Intelligent Multimedia News Networks (IMNN)

by

Shakil Khan

A Dissertation submitted to the Graduate Faculty in Computer Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

2000

UMI Number: 9986345

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Computer Science in satisfaction of the requirement for the degree of Doctor of Philosophy.

May 24, 2000
Date

Professor Syed V. Ahamed
Fellow IEEE
Chair of Examining Committee
The City University of New York

May 24, 2000
Date

Professor Theodore Brown
Executive Officer
The City University of New York

May 24, 2000
Date

Professor Michael Anshel
The City University of New York

Professor Michael Kress
Vice President of Technology Systems
College of Staten Island
The City University of New York

Dr. Victor Lawrence
Fellow IEEE
Vice President
Advanced Technologies
Bell Labs Innovations
Lucent Technologies

Supervisory Committee

The City University of New York

Abstract

Intelligent Multimedia News Networks (IMNN)

by

Shakil Khan

Advisor: Professor Syed V. Ahamed

In this study an intelligent network for the news has been proposed. The study includes detailed analysis of developing the means to produce, distribute, and retrieve multimedia news over the intelligent network. The goal is to build the intelligence in the network so that all the stages of news creation, storage, management and delivery to regular subscribers and occasional customers are smooth, efficient and error-free.

The study discusses an intelligent system concept for network-based multimedia news and information services. A flexible information structure to support composition, linkage, reuse, indexing, versioning, access control, and presentation control for live and stored multi-media has also been discussed.

As part of the study the existing news services of Associated Press, Reuters, AFP and Down Jones were studied and several COMNET III based simulation models had been developed. The model made efficient use of available multimedia and intelligent communications technologies, including but not limited to IP (Intelligent Peripherals) and SS7 (Signaling System Number 7).

We had analyzed in great detail and compared the commonly used current

technologies in the current news wire systems with the future trends of deployment of ATM backbone with and without intelligent concepts.

However, the model may be applicable to any other news and multimedia delivery system in the world. Besides its application to news and multimedia networks, somewhat similar model can also be used for wide range of applications, such as, distance learning, collaborative design, video catalog shopping, etc.

Acknowledgements

I wish to express my appreciation to Professor Syed Vickar Ahamed for his continuous supervision and encouragement throughout the course of this research. I highly appreciate his suggestions to develop the intelligence in the Multimedia News Networks. Special thanks to Professor Michael Anshel for his suggestions and advice during the entire course of the research work.

I am thankful to Professor Michael Kress, Vice President of Information Technology at the College of the Staten Island for his suggestions and guidance in the core research.

My special regards go to Paul Wu, the Director of Management Information Systems at the Associated Press for his continued support and guidance that has made it possible for me to complete this study. I would like to express my appreciation to Vince Heether, Senior Communications Executive, Communications and Technology Department of the Associated Press for providing me the appropriate access to the technical resources and his technical suggestions from time to time.

Special gratefulness is to be addressed to my wife Anisa and my children Omer, Yousuf and Fatima who spared me to concentrate to the research. I would also like to thank Dr. Victor Lawrence, Vice President of Advanced Technologies, Bell Labs Innovation; Lucent Technologies for his overall guidance and help.

Intelligent Multimedia News Networks (IMNN)
Table of Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

In this study an intelligent network has been proposed. The study includes detailed analysis of developing the means to produce, distribute, and retrieve multimedia news over intelligent network.

The goal is to build intelligence in the network so that all the stages of news creation, storage, management and delivery to regular subscribers and occasional customers are smooth, efficient and error-free.

The study discusses an intelligent system concept for network-based multimedia news and information services. A flexible information structure to support composition, linkage, reuse, indexing, multiple versions, access control, and presentation control for live and stored multimedia will also be discussed briefly.

The model will make efficient use of available multimedia and intelligent communications technologies, including but not limited to IP (Intelligent Peripherals).

The current model has been built after studying and using some of the core structure of wire systems of the Associated Press, the Dow Jones and The Reuters. However, the model may be applicable to any other news and multimedia delivery system in the world. Besides its application to news and multimedia network, a somewhat similar model can also be used for a wide range of applications, such as, distance learning, collaborative design, video catalog shopping, etc.

1.1 Features of a News System

A number of requirements distinguish this application from conventional multimedia publishing efforts:

- **Time critical:** There is always some activity-taking place around the globe, e.g. war, financial news, and weather changes. Therefore, the timeliness of information is critical.

It is absolutely necessary that as many events be made available as quickly as possible.

- **Volume:** Several hours of program material will be produced and distributed on a daily basis. Multi-gigabytes of highly compressed information will accumulate every day.
- **Live and stored content:** The application calls for live "broad-cast" video to be presented along with material that has been digitized, compressed, and stored as a data file.
- **Prioritization:** The information provider will prioritize content by establishing criteria to determine its urgency.
- **Alerts:** Users will be notified of important new information as soon as it becomes available.
- **Updates:** Since news is inherently dynamic, numerous revisions of the same news story may be published within minutes of each other.
- **Aging:** Certain news stories are discarded at the end of the day, while others are retained for months.
- **Archiving:** Some information will be archived for permanent search and retrieval.
- **Dynamic user needs:** Users have ever-changing information needs and need to be able to vary their "interest profiles" at any time. The information must be coded and categorized so users can retrieve it at any time.

- **Distribution control:** The information provider must be able to target each news story to a particular group of authorized subscribers, and be able to monitor the delivery.

1.2 System Concepts

It is my assertion that newswire systems are essentially file transfer systems, that is, stories (files) created by editors on word processing systems are transmitted over wide-area-networks to computers at client sites, where they are copied as files, displayed on a terminal, or printed. A major part of the system described here is involved in the storage and forward transfer of files. The other major function of the system is the allocation of network bandwidth and system facilities for the transmission of isochronous (constant bit rate) streams of data for which there is no intermediate storage, and for which transmission resources must be allocated to permit synchrony between the creation and reception of the data stream. System components include the Production Center, a wideband network, the Media Service Center, and customer sites. At the news Production Center, media files are stored on one of two file servers. When a story is published, the file(s) are transferred to the multimedia storage subsystem of the Media Service Center. Then, they are forwarded to similar file servers located at the customer sites. Stories can also originate as time-dependent streams of data from a video codec. When an appropriate command to the system is given, the output stream from the codec at the Production Center is passed through the network and delivered to the codecs at the customer sites.

1.2.1 Information Content and Structure

The basic information unit of the system is a file. A file can contain data like ASCII characters, bitmaps or structured graphics, a stock market quotes or the results of a database search. Files can also contain time-based information such as digitized streams of audio or video, or a series of frames stored as animation.

Normally, each file contains only one media type. For example, a text file only contains text. But stories may be collections of files. For example, a simple story might contain only a single file, while a complex story might contain dozens of files with a varying mix of all media types. The decision for which types of media to use for each story is left to the editors. Complex stories may use text, graphics or streams of data from a video codec. It is important to distinguish the codec stream from the digitized video files because files can be delivered over a packet network (though the delivery speed or time is not guaranteed), but codec streams cannot. If there is insufficient bandwidth in the network to deliver the stream, the codec at the customer site will become "data starved" and the video will appear broken up and erratic, or may not appear at all. A file, on the other hand, may simply arrive later than expected, but can be played back without distortion. Control information for both files and streams are stored in a data structure called an Object Control Structure (OCS). For files, the OCS contains a "pointer" (full path name) to the data on the file server. For the codec stream, the OCS contains a pointer to the network channel on which the stream will appear, the time the stream will start, and its approximate duration. OCSs can be combined in varying hierarchies (i.e. stories can contain other stories) to permit the most efficient packaging and reuse of the information content as well as the network and system resources. It is this ability to combine both stored and isochronous data in a structure that permits such a rich blending of these two dramatically different media forms that makes this system unique.

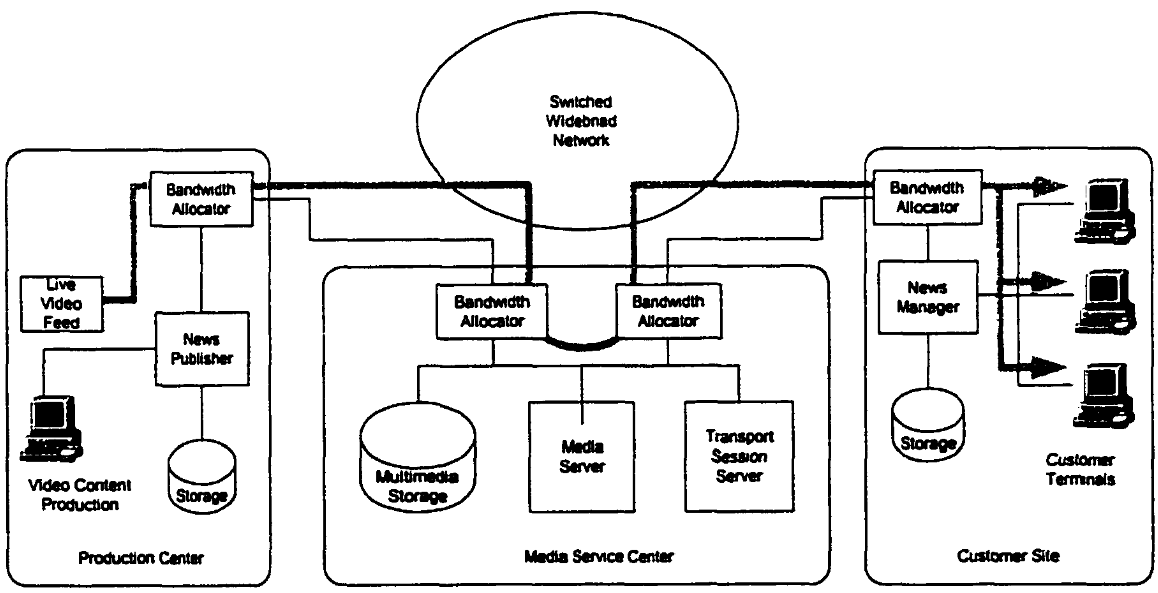


Figure 1.1. Overall Architecture

1.2.2 Transport Facilities

A wide area network connects the Production Center and customer sites to the Broadcast Media Service Center. The current implementation of this network is either over private line T1 circuits, or wireless broadcasts.

1.2.3 Media Service Center

The Media Service Center provides centralized storage as well as communications and information management for the system. These services are designed to be generally applicable to a broad range of information including the product offering.

The Transport and Session Server controls and monitors communication paths used by the system. The Broadcast Service Center provides central database, notification, transaction, and distribution services. It controls the transfer of media files into the multimedia storage subsystem, and allows storage, retrieval, searching, modification and deletion of the OCSs and media files.

1.2.4 Customer Site

It is customary that the news service provider usually operates the customer site equipment. The customer site equipment is usually a mirror image of the file servers and communication equipment used at the Production Center, except that the custom application called News Manager is present to coordinate end user requests and notifications.

1.3 Information Model and Structure

A key objective of the system is to provide a common information model that would support the various kinds of live and stored data produced by multimedia applications while still providing uniform access and control of the data at the Broadcast Service Center. In addition, the model should not only be adaptable to one's changing requirements, but also fit the needs of other information providers. Looking ahead, this same model should also be suited for the wide range of other applications, such as distance learning, collaborative design, video catalog shopping, etc. Thus, flexibility and extendibility of the data model are important. Other characteristics that should be supported by the information model include efficient transmission, information reuse; indexing for fast search and retrieval, update control, privacy and security, and presentation control for multimedia.

1.3.1 The Media Object

The main entity in the information model is the Media Object, an abstract entity that encapsulates numerous data formats and mediums and is uniformly accessible by upper layers of the system. A Media Object contains two kinds of information:

1. An Object Control Structure (OCS).
2. An optional media data set (file or stream), which is the raw data payload of the Media Object.

A Media Object may optionally point to other media objects to provide object composition.

1.3.2 Composite Objects

In order to support the creation and representation of compound stories composed of multiple possibly independent components that can be manipulated separately and perhaps added incrementally. The information model must support object composition. Moreover, multiple composite objects may use a single object. For example, the same video segment can be used in several different news stories. This requires the ability to share objects among multiple composite objects.

1.3.3 Object Versions

Versioning of objects is desirable for any design environment. Authors may need to revise their artifacts several times, or reuse pieces to create new artifacts. In addition, versioning can reduce the amount of digital storage space required.

Within the context of the news application, versioning is imperative. Late breaking news is frequently updated as new information is obtained, or as Metadata codes are added. In a typical model, each object has a special attribute called a modification (mod) number, which is maintained by the system and keeps track of multiple versions of an object. Object modifications are considered non-destructive. A later mod can safely replace an earlier mod. For example, the event list in a later mod is normally a superset of the event list in the earlier mod. This assures that the binder object can safely reuse child objects.

1.3.4 Multimedia Data Sets

The raw materials of a news story are data sets, each containing data of a specific format. The data set can be a stored file or a continuous stream that occurs on a channel during a particular time interval. A data set may contain raw device control data, e.g., pixel values or sound samples, as well as auxiliary format and structural information. Data sets are usually associated with existing data handling software or hardware systems. Examples include: An ASCII text file; A Macintosh pict file; An AVS compressed video file; A Px64 compressed video stream. Each data set belongs to a general media type, (e.g. video) and a specific media type (e.g., dvi), and has specific attributes (e.g., rows, columns, frames per second, bit rate). To optimize for digital storage, Data sets may be compressed or uncompressed.

1.3.5 Object Control Structure (OCS)

The OCS for a media object is the collection of attributes for that object. OCSs contain both the system attributes as well as the publisher attributes. System attributes are mandatory and are contained in each OCS. Publisher attributes are specifically defined to meet the needs of each information provider. System attributes are used for access control, version management, storage management and distribution management.

1.3.6 Event Lists and Multiple Views

A single object can be accessed as a set of logical sub-components. For example, a video file may be defined to have 5 logical "scenes," where each scene is tagged with a start and end frame number. This information is currently encoded in an object's event list. With this representation, it is possible for a story to "play" scenes out of order (5,1,3,2,4). Various stories may access different scenes in the same object, thus implementing multiple views of the same underlying media objects. In addition, a parent object may inherit the play list of all of its children.

1.3.7 Object Consistency

Having defined an object model, leads me to define a set of properties that define the consistency model. Once defined, the transaction management layer has to ensure that the system is always in a consistent state. The following properties must all hold invariably in Media Object Server and Media Data Server:

1. A single file must always exist in its entirety or not exist at all in the Data Server. Likewise, a single OCS must always exist in its entirety or not exist at all in the Object Server.

2. For a FILE object, the OCS and its associated file must both exist in the Object Server and Data Server, respectively.

3. A binder must have all of its child objects present at any time (recursively).

4. The modification (mod) number of a child object must be greater or equal to the mod number referred to in the object list of its binder. Note that any object (including binder objects) may reside in the Media Server without its parent binder, even if it originally had one. It may, or may not, be joined at a later point by its original or by new parent(s). The reason for this relaxation is to enable more efficient and consistent management.

1.4 Production Center

The Production Center is the source of all of the news material for the system. It consists of several components: several Production Stations (a Capture station, an Editing station, a Text/Graphics station, an Authoring station), an Audio-Video File Server, a Text/Graphics file server, and a specialized application called News Publisher, which is the repository for all control information in the Production Center. All of the stations are usually Intel PCs and run on an Ethernet Local Area Network. The Production Stations are equipped with multimedia features. The Capture Station digitizes and compresses the analog audio and video and stores the "raw" media files on the File Server. Video is compressed in RTV 2.1 format. Audio and video can be input to the system from any live source (satellite, cable, etc.) or from standard analog tape. The Editing station is an off-the-shelf DVI-compatible Digital Video Editing (DVE) system used to post-produce the raw audio and video files created by the Capture station. The Text/Graphics Station is a standalone editing station used to create or import all of the text and graphics-based media files for the system. Both the text package and the graphics package are off-the-shelf. The Authoring station is a custom software application that allows an editor to create a multimedia story by reviewing and selecting the text, graphics, audio, and video-based media files on the file servers. The files are combined under display formats or templates, a headline is written, and the Metadata and authorization codes are attached, and a command is sent to News Publisher to begin transferring the story to the Media Server.

The Audio-Video File Server is an off-the-shelf server that combines special hardware and software to provide guaranteed delivery of isochronous streams of audio and video to and from the Production Stations. News Publisher is the station that controls all of the activities in the Production Center. All of the Production Stations communicate with News Publisher to negotiate the creation, deletion, and modification of Object Control Structures for the media files, and to interact with the Media Service. News Publisher stores all of the OCSs in a small local database, and assigns all system-defined values in the OCS.

1.4.1 Story Flow

Like other established editing environments, the production of multimedia news stories is, at first glance, hierarchical. A story passes through several levels of editing until it reaches the most senior level editors or their deputies, who have final responsibility for publication of the article. The story appears to flow in linear fashion from one person to the next. However, a more detailed analysis of the traffic flow shows it to be quite serpentine. Editors often send stories back to reporters for minor changes or complete revision. Two or more reporters may collaborate on an article and frequently share incomplete drafts. A reporter checks the figures on a chart or graph to compare them against the text in an article. Senior editors review work-in-progress at all levels. To accommodate for these traffic flows, the Production Center is capable of maintaining a collaborative editing environment, under control of the News Publisher application.

Once a media file has been edited and is ready for inclusion in a story, the operator of the station requests the creation of an Object Control Structure to News Publisher. News Publisher creates the OCS, stores it in the local database, and passes an object identifier (OID) back to the requesting application. News Publisher also makes available the OID and all relevant information about the article (headline or caption, date, revision, etc.) to all Production Stations. Applications can then retrieve the caption for that media file and display it to their users.

1.4.2 Modification

For each OCS, News Publisher maintains a count of the number of times the OCS has been modified since it was last published. This is known as the operational modification (op-mod) number, and is different from the mod number used in the transfer of the media files to the Media Server. From the perspective of the Production Center, the mod number indicates the number of times that an OCS and its related media file have been published to the

Media Server. Conversely, the op-mod number is used to indicate the number of times that the OCS is modified between publications. The Production Station applications compare op-mod numbers for each OCS to determine whether they have been modified since the last poll.

1.5 Media Transport and Session Server

The Media Transport and Session Server provide application independent transmission management and session services. The underlying transport is provided by private line services (T1/E1) or switched wideband service, such as primary rate ISDN.

The Transport Manager layer allocates logical channels within the transport medium to allow media with differing performance requirements to be transmitted appropriately. The quality of service characteristics includes bandwidth, error rate, and latency. The basic function used for the current application is the dynamic allocation of isochronous stream and packet channels. The Session Server layer prioritizes data transfers and manages connections. A key component is an inverse multiplexer, called the Bandwidth Allocator, that can dynamically call up multiples of D0 circuits and aggregate them into higher bandwidth channels for such functions as signaling, rapid file transfer, and isochronous video streams. Multiplexers are at each end of the communications links, and operate under the control of the transport and session server.

1.6 Media Server

The Media Server is responsible for the storage and distribution of all media objects, including the OCSs, and the underlying FILE and LIVE multimedia data sets. The Media Server usually consists of three server modules: Transaction Server, Object Server and Data Server.

1.6.1 Transaction Server (TS)

This is a central component of the Media Server used properly and efficiently to distribute objects from the publishing site(s) to the Media Server to the multiple customer sites that are authorized to receive the information. In addition, it provides scheduling services including queued task management, prioritization, and coordinated information transfer. Finally, it interfaces with the Object Server for information access, and provides resource management. In order to ensure the consistency of the object-base (as defined in section all operations involving media objects are executed within a context of transactions.

1.6.2 Object Server (OS)

The OS provides persistent and consistent storage and retrieval of OCS information, and maintains the consistency between the OCS and the files in the Data Manager. In addition to stored data, OS manages OCSs of live objects. OS is called from the Transaction and Distribution server and calls the Data Server and the Video Stream Distributor. The transaction and distribution server manages the multi-step process of adding objects. OS guarantees the incorporation of a single object addition, and assures the consistency of the OCS and the corresponding media file.

The object delete service primitive marks an object as available for deletion. Deletes are not necessarily recursive, that is, deleting a binder does not necessarily purge its contents. An expiration mechanism normally collects and deletes orphan objects. The Object Server maintains an expiration date on all objects. When an object is expired, it is automatically deleted, but only if it is not contained in any binder.

1.6.3 Data Server (DS)

DS is responsible for managing; storing, and retrieving multimedia data sets. DS is shielded from knowledge of the object model, and deals only with raw multimedia data. DS interfaces only with the Object Server.

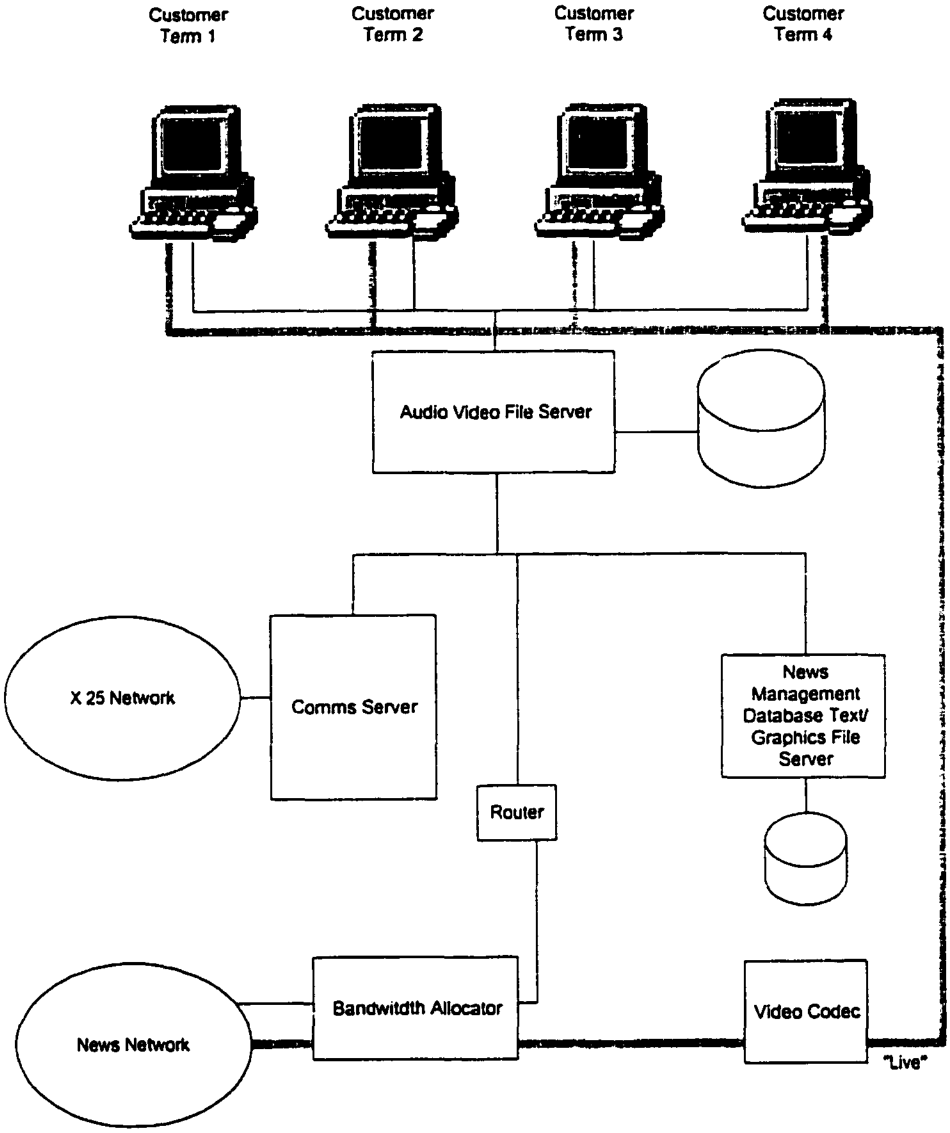


Figure 1.2. A Typical Customer Configuration

1.7 Customer Site

A customer of the multimedia service will be provided access to multimedia information (as published by the Production Center). Although the two services utilize vastly different delivery and transaction mechanisms, and are currently delivered on logically separate feeds, the two information streams share several important characteristics: every story event contains publication date (for aging control), modification information (for database update synchronization), metadata codes (for content indexing), and headline captions (posted as an event notification to "interested" users).

1.7.1 Delivery System

The customer site delivery system, Multimedia News Manager, consists of five components. The Communications Server, a specialized application called News Manager, an Audio-Video File Server, a Text/Graphics file server, and Client Workstations.

- **Communications Server** provides X.25 feed termination.
- **News Manager** is the station that controls all of the activities at the Customer Site. The News Manager is a client of the Media Service and receives feeds from the X.25 Communications Server. News Manager communicates with all the client workstations to send alerts, and provide headline feeds. News Manager stores all of the OCSs in a local database, and manager's files stored in the Audio/Video File Server and the Text/Graphics File Server. It also stores all the data in a local database.
- The **Audio/Video File Server** is a media server that provides guaranteed delivery and storage of all text/graphic files and multiple isochronous streams of audio and video to the production stations. In order to

satisfy real-time constraints on each of several LAN segments, it also serves as a bridge between these segments. The client workstations include digital video/audio hardware, run special retrieval software and communicate with all other components at the customer site. Additional components include Router, Codec, and Bandwidth Allocator.

The News Manager application provides broadcast alerts (headline captions with metacode) notifying all client stations when a multimedia story has been downloaded and is available for viewing. At the client workstation, the user application filters these broadcast alerts based upon attached metadata codes and local user-defined preference profiles. Depending upon this metacode analysis, the alert may be routed to one or more view windows, or it may simply be discarded. Additionally, multimedia story data is always accessible by means of database search.

Chapter 2

USER INTERFACES

2.1 Integrated Services Digital Network (ISDN)

The Integrated Services Digital Network may be viewed as a network evolved from a telephony Integrated Digital Network, which provides end-to-end digital connectivity to support a wide range of services, including voice and non-voice services, to which users have access by a limited set of standard multi-purpose user-network interfaces.

In a more comprehensive view, an ISDN is a network that can provide a myriad of data and telecommunications services. It is a fully digital network, where all devices and applications present themselves in a digital form. That is, information from the telephone, personal computer, stereo, television, Private Branch Exchange (PBX), mainframe, mini, workstations, Personal Computers (PCs) are all seen as bit streams by the network switch. Therefore, all of the information can be switched and transported by the same network equipment.

2.1.1 Forces Behind ISDN

There are, in fact, several forces driving the move to ISDN. Among them are:

- The growing demand for domestic and international digital communications service for voice and data, particularly as businesses become more geographically disperse and dependent upon international cooperation.
- The demand for increased network simplicity, flexibility, and control at economical prices.
- Separating evolving customer premises equipment from evolving network equipment allows users to purchase the equipment best suited to their applications regardless of the network hardware. The integrated network is flexible enough to adapt to multimedia and other emerging technologies while protecting the users current investment.

2.1.2 Switched wide-area network technology

Bearer or B channels carry phone calls or the telecom operators to carry numbering, signaling and maintenance information at 16 kb/s use user-defined data at 64 kb/s, while the delta or D channels.

Two types of ISDN services are available: basic rate and primary rate. Basic-rate ISDN has two B channels and one D channel and is targeted at the mass-market end user. The high-speed primary-rate ISDN is intended more for corporate users and network-based information servers. In the United States and Japan, it has 23 B channels and one D channel; in Europe, it has 30 B and two D channels.

ISDN belongs to the wide-area networking category of communication services, as opposed to local-area technologies. As the name implies, local-net technologies are designed to operate within a small geographical area and typically at high speeds. Examples are the universally used Ethernet, Token Ring, Token Bus, and Fiber Distributed Data Interface (FDDI). Such networks are usually installed and managed by the company that uses them and are constrained by external regulations or restrictions.

In contrast, a wide-area network imposes certain constraints on users and equipment. The operator that installs and manages the network defines the rules and regulations for the users. Since a large geographical region is covered, the data network operator is generally the telephone company or some closely related entity.

As for wide-area network equipment, it has to comply with international standards; otherwise, gear from different vendors could not work together easily, if at all.

Wide-area network services may use switched or nonswitched lines. A nonswitched or leased line is simply a line or a virtual link rented out for a fixed price to a specific network user. On digital lines, it is an on-demand dial-up service that makes a portion of the public network available to the customer for that customer's internal needs. In the case of a switched line, users pay a subscription for connection to the network. They may then connect to any other user of the network, typically by dialing a particular number. Because, ISDN's D channel allows complete numbering information to be carried across the network, it is ideally suited for switched applications.

2.1.3 ISDN and other protocols

Carrier-provided communication services are currently available in the United States across a wide range of line speeds, using several types of access technologies.

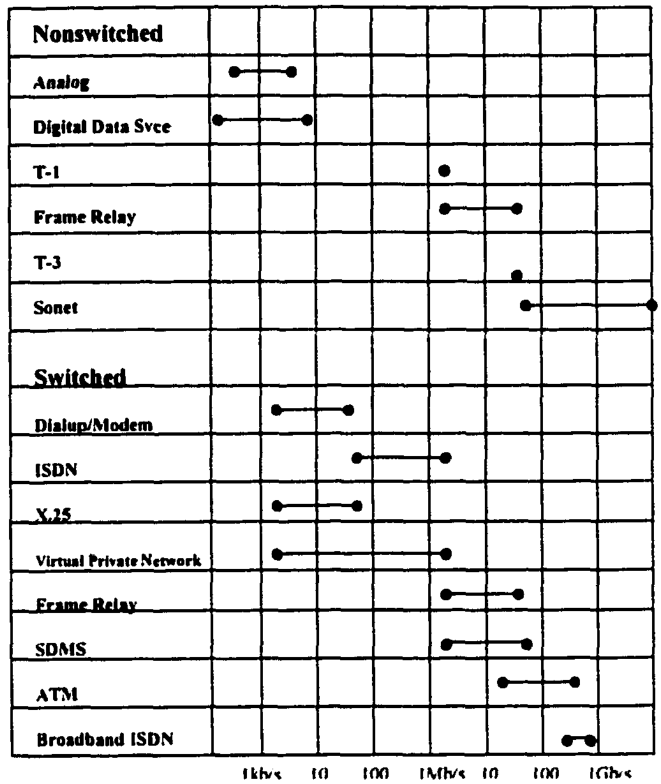


Table 2.0 Comparative Chart of Media

Remote data communications equipment is still mostly using analog modems with data rates of 1200 baud to 56 kb/s. Modems are popular with consumers due to their availability, low cost, and ease of use.

But within companies and other large organizations in the U.S., services based on the T-carrier (the T-1 is the most popular) often are used on dedicated lines for the transmission and distribution of images, messaging, conferencing and many other applications.

In Europe, with its patchwork of international lines, packet switching is favored, and the ITU-T X.25 service has proven the ideal network protocol and service for pure data communication. X.25 is complete series of well-established international standards that employ reliable error-correction protocols. Tariffs are based on the volume of data transferred, whereas ISDN charges depend on the duration of the call regardless of how much or how little data is transferred. Still, X.25 and ISDN networks are interconnected, so that users may send ISDN messages across to any connected X.25 equipment and vice versa.

Frame relay may be viewed as T-1's expansion into the world of switched technologies. Although it is not suited for voice applications, the low overhead of its protocol stack-error-check sequences and such are minimized. High line quality is assumed with T-1s, which makes it very well suited for high-speed internetworking between local and wide-area networks.

In the future, Asynchronous Transfer Mode might well become the heart of the high-speed network fabric, in which the borders between local- and wide-area networks start to disappear.

2.1.4 B & D channels

On average, the time required for call connection (the stretch between the last digit dialed and the first ring) is 10-20 seconds for non-ISDN phones.

During this period, the line is monopolized and cannot be used for other purposes. Using ISDN equipment, the call connection time is slashed to just 3 seconds or less. At the root of this dramatic improvement is ISDN's D-channel, which handles all information regarding call set up, call control, call-by-call service selection, and other types of call intelligence.

The D channel can additionally be employed for control purposes during conversations over the B channels and so may carry information in real time about the on-going phone call-back to the calling party. This feature is handy for communicating pricing and billing data to the person making the call.

The advantage of ISDN and the D channel are perhaps clearest in the context of facsimile transmission, which of course has been growing by leaps and bounds. In an average company, roughly one third of long-distance phone charges are now attributable to fax traffic.

ISDN-connected Group 4 fax machines (the first to be custom-made for digital lines) are about 10-20 times as fast as the fastest Group 3 machines (the best of those made for analog fax lines) in use today. Reckoning in the short ISDN call-setup time of 3 seconds, a simple calculation shows that an eight-page Group 4 fax could be sent in full while an average Group 3 fax machine would be still trying to set up the call. Further, the quality of Group

4 faxes is astonishingly like the printing quality of expensive laser printers.

2.1.5 High-Speed Multimedia Usage

To start with, ISDN will facilitate public video transmission and video conferencing services through "bonding": concatenating B channels. Aggregating just two channels enables the latest generation of video devices to transfer low-speed live pictures at 128 kb/s. If bonding is then combined with the use of new video compression algorithms, such devices can transmit full-speed video across the network.

Besides affecting video, the ISDN high-speed lines will add fuel to the personal computer explosion, which conversely, by synergistic effect, has been a key to ISDN's inception and evolution. This is because a PC, together with a telephone, has the built-in capacity to simultaneously handle voice, data, and images.

Industry analysts predict that, for the rest of this decade, PC sales will grow fastest in the small office/home office market segment. Typical home users will use ISDN products for purposes such as:

- Workplace applications, where the network interface is used to communicate with all business partners and to access remote data banks and other external on-line services, such as those available over the Internet.

- Telecommuting, where the network interface will mainly be connected in reverse-charge mode to the corporate servers containing applications, databases, and so on. Telecommuters will also consult services available over the Internet.
- Entertainment, bank services, travel arrangements, electronic phone books, and the like.

Several large corporations have started encouraging employees to work at home. These virtual commuters or telecommuters, whose numbers are growing rapidly, are for the most part remotely connected to a head-office having data banks and application servers. The office normally would pay all toll charges of remote users phone calls, and so the ability to reverse phone charges is an obvious need. This feature is simple to implement with ISDN's D channel capabilities, far simpler than it is with conventional methods using automatic callback procedures running only on dedicated servers.

2.1.6 Network Interface

Once small and home office users and telecommuters start being connected over a wide area, it immediately becomes vital for them to conduct mixed communications over high-speed digital phone lines. A dedicated ISDN card added to the PC is usually the most cost-effective solution.

For users who prefer not to invest in PC cards, there are low-cost pocketsize terminal adapters. These versatile devices can unobtrusively supply any standard asynchronous application with easy ISDN access.

The market penetration of these products varies in different parts of the world. Most European basic-rate ISDN users integrate cards into their PCs, but terminal adapters tend to be more popular in the United States. The ideal ISDN Network Interface Card or terminal adapter covers many features.

Worldwide compatibility is required in order to reach the manufacturing volumes that will reduce product prices to affordable levels. Accordingly, hardware must be designed to accommodate any remaining regional differences in physical interfaces. A comprehensive application program interface (API) must support Commodity operating systems, such as Microsoft's Windows so that many applications may utilize the same card. These cards should be designed to the most recent bus standards; for new designs, the high-performance Peripheral Chip Interconnect bus seems the right choice, especially for multimedia PCs.

2.1.7 Special Services, Tariffs

Services and applications are now available in all fields where communication is needed over great distances. The applications may be classified by the types of traffic carried and by the service fields that will use them. In the field of high-volume home-based personal services, the most important applications to telecommuters and the small offices is access to Internet services. Electronic phonebooks, travel agencies, home banking, and many other on-line services are great assets to small businesses.

The latest generation of private branch exchanges is capable of transparently rerouting phone calls through ISDN to remote users. Business people, for instance, can get calls forwarded to their homes when they are away from the telephone extension at the office. In case of an incoming call from a customer, the relevant business file can then be quickly brought up from the office database and onto the home computer or the videotext screen through the fast ISDN network.

Individual end-users with limited budgets will not be adopting ISDN quickly unless operators lower their service tariffs soon. Currently, prices vary widely from country to country, which may suggest that telephone companies have a good deal of leeway in what they can afford to do to make ISDN attractive to customers.

Installation of one basic-rate interface connection can cost US $25-$95 plus subscriber charges of $35-$120 per channel per month and per-hour usage charges per channel of $3-$6. It should also be noted that those Internet providers that now offer ISDN access typically charge four times as much as for a 56 kb/s modem connection.

All telecom operators are now offering useful services that require ISDN support for some features. In the PC market, access to such services ought to be standardized with basic application program interfaces to support commodity operating systems.

Efforts are underway to define standard interfaces of this kind for ISDN. All the same, several proposed APIs are competing to become the one and only standard recognized. Two of the proposals are the German ISDN industry's Common Application Programming Interface (CAPI), and Programming Communication Interface (PCI), which was first proposed by ISDN software vendors in France.

The German interface seems currently to be getting the most support from large vendors. Meanwhile, the European Telecommunications Standards Institute has adopted the other.

This type of conflict between two essentially similar proposals, both attempting to standardize the interface to ISDN software, is regrettable. In some cases, ISDN vendors have even implemented both versions in order to cover as much of their market as possible. It would be better if ITU-T, together with the leading vendors, would settle on one of these software interfaces as soon as possible.

Vendors have yet to grasp the full range of commodity services that can be offered through ISDN, let alone actually offer end users the full range of potentially high-volume services. Generally speaking, ISDN needs to find acceptance as the preferred means of access to multimedia applications through the telecom network. Yet this is still not the case even if telecom operators propose multimedia services.

The introduction of Europe's Vemmi may provide an opportunity to obtain greater appreciation of ISDN. Vemmi is a proposed communications standard for on-line multimedia PC services, targeted for use by both the European Telecommunications Standards Institute and the International Telecommunications Union. One of its main goals is to grant videotex terminals access to multimedia services. These terminals link a keyboard and a standard display screen to the phone network. Vemmi will boost the multimedia services and at the same time protect the large installed base of videotex equipment, especially in Europe and above all the Minitel network in France.

2.1.8 Factors inhibiting ISDN

To repeat, ISDN has been slow to take off. Telecom operators have not adequately shown the advantages of ISDN to potential users the service is not available in some areas and current tariffs are too high.

Besides, for a long time it has been ridiculously costly and difficult to complete product homologation procedures, the term for the testing of new products for network compatibility. Having a different form of homologation in each country fragments the market, so that each version of a given product is manufactured in smaller volume and is therefore higher in price.

Worse yet, the major countries and regions of the world have been slow to adopt generic international ISDN standards. In the United States, until recently, each switch vendor implemented ISDN in a slightly different manner, so that testing had to take place against each switch supplier before a product went live on the network.

The venerable TCP/IP protocol stack (Transmission Control Protocol plus Internet Protocol) has become the worldwide de facto standard, even for the expansion of local network traffic across wide areas. Most ISDN vendors offer their products with the support of these protocols.

Nevertheless, some industry experts have voted for the full implementation of a pure Open Systems Interconnect (OSI) protocol stack as the only way to ensure seamless vendor interoperability. One such program is already in place in Japan, after being pushed by the Promotion Conference for the Harmonization of Advanced Telecommunications Systems, an organization sponsored by the Japanese telecommunications industry and supervised by the Ministry of Posts and Telecommunications. This type of program will not benefit the acceptance of ISDN products, since for end-user applications have never accepted OSI as a commodity protocol stack.

2.2 SONET Networks

SONET stands for the Synchronous Optical Network. It is also associated with the well-accepted standards for the transport of pre-defined frames through the network. The SONET standards, originally proposed by Bell Communications Research, specify the standardized formats and interfaces for optical signals. Historically, the work started as far back as 1985 to provide standard optical interface between major carriers (such as NYNEX, MCI, Sprint, etc.) At the outset, the broad goals of the SONET standard are:

- To provide a broad family of interfaces at optical rates (i.e., OC-1 through OC-192)

- To provide easy and simple multiplexing/demultiplexing of the signals (such as headers and play loads) within the network components

- To exploit the growing trend toward the communication networks becoming synchronous

- To provide ample overhead channels and functions (via the header blocks) to perform and support the maintenance of the network facility

Total synchronism of all networks is unachievable and for this reason the standard committees have accepted the interfacing signals to be plesiochronous. The plesiochronous nature permits a small variation in the significant timing instants of the nominally synchronous signals, i.e., the recovered signals may be permitted to have small, but controlled, variations in their instantaneous rates or their instantaneous frequencies.

ITU-T and the European Telecoms in the 1987-1988, after considerable debate (9-row/13-row debate), about the number of rows in each of the frames established a standard. Subsequently, ANSI's committee for standard fiber optic transmission originally approved these standards. Some of the European Telecoms favored the 9-rows x 90 columns frame and the United States favored the 13-row x 60 columns frame. In February 1988, the Synchronous Transport Signal Level 1 was finally accepted with 9 rows x 90 columns and at a bit rate of 51.84 Mbps. At this rate, the time duration of the frame is exactly 125 μs and this duration is necessary to get 810 bytes (or octets with 8 bits per byte or octet) at a rate of 51.840 Mbps. Thus, one entire frame gets into the network in 125 μs (at OC-1 rate) and this is the same as the sample time of the Pulse Code Modulation (PCM) encoded speech of one byte (8 bits) at 8 kHz, thus maintaining the 125 μs clock duration for both the traditional voice circuits and the new SONET network.

The approvals by ITU-T have been granted with the complete standards issued in 1992. SONET standards are being incorporated for asynchronous transfer mode (ATM) in broadband ISDN applications for voice and data. The standards for fiber optic digital hierarchy range from the OC-1 rate of 51.84 Mbps through OC-3 (widely accepted) rate of 155.52 Mbps to the OC-96 rate at 4.97664 Gbps. More recently the specialized fibers fabricated with self-focusing properties have opened the OC-192 rate for possible applications.

SONET Transport	Optical Carrier	Data Rate	Signal (STS)
STS-1	OC-1	51.840 Mbps	X 1
STS-3	OC-3	155.520 Mbps	X 3
STS-9	OC-9	466.560 Mbps	X 9
STS-12	OC-12	622.080 Mbps	X 12
STS-18	OC-18	633.120 Mbps	X 18
STS-24	OC-24	1.244160 Gbps	X 24
STS-36	OC-36	1.866240 Gbps	X 36
STS-48	OC-48	2.488320 Gbps	X 48
STS-96	OC-96	4.976640 Gbps	X 96
STS-192	OC-192	9.953280 Gbps	X 192

Table 2.1 Rate Hierarchy for SONET Date Transport

Only the OC-1, OC-3, OC-9, OC-12, OC-18, OC-24, OC-36 and OC-48 are allowed by the American National Standard as currently accepted OC rates. The ITU-T has the STM-i designation, with i = *N/3* for the STS-N designation. The numerical value of i ranges *from* 1 to 64 corresponding to SONET rates from STS-3 to STS-192 with i limited to 1, 4, 16 or 64 (or with N = 3, 12, 48, or 192) for the currently accepted rates. The play load capacity for each STS rate is exactly 87/90 times the data-rate tabulated.

The multiplicative relation [1] between the hierarchical rates greatly facilitates the framing format for SONET synchronous data transport. The STS-N signal is scrambled and converted to the OC-N signal and this line rate is exactly N times the OC-1 rate. Multiplexing of the STS signals needs intricate byte encoding in the header blocks, and this becomes a crucial concern in the "SONET ready or compatible" switches that link [1] various types of networks (such as DS-1, DS-3, all-digital synchronous networks, all optical networks, etc.) via the Digital Cross Connect Systems (DCS). The standard 125 μs time slot is chosen for all the hierarchical OC rates. This leads to simplistic calculations for the number of bytes for the OC-1 through OC-192 rates as follows:

Number of bytes (or octets) in a 125 μs for the OC-1 rate is

$= 51.84 \times 1.E+6 \times 125 \times 1.E-06 / 8$

$= 6{,}480$ bits / 8 or 810 bytes

$= 9 \times 90$ bytes (i.e., 9 rows × 90 columns)

$= 9 \times (3+87)$ bytes

= (9 rows) × (3 bytes OH + 87 bytes PL) /frame. [1]

Here OH signifies overhead for the SONET header block and PL indicates play load for carrying the customer data. The data transport function uses all the 9x3, or 27, bytes for overhead and the 9x87, or 783, bytes for play load. The ATM network retains the byte integrity as the rows are transmitted one by one. It is thus implicit that when the network is operating in its normal synchronous mode, the 8 sequential bits are also the bits of the same byte. The byte (or octet) classification for other rates follows the same pattern. For example, for OC-N rate the byte classification is as follows:

Number of bytes (or octets) in a 125 μs for the OC-N rate is

$= 51.84N \times 1.E+6 \times 125 \times 1.E-06 / 8$

$= 6{,}480\,N$ bits / 8 or 810 bytes

$= 9 \times 90N$ bytes

$= 9 \times (3N + 87N)$ bytes. [1]

The breakdown between the overhead and the play load is similarly segmented with 27 x *N* bytes for overhead and 783 x N bytes for play load. The arrangement of the SONET frame can be simply derived from the considerations listed above. For example, the SONET frame for the commonly used OC-N rate is shown in table 2.1.

The starter frame Synchronous Transport Systems (STS) is of an *N-th* level STS-N frame needs special attention. Byte interleaving and frame alignment are both necessary to form the STS-N frame. Byte integrity at the OC-3 level (to meet the ITU-T standards) is thus implemented; and it also permits STS-N to carry broadband play load at 150 and 600 Mbps. Multiplexing of various SONET frame and digital cross connects are handled by the SONET Central Offices.

Here, the various services such as T1, T3, HDSL (High-speed Digital Subscriber Line), VHDSL (Very High Speed Digital Subscriber Line), and other STS-N (Synchronous Transport Signal Level N) based services are reduced to the common basis of their synchronism (or the 125 μs), appropriate switch and then the information is reconstituted in the appropriate signaling format and sent to the customers. The coexistence of broadband ISDN and ATM was proposed as early as 1990 via the ATM cell mapping into the SONET play load for easy, elegant transport via the optical information highways of the future. This will be necessary until all the systems assume one standard format, preferably the ATM format (see next Section), and then the ATM switch in conjunction with the bridges and routers will perform all the switching. Meanwhile, a number of the digital cross connect systems, such as AT&T's DDM-2000, FT-2000 SONET multiplexers and many versions of Digital Access Cross connect Systems (DACS) and SONET DCS are available to perform high-speed optical-rate switching. Newer services such as broadcast TV facilities, interfacing HDTV (High Definition Television) quality Computer Aided Design/Computer Aided Manufacturing (CAD/CAM) terminals, virtual reality devices, medical image transfer, super-computer applications, etc. will be the prime beneficiaries of these DCS and DACS switches.

2.2.1 SONET Architectures

The deployment of SONET occurs in numerous architectures and applications. The broadband network uses the point-to-point linear deployment for access and interoffice capabilities, integrated access, tree configurations with limited DS-1, DS-3 switching, and hub sites architectures with extended switching.

Rings using Add-Drop Multiplexers (ADMS) can also be constituted for broadband data traffic, and hubs with Digital Cross Connect Switch (DCS) provide a localized broadband switching capability. The DCS are generally built with wideband (DS1 and VT) and broadband (DS3 and STS-1) interfaces. SONET interfacing with both the WDCS and BDCS is already available. The virtual terminals (VT) have two modes of operation: floating and locked. The floating mode is suitable for bulk transport of channelized or unchannelized DSN and for distributed VT grooming. The locked VT mode is better suited for integral numbers of STS-1s transport and for distributed DSO grooming between DSO paths terminating equipment.

2.2.2 Deployment of SONET

SONET deployment in phases can enhance the capacity and the switching of the modem information highways quite efficiently and inexpensively since the ADMs and DCSs are relatively inexpensive compared to the traditional ESS (Electronic Switching System) type of switches. SONET may also be deployed for DS-1 and DS-3 transport in the trunk environment using the Bellcore developed TIRKS (Trunk Signaling System) provisioning system. SONET is also a contender in the Integrated Digital Loop Carrier (IDLC) environments. The IDLC applications can provide access to the DS1, baseband, broadband, (OC-3) services from any broadband network and Local Digital Switch (LDS), thus facilitating broadband and baseband networks to coexist via ADMs and remote data terminals.

Low Power Complementary Metal Oxide Semiconductor (CMOS) technology devices at OC-3 and higher rates make these architectures easy to implement for generic and programmable data distribution systems to reach homes, businesses, and special services customers. SONET offers direct mapping of the SONET frame to the byte synchronous DSO for encoded voice/speech applications.

This advantage in the IDLC environment removes the need for DS1 realignment, thus offering these carrier systems better utilization of the switching facilities. The floating VT modes of the Remote Digital Terminals (RDTs) make feasible distributed VT cross connections, bulk transport/switching of channelized and unchannelized, synchronous or asynchronous DS1s (mapped asynchronously), unchannelized bitsynchronous DS1 transport, DSO circuit-switched traffic and IDLC byte synchronous mapping.

