NEAR EAST UNIVERSITY

Faculty of Engineering

Department of Computer Engineering

THE TECHNICAL TOPOLOGY OF A SAN

Graduation Project
COM- 400

Student: Kayied Hamouda (20033370)

Supervisor: Mr. jamal fathi

Nicosia - 2007

ACKNOWLEDGMENTS

My utmost thanks to my Lord Allah that i could complete my graduation project.

I could not have prepared this project without the generous help of my supervisor, colleaques, friends, and family.

First, I would like to thank my supervisor Mr. Jamal Fathi for his invaluable advice, and belief in my work and myself over all the courses of this Degree. Mr. Jamal supplied the warmth, enthusiasm, and clarity of judgement that every student hopes for. Going beyond the limited role of literary agent, he provided valuable advice at each stage of the preparation of this project.

I would like to express my gratitude to Assoc. Prof. Dr. Adnan Khashman for him because he provided valuable advice at each stage of the preparation of this project also.

I will never forget the help that i got from this university for continueing my education especially from Prof. Dr Şenol Bektaş, so my regards and my love
to him.

My deppest thanks are to my family. I could never have prepared this project without the encouragement and support of my parents, brothers, and sister.

The root of this success lies under the most affectionate wish of my loving FATHER. I am grateful to him to assist me to grow in knowledge. I salute you, my father.

I would also like to thank all my friends for their help and for their patience also for their support, Mr Tayseer Alshanableh.

Dedicated to my Mum

CONTENTS

1. INTRODUCTION

1.1 Introduction to Wireless Networks Moments

Wireless networks are an emerging new technology that will allow users to access information and services electronically, regardless of their geographic position. Wireless networks can be classified in two types: - infrastructured network and infrastructure less (ad hoc) networks. Infrastructured network consists of a network with fixed and wired gateways. A mobile host communicates with a bridge in the network (called base station) within its communication radius. The mobile unit can move geographically while it is communicating. When it goes out of range of one base station, it connects with new base station and starts communicating through it. This is called handoff. In this approach the base stations are fixed.

In contrast to infrastructure based networks, in ad hoc networks all nodes are mobile and can be connected dynamically in an arbitrary manner. All nodes of these networks behave as routers and take part in discovery and maintenance of routes to other nodes in the network. Ad hoc networks are very useful in emergency search-and-rescue operations, meetings or conventions in which persons wish to quickly share information, and data acquisition operations in inhospitable terrain.

Figure 1.1 shows the different between Ad hoc networks and infrastructure wireless networks.

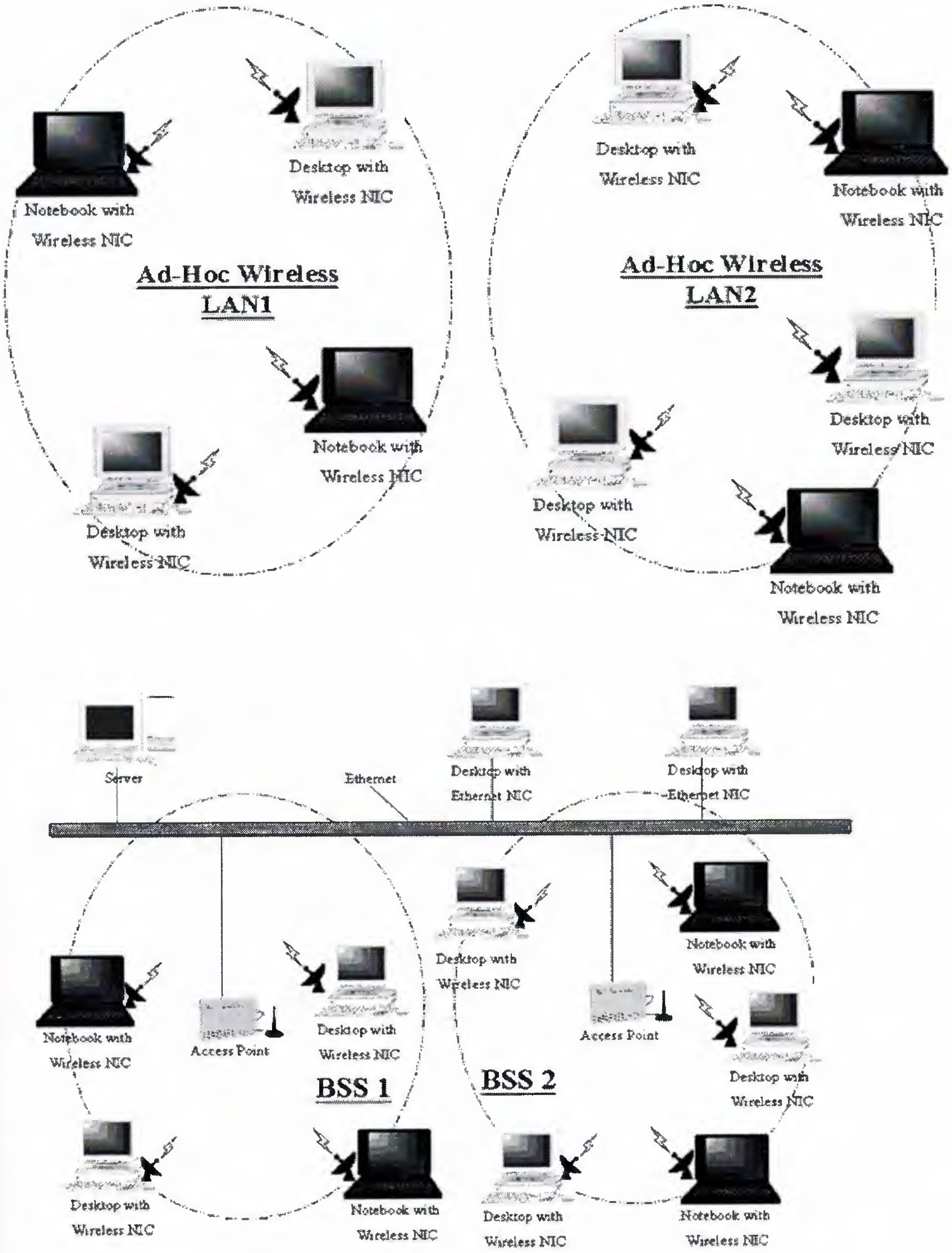

Notebook with Wireless NIC
Desktop with Wireless NIC
Ad-Hoc Wireless LAN1
Notebook with Wireless NIC
Desktop with Wireless NIC
Desktop with Wireless NIC
Notebook with Wireless NIC
Ad-Hoc Wireless LAN2
Desktop with Wireless NIC
Desktop with Wireless NIC
Notebook with Wireless NIC
Server
Ethernet
Desktop with Ethernet NIC
Desktop with Ethernet NIC
Notebook with Wireless NIC
Access Point
Desktop with Wireless NIC
BSS 1
Desktop with Wireless NIC
Notebook with Wireless NIC
Desktop with Wireless NIC
Notebook with Wireless NIC
Access Point
Desktop with Wireless NIC
BSS 2
Desktop with Wireless NIC
Notebook with Wireless NIC

Figure1.1 Ad Hoc Networks and Infrastructure Wireless Network

1.2 Introduction to GPS

The Global Positioning System (GPS) is a radio based navigation system that gives three dimensional coverage of the Earth 24 hours a day in any weather conditions. The satellites orbit the Earth every 12 hours at approximately 12,600 miles above the Earth. The GPS system is passive, meaning that the satellites continuously transmit information towards the Earth. If someone has a GPS receiver they can receive the signal at no cost.

If the signals from three or more satellites are received, simple triangulation will make it possible to determine the location of the user. Up to 30 GPS satellites fly, mostly in highly inclined (polar) orbits this means that there will be between four and eight of them reasonably high in the sky above any site on the Earth at any time. The GPS satellites around the earth are shown in figure1.2

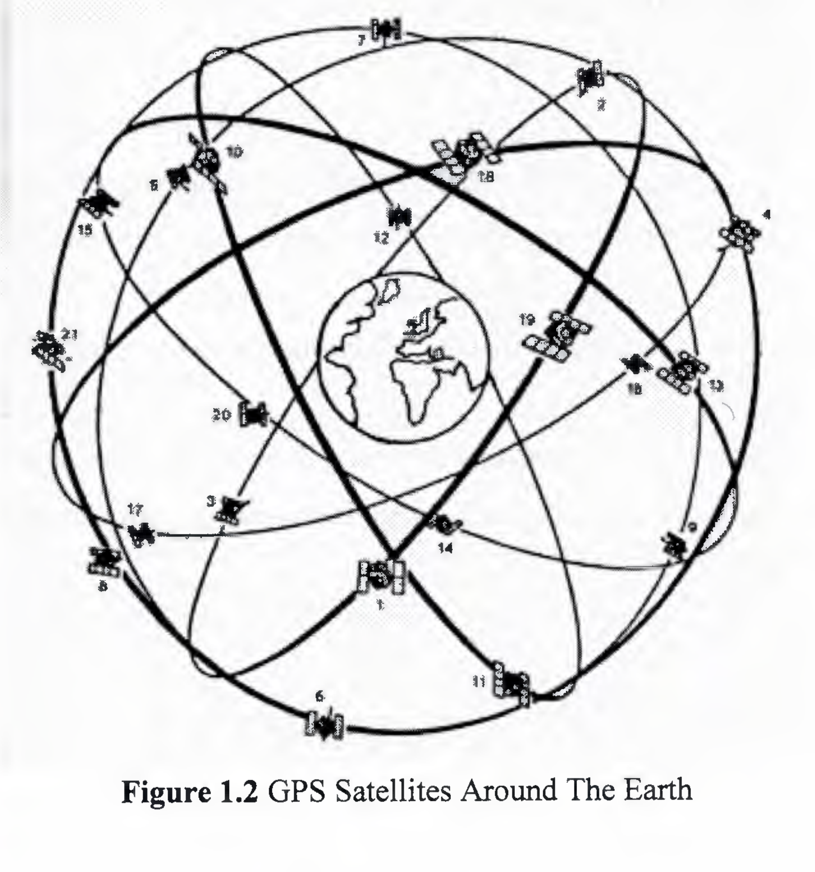


Figure 1.2 GPS Satellites Around The Earth

The achievable accuracy depends on the status of the user. For military purposes (and some specific civil ones), one meter or better in all three coordinates (longitude, latitude, altitude) can be reached. For common civil users, the full accuracy of the coded satellite signal cannot be exploited, but it is still possible to reach an accuracy of about 15 meters in the best cases.

The applications of the Global Positioning System fall into five categories: location, navigation, timing, mapping, and tracking. Each category contains uses for the military, industry, transportation and science.

The technology of the Global Positioning System is allowing for huge changes in society. The applications using GPS are constantly growing. The cost of the receivers is dropping while at the same time the accuracy of the system is improving.

1.3 Introduction to GIS

GIS stands for Geographic Information Systems, GIS is a mapping technology that allows the user to create and interact with a variety of maps and data sources. GIS integrates databases with georeferenced spatial data (maps tied to specific known locations). In other words GIS allows the user to create visual displays of tabular information. [1]

A GIS can be regarded as a computer system that can capture, store, query, analyze and display geographical information. Figure 1.3 shows the GIS functions.

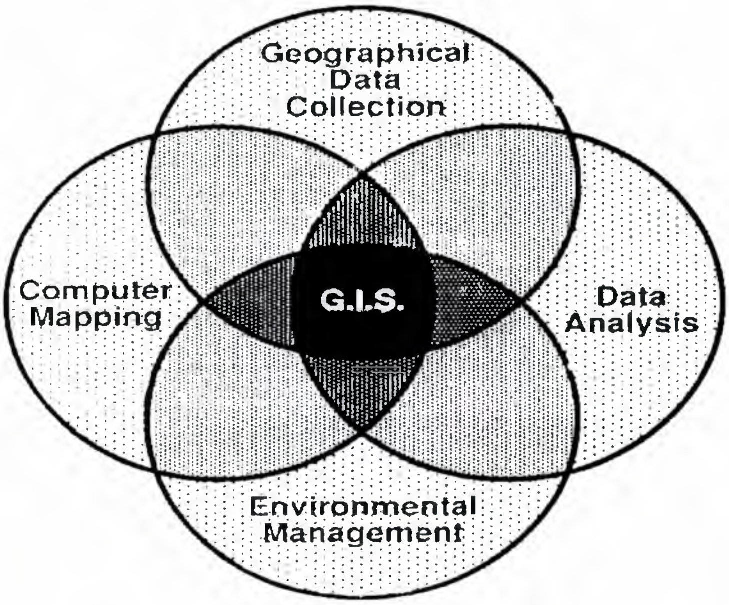


Figure 1.3 GIS Functions

Information in a GIS can be conceived as a series of layers, linked together by their common geographical framework, figure 1.4 show example of GIS layers.

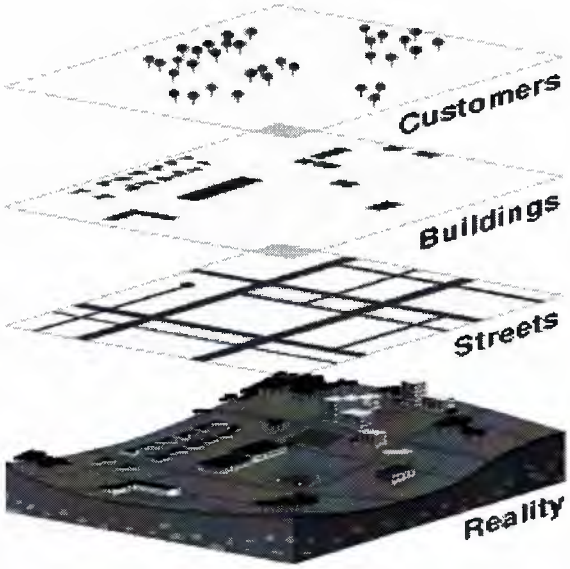


Figure 1.4 Example of GIS Layers

The data in a GIS is stored in one of two different ways, each representing a separate view or model of the real world. The vector model represents geographical reality as a series of discrete objects or features, classified as points, lines or areas (polygons). The geographical co-ordinates describing the locations of these features are stored in the computer database which lies at the heart of the GIS. Linked to these features will be associated attribute information recording their characteristics. In the raster model a regular grid of cells, or pixels, is used to encode the features found on the earth's surface. Each pixel has a number associated with it representing the value of a geographical phenomenon, such as terrain elevation or soil type.

Current commercial GIS products use both the vector and raster models although transferring data from one format to the other is not always straightforward. They both have advantages and drawbacks. The vector model is good for querying database records associated with distinctive geographical features, while the raster model is suited for complex overlay operations (for example in examining vegetation change over time).

GIS now provides the ability to integrate information, solve problems, develop effective solutions and visualize scenarios in a way that was simply not possible even a decade ago.

1.4 Problem Identification and Scope of Work

In the fast growing computer network market the demand for wireless computer network increases. An even more extreme case is one in which the network is wireless and mobile at the same time. Among the possibilities are:

- Military vehicles on a battlefield with no existing infrastructure.
- A fleet of ships at sea.
- Emergency workers at an earthquake site that destroyed the infrastructure.
- A gathering of people with notebook computers.

In all these cases, and others, each node consists of a router and a host, usually on the same computer. Networks of nodes that just happen to be near each other are called ad hoc networks.

What makes routing in ad hoc networks difficult from wired networks is that all the usual rules about fixed topologies, fixed and known neighbors, fixed relationship between IP address and location, and more are suddenly tossed out the window. Routers can come and go or appear in new places at the drop of a bit. With a wired network, if a router has a valid path to some destination, that path continues to be valid indefinitely (barring a failure somewhere in the system). With an ad hoc network, the topology may be changing all the time, so desirability and even validity of paths can change spontaneously, without warning.

These circumstances make routing in ad hoc networks more difficult from routing in their fixed counterparts.

A variety of routing algorithms for ad hoc networks have been proposed. The scope of this study is to implement simulation for ad hoc wireless networks routing utilizing GPS and GIS information.

1.5 Summary

In this chapter we give a quick introduction to the wireless technology, global positing system (GPS) and the geographic information systems (GIS), which we will focus on in solving our problem.

2. AN OVERVIEW OF COMPUER NETWORKS

2.1 Introduction

Networks are an extremely important aspect of Computer Science in today's computing environment. Currently, 80% of all computers function in some way on a network. Within the next decade, virtually every electronic gadget will be networked in some fashion. Your cars will uplink with a satellite to find directions. Your TV will give out cookies as to your viewing habits, and what commercials should be shown during your favorite programming. These applications are not far-fetched, and the implementation is imminent. Everyone who works with computers will be faced with some, if not many aspects of networks. Having a basic understanding of how they work and where they came from is essential. [3]

2.2 Computer Networks

Computer network means an interconnected collection of computers such that they can:

- Exchange information
- Share resources with each other:

To make all programs, data and equipment available to anyone on the network without regard to the physical location of the resource.

- Increase computation power:

Since resources are located at specific servers instead of each computer, it saves more storage space and allows for ease in data backup.

- Maintain a high reliability:

If some computers go down, others may be able to take over its work. It is very important for example for military and air traffic control systems to continue operating when some hardware problems are encountered.

2.3 Classification of Computer Networks

Computer network can be classified by the geographical coverage of the collection of computers, in the coming sections we discuss these classes.

2.3.1 Local Area Networks (LAN)

A LAN is a high-speed data network that covers a relatively small geographic area. It typically connects workstations, personal computers, printers, servers, and other devices. LANs offer computer users many advantages, including shared access to devices and applications, file exchange between connected users, and communication between users via electronic mail and other applications.

2.3.1.1 LAN Topology

LAN topologies define the manner in which network devices are organized. Four common LAN topologies exist:

(a) Star:

A star topology is a LAN architecture in which the endpoints on a network are connected to a common central switching element (hub or switch), by dedicated links. Two stations that want to communicate will have to setup a dedicated pair via the central switch.

The primary advantage of this type of network is reliability - if one 'point-to-point' segment has a break, it will only affect the nodes on that link; other computer users on the network will continue to operate as if that segment were non-existent.

(b) Ring:

A ring topology is a LAN architecture in which all devices are connected to one another in the shape of a closed loop, so that each device is connected directly to two other devices, one on either side of it. The data is injected into the ring and circulated around the ring until it reaches the destination.

(c) Bus:

A bus topology is a linear LAN architecture in which transmissions from network stations propagate the length of the medium and are received by all other stations. Many nodes can tap into the bus and begin communication with all other nodes on that cable segment. A break anywhere in the cable will usually cause the entire segment to be inoperable until the break is repaired.

(d) Tree:

A tree topology is a LAN architecture that is identical to the bus topology, except that branches with multiple nodes are possible in this case.

