

Washington University in St. Louis

## Washington University Open Scholarship

---

All Computer Science and Engineering  
Research

Computer Science and Engineering

---

Report Number: WUCS-95-08

1995-01-01

### A Survey of Network Signaling

Dakang Wu

Abstract Network signaling is the process of transferring control information among components of a communication network to establish, maintain, and release connections, and to pass the network management information. The rapid evolution in the field of telecommunications has led to the rapid evolution of network signaling. In this paper, we review the evolution of network signaling. We emphasize the concepts and protocols used in modern fast packet switching networks especially in emerging ATM networks.

Follow this and additional works at: [https://openscholarship.wustl.edu/cse\\_research](https://openscholarship.wustl.edu/cse_research)



Part of the [Computer Engineering Commons](#), and the [Computer Sciences Commons](#)

---

#### Recommended Citation

Wu, Dakang, "A Survey of Network Signaling" Report Number: WUCS-95-08 (1995). *All Computer Science and Engineering Research*.

[https://openscholarship.wustl.edu/cse\\_research/368](https://openscholarship.wustl.edu/cse_research/368)

Department of Computer Science & Engineering - Washington University in St. Louis  
Campus Box 1045 - St. Louis, MO - 63130 - ph: (314) 935-6160.

**A Survey of Network Signaling**

**Dakang Wu**

**WUCS-95-08**

**April 1995**

**Department of Computer Science  
Washington University  
Campus Box 1045  
One Brookings Drive  
St. Louis MO 63130-4899**



# A Survey of Network Signaling

WUCS-95-08

Dakang Wu

Department of Computer Science

Washington University

dw@dworkin.wustl.edu

April 20, 1995

**Abstract** Network signaling is the process of transferring control information among components of a communication network to establish, maintain, and release connections, and to pass the network management information. The rapid evolution in the field of telecommunications has led to the rapid evolution of network signaling. In this paper, we review the evolution of network signaling. We emphasize the concepts and protocols used in modern fast packet switching networks especially in emerging ATM networks.

## 1.0 Network Signaling and Major Functions

A telecommunication network can be modeled as a graph with three types of nodes and two types of links, as shown in FIGURE 1.

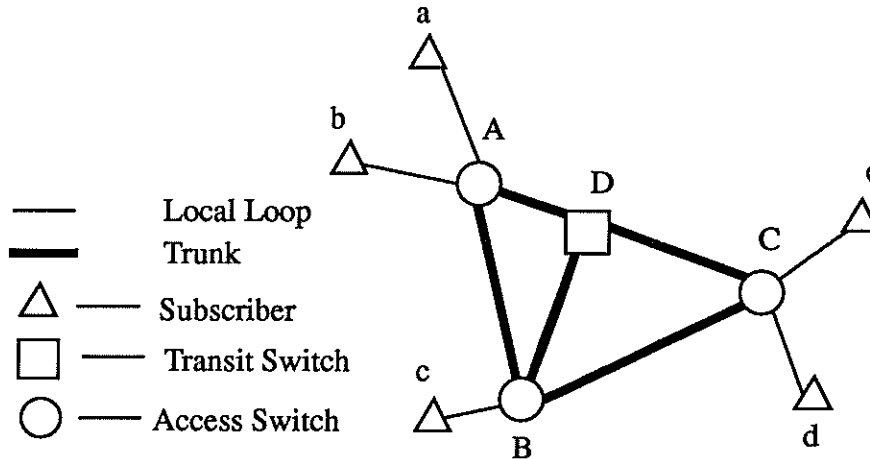


FIGURE 1. A Telecommunication Network

A triangle node is a subscriber which initiates and terminates calls. A subscriber can be a conventional telephone, a computer terminal, or any user equipment that can terminate a network connection. The circle nodes are access switches which are directly connected to some subscribers. The ones not connected to any subscribers are called transit switches. The links between an access switch and the subscribers are called local loops. The links between switches are trunks. Usually trunks have larger capacities than local loops with multiplexing techniques employed at switches to combine many subscriber channels on a single trunk.

