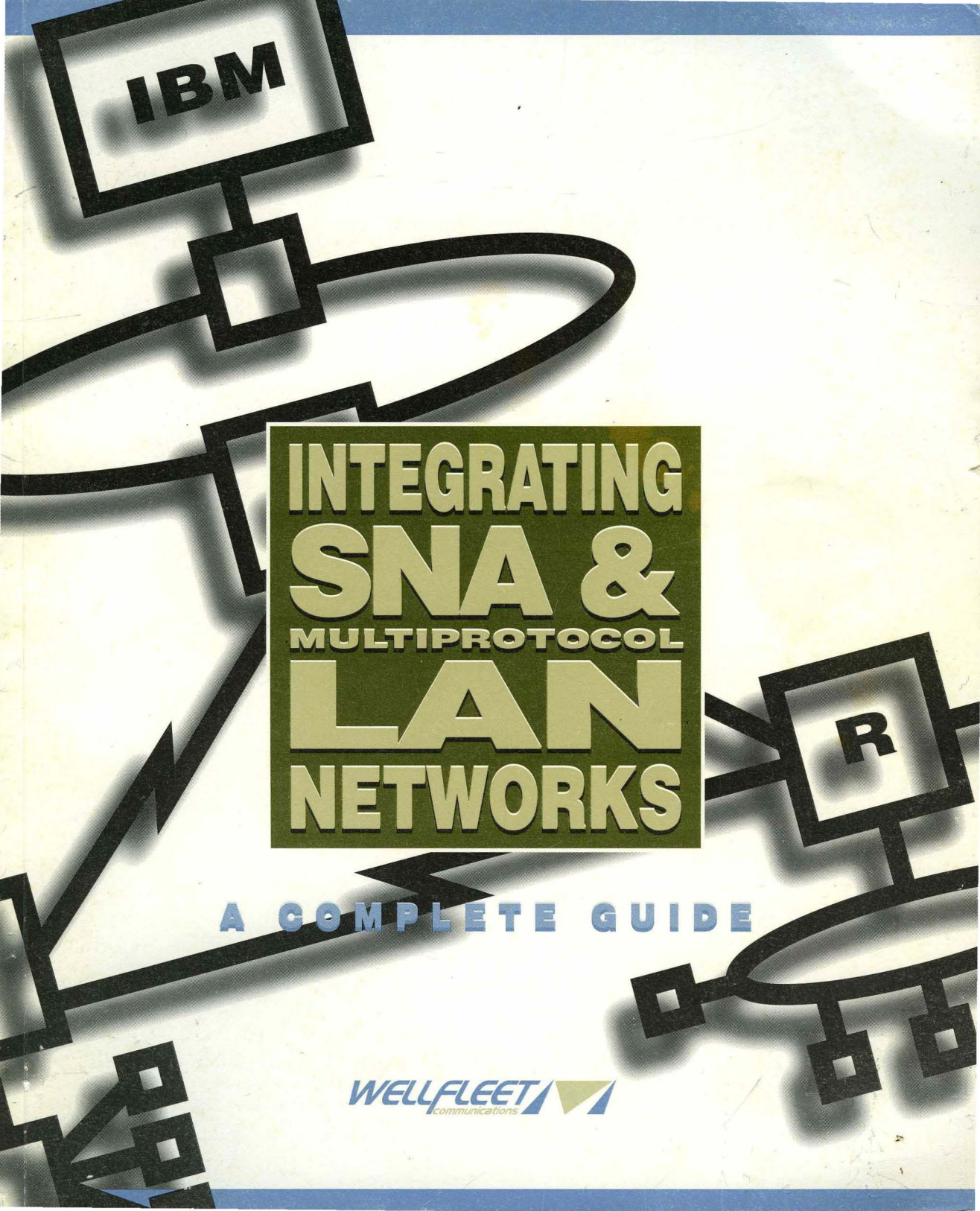


INTEGRATING SNA & MULTIPROTOCOL LAN NETWORKS

A COMPLETE GUIDE

WELLFLEET communications

INTEGRATING SNA AND MULTIPROTOCOL LAN NETWORKS

A COMPLETE GUIDE

June 1993 Edition

WELLFLEET COMMUNICATIONS, INC.
8 Federal Street, Billerica, Massachusetts 01821

Printed in USA

Integrating SNA and Multiprotocol LAN Networks was written by Steven S. King, with assistance from Karen Barton, Scott Barvick, John Brewer, Jennifer Davies, Jim Hourihan, Alan Rosenberg, and Sheryl Schultz.

The author would like to thank the following network professionals for their gracious assistance:

Michael E. Bowman, Netlink Inc.

Louise Herndon Wells, SNA Perspective

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE LIST CONTINUED

FIGURE LIST CONTINUED

INTRODUCTION

CAN THINGS EVER BE BETTER THAN THEY WERE IN THE GOOD OLD DAYS OF NETWORKING?

It used to be, if you wanted to build an enterprise network, you called a single vendor and you got a single solution — typically a host-based, SNA architecture. In centralized transaction-oriented applications, SNA networks deliver data reliably and predictably. During the seventies and eighties, SNA's heyday, there wasn't much to compare to the monolithic corporate network, so network managers and IS executives probably didn't realize how easy they had it.

In the nineties, multivendor, multiprotocol internetworking has all but shattered the single-vendor, single-protocol paradigm. This is as it should be, for the potential of internetworking far outweighs the complexities it introduces. Internetworking brings an ever increasing wealth of network resources to the enterprise, but for network professionals who work daily in the present climate of diversity and change, the "good old days" can start to look pretty good.

The growth of traditional SNA networks has been driven by the connectivity needs of host-based corporate applications — order processing, accounting, claims handling, reservations, inventory management, and a large array of key financial and planning software programs. Internetworks, in contrast, are driven by PCs and workstations running office-automation or technical applications — word processing, spreadsheets, local databases, email, CAD/

CAE, and sophisticated engineering programs. Originating in workgroups and departments, internetwork infrastructures typically are independent of existing SNA architectures.

In the process of internetworking client/server systems, an organization can end up with a separate physical network for each protocol — Novell, DEC, Sun, Apple, and so on. Multiple parallel networks meet the specialized needs of each client/server protocol, but are less than ideal in terms of configuration, management, cost, and support. What networked organizations really need is a single, integrated architecture that unites all client/server and SNA applications — a highly available, high-performance, multiprotocol backbone that's managed as a single entity.

Few in the industry would question the tangible benefits of a universal architecture that embraces all installed end stations. The real question is: Can current network technology support SNA and client/server applications on the same physical network without reduced reliability, performance, or functionality?

A New Network Paradigm

The goal of this guide is to show conclusively that the technology exists today to integrate SNA and internetwork traffic, while exceeding traditional FEP and host networks in terms of performance, reliability, cost, and ease of management. This emerging architecture is here, it works, and it is...

...The Integrated SNA Internetwork

The term internetwork denotes multiprotocol capabilities. So an Integrated SNA Internetwork (ISI) integrates SNA and multiprotocol traffic into a single physical infrastructure. By any name, this new, all-encompassing network is key to achieving enterprise-wide cooperative processing for all types and sizes of computing platforms. To accomplish its goal, the ISI must support the native protocols of all networks:

- SNA/SDLC
- NetBIOS
- Novell NetWare
- Token ring
- TCP/IP
- Ethernet

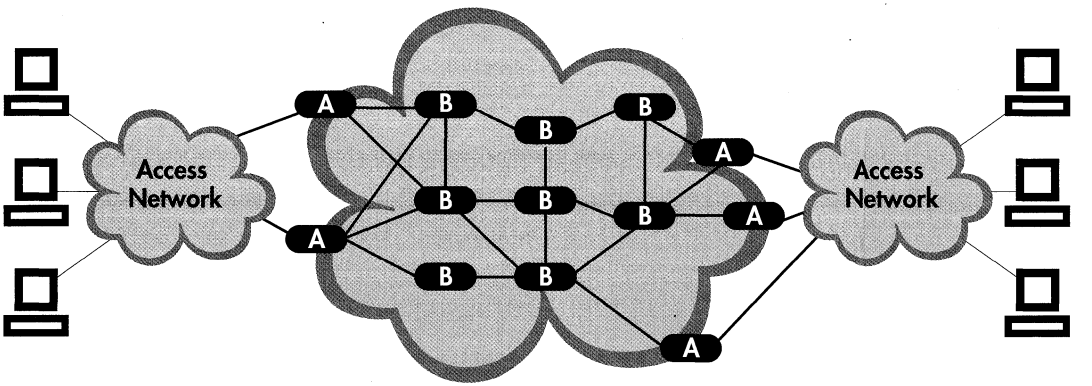
- DECnet Phase IV
- AppleTalk
- Banyan VINES
- OSI
- X.25
- XNS
- Others

The most suitable network-hardware platform for building an ISI is the multiprotocol router. High-end multiprotocol routers currently can support all major LAN-based protocols, as well as SNA and APPN. Routing data through a large ISI is the most demanding of networking tasks — not all routers can rise to the challenge. Protocol support is an evolving process; ISI-capable routers will need incremental enhancements over the next few years before the highest level of SNA-multiprotocol integration is realized.

BACKBONE AND ACCESS NETWORKS

Conceptualize an enterprise ISI network by focusing on two major spheres of functionality — the backbone network and the access network (see Figure 1A). An access network is the realm of end-user devices — PCs,

FIGURE 1A. INTEGRATED SNA INTERNETWORK

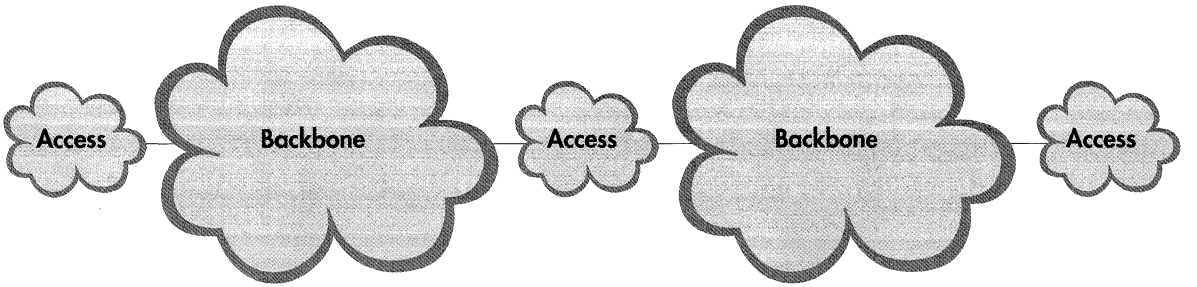


workstations, terminals, cluster controllers, file servers — LANs that connect workgroups within a building. A backbone network is the wide-area infrastructure that connects access networks across a campus, a region, a country, or the globe.

Backbone and access networks make different demands on routers. Access routers (sometimes called "edge" or "boundary" routers) interface with local SNA and internetwork devices via a LAN, SDLC, or other local link. Access routers use protocol spoofing, encapsulation, and conversion software that compensates for the wide-area deficiencies of native end-station protocols (e.g., NetBIOS and NetWare IPX). Once an access router concentrates traffic from local devices, a backbone routers delivers this traffic via Metropolitan Area Network (MAN) and WAN links to all points of the enterprise. Working together, the ISI routers provide:

- High network availability and reliability
- High performance levels
- The full spectrum of wide-area technologies
- Scalability from small workgroups to global networks
- Ease of network configuration and management
- Migration to future architectures
- Cost-effective hardware and software

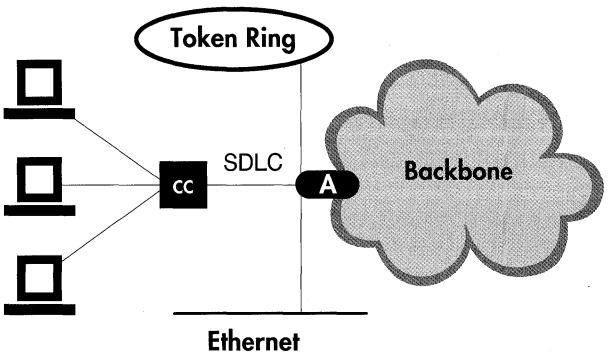
FIGURE 1B. MIX OF ACCESS & BACKBONE NETWORKS



ORGANIZATION OF THE ISI GUIDE

Throughout this Guide, ISI principles are demonstrated with a simplified topology of two access routers separated by a backbone "cloud." This configuration is ideal for examining the details of end station and router interactions. But larger enterprises have a more complicated mix of backbone and access networks (see Figure 1B). Also, keep in mind that multiprotocol routers can have more than 50 ports (see Figure 1C). Consequently, the number of LAN and SDLC segments that connect to an access router can be considerably more than those shown in the Guide's tutorial diagrams.

FIGURE 1C. MULTIPORT ACCESS ROUTER



The ISI Guide starts by attacking network issues at the link level (SDLC, token ring, source route bridging, LLC), and later moves to issues at the routing and transport levels (transmission priorities, class-of-service, LAN/WAN congestion control, adaptive route calculation). The Guide examines each issue from two perspectives:

- SNA Realities
- ISI Solutions

For each major topic, an SNA Realities section first explains the workings of current technology and identifies its pitfalls. Following each Realities section, ISI Solutions detail how to overcome common SNA problems with a robust integrated internetwork architecture.

For example, SNA Realities: SDLC explains how SDLC works in SNA environments and some of its difficulties (e.g., polling overhead). ISI Solutions: SDLC then reveals techniques that successfully integrate SDLC links into an internetwork (e.g., local polling). The Realities section contains a tutorial on SNA/SDLC basics. This primer can be passed over by readers who are familiar with the subject.

In addition to data-link and routing issues, the ISI Guide presents SNA Realities and ISI Solutions for other major topics that affect network design — SNA gateways, NetBIOS, APPN, network management, and more. It's a lot of reading, but if you invest the time, the Guide will serve as a comprehensive reference as your SNA internetworking strategy unfolds. And, yes, fellow network adventurers, it really is possible for things to be better than they were in the good old days!

TERMINOLOGY

Widely differing terminologies describe the diverse computing environments integrated by an ISI. This is particularly true between SNA and OSI-style internetworks. For clarity, this Guide uses consistent network terms. This by no means rules out the validity of alternative terms. While reading, consider the following terms equivalent:

- LAN segment, LAN, ring, token ring
- Connection, link, line
- FEP, NCP, PU4
- Cluster Controller, CC, PU2, 3174
- Host, mainframe, S/370, PU5
- Terminal, display, 3270 display
- End station, end node, computer
- Intermediate node, router, bridging router
- Packet (used in TCP/IP), message (used in SNA)
- Explorer frame, discovery frame, test frame
- Pacing, flow control, congestion control

Frame, in all cases, refers to low-level units of data at the link layer (802.5, SDLC). *Packet* and *message* are interchangeable and refer to network data generated by higher layer protocols (e.g., IP, SNA Path Control). The Guide uses the term *host* to refer exclusively to a mainframe class computer. This is not the TCP/IP definition of host, which refers to any end station. SNA experts tend to label devices by their architectural names (LU, PU). This practice has been avoided until the later stages of the guide where the terms have been defined.

Compared to SNA, internetworking terminology is reasonably straightforward. An internetwork connects any number of LAN-based *end stations* through a mesh of *intermediate nodes*. TCP/IP internetworks use this model to route traffic between PCs, workstations, and larger systems that converse as peers. Architecturally, internetwork intermediate nodes use layer-3 network protocols to route traffic across a mix of LAN or WAN links.

SNA terminology is conceptually complicated by multiple classes of end stations and intermediate nodes and by a hierarchical master/slave structure. For instance, hosts, front-end processors, and cluster controllers can all play an intermediate networking role in SNA. The basic SNA configuration links host front-end processors with other front-end processors and with downstream cluster controllers that support local end-user terminals and I/O devices.

1

SYNCHRONOUS DATA LINK CONTROL (SDLC)

SNA REALITIES: SDLC

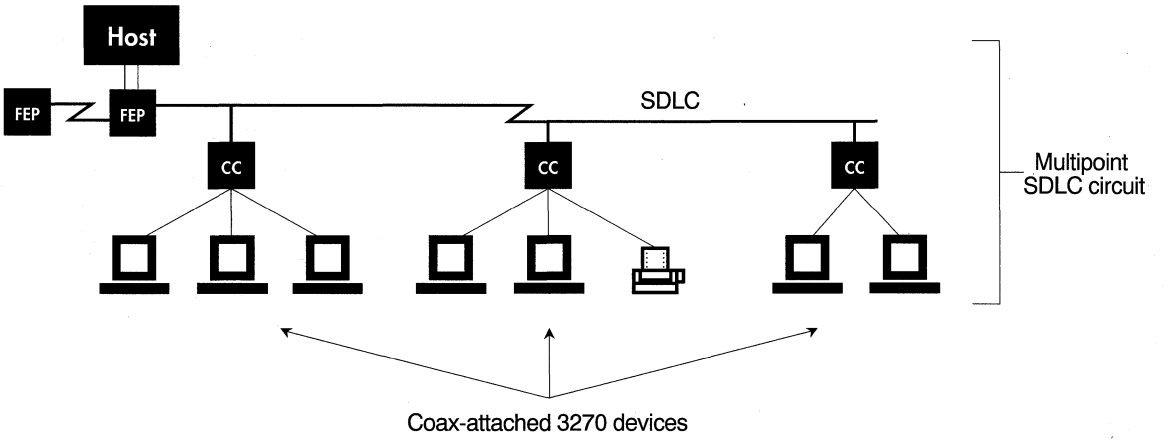
Traffic in an SNA network is generated by sessions between a host application and end-user display terminal (e.g., 3278) or I/O device (e.g., 3287 printer). The concentration point for traffic from these devices is typically a cluster controller (e.g., IBM 3174 Enterprise Controller). Cluster controllers have ports for up to 32 end-user devices, or with port expansion hardware or third-party products, more than 64 ports.

Display terminals in an SNA environment usually have a direct coaxial cable connection to the cluster controller. In the early eighties, PCs with 3270 coax cards and emulator software began to replace displays. Today it is estimated that over half of all coax-attached 3270 devices are actually PCs emulating 3270-type terminals.

The connection between the cluster controller (CC) and the front end processor (FEP) is often a Synchronous Data Link Control (SDLC) point-to-

point, SDLC multipoint (see Figure 2), or LAN link. SDLC links are switched or non-switched, and between CCs and FEPs, typically run at 9.6 or 19.2 Kbps speeds. FEP-to-FEP links run at up to T1/E1 speeds. With the advent of LANs, SNA devices are increasingly interconnected with Logical Link Control (LLC) sessions over 802.5 token rings; X.25 packet switching, ISDN, and bisynchronous protocols are also used. But for many organizations, SDLC is still the dominant SNA link protocol.

FIGURE 2. HIERARCHICAL SNA



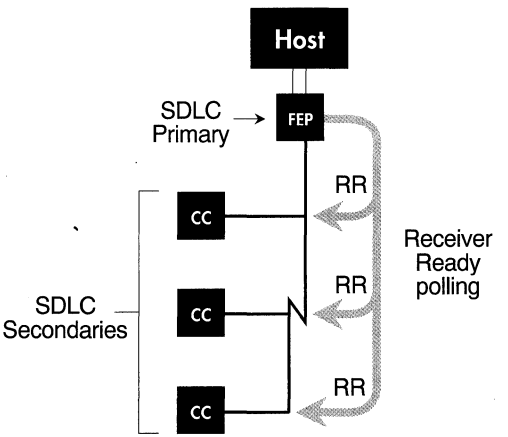
The coax (in some cases, twisted pair) connections between 3270 devices and CCs are local, and hence, not an issue in enterprise internetworking. What is an issue is the limited throughput and limited multiprotocol support of CC-to-FEP and FEP-to-FEP SDLC links. It is not uncommon for a large or mid-size corporation to maintain hundreds or thousands of SDLC links for its production SNA network. Clearly, there is a great need to integrate SDLC devices into the ISI.

SDLC (Synchronous Data Link Control) is the senior member of a family of bit-oriented data link protocols that include IEEE 802.2 LLC, HDLC LAP-B, and ISDN LAP-D. Bit-oriented links carry blocks of binary data (regardless of content), whereas character-oriented protocols (e.g., asynchronous) must convey a single character at a time. The common lineage of SDLC and other bit-oriented protocols has given them similar frame fields and command syntax (e.g., RECEIVER READY, XID, TEST).

LLC and IEEE 802.x link methods are balanced protocols that allow end stations to freely initiate a peer-to-peer conversation with one or more partners. In contrast, SDLC is an unbalanced, master/slave protocol that gives a "primary" station complete control over "secondary" stations. In SNA, the FEP is typically the primary station while downstream CCs are secondary.

Secondary stations can only send data when they are polled by their primary station. When secondary stations are not being polled, they cannot use the network (see Figure 3). The FEP uses SDLC to poll its CCs several times a second (a typical polling interval is 5 times/second). In some cases the FEP may poll a single device on a line; with multipoint (sometimes called multidrop), the FEP polls each CC in turn, round robin.

FIGURE 3. SDLC POLLING ON MULTIPOINT LINE



The poll is conveyed from primary to secondary station by a Supervisory frame containing a Receiver Ready (RR) command. SDLC uses RR frames for a number of polling and acknowledgment operations:

- Primary polls secondary
- Primary acknowledges frame(s) from secondary
- Secondary acknowledges frame(s) from primary
- Primary or secondary is ready to receive after pause

SDLC Supervisory frames also can convey a Receiver Not Ready (RNR) command. Used during congestion conditions by either primary or secondary stations, RNR indicates that no further frames should be sent. After the congestion or other delay has passed, an RR is sent to resume data flow.

When a secondary or primary SDLC station wants to send data (as opposed to control or commands), an Information frame is used. Each Information frame contains Send and Receive sequence-number fields that enable end-to-end sequencing, error checking, and flow-control. SDLC stations maintain send and receive sequence counters that increment with each transferred packet. By comparing their sequence counters to the frame sequence-number fields, stations can detect out-of-sequence packets.

A station can acknowledge received frames individually or in groups. To keep track of outstanding frames, SDLC frames have either 3-bit or 7-bit sequence-number fields. With 3-bit sequence fields, stations can send up to 7 frames before requesting an acknowledgment (and starting the count over again). With 7-bit sequence numbers, stations can send up to 127 frames before acknowledgment. Not all SDLC devices can support 7-bit sequence-number frames. FEP-to-FEP circuits typically have the larger sequence-number capabilities.

On any SDLC link, primary and secondary stations must agree (or be configured to) a maximum number of frames that can be sent before an acknowledgment. The number of allowable outstanding frames is called a window. If a station can send 4 SDLC frames before receiving an acknowledgment, it is said to have a window size of 4.

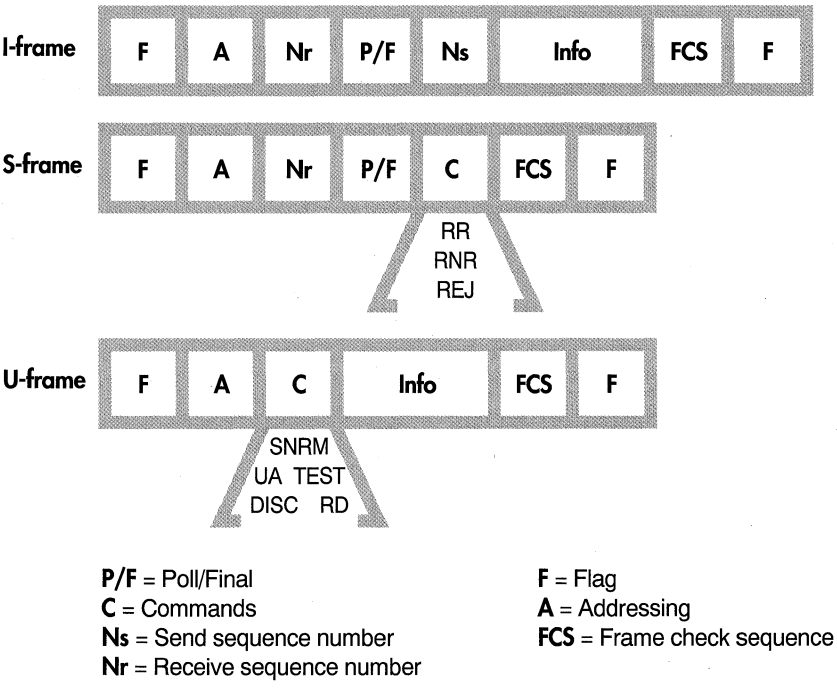
On clean lines or high-speed lines (e.g., T1) the window size should be large, allowing many frames to flow before an acknowledgment frame is required. On error-prone lines, the window should be smaller so that a sending station can quickly retransmit lost frames when it does not receive acknowledgment. (Defining a "large" or "small" window size is relative to the number of frames accepted before acknowledgment.) In addition to the RR and RNR, a third SDLC Supervisory frame, Reject, is used to reject a frame that is bad or out of sequence.

