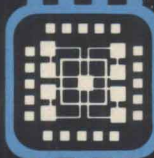
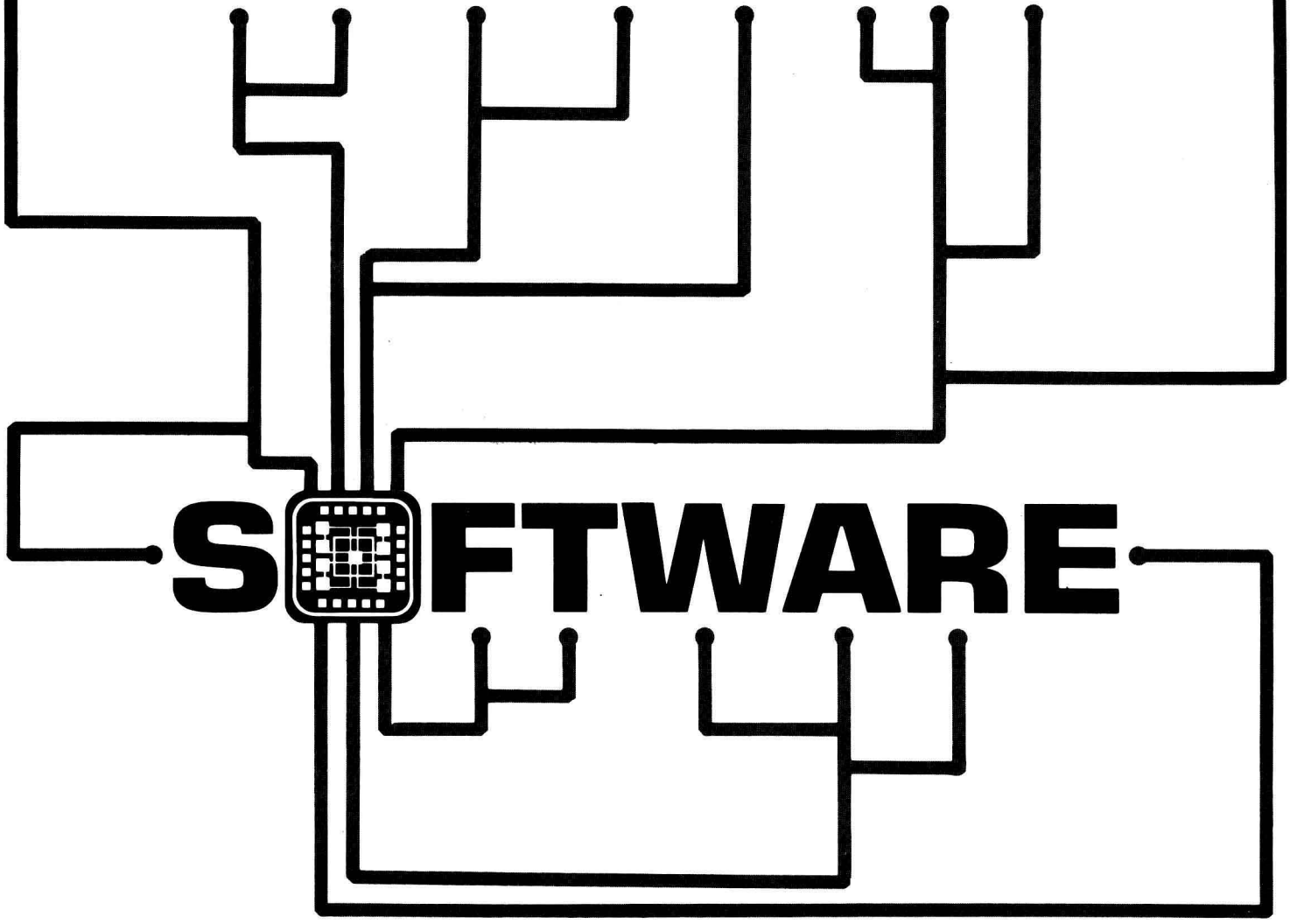
NETWORKING SOFTWARE



EDITED BY COLIN B. UNGARO

McGraw Hill
Data Communications
BOOK SERIES

NETWORKING



SOFTWARE

EDITED BY COLIN B. UNGARO
EDITOR IN CHIEF
DATA COMMUNICATIONS

McGraw-Hill Data Communications Book Series

McGraw-Hill Information Systems Company
1221 Avenue of the Americas
New York, New York 10020

McGraw-Hill Data Communications Book Series

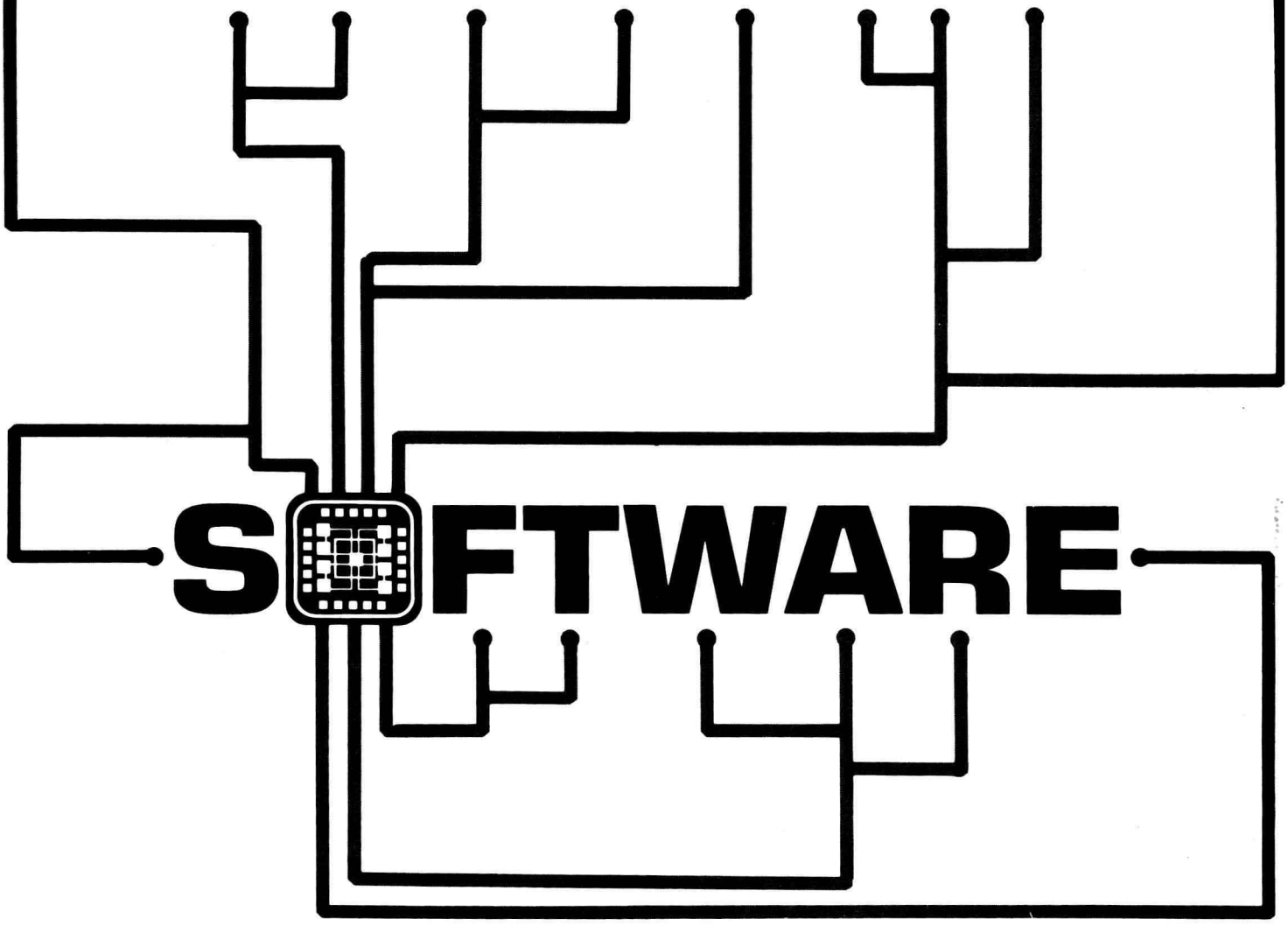
- **Basic Guide to Data Communications.** Edited by Ray Sarch, Executive Technical Editor, Data Communications. 1985, 360 pages, softcover.
- **Cases in Network Design.** Edited by William E. Bracker, Jr. and Ray Sarch. 1985, 275 pages, softcover.
- **Computer Message Systems.** By Jacques Vallee. 1984, 176 pages, clothbound.
- **Data Communications: A Comprehensive Approach.** By Gilbert Held and Ray Sarch. 1983, 441 pages, clothbound.
- **Data Communications: Beyond the Basics.** Edited by Ray Sarch, Executive Technical Editor, Data Communications. 1986, 307 pages, softcover.
- **Data Network Design Strategies.** Edited by Ray Sarch, Executive Technical Editor, Data Communications. 1983, 273 pages, softcover.
- **Integrating Voice and Data.** Edited by Ray Sarch, Executive Technical Editor, Data Communications. 1987, 280 pages, softcover.
- **Interface Proceedings.** Edited by Data Communications Magazine. Annual publication, softcover.
- **Linking Microcomputers.** Edited by Colin B. Ungaro, Editor-in-Chief, Data Communications. 1985, 310 pages, softcover.
- **The Local Network Handbook.** Edited by Colin B. Ungaro, Editor-in-Chief, Data Communications. 1986, 387 pages, softcover.
- **McGraw-Hill's Compilation of Data Communications Standards, Edition III.** Edited by Harold C. Folts. 1986, 4,300 pages, clothbound.
- **Networking Software.** Edited by Colin B. Ungaro, Editor-in-Chief, Data Communications. 1987, 540 pages, softcover.
- **Teleconferencing and Beyond: Communications in the Office of the Future.** By Robert Johansen, with others contributing. 1984, 206 pages, clothbound.

