15 DATA LINK AND PHYSICAL LAYERS

Below the network layer, the architectural uniformity of the OSI reference model partially breaks down. According to the OSI reference model, the role of the data link layer is to "provide for the control of the physical layer, to detect and where possible, correct errors which may occur in a physical connection established for bit transmission" (ISO/IEC 7498: 1993). There is, however, a small problem: at the data link layer, things get very technology-specific. All sorts of technologies are used to connect computers, and there is an astonishingly large number of standards for the data-link and physical layers that describe a trillion or so modem and electrical interface standards. Some have been around for decades. Consider the venerable high level data link control (HDLC). Having provided nearly two decades of wide-scale deployment and service over kilobit transmission facilities (the equally venerable V-series modems), HDLC is enjoying something of a rejuvenation in the form of frame relay services.

Later technologies—notably, local area networks—improved on earlier technologies by introducing shared-medium, multipoint communication; significantly higher bandwidth; low latency; and very low error rates. Metropolitan area network (MAN) technologies, the *asynchronous transfer mode* (ATM) platform for broadband integrated services, and the *synchronous optical network* (SONET) technology are expected to further expand the role of the data link layer, introducing constant bit rate and isochronous services (for voice and real-time video) and even greater (potentially gigabit) bandwidth.

This technology advance wreaked havoc on the poor souls responsible for providing service definitions for the data link and physical lay-

ers of OSI (ISO/IEC 8886: 1992; ISO/IEC 10022: 1990). Ultimately, the parties involved quite literally threw their hands up in exasperation and provided definitions that can be viewed as at best a "for completeness's sake" effort. Truthfully, this is nothing to lose sleep over: given how frequently technologies are introduced, the data link and physical layer service definitions offer a snapshot of what these layers looked like at the time of standardization, but they absolutely should not be interpreted as the final word on what the data link and physical layers should be.

Some technologies—for example, point-to-point and HDLC-based links—fit conveniently into the OSI reference model's notion of a data link. Advocates of local area networks, however, opined that LANs were richer in function than mere data links—they had globally unique addressing, and it could even be argued that Medium Access Control included routing functions. For a brief time, LANs led a truly schizophrenic existence: eventually, the "it's a subnetwork technology" folks were beaten into submission by the "it's a data link" folks, and LANs were placed at the data link layer, along with point-to-point protocols like ISO/IEC 7776, *high level data link control*, and CCITT Recommendation X.21*bis*, *Use on Public Data Networks of Data Terminal Equipment Designed for Interfacing to Synchronous V-series Modems*. Metropolitan area networks—fiber distributed data interface (FDDI) and the distributed queue dual bus (DQDB)—will undoubtedly share the same fate as LANs, and since politics will most certainly play a role in *broadband integrated services digital networks* (BISDN), it is inevitable that one or more broadband services will contend for roles in the network layer.

Not to worry: it will be possible to run IP and CLNP over *all of them*.

Taxonomy of Data Link Standards

Consistent with Tanenbaum's (1988) taxonomy, data link and physical layer standards for OSI generally fall into two categories:

1. *Point-to-point connection standards* describe the use of HDLC procedures (ISO/IEC 3309: 1991; ISO/IEC 4335: 1991; ISO/IEC 7809: 1991) as a means of framing data for transmission over various physical media in single- and multilink configurations (ISO/IEC 7478: 1984; ISO/IEC 7776: 1986).[1]
2. *Multi access channel standards* describe logical link control proce-

1. There are more data link and physical layer standards that apply to OSI, to be

dures (ISO/IEC 8802-2: 1990) as well as physical access methods and medium specifications for local and metropolitan area networks: *carrier sense multiple access with collision detection* (CSMA/CD; ISO/IEC 8802-3: 1992), *token-passing bus* (ISO/IEC 8802-4: 1990), and *token ring* (ISO/IEC 8802-5: 1990). *Fiber distributed data interface* (ISO/IEC 9314-2: 1989) and *distributed queue dual bus* (IEEE 802.6 1990) MAN complete the list.

These categories apply equally well to the Internet architecture. Although the nomenclature differs slighty—Internetters prefer the term *network interfaces* over *data link layer*—the Internet community wisely elected to treat anything and everything that IP could conceivably be run over as a network interface.

Point-to-Point Connection Standards

HDLC-based protocols remain the most common form of link-level framing for point-to-point connection technologies; connection-oriented and datagram transmission (a.k.a. *unnumbered information* frames, or UI) are described among the HDLC classes of procedures. Both perform error detection using a 16- or 32-bit cyclic redundancy check.

Point-to-point subnetworks play an important role in both TCP/IP and OSI network connectivity. The standard for encapsulation of Internet datagrams over point-to-point links is RFC 1331, the *point-to-point protocol* (PPP). RFC 1331 describes a convention for encapsulating network-layer protocols in full-duplex, asynchronous or synchronous links, using HDLC framing. The default 8-octet frame format defined in the PPP (Figure 15.1) accommodates a maximum frame size of 1,500 octets and specifies a 16-bit *frame-check sequence*. PPP uses an address-extension mechanism available in HDLC to specify a protocol field, which is used to identify network-layer protocols in a multiprotocol (OSI, TCP/IP, AppleTalk®, XNS/IPX, etc.) environment. PPP also describes a *link control protocol* that can be used, for example, to negotiate maximum receive frame size, to indicate that authentication must be performed using the specified authentication protocol, and to indicate to a peer that the link quality is to be monitored using the specified protocol. Finally, PPP provides a framework for the development of a set of *net-*

sure. Folts (1991) devotes entire volumes of a multivolume compendium of OSI standards to the CCITT V-series and X-series recommendations. Halsall (1988) provides a college-text-level primer on these layers. And of course, there is always Tanenbaum (1988).

PPP Field	No. of Octets
Flag (01111110)	1
Address	1
Control	1
Protocol ID	2
User data	<Variable, up to 1,500 octets>
Frame-check sequence	2
Flag (01111110)	1

FIGURE 15.1 Point-to-Point Protocol Frame

work control protocols (NCPs), which deal with the behavior and idiosyncracies of individual network-layer protocols operating over point-to-point links. (At the time of this book's publication, network control protocols for OSI, DECnet, AppleTalk®, IP, and IPX were available only as Internet drafts and RFCs in preparation and thus did not yet have any official standards status.)

Systems using PPP establish and negotiate the characteristics of the link and specify the set of network-layer protocols that will share the link using configuration packets of the link control protocol, then proceed through authentication (if specified), and finally process the network-centricities of the protocols that will share the link using the appropriate network control protocols (if specified). The OSI standards do not describe multiprotocol issues; thus, it should be expected that PPP will eventually be used in many multiprotocol environments that include IP and CLNP.

The corresponding description of encapsulation of CLNP into point-to-point subnetworks was originally drafted as an addendum to ISO/IEC 8473 and was then incorporated into ISO/IEC 8473 prior to its publication as an international standard. The standard only describes the minimum maximum service data unit size required and assumes that network layer protocol identification is sufficient to distinguish one network layer protocol from another; i.e., it relies on some other (data link) demultiplexing mechanism to distinguish protocols that have as their initial octet an initial protocol identifier (see "Network Layer Protocol Identification," Chapter 13) from those that do not.

Multiaccess Channel Standards

Several local area network technologies—the Ethernet/IEEE 802.3 CSMA/CD MAC, the IEEE 802.4 token-bus MAC, and the IEEE 802.5 token-ring MAC—have been around for better than a decade: descriptions abound, and there is little left to say that has not already been said. One aspect of the IEEE 802/ISO 8802 LAN architecture that should be considered within the context of OSI and TCP/IP, however, is IEEE 802.2/ISO 8802-2, *logical link control* (LLC).

Logical Link Control—The Nether-Layer

A perturbation introduced in the IEEE 802 LAN architecture, perpetuated by OSI, and eventually finding its way into TCP/IP, logical link control is primarily used to demultiplex higher-layer protocols (well, to placate frame zealots, there *are* type-1 connectionless and type-2 connection-oriented LLCs, and even a type-3 acknowledged connectionless LLC, but thankfully, OSI and TCP/IP both use type 1). To indicate that the protocol encapsulated in the LLC information field is an OSI network layer protocol, the originator of an LLC frame sets the destination and source service access points (DSAP, SSAP; see Figure 15.2) to the hexadecimal value FE. The receiver examines the first octet of the information field (the initial protocol identifier; see Chapter 13) to distinguish which of the many OSI network-layer protocols this might be.

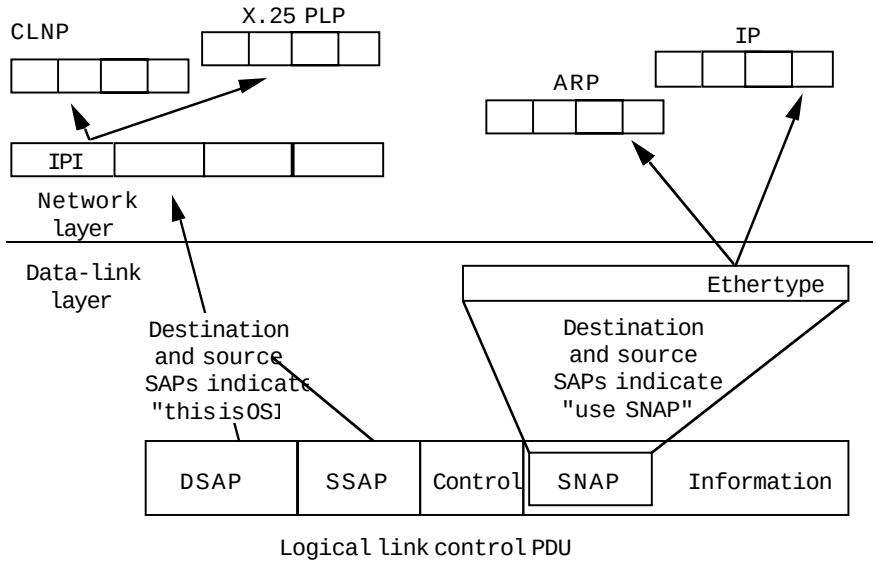


FIGURE 15.2 LLC and Protocol Identification

When Ethernet was first developed, the last 2 octets of the MAC frame (the ethertype field) were used to identify higher-layer protocols. When the technology was introduced into the IEEE 802 committee so that it might become a standard LAN technology, the ethertype field of the MAC frame was abandoned in favor of a MAC frame-length field (ostensibly, the ethertype field was specific to Ethernet; given that the IEEE 802 committee was attempting to standardize multiple MAC technologies, it seemed more appropriate to have a MAC-independent means of identifying higher-layer protocols—i.e., logical link control). The dual role of these octets would cause considerable difficulty if stations with IEEE 802.3 CSMA/CD and Ethernet MACs were to attempt to communicate if not for the fact that the values of the field are nonoverlapping: the maximum value of the length field in the IEEE 802.3 MAC frame is slightly more than 1,500 octets, and the ethertype values are, conveniently, integers greater than this value. Most implementations now recognize both techniques. Even though interoperability is accommodated, a rift was formed between the Ethernet and IEEE 802 communities that persists today.

