:	1 613 831 6120

You can also contact product support staff by sending an e-mail message on the Internet to:

```
techsup@plntree.isis.org
```

Bulletin board system (BBS)

The BBS contains software updates, test tools, product documentation, and other information you can download using the menus.

You can contact the Plaintree Systems BBS at (613) 831-8312. The BBS settings are 8 data bits, 1 stop bit, and no parity. The preferred file transfer protocol is ZMODEM. The BBS transfers data up to 14,400 baud.

The latest *WaveSwitch 100* user documentation is available from the BBS. Most documents are contained in .EXE compressed self-extracting files.

To download an item, select Main Menu/Files Menu/Download a File and choose a file. After downloading, enter *filename.EXE* on the command line to uncompress the contents into a PostScript file.

Product information

On the Internet

For up-to-date information about Plaintree products, see the Plaintree Systems Home Page on the World Wide Web at the following URL:

```
http://www.plaintree.com/plaintree
```

By telephone

Call for information about Plaintree Systems products.

In the United States:	1 800 370 2724 (toll free)
In Canada:	1 800 563 1178 (toll free)
In the U.K. and Europe:	+44 491 57 9600
In the rest of the world:	1 617 290 5800 1 617 239 7570 (Fax)

Contents

Figures

Introduction

The WaveSwitch 100 is a high performance switch for Ethernet LANs.

It transfers data packets between

- Ethernet LANs
- Ethernet LANs and higher speed LANs, such as FDDI or Fast Ethernet
- higher speed LANs

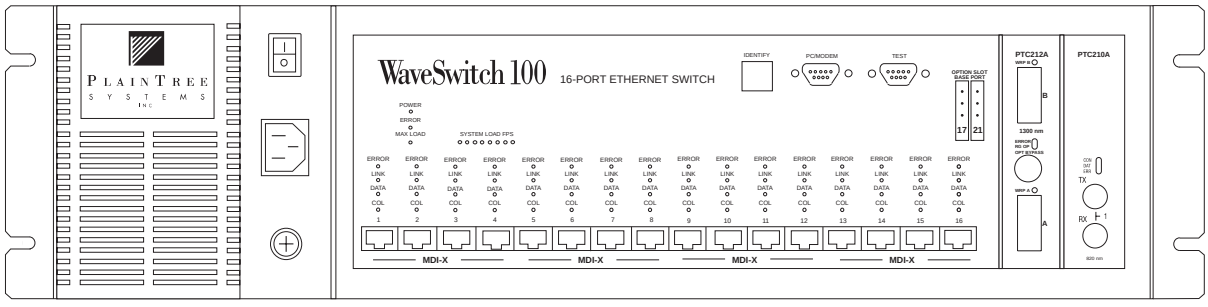
Adding a WaveSwitch 100 to an Ethernet LAN increases the bandwidth available to each workstation or fileserver, eliminating congestion or preventing the onset of congestion.

The WaveSwitch 100 increases LAN bandwidth by

- subdividing (microsegmenting) Ethernet networks to increase the bandwidth available to workstations
- supporting both FDDI and WaveBus Fast Ethernet (100 Mbps) for efficient access to high performance fileservers or connection to a campus backbone

The WaveSwitch 100 automatically learns the addresses of all stations on attached LANs, and directs packets accordingly. The manager of the network does not manage addresses in the WaveSwitch 100 unless special treatment of particular destination addresses is desired.

Figure 1
The WaveSwitch 100 Ethernet Switch



The WaveSwitch 100 supports

- full compatibility with the IEEE 802.1D transparent bridging standard, including the spanning tree procedure
- redundancy for mission-critical applications
- IP fragmentation for FDDI traffic destined for Ethernet networks

This guide describes how to connect a WaveSwitch 100 to enhance LAN performance. Because most installations require no parameters to be set within the WaveSwitch 100, this guide emphasizes network topology and interconnection with other LAN devices. It describes configurations suitable for small and large LANs, with an emphasis on mission-critical LANs.

General principles

The WaveSwitch 100 has the attributes of a switch: many ports and a large capacity for data traffic. It also has the attributes of a standard multiport transparent bridge.

The WaveSwitch 100 achieves configuration flexibility by incorporating optional interfaces for multiple high speed LAN types, and by conforming to standards for transparent bridges.

WaveSwitch 100 base units come with a fixed number of Ethernet ports—eight or sixteen 10Base-T ports for connection with unshielded twisted pair (UTP) copper cable, or twelve 10Base-FL ports for connection with fiber-optic cable. Each Ethernet port connects to a separate Ethernet network. (A separate Ethernet network is also called an Ether, or an Ethernet segment, or a collision domain).

> ***Note:*** The examples in this guide are based on the WaveSwitch 100 Ethernet Switch base unit with sixteen 10Base-T Ethernet ports.

Each base unit has two option slots that accept optional feature modules. The feature modules contain high speed ports for interconnection with FDDI or Fast Ethernet LANs.

Each Ethernet port of the WaveSwitch 100 filters up to 14,880 packets per second, for a total Ethernet filtering capacity for the whole WaveSwitch 100 of 238,000 Ethernet packets per second. Each FDDI or WaveBus Fast Ethernet port filters up to 150,000 packets per second. The total filtering capacity of 16 Ethernet ports and two high speed ports is 538,000 packets per second.

The *forwarding* rate of the WaveSwitch 100 is 150,000 packets per second for short packets, and 100 Mbps for long packets.

The WaveSwitch 100 translates packet formats while transferring data traffic between FDDI and Ethernet. The packet transformations required for the translation are done entirely by hardware during the reception of each packet from an FDDI LAN, or during transmission of each packet to an FDDI LAN. The transformations are those specified by IETF RFC 1042 and recommendation 802.1H of the Institute of Electrical and Electronics Engineers (IEEE), including the accommodation for AppleTalk.

The WaveSwitch 100 conforms to the transparent bridging standard, IEEE 802.1D, including the spanning tree procedure. The switch management protocol is SNMP. Switch management conforms to the Internet Engineering Task Force (IETF) bridge management standard, RFC 1493 (the bridge MIB for SNMP MIB II, RFC 1493 replaces RFC 1286).

Any configuration that is valid for a standard multiport transparent bridge is valid for the WaveSwitch 100. Network administrators can apply knowledge gained managing standard bridges to the WaveSwitch 100. Third party management applications that have embedded knowledge of the standard bridge MIB (RFC 1286/1493) can be used with the WaveSwitch 100. The standard MIB includes reading port and switch statistics, and specifying special filtering or routing for particular individual destination addresses. Additional management capability can be obtained by incorporating private MIB extensions into the management station.

Figures 2 and 3 show the general connection capabilities of the WaveSwitch 100. Note that each port connects to a separate network of multiple stations and fileservers.

The WaveSwitch 100 has several content addressable memories (CAMs) of 1024 entries each. The CAMS hold MAC addresses of stations on attached LANs. The multiple CAMs provide higher CAM-access bandwidth than one CAM. Most addresses stored in each CAM are stored in all CAMs. For this reason the total memory of the WaveSwitch 100 for different MAC addresses is at least 1024 addresses, and sometimes slightly exceeds 1024 addresses. Each CAM, when presented with a 1025th address, automatically replaces the oldest addresses in the address memory.

More than one thousand stations can be simultaneously active in all attached networks before packets leak onto LANs other than their destination LANs. The stations can be distributed anywhere in the networks shown. Several thousand stations can be present in all attached networks, as long as there is a very small probability that more than a thousand stations will be active within a period of a few seconds. Whether bridged networks this large are practical for particular situations will depend on the protocols used in the network.

Each Ethernet network in Figure 2 and 3 is connected to one of the 16 Ethernet (10Base-T) ports of the WaveSwitch 100.

Figure 2 shows two FDDI networks connected to the WaveSwitch 100. Each of the two option slots of the WaveSwitch 100 can support one single-attachment station (SAS) connection, or one dual-attachment station (DAS) connection to an FDDI network.

Figure 2
A network configuration using optional FDDI feature modules

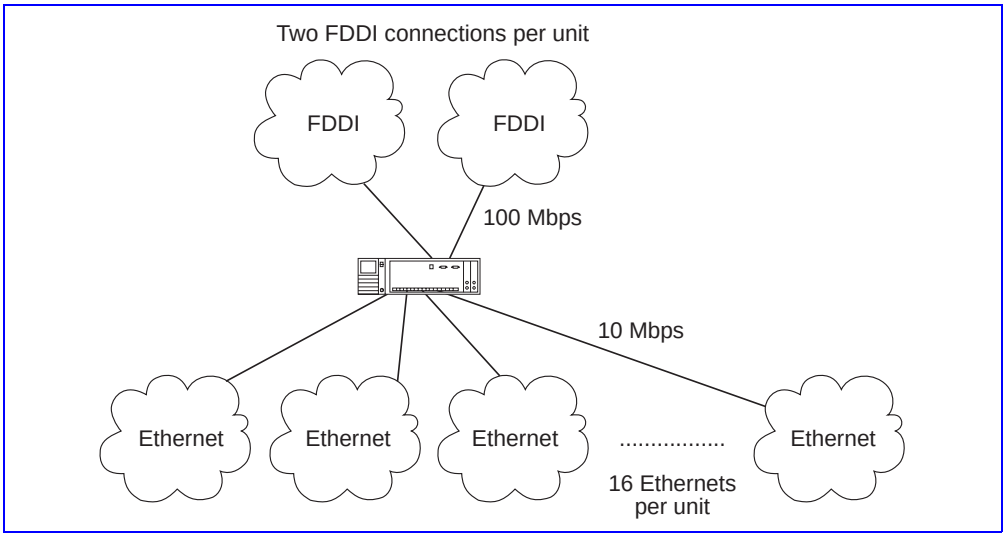


Figure 3 shows four WaveBus Fast Ethernet networks connected to the WaveSwitch 100. Each of the two option slots of the WaveSwitch 100 can support one or two WaveBus ports, for a maximum of four WaveBus connections. There are separate one- and two-port WaveBus feature modules.

Figure 3
A general configuration using optional WaveBus feature modules

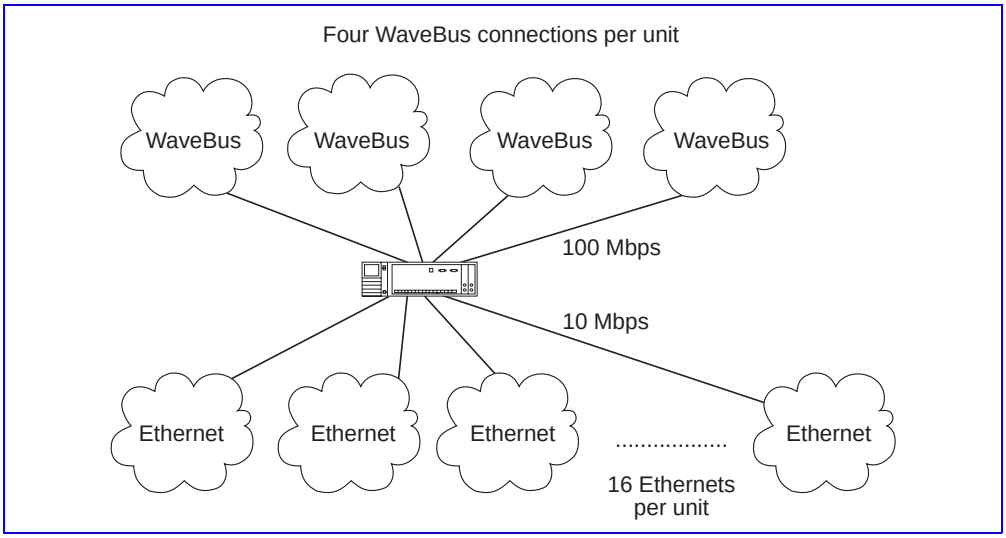
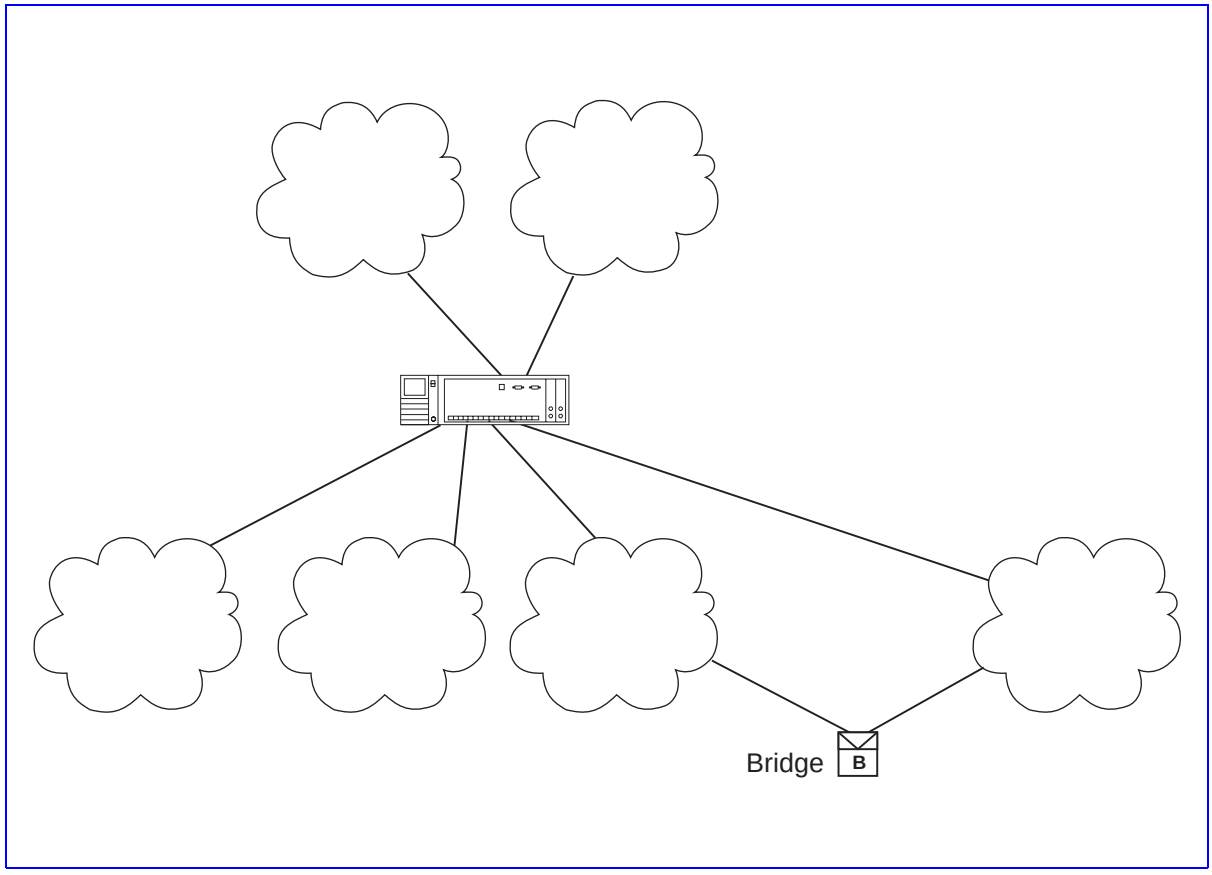


Figure 4 shows that networks connected to different ports of a WaveSwitch 100 can be connected by a bridge (it could be another WaveSwitch 100). The WaveSwitch 100 performs the standard spanning tree procedures specified by IEEE 802.1D to prevent traffic loops in the LAN, making such configurations possible. The WaveSwitch 100 makes use of this feature in redundant, or mission-critical, configurations.

Figure 4
A network configuration with a bridge between Ethernet ports of the WaveSwitch 100

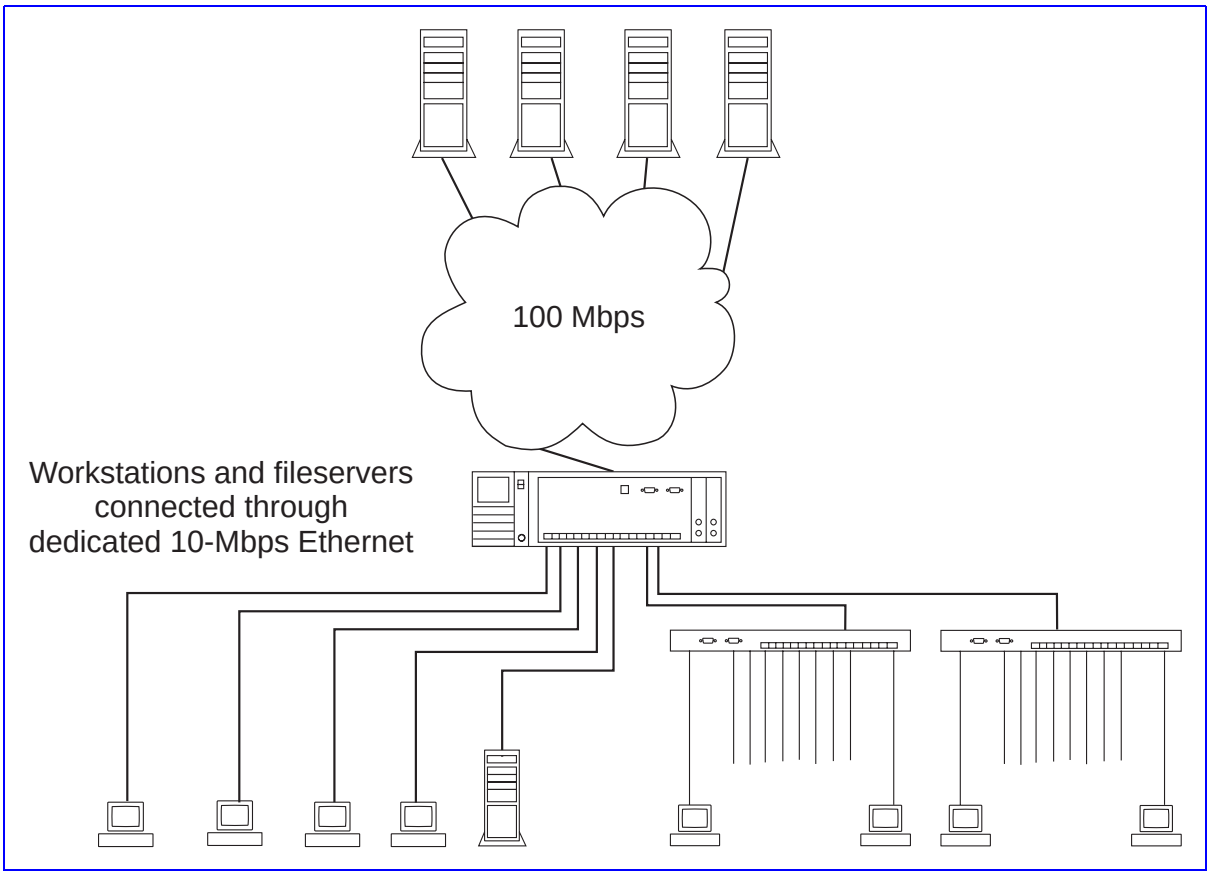


Dedicated (Private, Personal) Ethernet

Some workstations often exchange large files with a fileserver. Image-intensive work, such as computer-aided design, generates such traffic. Each workstation may need the full 10-Mbps capacity of an unshared Ethernet.

Figure 5 shows several workstations, each connected directly to an unshared 10Base-T port of the WaveSwitch 100. This arrangement is called dedicated Ethernet, private Ethernet, or personal Ethernet.

Figure 5
A network configuration using dedicated Ethernet for workstation connections



The Spanning Tree Procedure

The IEEE 802.1D spanning tree procedure provides support for mission-critical configurations. In mission-critical configurations, two ports of different WaveSwitch 100s—or two ports of the same WaveSwitch 100—are connected to a single LAN to create redundancy needed for fault tolerance. The spanning tree procedure prevents data traffic from circulating endlessly in the closed loops created by these connections.

A spanning tree is a subset of a physical network of bridges and LANs. Only the spanning tree portion of the network carries data traffic. The spanning tree *spans* the physical network—it connects all LANs and bridges. The spanning tree has the shape of a tree, with a root and dividing branches. The tree shape has no loops in which data traffic can circulate.

A bridge or port of the WaveSwitch 100 is in *forwarding state* if it is allowed by the spanning tree procedure to receive and transmit packets. The spanning tree procedure is performed simultaneously and cooperatively by all bridges to select a subset of bridge or switch ports that can enter or remain in forwarding state without causing traffic loops. The spanning tree procedure selects bridge ports to be in forwarding state in such a way that there is only one path a packet can follow between any pair of LANs. This single-path condition is incompatible with loops. A special case of loop elimination allows at most one port of each bridge to be in forwarding state for each LAN.

Figure 6 shows a physical network with a potential loop through two bridges. Using the spanning tree procedure, the bridges agree to prevent one of the bridge ports from forwarding traffic, as shown in Figure 7.