In addition to Information and Supervisory frames, SDLC uses an Unnumbered frame (U-frame, no sequence numbers) to initialize, disconnect, diagnose, and control SDLC link stations. When a primary station brings a secondary station on-line, it sends a U-frame that contains the Set Normal Response Mode (SNRM) command. When an SDLC secondary station receives an SNRM, it changes from offline to normal operating mode and is ready to send/receive data. The secondary acknowledges the SNRM command with an Unnumbered Acknowledgment (UA) response.

A secondary station goes offline when it receives a U-frame containing the Disconnect (DISC) command. Other U-frame commands include Exchange Station Identification (XID), used to locate stations and determine their characteristics; Test (TEST), used during diagnostic procedures and also to locate stations; and Request Disconnect (RD), used by a secondary link station to request offline status.

Figure 4 shows SDLC Information, Supervisory, and Unnumbered frames, some of their key fields, and their command/response contents. I- and S-frames have an important poll/final (P/F) field containing a single bit that toggles on or off to indicate whether a frame should be immediately acknowledged. The primary sets the P/F bit to indicate that it is polling the secondary (immediate response required). The secondary can use the bit to indicate that a frame is the final member in a sequence of frames requiring group acknowledgment by the primary. Figure 5 shows (chronologically from top to bottom) 3 frames sent by the primary, a secondary acknowledgment, and then three frames sent by the secondary.

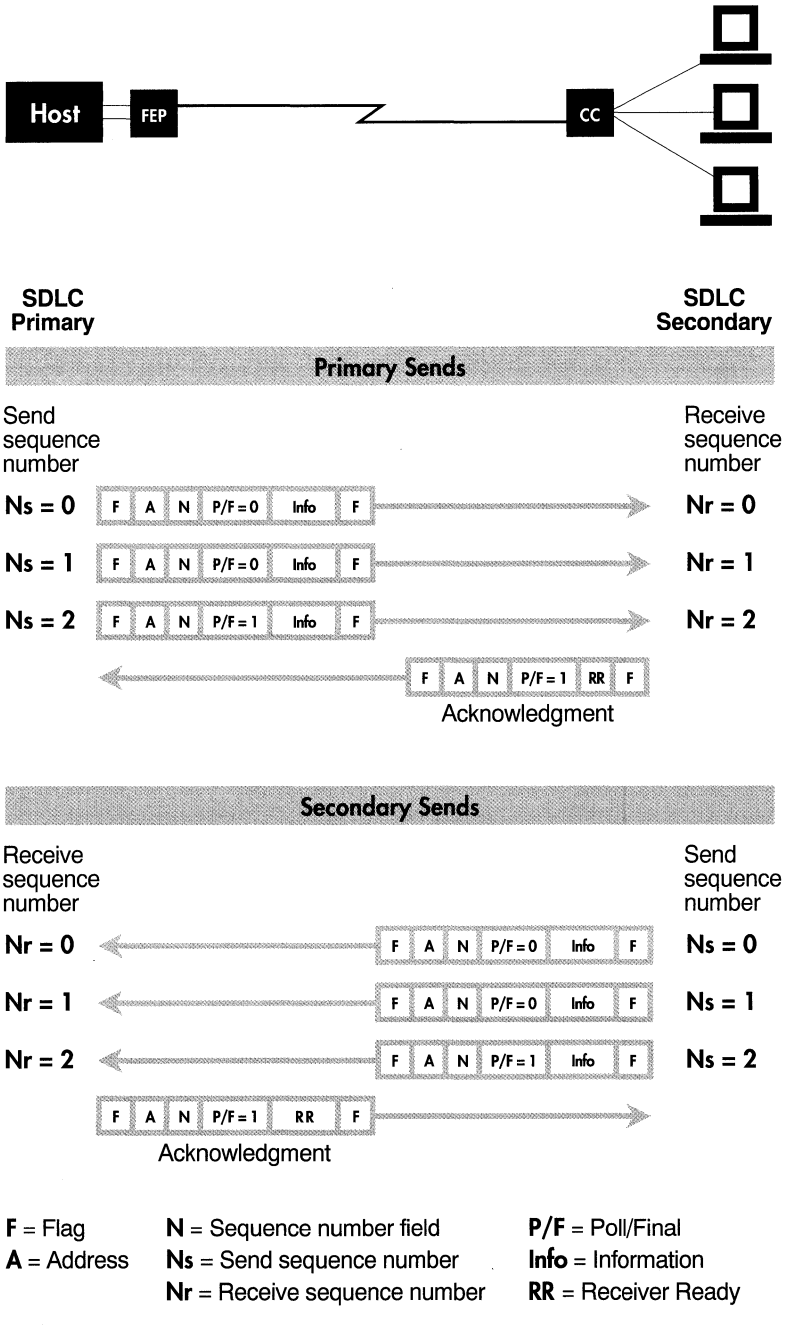
FIGURE 4. SDLC FRAME FORMATS



P/F = Poll/Final
C = Commands
Ns = Send sequence number
Nr = Receive sequence number

F = Flag
A = Addressing
FCS = Frame check sequence

FIGURE 5. SDLC DATA FLOW



F = Flag
A = Address
N = Sequence number field
Ns = Send sequence number
Nr = Receive sequence number
P/F = Poll/Final
Info = Information
RR = Receiver Ready

To detect a link or station failure, each SDLC circuit is monitored by end station timers that measure time between frames. The idle timer in a FEP can be set as high as 10 seconds or more. If the FEP does not receive a poll response within the timer's limit, it sends another poll. After a number of retries, the secondary station is assumed to be out of service. A FEP's inactivity timer and maximum retry count are set during SysGen.

On a dedicated FEP/CC link, SDLC's constant polling does not pose response-time problems. In fact, many 3270 users have grown accustomed to SDLC's predictable response times. But polling does take up a considerable amount of the bandwidth, particularly on multipoint lines. Rerouting the FEP/CC path through a multiprotocol internetwork magnifies this problem because polling overhead hampers the throughput of end user data on a shared LAN/WAN medium. Integrating SDLC into an internetwork and eliminating unnecessary SDLC frames is discussed in the next section.

ISI Solutions: SDLC

The goal of many network managers is to migrate wide-area SDLC links to an internetwork infrastructure while maintaining in-place SDLC equipment. The techniques discussed below are explained with respect to FEP/CC traffic, but the same principles apply to SDLC traffic between nearly all of the many devices that support SDLC:

- AS/400 minicomputers
- System 36/38 minicomputers
- Banking controllers (3600, 4700)
- Series/1 minicomputers
- Various SDLC printers and terminals
- Mini- and microcomputer-based SNA gateways

The ISI integrates SDLC by replacing portions of the SDLC path with an internetwork topology, while maintaining the appearance of native SDLC to FEPs, CCs, and other devices. There are a number of ways to fold SDLC traffic into the ISI, including:

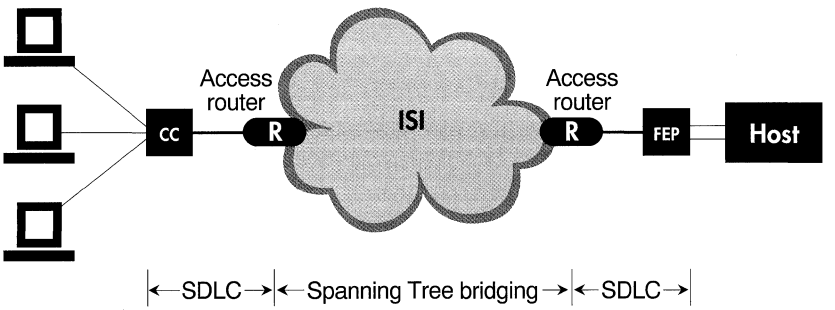
- SDLC via synchronous pass-through
- SDLC/LLC conversion via source route bridging (SRB)
- SDLC/LLC conversion via extended source route bridging
- SDLC/LLC conversion via TCP/IP

SYNCHRONOUS PASS-THROUGH

The most straightforward SDLC internetworking technique is synchronous pass-through, a form of protocol encapsulation that lets one protocol's frames pass through another's without conversion or modification. With synchronous pass-through, SDLC frames tunnel through a MAC-layer bridging topology.

To deploy synchronous pass-through, a downstream SDLC device (e.g., CC) is physically connected to a serial port on an adjacent access router. The target FEP is also physically connected, via SDLC, to an access router (see Figure 6). In the CC-to-FEP direction, SDLC frames from the downstream CC are encapsulated in a MAC-layer frame by the router and bridged across the ISI infrastructure using the Transparent Spanning Tree protocol. This provides automatic best-path selection through the network, and dynamic

FIGURE 6. SYNCHRONOUS PASS-THROUGH



recovery in the event of intermediate link failures. Once traffic makes it through the ISI, the upstream access router strips off the MAC header/trailer and passes the unmodified SDLC frame to the FEP via a serial port. From FEP to CC, the process happens in reverse.

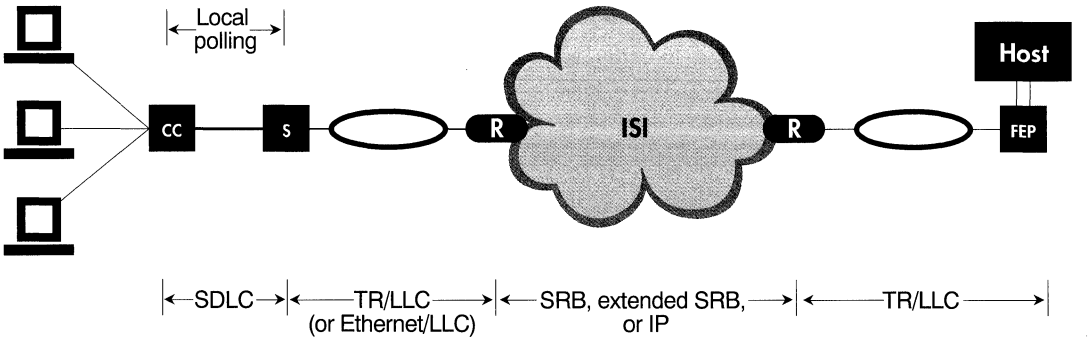
Because it uses MAC-layer bridging, synchronous pass-through has little encapsulation overhead and therefore makes efficient use of WAN links. As with any layer-2 bridging technique, synchronous pass-through is well-suited to simple end-to-end paths that have a limited number of hops. Since many hops always increases delay, limiting the number of hops protects end-user response times. But if the internetwork portion of an SDLC path must have many hops, an IP encapsulation method may be preferable to synchronous pass-through.

The pass-through method has little impact on existing SDLC equipment. If access routers have robust configuration options, FEP hardware and software does not require regeneration or modification. Pass-through is highly compatible with NetView network management because frames flow between end devices without change to SNA headers, trailers, or contents. Pass-through and other encapsulation techniques are bandwidth-efficient and largely transparent to SNA applications, management, and configuration. They do, however, place a heavy load on the internetwork due to the constant stream of SDLC polling and acknowledgment frames.

SDLC/LLC Conversion

In cases where upstream FEPs have token ring (or ethernet) connections, an increasingly popular approach is SDLC/LLC conversion. Like synchronous pass-through, this method preserves the downstream SDLC interface on the CC. To the FEP, SDLC/LLC conversion makes SDLC devices look as if they are natively attached to a LAN. In one conversion approach, a standalone SDLC/LLC "server" connects to the CC and provides SDLC-to-LLC conversion (see Figure 7). The SDLC server uses token ring/LLC or ethernet to connect with the access router.

FIGURE 7. SDLC/LLC CONVERSION



S = SDLC/LLC server

In other cases, the SDLC server function can be integrated into the router. However, the additional protocol processing overhead may introduce a performance liability. A stand-alone SDLC server can support over a dozen SDLC ports, as well as such features as:

- Full-duplex and half-duplex operation
- Leased and dial-in connections
- Point-to-point and multipoint connections
- External and internal clocking
- NRZ/NRZI encoding
- RS232, V.35, X.21, or RS449 interfaces
- Speeds of 1200 to 64 Kbps
- NetView management

The multipoint feature allows a number of CCs to share a single SDLC port on the SDLC server. SDLC devices encode frames onto the line using a Non-Return to Zero (NRZ) or NRZ-Inverted bit pattern. NRZ and NRZI bit transitions don't return to a zero voltage level during one bits, only during

zero bits. The differences between NRZ and NRZI are minor but IBM devices tend to default to NRZI. SDLC servers should support both. Another variable is the maximum frame size of SDLC devices. CCs often use 265 byte frames but some devices use 1 Kbyte or larger, so conversion servers must accommodate a variety of frame sizes.

Unlike the encapsulation of synchronous pass-through, LLC conversion actually strips off the SDLC headers and trailers and replaces them with 802.2 LLC frame formatting. After the SDLC is converted, the LLC frames are bridged or routed through the LAN/WAN internetwork to a token ring-attached FEP or other target device. Synchronous pass-through lets polls travel end-to-end, but conversion servers "remotely" poll CCs with RRs that only move locally between the server and its CCs, as shown in Figure 7.

CONFIGURING FOR CONVERSION

When an SDLC circuit is migrated to LLC conversion, the FEP is configured to support downstream CCs as token ring devices. This is accomplished by the routine process of defining controllers as VTAM-switched major nodes (VTAM's standard definition for downstream LAN and dial-up devices). Because CCs aren't actually on a ring and don't have a MAC address, a MAC address is assigned to the SDLC server.

The SDLC server converts MAC-addressed frames from the FEP into SDLC addresses for CCs, and vice versa. To do this, the server reads unique LLC Service Access Point (SAP) numbers in frames coming from the FEP to differentiate between the CCs it supports. Each CC has its own SAP address. (More on LLC SAPs in the LLC section, following.)

On the downstream side, CCs that are attached to an SDLC server generally don't have to be reconfigured. Considering that older CCs (e.g., 3274) have arduous configuration procedures, the server's ability to accommodate CCs of all types and ages is important. Normally, CCs on end-to-end SDLC lines often run at 9.6 Kbps. With the SDLC server, CCs may have their speeds increased to 19.2 or 64 Kbps in some cases.

SDLC servers support multipoint CC configurations, but when CCs are given their own port on the server, additional performance increases are realized. Because devices don't have to wait in queue to communicate, CCs can send and receive data from the server whenever necessary. Because the server uses balanced-mode LLC on its upstream link, it can send data to the FEP without waiting to be polled. With high-end SDLC servers, performance can be tuned both on the SDLC and the LLC portion of the end-to-end path, without disturbing the default settings of FEPs and controllers.

CONVERSION BENEFITS

The conversion process is non-intrusive and simple, and it reduces the traditional requirement for parallel SDLC lines to remote CCs. Remote FEPs can be eliminated because traffic is concentrated by the SDLC server at the downstream CCs. Additionally, SDLC/LLC conversion reduces the number of required ports on the upstream FEP.

When migrated to token ring, a FEP can receive traffic from multiple downstream servers via a single token-ring port. This multiplexing effect greatly reduces the number FEP ports required, and in the process, reduces FEP software fees, which are generally keyed to the number of ports. A token ring/LLC-capable FEP also improves recovery from link failures. When a 3745 FEP detects a failed LLC session it can send SRB explorer packets through the internetwork to identify an alternate route. In some cases, this rerouting process can be invisible to end-node applications. In contrast, if an SDLC link fails, applications are interrupted and must explicitly reestablish new sessions.

With SDLC/LLC conversion, the end-to-end SDLC circuit no longer exists, but CCs, FEPs, and links are still visible to NetView. The SDLC server itself can have NetView visibility by deploying its own SNA PU2 management component (in addition to SNMP). The SDLC server can also provide a PU2.1 capability for peer-to-peer traffic originating on downstream SDLC links. (More on network management, PU2, and PU2.1 in later sections.)

Once SDLC is converted to a LAN-based LLC format, it can be transported through the internetwork with standard or extended source route bridging, or as TCP/IP packets. The network issues for SDLC/LLC traffic over token ring are the same as they are for LLC traffic generated by NetBIOS and other LLC LAN nodes. Options for transporting LLC through an internetwork will be addressed in the coming sections on token ring, SRB, and LLC.

2

TOKEN RING AND SOURCE ROUTE BRIDGING (SRB)

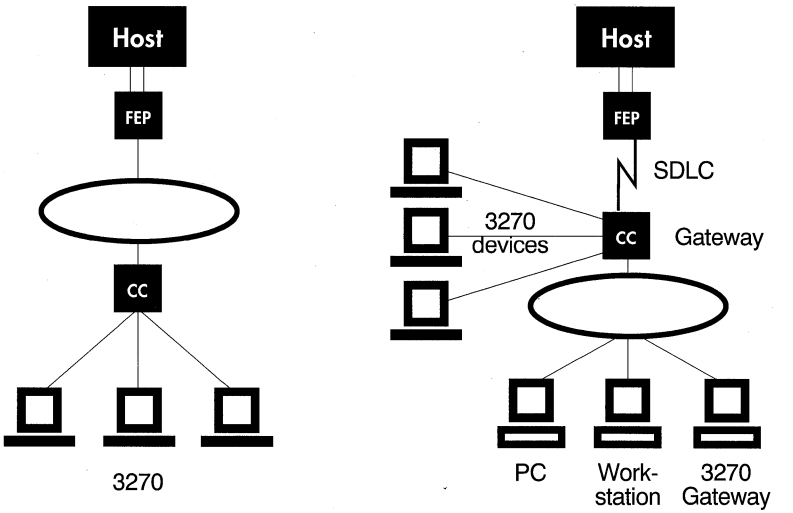
SNA REALITIES: TOKEN RING AND SRB

The advent of PCs, LANs, and client/server computing has introduced a multitude of protocols that do not run across traditional SDLC links. Fortunately, SNA devices became part of the LAN revolution when IBM gave its platforms token ring interfaces. Token ring links can simultaneously accommodate traffic from SNA and a variety of LAN-based end stations (e.g., NetWare, NetBIOS, Apple, Banyan, etc.). In addition to its multiprotocol aspect, token ring gives FEPs, controllers, and other traditional SNA nodes a fast, shared access communications method that transfers data at 4 or 16 Mbps.

In addition to X.25, synchronous and bisynchronous, FEPs can be outfitted with up to 8 token ring interfaces (up to 17, with special expansion hardware). These interfaces are sometimes called TICs in the IBM environment, which is short for Token-Ring Interface Coupler. CCs are typically limited to one token ring interface, and don't have to give up their SDLC or X.25 interfaces in the process. IBM minicomputers, RS/6000, and PS/2 all support one or more token ring connections. IBM hosts can channel-attach to a FEP (e.g., 3745) or cluster controller (e.g., 3174) for LAN access. The 3172 Interconnect Controller also gives host access to LANs. Some of the newer host platforms (9370, ES/9000) have direct token ring attachments.

In the most simple local FEP-to-CC configuration, devices are connected by a single ring as seen in Figure 8. In this configuration, the 3174-style CC forwards data from 3270 devices to the FEP via token ring. In the other configuration in Figure 8, the CC is receiving traffic from downstream token ring PCs and workstations (with 3270 emulation software) and forwarding this traffic to the FEP via SDLC. A PC running 3270 gateway software can also provide this remote token ring function. Token ring is often used for direct host connections and is frequently extended with local and remote source route bridges.

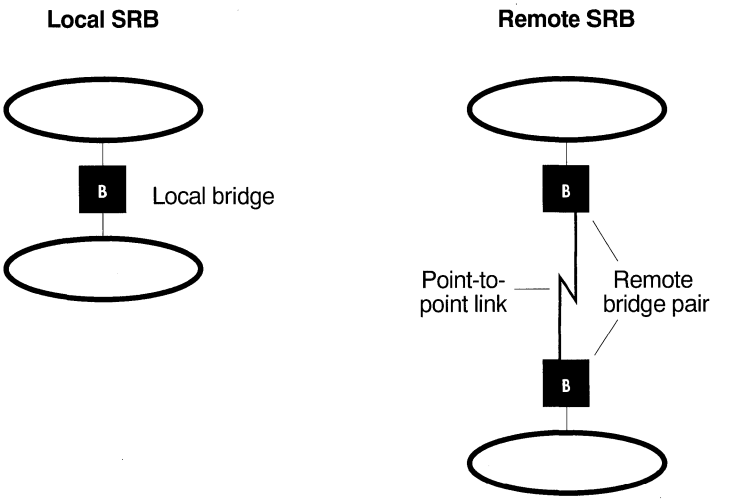
FIGURE 8. SNA TOKEN RING



SOURCE ROUTE BRIDGING PRIMER

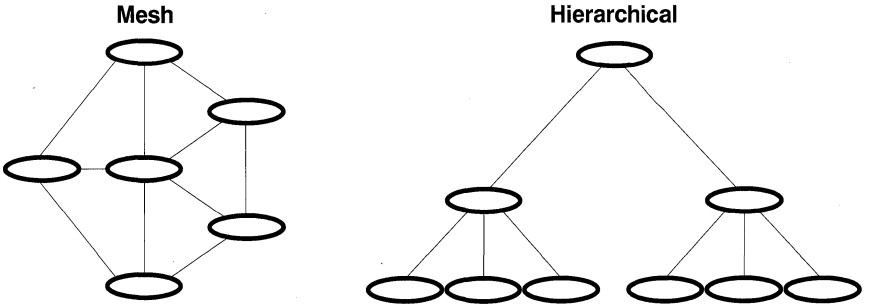
The predominant method of moving traffic between local and/or remote token rings is source route bridging (SRB). SRB is used extensively for FEP-to-CC and FEP-to-FEP, and also between NetBIOS clients and servers. In the original IBM SRB product, two local rings are bridged by a PC installed with two token ring interfaces and IBM's Token Ring Network Bridge Program (see Figure 9).

FIGURE 9. SOURCE ROUTE BRIDGING



The same software is used when two remote PCs are outfitted with serial ports and connected with a point-to-point WAN link. Router vendors have gone far beyond the original IBM two-port bridges to supply routers with dozens of SRB-capable ports. Many different token ring configurations are possible with SRB, including mesh, star, ring, and hierarchical architectures (see Figure 10).

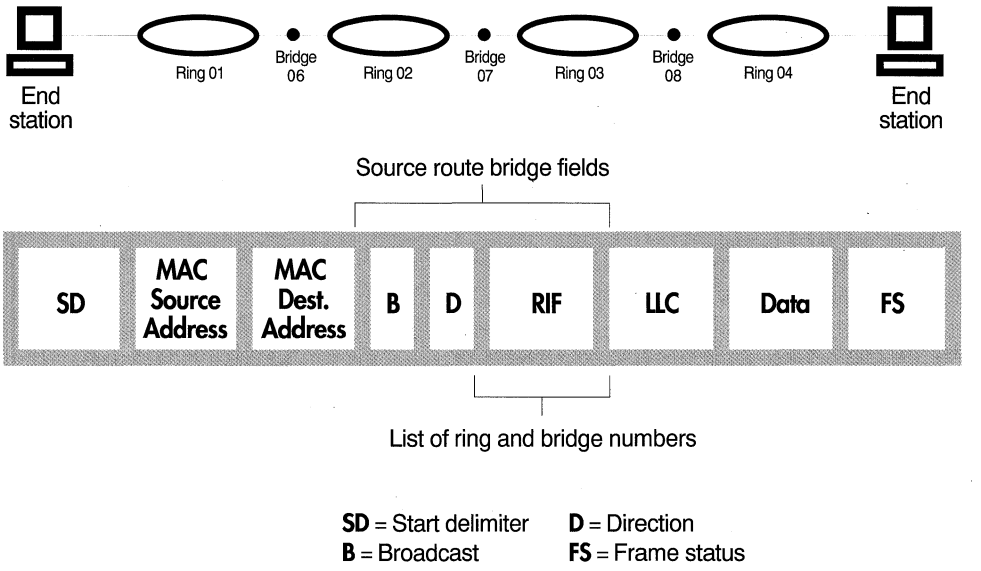
FIGURE 10. SOURCE ROUTE BRIDGING TOPOLOGIES



SRB OPERATION

Token ring 802.5 frames contain fields for source and destination address, start and end delimiters (flags), and frame check sequences. To support SRB, token ring frames also have several optional source route bridging fields. The most important of these is a Routing Information Field (RIF) that tells source route bridges which path a frame should take through the network. The token ring topology assigns numbers to each ring and bridge. A SRB RIF field is simply a list of consecutive bridge and ring numbers that define a path through the topology (see Figure 11).