Initially, IEEE intended that values other than hexadecimal FE would be assigned to deserving organizations; it didn't take long, however, to realize that a single octet (255 values) would not suffice to identify all organizations requiring protocol/organization identifiers. Rather than expand the length of the LLC SAP fields, the IEEE 802.1 committee adopted what is called the *subnetwork access protocol* (SNAP). The SNAP is itself identified by populating the destination and source service access point fields with the hexadecimal value AA. This value indicates that an additional 5 octets are appended to the 3-octet LLC type-1 frame to convey a 3-octet *protocol identifier/organization identifier* field and a 2-octet organization-specific field. The existence of the SNAP doesn't come close to smoothing the feathers ruffled during the "ethertype versus IEEE 802.3 frame length" debates; however, SNAP does have two redeeming virtues:

1. For IP encapsulation in IEEE 802.x LANs (RFC 1042, 1988), switched multimegabit data service (RFC 1209, 1991), and fiber distributed data interface (FDDI) (RFC 1188), the protocol identifier/organization identifier field is set to 0, and the *ethertype* field is conveyed in the remaining 2 bytes.
2. The addition of 5 octets to the 3-octet LLC type-1 field word-aligns the header of the encapsulated network protocols for both 16- and 32-bit machines.

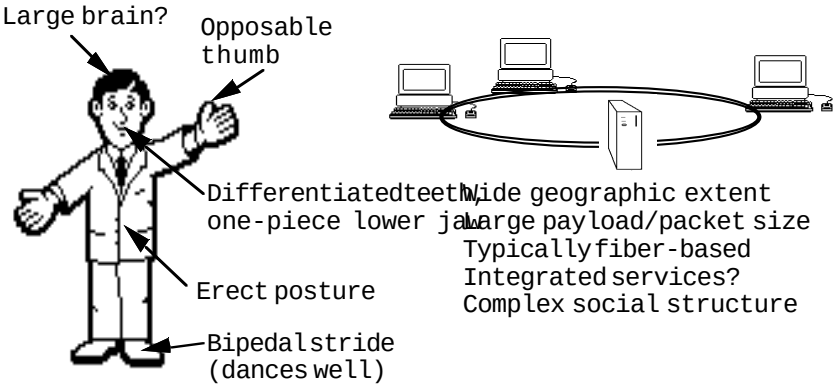
Emerging Digital Technologies

Several new technologies—in some cases, applications of these technologies—have emerged only recently, among these the FDDI and DQDB metropolitan area networks, the HDLC-based frame-relay technology, and broadband ISDN. These technologies are discussed here because they are expected to provide much-needed high-bandwidth/low-latency platforms for Internet and OSI-based applications over a wide geographic extent. A related physical layer technology, synchronous optical network (SONET), is also discussed.

Metropolitan Area Networks: FDDI and IEEE 802.6 DQDB

Metropolitan area network technologies extend the characteristics of LANs that provide a favorable environment for distributed computing—high bandwidth, low latency, large information payloads—beyond the distance constraints of existing LAN technologies. MAN services are likely to be extended over very wide areas and hence encompass more than what intuitively comes to mind when the term *metropolitan* is applied.

Another characteristic that continues to be associated with MANs is the ability to integrate services—particularly real-time voice and data—over the same physical transmission facilities. Also, with the virtually boundless promise of bandwidth offered by fiber-optic transmission systems, real-time video will inevitably enter the picture. Services other than real-time broadcast television and video conferencing—image transfer and recognition systems, collaborative work and education, virtual reality—are more likely to play a prominent role once their potential as ubiquitous applications is realized. In many respects, the term *MAN* is apropos; like *homo sapiens*, a MAN is a multifaceted beast:



Fiber-Distributed Data Interface

FDDI is a very-high-bandwidth (100 Mbps) LAN technology. It is typically deployed as a dual ring, with one ring enabling transmission "clockwise," the other "counterclockwise" (see Figure 15.3).

Stations may be attached to one or both rings, or via a concentrator to allow "trees" to be branched off a central dual ring (see Figure 15.4). Attachment to both rings is a reliability consideration:[2] when a link be-

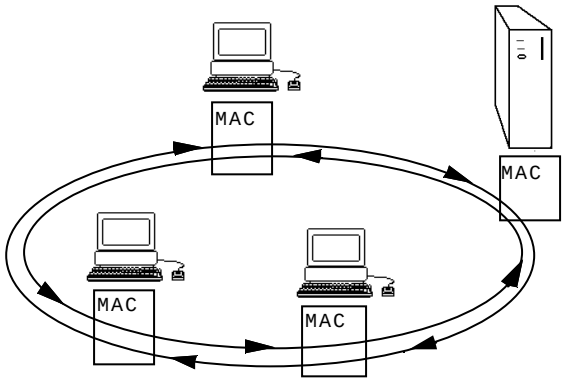


FIGURE 15.3 FDDI Deployed as a Dual Ring

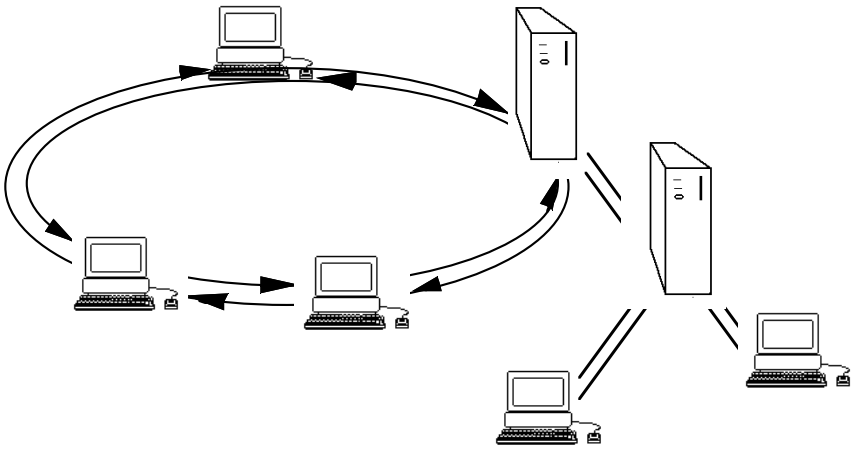


FIGURE 15.4 FDDI Trunk Ring and Tree

2. Contrary to claims made by an overzealous (albeit well-intentioned) sales force during an early FDDI marketing presentation, dual attachment does not double the available bandwidth.

tween two stations fails, dual-attached stations can reverse their direction of transmission and take advantage of the healthy ring; if a station fails, stations adjacent to the failed station can reconfigure (wrap) the dual ring so that the single logical ring survives (see Figure 15.5).

Like the IEEE 802.5 token ring LAN, the FDDI MAC protocol uses a token-ring medium access method. A station must acquire a MAC control frame called a *token* before it can transmit data. Once a station has transmitted a MAC data frame, it generates and writes a new token to the ring from which it acquired the token. Each station reads every frame off the ring and checks to see whether the destination address in the MAC frame is the same as its respective MAC address. If the address is not a match, a station repeats the frame (downstream); if it is a match, the station copies the frame off the medium and passes it up to the logical link control function for protocol demultiplexing, etc.

Each 4 bits of binary data in a MAC data frame are encoded into a 5-bit pattern called a *symbol* prior to submission to the physical layer. A 32-bit *cyclic redundancy check* (CRC) is computed over the frame-control, information, and CRC fields to detect bit errors on the data symbols. Control frames (e.g., the token) are also encoded as a (separate) sequence of symbols (see Figure 15.6).

Priorities are implemented by restricting a station's ability to acquire the token. A station is responsible for determining when a frame it has generated has traversed the entire ring (i.e., it has returned to its origin) and for *stripping* these frames; a *frame status* encoded in the MAC

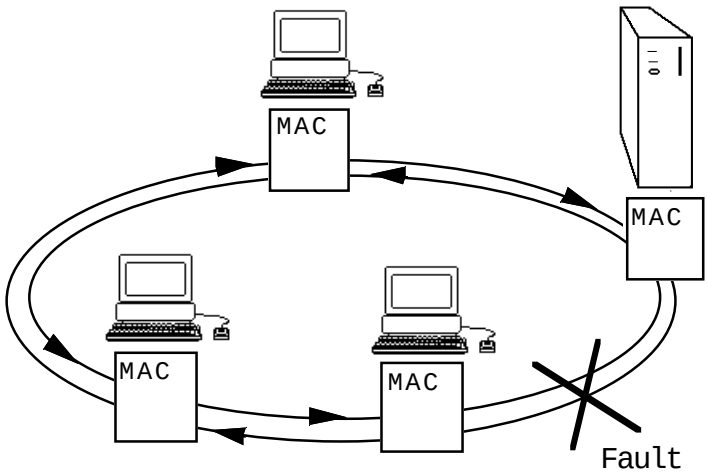


FIGURE 15.5 FDDI Ring Wrap

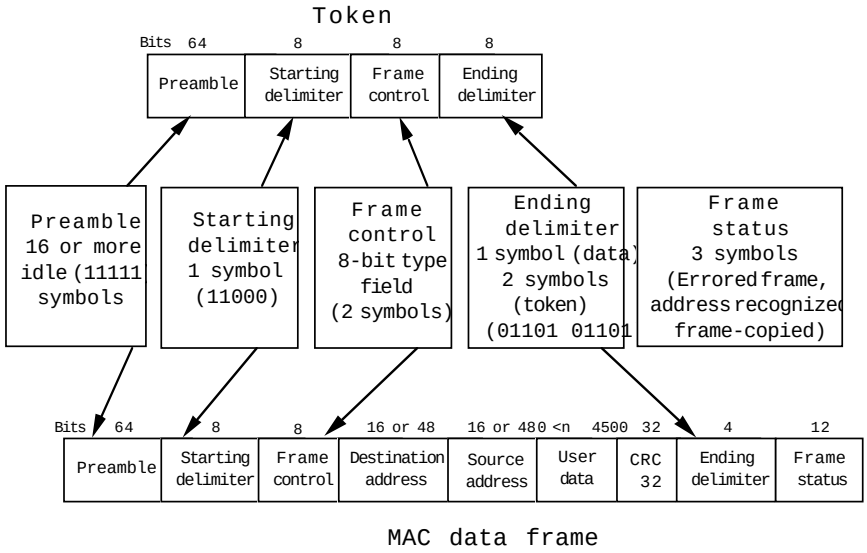


FIGURE 15.6 FDDI MAC Frame Formats

frame header indicates whether the transmission has been successful.