Figure 6
A physical extended LAN with two bridges

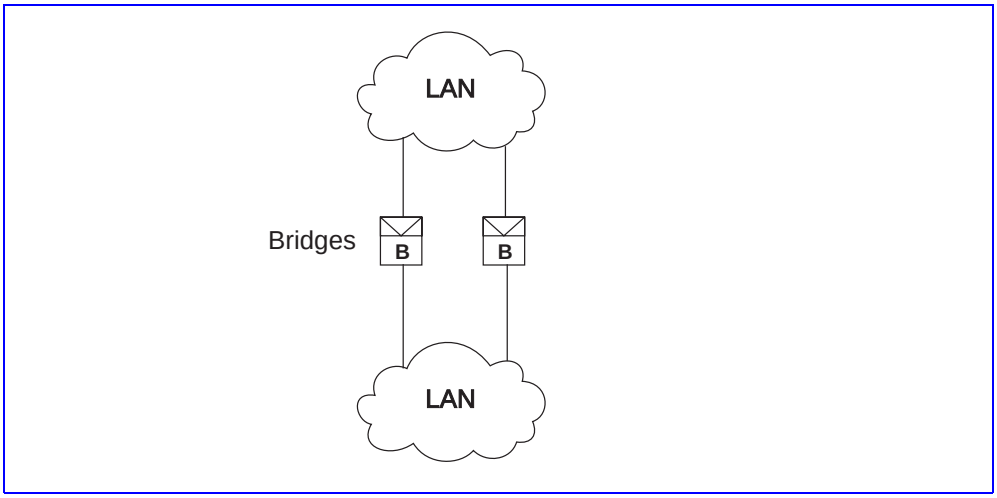
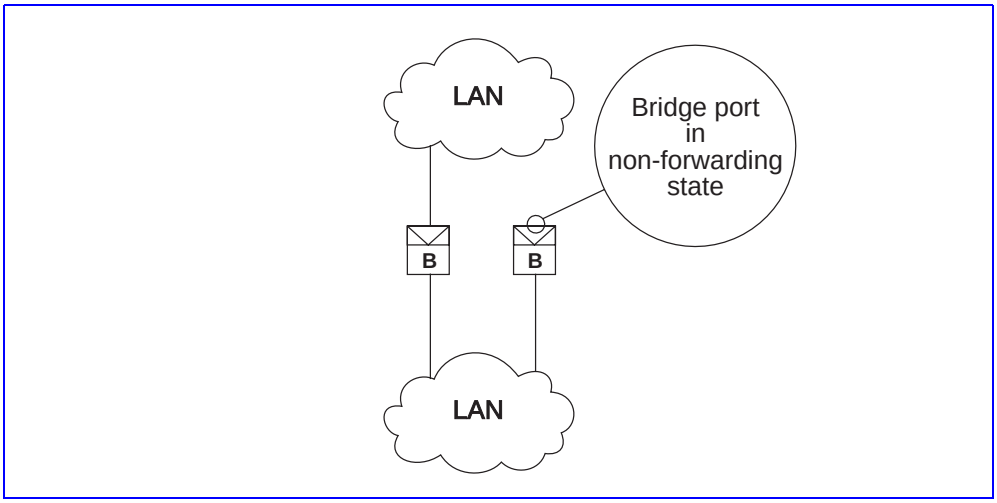


Figure 7
A spanning tree for the same LAN



A bridge can become idle if all but one of its ports are disabled by the spanning tree procedure (as in the example), but the bridge remains ready for service and continues to participate in the spanning tree procedure. The spanning tree procedure continually polls to detect changes in the network topology that might require a new set of forwarding bridge ports to be chosen.

Figure 8 shows a more complicated network with several possible physical loops.

Figure 8
A physical network

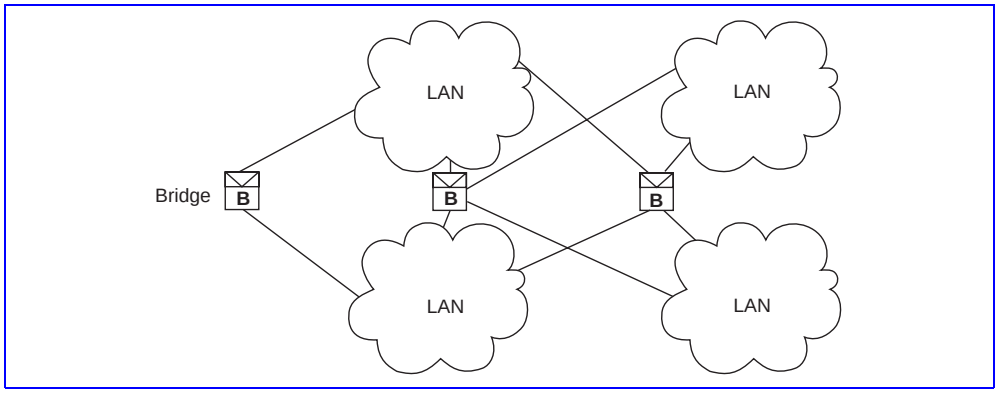


Figure 9 shows a possible spanning tree, with a fault about to occur in one of the media connecting a bridge to a LAN.

Figure 9
The first spanning tree is computed for the network

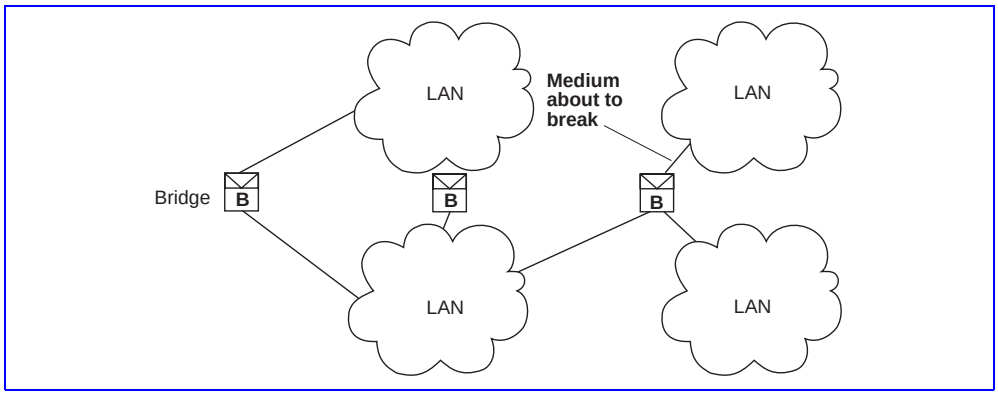
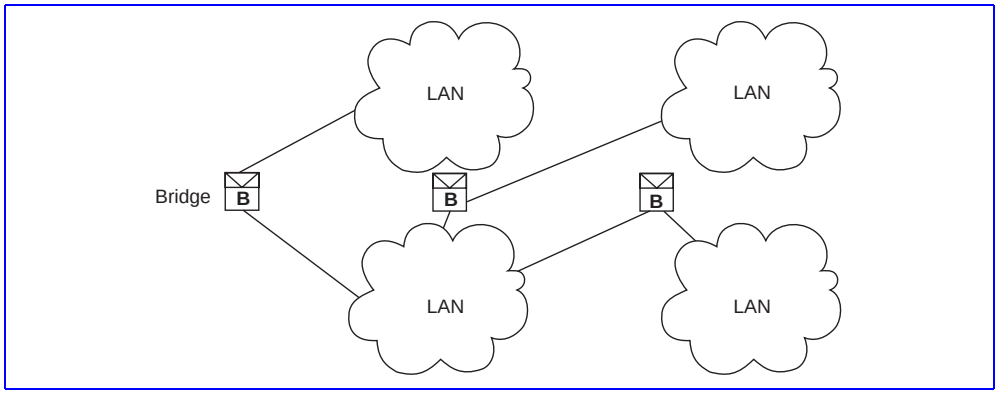


Figure 10 shows a different spanning tree computed after the fault.

Figure 10
A second spanning tree is computed after a fault occurs in the first spanning tree



The spanning tree procedure is optional if the physical connection of bridge ports has no loops — if there is only one physically possible path from each LAN to every other LAN. Such a LAN is vulnerable to the failure of any bridge or bridge port, which will partition the LAN into two or more pieces which cannot intercommunicate. Multiple possible paths between LANs, of which the spanning tree procedure chooses a single-path subset, are required to obtain tolerance of faults.

Even in deliberately non-redundant networks, the spanning tree procedure prevents disastrous traffic loops when bridges are connected accidentally to form physical loops.

The selection of the spanning tree is deterministic—the final spanning tree does not depend on the sequence in which bridges power up, or the sequence in which ports are enabled and disabled by management or by faults.

The spanning tree depends entirely on

1. the values of bridge and port identifiers
2. the root path cost increment of each bridge port
3. the set of bridge ports that are made available (enabled) by management (all ports are enabled by default)

Each bridge is known to the spanning tree procedure by its *bridge identifier*. A bridge identifier has two components, a 16-bit priority value, which can be set by management using SNMP, and a 48-bit value equal to the address of the bridge. The spanning tree procedure treats the two components as a single unsigned 64-bit integer value, with the priority field as the high order bits. The priority value defaults to the middle of the priority value range.

Each port is known to the spanning tree procedure by its *port identifier*. A port identifier has two components, an 8-bit priority value, which can be set by management using SNMP, and an 8-bit value specifying the unique and constant number assigned by the bridge to designate the port within the bridge. The spanning tree procedure treats the two components as a single unsigned 16-bit integer value, with the priority value as the high order bits. The priority component defaults to the middle of the priority value range.

The root path cost increment of a port may be set by management action using SNMP, but defaults to the path cost of the LAN attached to the port. The path cost of a LAN is 1000 divided by the speed of the LAN in Mbps. The spanning tree procedure selects paths that minimize the sum of the root path cost increments of ports along the path to each bridge from the root bridge. The path cost of an Ethernet LAN is 100. The path cost of an FDDI or WaveBus LAN is 10.

The standard spanning tree procedure is a fairly complicated distributed process performed simultaneously by all bridges and switches. The workings of the procedure can be found in references [1] and [2] on page 61. The result of the procedure is described here.

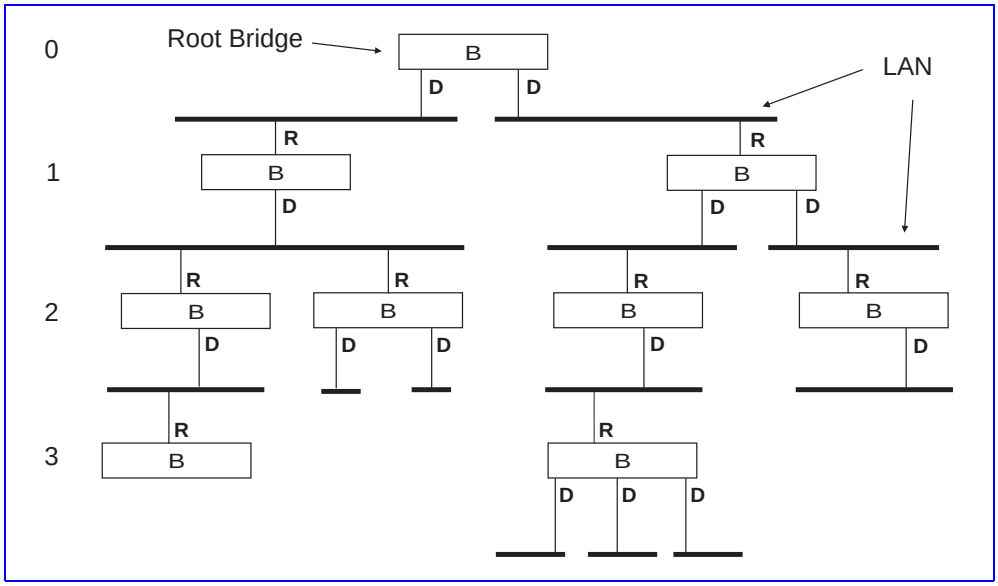
The bridge with the numerically lowest bridge identifier is the *root bridge* of the spanning tree. If more than one port of the root bridge is connected to the same LAN, the port with the numerically lowest identifier is chosen for forwarding state; the other ports are kept in a non-forwarding state. Each port of the root bridge in forwarding state is the *designated port* of the LAN to which it is connected.

The spanning tree can be drawn in layers, with the root bridge in the top layer. For reference, the layers are numbered according to their proximity to the root bridge. Figure 11 shows a spanning tree. The root bridge is the only bridge in layer 0. Each subsequent layer of bridges further from the root bridge has a layer number one higher than the preceding layer.

Each bridge in layer N, has a single port in forwarding state that connects it to the whole network formed by all the bridges in layers from 0 to N-1. This port is called the root port of the bridge. The ports labeled R in Figure 11 are the root ports for each bridge.

Each LAN between layers N-1 and N is connected to a single port in forwarding state that connects it to the whole network formed by all the bridges in layers 0 to N-1. That bridge is in layer N-1, and is called the *designated bridge* for the LAN. The forwarding port of the designated bridge to which the LAN is connected is called the *designated port* for the

Figure 11
A spanning tree structure



LAN. The ports labeled D in Figure 11 are the designated ports for each LAN.

Each bridge in the network has a *root path cost*. The root path cost of the root bridge is 0. The root path cost of another bridge can be computed by starting at the root bridge with root path cost = 0, and tracing a path through the spanning tree to the bridge. At each root port, add the root path cost increment of that root port. Each subsequent layer of bridges in the tree thus has a higher root path cost than the preceding layers.

In the spanning tree, the root path cost of each bridge is the minimum possible that can be obtained by selection from the physical network.

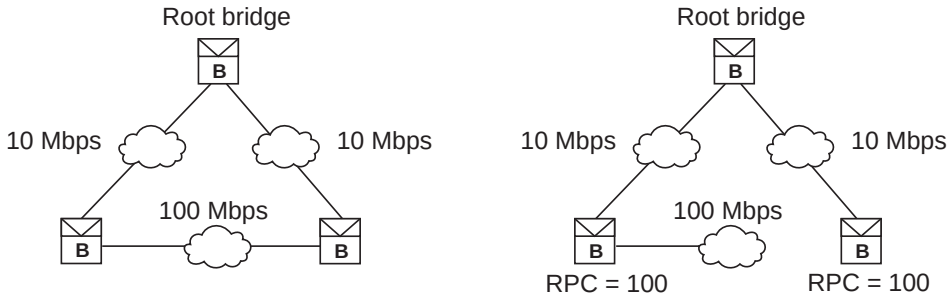
Controlling spanning tree selection

The spanning tree path from the root bridge to each other bridge is the shortest possible path between the two bridges; shortest means having the lowest possible value of root path cost. The spanning tree path between two non-root bridges is often *not* the shortest path, which sometimes produces a surprising result.

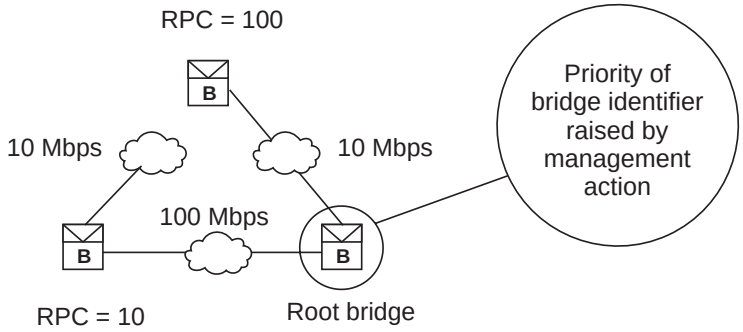
A network manager can cause the spanning tree procedure to select any spanning tree of the network by controlling bridge identifier priority values, port identifier priority values, and path cost increments. In most networks, the default spanning tree is acceptable, and management control of spanning tree selection is not required. In some networks, control of spanning tree selection is indispensable.

Figure 12 shows a physical network and its default spanning tree. Each bridge is labeled with its root path cost (RPC) in the spanning tree. It is unlikely that the network manager intended the spanning tree path between the two lower bridges to be through the two 10-Mbps LANs, as in the default spanning tree. It is far more likely that the 100-Mbps LAN was intended to be a spanning tree path. The network manager may be surprised by the consequences of this default spanning tree when one of the Ethernet LANs carries much more traffic than planned.

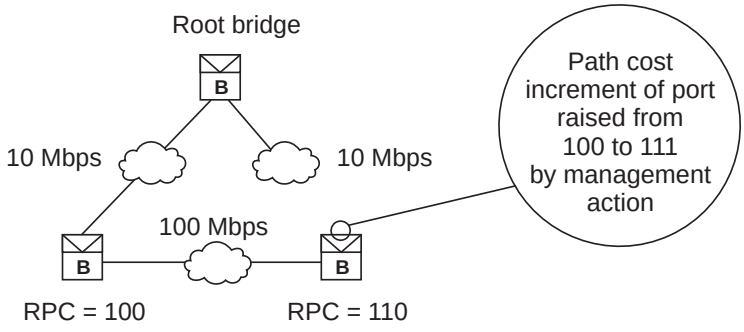
Figure 12
Obtaining non-default spanning tree selection



Physical network **Default spanning tree**



Spanning tree selection control by specifying the root bridge



Spanning tree selection control by raising the root path cost increment of a port

In this example, there are two ways the network manager can ensure that the 100-Mbps LAN forms a spanning tree link between the two bridges it connects directly:

1. by raising the priority (that is, lowering the value of the priority field) of one of the identifiers of the bridges connected to the 100-Mbps LAN to ensure that it becomes the root bridge
2. by raising the root path cost increment of one of the bridge ports above the default value. The default value is the path cost of the attached LAN, equal to 1000 divided by the speed of the LAN in Mbps.

Raising the root path cost increment of a port is often preferable because it can be applied locally and independently at several places in a network. There can be only one root bridge in a network. Moving the root reshapes the whole spanning tree, and may simply move or create spanning tree surprises.

Practical reasoning about spanning trees

You cannot determine which spanning tree of a network the spanning tree procedure will select by looking at the network. You must know the bridge identifiers, the port identifiers, and the root port cost of each bridge port, and you must simulate the spanning tree procedure. Usually, when planning or analyzing a network, you don't know the identifiers, and simulating the spanning tree procedure is too complicated. You usually know the root port costs, because they are defaulted to the LAN path cost.

The following observations will help you reason about spanning trees without going into extreme detail.

1. If there is a physical path between two LANs—a path consisting of a sequence of bridges and LANs—there will be a spanning tree path between the two LANs.
2. The spanning tree will have only one path between any two LANs. If it appears that more than one path is possible in the physical network, all but one of the paths will be blocked by bridge ports in non-forwarding states.
3. If a fault occurs that breaks the spanning tree, the bridges of the network will compute a new spanning tree in about 30 seconds and put it in force.

4. Faster LANs are given preference over slower LANs in the selection of spanning tree paths between the root bridge and other bridges. A 100-Mbps LAN has a much lower path cost than a 10-Mbps LAN. By default, as recommended by 802.1D, the WaveSwitch 100 gives FDDI and WaveBus LANs a root port path cost increment of 10, and gives Ethernet LANs a path cost of 100.

5. Even if two bridges are connected to the same LAN, there may be no single-LAN spanning tree path between them. (The spanning tree path may go through other bridges.) This will happen if each of the two bridges has a lower cost path to the root bridge than any path through the other bridge. See Figure 12 on page 14.

6. If several LANs of a physical network are each connected to two bridges, and only to those two bridges, one of the bridges will be the designated bridge for all the LANs. This means that each of the LANs, except perhaps one of them, will be connected to one forwarding bridge port only, and all these forwarding ports will be on the same bridge. If there is a LAN with a second forwarding port, that port is the root port of the other bridge, and that LAN is the spanning tree path between the two bridges.

7. If several LANs are connected to two bridges, and only to those two bridges, at most one of the LANs can form a spanning tree path between the two bridges. If the LANs have different speeds, the faster LANs are given preference in the selection of the spanning tree path. If all of the LANs have the same speed, and you want to give preference to one of them for spanning tree path selection, connect it to a port with a lower port identifier on each of the bridges than the ports to which the other LANs are connected. The lower port identifier can be created by using the SNMP bridge MIB to set the priority value of the port identifier of any port, or you can take advantage of the default case, in which ports with lower bridge port numbers have lower port identifiers, and connect the preferred LAN to a port with a lower number. Be aware that even when preference has been arranged, the spanning tree procedure may not choose *any* of the LANs directly connecting the two bridges as a spanning tree path. This will happen if the two bridges each have a lower cost path to the root than through the other bridge. See Figure 12 on page 14.

Networks with Ethernet fileserver connections

One of the main advantages of using the WaveSwitch 100 is the way it combines Ethernet for economical workstation connections and FDDI or WaveBus for fileserver connections. Some users may find, however, that 10-Mbps Ethernet connections to a WaveSwitch 100 are sufficient for their fileservers.

Two low-speed backbone configurations

The two typical network configurations shown in Figures 13 and 14 need the WaveSwitch 100. The figures show two low-speed backbone configurations commonly found in medium-size LANs (100-500 stations). Many of these LANs are outgrowing their capacity as they add new users and applications. Introducing the WaveSwitch 100 to these LANs greatly increases their performance.

Figure 13 shows a configuration that is common in Novell NetWare installations. Each workgroup has its own fileserver. Each of the fileservers does double duty as a router connecting the workgroup to a 10-Mbps Ethernet backbone. Traffic between the members of each workgroup and their fileserver is confined within the local workgroup LAN, while traffic between a member of a workgroup and a fileserver belonging to another workgroup appears on the backbone Ethernet LAN.

This configuration works well as long as access to the fileservers is dominated by members of each workgroup accessing their local fileserver. When users begin to make regular use of other fileservers, the limited backbone bandwidth becomes overloaded, and too much of the capacity of the fileservers is devoted to routing packets between LANs rather than serving files. The resulting delays frustrate users of the LAN.