FIGURE 11. SOURCE ROUTE BRIDGING



When a token ring station prepares to send a frame, it knows the destination MAC address of the target station but not the path to it. To locate its partner, the source station sends out an explorer frame (addressed to the target MAC address) that propagates throughout the network to every local and remote ring. Bits in the source route bridging broadcast field tell bridges that the explorer frame should be copied to all rings. As an explorer traverses the network, bridges add their ring/bridge numbers to the RIF field.

The target station eventually receives an explorer frame and sends back a response using the new RIF. When this response frame gets back to the source station, it records the RIF information and uses it for subsequent frames in the session, as does its partner. Bridges know which direction a frame is headed by the SRB direction bit. When multiple explorer-response frames are received by the source station, the first one back is used for RIF information, ensuring the fastest path. The size of the source routing RIF field limits to 7 the number of bridge hops a frame can take.

Source route bridging automatically determines a route through an arbitrary topology of rings, adapting to failed links and changes in the topology. SRB tolerates parallel paths and (in most cases) paths with loops. In contrast, SNA topologies must have all routes defined manually in the arduous SysGen process. When SNA links or nodes change, path tables must be manually updated.

Because of its simplicity and ease of use, token ring source route bridging is widely deployed on SNA and NetBIOS end stations. But in spite of its advantages, SRB has only layer-2 bridging intelligence and relies heavily on broadcasts, not routing tables, to choose its routes. Route discovery broadcasts can multiply as they pass through a mesh topology and as the number of end stations increases. This effect can flood wide-area communication links and ultimately result in broadcast storms in cases of major link or node failures.

SRB Broadcast Overload

Because duplication and propagation of broadcasts in large SRB environments is a significant problem, the IEEE has added a single-route broadcast feature to the SRB standard. This ensures that broadcast frames traverse a given ring one time only. To establish a broadcast route for explorer frames, a spanning tree topology is set up between SRB bridges. The spanning tree algorithm relationship between SRB bridges guarantees that each ring has a unique broadcast route to every other ring. To use a single route during route discovery, an end station sets the SRB broadcast bits to indicate single-route broadcast.

The spanning tree broadcast topology can be established by manual definition or automatically. The spanning tree used by SRB is only used for broadcast packets. Once a route is discovered, subsequent packets are routed end-to-end with the RIF information supplied by the source end station. Hence "source" route bridging.

When SNA messages are carried via SRB, the SNA network and transport layer capabilities (pacing, routing, load balancing) are not used. SNA routing nodes see the SRB bridge topology as a single link, regardless of hopcount. So in the process of migrating from SDLC to token-ring/SRB, SNA devices have gained speed and multiprotocol links but have lost reliable layer-3 routing capabilities. (The ISI restores this capability, as discussed in later sections.)

Bridges and bridging routers can support the Source Route Transparent (SRT) protocol, which allows them to bridge token ring traffic with Transparent Spanning Tree when the frames don't have SRB fields (e.g., no RIF). SRT bridging is desirable when there is a mix of SRB and non-SRB end stations sharing token rings or ethernets. For small- or medium-sized topologies, SRB or SRT may be adequate. But even with single route broadcast, SRB can be problematic in large or complex internetworks.

ISI Solutions: Token Ring and SRB

The trend in recent years has been to extend token ring topologies with local and remote source route bridging. Connections between remote bridges are typically low- or medium-speed links, 19.2 or 56 Kbps, for instance, or in some cases T1. Token ring networks bridged with SRB pass SNA, NetBIOS, Novell, and other client/server traffic, but as they grow, the negative effects of SRB become apparent. An ISI addresses the two major deficiencies of token ring/SRB protocols:

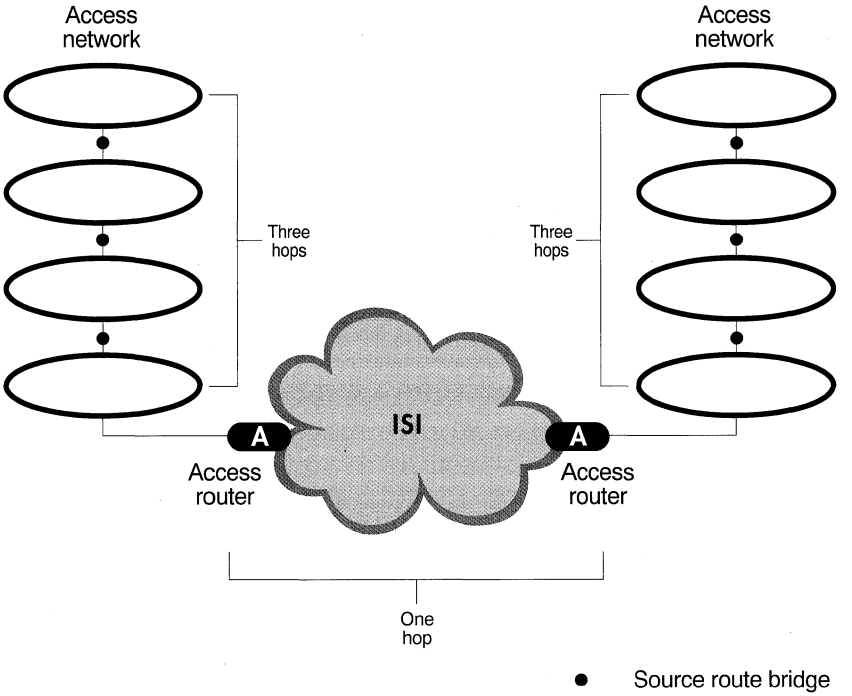
- Limited hop counts
- Excessive broadcasts

Extended SRB

To overcome the 7-hop limit imposed by SRB, ISI routers manipulate the RIF field to allow more hops. This enhanced or "extended" source route bridging is typically accomplished by making the ISI backbone look like a

single ring to nodes on SRB access topologies (see Figure 12). With extended SRB, a frame can take up to 3 hops on one side of the ISI and 3 hops on the other side (or any two-way split of 6 hops). Although the ISI itself may consist of many intermediate nodes and links, these appear as a single virtual hop to the SRB protocol.

FIGURE 12. EXTENDED SOURCE ROUTE BRIDGING



SNA/IP ENCAPSULATION

The ISI supports native SRB and extended SRB throughout the enterprise. But in medium and large topologies where there's a need to control the effects of LAN protocols, ISI access routers encapsulate SRB traffic in IP packets for the backbone portion of the trip. This technique, called SNA/IP encapsulation, facilitates SNA/LLC, SNA/SDLC, NetBIOS, and other SRB-based protocols. SNA/IP is key to broadcast reduction, hop count extension, LAN/WAN congestion control, and rapid link-failure recovery, as well as the elimination of excessive time-outs and acknowledgments in end station software.

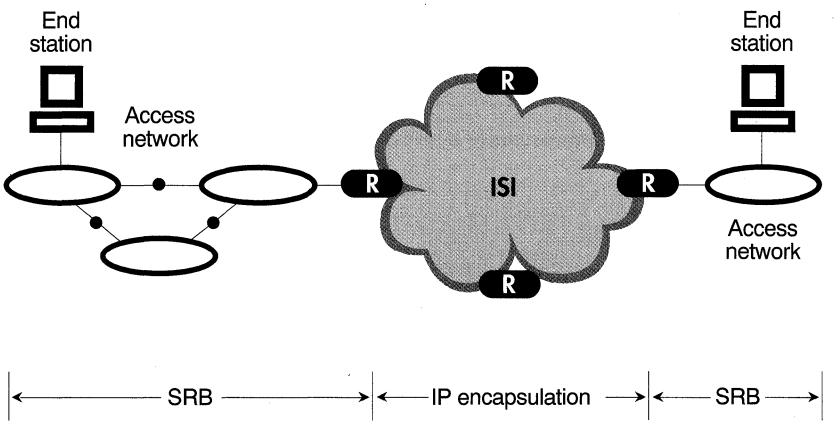
SNA/IP encapsulation allows traffic to be routed across the ISI backbone with layer-3 and -4 intelligence, confining source route bridging to access networks. To native SRB end stations and bridges, the whole enterprise looks like an SRB topology, even though SRB is not used in the ISI backbone portion of the network. To provide reliable transport through the ISI, access routers can run TCP software on top of IP. For less overhead, UDP/IP is deployed.

Because the ISI is based on a robust IP and OSPF backbone, it can provide an efficient, reliable network infrastructure for SNA traffic. When encapsulated, SNA and NetBIOS leverage the advantages of high-performance multiprotocol routing — automatic reroute around failed links, fault tolerance, load balancing, congestion control, cost-effective hardware platforms, and complex mesh configurations. After minimal configuration, SNA/IP is self-learning and almost completely transparent to the management of SNA/SRB topologies.

SNA/IP Operation

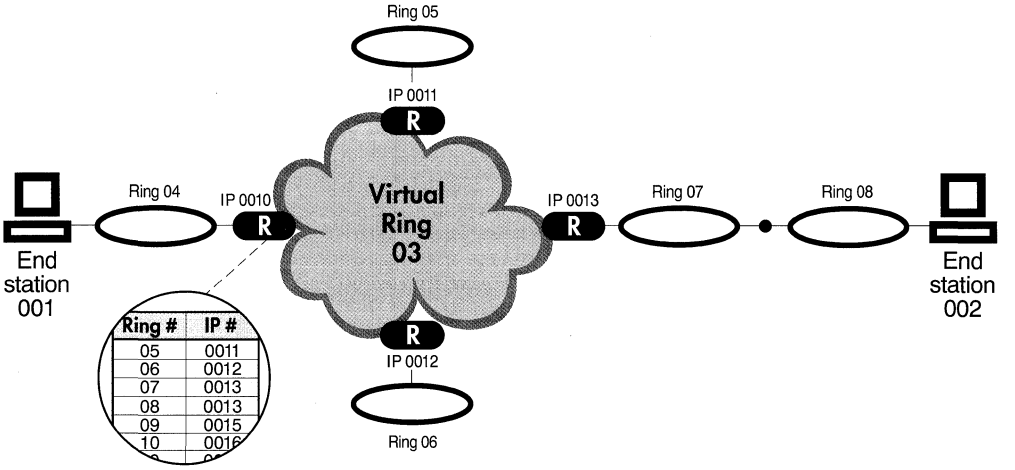
As shown in Figure 13, SNA/IP encapsulation begins when token ring frames leave an end station and are bridged through the local SRB access network. When the frames reach the first access router, they are encapsulated in IP packets and forwarded through the ISI using layer-3 routing. After traversing the ISI, the frames are de-encapsulated and output onto the target access network in native SRB format.

FIGURE 13. SNA/IP ENCAPSULATION



To locate each other, access routers keep tables that list the IP addresses of other access routers connected to SRB networks (see Figure 14). Access routers automatically build these tables by monitoring the contents of SRB RIF fields as explorer frames pass through. There are two sets of addresses in each table. The first value is the ring number of the target token ring. The second value is the IP address of the access router associated with the target ring. Each table entry maps a ring number to an access router's IP address.

FIGURE 14. SNA/IP ADDRESSING



In Figure 14, the RIF describing the path from end station 001 to end station 002 would include the ring numbers 04-03-07-08. The ISI virtual ring number (03) acts like any other ring number in the RIF. By consulting its table, router 0010 knows it can get to ring 07 by sending packets to IP 0013. When the packets reach 0013 they are output onto ring 07 where the SRB access network uses the remainder of the RIF to forward the frame to ring 08 and end station 002.

If all token ring frames (including broadcasts) are encapsulated and tunneled through the ISI, then a large amount of extraneous traffic is propagated throughout the enterprise. The ISI goes beyond simple tunneling by providing three services that restrict excessive broadcasts:

- Selected broadcast networks
- Directed explorers
- MAC address caching

The first two methods are enabled by SNA/IP, the third requires Data Link Switching (DLS), an alternative IP encapsulation architecture described below.

In the selected broadcast network approach, access routers are configured to forward SRB explorer frames only to certain access networks. The target networks are administratively defined by a simple process of numbering router ports that output broadcasts. When an end station sends a broadcast, it is only forwarded to defined networks. Broadcast frames are routed through the ISI with IP so there is no chance that extraneous or redundant explorers loop around the topology.

The selected broadcast network method is particularly useful for managing the route discovery process in networks where traffic patterns are well defined. In this case, certain access networks may be configured to receive no broadcasts, while others (e.g., networks with FEPs and servers) are configured to receive broadcasts from many clients and gateways.

The second method, directed explorers, is a targeted approach that focuses broadcasts on specific often-accessed nodes. Routers are administratively configured with the MAC addresses of heavily used nodes and the IP addresses of the closest access routers. When a broadcast is generated, the access router looks up the frame's destination MAC address in its table and if found, the router forwards the explorer (via IP) to the destination node's access network. The returning explorer response is also encapsulated in IP (as is all subsequent data traffic). Directed explorers works particularly well in a situation where CCs, gateways, or PCs regularly access centralized FEPs or servers.

MAC Caching and Data Link Switching

The third method of broadcast control, MAC address caching, automatically caches MAC-to-IP address mappings for all active token ring devices in the enterprise. After initial configuration, ongoing manual definitions are not required. Automatic MAC caching originated in IBM's Data Link Switching (DLS), a published SNA encapsulation specification. Like SNA/IP, DLS encapsulates SNA, LAN, and SDLC traffic for transport across an IP infrastructure. When building MAC/IP address tables in access routers, DLS uses a special CAN YOU REACH/I CAN REACH protocol.

As shown in Figure 15, access router 0010 receives an explorer packet from end station 001. Router 0010 then sends a CAN YOU REACH request via IP to access routers 0011, 0012, and 0013. These routers in turn broadcast (to their attached rings) a local explorer frame containing the destination MAC address. In Figure 16, router 0013 locates end station 002 when it receives an explorer response frame from 002. At this point, router 0013 sends an I CAN REACH message via IP back to router 0010.

FIGURE 15. LOCALIZED EXPLORERS

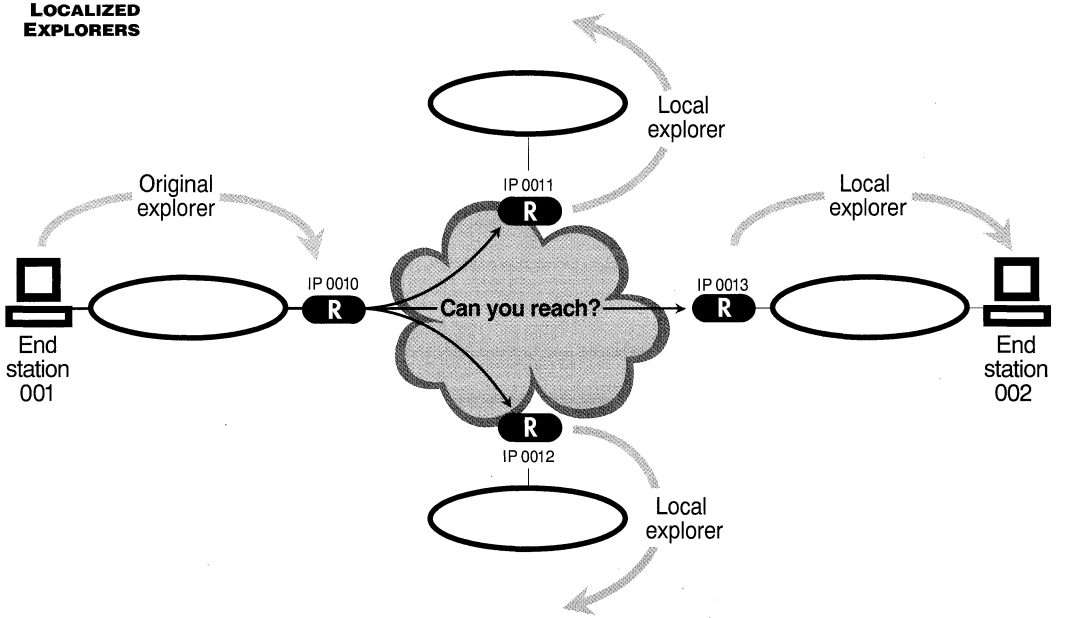
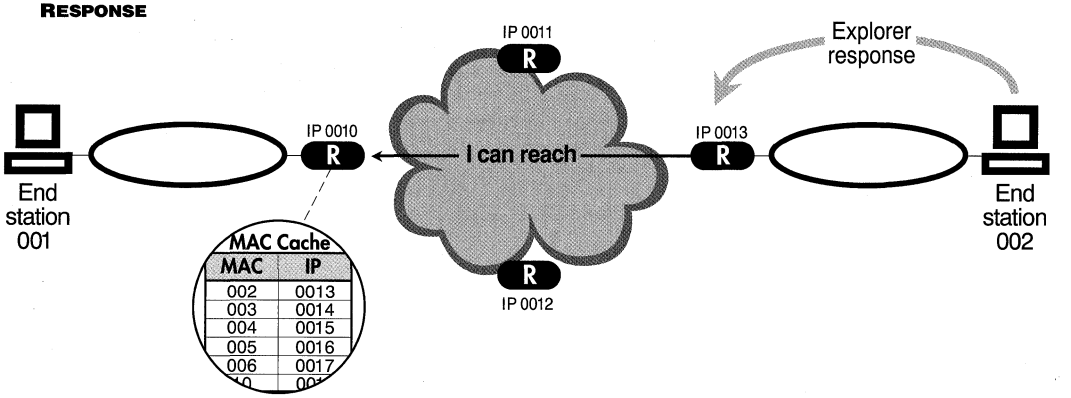


FIGURE 16. EXPLORER RESPONSE



Router 0010 receives the I CAN REACH message and automatically builds a table entry that maps the destination MAC address (002) to the IP address of its access router (0013). Router 0013 can also build a table entry because it now knows that router 0010 is the path to end station 001. To complete the process, router 0010 sends an explorer response frame back to end station 001 containing a RIF field that describes the local path between end station 001 and router 0010. From this point on, router 0010 knows to forward all frames for end station 002 to router 0013.

Every time a new end station sends an explorer, DLS creates a new MAC/IP address entry for access routers in the path. Even after a MAC address has been cached, DLS sends a directed CAN YOU REACH message to verify that the destination end station exists.

CACHING TRADEOFFS

Once DLS has built up MAC/IP tables for active SRB end stations, broadcast traffic is kept to a minimum. This conserves bandwidth, but there is a price to pay in memory and processing overhead on access routers. The more addresses that are cached, the more routers are taxed to maintain and sort through tables in real time. The size of the MAC tables can be limited somewhat by aging and discarding inactive MAC addresses. But the cost of DLS in router performance is still considerably more than with selected broadcast networks and directed explorers. The added value of DLS is justifiable in certain client/server environments that exhibit substantial peer-to-peer traffic patterns throughout the enterprise.

None of the above methods is ideal for all SNA or token ring environments. For each enterprise, a trade-off must be made between the amount of broadcast control and hop-count extension desired, versus how much router overhead can be tolerated.

SRB Hop Count Extension

In addition to broadcast reduction and IP encapsulation, DLS brings another advantage in the area of hop-count maximums — because DLS RIFs only represent source or destination access networks (not end-to-end routes), up to 7 hops can be supported on either side of the ISI. The DLS network appears as one virtual hop, so end-to-end paths of 7 + 1 + 7 = 15 hops can be achieved. This localized RIF technique is sometimes called resetting or terminating the RIF. SNA/IP encapsulation does not reset RIFs because it uses the same RIF information end-to-end. This yields the SRB standard 7-hop maximum with the ISI counting as one hop.

Resetting the RIF may be valuable in a topology of many bridged rings. Alternatively, connecting rings directly to a multiport router that serves as a high-performance "backbone-in-a-box" can eliminate local bridges. For performance and management reasons, dual-port bridge infrastructure with large numbers of hops should be avoided.

3

LOGICAL LINK CONTROL (LLC)

SNA REALITIES: LLC

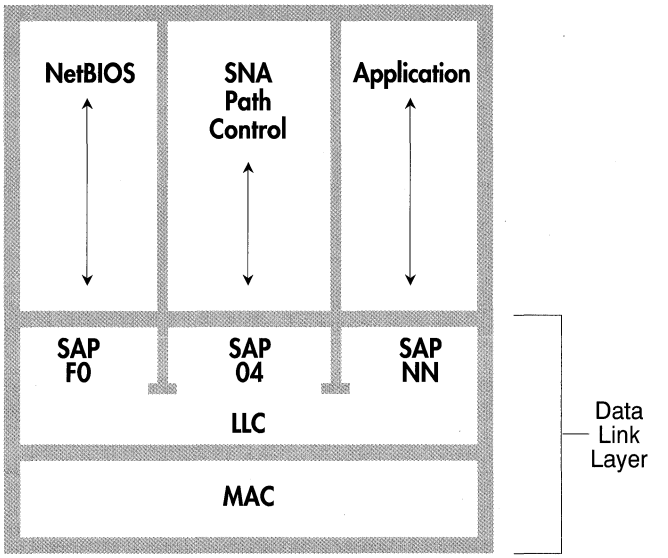
SNA network protocols are designed to run on a connection-oriented data-link layer that provides end-to-end sequencing and error control. Traditionally this service has been furnished by SDLC, but in LAN environments, SNA's link method is 802.2 Logical Link Control (LLC). As a sublayer of data link control, LLC resides above the Media Access Control (MAC) layer (see Figure 17). LLC is a significant factor in an ISI because of its wide deployment on FEPs, controllers, and NetBIOS client/server stations.

LLC can be implemented on a number of LAN types, including FDDI and ethernet. On token ring, LLC uses the SRB route discovery process, so all of the above-mentioned SRB issues apply. As organizations have added remote bridges to their networks in an attempt to scale up SRB topologies, LLC has become an increasing liability. After the following discussion of LLC operation, an ISI Solutions section will explain how LLC can be adapted to medium- and large-scale internetwork environments.

LLC PRIMER

There are several versions of LLC. Type 1 is a connectionless, datagram method; Type 2 is based on connection-oriented virtual circuits that require end stations to negotiate a logical connection before data is sent. LLC Type 2 is a point-to-point method, not a routing protocol, but it does provide sequencing of MAC layer frames, error correction, and flow-control between end stations. LLC Type 3 is an acknowledged, connectionless service that is not widely used.

FIGURE 17. LLC SERVICE ACCESS POINTS



SNA devices and NetBIOS PCs use LLC Type 2 when they establish sessions through a token ring topology. SNA and NetBIOS need the LLC2 connection-oriented circuits to provide higher-layer sequencing and error control in bridged token ring environments. In contrast, TCP/IP, NetWare, DECnet, and other internetwork protocols don't need a connection-oriented layer 2 because this is provided at the routing and transport layers. These protocols may use LLC1.

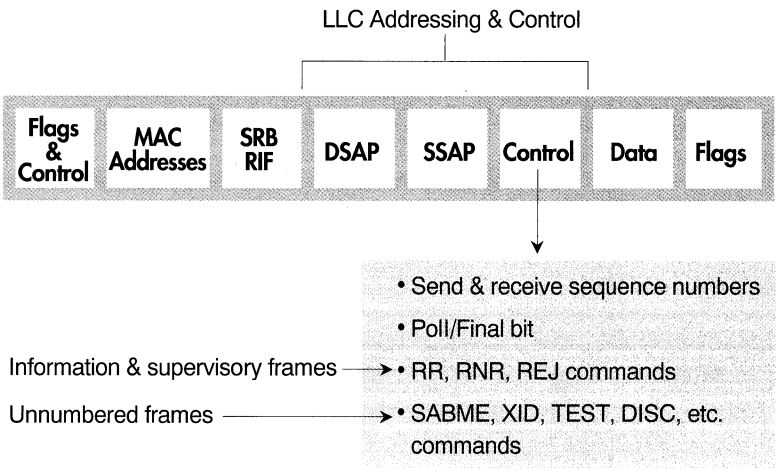
Note: For the duration of this document, the term LLC implies LLC Type 2 in SNA and NetBIOS context, unless otherwise noted.