These architectures indicate the strategies for the incorporation of the evolving SONET and ATM components into the existing framework of telecommunication networks. Over a period of time the older network rates such as DS1 through DS3 will be substituted by the more accepted OC-N rates.

2.2.3 SONET Frame

The optical rates currently available (up to 4 to 8 Gbps) for the transport of data are much too high for the existing telecommunication networks to directly integrate them. The SONET frame, lasting for 125 μs, can carry a large amount of information as the playload (2,349 bytes) even at the OC-3 rate. While this size may appear perfectly reasonable in the LAN/MAN/WAN data environment, it can be perceived as being too large for other similar Switched Virtual Service (SVC), or Permanent Virtual Service (PVC), types of application in business and industry. In the trunk environment, numerous low rate data channels are multiplexed onto the fiber trunk with its own particular protocol, framing and bit designations for rates ranging from T0 to T3 or E0 to E3. However, recent applications are not always circuit switched, but are becoming more and more packet and message switched. Hence, the idea of having the SONET carry smaller size packets or "cells" becomes more appealing. In a sense, Asynchronous Transfer Mode (ATM) is an extension of the signal transport concept that SONET itself relies upon. Cells provide user level interfacing packets and SONET frames become the photonic network transport packets. ATM standards outline the efficient (and perhaps elegant) techniques to bridge the gap between multiplicity of users and the vast variety of lightwave devices that make up the network.

However, until such networks are a reality, ATM can be used to supplement DS-2 and DS-3 networks.

ATM of data transfer is a worldwide standard. It stems from the availability of a high-speed backbone in the national or global environment. It has started an era of cell relay technologies and falls under a wide umbrella of packet switching. The technological fit of ATM can be seen as a progression of X.25 technologies to frame technologies and then into the cell technologies. It also depends upon the SONET standard for the transport of data. ATM is a standard for most networks to directly communicate with each other provided the SONET and ATM protocol and header block information is consistently followed.

ATM is based upon packet switched technology. The ATM packets or cells are relayed and routed throughout the network and thus get communicated between the source to the appropriate destination. Since the protocols and functions are unique to ATM, the technology for these ATM based networks is called the "cell relay" technology as opposed to the "frame relay."

ATM functionality has been embedded within a fixed size link-layer (i.e., the OSI layer six, data link-layer) entity. This link-layer entity is the "cell,' and the size of this fixed length entity is 53 octets (bytes). There are 5 octets for the header block and 48 octets for the playload (or the data to be transported). Similar to the header of most packet networks, ATM cell header carries information unique to the nature of data, its functionality and disassembles and reassembles cells and their playloads to ATM Based Cell Relay Technologies.

ATM cells are routed and relayed by virtual circuits. Such circuits may be fabricated, depending upon the traffic demands, and switched and scaled by the ATM switches. The ATM switches and/or nodes may also switch these virtual networks in the hardware.

ATM also offers high-speed data transport capabilities. Since the ATM cells "ride" within the SONET frame, they carry all the advantages offered by transport systems embedded within SONET. The speed of transport is derived from optical rates (and thus low buffering delays), optical speeds and low switching times (because of the elegant simplicity in the SONET header information).

There are five functions at the Transmission Convergence Sublayer (TCS) sublevel:

- Cell rate decoupling
- Header error check (parity) header sequence generation/verification

- Cell delineation
- Transmission frame adaptation
- Transmission frame generation/recovery

Each of these functions is tailored to fit the cell structure (5 octet header plus 48 octet playload) and the "frame" that carries these cells. Though tailored to map efficiently into to SONET/SDH frames, these cells may also be mapped into DS-3 or even DS-1 frames.

2.2.4 SONET Signals

SONET can carry data from a variety of existing carriers such as DS-1, DS-3. It is also well suited to carry the ATM cells. However, when the data/cells arrive, the information generic to the data type or the cells has to be made available to the terminal, SONET repeaters and multiplexers. This information is mapped into a new header by three new overhead blocks: the path overhead, the line overhead and the section overhead.

Accordingly, the data/cells flow into the path layer that generates the path overhead, the line layer that generates the line overhead, and then the section layer that generates the section overhead. It is the data/cells that are packed into the SONET frames with an additional header having these three (section, line and path) overhead blocks. This data structure becomes the

SONET signal and it is ready to be carried by the photonic layer of the terminal, the SONET repeater or the SONET multiplexer.

2.2.5 Impact of SONET and ATM

The impact of the SONET and ATM on the older modes of digital communication is profound. In a sense, the newer fiber optic networks and the ATM switches break into the architectures. Typically, the DSN (n=2, 3 or 4) information is absorbed into 48 byte ATM cells and the DSN signaling is re-encoded into the 5-octect header of each ATM cell. Similar remapping occurs from the DSN signals to the SONET frames. The cheaper and more versatile ATM switches replace the bulkier cross connects. Still extreme care and attention is necessary to determine the viability of such conversion in view of the application at hand. Chaos is likely to clutter the customer information especially if real-time multimedia or video information is being carried by the DS-n signals.

Ideally, the composite architecture of the telephone network would evolve as the all-digital broadband facilities and the circuit/packet switches would become integrated as a fast packet relay technology frame. Whereas the private line, software defined networks, ISDN and switched-digital networks need to keep their identity for reasons of service requirements, it is possible to see other packet-oriented digital and computer networks being unified into the ATM backbone network. Routers or ATM hubs accomplish these functions. The ATM cross connect (AXC) switch permits the network both to switch at very high rates (in the order of tens of Gbps) and to route the cells to the appropriate address. It is also feasible to foresee the ESS facilities being integrated with AXC to makeup the universal switch architecture.

A new broadband network will evolve with both network service capabilities: traditional digital hierarchical (DSO-DS4) network and the SONET synchronous digital hierarchy WAN. This broadband switch would be the key integrator of almost all digital network services with one unified voice/data/video services network.

ATM to the desktop is more than a proposition. ATM adapters at 155 Mbps full duplex campus networks bring in more bandwidth and flexibility. The other alternatives such as Ethernet, full-duplex Ethernet or token ring, fast Ethernet (i.e., 100 Mbps half duplex mode), or FDDI can only become obsolete in a few years after the ATM cross connects and adapters, routers and brouters become common and inexpensive.

The OC-3c, 155 Mbps adapters uses category-5 unshielded twisted wire pair (UTP-5) media compared to the more frequently used fiber media for ATM

LANS. Complete standardization can only bring down the cost and point-to point ATM is likely to bring the information highway to the desktop in the next few years.

But more than that, wireless personal communication service (PCS) networks permit portables laptops to exchange data at rates ranging from 2 to 20 Mbps. They may be also connect via a server or peer-to-peer to a Portable Base Station (PBS).

The link rate between these PBSs is in the order of a few Gbps by using the cell relay ATM technology. Typical rates of 1.2 Gbps over radio links or free-space optic (line of sight optical) links have been functional in the laboratory environment. Finally, a gateway-PBS permits the access to and from the wireless ATM LAN to the wired/fibered/cabled LAN, MAN or WAN. Personal and mobile communication networks (PCNS) at multi-megabits can be elegantly switched via wireless ATM networks since ATM supports the Switched Multimegabit Data Service (SMDS) capability. Indoor, wireless ATM LANs complete the promise of multimedia services to portable "plug-n-play" laptops.

On a realistic note, most rate requirements are modest. For example, a word processing application with 6400 byte text object retrieval needing 30%

transmission time, will need a peak data rate of 0.7 Mbps to still "get" the object in less than 250 (i.e., 244) msecs. CAD/CAM applications with 50 kB manufacturing objects need about 5 to 6 Mbps, and a fairly detailed 144 kB office image needs 15 to 16 Mbps to meet similar object retrieval objectives over LAN networks. In computing environments the requirements are easily estimated. For example, a 200 bytes message (1,600 bits) is furnished in one-tenth msec (10^{-4}) over a 16 Mbps link. Time for file transfers may be calculated quite readily at different rates but the slowest element is generally the human response.

Very detailed X-ray images for real-time diagnostic analysis need about 45 Mbps. Transmission of HDTV quality, imaging applications with Motion Pictures Expert Group (MPEG++) video compression techniques need about 18 Mbps. With National Television Standards Committee (NTSC) TV quality of pictures (3 to 4 Mbps rate) on personal computers, video conferencing with 6 coworkers needs a 25 Mbps ATM LAN.

2.3 ATM Networks

Asynchronous Transfer Mode (ATM) is changing the way we think about traditional data communications. Although ATM originated in the telecommunications industry, its application to datacom networks is more imminent. What started as simply another proposal for moving signals across B-ISDN networks, has turned into a core networking technology - a uniform method of transporting, multiplexing, and switching all types of traffic at unprecedented speeds. ATM can be integrated into every facet of networking - from LANs to WANS, from the backbone to the desktop. To appreciate the magnitude of ATM's impact, we will first review current datacom and telecom networks before looking at the development of ATM.

2.3.1 Datacom Networks

Various factors influence the changes in the datacom industry. The most significant factor is the enhancement in computing capabilities. Desktop computers provide processing power available only in mainframes just a short time ago. They support full color images, complex 3-D graphics, and some full-motion video. These new capabilities coupled with rapidly dropping prices impact datacom networks in many ways.

The number of network users is exploding. Users are moving to new classes of applications, which use visual imagery to facilitate processing, and presentation of complex data. These new applications are more bandwidth intensive. Examples include visualization to model complex real-time phenomena; multimedia integration of voice, video, and data; and electronic document imaging. All these enhancements are leading users to expect more access to bandwidth-on-demand.

All of these changes are creating an interesting problem for traditional data networks. Popular LAN protocols are based on a shared media concept. While this makes network design easier and less expensive, it does so only to a portion of the total network bandwidth to each node.

The bandwidth available to each node is fast becoming insufficient to support the growing number of users and their new applications.

The datacom industry has responded to the challenges by creating smaller, more highly segmented networks (micro-segmentation) using bridges and routers to interconnect the many segments. This was a logical approach because it allowed the use and maintenance of existing technology and products. However, carving up the network in this manner is an interim solution that will not keep up with the increasing bandwidth demand. The network must evolve into a more flexible and more efficient network - one that can deliver both the bandwidth and the services required to fully support emerging applications.

2.3.2 Telecom Networks

The telecommunications industry is also facing significant challenges. The shift from analog to digital switches and increased availability of fiber optic cable have dramatically changed the nature of these networks - allowing an increase in the size of the networks, the number and kinds of services that can be offered, and the quality of service. In turn, the greater volume of non-voice traffic has significantly increased the overall bandwidth demand.

The telecom industry has responded to the challenges by upgrading to faster, digital switches and consolidating current transmission lines over the higher bandwidth fiber media, still, these improvements do not solve all of their problems. For one, carrying non-voice traffic over current telecom networks is very inefficient. And two, customers are often required to purchase more dedicated bandwidth than they use most of the time.

2.3.3 ATM Standardization

The standardization process for ATM is both similar to and different from that of a typical data network standardization process. It is similar because international committees are developing it with members from a multitude of companies. It is different because the process involves two very large and distinct industries - datacom and telecom - each accustomed to different processes, rules, regulations, and timing. As a result of these similarities and differences, several organizations have been formed to assist in the development and standardization of ATM technology.

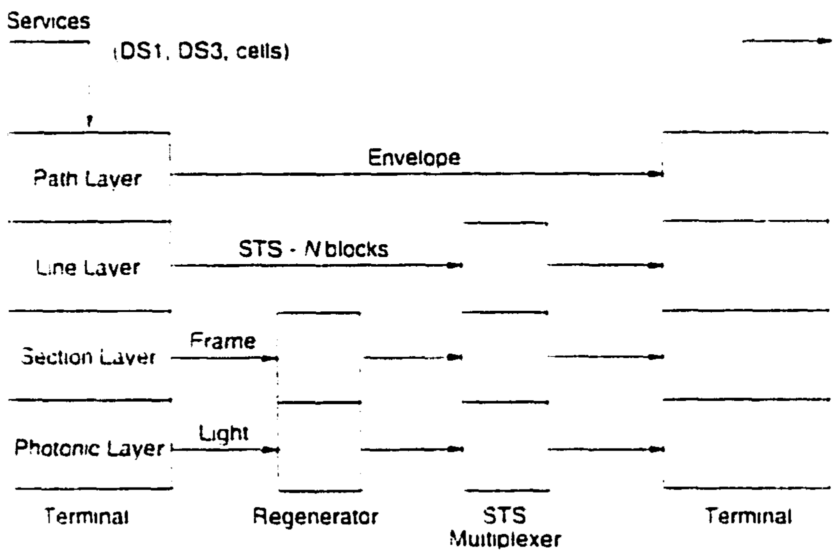


Figure 2.1. Logical Hierarchy of DS-1, DS-3 and ATM Cells mapped into The STS Signal (Frame)

2.3.4 ATM Forum

The ATM Forum was founded to ensure interoperability between public and private ATM implementations. It is focused on accelerating the use of ATM products and services through the convergence of standards, interoperability demonstrations, and promotion of industry cooperation and awareness. This international consortium has grown quickly to include more than 400 member companies representing a variety of industries and perspectives. It is comprised of public and private network equipment vendors, datacom and telecom network service providers, software companies, consultants, and end users.

Early 80s	CITT adopted Narrowband Integrated Service Digital Network (N-ISDN). - Basic Rate = 144 Kbps - Primary Rate= 1.544 Mbps
Mid 80s	Study Group formed to develop Broadband ISDN (B-ISDN). Study Group goals: Define technology to support multiple traffic types Define technology scalable to Gbps bandwidths Define efficient and flexible methodology for bandwidth allocation
1988	ATM chosen as the core technology for B-ISDN
1991	The ATM Forum founded
1992	The ATM Forum released UNI V2. 0 draft standard
1993	IETF published RFC 1483, Mulitprotocol Encapsulation over ATM Adaptation Layer 5 The ATM Forum released UNI V3.0 and DXI VI.O draft standards
1994	IETF published RFC 1577, Classical IP and ARP over ATM The ATM Forum publishes B-ICI VI. 0 draft standard
In Progress	The ATM Forum Sub-Working Groups: P-NNI Signaling, Physical Layers, LAN Emulation, Traffic Management, Service Aspects and Applications, etc. continue defining implementation agreements

Table 2.2. Milestones in the Standardization of ATM

The ATM Forum works with the ITU-T and other standards bodies to accomplish these objectives and to offer guidance and knowledge when appropriate. The ATM Forum has provided rapid and significant assistance in developing the definitions for the fundamental standards.

Other organizations focused on ATM include the Enterprise Network Roundtable (ENR) within The ATM Forum, dedicated to development and communication of user requirements between The ATM Forum technical committee and users, and the Internet Engineering Task Force (IETF) Working Groups, which make contributions to the evolution of the Internet and its technologies. There are three IETF workgroups focusing on particular aspects of ATM technology.

	Datacom	**Telecom**	**ATM**
Traffic Type	Data	Voice	Data, Voice, Video
Transmission length and Unit	Variable Packet	Fixed Frame	Fixed Cell
Switching Type	Packet	Circuit	Cell
Interval	Bursty	Periodic	Periodic
Timing	Indeterminate	Time-Sensitive	Some Time Sensitive
Connection Type	Connectionless	Connection-Oriented	Connection-Oriented
Delivery	Best Effort	Guaranteed	Defined Classes
Media & Rate	Defined by Protocol	Defined by Channel	Scalable to Application
Access to Media	Shared	Dedicated	Dedicated

Table 2.3. Technology Comparison

2.3.5 ATM versus other Technologies

ATM is significantly different from the popular datacom/LAN and telecom/WAN technologies in use today. ATM technology unifies these worlds by borrowing attributes from both. Technologies used in data communications - such as Ethernet, Token Ring, FDDI, X.25, and Frame Relay - employ variable length packets for carrying data. They generate variable length traffic at irregular "bursty" intervals, which can tolerate indeterminate delays or latencies. Technologies used in telecommunications employ small, fixed-size frames for carrying voice. They generate uniform length data at regular intervals, which is very time-sensitive. Time division multiplexing (TDM) is a good choice for supporting this kind of traffic.

ATM technology supports multiple kinds of traffic, including data and voice. To do this, it employs small, fixed-size cells, cell switching, and bandwidth-on-demand. An ATM connection's route, medium and transmission rate is based upon the specific application.

2.3.6 ATM Key Characteristics

Let's look more closely at the key characteristics of cell switching, connection-oriented transmission, and media and rate independence.

2.3.6.1 Cell Switching

ATM's fixed-size cells contain only 53 bytes. This is relatively small when compared to typical LAN packets, which are often several kilobytes in size. The 5-byte header contains identification, control, priority and routing information. The 48-byte payload contains data. Because ATM cells are small and fixed in size, the delay between cell transmissions (or latency) is reduced greatly compared to typical LANS. And because cell transmission is asynchronous, ATM cells can send delay-tolerant traffic (like data) intermixed with time-sensitive traffic (like voice and video).

ATM's use of cell switching enables scalable user access to the network, from a few Mbps to several Gbps as appropriate to the application. And the fixed-size cell format enables ATM cell switching via high-speed hardware as opposed to the slower software or firmware used by traditional LAN routers.

2.3.6.2 Connection-Oriented Transmission

ATM is fundamentally a connection-oriented technology, meaning that a connection needs to be established between two stations before data can be transferred between them. An ATM connection specifies the transmission path, allowing ATM cells to self-route through an ATM network. Being connection-oriented also allows ATM to specify a guaranteed quality of service (QoS) for each connection.

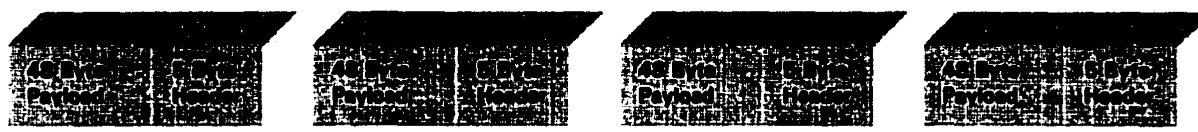
By contrast, most LAN protocols are connectionless, which means that nodes simply transmit data when they need to, without first establishing a specific connection or route with the destination node.

Since ATM is a connection-oriented protocol, bandwidth is allocated only when an end station requests a connection. This allows ATM to efficiently support a network's aggregate demand by allocating bandwidth-on-demand based on immediate user need. This allocation can be accomplished without administrative intervention.

2.3.6.3 Media and Rate Independence

By definition, the ATM protocol is not tied to a particular transmission rate or physical medium. It can operate at whatever rate is appropriate for whichever physical layer technology is being used.

Features	Benefits
Cell Switched Small, fixed size cells 53 Bytes	Scalable to high speeds Low Latency Multiple traffic types supported
Connection Oriented Connection setup required Dedicated Channel	Simplified switching/routing Bandwidth on-demand Defined quality of service for data delivery
Media and rate Independent Mbyte speeds Gbyte speeds	One technology supports a variety of user requirements Scalable price vs bandwidth curve



ATM Cells

Figure 2.2. ATM Features and Benefits

2.3.7 ATM's Impact

We began this discussion by comparing datacom and telecom networks to ATM. ATM works for data networks because it offers reasonable access time, the ability to send short or long messages, and the capability for higher speeds, auto-routing, and support for emerging applications. ATM works for voice communication because it offers guaranteed access intervals, quality of service definitions, and support for new applications. Traditional network users can benefit from integrated networks as they:

- Adopt multimedia and other emerging technologies while protecting existing network investments
- Use the network when needed at the rate needed instead of buying extra bandwidth to cover every worst case scenario
- Utilize available bandwidth more effectively by intermixing traffic types and services
- Improve network performance due to scalable services in combination with faster, hardware-based switching equipment
- Converge Local and Wide Area Networks

These capabilities make ATM the most flexible integrated general networking standard to date.

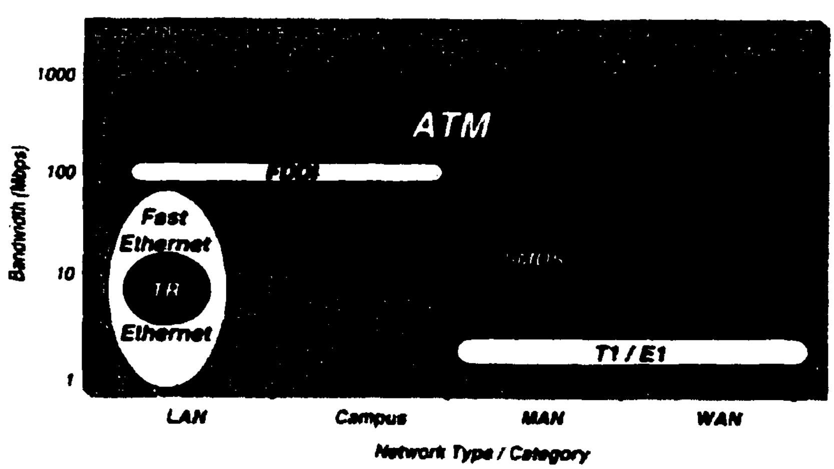


Figure 2.3. ATM Solves Your Networking Needs

2.3.8 ATM's Protocol Suite

Like other data communication protocols, communication from higher layers is adapted to these lower ATM defined layers. The ATM protocol layers then pass the information onto the physical layer for transmission over a selected media. Note that the ATM protocol model is further divided into sub layers.

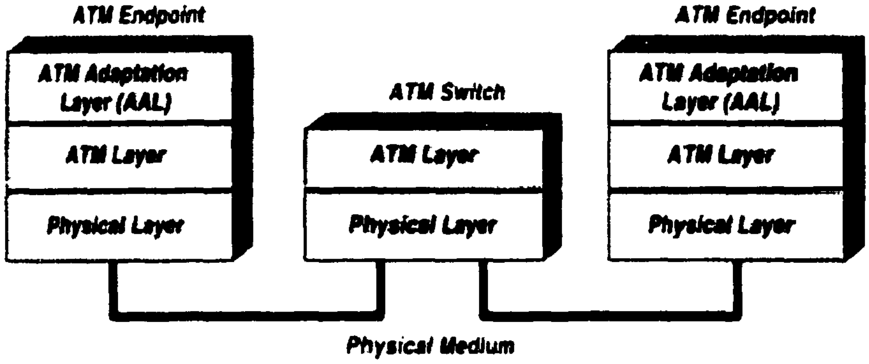


Figure 2.4. ATM Protocol at work

2.3.8.1 The ATM Adaptation Layer

The ATM Adaptation Layer (AAL,) interfaces the higher layer protocols to the ATM' *Layer,* generally at endpoints. Thus, it relays cells both from the upper layers to the ATM Layer and vice versa, when the AAL receives relaying information from the higher layers to the ATM Layer, the AAL must segment the data into ATM cells. When relaying information received from the ATM Layer to the higher layers, the AAL must take the cells and reassemble the payloads into a format the higher layers can understand. This is called Segmentation and Reassembly (SAR). There are five types of AALs, which support the different types of traffic or services expected to be used on ATM networks.

2.3.8.2 The ATM Layer

The ATM Layer interfaces between the AAL and the Physical Layer. It is responsible for relaying cells both from the AAL to the Physical Layer for transmission and to the AAL for use in an endpoint after receiving cells from the Physical Layer. How the ATM Layer performs this function depends on whether it is inside an endpoint or a switch. If the ATM layer is inside an endpoint, it receives a stream of cells from the Physical Layer and simultaneously transmits either cells with new data or empty ("idle") cells, if there is no data to send. In a switch, the ATM Layer determines where the incoming cells should be forwarded to, resets the corresponding connection identifiers for the next link, and forwards the cell. The ATM Layer also handles traffic management functions and buffers incoming and outgoing cells. It indicates to the AAL whether or not there was congestion during transmission. The ATM Layer monitors both the rates of transmission and conformance to the service contract – called traffic shaping and traffic policing.

2.3.8.3 Physical Layer

An ATM-specific physical layer is not defined. However, proper ATM operation requires an interface to existing physical layers defined in other networking protocols. Because the ATM Layer does not depend on one particular Physical Layer, it is possible to build ATM networks on a variety of physical interfaces (or media types) including, but not limited to, the fiber based Synchronous Optical Network (SONET) specifications developed by the telecom industry.

2.3.9 ATM Network Interfaces

ATM is a connection-oriented, cell-switching technology that can use many different physical layers and can operate at speeds from a few megabits to several gigabits per second. ATM will also support many different types of traffic, including emerging multimedia applications, and allow allocation of different amounts of bandwidth-on-demand. Implementing an ATM network will require well defined ATM interfaces, signaling protocols, addressing schemes, routing schemes, traffic management, network management, and mechanisms for integrating legacy network technologies and their applications to name a few.

ATM networks have the following key components: ATM network interface cards for workstations, ATM network interfaces for legacy LANs, one or more ATM switches to relay each cell to its destination, and intelligent entities in the network to help setup connections and determine their routes. In a large network, these components typically will be connected together in a combination of collapsed backbones and mesh topologies over local area and wide area networks. The kind of information that will be passed between LANs and WANs and between users and local switches will be different. As such, there are several different ATM interfaces defined.

Current regulations forbid American public and private networks to exchange information other than that required for connection setup. Consequently, public and private ATM interfaces will exchange very little information and may have different implementations within their own captive environments. In fact, public and private ATM interface implementations are being defined by different sub-working groups within The ATM Forum.

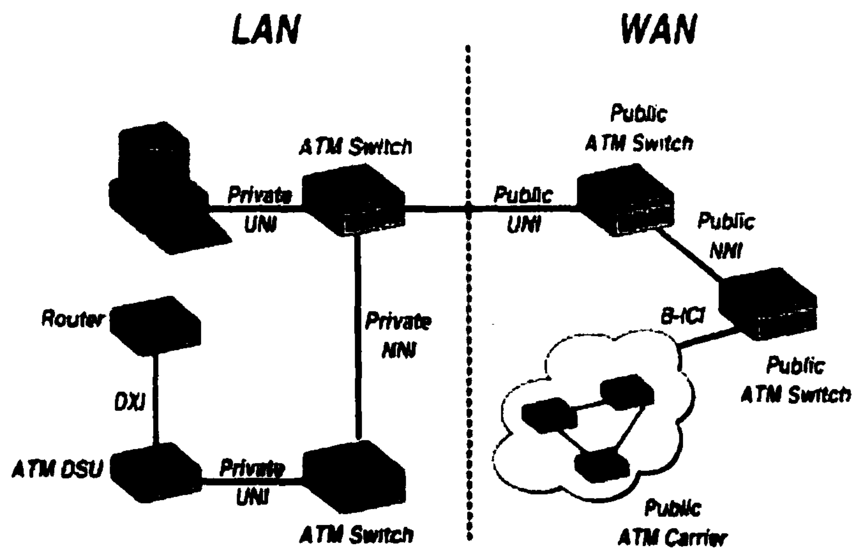


Figure 2.5. ATM Network Interface

2.3.9.1 User-to-Network Interface (UNI)

The interface between an ATM switch and an ATM endpoint is called a User--to-Network Interface (UNI). Ironically, where the customer's network crosses the regulatory boundary into a public network is also called a UNI. The terms Public and Private UNI have been adopted to differentiate between the two types of UNI, where Public UNI refers to the interface between a customer's ATM endpoint or switch and a public ATM network.

2.3.9.2 Network-to-Node Interface (NNI)

As the information passed between a user and a switch via the UNI is simpler than the information required between two switches, a slightly different interface was defined for switch-to-switch connections. Switches, for example, will need to update each other on the utilization of their respective links and participate in network-wide connection setups to support an endpoint's request. This interface is called the Network-to-Node or Network-to-Network Interface (NNI).

The NNI originally described the interface between two public ATM switches. Currently, Private NNI (P-NNI) is used to describe the ATM switch-to-switch interfaces on a customer's premise and Public NNI describes the interface between public ATM switches.

Public NNI is also known as the Inter-Switching-System-Interface (ISSI), which is further categorized into the Public and Private ISSI. The Public ISSI can be further categorized into the Intra-IATA ISSI and Inter-LATA ISSI. In an Intra-LATA ISSI, both of the connected public switches belong to the same Regional Bell Operating Company (RBOC). Inter-LATA ISSI describes the link between the ATM switch of an RBOC and that of an inter-exchange carrier such as AT&T, Sprint, or MCI. Inter-LATA ISSI is now called Broadband Inter-Carrier Interface (B-ICI).

2.3.9.3 Data Exchange Interface (DXI)

The DXI describes a local interface between packet-based routers and ATM capable Digital Service Units (DSUs). Typically, the DXI port and the High-Level Data Link Control (HDLC) port of a legacy router can be connected. The router would then pass packets, encapsulated as defined in RFC 1483, to the ATM DSU. The DSU would then perform cell-level processing, Segmentation and Reassembly (SAR), and virtual connection termination. This allows legacy routers to connect into ATM LANs via the UNI interface of the ATM DSU.

2.3.10 ATM Connection Types

Since ATM is a connection-oriented technology, a connection has to be established between two ATM endpoints before any data stream is sent. An ATM connection is established as either a Permanent Virtual Circuit (PVC) or as a Switched Virtual Circuit (SVC). A PVC is usually set up by a network administrator and will remain up until it is torn down manually. An SVC can be setup and torn down by an ATM network on demand without manual intervention.

As ATM networks will be able to set up SVCs without intervention why would anyone use PVCS? PVCs may be more reliable for specific applications. For example, an SVC supporting a mission-critical, distributed application could time out and tear down after a brief period without traffic between resources. Later, when the application tries to request a replacement connection, the network may not be able to provide the same guaranteed bandwidth it had previously allocated to that application.

Each connection's route (from the requester to the desired destinations) is determined only once. Afterwards, cells (from the source to the destinations) self-route through the network based on the connection identifiers in their header. This is considerably different from today's LANS, where devices such as routers must calculate/determine every packet's route through a network.

2.3.10.1 Connection Identifiers

There are two connection identifiers in the ATM cell header: the Virtual Path Identifier (VPI) and the Virtual Channel Identifier (VCI). They form a two-level hierarchy where multiple VCIs can be aggregated into a single VPI value. ATM network operation benefits from this arrangement because it can more efficiently manage, route, and allocate network resources to connections with similar requirements if they are bundled together in one virtual path.

It is important to note that VPI and VCI values, or both together, do not represent addresses. While only one unique VPI/VCI combination is allowed to pass through a given switch, these individual values could be reused elsewhere in the network. In fact, the same VCI value could appear in different cells at the same switch if they were aggregated in different VPIs. Reusing VPI and VCI values allows ATM to support large networks while keeping the coni-iectioii identifiers short enough so switches can relay cells with minimal latency.

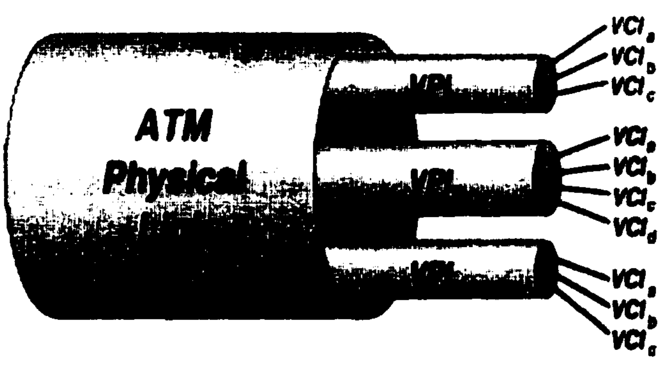


Figure 2.6. VPI / VCI Trunking

2.3.10.2 Cell Switching Using Connection Identifiers

An ATM switch will read an incoming cell's VPT, VCI, or both, and then determine the correct output port to which to relay or switch the cell based on its lookup table. (The look up table is updated when connections are established.) The switch must also determine what the new VPI or VPI/VCI values should be and substitute these new values into the cell's header.

A switch, which only reads, and substitutes new VPI values Is often referred to as a Virtual Path Switch (VP Switch) and a switch that reads and can substitute both VPI and VCI values is referred to as a Virtual Channel Switch (VC Switch). A VP Switch is analogous to a central office telephone switch in which trunk lines with a number of voice channels are grouped together and switched between several other central offices closer to the destination's voice channel(s). While an ATM network will probably contain several VC Switches, VP Switches could be used in the backbone where a number of connections are aggregated and switched together as one unit.

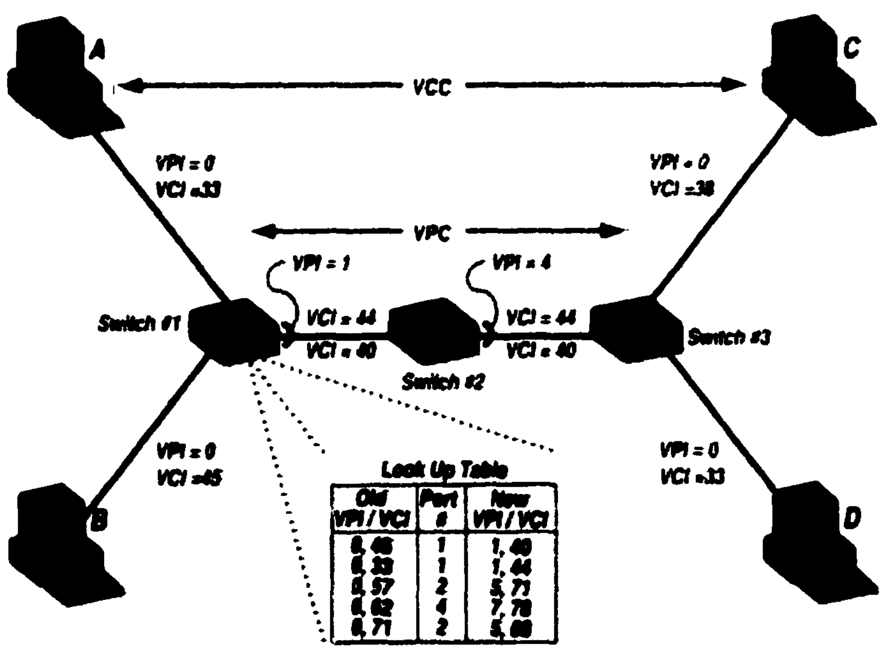


Figure 2.7. Cell Switching in ATM Networks

Figure 2.7 illustrates cells self-routing through the network: When switch #1 receives a cell with VPI=0/VCI=33 in port #3, it compares this to its look up table, assigns VPI= 1/VCI=44 for the cell header, and switches the cell to port #1. In switch #2, the process is repeated except that only the VPI value is read and reassigned because it is a VP switch. Finally, all the switches between A and C repeat the same process until the cell arrives at its destination. While ATM standards do not require VPI/VCI values to be reassigned anywhere in the network for a given connection, dedicating VPI/VCI values to each connection would not be possible in larger networks.

2.3.10.3 Virtual Channels and Paths

There are two basic types of connections in ATM networks: Virtual Channel Connections (VCC) and Virtual Path Connections (VPC). A VCC is a concatenation of virtual channel links extended between two ATM points where higher layer protocols are accessed, such as IP or a workstation's application. By definition, the cell sequence must be preserved over a VCC. On the other hand, a VPC is a concatenation of virtual path links between the points where a VCI value is first assigned and the point where that VCI value is either reassigned or terminated. In Figure 2.7, the VCC between station A and C is supported by a VPC between switches #1 and #3.

2.3.11 Making an ATM Connection

To set up an SVC, an ATM endpoint must request a connection setup via the UNI using a signaling protocol. The ATM network passes on this request to the destination using the destination's address, specified in the signaling request. If the destination agrees to form a connection with the requester, a virtual connection is set up across the network between the two endpoints. VPI/VCI values are assigned for both UNIs and the look-up table in all the switches in the network that will support the connection are updated to support their respective NNI connections.

An ATM endpoint may also request a specific Quality of Service (QoS) from the network when requesting its connection. For example, the QoS parameters quantify the desired bandwidth, the traffic type in the cell's payload, the maximum acceptable jitter, and the cell loss priority. A connection may also be requested without a specific QoS, which the network will support on a "best-effort" basis.

Actually, an ATM end-station may establish several virtual connections where each one carries different types of data and has different amounts of bandwidth and priorities allocated to it. In other words, each virtual connection could have a unique QoS, which was negotiated and agreed to during connection setup. To illustrate: a physical link to an ATM endstation could simultaneously support several virtual connections, such as a TCP/IP-based distributed computing application, a video conferencing session, a session receiving a different broadcast video, and another supporting an occasional email.

2.3.11.1 UNI Signaling

The ATM Forum finalized an initial UNI signaling implementation, which was called Q.93B, in September 1993. It has since been renamed Q.2931 by the ITU-T Q.2931 protocol borrows many attributes from the Q.931 signaling protocol used in Narrowband-ISDN (N-ISDN). Information included in a Q.2931 signaling request includes the source and destination ATM addresses and parameters for a desired QoS. Either the ATM network or the destination endpoint can accept or reject the request for a connection.

The Q.2931 protocol operates over its own Signaling ATM Adaptation Layer (SAAL), which includes a Service Specific Convergence Protocol (SSCOP) similar to a data link layer protocol to ensure reliable delivery of signaling requests. For example, the SSCOP preserves cell sequence integrity, ensures retransmission of lost signaling Protocol Data Units (PDUs), and performs "keep alive" functions.

The basic capabilities supported by Phase 1 of the UNI Signaling specification include:

- VPI/VCI Assignment
- Public and Private Addresses
- Point-to-point Connections
- Bi-directional Point-to-Point Connections with symmetrical or asymmetrical service requirements
- Unidirectional Point-to-Multipoint Connections
- Root can add/drop leaves in a Point-to-Multipoint Connection
- Leaves can drop from a Point-to-Multipoint Connection
- Network can tear down a connection
- Calling party can request a Quick Restart

The basic capabilities proposed for Phase 2 of the UNI Signaling specification, due in late 1995, include these additions:

- QoS/Bandwidth haggling during connection setup
- QoS/Bandwidth modification during connection
- Common/Last Route Connection
- ATM Address Groups (e.g. Multicast Addresses)
- Leaves join Point-to-Multipoint connection without root
- Optional Multipoint-to-Multipoint connections
- VPI/VCI negotiation

2.3.11.2 Addressing

The ATM Forum has agreed that all ATM equipment will identify ATM endpoints using the OSI Network Service Access Point (NSAP) address format. While other ATM addressing formats may be included in the future, three (3) ATM endpoint identifiers based on NSAP address formats have been specified to date.

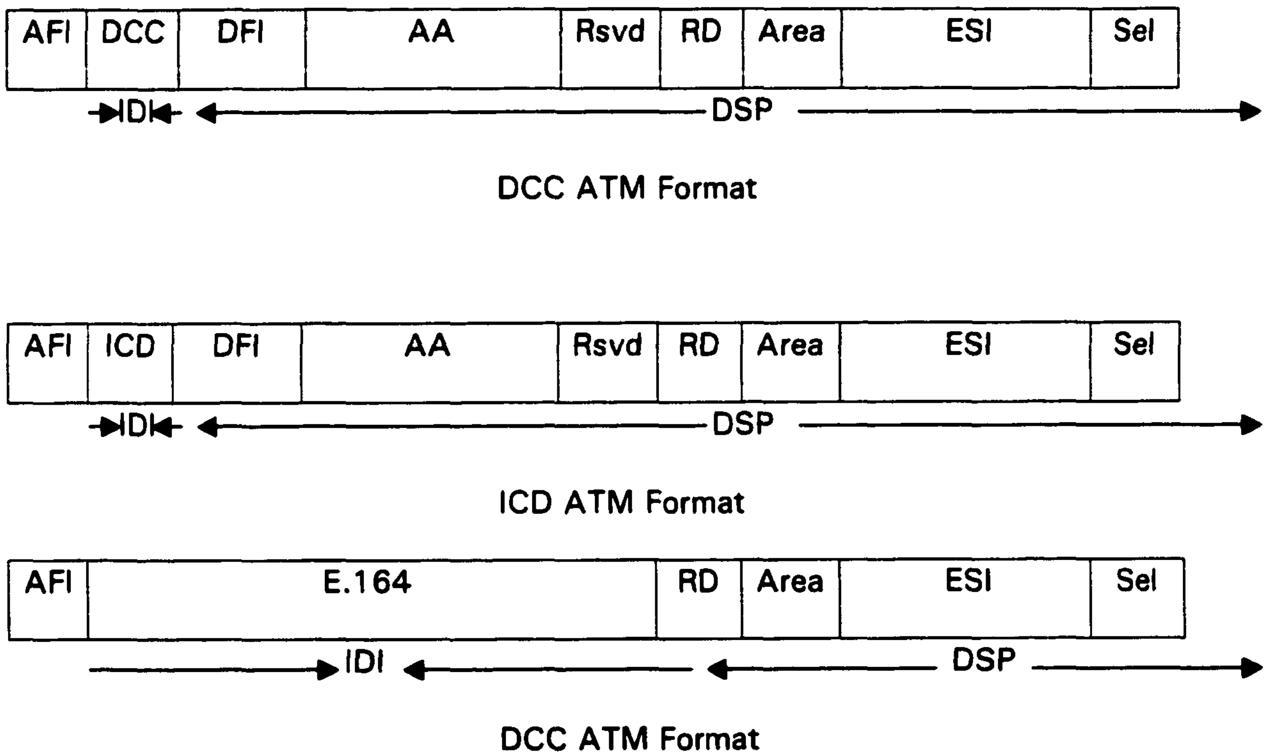
The Authority and Format Identifier (AFI) indicate the syntax of each Network Service Access Point format. This is followed by an Initial Domain Identifier (IDI), which includes an address field that can be assigned by different organizations. These different private NSAP addresses can be either specific to a local country domain or globally unique. Public ATM networks may use either E. 1 64 addressing - the same format used today in public telephony networks - or the NSAP address format, which can incorporate E.164 addresses.

In ATM LANS, an endpoint's IEEE MAC address will most likely be used as the End System Identifier (ESI). Therefore, when an ATM end station connects to a network for the first time, it will have to register its MAC adaress with an Address Registration service provided by the ATM network. The ATM network's Address Registration service will then respond to the endpoint with its assigned NSAP address and store the MAC-to-ATM address pairs with the respective switch and port number.

Address Registration will be accessed and completed via the Interim Local Management Interface (ILMI) - facilitating auto-configuration of an ATM endpoint's NSAP format. This address registration procedure will provide information to the ATM network which will later be used to support address resolution and connection routing procedures. The embedded hierarchical structure of RD/Area sub fields plus the ESI in the Domain Specific Part (DSP) actually makes the NSAP ATM address format both scalable and routable in large networks.

2.3.11.3 Routing ATM Connections

The UNI signaling scheme defines how end stations will set up SVCs between a user and a switch. However, a robust signaling and routing protocol defining how a local infrastructure of ATM switches will setup SVCs to support those end-to-end signaling requests has not been completed. The ATM Forum's P-NNI Sub-Working Group is tasked with completing both a signaling and routing protocol which will be used by an infrastructure of ATM switches. Currently, the P-NNI signaling implementation is built upon UNI signaling protocol. The routing implementation will likely be based on a blend of existing routing protocols and require a well-defined administrative and topological architecture.



IDI = *Initial Domain Identifier*
DSP = *Domain Specific Part*
AFI = *Authority and Format Identifier*
DCC = *Data Country Code*
ICD = *International Code Designator*
E. 164 = *Public Telephony Address Format*
DFI = *DSP Format Identifier*
AA = *Administrative Authority*
Rsvd = *Reserved*
RD = *Routing Domain*
Area = *Area Identifier*
ESI = *End Station Identifier*
Sel = *NSAP Selector*

Figure 2.8. ATM NSAP Address Formats

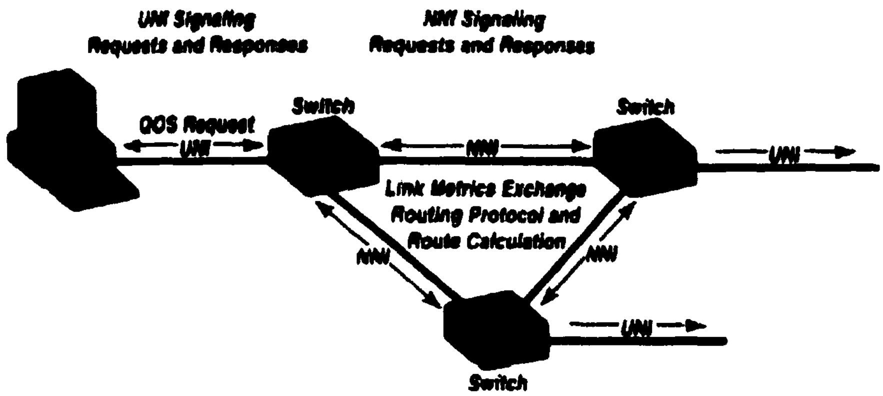


Figure 2.9. UNI and NNI Signaling

Administrative and Topological Architecture

Establishing an administrative and topological architecture to support the P-NNI routing implementation will be required. Specifically, routing within a local domain between individual switches will typically require that each switch have a complete topological map of all the switches and their links within that same domain. Routing between domains will typically not be governed by topological maps, but more by administrative policies where routes are aggregated and inserted into the other domains.

Current thinking on the P-NNI's administrative and topological architecture is based upon a multi-level routing hierarchy. The lowest level consists of a peer grouping of switches and/or routing engines, which make up a domain. Within this domain, link state updates will allow all the switches and/or routing engines to obtain timely, accurate topology and reach ability information. In the next highest level of the routing hierarchy, the entire peer group of switches is represented as a single node or Switching System (SS), at which point all administrative and reach ability information is summarized for each member of that domain. An unlimited number of hierarchical routing levels can be established depending on size of the network and the desired maximum size for each SS.

Routing Protocol

The network's switches will use the P-NNI routing protocol and/or routing engines to exchange routing information between them. The actual P-NNI routing protocol will likely be based on attributes borrowed from OSPF, IS-IS, BGP or IDR-P and adapted to meet ATM's unique requirements.

Finding a route through an ATM network topology to an endstation that will support the requested Quality of Service is not an easy task. Thus, route calculations will also have to be adapted to consider link and endpoint metrics to help determine the appropriate route that will support the requested QoS.

Static	Dynamic
Cost	Cell Delay (Latency)
Link Capacity and Constraints	Cell Delay Variance (jitter)
Propagation Delay	Current Sustained Capacity

Table 2.4. Minimum Link Metrics Exchanged by P-NNI

It is likely that the P-NNI routing protocol will generate partial-source routes. This will be better suited for accurately determining a route that will support a desired QoS than hop-by-hop routing schemes. A partial-source route is also more attractive since each switch or routing engine will need to source route requests only within its reach ability or level in the multi-level routing hierarchy, allowing route engines at higher or lower levels to further route the request. A hop-by-hop routing scheme may be defined as a default option only. In spite of the complexity routing ATM connections may require, overall routing overhead will be lower than it is in today's networks since the route is determined only once.

2.3.12 IP and ARP over ATM

Today's network layer protocols, such as IP, were not designed to interface with ATM's addressing or connection setup procedures. The Internet Engineering Task Force (IETF) has formed a working group with the purpose of drafting new Request For Comment (RFCs) which will define Classical IP and ARP implementations over ATM networks. Two other working groups in the IETF are focusing on Routing Over Large Clouds (ROLC) and defining ATM Management Information Base (MIBs).

One method for using Classical IP and ARP directly over ATM was recently described in RFC 1577. Conceptually, the fundamental operation of Classical IP and ARP over ATM will be similar to their use in today's LANS. ATM Address Resolution Protocol (ATMARP) is used to resolve an IP address to an ATM address. IP is used to route Protocol Data Units (PDUs) between ATM networks configured as Logical IP Subnetworks (LIS). However, both IP and ATMARP will now have to be concerned with the connection status of Logical IP Subnetworks (LIS) members and be able to request new inter-LIS connections when needed and be able to use ATM's addressing scheme.

2.3.12.1 ATMARP

Each LIS must have one ARP service and all members of that LIS must be configured with the ARP service's ATM address. The ARP service could be a dedicated server or a server shared or multiplexed among multiple LISs. The ARP service will keep a point-to-multipoint connection to all the other LIS members and add each new LIS member after it has sent an initial connection request. Then, each LIS member must be able to respond to any ATMARP or inverse ATMARP request on any virtual connection of which it is a member.

2.3.12.2 IP Over ATM

Operation of the classical IP layer doesn't really change, as all inter-LIS traffic must still pass through a router connection, even if a direct connection across the ATM network exists. While this has the benefit of being quick to implement, having all inter-LIS traffic go through a router could create a bottleneck. The IETF's Routing Over Large Clouds Sub-Working Group has been focused on developing solutions to overcome this potential problem by allowing direct connections across Non-Broadcast Multi-Access (NBMA) networks such as ATM. To date, the ROLC group's work has been based on a scheme called Next Hop Resolution Protocol (NHRP), which may one day emerge to replace the Classical IP and APP over ATM implementation described in RFC 1577.