Figure 2.1 shows the four LAN Topology:

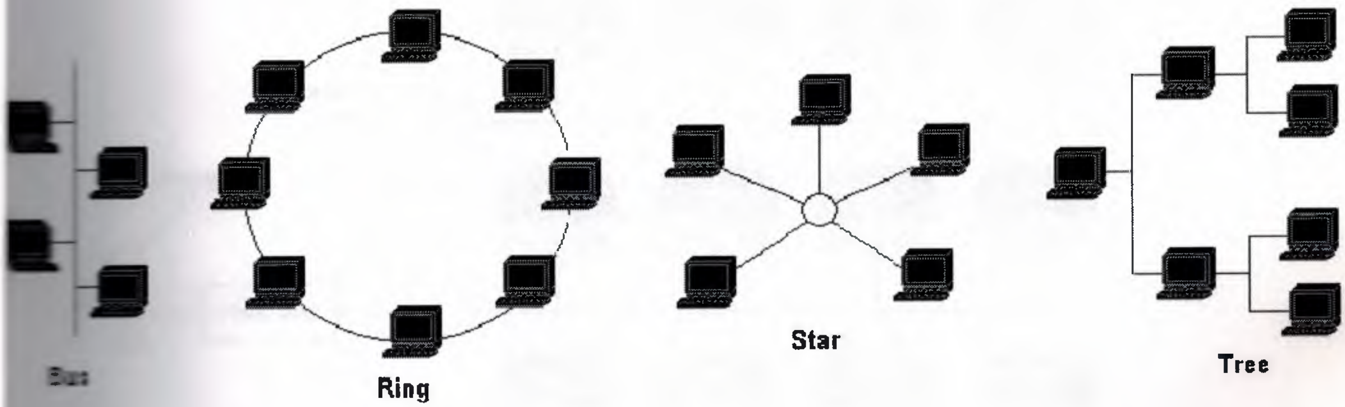


Figure 2.1 The Four LAN Topology

2.3.2 Wide Area Network (WAN)

A WAN is a communications network that covers a wide geographic area, such as state or country. As shown in figure2.2

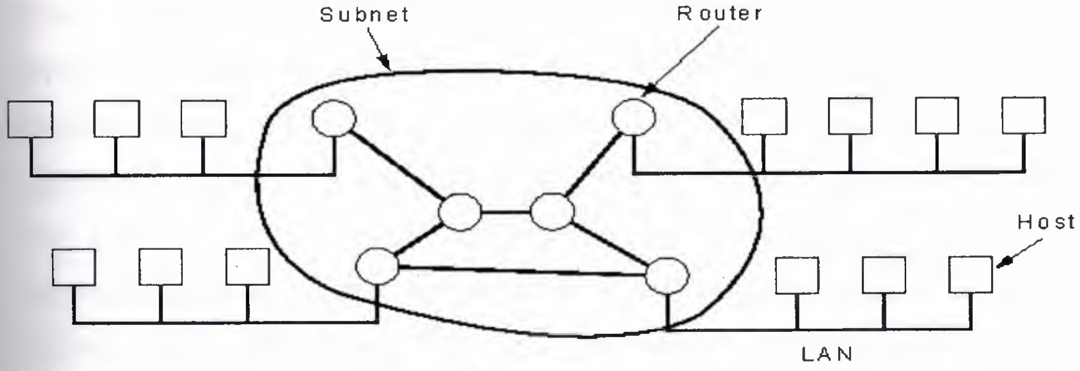


Figure2.2 WAN Network

2.3.3 Metropolitan Area Networks (MAN)

Metropolitan area networks covers up to a city (<= 10 km), Example: Cable TV network. The metropolitan area network is shown in figure 2.3

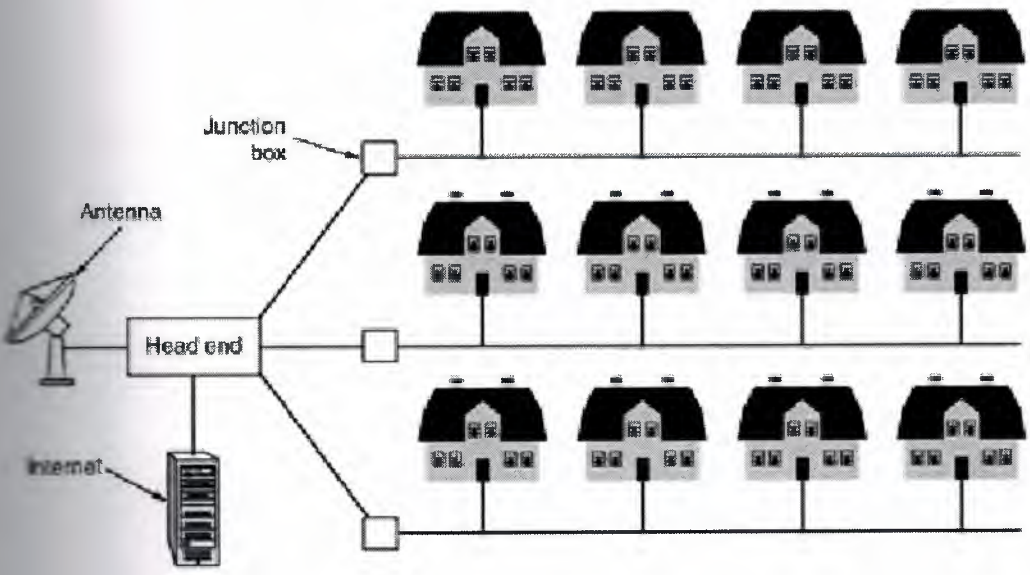


Figure 2.3 MAN Network

2.3.4 Wireless Networks

Wireless technologies, in the simplest sense, enable one or more devices to communicate without physical connections – without requiring network cabling. Wireless technologies use radio transmissions as the means for transmitting data, whereas wired technologies use cables. Wireless technologies range from complex systems, such as WLANs and cell phones, to simple devices such as wireless headphones, microphones, and other devices that do not process or store information. They also include infrared (IR) devices such as remote controls, some cordless computer keyboards and mice, and wireless hi-fi stereo headsets, all of which require a direct line of sight between the transmitter and the receiver to close the link. Wireless technology aims to provide users access to information anywhere – it allows mobility.

2.4 Reference Models

In the next two sections we will discuss two important networks architectures, the OSI reference model and the TCP/IP reference model.

2.4.1 OSI (Open System Interconnection) Reference Model

The OSI model is shown is figure2.4

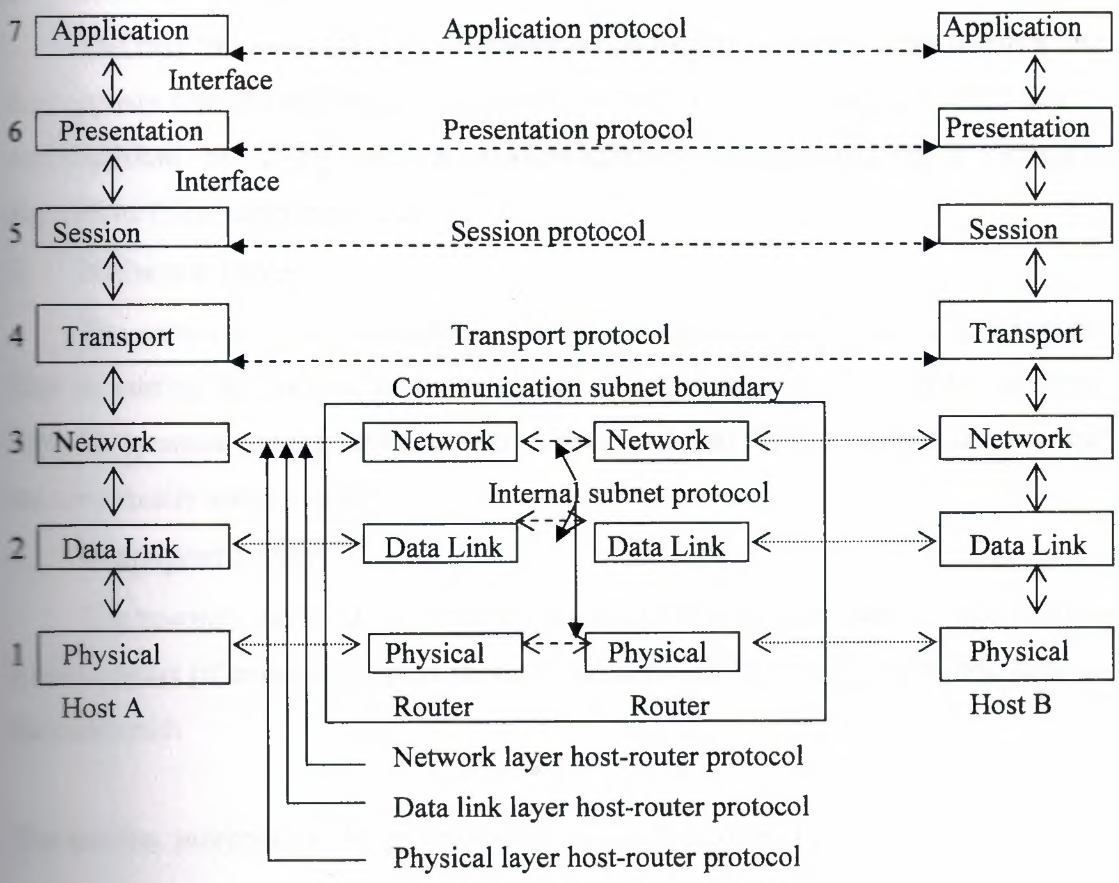


Figure 2.4 OSI Model

This model deals with connecting open systems – that is, systems that are open for communication with other systems. Note that the OSI model itself is not a network architecture because it does not specify the exact services & protocols to be used in each layer. However, common standards were produced by ISO (International Standards Organization) for each layer:

1. Physical Layer

The physical layer is concerned with transmitting raw bits over communication channel.

2. Data Link Layer

The data link layer takes a raw transmission facility and transforms it into a line that appears free of undetected transmission errors to the network layer. This task is accomplished by using data & acknowledgment frames and error detection algorithms (like humming code).

3. Network Layer

The network layer is concerned with controlling the operation of the subnet. That is routing the packets from the source to destination. Routes can be based on static or dynamic routing tables as will be reviewed later. *(This layer is the one that we are actually interested in)*.

4. Transport Layer

The transport layer basic function is to accept data from the session layer derive it into packet (if necessary), pass these to the network layer and restore the data on the other end.

The session, presentation & application layers are less interesting for us.

2.4.2 TCP/IP Reference Model

This model was developed on the base of first computer networks and has only four layers.

1. Internet Layer

The internet layer is the linchpin that holds the whole architecture together. It allows hosts to inject their packets into any network and have them travel independently to their destination. This layer defines official protocol called IP.

2. Transport Layer

The transport layer lies above the internet layer and its functionality is much alike to the same layer in OSI model – it allows peer entities on the source & destination hosts to carry on a conversation (2 end-to-end protocols were defined here: TCP & UDP).

There are tow more layers **application & host-to-network** that less interest us (the host-to-network layer plays minor part in TCP/IP protocol, steel being significant enough by itself).

2.4.3 OSI verses TCP/IP

- The OSI model supports both connectionless & connection-oriented communication in the network layer, but only connection-oriented communication in the transport layer.

- The TCP\IP model has only one mode in the network layer – connectionless, but supports both modes in transport layer, giving the user a choice.

Generally the OSI model has proven to be exceptionally useful for discussing computer networks, but OSI protocols did not become popular .The reverse is true of TCP/IP: the model is practically nonexistent , but the protocols are widely used.

2.5 Routing

Routing is the act of moving information across an internet work from a source to a destination. Along the way, at least one intermediate node typically is encountered. Routing is often contrasted with bridging, which might seem to accomplish precisely the same thing to the casual observer. The primary difference between the two is that bridging occurs at Layer 2 (the link layer) of the OSI reference model, whereas routing occurs at Layer 3 (the network layer). This distinction provides routing and bridging with different information to use in the process of moving information from source to destination, so the two functions accomplish their tasks in different ways.

2.5.1 Routing Algorithms

The main function of the network layer is routing packets from source to destination. The algorithms that choose the routes are a major area of network layer design.

The routing algorithm is that part of the network layer software responsible for deciding which output line an incoming packet should be transmitted on.

Routing algorithms can be grouped into two major classes: nonadaptive and adaptive:

1. Nonadaptive

Algorithms do not base their routing decisions on measurements or estimates of the current traffic and topology. Instead, the choice of the route to use to get from I to J is computed in advance, of-line, and downloaded to the routers when the network is booted. This procedure is sometimes called static routing.

2. Adaptive

Algorithms, in contrast, change their routing decisions to reflect changes in the topology, and usually the traffic as well. Adaptive algorithms differ in where they get their information, when they change the routes, and what metric is used for optimization. They are also called dynamic.

2.5.2 Design Goals

Routing algorithms often have one or more of the following design goals:

1. **Optimality**

Optimality refers to the capability of the routing algorithm to select the best route, which depends on the metrics and metric weightings used to make the calculation. For example, one routing algorithm may use a number of hops and delays, but it may weigh delay more heavily in the calculation. Naturally, routing protocols must define their metric calculation algorithms strictly.

2. **Simplicity and low overhead**

Routing algorithms also are designed to be as simple as possible. In other words, the routing algorithm must offer its functionality efficiently, with a minimum of software and utilization overhead. Efficiency is particularly important when the software implementing the routing algorithm must run on a computer with limited physical resources.

3. **Robustness and stability**

Routing algorithms must be robust, which means that they should perform correctly in the face of unusual or unforeseen circumstances, such as hardware failures, high load conditions, and incorrect implementations. Because routers are located at network junction points, they can cause considerable problems when they fail. The best routing algorithms are often those that have withstood the test of time and that have proven stable under a variety of network conditions.

4. Rapid convergence

In addition, routing algorithms must converge rapidly. Convergence is the process of agreement, by all routers, on optimal routes. When a network event causes routes to either go down or become available, routers distribute routing update messages that permeate networks, stimulating recalculation of optimal routes and eventually causing all routers to agree on these routes. Routing algorithms that converge slowly can cause routing loops or network outages.

5. Flexibility

Routing algorithms should also be flexible, which means that they should quickly and accurately adapt to a variety of network circumstances. Assume, for example, that a network segment has gone down. As many routing algorithms become aware of the problem, they will quickly select the next-best path for all routes normally using that segment. Routing algorithms can be programmed to adapt to changes in network bandwidth, router queue size, and network delay, among other variables.

Routing Metrics

Routing tables contain information used by switching software to select the best route. But how, specifically, are routing tables built? What is the specific nature of the information that they contain? How do routing algorithms determine that one route is preferable to others.

Routing algorithms have used many different metrics to determine the best route. Sophisticated routing algorithms can base route selection on multiple metrics, combining them in a single (hybrid) metric. All the following metrics have been used:

1. Path length

Path length is the most common routing metric. Some routing protocols allow network administrators to assign arbitrary costs to each network link. In this case, path length is the sum of the costs associated with each link traversed. Other routing protocols define hop count, a metric that specifies the number of passes through internetworking products, such as routers, that a packet must take en route from a source to a destination.

2. Reliability

Reliability, in the context of routing algorithms, refers to the dependability (usually described in terms of the bit-error rate) of each network link. Some network links might go down more often than others. After a network fails, certain network links might be repaired more easily or more quickly than other links. Any reliability factors can be taken into account in the assignment of the reliability ratings, which are arbitrary numeric values usually assigned to network links by network administrators.

3. Delay

Routing delay refers to the length of time required to move a packet from source to destination through the internetwork. Delay depends on many factors, including the bandwidth of intermediate network links, the port queues at each router along the way, network congestion on all intermediate network links, and the physical distance to be traveled. Because delay is a conglomeration of several important variables, it is a common and useful metric.

4. Bandwidth

Bandwidth refers to the available traffic capacity of a link. All other things being equal, a 10-Mbps Ethernet link would be preferable to a 64-kbps leased line. Although bandwidth is a rating of the maximum attainable throughput on a link, routes through links with greater bandwidth do not necessarily provide better routes than routes through slower links. For example, if a faster link is busier, the actual time required to send a packet to the destination could be greater.

5. Load

Load refers to the degree to which a network resource, such as a router, is busy. Load can be calculated in a variety of ways, including CPU utilization and packets processed per second. Monitoring these parameters on a continual basis can be resource-intensive itself.

6. Communication cost

Communication cost is another important metric, especially because some companies may not care about performance as much as they care about operating expenditures. Although line delay may be longer, they will send packets over their own lines rather than through the public lines that cost money for usage time.

2.6 Summary

Computer networks can be used for numerous services, both for companies and for individuals. Networks can be divided up into LANs, MANs and WANs, with there own characteristics, technologies and speeds. LANs cover a building and operate at high speeds. MANs cover a city, for example, the cable television system, which is now used by many people to access the internet. WANs cover a country or continent. Wireless networks are becoming extremely popular, especially wireless LANs.

3. INTEGRATION OF GPS AND GIS IN ROUTING PROTOCOL

3.1Introduction

In the previous chapter we show how the network determines the route but still we didn't make the decision which route to choose, in this chapter we will discuss ideas for integrating GPS and GIS technique in route decision in ad hoc networks.

3.2Integration of GPS in Routing Protocol

In this section we will discuss integrating the GPS technique in routing protocol in order to determine the distance between nodes which we will use as a route metric. We assume that each node has a GPS receiver to provide the node with X, Y and Z coordinates to build the coordinate table.

3.2.1 Building the Tables

In addition to the routing table shown in table 4.1, for each node we have a coordinate table in which we store the coordinates of each node and the distance to each destination node. The coordinate table is shown in table 3.1.

Table 3.1 Coordinate Table For Each Node

Destination	Coordinate			Distance	Unit Distance
	X	Y	Z		

When new node enters, it will construct a hello message and broadcasts it, the form of hello message will be as shown in figure 3.1.

Hello	Node Address	New entered Node Coordinate		
		X	Y	Z

Figure 3.1 Hello Message

When Hello Message received by any node in the range of the sender, the received node will take the Node address from the Hello Message and put it in the Destination address field in the coordinate table and in the Destination address field in the routing table, and put the Next Hop for that destination equal to the Node address in the Hello Message, but if the node address already exists in the route table the node just will change the coordinate in coordinate table with the new one. Also it will take the coordinates in this message and put them in the coordinate field in the coordinate table for that destination.

Now the node will calculate the distance between the receiver of the Hello Message and the sender of the Hello Message and put it in the Distance field for that destination and calculate the distance in route unit and put it in unit distance field, where distance and unit distance are calculated as shown below:

$$\text{Distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \tag{3.1}$$

$$\text{Unit Distance} = \text{Distance} / K_{(\text{meter})} \tag{3.2}$$

Where K constant for the network in meter

In this case each K meter will represent one unit distance, and this distance is not the real (geographic) distance between nodes, its called the communication distance.

Node in ad hoc network can move as we mention early, so the sender of the Hello Message will save the coordinate which he sends in the Hello Message in a coordinate History table and compares it with its coordinate from the GPS receiver as node moves. If the difference is greater than one unit route the node will send the Hello Message again containing the new coordinate.

The received node will construct now a Reply Message for the Hello Message and sends it back to the sender of the Hello Message; the Reply Message will be as shown in figure 3.2 and Figure 3.3 shows how this process done

ess of e msg nder	Address of the sender of reply msg	Coordinate of sender of Reply msg			Neighbor Address	Neighbor Coordinate			Distance from each Neighbor and units distance
		X	Y	Z		X	Y	Z	

Figure 3.2 Reply Message

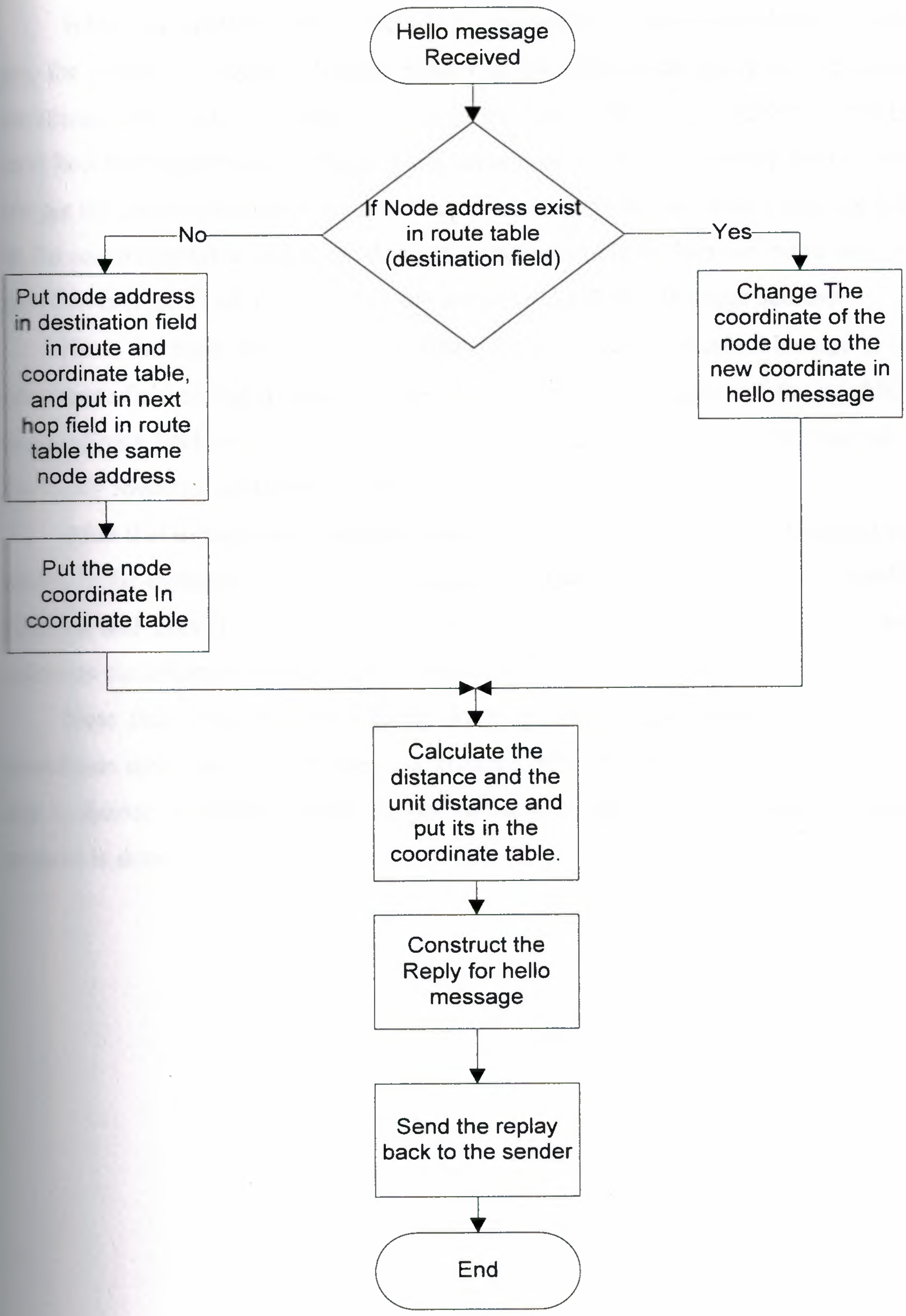

Hello message Received
If Node address exist in route table (destination field)
No
Yes
Put node address in destination field in route and coordinate table, and put in next hop field in route table the same node address
Put the node coordinate In coordinate table
Change The coordinate of the node due to the new coordinate in hello message
Calculate the distance and the unit distance and put its in the coordinate table.
Construct the Reply for hello message
Send the replay back to the sender
End

Figure 3.3 Hello Message Process When Received By Any Node

When the Reply Message received by the sender of the Hello Message it will put the address of Reply Message sender in the destination address field in the coordinate table and in the Destination Address field in the routing table and puts the next hop for that destination equal to the address of the Reply Message sender. Also we put the neighbors address from the Reply Message in the destination address field in the coordinate table and in the destination address field in the route table and puts the Next hop for all of them equal to the sender of the Reply Message address.

Now, the node will put the coordinate for the sender of Reply Message in the coordinate field to that destination. Also it will put the coordinates of the neighbors in coordinate field as specified in the Reply Message. And it will put the address of the Reply Message sender in neighbor's field also.

After that the node will calculate the distance from the source to destination and puts it in the distance field for that destination and adds this distance to the neighbor distance and puts it in the distance field for each of the neighbors. Also it will calculate the distance in unit for each destination.

Note that if we receive a Reply Message that contain address exist in the coordinate table, we compare the new distance with the one in the table if the new one is shorter we replace it with the one in the table and Figure 3.4 shows how this process is done.