Library of Congress Cataloging-in-Publication Data

Networking software.

(McGraw-Hill data communications book series)
Includes index.
1. Computer networks. 2. Computer software.
I. Ungaro, Colin, 1952- . II. Series.
TK5105.5.N468 1987 005.7 87-2728
ISBN 0-07-606969-9

NETWORKING



SOFTWARE

EDITED BY COLIN B. UNGARO
EDITOR IN CHIEF
DATA COMMUNICATIONS

Preface

Software is the glue that holds communications networks together. Without it, of course, one would be left with only a maze of interconnections that may be interesting to look at, but would be virtually useless. Almost every piece of the networking puzzle utilizes software in one way or another. The importance and complexity of software can be seen in the high price that networking software experts can command.

Virtually every facet of networking embraces software—from network design to network security. Today such relatively new areas as artificial intelligence and expert systems hold great promise for information networking users. And these software topics are covered in these pages under the appropriate sections. No discussion of software, of course, would be complete without some practical applications. These, too, are included here.

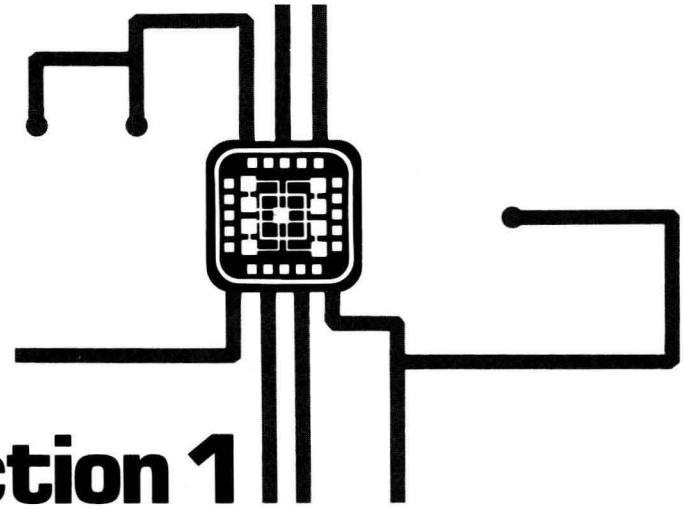
This book—a collection of articles from DATA COMMUNICATIONS magazine—is divided into eight sections: Technology, Architectures, Protocols, Operating Systems, Network Management and Design, Security, Software Selection, and Case Studies. Perhaps more so than in any other book in this series, it would have been very easy to interchange articles from section to section. This says more about the pervasiveness of the software discussed here than the actual difficulty in categorizing it. What differentiates an article in the Technology section from one in, say, the Architecture section is its main focus on tutorial material or the generality of the article's content. Similarly, other sections in this book are categorized by article focus. In many cases this became a difficult judgment call because of the scope of a particular piece.

At any rate, in the Architectures section, you will find such provocative articles as "Problems and Opportunities in Advanced Net Architectures" by John M. McQillan and James G. Herman. This article serves as an ideal overview to the rest of the section. The other chapters are set up in much the same manner.

It is important to note that cost and vendor data is subject to change, so the pricing information in these articles should be treated as sample guidelines. For more up-to-date vendor and product listings, consult such sources as the DATA COMMUNICATIONS Buyers' Guide.

After reviewing this book, the reader should have more than a basic understanding of one of the most complex and important elements of information networking: software.

Table of Contents

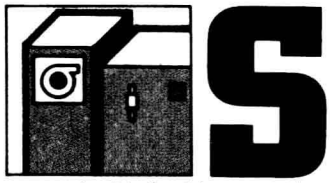


Section 1
Technology

David M. Piscitello, Burroughs Corp., Southeastern, Pa., Alan J. Weissberger, Teledimensions Inc., Santa Clara, Calif., Scott A. Stein, Honeywell Information Systems, Phoenix, Ariz., and A. Lyman Chapin, Data General Corp., Westborough, Mass.

Internetworking in an OSI environment

The authors, who are active in OSI standards groups, provide authoritative information on protocols and connection methods.



tandards bodies concerned with fleshing out the Open Systems Interconnection (OSI) model are also tackling the associated problems of internetworking—that is, communications between an interconnected set of networks. The seven-layer model from the International Organization for Standardization (ISO) has become familiar within the industry. OSI provides a framework for the interaction of users and applications in a distributed data processing environment that may include a wide variety of both computer and terminal equipment as well as many different communications technologies.

The term OSI refers to the seven-layer architectural reference model and to a set of standards that describe how to provide communications among computers and terminals. To examine internetworking, one must understand OSI terminology. For each of the seven layers, a layer service is defined that identifies the set of functions that the layer provides. Layer services in OSI are of two general types: connection-oriented, which allow the service users (entities in the next higher layer) to establish and use logical connections; and connectionless, which allow the service users to exchange information without having to establish a connection (see "Connections vs. connectionless?").

Within each layer, protocols operate to provide the services defined for that layer. As a number of protocol-selection options exist in some layers of the reference model (for example, the Transport Layer defines five distinct connection-oriented protocol classes), conformance requirements are specified by each of the layer-protocol standards. When in compliance with the required suite of standard protocols prescribed for OSI, configurations are considered to be open.

The lowest two layers, Physical and Data Link, provide technology-specific access to the media that is used to interconnect the network equipment. The third layer (Network) performs data routing and relaying.