Figure 13
Each fileserver doubles as a router connecting its workgroup to the backbone

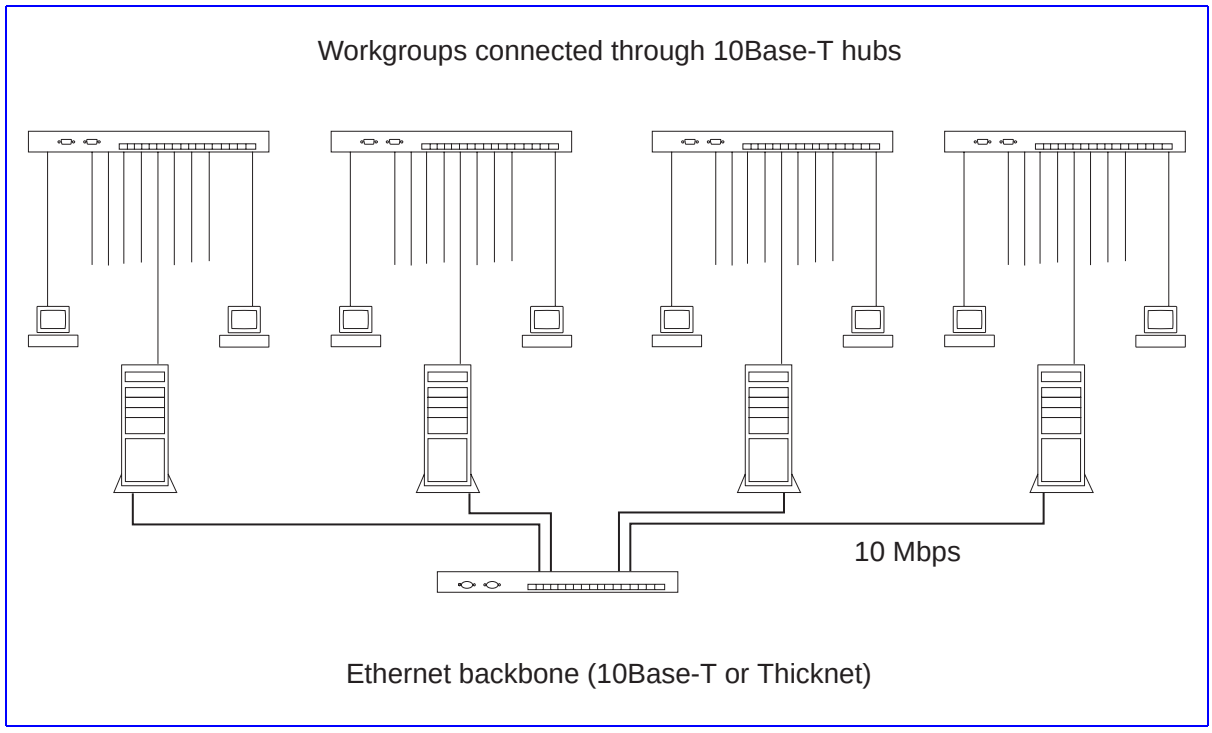
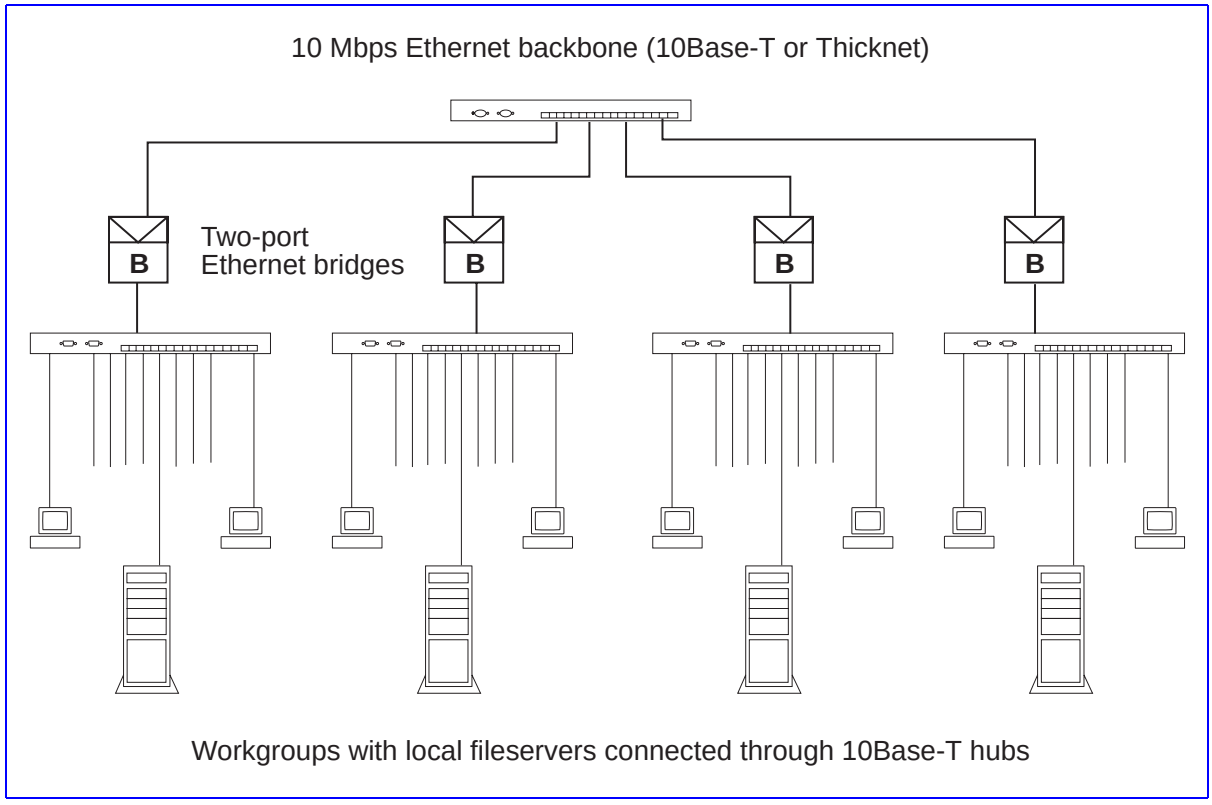


Figure 14 shows a configuration that overcomes one of these two performance limitations. Two-port bridges, rather than the fileservers, perform the function of dividing the LAN into workgroups. This configuration has the same properties as that of Figure 13, and requires the same general conditions to work well, but can carry a greater load, since the fileservers are devoted to serving files rather than routing packets.

These configurations are overloaded by backbone loads greater than 10 Mbps. The performance of the individual fileservers is limited by their 10-Mbps LAN connections.

Many network managers would like to manage fileservers at centralized locations, but find that the bandwidth limitations of configurations like these require the dispersion of the fileservers to the workgroup LANs to obtain sufficient LAN capacity for all workgroups.

Figure 14
Each bridge localizes workgroup traffic away from the low-speed backbone



Backbone replacement and fileserver centralization with the WaveSwitch 100

Figure 15 shows the simplest use of the WaveSwitch 100 to remove performance limitations of the configurations in Figures 13 and 14. The WaveSwitch 100 replaces the backbone network, providing a 100-Mbps collapsed backbone. The fileservers in Figure 13 are no longer interposed between each workgroup and the backbone. In these configurations, the network designer has decided that 10-Mbps access, either shared or dedicated, is sufficient for the fileservers; there is no use of FDDI, Fast Ethernet, or any other high-speed LAN, and the option slots of the WaveSwitch 100 are not occupied.

In the configuration of Figure 15, traffic resulting from access by each workgroup to its own fileserver still does not appear on the collapsed backbone (does not use the switching capacity of the WaveSwitch 100). This is less important than in the previous configurations because of the greater capacity of the WaveSwitch 100 as a backbone.

Figure 15
Straight backbone upgrade preserving workgroup traffic localization

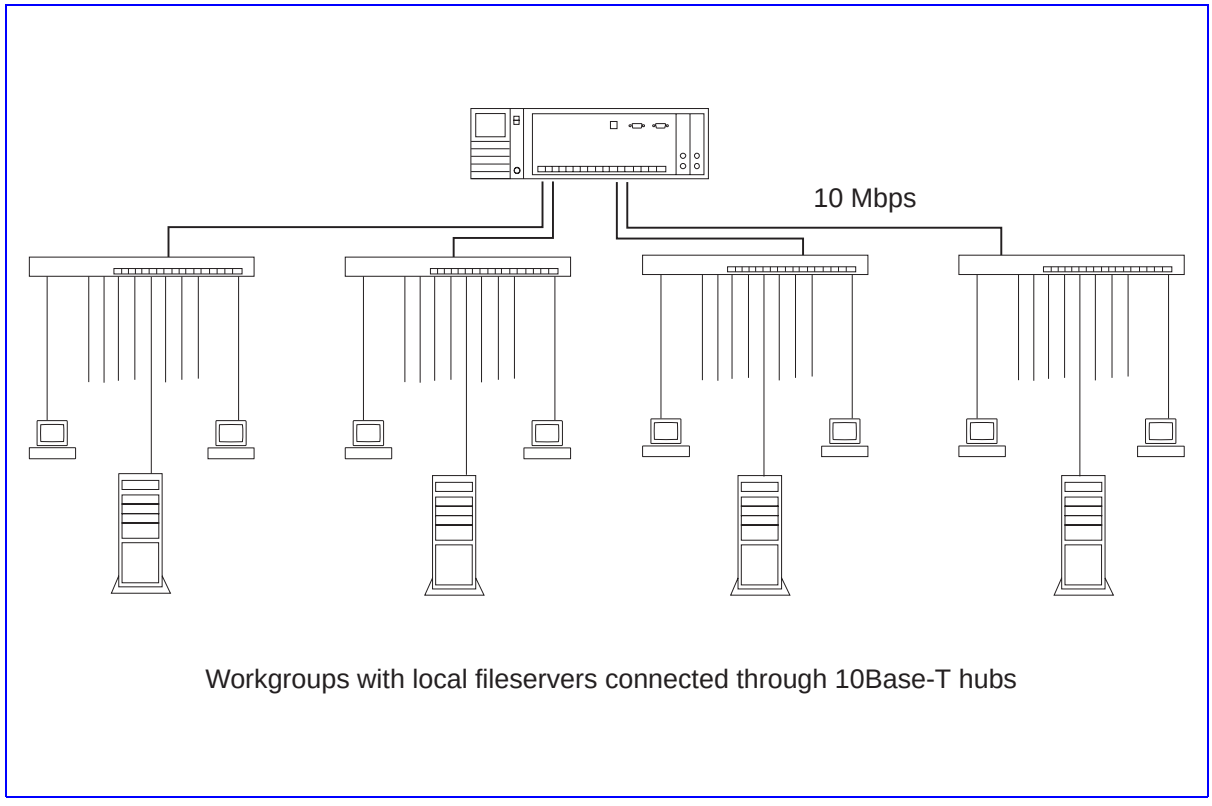
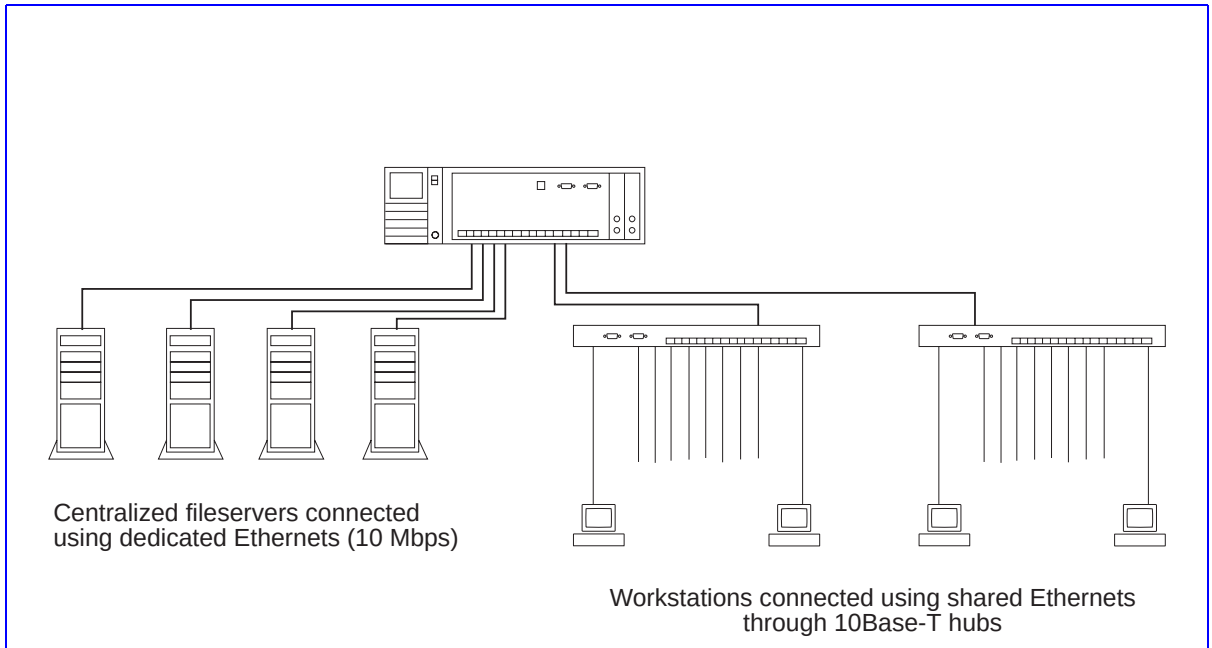


Figure 16 shows how a network can use the switching capacity of the WaveSwitch 100 to centrally manage the fileservers without using a 100-Mbps LAN.

In Figure 16, each fileserver has been given a dedicated 10Base-T port on the WaveSwitch 100. Direct connection of the fileservers to the WaveSwitch 100 keeps the fileservers in a single room for backup supervision and maintenance.

The configuration in Figure 16 may be easier to manage because of the centralized fileservers. Its performance can be better or worse than the configuration of Figure 15, depending on circumstances. If many workgroup LANs distribute their traffic uniformly over a smaller number of fileservers, the centralized configuration will provide greater throughput, because each fileserver is connected by a 10-Mbps LAN that carries only its traffic. If the traffic for each fileserver comes entirely, or almost entirely, from its own workgroup in the case of Figure 15, and from a single workgroup in the case of Figure 16, the file transfer rate of the configuration of Figure 16 will be lower that of Figure 15, because of the propagation delay (latency) through the switch between the workgroup and the fileserver.

Figure 16
Centralized fileservers



Latency can be reduced or eliminated, however, by the use of a windowed file transfer protocol such as Novell's packet burst version of the NetWare Core Protocol (NCP). A more general way of avoiding delay problems is to provide more access bandwidth to each fileserver, greatly improving the performance of the fileservers.

Access bandwidth for fileservers can be increased by upgrading to 100 Mbps for fileserver connections, or by providing multiple 10-Mbps connections for each fileserver, as described in the next section.

Novell NetWare version 3.12 version or higher is recommended for WaveSwitch 100 networks with Ethernet fileserver connections. Version 3.12 has a much more efficient implementation of the packet burst NCP than version 3.11 (the first version in which packet burst appeared).

Fileserver multiconnection

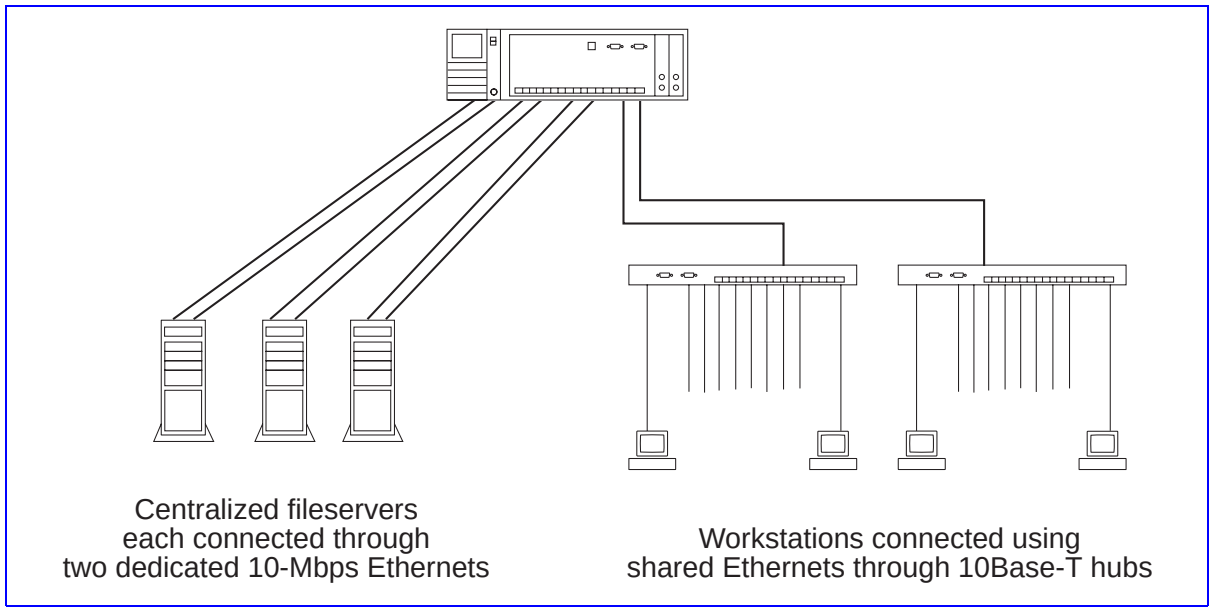
The fileservers in Figures 15 and 16 are performance-limited by their single 10-Mbps connections to the LAN. The performance of a fileserver can be improved by upgrading fileserver connections to 100 Mbps, or by providing multiple 10-Mbps Ethernet connections to each fileserver. A 100-Mbps LAN connection permits a high-performance fileserver to reach its full performance potential. Multiple 10-Mbps LAN connections to a single fileserver are considerably less expensive, however, and provide a throughput enhancement approximately proportional to the number of connections.

Figure 17 shows each fileserver connected to a WaveSwitch 100 by two 10-Mbps connections operating in load-sharing mode. These two connections require two Ethernet interface cards in each fileserver, and fileserver software that supports load-sharing multiconnections. The software makes two or more 10-Mbps connections appear to the operating system of the fileserver as a single higher bandwidth connection.

NSI (Network Specialists, Inc.) of Lyndhurst, N.J., offer fileserver multiconnection software called Balance for Novell NetWare version 3.11 and later.

The NSI fileserver multiconnection software also offers a redundancy option called Redundancy that redistributes the fileserver load over the connections that remain functioning after some connections fail.

Figure 17
Multiconnection improves fileserver performance



Mission-critical WaveSwitch 100 backbones

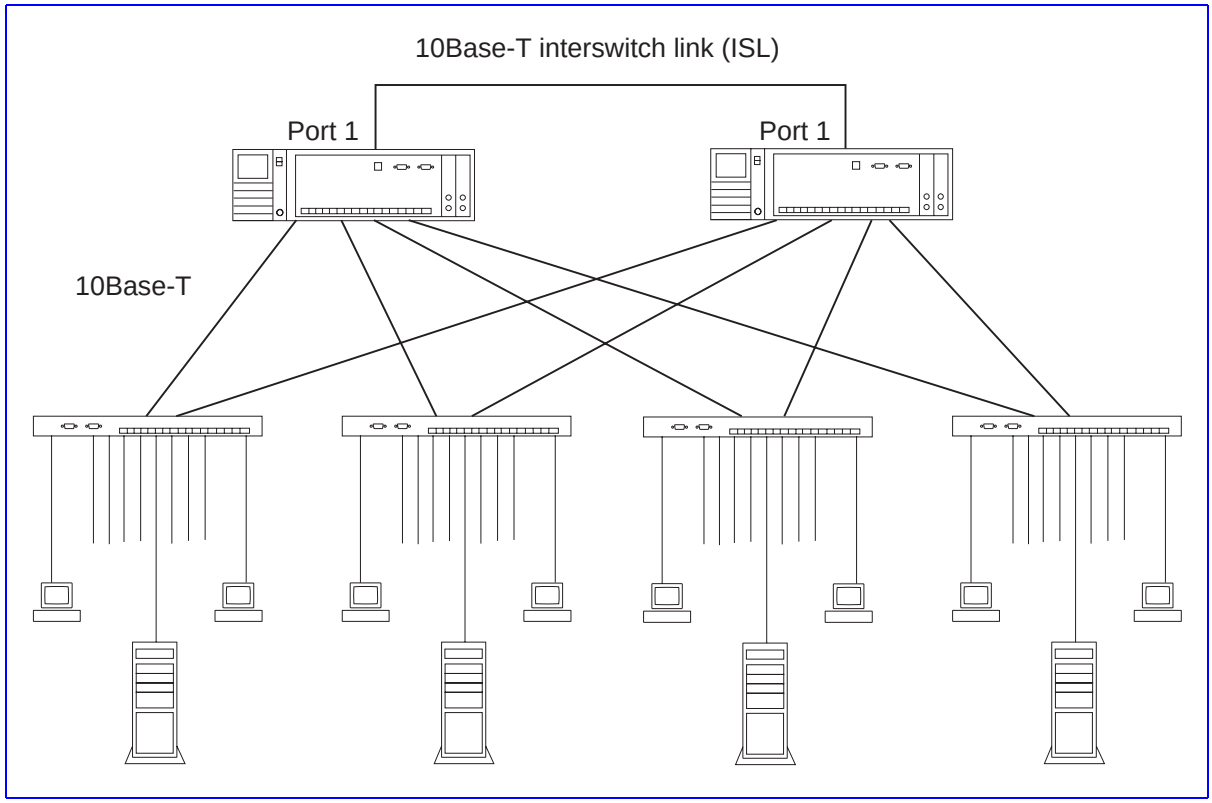
As an organization becomes more dependent on the reliable performance of its LANs, it becomes more involved in contingency planning for equipment failure.

One level of contingency planning allocates sufficient spare equipment to ensure that repairs can be performed quickly.

A further level of contingency planning designs the network so that it continues to perform its critical functions even when some parts of the network fail. Network configurations that continue to perform in the face of equipment failures—so-called mission-critical configurations—contain redundant equipment which can take over from failed equipment.

Figures 18 and 19 show how a redundant WaveSwitch 100 can be added to two configurations that were presented before. Figure 18 shows a dispersed fileserver configuration evolved from that of Figure 15; Figure 19 shows a centralized fileserver configuration evolved from that of Figure 16. These evolved designs assume that 10-Mbps connections, or dual load-sharing 10-Mbps connections, are sufficient for the required level of fileserver

Figure 18
Mated WaveSwitch 100s as a higher speed backbone replacement



performance. The dual fileserver connections in Figure 17 operate in load-sharing mode.