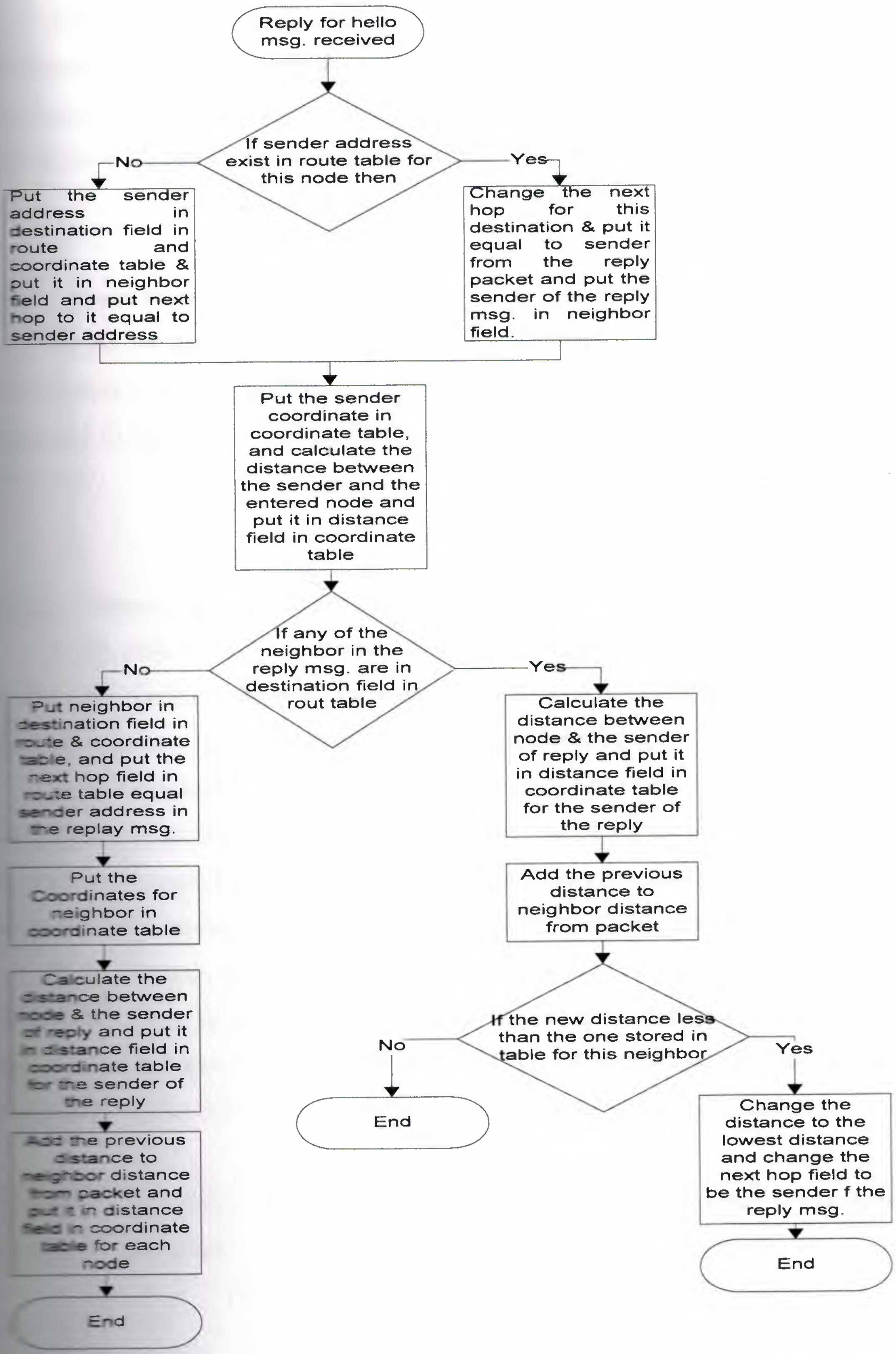

Reply for hello msg. received
If sender address exist in route table for this node then
No
Yes
Put the sender address in destination field in route and coordinate table & put it in neighbor field and put next hop to it equal to sender address
Change the next hop for this destination & put it equal to sender from the reply packet and put the sender of the reply msg. in neighbor field.
Put the sender coordinate in coordinate table, and calculate the distance between the sender and the entered node and put it in distance field in coordinate table
If any of the neighbor in the reply msg. are in destination field in rout table
No
Yes
Put neighbor in destination field in route & coordinate table, and put the next hop field in route table equal sender address in the replay msg.
Put the Coordinates for neighbor in coordinate table
Calculate the distance between node & the sender of reply and put it in distance field in coordinate table for the sender of the reply
Add the previous distance to neighbor distance from packet and put it in distance field in coordinate table for each node
End
Calculate the distance between node & the sender of reply and put it in distance field in coordinate table for the sender of the reply
Add the previous distance to neighbor distance from packet
If the new distance less than the one stored in table for this neighbor
No
Yes
End
Change the distance to the lowest distance and change the next hop field to be the sender f the reply msg.
End

Figure 3.4 Reply Message Process When Received By Any Node

If any node wants to quit, it will send Quitting Message to all the Destinations Addresses in its routing table. When the Quitting Message received by any node, the received node will delete the Quitting node address from its routing table and all the record for that address. The form of Quitting Message is shown in figure 4.4

3.2.2 Route Discovery

As in the previous chapter we have classified the node to Source node, Intermediate node and Destination node. And we will discuss now in details what happened to the ROUTE REQUEST packet and to the ROUTE REPLY packet at each node.

Case 1: when a Source node send ROUTE REQUEST

If any node wants to send a packet to another node it must first know the route to that node. To know the route the node must first look at its Routing Table to see if the destination address exists in the Destination Address field.

If the Destination Address exists in the routing table there is a known route to the destination required so we use this route. Note that if we use the known route and no reply come back this mean that the destination node had moved out of the range of the sender so we have to make a ROUTE REQUEST packet and broadcast it.

If the required destination address does not exist in the routing table for the source node, there is no known route to that destination so we have to discover the route to that destination by constructing a ROUTE REQUEST packet and broadcast it. The form of the ROUTE REQUEST packet will be as shown in figure 3.5.

Source address	Source Coordinate			Request ID	Destination address	Sender Address	Sender coordinate			Distance
	X	Y	Z				X	Y	Z	

Figure 3.5 ROUTE REQUEST packet

Distance is set to zero when sending this packet and the source broadcast this packet. Request ID is a counter that is incremented each time the same node wants to send a ROUTE REQUEST, so we increment the request ID and put it in the Request ID field in the packet. Figure 3.6 shows how this process is done.

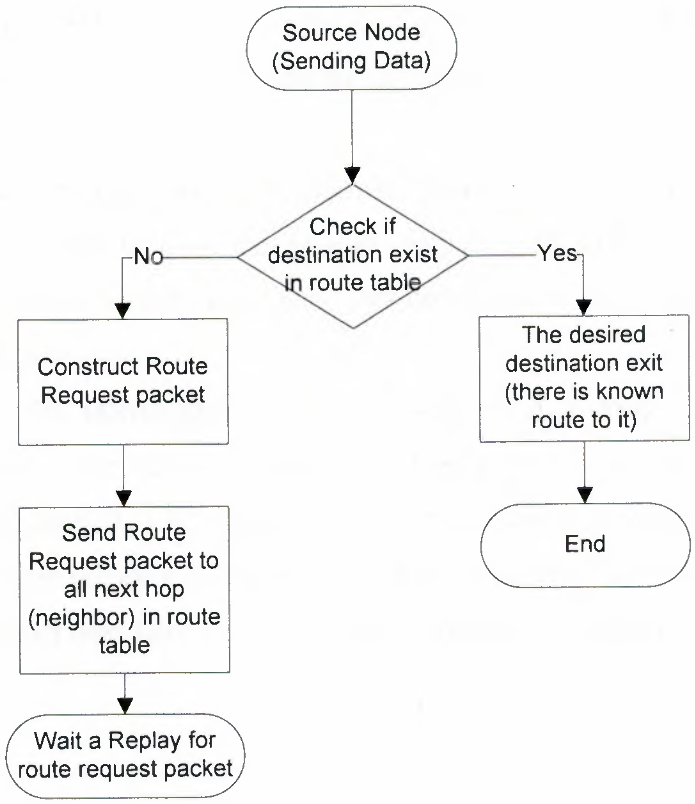

Source Node (Sending Data)
Check if destination exist in route table
No
Yes
Construct Route Request packet
Send Route Request packet to all next hop (neighbor) in route table
Wait a Replay for route request packet
The desired destination exit (there is known route to it)
End

Figure 3.6 Source Node When Sending ROUTE REQUEST

Case 2: when Intermediate node received ROUTE REQUEST

When intermediate node receives a ROUTE REQUEST packet it compares the history table with Source Address & Request ID in the ROUTE REQUEST packet received, this way will prevent processing duplicated packet. The History Table at each node will be as shown in Table 4.2

If the comparison result is a match, ignore packet because it is a duplicated packet. If the comparison result is not match we have a new ROUTE REQUEST packet, the node will copy the Source address & Request ID from the ROUTE REQUEST packet and put them in the History Table so future duplicated packet will be ignored.

The intermediate node also take the Source Address and the Destination Address from the ROUTE REQUEST packet and put them in the Reverse Table, and put the next hop equal to the Sender Address. The reverse table will help the ROUTE REPLY packet in finding its way to the source node. The Reverse Table is shown in Table 4.3

Now, the intermediate node will calculate the distance between the intermediate and the sender and add to it the distance in the packet. Also the intermediate node will change the sender address and the sender coordinate to the intermediate address and intermediate coordinate.

After that the intermediate node will search for a route for the required destination in its route table. If the destination exists in the routing table the intermediate will send the route request packet to the next hop for that destination. If the destination does not exist in the routing table the intermediate will send the route request packet to all next hops in its route table. Figure 3.7 shows how this process is done.

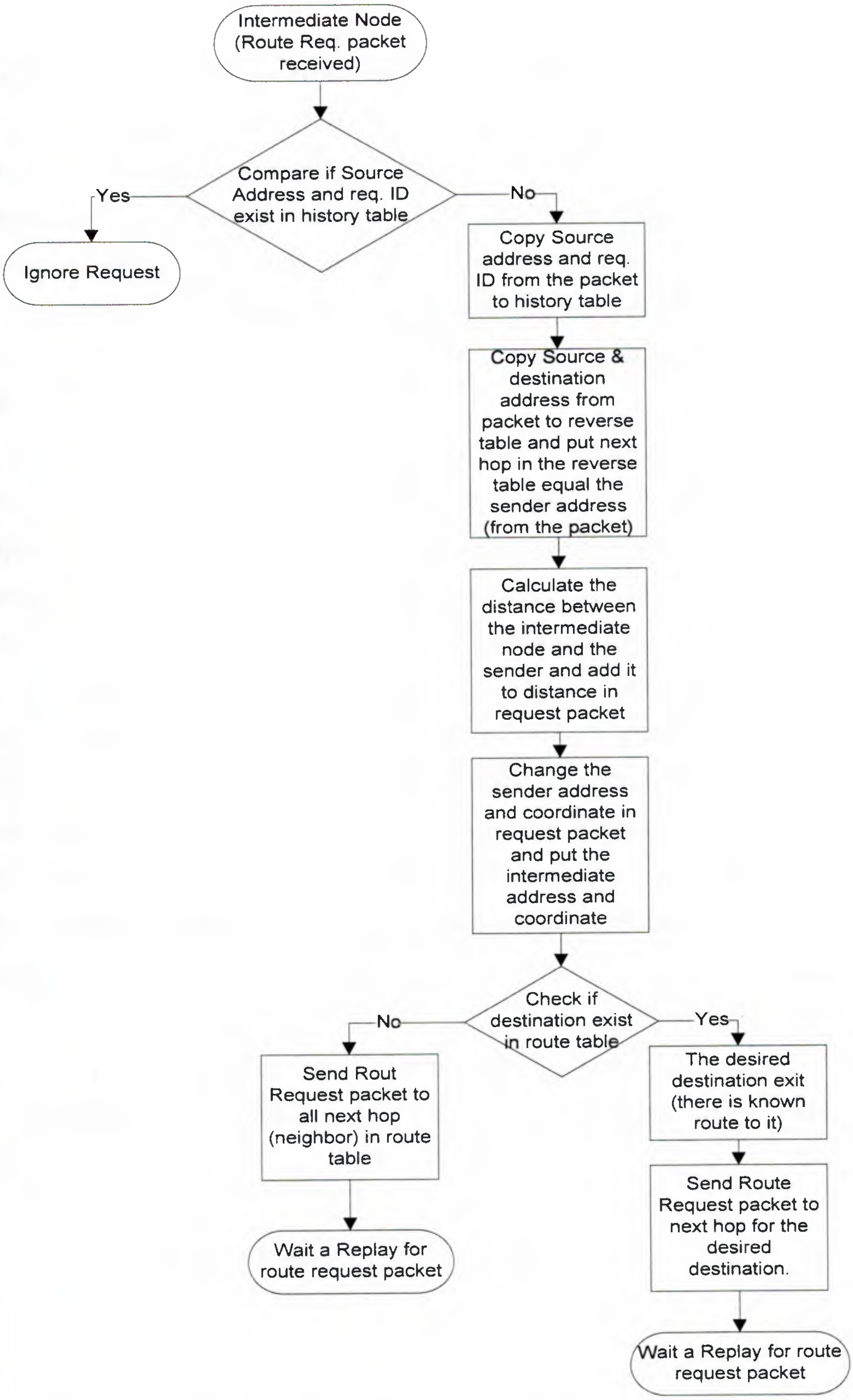

Intermediate Node (Route Req. packet received)
Compare if Source Address and req. ID exist in history table
Yes
Ignore Request
No
Copy Source address and req. ID from the packet to history table
Copy Source & destination address from packet to reverse table and put next hop in the reverse table equal the sender address (from the packet)
Calculate the distance between the intermediate node and the sender and add it to distance in request packet
Change the sender address and coordinate in request packet and put the intermediate address and coordinate
Check if destination exist in route table
No
Send Rout Request packet to all next hop (neighbor) in route table
Wait a Replay for route request packet
Yes
The desired destination exit (there is known route to it)
Send Route Request packet to next hop for the desired destination.
Wait a Replay for route request packet

Figure 3.7 Intermediate Nodes When Receiving ROUTE REQUEST

Case 3: when Destination node has received ROUTE REQUEST

When the ROUTE REQUEST packet finally is received by the destination node it will put the Source Address from the ROUTE REQUEST packet in the Destination Address field in the Routing Table if not exists. And it will put the source coordinate from the route request packet in the coordinate field for that destination.

The destination node will calculate the distance between it and the sender and add to it the distance in the ROUTE REQUEST packet, and puts it in the distance field for the source in the coordinate table.

If the destination receives a ROUTE REQUEST packet from the same source from different intermediate node it will calculate the distance and makes a comparison between them and choose the minimum distance and puts it in the distance field for the source in the coordinate table, and puts sender address with minimum distance in the next hop field for the source address in routing table.

Now, the destination node will construct a ROUTE REPLY packet as shown in figure 3.8. The source address and source coordinate are copied from the ROUTE REQEST packet to the ROUTE REPLY packet and the node will put its address and coordinate in the ROUTE REPLY packet. Also the destination node will put the overall distance between the source node and destination node in the ROUTE REPLY packet and send this packet to the next hop for the source address in the destination node routing table.

Source address	Source coordinate			Destination address	Destination address			Sender address	Sender coordinate			Distance
	X	Y	Z		X	Y	Z		X	Y	Z	

Figure 3.8 ROUTE REPLY packet

Case 4: when Intermediate node has received ROUTE REPLLY

When the ROUTE REPLY packet is received by the intermediate node it will compare Source Address and Destination Address in ROUTE REPLY packet with reverse table. If they are equal put the intermediate node address and its coordinate in the Sender address field and coordinate field in the reply packet and sends it to the next hop in the reverse table for that pair. If they are not equal ignore the packet. Figure 3.9 shows how this done.

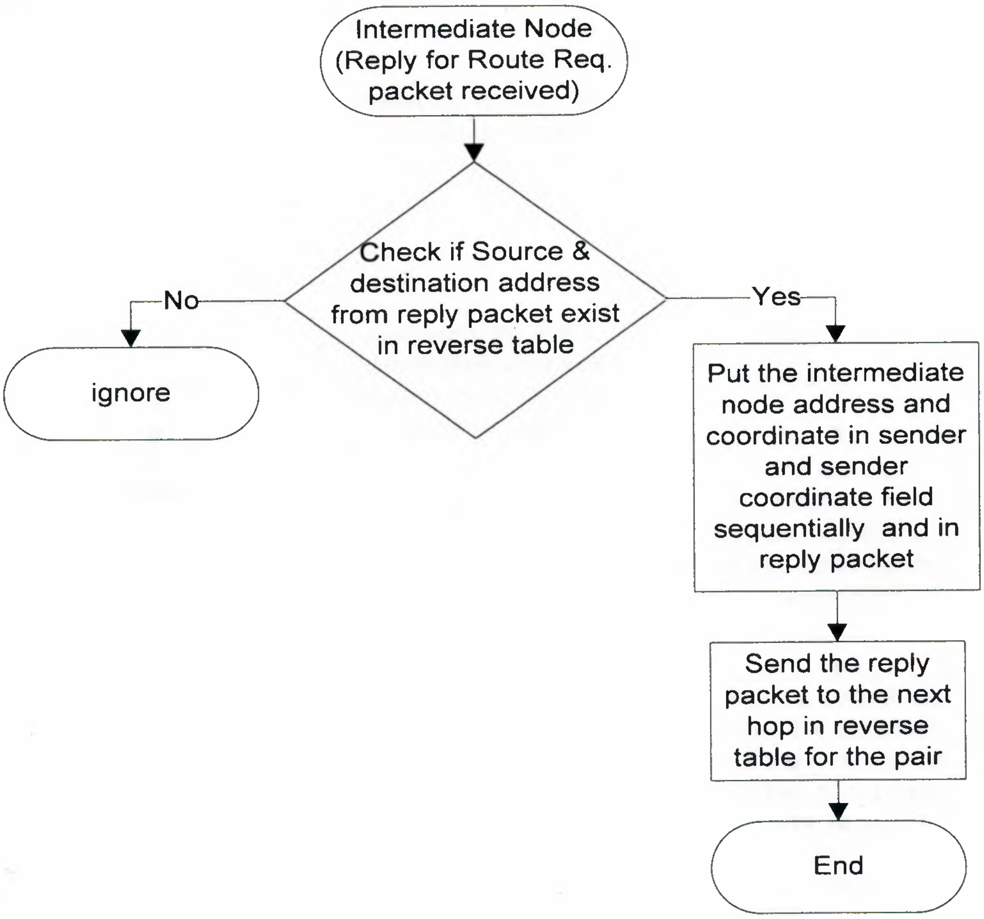

Intermediate Node (Reply for Route Req. packet received)
Check if Source & destination address from reply packet exist in reverse table
No
ignore
Yes
Put the intermediate node address and coordinate in sender and sender coordinate field sequentially and in reply packet
Send the reply packet to the next hop in reverse table for the pair
End

Figure 3.9 Intermediate Nodes When Has Received ROUTY REPLY

Case 5: when Source node has received ROUTE REPLY

When the ROUTE REPLY packet is received by the destination node it will compare Source address and Destination address between ROUTE REPLY packet and ROUTE REQUEST packet. If equal, put the destination address and the destination coordinate from the route replay in the coordinate table and put the distance from the route reply to the distance field in the coordinate table for that destination. Also put the next hop for that destination in routing table equal to the sender address. If not equal ignore. Figure 3.10 shows how this is done.

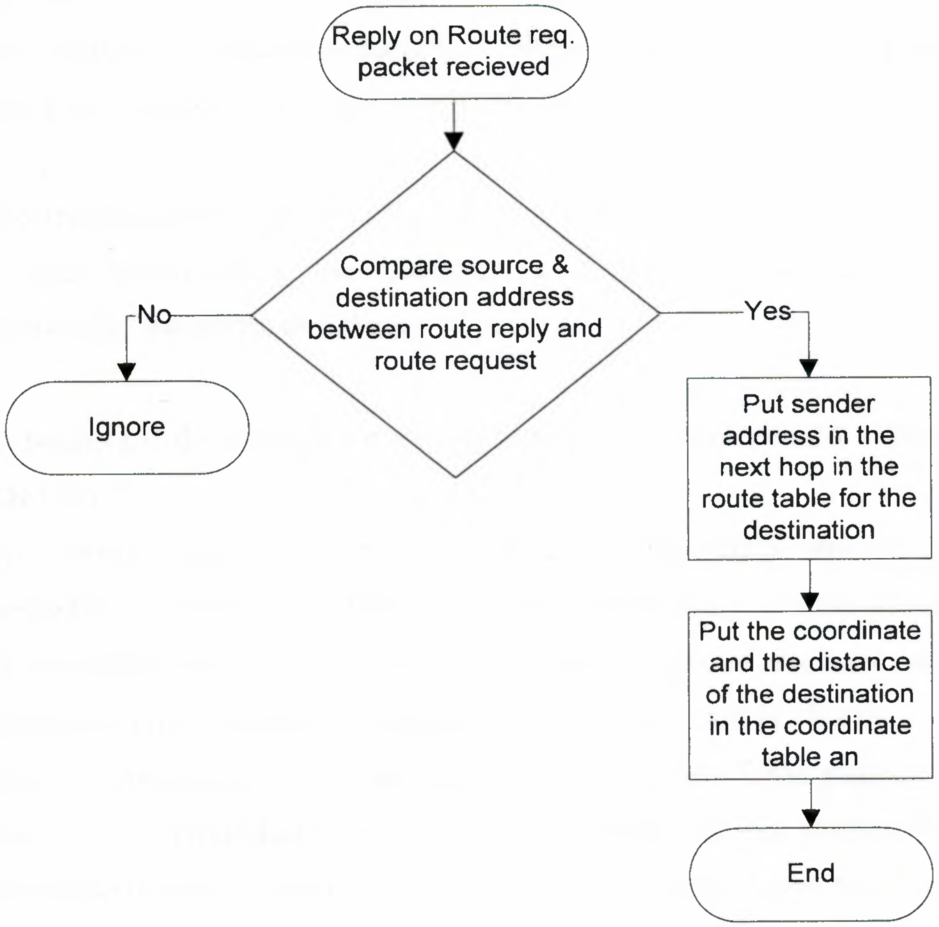

Reply on Route req. packet recieved
Compare source & destination address between route reply and route request
No
Yes
Ignore
Put sender address in the next hop in the route table for the destination
Put the coordinate and the distance of the destination in the coordinate table an
End

Figure 3.10 Source Node When Has Received ROUTE REPLY

3.3Integration of GIS in Routing Protocol

So far we show how to find the route in ad hoc network and how to take the decision according to time delay metric using GPS, now we will discuss ideas for integrating the surface information (Geospatial Information) as metric in the routing process.