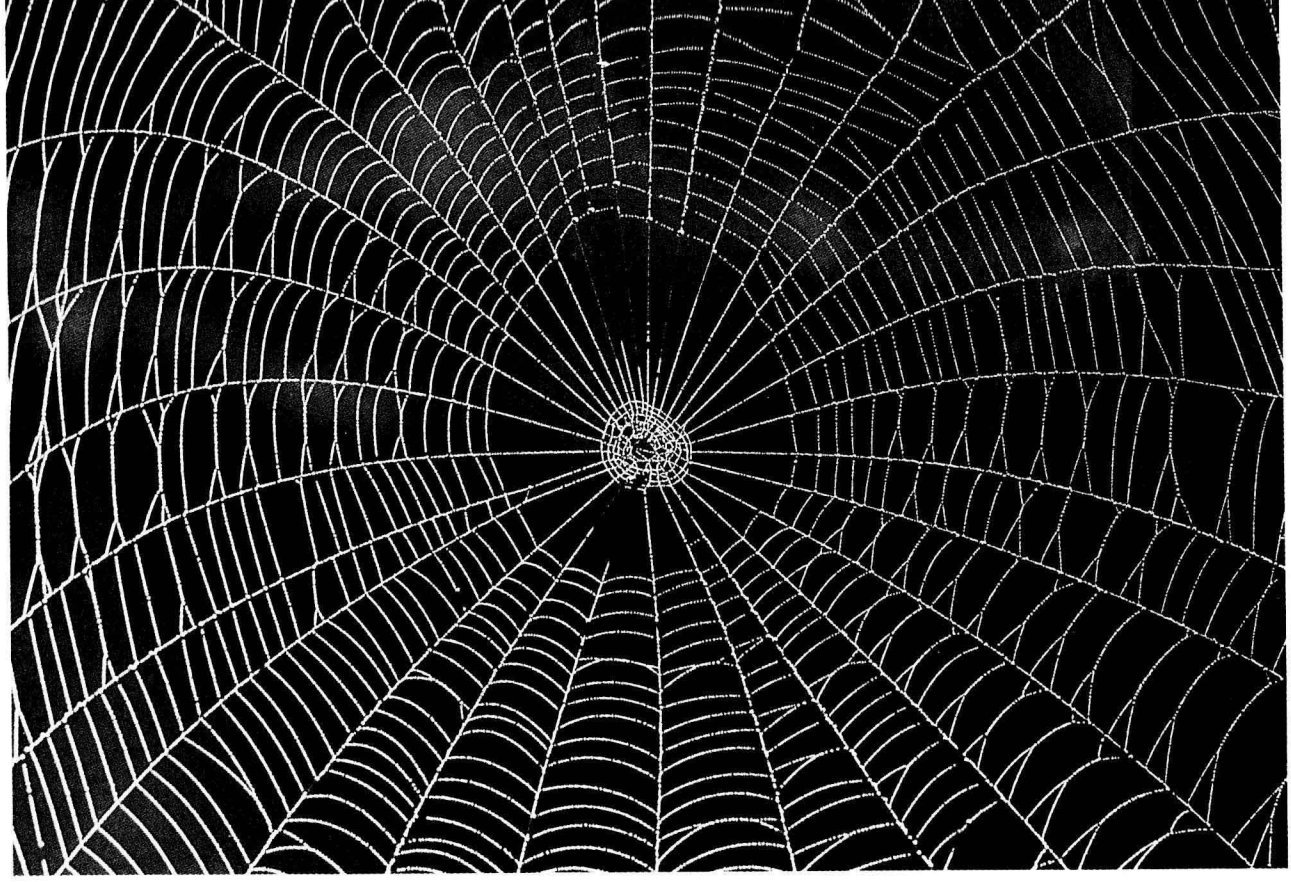
The fourth layer (Transport) provides for the reliable, error-free, end-to-end delivery of user data. An equally important and often overlooked function of the Transport Layer is to determine the most cost-effective means of providing data-transport service. Given specific quality-of-service (QOS) constraints by the higher layers, the Transport Layer matches a Transport protocol to the Network Layer service provided. (The QOS is defined by the characteristics of a connection-oriented or connectionless transmission as observed between the end points.) Thus the QOS requested is satisfied in the most expedient and cost-effective manner. Note that any meaningful analysis of the problems related to network interconnection often requires consideration of the Transport Layer as well as the Network Layer.

Internetworking is important

To be a successful standard, OSI must provide a homogeneous environment in which information may be accessed and exchanged independent of the immediate network to which corresponding users are attached. Network interconnection, or internetworking, has thus become a most important issue.

The advent of local area network (LAN) technologies has lent new urgency to this effort. Part of the world treats LANs as a sophisticated, multipoint data link; the remainder, as merely one more type of subnetwork (OSI parlance for a network as part of an internetwork) that must be accommodated by OSI. Therefore, it is inevitable that internetworking solutions involving LANs would be similarly divided.

The problem of interconnecting networks to form a single, "global" network is an inevitable consequence of the recent explosion of network technology. Network



designers have investigated and implemented a number of interconnection strategies that attempt to facilitate communications among computers and terminals connected to different networks. While all create a homogeneous networking environment by resolving differences in network technology, access method, address structure, and administration, two appear to be most applicable to an OSI global network.

Strategy 1. In this case, the networks to be connected:

- Offer predominantly connection-oriented services.
- Exist where close cooperation among the network administrations can be achieved and enforced.
- Exist where the extent to which the individual network services differ is limited.

With this approach, connection-oriented internetworking may be achieved by relaying the services of one network directly onto corresponding services of the other networks.

An underlying assumption of this network interconnection strategy is that it is easier to solve the problems associated with subnetwork interconnection when the services that the networks offer are the same than when they are different. Hence, to ensure that the services presented on each side of a relay point are sufficiently similar so that a direct mapping of one set of services onto the other is possible, enhancement of individual networks on a "hop-by-hop" basis (see below) may be necessary. This approach is particularly attractive with public data networks and in countries that centralize provision of network services under a single government-controlled administration, such as a Postal, Telegraph, and Telephone (PTT) organization. In such environments, strong regulatory constraints operate to limit network diversity.

Strategy 2. Here, the networks to be connected offer a mix of connection-oriented and connectionless services and exist where network administration is largely autonomous. The extent to which the individual network services differ cannot be predicted or controlled.

In such configurations, connectionless internetworking preserves individual network autonomy and service characteristics of the networks to be connected. This is achieved by conveying the information necessary to support a uniform network service in an explicit Internetwork Protocol (IP). This protocol makes minimal assumptions about the services available from each of the interconnected networks.

One important observation about these two network interconnection strategies: The service that is ultimately provided to the user at the Transport service boundary is the same in both cases. The only real difference between them is where the end-to-end reliability functions are performed. Hop-by-hop enhancement tries to combine both internetwork routing and end-to-end reliability functions in the Network Layer, thereby making each network connection (hop) a miniature end-to-end transport connection. An internetworking protocol, on the other hand, concentrates on dynamic routing (the proper province of a Network Layer protocol) and leaves to the Transport protocol the responsibility for ensuring end-to-end reliability (the proper province of a Transport Layer protocol).

In practice, hop-by-hop enhancement may provide a reliable Network Layer service. However, the users most concerned with reliability, security, and data integrity—most notably the military and the banking industry—require a Transport service that guarantees those properties from end system to end system. (An end system is defined as a seven-layer configuration. For more detail on this and other OSI topics, see DATA COMMUNICATIONS, "The status and direction of open

Connections vs. connectionless?

In the earliest work on Open Systems Interconnection, communications was modeled exclusively in terms of connection-based interactions. These interactions proceeded through the three familiar phases of connection establishment, data transfer, and connection release. Traditional monolithic networks—where applications are the users—are the source of this connection model. The classic example is the voice telephone network, which is operated directly by human users who establish connections (call), transfer data (talk), and release connections (hang up the telephone).