2.3.13 LAN Emulation

While the IETF has just finished an initial implementation of IP and ARP over ATM, and Novell is working on a second generation IPX that will be ATM-aware, the modification of other network layer protocols is believed to be much further behind. More importantly, these revised protocols will require ATM users to change their network operating software - a potentially disruptive and costly undertaking.

LAN Emulation promises to at least postpone, if not eliminate, the need for changing network operating protocol software. While this can provide a huge cost benefit, it does prevent higher layer applications from accessing ATM's unique services. When finished, the LAN Emulation specification will define how an ATM network can emulate enough of the MAC protocol of an existing LAN technology, such as Ethernet and Token Ring, so that higher layer protocols can be used without modification. LAN Emulation would be implemented as device drivers below the network layer in ATM-to-Legacy LAN bridges and ATM endstations. In an ATM endstation adapter, LAN Emulation device drivers would interface with widely accepted driver specifications, such as Network Driver Interface Specification (NDIS) and Open Data link Interface (ODI) used by TCP/IP, IPX, etc.

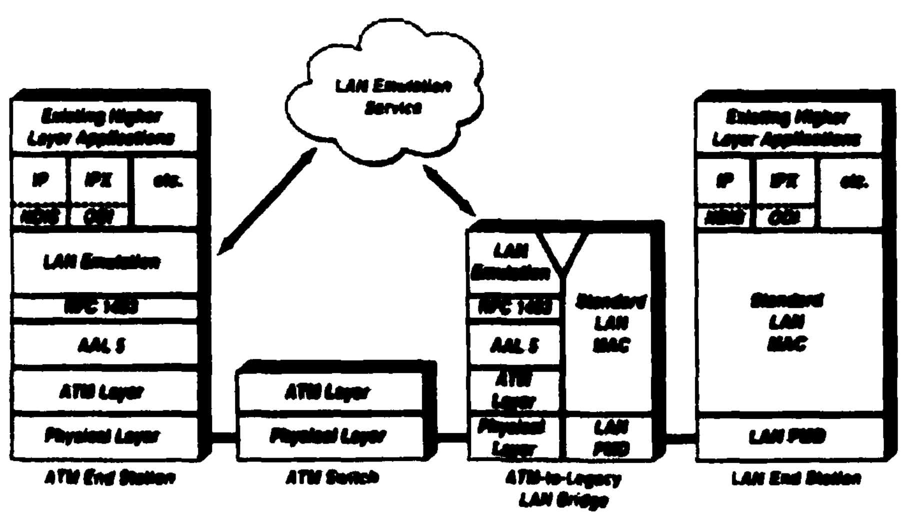


Figure 2.10. LAN Emulation Structure

Note that LAN Emulation is not designed to provide translational bridging between legacy LANS. A bridge or router is still required to inter network the different legacy LANS. Nor will LAN Emulation perform all MAC-layer functions since, for example, there will be no reason to perform collision detection on an ATM network. At this time, The ATM Forum's LAN Emulation Sub-Working Group has decided to define the emulation for only two legacy LAN protocols, Ethernet and Token Ring. All other legacy LAN technologies, such as FDDI, will have to be routed over ATM or bridged/routed to either Ethernet or Token Ring.

2.3.13.1 LAN Emulation Service

An ATM network could have several different emulated LANs which could not communicate directly between each other without bridging or routing - the same requirement as connecting individual segments in today's LANs. A LAN Emulation service can be augmented to allow the configuration of one or more autonomous workgroups, better known as Virtual LANs (VLANS), where no broadcast, multicast, or unicast traffic could enter or leave that group.

Each emulated LAN is composed of LAN Emulation Clients (LEC), generally located in ATM endpoints, and a single LAN Emulation Service (LES). The LES can be implemented in an ATM-attached server or in one or more ATM switches, or both. The LES provides initialization, registration, address resolution, and data forwarding services to the LAN Emulation Clients over what is called the LAN Emulation UNI (LUNI). Multiple LESs will exchange information over the yet-to-be-fully-defined LAN Emulation NNI (LENNI).

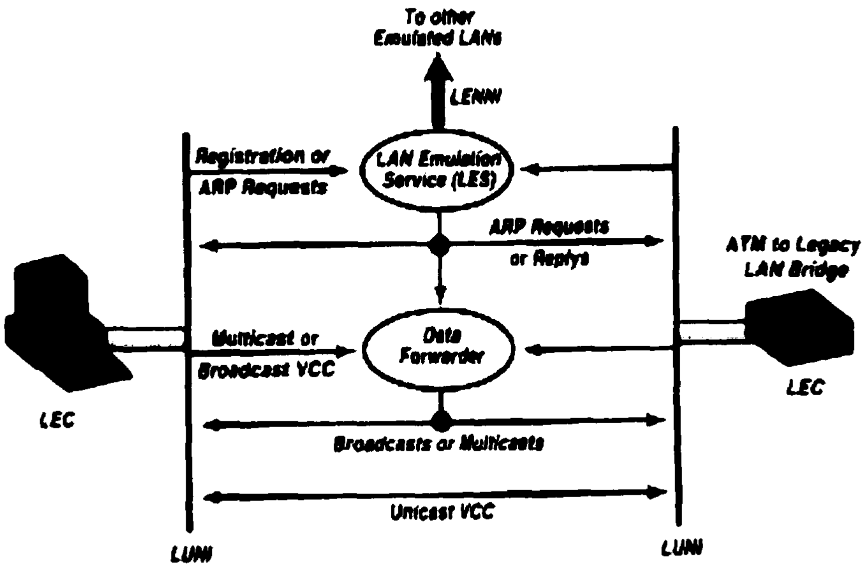


Figure 2.11. LAN Emulation Service Operation

Initialization: During initialization, a LEC requests the ATM addresses of all the LESs that are available (or the LESs that the client has been authorized to join) in the ATM network via a well-known ATM address. After receiving the ATM addresses, the client can contact the LES of the emulated LAN it wants to join.

Registration: The LEC gives the LES the list of individual NIAC and ATM addresses that it represents. In the case of a bridge, it does not have to give all the MAC addresses that may be attached to it. The LES also gives important parameters about the emulated LAN, such as the maximum transfer unit size, to the client during registration.

Address Resolution: The LEC sends an ARP request to the LES to discover the data forwarder's ATM address. After receipt of response, the client will establish a connection with the data forwarder. The client can also send ARP requests to the LES to discover the corresponding ATM address of the MAC address it wants to send unicast frames to. The LES will either have the information already in its table or pass the ARP request to all the other clients via the data forwarder.

Data Forwarder Sends all multicast and broadcast frames to every client in the emulated LAN. The data forwarder can be implemented as an ATM-attached station used as a cell-replicater or it can be distributed among one or more ATM switches that have architecture, which support the multicasting of cells.

2.3.14 Traffic Management

Effective traffic management mechanisms will allow an ATM network to deliver guaranteed QoS on demand while maximizing the utilization of available network resources. In fact, traffic management mechanisms will be inserted into nearly every aspect of ATM network operation, from signaling requests and routing to network resource allocation and policing. In addition to preventing congestion from over subscription to network resources, traffic management mechanisms will also be designed to recover from unexpected congestion conditions to ensure good ATM network performance.

ATM virtual connections can have either a specified or unspecified QoS. Why the choice? If an ATM network supported only specified QoS connections, a large percentage of the network resources would go unused. This waste would result from periods when one or more connections were not using the full capacity of their QoS contracts or because a conservative network policy was over-provisioning resources.

Therefore, unspecified QoS contracts can be supported by an ATM network on a "best-effort" basis - which is more than adequate for supporting typical datacom needs in today's LANS. Since unspecified QoS traffic patterns will be unpredictable at best, fair and predictable congestion control and feedback mechanisms become important ATM traffic management responsibilities.

Parameter	**Indicates**
Cell Error Ratio	Accuracy
Severely Errored Cell Block Ratio	Accuracy
Cell Loss Ratio (CLR)	Dependability
Cell Misinsertion Rate	Accuracy
Cell Transfer Delay (CTD) (latency)	Speed
Cell Delay Variation (CDV) (jitter)	Quality
Mean Cell Transfer Delay	Speed

Table 2.5. ATM Link Performance Parameters

2.3.14.1 Traffic Contract

At connection setup, the traffic contract, which supports the QoS requested, contains a Connection Traffic Descriptor and a Conformance Definition. Traffic contracts will be able to specify the following types of traffic flows: Constant Bit Rate (CBR), Variable Bit Rate (VBR), Available Bit Rate (ABR), VBR+ (combination VBR plus ABR), and Unspecified Bit Rate (UBR). An ATM network when providing a "best-effort" service will use ABR and/or UBR traffic contracts.

	CBR	VBR	VBR+	ABR	UBR
CLR	Specified	Specified	Specified	Specified	Unspecified
CTD & CDV	Specified	Specified	Specified	Unspecified	Unspecified
PCR & CDVT	Specified	Specified	Specified	Specified	Specified
MBS & SCR	N/A	Specified	Specified	N/A	N/A
Feedback	No	No	Specified	Yes	N/A

Table 2.6 Connection Traffic Descriptors

Conformance Definition

The Conformance Definition for ATM is based on a Generic Cell Rate Algorithm (GCRA). The GCRA actually performs a traffic shaping function and is roughly equivalent to a continuous-state Leaky Bucket Algorithm (LBA). Figure 2.12 illustrates this traffic shaping function using the simple Leaky Bucket Algorithm.

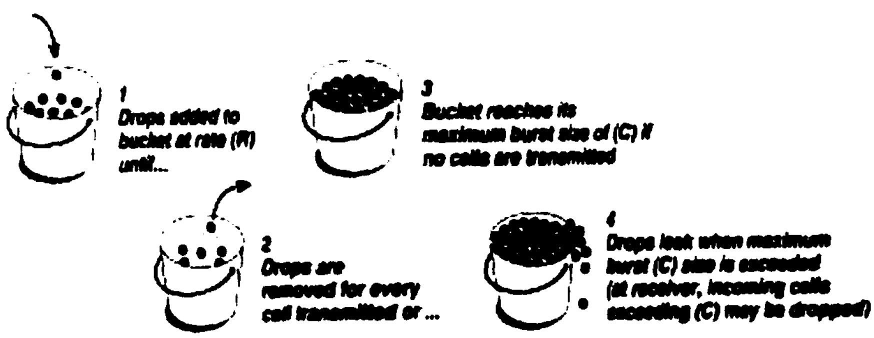


Figure 2.12. Simple Leaky Bucket Algorithm

The source is left to conform to the GCRA on its own, inserting idle cells into its traffic stream as needed - since the network does not actually issue these "drops." On the receiving end of the ATM link where conformance may be

monitored or policed, every cell that exceeds the bucket's capacity upon arrival can either be dropped or have its Cells Loss Priority (CLP) lowered. If extra bandwidth is available, changing the CLP will allow the subsequent ATM links to support the nonconforming cells on a "best-effort" basis. Policing of connections will probably be of greater interest to public ATM WAN providers where billing is based on the Quality of Service agreed to during connection setup.

2.3.14.2 Traffic and Congestion Control Functions

Connection Admission Control (CAC): The set of actions taken by the network during the connection setup phase to determine whether the requested QoS can be accepted or should be rejected. CAC is also used when routing a connection through an ATM network.

Usage Parameter Control (UPC): The set of actions by the network to monitor and control traffic. Its main purpose is to protect network resources from malicious as well as unintentional misbehavior, which may adversely affect the QoS of already established connections. UPC detects violations of negotiated QoS parameters and takes corrective policing actions.

Priority Control.- ATM endpoints can set one of two priority levels for traffic flows. ATM switches may selectively discard cells with low priority to avoid congestion or to preserve overall network performance.

Traffic Shaping.- Used to achieve or modify traffic characteristics in order to match a desired QoS contract.

Feedback Controls.- Used by the network and/or by end stations to regulate or alter the traffic transmitted on ATM connections in response to changing network conditions - such as congestion.

2.3.14.3 Congestion Control Mechanisms

Proposals to The ATM Forum on various congestion or flow control mechanisms are currently focused on data traffic serviced by Variable Bit Rate (VBR) and/or Available Bit Rate (ABR) traffic contracts. These control mechanisms will be used to regulate and alter those traffic contracts, perhaps only temporarily, when congestion is experienced. The three mechanisms under consideration include: Forward Explicit Congestion Notification (FECN), Backward Explicit Congestion Notification (BECN), and Virtual Connection Flow Control (VCFC).

FECN.- Forward Explicit Congestion Notification. A connection that experiences congestion will have the congestion bit set in the cell's header by the switch at the point of congestion. Those cells continue to their destination, unless they are dropped. The destination detects the congestion bit and sends a congestion notification back to the source. If the distance between the source and the destination is large enough, the delay in receiving feedback can make this scheme ineffective.

BECN.- Backward Explicit Congestion Notification. The switch at the point of congestion sends a congestion notification directly back to the source. While feedback time can be considerably shorter, this scheme adds to the complexity of the switch.

VCFC.- Virtual Connection Flow Control. VCFC is very different from BECN and FECN as it uses a credit-based approach. Specifically, credit messages are issued between every ATM link to ensure that cells are transmitted only when there is sufficient buffer capacity in the next switch. While this is a very reliable feedback scheme, it does place a huge burden on all the switches in the network as well as adding more overhead to the network's operation.

2.3.15 Network Management

Once all the specifics of ATM network operation are finalized, a robust network management mechanism can be defined. In the meantime, the ATM Forum has defined an Interim Local Management Interface (ILMI) and Operation and Maintenance (OAM) principles have been recommended by the ITU-T.

2.3.15.1 Interim Local Management Interface (ILMI)

The Interim Local Management Interface uses the Simple Network Management Protocol (SNMP) and an ATM UNI MIB to provide an administrator with status and configuration information. Specifically, the ILMI supports bi-directional exchange of management information between management entities related to the UNI ATM Layer and physical layer parameters. The types of information available in the ATM UNI MIB are:

- Physical Layer
- ATM Layer Statistics
- VPC status
- VCC status
- Address Registration Information

2.3.15.2 Operations and Maintenance (OAM)

The ITU-T has already issued a number of recommendations for OAM principles for B-ISDN. The primary OAM goal is to prevent the need for corrective actions. OAM principles consist of supervision, testing, and performance monitoring. OAM is further organized into five hierarchical levels categorized as information flows (Fn).

Virtual Channel (F5): Operations and Maintenance functions are performed on a VCI level and indicate if a channel is not available or if performance has degraded.

Virtual Path (F4): OAM functions are performed on a VPI level and indicate if a path is not available or if performance has degraded.

Transmission Path (F3): OAM functions indicate if cell delineation has been lost, if non-correctable headers are being experienced, and if insertion or suppression of idle cells has failed.

Digital Section (F2): OAM functions are capable of generating and transporting OAM information between SONET digital sections.

Regenerator Section (F1): OAM functions indicate if there has been a loss of signal over the physical medium of the transmission and if bit error rates have degraded.

2.3.16 Layers and Sublayers

The ATM protocol uses a layered structure similar to the Open Systems Interconnect (OSI) model. The ATM Protocol Reference Model is different from the OSI model in the use of planes, which operate across all three layers. These are: the User Plane, the Control Plane, and the Management Plane. The User Plane is used for end-to-end data transfer. The Control Plane supports network-wide functions such as the signaling used to establish Switched Virtual Circuits (SVC) between endpoints. The Management Plane supports management and control functions for the network and its endpoints.

In ATM Networks, communication from higher layer applications and protocols to the network media occurs through three layers: the ATM Adaptation Layer (AAL), the ATM Layer, and the Physical Layer. The ATM Protocol Reference Model is further divided into sub layers that perform specific functions associated with accepting information from or preparing information for another layer. The AAL Layer includes the Convergence Sub layer and the Segmentation and Reassembly Sub layer. The Physical Layer includes the Transmission Convergence Sub layer and the Physical Medium Dependent Sub layer.

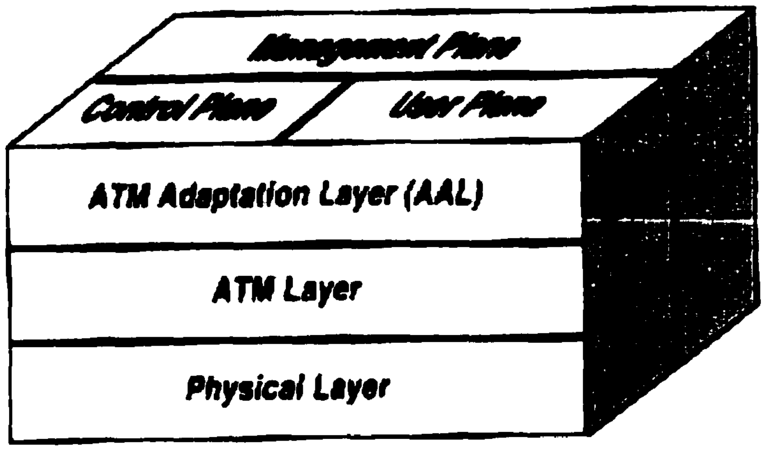


Figure 2.13. ATM Protocol Reference Model

Layer	Sublayer	Functional Description
AAL	CS SAR	ATM Adaptation Layer Convergence Segmentation and Reassembly
ATM		ATM Layer Cell Header Generation/Removal Cell Address Translation Cell De/multiplexing
PHY	TC PMD	Physical Layer Transmission Convergence HEC Generation/Verification Cell Deliniation Physical Medium Dependent Bit Timing

Table 2.7. Layer / Sublayer Function

2.3.17 ATM Layer Operation

Generally speaking, the ATM Layer performs cell multiplexing, cell header generation or removal, and VPI/VCI translation. Although ATM Layer operation is relatively uniform across the network, it does vary depending on the kind of ATM network device. For example, the ATM Layer has to generate or remove cell headers in an endstation. But in a switch, the ATM Layer must multiplex/demultiplex cells from several different connections at once and translates cell VPI/VCI values for transmission to the next link for each connection. In any case, the ATM Layer is responsible for the successful transmission of 53-byte cells and must communicate with the ATM Layer of other network devices only through the cell's 5-byte header.

In an ATM endpoint or station, the ATM Layer exchanges a cell stream with the physical layer, inserting idle cells if no higher layer information is waiting to transmit or if empty cells are needed to comply with Quality of Service (QoS) parameters. Upon receiving cells from the physical layer, the 48-byte payload is passed to the AAL along with a few parameters, such as the Payload Type Indicator (PTI) value and the Cell Loss Priority (CLP) value.

Upon receipt of an ATM cell at one port of an ATM switch, the ATM Layer will determine from the VPI/VCI value the port to which it should relay the cell and the new VPI/VCI value. It will then forward the cell to this new port, change the VPI/VCI value, and pass the cell down to the physical layer of the outgoing port for transmission. The ATM Layer may also set a bit in the PTI field if it experienced congestion and change the CLP field when enforcing the appropriate traffic shaping and policing algorithms, such as the leaky bucket algorithm.

In a switch, the ATM Layer also ensures that cells from the same virtual circuit do not get misordered and that system requirements, such as maximum end-to-end latencies, are met. The ATM Layer must also provide adequate buffering and other congestion control mechanisms.

2.3.17.1 Cell Header Generation or Removal

The structure of the ATM cell is identical everywhere within an ATM network. Only a small variation appears in the UNI header compared to the NNI header. Part of the VPI field in the UNI header is reserved as a Generic Flow Control (GFC). The NNI has a larger range of VPI values which correctly reflects the greater use of Virtual Paths (VP) for trunking purposes across inter-switch and inter-network interfaces. ATM cells are transmitted serially, starting with bit number 8 in the first byte of the cell header.

Generic Flow Control (GFC): Reserved only for the UNI. The GFC is intended for use by the UNI to control traffic flow and alleviate short-term overload conditions. To date, the GFC has not been defined and remains an unused field. The NNI cell utilizes this space not for GFC, but instead for a larger VPI value.

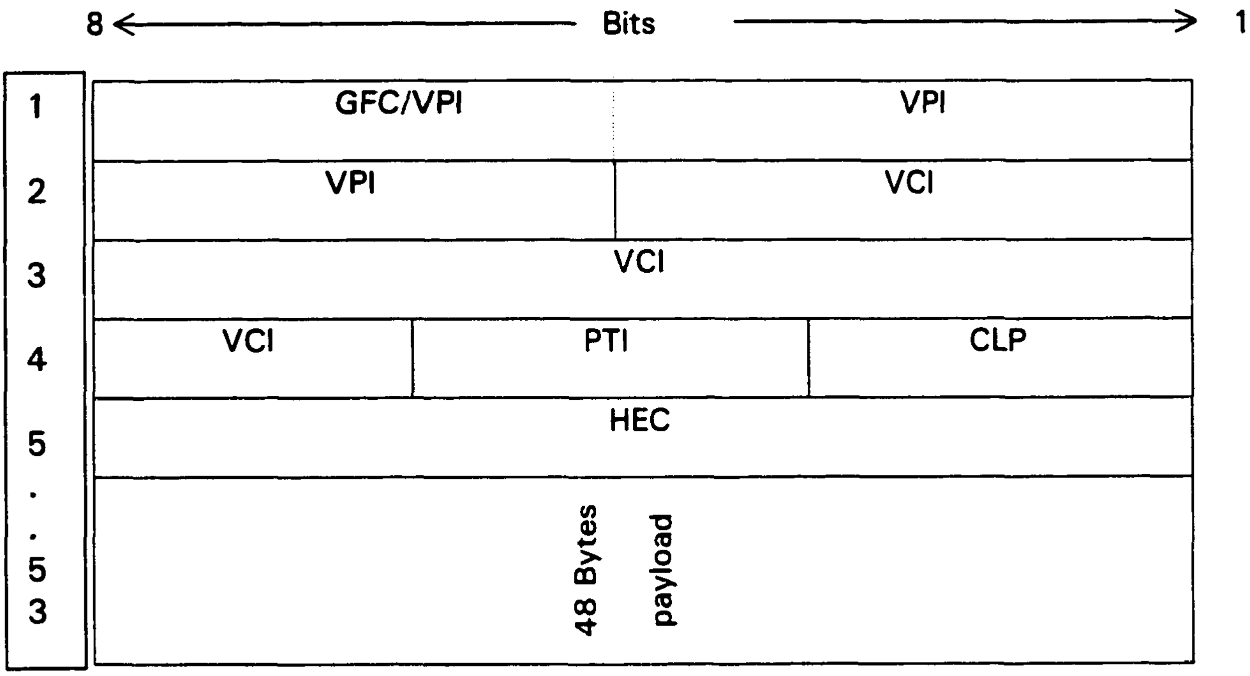


Figure 2.14. UNI/NNI TM Cell Header Format

Virtual Path Identifier (VPI): Identifies virtual paths. In an idle or null cell, the VPI field is set to all zeros. (A cell containing no information in the payload is called either "idle" or "null.") Other nonzero values of the VPI are reserved for various meta-signaling and OAM procedures.

Virtual Circuit Identifier (VCI): Identifies virtual circuits. In an idle or null cell (one containing no information in the payload) the VCI field is set to all zeros. Other nonzero VCI values are reserved for special purposes. For example, the VPI=0/VCI=5 value is used when sending a signaling/connection request.

Payload Type Identifier (PTI): Identifies the payload type carried in the ATM cell and/or to identify control procedures for the OAM. A particular combination is used by the AAL if the cell is a part of an AAL 5-type connection. Another combination is used to indicate that the cell has experienced congestion.

Cell Loss Priority (CLP): This is set by the AAL layer and used by the ATM Layer throughout the network to determine the relative importance of cells. This bit is set to 1 when a switch experiencing congestion can discard a cell. If a cell should not be discarded in order to support a specified QoS, this bit is set to 0. The CLP bit may also be set by the ATM Layer in the network if a connection exceeds the QoS parameters agreed to during connection setup.

Header Error Check (HEC): This 8-bit Cyclic Redundancy Code (CRC) is computed only over the ATM cell header and is capable of detecting all single bit errors and certain multiple bit errors. It can also correct single bit errors and is actually computed by the physical layer, which also uses it for cell delineation.

2.3.18 ATM Adaptation Layer (AAL) Operation

The AAL Layer is responsible for adapting higher layer data with different service requirements into cell payloads for the ATM Layer and vice versa.

While the ATM Layer is designed to make ATM more reliable, flexible, and easier to operate than today's networks, it is the AAL that provides the ability to support many different types of traffic and applications over an ATM network.

2.3.18.1 Sublayers

The AAL protocol performs two important functions - a Convergence function and a Segmentation and Reassembly (SAR) function. After a connection with the appropriate QoS is established, the Convergence Sublayer (CS) accepts higher layer traffic for transmission across the network. The SAR segments each packet from this higher layer traffic into smaller units (also known as CS-PDUS) and adds a header and/or trailer field depending on the AAL type to form 48-byte payloads (also known collectively as SAR-PDUs). Upon receipt of cell payloads, the AAL removes any AAL-specific information from each payload and reassembles the entire packet before passing it to a higher layer.

For some AAL types, the CS is further subdivided into a Service Specific Convergence Sub layer (SSCS) and a Common Part Convergence Sub layer (CPCS). Traffic using these AALs have their packets encapsulated with another header and/or trailer field before being segmented into smaller units (the encapsulated packets are known as CPCS-PDUs). After being segmented into smaller units, these CS-PDUs may have another header and/or trailer field added to them to form the 48-byte SAR-PDU.

2.3.18.2 Classes of Traffic

These classes of traffic were defined during B-ISDN development and are supportable by ATM:

Class A: Constant bit rate (CBR) - connection-oriented, synchronous (e.g. uncompressed voice or video)

Class B: Variable bit rate (VBR) - connection-oriented, synchronous traffic (e.g. compressed voice and video)

Class C: Variable bit rate (VBR) - connection-oriented, asynchronous traffic (e.g. TCP/IP, IPX, X.25, Frame Relay)

Class D: Variable bit rate (VBR) - connectionless asynchronous traffic (e.g. SMDS, etc.)

2.3.18.3 Types of AALs

Different types of AALs provide service to the higher layer protocols or applications. The AALs correspond to the four classes of traffic defined by

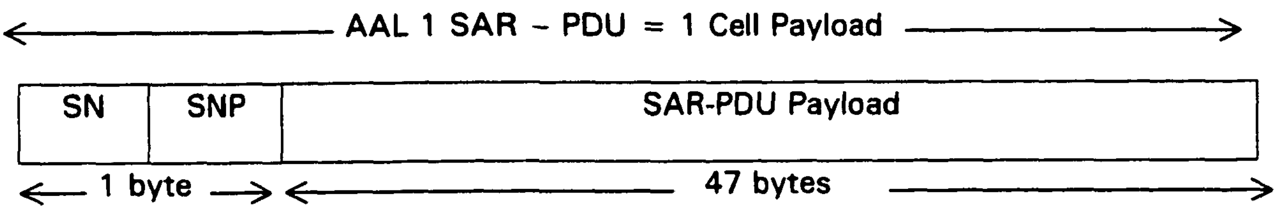
the ITU-T when B-ISDN capabilities were originally determined. To date, there are a total of five different AALs defined to carry one or more of the four classes of traffic. The operation of each AAL varies depending upon the type of traffic that it is optimized to carry. There is no stipulation in the standards that AALs designed for one class of traffic cannot be used for another. For example, it may be common to see AAL 5 carry Class B traffic or AAL 1 carry Class C traffic.

	Class A	*Class B*	*Class C*	*Class D*
Timing relation between source and destination	*Required*		*Not Required*	
Bit Rate	*Constant*	*Variable*		
Connection Mode	*Connection Oriented*			*Connectionless*
ATM Adaptation Layer	*AAL 1*	*AAL 2*	*AAL 5, AAL 3/4*	*AAL 3/4*
Examples	*Voice, Circuit Emulation*	*Compressed Video*	*TCP, X.25*	*SMDS*

Figure 2.15. Classes of Traffic and Associated AALs

2.3.18.3.1 AAL Type 1

AAL 1 is designed to carry Class A traffic - CBR traffic from a higher layer that must be delivered to the destination at the same rate and at equal intervals. A connection supporting voice traffic or telephony circuits is a prime example of this kind of service. Since misordering of cells is worse than losing cells in this kind of service, a sequence number (SN) is added when forming the SAR-PDU to help with the detection and correction of lost or misinserted cells.



SN = Sequence Number
SNP = Sequence Number protection

Figure 2.16. AAL Type 1

Some aspects of AAL 1 operation are not fully specified. For example, it is not clear how synchronization will be maintained between the source and destination since AAL 1 does not allow a clock to be carried.

It is believed that using priorities and provisioning, a clock could be generated at the destination from the cell inter-arrival time using a phased locked loop and an elastic buffer. The ITU-T recommended a technique named Synchronous Residual Time Stamp (SRTS) which can be used over SONET. SRTS uses special synchronization information generated and transmitted to the destination by the AAL 1 at the source and the network clock extracted from the Physical Layer to generate a clock at the destination.

2.3.18.3.2 AAL Type 5

At the other end of the spectrum, AAL 5 has been designed to carry data traffic typically found in today's LANS. AAL 3/4 was originally designed to carry this kind of traffic. Concerns over the poor efficiency of AAL 3/4 led to the definition of AAL 5. Only a small amount of overhead is added to the CPCS-PDU and no extra overhead is added when the AAL 5 segments them into SAR-PDUS. AAL 5 is also known as the Simple and Efficient AAL (SEAL).

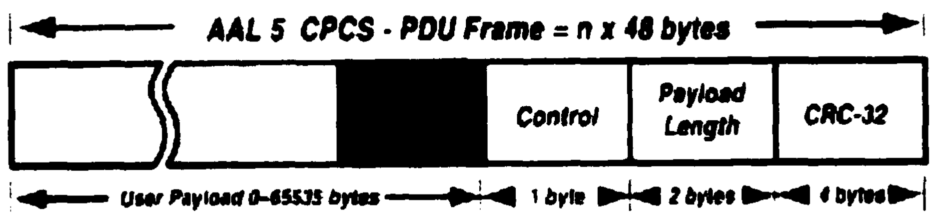


Figure 2.17. AAL Type 5

Another attribute contributing to AAL 5's efficiency is the use of the PTI field to indicate that the cell is supporting AAL 5 traffic, rather than using some of the payload for this purpose. Also, a 32-bit CRC is calculated over the entire CPCS-PDU allowing for the detection of cell loss and the misordering or misinsertion of a cell. One compromise made in defining AAL 5 was the decision that the CS-PDU will not carry a length field in its header to indicate the length of the entire data packet to be transmitted. While this simplifies the operation of the transmitter, it complicates buffer allocation in the receiver. This is no worse or better than it is in today's LANS. A well-defined maximum transfer unit (MTU) size will at least make buffer allocation more deterministic.

2.3.18.3.3 AAL Type 2

For Class B traffic it is just as important not to lose cells as it is not to misorder them. This is because some compression algorithms use reference cells (e.g. MPEG uses reference frames for differential coding of video signals) which if lost can result in a simple, single-bit error being multiplied and propagated many times. The operation specifics of AAL 2 are not well defined at this time. ITU-T suggested the following SAR-PDU structure for AAL 2.

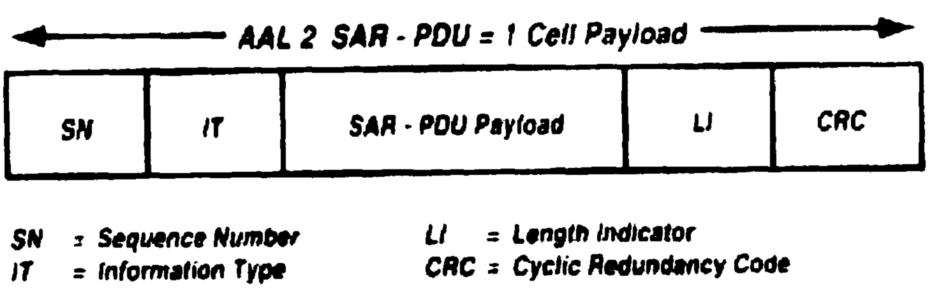


Figure 2.18. AAL Type2

2.3.18.3.4 AAL Type 3/4

AAL 3/4 is fairly complex because its PDU formats were designed to support both Class C and Class D services. AAL 5 was defined after AAL 3/4 to provide only a subset of Class C service. AAL Type 3/4 can operate in a number of different modes. In Message Mode, fixed or variable size packets are passed to the AAL and transmitted in a single CPCS-PDU. In Streaming Mode, a single packet is passed to the AAL layer and transmitted in multiple CPCS-PDUS when pieces of the packet are received. Streaming mode may be used in intermediate switches or ATM-to-SMDS routers so they can begin re-transmitting a packet being received before the entire packet has arrived - reducing the latency experienced by the entire packet. Both modes of operation can be run in either assured operation (where the AAL layer will assure the correct end-to-end delivery of user data through such mechanisms as retransmission and flow control) or in non-assured operation (where delivery of CPCS-PDUs is not assured and there is no retransmission of lost or corrupted CPCS-PDUs).

2.3.19 Physical Layer Operation

The Physical Layer defines the bit timing and other characteristics specific to the physical media used and passes the cells from the media to the ATM Layer and vice versa.

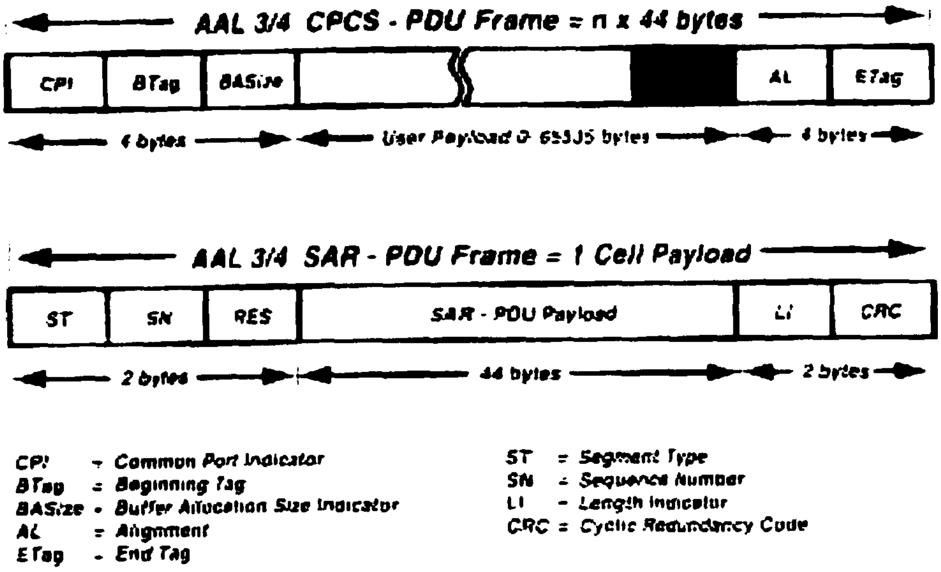


Figure 2.19. AAL Type 3/4

The interface from ATM to a large variety of physical layers is possible because of the segmentation of the ATM physical layer into two sub layers: PMD (Physical Medium Dependent) and TC (Transmission Convergence). Each PMD is specific to a particular physical medium and includes definitions for proper cabling as well as bit timing. The TC sub layer performs a convergence function by receiving a bit stream from the PMD and extracting cells using the HEC for cell delineation.

While PMD operation depends on the physical medium, a number of TC functions remain common to all the Physical Layers. These include cell delineation, cell rate decoupling, HEC generation and checking, and various OAM functions. Cell delineation is the extraction of cells from the bit stream received from the PMD, while cell rate decoupling is the adaptation of the speed of the ATM Layer cell stream to the cell rate of the physical interface. The TC checks where cells start and finish by calculating the HEC for every cell received and checking this against what it expects to be the HEC of the received cell.

2.3.19.1 Physical Layers Proposed by The ATM Forum

The ATM Forum has proposed a variety of physical layer draft standards in addition to the optical media and line rates defined for SONET. Currently, the ATM Forum's Physical Layer Sub-Working Group on specifications is making considerable effort for ATM over twisted-pair wire. They are focusing on implementing ATM over UTP-3, the prevalent twisted-pair cabling used in legacy networks and UTP-5, which is recommended for new twisted-pair installations. Once these standards are completed, ATM's acceptance as a desktop connection technology is expected to accelerate since existing cabling plants can be retained and the cost per connection will drop.

The ATM over UTP-3 specification is being based on the STS-1 rate (51 Mbps) and frame format using a CAP-16 coding scheme. The completed solution is also expected to scale to lower speeds that are fractions of these rates (e.g. STS-0.5, STS-0.25), so that longer cable lengths and lower quality cable can be accommodated in some environments. The specification for ATM over UTP-5 will use the STS-3 rate (155 Mbps) and frame format and the NRZI coding scheme.

2.4 Satellite Networks

Satellite links use an orbiting satellite as a repeater for the transmitted signal to amplify and rebroadcast to the receiver. Limited amount of channel switching is also accomplished. Communication satellites have a geostationary orbit that places it in the plane of the equator at an altitude of about 22,300 miles (35,784 km) above the earth. The direction of rotation is the same as that of the earth. The minimum spacing of different satellites is four degrees for the directional antennas. Geostationary satellites remain synchronous with the rotation of the earth thus being able to "view" a predefined area throughout the day without the need of the earth antennas to track the satellite. Station-keeping of the satellites is necessary since gravitation from the sun, the moon and other celestial forces causes a slow drift from their ideal positions. Both north-south and east-west drifts are encountered and the angular drifts are regulated to be within ± 0.1 degrees by timely rocket thrusts. The intervals for the correction may be several weeks. Due to the slow wander of the satellite, the communication path length also changes slightly causing perturbation in the delay. Highly directional antennas to the earth stations are used and their stabilization becomes crucial. Two techniques (double spin configuration and three axis stabilization) can be used. In the former configuration the angular drift is much less than 0.1 degree and in the later stabilization the drift can be under 0.001 degree, thus the exterior of the satellite appears fixed with respect to any point on the surface of the earth.

Satellite design is generally considered from *two* distinct considerations: *one* for its tracking, telemetry and control, and the *other* for its role as a communication unit. Many designs separate the requirements, functions and the components that are isolated. Earlier non-geosynchronous low-orbit satellite systems that forced spin-stabilization about a fixed axis maintained a fixed angle with respect to the earth's axis. Tumbling was thus eliminated, however the antennas could not be properly aligned to the earth station.

To permit the use of highly directional antennas aboard, two techniques (double-spin configuration and three-axis stabilization) were later developed. In the first method, the satellite and antennas are spun in opposite directions at low speed (75 rpm with ± 25 rpm). The antennas would retain their orientation with respect to the earth station if the angular damping could be made accurate. In the second method, a flywheel rotating at high speed would provide stability in all three axes and the exterior of the satellite (including the antennas) would be stationary with respect to the earth station. Two specific directionalities are necessary, the direction of the antenna and the direction of the solar panel.

Earth stations monitor, at regular intervals, the satellite position and functions, the antenna, and solar panel orientations. These earth stations then relay appropriate control signals to the satellite. Most satellite systems have certain redundancy of critical components and subsystems, and rigorous preflight testing. Solar cells (including back-up batteries for earth eclipse) and thrust fuel restrict the weight. Solar cells die an exponential death and the minimum wattage at "end of life" is balanced against the weight of thrust fuel that satellites carry. Component failure (typically occurring over a span of seven to ten years) rather than satellite failure due to decay of solar cells or fuel exhaustion, is the chief cause of "dead" satellites.

Even though non-synchronous satellites may be used (e.g., three Soviet launched "Molniya" satellites in elliptical orbit with 12 hour periods), a 24-hour continuity of service is possible by a timed ritual of daily handover procedures from the setting satellite to the rising satellite. They are particularly important in order to reach the polar regions of the earth where geostationary satellites do not provide good coverage due to extreme latitudes of these areas.

The up-link from the transmitter and the down-link to the receiver are each about one thousand times longer than a corresponding terrestrial hop. The cumulative time and signal distortion of the multiple hops in the terrestrial systems are non-existent in the satellite systems. Selective fading of the signal, common in terrestrial microwave systems, does not exist in satellite links. A single satellite covers a wide area of the earth. The problems of crossing rough terrain for humans and fading microwaves is thus resolved.

Due to limited satellite capabilities, *three* effects emerge:

I. Longer delays

II. More transmit power in the up-link and

III. Amplification of the down-link.

The weight, size and power of the satellite are major restrictions. The variation of media characteristics (especially the atmospheric region) and its loss place upper bounds upon what satellite links can accomplish as communication pathways.

Repeater satellite channels called transponders are "visible" to the transmitters and receivers by specially designed antennas. Antennas covering very wide continuous areas cannot offer enough power and cause interference with other satellite systems. This effect dictates the fact that certain orbital positions are more desirable than others to cover the Western Hemisphere. For example, the COMSTAR-2 satellite occupies 95 degrees West Longitude. It uses a 3.2' X 8' elliptic antenna beam to cover the continental United States; for Alaska, a 2' X 4' elliptical beam, and two 3' circular beams for Hawaii and Puerto Rico. Other communication satellites have specialized and focused antennas to serve that particular region of the globe.

2.4.1 Spectral Groups and Bands of Satellite Links

Three major frequency groupings 4/6 GHz, 12/14 GHz, 19/29 GHz exist. The up-links and down-links are separated by about 2 GHz for the 4/6 and 12/14 GHz groups and about 10 GHz for the 19/29 GHz group. The Federal Communications Commission (FCC) also imposes frequency tolerance. For example, frequency tolerance is restricted to be 0.001 and 0.002 percent of nominal value, respectively for both the earth station transmitter and space station. The up-links use a higher frequency band in each group. Also, a spectral band is necessary to carry numerous conversations or channels in each satellite link. This spectral bandwidth depends upon the number of channels carried on each link. Typically the 4/6 and 12/14 groups each have a 500 MHz bandwidth and the highest group has a bandwidth of 3500 MHz. Largely, the bandwidth is too large for individual use and is subdivided into a series of lower bandwidth transponder channels. Each transponder channel is at least adequate to carry one standard television transmission requiring a 36 MHz band. These transponder channels need suitable (about 4 MHz) guard space, thus requiring a separation of about 40 MHz. These frequency groupings and their spectral bands are presented in Table 2-1.

Table 2.8 Stationary Satellite Antenna Coverage

Area Covered	Bandwidth in Degrees	Gain in dBi (#)	Aperture at 4 GHz (Meters)
Entire World	18 circular	17	0.3
48 Cong. U.S.	8x4(elliptical)	27	0.8x1.6
1 hour time zone	1.5x3 (elliptical)	36	4.2x2.1

- 3 dB down at band edge. # dBi = dB with respect to isotropic radiator.

2.4.2 Interruptions in Satellite Communications

The satellite, having an orbit of its own, experiences the effects of eclipses. Of particular importance to communications via satellite are *two events. First,* during the spring and autumn seasons, the receiving antenna, satellite and sun align themselves, creating an eclipse. Any signal from the satellite is completely obliterated against thermal noise of the sun. These two annual eclipses are predictable and occur for about six successive days, for a few minutes each day, at noon at the location of the earth station. Another standby earth station takes over the communication from the satellite. At the end of the eclipse, the standby earth station hands back the function to the first earth station.

Second, the satellite also experiences an eclipse due to the earth's alignment between the sun and the satellite. During this time, the solar energy from the sun is blocked, making it necessary for storage batteries to power the transponders. The occurrence and duration (46 days) of these two annual eclipses are also predictable. The sudden change of temperature during the eclipse, lasting a maximum of about 72 minutes every day, can be of serious concern to the transponder components.

2.4.3 Impairments in Satellite Communications

The four major impairments to the propagation of the signal arise due to attenuation in the earth's atmosphere, effects of rain, ionosphere and low angles of satellite elevation from the earth station. Two important device limitations (antenna and system noise) also exist. Hence, it becomes essential to examine the design and operability of the satellite links in view of the propagation impairments and the device limitations. The effects of these have been documented [12.1] and the channel performance is crucial in maintaining its quality. Maintaining an acceptable ratio of the signal-to-noise ratio within the system regulates these impairments.

Generally, the capacity quotient computed (in dB) as the ratio of the carrier power to the noise power density of the channel is calculated. The noise density is determined as the sum from all the noise components arising from thermal, inter modulation, interference, and distortion. Both up and down links need to be considered in the overall channel performance. Losses of the channel consist of space propagation loss, atmospheric loss, earth station antenna tracking loss and rain loss. Six system gains compensate the losses and reestablish the signal (and introduce some noise in the process). Three such gains on the up-link are earth station transmitter amplification, transmitting antenna, and satellite receiving antenna. The remaining three gains on the down-link are the satellite-transmitting antenna, earth station receiving antenna, and receiver. The earth station gains are sometimes broken up into the IF (intermediate frequency) gain that amplifies the raw baseband signal and the main link amplifier that amplifies the entire satellite link. The channel quality is ascertained by monitoring the capacity quotient accurately.

To limit the weight of the satellite, the up-link receiver and down-link antenna gains at the satellite are roughly half of the earth stations up-link transmitter and down-link receiver antenna gains. The greatest loss in the satellite systems occurs due to signal propagation in space in the order of about 200 dB in either direction with the 30-meter diameter antennas. This value is for the 4/6 GHz systems, 1200-voice channel transcontinental communication via COMSTAR satellite in clear weather. The atmospheric and antenna tracking losses are less than 0.25 percent in clear weather. The rain loss, being frequency dependent, is low for the 4/6 GHz systems and can become significant for the higher frequency bands, causing a total system outage.

Chapter 3

INTELLIGENT NETWORK

The network environment may change because of a large number of internal and external conditions. The adaptive characteristic is considered to be one of the basic requirements of the intelligent network. The network may become overloaded or faulty: it may experience switching delays or inadequate standby channel capacity or any other network condition. Further the user of the source and the destination of the information may lead to extraneous searching before the right information is conveyed to the consumer. It is here that the built in algorithmic intelligence should monitor the network performance without the user or operator or any other human intervention. In essence the network adaptively responds to the commands that control and execute the entire range of communications functions.

Intelligent networks perform in two distinct directions. First, they have to actively process information to respond to the query of the user. Second, they have to adapt and fulfill the switching and transmission requirements to convey information from its source to its destination, wherever either one may be geographically or logically located within the network. The network intelligence can thus be grouped into its information processing aspect and its switching and transmission aspect.

3.1 Information Processing in IN

Networks are designed to serve a large number of users with a large number of queries, seeking a wide variety of answers. The queries may be in-depth or peripheral, they may have constraints and modifiers, they may seek solutions and modify the subsequent queries depending upon the previous answers, etc. For this reason, the network functions require a certain amount of sophistication to comprehend the query, seek the answer, and convey the right answer to the right user within a reasonable amount of time.

The front-end processors of the computers, which handle the user queries, need natural language processing capabilities. The software for processing natural language queries would exist at the user network interface. These processors require elaborate rules of grammar to comprehend the complex queries. Next, the required information needs to be identified and accessed. The information may be available locally within the local database environment of the computer handling the user query or may be available elsewhere in the network. In the former case, a search within the local database is initiated. In the later case, a query is dispatched to another node where the answer is available. Information that is supplied to the customer through a public domain intelligent network (PDIN) may be from an information vendor. Such information vendors are abundant. For example, a hospital may be searching for an organ donor. In this case, a national database with a list of donors may be maintained by a private facility, which dispenses its information for a fee. In this case, the user would incur a cost in receiving the information. The network has to determine the appropriate information needed by both parties (the customer and the vendor) so that the transaction may take place.

Such transactions have to be monitored by the unattended network via its own follow-up sequence of programs. The functions of locating the information sought by the user, the choice of the programs to execute the network functions, etc., become the software environment that drives networks. Thus, the nature of intelligence required in the networks to handle information access and retrieval, the comprehension of the user query, and the integration of the steps necessary in the procurement of the information sought become necessary sub functions of an IN. Concepts evolved from formal language theory provide the software tools to comprehend the user queries. Concepts evolved from knowledge engineering provide software tools to tackle the retrieval, the design, and the fabrication of the answers sought by the users.

3.2 Switching and Transmission in IN

The switching of information-bearing channels is essential in order to use a specific channel to perform a variety of service functions for the customer for a finite length of time. The switches perform in three distinct ways.

- First, the channel may be circuit switched, that is. switched by the network and allocated to a certain user for as long as the user may need it.
- Second, the channel may be packet switched. That is, the information may be packetized and dispatched to the appropriate destination, either as an individual packet or as a sense of packets in an appropriate sequence to complete an informational transaction.
- Third, the channel itself may perform the switching; in this case, the information is known to be channel switched.