3.3.1 Theory and Our Approach

As we know the surrounding environments will affect the communication channel, and thus will make the channel unreliable, what we are after here is to use the surface information to find the best reliable route.

Geospatial information can offer intensive information to support the routing metric, after understanding the foundation of Geospatial information we can summarize using the metric as follow:

- A segment is the physical connection between two neighborhood segments on the route.
- A network segment crosses the geospatial surroundings and causes the degradation of the quality of the communication channel across the segment.
- A degradation factor has to be calculated for each segment across the route and finally the route, which has the minimum degradation, will be chosen.
- Some references suggest using Fade margin, which is a GSM planning factor that uses the terrain factor to generate a fade margin factor at specific location. But again the GSM communication model assumes fixed access point in which a terrain analysis can be predefined while in our case each point is a mobile multiple access point of communication.

o For more information about the Fade, readers can refer to the following [6] for more information.

Note: Even though finding this metric sounds process intensive, but this metric can be applied after using the time delay metric.

3.3.2 Problems in Implementation

Routing is a real time task, using GPS can be fast in decision while using GIS needs more complicated issues these can be summarize in:

1. Availability of data (surface information).
2. Size of data.
3. Processing of data.

To overcome of the problem of availability of data we can build the surface by interpolation using triangulation techniques of history logged GPS coordinate from nodes. There are two types of surfaces:

1. Digital Elevation Model (DEM)

In this type we represent the surface of earth only without any human inputs (building) as shown in Figure 3.11. This type of data is useful for network in inhabited places.

Figure 3.11 Digital Elevation Model

2. Digital Terrain Model (DTM)

In this type we represent the surface of earth with the human inputs (building) as shown in Figure 3.12. This type of data is useful for network in urban places.

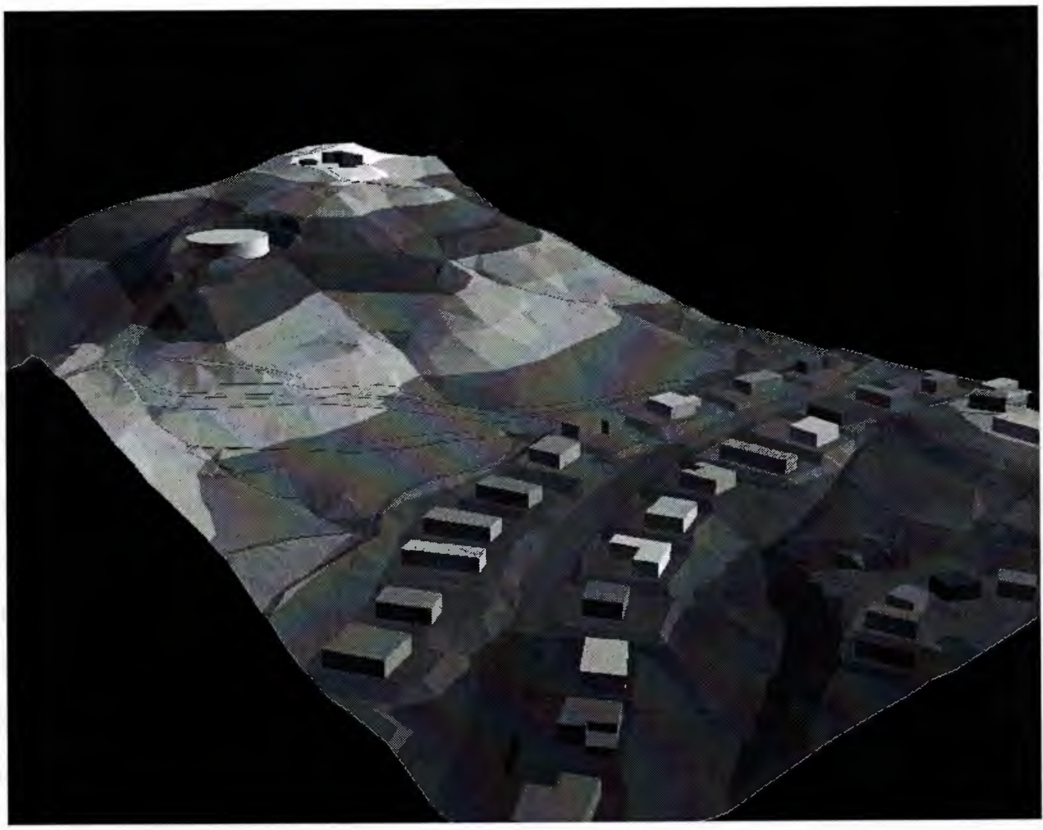


Figure 3.12 Digital Terrain Model

By this way we theoretically solve the problem of availability of data but still there is a big problem, the size of data is very large and the processing of this data.

To overcome of these problem experimental tests must be done to see how efficient the use of this information can be.

3.4 Summary

In this chapter we have discussed how GPS technique is used in routing protocol in ad hoc networks, Using GIS technique is more complicated than the use of GPS and can cause a big delay. Also some quantitative test must be done to see the efficiency of using GIS in routing protocol.

4. FIBRE CHANNEL BASICS

4.1 Overview

Fibre Channel (FC) is a technology standard that allows data to be transferred from one network node to another at very high speeds. Fibre Channel is simply the most reliable, highest performing solution for information storage, transfer, and retrieval available today. Current implementations transfer data at 100 MB/second, although, 200 MB/second and 400 MB/second data rates have already been tested.

This standard is backed by a consortium of industry vendors and has been accredited by the American National Standards Institute (ANSI). Many products are now on the market that take advantage of FC's high-speed, high-availability characteristics. In the topics that follow, we introduce Fibre Channel basic information to complement the solutions that we describe later in this redbook. We cover areas that are internal to Fibre Channel and show how data is moved and the medium upon which it travels.

4.2 SAN components

The industry considers Fibre Channel as the architecture on which most SAN implementations will be built, with FICON as the standard protocol for S/390 systems, and Fibre Channel Protocol (FCP) as the standard protocol for non-S/390 systems.

Based on this implementation, there are three main categories of SAN components:

- SAN servers
- SAN storage
- SAN interconnects

We show the typical SAN components that are likely to be encountered in Figure 4.1.

Fibre Channel Basics

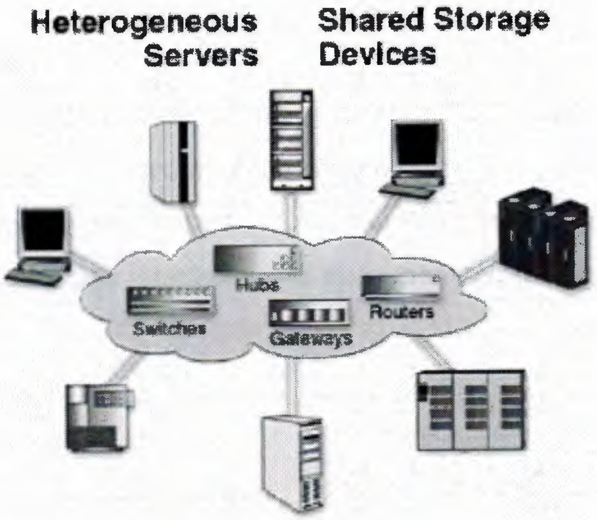


Figure 4.1. SAN Components

4.2.1 SAN Servers

The server infrastructure is the underlying reason for all SAN solutions. This infrastructure includes a mix of server platforms, such as Windows NT, UNIX and its various flavors, and mainframes. With initiatives, such as server consolidation and e-business, the need for a SAN has become very strong.

Although most current SAN solutions are based on a homogeneous server platform, future implementations will take into account the heterogeneous nature of the IT world.

4.2.2 SAN Storage

The storage infrastructure is the foundation on which information relies, and must support the business objectives and business model. In this environment, simply deploying more and faster storage devices is not enough; a new kind of infrastructure is needed, one that provides network availability, data accessibility, and system manageability. The SAN meets this challenge. It is a high-speed subnet that establishes a direct connection between storage resources and servers. The SAN liberates the storage device, so it is not on a particular server bus, and attaches it directly to the network. In other words, storage is externalized, and functionally distributed to the organization. The SAN also enables the centralization of storage devices and the clustering of servers, which makes for easier and less expensive administration.

4.2.3 SAN Interconnects

The first element that must be considered in any SAN implementation is the connectivity of components of storage and servers using technologies such as Fibre Channel. The components listed here are typically used in LAN and WAN implementations. SANs, like LANs, interconnect the storage interfaces into many network configurations and across long distances.

- Cables and connectors
- Gigabit Link Model (GLM)
- Gigabit Interface Converters (GBIC)
- Media Interface Adapters (MIA)
- Adapters
- Extenders
- Multiplexors
- Hubs
- Routers
- Bridges
- Gateways
- Switches
- ESCON Directors
- FICON Directors.

4.3 Jargon Terminology Shift

Much of the terminology used for SAN has its origin in Internet Protocol (IP) network terminology. In some cases, companies in the industry use different terms that mean the same thing, and in some cases, the same terms meaning different things. In this book we will attempt to define some of the terminology that is used and its changing nature among vendors.

4.4 Vendor Standards and Main Vendors

This section gives an overview of the major SAN vendors in the industry:

- **Systems/storage SAN providers**

  IBM (Sequent), SUN, HP, EMC (DG Clariion), STK, HDS, Compaq, and Dell
- **Hub providers**

  Gadroon, Vixel and Emulex
- **Switch providers**

  Brocade, Ancor, McDATA, Vixel, STK/SND and Gadzoox
- **Gateway and Router providers**

  ATTO, Chaparrel Tech, CrossRoads Tech, Pathlight, Vicom
- **Host bus adapters (HBA) providers**

  Ancon, Compaq, Emulex, Genroco, Hewlett-Packard, Interphase, Jaycor Networks, Prisia, Qlogic and Sun Microsystems.
- **Software providers**

  IBM/Tivoli, Veritas, Legato, Computer Associates, DataDirect, Transoft (HP), Crosstor and Retrieve.

4.5 Physical Characteristics

This section describes the components and technology associated with the physical aspects of Fibre Channel. We describe the supported cables and give an overview of the types of connectors that are generally available and are implemented in a SAN environment.

4.5.1 Cable

As with parallel SCSI and traditional networking, different types of cable are used for Fibre Channel configurations. Two types of cables are supported:

- Copper
- Fiber-optic

Fibre Channel can be run over optical or copper media, but fiber-optic enjoys a major advantage in noise immunity. It is for this reason that fiber-optic cabling is preferred. However, copper is also widely used and it is likely that in the short term a mixed environment will need to be tolerated and supported. Figure 4.2 shows fiber-optical data transmission.

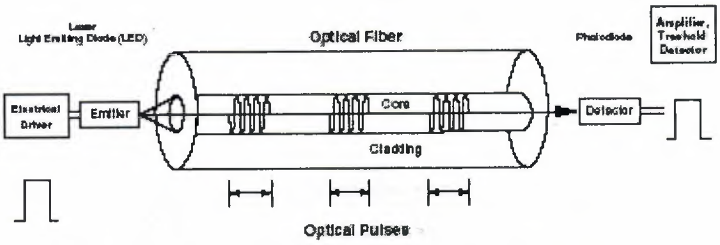


Figure 4.2 Fiber Optical Data Transmission

In addition to the noise immunity, fiber-optic cabling provides a number of distinct advantages over copper transmission lines that make it a very attractive medium for many applications.

At the forefront of the advantages are:

- Greater distance capability than is generally possible with copper
- Insensitive to induced electro-magnetic interference (EMI)
- No emitted electro-magnetic radiation (RFI)
- No electrical connection between two ports
- Not susceptible to crosstalk
- Compact and lightweight cables and connectors

However, fiber-optic and optical links do have some drawbacks. Some of the considerations are:

- Optical links tend to be more expensive than copper links over short distances

- Optical connections don't lend themselves to backplane printed circuit wiring
- Optical connections may be affected by dirt and other contamination

Overall, optical fibers have provided a very high-performance transmission medium which has been refined and proven over many years.

Mixing fiber-optical and copper components in the same environment is supported, although not all products provide that flexibility and this should be taken into consideration when planning a SAN. Copper cables tend to be used for short distances, up to 30 meters, and can be identified by their DB-9, 9 pin, connector.

Normally fiber-optic cabling is referred to by mode or the frequencies of light waves that are carried by particular cable type. Fiber cables come in two distinct types, as shown in Figure 4.3.

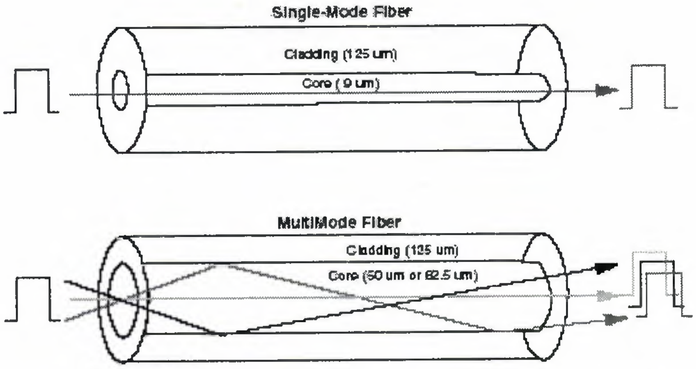


Figure 4.3. Multi-mode and single-mode propagation

- **Multi-mode fiber (MMF)** for short distances, up to 500m using FCP Multi-mode cabling is used with shortwave laser light and has either a 50 micron or a 62.5 micron core with a cladding of 125 micron. The 50 micron or 62.5 micron diameter is sufficiently large for injected light waves to be reflected off the core interior.
- **Single-mode fiber (SMF)** for long distances Single-mode is used to carry longwave laser light. With a much smaller 9 micron diameter core and a single-

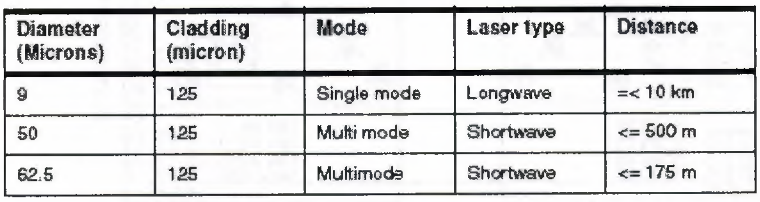
mode light source, single-mode fiber supports much longer distances, currently up to 10 km at gigabit speed.

Fibre Channel architecture supports both short wave and long wave optical transmitter technologies, as follows:

- **Short wave laser** — this technology uses a wavelength of 780 nanometers and is only compatible with multi-mode fiber.
- **Long wave laser** — this technology uses a wavelength of 1300 nanometers. It is compatible with both single-mode and multi-mode fiber.

IBM will support the following distances for FCP as shown in Table 1.

Table 4.1. FCP distance.



Diameter (Microns)	Cladding (micron)	Mode	Laser type	Distance
9	125	Single mode	Longwave	=< 10 km
50	125	Multi mode	Shortwave	<= 500 m
62.5	125	Multimode	Shortwave	<= 175 m

Campus

A campus topology is nothing more than "cabling" buildings together, so that data can be transferred from a computer system in one building to storage devices, whether they are disk storage, or tape storage for backup, or other devices in another building. We show a campus topology in Figure 4.4.

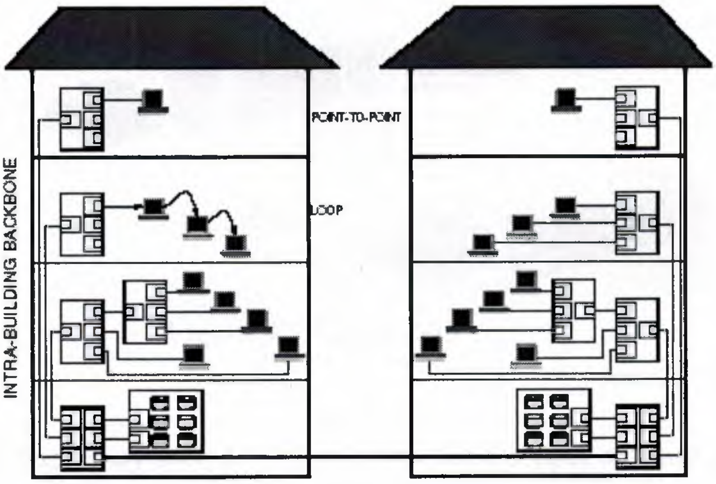


Figure 4.4. Campus Topology

4.5.2 Connectors

Three connector types are generally available. Fiber-optic connectors are usually provided using dual subscriber connectors (SC). Copper connections can be provided through standard DB-9 connectors or the more recentlydeveloped high speed serial direct connect (HSSDC) connectors. We show a selection of connectors in Figure 4.5.

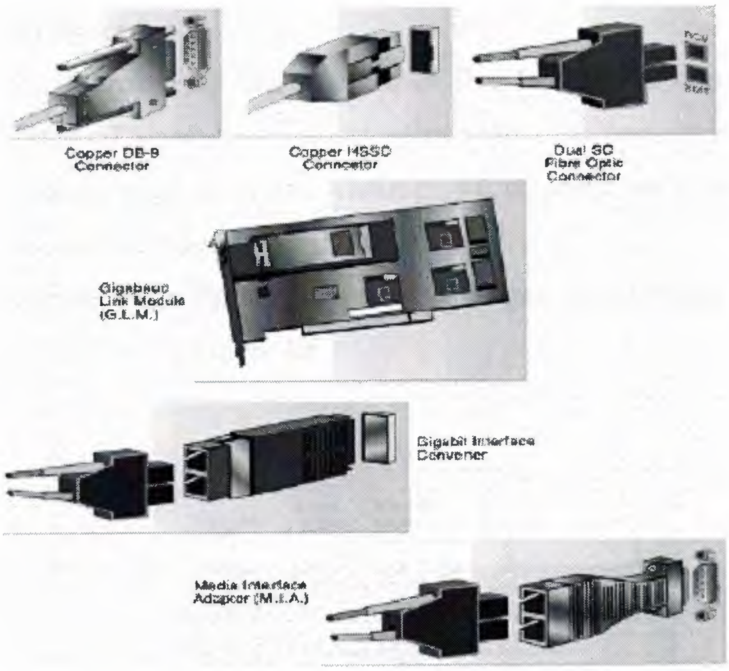

Copper DB-9 Connector
Copper HSSD Connector
Dual SC Fibre Optic Connector
Gigabaud Link Module (G.L.M.)
Gigabit Interface Converter
Media Interface Adaptor (M.I.A.)

Figure 4.5. Connectors

Fibre Channel products may include a fixed, embedded copper or fiber-optic interface, or they may provide a media-independent interface. There are three media-independent interfaces available:

- **Gigabit Link Modules (GLMs)** — convert parallel signals to serial, and vice versa. GLMs include the serializer/de-serializer (SERDES) function and provide a 20-bit parallel interface to the Fibre Channel encoding and control logic. GLMs are primarily used to provide factory configurability, but may also be field exchanged or upgraded by users.
- **Gigabit Interface Converters (GBICs)** — provide a serial interface to the SERDES function. GBICs can be hot inserted or removed from installed devices. These are particularly useful in multiport devices, such as switches and hubs, where single ports can be reconfigured without affecting other ports.
- **Media Interface Adapters (MIAs)** — allow users to convert copper DB-9 connectors to multi-mode fibre optics. The power to support the optical transceivers is supplied by defined pins in the DB-9 interface.

4.6 Fibre Channel Layers

Fibre Channel (FC) is broken up into a series of five layers. The concept of layers, starting with the ISO/OSI seven-layer model, allows the development of one layer to remain independent of the adjacent layers. Although, FC contains five layers, those layers follow the general principles stated in the ISO/OSI model.

The five layers are divided into two parts Physical and signaling layer and Upper layer The five layers are illustrated in Figure 4.6.