In addition to sequencing, LLC (all types) provides a Service Access Point (SAP) addressing scheme that lets multiple applications and protocol entities in a single machine share a MAC address. Popular network protocols (NetWare, NetBIOS, SNA, etc.) all have published SAP addresses, but any

application can use a SAP to send or receive data via LLC. The LLC SAP function "de-multiplexes" frames coming up from the MAC layer and directs them to the appropriate protocol or application software. The published SAPs for NetBIOS and SNA Path Control are shown in Figure 17.

Within a MAC frame, LLC has its own fields for destination SAP address (DSAP) and source SAP address (SSAP) as well as a control field that conveys LLC sequence numbers, poll/final bit, commands, and responses (see Figure 18). LLC works much like SDLC in terms of sequencing and

FIGURE 18. LLC FRAME FIELDS

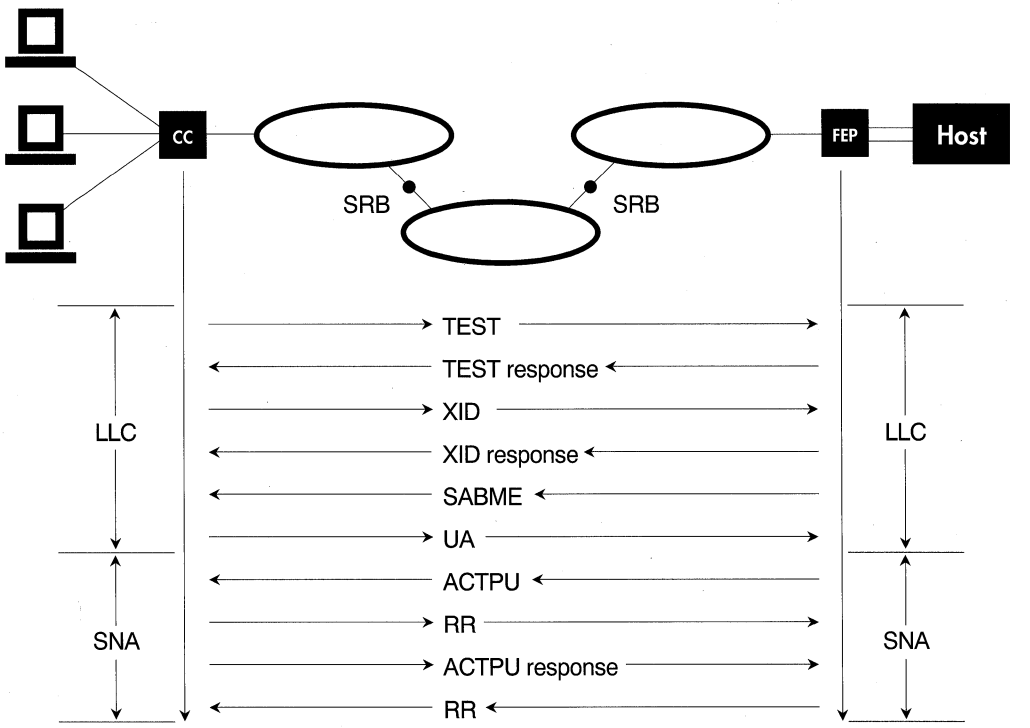


acknowledging, but does not impose unbalanced, primary/secondary relationships. Any LLC station can initiate a peer-to-peer conversation with any other station. Just like SDLC, LLC frames with the P/F bit set are acknowledged immediately. Groups of sequenced frames can be acknowledged with a single Receiver Ready response. When an LLC station gets into a busy condition, it can send its partner a Receiver Not Ready frame to temporarily suspend data flow.

LLC CIRCUIT ESTABLISHMENT

SNA and NetBIOS nodes establish LLC circuits using the SRB route discovery process. To find another station's location, LLC stations typically send out a TEST or XID command contained in an SRB frame with the broadcast bits set. Once this explorer returns, the standard SRB RIF guides frames in both directions for the duration of the LLC session. The following text gives a step-by-step description of how this takes place in a typical SRB application, as shown in Figure 19.

FIGURE 19. ESTABLISHING AN FEP/CC CONNECTION



SNA OVER LLC

When a CC (or PC gateway) sets up an LLC session with a TIC-attached FEP, it must know the FEP's MAC address, and the SAP address for the FEP's SNA routing software (Path Control). To locate the FEP, the CC first sends out a MAC frame containing the LLC TEST command on its local ring. If the CC doesn't get a response locally, it sends out a second TEST frame with the broadcast bits set. This frame propagates throughout the SRB topology and reaches the FEP, which sends back a TEST Response frame with a completed RIF field.

The CC now sends an LLC Exchange Station Identification (XID) frame with the destination SAP address set to SNA Path Control. This frame verifies that the FEP's Path Control SAP is active and it also tells the host that a peripheral node wants a connection. After an XID exchange of identities and capabilities, the FEP sends an LLC SABME (Set Asynchronous Balanced Mode Extended) command that initiates an LLC circuit. This is the equivalent of the SDLC Set Normal Response Mode (SNRM) command. The CC then acknowledges the SABME with an Unnumbered Acknowledgment (UA) frame.

After the LLC circuit is established, the FEP sends the CC a series of I-frames that contain higher level SNA commands, e.g., ACTPU and ACTLU. The ACTPU command activates the physical unit management software in a CC or gateway; ACTLU activates the logical unit session-layer software. When a session is completed, the FEP sends a DISC command and the LLC circuit is torn down.

In simple WAN or LAN configurations, this route discovery and connection process works well. But in large remote-bridged networks, broadcasted TEST frames and RR acknowledgments can overburden WAN links. This is particularly the case when a FEP's token ring interface fails, forcing large numbers of CCs or PCs to reestablish LLC connections by simultaneously broadcasting TEST frames. This can create a broadcast storm that monopolizes bandwidth and degrades response times throughout the network.

RECEIVER READY OVERHEAD

LLC does not create the constant polling traffic that typifies SDLC. But LLC is an acknowledgment-intensive protocol that generates a considerable number of Receiver Ready frames. In a typical LLC application, when a higher level protocol sends a command (e.g., ACTPU), the receiving node generates an LLC RR in response. When the receiving node responds to the higher-level protocol command (e.g., ACTPU response), another LLC RR goes back the other way (see lower flows in Figure 19 on page 3–4). Excessive RRs are generated when acknowledgments and responses take place at different levels of the protocol stack, particularly in SNA and NetBIOS implementations.

Even when there is no data flowing, LLC stations send RRs back and forth periodically to "keep alive" the idle session. LLC stations run inactivity timers (Ti timers) during idle periods. When an end station's Ti timer expires, it sends its partner station a keep-alive RR, and expects a response back. The typical default value for inactivity timers is 30 seconds.

On a LAN, RR acknowledgments and keep-alives do not threaten network performance. But in larger bridged architectures with hundreds or thousands of stations, RRs can take up valuable bandwidth, particularly on low-speed links. As with SDLC, LLC RRs must be reduced when LLC circuits are integrated into an ISI.

TIMING SENSITIVITIES

As end-to-end transit times increase, LLC displays another weakness that relates to its response-timer parameters. LLC response timers — T1 timers — are independent from the Ti inactivity timers. Each time an LLC station sends a frame with the poll bit set, it starts its T1 timer and expects an acknowledgment back within the timer's period. The recommended value for T1 timers is typically 1 second. Most network devices allow adjustment of T1 timers, but CCs have 1- to 2-second default timer values that often cannot be adjusted. FEP timers are adjustable and default to 1 or 2 seconds.

If a station does not receive an acknowledgment frame on time, it enters a recovery condition and immediately sends out an RR with the poll bit set. The station continues to send out RRs until it receives a response or until it's maximum retry count is reached. After the maximum number of retries, the session is considered lost and the connection must be reset (reestablished) with a SABME frame.

Because LLC was designed as a LAN protocol, its timers assume LAN (sub-second) response times. But LAN throughputs are not generally attainable on SRB topologies that have many hops and remote-bridge links. Each bridge that a message passes through introduces a delay. This delay is a function of the speed of a bridge, the depth of its queues, the amount of traffic it is processing, and other variables. Slow WAN links also introduce congestion that causes delay.

When bridges become congested, message delivery and response slows and large numbers of LLC response-timers can time out. The resulting RRs, retries, and resets create even more congestion. The worst case in this self-perpetuating scenario is a halt to production application processing. Timers can, in some cases, be set to a longer interval, but if they are set too long, stations take an excessive amount of time to recover from lost or bad packets, in which case, higher level protocol or application timers may expire. Therefore, T1 timers must have an interval shorter than that of higher level timers. Fine-tuning a large number of end station timers on LAN/WAN topologies is problematic.

CONGESTION CONTROL QUESTION

One additional issue for major LLC deployments is congestion control. LLC stations use RR/RNR signalling to indicate busy or congested states. On a single LAN segment, this method can control congestion adequately because only LLC end stations are involved. But when a wide-area SRB topology is inserted between LLC end stations, congestion control becomes an issue because bridges do not participate in RR/RNR signalling. Since bridges cannot throttle end stations when congestion occurs in the WAN, they drop packets.

ISI Solutions: LLC

Because LLC is used by a very large number of PCs, workstations, and SNA devices, it is critical that internetworks optimize LLC traffic. Four major LLC liabilities are addressed by the ISI architecture:

- Broadcast overhead during route discovery
- Excessive RR acknowledgments and keep-alives
- Retries and resets due to time sensitivities
- Inadequate congestion control in the WAN

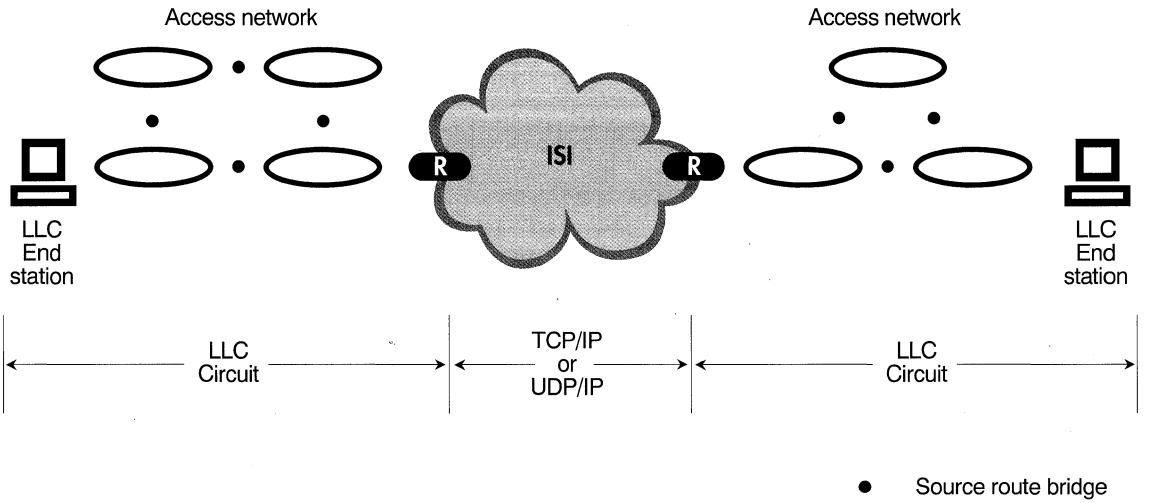
LLC Broadcast Control

The essential ingredient for successful LLC integration is SNA/IP encapsulation. The first issue, LLC broadcast overhead, stems from the SRB route discovery process. The SNA/IP broadcast strategies described earlier — selected broadcast networks, directed explorers, and DLS MAC caching — can control LLC TEST frame propagation.

LLC Circuit Termination

To address other LLC issues — RRs, timers, and congestion - SNA/IP-capable access routers divide each end-to-end LLC circuit into two localized segments that are separated by an IP delivery route (see Figure 20). As far as the LLC end stations are concerned, there still exists an end-to-end LLC Type 2 circuit.

FIGURE 20. LLC TERMINATION



TAMING LLC TIMERS

In the process of segmenting LLC circuits, access routers create the appearance of a local LLC link station that generates LLC Supervisory frames, (e.g., RR, RNR, REJ). Once data starts flowing, frames sent by LLC end stations are acknowledged, individually or in groups, by the local access router. Keep-alives are exchanged locally as well. With this LLC termination or local acknowledgment approach, excessive RR frames do not impact the ISI backbone and its WAN links.

LLC T1 response timers in FEPs, controllers, and gateways cease to be a major issue when SNA/IP encapsulation takes place. In terminating the LLC circuit on either side of the ISI, access routers also terminate the LLC timers. This essentially creates two sets of timers, one set for each local LLC circuit. When timers are terminated in this way, LLC timeouts are rare because acknowledgments take place in the time it takes a frame to cross the LAN topology.

This is an important ISI benefit because fine-tuning timers to accommodate WAN transit times is a difficult trial-and-error chore. By keeping timers local, the default T1 response-time values (e.g., 1 second) can be preserved and applications do not time out. In the case of FEP's, there is no longer a great need to modify the timer values in Network Control Program (NCP) software.

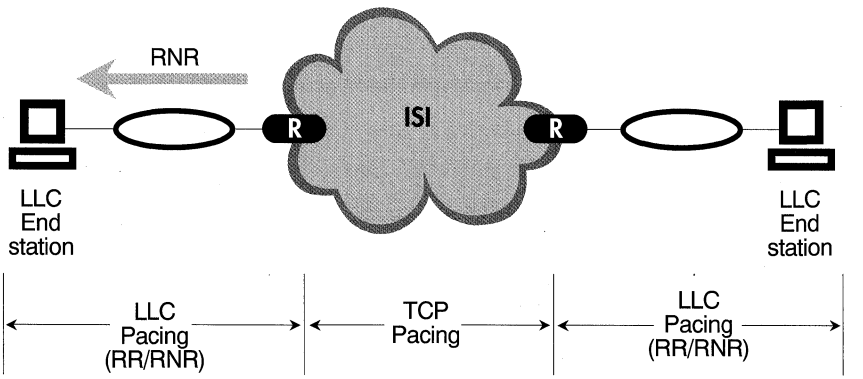
CONQUERING CONGESTION

When LLC circuits are terminated locally, an advanced two-stage congestion control technique is possible. To accomplish this, ISI access routers use TCP congestion control in the backbone and LLC congestion control in the access network. Although it is usually run on end stations, TCP protocols can be deployed in access routers to provide congestion control and sequenced, error-free transport.

TCP uses sequence numbering and sliding windows that are similar to SDLC and LLC, but more sophisticated. Using an adaptive "sliding" window algorithm, TCP-equipped routers slow down network traffic by forcing sending nodes to wait for frequent TCP acknowledgments before continuing. TCP is well suited to a complex ISI because it can reorder packets received out of sequence. SDLC and LLC, in contrast, must discard misordered packets, requiring retransmission.

To control congestion on access networks, routers conduct RR/RNR flow control with end stations via the localized LLC circuits (see Figure 21). ISI

FIGURE 21. INTEGRATED CONGESTION CONTROL



routers coordinate LLC and TCP congestion control by translating LLC pacing commands into TCP pacing commands, and vice versa. Integrated congestion control achieves end-to-end traffic management and has many benefits. For example:

- A congested LLC end station can exert "back-pressure" on the ISI by sending an RNR to its access router which in turn slows down ISI traffic with TCP.
- A congested access router can send an end station an RNR to stop it from feeding too much data into the access network. When the access router catches up, it sends the end station an RR.

Like SDLC and LLC, TCP pacing and sequencing adds extra fields to each frame. TCP control fields can add as many as 20 bytes or more overhead to each encapsulated LLC frame. This TCP overhead is in addition to the roughly 20-byte IP fields added during encapsulation. Although TCP brings many benefits to the ISI, there are cases when its overhead is too much for access routers.

If the services of TCP aren't required, a connectionless, non-guaranteed transport protocol like UDP can be employed on top of IP. With UDP, encapsulation overhead is much less, but access routers or end applications have to duplicate at least some of TCP's services.

4

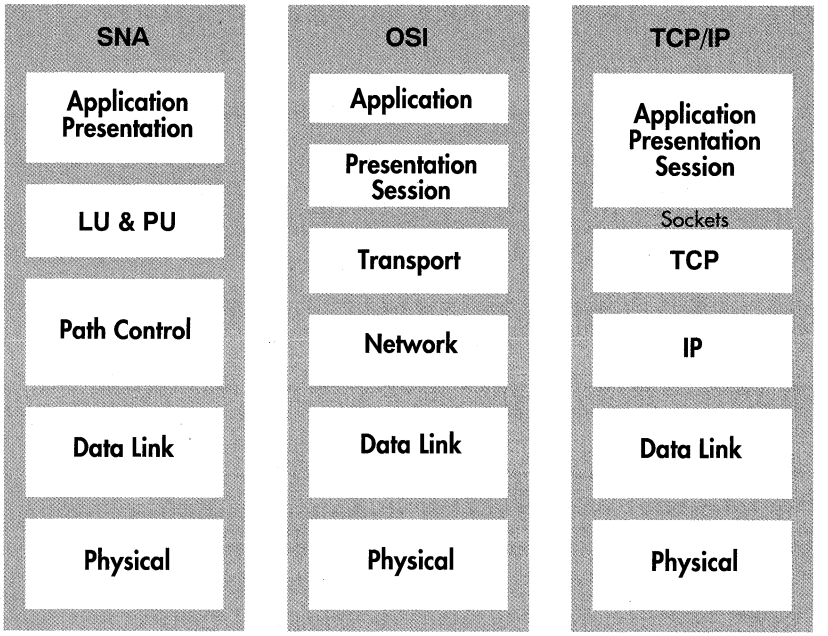
ROUTING INFRASTRUCTURE

SNA REALITIES: ROUTING INFRASTRUCTURE

Above the data-link layer, SNA's Path Control layer provides routing and transport services that are analogous to the layer-3 and -4 services of TCP/IP. Path Control is responsible for routing messages between SNA nodes — FEPs, hosts, and CCs. Above Path Control, SNA defines session-layer software components (LU, PU) that provide a network interface for end users, applications, and management utilities (see Figure 22).

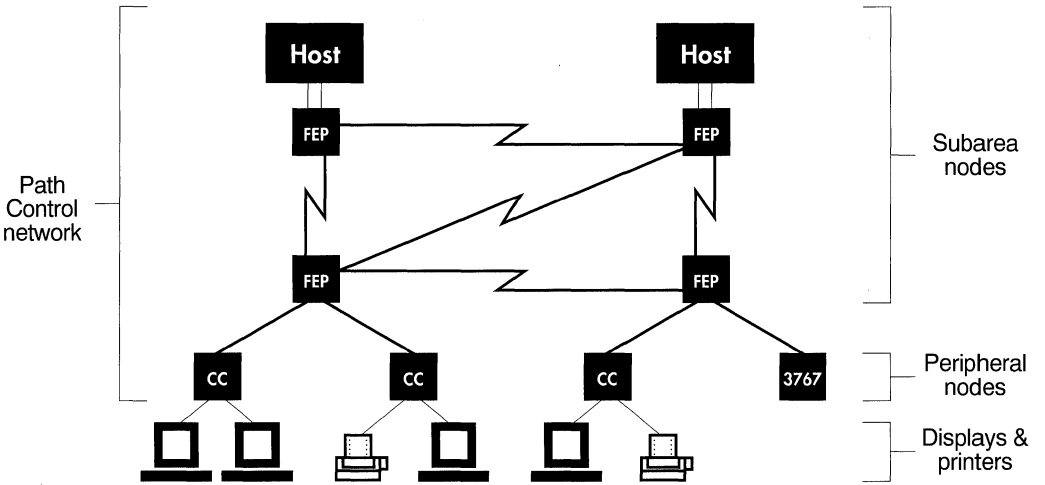
In SNA, "Path Control layer" refers specifically to SNA routing and transport services. "Path Control network" refers more generally to all the physical links and networking and transport software in an SNA network — layers 1 through 4, roughly. The SNA Path Control network connects FEPs, hosts, and CCs in the same way that TCP/IP and ethernet connect routers and end stations in an internetwork.

FIGURE 22. NETWORK LAYER COMPARISON



The SNA Path Control network interconnects two kinds of nodes — subarea nodes and peripheral nodes. FEPs and hosts are considered subarea nodes, while CCs and certain stand-alone terminals are considered peripheral nodes. Display units (e.g., 3278) attach as slaves to CCs and are not considered nodes or part the Path Control network (see Figure 23). An SNA enterprise network is divided into logical subnetworks called subareas. A subarea node and all its downstream peripherals is considered an SNA subarea; typically a subarea comprises a FEP and its CCs. A host (which can have its own channel-attached CCs) also constitutes an SNA subarea.

FIGURE 23. SNA NODES

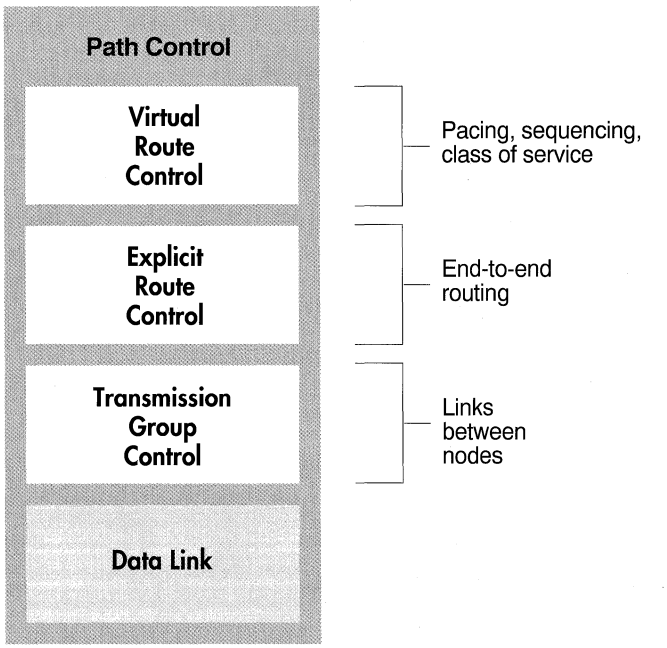


The Path Control network provides two very different kinds of connectivity throughout a subarea network: backbone routing and peripheral routing. In the SNA backbone network, Path Control routes data between FEP subarea nodes with a reliable mesh network. Peripheral nodes do not participate in this routing. Path Control conducts simple point-to-point or multipoint communications in the peripheral links between a CC and its FEP. (Multipoint connections link a number of CCs on a single circuit.) The differences between subarea and peripheral networking are considerable, so they will be discussed separately.

SUBAREA NODES

Between subarea nodes, the Path Control network provides three sublayers of functionality — Transmission Group Control, Explicit Route Control, and Virtual Route Control. These three services are layered on top of each other as sub-components of the Path Control layer (see Figure 24). Transmission Groups (TG) are one or more direct logical connections between two nodes. Explicit Routes (ER) are end-to-end physical paths that can include intermediate nodes. Virtual Routes (VR) add pacing, sequencing, and related services to Explicit Routes. In a loose analogy, TGs, ERs, and VRs correspond to data links, IP, and TCP in the internetwork realm.

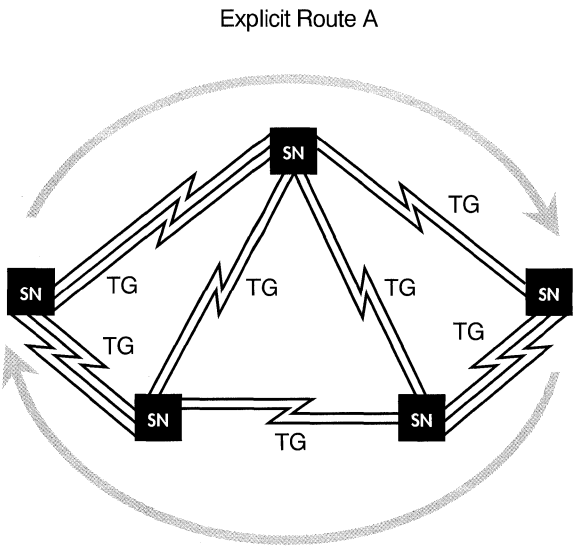
FIGURE 24. PATH CONTROL COMPONENTS



SNA Transmission Group Control defines a direct logical connection between two subarea nodes (e.g., two FEPs). A TG consists of one or more parallel, physical links. A TG makes parallel links with the same characteristics (e.g., multiple SDLC links) look like a single logical pipe to higher layers. TGs perform dynamic load-balancing across links and if a link in a TG fails, traffic is automatically routed to another link in the group.