Network Signaling can be defined as the exchange of control information among elements of a telecommunication network to initiate, maintain, and release connections and for network management [REY83]. A signaling protocol is a set of message definitions and the procedures at the processing entities involved in the signaling message processing.

Network signaling can be classified into user-network, network-node, and user-user signaling, as shown in Figure 2. The User-Network signaling concerns control messages passed between user equipment and a network node. Network-Node signaling is the control signals among network nodes. The User-User signaling is the control signals between end users.

Historically, in public networks, signaling was the process of controlling the connections between pairs of subscribers. As networks become more and more complex, the number

of functions performed by control signaling necessarily grew. The following are the most important ones [STAL89].

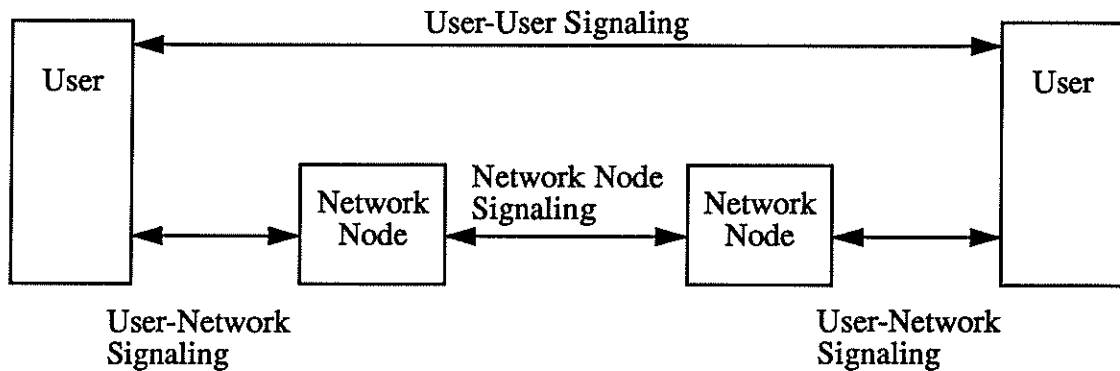


FIGURE 2. User-Network, Network-Node, and User-User signaling.

1. Signals to indicate the status of a connection, including audible signals like dial tone, ringing tone, busy signal and so on. These signals can also be translated into texts on a computer terminal or visible icons on any sort of visual display.
2. Transmission of the number dialed to switching offices that will attempt to complete a connection which can be either point-to-point or multipoint.
3. Transmission of information between switches indicating that a call can not be completed.
4. A signal to get the attention of a user, such as making a telephone ring or to change an icon on a screen.
5. Transmission of information used for billing purposes.
6. Transmission of information giving the status of equipment or trunk in the network. This information may be used for routing and maintenance purposes.
7. Transmission of information used in diagnosing and isolating system failures.
8. Control of special equipment such as satellite channel equipment.

The functions performed by signaling systems are traditionally grouped into four categories: supervisory, address, call information, and network management. Supervisory signals include the “on-hook/off-hook” signal used to request services, answer an incoming call, or disconnect a call. Address signals usually carry address numbers of the called parties. Information signals are audible tones and announcements, such as dial tone and busy signals, that convey progress in completing a call. Network management signals carry the network management information such as status, billing information etc.

## 2.0 Signaling in Circuit Switching Network

In a circuit switching network, a channel or multiple channels are dedicated to a connection. Communication via circuit switching involves following three phases.

1. **Circuit establishment.** When a call request is received from a subscriber, the network nodes work together to allocate the required resources to satisfy the request. If there are enough network resources and the called party responds positively, a connection is set up, which will be dedicated to that call for the whole duration of the call.
2. **Data transfer.** The communicating parties transfer data on the circuit established in the previous phase.
3. **Circuit disconnection.** When the users finish their communication session, they terminate the connection. The network nodes deallocate all the dedicated resources.

The signaling system plays the most important role in the circuit establishment and circuit disconnection phases. Based on the path the control signal goes through, signaling in a circuit switching network can be divided into two categories: in-channel signaling and common channel signaling.