The third type of switching generally occurs in private networks, and the customer (private branch exchange, PBX) exchange carries out the localized switching. In the first and second cases, the external network has the option to choose and switch between various channels, depending upon the availability, type of service, and source and destination of information. The transmission of the information also needs intelligence. The network can choose between a large number of paths and circuits. The network conditions strongly influence the path selected. These paths may be quite physical, such as a pair of wires, or they may be one or more channels multiplexed over one physical media, such as a wire pair, a coaxial cable, an optical fiber, a digital radio circuit, or any viable medium for data transmission.

Fortunately, a large number of adaptive algorithms have been developed during the evolution of the conventional telephone network. The newer and truly sophisticated networks offer a new breed of intelligence. The capacity to incrementally modify the network at the customer command would be impossible by the older Plain Old Telephone System (POTS) network hardware and its rudimentary software. The choice of the path may also satisfy certain other user-defined constraints, such as minimal cost, minimal delay, and low error. A different layer of the widely accepted network model handles the path routing intelligence. Processing of address information for communication also needs certain specialized capabilities. For example, the intelligence to wake up an idle channel and to test it before the actual communication needs very different procedures than those for cost minimization. Thus it becomes necessary to classify the types of intelligence necessary for the smooth error-free transmission of information. The smooth functioning of these networks becomes a fundamental requirement in light of the possible faulty conditions that may exist in the network, the transmission and the associated control computers. Intelligent networks have widely dispersed intelligence modules, software routines, micro coded read only memories, programmable read only memories and information modules (databases). It becomes necessary to provide access to the right module at the right time. Hence, some of the questions addressed by the network architects pertain to the types of networks and their environments, local and global intelligence, local and global data-bases, and accessibility in the communications, interfaces, sources, and destinations of the information communicated. In a typical layout of an exchange facility within an IN is depicted. This leads to the architectural issues, which govern the network design.

3.3 Components of Intelligent Network

3.3.1 Basic Components

Intelligent networks consist of sources and destinations of information, network nodes, transmission facilities, and networks to control the switching and transmission through the network. The need for a common channel-signaling network to support and control the information bearing channels has been recently recognized, and most intelligent networks have this additional network component in their architecture. These controlling networks adopt out-of-band signaling. It should be recognized that this mode of signaling (as opposed to in-band signaling) is only one of the possible ways to control intelligent networks.

3.3.2 Data Source and Destination

Customer, vendor data bases, billing systems, public domain information banks, and any information storage or handling facility are data sources, provided they can be interfaced with the network. The sources of intelligence, such as programs, routines, and modules, also become part of the control information. This information may reside as databases, programs, or operating systems for the interconnected computers within the network. However, the difference between information that passes through the network and the intonation to control and monitor the flow of customer information remains fundamental. Such a difference also exists in computer systems. The flow of data signals is distinct from the flow of the control signals within the computer architecture.

3.3.3 Network Nodes

Nodes of an IN perform a wider variety of sophisticated functions in comparison with the functions of the nodes within the telephone or computer network. Nodes may be designed as one of four types:

- Nodes in which no through data traffic is permitted (e.g. the end office or a computer port)
- Nodes through which only, through traffic is advised (e.g. a satellite relay station or a digital repeater bank)
- Nodes through which both through and terminal traffic are allowed (e.g., a metropolitan switching office that forwards information to other nodes and has spare capacity and links terminating at a local area customers)
- Nodes where numerous links meet to permit local and global exchanges of channels and where local and through traffic is permitted. (e.g. Nodes in a typically mountainous region of a national network)

In addition to the switching of information bearing channels, a typical node in IN manages and controls the flow of information in context of the services it is providing. Vendor services need the correct querying, retrieval and follow up. Administrative, maintenance, and billing information is appropriately exchanged. Diagnostics are administered and follow-up operator actions are communicated. In the context of the ISDN, the nodes serve to manage the B channels. As presented earlier the control and signaling information for switched channels is received on the D channel or in the appropriate X.25 protocol for the packet switched channels. The network control module responds to this information and acts to establish the digital connectivity of the B channels via the switching or the connection control module. Further the node may be called upon to send signaling information to the next ISDN central office regarding the activated and active B channels.

The node also provides access to the information vendors to supply the necessary information to the network users. It serves to alert the office administration and management network regarding the call progress, charges, rates, billing information, etc. The node can serve as a gateway office by providing access to other networks, such as packet switched or private networks. The resident modules are necessary to permit a dependable flow of voice, data, or video signals through the node. Diagnostic and maintenance functions are also undertaken by nodes to check out the network modules and their functionality.

In the United States where the subscriber loop carrier systems, private networks, and carrier serving areas are plentiful, the nodes of the intelligent networks serve as a greater diversity of functions. Private network interfacing calls for certain specialized functions from the node. Such functions are individually tailored to specific node or the central office. The more recent ESS has made such customization possible by the modular hardware and software approach used by the ESS architecture designers. In pursuing a modular approach these nodes are divided to pursue four basic functions:

- First, the peripheral switching modules interface with a 4-wire or 2-wire digital subscriber loop at any one of the basic or wide-band ISDN rates.
- Second, the main switching modules within the node for the local traffic carry out the switching function for the numerous data channels or the call is forwarded via the inter-exchange carrier network and the interoffice signaling network.
- Third, the communications module that provides a cross-connect facility between the switching modules carries out the localized communications between the various switching modules.
- Finally, the administrative module carries out the office administration and network diagnostic and routine housekeeping functions.

3.3.4 Transmission Facilities

Digital information may be transmitted over any number of physical media. The line coder, the transmit filter, the channel media (with or without repeaters, the regenerator and the line decoder constitute an entire digital facility. In the regional, national, and global networks, a large number of digital carrier systems are in place.

Metallic media has been used extensively to carry digital data in most of the local, metropolitan and regional networks. Transmission rates depend upon the distances and the coding techniques used. Local loops to the subscribers can carry data effectively at the basic access rate of 144kb/sec. The line rate can be as high as 192 kb/sec. With the proposed 2BIQ (2 binary bits convened to 1 quaternary level) code with refined digital echo cancellation techniques, the line rate can approach 1 Mb/sec over the shorter loops that span the Carrier Serving Area (CSA) location and subscribers at a maximum distance of 12,000 ft. The T1 carrier facility carries digital information at 1.544 Mb/sec extremely dependably.

Other rates for digital transmission facilities are 44.736 Mb/sec in the United States and 34.368 Mb/sec in Europe. The physical media at this DS3 rate are usually light-wave transmission facilities and free space with the 11 GHz digital radio carrier facilities. Higher bit rates at 274.176 Mb/sec (United States) and 139.264 Mb/sec (Europe) are also in service. The intelligent networks can be interfaced with some or all of these digital carrier systems depending upon the type of the node and its switching facility. These data rates are derived from the digital hierarchies of the different countries around the world.

These carrier systems are used in context of the intelligent networks to carry digital information over any given distance from one node to the next. The switching and transmission facility does not constitute an intelligent network unless its own control packets or its control network architecturally controls it.

Chapter 4

SIGNALING ENVIRONMENT

Vital differences exist between environments for human communication and those for computer communication. Human communication has been studied by social scientists in great detail and is of no direct consequence in the context of this chapter. However, almost all network data communication (except that used to control the network functions) is for eventual use in human interpretation and decision-making. The users exchange information to and from the network in a multiplicity of modes, (for example, voice, visual graphic, touch-tone, keyboard, scanners, fax, or even dialing). A degree of excellence in network service is a realistic expectation, as well as accuracy and rapid response from the network. The quality and response that the Intelligent Networks (INs) can provide exceeds far more than that of the conventional communication networks. For example, if visual communication in the High Definition Television (HDTV) mode is needed from a broadband intelligent network, the network must allocate about two orders of magnitude more bandwidth compared to the bandwidth for voice communication. Hence, the networks that provide integrated service need much greater sophistication to accommodate users and their individual needs.

Due to the complexity of rather sophisticated information communicated between the network and its users, both the audio and visual modes of multimedia communication for conveying the concepts are significant issues. Monitoring the rate, relevance, and quality facilitates the exchange of ideas, and accuracy of information presented by the network to the user. Flexibility of use and mixing numerous modes of communication (video and voice icon and real-life graphic, and mouse and keyboard, etc.) also remain key issues in multimedia communication. The enhanced demand and quality of services expected from the network places increased load upon its capacity to carry more information, more accurately, over longer distances, and over prolonged sessions of interaction. The functional requirements of the WAN place certain stringent demands on the components, algorithms, and interconnections that make up the network. Yet, the network functions have to be adaptive, intelligent, and command-language dependent. The unification of the requirements, components, and adaptation can be most easily achieved in a computer-like environment, thus forcing networks to look more and more like distributed processor computer systems that may span the entire globe as one multimedia communication environment.

4.1 The Signaling Network

The signaling in the switches (within the network) is analogous to the operation code (opcode) in the central processors (within the computers). Signaling conveys to the switches the commands and data to process and monitors calls much the same way as opcode conveys what to do with the data in the data field of an instruction to the control circuits of the central processing unit (CPU). The breadth and flexibility of the command structure in opcodes also exist in signaling in the signaling networks.

The signaling network permits the nodes within the network to communicate with each other, such that they cooperate with each other to execute the functions necessary for call processing. Examples of such call processing might be as simple as:

a) Seize (capture) an available channel,

b) Establish voice path,

c) Hold during call progress,

d) Release channel after call completion, and

e) Reallocate channel for future use.

In most modern networks, the variety of such commands and the flavor of the services provided by the switching systems is dramatically more elaborate and complex. And, for this reason, the standards, as they apply to signaling, are not only necessary within an individual country, but also throughout the globe. In the past era, human operators were deployed for completing message decoding, call completion, call monitoring, and call disconnect. Billing was a human function. The signaling network was the same as the communication network. From our modern perspective, the status of the network was a recipe for disaster.

Automation has taken network signaling through at least two major events: Common Channel Interoffice Signaling 6 (CCIS6) and Signaling System Number 7 (SS7), even though signaling systems 3, 4 (in Europe), and 5 in the (United States) existed transitionally. The modern signaling standards (CCIS6 and SS7) are complex and elaborate. Such topics are generally standardized after exhaustive discussions and detailed analysis by international standards committees. The signaling network is an essential pathway to carry standard signals. It is highly dependable and almost totally duplicated to ensure that its functionality is disaster proof. The signaling functions are vital to all network functions. In under developed nations, the same network is shared to carry signaling and voice or data between customers. Single and multiple-frequency signaling was used before Common Channel Signaling (CCS), which was approved by the ITU-T in 1968 as System 6. But in most of the modern telecommunications environments, the signaling network is a distinct and well-developed network in its own right and has a very secure existence in the background. Sometimes referred to as the "backbone-signaling network." It conveys and carries the control, Operations, Administration, and Maintenance (OA&M) signals from node to node. Now, in context to the INs, the OA&M becomes OAM&P (P for Provisioning).

Also in 1981, a signaling category, distinct from the network control signals, was initiated to facilitate the most rudimentary of the IN functions, that is the 800 service and the calling card service. This new category was in a query rather than a signal. The query would be for billing authorization (for calling card services) or call completion information (for the 800 services).

4.1.1 CCIS Network

The implementation of Common Channel Interoffice Signaling (CCIS) started in the United States in May 1976. Hardly a network then, CCIS was simply a link connecting two toll offices: the 4A toll crossbar office and the first deployed 4ESS switch toll center. After having proved its worth and potential, the link became a network over the eight years that followed until the divestiture of the Bell System. The signaling standard used was the ITU-T approved System 6 and was crucial to proving the advantage of the digital common channel signals over the single-frequency and multi-frequency techniques used earlier. Typical of the advantages that have been documented are the reduction of call setup time, from about 10 seconds to 2 seconds, thus improving the utilization of network resources. Additional benefits included improved reliability, reduced fraud, call tracing, call process acknowledgment, and optimal call routing.

The increasing network topology and its geographical expanse in the United States called for nodes in the network. These nodes are called the signal transfer points (STPs). Germane to the CCIS network, the STPs prevailed well before any assemblance of an IN was implemented anywhere (not until 1981). Before the divestiture of the Bell System, the CCIS network operated with ten regional signaling sectors in the United States.

Each region contained two STPs. The locations of these individual STPs within the regions were chosen with due consideration to survival and recovery of the network after natural disasters and disastrous network conditions.

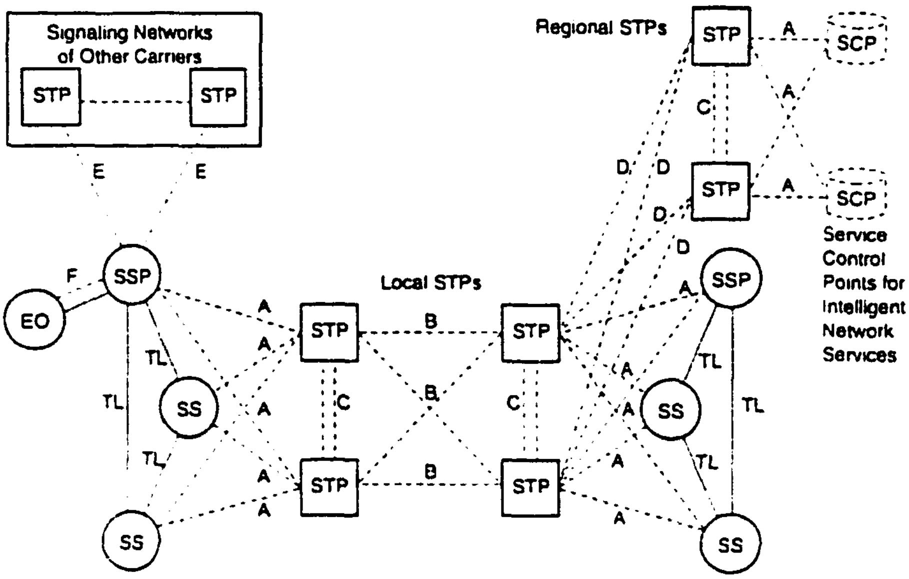


Figure 4.1. Typical Interconnection of STPs in SS7 Signaling Network

STP = Signal Transfer Point	SCP = Signal Control Point
SSP = Service Switching Point	EO = End Office
SS = Switching System	TL = Trunk Lines

Three types of links span these STPs. Type A links connect the switching offices with the STPs. Type B links connect the STPs of neighboring (adjoining) regions. Type C links are specially balanced, mated pairs that interconnect STPs within regions. In a sense, the links, STPs, and the switching nodes that communicate with the STPs within this network all have unique logical and geographical addressing capabilities.

Each link synchronously transmits a continuous stream of data in both directions. The data stream is divided into various signaling units (SUs), each 28 bits with 20 information/data bits and 8 parity bits. Trunk signaling messages are routed through the STPs by using a band and a label code that uniquely identifies the virtual circuit. Each SU can hold up to 512 bands and 16 trunks to be able to identify and address a maximum of 8192 voice channels. Like the opcodes of complex computer systems, the SUs have predefined code formats to permit a variety of functions to be invoked at the switching offices. Dynamic overload control permits the switching systems to function more optimally with the CCIS network because it anticipates traffic congestion and permits automatic rerouting and out-of-sequence transmissions of SUs and message units.

Twelve signal units make up one block and the last or the twelfth signal unit is the acknowledge unit (ACU). Retransmission under unrecoverable errors is thus built into the CCIS network. Analog or digital transmission is feasible in the network. In the case of analog transmission, the data modems at each terminal operate over standard analog voice band channels. For digital transmission, channels are derived from the bit stream of pulse code modulation (PCM) systems, like subframing of T1 channels or separate digital channels. The initial data rate of the CCIS was 2400 bps (analog) or 4000 bps (digital) with possibilities to 56 or 64 kbps. The signaling format on the CCIS has a well-standardized format in most countries, though not on an international basis.

CCIS expanded rapidly to encompass over 100 toll centers in the United States by 1980. Increased traffic volume called for increasing the bit rate in the CCIS network from 2.4 kbps to 4.8 kbps, and the number of wire pairs in the CCIS network was updated from 10 to 16. Message encryption was also included for security when the data rate was enhanced to 4.8 kbps.

4.1.2 Signaling System 7 Network

CCIS6, approved in 1976 by ITU and deployed in the United States, underwent yet another scrutiny starting in 1980. International standards for implementation using a layered protocol structure were the ultimate goal for voice, data, and any other multimedia communication around the world. This quest also resulted in the Open Systems Interconnection (OSI) seven-layered reference model with the higher layers geared toward applications and the lower layers geared toward network and transport of data The layered or modular approach to specifying the communication protocol is the framework of SS7. The standard was approved by ITU in 1984 and published in 1985.

In a sense, the Signaling System 7 (SS7) has an evolutionary architecture of its own. The exchange of signaling information also follows the OSI model. Thus, all the features inherent in the OSI data exchange model exist in the exchange of signaling information in the SS7 network. The SS7 network architecture in the United States has undergone considerable enhancement during the 1980s. In divestiture of the Bell System, the resulting numerous local exchange carriers have increased the number of STPs dramatically. Provisioning for the access and interconnection between the STPs needs more numerous links and more types of links. The increased numbers of IN and ISDN networks being installed have required a far greater sophistication in the signaling environment. The SS7 environment addresses these needs effectively, resulting in a much more versatile architecture and protocol compared to the earlier signaling systems. In the composite functions of signaling and communication, the signaling flows through the SS7 network and the user information flows through the communication network (in all circuit-switched configurations). The interdependence is as essential as the interdependence of control circuits and data paths in computers. The analogy is complete, and the composite network can be built like a cosmic computer with distributed processing and switching capabilities. There is no single switching system (or CPU) that controls all functions; instead, many localized switches (or CPUs) seek mutual cooperation.

Being slightly adapted in the United States to include the American National Standards Institute (ANSI) version of the protocol, the original ITU-T standard has the widest acceptance and perhaps has been the single, most dominant influence on the unification of the communication industry. From a global point of view, there are three layers to this protocol. The lowest layer, called the Message Transfer Part (MTP), corresponds to the lower three layers of the OSI model, that is, Network (Layer 3 of the OSI Model), Data Link (Layer 2 of the OSI Model, and the Physical (Layer 1 of the OSI Model). It forces any two signaling points in the SS7 network to route information accurately, dependably, and quickly. The next higher level has the signaling connection control part (SCCP) that corresponds to the Transport Layer (or Layer 4 of the OSI model). It finds a means for the transportation of the messages between the signaling points. The telephony user part (TUP) is essential for call set up and call disconnect in the regular telephone networks. The Transactions Capabilities Applications Part (TCAP) is used to find the applications with data bases like 800 service, alternate billing service, or credit call service deals with ISDN Users Part (ISUP), and OMAP deals with the "O" perations, "M"aintenance, and "A"dministration " P"art.

The SS7 signaling model is derived from the more generic and universal OSI model to represent its own seven layers:

1. OMAP and Application Service Elements (ASEs) with the TCAP for OSI Layer 7.
2. ISDN User Part (ISUP) for OSI Layers 7, 6, 5, 4, and 3 part.
3. SCCP and the signaling network for OSI Layer 3 part.
4. Signaling link for OSI Layer 2.
5. Signaling data link for OSI Layer 1.

The layout of the SS7 network is more elaborate than that of the CCIS6 network because of much greater latitude in services and signaling carriers, and separation of the United States into the seven regions that the Regional Bell Operating Companies (RBOCs) serve. Further, the local STPs and regional STPs need communication because the service control points (SCPs) are maintained on a region-wide basis. Still further, the access tandem or the service switching point (SSP) of an IN needs communication with other carrier signaling networks. In most cases, the architecture can be much more complex. Six major links are identified in the SS7 network. The first three links (types A, B, and C) correspond to those discussed earlier in the section "CCIS Network" that describes the CCIS6 network. Type D links interconnect local STPs with the Regional STPs to query IN-type of information from the regional SCP. Type E links interconnect the other independent carrier with the access tandem (AT) or the SSP node of the IN. Finally, the type F link interconnects the non-SSP type of end office with the SSP. The need for these additional links becomes evident from the IN architectures discussed previously.

The link rate for SS7 is normally at 56 kbps, although 64 kbps is likely when data transparency in ISDN is clearly available. These links can also operate at 9.6 kbps. When ISDN is completely implemented, its bearer channels may serve as the SS7 links. With a total transparency in the SS7 network, most of the IN functions, such as 800 Service, CLASS, and interoffice data transport, become realizable. Selective call rejection, caller identification, and automatic callback features can he introduced by simple programming steps within the SSPs of N/1 + and IN/2.

4.2 Data and Communication Networking

From the current perspective of data exchange, there is little that is different between the two types of networks. Transmission protocols, signaling, switching, and the applicability of the OSI reference model prevail in both. Each of these networks has evolved because of the intense and immediate effect of social demands, such as the need for computers in the society and the need for human communication. The computer-to-computer connectivity, the bursty nature of data transfer, the differences in accounting and billing, differences in security requirements, and the prototypical aspects of the data exchange process make the computer networks easier to implement, monitor, and control. However demanding the task, computers rarely elicits any emotional response from another computer. These reasons account for the meteoric growth of data networks over the last two decades compared to the more subdued, controlled, and more-or-less optimal growth of the communication networks.

Historically, the emergence of each of the two networks has followed a different trajectory. The computer industry has been the chief proponent of the SNA concept (IBM's System Network Architecture) and the DECNET concept (Digital Equipment Corporation's Network Architecture). Being unregulated and responding to the growth of multi computer and distributed processor systems, the industry developed a set of de facto standards and protocols.

The communication industry, though comparatively sluggish, has a history of facilitating data communication. The transition to all-digital telephony was initiated when the concept of short-haul digital carrier, T1 systems, program control of switching systems (No. 1 ESS), and channel banks, D1 controllers were introduced in the telephone systems in the early, mid, and late 1960s, respectively. The environments and constraints are quite different from those in the computer networks; however, recent strides are more impressive. For example, the Public Packet Switched Network (PPSN), the Integrated Network Access (INA), and the Integrated Special Services Network (ISSN), the provisioning of Digital Services Nodes (DSN), the Circuit Switched Digital Capability (CSDC), and most recently the ISDN are good indicators of two major trends:

- First, that all digital networks can be elegantly integrated.

- Second, that shared networks are the optimal networks from the point of view of cost, control, resource allocation, service provisioning, error control and economic survival.

It is important to realize that these networks do not meet the requirements of being intelligent. Instead, these networks are scenarios for the IN concepts in which the computer architecture and communication networks are intellectually wedded, going as far back as 1980 [1]. The influence of universities, United States federal agencies, and the entertainment industry is also noteworthy. For example, the ALOHA network from the University of Hawaii and the ARPANET from the Advanced Research Project Agency began the concept and implementation of the packet-switched networks. The LOCALNET using some of them CATV standards for frequency allocation, and the WAGNET, using some of the CATV components, provide data communication over cables.

During the early 1980s, a subcommittee formed by ANSI and ITU-T committees meticulously executed the groundwork for the sophisticated ISDN. As a result, the Bell operating companies in the United States and the various networks in Canada, Japan, Europe, and Australia have already offered preliminary ISDN-type services to customers. During the same time, a certain amount of network intelligence was placed in the network elements to substantiate the newer services, such as, call forwarding, call waiting, automated re-dialing, and the 800 and 900 services. This embedded intelligence, in the form of the system utility programs for the Electronic Switching Systems (ESSs), is finding its way in the networks. These simultaneous demands of ISDN and INs on the modern networks have had a profound impact on the functionality of the switching systems and their controlling software. These new switching systems blend their three basic functions (switching, control and administration) with service provisioning in the most innovative and unique compositions.

Such capability, however, can only evolve slowly and become more encompassing in the networks. Achieving sophisticated service functions with computational accuracy, elegance, and optimality can only be gradually introduced. The maturity of the computer industry through the fourth and fifth generation machines must find an expression in the network industry over the next one or two decades. The dimensions of Artificial Intelligence (AI) now being evolved for computers can be brought home in the network environment. The more sophisticated network services, now demanded by a large fraction of the customer population, can be routinely and inexpensively handled by the AI embedded in the micro code of the network computers. If the society demands inexpensive, high-bandwidth communication for media interfaces between the customers of the network, then designers develop a synergy between the fourth-generation computers and the high-speed optical media now. Only an innovative and conducive socioeconomic atmosphere can cross-fertilize the concepts and foster the technology between the two industries to have a new discipline for network sciences to emerge.

Consumers have created the need for intelligent networks and such intelligence is being actively brought into the network environment. The migration of concepts and algorithms is presently underway from the computer field, optoelectronics, and circuit and system sciences. It is interesting to note that the migration was in the opposite direction from electrical network theory to computers while computers were in their infancy. The growth of these intelligent networks appears as imminent as wafer-level integration in the semiconductor industry. In fact, all the technological roads lead to expansion and excellence in communication.

Numerous examples of INs exist in Western countries. These networks are mentioned here in this chapter but explored further in Part II. Most of these networks are mature and functioning in the United States and in Europe. In 1991, the 800 service was already commissioned in 33 countries around the world, and video conferencing (a form of private virtual network-PVN) was available in 8 countries. In the United States, the 800 service or IN/1, and in West Germany the 0130 service, both function via a data base query. In Europe, these services are also called "Green Number Services." When any user requests this service, the call is dispatched to the nearest SSP, which may be an end exchange office or a tandem office. Here, a data base query is initiated to the service control point. This query-setup procedure uses the SS7 network. Signal transfer to the appropriate database would occur via the STP should it be necessary. The target number of the called party would be recovered and the call would be completed. In the enhanced operation, the network service may depend upon additional factors, such as the calling party number identified with automatic number-identification capability, time and date of the call, working hours for the office to which the call is targeted, and caller interaction if necessary.

Another version of the IN is implemented in Alternate Billing Service (ABS). In this case, the charges for the call may be made to a number different from the calling number, or to a calling card number or an authorized account given by the RBOCs in the United States. A data base query is established to validate the authorization to charge the call. Operator intervention is eliminated, and call completion occurs by the network.

The Emergency Response System (ERS) is implemented in two ways. One way uses an operator-assisted system. An emergency response center is alerted when the call occurs from predefined areas. An abbreviated dial code, such as 911 in the United States and 110 or 0110 for police in Germany, or 0112 for fire/ambulance, also in Germany, alerts a human operator in the emergency response center for that particular area. Then, a human dialog precedes the actual response to the emergency.

In the second, more intelligent environment, the call triggers an Automatic Number Identification (ANI) system for the calling number. A data base query is initiated, and the real target number of the Public Safety Answering Point (PSAP) is connected through the public switched telephone network. In this case, the flow of information becomes somewhat complex because the ANI of the calling number and the PSAP both are needed. The PSAP attendant receives the caller phone number, the caller's name, and the caller's location. This entails two data base searches, first to find the nearest PSAP, and second to access the calling party's phone number, name, and address.

Packet network is used for quick response with X.25 messages between the SCP and SSP nearest to the PSAP. In the United States, the SSP generally responds within 3 seconds, and the SCP response is targeted for 1.5 seconds for 99% of the calls. At the present time, there are three versions for the implementation of the IN-handled ERS. All three work satisfactorily and used to suit local traffic and geographic locations of the PSAPs.

PVNs also manifest themselves within IN environments. Typically, these networks aim to replace the privately leased lines that most of the governments and major organizations hold for their specialized needs and specific requirements. These networks take on the responsibility of being fabricated on demand, rather than being totally dedicated to the user. The intelligence built into the network switch of INs permits them to create a virtual network between predefined points within any network quickly and inexpensively. The parallel exists between the virtual memory concepts within the computer systems and virtual paths between nodes of an IN. In the former case, the operating system provides more pages of a physical memory than what is actually available by emptying out those pages of physical memory not in use. Users believe that the machine is giving them all the memory they need regardless of the physical size of the memory. In the later case, the networks assign the channels not currently in use to the PVN-user quickly and efficiently for as long as the user needs them, with the features specifically suited to his or her individual use. Thus, users think that they lease these channels, whereas in fact these channels are virtual.

4.3 Network Types

Public domain networks and totally private networks constitute the extremes of network ownership and, thus, the organization and management of networks. Whereas considerable freedom exists in the architecture and operation of the private network, the architecture and operation of public domain networks tend to be standardized and streamlined. In our discussion presented here, acknowledged international standards are used in context to the public domain intelligent networks (PDIN), especially for the ISDN.

ISDN has evolved in the same time frame, as INs were being developed and implemented. ISDN is not an overall services-oriented network even though it has attempted to make the telephone network all digital and provided digital services to the end user by providing access to its circuit-switched, packet-switched, and message-switched facilities. For instance, there are no 800 (Freephone service, an IN/1 feature), call forwarding (an IN/1+, IN/2 service), etc. in ISDN. To this extent, ISDN is a digital services network and not a digital network that is service oriented.

From a purely ISDN perspective, the evolution of the public domain network has only two well-planned steps. The first step calls for the provisioning of all customer interfaces to be at the standard subrate (B+D), basic rate (2B+D), H0 (5B+D; 384 kbps), H1 (11B+D; 768 kbps), primary rate (23B+D; 1.544 Mbps) or (30B+D; 2.048 Mbps), or broadband ISDN. The CCS will be universal throughout the region, nation, and the world. Thus, digital connectivity can be provided to the user.

The second step calls for a fully integrated network at the Central Office level calling for the digitization of all data transport facilities and full interconnectivity from any node to any other node. Both the circuit and packet facilities would be fully deployed in this integrated network for high-capacity circuit switching and wide band packet switching. This includes the SONET network at the "Packet" level and the ATM network at the "Cell" level.

There are three major architectural variations in the networks used in the public domain for the transfer of data. In this connection-oriented data transfer, there are five stages for the data transfer to take place through a switched channel. The sequence is summarized as follows: an idle channel is identified and tagged, a connection is established, the data is transferred, the channel is released, and the channel resumes idle status once again.

This sequence of stages closely parallels the well-established network steps in a typical voice call that follows the five network functions: call setup, alert, connect, disconnect, and release.

In context to ISDN, the information necessary to accomplish these individual steps is incorporated in the ITU-T Q.931 protocol at the network layer used over the D (Delta) channel to set up the B (Bearer) channel. Circuit switching is preferred for its simplicity, constant-delay capacity, and constant-link capacity.

In the circuit-switched mode, the circuit setup time can be excessive, and single-user numbers can become busy or error-prone over prolonged periods. There is no error-correction capability. In the modern telephone systems, which are circuit-switched, additional software layers that monitor the lines and provide operator/network intervention, call waiting, and call blocking have corrected most of these drawbacks.

In this connectionless-oriented data transfer, a packet of information is assembled with its own address such that it can be forwarded to its destination. The packet may transit through any number of nodes until it reaches its destination. The exact physical routing is not known, but the innate functioning of the participating nodes through the network assures the correct transmission of the packet. In many ways, the packet can be compared to a letter dropped in a mailing system that assures the delivery of the letter to the right destination.

The receiving host sends an acknowledgment back to the sending host when a packet is received error free. The time of arrival of the original packet at the receiving host, or the acknowledgment packet at the transmitting host, are both a matter of chance, network topology, and traffic dependence. In context to packet switching, ITU-T has specified the widely accepted X.25, X.75, X.28, and X.29 protocol at the network layer. Typically, Ethernet standards dominate the packet-switching usage for LANs, and X.25 for WANs. With Fiber Distributed Data Interface (FDDI) standards, chips, and interface, the availability of the packet-switched networks can get extremely widespread. SONET and ATM technologies provide additional applications. Typically, the packet networks are controlled for their flow and for any possible errors. Being highly multiplexed and time-shared, they are highly efficient and very inexpensive.

In this message-oriented data transfer, the "user agent," such as a computing facility, a node, or any terminal, first receives a message of information in its entirety. The sender address and identification are also taken at the same time such that they can be forwarded to their recipient.

Both addresses are delivered together with the message and the message header (or the subject). The agent receiving the message can choose the message path and its orientation, that is, circuit or packet. The message itself may transit through any number of nodes until it reaches its destination. The final node puts the message received in an appropriate queue to be delivered to the recipient. The exact physical routing does not concern the sender or the recipient. The user agent assumes the responsibility of delivering the entire message and bills the users of this type of messaging facility. In a sense, the message switched facility is not truly independent of the two basic switching orientations, but rather a form of value-added service provided to the end user. Message accountability, interoperability with numerous networks, and message acceptance at peak periods are some of the advantages of the message-switched systems. At extreme conditions, the message files can overload the allocated disk space at the intermediate nodes and some form of temporary "dump" becomes necessary. Variable message delay is also a feature of such networks.

These networks can work cooperatively. One of the objectives of the network switches and their nodes is to assure the compatibility of the two major network ideologies of data transfer, that is, circuit-switched and packet-switched. Network switches have an added burden in the evolution of newer combined networks. Whereas the ISDN perspective provides integrating the two types of data transfer, it does not provide for network intelligence. From a global perspective, the demands on the switching functions for IN implementation are beyond the requirements for ISDN. For this reason, the switch designers contend with two types of enhancements for future switches. First, they should provide the fully integrated network service and wideband packet switching. Second they must provide the SSP function in the IN environment.

Switches perform differently for IN functions than for ISDN functions. From a pragmatic point of view, the public networks of most countries need to evolve around IN, ISDN, or both, in addition to being a "plain old telephone" network. Concepts culminating in the present standardization and implementation of the intelligent ISDN have initiated the evolution of new breeds of networks whose future characteristics can be tailored by switch designers now.

A fourth supporting network, which is essential because of the need to control and monitor the flow of information, is the CCS network. In a traditional sense, the ideology behind Common Channel Signaling was implemented in the pre-divestiture Bell System in 1976. During this time, an interoffice network specifically named CCISS6 (for "Common Channel Interoffice Signaling System") based upon ITU-T Signaling System 6 with various STPs was introduced. This signaling system with its own signal transfer points provides the network with the ability to selectively accomplish the control functions on the information pathways through the network. To this extent, the programmed I/O handling of computer systems was being implemented in the communication networks.

In a broad and general sense, signaling is essential to the functioning of any IN. In both INs and ISDN, the standard SS7 network adopted in the United States and Europe carries out the signaling. The transition from the older modes of signaling is gradual, if not untidy and expensive (such as, in-band multi-tone, or trunk signaling practiced in the older switching centers). The current SS7 network uses the out-of-band information (that is, the information on the D channel for ISDN) to control and signal the various switches to complete and monitor the B channels. This network becomes essential in the circuit-switched context because the B channels provide transparent end-to-end digital connectivity for the network users.

If the user invokes a transition to the packet mode, the control information is also passed onto the network via the signaling bits on the D channel. These bits are then communicated to the packet network and the channel initiation procedure starts. The entire network signaling thus takes place over this D channel closely tied into the SS7 network.

Private and semiprivate networks (sometimes having the features of channel-switched networks) may also use the standard components and the associated interfaces. Generally, these operations include private branch exchanges (PBXs), private and dedicated lines, and a digital distribution facility. Any amount of computerized information-handling capacity may be encountered in private (intelligent) networks. These networks are emerging with surprising amounts of sophistication and intelligence. Typically, such networks are found in the scientific community (for example, the National Science Foundation's NSFNET [now obsolete], Defense Advanced Research Projects Agency's ARPANET, and Carnegie Mellon & Bell of Pennsylvania's Metropolitan Campus Network [MCN]). Other examples from industry (for example, VLSI vendor's networks) and national laboratories also prevail (for example, Lawrence Livermoore National Laboratory's private network for their 12,000 employees, spanning 500 buildings). The Internet has also emerged as a major provider for Wide Area Network (WAN) services. The major feature of this network is that it uses effective dynamic bandwidth allocation algorithms and provides frame relay, switched multi-megabit data transfer, high-speed circuit switching, fractional T1/T3, and the international X.25 capabilities. Its generalized architecture enables tying into existing LAN or MAN, with token ring, FDDI, and even PBX architectures.

In business environments, two emerging networks in the competitive telecommunications environment include the Private Virtual Network (PVN) and Area Wide Centrex (AWC). In the former type of network, users retain a certain amount of control and management over their own PVN. It is tailored to the specific needs of the business by using the resources of the Regional Bell Operating Companies (RBOCs) in the United States together with the services offered by the Interexchange Carriers (ICs). The user pays for the resources, and the RBOC offer their availability. Thus, the physical allocation of channels or resources is not committed to the user on a permanent basis. Presently, the leased lines may be used for voice (private tie lines), data (leased 1200-9600 kbps lines), high-speed data (leased 56 or 64 kbps lines), and broadband data.

This type of option has become feasible only because of programmed switching throughout the network and the topology of the PVN, which can be modified. It offers certain attractive features to the users, RBOCs, and the ICs. The user pays for the resources used, with the network's availability to the entire company network, and retains multiple-location access to the PVN. The RBOCs retain their market share, pass the management responsibility to the user, and share investment in new technology among numerous users. The ICs benefit because of the increased use of their carrier services.

AWC offers users some of the private exchange features via a Central Office. Though not truly intelligent, some interesting features already exist in the systems, such as a centralized number plan, abbreviated dialing & call forwarding, and conference calls. These features of any Centrex system are extended to many exchanges in AWC. The SS7 is activated in this network, and the application program, tailored to the specific AWC, translates the dialed number to the actual number according to the defined numbering plan for that service subscriber. The target number and the service logic instructions for handling the call are forwarded to the SSP via the SS7 network. The STP also participates in this transfer to the SSP. The protocol used in this transfer between the SCP and the SSP is part of the SS7 protocol that is, the transaction-capabilities application.

4.4 Intelligence in American Public-Switched Networks

This section presents the history of the evolution of phased intelligence in the public-switched networks in the United States of America. It is important to note that the prolonged divestiture of the Bell System (1982 through 1984) was effective during the prime evolutionary period of the intelligent networks (since 1980). Hence, the origin of the seminal concepts can be traced back to the pre-divestiture Bell System of the early 1980s. Some of the techniques were specifically evolved toward bringing the newer services to the local telephone customers. Later, such facilities and plans fell under the jurisdiction of the Bell operating companies and were implemented at the regional level. Some of these services were introduced by the individual RBOCs. The concepts leading to refined intelligent networks to serve the RBOCs were evolved at Bellcore (Bell Communications Research). It is here that the concepts leading towards IN/1, IN/1+, and IN/2 were developed. One of the key (and socially accepted) group of functions of the intelligent network had been introduced in 1984. The custom local area signaling services (CLASS) actually perform the intelligent functions in the SPC-based, No. 1 ESS Central Offices by being able to selectively reject calls, perform auto recall or callback service, etc. Once again, the CCIS capability (to relay to the called party the specific caller information, that is, caller number) was used by programs stored in the Central Offices to force an intelligent network response. Call-related information could be exchanged between Central Offices via the CCIS6 extended message capabilities, thus providing the automatic number identification (ANI) feature between switches. At the rudimentary level, most of the CLASS services are based upon the ANI feature and CCIS6 extended-message facilities between the Central Offices.

The growth of preprogrammed algorithmic intelligence into communication network fields has been steady. In 1985 (after divestiture), the Software

Defined Network Service (SDNS) was introduced. This service permitted many private network owners to take advantage of the more dramatic advances taking place in the switched network operations and yet be able to keep the advantage of their prorated network. The virtual network defined by software commands permits private network owners, such as corporations, to use their own network in virtual cooperation with the switched public network.

The business and/or residence custom services, also introduced in 1985, permit users to access the generic programs embedded in the Central Offices. With access and options to exercise these powerful Central Office routines (functioning much like computer system utility routines with appropriate input data), the customer has the ability to use Centrex, custom calling, and electronic tandem switching capabilities.

Numerous developments in other directions, such as billing and authorization, also facilitated the newly enhanced intelligent functions to be commercially viable. While the network may respond functionally to signals and codes forwarded by the Central Offices and users, the collection of tolls and revenues for the service provided is also crucial. To facilitate this aspect, billing information was validated by authorization in the database administration systems distributed within the network. Initially introduced for the 800 calls in 1981, this facility also paved the way for other services, such as calling-card services. The human (operator) intelligence in handling the queries and call processing was gradually replaced by the programs. These programs queried the distributed databases for authorization to connect and process calls by algorithmically decoding the long strings of digits to the called numbers through the networks.

The concept of call routing was also introduced as far back as 1981. The CCIS6 provided the 800 service facility to respond to these calls by matching the location of the user to the nearest or the most logical service provider. Intermediate databases to provide the routing information were distributed throughout the network. Thus, the functions of obtaining the appropriate authorization, call routing, switching, and billing were all streamlined into one programmed and advanced network operation.

The acceptance of SS7 by ITU-T standards committees permitted the eventual integration of the numerous procedures under one common conceptual and architectural umbrella dealing with intelligent networks, their services, and their protocols. Generally SS7 becomes central in streamlining the flow of information between three basic modules, the Service Control Point (SCP), Service Switching Point (SSP), and the Tandem Access Office (TAO) of most intelligent networks via the Signal Transfer Point (STP).

SSPs sense the unique set of circumstances at a switch and query a database at the SCP. Functions, such as routing control, digit collection, billing and recording, and network management commands, take place at this switch. Incomplete information at the SSP causes the query to the SCP. The database resident at the SCP contains a wide variety of network service information. Typical examples of the databases are the line information database, Bell operating company 800 database, and business services database.

The various service-switching points provide basic access to the SS7 signaling network and switching facility within the network. They respond, by completing the original command from the switch, or they furnish the information queried by the customer. In their functionality within the more advanced networks, they may also access their own databases and provide localized services. In the 800 networks, the remote database is queried after a trigger condition (such as an 800 number, or a 911 number) is determined.

The TAOs provide the exchange of information (data base, network and/or service) to distant locations via interexchange carriers. If the SS does not have direct access to the SS7 network, or a SCP, then the tandem access office provides access.

The goal of most network designers is to create architecture of the breed of intelligent networks to make the provisioning of any service total independent of the type of service they provide. The transport of relevant information is the prime objective of ingrained network intelligence. From the perspective of the network designer, the intelligent network of the future should be able to provide any network service, at any time, at any place, any customer authorized to call upon the service. Most of these features are incorporated in the IN/1+ and IN/2 initiated from 1988 to 1989. The aforementioned chronology indicates the framework for introducing preprogrammed modular procedures in the American public-switch network. The concepts and technology to bring network programmability most existing networks prevailed a long time prior to these dates. The public-switched network in the United States, being as extensive and dependable as it is, needs a considerable amount of time to be altered and enhanced. Truly intelligent smaller private networks have been introduced earlier to the communication facilities. Other more powerful and sophisticated networks also exist in the corporate, banking, and computer environments.

4.5 Public Domain INs

This section discusses the viewpoint of AT&T and that of the Bell Operating Companies concerning INs in the public domain.

4.5.1 Viewpoint of AT&T

The IN/1+ provides for the use of intelligent peripherals to interface the intelligent networks of the American RBOCs and/or other intelligent nodes. Service independence, faster feature introduction, and standard node interfaces at various switching centers will also be the features of this IN environment. This facility has been in operation during the early 1990s. By the mid-1990s, both the Universal Information Service network and Intelligent Network 2 (UIS/IN-2) are proposed. Voice, visual, and computer data transportation is the ultimate goal of these networks. Independence from the origin and destination of services and the service provisioning at any time is a willful objective of this network. The implementation of this policy permits interfacing with any third-party vendor/user node in the 1990s as off-network nodes (ONNs). Enriched services to the customers will be provided by the use of intelligent peripherals, tandem access offices, and the SCPs for the UIS. The proposed architecture will be able to function in a new type of service creation environment tailored to the long distance and national customers. This environment also provides the framework in which services (such as, national polling service, 700 number services, and global educational services), concepts, and offerings may be quickly and easily tested before fully introducing them in the marketplace.

Localized network control information may be derived for users (provided the authorization exists) or network management programs. This permits the use of software-defined network functions and configurations. Such services become extremely feasible, and they could be made for a short duration and for immediate availability. The goal is to provide new nationwide IN/2, AIN-type of telecommunication services by offering a PVN-type of flexibility to the customers.

4.5.2 Viewpoint of Bell Operating Companies

The 800 services network represents the first generation of a public domain network. Being functional and well received in the public domain, it provides the three basic features of most INs. First, stored program control of switches and nodes are used to provide a software-controlled environment. Second, the interoffice signaling is accomplished by a Common Channel Signaling System (CCIS6 and/or SS7). Third, few service-independence features can be incorporated later, such as, transfer of control, connection control, etc.

Compared to the potential and the capabilities of these networks (as they are understood in the late 1980s), the 800 services are truly a first generation IN. Architecturally, the 800 network uses five of the seven building blocks that make networks truly flexible and intelligent. A detailed description of the architecture and the components of IN/2 are presented in Part II of this volume. The evolution towards IN/2 through the mid-1990s is the objective of many Bell Operating Companies (RBOCs).

The 800 network uses SCP/1 extensively (the "/1" indicates that it refers to IN/1 for all the building blocks). The SCP hardware and data base configuration that resides in these SCPs may be duplicated throughout the country. In the RBOC's environment, the data that is pertinent to that local region is stored, generally. The data is managed and maintained by a single service management system (SMS/1). Since the features of the 800 network are not completely evolved throughout the country, the SMS/1 may be used to control one or several SCPs. Updates and modifications can be carried out using simple protocols for the SCPs. In the RBOCs, the packet network protocol user is termed as BX.25 and is slightly different from the normal X.25 protocol. Compatibility with the other protocol within the SSP, STP, and SCP communication has prompted the RBOCs to use this rather specialized BX.25 protocol.

Alternate Billing Service (ABS) and Private Virtual Networks (PVNs) are additional IN features evolved for the RBOCs. These networks and their specialized features make newer services available more quickly, more economically, and on a more customized basis. The incorporation of INs into the switching environment, together with ISDN in the customer and business environment, opens new opportunities for the sophisticated user.

4.6 Technological Feasibility

The system and architectural proposal for the future INs do not need any major technical breakthrough. The detailed ITU-T recommendations are adequate and clear to implement most of the IN functions with current technology. The features of the public domain networks are being offered to individuals and businesses by numerous RBOCs. Other private networks, such as those owned by financial networks and the travel industry, have many advanced features not discussed here. However, the major changes, such as the acceptance of the Optical Carrier (OC) rates and ATM are also emerging in the telecommunications networks. It is important to understand that fiber can enhance the data rate by two or three orders of magnitude. The control logic and the programmed logic for switching, interfacing, protocol, error control, and so forth, are also evolving rapidly as fibers are introduced into public networks. The emergence of fiberonics may eventually facilitate entirely new networks and switches for a distant tomorrow. But in the meantime, the popular silicon technology provides all the tools for the INs of next few years.

Chapter 5

NEWS NETWORKS

The world has never been more abundantly informed about itself. Computers, satellites, and high-speed data links have revolutionized communications in the past four decades. Information is transmitted at ever-greater speed and volume across national frontiers thanks to scientific progress. If we imagine the thousands of communication lines we can conjure up the image of a global village whose inhabitants are linked.

Press and broadcasting organizations use new technologies to send and receive reports from all over the globe, free from the wordage restraints and delays that hampered reporting in the days when dispatches were transmitted laboriously via cable. A single high-speed burst over a satellite circuit can move tens of thousands of words in the time a telex machine used to take to send a hundred. Live television broadcasts show events as they are happening on the other side of the globe; news photographs are in print around the world within a few hours of having been shot.

5.1 Historical Perspective

Western news organizations make up the bulk of the world's commercial media. Four major international news agencies are leading in collection and distribution of news on a worldwide scale. It has been estimated that they account for 80 percent of the immediate international news circulating around the planet each day. Their services are indispensable to the biggest as well as the smallest newspapers, to major television networks as well as tiny local radio stations. They distribute their news to rich and poor countries alike. They are little-known to the general public, and usually work on a wholesale basis, selling their services to newspapers and broadcasters that select what they want to retail to their reading, viewing, or listening audiences. The agencies disclaim the exercise of power; they owe responsibility only to their clients and to their own standards

5.1.1 Four Major News Companies

The four major international news services Reuters, the Associated Press (AP), United Press International (UPI), formerly known as United Press (UP), and Agence France-Presse (AFP) have the same basic purpose as the 103 national news agencies around the world: the collection of news on behalf of recipients who are thus able to get more information more quickly and more cheaply than they could if they depended upon their own resources. What sets the "Big Four" apart from the rest is the worldwide scope of their operations and their independence from external control-either from governments or from powerful clients.

Based in the United States and Europe, Reuters, the AP, UPI, and AFP span the whole range of general-interest news reporting from summit conferences to college sports, from airline crashes to humorous fillers. Each works independently of the others, but much of their information covers the same events in broadly similar style. Their main concern is for the speedy reporting of immediate news, known in the trade as "spot news." In addition, they provide feature articles, explanatory background pieces, descriptive accounts of places and people, and other material ranging from Hollywood gossip to weather forecasts. In their wholesale role, they supply news and features to the press, radio, and television, but they also sell their services directly to government departments, companies, international organizations, and anyone else who needs a twenty-four-hour service of immediate news and is willing to pay the price.