Like IP, the Explicit Route Control layer provides hop-by-hop routing along an end-to-end path between subarea nodes. ERs are manually defined by SNA system programmers in path tables located on hosts and FEPs. An ER is unidirectional, so it takes two opposing ERs to create a two-way traffic stream. In a mesh subarea network, two subarea nodes can have multiple ERs between them. Figure 25 shows two different ERs and their intermediate nodes. If an ER fails, traffic can be routed to another ER, but rerouting is not transparent to applications (more on this later).

FIGURE 25. EXPLICIT ROUTES



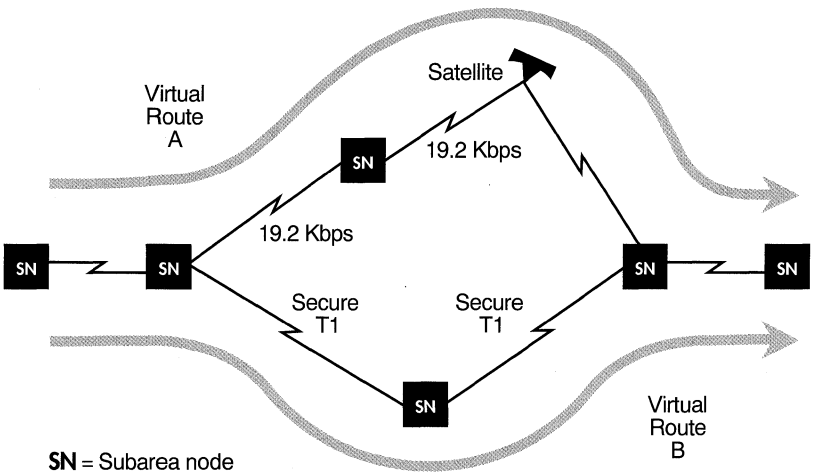
Layered on top of ERs are Virtual Routes that provide two-way reliable transport between subarea nodes. VRs are somewhat analogous to TCP virtual circuits, which exist between two IP end stations. It takes a pair of ERs to support a VR. Multiple VRs can exist between a pair of subarea nodes. A single pair of ERs can support multiple VRs that are differentiated by their transmission priorities (high, medium, low). Consequently, the same physical route can have high-, medium-, and low-priority Virtual Routes. Intermediate nodes forward traffic for routes with the highest priority first.

When an application starts up, a transmission priority can be specified for its traffic. For example, interactive traffic generated by real-time user-to-host sessions may be a high-priority VR while batch transfer traffic is a low-priority VR. SNA Path Control also defines a network priority that routes critical system control messages before end-user messages. This is the highest level of priority.

CLASS-OF-SERVICE ROUTING

In addition to transmission priorities, Virtual Route Control allows applications to select a specific type of network service that's analogous to the OSPF/IP Type of Service routing. Classes of service (COS) can be based on route security, cost, bandwidth, propagation delay, and so on. For instance, with COS routing, an application can be provided with a route that has low security, high propagation delay, and low cost, while another application gets high security, high throughput, and high cost (see Figure 26). A network's service classes are defined in tables on the host. A different class of service can be selected each time a different application session is established across the network.

FIGURE 26. VIRTUAL ROUTE SELECTION



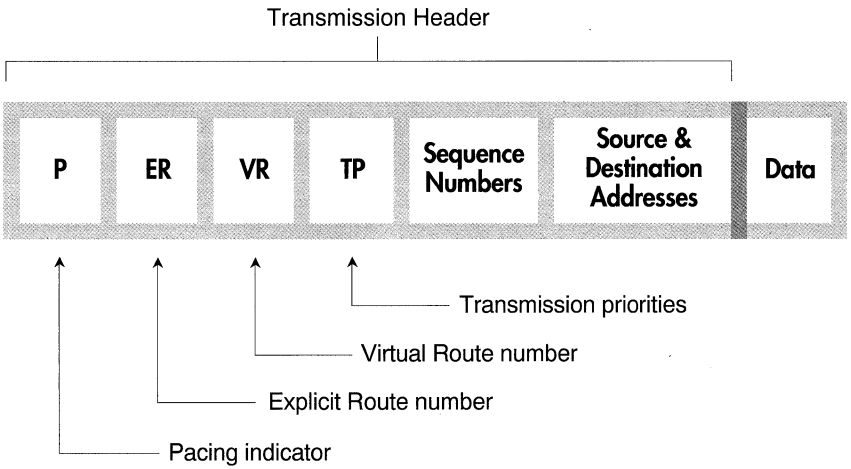
Virtual Route Control also provides error and congestion control. VR congestion control uses acknowledgments and adjustable windows to control the rate of traffic. Although end-to-end pacing windows are maintained by SNA end nodes, any node along the Virtual Route can participate in congestion control by changing the pacing bits in Path Control message headers.

The two subarea nodes at either end of a VR keep sequence counters that are used to check the order of messages sent and received. Each Path Control message has a sequence number field that end nodes compare with their counters. When SNA messages are received out of sequence, they are discarded and retransmission is requested. VR sequencing works much the same as SDLC sequencing. Above the VRs and Path Control, SNA defines session-level pacing that is conducted by LU and PU session-layer software. VR and session pacing are independent. VR pacing controls the rate for all the messages on a Virtual Route; session pacing controls the rate of traffic for individual sessions between applications.

PATH CONTROL HEADERS

The Path Control layer appends a 26-byte transmission header to each SNA message before it leaves a subarea node. Transmission headers have fields for source and destination addresses, Virtual Route number, Explicit Route number, pacing, sequencing, transmission priorities, and other Path Control functions. Figure 27 shows some of the fields in a typical FID4 subarea transmission header.

FIGURE 27. PATH CONTROL FID4 TRANSMISSION HEADER



Path Control uses several types of transmission headers, each with its own format identifier (FID) number. FID4 headers are used to route traffic between subarea nodes that have Explicit and Virtual Routes defined. FID1 headers are used in older subarea networks that do not support Explicit and Virtual Routes. FID2 headers route traffic between subarea and peripheral nodes. FID2- and FID4-style messages are the most common in modern SNA networks but other header types are still in use. FID0, for instance, is used to connect non-SNA machines into a subarea.

Source and destination address fields in FID4 headers convey two-part SNA network addresses as *subarea.element*. The subarea address identifies a logical subarea; the element address identifies SNA nodes and other unique architectural elements (LU, PU) within a subarea. The length of SNA network addresses varies from 16 to 32 bits depending on which version of SNA is deployed.

Not all SNA messages use full subarea.element network addresses. The FID2 headers have only a short local address that is used on links between subarea and peripheral nodes. For instance a 3174 CC is assigned an 8-bit address that only its FEP knows. Local addresses are translated into full network addresses in the CC's FEP subarea node.

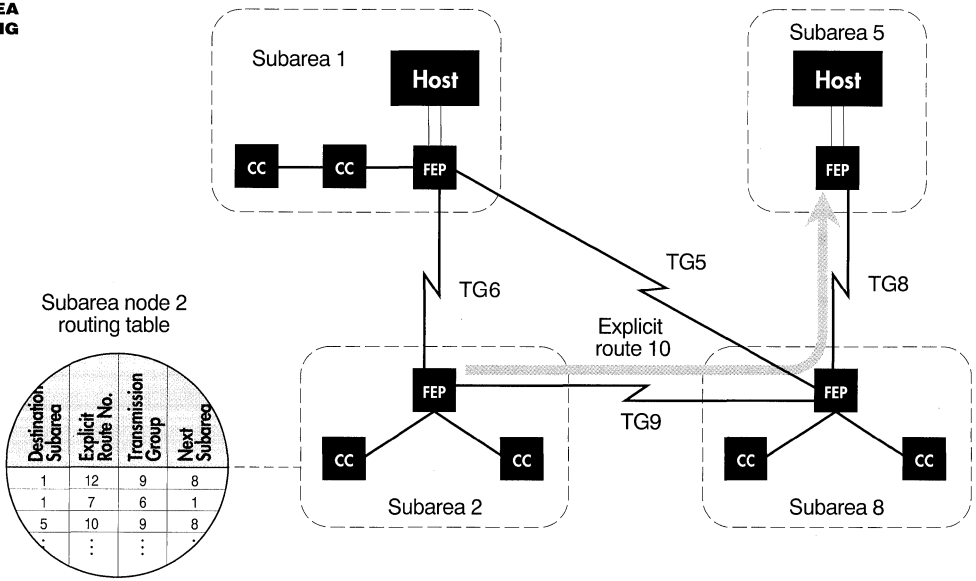
ROUTE DEFINITION

Before routing can take place, an SNA systems programmer must define all possible routes between subarea nodes. This process of configuring physical and logical connections is called *system generation* (SysGen). As changes are made to the network, path tables must be manually updated to remain consistent between routing nodes. Paths must be defined on both FEPs and hosts.

In addition to Explicit Routes, SNA host tables contain information on Virtual Routes, transmission priorities, service classes, logon privileges, and other network details. In host Class-of-Service tables, system programmers give names to the various service classes and map Virtual Routes to them. End users can use COS names to request a certain level of network service when they start an application session. The host network software, VTAM (Virtual Telecommunications Access Method), uses the COS name to determine the appropriate VR. The VR number is then used to select an available ER. Once an ER number has been determined, it is used to route messages from end to end for the duration of a session.

FEP tables do not contain information on the entire topology. Path tables in the FEP contain definitions for all possible subarea destinations from that FEP (see Figure 28). Table entries only give the TG link to the next hop in an Explicit Route. When an subarea SNA message comes into a FEP, NCP software in the FEP looks in the FID4 transmission header to find an ER number and a destination subarea node address. By consulting its routing table, NCP learns the next TG and subarea in the route. The message is output on that TG.

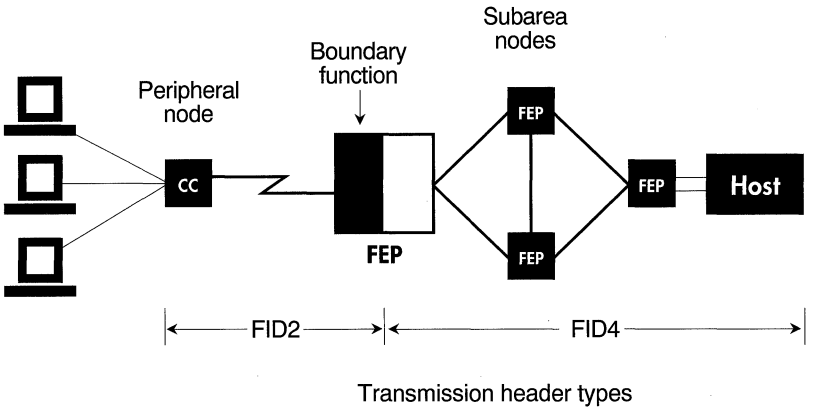
FIGURE 28. SUBAREA ROUTING



In Figure 28, subarea node 2 has a routing table with definitions for routes to subareas 1 and 5. The diagram shows ER 10, which is used for all messages with an ER number 10 in their transmission header. Note that there are two possible ERs from subarea 2 to subarea 1. Hence, subarea 2 has two different table entries for destination subarea 1.

The link between a subarea node and a peripheral node is a simple point-to-point or multipoint connection that doesn't require sophisticated mesh routing. Messages travelling upstream from a CC to a FEP use a small 6-byte FID2 transmission header with 1-byte address fields. When a FID2 message reaches a FEP, it is converted to the 26-byte FID4 format by the FEP's boundary function. Hosts can also provide the boundary function for their channel-attached CCs. On the return trip, FID4-style messages are converted to FID2 before being sent to the CC (see Figure 29).

FIGURE 29. THE BOUNDARY FUNCTION



The boundary function keeps tables that map local addresses to network addresses. This shields peripheral devices from the overhead of subarea addressing and from Explicit Route complexities. SNA peripheral communications is designed to limit the demand put on CCs, minimizing the need for large buffers and processing capabilities. The portion of an end-to-end SNA path that resides between a subarea and a peripheral node is referred to as the route extension.

In summary, SNA subarea networks provide users with reliable and deterministic transport services. With Virtual Routes and transmission priorities, SNA accommodates a wide variety of route characteristics. Unfortunately for SNA network managers, this functionality comes at the cost of arduous manual path definitions. The specification of Explicit Routes, Virtual Routes, and Transmission Groups is a time-consuming, error-prone task that is

typically repeated each time a link or node changes. In many cases, a production network must be taken out of service for a new SysGen to take place.

Excepting the link-layer redundancies between adjacent nodes, the Path Control network has little in the way of automatic fault tolerance. Backup routes are possible, but if a route fails, its sessions are terminated and applications must explicitly request a new session before the new route is selected. As discussed in the following sections, TCP, APPN, and other internetwork protocols can automatically reroute sessions without disrupting applications.

ISI Solutions: Routing Infrastructure

SNA has for many years provided backbone routing for hierarchical enterprise networks. If an Integrated SNA Internetwork is to replace the SNA infrastructure, it must provide everything SNA does and more. Inherent in ISI technology is support for dynamic adaptive routing and complex mesh topologies, and compensation for the following SNA weaknesses:

- SNA/SDLC links are not multiprotocol
- SNA/SRB topologies are bridged, not routed
- SNA Path Control routing is not used in SRB networks
- SNA subarea routes are manually configured
- SNA rerouting interrupts application processing
- SNA nodes are bandwidth-constricted (limited throughput)

The backbone of an ISI is constructed with high-speed multiprotocol routers that learn complex mesh-network topologies automatically, without manual path definitions. A multiprotocol router maintains a separate routing table for each protocol it supports: IP, NetWare, DECnet, OSI, AppleTalk, etc. In addition to routing, routers also support transparent, translation, and source route bridging, and related protocols. This combined router-bridge functionality facilitates migration from existing bridged SRB networks to the routed multiprotocol ISI.

ISI routers convey SDLC traffic by encapsulating it in MAC frames or IP packets. In smaller networks, routers use SRB protocols to bridge SNA and NetBIOS in their native LLC frame formats. On the enterprise backbone, routers encapsulate and transport SNA and NetBIOS traffic via IP. To ensure sequencing and end-to-end reliability for encapsulated traffic, an ISI deploys TCP protocols on access routers.

Traditionally, IP networks have relied on RIP (Routing Information Protocol) to build and update routing tables in each router. But RIP is inefficient because it sends complete routing tables with each update. When links or routers in a network change states (e.g., add, delete, failure), RIP doesn't learn the new topology very quickly (convergence). RIP is being superseded by the OSPF (Open Shortest Path First) protocol standard. OSPF is a sophisticated link-state route update protocol that allows routers to dynamically exchange specific information about their links, enabling rapid, bandwidth-efficient topology convergence for internetworks of all sizes.

OSPF PRIMER

OSPF is well suited to play a key role in the ISI backbone because it has the following characteristics:

- Automatic calculation of routing paths
- Automatic routing around link/node failures
- Highly scalable topologies (1-to-25 or more hops)
- Full spectrum of WAN link operations
- Least-cost routing metric
- Traffic prioritization
- Type-of-Service route calculation
- Fast topology convergence
- Automatic load balancing

Each OSPF router keeps a database that describes paths to every other router in the topology. When a router or link is added to or deleted from the network, Link State Advertisements (LSA) are broadcast to every router, updating topology databases. (RIP typically sends complete routing tables to routers every 30 seconds whether there have been topology changes or not.) OSPF routers update each other with small update messages that only reflect changes to link states, not entire routing tables. OSPF LSA updates are typically sent at 30-minute intervals or when there is a link or router state change.

Once links and routers are assigned IP numbers, OSPF route determination requires no manual effort. OSPF/IP backbones have few topology constraints and can by fully meshed with many redundant paths between routers. The OSPF routing algorithms can adapt to the full range of network sizes, from small workgroups to global internetworks. When parallel links exist, traffic can be load-balanced across them. OSPF is link-protocol independent, so all of the latest LAN/WAN communications methods (frame relay, SMDS, etc.) can be deployed on backbone or access routers.

OSPF METRICS

RIP and other unsophisticated routing techniques choose routes by applying a single, simplistic routing metric, for instance, least-number-of-hops. OSPF, in contrast, can choose routes by a user-defined metric that assigns different costs to links depending on bandwidth, transit times, or other values. Once links are assigned values, OSPF automatically computes the optimum route on a least-cost basis.

The OSPF route selection process can be further fine-tuned with the Type of Service (TOS) feature. TOS allows user applications and network software to request a specific type of end-to-end route for traffic. When TOS is deployed, routers keep separate routing tables for routes that are differentiated by throughput, reliability, and propagation delay.

For instance, an interactive, time-sensitive application can use the TOS feature to request a high-bandwidth connection that introduces little delay. In contrast, a file transfer or batch-oriented application can request a low-cost connection with high delay and low throughput. With a powerful multipro-

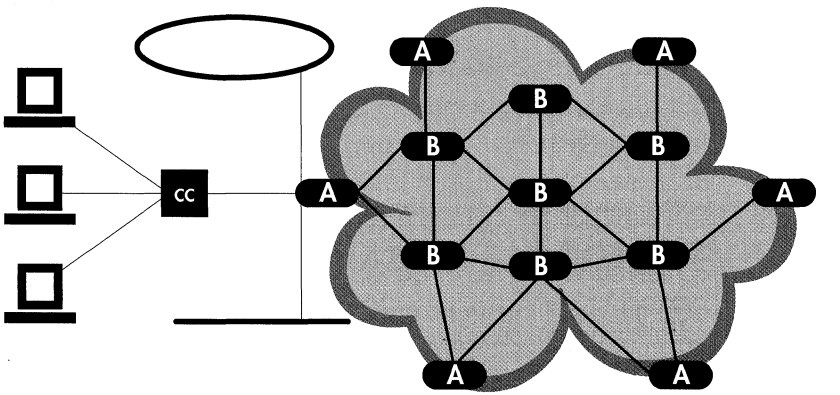
tocol router platform and a TOS implementation, there is no reason why SNA COS routes couldn't be mapped to OSPF TOS routes, allowing the same metrics to apply. This feature will mature as ISI technology advances. OSPF routers can also give priority to time-sensitive traffic on a packet-by-packet basis, in the same way SNA Path Control does.

ADDITIONAL ISI BENEFITS

In addition to the advantages of a high-performance multiprotocol OSPF backbone, the ISI has additional features that ensure good response times for SNA applications (see Figure 30):

- Traffic priorities
- Packet filtering
- Port-by-port protocol configuration
- Redundancy and fault tolerance

FIGURE 30. ISI INFRASTRUCTURE



B = Backbone router

A = Access router

Access routers

- Interface and protocol configuration
- Filtering

Backbone network

- Adaptive routing
- Fast convergence
- Fault tolerance
- Type of service
- Transmission priorities

ISI routers allow network managers to set high, medium, and low traffic priorities that can be tied to source addresses, destination addresses, packet lengths, protocol types, as well as other user-defined fields. This means that traffic destined to a certain node or application can be given high priority based on MAC, SAP, LU, or PU address. Broadcasts and other non-data packets can be given a lower priority.

With administrative traffic priorities, interactive and time-sensitive applications can be given an expedited delivery service without involving application layers. An ISI's ability to prioritize traffic is related to the size and latency of router queues (buffers). To fine-tune response times, maximum queue size and queue latency values can be adjusted on a router-by-router basis.

CONFIGURING PORTS, PROTOCOLS, AND FILTERS

Folding SNA, LLC, SDLC, and NetBIOS into an ISI requires a great deal of configuration flexibility on the part of routers. A typical multiport, multiprotocol router has anywhere from a few to fifty or more ports. To ensure optimum traffic management, each routing and bridging function that a router supports is configurable on a port-by-port basis.

For instance, a multiport router can be configured to bridge SRB traffic on one port and encapsulate SRB in IP on another port. In a complex enterprise ISI, some routers may act as multiport bridges while others act as multiport routers or hybrid bridge/routers. The ports on access routers can be enabled for the specific protocols that are allowed to pass from the access network to the backbone — all other protocols remain local. Because backbone routers handle traffic from many access networks, their ports can be enabled for all enterprise-wide protocols.

Similarly, routers can provide traffic filters which either block or pass packets based on user-defined packet fields. For example, routers can selectively forward or drop packets as a function of protocol type, source or destination addresses, or packet length. Packets in both bridged and routed traffic streams can be filtered, including SNA, NetBIOS, NetWare, DECnet, OSI, TCP/IP, and so on. High-end routers allow network managers to set dozens of filter settings for each port on a protocol-by-protocol basis.

The ability to apply filters uniformly to all traffic brings benefits in several areas. For instance, security can be enforced throughout the enterprise by blocking unwanted packet types from entering backbone or access net-

works. For unsecured areas, filters can be relatively open. For secure areas, filters can be tightly defined to pass only authorized traffic from devices with approved addresses.

Uniform traffic filters can fine-tune backbone performance by determining which packets pass into the backbone and which must remain on access networks. This benefits bandwidth conservation because in many cases ISI access networks have large amounts of local-only workgroup traffic. For instance, filters can be set so that only SNA, 3270 gateway, and critical client/server traffic gets through to the backbone. This conserves backbone resources and ensures efficient use of WAN links.

Filters also help enforce network usage policies. Because traffic can be filtered by source and destination addresses, users can be restricted in their network usage patterns. For instance, the members of a workgroup can be given access to certain remote LAN segments but not others. This control greatly facilitates network capacity planning and management.

SNA backbones are based on technology and products that are mature and relatively slow to change. For instance, the pace at which fault tolerance has been added to SNA has been particularly slow and frustrating for many network managers. ISI routers, in contrast, are based on advanced software and hardware technologies that have built-in redundancy and fault tolerance, including:

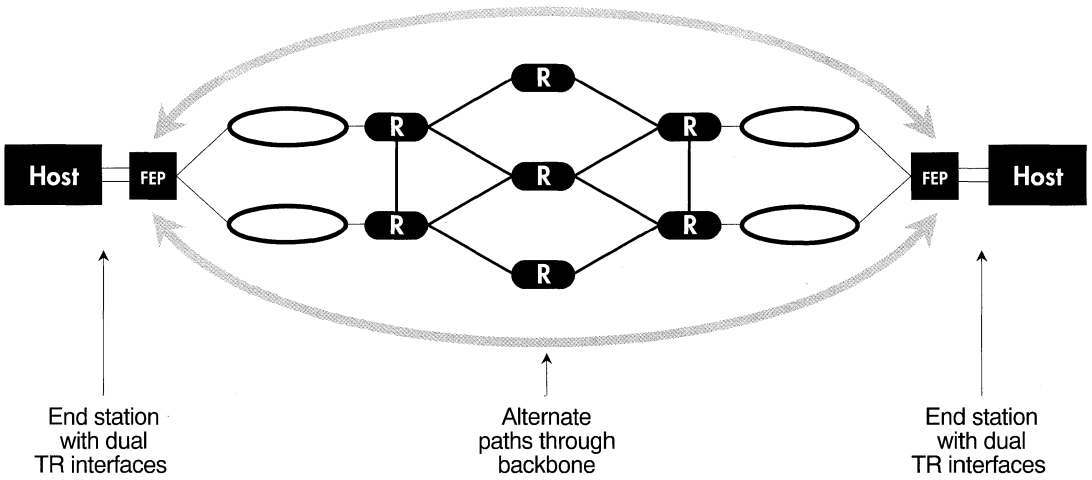
- Redundant WAN links
- Redundant LAN links
- Redundant routing nodes
- Fault tolerant router components

Large SNA networks typically connect major communications nodes — FEPs — in a mesh configuration. But due to the cost of FEPs (hundreds of thousands of dollars), redundant nodes and links in a subarea network have limited practicality. With such high costs and limited functionality, it is generally not an option to buy an extra FEP wherever fault tolerance would

be useful. And while subarea nodes may be interconnected with meshed T1 or fractional T1 lines, downstream FEP-to-CC links are simple low-speed connections that are not typically redundant.

Because the ISI is built from relatively inexpensive multiprotocol routers, redundant intermediate nodes and links are quite practical. In mission-critical applications it is not unusual for both end stations and routers to have multiple backbone connections (see Figure 31). If an interface card or link

FIGURE 31. LAN/WAN FAULT TOLERANCE



fails, another is immediately available. OSPF routing software fully exploits a highly meshed configuration by switching automatically to alternate routes when a link or node fails. With a highly meshed LAN/WAN topology, the chance that all possible paths will fail is effectively nonexistent.