When two WaveSwitch 100s are interconnected like this to back each other up they are called mated. If one of the two mated WaveSwitch 100s fails, the other takes over completely, with a loss of service of about 30 seconds to recalculate the spanning tree.

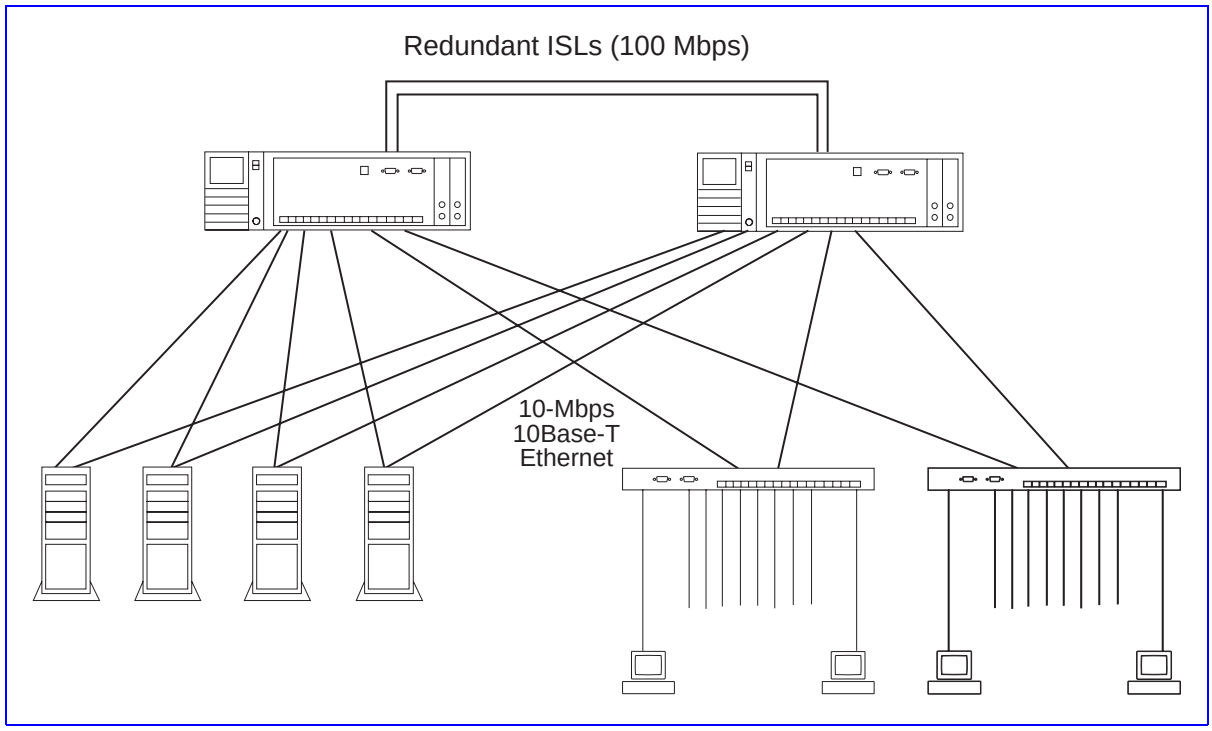
In both configurations, each 10Base-T concentrator is connected to both WaveSwitch 100s. The dual connection of the 10Base-T hubs creates multiple closed loops of LAN media. The existence of loops of media means that the spanning tree procedure allows only a subset of ports of the WaveSwitch 100 to enter the forwarding state—the state in which the port passes payload traffic between the WaveSwitch 100 and the LAN. The spanning tree procedure chooses a subset of ports that does not contain forwarding loops in which traffic would circulate indefinitely. The ports not permitted to enter forwarding state neither receive nor forward payload traffic.

The spanning tree procedure gives the ports of one WaveSwitch 100 complete preference over the ports of the other WaveSwitch 100 in the election of the forwarding ports. (One of the WaveSwitch 100s will be the designated bridge for all LANs connected to the two WaveSwitch 100s.) If all equipment is working correctly, the spanning tree preference means that one of the WaveSwitch 100s in Figure 18 carries no payload traffic. When a fault occurs—say someone unplugs one of the ports on the designated WaveSwitch 100—the spanning tree procedure computes a new set of forwarding ports, with the result that the other WaveSwitch 100 is enlisted to pass packets between the faulted LAN and all other LANs.

The 10-Mbps point-to-point interswitch link (ISL) that joins the mated WaveSwitch 100s in Figure 18 protects the workgroup LANs in the event of a fault. Without the ISL, the interLAN traffic to the faulted LAN would pass through one of the other workgroup LANs, multiplying its load, and perhaps creating network instability. Properly configured, the ISL carries the interLAN traffic to and from the faulted LAN. Because the link is used only in fault conditions, and will rarely carry more than the traffic of one workgroup LAN, a 10-Mbps link is sufficient in this configuration.

In Figure 19, the dual connections to each fileserver are distinct LANs, each of which is connected to a different WaveSwitch 100, and each of which must be included in the spanning tree. As a result of the load-sharing mode of operation of the dual fileserver connections, each WaveSwitch 100 carries about half the fileserver access traffic in normal operation. The dual connections protect each fileserver from failure of one WaveSwitch 100, and significantly enhance fileserver performance as long as both WaveSwitch 100s are functioning correctly. The priority given to the ports of one ensures that only the higher priority WaveSwitch 100 has forwarding ports on the workgroup LANs when all equipment is functioning correctly. This means that half of the fileserver access traffic passes over one of the ISLs, even during normal operation when all equipment is working. A 10-Mbps link is unlikely to be sufficient for this load, which is why Figure 19 shows 100-Mbps WaveBus links between the two WaveSwitch 100s. The absence of an ISL would cause the spanning tree procedure to elect a single workgroup LAN to carry half of the fileserver traffic, which would probably destabilize the whole LAN. If this configuration had only one WaveBus ISL joining the two WaveSwitch 100s, the single ISL would form a single point of failure for the whole LAN. As a result, only a dual redundant high-speed ISL is compatible with the mission-critical nature of this network.

Figure 19
Mated WaveSwitch 100s with fileserver centralization



An ISL between the mated WaveSwitch 100s in Figures 18 and 19 must form part of the spanning tree at all times. No special arrangements are required for the WaveBus ISLs in Figure 19, because their low path cost, which is due to their high speed, means that one of them will be present in any spanning tree. The Ethernet ISL of Figure 18, however, must be connected to the Ethernet port with the highest spanning tree priority in each WaveSwitch 100, otherwise the ISL would lose the spanning tree election to one of the workgroup LANs.

The network manager may set the priority of a port with SNMP, but, by default, the spanning tree procedure uses port numbers to determine port priority within each WaveSwitch 100, with port 1 having the highest priority. Both Figures 18 and 19 show the ends of the ISL connected to port 1 of each WaveSwitch 100, which avoids the need to use SNMP to set the port priority.

An FDDI point-to-point dual ring could be used, at higher cost, instead of the dual redundant WaveBus ISLs of Figure 19.

Fileserver multiconnection in a mission-critical network

A fileserver must be connected to more than one WaveSwitch 100, as shown in Figure 19, if fileserver access is to survive failure of a WaveSwitch 100.

A multiconnected fileserver incorporates software that permits at least two connections to a single LAN.

The multiconnection software manages two or more connections, either as

- alternates, so that only one correctly-functioning connection is active
- load-sharing partners, having the capability to fall back to fewer connections if some should fail

Load-sharing multiconnection is better for 10-Mbps connections, because it has the additional benefit of improving performance dramatically while all connections are working. Load-sharing multiconnection may provide no perceptible performance improvement for fileservers accessed by 100-Mbps LANs.

FDDI dual homing provides a form of dual connection in which one connection is active at a time. Dual homing is a standard feature of the FDDI MAC procedures; for Ethernet, WaveBus, and other LAN technologies, separate multiconnection software is required.

The NSI file-server multiconnection software described on page 23 provides load-sharing multiconnection software suitable for mission-critical configurations involving Novell NetWare version 3.11 and later. This software works with Ethernet and WaveBus.

Two connections, at least, are required for mission-critical fileservers. For some fileservers, two 10-Mbps connections in load sharing mode may provide sufficient access bandwidth at lower cost than two 100-Mbps connections.

Networks with FDDI fileserver connections

One of the strongest features of the WaveSwitch 100 is its support of high speed connections—FDDI or WaveBus Fast Ethernet. Shared or dedicated Ethernet provides sufficient bandwidth for almost all desktop computers. Fileservers, however, often require higher bandwidth to reach their full throughput potential. FDDI is a widely supported 100-Mbps standard. Many companies manufacture FDDI concentrators and network interface cards.

A general introduction to FDDI networking is beyond the scope of this guide. An excellent elementary introduction to FDDI [4] is available from Digital Equipment Corporation. Mirchandani (ed.) [3] provides a more comprehensive treatment that is suitable for technical network planners.

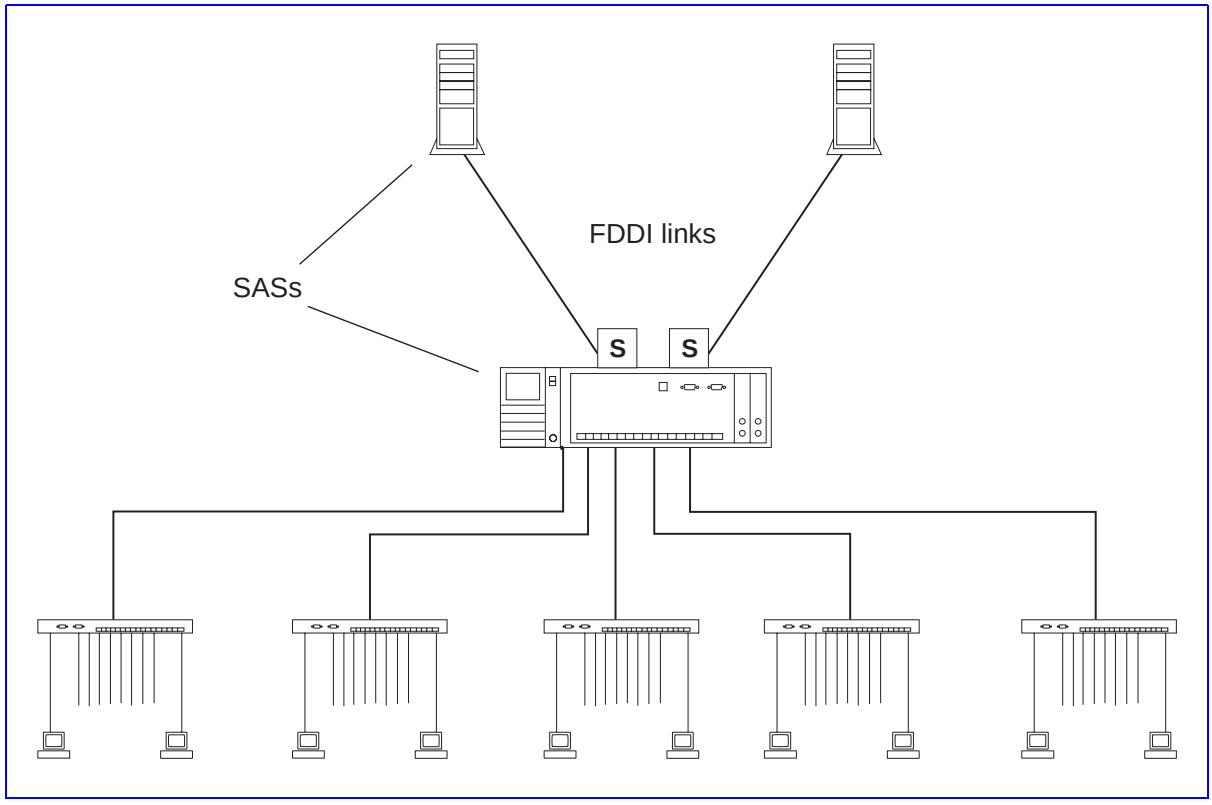
One or two fileservers

This example shows one way of providing high bandwidth access for one or two fileservers. The configuration in Figure 20 has a WaveSwitch 100 into which two optional FDDI SAS (single attachment station) feature modules have been installed, each connected to a single FDDI ring. There is a high-performance fileserver connected to each FDDI ring.

For configurations of one or two fileservers, the cost, complexity, and other attributes of several configurations should be compared, for example see the dual ring configuration on Page 31, and the WaveBus configuration described on page 47.

The next few pages show configurations for networks with more than two fileservers.

Figure 20
A network with two optional FDDI SAS feature modules installed



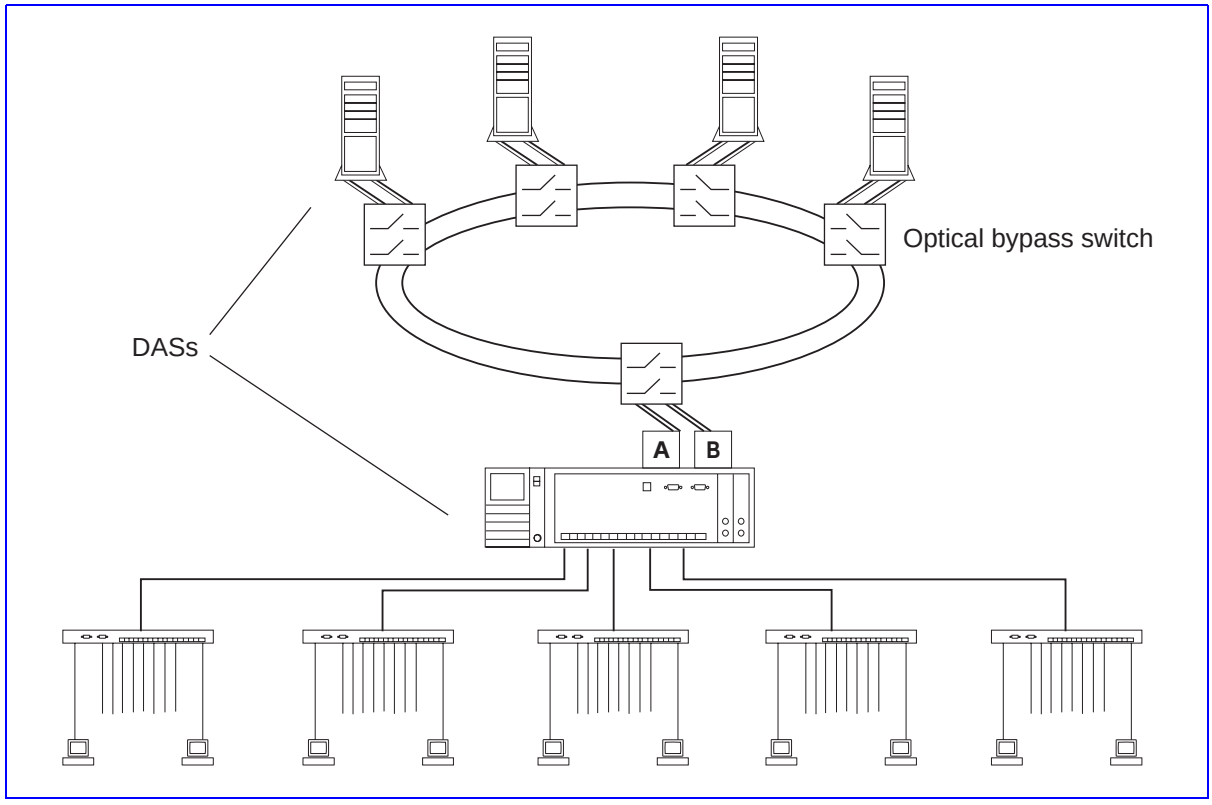
Dual ring of switches and DAS servers

Dual FDDI rings are necessary for some of the mission-critical configurations discussed in later sections. For many applications, including the one shown here, a concentrator providing SAS connections to the fileservers would probably be preferable to the dual ring, because it would be easier to manage. This configuration does, however, provide a simple introduction to dual rings.

Figure 21 shows a WaveSwitch 100 connected to a dual ring of fileservers. The WaveSwitch 100 is a dual attachment station (DAS). This means that one of the option slots of the WaveSwitch 100 is occupied by an optional FDDI DAS feature module. The A and B ports of the WaveSwitch 100 DAS module are shown explicitly. Each of the fileservers is also a DAS.

An optical bypass switch is shown interposed between each DAS and the dual ring. The optical bypass switch allows the DAS to be powered off, or otherwise go down, without wrapping the dual ring. The bypass switches are optional. If they are not present, faults in two DAS stations—turning the power off, or reloading the software, counts as a fault—will divide the dual ring into two isolated segments as the ring wraps at each failed station.

Figure 21
A network with two optional FDDI DAS feature modules installed



Optical bypass switches provide additional security against partitioning of a dual ring when there are more than three DAS connections on the dual ring. A two-fileserver, one-WaveSwitch 100, dual ring would not benefit from the inclusion of optical bypass relays. As a dual ring becomes significantly larger, it becomes difficult to manage without optical bypass switches. Large dual rings should be preconfigured with unused optical bypass switches installed in server rooms to allow for network expansion.

Each WaveSwitch 100 FDDI DAS feature module has a connector that supplies electrical signals to control an external optical bypass switch. Suitable optical bypass switches are manufactured by AMP and Dicon Fiber Optics.

Configurations based on FDDI concentrators are easier to manage than dual rings, and are equally secure in applications like this one. Dual rings, usually in conjunction with concentrators, are indispensable in some mission-critical configurations, however.

In a dual ring, all DAS stations have equal responsibility for the integrity of the dual ring. One DAS station cannot relieve other DAS stations of that responsibility. Concentrators are intended to provide just such a transfer of responsibility for ring integrity away from individual stations.

Single attachment concentrators and SAS servers

Figure 22 shows a simple, recommended configuration with an FDDI concentrator that provides high-speed interconnection between the WaveSwitch 100 and multiple fileservers. This configuration is suitable for many applications.

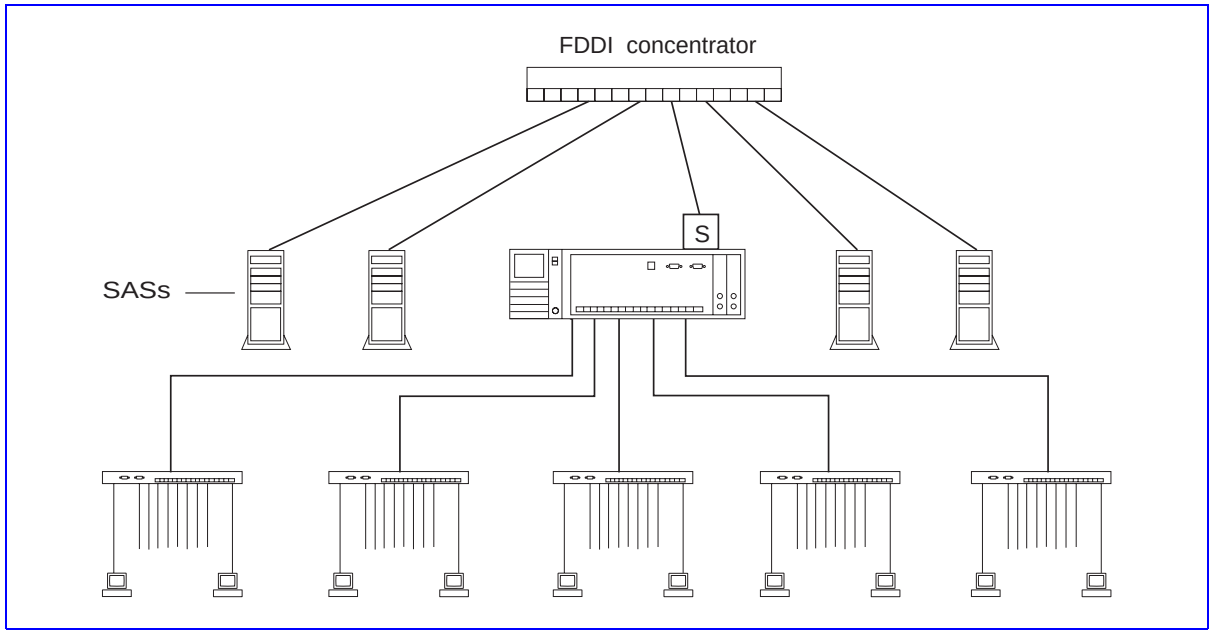
The WaveSwitch 100 operates as an FDDI single attachment station (SAS) connected to an M port of the concentrator. See the summary of FDDI connection rules on page 40.

This means that one of the two option slots of the WaveSwitch 100 is occupied by an FDDI SAS feature module, shown as S. The other option slot is unoccupied. Each of the fileservers also operates as a SAS connected to an M port of the concentrator.

Suitable FDDI concentrators are available from many vendors.

This configuration offers several management advantages compared to the configuration discussed on page 31. The concentrator operates somewhat like a collapsed dual ring. The concentrator verifies the correct operation of each attached station, and automatically takes failed stations, or stations that are powered off, out of the ring. The concentrator also operates as a manageable entity visible through SNMP; it provides a central point of control for the whole FDDI ring. The network manager can direct the concentrator to remove from the ring a station that is causing trouble.