An alternative model of connectionless interactions, which begin and end with the transmission of a single self-contained data unit (often called a datagram), was developed as an extension to the OSI architecture. Standards developers, working in today's much more complex multiple-network environment, encountered interconnection scenarios that could not be modeled satisfactorily in entirely connection-oriented terms. Both models have been applied successfully to the specification of OSI layer services and protocols.

The essential difference between these two models is that a connection preserves the state of peer-to-peer communications from one data transfer to the next, storing and distributing information regarding the connection within the service provider; connectionless transmission does not. The components of a connection-oriented network—such as gateways, switches, and interface processors—operate collectively over time to create and maintain state information for each connection that is established between one point on the network and another. (Examples of state information include the locations of the communicating peers, the sequence number of the last data unit forwarded, and the status of peer-to-peer flow control.) This information relates each new data transfer to previous transfers on the same connection, so that the network deals with data units containing shorthand "pointers" to the information base (such as a connection identifier) rather than the information itself (such as a full destination address).

A connectionless network, on the other hand, does not establish or maintain any relationship between individual data transfers. All of the addressing and other information needed to convey data from source to destination is included explicitly in each data unit. Connection-oriented networks recover from errors through global-state resynchronization; connectionless networks, with no state to preserve, use time-outs and retransmissions. Intuitively, connection-oriented networks are "deterministic"· connectionless, "probabilistic."

The obvious question that arises from the existence of two different interaction models in the OSI environment is how to choose between them. Unfortunately, the connections vs. connectionless debate has often been carried on as if the question were simply, "Which model is 'better' "—without regard to the context in which it is applied. Those arguing in such terms would have one believe that because a connection seems to be the best model for a particular application, it must also be the best model for the entire data communications structure that supports the application. Or they argue that because connectionless transmission is the most efficient mode of operation for local area networks that use a particular contention-based access method, it must also be the best model for the operation of all types of networks.

A less dogmatic—and much more practical—approach would consider the individual characteristics of the two types of operation in terms of the applicability of one model or the other to specific interconnection scenarios. This approach recognizes that OSI can be viewed from a number of different perspectives, and that it is often necessary to apply different techniques to the solution of different problems.

In general, when a service that processes a large number of data units in essentially the same way—while keeping track of the data transfer state on behalf of the communicating peers—is desired, the connection model is more useful. When the dynamic flexibility of a service that processes data units independently—or the simplicity of a service that does not maintain state on behalf of the communicating peers—is desired, the connectionless model is more useful. In the real world, of course, there are almost

systems interconnection," February 1985, p. 177.) In any configuration in which end-to-end reliability is important, the most reliable Transport Layer protocol available must be used, and any effort to enhance the subnetworks is largely wasted (although the user, of course, is expected to pay for the enhancement).

Hop-by-hop: For simplicity?

In strategy 1, gateways (Network Layer relays) perform a mapping of the service offered by one network onto another. In general, the gateways do not add services. Rather, they perform the relaying and switching functions necessary to bind the individual subnetworks into a unified or global network. A consequence of this approach is that either all of the subnetworks must inherently provide equivalent services, or each must be enhanced to some common level of service.

The ISO has included this interconnection strategy in its "Internal Organization of the Network Layer" standard (ISO 8648). Called hop-by-hop (subnetwork) enhancement, the approach may be summarized as follows (Fig. 1): All subnetworks that are to be interconnected must provide exactly the Network Layer service (usually, connection-oriented), either directly or through enhancement. Any subnetwork that does not provide this service must be enhanced or modified to

always practical trade-offs rather than absolute positions, and they must be evaluated as such.

From the perspective of the OSI user, the only relevant concern is whether the application's operation is connectionless or connection-oriented. From the perspective of the OSI builder, the important issue is the way in which a particular combination of connectionless and connection-oriented layer services and protocols solves the technical, administrative, and economic problems of interconnecting OSI networks. These two perspectives are essentially independent. The user's choice is based on the characteristics of the application. The builder's choices depend on the available network technologies, an overall strategy for achieving network interconnection, and the economic and administrative constraints of a particular operating environment. The clearest illustration of these choices is found in the nature of applications (the ultimate users of OSI) and the different ways in which multiple networks can be interconnected.

Beneficial examples

Some interactions between peer application entities involve the exchange of many related data units. These interactions must be performed in an explicit application context that defines the intent of the entities' exchange. Such applications can benefit from a connection-oriented service. Examples include bulk file transfer (particularly when checkpoint/recovery features are implemented); virtual terminal usage (long-term attachment of a terminal, workstation, or other interactive device to a host computer, for which the security established during an initial log-on procedure is an important part of the context associated with a connection); and access to distributed network components, such as print servers and remote-job-entry stations.

Other application interactions are either entirely self-contained or do not benefit from the context characteristics of a connection. Examples include "inward" data collection (periodic sampling of a large number of data sources, as in a sensor network); "outward" data dissemination (the distribution of a single piece of information to a large number of destinations); broadcast and multicast (group-addressed) communications; and a variety of other request/response applications (such as directory and time-of-day servers) in which a single request is followed by a single response, with no significant relationship between one message and the next.

Interconnection realities

If all networks were of the same type, used for the same purposes, administered by the same organization, and operated under the same tariffs and regulations, interconnecting them would pose few major technical problems. Clearly, not many of these conditions—much less all of them—are met in the real world.

There are excellent reasons for the providers of public network services to prefer that their interactions with users and with each other be connection-based. These service providers must deal with unpredictable and widely fluctuating loads; must limit variations in quality of service to a relatively narrow range, according to the terms of a legal contract; and must charge for their services on a fair and auditable basis in accordance with that contract. Deterministic global resource allocation is of paramount importance. On the other hand, private networks, such as LANs, tend to be owned and used by the same organization. Their operating costs are generally recovered in ways that are only indirectly related to individual instances of use.