The FDDI MAC frame can accommodate 4,500 octets of user data. Typically, 48-bit (Ethernet) addressing is used (see IEEE 802.1; ISO/IEC 8802-2); 16-bit addressing may also be used. A *timed token-rotation* (TTR) *protocol* is used by stations to establish a uniform token-rotation time. The token-rotation time can be set to a large value to allow very high ring utilization under heavy load (e.g., many stations distributed over a large ring perimeter), or it can be set to a small value to guarantee bandwidth for delay-sensitive applications such as packetized voice or video.

FDDI was originally designed to operate over multimode optical fiber. This limited the distance between stations to 2 kilometers. FDDI has since been extended to operate over single-mode fiber-optic cable; with single-mode fiber, a maximum perimeter of 60–100 kilometers can be achieved for the dual ring. The specification of a *single-mode fiber physical-layer medium-dependent* (SMF-PMD) and an FDDI-to-SONET physical-layer mapping (FDDI-SPM) function is significant, since single-mode fiber is used by telecommunications carriers and can be leased and used between facilities to extend individual links between FDDI stations over metropolitan area networking distances.

IP encapsulation over FDDI is described in RFC 1188; like encapsulation of IP in IEEE 802 subnetworks, RFC 1188 prescribes the use of the LLC/SNAP, described earlier in this chapter, and both ARP and routing

protocols are operated over FDDI networks in the same manner as IEEE 802 LANs. Encapsulation of OSI CNLP in FDDI frames is not the subject of a standard but in practice is accommodated by using LLC type 1 in the fashion described earlier in this chapter for IEEE 802.3 and IEEE 802.5 subnetworks.

Distributed Queue Dual Bus

The *distributed queue dual bus subnetwork* (IEEE 802.6 1990) is a high-speed MAN technology that operates over two unidirectional, contraflowing buses over a variety of transmission rates. The MAN can be deployed in two topologies (see Figure 15.7):

1. In an *open* topology, stations providing a *head-of-bus* function generate fixed-length data *slots* and management control information over both buses. The slots are used to compose variable-length MAC frames.

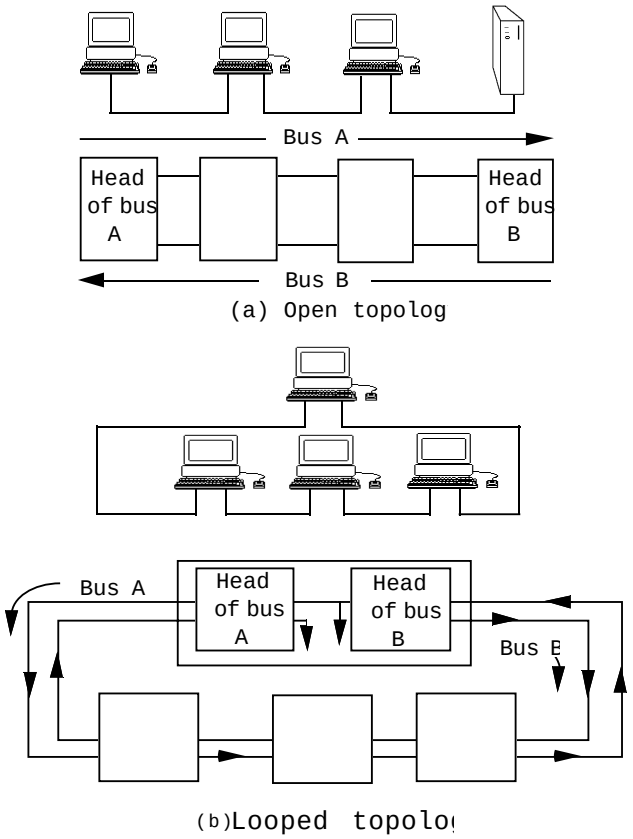


FIGURE 15.7 Topologies of a DQDB MAN

2. In a *looped* topology, the heads of bus are colocated in a single station, and the topology can be configured to be self-healing in the face of failure.

The DQDB MAN provides:

- A connectionless data-transmission service
- A connection-oriented, isochronous service, suitable for voice and video applications
- A connection-oriented, nonisochronous service alternative for data communications

The functional model of a DQDB node to support these services is depicted in Figure 15.8. Both the connection-oriented and connectionless MAC services submit data through a segmentation and reassembly (SAR) function to "atomize"[3] variable-length frames into fixed-length *slots* consisting of an access-control octet, a 4-octet segment header, and a 48-octet segment payload (for the connectionless data service, 4 octets are taken from the segment payload for "adaptation" functions; see Figure

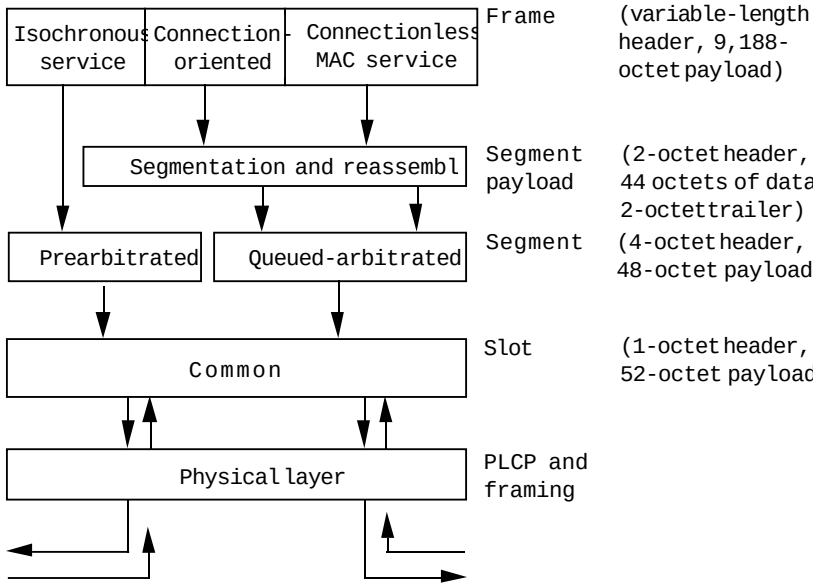


FIGURE 15.8 DQDB Functional Model

3. Both the DQDB MAN and broadband ISDN use fixed-length micro-packets (slots, cells). The term "atomize" is pejorative, derived from "ATM-ize," or "to make into ATM-like cells" (origin unknown).

15.9). Standards for call signaling and bandwidth management for isochronous services have yet to be developed.

In the *queued-arbitrated* mode of operation, MAN stations can send and receive variable-length, connectionless MAC frames (up to 9,188 octets). The *prearbitrated* mode of operation, still under study in IEEE 802.6, provides support for digitized voice and other isochronous services. The DQDB connectionless MAC frame permits the use of 16-, 48-, and 60-bit MAC addresses. Addresses are encoded in fixed-length, 8-octet source and destination address fields; the most significant "nibble" of the address fields identifies the address type, followed by 60 bits of padding and addressing (44 bits of padding precede a 16-bit address, 12 bits of padding precede a 48-bit address, and 60-bit addresses require no padding). An optional 32-bit *cyclic redundancy check* is selected by setting the *CRC indicator bit* in the *initial MAC protocol header* (see Figure 15.9). The *beginning-end tags, buffer allocation size*, and *length* fields of the initial MAC protocol header and trailer are used by the sending and receiving stations for error detection and control functions (see Figure 15.9).

IMPDUs are segmented into 44-octet segment payloads, and each payload has a 2-octet header and trailer (see Figure 15.10). These 4 octets are aligned with the broadband ISDN ATM adaptation layer, type 4 (see "Asynchronous Transfer Mode and Broadband ISDN," later in this chapter). The segment header contains a segment type indicator—beginning of message (BOM), continuation of message or "middle" (COM), and end of message (EOM)—and a single-segment message type indicator (SSM). A sequence number is used to detect lost, misordered, or inserted segments.[4] A *message identifier* (MID) assists in the identification of seg-

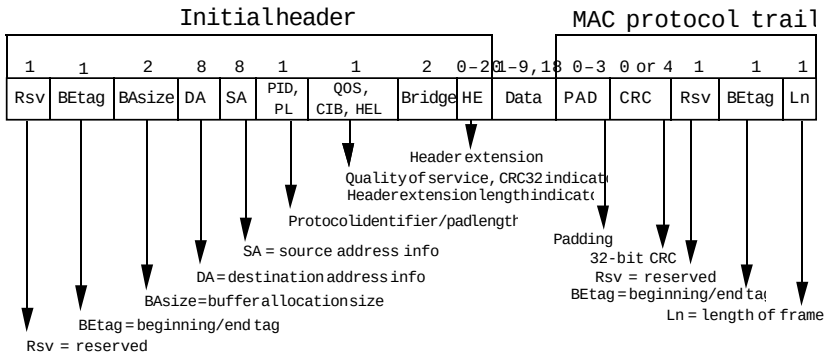


FIGURE 15.9 DQDB Initial MAC Protocol Data Unit (IMPDU)