### 2.1 In-Channel Signaling

With in-channel signaling, the same physical path and the same channel is used to carry signals as is used to carry the call to which the control signals are related. Let's take an example to show how in-channel signaling works. In the example, we assume the subscriber devices are telephones. They can be replaced by any other equipment. Figure 3 [REY83] shows the steps in a successful complete call.

The signaling sequence starts when the calling customer lifts the phone, an off-hook signal is sent to the central office directly connected to the customer. When the switch at the central office detects the off-hook signal, it sends a dial tone to the customer to indicate that it is ready to receive the address information. The numbers the customer dialed are transferred to the central office through address signals. When the address information is received by the central office, the central office tries, by calling a routing subroutine at the switch controller, to find an outgoing link towards the called party. If there is a spare channel on the link, it reserves the channel and forward the off-hook signal along the link to the next switch. When the off-hook signal is received by a switch, the switch send an off-hook followed by an on-hook signal, known as a "wink" back. Receiving this signal, the sender sends the address signal in the channel reserved. Every switch on the path does the same thing until the destination central office is reached or there is no way to forward the signal. In the second case, a busy tone is sent back to the calling customer and all the reserved resources are released. If the off-hook and address signal successfully reached the destination central office and the called subscriber is not busy, the destination central office rings the subscriber called, and at the same time, it sends a ringing tone signal to the calling party. The call set-up process ends when the called customer lifts the phone.

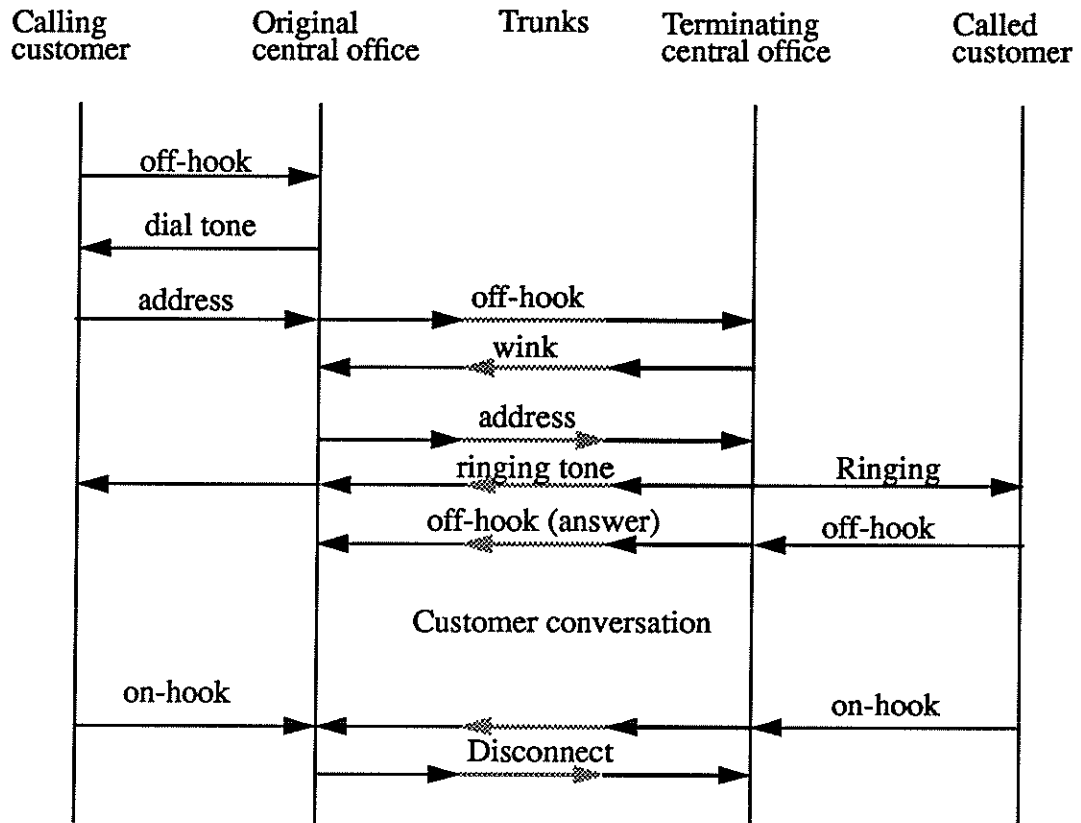


FIGURE 3. Signaling on a successful complete call

When one of the communication parties hangs up, an on-hook signal is sent to the central office. This on-hook signal will be propagated along the path. All the switches release the resources dedicated to the call when this on-hook signal is received and the call is disconnected.