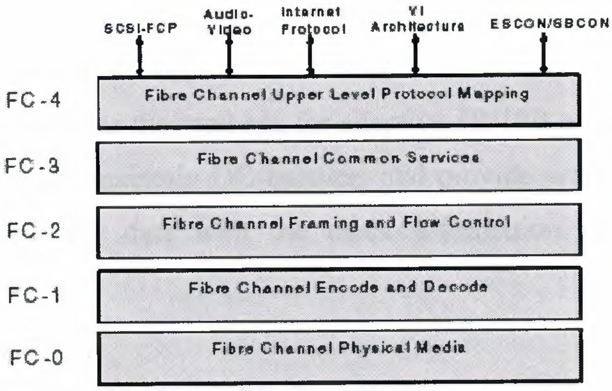

SCSI-FCP
Audio-Video
Internet Protocol
VI Architecture
ESCON/SBCON
FC-4
Fibre Channel Upper Level Protocol Mapping
FC-3
Fibre Channel Common Services
FC-2
Fibre Channel Framing and Flow Control
FC-1
Fibre Channel Encode and Decode
FC-0
Fibre Channel Physical Media

Figure 4.6. Fibre Channel layers

4.6.1 Physical and Signaling Layers

The physical and signaling layers include the three lowest layers: FC-0, FC-1, and FC-2.

4.6.1.1 Physical Interface and Media: FC-0

The lowest layer (FC-0) defines the physical link in the system, including the cabling, connectors, and electrical parameters for the system at a wide range of data rates. This level is designed for maximum flexibility, and allows the use of a large number of technologies to match the needs of the desired configuration.

A communication route between two nodes may be made up of links of different technologies. For example, in reaching its destination, a signal may start out on copper wire and become converted to single-mode fibre for longer distances. This flexibility allows for specialized configurations depending on IT requirements.

Laser safety

Fibre Channel often uses lasers to transmit data, and can, therefore, present an optical health hazard. The FC-0 layer defines an open fibre control (OFC) system, and acts as a safety interlock for point-to-point fibre connections that use semiconductor laser diodes as the optical source. If the fibre connection is broken, the ports send a series of pulses until the physical connection is re-established and the necessary handshake procedures are followed.

4.6.1.2 Transmission Protocol: FC-1

The second layer (FC-1) provides the methods for adaptive 8B/10B encoding to bind the maximum length of the code, maintain DC-balance, and provide word alignment. This layer is used to integrate the data with the clock information required by serial transmission technologies.

4.6.1.3 Framing and Signaling Protocol: FC-2

Reliable communications result from Fibre Channel's FC-2 framing and signaling protocol. FC-2 specifies a data transport mechanism that is independent of upper layer protocols. FC-2 is self-configuring and supports point-to-point, arbitrated loop, and switched environments. FC-2, which is the third layer of the FC-PH, provides the transport methods to determine:

- Topologies based on the presence or absence of a fabric
- Communication models
- Classes of service provided by the fabric and the nodes
- General fabric model
- Sequence and exchange identifiers
- Segmentation and reassembly

Data is transmitted in 4-byte ordered sets containing data and control characters. Ordered sets provide the availability to obtain bit and word synchronization, which also establishes word boundary alignment.

Together, FC-0, FC-1, and FC-2 form the Fibre Channel physical and signaling interface (FC-PH).

4.6.2 Upper Layers

The Upper layer includes two layers: FC-3 and FC-4.

4.6.2.1 Common Services: FC-3

FC-3 defines functions that span multiple ports on a single-node or fabric.

Functions that are currently supported include:

- **Hunt groups:** A hunt group is a set of associated N_Ports attached to a single node. This set is assigned an alias identifier that allows any frames containing the alias to be routed to any available N_Port within the set. This decreases latency in waiting for an N_Port to become available.
- **Striping:** Striping is used to multiply bandwidth, using multiple N_Ports in parallel to transmit a single information unit across multiple links.
- **Multicast:** Multicast delivers a single transmission to multiple destination ports. This includes the ability to broadcast to all nodes or a subset of nodes.

4.6.2.2 Upper Layer Protocol Mapping (ULP): FC-4

The highest layer (FC-4) provides the application-specific protocols. Fibre Channel is equally adept at transporting both network and channel information and allows both protocol types to be concurrently transported over the same physical interface.

Through mapping rules, a specific FC-4 describes how ULP processes of the same FC-4 type interoperate. A channel example is sending SCSI commands to a disk drive, while a networking example is sending IP (Internet Protocol) packets between nodes.

4.7 The Movement of Data

To move data bits with integrity over a physical medium, there must be a mechanism to check that this has happened and integrity has not been compromised. This is provided by a reference clock which ensures that each bit is received as it was transmitted. In parallel topologies this can be accomplished by using a separate clock or strobe line. As data bits are transmitted in parallel from the source, the strobe line alternates between high or low to signal the receiving end that a full byte has been sent. In the case of 16- and 32-bit wide parallel cable, it would indicate that multiple bytes have been sent.

The reflective differences in fiber-optic cabling mean that modal dispersion may occur. This may result in frames arriving at different times. This bit error rate (BER) is referred to as the jitter budget. No products are entirely jitter free, and this is an important consideration when selecting the components of a SAN.

As serial data transports only have two leads, transmit and receive, clocking is not possible using a separate line. Serial data must carry the reference timing which means that clocking is embedded in the bit stream.

Embedded clocking, though, can be accomplished by different means. Fibre Channel uses a byte-encoding scheme, which is covered in more detail in 4.7, “Data encoding” on page 56, and clock and data recovery (CDR) logic to recover the clock. From this, it determines the data bits that comprise bytes and words.

Gigabit speeds mean that maintaining valid signaling, and ultimately valid data recovery, is essential for data integrity. Fibre Channel standards allow for a single bit error to occur only once in a trillion bits (10-12). In the real IT world, this equates to a maximum of one bit error every 16 minutes, however actual occurrence is a lot less frequent than this.

4.8 Data Encoding

In order to transfer data over a high-speed serial interface, the data is encoded prior to transmission and decoded upon reception. The encoding process ensures that sufficient clock information is present in the serial data stream to allow the receiver to synchronize to the embedded clock information and successfully recover the data at the required error rate. This 8b/10b encoding will find errors that a parity check cannot. A parity check will not find even numbers of bit errors, only odd numbers. The 8b/10b encoding logic will find almost all errors.

First developed by IBM, the 8b/10b encoding process will convert each 8-bit byte into two possible 10-bit characters.

This scheme is called 8b/10b encoding, because it refers to the number of data bits input to the encoder and the number of bits output from the encoder.

The format of the 8b/10b character is of the format Ann.m, where:

- A represents 'D' for data or 'K' for a special character
- nn is the decimal value of the lower 5 bits (EDCBA)
- '.' is a period
- m is the decimal value of the upper 3 bits (HGF)

We illustrate an encoding example in Figure 4.7.

In the encoding example the following occurs:

1. Hexadecimal representation x'59' is converted to binary: 01011001
2. Upper three bits are separated from the lower 5 bits: 010 11001
3. The order is reversed and each group is converted to decimal: 25 2
4. Letter notation D (for data) is assigned and becomes: D25.2

As we illustrate, the conversion of the 8-bit data bytes has resulted in two 10-bit results. The encoder needs to choose one of these results to use. This is achieved by monitoring the running disparity of the previously processedcharacter. For example, if the previous character had a positive disparity, then the next character issued should have an encoded value that represents

negative disparity.

You will notice that in our example the encoded value, when the running disparity is either positive or negative, is the same. This is legitimate. In some cases it (the encoded value) will differ, and in others it will be the same.

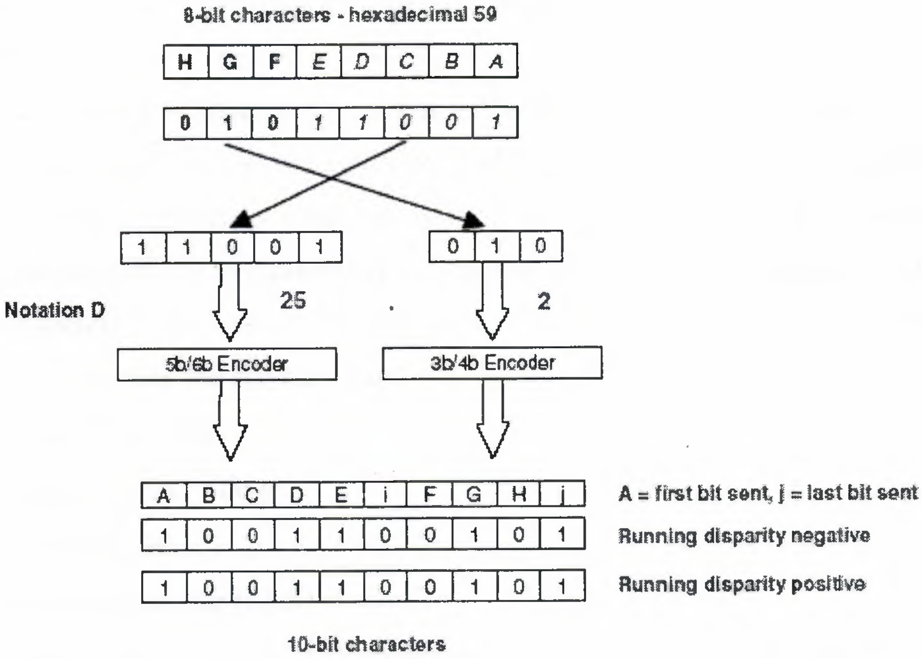


Figure 4.7. 8b/10b Encoding Logic

4.9 Ordered Sets

Fibre Channel uses a command syntax, known as an ordered set, to move the data across the network. The ordered sets are four byte transmission words containing data and special characters which have a special meaning.

Ordered sets provide the availability to obtain bit and word synchronization, which also establishes word boundary alignment. An ordered set always begins with the special character K28.5. Three major types of ordered sets are defined by the signaling protocol.

The frame delimiters, the start-of-frame (SOF) and end-of-frame (EOF) ordered sets, establish the boundaries of a frame. They immediately precede or follow the contents of a Frame. There are 11 types of SOF and 8 types of EOF delimiters defined for the Fabric and N_Port Sequence control.

The two primitive signals: idle and receiver ready (R_RDY) are ordered sets designated by the standard to have a special meaning. An Idle is a primitive signal transmitted on the link to indicate an operational port facility ready for frame transmission and reception.

The R_RDY primitive signal indicates that the interface buffer is available for receiving further frames.

A primitive sequence is an ordered set that is transmitted and repeated continuously to indicate specific conditions within a port or conditions encountered by the receiver logic of a port. When a primitive sequence is received and recognized, a corresponding primitive sequence or Idle is transmitted in response. Recognition of a primitive sequence requires consecutive detection of three instances of the same ordered set. The primitive sequences supported by the standard are:

- Offline state (OLS)
- Not operational (NOS)
- Link reset (LR)
- Link reset response (LRR)

Offline (OLS): The offline primitive sequence is transmitted by a port to indicate one of the following conditions: The port is beginning the link initialization protocol, or the port has received and recognized the NOS protocol or the port is entering the offline status.

Not operational (NOS): The not operational primitive sequence is transmitted by a port in a point-to-point or fabric environment to indicate that the transmitting port has detected a link failure or is in an offline condition, waiting for the OLS sequence to be received.

Link reset (LR): The link reset primitive sequence is used to initiate a link reset.

Link reset response (LRR): Link reset response is transmitted by a port to indicate that it has recognized
a LR sequence and performed the appropriate link reset.

4.10 Frames

Frames are the basic building blocks of an FC connection. The frames contain the information to be transmitted, the address of the source and destination ports, and link control information. Frames are broadly categorized as Data frames and Link_control frames. When the frame is defined as a link control frame the length of the data field is zero bytes. If the frame is defined as a data frame, the data field may be any number of words between zero and 528 (0 and 2112 bytes). Data frames may be used as Link_Data frames and Device_Data frames. Link control frames are classified as Acknowledge (ACK) and Link_Response (Busy and Reject) frames.

The primary function of the fabric is to receive the frames from the source port and route them to the destination port. It is the FC-2 layer's responsibility to break the data to be transmitted into frame size, and reassemble the frames. The frame structure is shown in Figure 4.8.

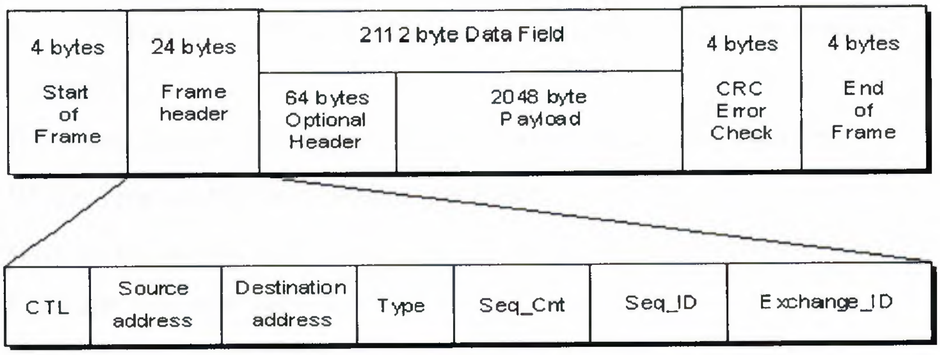


Figure 4.8. Frame Structure

Each frame begins and ends with a frame delimiter. The frame header immediately follows the SOF delimiter. The frame header is used to control link applications and control device protocol transfers, and to detect missing or out of order frames. An optional header may contain further link control information. A maximum 2112 byte long field contains the information to be transferred from a source N_Port to a destination N_Port. The 4 bytes cyclic redundancy check (CRC) precedes the EOF delimiter. The CRC is used to detect transmission errors.

4.11 Framing Classes of Service

Fibre Channel provides a logical system of communication called class of service that is allocated by various login protocols. Fibre Channel provides six different classes of service:

- Class 1: Acknowledged connection service
- Class 2: Acknowledged connectionless service
- Class 3: Unacknowledged connectionless service

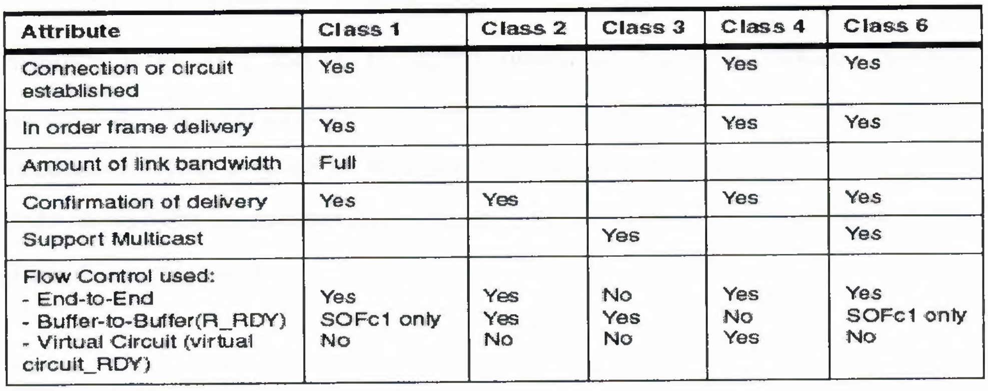
- Class 4: Fractional bandwidth connection-oriented service
- Class 5: Reserved for future development
- Class 6: Uni-directional connection service

Each class of service has a specific set of delivery attributes involving characteristics, such as:

- Is a connection or circuit established?
- Is the in-order delivery of frames guaranteed?
- If a connection is established, how much bandwidth is reserved for that connection?
- Is confirmation of delivery or notification of non-delivery provided?
- Which flow control mechanisms are used?

The answers to the above questions form the basis for the different classes of service provided and are shown in Table 4.2.

Table 4.2 Classes of service



Attribute	Class 1	Class 2	Class 3	Class 4	Class 6
Connection or circuit established	Yes			Yes	Yes
In order frame delivery	Yes			Yes	Yes
Amount of link bandwidth	Full				
Confirmation of delivery	Yes	Yes		Yes	Yes
Support Multicast			Yes		Yes
Flow Control used: - End-to-End - Buffer-to-Buffer(R_RDY) - Virtual Circuit (virtual circuit_RDY)	Yes SOFc1 only No	Yes Yes No	No Yes No	Yes No Yes	Yes SOFc1 only No

Class 1: Acknowledged connection service

Class 1 provides true connection service. The result is circuit-switched, dedicated bandwidth connections.

An end-to-end path between the communicating devices is established through the switch. Fibre Channel Class 1 service provides an acknowledgment of receipt for guaranteed delivery. Class 1 also provides full-bandwidth, guaranteed delivery, and bandwidth for applications like image transfer and storage backup and recovery. Some applications use the guaranteed delivery feature to move data reliably and quickly without the overhead of a network protocol stack. Camp On is a Class 1 feature that enables a switch to monitor a busy port and queue that port for the next connection. As soon as the port is free, the switch makes the connection. This switch service speeds connect time, rather than sending a "busy" signal back to the originating N_Port and requiring the N_Port to retry to make the connection.

Stacked connect is a Class 1 feature that enables an originating N_Port to queue sequential connection requests with the switch. Again, this feature reduces overhead and makes the switch service more efficient.

Another form of Class 1 is called dedicated simplex service. Normally, Class 1 connections are bi-directional; However, in this service, communication is in one direction only. It is used to separate the transmit and receive switching. It permits one node to transfer to another node while simultaneously receiving from a third node. We show this in Figure 4.9.

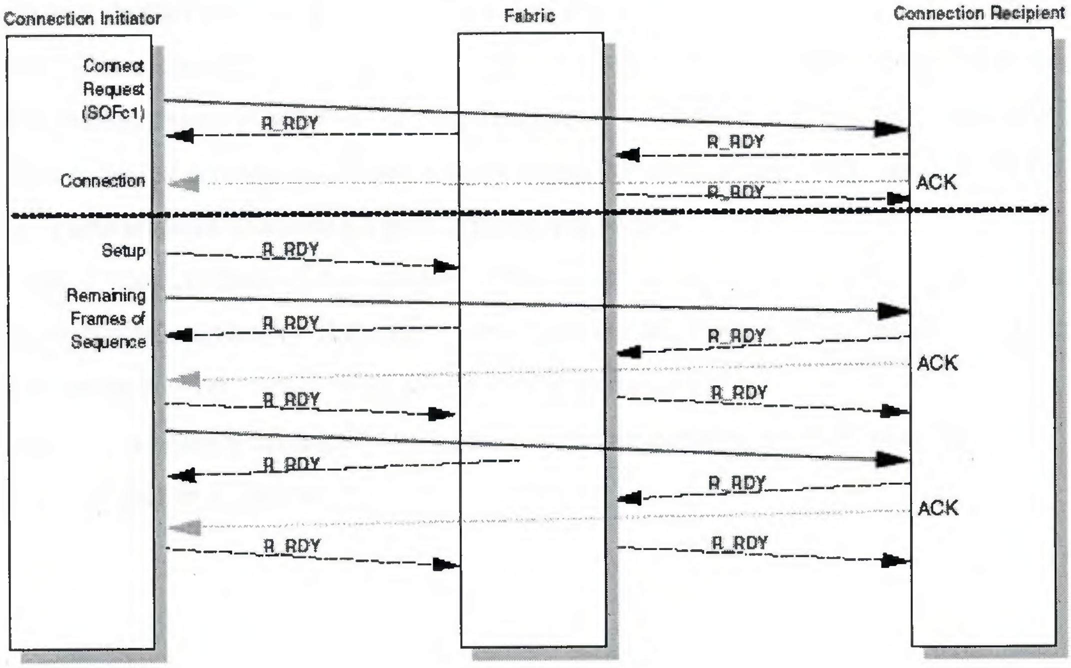


Figure 4.9. Class 1 flow control

Class 2: Acknowledged connectionless service

Class 2 is a connectionless service, independently switching each frame and providing guaranteed delivery with an acknowledgment of delivery. The path between two interconnected devices is not dedicated. The switch multiplexes traffic from N_Ports and NL_Ports without dedicating a path through the switch.

Class 2 credit-based flow control eliminates congestion that is found in many connectionless networks. If the destination port is congested, a "busy" signal is sent to the originating N_Port. The N_Port will then resend the message.

This way, no data is arbitrarily discarded just because the switch is busy at the time.

We show this in Figure 4.10.

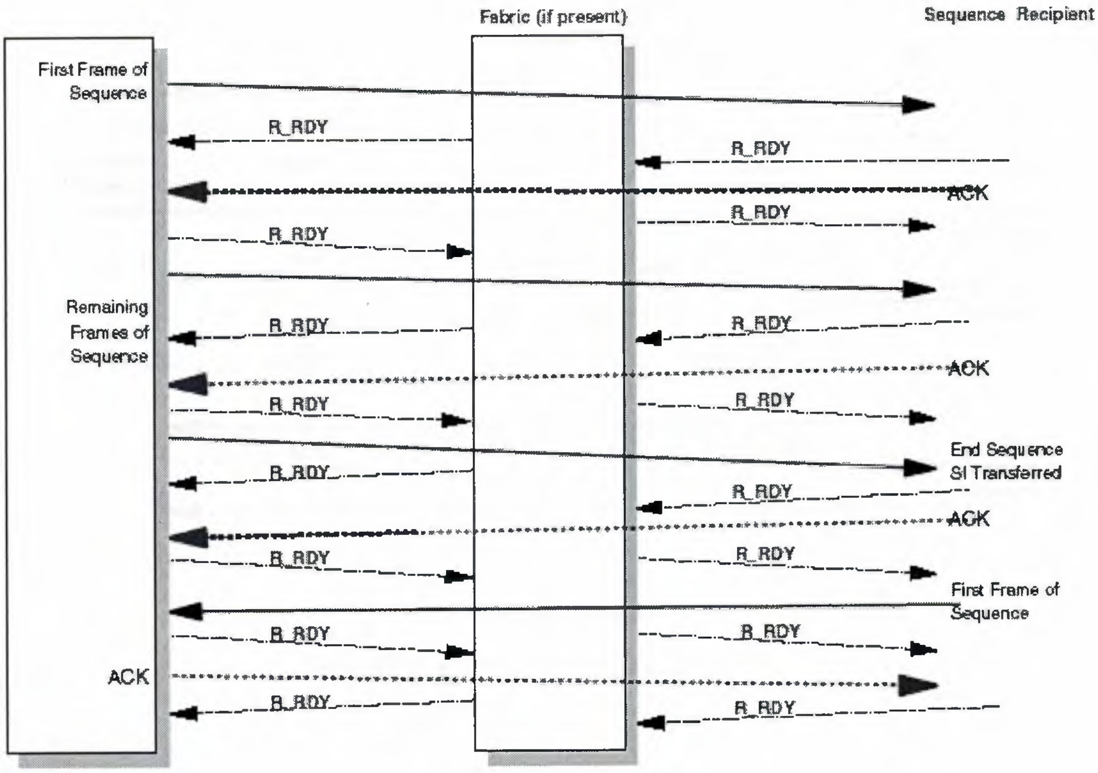


Figure 4.10 Class 2 Flow Control

Class 3: Unacknowledged connectionless service

Class 3 is a connectionless service, similar to Class 2, but no confirmation of receipt is given. This unacknowledged transfer is used for multicasts and broadcasts on networks, and for storage interfaces on Fibre Channel loops.