4. The usefulness of performing cell sequence checking, particularly over dual-bus

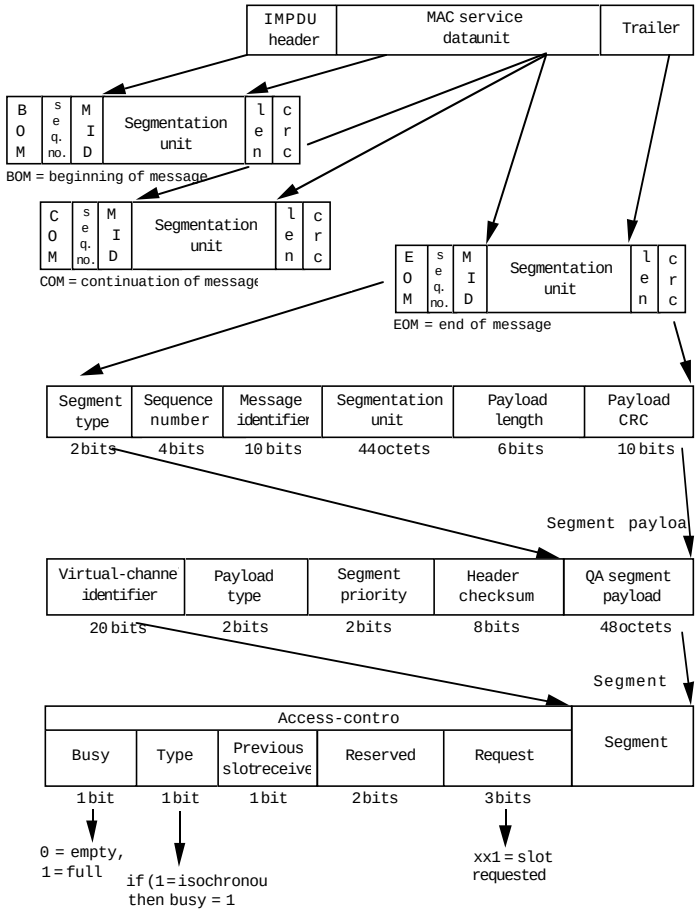


FIGURE 15.10 DQDB Segmentation, Slot Decomposition

ments of the same IMPDU during reassembly. The payload-length field contains either the value 44 (BOM or COM) or the number of octets in an EOM or SSM, and the *payload cyclic redundancy check* is optionally used for single-bit error detection on the segment payload.[5]

topology, is questionable, especially since the sequence space is too small to uniquely identify all segments of an IMPDU that is segmented into 16 or more slots. This bit twiddling is another fine example of last-minute standards chicanery: rather than reopen the question of whether the length and composition of the ATM cell/DQDB slot was correct, the IEEE 802 and CCITT folks decided it would be easier and subtler to whittle away 4 octets from the segment payload for adaptation-layer functions.

5. The DQDB MAC protocol stack is replete with checksums of one sort or another (see Figures 15.9 and 15.10). During the joint development of the ATM cell/DQDB slot by CCITT and IEEE 802.6, the network-centrics viewed the slot as the unit of significance for

Connectionless MAC stations operate in what is called *queued-arbitrated* (QA) mode. Heads of bus A and B continuously generate/forward 53-octet queued-arbitrated slots on their respective buses, composed of a 1-octet *access control field* (ACF) and a 52-octet payload (see Figure 15.10). The operation of the distributed queue access protocol is predicated on the values of two fields of the ACF: the *busy* bit indicates whether the slot is empty (available) or full (taken). The *request* bits indicate whether slots have been queued for access (i.e., that a station has requested a slot). Each station maintains two *request counters*, one for each bus (logically, one for each direction of information flow). A station in the idle state (i.e., one with nothing to send) decrements the request counter for each empty slot it sees on bus A and increments the same request counter by 1 for each request bit set on bus B; i.e., it constantly checks how many slots have been queued on bus A. In so doing, a station establishes its position in the distributed queue, relative to other stations wishing to send on bus A (a similar computation occurs simultaneously for bus B).

When a station wishes to send on bus A, it writes a request on bus A; i.e., it writes a binary 1 to the request bit in the access-control field of the first slot that has not already been used by downstream stations to request slots. The station copies the current value of the request counter from bus B into a *countdown counter*, then decrements the countdown counter each time an empty slot passes by on bus A; i.e., a station waits until all the stations that have queued requests to send on bus A ahead of it have satisfied their requests (sort of like children in a classroom raising their hands and waiting their turn). When the countdown counter reaches 0, the station copies the next empty slot from bus A, fills the payload with user data, marks the slot busy, and writes it back to bus A. The process is the same if a station wishes to send a slot on bus B; here, the station queues a request on bus B, copies the current value of the request counter from bus A into the countdown counter, and waits until all the stations that have queued requests to send on bus B ahead of it have satisfied their requests.

The DQDB MAN MAC protocol was developed in a physical-transmission-facilities-independent fashion; i.e., the MAN can operate over a variety of physical transmission systems, including those currently used

---

all broadband services and insisted that error detection and recovery be performed on a cell-by-cell basis. Host-centrics pointed to the IEEE 802.1a LAN architecture standard and claimed that to be consistent with *all* the MACs that preceded IEEE 802.6, error detection—in the form of a 32-bit CRC—should be performed on the initial MAC protocol data unit. Who won? Everybody . . . sort of. There are *optional* error-detection functions for the entire IMPDU and optional error-correction functions for the slot header.

by telephony providers in the United States, throughout Europe, and Asia, as well as the emerging SONET transmission system, described later in this chapter. Borrowing (perhaps subliminally) from the principles of convergence described in the *Internal Organization of the Network Layer* (see Chapter 13), *physical layer convergence procedures* (PLCPs) have been defined to describe how to map DQDB slots onto physical-layer framing provided by a variety of standard telephony transmission systems (Brandwein, Cox, and Dahl 1990). PLCPs enable the DQDB MAN MAC to operate over existing telephony digital transmission hierarchies at rates of 1.544 Mbps (DS1) and 44.736 Mbps (DS3) in North America and at rates of 2.048 Mbps (E1), 34.368 Mbps (E3), and 139.264 Mbps (E4) in Europe and Asia, as well as over SONET (discussed later in this chapter). This, combined with the fact that the length and encoding of DQDB segments is intentionally aligned with the ATM "cell" of the broadband ISDN under study in CCITT, makes IEEE 802.6 attractive for a public network service. Two such services are currently under trial and early deployment in the United States and several European countries: switched multimegabit data service (SMDS) and the European Telecommunications Standards Institute's MAN project (ETSI MAN).

Fast Packet Services and Technologies

The term "fast packet" has emerged from the telecommunications industry as a way to collectively refer to transmission technologies that may be used in wide area network as well as local area network and campus configurations. These technologies may be operated over existing telecommunications digital transmission facilities to provide public data networking services that offer higher bandwidth and lower delay characteristics than their "narrowband" predecessors, ISDN and X.25. In the following sections we describe the most prominent of these emerging "broadband" services—SMDS and its European cousin, ETSI MAN, frame relay, and ATM—and a relatively new and promising fiber-optic transport system, SONET.

Switched Multimegabit Data Service

SMDS is a public, packet-switched datagram service. The service is often described as "LAN-analogous" (Bellcore Technical Requirement TR-TSV-000772 1991), meaning that the features of the public service—high bandwidth, low delay, large packet sizes, multicasting, address screening/filtering—emulate characteristics of LANs. The interface protocol to the public network (SMDS interface protocol, or SIP) is based on the IEEE

802.6 MAN. The SMDS interface protocol is described as a three-level protocol exchanged across the *subscriber network interface* (SNI) between a switching system within a public carrier network—in the United States, a local exchange carrier, an independent exchange carrier, or an interexchange ("long-distance") carrier network—and the customer communications equipment (in Figure 15.11, the router and host B). In the United States, an open DQDB topology will be applied; i.e., separate physical transmission facilities will be dedicated to each subscriber site.

The SIP level-3 packet corresponds to the IEEE 802.6 initial MAC frame. Most of the IEEE 802.6 protocol control information is interpreted exactly as defined in the IEEE standard. SMDS uses 60-bit publicly administered addresses. The most significant 4 bits of the destination address field are used to indicate whether the 60-bit address is an individual or a group address (a group address is functionally similar to a multicast Ethernet address); the remaining bits of the source and destination address fields are used to convey the SMDS address. The address format is ten BCD-encoded decimal digits, imitating the format used for telephone numbers in the United States. In hosts, bridges, or routers attached to the dual bus, these addresses are used the same way that 48-bit Ethernet addresses are used for communication over LANs; in the public

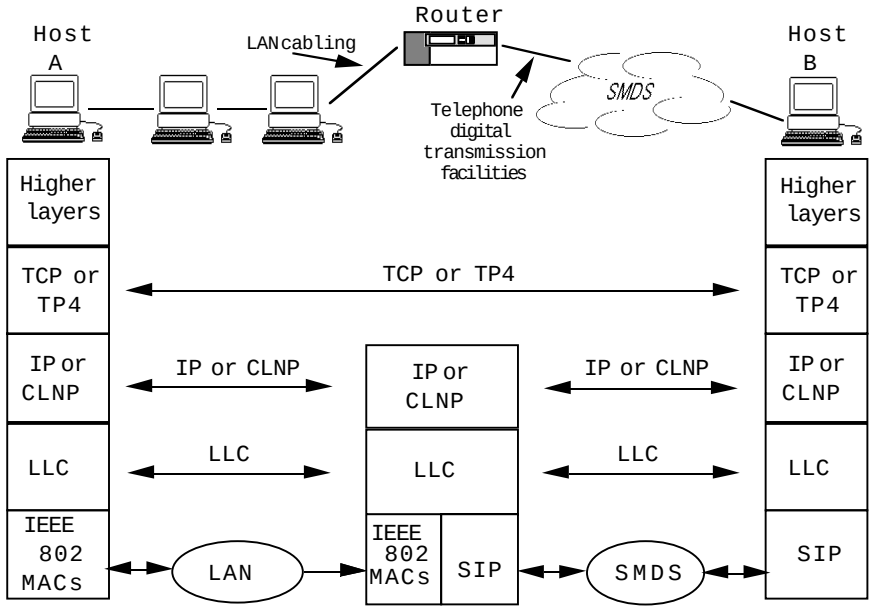


FIGURE 15.11 SMDS

network, however, the addresses are used for routing between switches. The header extension field provides a means to support carrier selection for interexchange access (to long-distance carriers).