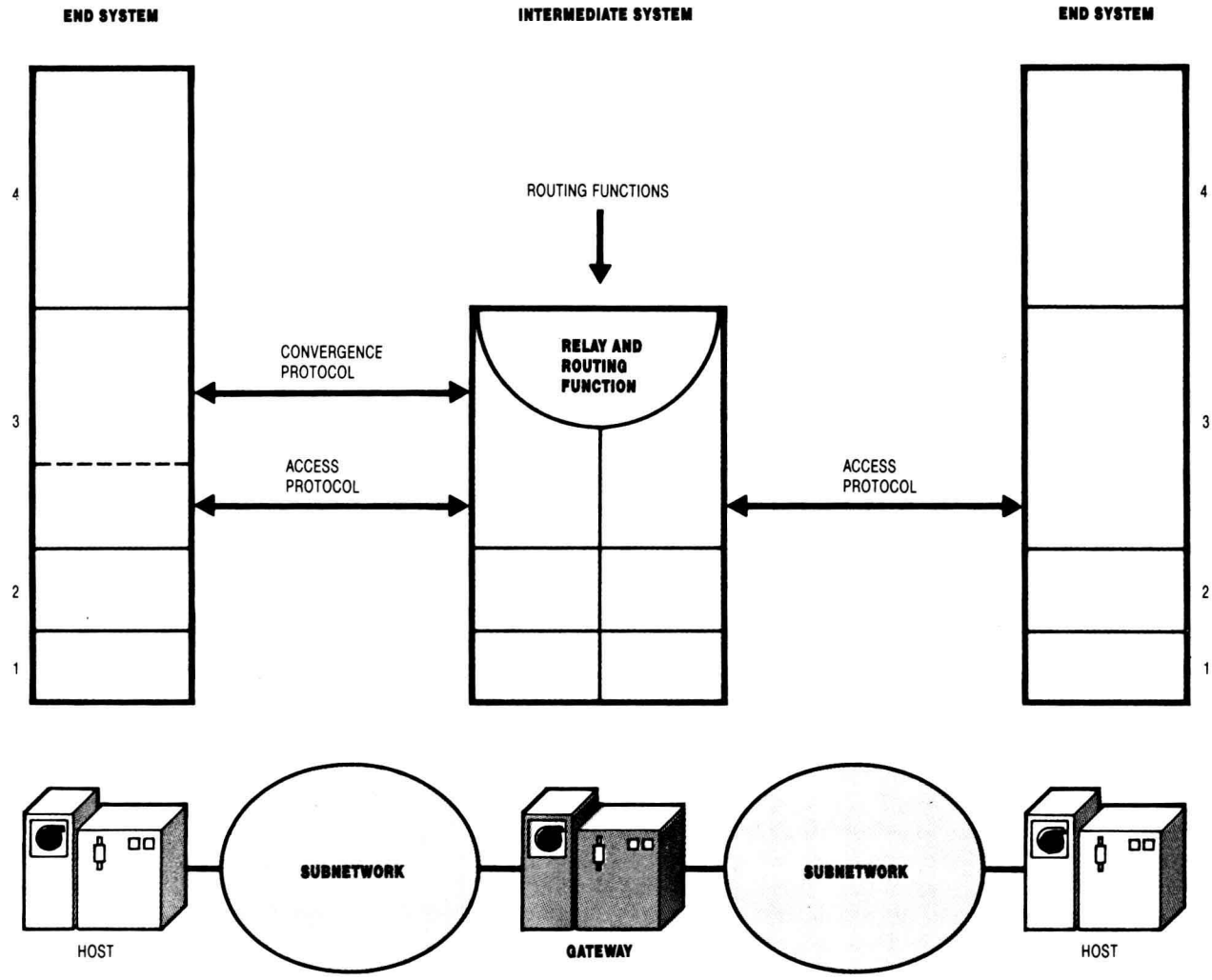
Public-network administrators exercise greater global control over their configurations than do private-network administrators, to the extent that public networks are operated (and perceived by their users) as one global network. Private networks, however, usually consist of a number of individual, interconnected smaller networks, forming an internetwork topology in which boundaries persist that are related to administration, local control, and mode of use. The managers of private internetwork topologies are concerned with the flexible and reconfigurable interconnection of a variety of individual networks, and they are correspondingly reluctant to make too many assumptions about the nature of individual underlying services. These networks tend to be connectionless.

do so. Relays are used to passively map the connection establishment, data transfer, and connection release facilities of one subnetwork onto another whenever network connections cross subnetwork boundaries.

The enhancement of subnetworks that do not provide the OSI network service is usually accomplished in either of two ways: through direct modification of the protocol used to access the subnetwork, as is the case for the 1984 version of CCITT Recommendation X.25, or through the use of what is called a subnetwork dependent convergence protocol (SNDCP). An SNDCP operates on top of a subnetwork access protocol (SNACP), such as X.25, to provide the elements of the OSI network service that are missing from the access protocol. Such a protocol (ISO 8878) has been developed for operation with X.25's 1980 version.

One of the consequences of the hop-by-hop approach is that when a subnetwork administration revises its access protocol to provide the OSI network service, all data terminal equipment (DTE) attached to that subnetwork must often be revised. If the SNACP is modified principally to accommodate internetworking, the user and the DTE manufacturer must absorb the migration expense regardless of whether the DTE is to be used for internetworking—or even whether it will be used for OSI at all. Often such a revision does not result

1. Enhancing hop-by-hop. *All subnetworks to be interconnected that do not provide the Network Layer service (usually connection-oriented) must be enhanced to do so. Gateways are then used to map the connection establishment, data transfer, and connection release facilities of one subnetwork onto another.*



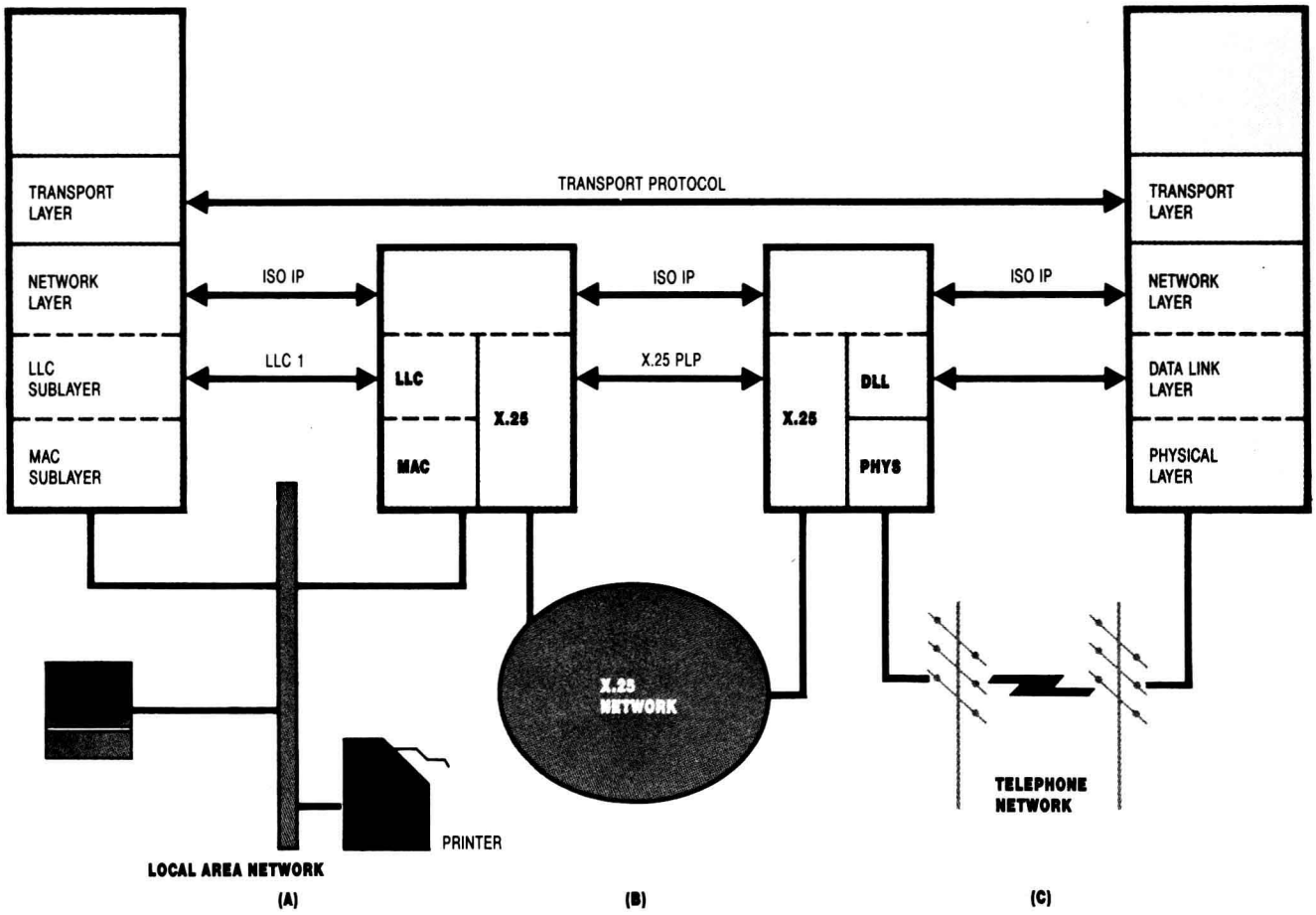
in any significant improvement in the service provided in the local environment. In fact, accommodations for connection-oriented internetworking often result in lower performance (due to, for example, additional protocol overhead) for local information exchanges.