Routers are based on state-of-the-art hardware technologies that are more sophisticated than the older FEP and CC platforms. High-end routers with symmetric multiprocessing/processing architectures have no single point of hardware or software failure because they duplicate critical management and protocol processing functions across multiple network modules. Failed

modules can be replaced or "hot-swapped" without bringing down the router. With redundant routing nodes, power supplies, LAN links, and WAN links, the ISI can exceed the reliability of SNA networks without incurring excessively high costs.

And then there's the issue of performance. FEPs generally forward SNA traffic at a rate that's considerably less than 1000 packets per second. In comparison, high-end multiprotocol routers are designed to sustain 15,000 to 500,000 packets per second. Consequently, routers vastly outperform FEPs and controllers in both backbone and access roles. In any reasonable comparison, ISI routers have a large price/performance advantage.

EXPLOITING WAN RESOURCES

SNA devices often rely on multiplexers to manage wide-area enterprise bandwidth. Typically a T1 multiplexer divides WAN lines into 56 Kbps or other increments that are used for SDLC links among FEPs and CCs. Like SNA devices, routers can use channelized bandwidth, but they do not necessarily require muxs for connection to subrate, T1, or T3 services. ISI routers with D4 framing capabilities can connect directly to DSU/CSU equipment, enabling bandwidth management that's optimized for internetwork traffic. SNA traffic can be merged with internetwork traffic or any number of a T1's 24 DS0 channels can be dedicated exclusively to SNA traffic — further guaranteeing fast response times for interactive terminal users.

Routers lead the communications industry in support for new wide-area protocols and services. Routers were early adopters of frame relay and they can be configured with built-in 802.6 interfaces for direct connection to SMDS. Because router platforms don't have the development and performance bottlenecks of older FEP and CC devices, emerging WAN technologies like ATM and SONET can be rapidly incorporated into the ISI architecture.

5

SNA's Upper Layers

SNA Realities: The Upper Layers

SNA Path Control layers (TG, ER, VR) map fairly well to the TCP/IP internetworking model. The same cannot be said for the upper layers of SNA, which have their own unique structure and terminology bearing little resemblance to the upper layers of internetwork or OSI architectures.

The way applications use TCP/IP or similar internetwork protocols is quite straightforward — PCs and workstations have local operating systems that service application and utility programs. To access the network, programs interface with a session- or transport-oriented API (application program interface) at the top of the stack. Intermediate internetwork nodes — routers and bridges — run layer 2 or 3 protocols and don't generally get involved in applications or high-level protocols. In some special cases (an ISI, for instance), routers run layer 4 TCP for error and congestion control of encapsulated traffic between access routers.

In contrast to internetwork routers and bridges, SNA devices get involved in both high- and low-level network protocols. SNA's architectural definitions extend from the lower routing layers all the way up to application-layer data streams of specific display terminals (e.g., 3270, 5250). SNA display terminals don't have the benefit of a local operating system, hence the need for upper layer functionality in cluster controllers. Above Path Control, SNA provides session, presentation, and application layer services for data generated by terminal users, I/O devices, and host applications.

In addition to architectural differences, the terminologies of SNA and internetworking are quite divergent. In the internetworking realm, users are people who access local PC or workstation application programs that in turn access network protocols. (The exception to this is LAT or similar terminal servers, which use the network to convey dumb terminal I/O data.)

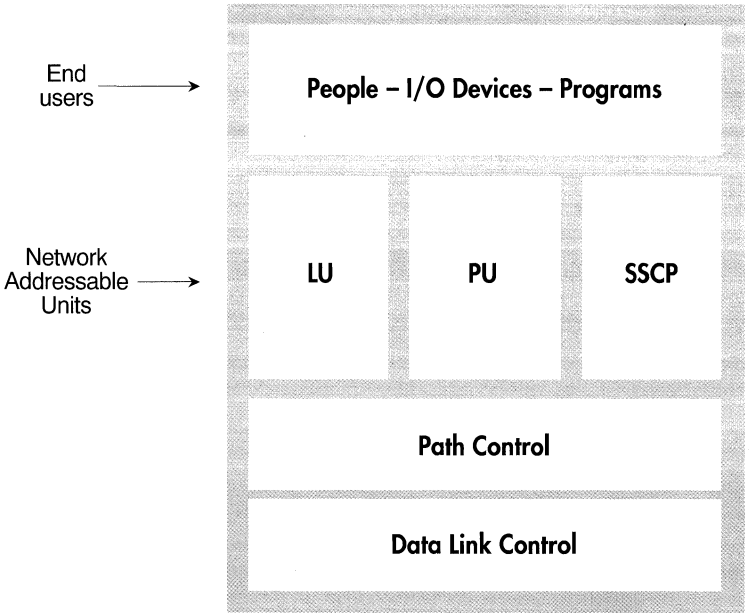
In SNA, network end users are defined as "people, programs, or I/O devices." This generalized definition stems partly from the fact that human operators can directly access SNA protocol services (LU, PU) through a terminal. In contrast, internetwork protocols (e.g., TCP/IP, IPX/SPX, DECnet) do not provide higher level services and hence require an application or utility layer between end users and protocol elements. Considering that SNA has its own special definition for "user," it is clear that the process of integrating SNA traffic into an ISI must include an effort to resolve differences in terminology between these two worlds.

The task of an SNA network is to join a terminal operator or I/O device with a host application program. SNA can also connect two remote or local host applications. SNA end users communicate with each other by way of communication software/firmware elements defined as network addressable units (NAU). Residing above Path Control, NAUs provide various session, presentation, and application layer services. All SNA sessions take place between NAUs (see Figure 32).

SNA defines three types of NAUs: logical units (LUs), physical units (PUs), and System Services Control Points (SSCP). NAUs are all located at the same level of the stack and are differentiated by the functional role they play in the SNA network. Comparing NAUs to internetwork protocols is difficult because the OSI model differentiates protocol layers, not functions within layers. For example, the OSI session layer is always called the session layer, no matter what function it provides to applications, utilities, or management software.

Internetworking aficionados should think of NAUs as upper layer (roughly session and above) software or firmware modules that reside in SNA nodes,

FIGURE 32. UPPER LAYER SNA



providing end users with a high-level network interface to logical end-to-end sessions. In providing SNA end users with sessions, NAUs conduct session initialization, synchronization, sequencing, pacing, and related services. The pacing and sequencing of NAU sessions is in addition to — and independent of — lower layer SNA Path Control services. In addition to coordination of application dialogues, NAUs have a Function Management sublayer that defines data format and display for terminal operators.

LU, PU, SSCP

LUs reside on hosts and CCs, and give users access to applications by setting up end-to-end sessions and formatting data streams. Hosts and CCs typically contain multiple LUs, each capable of supporting one or more sessions with other NAUs. FEPs do not run end-user applications and therefore don't generally support LUs. But in some cases FEPs run special-purpose programs that require LUs.

LUs located in CCs are called dependent LUs because they must rely on the host for session establishment. The most common dependent LU is LU Type 2 (LU2) but there are a variety of LU types that correspond to different SNA devices. LU2 and LU3 carry the 3270 data stream, whereas LU7 and LU4 carry the 5250 data stream used by IBM midrange systems (S/38, AS/400). Dependent LUs can also reside in certain stand-alone IBM terminals, and in PCs or workstations. During the eighties, SNA enhancements defined an independent LU — LU6.2 — that can establish network sessions without host control. Unlike the other specialized LUs, LU6.2 can carry any type of data stream.

A single PU resides in each SNA node. A PU's function is to administer a node's physical links, ports, and communications software/firmware. Operators use PU services to configure a node's resources. PUs facilitate the initialization of their node and they also format and synchronize management data that flows to and from the host's centralized management software. LU6.2 can run on a number of transports but in APPN environments it uses a special Type 2.1 PU that provides directory and routing services, enabling peer-to-peer sessions with or without host involvement (detailed in the upcoming APPN section).

SNA nodes are identified by the type of PU they contain. The PU in a host is called a PU Type 5 (PU5); the PU in a FEP is a PU Type 4 (PU4); the PU in peripheral nodes are either PU Type 1 or more commonly PU Type 2 (PU2). The ubiquitous 3174 CC is often referred to as a PU2.

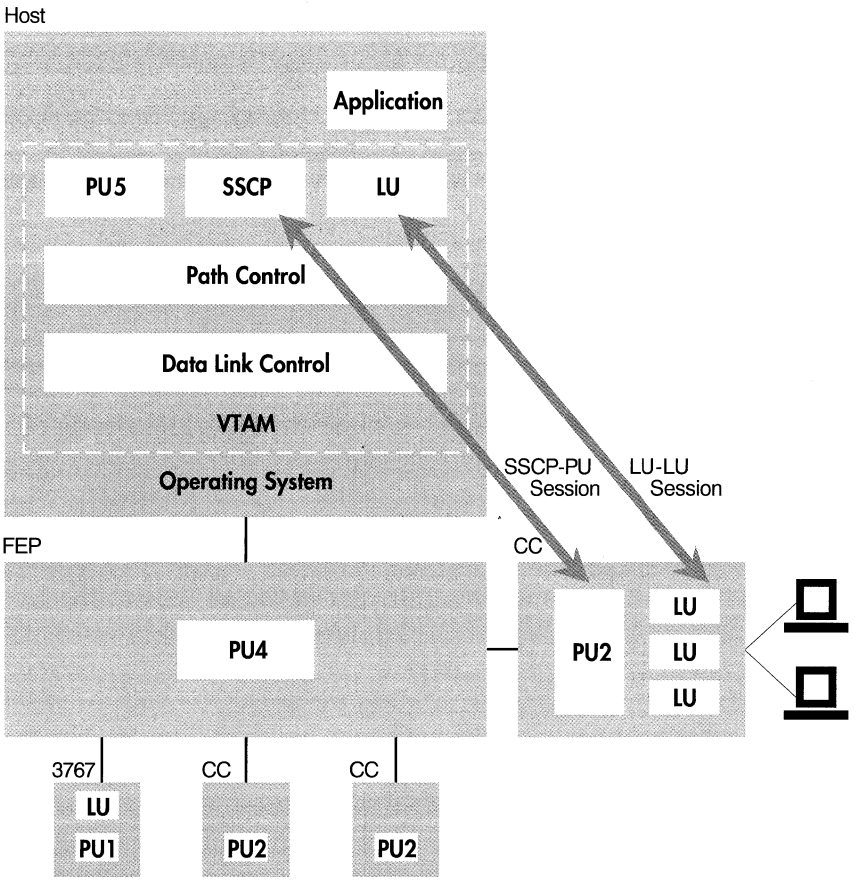
The SSCP is a highly specialized NAU that resides only on host nodes. Implemented in VTAM software, SSCP is the master SNA control point, providing centralized control of links, nodes, and communication activities throughout the host's domain. An SNA domain consists of all the subareas downstream from a VTAM SSCP.

NAU SESSIONS

There are several types of sessions possible between the various NAUs. LU-to-LU sessions support end-user applications. Sessions between PUs and the SSCP handle various management and administrative functions. For instance, NetView management information is conveyed to and from the host

via SSCP-to-PU sessions. SSCP-to-LU sessions are used when a user first logs on to set up an LU-to-LU session. Figure 33 shows the PU types (1, 2, 4, 5) and some of their interactions. In the diagram, the lower right PU2 is shown with its resident LUs and terminal connections.

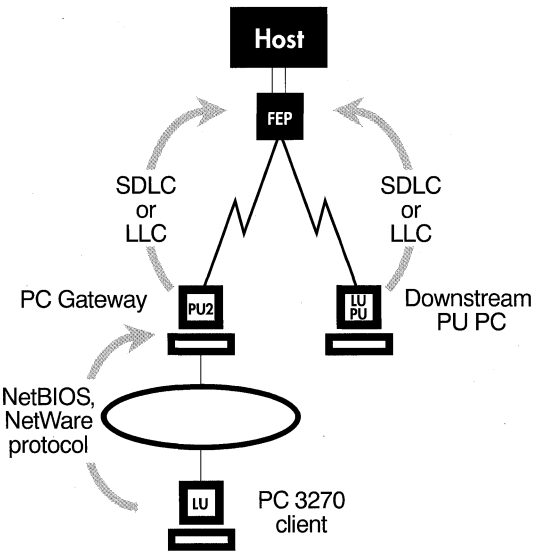
FIGURE 33. SNA INTERNALS



In today's SNA networks, peripheral CCs and display terminals have been supplemented by large numbers of PCs and 3270 gateways. As far as host VTAM software is concerned, an SNA gateway looks and acts just like a PU2. A PC with 3270 emulation software looks like a 3270 terminal. IBM and third parties implement SNA gateways in a variety of ways, but they all basically mimic LU and PU functions.

The most basic 3270 gateway consists of a DOS or OS/2 PC or a UNIX workstation that emulates a PU2 cluster controller. The PC or workstation can be dedicated solely to the gateway function or can also support end-user applications. Gateways link to an upstream controller or FEP with SDLC or LAN connections. PC and workstation clients typically link to the gateway via a LAN (see Figure 34). The protocol between the PC and the gateway is

FIGURE 34. PC 3270 GATEWAYS



typically supplied by NetBIOS, NetWare IPX/SPX, or similar LAN protocols. The PC's 3270 software emulates all or part of a dependent LU2. Software on the gateway machine emulates a PU2 and the SNA peripheral node Path Control function. This division of LU and PU functions on gateways and clients is called the split-stack gateway approach.

As PCs gained processing and I/O power, it became possible to put the entire LU/PU/Path Control stack on a single client machine. Available from IBM and other vendors, this configuration is sometimes called a Down Stream PU (DSPU) and is implemented in programs for OS/2 and other operating systems. PCs acting as DSPUs can talk directly to FEPs and other upstream nodes through a LAN or SDLC connection. Some vendors provide multipurpose gateways that support host connectivity for both DSPUs and split-stack clients.

ISI Solutions: The Upper Layers

Because 3270 gateways implement SNA link station, Path Control, and PU functions, all the previously covered ISI issues concerning polling, broadcasts, and acknowledgments apply. Token ring-based SNA gateways and DSPUs locate upstream nodes with source routing broadcasts. Once an SNA node is located, an end-to-end LLC2 session is established. Consequently, the need to control SRB broadcasts and LLC acknowledgments in the gateway environment is just as great as it is with CCs.

Although PCs emulate 3270 protocols, their traffic patterns are not always identical to those of terminals. In addition to interactive terminal-to-host sessions, PCs and workstations can conduct file or batch transfer operations with a host. These bulk data transfer operations put a greater load on the network than interactive screen traffic.

The ISI is well suited to support 3270 gateways in both interactive and bulk transfer operations. With an ISI, response times for interactive PC traffic can be optimized with IP encapsulation, packet prioritization, and packet filtering methods. For bulk data transfer, the ISI has the advantage of flexible bandwidth that can be allocated on an as-needed basis. With high-performance routers and load balancing, the ISI backbone readily adapts to random bursts of 3270 data transfer traffic.

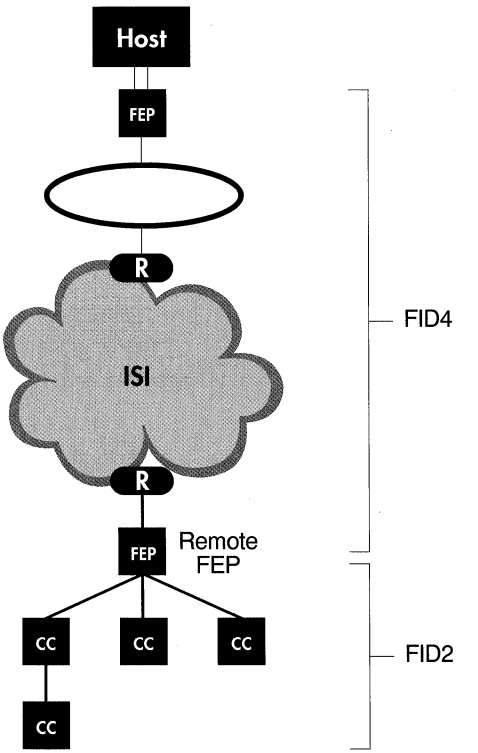
Traffic Management

SNA application sessions are configured and managed by their assigned LUs and PUs. To facilitate SNA integration, the ISI can filter and prioritize SNA traffic by LU or PU address fields. This means that an ISI's responsiveness to SNA traffic can be fine-tuned at the LU and PU level. For instance, a specific LU-to-LU session can be given a high priority and/or a preferred path through the network. Similarly, all messages to a specific PU can be prioritized, based on address.

The degree to which the ISI can manage NAU session traffic is largely dependent on which part of the SNA network is involved. If the ISI replaces peripheral links (CC-FEP), then filtering and prioritization is limited to the fields in FID2 headers. (Because full network addresses aren't present in FID2 headers, filtering and prioritization can only act on local route extension addressing.)

If the ISI replaces subarea links (FEP-FEP), then the robust address and control information in FID4 message headers can be used for traffic management. In addition to NAU element addresses, FID4 packets contain subarea addresses, Virtual Route data, and SNA priority bits. With this information, the ISI can filter and prioritize traffic by LU, PU, subarea, transmission priority, and class-of-service values. It is important to note that the FEP PU4 boundary function translates message headers from FID2 to FID4 format. If the ISI is to take advantage of FID4 headers, a remote FEP must be located between downstream peripheral nodes and the ISI access routers (see Figure 35).

FIGURE 35. ISI WITH REMOTE FEP



NETBIOS

3270 gateways often communicate with PC clients via the NetBIOS protocol. NetBIOS is also popular for client/server applications and is used by Microsoft and IBM file servers. All major LAN operating system vendors (Novell, Banyan, etc.) support NetBIOS interfaces in their client/server software. Numerous email, database, and workgroup applications use NetBIOS to set up peer-to-peer sessions among token ring and ethernet LAN stations.

On token rings, NetBIOS is typically deployed on top of LLC Type 2 but it can also run on a connectionless datagram link such as LLC1. Though it does not conform well to OSI definitions, NetBIOS is basically an end-to-end session-layer protocol without layer 3 routing capabilities. Because it does not support intermediate nodes or subnet routing, NetBIOS is essentially a local- or remote-bridged protocol.

NetBIOS applications identify each other with a 16-character NetBIOS name. To locate a partner, an application broadcasts an SRB frame containing the NetBIOS name of a target application. When the target station receives the broadcast, it sends back a frame containing its MAC address and a completed RIF field which is used for the duration of the session. The broadcast takes places each time a NetBIOS session is established.

NetBIOS works well for small office-automation applications. But when numerous NetBIOS client/server workgroups are connected in an SRB enterprise network, the broadcast of name-location frames can affect performance, particularly where there are slow WAN links in place. As broadcasts reduce performance, end station time-outs may increase. The LLC T1 response timers, in DOS and OS/2 NetBIOS clients, default to around 1 second and are adjustable.

The ISI controls broadcast frames for NetBIOS in the same way as LLC test frames — IP encapsulation, directed explorers, etc. These techniques greatly reduce the need to adjust timers on NetBIOS stations that use the ISI. In some cases, ISI routers may provide caching of NetBIOS names in access routers. This approach sets up tables that relate NetBIOS names to a location within the ISI topology based on MAC or IP addresses. This allows NetBIOS broadcast frames to be focused on a target access network, not the whole enterprise.

6

ADVANCED PEER-TO-PEER NETWORKING (APPN)

SNA REALITIES: APPN

Traditional host-centric SNA is not well suited to the communications needs of PC file-servers, UNIX workgroups, distributed databases, remote procedure calls, and other cooperative processing methods. In response to new network paradigms, IBM has defined the Advanced Peer-to-Peer Networking (APPN) architecture. With its host-independent routing services, APPN enables peer-to-peer communications between heterogeneous populations of mini, micro, and mainframe platforms. APPN is IBM's answer to internetworking.

APPN and its LU6.2 session protocol are highly strategic elements in IBM's networking plan and are increasingly supported by third-party network vendors. Many of IBM's key products for the nineties are based on APPN/LU6.2, particularly in the areas of client/server software, distributed databases, network management, and dependent LU support. Consequently, ISI routers should deploy the full suite of IBM-defined APPN protocols if they are to fully support future internetwork software and corporate applications.

Although APPN is in many ways derived from SNA, it is a fully independent architecture that is engineered for sophisticated peer-to-peer applications. APPN is also an excellent transport for dependent-LU (3270) traffic. Ultimately, APPN may supplant hierarchical SNA altogether, routing both peer-to-peer and terminal-to-host network traffic. But for the foreseeable

future, the two architectures will coexist in many sites. As APPN takes hold, the tremendous computing power of the host processor will evolve to a role of back-end network server for user-friendly front-end desktop computers.

In the mid eighties, APPN was first introduced on IBM midrange systems as PU Type 2.1. Since then, the PU term has been dropped and APPN nodes are now known as Type 2.1 nodes. A fully functional Type 2.1 node consists of APPN routing and directory services, Type 6.2 LUs, and various administrative and management software elements.

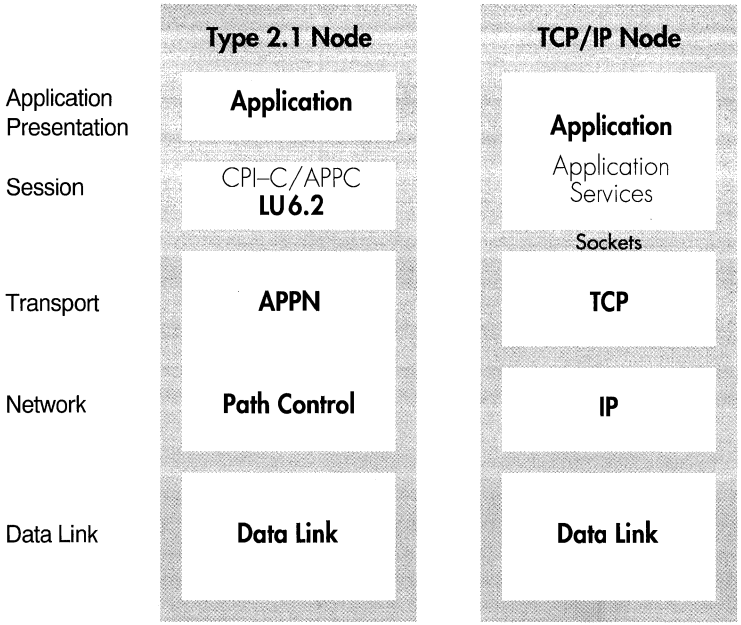
By the end of 1993, APPN will be available in full force on S/370, PS/2, RS/6000, AS/400, and many other third-party platforms. Also widely supported is the session-layer application interface associated with APPN — Common Programming Interface for Communications (CPI-C) — the standard interface to LU6.2. (This was previously called Application Program-to-Program Communications.)

APPN functionality is in many ways analogous to TCP/IP. Both protocol suites operate in the area of OSI layers 3 and 4; both supply routing and transport services (see Figure 36). In addition to routing, APPN provides directory services that locate resources throughout an enterprise network. There are three basic levels of support for APPN:

- Low-Entry Networking (LEN) node
- End node
- Network node

APPN LEN and end nodes can be deployed on all sizes of computers, from PCs to S/370 hosts. Network nodes provide the intermediate routing function and can be located in routers, computers, or IBM communication controllers (FEP, CC). All three node types support LU6.2 sessions and are based on the Type 2.1 node definition.