SMDS provides several supplementary services: *source address screening* allows subscribers to filter traffic originating from unwanted sources on the public network, and *destination address screening* allows subscribers to limit the destinations to which traffic may be forwarded across the public network. *Group addressing* allows subscribers to create the equivalent of a multicast capability across the public network. Essentially, a subscriber identifies a set of SMDS addresses that should all receive copies of a group-addressed packet, and the SMDS network provider assigns the subscriber a 60-bit group address. Thereafter, any SMDS packet that is submitted to the SMDS network having that group address as its destination address will be copied to all members of the group address list submitted by the subscriber.

The SIP level-2 protocol data unit corresponds to the IEEE 802.6 DQDB slot. SIP level 2 provides framing and segmentation for the variable-length SIP level-3 packets through the use of fixed-length slots and also provides an error-detection capability through a 10-bit payload CRC in the level 2 trailer (see Figure 15.12).

SMDS will be deployed over existing digital transmission facilities of the public telephone network over a metropolitan or wide area by telecommunications providers. SIP level 1 describes a physical interface

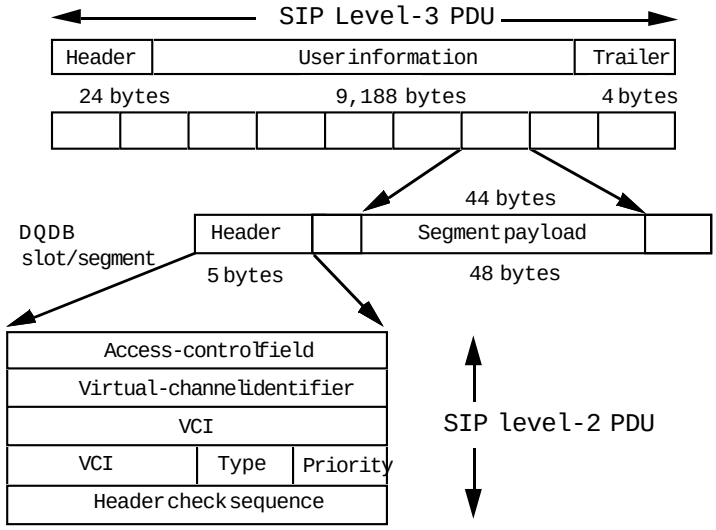


FIGURE 15.12 SIP Level-2 PDU

to the digital transmission network. Currently, two access rates are defined. A PLCP is defined for DS3, which provides a transfer rate of 44.736 Mbps, and for DS1, which provides a transfer rate of 1.544 Mbps. The DS1 physical interface uses the *extended superframe* format (ESF). SMDS is described in a series of Bellcore Technical Requirements (TR-TSV-000772 1991; TR-TSV-000773 1991).

There will be a number of subscribers who may want to benefit from the low latency offered by the DS3 access for bursty data applications but whose sustained information rates are expected to be much lower than 45 megabits per second. *Access classes* allow customers to subscribe to what are essentially "fractions" of bandwidth for both bursty and sustained data transmission. The mechanism is straightforward. A customer selects an access class (separate access class tariffs are expected for sustained information rates of approximately 4, 10, 16, and 25 Mbps, accompanied by maximum burst rates of approximately 10, 13, 18, and 34 Mbps, respectively). A *credit manager* monitors the rate at which SIP level-2 protocol data units are sent to the network. If the rate at which the subscriber sends SIP level-2 protocol data units into the network exceeds either the subscribed-to *sustained information rate* or the *maximum burst rate,* the network provider will discard the excess (see Bellcore Technical Requirement TR-TSV-000772 [1991] for more details on this "leaky bucket" or credit manager algorithm).

SMDS will be used primarily for local area network interconnection. The unusual (unprecedented) combination of local area network characteristics and telephony-style numbering in SMDS makes the service suitable for a number of interenterprise (multiorganizational) as well as intraenterprise applications (this is touted as "any-to-any communication"). Early trial and field experience has demonstrated that SMDS can be used as a "gathering net" for commercial IP networks (it may also be used as part of the backbone network). In this scenario, the commercial IP provider subscribes to "fat" DS3 SMDS pipes and recommends SMDS access at DS1 rates to its customers. The DS3 SMDS access provides an aggregation point for the IP provider and reduces the number of interfaces the IP provider must manage and pay for.

SMDS can also be used by disaster recovery service providers (people who provide data center redundancy—i.e., exact duplicates of an organization's data-processing environment and database) to simplify and reduce the cost of cutover procedures following a catastrophic failure. In this scenario, the disaster recovery service provider recommends SMDS for wide area service; if an earthquake destroys a company's data center, cut-over to the duplicate data center is a matter of "entering a different

telephone number" in routing tables across the company's internetwork. This is considerably simpler and faster than reprovisioning private lines and virtual-circuit services and cheaper than subscribing to additional private lines to the disaster recovery service provider.

SMDS makes it convenient and economical for organizations to have many temporary communications partners (commercial IP providers do this as well, but incredible as it may seem, there are folks who do not use them). For example, medical centers may routinely distribute X-ray, magnetic resonance, and positron tomography images to affiliated medical schools and attending physicians, but on occasion, they may wish to distribute them to specialist and other medical centers. In these circumstances, SMDS may offer an attractive alternative to constructing and operating a private network.

To accommodate those enterprises that want SMDS but cannot justify the cost of DS1 access to all locations, *frame-based access* at 56 and 64 Kbps—in many instances, on frame-relay interfaces—will be offered by local and interexchange carriers. In these configurations, the SMDS level-3 packet is encapsulated in an HDLC frame rather than segmented (at such low rates, the packetization into IEEE 802.6 slots is impractical). SMDS will eventually be offered over SONET interfaces (discussed later in this chapter) and, as broadband switching is introduced into public networks, over ATM as well.

ETSI MAN

Like SMDS in the United States, the ETSI MAN responds to the European consumer demand for multimegabit communication services. An enabling vehicle for broadband ISDN, ETSI MAN is slightly more ambitious than SMDS, encompassing

- A "packet-oriented" MAC service of ISO 8802 LANs, suitable for bursty data communications
- A connection-oriented isochronous service, suitable for voice and video applications
- A connection-oriented, nonisochronous service alternative for data communications

Like SMDS, ETSI MAN applies the DQDB MAN MAC protocol for the connectionless MAC service. Because the regulatory environment in many parts of Europe differs from that in the United States, single- or multisubscriber access facilities may be provided, and both open and looped topologies are anticipated. Two forms of interface are described in the ETSI MAN architecture ("ETSI Metropolitan Area Network" 1991). The *user MAN interface* is similar to SMDS; the public network provides a

physical transmission facility with access rates of 2.048, 34.368, and 139.264 Mbps (i.e., over E1, E3, and E4 transmission facilities), and the subscriber attaches a host, bridge, or router. Network providers may also offer encapsulation, bridging, or routing via a *user service interface* at access rates of 4, 10, or 16 Mbps; in such configurations, subscribers attach their IEEE 802.x LANs directly into network equipment (see Figure 15.13).

Many SMDS features are present in the ETSI MAN: 60-bit, telephony-style addressing and supplementary services (e.g., address screening). Since both SMDS and ETSI MAN are early broadband ISDN applications, compatibility between the service offerings is both desirable and essential.

RFC 1209 describes IP encapsulation and operation of the address resolution protocol (ARP) over SMDS. IP datagrams are encapsulated in the 8-octet LLC/SNAP frame in exactly the same manner as for IEEE 802 LANs. ARP is performed over what are called "logical IP subnetworks"; essentially, a set of hosts whose IP addresses share a common IP network/subnet number and whose SMDS addresses are all identified as recipients of the same group address send ARP requests using the group address in the same manner as they would use a 48-bit broadcast address over an Ethernet or IEEE 802 LAN. ARP replies are directed back to the source using the 60-bit address of the ARP request originator.

RFC 1209 is suitable for IP over ETSI MAN as well; only the aspects of group addressing for the purposes of using ARP require additional consideration. Encapsulation of CNLP (ISO/IEC 8473-1: 1993) is not the subject of a standard but, in practice, is again accommodated by using LLC 1 in the fashion described for IEEE 802.3 and IEEE 802.5 subnetworks in ISO/IEC 8802-3.

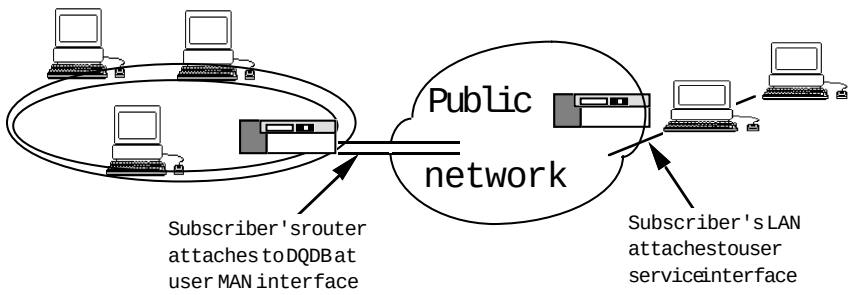


FIGURE 15.13 ETSI MAN, UMI, and USI

Frame Relay Technology

By OSI standards, frame relay is a hybrid. It offers aspects of connection-oriented services that hint of the X.25 packet-level protocol—for example, frame relay service is provided over permanent and switched virtual circuits—but forgoes the protocol and processing overhead of the level-3 error handling of X.25. Since the X.25 reliability mechanisms are omitted, frame relay offers higher throughput than X.25 and is less expensive than X.25 and equivalent private-line solutions.

From a technology standpoint, frame relay is nothing new: it is a juiced-up or "fast-packet" technology based on HDLC. It can be deployed as a shared, common-access WAN in a public or private networking environment. From a services perspective—applying nomenclature from the architecture of the *integrated services digital network*, CCITT Recommendation I.122-(1989)—frame relay is a packet-mode bearer service. The access protocol is based on HDLC and on the link access protocol developed for signaling over the D channel of narrowband ISDN (LAP-D; CCITT Recommendation Q.921 1989). Variable-length HDLC frames (up to 1,600 octets) are packet-switched on the basis of a 10-bit *data link connection identifier* (DLCI; see Figure 15.14). Although the data link connection identifier can be locally or globally administered, a 10-bit DLCI is already viewed as a major shortcoming of the protocol; an address extension indicator in both octets of the header can be used in later versions of frame relay to extend the header to 3 or 4 bytes. A *frame check sequence* (FCS) is computed on the entire frame prior to its submission to the frame relay network.