In addition to their general-interest textual files, the major agencies have built up a series of other services during their 150 years of existence. The two American agencies, the AP and UPI, have dominated the supply of international news pictures around the world. The AP, UPI, and the French AFP are the basic domestic news-reporting organizations in their respective countries. Reuters, superior to the American agencies, has developed into a major supplier of specialized financial and commodity news to banks, foreign exchange dealers, companies, and official institutions, which has increasingly set it apart from its competitors.

Despite this wide range of activities, the supply of textual news to the international media remains the core of the agencies' international operations as a whole

The four agencies exist to serve markets, whether it is the global market for news of an event of international interest or the restricted market for coverage of a story of local or national significance. Their international operations are conducted on a commercial basis, although-paradoxically-they frequently operate in ways not calculated to maximize profits.

The two U. S. agencies, the AP and UPI, compete fiercely in domestic as well as in international news services. In France, AFP dominates the provision of domestic news to the media. Abroad, all four agencies compete energetically in Europe, Asia, and South Africa. In the Middle East, Reuters and AFP run Arabic-language services, while the two U.S. agencies confine themselves to English language distribution.

Apart from such geographical differences, the agencies also differ in the activities they undertake that are not directly concerned with supplying textual general news to the media. Reuters, is the leading supplier of news to the business world. Reuters earns the vast majority of its revenue outside Britain, while the other agencies draw most of their income from their home countries. News photographs have long been an integral part of AP and UPI operations, but neither European agency went into international pictures on a major scale until 1984. The two U.S. services provide comprehensive text and audio reports for radio stations; AFP does so only on an ad hoc basis; Reuters closed down its service of voice reports for radio stations at the end of the 1970s.

There are differences, too, in forms of ownership. The AP is a cooperative, while UPI is a private company. UPI changed hands in 1982 and again moved to modify ownership in 1985. Reuters is a limited liability British company that was owned mainly by the national and provincial press of the United Kingdom until June 1984. The big profits earned by its business news services led its owners to decide to float the firm on the London and New York stock exchanges, but with stiff restrictions to ensure that control remained with the original proprietors. AFP enjoys the sole status of a public French organization guaranteed by a parliamentary statute.

5.1.2 Origin of News Companies

The nature of the major news agencies was fixed at the start of their history. Their appearance has changed radically over 150 years, but their basic characteristics were established in the era of carrier pigeons and hand-delivered bulletins. An editorial philosophy that professed objectivity and neutrality was a precondition for survival; from the beginning, the agencies had to be one thing to all people, operating within the status quo and avoiding involvement in the events they reported [33].

In New York, the founders of the AP were spurred to action by a recurrent item on their newspapers' expenditure sheets: the costs of bringing foreign news down from Nova Scotia and the far northern U.S. ports, where it arrived packed in canisters aboard boats sailing from Europe. The founders were also concerned with telegraph tolls, which were high. Since the news each paper got was similar, the competitive advantages were minute. In 1848, thirteen years after the establishment in Paris of Agence Havas, the first European agency, the leaders of the New York press pooled their resources in a cooperative venture to save money by transmitting foreign news jointly from the northern ports.

Their Associated Press, which soon began to handle domestic news as well, belonged to the newspapers that used its services. Member papers exchanged news and jointly met the association's costs. Competing nonmembers were not permitted access to the service. Later generations of AP executives would hail this as the purest form of news agency, in which the press served itself cooperatively with no concern other than mutual professionalism.

In Europe, on the other hand, the founders of the major agencies came from outside the press. Charles-Louis Havas was a translator, Bernhard Wolff, a doctor, and Paul Julius Reuter, a book publisher, before they embarked on news-agency work [33]. The services they created were sold to newspapers, but their companies remained independent of the press. Subscribers enjoyed none of the control that AP members exercised over the running of their agency. Far from excluding possible subscribers, the Europeans' aim was to sell their news as widely as possible. Lacking the AP's cooperative financing, they were constantly in search of fresh revenues and displayed all the diversity typical of nineteenth-century venture capitalism. Charles Louis Havas had made and lost a fortune as a contractor to Napoleon's armies before his translation work led him to found the first modern news agency.

Despite the different forms that their agencies took, the founders in New York and Europe shared some basic assumptions. Without these, the agencies either would not have been launched at all or would have taken on a very different shape. Basic to the establishment of the nineteenth-century news services was the belief that a single satisfactory report could be provided to a large number of recipients whose politics, editorial values, and publication schedules differed widely. It is a belief that remains critical to the agencies today, and it lies at the root of much of the criticism of the major news services in the 1980s. The agencies do not issue one service for conservative newspapers and another for liberal newspapers. They distribute a single account of each event. This has been a constant of agency operations for 150 years. When Charles-Louis Havas set up his agency in 1835, he supplied the same items to reactionaries and reformists alike. Today, his successors at Agence France-Presse provide the same coverage for subscribers in Peking and Pretoria.

AFP

> AFP / History

HISTORY
PRESENTATION
NEWS
AFP WORLDWIDE
SUBSIDIARIES

A Long Tradition of Newsgathering

Founding father Charles-Louis Havas

AFP is the world's oldest established news agency, founded in 1835 by Charles-Louis Havas, the father of global journalism.

Today, the agency continues to expand its operations worldwide, reaching thousands of subscribers (radios, TVs, newspapers, companies) from its main headquarters in Paris and regional centers in Washington, Hong Kong, Nicosia and Montevideo. All share the same goal: to guarantee a top quality international service tailored for the specific needs of clients in each region.

Some Key Dates

1835 Charles-Louis Havas creates the first worldwide news agency.

1852 In addition to news, Havas launches into advertising, then in its infancy.

1940 The advertising and newsgathering operations of the Agence Havas are split up, leading to the creation of the Office Français d'Information (OFI, or French Information Office).

1944 After the Liberation of France from the Nazis, a group of Resistance fighters trade in their weapons for typewriters, and turn the former OFI into Agence France-Presse.

1957 The agency gets a new legal structure.

1985 AFP decentralises, with the creation of regional headquarters for both news production and distribution in Hong Kong, Nicosia and Washington.

Today's AFP employs some 1,200 reporters, 200 photographers and 2,000 stringers working in 165 countries.

Figure 5.1. AFP's History from its web site

5.2 Historical Perspective of AFP

Foreign news, appropriately enough, inspired Charles-Louis Havas to set up his agency. He had enjoyed a connection with the press in his early days of glory under Napoleon, owning shares in the major newspaper of the day, the Gazette de France, but this had been a strictly secondary interest compared with his moneymaking activities as a financier and supplier to the imperial army. After Napoleon's defeat at Waterloo ruined him, Havas was thrown back on using his extensive linguistic knowledge to translate extracts from foreign newspapers for the French press and private clients. By 1832, he had built up enough of a reputation to launch a regular news bulletin under the motto "Fast and Good" Three years later, having absorbed several competitors, he opened the Agence Havas.

The idea of distributing information bulletins was by no means new. In the middle ages, European merchant houses had sent their clients newsletters reporting the latest significant events. What was different about Havas was that he supplied his information to the press, to be passed on to the public as the newspapers saw fit. He maintained deliveries to some private clients, but his prime targets were the French newspapers, which were enjoying a period of expansion and international-mindedness in the early, liberal years of King Louis Philippe's rule. Then, as now, most newspapers lacked the resources and organization to collect foreign news on a regular basis. But they could not let their coverage of the world fall delinquent. With his daily service of foreign news, to which he added reports of official French pronouncements, Havas filled the gap.

REUTERS

Media Pack

Company Information

Reuters Board
Reuters History
Subsidiaries
Reuters Independence
Reuters Foundation
Greenhouse Fund
Word* Versions

Company Information: General

Word Version

Reuters vision: "to make the financial markets really work on the Internet". The Group is focused on three business areas – **Reuters Financial**, which consist of **Reuters Information** and **Reuters Trading Solutions**, **Reuterspace** and **Instinet**.

The strategy is to link Reuters Information to Reuters Trading Solutions in order to anticipate and meet the e-commerce needs of the financial markets. Reuterspace will rapidly penetrate into wider markets in business-to-business, consumer finance and media. Instinet will expand both globally and into the retail and fixed income markets.

Reuters Information supplies the global financial markets and the news media with the widest range of information and news products including real-time financial data, collective investment data from Lipper, numerical, textual, historical and graphical databases plus news, graphics, news video, and news pictures. It makes extensive use of internet technologies for the wider distribution of information and news.

Over 521,000 users in 52,800 locations access Reuters information and news worldwide. Data is provided on over 940,000 shares, bonds and other financial instruments as well as on 40,000 companies. Financial information is obtained from 260 exchanges and over-the-counter markets and contributed by 5,000 clients.

Reuters services are delivered globally over one of the world's most extensive private satellite and cable communications networks. Typically, Reuters updates 6,000 prices and items of other data per second and handles 65 million changes daily.

Reuterspace includes:

- Reuters Media serving both traditional and new media. Together with other parts of the Group, it provides news and information to over 900 internet websites reaching an estimated 40 million viewers and generating approximately 140 million pageviews per month, in addition to serving the traditional print and TV media. Some 330 subscribers plus their networks and affiliates in over 90 countries use Reuters television news coverage.

  Reuters is the world's largest international news and television agency with 2,100 journalists, photographers and camera operators in 184 bureaux serving 154 countries. News is gathered and edited for both business and media clients in 23 languages. Approximately 30,000 headlines, including third party contributions, and over three million words are published daily.

- Reuters Enterprise, serving the business-to-business e-commerce markets. It includes the newly acquired market research subsidiaries ORT and TowerGroup as well as the health and medical news services, Reuters Health Information and APM International. Reuters also has a 27% investment in the Datamonitor, a UK-based market research company.

- Reuters Personal providing services for the consumer retail finance markets. It plans to launch a joint venture with Multex.com to offer a financial portal for the European retail investor.

FINANCIAL

	Year ended 31 December				
	1999	**1998**	**1997**	**1996**	**1995**
Revenue (£m)	3,125	3,032	2,882	2,914	2,703
Profit before tax (£m)	632	580	626	652	558
Earnings per ordinary share (p)	30.2	26.7	24.0	27.3	23.2
Dividends per ordinary share (p)	14.65	14.4	13.0	11.75	9.8

Contact: Peter V Thomas/Adrian Duffield
Tel: +44 (0) 20 7250 1122
Media Relations

April 2000

Strategic Directions Webcast   Reuters on your Vodafone   Decimal Trading

REUTERS HQ

Figure 5.2. Reuters financial statement from its web site

Havas used a combination of mail coaches, couriers, carrier pigeons, and, after its introduction in France in the 1840s, the telegraph. His pre-cable arrangements enabled him to supply the Paris papers at noon with information from the morning's Belgian press. News from London followed at 3:00 P.M.

5.3 Historical Perspective of Reuters

Reuters in its early stages set up the carrier pigeon system between the German and Belgian cable heads, which became the legend. Currently, Reuters employs 16,119 people of 40 nationalities and has regional offices in London, New York, Geneva, Hong Kong and additional offices in another 217 cities. It supports customers in 163 countries, and creates news services in 25 languages.

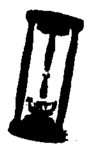
Reuter employees affect 277 markets in 163 countries, which rely particularly on their financial news 24-hours a day.

5.4 Historical Perspective of AP

In 1848, six of New York newspapers decided to share both the expenses and the news. The cooperative newsgathering was born and their innovation was called The Associated Press.

AP Associated P

AP HISTORY   Complete History



Origins

On an early morning in May 1848, 10 men representing six New York City newspapers sat around an office table of the New York Sun. They had been in session for more than an hour and all that time they had been in stubborn argument.

At issue was the costly collection of news by telegraphy. The newly invented telegraph made transmission of news possible by wire but at costs so high that the resources of any single paper would be strained.

David Hale of the Journal of Commerce argued that only a joint effort between New York's papers could make telegraphy affordable and effectively prevent telegraph companies from interfering in the newsgathering process To get news from the west and from abroad, Hale argued, newspapers had to work together if the public was to be served with increasingly wider coverage of the United States and the world.

Although reluctant at first, the six highly competitive papers agreed to the historic plan, and The Associated Press was born.

Today, that six-newspaper cooperative is an organization serving more than 1,500 newspapers and 5,000 broadcast outlets in the United States. Abroad, AP services are printed and broadcast in 112 countries.

Worldwide, the AP serves more than 15,000 news organizations.

Figure 5.3. AP's History from its web site

The early AP was much different than the organization that exists today, but the mission of that pioneering effort has not greatly changed.

One of the AP papers' foremost interests was foreign news, and in 1849 Daniel Craig became the AP's first foreign correspondent. His primary duty was to meet incoming steamer ships at Halifax, Nova Scotia, and forward European news to Boston and New York. His dispatches moved first via carrier pigeon and pony express and later by wire as telegraph lines extended north.

The founders of the AP were motivated by economic, rather than journalistic, principles. Within months of its creation, the AP began forwarding its news to papers in Philadelphia and Baltimore. These papers were not members of the AP – membership was restricted to New York papers – but they were its first paying clients. Others followed, and gradually the subscriber papers began to form loosely defined geographic groups. They referred to themselves by such names as the Southern Associated Press, the Western Associated Press and the New York State Associated Press.

To distinguish the parent organization from those regional groups, the newspaper world began to speak of it as the New York Associated Press.

General Agent Craig arranged for the purchase of news from correspondents in most major cities, but in 1851 the AP set up its first permanent bureaus or "agencies" in Washington and Albany. Gobright headed the Washington

Bureau, and it was he who wrote the succinct bulletin telling the world of the shooting of Abraham Lincoln.

5.5 Historical Perspective of UPI

Unite Press International (UPI), formerly known as United Press (UP) was founded in the United States in 1907 by E. W. Scripps. In 1935, United Press became the first major American news service to supply news to radio stations. In 1958, United Press merged with William Randolph Hearst's International News Service to become UPI.

NEWS SERVICES
- Global Impact Net
- WebLine
- K@W
- WebNewsletters
- Arabia 2000
- UPI Newspictures

SPECIAL REPORTS
- Egypt Air Flight 990
- The Balkan Files
- JFK Jr. Tragedy
- Helen Thomas

COMPANY INFORMATION
- Corporate News
- Resource Links
- Contacts

UPI's history and heritage

"I regard my life's greatest service ... to be the creation of the UP ... I have made it impossible to suppress the truth or successfully disseminate falsehood. The mere fact that the UP can be depended upon to disseminate news ... makes it not only worthwhile to put out such information, but positively dangerous to withhold it."

- E.W. Scripps, UPI

A news organization with a proud and colorful history going back 91 years, United Press International is focused on being a leader in the Knowledge Age. The company provides news and information for broadcast, print, online and a variety of other subscribers.

E.W. Scripps founded the United Press Associations in 1907 to cover news from around the world. In 1935, United Press became the first major American news service to supply news to radio stations. In 1958, United Press merged with William Randolph Hearst's International News Service to become UPI. That also was the year the company began the first wire service radio network, providing radio stations with voice reports from correspondents around the globe. In 1982, the news service put into operation the industry's first method to let subscribers choose to receive copy by topic and subtopic, rather than only by broad category of news. By 1995, the company completed a system for global satellite transmission that virtually eliminated the need to send news over telephone land lines.

The world's largest privately owned news service, UPI remains a leader in the coverage and delivery of news text, audio reports and photographs -- and is making its presence felt on the Internet and the World Wide Web. UPI is the interface of journalism's future, committed to unbiased, free and fair reporting of the world's events.

UNITED PRESS INTERNATIONAL

Figure 5.4. UPI's History from its web site

NEWS SERVICES
- Global Impact Net
- WebLine
- K@W
- WebNewsletters
- Arabia 2000
- UPI Newspictures

COMPANY INFORMATION
- Corporate News
- Resource Links
- Contacts

Arabia 2000

We are your window to the Middle East.

Arabia 2000 is a unique concept, grouping together all available news sources in Arabic on the existing UPI WorldView software platform. Arabia 2000 will be delivered by satellite and through the Internet to clients who have a need to keep abreast of developments in the region and in the Arabic language. The primary market for us is in the Middle East and Europe. This is a revolutionary concept and presents UPI to the market as a "one-stop-shop" for all available news information in Arabic. It will save clients wanting to take a range of separately-available news sources immense amounts of time, energy and money.

DELIVERY
Arabia 2000 will be delivered on the Arabic platform of UPI WorldView (where all the text and instructions are in Arabic). Each information provider will be segregated and a button allocated to each. On opening one service, a splash screen welcoming the user will appear. (If the client has not paid for this particular service, instructions about ordering the service will appear). Arabia 2000 will be delivered initially by VBI on the Arabsat and Eutelsat satellite footprints (The Middle East and Europe). It will be displayed on PCs and can easily be taken by anyone with a PC who also receives the MBC TV signal. A computer card will substitute for the current VBI decoder in use.

UNITED PRESS INTERNATIONAL

Figure 5.5. UPI's international services

5.6 Associated Press Applications

Associated Press is organized around a network of news offices located in all 50 of the United States and in over 70 foreign countries.

AP's national and international headquarters is in New York City. The Associated Press Building at 50 Rockefeller Plaza houses news and administrative offices.

5.6.1 News Broadcast Services

The AP's national broadcast report, radio network and video news services are prepared and broadcast from the Broadcast News Center in Washington D.C. More than 100 news managers, newsmen and newswomen report, write, produce and anchor the AP's news reports for more than 6000 radio stations, television stations and cable systems.

5.6.2 Databases, Client Server Services

AP's MIS department provides a variety of client-server databases to the sales staff of different departments of the company for sales and support of a variety of services. The company and its subsidiaries use a large number of databases for billing different services.

5.6.3 WWW Servers

A variety of World Wide Web servers provide information about different AP services worldwide to the entire Internet and to Intranet users.

5.6.4 Multimedia Services

This department fulfills audio and video needs of the members and non-members. Variety of SUN Microcomputer systems and Apple computers are used to develop and deliver the services.

5.6.5 News Photo Services

This department is located at the headquarters and is the hub of a network of photographers and photo editors who make and transmit photos around the world.

5.6.6 Graphics Services

Artists in the graphics department use both pen and computer to illustrate the AP's news reports. Artists generate both traditional drawings and computer generated charts, graphs and maps for breaking news stories and longer-range enterprise work.

5.6.7 News Features

AP news features department produces a series of in-depth feature stories for Sunday newspapers. The executive editor coordinates the production of a weekly package of stories written by reporters based at AP network and those contributed by news staffers in bureaus around the world.

5.6.8 Financial Applications

The Treasury is responsible for managing AP's fiscal affairs. The treasury is divided by functions into sections which audit and record financial transactions, produce financial reports, budgets, calculate payrolls, assess members and subscribers, collect funds and purchase supplies.

5.6.9 Press Association Services

Press Association is an AP subsidiary that markets both existing and specially created AP services to non-members in the United States. PA customers include businesses, embassies and political campaigns interested in reading the AP report on an "information only" basis.

PA also sells a variety of AP services to computer-based information retrieval services and archival database companies for use by researchers.

5.6.10 Sales

A team of full and part-time sales staff sells AP services to members and non-members all over the world. This sales force uses a variety of contact and billing databases to offer their services and follow-ups.

5.6.11 News & Photos Library

The news library's card catalog alone includes more than 10 million index cards. Microfilm of AP copy begins with the 1937 report. The library has a wealth of general and specialized reference works, plus access to hundreds of newspapers, magazines and specialty publications. A librarian can also access NEXIS, a database that includes every data stream story since 1977.

The Photos Library has about 50 million negatives on file for use by members and for sale through Wide World Photos, a commercial subsidiary of the AP.

5.6.12 World Services

The world services handles sales and marketing of AP services in more than 115 countries. Those who receive AP's reports include other news agencies and virtually every major newspaper, television station and radio network in the world, as well as non-media clients such as business and government installations, including embassies and military bases.

Chapter 6

SECURITY AND NEWS NETWORKS

The great Internet explosion during the past two years is largely fueled by the prospect of performing business online. Soon nearly any transaction can be handled through the Internet. The Internet can bring down physical barriers to Commerce, immediately giving even the smallest business access to untapped markets around the world. By the same token, consumers can conduct business and make purchases from organizations previously unavailable to them.

Keeping in mind these goals, individuals and small and/or large organizations have flocked to the Internet, and most businesses have set out to set up storefronts on the Internet and the World Wide Web. Just about every major business in the U.S., perhaps even in the world, has a home page on the Internet on which users can find information about their services and products. Despite the forecasts, however, consumers and businesses alike seem wary of this new medium for conducting business on a large scale. Given the potential for both consumers and businesses, why the apprehension?

6.1 Insecurity

The original Internet was designed for research, not as a commercial environment. As such, it operated in a single domain of trust, while provisions were made to allow remote users to access critical files on machines through the Unix Operating System's r-commands (e.g., rlogin and rsh). Security generally relied on users' mutual respect and honor, as well as their knowledge of conduct considered appropriate on the network. Minor security was made available in the form of password-protected hosts but was basically an after-thought in design.

As the Internet grew, the community expanded, and the existing security framework was insufficient. Over the past few years, we have seen evidence of this fact, in the form of Internet-based attacks on commercial systems:

- The Morris Worm of 1988

- The "Berferd" incident at AT&T in 1991
- The theft of passwords from service providers in late 1993 and early 1994
- The "IP Spoofing" attack on the San Diego Supercomputer Center in late 1994
- The theft of funds from Citibank in 1995
- Several virus attacks during 1999 and 2000

For the most part, such attacks take advantage of simple holes largely attributed to mis-configured systems, poorly written software, mismanaged systems, or user neglect. Awareness of these attacks has spawned the use of network security tools, such as firewalls, to protect individual networks from attack. More sophisticated attacks can take advantage of basic flaws in the Internet's infrastructure. For example, the TCP/IP protocol suite used by all computers connected to the Internet is fundamentally lacking in security services. The most glaring omissions occur at the lower layers of the protocol stack-within TCP, IP, and such transmission protocols as Ethernet - since they affect all applications that rely on the transport mechanism (see the sidebar, "Basic Flaws in Internet Infrastructure"). In the commercial world, these problems manifest themselves in a number of ways.

6.1.1 Eavesdropping

Eavesdropping attacks on a network can result in the theft of account information, such as credit card numbers, customer account numbers, or account balances and billing information. Similarly, such attacks can result in the theft of services normally limited to subscribers, such as information-based products. In some cases, even the fact that a transaction occurs can be used against one or, both of the parties involved. For example, if a merchant conducted business with a individual and then created and used a profile of the individual, this could border on an invasion of the individual's privacy. Or if a company is doing business with another company it could alert competitors to potential business partnerships or negotiations.

6.1.2 Password Sniffing

Password-sniffing attacks can be used to gain access to systems on which proprietary information is stored. As the use of stronger cryptographic algorithms becomes more and more commonplace, we would likely see a shift in focus away from attempts to break the protocol toward retrieving clear text information from poorly protected systems.

6.1.3 Data modification

Data modification attacks can he used to modify the contents of certain transactions (e.g., changing the payee on an electronic check or changing the amount being transferred to a bank account). Such attacks can also be used to modify certain orders over the network.

6.1.4 Spoofing

Spoofing attacks can be used to enable one party to masquerade as another party. In one such situation, a criminal can set up a storefront and collect thousands or even millions of credit card numbers, account numbers, or other information from unsuspecting consumers. This is similar to fly-by-night insurance companies or fraudulent financial institutions. But, since a criminal uses a virtual storefront, they can be beyond the reach of any regulatory body. In a more serious case, an individual could pose as a financial clearinghouse or acquirer (e.g., Visa or MasterCard), and collect payment from unsuspecting consumers or fees from merchants.

6.1.5 Repudiation

Repudiation of transactions can cause major problems with billing systems and transaction processing agreements. For example, if one party reneges on an agreement after the fact, the other party may incur the cost of transaction processing without benefiting from the transaction. Consider, for example, the real-world case of a bounced check. But rather than being rejected for insufficient funds, the check bounces due to the bank's inability to verify its authenticity.

6.2. Security Requirements for Commercial Transactions

While firewalls serve a valuable purpose in securing Internet-connected networks, they do not provide end-to-end transaction security and cannot be considered adequate security solutions for commercial Internet transactions. Other solutions, such as one-time passwords, solve part of the problem (e.g., password-sniffing attacks) but do not address the whole problem. A robust security solution for transaction processing satisfies the following fundamental security requirements:

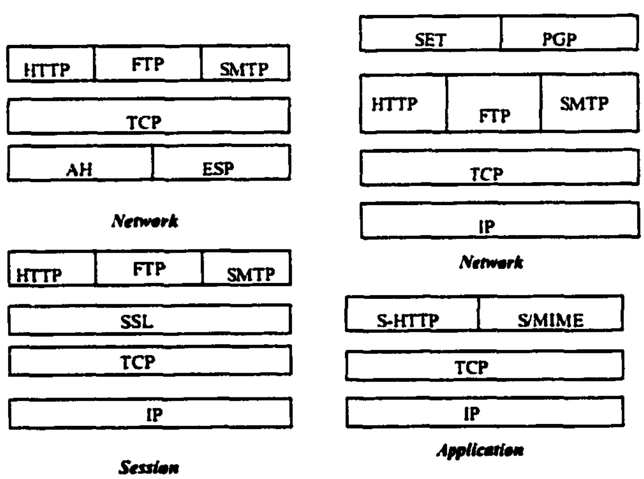


Figure 6.1 Relationship between various security initiatives and the protocol stack

6.2.1 Confidentiality

All communications between parties are restricted to the parties involved in the transaction. This confidentiality is an essential component in user privacy, as well as in the protection of proprietary information and as a deterrent to theft of information services.

6.2.2 Authentication

Both parties should be able to feel comfortable that they are communicating with the party with whom they think they are doing business. Authentication is usually provided through digital signatures and certificates.

6.2.3 Data integrity

Data sent as part of a transaction should not be modifiable in transit. Similarly, it should not be possible to modify data while in storage.

6.2.4 Non-repudiation

Neither party should be able to deny having participated in a transaction after the fact.

6.2.5 Selective application of services

It may be desirable for part of a transaction to be hidden from view while the remainder of the same transaction is not.

Confidentiality is usually provided through encryption. Authentication, data integrity, and non-repudiation are usually provided through digital signatures and public key certificates (see the sidebar, "Cryptographic Concepts"). These first four requirements are fairly standard for business communications. The fifth, however, raises an important issue. Consider the following scenario: A customer wants to buy something from a given merchant. Under the standard model, the customer hands his or her credit card to the merchant, who then sends the number to the financial institution to be processed. Assuming appropriate authorization, the merchant proceeds with the sale. However, this model unnecessarily gives the merchant access to the customer's credit card number.

Now consider the following alternative: A customer wants to buy something from a given merchant. The customer packages his or her credit card information and seals it in a digital envelope, so only the bank can see it. The envelope is sent to the merchant along with the purchase order. The merchant passes the envelope on to the bank, which provides the authorization for the purchase. Upon receipt of the authorization, the merchant concludes the sale.

6.3 Security Initiatives

In order to provide these services, a number of cryptographic protocols have been proposed. While these protocols are similar in the services they provide and in the cryptographic algorithms they use, they vary in the manner in which they provide the services and in their locations with respect to the rest of the TCP/IP protocol stack. Some initiatives endeavor to implement security at the network IP layer; others, just above TCP, at the session layer; still others, such as File Transfer Protocol (FTP), Hyper Text Transfer Protocol (HTTP), and telnet, within specific application protocols, and a whole class of solutions for securing the content of documents (e.g., application-independent payment protocols) residing above the existing application protocols.

The issue of where within the protocol stack to provide security is contentious. Proponents of placing security lower in the stack argue that lower-layer security solutions can be implemented transparently to end users and application developers, effectively killing many birds with a single stone. Proponents of higher-layer solutions argue that lower-layer solutions attempt to do too many things and that application-layer solutions allow application-specific security services (e.g., protecting selected fields within a protocol, or individual methods, as in HTTP).

6.3.1 Network-Layer Security Solutions

Since 1993, efforts have been under way to develop IP security architecture to provide cryptographic protection for Internet traffic. Although the efforts were, at least in part, initiated to help define a next generation protocol, such as IPv6, the concepts have been incorporated into implementations of the existing Internet protocol. This architecture includes two mechanisms for providing security services:

- Authentication Header (AH), which provides authenticity and integrity rising the MD5 message-digest algorithm.
- Encapsulating Security Payload (ESP), which provides confidentiality rising the Data Encryption Standard (DES) algorithm.

6.3.2 Session-Layer Security Solutions

By far the most prevalent session-layer protocol is the Secure Sockets Layer (SSL), first introduced by Netscape in late 1994. SSL provides security services just above the TCP layer, using a combination of public-key and symmetric cryptosystems to provide confidentiality, data integrity, and authentication of the server and (optionally) the client. Although it may seem that not providing client authentication goes against the principles that should be espoused by a robust system, an argument can be made that the decision to optionally support it helped SSL gain widespread support and use. Support for client authentication requires individual asymmetric key pairs for each client, and since support for SSL is embedded in Netscape Navigator (the most popular Internet browser application software), requiring client authentication would involve distributing public-key certificates to every Netscape user on the Internet. In the short term, it was believed to be more crucial that consumers be aware of with whom they are conducting business than to give the merchants the same assurances. Furthermore, since the number of servers on the Internet is much smaller than the number of clients on the Internet, it is much more practical to outfit servers with the necessary capabilities to handle digital signatures and key management. However, support for client-specific key pairs is growing. This support is visible in recent releases of Netscape and through the recent growth of client authentication in SSL implementations. Other Web browser and server vendors also currently support SSL. An Internet Draft describing SSL was released in late 1995.

Also in late 1995, Microsoft introduced a protocol similar to SSL. This protocol called Private Communication Technology (PCT) attempted to solve some flaws within SSL but is otherwise similar in nature to, and completely interoperable with, SSL.

While SSL burst onto the scene without first proceeding through the standards process, a number of session-layer protocols have been proposed within the Internet Engineering Task Force (IETF) to support distribution of keys for use with almost any TCP/IP application. Fundamental to each of these protocols is establishment of "security associations" between two users or systems. Although there have been many efforts to standardize this idea, three protocols remain under consideration:

Simple Key Exchange for Internet Protocols (SKIP)

SKIP makes use of public-key certificates to exchange long-term symmetric keys between two communicating parties. These certificates are obtained through the use of a separate protocol that runs over the User Data Protocol (UDP).

Photuris

The chief complaint about SKIP is its lack of "perfect-forward secrecy"; that is, if someone can obtain a long-term SKIP key, he or she can decrypt all messages previously generated using that key. Photuris is one alternative that does not have this problem; it uses long-term keys only to authenticate session keys, thereby offering perfect forward secrecy. However, to provide perfect-forward secrecy, Photuris necessarily sacrifices a certain amount of SKIP's efficiency.

Internet Security Association and Key Management Protocol (ISAIKMP) Unlike SKIP and Photuris, ISAKMP provides a generic framework for key management, rather than a specific key management protocol. By not binding itself to specific cryptographic algorithms or protocols, ISAKMP offers more flexibility with regard to use and policy issues than Photuris or SKIP.

It is too early to determine what the future holds for session-layer security. One group of people claims that SSL is already widely deployed and that its embedded base will cause it to supersede any other session-layer protocol. Another school of thought states that SSL, provides a convenient short-term solution, since it does not involve the modification of individual application protocols, and that it will just "drop out" once the true Internet security architecture is implemented. More than likely, we will see SSL and the Internet security architecture coexist, with each used for different purposes.

6.3.3 Application-Layer Security Solutions

Paralleling the session-layer protocols, a number of efforts have been aimed at securing individual applications. From the point of view of electronic commerce, two sets of applications are of concern: email and the World-Wide Web.

The IETF Privacy-Enhance Mail (PEM) Working Group defined a standard for securing email in February 1993. However, due to some limitations in scope, this proposal never really took off. A second proposal, called MIME Object Security Services (MOSS), which incorporates support for non-textual messages using MIME (Multipurpose Internet Mail Extensions), has since gained support. However, the prevalence and anticipated growth of Pretty Good Privacy (PGP), a widely available encryption program on the Internet, due to the resolution of patent infringement claims and export issues, as well as the existence of commercial versions, makes the future of MOSS unclear. Further clouding the issue, a third standard, S/MIME, was introduced by RSA Data Security, Inc. in July 1995, with announced support from many major email vendors. Help is on the horizon. The Internet Mail Consortium convened in February 1996 to begin addressing the resolution of differences between the various secure mail protocols.

With regard to the World-Wide Web, there are two major competing standards for securing HTTP transactions. The first, HTTPS, is merely an implementation of HTTP over SSL, as described earlier. The second, appropriately called Secure HTTP (S-HTTP), was first proposed in 1994 by

CommerceNet, a consortium of organizations interested in electronic commerce, and shortly thereafter was taken under consideration by the IETF Web Transaction Security Working Group. Like SSL, S-HTTP provides confidentiality and integrity of data and authenticates the server and (optionally) the client. However, S-HTTP defines an extension to the existing HTTP protocol, thereby embedding security within the existing HTTP framework. S-HTTP is also more flexible than SSL in terms of the algorithms and security mechanisms it supports. While this flexibility bodes well for interoperability, critics claim it makes it too difficult to develop a working implementation of S-HTTP; they point to the lack of widely available reference implementations of S-HTTP.

Initially, there seemed to be a divergence of support for S-HTTP and SSL, with most vendors aligning themselves with one standard or the other. More recently, we have begun to see some unification of support for both protocols. For example, to facilitate the use of security services for Web transactions, most World Wide Web server manufacturers have announced support for both S-HTTP and SSL.

Terisa Systems, the organization behind the original S-HTTP proposal, also released a Web Sever Toolkit, allowing developers to incorporate both SSL and S-HTTP into applications.

6.4 Internet Payment Protocols

All of the solutions discussed up to this point attempt to incorporate security into the TCP/IP stack, with the intent of fostering widespread use. However, a whole class of protocols is currently being investigated to provide secure payment systems on the Internet. For the most part, these protocols focus on methods for securing transmission of private information (e.g., credit card numbers and account information) among four parties:

- A purchaser (or consumer) buying something
- An issuer who issued a credit-card account to a purchaser
- A merchant making a sale to a purchaser
- An acquirer who maintains a financial relationship with a merchant

These protocols are specific to the secure transmission of private information and are independent of the underlying transport media, that is, they can be implemented within Web browsers using HTTP, or email programs using Simple Mail Transport Protocol (SMTP), or other applications

Using some other application protocol. The important thing to remember is that the data involved in the transaction is secured, regardless of the medium. In the event that the unsecured channel is attacked, the attacker gains nothing more than an indecipherable block of data.

Although many payment mechanisms were proposed, two have emerged as clear leaders in this area, largely due to their corporate backing:

- Secure Electronic Payment Protocol (SEPP), supported by MasterCard, IBM, and Netscape.
- Secure Transaction Technology (STT), backed by Visa and Microsoft

In early 1996, Visa and MasterCard agreed to consolidate their standards into a single payment system called Secure Electronic Transactions (SET).

One area in which there is bound to be a great deal of additional research in the near future is that of "micropayments," or transactions in which the value of the data being purchased is less than the cost of the transaction. Though efforts are under way to define low-cost transaction systems, such as Millicent PayWord, and MicroMint micropayments are still a research-grade problem, and implementations are not likely to be seen this year.

6.5 The Future

By almost all accounts, security is the major concern to widespread commerce on the Internet. A number of proposals have been made to secure

Internet communications, although it is difficult to predict what the future holds for secure Internet commerce. In addition to the security requirements outlined here, the usability of systems must also be considered. In order for a public-key based system to be widely adopted, it is necessary to make the entire key retrieval and distribution process transparent, as well as to provide transparent support for a certificate infrastructure. Although no such infrastructure currently exists, the efforts of the IETF Public Key Infrastructure Working Group (PKIX) are encouraging.

At the current pace, however, it is unlikely that the market will be able to wait for a solution to proceed through the full standards process before adopting it. We are more likely to see the immediate use of such protocols as SSL and S-HTTP, as well as such payment mechanisms as SET. Though compatible with the existing Internet protocols, the IP security architecture may not be adopted until the rollout of IPv6. Even though, an IPv6 developer admits, the full-scale deployment of IPv6 is still a number of years away. It will be interesting to see how the introduction of this architecture affects the use of protocols already deployed.

Chapter 7

SIGNALING SYSTEM NUMBER 7

Signaling System Number 7 (SS7) is a well-suited protocol to support the requirements of integrated digital networks. The standard was first approved by ITU-T in 1980, with revisions in 1984, 1988, 1992 and 1997. Since 1997 there has been also several revisions. SS7 is designed to be an open-ended common channel signaling standard that can be used over a variety of digital circuit-switched networks.

SS7 provides the common channel signaling system with the following primary characteristics:

- Optimized for use in digital telecommunication networks in conjunction with digital stored program-control exchanges, utilizing 64-kbps digital channels

- Designed to meet present and future information transfer requirements for call control, remote control, management and maintenance.
- Provides a reliable means for the transfer of information in the correct sequence without loss or duplication.
- Suitable for operation over analog channels and at speeds below 64 kbps.
- Suitable for use on point-to-point terrestrial and satellite links.

The scope of SS7 is immense, since it must cover all aspects of control signaling for complex digital networks, including the reliable routing and delivery of control messages and the application-oriented content of those messages.

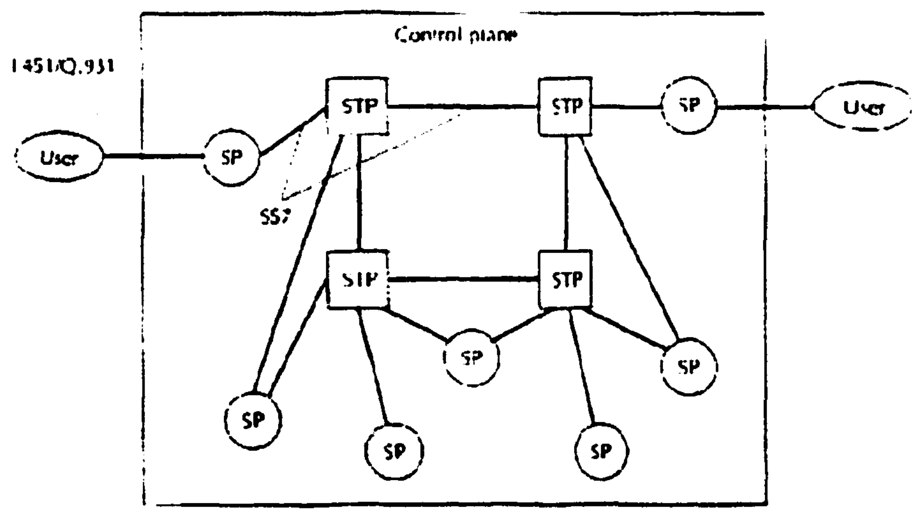
7.1 SS7 Architecture

Signaling System Number 7 (SS7) architecture can be by two separate criteria; the functional architecture and the protocol architecture.

7.1.1 Functional Architecture

With common-channel signaling, control messages are routed through the network to perform call management (setup, maintenance, termination) and network management functions. These messages are short blocks or packets that must be routed through the network. Thus, although the network being controlled is a circuit-switched network, the control signaling is implemented using packet-switching technology. In effect, a packet-switched network is overlaid on a circuit-switched network in order to operate and control the circuit-switched network.

SS7 defines the functions that are performed in the packet-switched network but does not dictate any particular hardware implementation. For example, all of the SS7 functions could be implemented in the circuit-switching nodes as additional functions; this approach is the associated signaling mode. Alternatively, separate switching points can also be used with some restrictions. Even in this case, the circuit-switching nodes would need to implement portions of SS7 so that they could receive Control signals.



STP = Signaling Transfer Point
SP = Signaling Point
Figure 7.1. Signaling and Information Transfer Network

Signaling Network Elements

SS7 defines three functional entities: signaling points, signal transfer points, and signaling links. A Signaling Point (SP) is any point in the signaling network capable of handling SS7 control messages. It may be an endpoint for control messages and incapable of processing messages not directly addressed to it. The circuit-switching nodes of the network, for example, could be endpoints. Another example is a network control center. A Signal Transfer Point (STP) is a signaling point that is capable of routing control messages; that is, a message received on one signaling link is transferred to another link. An STP could be a pure routing node or could also include the functions of an endpoint. Finally, a signaling link is a data link that connects signaling points. We can consider that there are two planes of operation. The control plane is responsible for establishing and managing connections. The user requests these connections over the D channel using Q.931. The Q.931 dialogue is between the user and the local exchange. For this purpose, the local exchange acts as a signaling point, since it must convert between the dialogue with the user (Q.931) and the control messages inside the network that actually perform user-requested actions (SS7). Internal to the network, SS7 is used to establish and maintain a connection; this process may involve one or more signaling points and signal transfer points. Once a connection is set up, information is transferred from one user to another, end-to-end, in the information plane. A circuit is set up from the local exchange of one user to that of another, perhaps being routed through one or more other circuit-switching nodes, referred to as transit centers. All of these nodes (local exchanges, transit centers) are also signaling points, since they must be able to send and receive SS7 messages in order to establish and manage the connection.

Signaling Network Structures

A complex network will typically have both Signaling Points (SP) and Signal Transfer Points (STP). A signaling network that includes both SP and STP nodes could be considered as having a hierarchical structure in which the SPs constitute the lower level and the STPs represent the higher level. The latter may further be divided into several STP levels. Figure 7.1 is an example of a network with a single STP level.

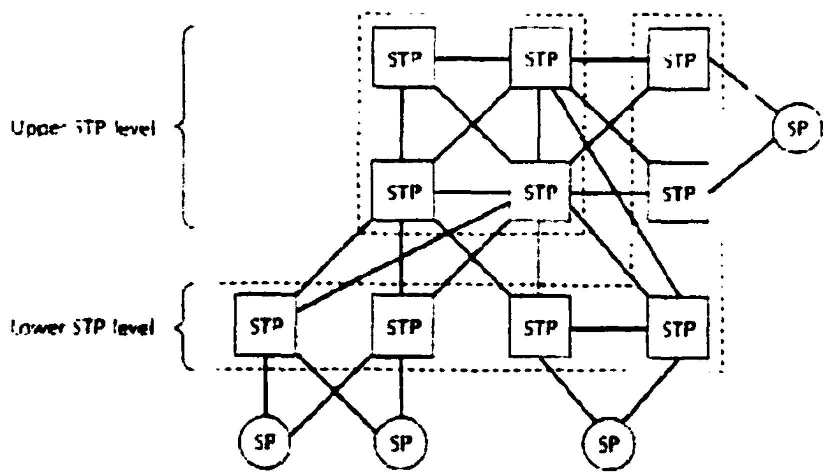


Figure 7.2. Hierarchical Signaling Network with Two STP Levels

Following parameters could influence the decisions concerning design of the network and the number of levels to be implemented:

- STP capacities: Includes the number of signaling links that can be handled by the STP, the signaling message transfer time, and the message throughput capacity.
- Network performance: Includes the number of SPs and the signaling delays.
- Availability and reliability: Measures the ability of the network to provide service in the face of STP failures.

When considering the network constraints in terms of performance, one STP level seems preferable. However, considerations of reliability and availability may dictate a solution with more than one level. The following guidelines are suggested by ITU-T:

In a hierarchical signaling network with a single STP level:

- Each SP that is not an STP at the same time it is connected to at least two STPS.
- The meshing of STPs is as complete as possible (full mesh: every STP has a direct link to every other STP).

In a hierarchical signaling network with two STP levels.

- Each SP that is not an STP at the same time is connected to at least two STPs of the lower level.
- Each STP in the lower level is connected to at least two STPs of the upper level.
- The STPs in the upper level are fully meshed.

The two-level STP hierarchical design would be typically configured such that the lower level is dedicated to traffic in a particular geographic region of the network, and the higher level handles inter-region traffic. SPs and STPs are connected by links that are defined by function in Table 7.1. STPs are configured in pairs for redundancy and linked by cross (C) links. Circuit-switching nodes hook into the SS7 packet-switching network by means of access (A) links to paired STPS. Bridge (B) links are provided between STP pairs in different regions and D links between STP pairs at different hierarchical levels. The remaining link types (E and F) provide additional paths to and from circuit-switching nodes to reflect particular high traffic demands.

It can be seen that this design combines good performance with high availability. Between any pair of signaling points, messages must ordinarily traverse only one or two STPs. This provides low message transit delay.

At the same time, the loss of a critical STP or signaling link does not prevent communication, although a somewhat longer route may need to be followed.

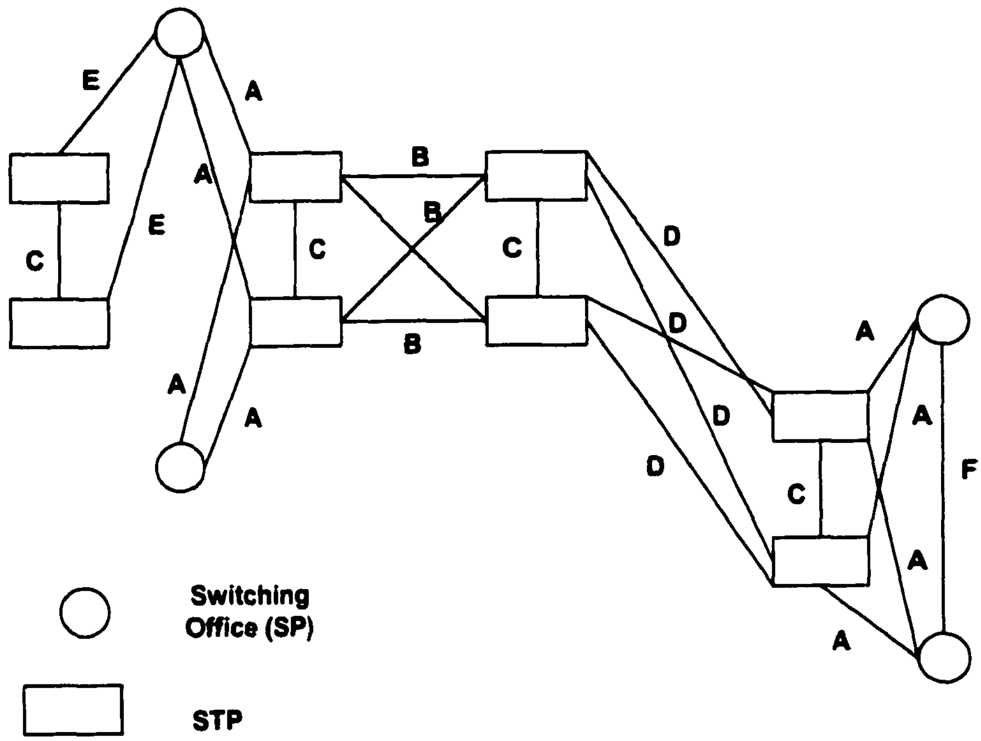


Figure 7.3. Links Used in an SS7 Network

7.1.2 Protocol Architecture

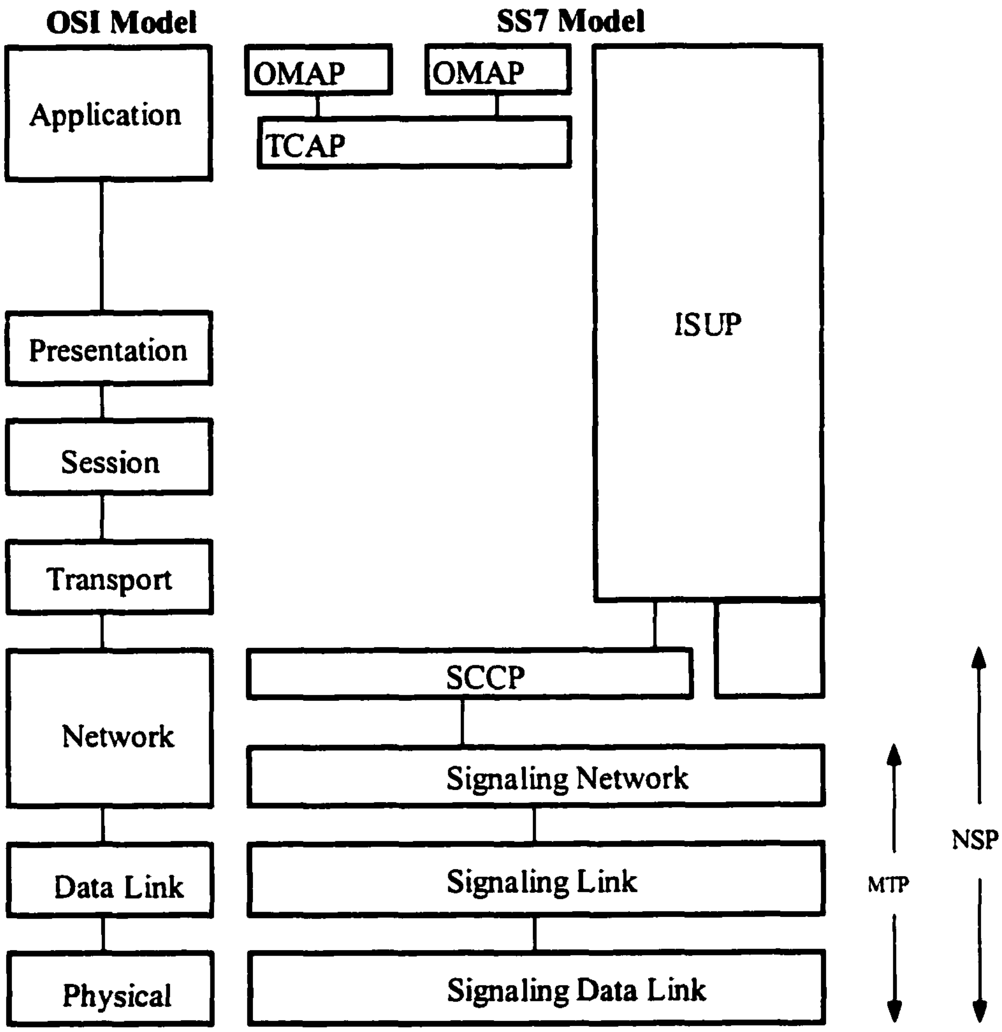
So far, we have been discussing SS7 architecture in terms of the way in which functions are organized to create a packet-switching control network. The term architecture can also be used to refer to the structure of protocols that specify SS7. As with the Open Systems Interconnection (OSI) model, the SS7 standard is a layered architecture. Figure 7.4 shows the current structure of SS7 and relates it to OSI.