FIGURE 36. APPN vs. TCP/IP

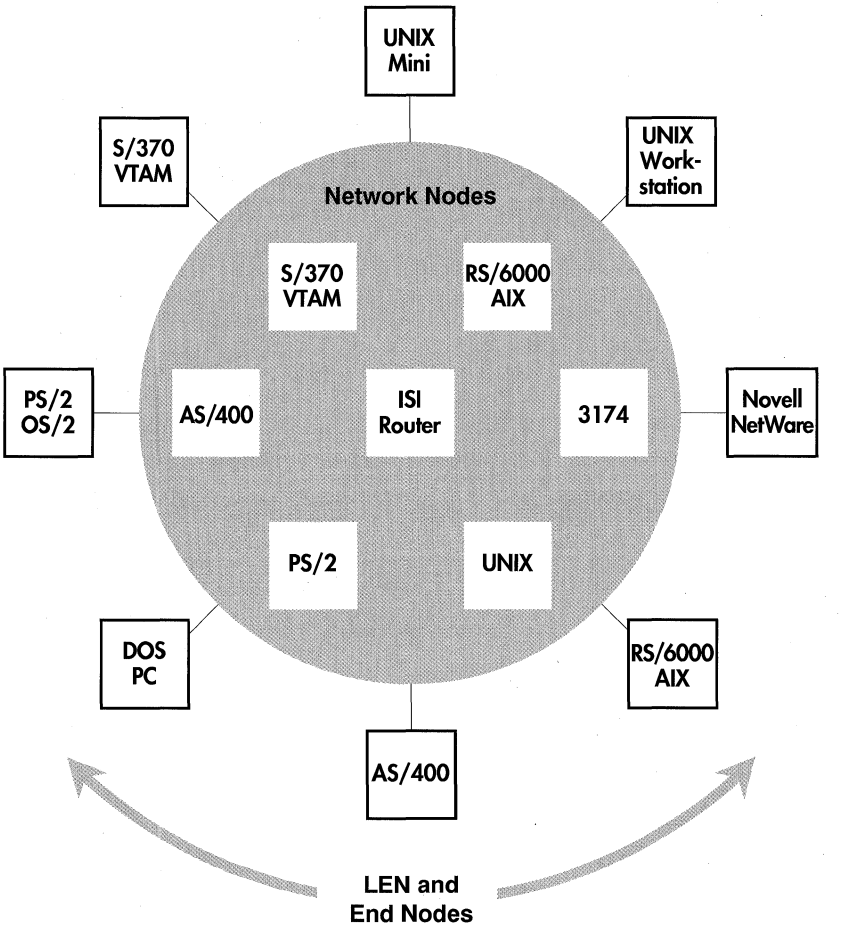


LEN nodes are a class of end node that have only the most basic level of peer-to-peer functionality. They communicate with each other on a direct link, but if they are to use the enterprise-wide topology they must rely on a local APPN network node for all routing and remote resource location. Many systems and software vendors furnish some form of LEN as basic transport for the LU6.2 protocol.

APPN end nodes also support peer applications and the LU6.2 interface, but go beyond LEN by providing automatic directory and routing services that work in tandem with the services of APPN network nodes. End nodes can exchange configuration information with network nodes without operator assistance, reducing manual network configuration.

APPN network nodes service LEN and end nodes with full enterprise-wide directory and routing services. In internetwork terms, an APPN network node is roughly equivalent to a router with built-in directory services and other end node support features. When deployed on a computer, network nodes can also run application software. Network nodes have full routing and directory capabilities whether on an OS/2 PC, mainframe, or router platform (see Figure 37).

FIGURE 37. APPN LEN NODES, END NODES, AND NETWORK NODES



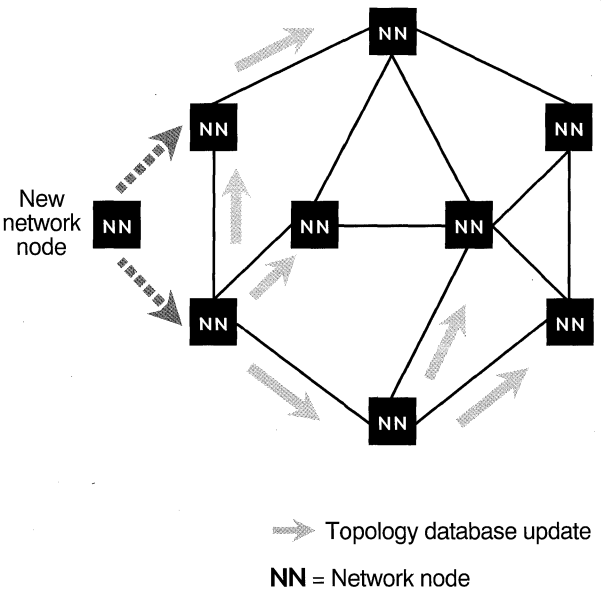
Each APPN end- and LEN-node is assigned a network node server that provides a local entry point into the enterprise network. Network node servers are architecturally the same as other network nodes; additionally, they keep information on the local links and resources of their served end nodes. When an end node wants to set up a session through the network, it contacts its network node server.

APPN ROUTING

Unlike SNA, APPN adaptive routing does not require repetitive manual network path definition in intermediate nodes. APPN network nodes use an OSPF-like link-state protocol to automatically update each other's topology databases. Each topology database contains a full representation of the routing topology (network nodes and their links). When network nodes or links are added or deleted, topology database updates are propagated throughout the network.

Normally, updates only contain state changes. The full database is sent to a newly added network node (see Figure 38). Every five days, network nodes attempt to reaffirm the links in their topology databases. If state information cannot be verified within 15 days, it is purged from the database. This controls the size and overhead of topology information.

FIGURE 38. APPN TOPOLOGY UPDATES



On any network, a major failure can cause excessive updates to flow between routers, impeding the reconstruction and synchronization of routing tables (convergence). To safeguard against router update storms, APPN restricts the flow of topology updates between network nodes with a flow reduction function that prevents duplicate updates from propagating throughout the network. Each network node keeps a sequenced history of the updates it exchanges. If a large APPN network has been divided into subnets, network nodes maintain a topology database for their own subnetwork only (more on APPN subnets later).

APPN ADDRESSING

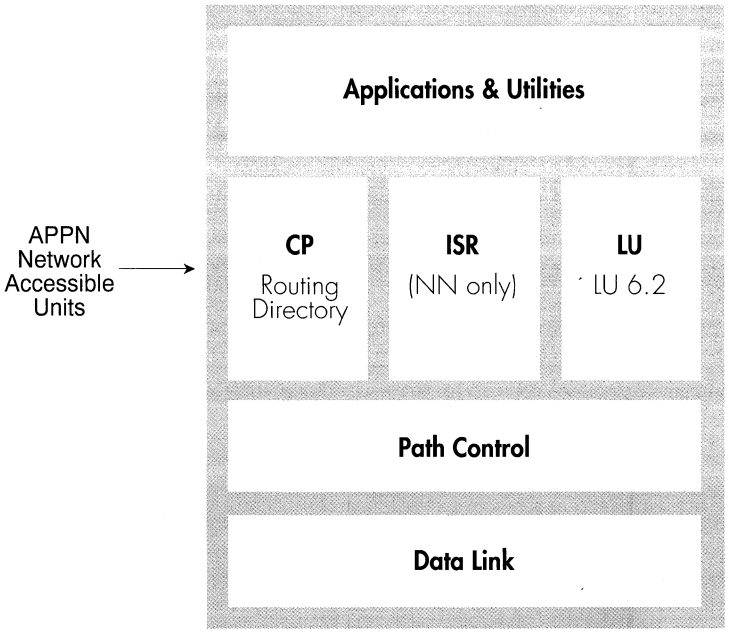
Of the original SNA addressable units, APPN only retains the LU, in the form of LU Type 6.2. The SNA PU and SSCP have been replaced by an APPN Control Point (CP) that provides topology, directory, session, configuration, and management functions in APPN end and network nodes. (APPN LEN nodes do not have the CP function.)

The SNA term "network addressable units" has been replaced in APPN by "network accessible units." APPN LUs and CPs are both considered network accessible units (NAU). APPN NAUs are addressed with a two-part identifier called a network qualified name. These names are alphanumeric and consist of a 1- to 8-byte network name (NETID) and a 1- to 8-byte NAU name — NETID.NAU.

An application can access multiple LUs simultaneously and a single LU can service multiple applications. For instance, a LAN gateway or server can receive requests from many local PC clients and forward them to the host via a pair of bidirectional LU6.2 sessions. This way each LAN station does not have to maintain a permanent session with the host.

An APPN node can contain many LUs but only one CP, so the CP name serves as a unique identifier for each node. In addition to LUs and CPs, APPN network nodes (only) contain a third NAU — Intermediate Session Routing (ISR) — responsible for network node intermediate routing tasks. As with SNA, APPN NAUs access the network via a Path Control layer, which roughly corresponds to the OSI network layer. Path Control in turn accesses the data link layer and the physical network (see Figure 39). APPN is link-independent and can run on top of a number of standard LAN and WAN protocols (e.g., 802.5, SDLC, frame relay).

FIGURE 39. APPN NODE INTERNALS



LUs, CPs, and ISRs are administered via a Node Operator utility, present on every node. This facility allows a node's operator to activate/deactivate links, define/delete LUs, and conduct other configuration and diagnostic tasks. Node operators can be people, transaction programs, or files containing a list of executable commands.

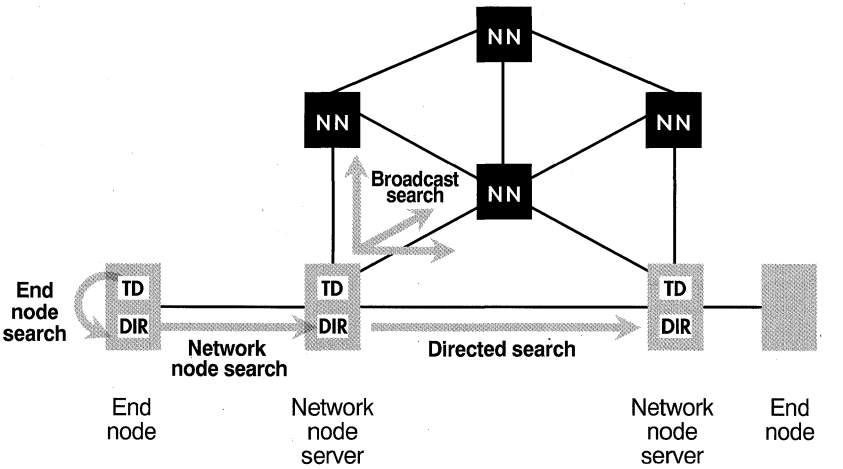
Once an end node's resources (LUs, CP, links) are defined, they can be automatically registered with the local network node server. (LEN nodes must be manually configured.) Unlike topology database information, local end node link and LU information is not shared among network nodes until it is explicitly requested. Network node servers use APPN directory searches to locate end node resources owned by remote network node servers.

APPN DIRECTORY SERVICES

APPN directory services are fully integrated into routing functions. This means that applications can interface with the network by using alphanumeric names (NETID.LU) that are understood by APPN services. Network-qualified names are passed directly to the LU interface and, in turn, to directory services. APPN end nodes and LEN nodes keep a limited list of local LUs but locate remote LUs through network node servers. Network node servers use several searching functions to locate LUs for end/LEN nodes (see Figure 40):

- End node search
- Network node server search
- Network node broadcast search
- Directed search

FIGURE 40. APPN DIRECTORY SEARCHES



TD = Topology database
DIR = Directory

Level 1 End node seach
Network node search
Broadcast search
Directed search

Level 2 Others

When an end node application wishes to establish an LU-to-LU session through the network, its LU asks the Control Point to check the local directory for an entry to the remote LU. If the LU is not in the end node directory, the adjacent network node server checks its directory for the target LU. If the server doesn't have an entry, it broadcasts a search request to other network node servers throughout the enterprise, requesting the target LU's location.

Broadcast searches are potentially bandwidth intensive, given the number of active end nodes on a large network. To limit broadcasts, network node servers store the locations of LUs they find. These cached directory entries contain the name and route to an LU, specified by a list of intermediate nodes and their links (similar to an SRB RIF). Cached directory entries allow end nodes to reach frequently accessed LUs without repeated broadcasts.

When an end node wishes to connect with an LU that is cached, the network node server first verifies the target LU's location by performing a directed search. A directed search uses the information stored with the LU name, avoiding additional broadcasts. This cached source route bridging information is only used to verify the target LUs end node and server. Once this is known, the actual route for session data is calculated using the topology database information stored in the network node server.

For large networks, APPN also supports a centralized directory service that allows network nodes to query a master directory containing enterprise-wide information. Based on a mainframe or powerful server, central directory services conduct searches and directory caching for network nodes. Once cache entries are built up, this technique minimizes broadcast searches and related network overhead.

ROUTE CALCULATION

When a target LU is located, the network node server uses its topology database to calculate the optimum route through the network to that LU. APPN supports class-of-service routing, so route calculation can be influenced by a service request from the application (e.g., low delay, high security, etc.). APPN also supports transmission priorities that dictate how intermediate nodes prioritize traffic (high, medium, or low). An application can specify a class-of-service and a transmission priority by passing a single APPN "mode" name to its local LU during session establishment.

After a route is calculated, an LU-to-LU BIND process sends a BIND message from end-to-end, initializing the routing tables of intermediate nodes along the new session's path. Intermediate nodes identify each link in a session's path with a 17-bit session ID or label that is also carried in each message.

Once the session's path has been initialized by the BIND, network nodes read message labels to determine on which link to forward a message. Labels are dynamically assigned during session establishment and a different label is used on every link in the route — hence the term label swapping, as applied to APPN routing. The overhead in each message of the 17-bit label is less than half that of the 48-bit IP address.

APPN PACING

All messages take the same path through the network after an APPN session begins. APPN can conduct congestion control on a hop-by-hop basis because pacing intelligence is built into both end nodes and network nodes. When an APPN node gets overloaded, it can signal other nodes along the path to send less traffic. Because all hops are involved in pacing, APPN congestion control exerts a high level of traffic management which ensures efficient use of low speed links.

APPN's adaptive pacing can be very different from the congestion control used on TCP/IP networks. IP routers often use a connectionless layer 3 based on unacknowledged datagrams. TCP at layer 4 only conducts flow and error control on end stations (not on intermediate nodes). When an IP router becomes congested or receives bad packets, it simply drops packets, relying on the end stations to retransmit.

TCP/IP is efficient when there are few line errors and little congestion. But when there are substantial line errors or congestion, IP wastes network bandwidth as packets drop and end stations retransmit. APPN's node-by-node adaptive pacing generally avoids packet-dropping, but the additional router overhead may limit router performance on clean high-speed lines.

AUTOMATIC REROUTE

A deficiency of APPN's current routing software is the lack of automatic reroute. If the end-to-end circuit fails, APPN terminates its sessions and applications must restart their sessions again. Upon application retry, APPN reroutes traffic to an alternative path — this is reroute, but not automatic reroute. TCP/IP's datagram-based layer 3 lets routers dynamically route around link or node failures without disturbing applications.

In 1993/1994, IBM plans to address this issue with APPN+, a high-speed routing extension to the current architecture. APPN+ is based on IBM's High Performance Routing (HPR) technology. HPR layer 3, Automatic Network Routing (ANR), provides a source-routed, connectionless service that can drop packets when over-congested.

Rapid Transport Protocol (RTP), a layer 4 end-to-end protocol, resides above ANR. RTP provides error handling with a rate-based flow-control that manages the rate that end nodes pump traffic into the network. Adaptive rate-based flow-control is an advanced technique for handling bursty traffic such as LAN interconnect, multimedia, and video servers at high speeds.

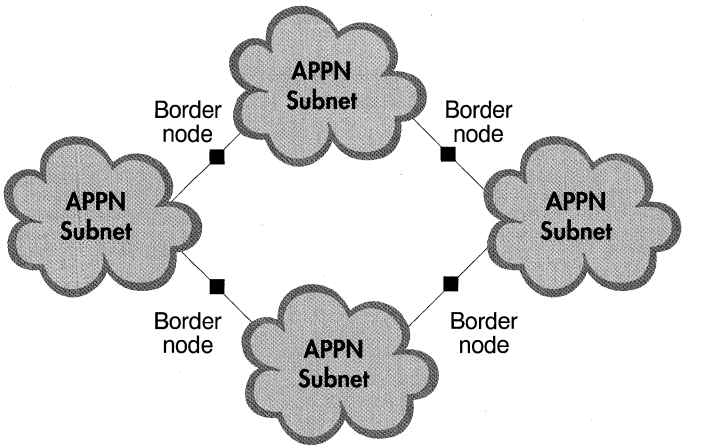
The beauty of HPR is its seamless integration into the APPN architecture. HPR and its connectionless ANR layer are interchangeable with the current connection-oriented ISR routing. In complex networks, HPR is used for links with low bit-error rates, typically high-speed fiber and digital links from fractional T1 up to T3. ISR is used for links with higher error rates — typically, slow or analog lines where hop-by-hop pacing and error control is valuable.

An APPN network can automatically adapt to any mix of ISR and HPR links. The benefits of HPR's connectionless layer 3 efficiencies become apparent with two or more consecutive links. Like ISR, HPR supports full-featured class-of-service routing. Unlike ISR, HPR provides automatic rerouting around link failures — transparent to applications. This is a powerful combination that TCP/IP can't match: connectionless or connection-oriented transport, on an as-needed basis.

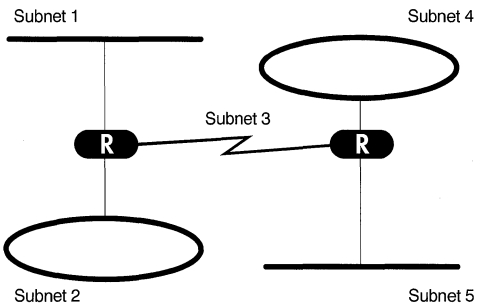
APPN Subnets

APPN supports a wide array of mesh, star, ring, and hierarchical topologies, and it can divide a large network into logical subnetworks. This ensures that traffic flows stay inside the subnet boundaries for the majority of sessions. APPN subnets are connected by border nodes that limit the flow of topology, directory, and user data between subnets (see Figure 41). TCP/IP networks can be subdivided by IP addresses that designate logical subnets, corresponding to a LAN segment.

FIGURE 41. APPN SUBNETS



IP Subnets



APPN addressing is not tied to router ports or LAN segments, so moving an APPN node generally does not require changing its NETID subnet address. APPN subnets — a campus, for example — can be very large with upwards of 10,000 nodes per subnet. For management and administrative reasons, smaller APPN subnets can also be defined.

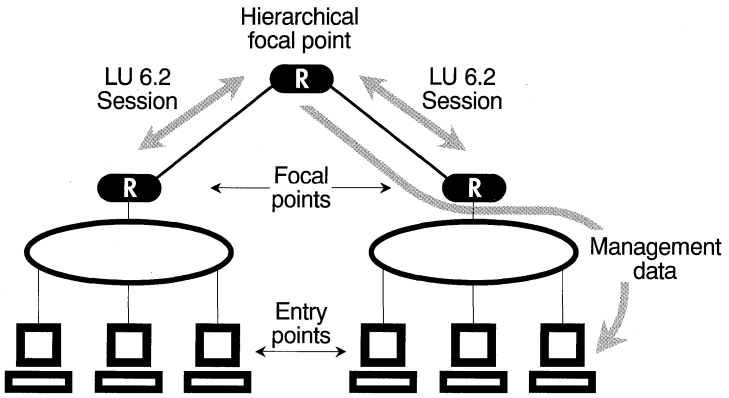
APPN MANAGEMENT

In traditional SNA, VTAM tables are the focal point for configuration, management, and control of enterprise resources. The operator's access to SNA Management Services is through NetView and supporting products. APPN, in contrast, distributes management and administrative functions to nodes throughout the network, eliminating the requirement for a single repository of configuration and management information. (Although architecturally defined, APPN Management Services are not yet fully available in deliverable products.)

APPN management relies on a flexible system of distributed focal points and entry points. Focal points are points of concentration for management information and control. Entry points monitor various network elements and, solicited or not, forward management data (statistics, alerts, etc.) to focal points. Network nodes and end nodes can be either focal points or entry points, but typically a network node provides a focal point for its served end nodes. Focal and entry points exchange management information via LU6.2 sessions.

APPN's management architecture distributes or centralizes management functions to meet the network administrator's needs. APPN Management Services can be hierarchical — a group of focal points can report to a higher-

FIGURE 42. NESTED FOCAL POINTS



R = Router with network node function

level focal point in a cascaded fashion. As shown in Figure 42, the hierarchical approach reduces the number of LU6.2 management sessions by eliminating the need for each entry point to have its own session with a high-level focal point. At the lower layers of a management hierarchy, a network node and its served end nodes can be viewed as a single logical reporting unit by Management Services.

APPN Management Services are divided into categories that include problem diagnosis, configuration, accounting, and operations management. A single focal point can provide all of these services, or services can be distributed among multiple focal points. For example, one focal point could handle network alerts and problem diagnosis; another could track serial numbers and software levels; and a third could monitor network statistics on utilization and availability. A focal point for one management category can be an entry point for another category and vice versa.

APPN Management Services are flexible enough to adapt to a wide range of topologies and management requirements. The internal structure of APPN management is based on OSI CMIP (Common Management Information Protocol). Although the majority of SNA management services are proprietary, OSI management is IBM's stated direction.

ISI SOLUTIONS: APPN

For the remainder of the nineties and beyond, APPN will be increasingly central to the transport of terminal and peer-to-peer traffic in IBM and related environments. Consequently, it is critical that the ISI architecture fully support APPN protocols and fully interoperate with IBM APPN nodes. When in complete compliance with the APPN network node specification, an ISI router assures the highest possible level of services to all types of APPN end nodes and LU6.2 programs.

Nearly all major network vendors are moving to support APPN, but there are many different approaches and time tables. In the router industry, a consortium of vendors and end users has formed to create an architecture that can run encapsulated APPN protocols over a TCP/IP backbone. Termed Advanced Peer-to-Peer Internetworking (APPI), this approach is intended to support APPN end nodes but not APPN network nodes.

APPI intermediate routing will be handled by consortium-selected protocols, (e.g., OSPF and RIP). The goal of APPI is for TCP/IP networks to provide all APPN routing and directory services to APPN end nodes. But because APPN is radically different from TCP/IP, the APPI approach could limit APPN functionality unless the APPN integrated routing and directory services are somehow duplicated by TCP/IP routers. Duplicating APPN services in a TCP/IP infrastructure is a substantial project but the APPI group hopes to publish a complete specification sometime in 1993.

Clearly, the ideal solution for end users is to run native, unencapsulated APPN on ISI routers. APPN is a fully-routable internetwork protocol that does not require TCP/IP transport. ISI access routers are perfectly suited to serving APPN end nodes; ISI backbone routers are perfectly suited to providing the APPN intermediate node function. Because many networks have a considerable mix of SRB, SDLC, TCP/IP, and APPN traffic, the ISI router should support all of these methods equally well.

By complying with the IBM network node specification, ISI routers can fully interoperate with IBM devices and support key corporate peer-to-peer applications. Additionally, full APPN support will ensure that IBM's emerging SNA/APPN integration products (e.g., APPC/3270 and the Dependent LU Server/Requester) will work successfully in the ISI environment.

7

NETWORK MANAGEMENT

SNA REALITIES: NETWORK MANAGEMENT

The effects of change and diversity in today's computing environments are clearly evident in the divergence of SNA and token ring network management products. Traditionally, SNA/SDLC devices have been centrally managed by host VTAM applications, in particular, NetView and its NMVT management protocol. IBM token ring PCs and bridges have their own OS/2-based manager. LAN Network Manager provides monitoring and control of 802.5 interfaces, LAN stations, hubs, and bridges.

In UNIX and TCP/IP realms, IBM offers AIX NetView/6000, an RS/6000-based management product that employs both SNMP (Simple Network Management Protocol) and OSI CMIP (Common Management Information Protocol). In general, internetwork vendors rely on the SNMP protocol to manage routers and other LAN devices. SNMP defines a repository of network device parameters and statistics (the Management Information Base or MIB) and an efficient means to retrieve information from devices. The SNMP GET command retrieves management data; the SET command configures devices.

Like all high-performance UNIX workstations, the AIX RS/6000 is well suited to demanding management tasks and is considered an important platform for future IBM SNA management products, including APPN. AIX

NetView/6000 is the IBM vehicle for translating SNMP and other non-SNA management data into NetView compatible format. With its support for CMIP, AIX NetView is positioned to manage OSI devices as well.

IBM has made a considerable commitment to the OSI management model and is integrating OSI CMIP into products at many levels, from wire hubs to hosts. IBM and 3Com worked together to define the CMOL (CMIP over LLC) IEEE 802.1b standard that runs CMIP protocols over a low-level LLC Type 1 transport. Because it doesn't need a full-blown transport stack, CMOL can be used for management of memory-constrained interfaces, wire hubs, and LAN stations.

With NetView/6000, IBM will provide distributed standards-based management stations with sophisticated graphical interfaces and expert system capabilities. In this scenario, processing and viewing management data migrates from the host console to the workstation, turning the host into a back-end management data server.