There are two forms of in-channel signaling: in-band and out-of-band. With in-band signaling, the control signal uses the same frequency band as the data transmission. One can not send control signals when there is data transferring on the channel. With out-of-band signaling, a separate narrow signaling band is allocated for control signals. Since the control band is narrow, one can only send the simplest control messages.

The term in-band is overloaded. Some author uses the term to refer to the concept of in-channel signaling described above, and uses the term out-of-band to refer to common-channel signaling.

In-channel signaling works fine for simple telephone services. As more and more services have been added into the telephone network, such as 800 number service, forwarding, multiparty call, the amount of control information needed has increased rapidly. In a modern communication network, an expanding set of control signals is needed. Traditional in-



channel signaling has to be replaced by more flexible and sophisticated signaling techniques.

## 2.2 Common Channel Signaling

Common channel signaling is a technology developed for modern telephone networks [APP86] [MOD90]. With common channel signaling, control signals are carried over control paths completely independent of the data (voice) channels. The control signals are transferred directly from one control processor to another and forwarding of control information can overlap the circuit-setup process, so that the call setup time can be greatly reduced. In this section, we mainly introduce Signaling System Number 7, the common channel signaling standard to support Integrated Service Digital Network (ISDN).

There are two forms of common signaling modes: associated mode, in which the control channel goes on the same trunk group as all the channels under its control, and nonassociated mode, in which the control signals are transferred on a totally separate network which controls the switching nodes that service the subscriber calls. Figure 4 shows an example of a nonassociated mode common channel signaling configuration.

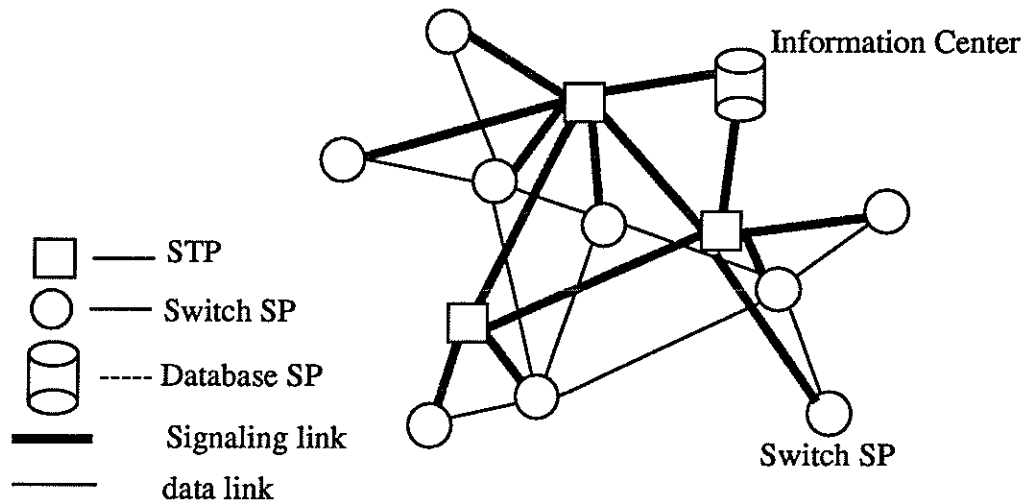


FIGURE 4. Nonassociated Common Channel Signaling Mode

Signal Transfer Points (STP) and Signaling Points (SP) are connected to form the signaling network. An SP is any point in the signaling network capable of handling control messages. In Figure 4, we show two kinds of SPs. One type of them is a switching point, the circle node with data link connected to it. A switch point SP is composed of a set of switches and a Control Processor who has control over the set of switches. We call this kind of SP an Exchange (EX). The other type of SP in the figure is an information center, which can hold a large database to serve the data network. An STP is a point capable of routing control messages. To the signaling network, the control signals are initiated and terminated at SPs.