This is both unreasonable and inefficient. In most subnetworks, local traffic predominates; hence, most subnetwork designs justifiably emphasize efficiency for local traffic. Protocols and addressing are often tailored for the local topology, traffic requirements, and underlying technologies (such as LANs) employed. It should not be necessary to change intranetwork operation—in particular, to modify its access protocol—solely for the purpose of internetworking. The imposition of additional overhead on intranetwork communications just to accommodate the less-frequent instances of internetworking is a poor bargain for most users. Moreover, problems of network interconnection cannot always be resolved by unilaterally imposing a single SNACP on all subnetworks. A more pragmatic approach to internetworking is to examine the interconnection scenarios, then base configuration design on those scenarios that are to be accommodated.

The use of an SNDCP permits existing terminal equipment to coexist in the intranetwork environment with equipment that supports the full OSI network service. However, since an SNDCP is designed to operate in conjunction with a specific SNACP, one must be developed for each subnetwork type to which a particular device may be attached. Such a proliferation of convergence protocols is hardly desirable.

Hop-by-hop subnetwork enhancement, then, is a practicable choice only where the type of service offered by the majority of the subnetworks to be interconnected corresponds very closely to the OSI network service. It is a poor choice for configurations in which the subnetworks provide dissimilar services, provide different or unpredictable qualities of service, or are managed and administered in different ways. In these types of configurations, an alternative method of

2. Connectionless internetworking. *The networks to be connected offer a mix of connection-oriented and connectionless services. The connectionless internetworking strategy is based on the use of a single end-to-end protocol to provide the Network Layer service over any combination of subnetworks.*



DLL = DATA LINK LAYER
IP = INTERNETWORK PROTOCOL
ISO = INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
LLC = LOGICAL LINK CONTROL PROCEDURE
LLC 1 = LOGICAL LINK CONTROL PROCEDURE, CLASS 1
MAC = MEDIUM-ACCESS COMPONENT
PHYS = PHYSICAL LAYER
PLP = PACKET-LEVEL PROTOCOL

network interconnection should be considered.

Strategy 2 is based on the use of a single end-to-end protocol to provide Network Layer service over any combination of subnetworks (Fig. 2). This IP is operated in a sublayer above the subnetwork-specific protocols in both the hosts and the gateways (see "Protocol relationships"). It performs the addressing and routing functions necessary for end-to-end communications independent of the subnetwork-specific functions operating in each of the interconnected subnetworks. The underlying subnetworks should provide only a data transmission service. No subnetwork enhancement is necessary; an IP can be operated directly over the Data Link Layer.

This second approach to network interconnection was successfully applied by a number of user communities and computer manufacturers long before the ISO began to examine the problems of internetworking. The Defense Data Network (DDN) of the Department of Defense (DOD) and such private networking architec-

Protocol relationships

The relationship of the ISO Internetwork Protocol to the Transport protocol, to others in the Network Layer, and to the Data Link Layer is illustrated in Figure 2 of the accompanying story. The ISO IP may operate (a) over a LAN that uses the IEEE 802.2 LLC (Logical Link Control) Type 1 Class 1 Data Link procedures; (b) over a Public Data Network that uses the X.25 Packet Level Protocol; or (c) directly over a leased or switched telephone line that provides the OSI Data Link service through data link procedures, such as ISO's HDLC or IBM's SDLC. An addendum to the ISO IP, a Draft International Standard, describes how to operate the IP over (a) and (b) above. The U. S. technical subcommittee on the Network Layer (ASC X3S3.3) introduced a description of (c) in April of this year.

tures as Xerox's XNS (Xerox Network Systems), Burroughs Corporation's BNA (Burroughs Network Architecture), and Digital Equipment's Decnet use internetwork protocols. These protocols provide an efficient and flexible means of interconnecting networks that differ in their types and qualities of service, but they do not impose unnecessary constraints and complexity in the local environment.

The ISO has formalized the second, or internetworking protocol, approach in its "Protocol for Providing the Connectionless-mode Network Service" (ISO 8473). It is designed to operate entirely within the computers (gateways and hosts) and terminal equipment attached to public-switched and private packet data networks, LANs, or other subnetworks. Like forerunners of its kind, the protocol ignores the idiosyncratic characteristics of the individual subnetworks, extracting from each subnetwork a data transmission service with no quality of service or other essential properties. The protocol demands so little from the underlying service (in the subnetwork sublayer or the Data Link Layer) that it may be operated directly over leased and switched telephone lines that provide the OSI Data Link service.

Here is how the ISO IP operates in the OSI environment. Computer and terminal equipment attached to different networks exchange data packets—called Internetwork Protocol Data Units (IPDUs)—in a connectionless (datagram) mode of operation. Each IPDU is routed independently. The route an IPDU takes is determined by the current state of the network links, rather than by the state that existed at some previous connection-establishment time.

Much of the information necessary to determine a route (such as source and destination addresses and quality of service) is conveyed in the protocol control information (PCI) of each IPDU. Additional information, which enables end and intermediate nodes to cooperate dynamically in determining the best route for each IPDU, is distributed and maintained by Network Layer management functions and protocols. In most cases, routing decisions are made independently by each forwarding node. However, the network entity (an active element within the Network Layer) of the end system that generates an IPDU may specify the route the IPDU should take by employing the protocol's optional "source routing" function.

The ISO IP assumes only that each subnetwork traversed is capable of serving as a "data pipe." The IP also harmonizes the service offered by each subnetwork into a uniform connectionless network service. Because subnetworks are required only to provide a data pipe, the mapping of the service required by the IP onto that provided by most SNACPs is extremely simple. For example, in the case of a LAN station operating under the procedures of IEEE 802.2 Type 1 Class 1 (unacknowledged connectionless), the mapping is a relatively trivial one-to-one process.

The ISO IP provides a connectionless network service to the Transport Layer. Unlike a connection-oriented service, a connectionless service dynamically allocates such resources as CPU, buffers, and data link availability on an as-needed basis rather than statically for the duration of a connection. This method of resource allocation is extremely efficient for normally encountered traffic conditions. Coupled with a deterministic or adaptive routing algorithm, it generally results in optimal use of network resources.

It is possible, however, that unusually high traffic volume (involving considerable delay), a temporary shortage of buffer space, the loss of a data link, or some other transient condition will cause an IPDU to be lost. In addition, the Network Layer may not be able to deliver IPDUs in the same order in which they were generated since alternate or parallel routes may be used for the transmission of a particular sequence of IPDUs. (These other routes would be used to enhance throughput or service quality, or in response to changes in the topology or service characteristics of the underlying networks.)

Thus where end-to-end reliability is important, the Transport Layer must provide reliable end-to-end connections that preserve data integrity and sequence. This is achieved by operating a Transport protocol that is able to recover from the loss, duplication, corruption, or reordering of data. In most internetworking configurations, service interruptions such as those mentioned above are relatively infrequent and short-lived. The Transport protocol design must also ensure that any penalty associated with the reliability function is assessed only when the function is invoked. That is, the additional overhead for normal (error-free) operation should be negligible. The ISO Class 4 Transport Protocol (TP 4) was designed precisely for this purpose.

Of the five Transport protocol classes developed by OSI (see Table), Class 4 ensures data integrity and data-transfer reliability when there is a possibility that any of the underlying networks will lose, corrupt, or reorder data. TP 4 utilizes a number of sophisticated error-detection and recovery mechanisms. Detailed below, they include transport protocol data unit (TPDU) numbering, retention of TPDUs until acknowledgment, retransmission of TPDUs following a time-

Comparing transport protocol classes

FUNCTION	CLASS 0	1	2	3	4
ERROR RECOVERY	NO	YES	NO	YES	YES
EXPEDITED DATA TRANSFER	NO	YES	YES	YES	YES
EXPLICIT FLOW CONTROL	NO	NO	YES	YES	YES
MULTIPLEXING	NO	NO	YES	YES	YES
DETECTION AND RECOVERY FROM:					
LOST TPDUs	NO	NO	NO	NO	YES
DUPLICATED TPDUs	NO	NO	NO	NO	YES
MISORDERED TPDUs	NO	NO	NO	NO	YES
CORRUPTED TPDUs	NO	NO	NO	NO	YES

TPDU = TRANSPORT PROTOCOL DATA UNIT

out, resequencing of TPDUs, and TPDU checksums. TP 4 provides a reliable, end-to-end transport service no matter what happens in the underlying networks.