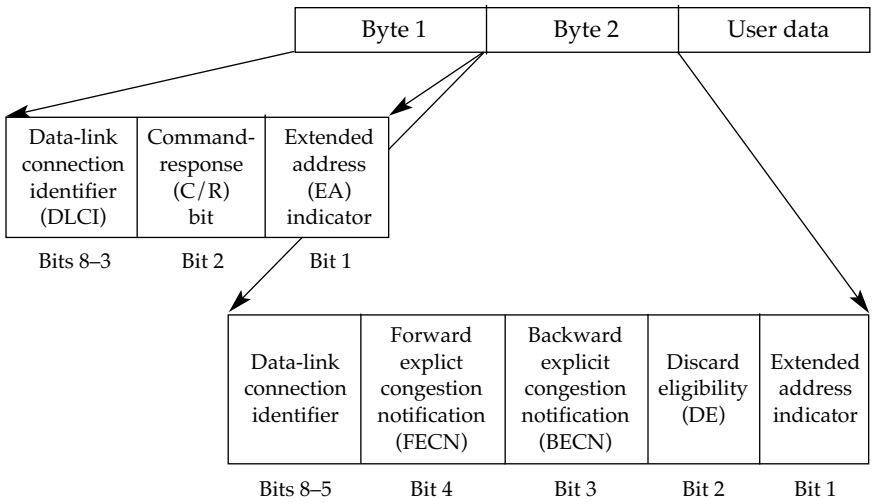


FIGURE 15.14 Frame Relay Header

Network (switch) processing of incoming frames is as follows. If the frame check sequence check is successful, the data link connection identifier is used as an index into a table of trunks over which frames are forwarded (relayed) to (toward) their destinations. If the FCS check fails, the frame is discarded. At the ingress point, the DLCI encoded in the incoming frame indicates the destination; however, the DLCI is modified by the network prior to transmission across the egress link so that the DLCI in the outgoing frame contains the source (the switch thus recomputes or adjusts the FCS as part of the processing of each frame).

Internal to frame relay networks, frame relay providers apply the concepts of *statistical multiplexing:* every user is not expected to use all of the bandwidth subscribed to every minute of every day, so resources (switching equipment, trunk lines between switches) are allocated to handle what is determined to be a reasonable load under normal conditions. As with highways, anticipated traffic loads can be exceeded, and congestion can occur. In a frame relay network, congestion is alleviated by discarding frames. Frame relay offers *explicit congestion notification* (ECN) mechanisms that are similar to that of OSI CLNP (see Chapter 13), but bidirectionally: forward and backward indications—FECNs and BECNs —are defined. Computers attached to frame relay circuits are thus notified that the network is discarding packets.

Frame relay provides another mechanism to maintain fairness among multiple senders. A *discard eligibility* indicator can be set by a sender or the network to indicate that certain frames are sacrificial lambs; i.e., if frames must be discarded, throw away the frames that have the DE flag set.

In theory, the DE bit may be used to guarantee some level of service (especially, throughput) to individual data link connections; i.e., the DE bit might be set in all frames of a given data link connection, but not set in the frames sent over a second data link connection, effectively assigning a higher priority of service to the users of the second data link connection. Under an arrangement such as this, an IP or CLNP network administrator can bias much of the *committed information rate* negotiated between the subscriber and network provider toward the (privileged) users of one data link connection. (Committed information rate is a subscription time parameter used by public frame relay providers that is an estimate of the user's "normal" throughput needs and a measure of the amount of data the public network promises to deliver.)

The "features" of congestion notification and discard eligibility are often promoted imprudently in frame relay marketing. As

an indicator in an end-to-end datagram network-layer protocol like CLNP, a congestion-experienced notification can be conveyed to the transport layer, where mechanisms exist to reduce the rate of traffic introduced into a network by a source end system. Beyond the obvious marketing value of being able to say, "We notify you of congestion," the usefulness of these indicators in a data link protocol is questionable. Unless additional mechanisms are introduced in the network layer protocol operating over a frame relay service to propagate congestion-status information of the frame relay link back to all sources—for example, mapping the value of a BECN/FECN received across a frame relay link into the congestion-experienced bit of the CLNP header—the notification is an annoying no-op. (Well, not quite; routers that have SNMP agents can now count the number of FECNs and BECNs received, so one at least knows whether the frame relay network is overworked). And unless transport protocol entities operating at traffic sources "behave" and actually slow down the rate at which they introduce traffic, FECNs and BECNs have absolutely no effect on the rate of packet flow into the frame relay network. Finally, in multiprotocol environments (OSI, TCP/IP, etc.), relying on single-bit congestion notification can be dangerous: unless all *network layer protocols propagate congestion notices, and* all *end system transport implementations react to this information by changing their rates of transmission uniformly, it is quite possible that those end systems that behave well will be penalized (they will slow down) and those that ignore the information will benefit (since others have slowed down, they may be able to continue at their present rate without penalty).*

It's easy to sum up the "truthful" aspects of advertisement regarding discard eligibility: discard eligibility relies on cooperation on the part of all senders, all of whom will responsibly identify which of their data aren't important.

Both CNLP and IP implementations will treat frame relay as a point-to-point network. RFC 1294 describes IP encapsulation and operation over frame relay as part of a multiprotocol identification and encapsulation scheme. Each protocol that is to be transmitted across a frame relay circuit is identified using a 1-octet network layer protocol identifier (see Chapter 13), which must be the first octet encoded in the frame relay frame. Values assigned by ISO and CCITT distinguish among the OSI network layer protocols, and code points in the NLPID are also used to identify the encapsulated protocol as IP and the IEEE 802.1 SNAP. The NLPID value for SNAP is used for bridging purposes and also to indicate that the ethertype value is to be used to distinguish among the protocols that do not have an NLPID assigned to them[6] (see *Assigned Numbers*, RFC

6. It seems that history is forever repeating itself. Standards makers, failing to heed the

1340). Encapsulation of CNLP over frame relay is not the subject of a standard, but in multiprotocol environments, it is expected that RFC 1294 will be applied. RFC 1294 also prescribes maximum packet sizes for frame relay networks and describes a means of negotiating a maximum frame size, a retransmission timer, and a maximum number of outstanding information frames, using HDLC exchange identification (XID) frames. Some frame relay networks (the switches, actually) support a maximum frame size of only 262 octets; since this is much smaller than the default maximum IP segment size of 576 octets, RFC 1294 defines a "convergence protocol" that must be used to segment IP packets before they are forwarded across such networks.

Address resolution over frame relay networks is a matter of associating a data link connection identifier with the IP address of a station on "the other side" of a permanent virtual circuit. The *inverse address resolution protocol* (INARP; RFC 1293) proceeds as follows. An IP station performs address resolution by (1) constructing an ARP request packet, (2) encapsulating the packet in a frame relay packet, and (3) sending it directly to the target DLCI (note that there is no notion of broadcast or group addressing of ARP requests over frame relay). The INARP request packet is essentially the same format as other ARP requests: the source IP address is provided, and the target IP address field is 0-filled; however, in the case of frame relay, the requesting station inserts DLCIs rather than conventional 48-bit LAN addresses in the source and target hardware address fields. Upon receiving an INARP request, the called, or target, station will typically place the requester's {IP address, DLCI} mapping into its ARP cache. The target station then composes an INARP reply using the source DLCI and IP addresses from the INARP request as the target addresses for the INARP reply and using its own DLCI and IP addresses to populate the source hardware and network protocol addresses, respectively. When the requesting IP station receives the INARP reply, it uses the {IP address, DLCI} address information from the INARP reply to complete its ARP cache.

Like SMDS, frame relay will be used primarily for local area network interconnection. It is a direct and cheap replacement for private lines. When switched virtual circuits are offered, frame relay will provide the virtual-circuit equivalent of any-to-any connectivity; until then, it is best suited for intraenterprise topologies.

---

lessons from the service access point debacle of the logical link control, also made the OSI NLPID a single-octet field and may again find themselves with insufficient code points to identify all the protocols that might be identified using the NLPID demultiplexing scheme.

Frame relay allows companies to supplant existing private lines with statistically multiplexed services at a lower cost per access line. This is attractive because it gives subscribers some interesting cost-performance trade-offs:

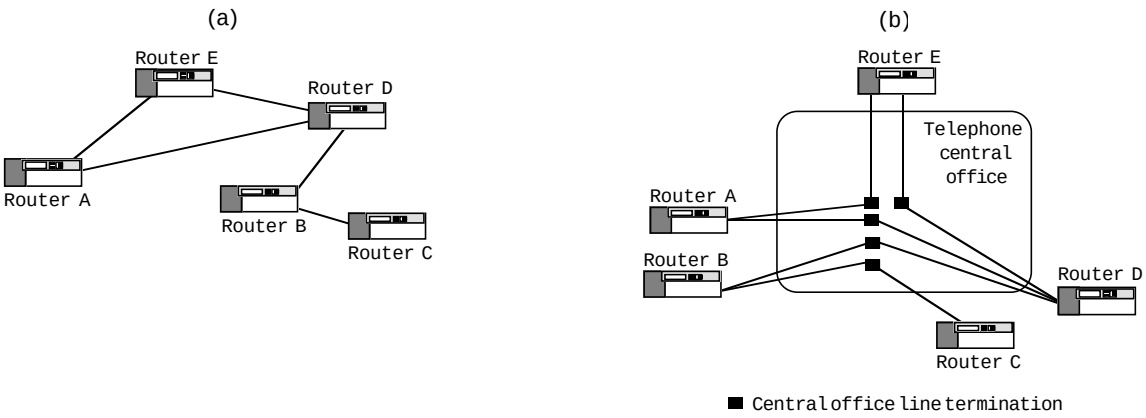
- Replace the existing private-line topology with corresponding frame relay permanent virtual circuits and reduce monthly communications costs. In this scenario, a network administrator would replace the private topology with permanent virtual circuits forming an identical virtual topology.
- Replace existing private lines with frame relay permanent virtual circuits and use the savings to enrich the topology by adding permanent virtual circuits. Frame relay tariffs often have an access-connection component, an information-transfer (switching) component, and per permanent virtual circuit charges. The permanent virtual circuit charges are relatively low, so some or all of the savings accrued by converting to a switched service could be used to add permanent virtual circuits between routers, especially between routers where no private lines existed. They would improve overall network performance by eliminating one hop.
- Replace existing private lines with frame relay permanent virtual circuits and use the savings to enrich the topology by subscribing to higher-rate access connections. The reduction in monthly communications costs may be significant enough over a given topology to justify the use of a DS1 facility (or fraction thereof) rather than a 56 Kbps or 64 Kbps access connection.