### 2.2.1 SS7

Signaling System No 7, SS7 for short, is a signaling network standard defined by CCITT. SS7 defines a layered signaling network structure. Figure 5 shows the SS7 protocol structure and the correspondence to the OSI model.

The Message Transfer Part (MTP) corresponds to the OSI's lowest two layers and a part of the network layer. The overall purpose of the Message Transfer Part (MTP) is to provide a reliable transfer and delivery of signaling information across the signaling network. It also provides the mechanism for error and flow control.

The Signaling Connection Control Part (SCCP) enhances the services of the MTP to provide the functionality equivalent to OSI's network layer. It expands the MTP's address mode to allow multiple users to access the network simultaneously. It provides four classes of services: 1) basic connectionless class; 2) sequenced connectionless class; 3) basic connection-oriented class; and 4) flow control connection-oriented class. The user of the SCCP can flexibly choose the service classes to meet his application requirements.

The MTP and the SCCP together is called SS7 network Service Part (NSP). The NSP provides a platform that control information can pass through.

The rest of the components are based on the NSP to provide the services for the end user, or for the network administrators. Transaction Capability Application Part (TCAP) provides the transaction capabilities in an SS7 network, which can be used as the basis of distributed databases. Other applications, such as the Operation Maintenance Administration Part (OMAP) which provides the application protocols and procedures to monitor, coordinate, and control all the network resources, can be built on top of the TCAP.

ISDN User Part is an application that manages the underlying switching network. Since, in this paper, we are concerned with the signaling for the underlying circuit switching network, we will examine ISDN-UP in more detail.

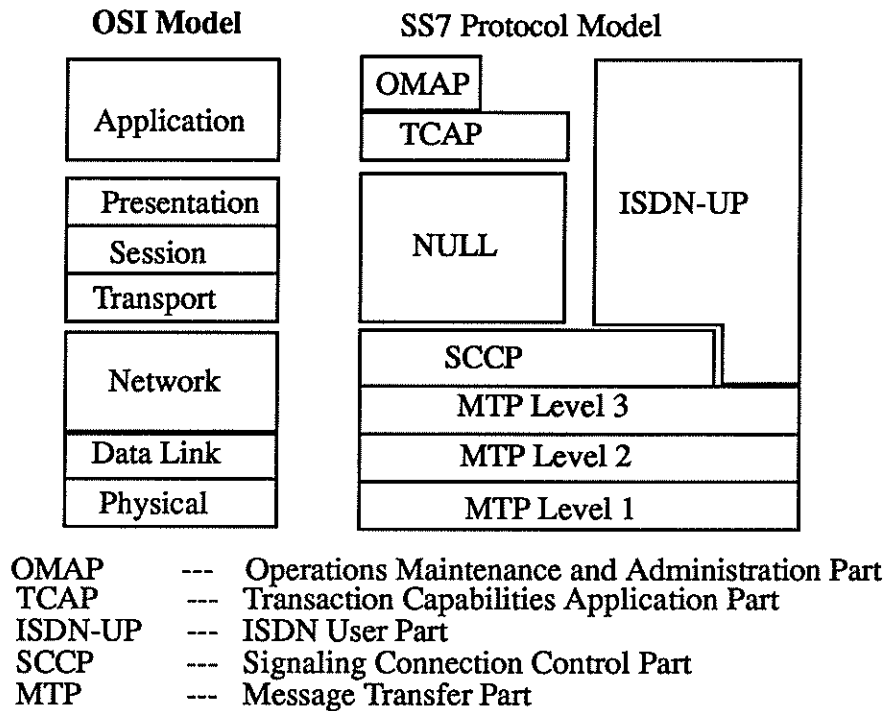


FIGURE 5. SS7 Protocol Architecture

### 2.2.2 ISDN-UP

The ISDN-UP protocol provides the signaling functions that are needed to support the basic bearer services, as well as supplementary services, for switched voice and non-voice (e.g. data) applications in an ISDN environment [APP86]. All ISDN-UP messages have a uniform format: a fixed length mandatory part, followed by a variable length parameter part, followed by an optional variable length parameter part. At the beginning, all ISDN-UP messages include a routing label identifying the origin and destination of the message, a Circuit Identification Code, and a message-type that defines the function and format of the message.