Figure 22
A single FDDI concentrator for fileserver connection to the WaveSwitch 100

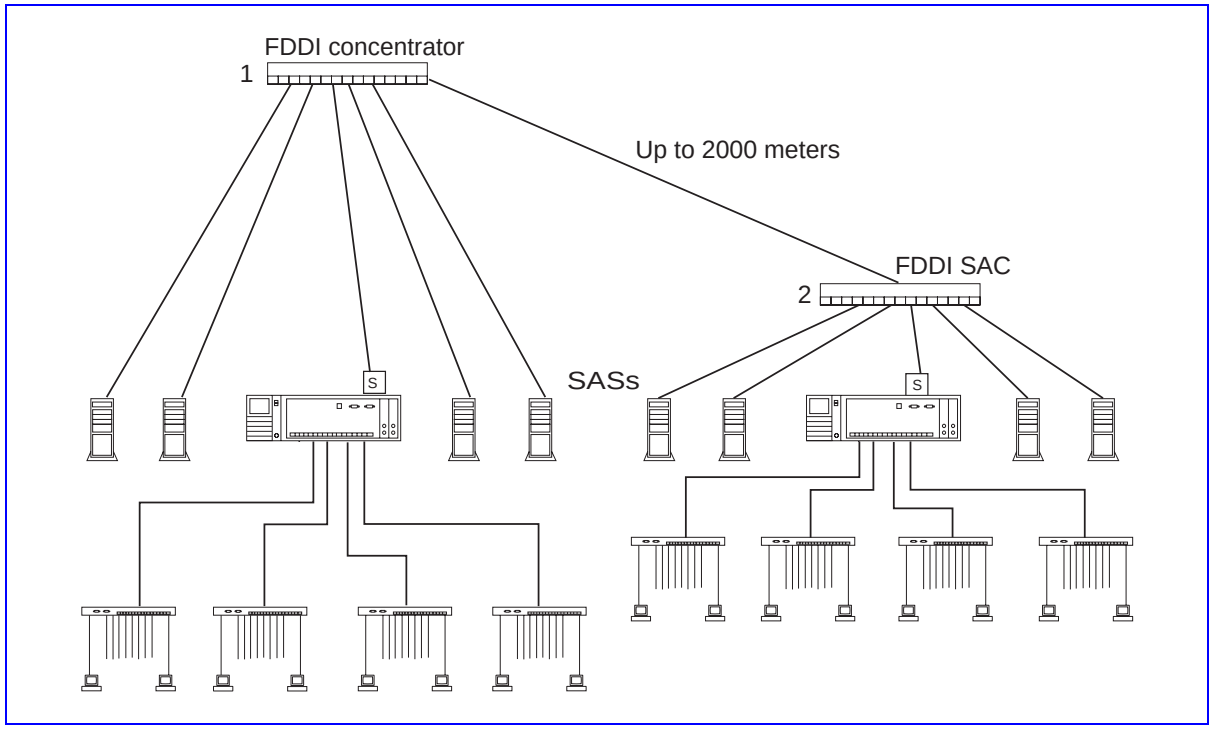


A concentrator can provide the means to include multiple FDDI media in one FDDI network to achieve an economical combination of media. The WaveSwitch 100 connects to the concentrator through 1300 nanometer multimode fiber-optic cable, but fileservers or other concentrators can connect by any standard FDDI medium, including unshielded twisted pair, shielded twisted pair, low cost fiber, or even single mode fiber, provided the concentrator supports these media types.

Figure 23 shows an application of the same principles in a slightly larger network, possibly for a campus, with two WaveSwitch 100s and two concentrators arranged in a tree-of-concentrators configuration. The S port of concentrator 2 is connected to an M port of concentrator 1. See page 40 for a summary of FDDI connection rules. This configuration is a simple extension of the configuration in Figure 22.

Although these are practical and recommended configurations, suitable for many applications, a network manager with extreme reliability needs might point with concern to the concentrator(s) and the WaveSwitch 100 as single points of failure. In some cases, such concerns can be addressed by keeping sufficient spares on site to ensure quick repair. In other cases, only configurations with redundant equipment ready to take over in the event of failure will meet the reliability requirement. The following pages present configurations that meet the most demanding reliability requirements.

Figure 23
A campus interconnection



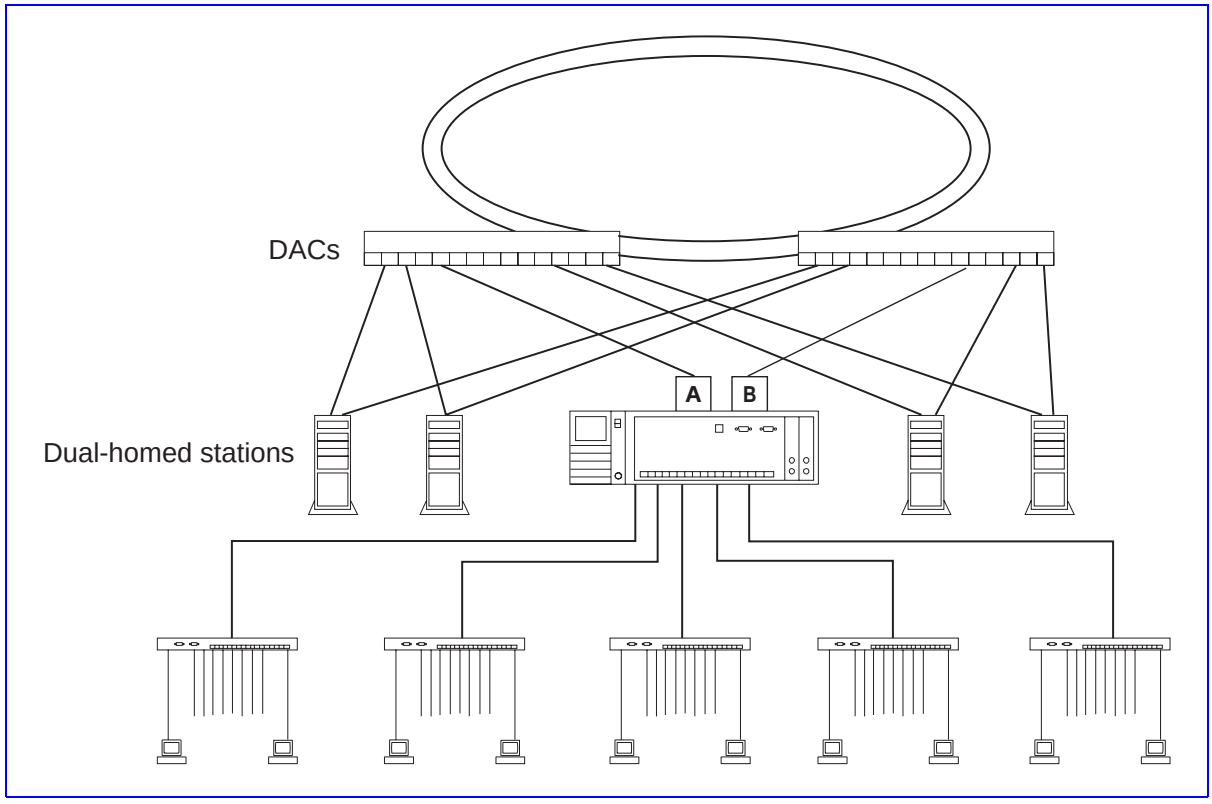
Redundant DACs, dual-homed servers, dual-homed WaveSwitch 100

The configuration in Figure 24 eliminates the FDDI concentrator as a single point of failure. The WaveSwitch 100 remains as a single point of failure which is eliminated in the configurations in the next section by adding a redundant WaveSwitch 100.

Each fileserver—and the WaveSwitch 100—is connected to two FDDI concentrators. The two FDDI concentrators (dual attachment concentrators, or DACs) are dual attached to a dual ring. The dual ring is recommended to harden the FDDI LAN against partitioning, even though only the two concentrators are attached to it. The investment in dual connections for the fileservers and the WaveSwitch 100 would be incompatible with a single ring connection between the two concentrators.

The fileserver connections to the concentrators, are neither SAS nor DAS connections; they are dual-homed connections. The connections between each fileserver and the two concentrators emanate from a single DAS network interface card in the fileserver. Dual homing is a standard (ANSI X3T9.5) mode of use of a DAS network interface card.

Figure 24
A dual-homed WaveSwitch 100 with redundant concentrators and a dual ring



The WaveSwitch 100 FDDI DAS feature module can also be configured for dual-homed connection. See page 40.

The FDDI Station Management (SMT) procedures allow only one of the two ports of a dual-homed station to be active at one time. The SMT procedures of the dual-homed DAS equipment monitor the state of the active connection and activate the inactive connection if the active connection fails. The dual-homed connection represents a single MAC entity, and the same MAC address is used on both connections. To procedures or applications that use the FDDI MAC in the dual-homed station, the dual-homed connection appears as a single connection.

Dual homing is a powerful redundancy option. It is widely used in mission-critical configurations. FDDI is the only LAN standard which provides dual homing as a standard feature of its MAC procedures. Other LAN technologies, like Ethernet and WaveBus, rely on additional software solutions above the MAC layer (for example, the spanning tree procedure) that manage two network interface cards to produce a redundant single attachment solution.

The WaveSwitch 100 is also dual homed to the two concentrators. The two connections shown between the WaveSwitch 100 and the two FDDI concentrators emanate from the A and B ports of a single FDDI DAS feature module occupying one of the two option slots of the WaveSwitch 100.

Extremely reliable network: mated WaveSwitch 100s, redundant DACs, dual-homed servers

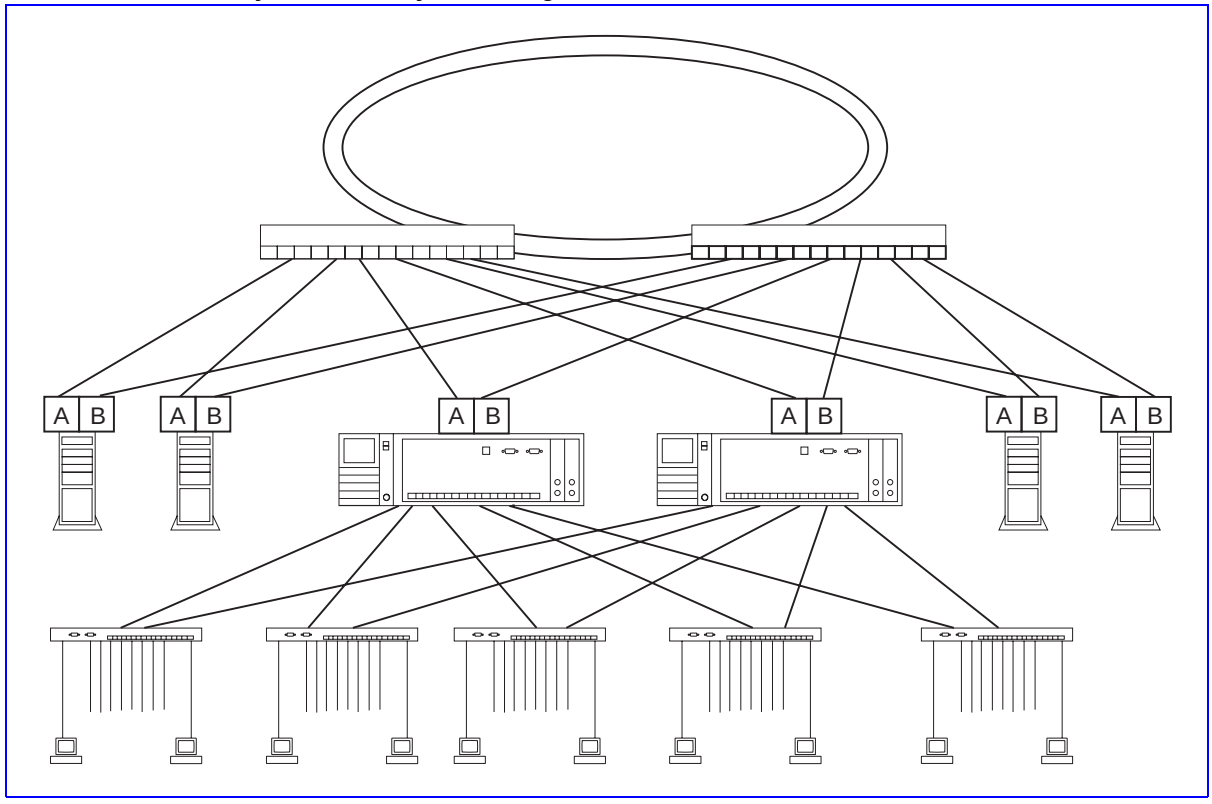
The configurations in Figures 25 and 26 are tolerant to the failure of a WaveSwitch 100 or a concentrator. Concentrators and WaveSwitch 100s are provisioned redundantly. Both configurations provide protection against failure of one concentrator, failure of one WaveSwitch 100, or simultaneous failure of one concentrator and one WaveSwitch 100.

In Figure 25, two dual attachment concentrators (DACs) are redundantly connected by a dual ring.

In Figure 26, one single-attachment concentrator (SAC) is connected by its S port to an M port of a stand-alone concentrator. This configuration is a low-cost upgrade of the non-redundant configuration in Figure 23 for mission-critical duty. This configuration is easily managed, but is not suitable for larger mission-critical networks, which require a dual ring for reliable interconnection of additional devices, such as another pair of mated WaveSwitch 100s.

In both configurations, each fileserver and WaveSwitch 100 is dual homed to two FDDI concentrators. The activity state of a dual-homed connection

Figure 25
Two DACs redundantly connected by a dual ring

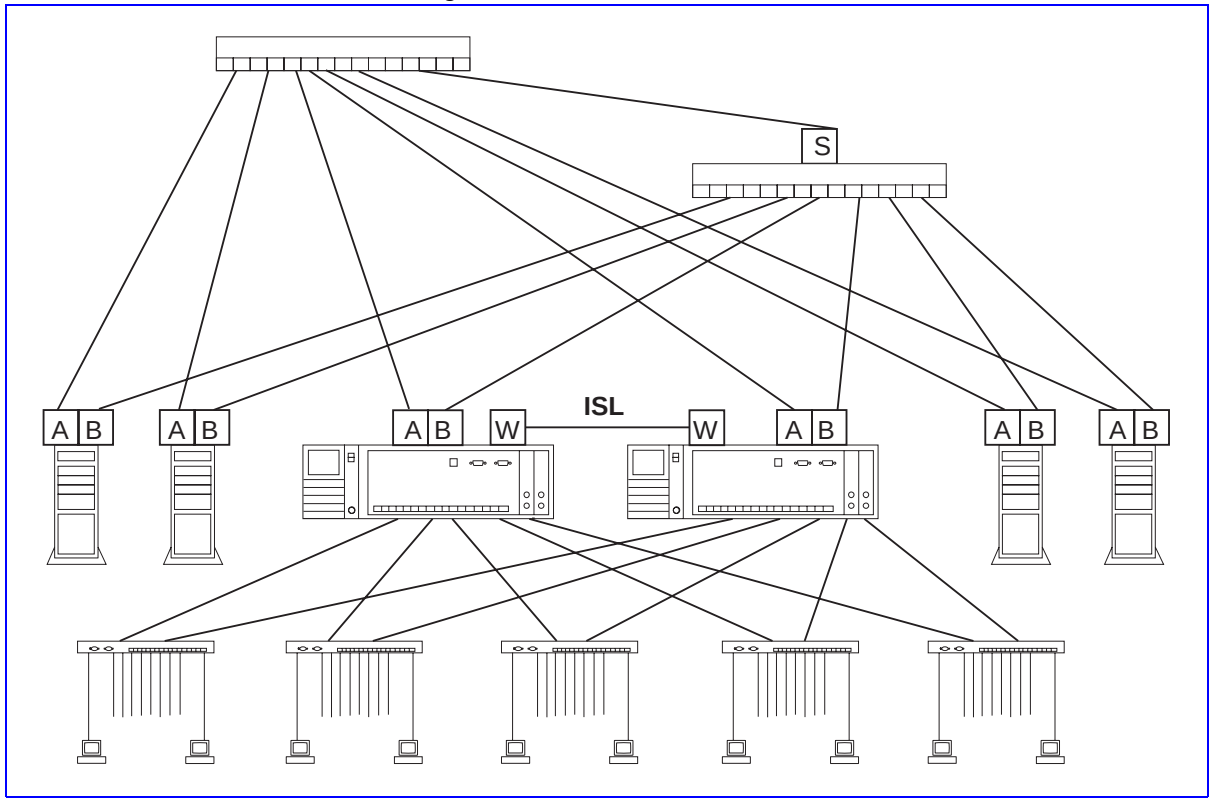


is controlled by the standard FDDI SMT procedures associated with each dual-homed DAS interface. The FDDI connection rules give priority to the B port. The WaveSwitch 100 management procedures, including the spanning tree procedures, respect the autonomy of SMT, and do not exert control over the AB activity state, but accept it as provided by SMT. The spanning tree procedure determines whether each dual home port of the WaveSwitch 100 will be allowed to enter forwarding state—the state in which the port forwards traffic in both directions between the medium and the WaveSwitch 100.

To analyze these networks, keep in mind that the spanning tree procedure will choose ports of the two WaveSwitch 100s to put into forwarding and non-forwarding states so that traffic can flow from one LAN to another only by a single path. When choosing which ports to put into forwarding state, the spanning tree procedure will give higher priority to root ports on FDDI and WaveBus LANs.

In addition to protecting against the failure of one concentrator or one WaveSwitch 100, the design of these networks reduces the risk that faults will cause the spanning tree procedure to ask an Ethernet workgroup to carry fileserver access traffic for other Ethernet networks, which could cause network instability when the Ethernet overloads.

Figure 26
A non-redundant mission-critical configuration



In Figure 25, the redundancy of the dual ring reduces the probability of a partition of the FDDI network. If a partition occurs between the two concentrators in spite of the dual ring, the priority of B ports, and the fact that they are connected to the same concentrator, still gives a high probability that the WaveSwitch 100s will be connected by the FDDI network, not by an Ethernet.

In Figure 26, there is a higher probability of partition of the FDDI (single ring) network between the two concentrators. Connecting the B ports to the same concentrator reduces the probability that a partition will cause the two WaveSwitch 100s to bridge the partition through an Ethernet.

The WaveBus interswitch link (ISL) is an optional refinement that compensates for the absence of a dual ring, and gives it further protection against bridging between the WaveSwitch 100s through an Ethernet. The ISL also protects against a cabling error in which the B ports of the WaveSwitch 100 are connected to different concentrators.

At least three faults would be required in either configuration before the spanning tree procedure would ask an Ethernet to carry fileserver access traffic for other Ethernets. This counts partition of a dual ring between the two DACs as two faults, the failure of the SAS connection between the SAC concentrators as one fault, the failure of a B port as one fault, and the failure of the WaveBus ISL as one fault.

Summary of FDDI connection rules

Figure 27 summarizes FDDI entities and their interconnection rules.

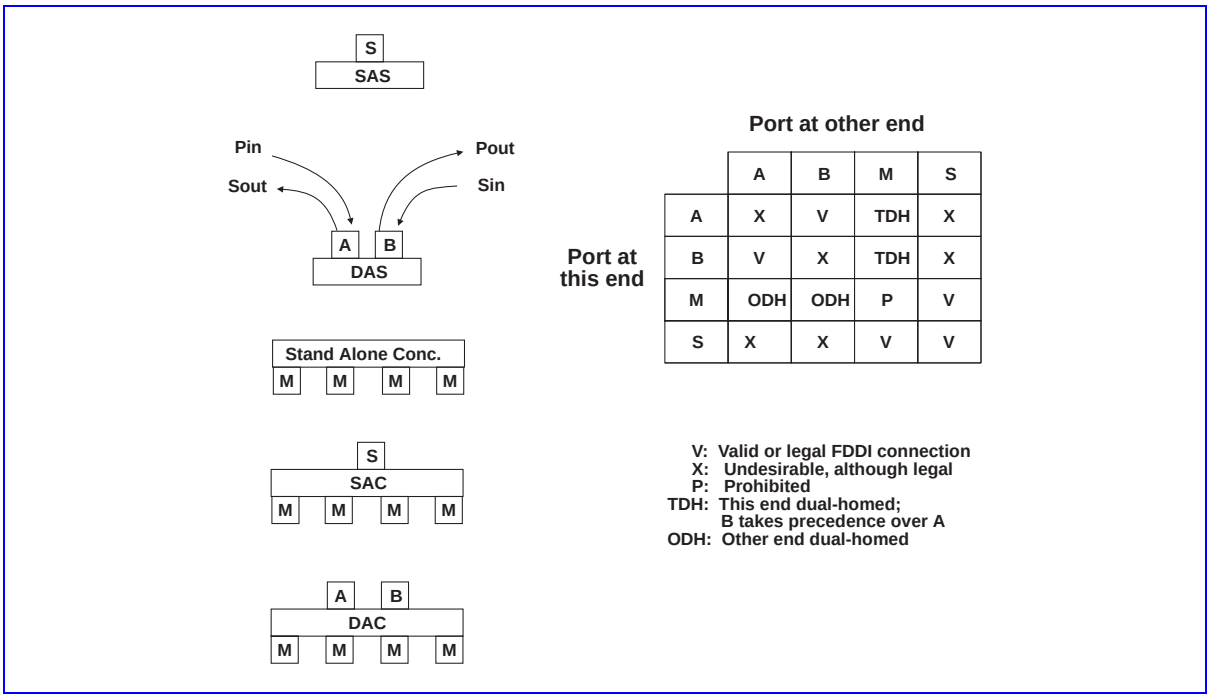
A single attachment station (SAS) connection has one port designated S. An S port can connect to an M port or to another S port. A single station can have more than one SAS connection. Examples: workstation, fileserver, bridge, router, WaveSwitch 100. The WaveSwitch 100 FDDI SAS feature module provides one S port.

A dual attachment station (DAS) connection has two ports designated A and B. The primary ring enters, and the secondary ring exits, at port A. When connecting to a dual ring, an A (B) port can be connected to a B (A) port, or to an M port when dual homing. A single station can have more than one DAS connection, such as a workstation, fileserver, bridge, router, or WaveSwitch 100. The WaveSwitch 100 FDDI DAS feature module provides one type A port and one type B port.