All TPDUs that contain user data (DT TPDUs) and are forwarded to a given destination are marked with a sequence number. TPDU sequence numbering is used by flow control, resequencing, and recovery mechanisms to ensure that each transmitted DT TPDU is acknowledged by its destination within a given time interval. If this time interval elapses without receipt of an acknowledgment TPDU, the DT TPDU is retransmitted with the same sequence number as the original transmission. The sending transport entity (an active software element in the Transport Layer) retains a copy of each transmitted DT TPDU until an acknowledgment TPDU is received from the destination transport entity. Once the acknowledgment is received, the resources used to hold the copy are released.

As indicated earlier, misordered (out of sequence) DT TPDUs may occur within the Network Layer. To ensure that data submitted by the source session entity is delivered in sequence to the destination session entity, the receiving transport entity uses the sequence numbers of the DT TPDUs to reconstruct the original user-data sequence. Only TP 4 has this capability.

TP 4 makes use of a checksum mechanism to detect the corruption of TPDUs that are not detected by the Network service. If the receiving transport entity determines that a TPDU has been corrupted, the TPDU is discarded. The retransmission mechanisms described earlier will cause the TPDU to be retransmitted by the sending transport entity.

TP 4 provides explicit flow control. Cooperating transport entities determine a maximum number of outstanding, unacknowledged TPDUs (called the window size) for both communications directions. During the data transfer phase, neither transport entity is permitted to transmit or retransmit more than this number of DT TPDUs until an explicit acknowledgment of previously transmitted DT TPDUs is received. This mechanism, coupled with congestion control procedures in the Network Layer, provides a means of maintaining a uniform service quality for users of the transport service.

Much ado about nothing

One of the most misunderstood features of TP 4 is its extensive use of timers and retransmission mechanisms. These mechanisms are widely recognized to be necessary when a Transport connection is constructed over one or more error-prone subnetworks. However, it is frequently assumed that the presence of the mechanisms results in inefficient operation when the underlying networks are more reliable than the error-prone subnetworks. This is not true. The protocol overhead associated with TP 4 is no greater than that associated with TP 2 and TP 3 during data transfer. In fact, the DT TPDU format is exactly the same for TPs 2, 3, and 4.

From a procedural standpoint, if TP 4's reliability mechanisms are not invoked (that is, TPDUs arrive in sequence, uncorrupted, without loss of data), TP 4 has no more overhead than does any other Transport protocol class. Careful implementation of the protocol and adjustment of the retransmission timers will ensure that no unnecessary delays are introduced during normal operation. When the underlying networks are highly reliable, therefore, the procedural cost of running TP 4 is the same as the cost of running any other Transport protocol class. However, if errors do occur—TPDUs could be lost, duplicated, or corrupted by the underlying networks—then clearly the reliability mechanisms invoked by TP 4 to recover are necessary. This can hardly be classified as overhead.

Proponents of the hop-by-hop approach claim that there will be simplification in Transport protocol operation as a result of the high degree of reliability provided by the enhanced subnetworks. The Transport protocol class of choice in this environment is usually TP 1. However, in the absence of errors, TP 1 and TP 4 operations are essentially equivalent. If an error does occur, the situation is quite different. TP 1 can recover after a failure is correctly signaled by the Network service but cannot detect or recover from errors that are unsignaled or incorrectly signaled, such as corruption of user data. Hence, any user community concerned with data integrity must provide recovery mechanisms at the user application. It is difficult to understand how this can be called simplification.

The protocol's advantages

Because the ISO IP operates in a sublayer above each subnetwork, all the subnetwork sees is data, which is precisely what it saw before there was a need for internetworking. Subnetworks, therefore, are completely unaffected by the presence of the IP (and, for that matter, of OSI). Subnetwork protocols are left intact, and users can continue to employ existing terminal and computer equipment and software. Since the subnetwork protocols continue to be concerned only with efficient operation in the specific environment for which they were optimized, local traffic is not penalized by the introduction of internetworking.

The simple data transmission service required by the ISO IP is readily obtained from the vast majority of subnetworks in existence today. The ISO IP provides segmentation and reassembly mechanisms to accommodate different subnetwork packet sizes. Each subnetwork can continue to use the packet sizes for which it is best suited. Note that the ISO IP was designed specifically to satisfy the OSI connectionless-mode Network service. Therefore, the Transport Layer need not rely upon the quality of service characteristics of any individual subnetwork to provide the QOS requested.

One of the benefits of using the Internetwork Protoçol is that the changes required to provide communications beyond the local subnetwork environment are restricted to the terminal equipment and gateways involved in internetworking. In all cases, subnetwork autonomy and local-traffic service characteristics are preserved. Terminal equipment can readily be designed so that the processing overhead associated with internetworking is incurred only when internetworking is actually performed. Routing functions within the Net-

work Layer determine whether the destination is in the local subnetwork. If so, the Inactive Network Layer Protocol (INLP) is used.

The INLP is actually a "null IP subset." It is used strictly to enhance performance, and it serves as an indicator to the destination end system that the complete ISO IP is not present, that the IPDU originator was on the local subnetwork, and that the source and destination subnetwork addresses have been mapped directly onto the corresponding OSI Network addresses for the purpose of expediting the transmission. Contrary to what has been stated in an earlier article (DATA COMMUNICATIONS, "Of local networks, protocols, and the OSI reference model," November 1984, p. 129), the routing functions that determine when to use the INLP are performed entirely within the Network Layer, and its use is completely transparent to the Transport Layer or to a user.

Finally, from the standpoint of product development, there is one significant advantage of using the ISO IP instead of a hodgepodge of subnetwork dependent convergence protocols. Since the IP can be operated over any subnetwork access or data link protocol, users and DTE manufacturers seeking OSI compatibility can limit the number of convergence protocols they must provide to one.

The tailor-made solution

The nature of most LAN applications ensures that a large proportion of the information exchange takes place within the confines of a single LAN or bridged-LAN topology. For other applications, however, data exchange beyond the local area must be made possible. There is an urgent need for the industry to consider a way to interconnect LANs to other LANs—and to other networks such as packet- or circuit-switched types. Primarily, the impact of providing interconnection of these subnetworks should be minimal. While nonlocal traffic may be generated infrequently, the information exchanged over long (nonlocal) distances is likely to be some of the most important. The ISO IP offers a tailor-made solution to this complex set of LAN interconnection requirements.

Since both the IP and the LLC1 (Logical Link Control Class 1) have relatively simple "state [of operation] machines" (essentially "send" and "idle"), Network Layer design in LAN workstations, as well as in LAN-to-LAN gateways, is quite simple. Each Network Layer service user request-to-send-data results in the generation of one or more IPDUs. The IPDUs are passed to the LLC sublayer along with the destination's LAN station address. The LLC sublayer encapsulates the IPDU into "Unnumbered Information" frames and submits them to the "Medium Access Component [device]" for transmission. When the frame is delivered to the destination LAN station, the receiving LLC sublayer strips the control information from the LLC1 frame and passes the IPDU to the internetwork sublayer for forwarding by the routing function (at a gateway) or for delivery to the transport layer (the "user" of the Network Layer service at the LAN station). The procedures for operating the ISO IP over LLC1 are described in greater detail in an addendum to the ISO IP: ISO 8473/DAD1.

Gateways between connection-oriented networks (particularly X.25 types) and LANs are the only devices that require the complexity of a connection-oriented state machine. Typical operations involve a call, reset, or disconnect that is pending or awaiting confirmation. However, the complexity is confined to the gateways: End systems on the LANs do not have to operate a connection-oriented state machine at the Network Layer. Also, in many scenarios a wide area network, such as an X.25 type, is but one "hop" among many. Particularly for internetworking scenarios involving many LANs, the number of hops on which the overhead of an additional protocol over a wide area network is incurred is relatively small.