When an ISDN end-user initiates a service request, say a call setup request, the basic sequence of actions is as follows.

The user request message is sent, through the D-channel, to the central office switch, which forwards all the control messages to the Control Processor (CP) who has the control over the switch. The switch and the CP together form an Exchange (EX). We call this EX the originating EX. If there is a free channel on the outgoing link towards the destination and the current switch has a free channel from the incoming link to the outgoing link, the EX will try to reserve the resources. In the meanwhile, a corresponding ISDN-UP message, called Initial Address Message (IAM), is sent to the next EX through the signaling network; This process continues until the destination is reached.

Figure 6 shows sequence of operations for a user call set-up request.

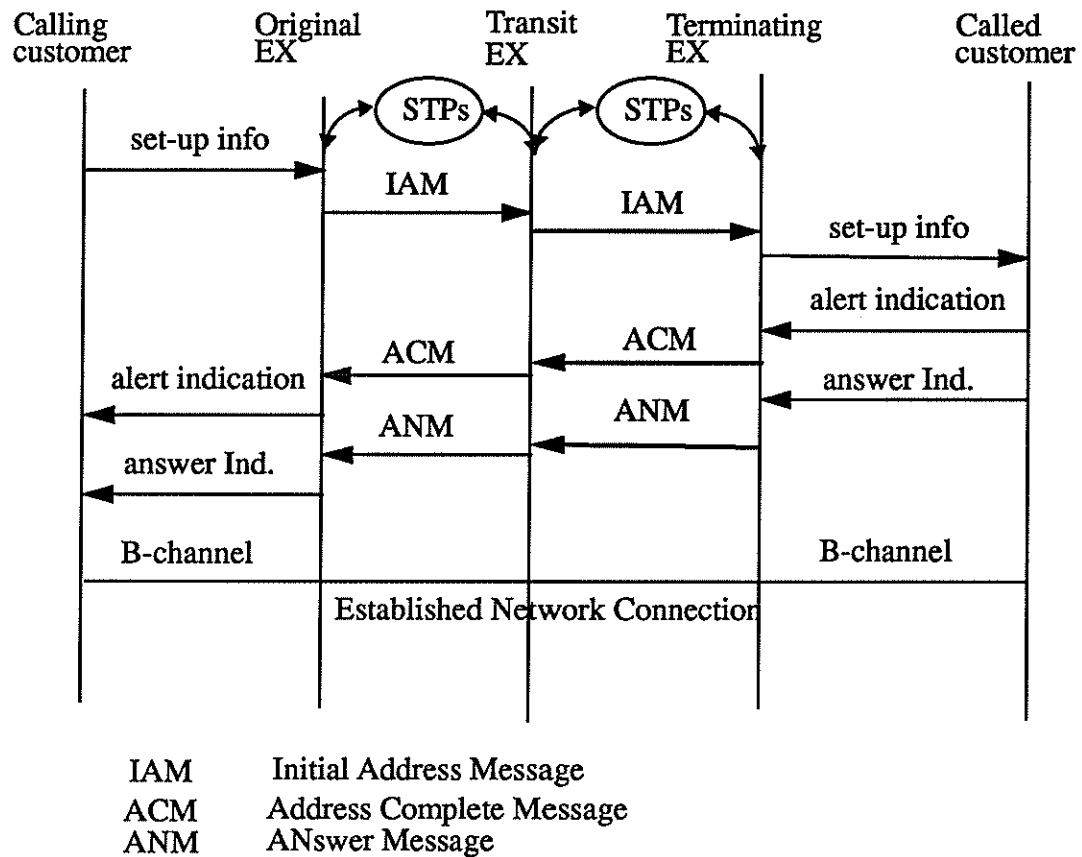


FIGURE 6. ISDN-UP basic call setup

- The calling customer sends a setup message through D-channel to the originating EX
- The originating EX checks its resources and finds the outgoing link. If the resources are available, the EX sets up the circuit in its switch. In the meantime, it sends an Initial Address Message (IAM) to the next EX through the signaling network.
- When an IAM is received by an EX, the EX does the same thing as the originating EX did when it received a setup message.
- The procedure of forwarding of IAM continues until the destination EX is reached. At the destination EX, a set up message is sent to the called party through D-channel.
- The called party equipment sends an Alerting Indication to the destination EX to indicate the setup message has been received.
- The destination EX generates an Address-Complete message when the alerting indication has been received, and sends the ACM on the reverse path. This message serves several purposes:
  - a) it serves as an acknowledgment to the originating EX that the requested network connection has been established,
  - b) it indicates that the called party was found to be idle and is being alerted,

- c) it indicates in the case where the originating EX has signaled the availability of the signaling connection control part protocol whether or not it is also available at the destination EX, and
  - d) it may contain a request to the originating EX to forward additional call-related information.
- When the destination EX receives the Answer indication from the called party, it generates an ANswer Message (ANM) and sends it to the originating EX to indicate the success of the setup request.

ISDN-UP provides more than the basic call control services. It also provides supplementary services such as closed user groups, user access to calling party or called party address identification, redirection of calls, connect when free, call completion to busy subscribers, in-call modification etc.