A stand-alone concentrator has multiple ports designated M. An M port can connect only to an A or B port (dual homing) or to an S port. M ports must never be directly cabled together. A single attachment concentrator (SAC) has one port designated S, and multiple ports designated M.

A dual attachment concentrator (DAC) has two ports, one designated A and the other B, and multiple ports designated M.

Figure 27
FDDI connection rules



The table in Figure 27 is used by the FDDI port management procedures of each port to decide whether the port at the other end of the point-to-point FDDI cable can be correctly cabled directly to it. The decision is made by the Connection Management (CMT) module of Station Management (SMT). (SMT is part of the ANSI X3T9.5 FDDI standard.) The connections marked X are legal but undesirable, since they wrap the ring. SMT can be instructed to alarm and disallow these connections.

IP fragmentation

Many TCP/IP implementations transmit the maximum length IP packets that FDDI permits, without regard to whether the destination of the packet is a bridged Ethernet station, which can receive only much shorter packets.

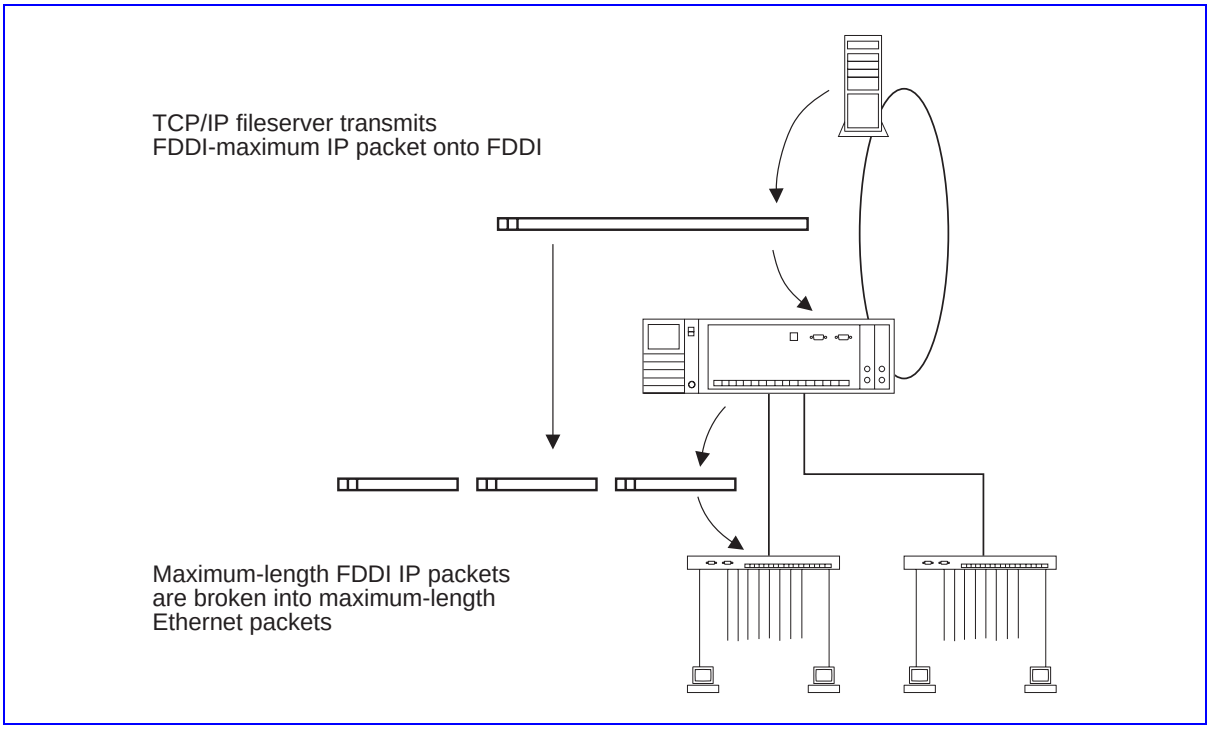
Fortunately, a bridge—or WaveSwitch 100—can emulate a TCP/IP router for the narrow purpose of intercepting such too-long packets and breaking them up, or fragmenting them, according to the rules of IP. This technique has become a standard for FDDI/Ethernet bridges.

Figure 28 shows a long packet transmitted by a TCP/IP application (probably NFS) in the fileserver connected to the FDDI network. The packet is shown being fragmented by the WaveSwitch 100 for consumption by a station on one of the Ethernets.

IP is the only protocol for which the WaveSwitch 100 performs this operation.

For all packets except long IP packets, the WaveSwitch 100 operates as a translating bridge, transforming packets between FDDI and Ethernet as specified by IETF RFC 1042 and recommendation 802.1H of the Institute of Electrical and Electronics Engineers (IEEE).

Figure 28
IP fragmentation



Networks with WaveBus fileserver connections

WaveBus is a simple, robust, and cost effective 100-Mbps fiber-optic Fast Ethernet. WaveBus was inspired by 10Base-T Ethernet, but uses fiber-optic cable where 10Base-T uses twisted pairs of copper wire. Because it has the same packet format as Ethernet, WaveBus Fast Ethernet is an ideal interconnection between fileservers and the WaveSwitch 100.

The high-speed interconnection provided by WaveBus permits fileservers to achieve their full throughput potential. The use of a 100-Mbps connection between a fileserver and WaveSwitch 100 cuts the total delay through the switch almost in half compared to 10-Mbps connections.

WaveBus is dramatically simpler than FDDI, requiring fewer new concepts for configuration and installation. There are no control parameters to be chosen during installation or operation. Network management is almost identical to that of 10Base-T; WaveBus Hubs are managed by the SNMP MIB which is standard for 10Base-T hubs (RFC 1368.)

WaveBus operates point-to-point, to connect two WaveSwitch 100s, or to connect a WaveSwitch 100 to a fileserver. WaveBus also operates as a shared LAN based on a hub or tree of hubs, to connect multiple WaveSwitch 100s and fileservers.

In point-to-point connections, WaveBus operates full duplex to provide a two-way bandwidth of 200 Mbps. Any WaveBus connection can extend to 500 meters (1640 feet) using multimode fiber-optic cable. Optional long-link port cards for the WaveSwitch 100 and WaveBus Hub permit cable lengths up to 2000 meters (6560 feet).

External link-extender equipment is available to permit WaveBus Hub connections up to 10 kilometers (6.2 miles) over single mode fiber, and point-to-point connections between the WaveSwitch 100 to 20 kilometers (12.4 miles) over single mode fiber.

Figure 29 shows four fileservers connected to a WaveSwitch 100 by four point-to-point WaveBus links. In this example each of the two option slots of the WaveSwitch 100 is occupied by a two-port WaveBus feature module.

Figure 29
WaveSwitch 100 network with four fileservers

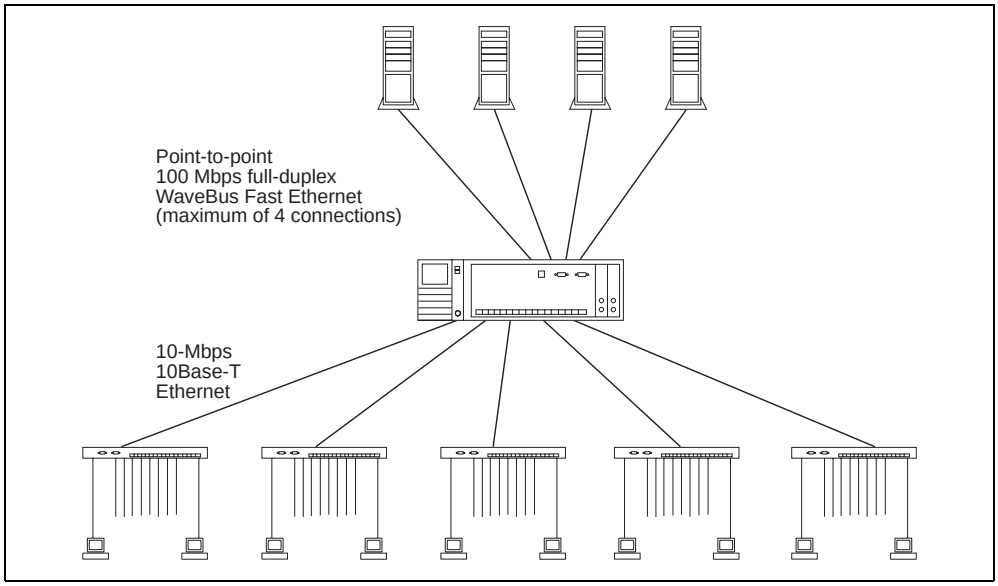
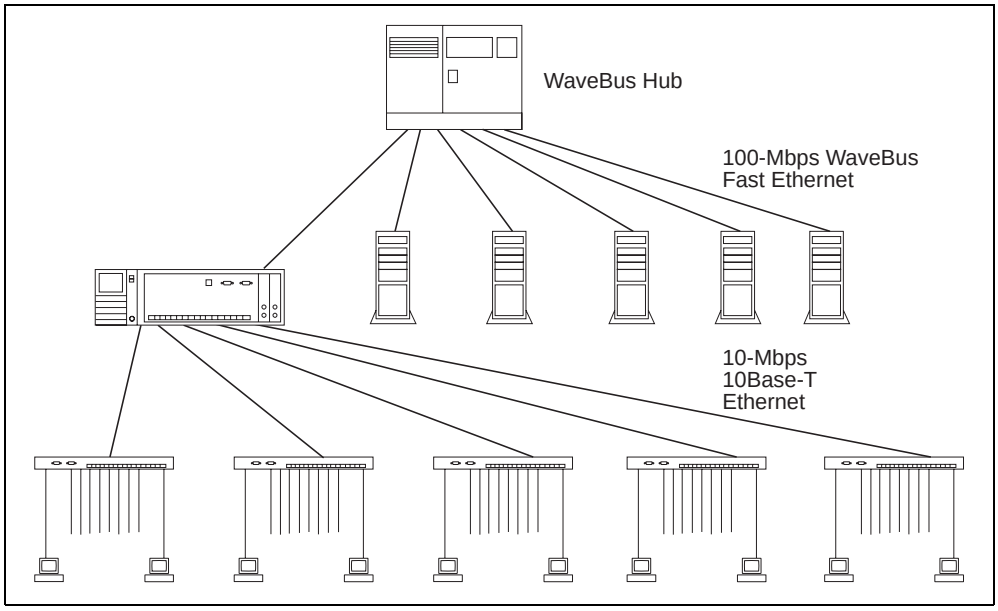


Figure 30 shows how to connect more than four fileservers directly to a WaveSwitch 100 using a WaveBus Hub. One hub can interconnect 16 devices, so this configuration could be extended to 15 fileservers connected to the WaveSwitch 100.

Figure 30
Connecting more than four fileservers to a WaveSwitch 100 with WaveBus

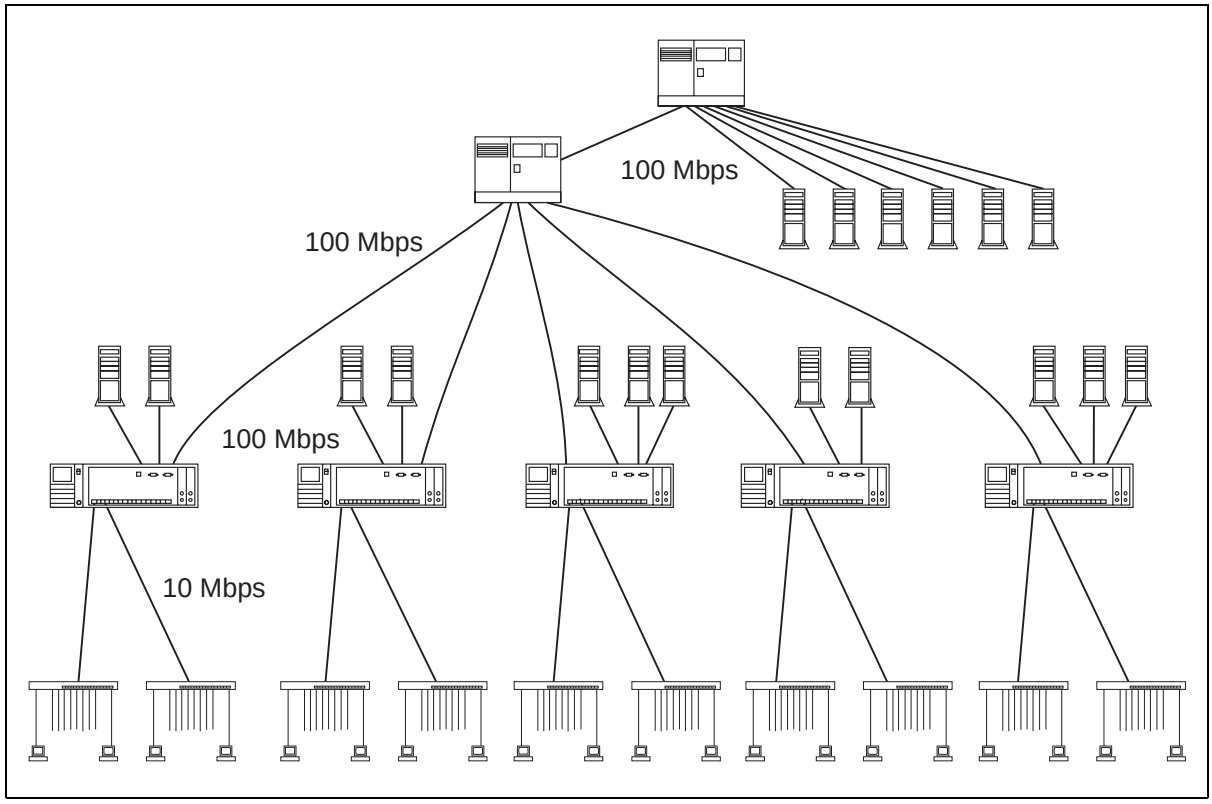


Large WaveSwitch 100/WaveBus networks

Figure 31 shows a larger WaveBus network connecting multiple WaveSwitch 100s and fileservers. This figure represents a campus network in which the local network at each site has one WaveSwitch 100. The WaveBus Hubs are arranged in an inverted tree, much like 10Base-T concentrators in an Ethernet network. Although two WaveBus Hubs are shown, many more hubs could be connected in such a network. The WaveBus long-link option permits each WaveBus link to extend to 2000 meters. With optional additional equipment to convert to single mode fiber for long distances, each WaveBus link can extend to 10 kilometers.

Fileservers are shown at each site, and at a central location. Fileservers attached directly to the WaveSwitch 100 at each site serve primarily the workgroups at the site of the WaveSwitch 100. Locating fileservers at each site permits the separate management of the fileservers by site personnel, and localizes the access traffic for the local fileservers, freeing the backbone network that interconnects the site, and increasing the total capacity of the network.

Figure 31
A campus network with WaveSwitch 100 and WaveBus



Servers attached directly to the WaveBus backbone contain databases shared equally by the whole enterprise, and would probably be managed centrally at one of the sites.

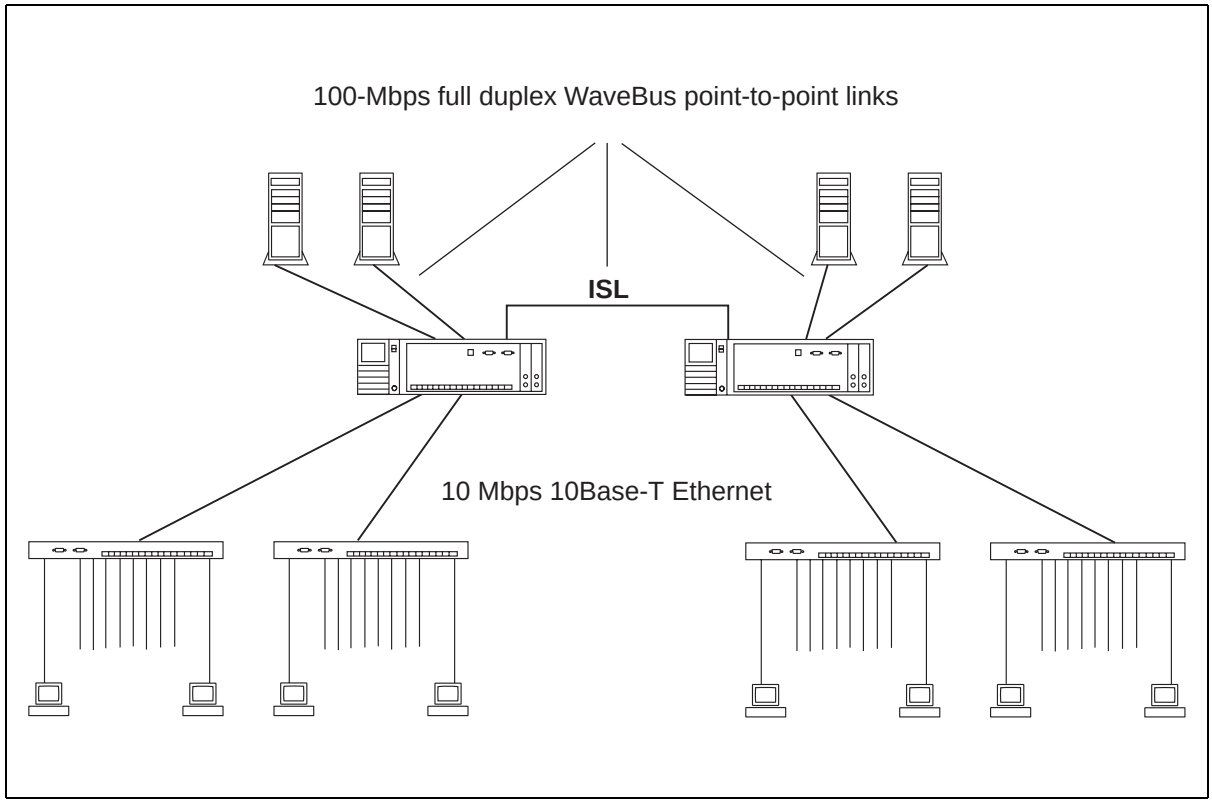
The manager of each site can deny access by other sites to one or more fileservers attached to the local WaveSwitch 100, and deny access to one or more central shared fileservers from one or more of the workgroups attached to the local WaveSwitch 100. The WaveSwitch 100 accepts management instructions—through the SNMP bridge MIB of RFC 1493—to discard packets addressed to a particular destination depending on the port on which the packet arrives at the WaveSwitch 100.

Point-to-Point WaveBus connections between WaveSwitch 100s

WaveBus is ideal as a point-to-point connection between WaveSwitch 100s. The full-duplex 100-Mbps WaveBus connection gives a two-way total bandwidth of 200 Mbps, which produces low delay and high throughput between the two WaveSwitch 100s. WaveBus feature modules are available in options that permit maximum link distances in point-to-point connections up to 500 meters (1640 feet) or up to 2000 meters (6560 feet). Additional equipment is available to extend these distances up to 20 kilometers (12.4 miles) over single mode fiber.

Figure 32, shows two sites, each based on a WaveSwitch 100. A point-to-point interswitch link (ISL) connects the two sites. The fileserver connections and the ISL are point-to-point WaveBus full duplex links.

Figure 32
A point-to-point WaveBus interswitch link



Mission-critical WaveBus networks

WaveBus networks can be provisioned redundantly to achieve mission-critical reliability. The root of a WaveBus network can be duplicated as shown in Figure 33. The two mated WaveBus Hubs in Figure 33 negotiate to determine which hub should occupy the root of the WaveBus network. WaveBus stations—such as WaveSwitch 100s—or other WaveBus Hubs are connected to both mated hubs.

The capability to operate as a mated root hub is present in all WaveBus Hubs. A switch on the controller card of the hub must be set to indicate that mate behavior is desired. Mated hubs must be cabled as shown.

The 18 ports of a WaveBus Hub—all of which may not be equipped—are numbered 1U, 2U, 3D, 4D...18D. The U ports are up ports; the D ports are down ports. U ports are used to create a hierarchy of hubs. The U ports of one hub (in a non-redundant network, each hub has only one U port) connect to D ports of a hub higher in the hierarchy. In a redundantly provisioned WaveBus network, the top of the hierarchy is formed by a pair of mated hubs, which are cabled in the special way shown in Figure 33.

An election cable joins the ports of the mated hubs numbered 2U. It provides a path by which the two hubs compete to occupy the root position. The two cables joining port 1U of each mate hub to port D3 of the other mate hub are cross cables. Only the cross cable connected to D3 of the elected root hub carries payload traffic. The other cross cable serves as a redundant election cable.