The major side-effect of so many different IBM management platforms is a plethora of management protocols (NMVT, CMIP, CMOL, SNMP) and management transports (SNA, TCP/IP, LLC, OSI, LU6.2). While IBM is providing a growing set of tools to interconnect NetView, LAN Network Manager, APPN, and AIX/NetView, there is no single management product that seamlessly spans all environments. IBM-related management methods and their transports include:

- NetView NMVT via SNA
- LAN Network Manager via LLC/SRB
- AIX NetView/6000 SNMP via IP
- CMIP via LLC/SRB (CMOL)
- CMIP via LU6.2/APPN
- CMIP via OSI
- CMIP via TCP/IP (CMOT)

Like the newer IBM products, ISI network management relies on SNMP to control and monitor routers. The ISI also facilitates translating management data into NetView, APPN, OSI, and LAN Network Manager formats. This translation can take place in ISI routers or in external management-protocol gateways from SUN, HP, and other vendors. After looking at traditional SNA management in the next section, integrated ISI management will be examined.

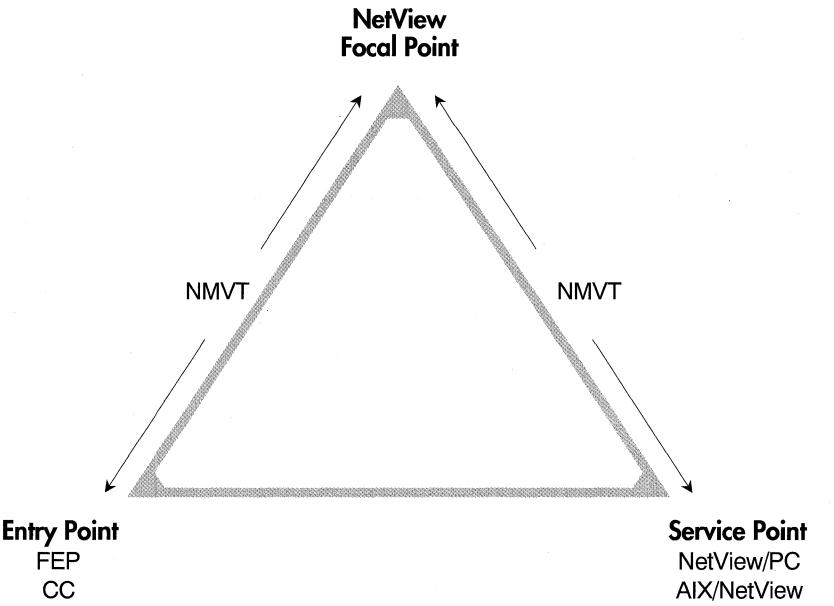
SNA management is traditionally highly centralized, relying on VTAM for collection of management data from FEPs, controllers, modems, and other network devices. In the mid eighties, IBM defined an SNA Management Services architecture that organizes tasks into:

- Problem Management — identifying and diagnosing failures
- Performance Management — tracking performance statistics
- Accounting Management — billing based on network usage
- Inventory Management — tracking software/ hardware components
- Change Management — planning and executing network changes

SNA network management operations are typically controlled from a host-based NetView console that displays both low-level hardware failures and higher-level session and response-time problems. Although NetView initially referred to the hands-on management-console aspects of SNA, the term has evolved to become a general prefix for a family of IBM management products, including utilities for trouble-ticket generation, network access control, and software distribution. (The delivery of NetView products that embody Management Services functions is an ongoing effort for IBM and third-party vendors.)

Like APPN, SNA distinguishes between management focal points and entry points. As shown in Figure 43, the focal point is supplied by NetView, or alternatively, Systems Center's Net/Master product. The entry point function manifests as software or firmware imbedded in network devices to provide direct monitoring and control. Non-SNA devices are integrated with NetView via a Service Point that translates management data into the NetView protocol format. Service Points are deployed on PCs, Unix workstations, and a variety of third-party platforms.

FIGURE 43. SNA MANAGEMENT SERVICES



NetView gets its SNA management data from host VTAM. FEPs and controllers maintain a permanent management session between their internal PU software and VTAM's SSCP. When the SSCP receives management data from downstream PUs, it is forwarded to NetView. NetView sends control commands downstream to PUs via the same SSCP-PU session. When a failure or abnormal network condition arises, devices can send unsolicited alerts to NetView via the SSCP.

Data on SSCP-PU sessions is formatted by a number of IBM management protocols, the foremost being Network Management Vector Transport (NMVT). The NMVT protocol has its own transmission header format and fields that carry alerts, statistics, trace data, response-time data, and related management information. The Service Point function converts third-party management flows into the NMVT format and vice versa.

NETVIEW'S FUTURE

Although NMVT via SSCP-PU is the principal protocol for SNA management today, IBM's stated direction is to move towards OSI CMIP protocols, LU6.2 transport, and HP OpenView interfaces. But with or without NMVT, NetView will continue to be the crown jewel of IBM management. A recent NetView version (2.2) can manage SNA, TCP/IP, Netware, and OSI networks (to varying degrees). LU6.2 connections to NetView have emerged and support for APPN is expected to solidify in 1993.

IBM and Novell have an agreement to improve NetView's ability to control NetWare servers. IBM and AT&T are developing connections between NetView and AT&T's WAN manager, the Accumaster Integrator. In spite of all this activity, NetView's strengths are mainly in the host realm. As computing moves increasingly away from hosts, the benefits of NetView become less apparent, particularly when compared to third-party object-oriented management products in UNIX and TCP/IP environments.

LAN NETWORK MANAGER

Born of the pressing need to manage growing LAN interconnect networks, IBM's LAN Network Manager (LNM) is an OS/2-based PC software product that uses proprietary protocols to monitor and control up to 255 LAN segments and attached bridges. (The LNM uses CMIP protocols to manage hubs and LAN stations, but its bridge protocol predates this effort.) To collect management data, LNM sets up LLC sessions with token ring bridges throughout the topology. Once a link is established, LNM can solicit status and statistics and activate or deactivate bridges, as well as change their parameters.

In operation, LNM can display statistics on the number of frames a specific bridge receives and transmits. Additionally, LNM can collect ring information from bridged token rings, including soft errors, interface beacons, ring status, discarded frames, ring data-rates, and so on. LAN administrators can define LAN performance and error levels that will generate alerts when maximum or minimum thresholds are crossed. LNM can list a ring's active LAN adapters and their status, and it can locate a LAN station by its adapter address. LNM uses the OS/2 Database Manager to track network configuration and statistics.

Until recently, there was only limited interaction between NetView and LNM. As of Version 2.2, NetView can fully control LNM from the host console. This means that any major function that LNM supports can be remotely executed from NetView. Alerts generated by LNM-monitored devices can be forwarded to NetView. Or, to limit the burden on the host, filters can be set allowing only certain types of alerts to be passed from LNM to the host.

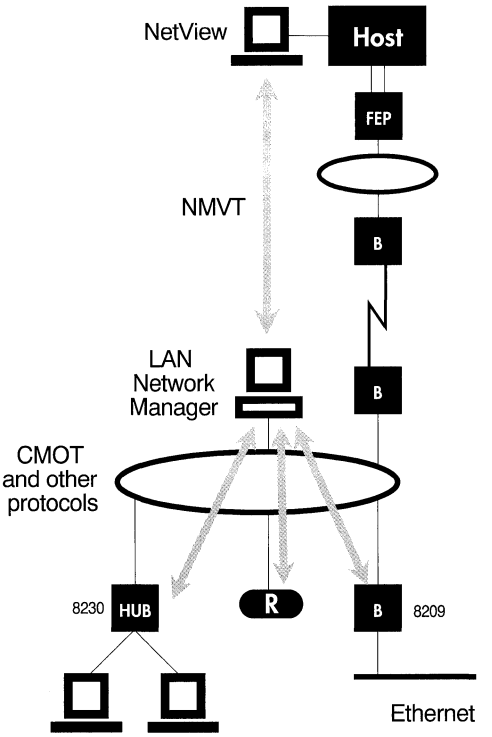
LNM typically uses the NetView/PC Service Point software to communicate with VTAM and NetView. OS/2 Communications Manager also supports this function. Figure 44 shows LNM managing devices such as an ISI router, PS/2 LAN station, 8209 token ring/ethernet bridge, 8230 wire hub, and IBM remote bridges.

ISI Solutions: Network Management

Because SNMP is the management protocol of choice for routers, full ISI support is critical. All of a router's management capabilities should be accessible from SNMP. In addition to its direct control of internetwork devices, SNMP management flows can be translated into other protocol formats. If a router can be fully configured and monitored with SNMP, it is highly likely that it can be managed by the native consoles of major systems and network vendors.

For instance, a number of products use SNMP to give host NetView access to routers, including AIX/NetView 6000, SunNet Manager, HP OpenView, and Brixton products. Increasingly, SNA sites will deploy SNMP consoles to

FIGURE 44. LAN NETWORK MANAGER AND NETVIEW



provide integrated management of multivendor routers, bridges, and other LAN-based gear. Also, IBM has committed to adding SNMP to LAN Network Manager. This will bring all IBM-compatible SRB bridges into the fold.

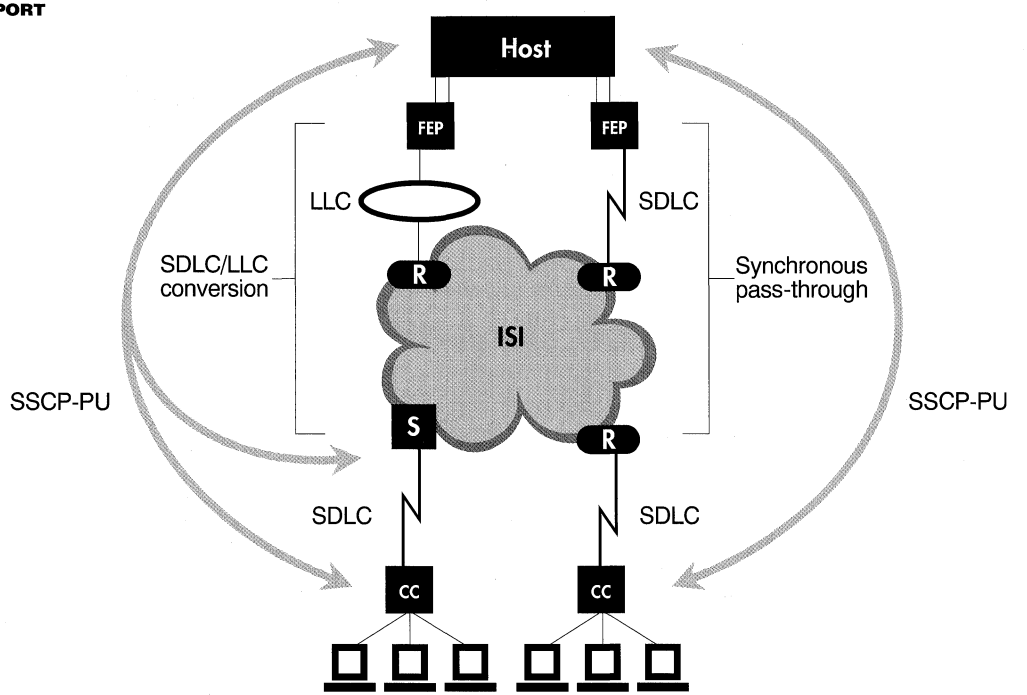
Because LNM does not yet support SNMP, ISI routers emulate LNM protocols so they can be managed by LNM as if they were native SRB bridges. This allows LAN administrators to monitor and diagnose ISI access networks with full visibility to bridges, routers, and LAN segments. Because LNM can be accessed by NetView, a router that emulates LNM can be managed directly by NetView, as long as a Service Point is present. (See Figure 44.) SNA system programmers can manage the SRB aspects of ISI routers with NetView and SNMP or with NetView and LNM.

NATIVE NMVT SUPPORT

The synchronous pass-through and SDLC/LLC conversion features of ISI are highly compatible with in-place SNA management tools (see Figure 45). Synchronous pass-through physically connects cluster controllers and other SDLC devices to ISI access routers, preserving pure SDLC on either side of the ISI. Consequently, SDLC link stations, PUs, and LUs are visible to NetView. Pass-through allows alerts and statistics to flow from downstream PUs to NetView; control messages flow in the reverse direction.

The SDLC/LLC conversion server also preserves NetView's visibility to SDLC devices. The PUs and LUs of devices attached to the SDLC server are all addressable by NetView. The SDLC server also contains its own PU that can interact with NetView via an SSCP-PU session. This PU generates

FIGURE 45. ISI NATIVE NETVIEW SUPPORT



S = SDLC Server

custom NMVT alerts that are passed to NetView to indicate anomalies and errors in SDLC lines and devices.

Support for NMVT protocols means that NetView operators can run commands to remotely control the SDLC server without leaving the data center. They can, for instance, use NetView DISPLAY and VARY commands to view and control the SDLC server's lines, interfaces, and software elements.

CONCLUSION

As evidenced by the preceding wealth of information, the ISI model provides an all-embracing approach to integrating and managing hybrid SNA/multiprotocol networks. With its high-speed backbone routers and intelligent access routers, the ISI gives SNA and token ring devices the benefits of a modern routing infrastructure without costly modifications to existing equipment populations.

Although a few elements of the ISI architecture will take a number of years to mature, this Guide has demonstrated that the technology exists today to realize many key benefits of the model. While new SNA internetworking products will increasingly come to market, not all will deliver high levels of availability, performance, WAN connectivity, IBM compatibility, and management. In many cases, low-end router products will not deliver adequate ISI access or backbone functionality, nor will they provide a good migration path to advanced multi-gigabit LAN/WAN internetworking. Consequently, an ISI requires careful planning and product selection to ensure that both short- and long- term SNA integration goals are achieved.

A

CONSIDERING WELLFLEET

A COMPANY PROFILE

MISSION AND FOCUS

Wellfleet Communications, Inc. was founded in June 1986 to design, manufacture, market, and support high-performance internetworking systems. Taking its name from the town on Cape Cod, Mass., where Guglielmo Marconi made the first transatlantic radio transmission in 1903, Wellfleet is focused on building large and complex enterprise internetworks that link multivendor, mixed-media LANs to other LANs within a building, throughout a campus, or across the globe.

Wellfleet, with its highly available, high-performance, multiprotocol router/bridge, was the first vendor to focus on enterprise internetworking. Early in its corporate history it identified five areas that, when building multiprotocol backbone networks, were critical to commercial organizations:

- Comprehensive Connectivity and Interoperability
- High Performance
- Reliability, Availability, and Maintainability
- Standards-based Solutions
- Commitment to Service

This agenda continues to set Wellfleet apart from others. Wellfleet's technological leadership in each of these areas has led to widespread acceptance of its products that is demonstrated by growth exceeding current market growth rates.

PRODUCTS

Wellfleet markets a family of scalable, highly flexible systems that are based on an advanced symmetric multiprocessor architecture specifically designed to ensure the highest levels of performance and availability, regardless of the number of network interfaces or protocols supported. The current family of six systems is offered in three performance/availability groups.

The Access Feeder Node (AFN) is a cost-effective system for small remote sites, with one ethernet or token ring interface and two synchronous interfaces in a single-board system.

Feeder, Link, and Concentrator Nodes utilize Motorola 68020/68030-based Advanced Communications Engine (ACE) processor modules and a 320 Mbps VMEbus processor interconnect. The versatile Feeder Node (FN) supports up to four LAN/WAN interfaces and offers forwarding performance to 14,500 pps. Next is the Link Node (LN), an expandable platform for small- to medium-size network sites. It supports up to four processor modules, 16 LAN/WAN interfaces and performance of 58,000 pps. The Concentrator Node (CN) is for large network sites. It supports 13 processor modules, 52 LAN/WAN interfaces and performance of 188,500 pps.

Backbone Link and Concentrator Nodes are designed to satisfy the throughput and/or availability requirements of the most demanding internetworks with FDDI, SMDS, and/or SNA. They utilize Motorola 68040-based Fast Routing Engine (FRE) processor modules and Wellfleet's Parallel Packet Express (PPX) processor interconnect with 1 Gbps bandwidth. The Backbone Node provides no single point of system failure with redundant processor interconnects, power supplies, and software image storage in combination with its symmetric multiprocessor architecture. The Backbone Link Node (BLN) supports four processor modules, 16 LAN /WAN interfaces (including four FDDI), and performance of 150,000 pps. The Backbone Concentrator Node (BCN) supports 13 processor modules, 52 LAN/WAN interfaces (including 13 FDDI), and system forwarding performance that scales to an industry-leading 480,000 pps.

Wellfleet router/bridges are designed to ensure interoperability based upon standards within multivendor environments. Every major LAN and WAN media and protocol is supported, including ethernet/802.3, 4 and 16 Mbps token ring/802.5, and FDDI. Complete routing and bridging support encompasses TCP/IP with OSPF, RIP, and EGP; DECnet Phase IV; Novell IPX; Xerox XNS; AppleTalk; Banyan VINES; NetBIOS; OSI with ES-IS and IS-IS; Transparent Bridge; Translation Bridge; Source Route Bridge; and SNA. PBX, video, and other D4/G.732-framed traffic are also supported. It all can be connected together in private, public, or hybrid wide area networks, supporting synchronous lines from 1200 bps to 45 Mbps, Fractional T1, T1/E1, T3/E3, X.25, Frame Relay, SMDS, Point-to-Point Protocol, and ATM.

Comprehensive node and network management are essential to successful mission-critical internetworks. Wellfleet fully supports SNMP — the industry standard for internetwork management. Every Wellfleet node contains a MIB II-compliant SNMP agent with numerous enterprise-specific extensions. Wellfleet supports all SNMP protocol data units (PDUs), including SNMP SET for dynamic configuration and control.

Wellfleet's nodes are managed efficiently from popular general-purpose SNMP management systems such as HP OpenView, SunNet Manager, IBM AIX NetView/6000, Cabletron Spectrum, and Digital DECmcc.

Wellfleet Site Manager is a platform-independent, SNMP-based application designed expressly for simplifying Wellfleet node management. It features an intuitive, windows-based, point-and-click user interface that hides the underlying complexity of SNMP. Site Manager offers central configuration management that simplifies network setup and expansion, real-time operations and monitoring, and real-time event and fault monitoring for efficient problem identification and isolation. Site Manager operates on popular computing platforms, including DOS-based PCs running MS Windows and Sun Microsystems' SPARCstations running the X Window System, OpenWindows, or Motif.

Because the technical requirements of internetworks are so critical, Wellfleet has staffed each of its offices with both sales representatives and network systems engineers. In this way, the company is providing more extensive technical resources locally in order to better understand user network environments.

Wellfleet is dedicated to providing strong customer service tailored to an organization's specific needs. Consulting services for internetwork design, installation, performance, and tuning; and education and training on internetwork concepts, operation, and maintenance are available. Complete maintenance service starts with 24-hour coverage, next-business-day shipment of spares, and a software subscription service for automatic access to the latest Wellfleet software and documentation. Technical support and diagnostic evaluation are provided through Wellfleet's Dial-In Diagnostic Center and Help Desk. Many options are available, including worldwide on-site service within four hours, available 24 hours a day, seven days a week.

Special support programs to meet the needs of multinational accounts and value-added resellers/integrators are also offered. The Multinational Account Partners (MAP) Program coordinates consistent service and support on a worldwide basis. The Wellfleet InterNetwork (WIN) Partners Program for value-added resellers/integrators offers many business-building and cost-saving services, including training, promotional assistance, maintenance support, and continuous timely information.

Wellfleet is dedicated to providing quality products that meet the demand for high-performance, standards-based networks both now and in the future. It is this dedication that ensures Wellfleet Communications will remain a strong and stable supplier.

B

Acronyms

ACTLU	Activate LU
ACTPU	Activate PU
ANR	Automatic Network Routing
APPI	Advanced Peer-to-Peer Internetworking
APPN	Advanced Peer-to-Peer Networking
CC	Cluster Controller
CMIP	Common Management Information Protocol
CMOL	CMIP over LLC
COS	Class of Service
CP	Control Point
DISC	Disconnect
DLS	Data Link Switching
ER	Explicit Route
FEP	Front End Processor
FID	Format Identification
HPR	High Performance Routing
I-frame	Information Frame
ISR	Intermediate Session Routing
LLC	Logical Link Control
LNM	LAN Network Manager

LSA	Link State Advertisements
LU	Logical Units
MAC	Media Access Control
MIB	Management Information Base
NAU	Network Accessible Units
NAU	Network Addressable Units
NCP	Network Control Program
NETID	Network ID (name)
NMVT	Network Management Vector Transport
NRZ	Non-Return to Zero
NRZI	Non-Return to Zero — Inverted
PU	Physical Units
RIF	Routing Information Field
RNR	Receiver Not Ready
RR	Receiver Ready
RTP	Rapid Transport Protocol
S-frame	Supervisory Frame
SABME	Set Asynchronous Balanced Mode Extended
SAP	Service Access Point
SDLC	Synchronous Data Link Control
SMDS	Switched Multi-Megabit Data Services
SNA	Systems Network Architecture
SNMP	Simple Network Management Protocol
SNRM	Set Normal Response Mode
SRB	Source Route Bridging
SSCP	System Services Control Point
SysGen	System Generation
TG	Transmission Group
TIC	Token-Ring Interface Coupler
TOS	Type of Service
U-frame	Unnumbered Frame
UA	Unnumbered Acknowledgment
VR	Virtual Route
VTAM	Virtual Telecommunications Access Method
XID	Exchange Station Identification

C

BOOKS AND REFERENCES

Black, Ulysses D.; *Computer Networks — Protocols, Standards, and Interfaces;* Prentice-Hall, Inc.

Black, Ulysses D.; *Data Communications and Distributed Networks;* Prentice-Hall, Inc.

Comer, Douglas E.; *Internetworking with TCP/IP — Volume I;* Prentice-Hall, Inc.

Dern, Daniel P.; *The Internet Guide for New Users;* McGraw-Hill

Cypser, R.J.; *Communications for Cooperating Systems — OSI, SNA, and TCP/IP;* Addison-Wesley Publishing

Guruge, Anura; *SNA Theory and Practice;* Pergamon Infotech Ltd.

Halsall, Fred; *Data Communications, Computer Networks and OSI;* Addison-Wesley Publishing

Haughdahl, J. Scott; *Inside Token Ring;* Architecture Technology Corp.

IEEE Computer Society; *Logical Link Control — IEEE Standards for Local-Area Networks;* John Wiley & Sons

IBM; *APPN Architecture and Product Implementations Tutorial;* IBM International Technical Support Center

Kapoor, Atul; *SNA—Architecture, Protocols, and Implementation;* McGraw-Hill, Inc.

Malamud, Carl; *DEC Networks and Architectures;* McGraw-Hill, Inc.

Martin, James; *SNA — IBM's Networking Solution;* Prentice-Hall, Inc.

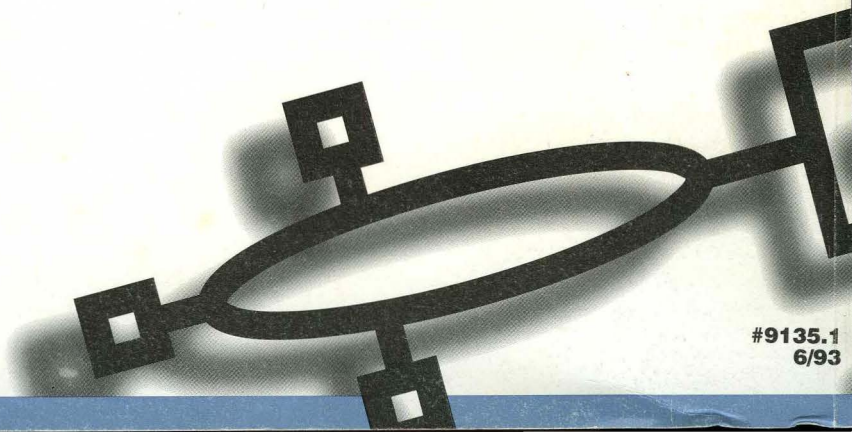
SNA Perspectives Newsletter; Communications Solutions Inc., San Jose, CA

INDEX

S

$19.95

Wellfleet Communications, Inc.
8 Federal Street, Billerica, MA 01821
(508) 670-8888 FAX: (508) 436-3658



#9135.1
6/93