The loop establishes a logical point-to-point connection and reliably moves data to and from storage.

Class 3 arbitrated loop transfers are also used for IP networks. Some applications use logical point-to-point connections without using a network layer protocol, taking advantage of Fibre Channel's reliable data delivery. We show this in Figure 4.11.

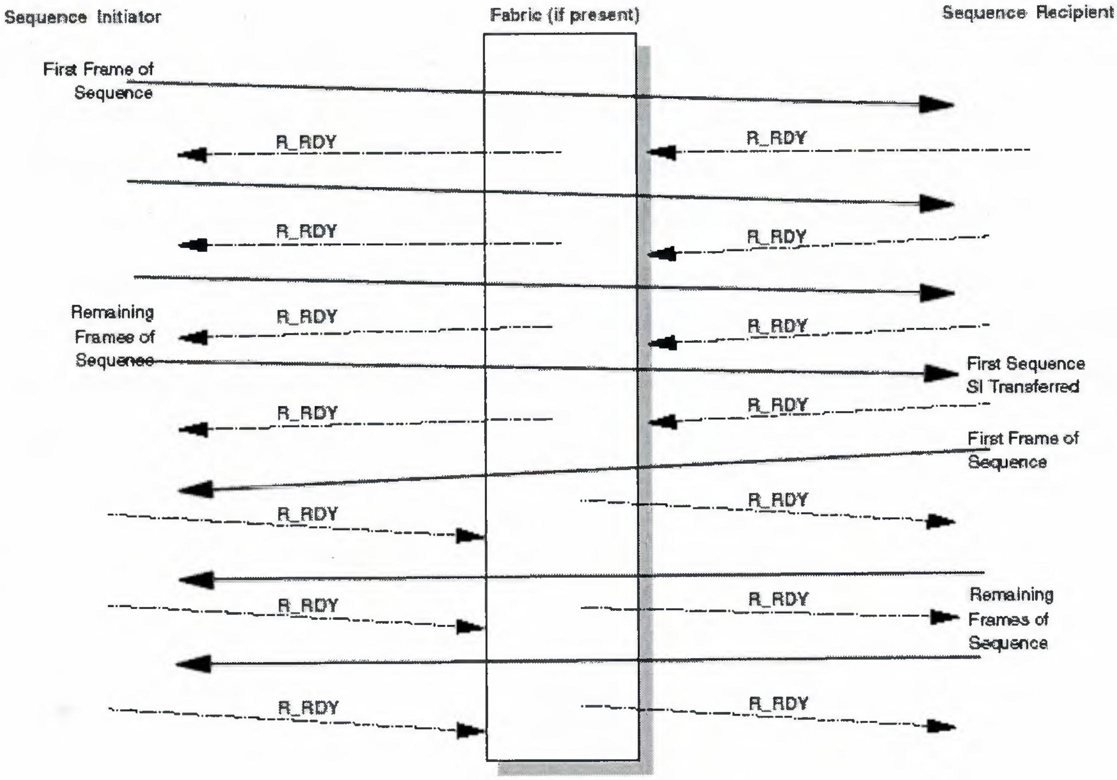


Figure 4.11. Class 3 flow control

Class 4: Fractional bandwidth acknowledged

Class 4 is a connection-oriented class of service which provides a virtual circuit. Virtual connections are established with bandwidth reservation for a predictable quality of service. A Class 4 connection is bi-directional, with one virtual circuit operational in each direction, and it supports a different set of quality of service parameters for each virtual circuit. These quality of service (QoS) parameters include guaranteed bandwidth and bounded end-to-end delay. A quality of service facilitator (QoSF) function is provided within the switch to manage and maintain the negotiated quality of service on each virtual circuit.

A node may reserve up to 256 concurrent Class 4 connections. Separate functions of Class 4 are the setup of the quality of service parameters and the connection itself.

When a Class 4 connection is active, the switch paces frames from the source node to the destination node. Pacing is the mechanism used by the switch to regulate available bandwidth per virtual circuit. This level of control permits congestion management for a switch and guarantees access to the destination node. The switch multiplexes frames belonging to different virtual circuits between the same or different node pairs.

Class 4 service provides in-order delivery of frames. Class 4 flow control is end-to-end and provides guaranteed delivery. Class 4 is ideal for time-critical and real-time applications like video.

We show this in Figure 4.12.

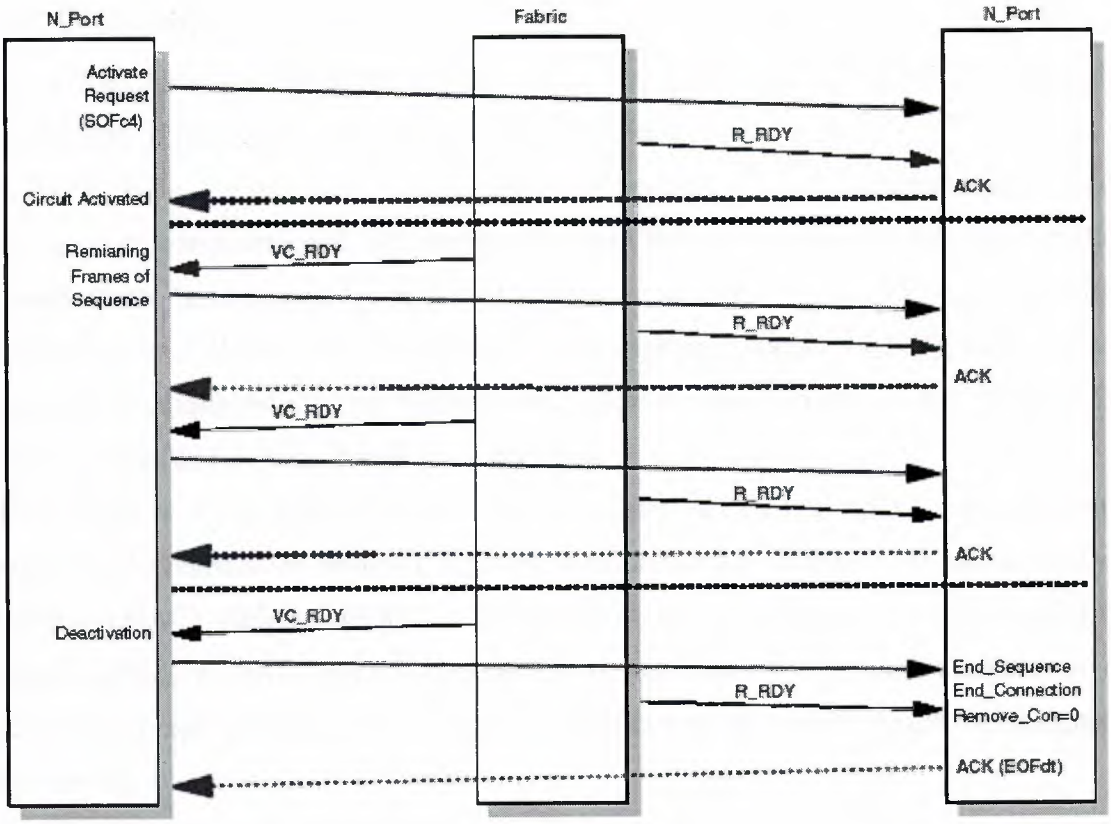


Figure 4.12. Class 4 Flow Control

Class 5: Still under development

Class 5 is still under development. This service allow for simultaneous (isochronous) data transfer to several participants and is especially applicable for audio and video servers in broadcast mode.

Class 6: Uni-directional connection service

Class 6 is similar to Class 1, providing uni-directional connection service. However, Class 6 also provides reliable multicast and pre-emption. Class 6 is ideal for video broadcast applications and real-time systems that move large quantities of data.

4.12 Naming and Addressing

In a Fibre Channel environment the unique identity of participants is maintained through a hierarchy of fixed names and assigned addresses identifiers.

In Fibre Channel terminology, a communicating device is a node. Each node has a fixed 64-bit Node_name assigned by the manufacturer. The node name will be unique if the manufacturer has registered a range of addresses with the IEEE, and so is normally referred to as a World-Wide Name. An N_Port within a parent (WWN) node is also assigned a unique 64-bit Port_Name, which aids the accessibility of the port and is known as the World-Wide Port Name (WWPN).

The WWN is a registered, unique 64-bit identifier assigned to nodes and ports. An example of a registration authority is the registration service support of the Media Access Control (MAC) address associated with the network interface card. In the IEEE understanding, a MAC address consists of 48 bits, 24 of which are assigned to a particular company through the registration process with the remaining 24 bits assigned by the user.

An example of the node and port name correlation is shown in Figure 4.13.

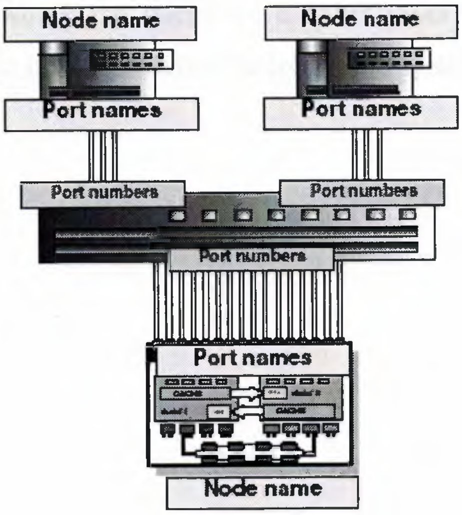


Figure 4.13. Nodes And Ports

For more information on the governing body and the WWN, go to: standards.ieee.org/regauth/oui/index.html This naming convention allows each node and its associated N_Ports to be unique and accessible, even in a complex SANs.

The Fibre Channel naming convention allows either global or locally administered uniqueness to be assigned to a device. However, the administered name or WWN is not used for transporting frames across the network. In addition to a Fibre Channel WWN, a communicating device is

dynamically assigned a 24-bit port address, or N_Port ID that is used for frame routing. As well as providing frame routing optimization, this 24-bit port address strategy removes the overhead of manual administration of addresses by allowing the topology to assign address.

In fabric environments, the switch is responsible for assigning a 24-bit address to each device as it logs on.

Allowing the topology to manage the assignment of addresses has the advantage that control of the addresses is now performed by the entity that is responsible for the routing of information. This means that address assignments can be made in a manner that results in the most efficient

routing of frames within that topology. This approach mimics the behavior of the telephone system, where the telephone number (address) of a particular telephone is determined by where it is attached to the telephone system.

Fibre Channel ports

There is more than one kind of port, though, and its designation represents the use which is being made of it. We show some port designations in Figure 4.14.

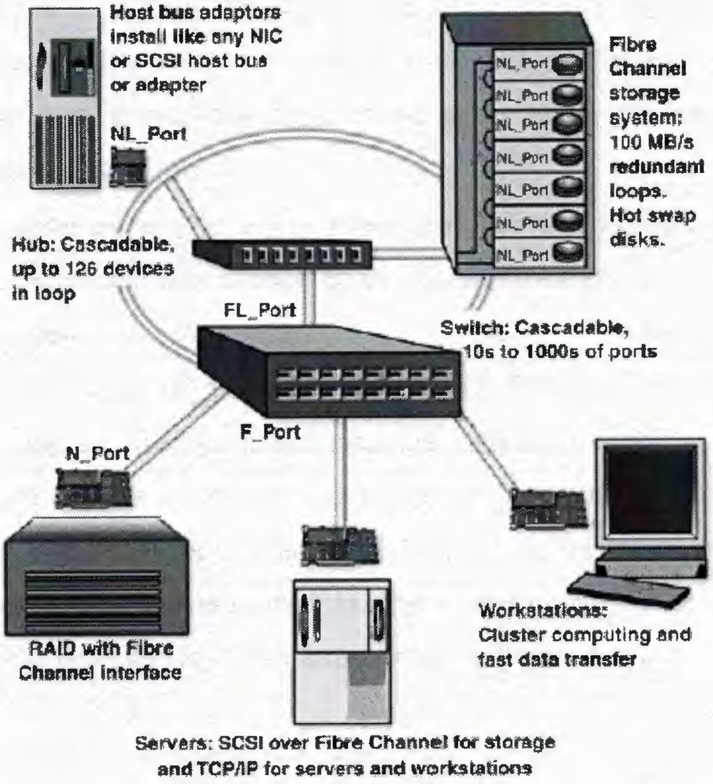

Host bus adaptors install like any NIC or SCSI host bus or adapter
NL_Port
Fibre Channel storage system: 100 MB/s redundant loops. Hot swap disks.
NL_Port
Hub: Cascadable, up to 126 devices in loop
FL_Port
Switch: Cascadable, 10s to 1000s of ports
F_Port
N_Port
RAID with Fibre Channel interface
Workstations: Cluster computing and fast data transfer
Servers: SCSI over Fibre Channel for storage and TCP/IP for servers and workstations

Figure 4.14. Fibre Channel Ports

There are six kinds of ports that we are concerned with in this redbook. They are:

- **Loop port (L_Port)** This is the basic port in a Fibre Channel arbitrated loop (FC-AL) topology. If an N_Port is operating on a loop it is referred to as an NL_Port. If a fabric port is on a loop it is known as an FL_Port. To draw the distinction, throughout this book we will always qualify L_Ports as either NL_Ports or FL_Ports.
- **Node ports (N_Port)** These ports are found in Fibre Channel nodes, which are defined to be the source or destination of information units (IU). I/O devices and

host systems interconnected in point-to-point or switched topologies use N_Ports for their connection. N_Ports can only attach to other N_Ports or to F_Ports.

- **Node-loop ports (NL_Port)** These ports are just like the N_Port described above, except that they connect to a Fibre Channel abritrated loop (FC-AL) topology. NL_Ports can only attach to other NL_Ports or to FL_Ports
- **Fabric ports (F_Port)** These ports are found in Fibre Channel switched fabrics. They are not the source or destination of IU's, but instead function only as a "middle-man" to relay the IUs from the sender to the receiver. F_Ports can only be attached to N_Ports.
- **Fabric-loop ports (FL_Port)** These ports are just like the F_Ports described above, except that they connect to an FC-AL topology. FL_Ports can only attach to NL_Ports.
- **Expansion ports (E_Port)** These ports are found in Fibre Channel switched fabrics and are used to interconnect the individual switch or routing elements. They are not the source or destination of IUs, but instead function like the F_Ports and FL_Ports to relay the IUs from one switch or routing elements to another. E_Ports can only attach to other E_Ports.

We show all these ports and how they interconnect in Figure 4.15.

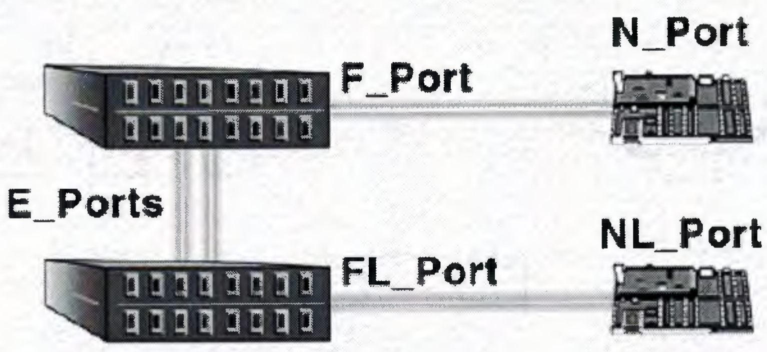


Figure 4.15. Port Interconnections

The Fibre Channel architecture specifies the link characteristics and protocol used between N_Ports, between N_Ports and F_Ports, an between NL_Ports and FL_Ports.

5. THE TECHNICAL TOPOLOGY OF A SAN

5.1 Overview

Fibre Channel provides three distinct and one hybrid interconnection topologies. By having more than one interconnection option available, a particular application can choose the topology that is best suited to its requirements. The three fibre channel topologies are:

- Point-to-point
- Arbitrated loop
- Switched — referred to as a fabric

The three topologies are shown in Figure 5.1.

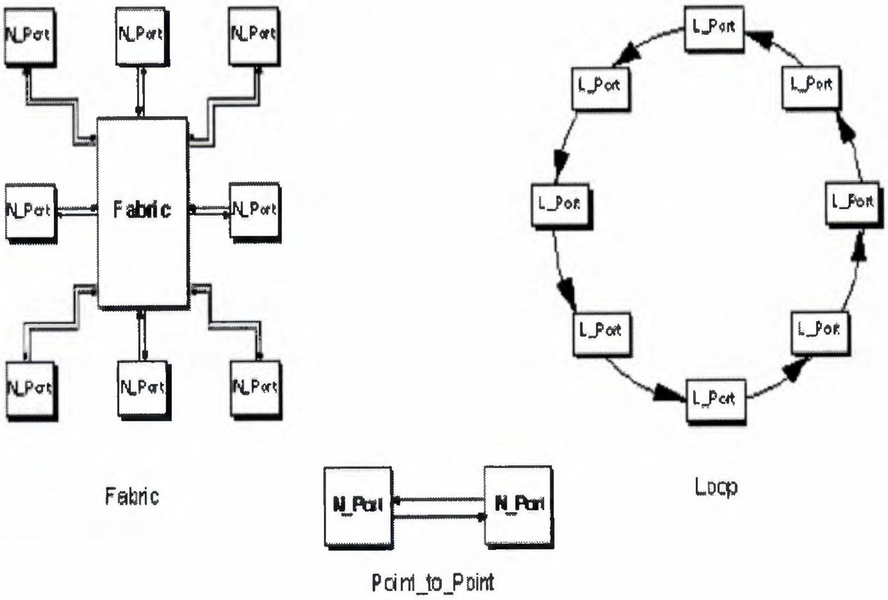


Figure 5.1. SAN Topologies.

5.2 Point-to-Point

A point-to-pointconnection is the simplest topology. It is used when there are exactly two nodes, and future expansion is not predicted. There is no sharing of the media, which allows the devices to use the total bandwidth of the link.

A simple link initialization is needed before communications can begin. We illustrate a simple point-to-point connection in Figure 5.2.

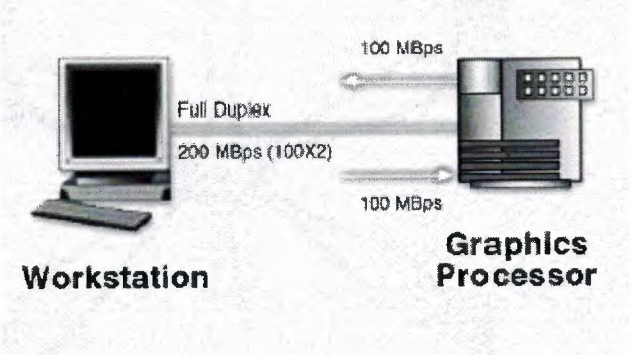


Figure 5.2 Point-To-Point

An extension of the point-to-point topology is the logical start topology. This is a collection of point-to-point topology links and both topologies provide 100 MB/s full duplex bandwidth.

5.3 Arbitrated Loop

The second topology is Fibre Channel Arbitrated Loop (FC-AL). FC-AL is more useful for storage applications. It is a loop of up to 126 nodes (NL_Ports) that is managed as a shared bus. Traffic flows in one direction, carrying data frames and primitives around the loop with a total bandwidth of 100 MB/s. Using arbitration protocol, a single connection is established between a sender and a receiver, and a data frame is transferred around the loop. When the communication comes to an end between the two connected ports, the

loop becomes available for arbitration and a new connection may be established. Loops can be configured with hubs to make connection management easier. Up to 10 km distance is supported by the Fibre Channel standard for both of these configurations. However, latency on the arbitrated loop configuration is affected by the loop size.
A simple loop, configured using a hub, is shown in Figure 5.3.