Designation	Connection	Use
A	SP to STP	Provides access to signaling network from a switching office
B	STP to STP at same level of a hierarchy	Primary routing of messages from one SP to another via multiple STPs.
C	STP to mated STP	Communication between paired STPs, also provides alternate route around failed B links.
D	STP to STP at different levels of a hierarchy	Routing of messages up or down in a hierarchy.
E	SP to STP	Provides direct connection to non home STP from a switching office.
F	SP to SP	Provides direct access between switching offices with a high community of interest

Table 7.1 Signaling Links

The SS7 architecture consists of four levels. The lowest three levels of the SS7 architecture, referred to as the Message Transfer Part (MTP), provide a reliable but connectionless (datagram style) service for routing messages through the SS7 network. The lowest level, signaling data link, corresponds to the physical layer of the OSI model and is concerned with the physical and electrical characteristics of the signaling links. These include links between STPs, between an STP and an SP, and control links between SPs. The signaling link level is a data link control protocol that provides the reliable sequenced delivery of data across a signaling data link; it corresponds to layer 2 of the OSI model. The top level of the MTP, referred to as the signaling network level or function, provides for routing data across multiple STPs from control source to control destination. These three levels together do not provide the complete set of functions and services specified in the OSI layers 1-3, most notably in the areas of addressing and connection-oriented service. In the 1984 version of SS7, an additional module was added, which resides in level 4, known as the Signaling Connection Control Part (SCCP). The SCCP and MTP together are referred to as the Network Service Part (NSP). A variety of different network-layer services are defined in SCCP, to meet the needs of various users. The remainder of the modules of SS7 is considered to be at level 4 and comprise the various users of NSP. NSP is simply a message delivery system; the remaining parts deal with the actual contents of the messages. The ISDN User Part (ISUP) provides the control signaling needed in an ISDN to deal with ISDN subscriber calls and related functions. The Transaction Capabilities Application Part (TCAP), first introduced in 1988, provides the mechanisms for transaction-oriented (as opposed to connection-oriented) applications and functions. The Operations, Maintenance, and Administration Part (O&MAP) specify network management functions and messages related to operations and maintenance.

In addition, other modules, referred to as Application Service Elements (ASE), may be defined to support other applications. The MTP was developed prior to SCCP and was tailored to the real-time needs of telephony applications.



OMAP = Operations, Maintenance, and Administration
ASE = Application Service Element
TCAP = Transaction Capabilities Application Part
ISUP = ISDN user part
SSCP = Signaling Connection Control Part
MTP = Message Transfer Part
NSP = Network Service Part

Figure 7.4. SS7 Protocol Architecture

The connectionless nature of MTP provides a low-overhead facility tailored to the requirements of telephony. In the context of ISDN, it became clear that there were other applications, such as network management, that needed the full services of the OSI network layer, such as expanded addressing capability and reliable message transfer. SCCP was designed to meet these requirements. The resulting split in OSI network functions between the signaling network layer and SCCP has the advantage that the higher-overhead SCCP services can be used only when required, with the more efficient MTP used for other applications.

7.2 Signaling Data Link Level

The signaling data link is a full-duplex physical link dedicated to SS7 traffic. SS7 is optimized for use over 64-kbps digital links. However, the recommendations allow for the use of circuit-switched connections to the data link, lower speeds, and the use of analog links with modems. The link can be routed via a satellite.

7.3 Signaling Link Level

The signaling-link level corresponds to the data link control layer of the OSI model. Thus, its purpose is to turn a potentially unreliable physical link into a reliable data link. Reliability implies

- All transmitted blocks of data are delivered with no losses or duplications.

- Blocks of data are delivered in the same order in which they were transmitted.
- The receiver is capable of exercising flow control over the sender.

The last point assures that blocks of data are not lost after delivery because of buffer overflow.

Many of the techniques found in better-known data link control protocols, such as Link Access Protocol for D Channel (LAPD) and Link Access Procedure-Balanced (LAPB), are used in the SS7 signaling link level. However, the formats and some of the procedures are different. The differences in some cases are matters of style rather than substance. In other cases, they arise from the performance needs of signaling that require the network to respond quickly to system-failure or component-failure events.

7.3.1 Signal Unit Formats

The blocks of data transmitted at the signaling-link level are referred to as signal units. There are three types of signal units:

- Message Signal Unit (MSU): Carries user data from level 4.
- Level Status Signal Unit (LSSU): Carries control information needed at the signaling link level.

- Fill-in Signal Unit (FISU): Transmitted when no other signal units are available. This allows for a consistent error-monitoring method so that faulty links can be quickly detected and removed from service even when traffic is low.

7.3.2 Operation

The key functions performed by the signaling-link protocol are flow control, error control, and error monitoring.

Flow Control

Both flow control and error control employ a sliding window technique, in which each Message Signal Unit (MSU) is numbered sequentially. Each new MSU is given a new Forward Sequence Number (FSN) that is one more (modulo 128) than the preceding sequence number. Link status signal units (LSSU) and fill-in signal units (FISU) are not numbered separately but carry the FSN of the last transmitted MSU. All three types of signal units carry piggybacked acknowledgments and negative acknowledgments, in the form of Backward Sequence Numbers (BSN). Figure 7.5 provides an example of an error-free exchange of signal units.

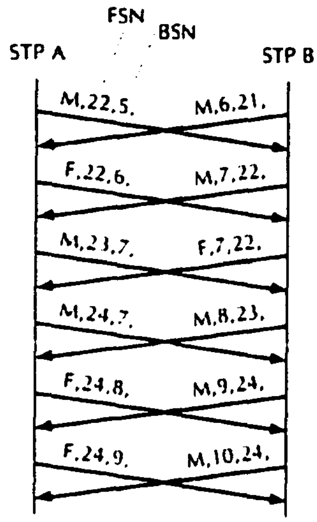


Figure 7.5. Error Free Signal Unit Exchange (M: Message, F: Filler)

Flow control is provided by the LSSU. When one side is unable to keep up with the flow of data from the other side, it transmits an LSSU with a busy indication in the status field. When such an indication is received, all transmission of MSUs must cease; the busy side will notify the other side that it can resume transmission by means of another LSSU. This activity is generally invisible to the next higher level (signaling network level), which may simply notice that throughput has declined. However, if a congestion condition persists and is not reported to the signaling network level, then the performance of the entire signaling network may be degraded. If the network level is aware of a congestion problem, then control packets can be routed around the point of congestion. For this purpose, tight timer control on the allowable duration of the busy condition is imposed. Three rules specify the time constraints:

1. If a receiver becomes overloaded, it must send a busy signal to stop transmission from the other side. The receiver withholds acknowledgment of the MSU that triggered the congestion-control condition and of subsequent MSUs received during the busy condition. If the overload condition persists, the node must repeatedly send a busy indication at intervals of T5 time units (suggested value: 80-120 ms). The other side suspends transmission of MSUs while the busy condition persists.
2. When congestion abates at the receiver, it signals the end of the busy condition by resuming the positive acknowledgment of incoming MSUS.
3. Even if repeated busy indications are received every T5 time units, a node will report to the network level that a link is out-of-service after a time interval of T6 (suggested value: 3-6 seconds).

Error Control

Two forms of error control are defined:

Basic method: applies for signaling links where the one-way propagation delay is less than 15 ms.

Preventive cyclic retransmission method: applies for signaling links with a one-way propagation delay greater than or equal to 15 ms; this would include signaling links established via satellite.

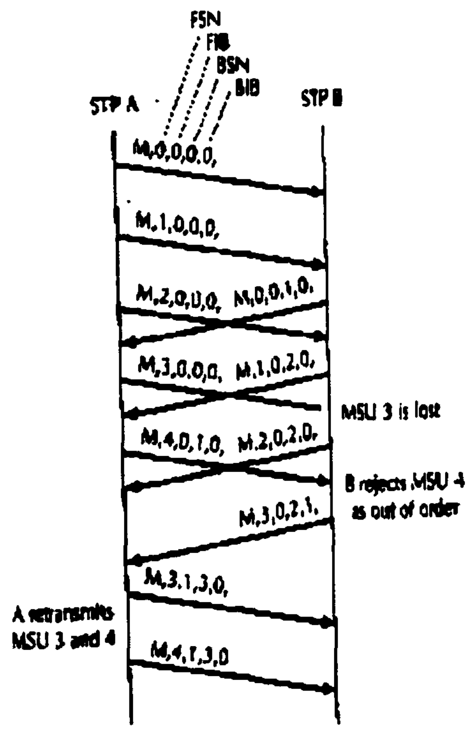


Figure 7.6. Transmission of MSUs with Error Correction

The basic method of error control is go-back-N ARQ. If a node receives a negative acknowledgment in an MSU, LSSU, or FISU, it will retransmit the specified signal unit and all subsequent signal units.

The alternative to go-back-N for long-delay links is preventive cyclic retransmission. For a link with a relatively long propagation delay, each message unit is comparatively short, and the link may be idle most of the time. In such a circumstance, it is not efficient to wait for a negative acknowledgment before retransmitting. Instead, whenever a node has no MSUs to send, it automatically retransmits unacknowledged MSUS, without waiting for a positive or negative acknowledgment. The other side sends only positive acknowledgments.

Because the other side sends only positive acknowledgments, there is a danger that a unit in error may go undetected and uncorrected for a considerable period of time. This is particularly true when the flow of traffic is heavy. In that case, one side may be so occupied sending new units that it rarely performs voluntary retransmission. Accordingly, when a predetermined number of outstanding, unacknowledged signal units exist, the transmission of new units is interrupted and the retained signal units are retransmitted cyclically until the number of unacknowledged signal units are reduced. This feature is known as the forced retransmission procedure.

Error Monitoring

There are two types of signaling-link error-rate monitoring: signaling-unit error-rate monitor and alignment-error-rate monitor. Signaling unit error-rate monitoring is employed while the signaling link is in service and provides a means for detecting when a link should be taken out of service due to excessive errors. This technique is known as a "leaky bucket" algorithm. It

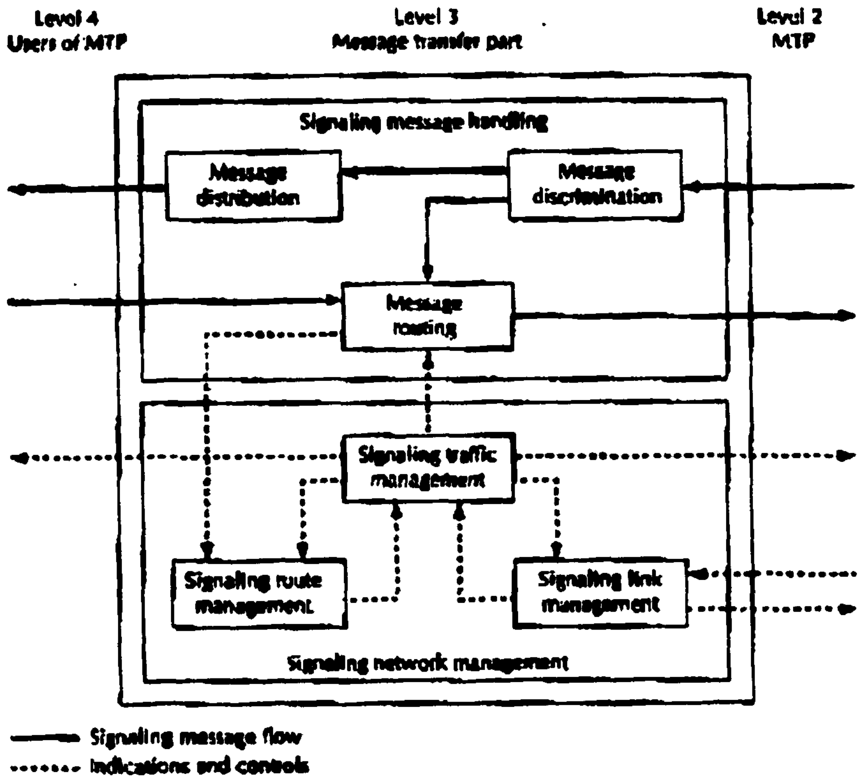


Figure 7.7. Message Transfer Part – Level 3

will ultimately detect a consistent error rate, but it will not be triggered by an occasional surge of errors, such as might be caused by a noise burst. Alignment error-rate monitoring is employed while the signaling link is being initialized and aligned. Alignment simply means that transmitters and receivers are aligned with respect to the opening flag field of each transmitted frame. The alignment error-rate monitoring procedure provides the criteria for rejecting a signaling link for service due to an excessive error rate. For this purpose a counter is used that is initialized to zero and is incremented by one for each signal unit received in error. If the counter exceeds a threshold before the end of an initial "proving period," the proving period is aborted. In the event of failure, the proving-period procedure may be tried up to five times. Five successive failures result in the link being declared unreliable.

7.4 Signaling Network Level

The signaling network level provides the functions and procedures for the transfer of SS7 messages between signaling points. The signaling network level includes functions related to message handling and functions related to network management.

7.4.1 Signaling Message-Handling Functions

The message-handling functions are performed at every signaling point (including signal transfer points). They are based on the parts of the message signal unit known as the routing label and the service information octet. The message-handling functions fall into three categories:

Discrimination:

Determines if a message is at its destination or is to be relayed to another node. This decision is based on analysis of the destination code in the routing label of the message. If this is the destination, the signal unit is delivered to the distribution function; otherwise, it is delivered to the routing function. The discrimination function is only needed in the STP.

Routing:

Determines the signaling link to be used in forwarding a message. The message may have been received from the discrimination function or from a local level 4 entity.

Distribution:

Determines the user part to which a message should be delivered. The decision is based on analysis of the service indicator portion of the service information octet.

The routing decision is based on the value of the Signaling Link Selection (SLS) field, which is assigned by a user part in level 4. For a given source/destination pair, several alternate routes may be possible; the value of the SLS field specifies which particular route is to be followed. With a 4-bit field, a total of 16 different routes through the network may be defined.

These different routes are, in effect, different internal virtual circuits. In general, all of the control signals associated with a single call will follow the same route; this guarantees that they will arrive in sequence. However, the

MTP needs to distribute traffic uniformly. This requirement can be satisfied if the user part varies the route selection from one call to the next.

7.4.2 Signaling Network Management Functions

The other main component of the signaling network level is signaling network management. The main objective of this component is to overcome link degradations (failures or congestion). To meet this objective, the signaling management function monitors the status of each link, dictates alternate routes to overcome link degradation and communicates the alternate routes to the affected nodes, and recovers from the loss of messages due to link failure. The goal for SS7 is no more than 10 minutes of unavailability per year for any route. This goal is achieved through redundancy of links and dynamic rerouting.

This emphasis on the internal management of the network is rare; virtually all other network protocols make no mention of network management. In most cases, it is preferable to leave network management details to the provider, so that the provider can pursue the most cost-effective approach and be responsive to changes both in customer expectations and advances in technology. However, in the case of SS7, there are strong reasons for the emphasis on network management:

1. The function being specified is critical. The performance of a network's control signaling architecture affects all subscribers to the network.
2. The various networks involved must support international traffic. Degradations in one nation's signaling system will have repercussions beyond that nation's borders. Thus, some international agreement on the degree of reliability of national networks is indicated.
3. Recovery and restoration actions may involve multiple networks (e.g., in the case of international calls). If SS7 did not include failure and congestion recovery procedures, it would be necessary for the administration of each public network to enter into bilateral agreements with a number of other networks.

Three categories of the signaling network management functions

- Signaling traffic management.
- Signaling link management.
- Signaling route management.

The signaling network management component as a whole is quite complex. All of the procedures relating to signaling network management involve the monitoring and control of the status of various entities, including signaling links, signaling routes, signaling points, and signaling route sets.

Signaling Traffic Management

Signaling traffic management is used to divert signaling traffic (without causing message loss or duplication from unavailable signaling links or routes to one or more alternative signaling links or routes) or to reduce traffic in the case of congestion. As an example of the functions performed in the category of signaling traffic management, let me consider the changeover procedure. The objective of the changeover procedure is to ensure that signaling traffic carried by a link that becomes unavailable is diverted to the alternative signaling link(s) as quickly as possible while avoiding message loss, duplication, or sending out of sequence.

1. Transmission and acceptance of MSU on the concerned signaling link is terminated.
2. Alternative links are determined to construct an alternative route.
3. Those messages in the retransmission buffer of the unavailable link that have not been received by the far end are identified.
4. The identified messages are transferred to the transmission buffer of the alternate link.

Steps 3 and 4 are accomplished in the following way. A signaling point that recognizes the unavailability of a link sends a Changeover Order (COO) message to the remote signaling point over some available alternate route. The message value field contains the forward sequence number of the last message accepted from the unavailable signaling link. When the other side receives the COO, it responds with a Changeover Acknowledgment (COA) message, which contains the forward sequence number of its last accepted MSU. The two sides are now able to resume the exchange of MSUs containing user-part information over an alternate route, maintaining the proper sequence with no losses or duplications. The new route is decided by the signaling message-handling function of the two endpoints.

In this example, traffic between signaling points A and D is initially carried via signaling point B. When the link between A and B becomes unavailable, A sends a COO message to D via signal transfer point C. C responds along the same route with a COA. Subsequently, all MSUs follow the route through C.

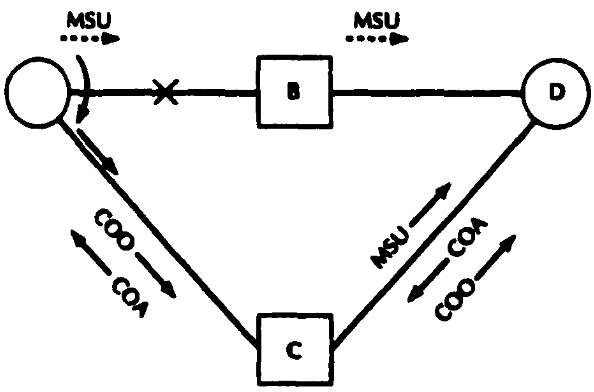


Figure 7.8. Changeover Procedures

Signaling Link Management

Signaling link management is used to restore failed signaling links, activate new signaling links, and deactivate aligned signaling links. There is a basic set of mandatory functions that perform signaling link management for links directly connected to a signaling point. There are additional, optional functions that allow for more efficient use of signaling equipment when signaling terminal devices have switched access to signaling data links.

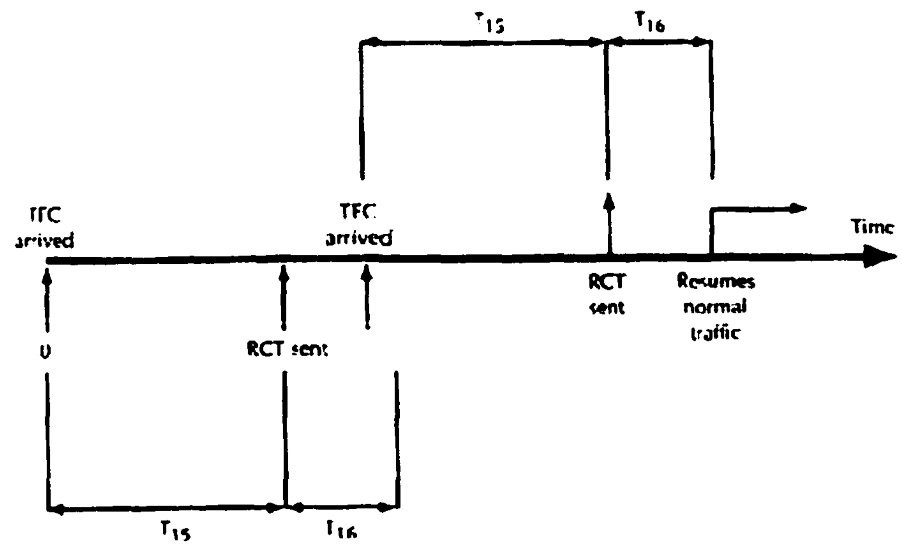
Signaling Route Management

Signaling route management is used to distribute information about the signaling network status in order to block or unblock signaling routes. As an example of the functions performed by signaling route management, consider the signaling-route-set congestion procedure: the procedure used by STPs to control congestion. Whereas congestion occurring between signaling points can easily be handled by flow control at level 2, when congestion occurs on a link emanating from an STP, the source SPs that send messages through that link must be controlled.

Each outgoing link of each signaling transfer point has a transmit buffer with three threshold levels: congestion onset (T), congestion abatement (A), and discard (D), where $A<T<D$. The discard threshold is equivalent to the buffer capacity. The set of routes through the link in question is either congested

or uncongested, depending on the buffer occupancy history. When the link is uncongested and the arrival of incoming message units causes the portion of the buffer that is filled to exceed T, then the link is considered to be congested, and a choking message, the transfer controlled (TFC) message, is sent to the source of the node that caused the threshold to be reached and to the source of all subsequent messages while the link is congested. When the buffer occupancy level decreases below A transmission of choking messages ceases and the link is considered uncongested. Thus, the status of the link when the occupancy of the buffer is between A and T can be in either state, depending on buffer occupancy history. This is a "hysteresis" effect, which prevents frequent changes of the link status.

When a signaling point receives a choking message, it stops generating messages for the route involved. Two timers are used to determine when to resume normal routing. When a node receives a choking message, it starts a timer with duration of T_{15} seconds. When the timer expires, the node sends a signaling route set congestion test (RCT) message to the same destination along the normal route and waits for a time period T_{16} to see if another choking message is returned in response to this RCT message. If a choking message is received within T_{16}, the timer is reset and the process begins again. If T_{16} expires, then the node assumes that the congestion has abated and resumes normal routing.



TFC = Transfer controlled message
RCT = Signaling route set congestion test message
T_{15} = Timer: waiting to start signaling route set congestion test
T_{16} = Timer: waiting for route set congestion status update

Figure 7.9. Behavior of an SP for the SS7 Congestion Control

Figure 7.9 illustrates the operation of the algorithm at a signaling point. At time t = 0, the SP receives a TFC message. The SP stops generating user-part messages for the indicated destination and starts the T_{15} timer. When this timer expires, the SP sends an RCT message and prepares to wait for time T_{16}. Before this timer expires, another TFC message is received, and the process begins again. In this example, after the second RCT message, timer T_{16} expires and the SP resumes normal transmission.

7.5 Signaling Connection Control Part

The Signaling Connection Control Part (SCCP) was developed as it became apparent that SS7 would have to support more than signaling and needed more explicit addressing and more sophisticated services between remote signaling points. The signaling network level does not provide all of the routing and addressing capabilities that the OSI model dictates for the network layer. As an example, the message distribution function provides only a limited addressing capability. For newer user-part applications, a more complex specification of the user of a message at a node is necessary; this can be provided by the SCCP. The SCCP enhances the connectionless sequenced transmission service provided by the MTP, to meet the needs of those user parts requiring enriched connectionless or connection-oriented service to transfer signaling information between nodes. Classic circuit-switched telephone call-related signaling does not use SCCP. For this application and those user parts for which MTP suffices, the extra overhead of SCCP can be avoided.

Thus, the enhancements provided by SCCP over those available in just the message transfer part are in the areas of addressing and message transfer services. The addressing capabilities of SCCP extend those found in MTP, which is limited to delivering a message to a specified node and using a 4-bit service indicator to distribute messages within a node. SCCP supplements this capability by providing addressing that uses destination point codes plus SubSystem Numbers (SSN). The SSN is local addressing information that identifies each of the SCCP users at a node. Another addressing enhancement is the ability to address messages with global titles, such as dialed digits, that are not in a form usable by MTP for routing. SCCP provides a mapping facility for translating global titles into an address of the form DPC + SSN.

SCCP also provides enhanced message transfer services, two connectionless and two connection-oriented. The four classes are:

0 Basic connectionless.

1 Sequenced (MTP) connectionless class.

2 Basic connection-oriented.

3 Flow-control connection-oriented.

In Class 0 service, a user provides a block of data, referred to as a Network Service Data Unit (NSDU), to SCCP for delivery to a user at another node. The NSDUs are transported independently and may be delivered out of sequence. This is a pure connectionless, or datagram service.

For Class 1 service, Class 0 is enhanced with the ability to specify that a particular stream of NSDUs should be delivered in sequence. SCCP does this by assigning a sequence number to each member NSDU and giving all messages in the stream the same signaling link code (Figure 7.4).

The remaining two classes of service operate over logical connections, called signaling connections. These connections are equivalent to virtual circuits through the signaling network. Each logical connection is given a unique signaling link code. Class 2 provides this basic connection-oriented service.

For Class 3 service, Class 2 is enhanced with the ability to perform flow control over a logical connection. Also, the detection of message loss and missequencing is provided. In the event of lost or missequenced messages, the signaling connection is reset and notification is given to higher layers.

As with most standards at the various layers of the OSI model, SCCP can be specified in terms of the services that it provides to higher layers and the protocol between peers SCCP entities in different nodes.

The structure consists of four functional blocks:

1 Connection-oriented control: Controls the establishment and release of signaling connections and provides for data transfer on signaling connections.

2 Connectionless control: Provides for connectionless transfer of data units.

3 Management: Provides capabilities beyond those of MTP to handle the congestion or failure of either the SCCP user or the signaling route to the SCCP user. With this capability, SCCP can route messages to backup systems in the event failures prevent routing to the primary system.

4 Routing control: Upon receipt of a message from MTP or from functions (1) or (2), SCCP routing provides the necessary routing functions to either forward the message to MTP for transfer or pass the message to (1) or (2). A message whose "called party address" is a local user is passed to (1) or (2), while one destined for a remote user is forwarded to the MTP for transfer to a distant SCCP user.

SCCP Services

As with any protocol service, SCCP can be expressed in terms of primitives, which can be viewed as commands or procedure calls, with parameters. Each type of primitive appears in one or more variations (request, indication, response, confirm), depending on the requirements of the service.

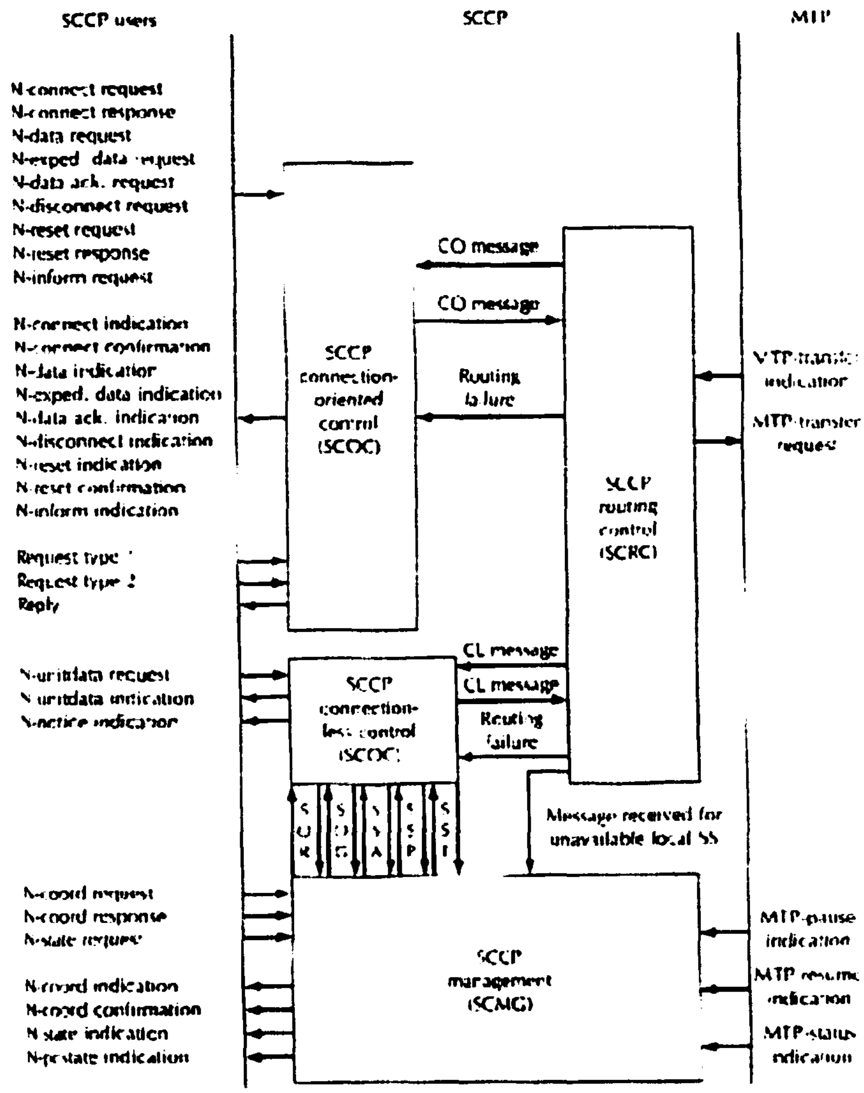


Figure 7.10. SCCP Overview

Connection-Oriented Services. The connection-oriented service provided by SCCP is based on the OSI network service, which is defined in ISO 8348 and ITU-T X.213.1 Key characteristics of the network service, as listed in the standard, are

- Independence of underlying communications facility: Network service users need not be aware of the details of the subnetwork facilities used.

- End-to-end transfer: The network layer performs all routing and relaying without concerning the network service user.
- Transparency: The network service does not restrict the content, format, or coding of the user data.
- Quality-of-service selection: The network service user has some ability to request a given quality of service.
- User addressing: A system of addressing is used that allows network service users to refer unambiguously to one another.

Primitives		
Generic Name	**Specific Name**	**Parameters**
N-CONNECT	Request Indication Response Confirmation	Called address Calling address Responding address Receipt conformation selection Expedited data selection Quality of service parameter set User data Connection identification
N-DATA	Request Indication	Confirmation request User data Connection identification
N-EXPEDITED DATA	Request Indication	User data Connection identification
N-DATA ACKNOWLEDGE (for further study)	Request Indication	Connection identification
N-DISCONNECT	Request Indication	Originator Reason User data Responding address Connection identification
N-RESET	Request Indication Response Confirmation	Originator Reason Connection identification

Table 7.2 Network Service Primitives for Connection Oriented Services

Connection establishment begins with a user request, contained in an N-CONNECT request primitive. In addition to specifying the called and calling user, the primitive can request certain services to be provided for the requested connection:

- Receipt confirmation selection: Ordinarily, the network service will not confirm that data has been delivered to the other side; it is assumed that the data is delivered. However, the user may request that explicit confirmation be provided.
- Expedited data selection: The user may also request that an expedited data service be available.
- Quality of service: The user may specify two quality-of-service parameters. First, the user proposes one of the four classes to be used for the connection. If appropriate, the user also proposes a flow-control window size.

The remainder of the connection establishment process involves confirming the setup of a connection and the negotiation of quality-of-service parameters. The negotiation proceeds as follows:

1 The calling user specifies Class 3 or Class 2 in the quality-of-service parameter of the N-CONNECT request. If Class 3 is selected, a flow control window size is specified.

2 The SCCP at the calling user's node, at any intermediate STPs, or at the called user's node may downgrade a Class 3 request to a Class 2 request. If Class 3 was requested and is not downgraded, any of these nodes may reduce the window size.

3 An N-CONNECT indication is issued to the called user with the resulting quality of service.

4 The called user responds with an N-CONNECT response. The called user may downgrade the class or reduce the flow-control window size.

5 The final quality of service is conveyed to the calling user in an N-CONNECT confirm.

Chapter 8

Intelligent Multimedia News Networks (IMNN)

Today many industries are using intelligent networks. Telecommunications is the pioneer in building and deploying the intelligence in the networks. In telecommunications there are two major directions of force on the evolution of the public domain Intelligent Networks (INs) in the United States. First, the carriers of major data transport facilities around the country are evolving a special type of network intelligence to monitor and implement the network performance requirement from a data-throughput point of view. Optimal utilization of network switching and transmission facilities to maximize revenues is the major design criterion. With the advent of high-speed fiber optic systems that deploy multiple gigabit capacity in digital pipes, it is essential to be able to add, drop, and cross-connect channels appropriately from the digital superhighways spanning the country. Second, regional operations in the more densely populated areas of the United States are developing the capability to provide new intelligent network services (IN/1, IN/1+, and IN/2 architecture). In addition, the ISDN capabilities for distributing data are being implemented by way of the ISDN switches, which are being implemented throughout the country.

In this study the IN deployment in the news network was explored and the implementation details were investigated. Following sections describe the architectural considerations of the individual components:

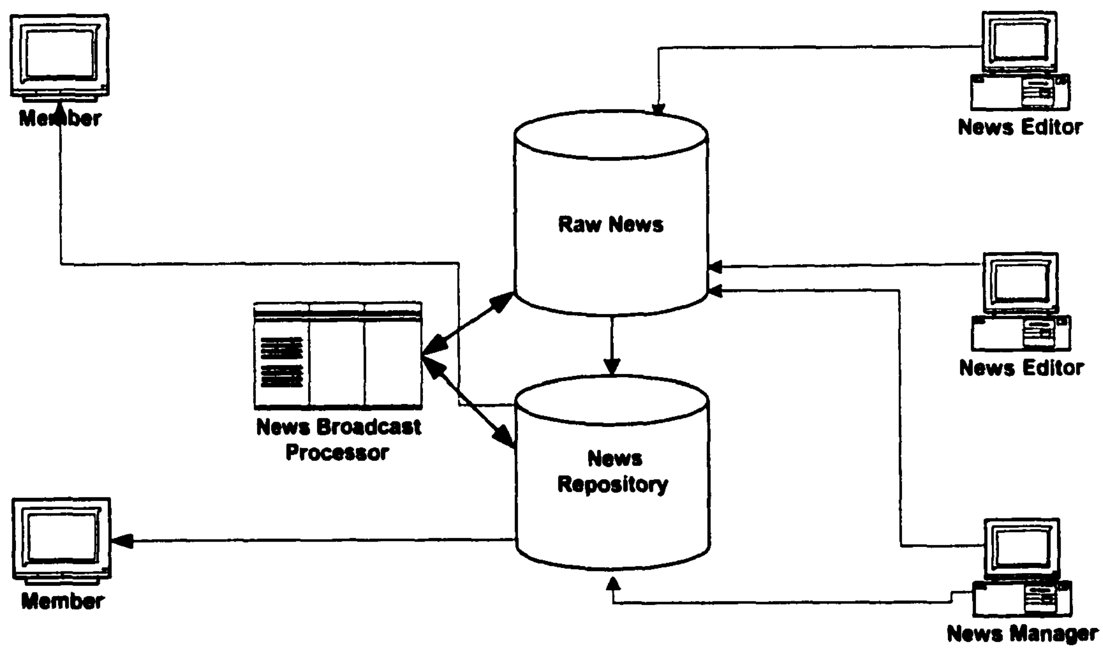


Figure 8.1. News Creation and Broadcast in current system

Similar concepts of the IN/2 (Intelligent Network 2) and AIN (Advanced Intelligent Network), can be deployed in the news network, which are becoming more and more important with the introduction of the intelligent components. In this research we are discussing the existing news network of the Associated Press and then incorporating the intelligent components to design the Intelligent Multimedia News Network (IMNN).

The creation of the news is triggered as soon as an event occurs, e.g. war, fire, elections, and crime, etc. A news reporter and or agency picks up the leads, collects the facts, supplements the history and or background and reports to its network. The process gets triggered to assign a reporter or a team based upon the importance and complexity of the news. It goes through several desks and levels before it is decided to be included as part of the news stream. From there on it becomes as part of the broadcast element in the news.

An architectural configuration of the IMNN and its components that have all the advantages and the features of any IN are discussed. The areas where this network differs from conventional intelligent networks, the need for syntactic rule-based "News Processing," and the need for a "News Control Point," rather than a Service Control Point (SCP) of the traditional intelligent network are also explored. Finally, we present how the IMNN is compatible with evolving networks, such as the SONET, campus networks, or any other speed digital network in the circuit and the packet mode. The Asynchronous Transmission Mode (ATM) or the hierarchy of the SONET rates does not constitute any problem to the evolution of the IMNN. We project how the

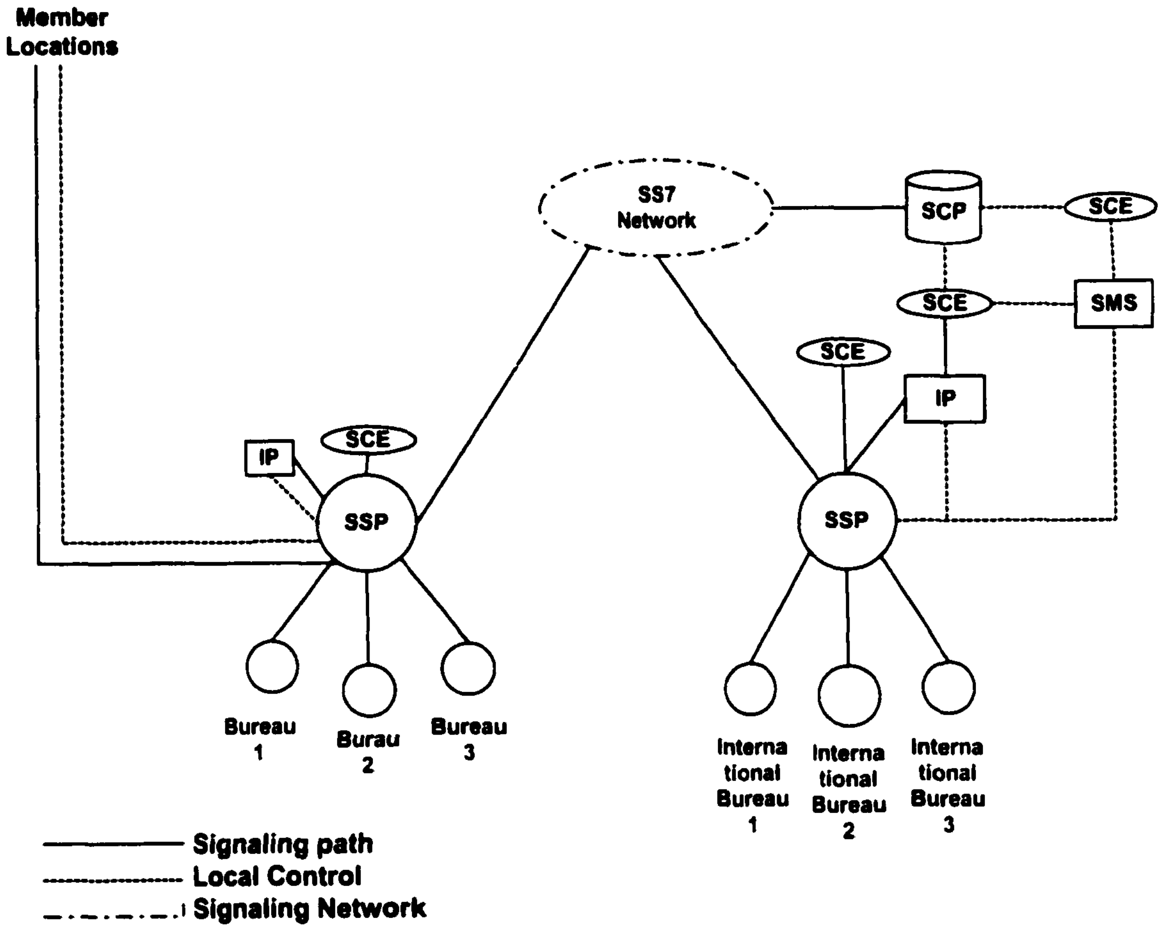
IMNN may evolve over the next decade to take full advantage of the already committed networks, such as IN/1, IN/1+, IN/2, AIN, UIN together with ISDN, SONET and ATM, already appearing on the horizons of the society [1].

8.1 Network Designer's Perspective

The perspective of the network designer centers around technology and topologies. Ease of access and processing speed are primary considerations. This section discusses the following areas of concern for network design:

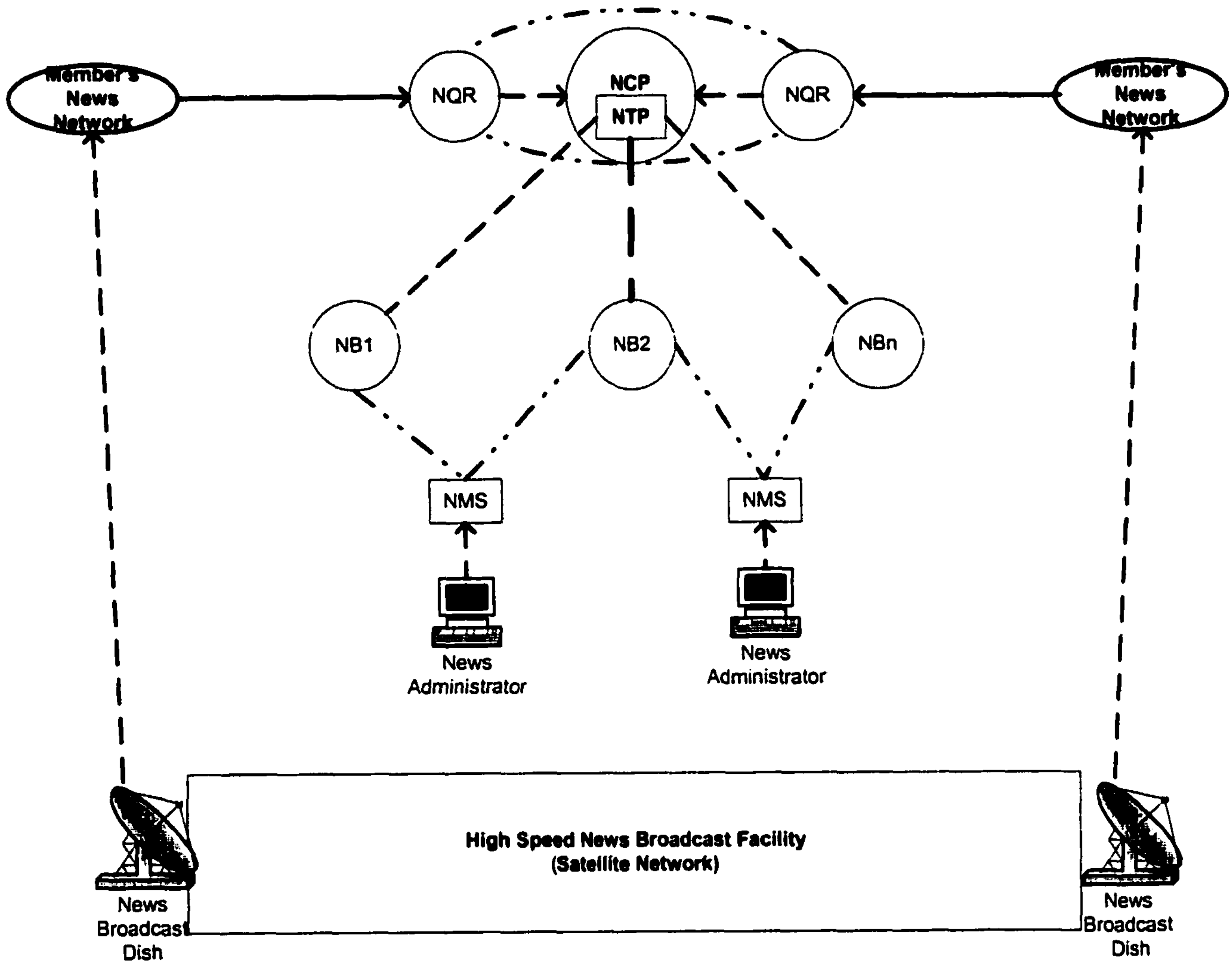
8.1.1 News Networks

IMNN is network based. It interconnects numerous news networks in the same manner that Metropolitan Area Networks (MANs) and Wide Area Networks (WANs) interconnect Local Area Networks (LANs). Network nodes and gateways located in IMNN monitor and direct the flow of information in and out of the IMNN to a knowledge query point and then on to a news ring. The querying network receives the outflow of queries or commands flow up to the news query point, and specifically compiled information from the news ring.



SSP = Service Switching Point
SCP = Service Control Point
SCE = Service Creation Environment
SMS = Service Management System
IP = Intelligent Peripheral
SS7 = Signaling System Number 7

Figure 8.2. IN/2 based Intelligent News Network



News Query
News Signaling
News Maintenance

NTP = News Transfer Point
NCP = News Contorl Point
NQR = News Query Ring
NMS = News Management System
NB = News Bases

Figure 8.3. Intelligent News Network with dual Broadcast Facility

8.1.2 Very High-Speed Data Networks

Classic circuit-switched, telephone type networks are bi-directional, generally with the same data rates. These are likely to become saturated (especially during the peak traffic hours) with the high volume of personalized data traffic in the IMNN. However, Very High-Speed Data Networks (VHSDN) dedicated to the transport of large blocks of data are evolving and show enormous promise. Members and/or non-members trying to query the old and current news require initial processing to remove inaccuracies of language or uncertainty of posed questions. However, when the information of the query is complete, a packet on the backbone packet-switched news network accomplishes communication to the database, which is equivalent to the SS7 network for intelligent networks. In the conventional sense, these SS7 features control and facilitate the new services and control the circuit switched channels within the Service Switching Points (SSPs). In the IMNN, the news ring can now transmit blocks of customized data in response to the query/command received also via a high-speed packet network linking the news ring and the numerous member networks. Fast packet transmission technology is also emerging through the national and international networks. It is interesting to note that normal flow of information becomes unidirectional in IMNN, as opposed to the bi-directional flow of control information in the conventional Ins. Thus, bandwidth allocation can be made optimal in IMNN.

8.1.3 News Bases (NB)

If the information contained in the news bases is highly specialized, current, and at the leading edge technology, a large number of networked or independent bases may become essential. In any case, and individual news management system for every news bases is necessary. Small, independent, dynamic databases exist, such as, stock market databases; air lines databases, or even patient databases. These can be suitably modified to store independent separate modules of news, called News Bases (NB).

8.2 Basic Elements of the Network

Two additional elements, the News Ring (NR) and the News Management System (NMS), are important considerations in the design of the network. This section discusses these elements and their functionality.

8.2.1 News Ring (NR)

Isolated News Bases (NBs) can become too expensive to dispense specialized modules of information. Conversely, widely documented subjects may need multiple terabytes of storage space. NR provides a viable compromise for containing and managing the storage of very vast bases. The ring, which can also manage small NBs, is capable of accessing numerous bases and caters to growing news and new areas of research through the addition of a node to the ring. New nodes may also be added to this ring as new news services emerge and evolve. Old nodes may be deleted, when it is no longer necessary to retain obsolete bases or when a particular base is no longer economically viable to sustain. The access and cost of access may then both be controlled by software and protocols as they are used in the nodes that access the long distance and trunk facilities of any communication system.

8.2.2 News Management System (NMS)

A group of news editors and their managers are responsible to create, update and modify the News Bases (NB). This functionality is currently performed manually and we do not foresee any change.