Figure 33
Cabling for a pair of mated WaveBus Hubs

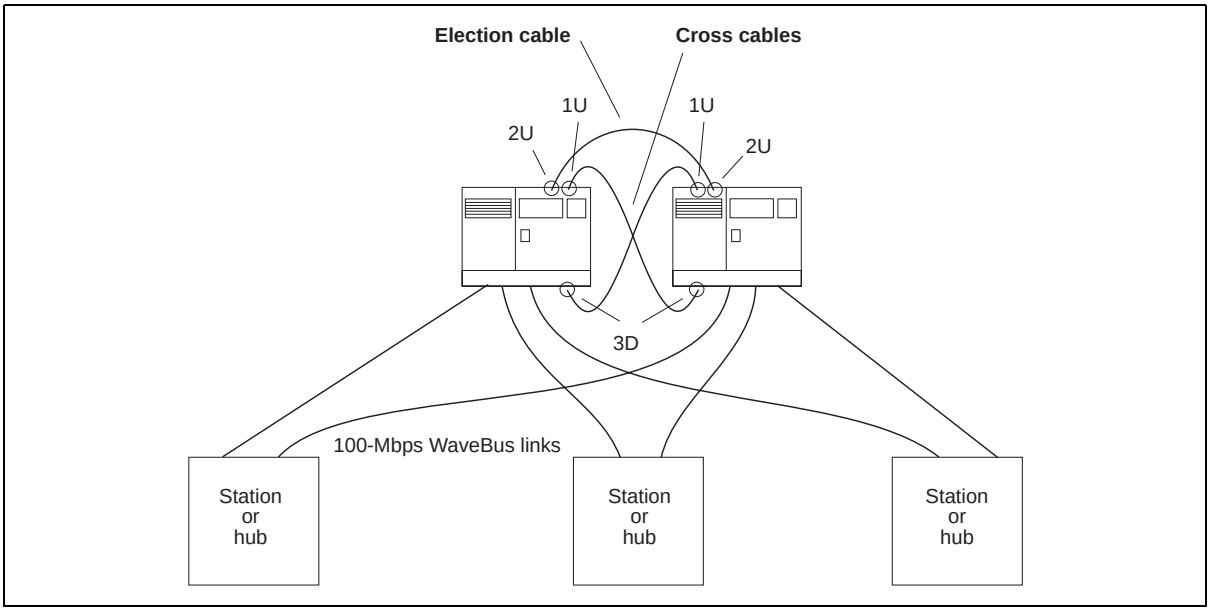
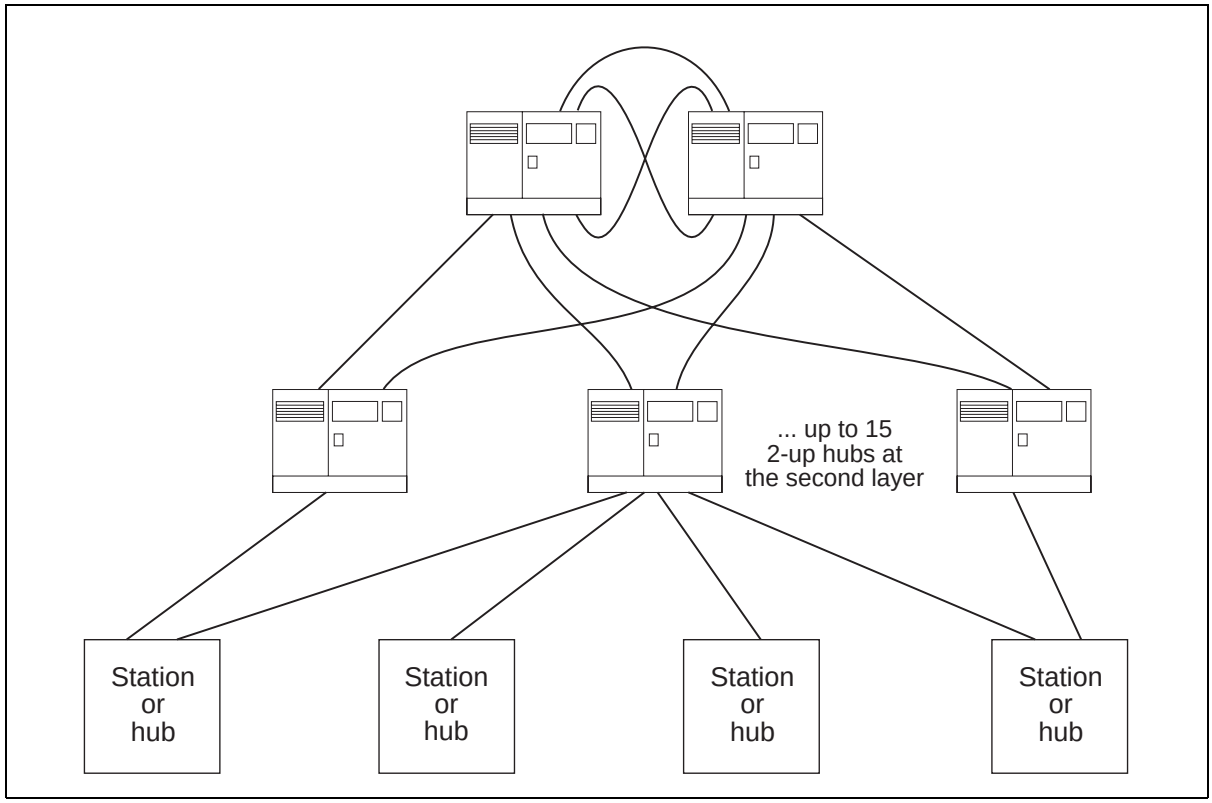


Figure 34 shows a large mission-critical WaveBus network formed from a hierarchy of WaveBus Hubs. A hub connected to two hubs above it by its up ports is called a *2-up* hub. Only one up port of a 2-up hub can be active at a time. The network automatically detects failures and selects new up ports to restore the network. Up to fifteen 2-up hubs can be connected to a pair of mated hubs. The next layer of the network could be composed of one-up or 2-up hubs, with a maximum fan-out of thousands of stations. Any layer of the network can have a mixture of hubs and stations.

Figure 34
A larger redundantly-provisioned WaveBus network

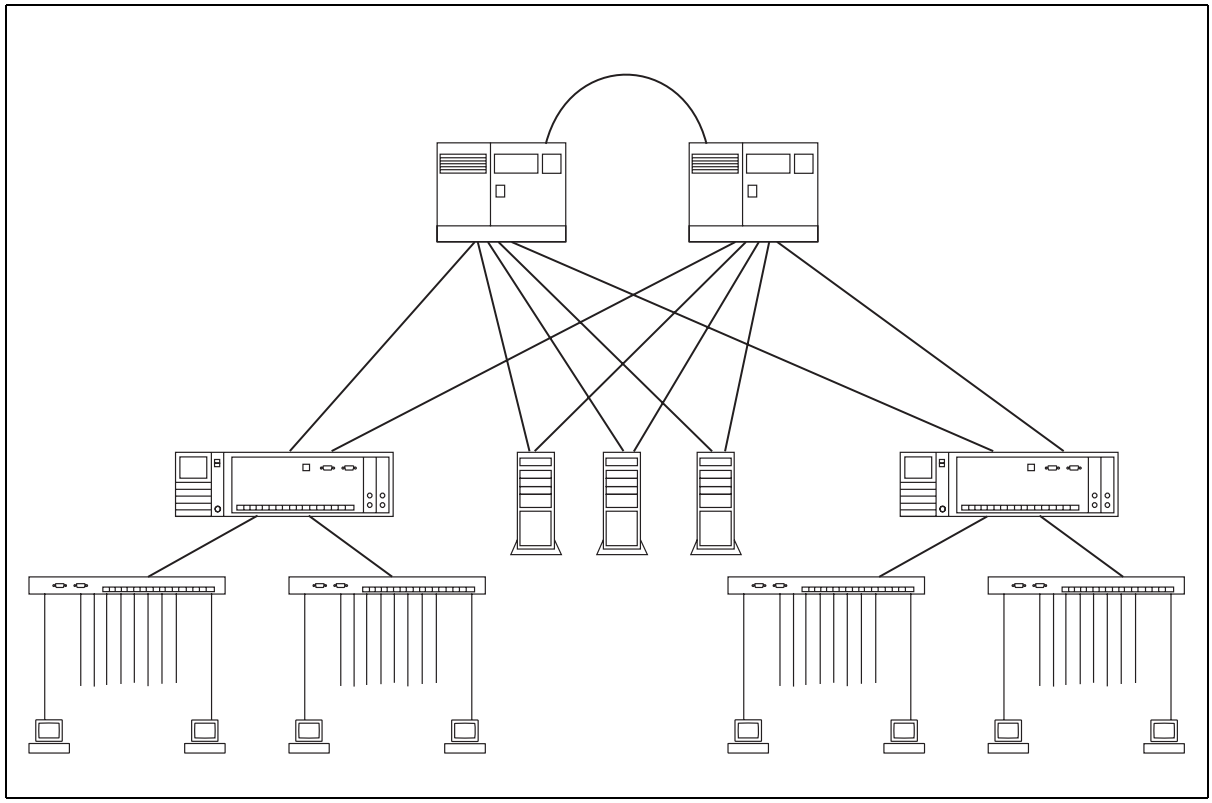


Mission-critical WaveBus/WaveSwitch 100 networks

Figures 35 and 36 show networks that combine WaveBus and WaveSwitch 100s to achieve extreme reliability. In both networks, the WaveBus Hubs and the WaveSwitch 100 are redundantly provisioned. Fileservers are dual connected to both WaveBus Hubs. The two connections of each fileserver are assumed to operate in load-sharing mode, which requires special software like the software from NSI described on page 23.

The network in Figure 35 provides protection against failure of one WaveBus Hub, but is vulnerable to the failure of one WaveSwitch 100.

Figure 35
A mission-critical WaveBus network connecting nonredundantly-provisioned WaveSwitch 100s

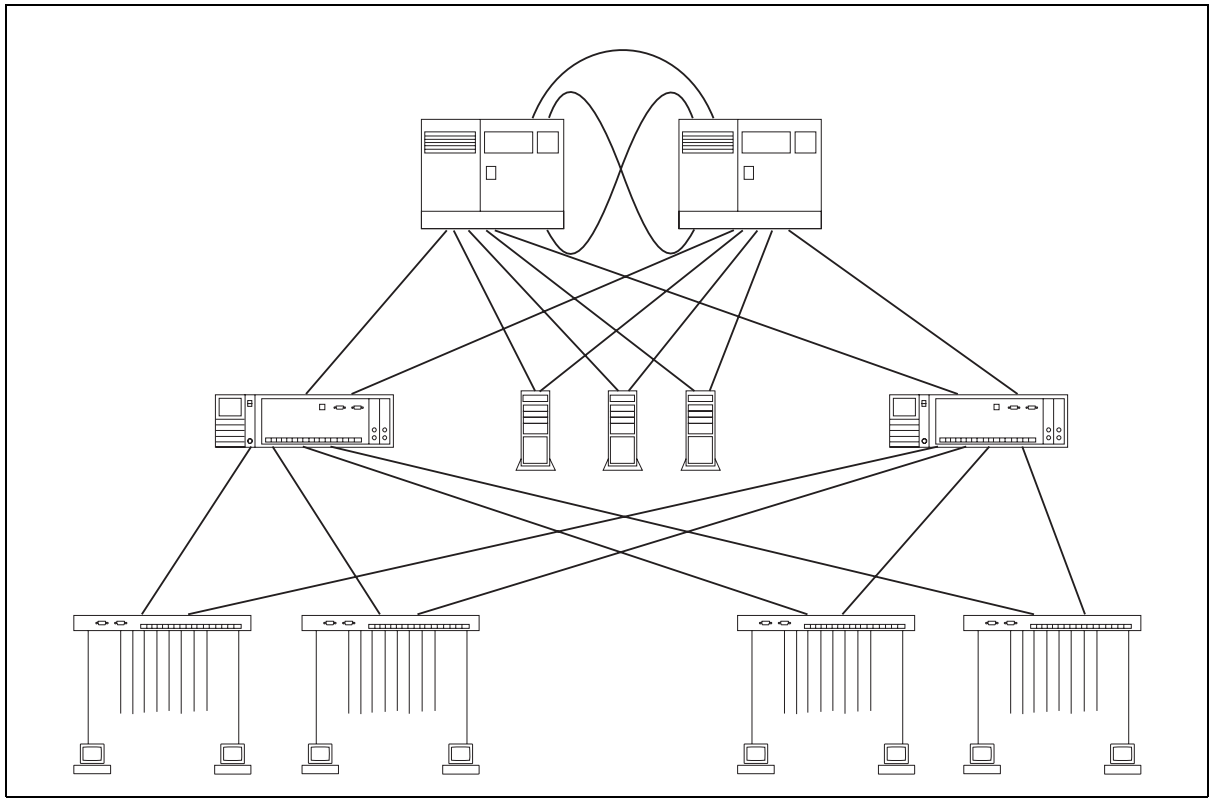


The network in Figure 36, provides protection against failure of a WaveBus Hub, failure of a WaveSwitch 100, or simultaneous failure of a hub and a WaveSwitch 100.

The low path cost of WaveBus, due to its speed of 100 Mbps, ensures that the WaveBus network of Figure 36 will form the spanning tree link between the two WaveSwitch 100s, in preference to one of the Ethernets.

A partition of the WaveBus network would cause the spanning tree procedure to provide a spanning tree path between the partitions through a one of the Ethernets. Because about half of the fileserver access traffic must flow between the two WaveSwitch 100s on the spanning tree link between them, such a partition would destabilize catastrophically the network of Figure 36. This possibility provides additional motivation for the redundancy of the WaveBus. This cannot happen in Figure 35, because each Ethernet is connected only to one WaveSwitch 100. The configuration in Figure 35 will be crippled but stable in the event of a partition of the WaveBus network.

Figure 36
A mission-critical WaveBus network connecting redundantly-provisioned WaveSwitch 100s



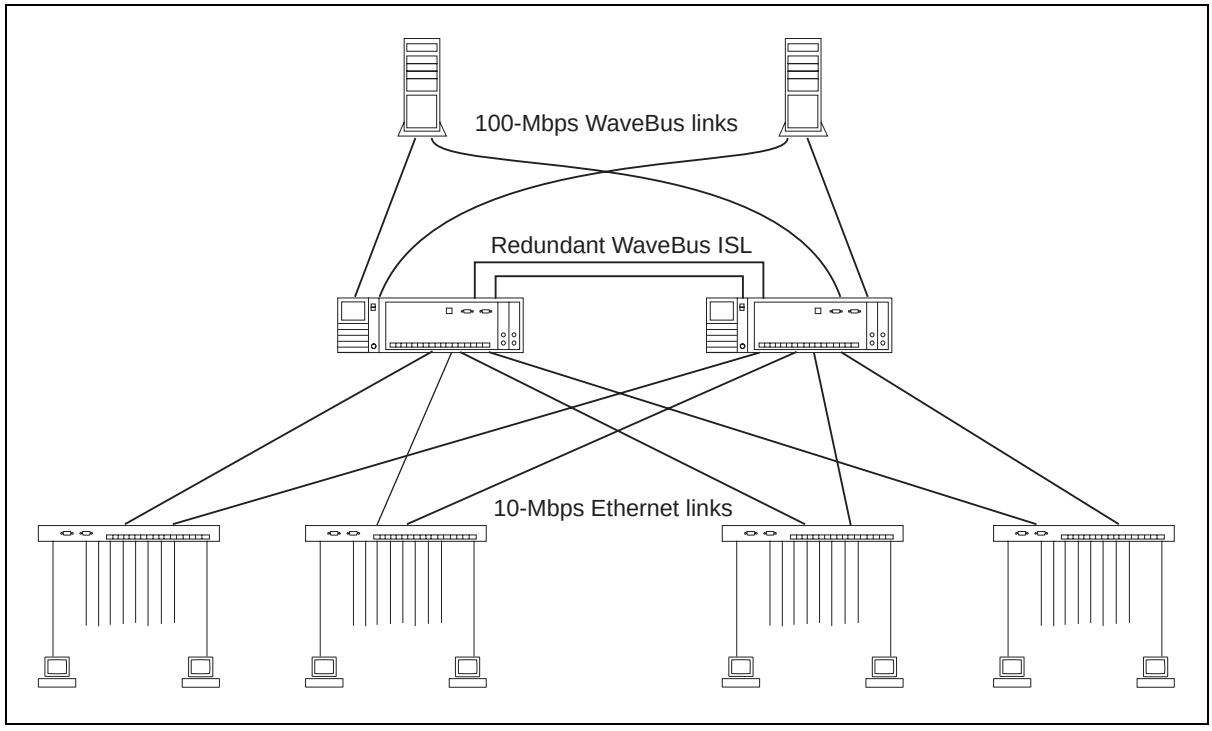
Mission-critical WaveSwitch 100 networks without WaveBus Hubs

Some mission-critical applications are based on one or two high performance fileservers at a site with many workstations. The configuration in Figure 37 provides extreme reliability and high performance for such applications. Each fileserver is dual-connected, one connection to each WaveSwitch 100, by 100-Mbps WaveBus links. The WaveSwitch 100s are connected directly to each other by redundant WaveBus interswitch links (ISLs). This configuration protects against failure of the links to the fileservers, or failure of one of the WaveSwitch 100s.

The dual fileserver connections operate in load-sharing mode.

One of the two WaveSwitch 100s is the designated bridge for all Ethernets while there are no faults. The ISL is required to ensure that one of the Ethernets does not form the spanning tree connection between the two WaveSwitch 100s. Each WaveBus connection to a fileserver is a separate LAN, forms a part of the spanning tree, and carries half the traffic of the fileserver. Half the fileserver traffic flows over the ISL while there are no faults. If there were no ISL, one of the Ethernets would be required to carry half the fileserver access traffic. Failure of a non-redundant ISL would require one of the Ethernet LANs to carry half the fileserver access traffic. To prevent this, the ISL is configured redundantly.

Figure 37
Redundant WaveSwitch 100s with two fileservers



Large mission-critical WaveSwitch 100/WaveBus networks

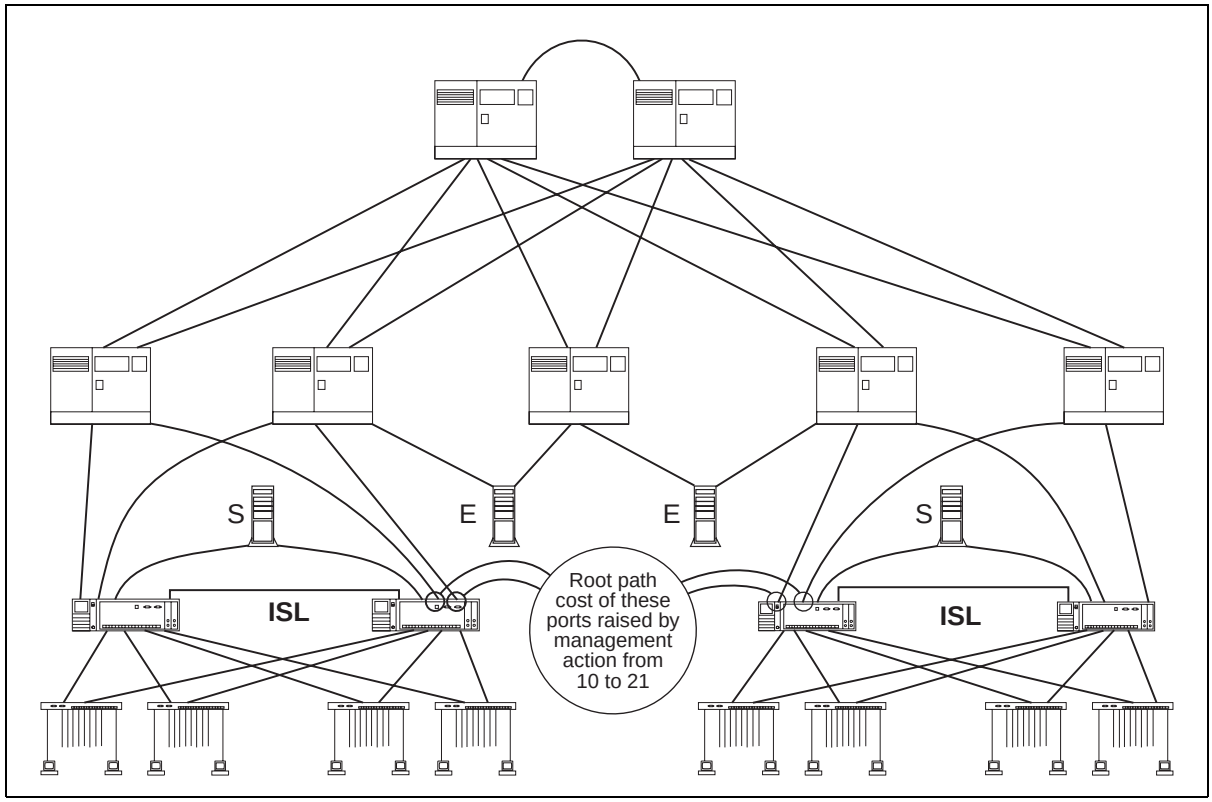
Figure 38 shows a mission-critical campus network in which each site is based on a pair of mated WaveSwitch 100s. The sites are connected by a mission-critical inter-site WaveBus backbone that interconnects the WaveSwitch 100.

This configuration protects against failure of WaveSwitch 100s and WaveBus Hubs.

Each site has a heavily used single fileserver (S, for site) which is used primarily or entirely by that site. There are also a number of fileservers shared equally by all sites in the enterprise (E, for enterprise.)

The network designer has chosen to maximize the capacity of the network by attaching each site's fileserver to the WaveSwitch 100s at the site. WaveBus point-to-point links attach this local fileserver directly to each WaveSwitch 100 at the site. The goal is to localize the traffic to the site, freeing the backbone for intersite traffic and access to the E fileservers.

Figure 38
A mission-critical campus network



If a site fileserver uses its two WaveBus links in load-sharing mode, both WaveSwitch 100s will carry traffic to and from the fileserver. Only one WaveSwitch 100 will have a port in forwarding state to each Ethernet LAN. One WaveSwitch 100 of the mated pair will be the designated bridge for all Ethernet LANs connected to both WaveSwitch 100s of the pair. This means that about half the local fileserver access traffic must be shuttled between the two WaveSwitch 100s on a spanning tree link between the two WaveSwitch 100s on the way between the Ethernets and the fileserver.

Assuming that a design goal of this network is to keep local access traffic for the S fileservers completely off the intersite WaveBus backbone, a 100-Mbps interswitch link (ISL) is required, as shown, to serve as the spanning tree link between the two WaveSwitch 100s. If the redundant WaveBus backbone is the spanning tree link between the WaveSwitch 100s of a mated pair, it will carry local fileserver access traffic, against the wishes of the network designer. But how does the network designer ensure that the spanning tree procedure will choose the ISL, rather than the redundant WaveBus backbone as the spanning tree link between the two WaveSwitch 100s?