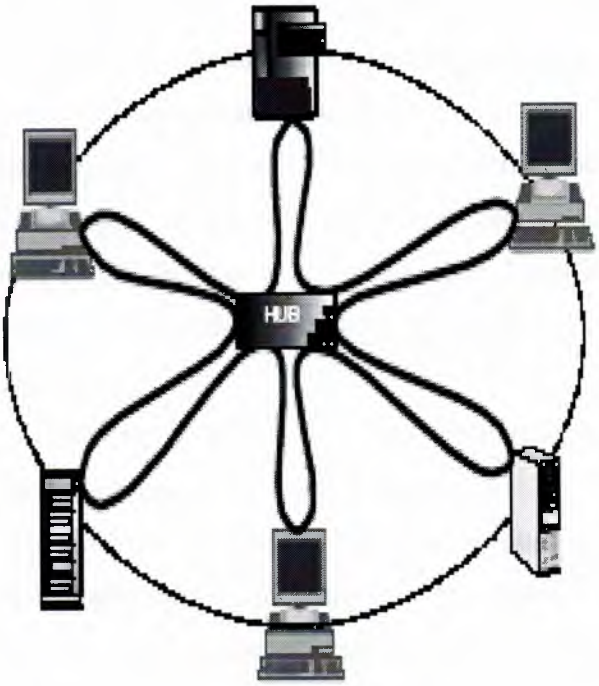


Figure 5.3. Arbitrated loop

5.3.1 Loop Protocols

To support the shared behavior of the arbitrated loop, a number of loop-specific protocols are used. These protocols are used to:

- Initialize the loop and assign addresses
- Arbitrate for access to the loop
- Open a loop circuit with another port in the loop
- Close a loop circuit when two ports have completed their current use of the loop

- Implement the access fairness mechanism to ensure that each port has an opportunity to access the loop

5.3.2 Loop Initialization

Loop initialization is a necessary process for the introduction of new participants on to the loop. Whenever a loop port is powered on or initialized, it executes the loop initialization primitive (LIP) to perform loop initialization.

Optionally, loop initialization may build a positional map of all the ports on the loop. The positional map provides a count of the number of ports on the loop, their addresses and their position relative to the loop initialization master.

Following loop initialization, the loop enters a stable monitoring mode and begins with normal activity. An entire loop initialization sequence may take only a few milliseconds, depending on the number of NL_Ports attached to the loop. Loop initialization may be started by a number of causes. One of the most likely reasons for loop initialization is the introduction of a new device.

For instance, an active device may be moved from one hub port to another hub port, or a device that has been powered on could re-enter the loop.

A variety of ordered sets have been defined to take into account the conditions that an NL_Port may sense as it starts the initialization process. These ordered sets are sent continuously while a particular condition or state exists. As part of the initialization process, loop initialization primitive sequences (referred to collectively as LIPs) are issued. As an example, an NL_Port must issue at least three identical ordered sets to start initialization. An ordered set transmission word always begins with the special character K28.5. Once these identical ordered sets have been sent, and as each downstream device receives the LIP stream, devices enter a state known as open-init. This causes the suspension of any current operation and enables the device for the loop initialization procedure. LIPs are forwarded around the loop until all NL_Ports are in an open-init condition. At this point, the NL_Ports need to be managed. In contrast to a Token-Ring, the Arbitrated Loop has no permanent master to manage the topology. Therefore, loop initialization provides a selection process to determine which device will be the temporary loop master. After the loop master is chosen it assumes the responsibility for directing or managing the rest of the initialization procedure. The loop master also has the responsibility for closing the loop and returning it to normal operation.

Selecting the loop master is carried out by a subroutine known as the Loop Initialization Select Master (LISM) procedure. A loop device can be considered for temporary master by continuously issuing LISM frames that contain a port type identifier and a 64-bit World-Wide Name. For FL_Ports the identifier is x'00' and for NL_Ports it is x'EF'.

When a downstream port receives a LISM frame from a upstream partner, the device will check the port type identifier. If the identifier indicates an NL_Port, the downstream device will compare the WWN in the LISM frame to its own.

The WWN with the lowest numeric value has priority. If the received frame's WWN indicates a higher priority, that is to say it has a lower numeric value, the device stops its LISM broadcast and starts transmitting the received LISM. Had the received frame been of a lower priority, the receiver would have thrown it away and continued broadcasting its own.

At some stage in proceedings, a node will receive its own LISM frame, which indicates that it has the highest priority, and succession to the throne of 'temporary loop master' has taken place. This node will then issue a special ordered set to indicate to the others that a temporary master has been selected.

5.3.3 Hub Cascading

Since an arbitrated loop hub supplies a limited number of ports, building larger loops may require linking another hub. This is called hub cascading. A server with an FC-AL, shortwave, host bus adapter can connect to an FC-AL hub 500 meters away. Each port on the hub can connect to an FC-AL device up to 500 meters away. Cascaded hubs use one port on each hub for the hub-to-hub connection and this increases the potential distance between nodes in the loop by an additional 500 meters. In this topology the overall distance is 1500 meters. Both hubs can support other FC-AL devices at their physical locations. Stated distances assume a 50 micron multimode cable.

5.3.4 Loops

There are two different kinds of loops, the private and the public loop.

5.3.4.1 Private Loop

The private loop does not connect with a fabric, only to other private nodes using attachment points called NL_Ports. A private loop is enclosed and known only to itself. In Figure 5.4 we show a private loop.

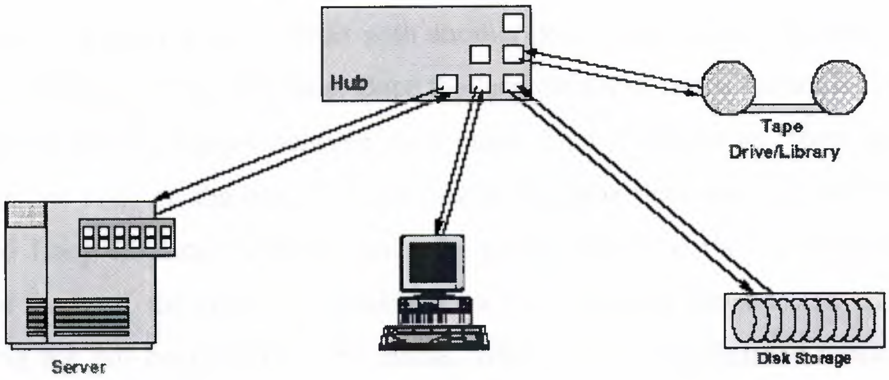


Figure 5.4. Private loop Implementation

5.3.4.2 Public Loop

A public loop requires a fabric and has at least one FL_Port connection to a fabric. A public loop extends the reach of the loop topology by attaching the loop to a fabric. Figure 5.5 hows a public loop.

The Technical Topology Of A SAN

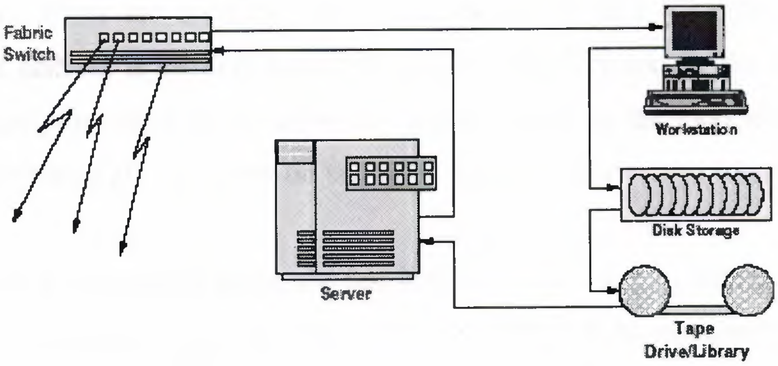


Figure 5.5 Public loop Implementation

5.3.5 Arbitration

When a loop port wants to gain access to the loop, it has to arbitrate. When the port wins arbitration, it can open a loop circuit with another port on the loop; a function similar to selecting a device on a bus interface. Once the loop circuit has been opened, the two ports can send and receive frames between each other. This is known as "loop tenancy". If more than one node on the loop is arbitrating at the same time, the node with the lower Arbitrated Loop Physical Address (AL_PA) gains control of the loop. Upon gaining control of the loop, the node then establishes a point-to-point transmission with another node using the full bandwidth of the media. When a node has finished transmitting its data, it is not required to give up control of the loop. This is a channel characteristic of Fibre Channel. However, there is a "fairness algorithm", which states that a device cannot regain control of the loop until the other nodes have had a chance to control the loop.

5.3.6 Loop Addressing

An NL_Port, like a N_Port, has a 24-bit port address. If no switch connection exists, the two upper bytes of this port address are zeroes (x'00 00') and referred to as a private loop. The devices on the loop have no connection with the outside world. If the loop is attached to a fabric and an NL_Port supports a fabric login, the upper two bytes are assigned a positive value by the switch. We call this mode a public loop.

As fabric-capable NL_Ports are members of both a local loop and a greater fabric community, a 24-bit address is needed as an identifier in the network. In the case of public loop assignment, the value of the upper two bytes represents the loop identifier, and this will be common to all NL_Ports on the same loop that performed login to the fabric.

In both public and private arbitrated loops, the last byte of the 24-bit port address refers to the arbitrated loop physical address (AL_PA). The AL_PA is acquired during initialization of the loop and may, in the case of fabric-capable loop devices, be modified by the switch during login.

The total number of the AL_PAs available for arbitrated loop addressing is 127. This number is based on the requirements of 8b/10b running disparity between frames.

As a frame terminates with an end-of-frame character (EOF) this will force the current running disparity negative. In the Fibre Channel standard each transmission word between the end of one frame and the beginning of another frame should also leave the running disparity negative. If all 256 possible 8-bit bytes are sent to the 8b/10b encoder, 134 emerge with neutral disparity characters. Of these 134, seven are reserved for use by Fibre Channel. The 127 neutral disparity characters left have been assigned as AL_PAs. Put another way, the 127 AL_PA limit is simply the maximum number, minus reserved values, of neutral disparity addresses that can be assigned for use by the loop. This does not imply that we recommend this amount, or load, for a 100MB/s shared transport, but only that it is possible.

Arbitrated Loop will assign priority to AL_PAs, based on numeric value. The lower the numeric value, the higher the priority is. For example, an AL_PA of x'01' has a much better position to gain arbitration over devices that have a lower priority or higher numeric value. At the top of the hierarchy it is not unusual to find servers, but at the lower end you would expect to find disk arrays.

It is the arbitrated loop initialization that ensures each attached device is assigned a unique AL_PA. The possibility for address conflicts only arises when two separated loops are joined together without initialization.

5.3.7 Logins

There are three different types of login for Fibre Channel. These are:

- Fabric login
- Port login
- Process login

Port login

Port login is also known as PLOGI.

Port login is used to establish a session between two N_Ports (devices) and is necessary before any upper level commands or operations can be performed. During the port login, two N_Ports (devices) swap service parameters and make themselves known to each other.

Process login

Process login is also known as PRLI.

Process login is used to set up the environment between related processes on an originating N_Port and a responding N_Port. A group of related processes is collectively known as an image pair. The processes involved can be system processes, system images, such as mainframe logical partitions, control unit images, and FC-4 processes. Use of process login is optional from the perspective of Fibre Channel FC-2 layer, but may be required by a specific upper-level protocol as in the case of SCSI-FCP mapping.

We show Fibre Channel logins in Figure 5.6.

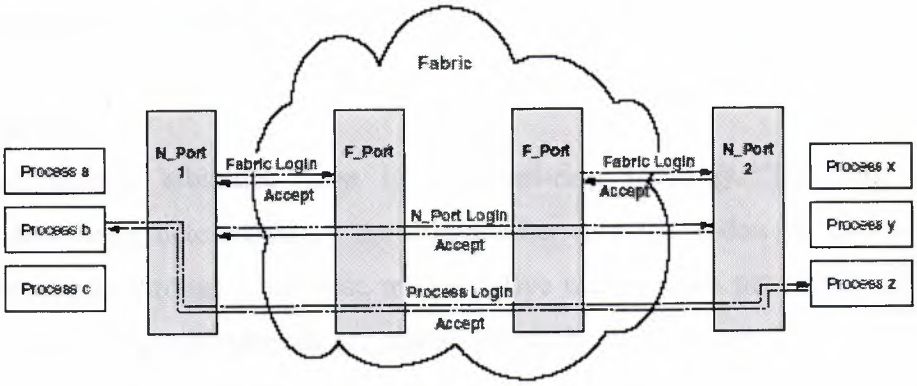


Figure 5.6 Fibre Channel logins

5.3.8 Closing a Loop Circuit

When two ports in a loop circuit complete their frame transmission, they may close the loop circuit to allow other ports to use the loop. The point at which the loop circuit is closed depends on the higher-level protocol, the operation in progress, and the design of the loop ports.

5.3.9 Supported Devices

An arbitrated loop may support a variety of devices, including HBAs installed in the following servers:

- Individual Fibre Channel disk drives
- JBOD
- Fibre Channel RAID
- Native Fibre Channel tape sub-systems
- Fibre Channel to SCSI bridges

5.3.10 Broadcast

Arbitrated loop, in contrast to Ethernet, is a non-broadcast transport. When an NL_Port has successfully won the right to arbitration, it will open a target for frame transmission. Any subsequent loop devices in the path between the two will see the frames and forward them on to the next node in the loop.

It is this non-broadcast nature of arbitrated loop, by removing frame handling overhead from some of the loop, which enhances performance.

5.3.11 Distance

As stated before, arbitrated loop is a closed-ring topology. The total distance requirements being determined by the distance between the nodes. At gigabit speeds, signals propagate through fiber-optic media at five nanoseconds per meter and through copper media at four nanoseconds per meter. This is the delay factor.

Calculating the total propagation delay incurred by the loop's circumference is achieved by multiplying the length — both transmit and receive — of copper and fiber-optic cabling deployed by the appropriate delay factor. For example, a single 10 km link to an NL_Port would cause a 50 microsecond (10 km x 5 nanoseconds delay factor) propagation delay in each direction and 100 microseconds in total. This equates to 1 MB/s of bandwidth used to satisfy the link.

5.3.12 Bandwidth

For optical interconnects for SANs, the bandwidth requirements are greatly influenced by the capabilities of:

- The system buses
- Network switches
- The interface adapters that interface with them
- Traffic locality

The exact bandwidth required is somewhat dependent on implementation, but are currently in the range of 100 to 1000 MB/s. Determining bandwidth requirements is difficult and there is no exact science that can take into account the unpredictability of sporadic bursts of data, for example. Planning bandwidth based on peak requirements could be wasteful. Designing for sustained bandwidth requirements, with the addition of safety margins, may be less wasteful.

5.4 Switched Fabric

The third topology used in SAN implementations is Fibre Channel Switched Fabric (FC-SW). A Fibre Channel fabric is one or more fabric switches in a single, sometimes extended, configuration. Switched fabrics provide full 100MB/s bandwidth per port, compared to the shared bandwidth per port in Arbitrated Loop implementations.

If you add a new device into the arbitrated loop, you further divide the shared bandwidth. However, in a switched fabric, adding a new device or a new connection between existing ones actually increases the bandwidth. For example, an 8-port switch with three initiators and three targets can support three concurrent 100 MB/s conversations or a total 300 MB/s throughput (600 MB/s if full-duplex applications were available). A switched fabric configuration is shown in Figure 5.7.

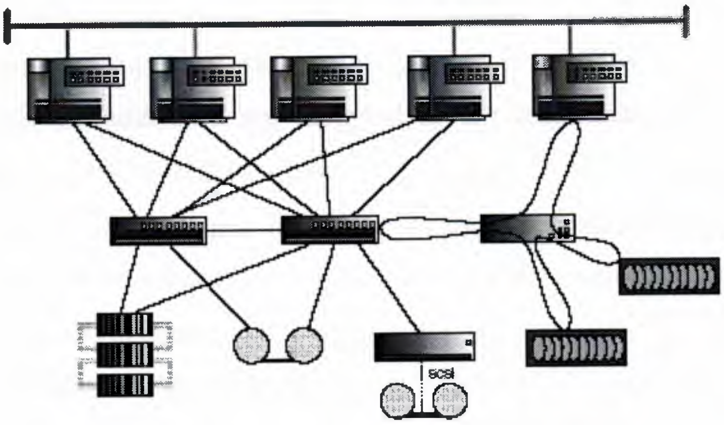


Figure 5.7 Sample Switched Fabric Configuration

5.4.1 Addressing

This ID is called the World Wide Name (WWN), This WWN is a 64-bit address and if two WWN addresses are put into the frame header, this leaves 16 bytes of data just for identifying destination and source address. So 64-bit addresses can impact routing performance.

Because of this there is another addressing scheme used in Fibre Channel networks. This scheme is used to address the ports in the switched fabric. Each port in the switched fabric has its own unique 24-bit address. With this 24-bit addressing scheme we get a smaller frame header and this can speed up the routing process. With this frame header and routing logic the Fibre Channel fabric is optimized for high-speed switching of frames.

With a 24-bit addressing scheme this allows for up to 16 million addresses, which is an address space larger than any practical SAN design in existence in today's world. Who knows what the future will bring? Maybe Fibre Channel addressing will have the same problems in the future as the internet does today, which is a lack of addresses. This 24-bit addressing has to be connected with the 64-bit addressing associated with World Wide Names. We explain this in the section that follows.

5.4.2 Name and Addressing

The 24-bit address scheme also removes the overhead of manual administration of addresses by allowing the topology itself to assign addresses. This is not like WWN addressing, in which the addresses are assigned to the manufacturers by the IEEE standards committee, and are built in to the device at build time, similar to naming a child at birth. If the topology itself assigns the 24-bit addresses, then somebody has to be responsible for the addressing scheme from WWN addressing to port addressing.

In the switched fabric environment, the switch itself is responsible for assigning and maintaining the port addresses. When the device with its WWN is logging into the switch on a specific port, the switch will assign the port address to that port and the switch will also maintain the correlation between the port address and the WWN address of the device on that port. This function of the switch is implemented by using a Simple Name Server (SNS). The Simple Name Server is a component of the fabric operating system, which runs inside the switch. It is essentially a database of objects in which fabric-attached device registers its values.

Dynamic addressing also removes the potential element of human error in address maintenance, and provides more flexibility in additions, moves, and changes in the SAN.

5.4.2.1 Port Address

A 24-bit port address consists of three parts:

- Domain (bits from 23 to 16)
- Area (bits from 15 to 08)
- Port or arbitrated loop physical address - AL_PA (bits from 07 to 00)

We show how the address is built up in Figure 5.8.

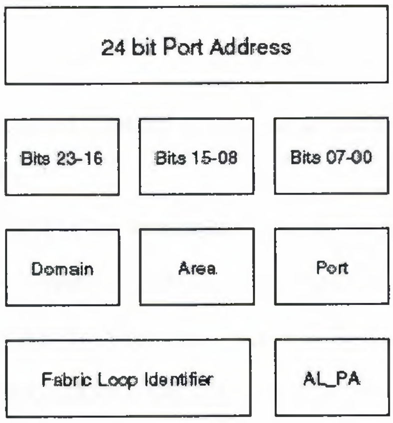


Figure 5.8. Fabric Port Address

We explain the significance of some of the bits that make up the port address in the following sections.

Domain

The most significant byte of the port address is the domain. This is the address of the switch itself. One byte allows up to 256 possible addresses. Because some of these are reserved (like the one for broadcast) there are only 239 addresses actually available. This means that you can have as many as 239 switches in your SAN environment. The domain number allows each switch to have a unique identifier if you have multiple interconnected switches in your environment.

Area

The area field provides 256 addresses. This part of the address is used to identify the individual FL_Ports supporting loops or it can be used as the identifier for a group of F_Ports; for example, a card with more ports on it. This means that each group of ports has a different area number, even if there is only one port in the group.

Port

The final part of the address provides 256 addresses for identifying attached N_Ports and NL_Ports. To arrive at the number of available addresses is a simple calculation based on:

Domain x Area x Ports

This means that there are 239 x 256 x 256 = 15,663,104 addresses available.

5.4.3 Fabric Login

After the fabric capable Fibre Channel device is attached to a fabric switch, it will carry out a fabric login (FLOGI).

Similar to port login, FLOGI is an extended link service command that sets up a session between two participants. With FLOGI a session is created between an N_Port or NL_Port and the switch. An N_Port will send a FLOGI frame that contains its Node Name, its N_Port Name, and service parameters to a well-known address of 0xFFFFFE.

A public loop NL_Port first opens the destination AL_PA 0x00 before issuing the FLOGI request. In both cases the switch accepts the login and returns an accept (ACC) frame to the sender. If some of the service parameters requested by the N_Port or NL_Port are not supported, the switch will set the appropriate bits in the ACC frame to indicate this. When the N_Port logs in it uses a 24-bit port address of 0x000000. Because of this the fabric is allowed to assign the appropriate port address to that device, based on the Domain-Area-Port address format. The newly assigned address is contained in the ACC response frame.

When the NL_Port logs in a similar process starts, except that the least significant byte is used to assign AL_PA and the upper two bytes constitute a fabric loop identifier. Before an NL_Port logs in it will go through the LIP on the loop, which is started by the FL_Port, and from this process it has already derived an AL_PA. The switch then decides if it will accept this AL_PA for this device or not. If not a new AL_PA is assigned to the NL_Port, which then causes the start of another LIP. This ensures that the switch assigned AL_PA does not conflict with any previously selected AL_PAs on the loop.

After the N_Port or public NL_Port gets its fabric address from FLOGI, it needs to register with the SNS. This is done with port login (PLOGI) at the address 0xFFFFFC. The device may register values for all or just some database objects, but the most useful are its 24-bit port address, 64-bit Port Name (WWPN), 64-bit Node Name (WWN), class of service parameters, FC-4 protocols supported, and port type, such as N_Port or NL_Port.

5.4.4 Private Devices on NL_Ports

It is easy to explain how the port to World Wide Name address resolution works when a single device from an N_Port is connected to an F_Port, or when a public NL_Port device is connected to FL_Port in the switch. The SNS will add an entry for the device World Wide Name and connects that with the port address which is selected from the selection of free port addresses for that switch. Problems may arise when a private Fibre Channel device is attached to the switch. Private Fibre Channel devices were designed to only to work in private loops.

When the arbitrated loop is connected to the FL_Port, this port obtains the highest priority address in the loop to which it is attached (0x00). Then the FL_Port performs a LIP. After this process is completed, the FL_Port registers all devices on the loop with the SNS. Devices on the arbitrated loop use only 8-bit addressing, but in the switched fabric, 24-bit addressing is used. When the FL_Port registers the devices on the loop to the SNS, it adds two most significant bytes to the existing 8-bit address.