Like SMDS, frame relay services are likely to be offered over increasingly higher-bandwidth access connections. If there is a market pull for it, 45 Mbps frame relay access will undoubtedly be offered.

There are currently two schools of thought regarding the evolution or migration of frame relay services to a broadband ISDN environment. School 1 recommends that the relatively straightforward process of mapping data from one virtual circuit onto another be used. Here, an interworking unit—a data link service bridge—will maintain a logical relationship between frame relay data link connection identifiers and broadband ISDN virtual circuit identifiers and will forward encapsulated data (e.g., IP or CLNP packets) between the two networks. School 2 recommends that the frame relay bearer service operate on the asynchronous transfer mode platform that supports all broadband ISDN services (see "Asynchronous Transfer Mode and Broadband ISDN," later in this chapter). Here, frame relay frames operate over asynchronous transfer mode and even-

tually SONET facilities as a native rather than an interworked service.

Why are fast packet services like SMDS, ETSI MAN, and frame relay more economical than private lines? The answer usually provided by packet-switching pundits is that they offer individual subscribers the economies of a switched environment, in which the resources are shared among the entire population of subscribers rather than dedicated to each individual subscriber. Historically, this sharing took place more often at what network-centrics call "the customer's premises"; private networks use routers to reduce the number of dedicated links they must pay for, handcrafting topologies that offer end users acceptable service while lowering the enterprise's monthly telecommunications costs (see Figure 15.15[a]). Fast packet services attempt to move more of the successful and profitable form of switching back into the telephone "central office" and apply the same logic. Providing a private line between two computers requires that the network provider dedicate and maintain a physical transmission facility from one customer site to the central office, then out again to a second customer site, for each private line; thus, computers that act as packet switches (routers) have multiple line terminations (see Figure 15.15[b]). With a fast packet service, a public service provider must dedicate only one link to each router—from the subscriber's site to the fast packet switch inside its central office (this link is often called a "tail"), thereby saving the cost of maintenance and provisioning for the "other tail" (see Figure 15.15[c]). Since maintenance of transmission facilities is a very large part of the cost of offering a telecommunications service, carriers can propose attractive rates to those who set tariffs. (Note that if the subscriber actually needs the aggregate bandwidth offered by multiple links, a higher access class [committed information rate] or higher-bandwidth transmission facility may be used instead.)



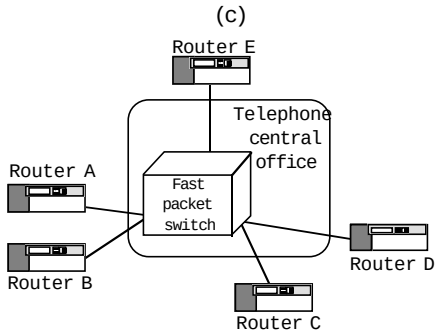


FIGURE 15.15 Anatomy of Fast Packet Services

Asynchronous Transfer Mode and Broadband ISDN

The DQDB MAN technology offers isochronous and queued-arbitrated services over the same physical medium; broadband ISDN promises this and more. Both DQDB MAN and BISDN accomplish this multiservice integration by using "cell" technology. In broadband ISDN, the chunks are called ATM cells (CCITT Recommendation I.361 1989); like DQDB slots, they are composed of a 5-octet header and a 48-octet payload. In a broadband integrated services environment, bandwidth can be flexibly allocated to support several forms of communications simultaneously, even when different amounts of bandwidth are required for each form. The network provides a steady stream of ATM cells, and stations attached to the network (multimedia stations) acquire ATM cells—on an as-needed basis, one at a time—to transmit voice, video, or data. From the available bandwidth, a station may acquire cells at fixed intervals to support voice calls and video (typically, 64 Kbps is required for a single voice call, whereas NTSC-quality video requires between 30 and 45 Mbps, and high-definition television may require anywhere from 140 to 500 Mbps) and on an as-needed basis to support a bursty datagram service like SMDS or a circuit-switched data service like frame relay.

Synchronous Transfer Mode and Asynchronous Transfer Mode To understand the current popularity of ATM, it is useful to compare it with its predecessor, synchronous transfer mode (STM, also known as synchronous time division multiplexing). Synchronous transfer mode networks divide the bandwidth of a transmission line into units of time rather than cells. Every connection *C* that is multiplexed over an STM link is given one or more fixed time slots *S* for transmission over the link. At any instant in time, there can be at most *C* times *S* slots used; thus, a large value of *S* for a connection *C* assigns a large fraction of the total bandwidth of the STM link to *C* and correspondingly reduces the total

number of connections that may share the link with *C*. Conversely, large values of *C* reduce the fraction of bandwidth that may be assigned to any given connection. If a single time slot represents 64 Kbps, for example, and 1 connection on a DS1 link is assigned 6 slots, then the bandwidth available for that connection is 384 KBps. The remaining bandwidth—approximately 1.1 megabits—can be assigned to other connections (there could be up to 17 connections having a bandwidth of 64 Kbps or 3 more connections having a bandwidth of 384 Kbps, etc.).

What's wrong with this scheme? Once time slots are assigned to a connection, the bandwidth represented by those time slots cannot be used to support other connections even if the connection is idle. This scheme is too inflexible if the transfer mode is to support a variety of applications (real-time data, voice and video, "bursty" data applications, etc.) that demonstrate different peak and average rates of transmission over a common medium; too much bandwidth must be reserved to accommodate the peaks, especially when the very high peak-to-average transmission ratio is considered.

Since synchronous transfer mode fell short of the expectations, folks turned to packet-switching techniques as a means of accommodating a wider range of transmission rates and unpredictable traffic loads. *Fixed-length* packets were selected so that constant bit rate channels could be emulated using a packet-switching technique. Small packets were selected to minimize delay. The resulting "microgram" is illustrated in Figure 15.16.

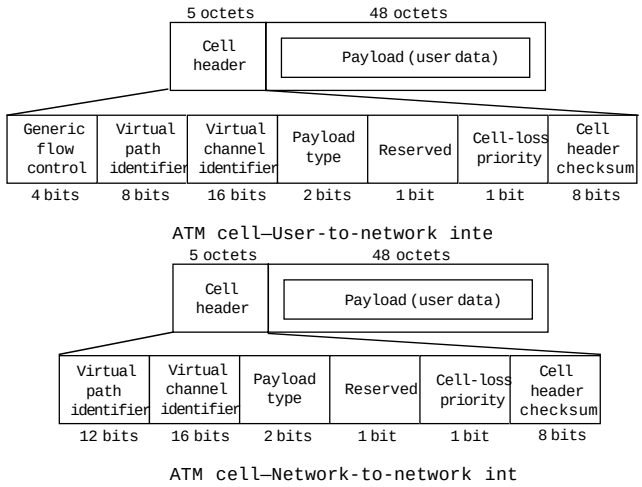


FIGURE 15.16 ATM Cell Structure

ATM Cell Structure Actually, there are two forms of the microgram: the cell submitted by and delivered to customer communications equipment (i.e., across the *user-to-network interface [UNI]* and the cell that is transferred betwen ATM networks (i.e., across a *network-to-network interface [NNI]*). The initial field of the ATM user-to-network interface cell is a *generic flow indicator* (4 bits), which will be used for end-to-end flow control between ATM stations. A *virtual-path identifier* (8 bits) identifies the path between a source and destination ATM station pair, and a *virtual-channel identifier* (16 bits) defines a connection within the virtual path. A *payload type* (3 bits) identifies the contents of the payload (currently only one value, 0, for user information) and provides congestion notification. A *cell loss priority* (1 bit) is like the discard eligibility bit in frame relay. An 8-bit *cell header checksum* can be used to detect and in some cases correct errors in the cell header. The ATM network-to-network interface cell is identical in all but one respect to the ATM user-to-network interface cell: the UNI generic flow indicator is not present in the NNI cell and the virtual-path identifier is extended an additional 4 bits.

To accommodate voice and perhaps real-time video without echo effects, the current wisdom is that fixed (or predictable) delay between cells/slots must be maintained. After considerable debate over packetization delay and packet length, CCITT, T1, and IEEE 802.6 converged on a chunk/cell/slot length of 53 octets (5 for the header, 48 for the payload). The number 48 *is notable because it is a fine example of the standards process in action. There is little technological substance in the decision to use 48 octets; it is a compromise between the 64-octet payload originally specified in the DQDB MAN standard and the 32-octet payload preferred by CCITT and T1. "I like 32, you like 64. Let's meet somewhere in the middle . . ."*

ATM Adaptation Layers Although they will be multiplexed/interleaved over the same cell/slot fabric, voice, video, and data services will use fixed-length cells in different ways; for example, a data service like SMDS could be implemented over a broadband ISDN platform and requires an IMPDU-to-cell mapping and SAR functions similar to those provided in the IEEE 802.6 architecture. The functions and protocols required to provide these services to the MAC layer (and similar services for voice and video) are placed at what is called the BISDN *ATM adaptation layer* (AAL; CCITT Recommendation I.362-1989), and there are several (CCITT Recommendation I.363-1989): AAL1 is provided for voice, AAL2 for video, AAL3 for mapping frame relay onto a cell fabric, and AAL4 is nearly identical to the IEEE 802.6 DQDB (see Figures 15.8 and 15.10). A

recent addition to the ATM adaptation layer famly, AAL type 5, is the encapsulate of choice for IP. This adaptation layer provides IP packet delimiting and a 32-bit CRC error check, effectively providing the same framing as a local area network MAC frame (see Clapp [1992] for a complete description of ATM adaptation layers). An RFC in preparation will describe multprotocol encapsulation over ATM, including CLNP.

ATM is likely to appear in several offerings (see Figure 15.17). Speculating a bit, the authors suggest that these offerings are likely to appear in the following chronological order:

1. *Local ATM, or ATM LANs*: In this configuration, an ATM switch acts as a hub for communication between workstations in ATM work groups. The switch will provide dedicated bandwidth to each station, initially in permanent virtual circuit arrangements and later, when signaling is available, in a switched environment.
2. *Campus ATM*: Here, ATM switches compete against FDDI and other MAN technologies. The campus ATM network is shared by dedicated-bandwidth users and users who wish to interconnect local area networks over high-bandwidth, low-delay networks.
3. *Seamless ATM*: Here, ATM hubs interconnect campuses using private lines and provide ATM services across a wide area topology.
4. *Broadband ISDN and cell-relay services*: Public network providers offer permanent virtual circuit and switched cell-relay services. Subscribers will have access to constant and variable bit rate services, distinguished at subscription or call-setup time by interpacket arrival guarantees and cell loss probability.