There is no default spanning tree that will achieve the goal. Consider any WaveSwitch 100 as the root bridge of a default spanning tree. Then all other WaveSwitch 100s in the network have a root path cost of 10, since they can all be reached by one hop over the 100-Mbps backbone. (Any path to the root through an ISL necessarily has a higher cost.) The ISL in the WaveSwitch 100 pair that includes the root bridge may not form the spanning tree link between the root and its mate, depending on the relative values of the port identifiers of the ISL and backbone connections to the WaveSwitch 100. Every other WaveSwitch 100 in the network must, however, have a root port on the backbone. As a result, for all WaveSwitch 100 pairs, except possibly for the pair that includes the root bridge, the backbone forms the spanning tree link between mated WaveSwitch 100s, contrary to the intention of the network designer.

The goal can be achieved by taking management action to raise the root path cost increment of all backbone connections of one WaveSwitch 100 of each mated pair. The minimum value of the root path cost increment for these ports that will achieve the goal is 21, since the default path cost through one WaveSwitch 100 to its mate over the ISL is 20. The reader can confirm that with the specified port costs, it does not matter which WaveSwitch 100 happens to be the root bridge. Each ISL will form the spanning tree link between the WaveSwitch 100s of each pair, and each WaveSwitch 100 pair will possess one forwarding port connected to the backbone.

Each of the two WaveBus point-to-point links between an S fileserver and a WaveSwitch 100 is seen by the spanning tree procedure as a separate LAN, so both will always be included in the spanning tree.

If, instead of load sharing, the fileserver uses one of the two WaveBus links exclusively while that link is operating correctly, the ISL may not be required, depending on whether the network manager can assign a priority to the fileserver connections. If the fileserver can be instructed to give priority to the WaveBus link connected to the WaveSwitch 100 with the highest spanning tree priority, then, without an ISL, the intersite backbone would be asked to carry traffic between the workgroups at a site and its S fileserver only in the event of a failure of the high priority fileserver connection. This strategy would require the network manager to identify and maintain the relationship between the priority given by the fileserver to its links and the higher bridge identifier of each pair of WaveSwitch 100s—an onerous requirement.

Novell NetWare SFT III in WaveSwitch 100/WaveBus networks

SFT III is a feature of Novell NetWare version 3.11 and later. SFT III provides tolerance for fileserver failure by keeping one server ready to take over from another.

SFT III defines one server as primary, and the other as secondary. The primary server responds to client requests, and keeps the secondary server up to date with all server activity and data. If the primary server fails, the secondary server takes over quickly, usually without loss or interruption of client sessions.

A direct, dedicated, high speed communication link is required between the two servers—the mirrored server link, or MSL. A few companies supply MSL equipment. Plaintree Systems, Inc. makes MSL kits for ISA, EISA, and Micro Channel servers.

Figures 39, 40, 41, 42, and 43 show a variety of networks based on the WaveSwitch 100, WaveBus, and SFT III servers.

Figures 39 and 40 show networks which tolerate faults only in the fileservers. As late as version 4.01 of NetWare, SFT III does not react to media failures which isolate the primary server. These are reasonable configurations because faults in the fileservers are more likely than faults in the other network equipment.

Figure 39
SFT III servers connected to multiple Ethernet networks by a WaveSwitch 100

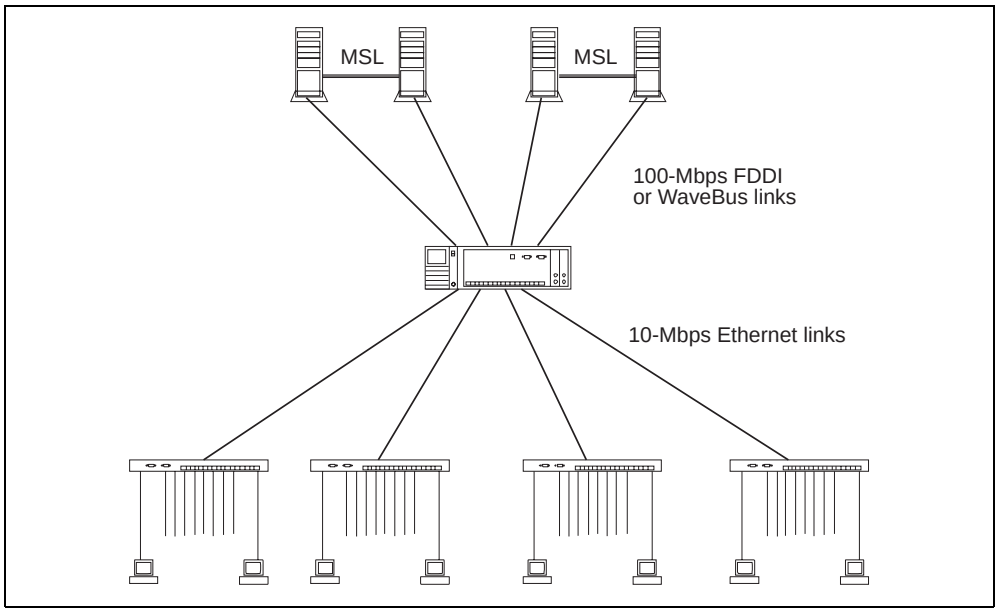
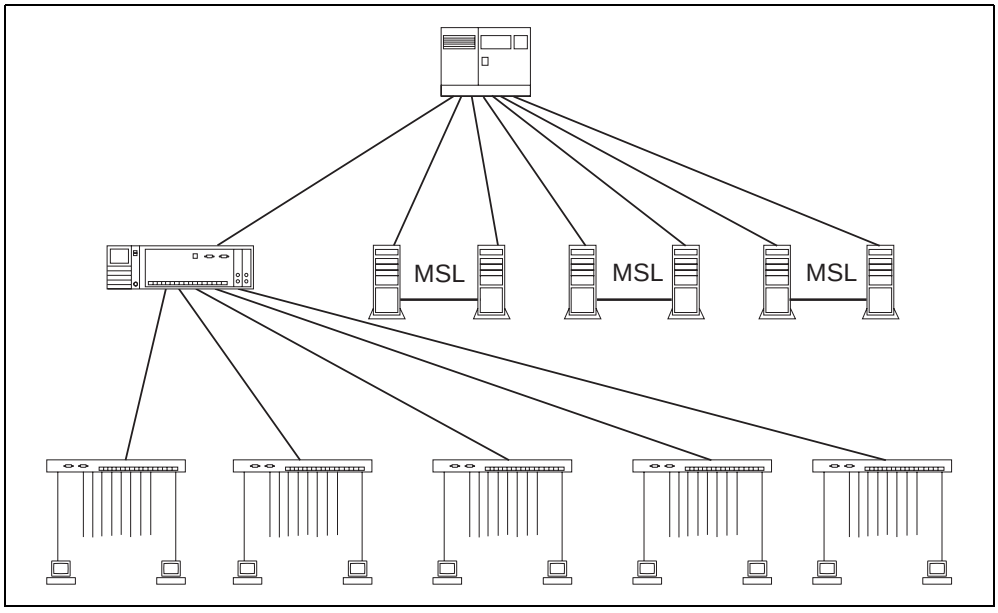


Figure 40
SFT III servers connected to Ethernet networks by WaveBus Hub/WaveSwitch 100

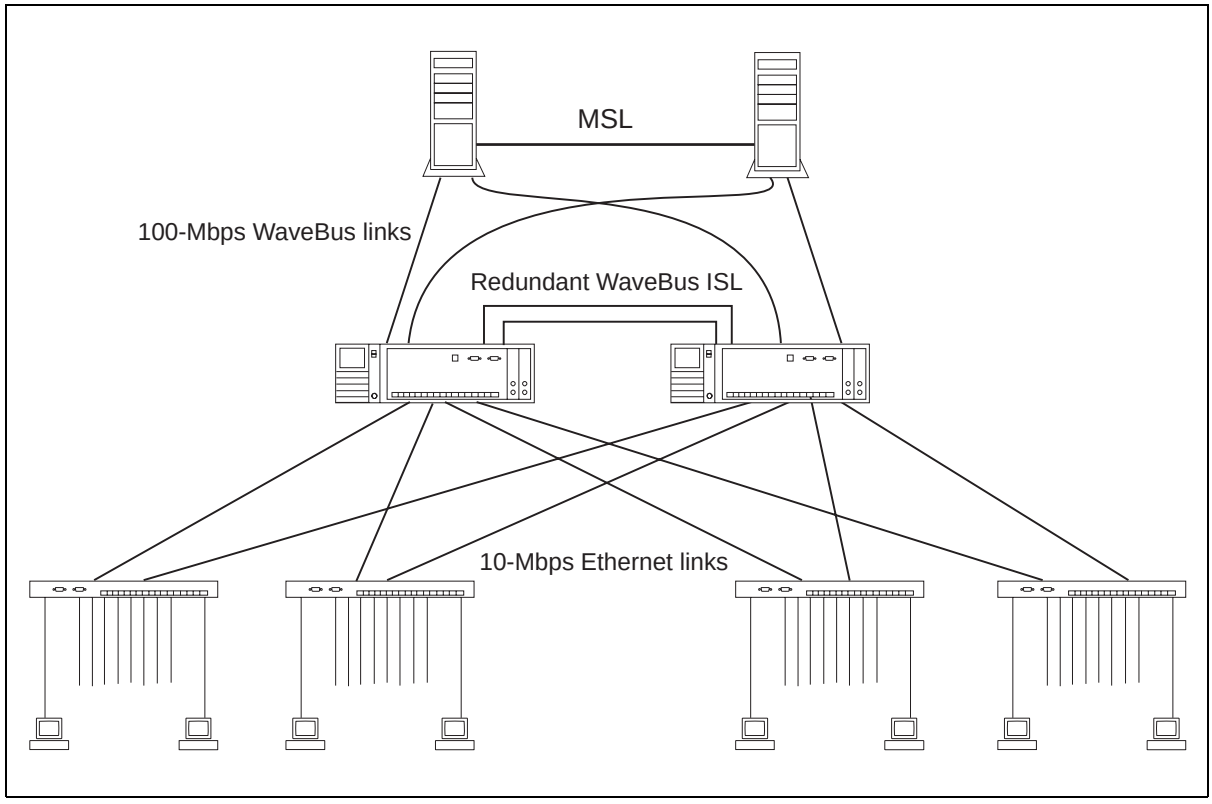


Figures 41, 42, and 43 show networks which tolerate faults in other network equipment as well.

SFT III is designed to cope well with fileserver faults, but makes no attempt to cope with media failures. For example, failure of the medium connecting users to the primary fileserver does not cause the primary to give up in favor of the secondary fileserver, leaving the users unserved. To add tolerance of LAN connection faults to SFT III servers requires dual connections between each servers and the single high speed LAN connecting the servers to the WaveSwitch 100. It is not worth duplicating WaveBus Hubs or WaveSwitch 100s unless four WaveBus connections are provided to each SFT III server pair. NetWare does not support such multiple connections, but the software from NSI described on page 23 does, and works well with SFT III. The fileservers in the configurations shown in Figures 41, 42, and 43 are assumed to have such multiconnection capabilities.

Figure 41 shows two pairs of SFT III servers connected by a WaveSwitch 100 to a number of Ethernet networks. This configuration allows up to two pairs of SFT III servers to achieve full performance. The two SFT III fileserver pairs shown require four WaveBus connections to the WaveSwitch 100, which is the maximum number possible. Although faults

Figure 41
A network with one pair of SFT III servers and redundant WaveSwitch 100s



in the WaveSwitch 100 will cause loss of service, such faults are less likely than faults in the fileservers, and the data stored on the fileservers may have great value. This configuration is practical, inexpensive, and suitable for many applications.

Figure 40 shows three pairs of SFT III servers connected by a WaveBus Hub and a WaveSwitch 100 to multiple Ethernet networks. The fanout capacity of the WaveBus Hub permits as many as seven SFT III server pairs to be connected in this way. The fault tolerance of the fileservers is compatible with the fault intolerance of the single WaveBus Hub and single WaveSwitch 100 because faults in these singly provisioned devices are less likely than faults in the fileservers.

Figure 41 shows a fully redundant system of one pair of SFT III servers and two WaveSwitch 100s. The configuration provides protection against media and WaveSwitch 100 faults, as well as fileserver faults. This network is more expensive per server pair than those of Figures 39 and 40 because of the large number of high speed interfaces required—six 100-Mbps WaveBus NICs (assuming the MSL is Plaintree MSL, made by Plaintree Systems, Inc.) for the fileservers, and eight 100-Mbps WaveBus ports for the WaveSwitch 100. It may, nonetheless be suitable for some applications. For an explanation of the redundant ISL, see page 37.

Figure 42
Three pairs of SFT III servers connected to a WaveSwitch 100

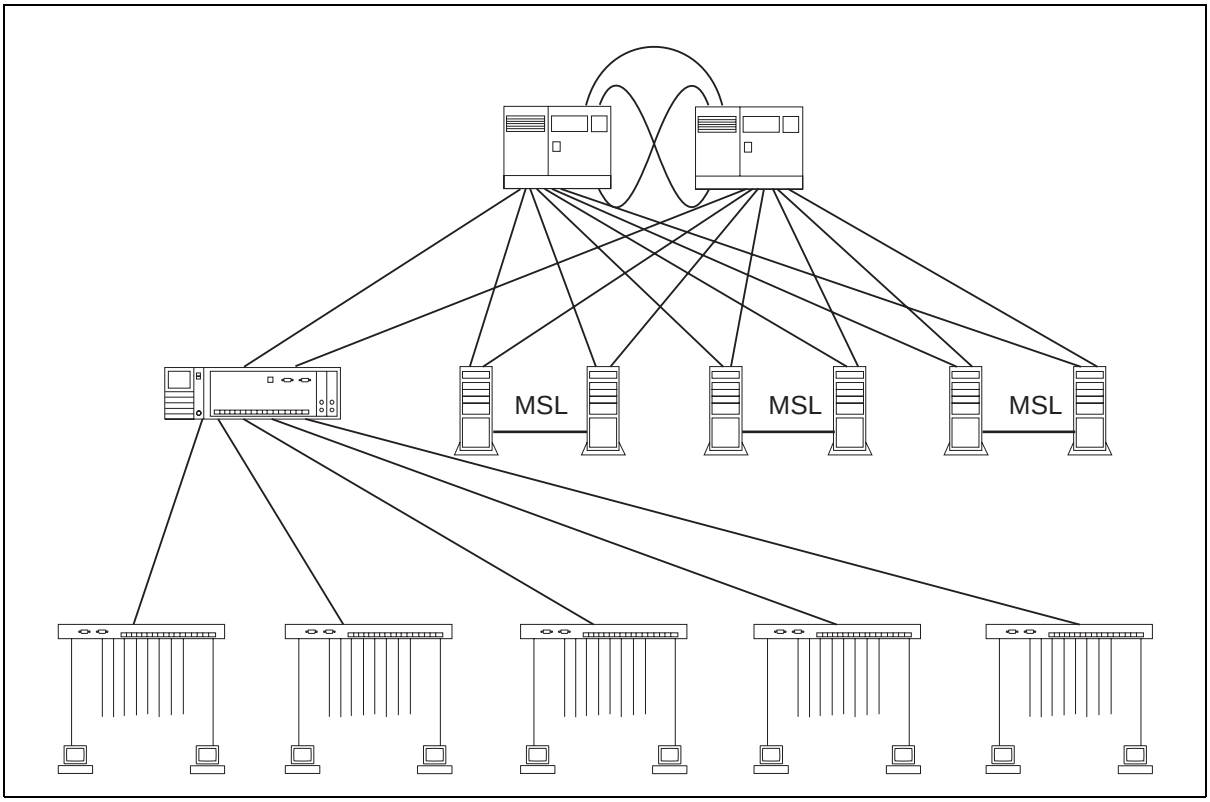
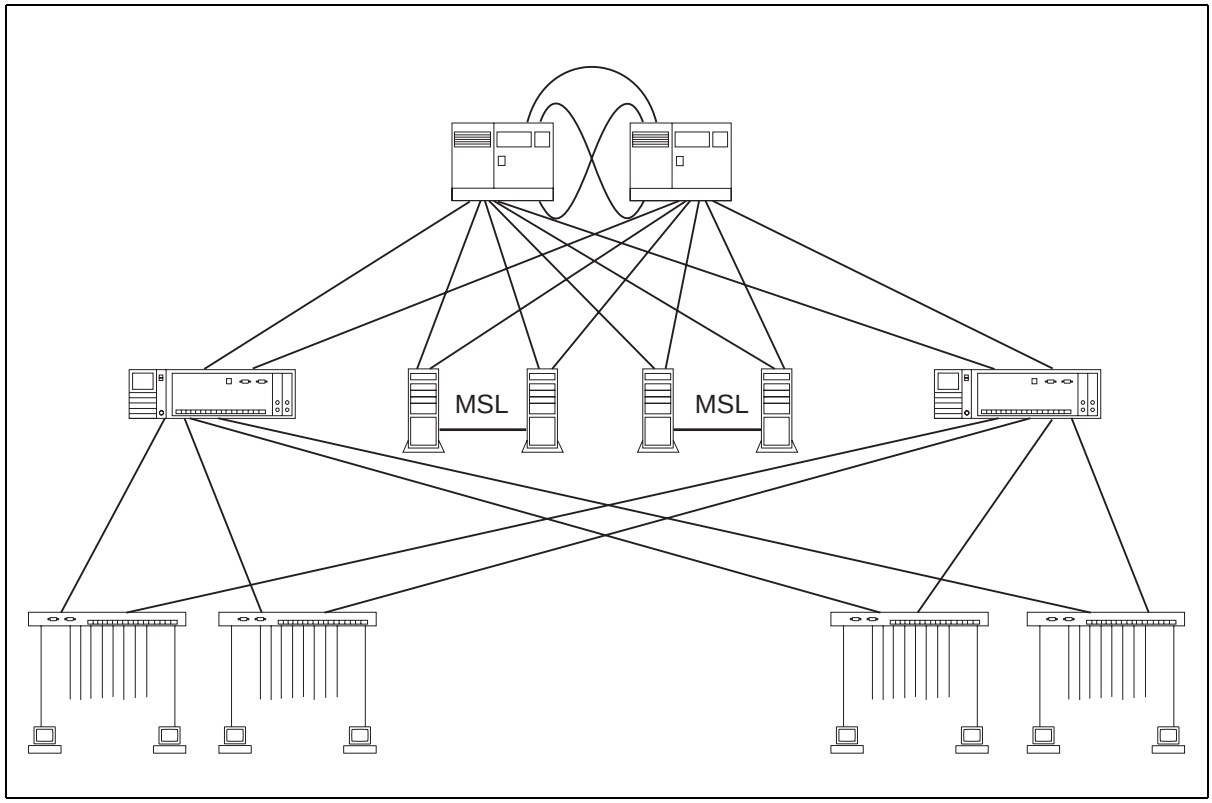


Figure 42 shows a redundant WaveBus network connecting three SFT III server pairs to one WaveSwitch 100. This network protects against faults in the redundant WaveBus network, as well as faults in the fileservers. The Ethernet networks are not protected against WaveSwitch 100 failures or failures of the Ethernet connections to the WaveSwitch 100. This degree of redundancy compensates nicely for the lack of notice taken by SFT III of faults in the media connected to the servers.

Each SFT III fileserver pair occupies two down ports on each mated WaveBus Hub; each WaveSwitch 100 occupies one down port on each mated WaveBus Hub. The 15 ports available on each mated WaveBus Hub can be split in any way between WaveSwitch 100s and SFT III server pairs.

Figure 43 shows a redundant WaveBus network connecting three SFT III server pairs to a pair of mated WaveSwitch 100s. This network protects against faults in the WaveBus network, in the WaveSwitch 100s, in the Ethernet connections to the WaveSwitch 100s, and in the fileservers.

Figure 43
SFT III servers connected to mated WaveSwitch 100s through a redundant WaveBus network



References

[1] IEEE Standard 802.1D-1990, *Media Access Control (MAC) Bridges*, The Institute of Electrical and Electronic Engineers, Inc. 345 East 47th Street, NY 10017-2394, U.S.A., 1991, 176 pages, ISBN 1-55937-055-6.

[2] Perlman, Radia, *Interconnections: Bridges and Routers*, Addison-Wesley Publishing Company, Inc., 1992, One Jacob Way, Reading, Massachusetts 01867, 389 pages, ISBN 0-201-56332-0.

[3] Mirchandani, S., and Khanna, R. (editors), *FDDI Technology and Applications*, John Wiley & Sons, Inc. New York, 1993, 359 pages, ISBN 0-471-55896-6.

[4] Digital Equipment Corporation, *A Primer on FDDI: Fiber Distributed Data Interface*, version 2.00, June 1992.

[5] American National Standards Institute (ANSI), *FDDI Station Management*, Document number X3.229 Rev. 7.3, Available from Global Engineering Documents, Irvine CA, (800) 854 7179

[6] Comer, Douglas E., *Internetworking with TCP/IP*, *Volume I, Principles, Protocols, and Architecture*, Prentice Hall, Englewood Cliffs, New Jersey 07632, 1991, 547 pages, ISBN 0-13-468505-9.

WaveSwitch 100
Ethernet Switch
Configuration Guide

Address comments to:

Documentation Manager
Plaintree Systems, Inc.
59 Iber Road
Stittsville, ON Canada K2S 1E7

Telephone: 1 800 461 0062
1 613 831 8300
Fax: 1 613 831 3283

Printed in Canada

PLAINTREE
SYSTEMS