The format of the address in the SNS table is 0xPPPPLL, where the PPPP is the two most significant bytes of the FL_Port address and the LL is the device ID on the arbitrated loop which is connected to this FL_Port. Modifying the private loop address in this fashion, all private devices can now talk to all public devices, and all public devices can talk to all private devices. Because we have stated that private devices can only talk to devices with private addresses, some form of translation must take place. We show an example of this in Figure 5.9.

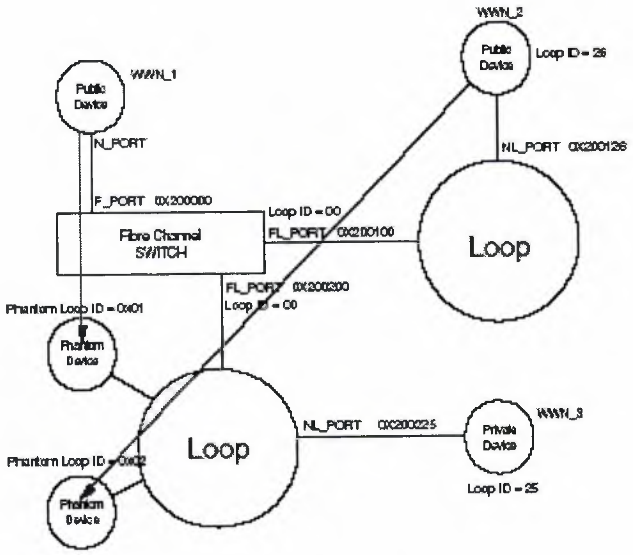


Figure 5.9 Arbitrated loop Address Translation

As you can see, we have three devices connected to the switch:

- Public device N_Port with WWN address WWN_1 on F_Port with the port address 0x200000
- Public device NL_Port with WWN address WWN_2 on FL_Port with the port address 0x200100. The device has AL_PA 0x26 on the loop which is attached on the FL_Port
- Private device NL_Port with WWN address WWN_3 on FL_Port with the port address 0x200200. The device has AL_PA 0x25 on the loop which is attached to the FL_Port

After all FLOGI and PLOGI functions are performed the SNS will have the entries shown in Table 5.3.

Table 5.3 SNS Entries

24 bit port address	WWN	FL_Port address
0x200000	WWN_1	n/a
0x200126	WWN_2	0x200100
0x200225	WWN_3	0x200200

We now explain some possible scenarios.

Public N_Port device accesses private NL_Port device

The communication from device to device starts with PLOGI to establish a session. When a public N_Port device wants to perform a PLOGI to a private NL_Port device, the FL_Port on which this private device exists will assign a "phantom" private address to the public device. This phantom address is known only inside this loop, and the switch keeps track of the assignments.

In our example, when the WWN_1 device wants to talk to the WWN_3 device, the following, shown in Table 5.4, is created in the switch.

Table 5.4 Phantom addresses

Switch port address	Phantom Loop Port ID
0x200000	0x01
0x200126	0x02

When the WWN_1 device enters into the loop it represents itself with AL_PA ID 0x01 (its phantom address). All private devices on that loop use this ID to talk to that public device. The switch itself acts as a proxy, and translates addresses in both directions.

Usually the number of phantom addresses is limited, and this number of phantom addresses decreases the number of devices allowed in the Arbitrated loop. For example, if the number of phantom addresses is 32 this limits the number of physical devices in the loop to 126 - 32 = 94.

Public N_Port device accesses public NL_Port device

If an N_Port public device wants to access an NL_Port public device, it simply performs a PLOGI with the whole 24-bit address.

Private NL_Port device accesses public N_Port or NL_Port device

When a private device needs to access a remote public device, it uses the public device's phantom address. When the FL_Port detects the use of a phantom AL_PA ID, it translates that to a switch port ID using its translation table similar to that shown in Table 5.4

5.4.5 Quick Loop

As we have already explained above, private devices can cooperate in the fabric using translative mode. However, if you have a private host (server), this is not possible. To solve this, switch vendors, including IBM, support a QuickLoop feature. The QuickLoop feature allows the whole switch or just a set of ports to operate as an arbitrated loop. In this mode, devices connected to the switch do not perform a fabric login, and the switch itself will emulate the loop for those devices. All public devices can still see all private devices on the QuickLoop in the translative mode.

5.4.6 Switching Mechanism and Performance

In a switched fabric, a "cut-through" switching mechanism is used. This is not unique to switched fabrics and it is also used in Ethernet switches. The function is to speed packet routing from port to port.

When a frame enters the switch, cut-through logic examines only the linklevel destination ID of the frame. Based on the destination ID, a routing decision is made, and the frame is switched to the appropriate port by internal routing logic contained in the switch. It is this cut-through which increases performance by reducing the time required to make a routing decision. The reason for this is that the destination ID resides in the first four bytes of the frame header, and this allows the cut-through to be accomplished quickly. A routing decision can be made at the instant the frame enters the switch.

An important criterion in selecting a switch is the number of frames that can be buffered on the port. During periods of high activity and frame movement, the switch may not be able to transmit a frame to its intended destination. This is true if two ports are sending data to the same destination. Given this situation, but depending on the class of service, the switch may sacrifice the frames it is not able to process. Not only does frame buffering reduce this likelihood, it also enhances performance.

Another great performance improvement can be realized in the way in which the 24-bit port address is built. Because the address is divided into domain, area and port, it is possible to make the routing decision on a single byte. An example of this would be if the domain number of the destination address indicates that the frame is intended for a different switch, the routing process can forward the frame to the appropriate interconnection without the need to process the entire 24-bit address and the associated overhead.

5.4.7 Data Path in Switched Fabric

A complex switched fabric can be created by interconnecting Fibre Channel switches. Switch to switch connections are performed by E_Port connections. This means that if you want to interconnect switches they need to support E_Ports. Switches may also support multiple E_Port connections to expand the bandwidth.

In such a configuration with interconnected switches, known as a meshed topology, multiple paths from one N_Port to another can exist. An example of a meshed topology is shown in Figure 5.10.

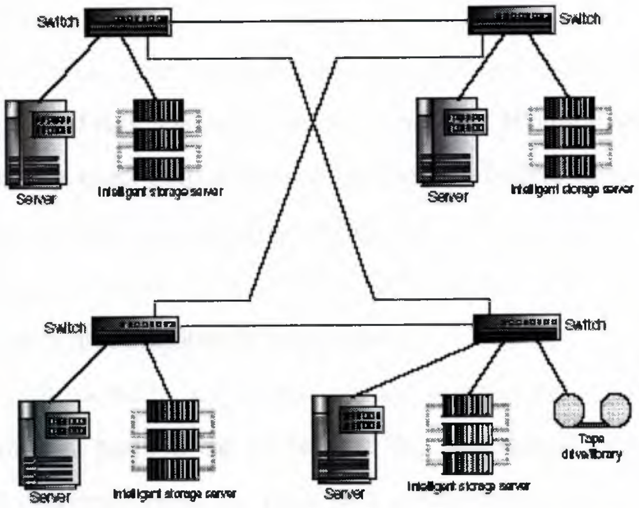

Switch
Switch
Server
Intelligent storage server
Server
Intelligent storage server
Switch
Switch
Server
Intelligent storage server
Server
Intelligent storage server
Tape drive/library

Figure 5.10 Meshed Topology Switched Fabric

5.4.7.1 Spanning Tree

In case of failure, it is important to consider having an alternative path between source and destination available. This will allow the data still to reach its destination. However, having different paths available could lead to the delivery of frames being out of the order of transmission, due to a frame taking a different path and arriving earlier than one of its predecessors.

A solution to this, which can be incorporated into the meshed fabric, is called a spanning tree and is an IEEE 802.1 standard. This means that switches keep to certain paths as the spanning tree protocol will block certain paths to produce a simply connected active topology. Then the shortest path in terms of hops is used to deliver the frames and, most importantly, only one path is active at a time. This means that all associated frames go over the same path to the destination. The paths that are blocked can be held in reserve and used only if, for example, a primary path fails. The fact that one path is active at a time means that in the case of a meshed fabric, all frames will arrive in the expected order.

5.4.7.2 Path Selection

For path selection, link state protocols are popular and extremely effective in today's networks. Examples of link state protocol are OSPF for IP and PNNI for ATM.

The most commonly used path selection protocol is Fabric Shortest Path First (FSPF). This type of path selection is usually performed at boot time and no configuration is needed. All paths are established at start time and only if the inter switch link (ISL) is broken or added will reconfiguration take place.

In the case that multiple paths are available if the primary path goes down, the traffic will be rerouted to another path. If the route fails this can lead to congestion of frames, and any new frames delivered over the new path could potentially arrive at the destination first. This will cause an out of sequence delivery.

One possible solution for this is to prevent the activation of the new route for a while, (this can be configured from milliseconds to a few seconds), so the congested frames are either delivered or rejected. Obviously, this can slow down the routing, so it should only be used when the devices connected to the fabric are not in a position to, or cannot tolerate occasional out of order delivery. For instance, video can tolerate out of sequence delivery, but financial and commercial data cannot.

But today, Fibre Channel devices are much more sophisticated, and this is a feature that is not normally required. FSPF allows a fabric still to benefit from load balancing the delivery of frames by using multiple paths.

5.4.7.3 Route Definition

Routes are usually dynamically defined. The fabric itself usually keeps only eight possible paths to the destination.

Static routes can also be defined. In the event that a static route fails, a dynamic route will take over. Once the static route becomes available, frames will return to utilizing that route.

If dynamic paths are used, FSPF path selection is used. This guarantees that only the shortest and fastest paths will be used for delivering the frames. We show an example of FSPF in Figure 5.11.

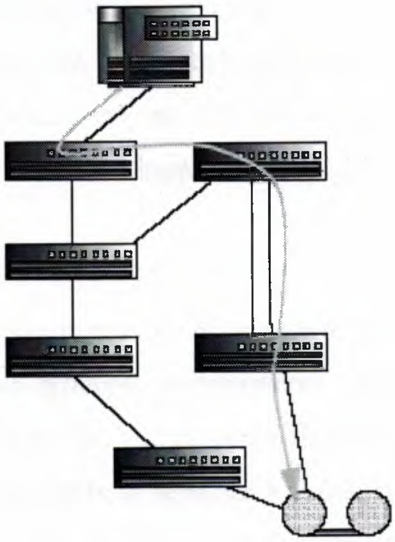


Figure 5.11 Fabric Shortest Path First

5.4.8 Adding New Devices

Switched fabrics, by their very nature, are dynamic environments. They can handle topology changes as new devices are attached, or previously active devices are removed or taken offline. For these reasons it is important that notification of these types of events can be provided to participants (nodes) in the switched fabric.

Notification is provided by two functions:

- State Change Notification - SCN
- Registered State Change Notification - RSCN

These two functions are not obligatory, so each N_Port or NL_Port must register its interest in being notified of any topology changes, or if another device alters its state.

The original SCN service allowed an N_Port to send a notification change directly to another N_Port. This is not necessarily an optimum solution, as no other participants on the fabric will know about this change. RSCN offers a solution to this and will inform all registered devices about the change.

Perhaps the most important change that you would want to be notified about, is when an existing device goes offline. This information is very meaningful for participants which communicate with that device. For example, a server in the fabric environment would want to know if their resources are powered off

or removed, or as and when new resources became available for its use.

Changed notification provides the same functionality for the switched fabric as loop initialization provides for arbitrated loop.

5.4.9 Zoning

Zoning allows for finer segmentation of the switched fabric. Zoning can be used to instigate a barrier between different environments. Only the members of the same zone can communicate within that zone and all other attempts from outside are rejected.

For example, it may be desirable to separate a Windows NT environment from a UNIX environment. This is very useful because of the manner in which Windows attempts to claim all available storage for itself. Because not all storage devices are capable of protecting their resources from any host seeking for available resources, it makes sound business sense to protect the environment in another manner.

Looking at zoning in this way, it could also be considered as a security feature and not just for separating environments. Zoning could also be used for test and maintenance purposes. For example, not many enterprises will mix their test and maintenance environments with their production environment. Within

a fabric, you could easily separate your test environment from your production bandwidth allocation on the same fabric using zoning.

We show an example of zoning in Figure 5.12.

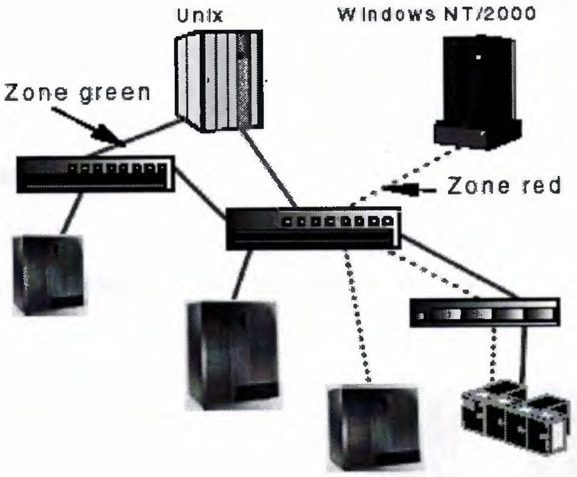


Figure 5.12 Zoning

Zoning also introduces the flexibility to manage a switched fabric to meet different user groups objectives.

5.4.10 Implementing Zoning

Zoning can be implemented in two ways:

- Hardware zoning
- Software zoning

Hardware zoning

Hardware zoning is based on the physical fabric port number.

The members of a zone are physical ports on the fabric switch. It can be implemented in the following configurations:

- One to one
- One tomany
- Many to many

A single port can also belong to multiple zones. We show an example of hardware zoning in Figure 5.13.

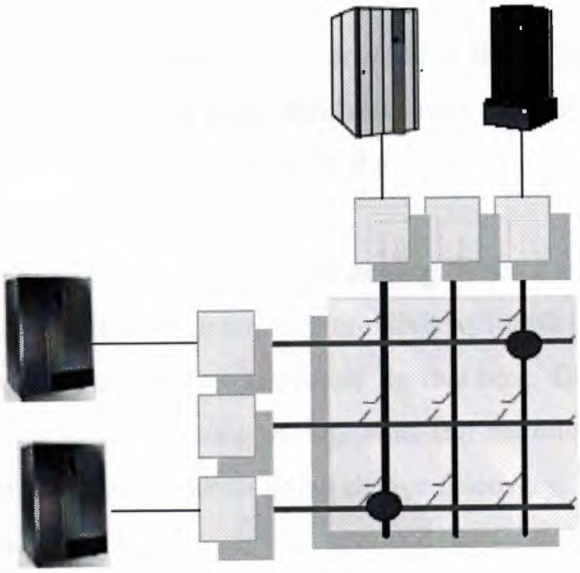


Figure 5.13 Hardware Zoning

One of the disadvantages of hardware zoning is that devices have to be connected to a specific port, and the whole zoning configuration could become unusable when the device is connected to a different port. In cases where the device connections are not permanent the use of software zoning is recommended.

The advantage of hardware zoning is that it can be implemented into a routing engine by filtering. As a result, this kind of zoning has a very low impact on the performance of the routing process.

Software zoning

Software zoning is implemented within the SNS running inside the fabric switch. When using software zoning the members of the zone can be defined with:

- Node WWN
- PortWWN

Usually zoning software also allows you to create symbolic names for the zone members and for the zones themselves.

The number of members possible in a zone is limited only by the amount of memory in the fabric switch. A member can belong to multiple zones. You can define multiple sets of zones for the fabric, but only one set can be active at any time. You can activate another zone set any time you want, without the need to power down the switch.

With software zoning there is no need to worry about the physical connections to the switch. If you use WWNs for the zone members, even when a device is connected to another physical port, it will still remain in the same zoning definition, because the device's WWN remains the same.

There is a potential security leak with software zoning. When a specific host logs into the fabric and asks for available storage devices, the SNS will look into the software zoning table to see which storage devices are allowable for that host. The host will only see the storage devices defined in the software zoning table. But the host can also make a direct connection to the storage device, while doing device discovery, without asking SNS for the informationit has.

5.4.11 LUN Masking

Another approach to securing storage devices from hosts wishing to take over already assigned resources is logical unit number (LUN) masking. Every storage device offers its resources to the hosts by means of LUNs. For example, each partition in the storage server has its own LUN. If the host (server) wants to access the storage, it needs to request access to the LUN in the storage device. The purpose of LUN masking is to control access to the LUNs. The storage device itself accepts or rejects access requests from different hosts. The user defines which hosts can access which LUN by means of the storage device control program. Whenever the host accesses a particular LUN, the storage device will check its access list for that LUN, and it will allow or disallow access to the LUN.

5.4.12 Expanding the Fabric

As the demand for the storage grows, a switched fabric can be expanded to service these needs. Not all storage requirements can be satisfied with fabrics alone. For some applications, the 100 MB/s per port and advanced services are overkill, and they amount to wasted bandwidth and unnecessary cost. When you design a storage network you need to consider the application's needs and not just rush to implement the latest technology available. SANs are often combinations of switched fabric and arbitrated loops.

5.4.12.1 Cascading

Expanding the fabric is called switch cascading. Cascading is basically interconnecting Fibre Channel switches. The cascading of switches provides the following benefits to a SAN environment:

- The fabric can be seamlessly extended. Additional switches can be added to the fabric, without powering down existing fabric.
- You can easily increase the distance between various SAN participants.
- By adding more switches to the fabric, you increase connectivity by providing more available ports.
- Cascading provides high resilience in the fabric.

- With Inter Switch Links (ISL) you can increase the bandwidth. The frames between the switches are delivered over all available data paths. So the more ISL you create, the faster the frame delivery will be, but careful consideration must be employed to ensure that a bottleneck is not introduced.
- When the fabric grows, the SNS is fully distributed across all the switches in fabric.
- With cascading, you also provide greater fault tolerance within the fabric.

5.4.12.2 Hops

As we stated in 5.3.2, the maximum number of switches allowed in the fabric is 239. The other limitation is that only seven hops are allowed between any source and destination. However, this is likely to change between vendors and over time.

We show a sample configuration that illustrates this in Figure 5.14.

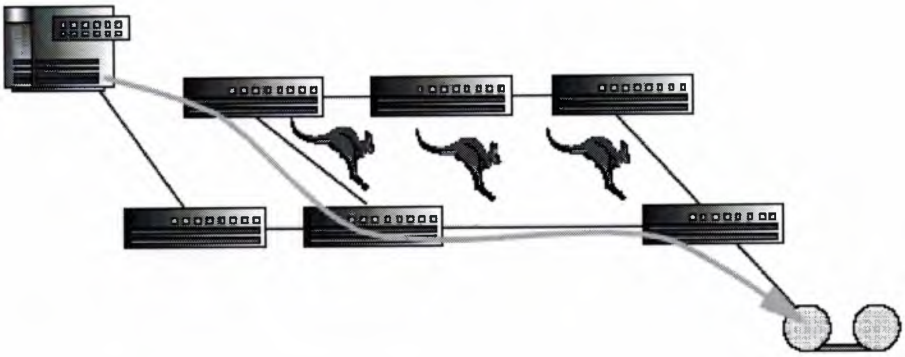


Figure 5.14 Cascading In Switched Fabric

The hop count limit is set by the fabric operating system and is used to derive a frame holdtime value for each switch. This holdtime value is the maximum amount of time that a frame can be held in a switch before it is dropped (Class 3) or the fabric is busy (F_BSY, Class 2) is returned. A frame would be held if its destination port is not available. The holdtime is derived from a formula using the error detect time-out value (E_D_TOV) and the resource allocation time-out value (R_A_TOV).

The value of seven hops is not 'hard-coded', and if manipulation of E_D_TOV or R_A_TOV was to take place, the reasonable limit of seven hops could be exceeded. However, be aware that this seven hop suggestion was not a limit that was arrived at without careful consideration of a number of factors. In the future the number of hops is likely to increase.

CONCLUSION

We have discussed how GPS technique is used in routing protocol in ad hoc networks, Using GIS technique is more complicated than the use of GPS and can cause a big delay. Also some quantitative test must be done to see the efficiency of using GIS in routing protocol.

In ad hoc network each node acts as a host and router at the same time, so each node must have a routing table that contain information about the network and to which node it must forward the packet. In this chapter we will discuss in details how route can be found in ad hoc networks.

The routing protocols meant for wired networks can not be used for mobile ad hoc networks because of the mobility of networks. The ad hoc routing protocols can be divided into two classes: - table-driven and on-demand. This chapter discusses routing protocols belonging to each category.

Computer networks can be used for numerous services, both for companies and for individuals. Networks can be divided up into LANs, MANs and WANs, with there own characteristics, technologies and speeds. LANs cover a building and operate at high speeds. MANs cover a city, for example, the cable television system, which is now used by many people to access the internet. WANs cover a country or continent. Wireless networks are becoming extremely popular, especially wireless LANs.

REFERENCES

[1] Berth Basch.E.E, Optical Fiber Transmission, McGraw-Hill, NewYork NY, 1989.

[2] Allen H. Cherin, An Introduction to Optical Fibers, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[3] John. M. Senior, Fiber Optics Cable, McGraw-Hill, New York NY, 1997.

[4] Academic writing. http://www.lascomm.com/tutorials/tut_fobasics.htm.

[5] Documentation, http://www.cisco.com/univercd/home.htm.

[6] Documentation, http://www.commspecial.com/introductiontofiberoptics.htm.

[7] Documentation, http://www.connectworld.net/computer/kitco/connectors.htm.

[8] Search, http://www.yahoo.com.

[9] Search, http://www.google.com.