ATM is most likely to succeed if (1) there is a LAN-based ATM "presence" in the workplace and an established need for very high bandwidth over the wide area; (2) the wide area cost is reasonable; (3) wide area ATM service is a ubiquitous and uniform offering (hard lessons learned from narrowband ISDN); and (4) there is a graceful migration from early broadband services like frame relay and SMDS to broadband ISDN. ATM is a seductive technology, but it is also very much a fledgling technology—a packet-switching technology having nearly all the complexities and baggage associated with scaling as IP and CLNP. To succeed, ATM will have to live up to some very ambitious market expectations (see especially Nolle [1993]).

Synchronous Optical Network (SONET)

SONET (ANSI T1. 105A-1990) is a fiber-optic transport system. The system combines a basic signal rate of 51.840 Mbps (synchronous transport signal—level 1, or STS-1) with a byte-interleaved multiplexing scheme to

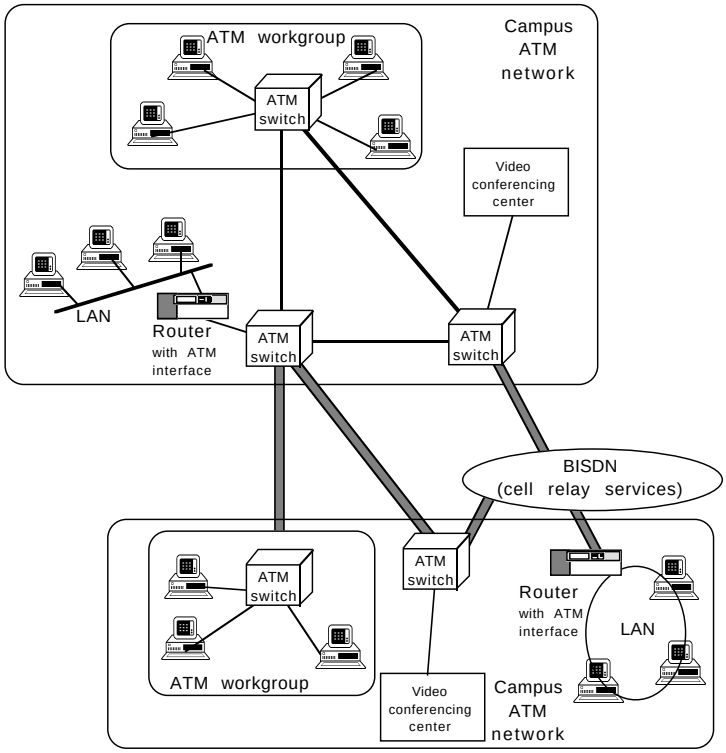


Figure 15.17 ATM Deployment

create a hierarchy of higher-rate signals, with each higher-rate signal providing an integer multiple of the STS-1 signal. The convention for denoting these higher rates is STS-*N*, where $N = \{1, 3, 9, 12, 18, 24, 36, 48\}$; with transmission rates *beginning at* 51+ Mbps and defined for up to 2.5 Gbps, the SONET physical layer will play an important role in the much-heralded gigabit networking strategies for the 1990s.

STS-1 frames (810 bytes, 6,480 bits) are transmitted at a rate of 8,000 frames per second (1 every 125 microseconds). *Transport overhead* bytes (those that deal with framing, synchronization, and the multiplexing of the STS-1 signal) and *payload overhead* bytes (those that deal with framing, synchronization, and the multiplexing of services) account for 36 bytes, leaving a payload of 774 bytes (6,192 bits) per frame, or 49.536 Mbps (see Figure 15.18). This is sufficient to convey an existing DS3 signal and will minimize the impact of this new technology on embedded digital transmission facilities.

STS-*N* signals are formed by interleaving bytes of *N* STS-1 signals.

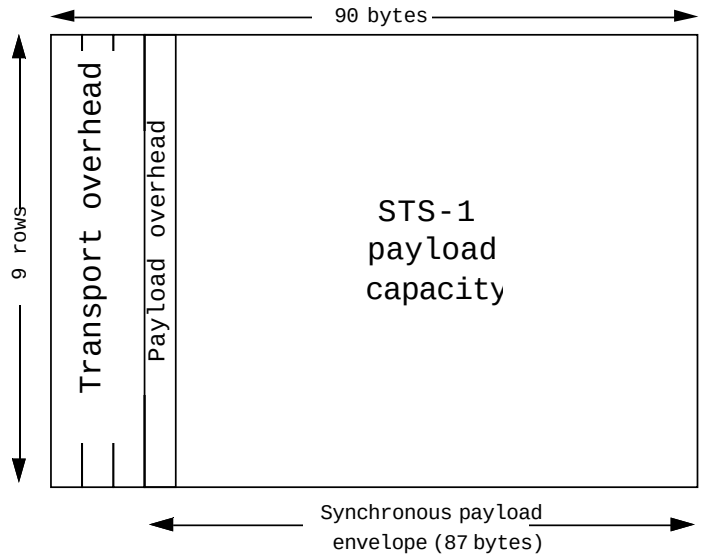


FIGURE 15.18 SONET STS-1 Frame

In the United States, STS-3 (155.52 Mbps) and STS-12 (622.08 Mbps) are expected to be offered as subscriber access rates. For broadband services such as SMDS (super-rate services), the synchronous payload envelopes of the constituent STS-1 frames will be concatenated (only one set of payload overhead bytes is used; see Figure 15.19, an example of an STS-3c frame).

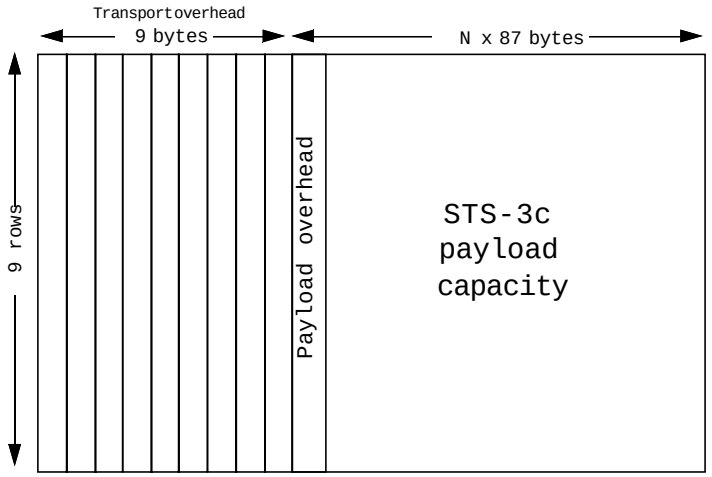


FIGURE 15.19 SONET STS-3c Frame

The impact of SONET will be more economical than technological. The existence of a standard for optical transmission systems will eventually make optical-to-electrical equipment a commodity, and commodity pricing will speed the deployment of fiber-based facilities by public telecommunications providers.

Very High Bandwidth as an Enabling Vehicle for OSI

It is often suggested that the protocol overhead associated with OSI's upper layers is so great that very high bandwidth is needed simply to establish an OSI application association; there are, for example, T-shirts that proclaim OSI to be "same-day service in a nanosecond world".[7] True, OSI's upper layers have lots of header bits, and the ASN.1 syntax hardly qualifies as a bit-economical form of data transfer, but (1) in a multimegabit environment, the overhead is not nearly so significant as it is in kilobit environments; (2) the flexibility introduced to the application writer may be worth the overhead; and (3) end users care about the net result of the application, not how it is encoded. And one simply cannot overlook the fact that ATM itself is hardly the model for packetizing efficiency. The fact remains that if ATM technology succeeds at providing very high bandwidth, and the cost is at or near the "willingness to pay," ATM platforms offer the potential to remove some encumbrances from distributed applications; if "bandwidth on demand" becomes the transmission norm, full-color imaging, recorded and real-time video, and high-fidelity audio can be integrated into distributed applications. OSI's upper layers provide standard tools to develop such applications, over OSI and Internet transport services; ATM/SONET facilities provide enough transport that standard tools may be worth the overhead.

Conclusion

Network architects sometimes imagine that networking happens from the "top down," because that is how network architectures are often constructed—the applications at the top are, after all, presumably the *raison d'être* of any network, so it makes sense to start with the requirements of

7. Van Jacobson came up with this particular poke at OSI, which appears on the T-shirt just below a graphic in which an elephant perches happily, if precariously, on a telephone wire stretched between sagging poles.

distributed applications and let those determine the functions that must be provided by successively lower layers. OSI demonstrates this approach in its alternation of service definition and protocol specification at each layer; in the OSI standards community, the generally accepted practice has been for the group working on one layer to ask for "requirements" from the group working on the layer just above and to use those requirements to decide what should and should not be in the service definition for its layer. In the real world, however, things are not nearly so simple.

The technologies that populate the data link and physical layers of a network architecture do not arise in response to "requirements" expressed by standards developers working in the network layer. The research and engineering that produce a new communications technology—fiber-optic transmission systems, for example, or token-ring local area networks—are driven by a broad range of scientific and economic forces that are far stronger than the voices of standards developers (or network architects) calling for the inclusion of a particular function in "the data link layer" or a particular feature in "the data link service" because it will help them to design network-layer protocol standards. It is important, therefore to pay attention not just to what the service definitions for the two lower layers say *ought* to be provided but also to the actual capabilities of the link-level transmission technologies that have been and are constantly being applied to the construction of new data networks. In an open system standards context, the right question to ask about SMDS or ATM, for example, is not "Does it (or can it) provide the OSI data link service?" but "Could it be used as a subnetwork with an appropriate internetwork protocol?" No one who discovers a promising new data link transmission or switching scheme is likely to abandon it because "it doesn't provide the OSI physical or data link service!" The real story of the data link and physical layers, therefore, is not how they accommodate "requirements" from above but how they embrace and express within a networking architecture the communication capabilities that are perpetually welling up from the real world.