A significant benefit of using the ISO IP is that gateway complexity is greatly reduced because of the way subnetwork connections—in particular, X.25 virtual circuits—are managed. ISO 8473/DAD1 identifies a set of mechanisms and timers for opening and closing X.25 logical channels. This greatly reduces the number of occasions on which state transitions associated with maintaining a connection interfere with the efficient operation of the gateway.

The data transmission service realized through the manipulation of X.25 virtual circuits is essentially the same as that offered by an OSI data link. The establishment and release of X.25 virtual circuits is governed by timer mechanisms. Logical channels are left open (available) to transmit IPDUs for as long as is economically and administratively efficient.

Some concrete examples

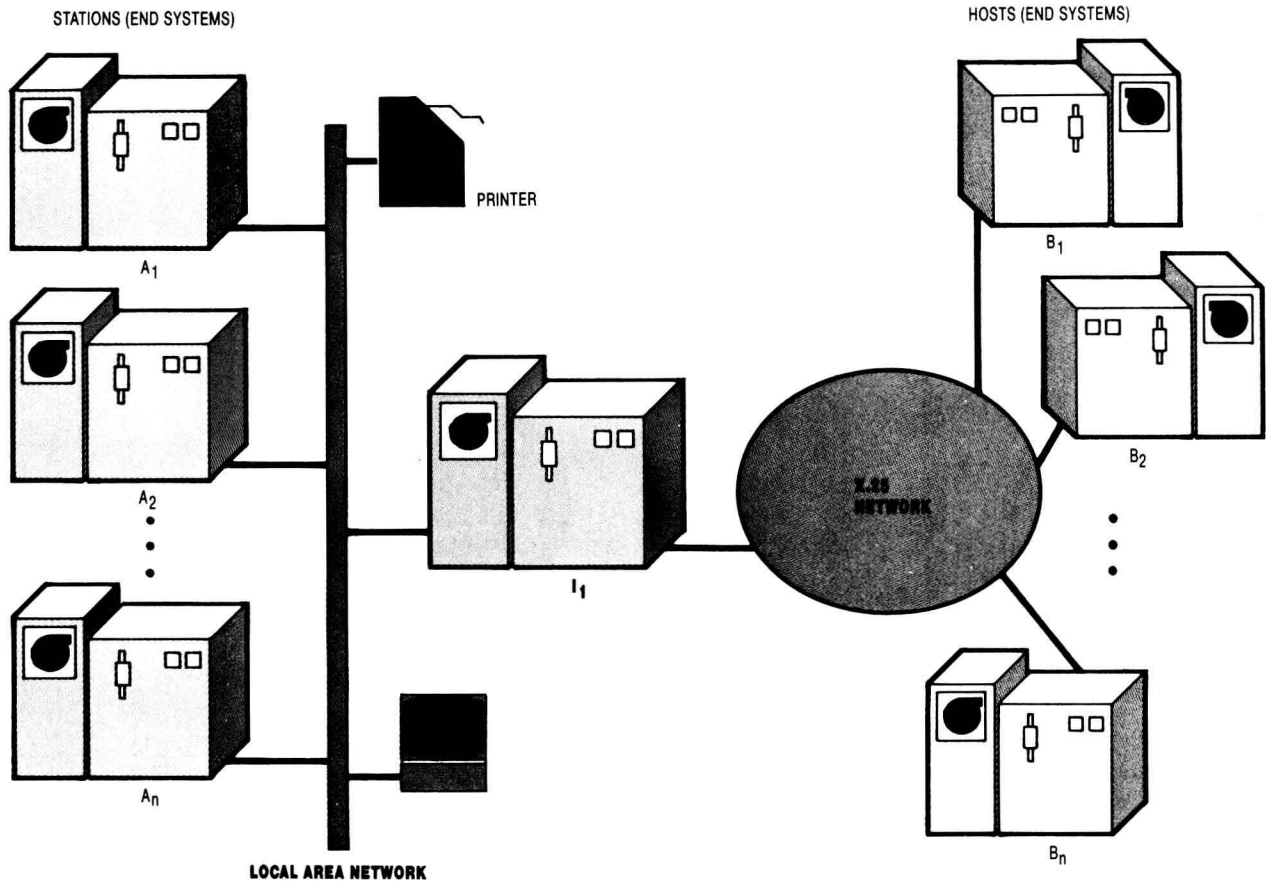
The ISO IP should be used where LANs are involved in internetworking; benefits are derived from resource optimization, throughput enhancement (through the use of load-splitting techniques), and redundancy and resiliency (the ability to adapt to redundancy).

The ways in which the ISO IP enables Network service users to get more for their money are considered next.

Resource optimization. Using the hop-by-hop approach (Fig. 3), each LAN station (end system), A, must establish a separate connection to each X.25-network device (end system), B, in order to transfer data. For each connection, resources (such as buffers, a connection-state-information base, and CPU) must be reserved in both end systems as well as in the relay or intermediate system, I_1, for the duration of the call. In particular, I_1 must have ample capacity to maintain all of these connections, even if no traffic is passed. Clearly, if connections remain idle for long periods of time, valuable network resources are wasted.

In contrast, if the ISO IP is used, the sending end system may free resources as soon as the data unit's transmission is completed. Normally, the receiving end system reassembles the IPDU prior to indicating its delivery to the destination Transport entity. I_1 processes each data unit separately, allocating buffer space as required and maintaining a simple state machine (send or idle). Any communicating pair of Network service

3. Hopping between networks. *Using the hop-by-hop approach, each station, A, must establish a separate connection to each X.25-network device, B, in order to transfer data. For each connection, resources must be reserved in both end systems (A and B) as well as in the intermediate system, I_1, for the call's duration.*



users (hereafter identified as A,B) that has long periods of inactivity between transmissions imposes no overhead. Therefore, the ability of the intermediate system to process transmission requests from any other communicating pair (A,B) remains unaffected. This typically results in highly efficient use of resources.

Throughput enhancement. In many internetworking scenarios, the ability to route IPDUs independently is particularly useful. Data exchanged between hosts attached to one subnetwork can be routed to hosts on a different (remote) subnetwork without the constraint that all data must be routed down the same path (and hence, through the same gateway). Using multiple paths to transmit data to the same destination typically improves throughput and response time.

The ISO IP provides an excellent means of equalizing a potentially vast throughput disparity through the use of load splitting techniques. Consider those configurations (Fig. 4) in which an IEEE 802-compatible LAN (operating at 4, 5, or 10 Mbit/s) must work with a lower-speed wide area network (the maximum throughput of which is normally 48 or 56 kbit/s). In a connection-oriented environment, a separate connection must be established to each network to transfer data. In addition, this separate connection must be relayed through either I_1 or I_2 relay systems.

A number of constraints can immediately be identified.

- Some knowledge of which intermediate system (I_1 or I_2) is to be used for which connections—as well as how many connections each intermediate system can maintain—must be known prior to establishing connections, to prevent intermediate-system overloading.
- Once a connection is established via one intermediate system, the resources of the other cannot be utilized for this connection, as no mechanisms exist for preserving the sequence of data to be transferred if routed via parallel paths.
- If an intermediate system experiences a failure, all of its connections are broken and must be re-established through a different intermediate system.

In contrast, use of the ISO IP resolves all of the above constraints. Each IPDU contains all of the information necessary to uniquely identify it, and each is processed and routed independent of all other IPDUs. Hence, either or both of the intermediate systems can be used to transmit data between end systems (A,B).

Routing functions performed as part of the ISO IP's operation make it possible to use any available intermediate system to transmit IPDUs. This is extremely