There is another side of the NBs and the NMS. Updating can be automatic when the sources of generating active news (such as weather stations and stock markets) are transparently networked to feed the NBs directly. Such automated environments exist. For example, a stock market transaction is automatically entered in the stock market database, weather satellites automatically update meteorological centers; and defense radar trackers automatically warn national defense networks. In a paperless society, we can foresee the sources of news being in direct communication with the consumers of news via the IMNN environments. For the accuracy, validity, integrity and authentication, checkpoints analogous to the News Query Point are essential only to the entry of the authenticated and authorized information in the NBs and in the NR.

8.3 New Key Elements in the Network

New key elements in the network involve checkpoints, transfer points, logic interpreters, and protocols, all necessary for the operation of the News Bases (NBs). This section discusses all these new elements.

8.3.1 News Query Point between IMNN and Member's News Network

In order for the IMNN to perform intelligent functions in responding to queries from member's networks and interpretive interfaces, it is necessary to accept structured queries and infer which news base to address in order to generate a response. In conventional intelligent networks, the network response is initiated when the Service Switching Point (SSP) detects a trigger condition. We see the News Query Point (NQP) as a new element in the general architect of the intelligent news networks. The functionality of the NQP is critical to the success of this network environment. It accomplishes the function of a preprocessor of the queries through the network; a preprocessor is the very first checkpoint of legitimate subscribers, here referred to as members, and their transactions in the form of questions/retrieval that may be initiated. In a sense, this is the software equivalent of the physical door placed at a money access center. If members are using the Internet for a service, the NQP validates members and their authorization codes (customer identification codes) at entry to the network. The NQP then accesses authorization and billing information, and verifies credit card information. Or the NQP will terminate transactions involving illegal users, hackers, or virus planters at this level by the network.

This software barrier at the NQP also has a query processor that identifies what a member asks about a certain subject matter. The parser in the NQP separates the query or solicitation and/or command from the subject matter in the query. The solicitation can be in the form of a "what," "how," "why," "when" question, or some combination thereof, about the particular subject matter in the query. The command can be in the form of a "lookup," "retrieve," and so forth. Every query and/or command is an operation code with its own customized protocol, and every subject matter can be viewed as an operand. Together they form and IMNN instruction. A transaction can be a single instruction or an assembly of such instructions depending upon what the user wishes to accomplish. This assembly is then executed or interpreted by the network. Every instruction accomplishes a finite step in "News Processing" in the network.

8.3.2 News Transfer Point

The News Transfer Point (NTP) is a software system that acts as a second check before any transaction with the network takes place. It identifies the subject matter via a modified English language dictionary. Valid dictionary words (including proper names of people, places, materials, and objects, about whom information exists in the NBs) are checked to see if they belong to the classification of news in the news ring. In a sense, the NTP identifies the subject matter in context to the structure of the overall news by identifying the "generalized news address" of the subject matter sought. The generalized news address is the Dewey Decimal Number, the Library of Congress identifier or some other coded identifier. Misspelled words, nonsense subject matters, and improper query in context to the subject matter are returned to the user. Unauthorized users of the particular subject matter, "sneakers," "browsers," and accidental information leaks are expelled in the interest of network security.

The modified dictionary identifies the subjects in the news bases within the news ring. The subject matter is itself classified and cataloged according to a subject hierarchy, such as the Dewey Decimal System, the Library of Congress system, the Asian Library system, or even the Princeton classification system. The two syntactical components of the initiator's query are reconstructed here. Only if the news ring can handle the interrogation from the campus network will the syntactical components of the query be allowed to proceed in the network. This procedure is similar to the operating system procedure of any computer system that does not permit the execution phase of a program, if there were fatal errors in compilation.

8.3.3 NQP-NTP Combination

Whereas the News Query Point (NQP) identifies what is to be done, the News Transfer Point (NTP) has a news address of the subject matter. Together, they are a complete, legitimate instruction to the network. Typically, the query is converted to a machine-language instruction, which has two components, namely, the operation code (identified by NQP) and the operand/ operands (identified by NTP). It is this single instruction, or a group of such instructions, that drives the IMNN like the basic commands or the micro program associated with the instructions that drive any computer. A lot of flexibility in the implementation of the NQP can be seen, such as that found in the implementation of a basic hard-wired or micro programmed central processing unit. This flexibility may be used to design an optimal NQP for any application by selecting the most suitable opcodes and/or network micro code for the application.

8.3.4 News Control Point (NCP)

The NCP will perform all the functions similarly defined as an SCP in IN/2. Besides that, it will also accommodate all incoming queries to the network for the specific news items. The NCP finally converts the "news address identifier" to the actual physical or logical address of locations in the news bases where that news is embedded. The news address in any of the news classification systems (such as the Dewey Decimal System, Library of Congress, and so forth) is converted to a network address of the news base. It is necessary to understand that this address translation is hierarchical; that is, the first segment of the address alone can identify the NB and the latter segment or segments can actually locate the exact track-sector address of the mass storage system. We foresee the possibility of building massive RAMS that hold memory addresses corresponding to the news address of the subject matter, thus further reducing the time for address translation.

8.3.5 News Logic Interpreter (NLI)

When the operation code (query/command) and operands (hierarchical number, for example, the Dewey Decimal Number of the subject matter) are both identified and validated, the News Logic Interpreter (NLI) becomes instrumental in dispatching (in any appropriate format or protocol) the executable binary language instruction to the network components. The network components carry out their respective function (as they are performed during call-processing in traditional telecommunication networks) to send the answer back to the initiator of the query. The information sought is extracted from the appropriate NB or news ring.

The high-speed digital network is used to burst this information to the campus network via its gateway to the Very High Speed Digital Network (VHSDN). At this stage the X.25, or any other suitable network protocol may be implemented. Once again, it should be noted that the function of the NLI is not exactly the same as that of the SLIP in IN/2.

8.3.6 News Control Point (NCP) Protocol

Wherever the physical or logical address of the actual news (the subject matter) may be located, the command (that is, opcode) for that piece of news must be communicated to the NB where that subject matter resides. This network instruction received at the appropriate NB is executed in its entirety. The protocol can be in the form of operation codes or it can be further condensed to make the best use of the network transmission and protocol processing facilities. If these NBs have CPUs that function on a standard micro code, then the News Control Point (NCP) needs to communicate the address of the micro program in the control memory of the CPU. Typically, we expect the complex queries to have more than one "instruction" to the network.

8.3.7 IMNN Protocol

We envision that the IMNN protocol will have its network functions to some extent similar to that of the Signaling System Number 7 protocol. However, there are several dissimilarities as well, though minor. For its implementation today, these dissimilarities can be ignored temporarily. As more and more IN based technologies will be developed, the new protocols and standards evolve, making IMNN implementation more viable in future.

8.4 Existing Platform For IMNN

To some degree, the functions and components of the conventional INs (such as, IN/1, and IN/2) and those of the IMNN are analogous. But we also foresee that forcing IMNN functions from IN/2 architecture is sub optimal, if not actually constituting misuse. The concept of recognizing the structure of news, and then being able to compose a response from the network, offers some distinct advantages. A properly designed IMNN can eventually combine all the advantages that distance learning and mass media systems have to offer. For these reasons, the implementation of the IMNN can be seen as an enhancement to the implementation of the conventional INs.

8.5 Architecture of IMNN

An architectural illustration is given in Figure 8.3. A transaction request is initiated from any member site. This request flows inward, via the News Query Point (NQP), into the News Query Ring (NQR). Here the user/member is identified and operational legitimacy (authentication) is established. The information is then fed into the NTP to establish the validity of the subject matter being transacted. The syntax of the command is formulated and verified to check who is doing what.

If such legitimacy is valid, the identifier number is translated to the logical or physical address of the NB on the NR at the News Control Point (NCP) at the center of the diagram. The NCP dispatches the function (in executable binary code) that the particular action has to be performed.

8.6 Conclusions

In this chapter we have proposed the infrastructure of a news network, which is not one-directional to only sending the news to the news outlets (i.e. to the member sites). Rather, keeping in mind to receive the feed coming in as well from the members. Even with a broader scope receiving the news from other news service providers too, for example Reuter, AFP and UPI. This process will enhance functionality, e.g. Reuters, with an employee work force of over 16,000, can specialize in financial market services and broadcast in multi-languages to enhance the overall services. Likewise, AFP has strong presence in the French speaking part of the Africa and has a geographical specialization there. The associated Press has its strength in new television quality digital products.

Also, by increasing the retention time of the current news and keeping the contents longer will allow the researchers to write the news features and material based upon the historical contents of the news. The storage technology has now provided the capability to store tera-bytes of data more economically.

Likewise, the search engines have provided the capability to query and analyze the required contents. The storage of news with proper indexing based upon several indicators will improve the search capabilities.

Several new services can be offered based upon the archived news to the researchers and the news feature writers for internal and external use to the company, to a direct subscriber of the service or through the internet.

Chapter 9

NETWORK SIMULATION

In this chapter we will present an overview of the use of simulation in the design and analysis of networks. It is often of interest to study a proposed or existing network to improve its performance. However, it is generally necessary to use a model for this purpose, since experimentation with the network itself is disruptive, expensive, or simply impossible (e.g., the network has not yet been built).

If the relationships that compose the model are simple enough, it may be possible to use mathematical methods (such as algebra, calculus, or probability theory) to obtain exact answers to the questions of interest; this is called an analytic solution. As a matter of fact, analytic queuing models have been used for years to study performance issues for networks and computer systems. However, as network topologies and protocols have become more complex, analytic methods have become increasingly inadequate. Additional shortcomings of analytic queuing models are as follows:

- Only steady-state results are typically possible
- It is difficult to obtain performance measures other than mean values (e.g., the 85th percentile of end-to-end delay)
- Original analytic solutions require considerable mathematical sophistication on the part of the analyst

Because of the drawbacks of analytic methods, there has been a considerable increase in the use of simulation for network analyses during the past ten years. This has in turn resulted in the introduction of a number of new simulation products specifically for networks. In a simulation a mathematical/logical model is numerically evaluated over a time period of interest, and performance measures are estimated from model-generated data. Simulation analyses are applicable to systems of almost any level of complexity. Perhaps the only impediment to the use of simulation is the potentially large amount of computer execution time required to process messages for high-traffic-rate networks.

9.1 Objectives of Simulation

The following are some of the benefits of using simulation to design and analyze networks:

1. Determination of the system-wide impact of making "local" changes to the network
2. Improved system performance (delays, throughput, etc.)
3. Reduced expenditures
4. Insurance that performance objectives are met before equipment is bought or leased
5. Identification of bottlenecks before system implementation
6. Reduced system development time

Simulation has addressed a number of specific communications issues, including:

1. How will my network perform when the traffic load increases?
2. What are the requirements for number/speed of links, number/speed of switches, and buffer sizes for my wide area network?
3. What will be the impact of a link failure?
4. Which protocols will provide the best network performance?
5. What is the best design for my new network?
6. What is the desirability of ATM or frame relay?

7. What will happen when additional PCs or workstations are added to my local area network (LAN)?
8. What will be the impact of additional applications on my network?
9. Will my client-server strategy provide the required performance?
10. How many satellites are needed to provide a certain level of service between two earth stations?

The following is a list of performance measures that are commonly used in simulation studies of networks:

Throughput (e.g., in kilobits per second)

End-to-end delay

Delay from point A to point B in a network

Number of "data units" in a queue or a buffer

Utilization of nodes or links

Probability of a blocked call

Probability of a lost call (mobile system)

Number of collisions and deferrals (LAN)

9.2 Simulation Software for Networks

One of the major tasks in building a simulation model of a network is that of converting a system description into a computer program. An analyst may use either a general-purpose programming language (e.g. C or C++) or a simulation software package for this purpose. Some advantages of a programming language are as follows:

1. Most modelers already know a programming language, but this is often not the case with simulation software.
2. C or C++ are available on virtually every computer, but a particular simulation software product may not be available for the analyst's computer.
3. Software cost will generally be considerably lower (but not necessarily project cost).

The major advantage of using simulation software is that they automatically provide most of the features needed in programming a simulation model, resulting in a significant decrease in programming time (and usually projects cost). Simulation software also provides a more natural framework for system modeling. In general, we believe that an analyst would be prudent to use simulation software to model a communications network.

There are three types of software for simulating networks. A general-purpose simulation language is a simulation package that is general in nature (e.g., it could also be used for modeling manufacturing systems or for combat modeling), but may have special features for communications such as explicit modules for Ethernet or token ring. Examples of simulation languages are

Arena, AweSim, BONES DESIGNER, GPSS/H, MODSIM III, SES/workbench, and SIMSCRIPT II (only BONES DESIGNER and SES/workbench have communications modules).

A model is developed in a simulation language by writing a program using the language's modeling constructs, which include entities (messages), attributes (message type or destination), resources (nodes or links), and queues (buffers). The major advantage of most languages is their ability to model almost any kind of network, regardless of its complexity or uniqueness. Possible drawbacks of languages, as compared to some simulators (see below), are the need for programming expertise and possibly the long time spent coding and debugging that is associated with modeling complex networks.

A communications-oriented simulation language is a simulation language that is specifically oriented toward networks OPNET Modeler is such a product. Advantages are possibly reduced programming time and modeling constructs oriented toward communications systems.

A communications-oriented simulator, in its most basic form, is a simulation package that allows one to simulate a network in a specific class of networks with no programming. Examples of basic simulators are COMNET

III, NETWORK II, OPNET Planner, and SES/strategizer. The particular network of interest (in the domain of the package) is selected for simulation by choosing items from menus (typically using a point-and-click approach), by filling in dialogue boxes (forms), and by the use of graphics. Typical modeling constructs for a LAN simulator are; LAN types (Ethernet, token ring, etc.), stations on a LAN (PCs or work stations), LAN interconnection devices (bridges and routers), and traffic (message) generators. The major advantage of a simulator is that program development time may be considerably less than for a simulation language. This may be very important given the tight time constraints in many business environments. Another advantage is that simulators have modeling constructs closely related to the components of a network, a very desirable feature for someone like a network manager. Also, people without programming backgrounds and those who use simulation only occasionally often prefer simulators because of their ease of use.

The major drawback of basic simulators is that they are limited to modeling only those network configurations allowed by the package's standard building blocks. Thus, if a communications system has some unique features, they might have to be modeled in an approximate manner when using certain simulators. This difficulty can be largely overcome in the Developmental Version of COMNET III, which allows existing modeling constructs to be modified and new constructs to be added. Also, OPNET Modeler can be used to create/modify modules for OPNET Planner. Note that an important feature for simulation software to be used for network modeling is fast model execution speed, because in some networks a very large number of messages will need to be simulated.

9.3 Developing Valid and Credible Simulation Models

A simulation model is a surrogate for actually being able to experiment with a communications system. Thus, an idealized goal in building a simulation model is for it to be valid enough so that any conclusions drawn from the model would be similar to those derived from physically experimenting with the system (if this was possible). It is also important for a model to be credible. Otherwise, its results may never be used in the decision-making process, even if the model is valid. The following are some important idea techniques for deciding the appropriate level of model detail, for validating a simulation model, and for developing a model with high credibility:

1. State definitively the issues to be addressed and the performance measures for evaluation at the beginning of the study.
2. Collect information on the network topology and protocols based on conversations with all-important people associated with the system.

3. Delineate all information and data summaries in an "assumptions document."
4. Interact with the manager on a regular basis throughout the study.
5. Perform a structured walk-through (before programming) of the conceptual simulation model as embodied in the assumption document before an audience of all key project personnel.
6. Use sensitivity analyses (see Law and Kelton 1991) to determine important model factors.
7. Compare performance measures (e.g., utilization) for the existing network (if there is one) to comparable performance measures for a simulation model of the existing network.

9.4 Statistical Issues in Network Simulation

Since random samples from input probability distributions (e.g., the distribution for inter departure times of messages) are used to "drive" a simulation model through time, basic simulation output data (e.g., end-to-end delays of messages) or an estimated performance measure computed from them (e.g., average end-to-end delay from the entire run) are also random. Therefore, it is important to model the random inputs to a simulation model correctly and also to design and analyze simulation experiments in a proper manner. These topics are briefly discussed in this section.

9.4.1 Modeling System Randomness

The most important source of randomness for network simulations is usually associated with message traffic. In general, one should model messages (or transactions) not packets. The messages are fragmented into packets by the network protocols employed in the simulation. Note also that messages may not be independent of each other (e.g., there are often acknowledgement messages).

The following methods of generating traffic are often used, with the first approach generally being the most statistically valid:

1. Message departure times and message sizes for a particular node are application based (file transfer, word processing, E-mail, etc.), and may (depend on the receipt of an acknowledgement message
2. Message inter departure times and message sizes for a particular node are each independent samples from respective probability distributions (usually exponential for inter arrival times)
3. Traffic data is read into the simulation model from a network analyzer

Because network components are generally quite reliable, equipment breakdowns are not typically modeled in a simulation. An exception is where one is interested in the transient response of the network, e.g., the ability of the network to reconfigure itself after a link failure. In such cases, the operational status of a component can be modeled as an "up" period of random duration followed by a "down" (or repair) period of random duration.

9.4.2 Design and Analysis of Simulation Experiments

Because of the random nature of simulation input, a simulation model produces a statistical estimate of the (true) performance measure not the measure itself. In order for a simulation estimate to be statistically precise (have a small variance) and free of bias, the analyst must specify for each network configuration appropriate choices for the following:

1. Length of each simulation run
2. Number of independent simulation runs
3. Length of the warm up period, if one is appropriate

We recommend always making at least three to five independent runs for each configuration, and using the average of the estimated performance measures from the individual runs as the overall estimate of the performance measure. (Independent runs mean using different random numbers for each run, starting each run in the same initial status, and resetting the model's statistical counters back to "zero" at the beginning of each run.) This overall estimate should be more statistically precise than the estimated performance measure from one run. Note that independent runs (as compared to one very long run) are required to obtain legitimate and simple variance estimates and confidence intervals.

When simulating certain types of communications systems, we are often interested in the long run (or steady state) behavior of the system, i.e., its behavior when operating in a normal manner. This results in the output data from the beginning of the simulation not being representative of the desired normal behavior of the system. Therefore, simulations are often run for a certain amount of time, the warm up period, before the output data is actually used to estimate the desired measure of performance. Use of this warm up period data would bias the estimated performance measure.

9.5 Simulation Analysis of a System

In the actual conference presentation, we will give a detailed simulation analysis of a wide area network. We will address the following performance issues:

1. Is the performance of the existing network satisfactory?
2. What impact will a link failure have on system performance?
3. How much traffic can be increased before the system "blows up"?
4. What impact will changing the message-size distribution and the form of the traffic have on system performance?
5. What will be the effect of faster links or processors?

9.6 Network Simulation Packages

To predict an end-to-end delays, throughputs, and utilization of links, buffers, and processors, two simulation packages have been taken into consideration in this study. The OPNET Modeler and Planner, and COMNET.

9.6.1 OPNET Planner and Modeler

OPNET can be utilized as a decision support tool to provide insight into the performance and behavior of existing or proposed networks, systems, and processes. To provide useful data, network models must combine accurate descriptions of topology, data flow, and control flow. Since no single paradigm of visual representation is ideally suited for all three of these model types, OPNET utilizes separate model format for each.

OPNET Network models define the position and interconnection of communicating entities, or nodes. Each node is described by a block structured data flow, or OPNET Node Model, which typically depicts the interrelation of processes, protocols, and subsystems. Each programmable block in a Node Model has its functionality defined by an OPNET Process Model which combines the graphical power of a state-transition diagram with the flexibility of standard programming language and a broad library of predefined modeling functions.

The OPNET Network Editor graphically captures the physical topology of a communication network. Networks consist of node and link objects, which are graphically assembled and parameterized via pop-up dialog boxes. To create node objects, users select node types from a library of example and user-defined models. Each OPNET Node Model has a specific set of attributes that are used to configure it.

Network models can be constructed in a dimensioned workspace with a user selected grid. Since users may define node abstractions, OPNET network models may represent LAN's, MAN's, WAN's, on-board vehicular networks, or any combination thereof. Sub-network objects can be used to structure an unlimited topology hierarchy within any network model.

The OPNET Node Editor graphically captures node architectures, which are diagrams of data flow between modules typically representing hardware and software subsystems. Module types include processors, queues, and traffic generators, receivers, and transmitters. Processors are general modules that provide complete flexibility in protocol and algorithm specification. The functionality of processors and queues objects is defined using OPNET Process Models.

The OPNET Process Editor uses a state-transition diagram approach to support specification of any type of protocol, resource, application, algorithm, or queuing policy. States and transitions graphically define the progression of a process in response to simulated events. Within each state, general logic can be specified using a library of over 350 pre-defined functions. The full flexibility of the C programming language is also accessible. As with other OPNET editors, users can construct entirely new process models, or modify those that are provided.

Once a set of OPNET Network, Node, and Process Models are fully defined, users can run simulation studies based via the OPNET Animation Viewer, and plot statistical performance measurements based on simulation studies of the OPNET Analysis Tool.

The OPNET Analysis Tool provides a graphical environment that allows users to view and manipulate data collected during simulation runs. Standard and user-specified probes can be inserted at any point in a model to collect statistics. Simulation output collected by probes can be displayed graphically, viewed numerically, or exported to other software packages. First and second order statistics in each traces and automatically calculates confidence intervals. OPNET supports the display of data traces as time series plots, histograms, probability density and cumulative distribution functions. Graphs may be output to a printer or saved as bitmap files to be included in reports or proposals.

OPNET is a comprehensive software environment for modeling, simulating, and analyzing the performance of communication networks, computer systems and applications, and distributed systems.

9.6.2 COMNET

COMNET is used to simulate the network's detailed operation producing dynamic animation and reports depicting network performance. It predicts end to-end delays, throughout, and utilization of link, buffers, and processors reproducing random, bursty traffic patterns to view peaks and valleys of traffic and not just snapshots and average. It pinpoints sources of delays and bottlenecks.

COMNET has been chosen to simulate the intelligent ATM network because of its capability to simulate the circuit-switched traffic and the packet-switched traffic on the same platform (e.g. circuit switching is used to send voice over digital transmission.). Regardless of whether it is used with packet-switching or alone, circuit-switching brings with it a number of challenges that are distinct from packet switching.

COMNET III object oriented framework offers the possibility of creation of a large number of scenarios. COMNET answers questions about link and node utilization, message delays, and congestion points in a large network. It can answer questions about the performance of the current network topology based on historical or projected traffic loads. As in any other communication network simulation package, the network topology, workload, and protocols need to be defined.

9.6.2.1 Topology

The topology is defined by switches (nodes) that are interconnected by transmission facilities (links).

Nodes

COMNET has three types of nodes: Processing Nodes, Router Nodes, and Switch Nodes. Each node type can switch packets or calls. Processing nodes model computer hosts as well as communication processing devices. Each processing node has an internal processor that executes software applications and processes packets. A processor node also has a disk storage device to simulate file reads and writes. A single Group Node can represent a number of identical computers. At simulation time, COMNET III expands the group node into the specified number of Processor nodes, each with its own independent traffic sources based on the initial group node. When the model schedules an application for execution at a processor node, the application waits in the node's pending queue until the internal processor is available. When it becomes available, the processor runs the application by executing in sequence each command requested by the application. The commands are defined in the node's Command Repertoire. There are commands for processing data, sending and answering messages, reading and writing files, establishing sessions, and suspending an application until a required message arrives. Router Nodes model routers, bridges, switches, hubs, and other devices that have shared internal back plane that moves packets between port buffers. A bus and a bus count characterize the back plane. In addition to the internal back plane, the router node has all of the capabilities of the processor node, so it can, for example, run applications that update routing table and distribute routing table information. Ports on a Router node can share a line card, which allows packets to move from an input port to an output port on the same line card without transiting a back plane bus. Non-blocking switch fabrics may also be modeled with Router node by specifying a bus count that is at least equal to the number of line cards on the node.

In the case of circuit switched networking, different routing algorithms are available for routing a call to its destination. Routing modeling capabilities include:

- The backbone and each of its top-level subnets have their own routing algorithm and routing tables.

- Routing is based on a shortest-path routing algorithm or user-defined routing tables.
- Shortest-path routing uses a flexible distance metric for each link that can vary by routing class and by link utilization.
- User-defined routing tables can also vary by routing class.
- User-defined table routing uses a route selection criterion when more than one route is listed; criteria include dynamic alternate routing, first available, maximum idle bandwidth, random list, and round robin.
- Calls traversing a node or a link that fails can be rerouted or simply disconnected.

Switch Nodes model switches as well as routers, hubs, and other communication devices that have an insignificant delay in moving packets from input port buffers to output port buffers. Packet in an input buffer always switches to the appropriate output buffer in zero time, provided buffer space is available. The switch node models head-of-line blocking for those switches that allow packets to wait at an input buffer for the required output buffer to become available.

Links

Links are used to model a variety of different transmission media. The physical characteristics of a link are defined in terms of bandwidth and propagation delay. The link physical transmission unit in COMNET is a frame. Packets are generated or assembled into frames for transmission on a link. Each link has a frame size (minimum and maximum), a frame overhead, and a frame error rate.

COMNET III provides a number of link types to model different kinds of link behaviors, primarily distinguished by the multi-access methods used:

- Point-to-Point link is used for dedicated digital serial links between nodes. It is primary for WAN circuits.
- DAMA Link is a simplex link where each station may transmit when the channel is clear.
- Dialup link (modem) for modeling a link that must be connected before it is used. Also includes a means to model an analog connection with varying bandwidth due to noise.
- Modem pool allows for a number of stations to compete for a limited number of modems (dialup connections).
- Aloha, CSMA, CSMA/CD, CSMA/CA, CSMA/CD-CAN: random access collision channels.
- Token Ring, Priority Token Ring, FDDI, and Priority FDDI: based on a rotating token around a logical ring.
- Polling: host polls each terminal.

- TDM, FDM: multiple access where each station is assigned a time-slot or bandwidth to transmit on, but all stations can hear the broadcasts.
- TDMA, FDMA: multiple access where many stations compete for a limited number of time slots or bandwidth and once connected get a full-duplex channel to a hub.
- Virtual: transmission-less link for modeling high-speed links (such as internal buses) or just for modeling convenience
- WAN link: to model a frame-relay virtual circuit with traffic policing rather than a transmission model

Wan Clouds

Used to model frame relay, cell relay, and packet switching services. The WAN cloud contains an access link for each node connected to the cloud. An access link models the connection from a user site to a network service's point-of-presence. Each access link is a point-to-point link with some number of channels, a channel transmission rate, and a propagation delay. The cloud has a delay attribute hat determine the time required for a frame to transit the cloud from one access link to another.

Subnets

The subnet object is used for modeling a topology hierarchically so that separate subnets may have a separate routing algorithm which can also be independent from the backbone.

Connections between the internal topology of the subnet and the backbone topology are through the Access Points. There may be as many access points as required. However, if the traffic will be routed into the subnet based upon the routing algorithm at the next higher level, then typically the backbone routing algorithm is used.

9.6.2.2 Workload

The workload in COMNET is defined by traffic sources representing voice and data communications over the network. In the case of circuit switching, for example, the workload is defined by call sources at each node that generates calls during the simulation. Each call originating at a particular node has a bandwidth requirement, a destination drawn from a probability-weighted list, a duration drawn from a call holding time distribution, a routing class, and other attributes. Workload modeling capabilities include:

- Arbitrary number of call sources per node.
- Circuit Switched call sources and packet switched message sources can be combined in the same model.

- Call scheduling can be based on times drawn from probability distribution, an external file of call arrival events, or a triggering event, such as delivery of a message.
- Combinations of randomly-generated and externally-scheduled calls are permitted.
- Blocked and preempted calls can try again later based on a call-retry distribution.
- High priority calls can preempt low priority calls.
- Probability distributions available for modeling call inter arrival times and holding times include the beta, erlang, exponential, gamma, hyper-exponential, log-normal, normal, pareto, triangular, uniform, and weibull.
- User-defined table distributions are available for empirical data not well-modeled by standard probability distributions.
- All probability distributions can be time-varying to model both time-of-day effects and networks spanning multiple time zones.
- When utilization levels exceed user-defined thresholds, alarms are displayed, trace options can be activated, and traffic sources can be triggered.

9.6.2.3 Traffic Sources

The application sources, call sources and their scheduling are important for the analysis of the traffic sources.

Application Sources

Generate applications that execute on Processor or Router nodes. Each application source specifies a sequence of commands. When an application begins execution on the processor of the node connected to the application's source, the node runs each command in the sequence. If the command is in the node's command repertoire, the node runs the local command, otherwise, it runs the command in the model's global command.

Call Sources

Generate circuit-switched calls. When a call originates, the model attempts to establish a dedicated, end-to-end circuit of the specified bandwidth. In order for a call to use a path, each node and link along the path must have enough idle bandwidth to satisfy the call's bandwidth requirements.

Scheduling

Traffic sources generate data using time-based scheduling or received-message scheduling. Call sources originate calls using time-based scheduling. With time based scheduling, a source data (or calls) periodically, with the interval between applications given by an inter-arrival time distribution.

9.6.2.4 Protocols

Physical and data link functions of the OSI reference model, including medium access control are performed in COMNET III by links. Network layer functions are performed in COMNET III at nodes, which make routing decisions for packets. Each packet has a routed protocol that can be used as one of the factors to determine both packet's port processing time and node processing time. The node's routing decisions are based on a routing algorithm.

The backbone and each of the subnets have separate and independent routing algorithms. The routing algorithms in COMNET III are based on a shortest path computation that automatically initializes and updates routing tables. The various shortest path algorithms differ in the metrics used to compute shortest path and in how the routing table updates are scheduled. Static algorithms compute the tables only at the start of the simulation. Dynamic algorithms periodically update the tables based on dynamic measures. Both static and dynamic algorithms update the tables whenever a node or link fails. Multi-path routing, which load balances traffic among routes with metrics within some specified percentage of the best metric, can be modeled. In addition to a number of shortest path routing algorithms that automatically populate the routing tables, there is a possibility to define a new table. When selecting user-defined routing tables, route selection criteria becomes possible.

Routing decisions can also vary by a packet's routing class, which allows different types of traffic to be routed over different kinds of facilities. To model connection-oriented routing of sessions (so the traffic associated with a session follows the same path through the subnet), some of the following algorithms can be used:

- RIP Minimum Hop,
- Shortest Measured Delay,
- Link-State Shortest Path First/OSPF,
- Minimum Penalty
- User-Defined Routing Table

End-to-end or transport layer function of the OSI model is modeled in COMNET III by the transport protocol assigned to each message. Messages are produced by Transport, Answer, and Setup commands. Different commands can reference different transport protocols. The transport layer model includes parameters for specifying packet size and overhead bytes, which flow control mechanism to use, and whether or not to retransmit blocked packets. Each transport protocol has a routed protocol identifier that is one determinant of the packet processing time at multi-protocol routers and nodes.

Each message produced during a simulation is segmented into packets by the message's transport protocol. The packets are transported, subject to flow control, to the destination node, where the message is reassembled. By explicitly modeling transport layer segmentation and Reassembly of messages, COMNET III can provide measures of end-to-end delays for messages such as files, e-mail, inquiries, or responses, in addition to the end-to-end delays experienced by the packets formed from such messages. When simply packet volumes characterize workload, more meaningful measures of message-level delay are not available.

The Setup Session command model events that are typically included at the session layer of the OSI model. The effects of presentation and application layer functions in the OSI model are accounted for in a COMNET III model by the specific commands executed by various applications at the nodal processors.

COMNET III includes traffic management features for improved ATM and frame relay modeling, including:

- Burst definitions in term of ATM cell rates or frame burst sizes

- Burst option for traffic policing
- Packets spaced by a fixed time between packets (CBR) or by burst limit (VBR, ABR).

9.7 Description of the News Network

For the sake of analysis and the development of real environment, we are taking the network topology of the Associated Press and analyzing and proposing the appropriate technology for the improvement.

9.7.1 Current Topology

Figure 9.1 represents the actual top-level topology of the Associated Press network. This view was extracted from the HP Openview software. The six areas are represented as London; Asia; Washington, D. C., Kansas City, Cranbury, New Jersey and New York City, U.S.A.

Figure 9.2 represents the overview of the Associated Press Topology from the link speed perspective.

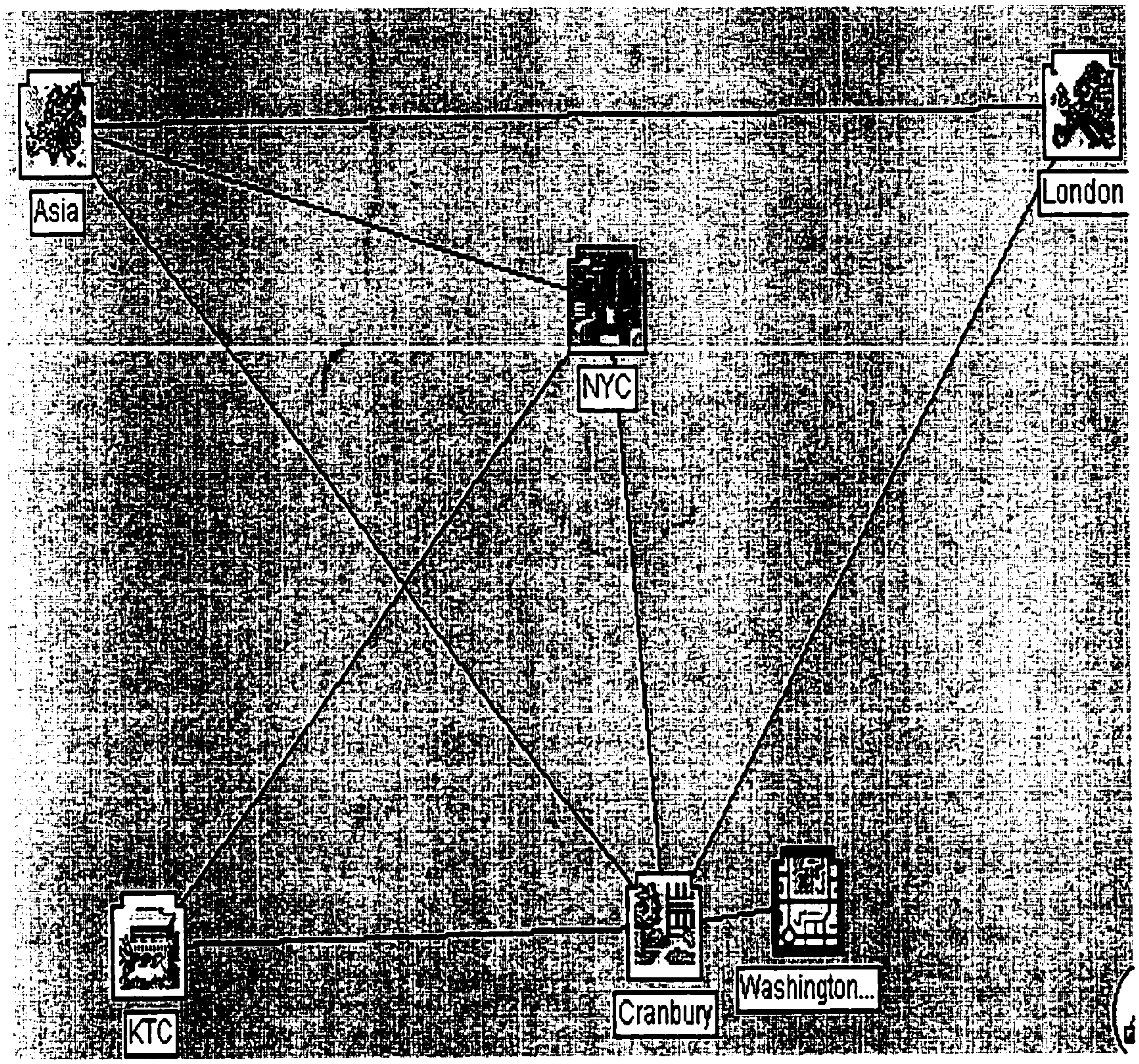


Figure 9.1. Network Topology of Associated Press. An HP Open View Diagram of the Network.

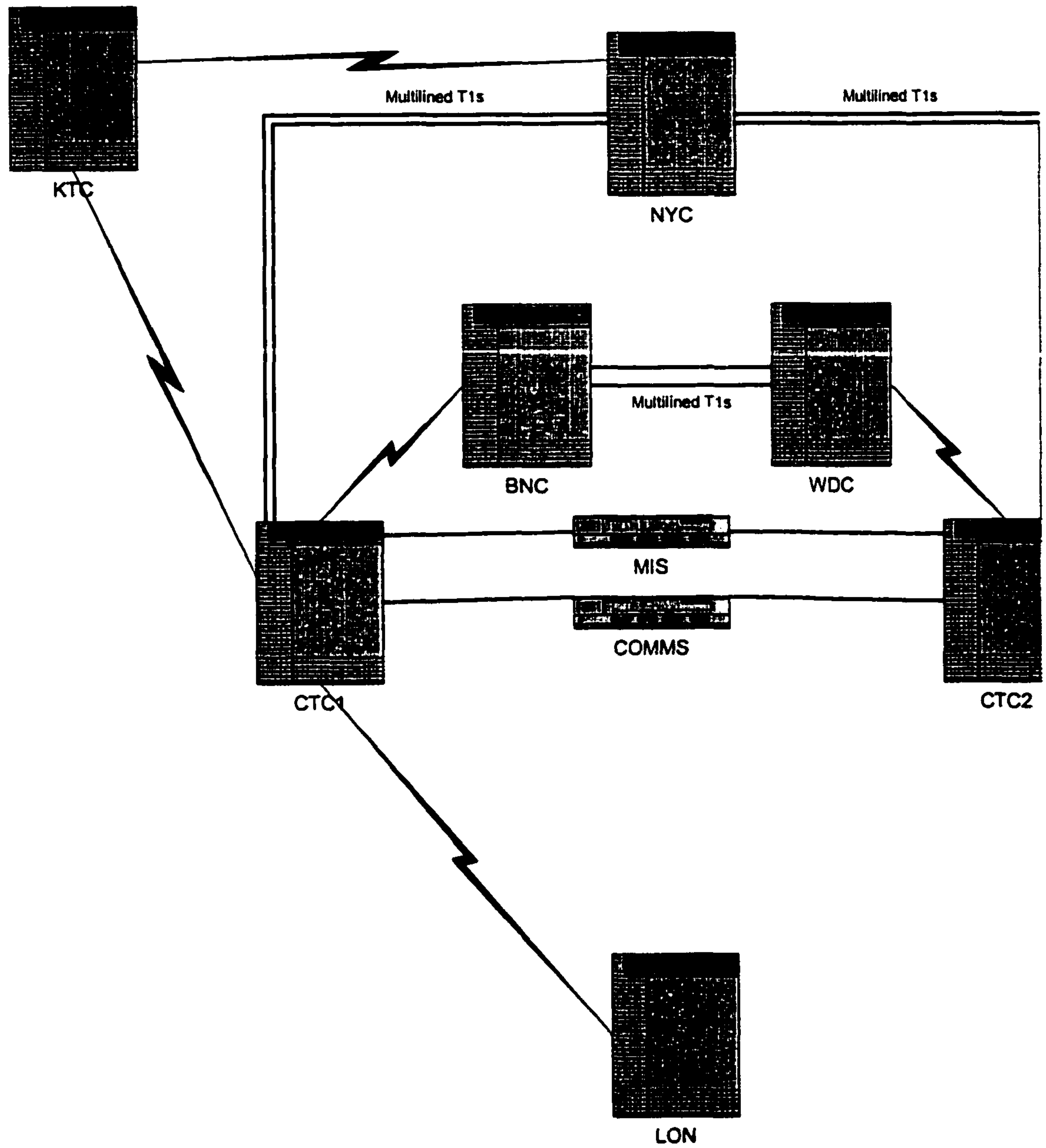
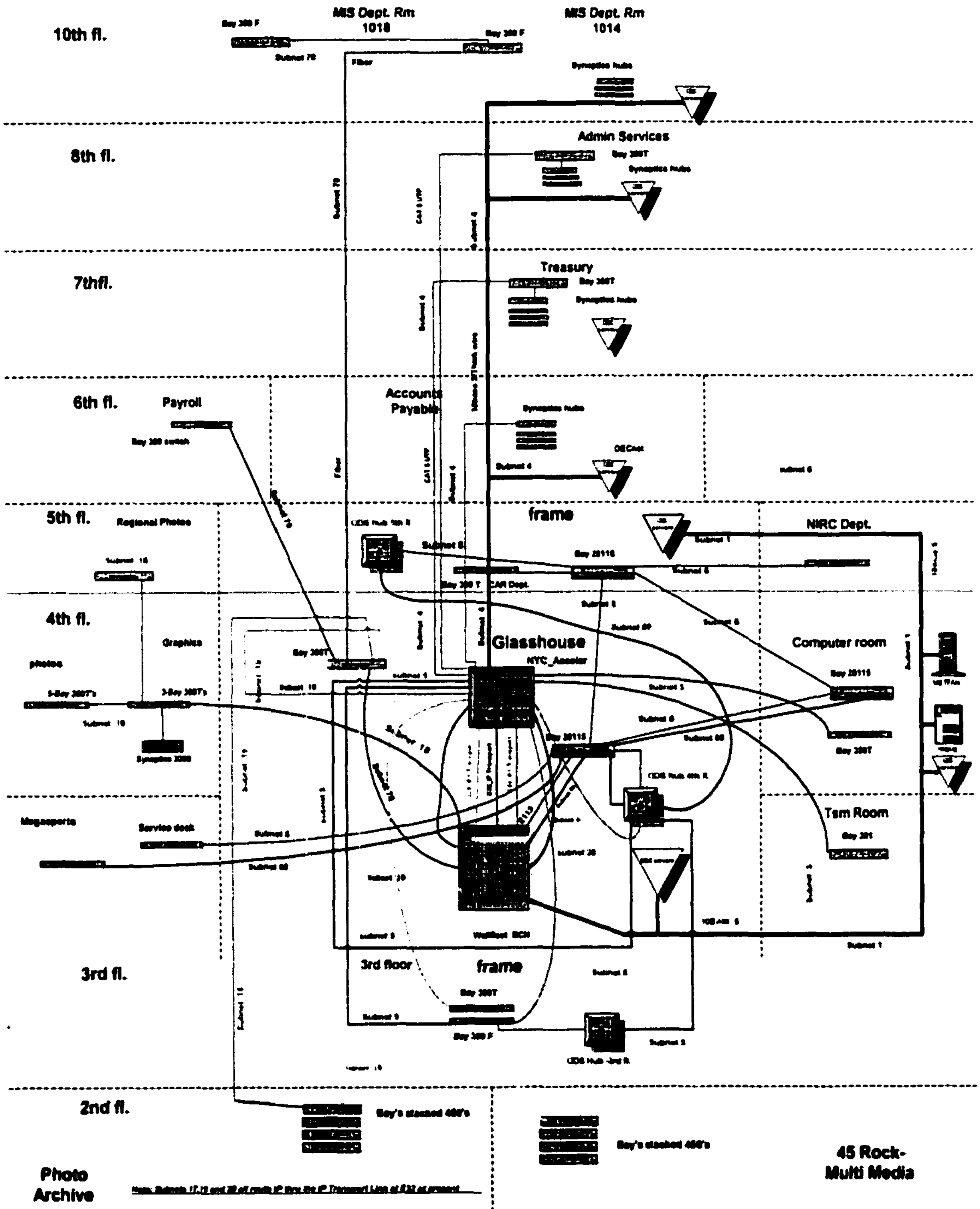


Figure 9.2. Associated Press Topology from link speed perspective

KTC	Kansas City, Missouri, USA
NYC	New York City, New York, USA
BNC	Broadcast Center, Washington, D.C., USA
WDC	Washington, D.C., USA
CTC1	Cranbury Technical Center, Cranbury, New Jersey, USA
CTC2	Cranbury Technical Center, Cranbury, New Jersey, USA
LON	London, UK

Figure 9.3. Network Topology of Associated Press at NYC



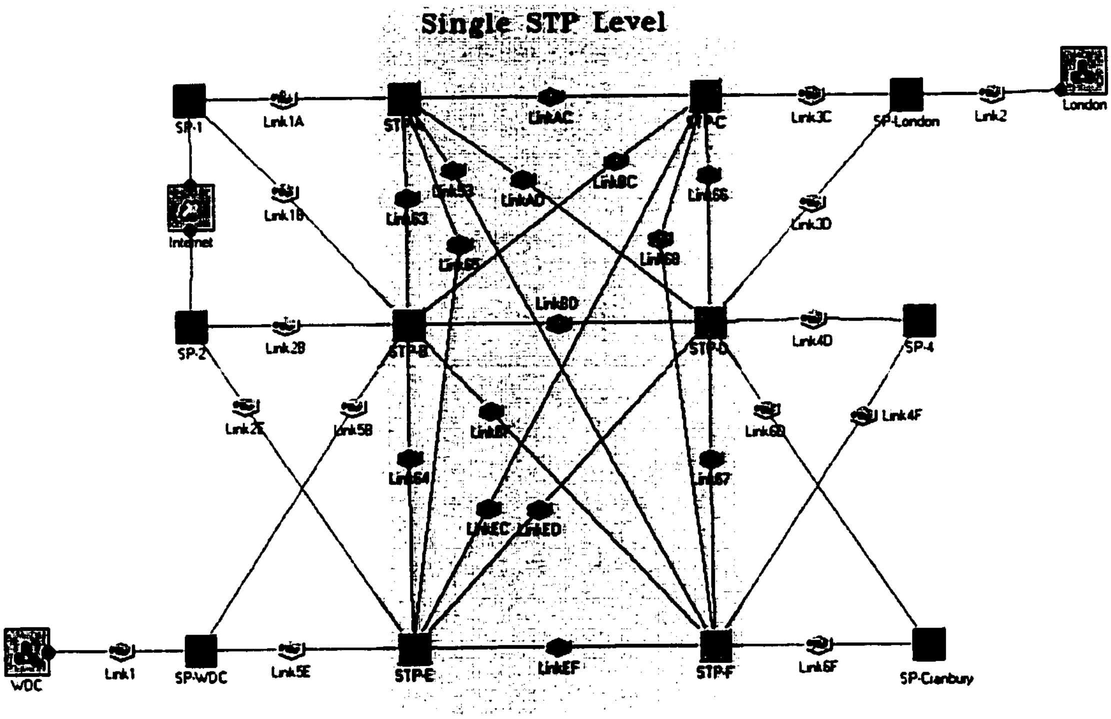


Figure 9.4. Simulation model based upon single STP level. All STPs are fully meshed.

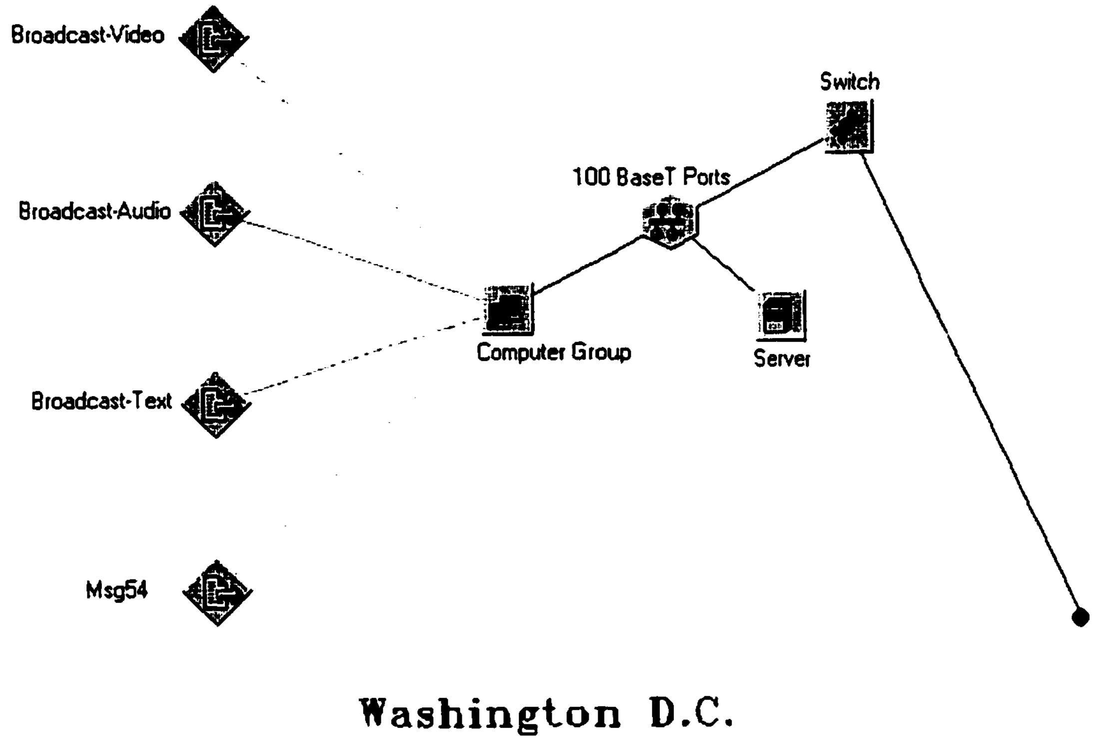


Figure 9.5. Washington, D.C. Broadcast Center LAN

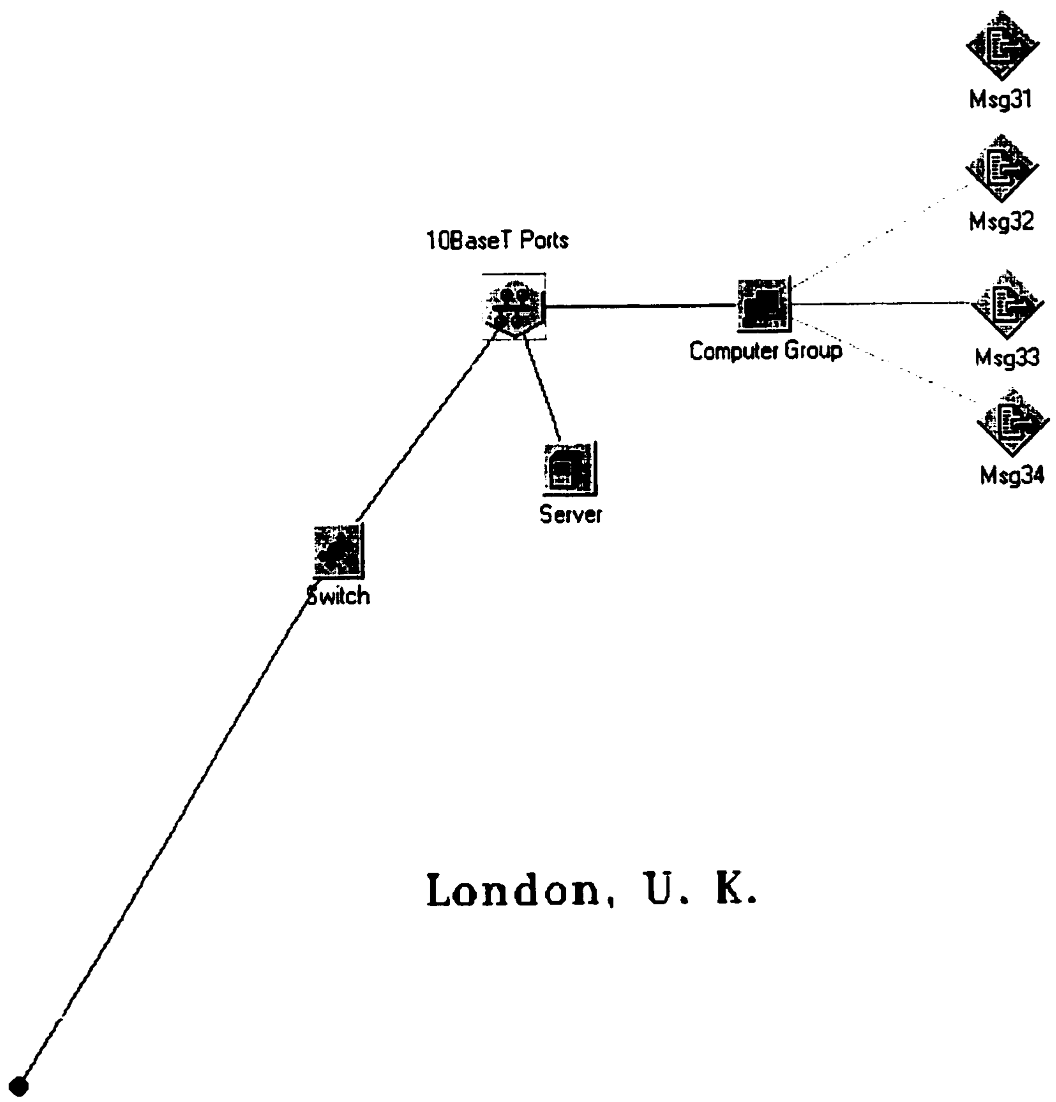


Figure 9.6. London, UK LAN

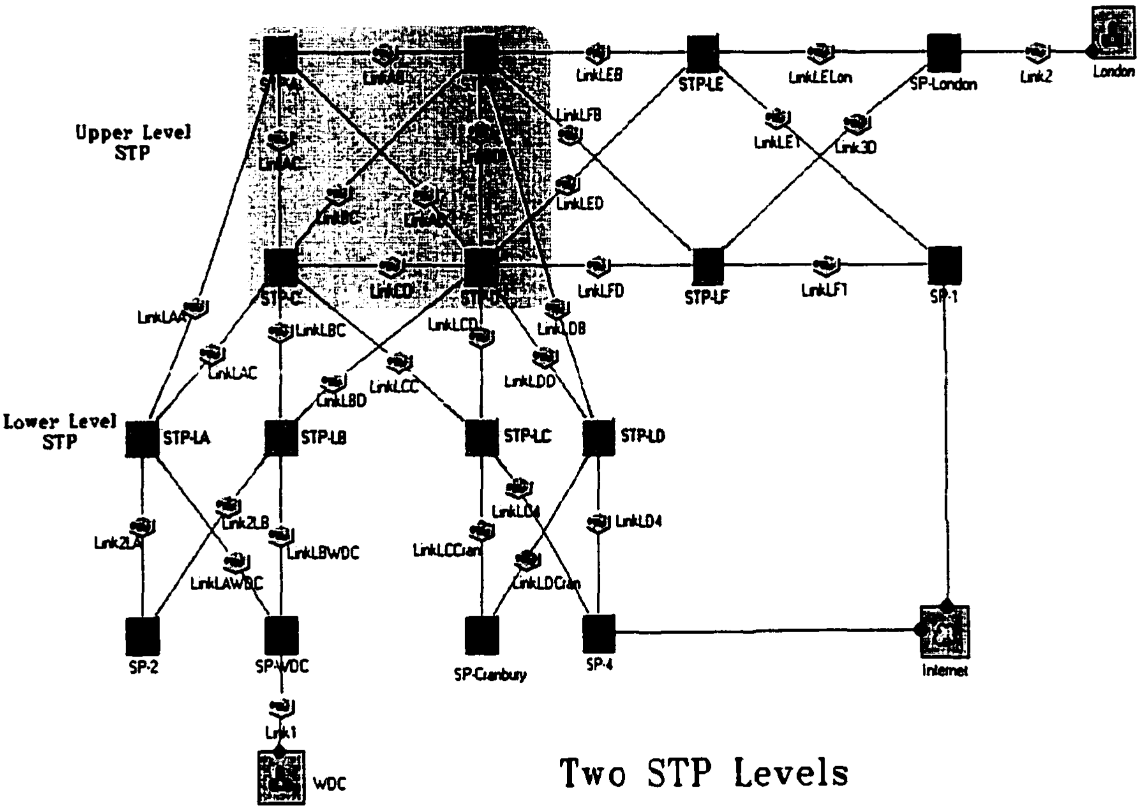


Figure 9.7. Simulation model in two STP levels. Lower level connects the regions, whereas inter-region traffic passes through higher STP level

9.8 Simulation Environment

Following was the hardware and software environment:

Hardware:

A Dell Precision 620 Workstation with dual Pentium III Processor of 860 MHz each, along with 1 GB RAM was used for the simulation purpose.

Software:

The operating system was NT 4.0 Workstation. The simulation software was Comnet III version 2.0.1, with Circuit Switching and Distributed modules loaded.

After analyzing the existing network of the Associated Press, an identical simulation model was developed in COMNET III. The Intelligent Network/2 components were introduced at several levels to further enhance the models.

There were fourteen models developed with varying characteristics to compare and view the implementation of different technologies. Those models are described as follows:

The first set of seven models was based upon the "**RIP Minimum Hop**" as the backbone characteristics of the network. Whereas, the second set of seven were based upon the **"Shortest Measured Delay."** Following was the detailed classification of the models:

Backbone Characteristics – RIP Minimum Hop

1. Single level STP where backbone protocol was 10/100 Base-T Ethernet and TCP/IP.
2. Two level STP where backbone protocol was 10/100 Base-T Ethernet and TCP/IP.
3. Single level STP where Washington, D.C. Broadcast Center LAN was ATM Base and the remainder network was 10/100 Base-T Ethernet using TCP/IP.
4. Two level STP where the Washington, D.C. Broadcast Center LAN was ATM Base and the remainder network was 10/100 Base-T Ethernet using TCP/IP.
5. Single level STP where Washington, D.C. Broadcast Center LAN was ATM Base and the remainder network was 10/100 Base-T Ethernet using TCP/IP and incorporates the IN/2 signaling characteristics.
6. Two level STP where Washington, D.C. Broadcast Center LAN was ATM-Base and the remainder network was 10/100 Base-T Ethernet, using TCP/IP and incorporates the IN/2 signaling characteristics.
7. Two level STP where entire network was based on fiber OC-12 connected through ATM switches and fully ATM compliant along with IN/2 signaling characteristics.

Backbone Characteristics – Shortest Measured Delay

8. Single level STP where backbone protocol was 10/100 Base-T Ethernet and TCP/IP.
9. Two level STP where backbone protocol was 10/100 Base-T Ethernet and TCP/IP.
10. Single level STP where Washington, D.C. Broadcast Center LAN was ATM Base and the remainder network was 10/100 Base-T Ethernet using TCP/IP.
11. Two level STP where the Washington, D.C. Broadcast Center LAN was ATM Base and the remainder network was 10/100 Base-T Ethernet using TCP/IP.
12. Single level STP where Washington, D.C. Broadcast Center LAN was ATM Base and the remainder network was 10/100 Base-T Ethernet using TCP/IP and incorporates the IN/2 signaling characteristics.
13. Two level STP where Washington, D.C. Broadcast Center LAN was ATM-Base and the remainder network is 10/100 Base-T Ethernet, using TCP/IP and incorporates the IN/2 signaling characteristics.
14. Two level STP where entire network was based on fiber OC-12 connected through ATM switches and fully ATM compliant along with IN/2 Base signaling characteristics.

9.9 Simulation Results

For the sake of analysis and comparison the three message sources are compared extensively. These are Video Broadcast, Audio Broadcast and Text Broadcast. These three sources cover the main news contents ready for TV stations, Radio Stations and the Newspapers respectively.

Video Broadcast

Description of Model	Messages Assembled	Avg Delay (Milli Sec)	Std Dev (Milli Sec)
Single Level STP, TCP/IP and Ethernet	17	7159	3718
Two Level STP, TCP/IP and Ethernet	17	8890	4660
Single Level STP, Broadcast LAN is ATM	67	4875	2441
Two Level STP, Broadcast LAN is ATM	0	0	0
Single Level STP, Broadcast LAN is based on ATM & IN/2	65	8478	4107
Two Level STP, Broadcast LAN is based on ATM & IN/2	0	0	0
Two Level STP, Fully ATM and all Fiber (OC-12) Links	206	.055	.002

Table 9.1 Message Delay from Washington, D.C. Broadcast Center to London (SP-London), where network backbone topology is "RIP Minimum Hop"

Video Broadcast (cont.)

Description of Model	Messages Assembled	Avg Delay (Milli Sec)	Std Dev (Milli Sec)
Single Level STP, TCP/IP and Ethernet	16	10790	3819
Two Level STP, TCP/IP and Ethernet	15	8155	4983
Single Level STP, Broadcast LAN is ATM	51	5254	2777
Two Level STP, Broadcast LAN is ATM	0	0	0
Single Level STP, Broadcast LAN is based on ATM & IN/2	75	7105	3999
Two Level STP, Broadcast LAN is based on ATM & IN/2	0	0	0
Two Level STP, Fully ATM and all Fiber (OC-12) Links	206	.055	.002

Table 9.2 Message Delay from Washington, D.C. Broadcast Center to London (SP-London), where network backbone topology is "Shortest Measured Delay"

Audio Broadcast

Description of Model	Messages Assembled (Number)	Avg Delay (Milli Sec)	Std Dev (Milli Sec)
Single Level STP, TCP/IP and Ethernet	98	9891	4313
Two Level STP, TCP/IP and Ethernet	99	9549	4683
Single Level STP, Broadcast LAN is ATM	405	5118	2458
Two Level STP, Broadcast LAN is ATM	41	10040	4886
Single Level STP, Broadcast LAN is based on ATM & IN/2	372	7866	3604
Two Level STP, Broadcast LAN is based on ATM & IN/2	7	11161	5413
Two Level STP, Fully ATM and all Fiber (OC-12) Links	260	61	248

Table 9.3 Message Delay from Washington, D.C. Broadcast Center to London (SP-London), where network backbone topology is "RIP Minimum Hop"

Audio Broadcast (cont.)

Description of Model	Messages Assembled	Avg Delay (Milli Sec)	Std Dev (Milli Sec)
Single Level STP, TCP/IP and Ethernet	68	9453	3902
Two Level STP, TCP/IP and Ethernet	73	9867	4311
Single Level STP, Broadcast LAN is ATM	363	5293	2541
Two Level STP, Broadcast LAN is ATM	10	12103	5093
Single Level STP, Broadcast LAN is based on ATM & IN/2	465	7478	4107
Two Level STP, Broadcast LAN is based on ATM & IN/2	8	11203	5504
Two Level STP, Fully ATM and all Fiber (OC-12) Links	260	61	248

Table 9.4 Message Delay from Washington, D.C. Broadcast Center to London (SP-London), where network backbone is "Shortest Measured Delay"

Text Broadcast

Description of Model	Messages Assembled	Avg Delay (Milli Sec)	Std Dev (Milli Sec)
Single Level STP, TCP/IP and Ethernet	138	9044	4402
Two Level STP, TCP/IP and Ethernet	135	10115	4444
Single Level STP, Broadcast LAN is ATM	565	4030	2096
Two Level STP, Broadcast LAN is ATM	200	8194	4166
Single Level STP, Broadcast LAN is based on ATM & IN/2	636	7962	3640
Two Level STP, Broadcast LAN is based on ATM & IN/2	93	8974	4895
Two Level STP, Fully ATM and all Fiber (OC-12) Links	958	26	170

Table 9.5 Message Delay from Washington, D.C. Broadcast Center to London (SP-London), where network backbone topology is "RIP Minimum Hop"

Text Broadcast (cont.)

Description of Model	Messages Assembled	Avg Delay (Milli Sec)	Std Dev (Milli Sec)
Single Level STP, TCP/IP and Ethernet	92	9533	3987
Two Level STP, TCP/IP and Ethernet	101	9520	4733
Single Level STP, Broadcast LAN is ATM	459	5581	3074
Two Level STP, Broadcast LAN is ATM	228	8932	4145
Single Level STP, Broadcast LAN is based on ATM & IN/2	664	6946	3744
Two Level STP, Broadcast LAN is based on ATM & IN/2	100	9536	4225
Two Level STP, Fully ATM and all Fiber (OC-12) Links	958	26	170

Table 9.6 Message Delay from Washington, D.C. Broadcast Center to London (SP-London), where network backbone is "Shortest Measured Delay"

Chapter 10

CONCLUSIONS

The major conclusions of this dissertation are divided into two parts. The first part describes the broad implications whereas the second part lists the specific findings.

10.1 Broad Implications

The three emerging technologies, transmission, switching and databases unified into one composite science of network architecture are modifying the course of IN (Intelligent Network) evolution, especially in the case of IMNN (Intelligent Multimedia News Networks). A widely and instantaneously available news services will have significant impact in the global lives and enhance our ability to transact business more efficiently. The political and military situations will be more easily disseminated. We expect that this IT platform will be significantly different from the typical IT platform of a corporation. In fact, we see a cooperative approach not only between the members who have subscribed for the services from an individual news organization like Associated Press, Reuters, AP or UPI, rather a more broad collective effort to gather and distribute the news to all the outlets in the world.

This work may evolve as a catalyst to unify all the independent news organizations, which can collectively contribute to this news repository and also benefit from it. It will provide a fast and economical coverage to all the news organizations with the best of collective features.

We also propose to have a repository of the global news, categorized similar to knowledge, like the Dewey Decimal System or Library of Congress Cataloging System along with date and time stamp so that it can provide the news on demand anywhere, any time, on any topic based upon any time in the history. This is technologically feasible today with the availability of high-speed networks, large, fast and economical storage, efficient relational databases, high performance search engines and html, xml and java based web technology.

This news repository will be a good source for all the researchers who are writing the news features by using and supplementing their articles with the historical facts from it. This collective news knowledge base may not be limited to the limited subscribers of a news company, rather open to the members of entire participating media giants, like Associated Press, Reuters, AFP and UPI.

10.2 Specific Findings

After carefully observing the simulation results collected from the fourteen models several conclusions can be drawn. However, a few worth mentioning findings are as follows:

Single Level vs. Two Level STP Model with TCP/IP backbone:

The backbone protocol is TCP/IP and the packet routing protocol is based upon the "Shortest Measured Delay". In the single level STP implementation there are 68 Audio Broadcast messages assembled from Washington D.C. Computer Group to SP-London during the simulation period and the message delay standard deviation is 3902 m sec. whereas, in the two level STP implementation, there are 73 messages assembled with the standard deviation of 4311 (Table 9.4). The results are consistent since inter-region the messages have to travel from one region to the other. Since the single level model is fully meshed, the intra region traffic in the two levels STP implementation does not outperform.

There are consistent results for the Video Broadcast as well as Text Broadcast messages in these two models.

Inter-Region Observation

In the single level STP environment there is 9453 m sec average delay, whereas in the two level STP environment, it is 9867 m sec (Table 9.4). This is consistent with respect to the overhead required.

Intra-Region Observation

In case of the intra region simulation the results are optimal since in the two level STP model the higher-level STPs are not involved. Hence reducing any overheads.

Single Level vs. Two Level STP Model with ATM Base Broadcast Center and 10/100 Base-T Ethernet using TCP/IP based remainder network:

The results are consistent and similar to those observed in the fully TCP/IP based backbone. The most striking observation is that delivery system works very efficiently within the Washington Broadcast LAN. It is due to the fact that there may not be any packet conversion required. However, in the case of inter-region environment the ATM packets need to be converted thus causing some packet delay. For optimal results, an entire ATM-Base network may be desirable.

Text Broadcast:

In case of two level STP for the above messages, the simulation results are consistent, however due to the packet conversion required the single level STP outperforms the two level STP.

There are two important factors in this comparison:

- Inter packet conversion while routing through various packet formats at different nodes. Hence packet compatibility and portability is the main concern.
- Additional overhead required for the connectivity of the two level STPs.

Single Level vs. Two Level STP

While reviewing the simulation results, we have observed that the messages received are consistent with the criteria of the network. That means extensive use of signaling and comparatively less overhead requirement in the single level STP makes it very efficient during the observation. Whereas in the two level STP environment the signaling is more extensive causing extra work.

Complete ATM Base Setup

There has been tremendous improvement in the throughput when the entire network is completely ATM Base. This implies significant investment in the network architecture however, achieving multifold improvement in the performance and high availability of the bandwidth. The improved performance is highly desirable to accommodate the proposed repository of global news and its delivery, the provision of search engines and the delivery of their results instantaneously.

GLOSSARY

A

AAL	ATM Adaptation Layer.
ACD	Automatic Call Distributor.
ACP	Action Point.
ADM	Asymmetric Digital Subscriber.
ADSL	Asymmetric Digital Subscriber Line.
ADSL/HDSL	Asymmetric Digital Subscriber Line/High-Speed Digital Subscriber Line.
AIN	Advanced Intelligent Network.
AL	Application Layer.
ALE	ATM Local Exchange.
ALI	Automatic Location Identification.
AM	Administrative Module; also Amplitude Modulation.
AMPS	Advanced Mobile Phone Service.
ANI	Automatic Number Identifier.
ANSI	American National Standards Institute.
AP	ATM Point.
ARPANET	Advanced Research Project Agency Network.
ASE	Application Service Elements.
ASN	Adjunct Service Node.
ASP	Adjunct Service Point.
ASP	Advance Services Platform.
ASP	ATM Signaling Point.
ASP	Adjuct Service Processor.
Asynchronous	Signals that are sourced from independent clocks. These signals generally have no relation to each other and have different frequencies and phase relationships.
AT	Access Tandem.
AT&T	American Telephone and Telegraph.
ATC	ATM Transit Exchange.
ATM	Asynchronous Transfer Mode. A form of LAN data transmission based on fixed-length packets, called cells, that can carry data, voice, and video at high speeds.
ATM endpoint	The point in an ATM network where an ATM connection is either terminated or initiated. This Includes ATM-attached workstations, ATM-attached servers, ATM-to-Legacy LAN bridges or routers, etc.
ATM Layer	The layer of the ATM Protocol that relays cells from one connection to another. The ATM Layer operates differently in endpoints and switches.

AUC	Authentication Center.
AXC	ATM cross-connect.
AXE	A Public telecommunications Switching System built by Ericsson.

B

B-NT	network terminator for broadband ISDN.
B-TA	terminal adapter for broadband ISDN.
B-TE	terminal equipment for broadband ISDN.
B CHANNEL	A 64-kbit per second bearer or information-carrying channel in the ISDN context.
BI Channel	the number one B channel of BRISDN service that is used to carry 64 kbps PCM voice data for telephone service.
B2 Channel	the number two B channel of BRISDN service that can be used to carry packet-switched data.
BECN	Backward Explicit Congestion Notification. Bit Error Rate.
BER	Bit Error Rate.
BHCA	Busy Hour Call Attempts.
BISDN	Broadband ISDN.
BRISDN	Basic Rate ISDN.
BS	Base Station.
BVA	Billing Validation Application.
BW	Bandwidth.
BX.25	a specialized Bellcore protocol to communicate service management information and provides an interface for maintenance, security, and operations of the SCP.

C

C-Link	the communications link for connecting STPs in the same region with specially connected wire pairs.
CAD	Computer Aided Design.
CATV	Cable Television.
CBR	Constant Bit Rate.
CCIS	Common Channel Interoffice Signaling System.
CCIS6	Common Channel Interoffice Signaling 6.
CCIS	Common Channel Interoffice Signaling.
CCITT	International Telegraph and Telephone Consultative Committee, now called ITU-T.
CCS	Common Channel Signaling.
CCS7	Common Channel Signaling 7.

CCSS7	Common Channel Signaling System 7.
CD-ROM	Compact Disk Read Only Memory.
Cell	The fixed-length transmission unit used by ATM. Each cell is 53 bytes long with a 5-byte header containing its connection identifier and a 48byte payload.
Centrex	a widespread telephone-company switching service that uses Central Office switching equipment and to which customers connect via individual extension access lines.
Channel	a transmission path between two points.
ChannelBank	Channel terminal equipment used for combining (multiplexing) channels on a frequency-division or time-division basis.
Circuit-mode	Type of switching that causes a one-to-one correspondence between a call and a circuit.
CLASS	Custom Local Area Signaling Services.
CLP	Cell Loss Priority. A one-bit descriptor found in ATM cell headers, indicating the relative importance of a cell. If set to 0, the cell should not be discarded. If set to 1, the cell may be discarded if necessary. The cell is set by the AAL.
CM	Control Memory; also, Communication Module.
CMAC	Customer Mobile Access Control, in UMTS.
CO	Central Office; a telephone company switching facility or center, at which a diversity of transmission channels, including subscribers' local loops terminate.
COS	Corporation for Open Systems.
CP	Control Point.
CPCS	Common-Part Convergence Sublayer.
CPE	Customer Premises Equipment.
CPI	Customer Premises Interface (with ISDN).
CPM	Central Processing Module.
CPU	Central Processing Unit.
CS	Convergence Sublayer.
CSDC	Circuit Switched Digital Capability.
CSMA/CD	Carrier Sense Multiple Access/Collision Detect.
CSPDN	Circuit Switched Public Data Network.
CT2	Second-generation Cordless Telephone system.
CT3	Third-generation Cordless Telephone system.
CTI	Centre de Transit Intemationale.
CTN	Centre de Transit nationals.
CTU	Centre de Transit Urbaine.

D

D Channel	The Delta Channel; in ISDN, a 16 kbps signaling channel for basic rate access, or a 64 kbps signaling channel with other access rates.
DCE	Data Circuit-terminating Equipment.
DECnet	Digital Electronics Corporation's Network.
DLCI	Data Link Connection Identifier.
DLL	Data Link Layer.
Domain	A subset of a larger network made up of endpoints and network devices. VLANs are a type of autonomous domain.
DQDB	Dual Queue Dual Bus (IEEE 802.6).
DSO	universal 64kbps channel (or rate); same in the CEPT and Japanese digital hierarchy.
DS1	Digital Signal 1, a formatted signal transmitted at 1.544 Mbps.
DS3	DS3 Carrier Systems; the standard third level digital carrier system used in the united states at 44-736 Mbps.
DS4	fourth level digital carrier system used in the North America at 274.176 Mbps; also known as T4 signal (Japanese equivalency at 397.2 Mbps).
DSDC	Direct Services Dialing Capabilities.
DSL	Digital Subscriber Line.
DSU/CSU	Data Service Unit/Channel Service Unit.
DTE	Data Terminal Equipment.
DTMF	Dual Tone MultiFrequency.
DXI	Data Exchange Interface. The local interface between a packet-based router and an ATM capable DSU. One of the ATM specifications.

E

E1	European transmission link with 2.048 Mbps of bandwidth capacity.
EDSL	Extended Digital Subscriber Lines.
EIA	Electronic Industries Association.
EIA/TIA	Electronic Industries Association/Telecommunication Industries Association.
EIR	Electronic Identifier Register.
EPD	Early Packet Discard.
ESS	Electronic Switching System.

ET	Exchange Termination Equipment.

F

FCC	Federal Communications Commission.
FCS	Frame Checking Sequence field.
FCs	Functional Components.
FDDI	Fiber Distributed Data Interface.
FDM	Frequency Division Multiplexing.
FECN	Forward Explicit Congestion Notification.
FEP	Front End Processor.
FIFO	First In First Out.
FM	Frequency Modulation.
FO	Fiber Optic.
FPD	First Packet Discard.
FR	Frame Relay.

G

G.C.	Global Control.
Gbps	Gigabits Per Second.
GFC	Generic Flow Control. The first 4 bits of the ATM UNI cell header; used when passing through the User-Network-Interface.
GHZ	GigaHertz.
GSM	Group Special Mobile; also, Global System for Mobile Communication.

H

H0-channel	in ISDN, a 384-kilobits per second information carrying channel (5B+D service; also know as C6).
H1-channel	in ISDN, a 1.536 Mbps information-carrying channel (23B+D service, also known as C24).
HDLC	High-Level Data Link Control.
HDSL	High-Speed Digital Subscriber Line.
HDTV	High Definition Television.
HEC	Header Error Control. The HEC field is an 8-bit Cyclic Redundancy Code (CRC) computed on all fields in an ATM Header; capable of detecting single bit and certain

	multiple bit errors. HEC is used by the Physical Layer for cell delineation.
HFC	Hybrid Fiber Coax.
HL	Higher layer of the OSI model.
HLR	Home Location Register.

I

IATMN	Intelligent ATM Networks.
IDB	INWATS Data Base.
ILMI	Interim Local Management Interface. The standard specification used to manage an ATM network. The ILMI uses the SNMP protocol and an ATM UNI MIB to provide the administrator with status and configuration information.
IM&M	Information Movement and Management.
IN	Intelligent Network.
IN/1	The initial Bellcore release of the intelligent network.
IN/I+	The evolution of the IN/1 architecture and some IN/1 elements.
IN/2	The final architecture of an intelligent network.
INAP	Intelligent Network Access Point.
INWATS	Inwards Wide Area Telecommunication Service.
IOP	Input/Output Processor.
IP	Intelligent Peripheral.
IP Address	An identifier for a node; expressed as four fields separated by decimal points. Example: 136.19.0.5. The IP address is site-dependent and is assigned by an administrator.
ISDN	Integrated Services Digital Network.
ISO	International Organization for Standards.
Isochronous	Signals which are dependent on some uniform timing or carry their own timing information embedded as part of the signal.
ISUP	ISDN User-Part.
ITU-T	International Telecommunications Union-Telecommunications Standard Sector; formerly known as CCITT.
IXC	Inter-exchange Carriers.

K

KB	Knowledge Base; also, Knowledge Bank.
KCP	Knowledge Control Point.
KD	Knowledge Domain.
KLI	Knowledge Logic Interpreter
KM	Knowledge Module; also, Knowledge Machine.
KMS	Knowledge Maintenance System or Knowledge Management System.
Kopcode	Knowledge Domain Opcode.
KQP	Knowledge Query Point.
KQR	Knowledge Query Ring.
KR	Knowledge Ring.
KTP	Knowledge Transfer Point.

L

L.C.	Local Control.
LAN	Local Area Network.
LAP	Link Access Procedure/Protocol.
LAPB	Link Access Procedures-balanced.
LAPD	Link Access Protocol for the D channel.
LASS	Local Area Signaling Service.
LATA	Local Access and Transport Area.
LEC	Local Exchange Carriers.
LIDB	Line Information Database.
LT	Line-Termination.
LTE	Line Terminating Equipment.

M

MAC	Media Access Control.
MAN	Metropolitan Area Network.
MBES	Machine Based Educational Network.
MBG	Multi-frequency Business Group.
MCPN	Mobile Customer Premises Network.
MF	Multi-Frequency.
MHZ	Million Hertz.
MPEG	Motion Picture Experts Group.
MPU	Medical Processor Unit.
MS	Mobile Station.
MSC	Mobile Switching Center.

MTP	Message Transfer Part.
MTS	Mobile Telephone Service.
MTSO	Mobile Telephone Switching Office.
Multicast	Similar to a broadcast, except that the receipts of a multicast message represent only a subset in the same broadcast domain. Multicast messages are not the same as a Point-to-Multipoint connection and, instead, should be supported by Multipoint-to-Multipoint connections.
Multipoint-to-Multipoint	A connection based on a full mesh of Point-to-Multipoint VCCs or VPCs between all the associated endpoints. In a Multipoint-to-Multipoint connection, all the endpoints are roots and can send cells to all the other endpoints.

N

N-ISDN	Narrow Band ISDN.
NAP	Network Access Point.
NCP	Network Control Point.
NCP	Network Service Part.
NCTE	Network Channel Termination Equipment.
NE	Near End.
NID	Network Information Database.
NL	Network Layer.
NM	Node Manager.
NNI	Network-to-Network Interface.
No. 1 ESS	the number one ESS switching System.
NRM	Network/News Resource Manager.
NSCX	Network Service Complex.
NSP	Network Service Control Point Corresponding to a SCP.
NT	Network Termination.
NT1	Network Termination 1, which typically consists of the network channel termination equipment and is located between the U and T reference points at customer premises.
NT2	Network Termination 2, which typically consists of PBX and/or terminal controller following the T interface.
NYNEX	New York telephone operating company (now merged with Bell Atlantic).

O

OA&M	Operations, Administration, and Maintenance.

OAM&P	Operations, Administration, Maintenance, and Provisioning.
OC	Optical Carrier.
OC-n	Optical Carrier nth level of multiplex.
OE&M	Operations, Engineering, and Maintenance.
OH	Overhead.
ONU	Off Network Unit.
OSI	Open System Interconnection.
OSI-RM	Open System Interconnection- Reference Model.
OSN	Operations System Network.

P

PA	Physical Access.
PABX	Private Automatic Branch Exchange.
PBX	Private Branch Exchange.
PCM	Pulse Code Modulation.
PCN	Personal Communication Network.
PCs	Personal Communication Services.
PHY	Physical layer Protocol.
PLCP	Physical Layer Convergence Protocol.
POH	Payload Overhead.
POTS	Plain Old Telephone.
PSN	Public Switched Network.
PSPDN	Packet Switched Public Data Network.
PSTN	Public Switched Telephone Network.
PT	Payload Type.
PVN	Private Virtual Network.

Q

Q.700	A standard Series of protocol definitions published by ITU-T.
Q.900	A standard Series of protocol definitions published by ITU-T.
QoS	Quality of Service.

S

S/N	Signal to Noise.
S-AAL	Signaling.

SAP	Service Action Point.
SAR	Segmentation and Reassembly.
SCCP	Signaling Connection Control Part.
SCE	Service Creation Environment.
SCN	Service Circuit Nodes.
SCP	Service Control Point.
SDDN	Software Defined Data Network.
SDH	Synchronous Digital Hierarchy.
SDLC	Synchronous Data Link Control.
SDN	Software Defined Network.
SF	Single Frequency.
SIP	SMDS Interface Protocol
SL	Session Layer.
SLC	Subscriber Loop Carrier.
SLEE	Service Logic Execution Environment
SLI	Service Logic Interpreter.
SM	Switching Module.
SMDS	Switched Multimegabit Data Service.
SMS	Service Management System.
SMS/2	Service management System/2.
SN	Service Node.
SNA	Systems Network Architecture (from IBM).
SNA/ATM	Systems Network Architecture/ATM Cell Relay.
SNA/FR	Systems Network Architecture /Frame Relay.
SONET	Synchronous Optical Network.
SPC	Stored Program Control.
SPE	Synchronous Payload Envelop.
SPM	Subscriber Processing Module.
SRAM	Stack Random Access Memory.
SS7	Signaling System 7.
SSCS	Service-Specific Convergence Sublayer.
SSP	Service Switching Point.
SSP/2	Service Switching Point/2.
STM	Synchronous Transport Module.
STM-1	Synchronous Transport Module level I (I=N/3=1, 4, 16, 64).
STP	Signaling Transfer Point.
STS	Synchronous Transport Signal /System.
STS-N	Synchronous Transport Signal level N (N= 1, 3, 12, 48, or 192).
SVC	Switched Virtual Circuit. A logical (not physical) connection between endpoints established by the ATM network on demand after receiving a connection request

	from the source/root, which it transmits using the Q.2931 signaling protocol.
Synchronous	Signals that are sourced from the same timing reference; and have the same frequency.

T

T1	A digital carrier used to transmit a DS-a formatted digital signal at 1.544 Mbps, also called TI carrier system.
T1 or E1	Primary rates for North American or European environment (see DS-n levels).
TA	Terminal Adapter.
TDM	Time Division Multiplex.
TDMA	Time Division Multiplex Access.
TE	Terminal Equipment.
TE1	Terminal Equipment specifically meeting ISDN equipments.
TE2	Terminal equipment type 2, not specifically meeting ISDN requirements, which is typically connected to the network at the R interface reference point separated from the S reference point TA.
TIA	Telecommunication Industry Association.
TOH	Transport Overhead.
TPM	Trunk Processing Module.
Traffic Policing	A mechanism used to detect and discard or modify cells/traffic that violate the traffic or Quality of Service contract agreed to at connection setup. Although applicable to both public and private networks, traffic policing will most likely be used by public ATM service providers where tariff may be based on guaranteed QoS.
Traffic Shaping	A mechanism used to achieve or modify traffic characteristics in order to match a desired Quality of Service (QoS) contract.
TWP	Twisted wire-pair.

U

UISN	Universal Information Service Network.
UPC	Usage Parameter Control. The set of actions used by the network to monitor and control traffic. Its purpose is to protect network resources from both malicious and unintentional misbehavior which may adversely affect the

	QoS of already established connections by detecting violations of negotiated Quality of Service parameters and taking policing actions.
UNI	User-to-Network Interface. Generally described as any connection that directly links a user's device to an ATM network, through an ATM switch.

V

VBR	Variable Bit Rate.
VCC	Virtual Channel Connection. A concatenation of virtual channel links between two endpoints where higher layer protocols are accessed. By definition, ATM cell sequence must be preserved over a VCC.
VCI	Virtual Channel Identifier. An identifying value found in the header of each ATM cell.
VFN	Vendor Feature Node.
VHSDN	Very High Speed Digital Network.
VLAN	Virtual Local Area Network. A logical collection of member endpoints and network devices grouped together in secure, autonomous domain where no broadcast and multicast traffic can enter or leave that domain.
VLR	Visitor Location register.
VPC	Virtual Path Connection. A concatenation of virtual path links between two points in which the VCI values are either reassigned or terminated. Several VCCs may be bundled into one VPC.
VPI	Virtual Path Identifier. An identifying value found in the header of each ATM cell.
VT	Virtual Terminal.
Vsm	Virtual Switching Machine.

W

WAN	Wide Area Network.
WATM	Wireless ATM.
WIN	Wireless intelligent Network.

X

X.21	Interface between data terminal equipment (DTE) and data circuit terminating equipment (DCE) for synchronous operation.
X.25	ITU-T recommendation that specifies the interface between user data terminal equipment and packet-switching data circuit-terminating equipment (DCE).
X.28	DTE/DCE interface for start-stop mode data terminal equipment accessing the packet assembly/disassembly facility (PAD) in a public data network situated in the same country.
X.29	Procedures for the exchange of control information and user data between a packet assembly/disassembly facility (PAD) and a packet-mode DET or another PAD.
X.75	Terminal and transit call-control procedure and data transfer mechanisms on typically international circuits between packet-switched data networks.

REFERENCES

[1] Ahamed, S. V., and V. B. Lawrence. 1997. *"Design and Engineering of Intelligent Communication Systems"*, Kluwer Academic Publishers, Boston.

[2] Atkinson, R. 1993. *"Default IP MTU for use over ATM AALS,"* Internet Draft, November 16.

[3] A. Iwata, N. Mori, C. Ikeda, H. Suzuki, M. Ott. 1995. *"ATM Connection and Traffic Management Schemes for Multimedia Internetworking,"* Communications of the ACM Magazine, February.

[4] Averill M. Law, Michael G. McComas. 1999. *"EXPERTFIT: Total Support for Simulation Input Modeling"*, Proceedings of the 1999 Winter Simulation Conference, 261-266.

[5] Averill M. Law, Michael G. McComas. 1995. *"Simulations of the Communications Networks"*, Proceedings of the 1995 Winter Simulation Conference, 46-50.

[6] Averill M. Law, Michael G. McComas. 1996. *"Simulations of the Communications Networks"*, Proceedings of the 1996 Winter Simulation Conference, 73-77.

[7] Antonio Franchi, Eyal Trachtman, Louis Christodoulides, Jay Sengupta. 1999. *"Multimedia via Inmarsat,"* IEEE Multimedia, October-December 1999, Vol 6, No4, 15-19.

[8] Averill M. Law, Michael G. McComas. 1999. *"EXPERTFIT: Total Support for Simulation Input Modeling"*, Proceedings of the 1995 Winter Simulation Conference, 261-266.

[9] Anu Maria. 1997. *"Introduction to Modeling and Simulation"*, Proceedings of the 1997 Winter Simulation Conference, 7-13.

[10] Armitage, G., and Adams, K. 1993 *"Packet Reassembly during Cell Loss,"* IEEE Network, September.

[11] Ahamed, S. V., and V. B. Lawrence. 1997. *"Intelligent Broadband Multimedia Networks"*, Kluwer Academic Publishers.

[12] Bellcore, *"Advanced Intelligent Network Release 1 Baseline Architecture,"* SR-NPL-001555, Issue 1, Mar. 1990.

[13] Bellcore. 1991. *"Advanced Intelligent Network (AIN) Release 1 - Switching Systems Generic Requirements,"* TA-NV,7-001123, Issue 1, May.

[14] Bellcore. 1993. *"Advanced Intelligent Network (AIN) - 0.2 Switching Systems Generic Requirements,"* TA-NWT-001298, Issue 2, May.

[15] Bellcore. 1992. *"Advanced Intelligent Network (AIN) - Service Control Point (SCP) Generic Requirements,"* TA-NV,7-001280, Issue 1, August.

[16] Bellcore. 1991. "*Advanced Intelligent Network (AIN) Release 1 - Service Control Point (SCP) Generic Requirements,*" TA-NWT-001125, Issue 1, May.

[17] B. G. Kim, P. Wang. 1995. *"ATM Network: Goals and Challenges",* Communications of the ACM Magazine, February.

[18] Bijan Jabbari. 1992. *"Routing and Congestion Control in Common Channel Signaling System No. 7,"* Proceedings of the IEEE, Vol 80, No.4, April, 607-617.

[19] Brian W. Unger, Greg A. Lomow. 1993. *"The Telecom Framework: A Simulation Environment for the Telecommunications",* Proceedings of the Winter Simulation Conference.

[20] B. Thuraisingham. 1992. *"On Developing Multimedia Database Management Systems Using the Object-Oriented Approach",* Multimedia Review, Vol 3, No.2, Summer.

[21] CCITT, Draft Recommendations. 1992. Q.1211-Q.1219, March.

[22] C. Alexopoulos, K. Kang, W. R. Lilegdon, D. Goldsman. 1995. *"A High Fidelity ATM Traffic and Network Simulator",* Proceedings of the 1995 Winter Simulation Conference, 996-1003.

[23] CCITT SG 1. 1992. "*Draft Recommendation F.850 - Principles of UPT Version 4,*" Geneva, March 24.

[24] CCITT SG 1. 1992. "*Draft Recommendation F.851 - UPT - Service Description. Version 7,*" Geneva, March 24.

[25] Claridge, P., and Holloway, D. 1994. *"LAN Emulation Tutorial,"* The ATM Forum Contribution 94-0225, March.

[26] Cole, R. 1994. *"IP over ATM: A Framework Document," Internet Draft,* January.

[27 Crowl, S. 1993. *"ATM for Multiservice Wide Area Networks,"* Business Communications Review Supplement, February, 11-15.

[28] Daniel Minoli. 1992. *"Telecommunications Technology Handbook",* Artech House, Boston. vol. 31, no. 3, March, 30-36.

[29] Daniel Stevenson, Nathan Hillery, Greg Byrd. 1995. *"Secure Communications in ATM Networks,"* Communications of the ACM Magazine, February.

[30] David Benham. 1994. *"ATM in Local Area Networks: A Tutorial,"* Hughes LAN Systems. (April 1994).

[31] D. J. Medeiros, E. F. Watson, J. S. Carson and M. S. Manivannaan, eds. 1998. *"A Tutorial with Advances in Database Access and HLA Support",* Proceedings of the 1996 Winter Simulation Conference, 199-204

[32] Estes, G. 1993. "*ATM Networking – What's Been Agreed,*" Business Communications Review Supplement, February, 4-7.

[33] Fenby, Jonathan. 1989. *"The International News Services":* Schocken Books, New York.

[34] Feltman, C. 1993. "*ATM: the Grand Unifier,*" LAN Magazine, November, 53-70.

[35] Fujioka, M., Ikeda, Y., and Norigoe, M. 1990."*Error Control Criteria in the Message Transfer Part of CCITT Signaling System No. 7.*" IEEE Transactions on Communications, September.

[36] Glen D. Johnson. 1999. "*Networked Simulation with HLA and Modsim III*", Proceedings of the 1999 Winter Simulation Conference,1065-1070

[37] Gene Miller, Greg Baber, Mark Gilliland. "*News On-Demand for Multimedia Networks*"

[38] Grossman, D., and Hluchyl, M. 1994. "Link Metrics for P-NNI," *The ATM Forum Contribution 94-0284,* March.

[39] Handel, R. and Huber, M.N. 1991. *Integrated Broadband Networks - An Introduction to ATM-Based Networks,* Addison-Wesley.

[40] Heinanen, J. 1993. "*RFC 1483: Multiprotocol Encapsulation over ATM Adaptation Layer 5,*" IETF, July.

[41] Jeffry Jones. 1995. "*SIMOBJECT: From Rapid Prototype to Finished Model – A Breakthrough in Graphical Model Building,*" Proceedings of the 1995 Winter Simulation Conference, 541-547

[42] J. M. Charnes, D. J. Morrice, D. T. Brunner, J. J. Swain. 1996. "*A Methodology for Integrating Business Process and Information Infrastructure Models*", Proceedings of the 1996 Winter Simulation Conference, 1305-1312

[43] J. M. Charnes, D. J. Morrice, D. T. Brunner, J. J. Swain. 1996. "*Injecting Simulation and Modeling Upstream in the System Engineering Procss*", Proceedings of the 1996 Winter Simulation Conference, 1421-1425

[44] John G. Goble, Robb Mills. 1994. "*COMNET III: Object Oriented Network Performance Prediction*", ", Proceedings of the 1994 Winter Simulation Conference, 443-445.

[45] Jabbari, B. 1991. "*Common Channel Signaling System Number 7 for ISDN and Intelligent Networks.*" Proceedings of the IEEE, February.

[46] J. Duran and J. Visser. 1992. "*Intemational Standards for Intelligent Networks CCITTs Capability Set-1 Defines the Scope and Content of INs for the Near Term Without Curtailing Subsequent Growth,*" IEEE Commun. Mag., vol. 30, No. 2, 3442, February.

[47] Lane, J. 1994. "*ATM Knits Voice, Data on any Network,*" IEEE Spectrum, February, 42-45

[48] Lucent Technologies, Bell Labs Innovation. 1997. "*Proceedings of the ATM Symposium 1997.*"

[49] Lang, L.1993. "*The Convergence of ATM with Internetworking,*" Business Communications Review Supplement, February, 8-10.

[50] Laubach, M. 1994. "*RFC 1577 - Classical IP and ARP over ATM,*" IEEE, January.

[51] McQuillan, J. 1991. "*Cell Relay Switching,*" Data Communications, Vol. 20, No. 10, September, 58.

[52] McQuillan, J., host. 1993. "*Getting Started with ATM,*" Videotape, Access Media Inc.

[53] Michael K. painter, Ronald Fernandes, Natarajan Padmanaban, Richard J. Meyers. 1996. *"A Methodology for Integrating Business Process and Information Infrastructure Models", ",* Proceedings of the 1996 Winter Simulation Conference,1305-1312.

[54] Manfield, D., Millsteed, G., and Zukerman, M. 1993. *"Congestion Controls in SS7 Signaling Networks."* IEEE Communications Magazine, June.

[55] McQuillan, John. 1993. "*Why ATM?,*" Business Communications Review Supplement, February,1-3.

[56] *"Managing Feature Interactions in Telecommunications Systems,"* IEEE Communications Magazine., vol. 31, no. 8, Aug. 1993.

[57] NCLATM Consortium. 1992. "*Network Compatible ATM for Local Network Applications,*" V1.01, IETF, October 19.

[58] Newman, P. 1992. "*ATM Technology for Corporate Networks,*" IEEE Communications Magazine, April, 90-101.

[59] Partridge, C., 1992. "*RFC 1363 - A Proposed Flow Specification,*" IETF, September.

[60] Ransom, M.N. and Spears, D. 1992. "*Applications of Public Gigabit Networks,*" IEEE Network, Vol. 6, No.2, March.

[61] Rathgreb, E. 1991. "*Modeling and Performance Comparison of Policing Mechanisms for ATM Networks,* " IEEEJournal on SelectedAreas in Communications, Vol. 9, No. 3, April.

[62] Ronald Vetter. 1995. *"ATM Concepts, Architecture, and Protocols"* Communications of the ACM Magazine, February.

[63] Ramaswami, V., and Wang, J. 1993. *"Analysis of the Link Error Monitoring Protocols in the Common Channel Signaling Network."* IEEE Transactions on the Networking, February.

[64] Sevclk, P. 1993. "*Edge Routing - the Next Routing Evolution,*" Business Communications Review, November 27-3 1.

[65] Smith, G. 1993. "*Planning for Migration to ATM,*" Business Communications Review, May, 54-59.

[66] "*Telecommunications: How Many Features Can You Add?,*" IEEE Computer, vol. 26, no. 8, August 1993.

[67] S. Andradottir, K. J. Healy, D. H. Withers, B. L. Nelson. 1997. *"Evaluating the Performance of Electronic Commerce Systems",* Proceedings of the 1997 Winter Simulation Conference, 1053-1056

[68] Shahar Yarden. 1996. *"Injecting Simulation and Modeling Upstream in the System Engineering Process,"* Proceedings of the 1995 Winter Simulation Conference,1412-1425.

[69] Testi, N., Naegle, J., and Gossage, S. 1994. "*Combining ATM and SONET at the LAN,*" Business Communications Review, January, 47-50.

[70] Wayne J. Davis, Duane Setterdahl, Joseph Marco, Victor Izokaitis, Bradley Chanpaign. 1993. *"Recent Advances in the Modeling, Scheduling and Control of Flexible Automation,"* Proceedings of the 1993 Winter Simulation Conference, 143-156

[71] Wojnaroski, L. 1994. "*Baseline Text for Bandwidth Management,*" The ATM Forum Contribution, January.

[72] Xinjie Chang. 1999. *"Network Simulation with OPNET, "* Proceedings of the 1999 Winter Simulation Conference, 307-314.