Since a whole signaling network is dedicated to signaling processing, there exists enough computing power and signaling bandwidth to provide more services than only for circuit switch control. Some databases can be installed in the signaling system to provide information services, such as 800 number service; even users can construct their own private data network by using the SS7 signaling network. Discussion of these additional services that SS7 can provide is beyond the scope of this paper.

### 2.2.3 Q.931

A user-network signaling protocol which is closely related to ISDN-UP is Q.931. Q.931 is specified by CCITT for ISDN. The main features of the protocol can be summarized as follows:

1. Out-of-band signaling: a channel, the D-channel of an ISDN link is used for signaling.
2. Support of both circuit switched and packet switched communication.
3. Signaling uniformly for either basic rate (2B+D) or primary rate (23B+D) interface.
4. Support point-to-point and point-to-multipoint line configurations. Multiple terminals can be connected to a bus which is connected to the local exchange. When the local exchange receives an IAM from the network, it broadcasts a setup message on the bus. When one of the terminals answers, the local exchange will send a release message to all the others. The protocol solves conflicts on First-Come-First-Serve basis.
5. Compatibility checking between the calling and called terminals. A called terminal only accepts calls from compatible calling terminals.
6. Symmetrical protocol for outgoing and incoming calls to enable the control of direct user-to-user connection, for example, direct PBX-to-PBX connection over leased-line circuits.
7. Modular message structure: each Q.931 message is composed of two parts: a common part and a message-specific part. The message-specific part is a set of message elements which follow common coding rules and makes the message formats very flexible.

8. alignment with the SS7 ISDN-UP.

In the next section, we will introduce Q.93B in more detail which is the counterpart of Q.931 in B-ISDN.

### **3.0 Signaling in Packet Switching Networks**

Circuit switching is efficient for constant data rate traffic such as voice. It is not efficient for bursty traffic such as computer based data communication since the communication resources are dedicated to a particular call. Packet switching networks emerged in the early 1970's, and were initially designed to allow more efficient use of line capacities for data communication.

In a packet switching network, user data is cut into fragments. Some control information is added to the data fragments, and the result is called a packet. At each node on route, the packet is received, stored and passed to the next node until the destination is reached. The control information serves, at a minimum, to guide the packet through the network [COM91].

There are two technologies to deliver data packets to the destination: datagram and virtual circuit.

#### **3.1 Datagram**

With datagram service, there is no signaling involved. Packets are sent independently. Each packet chooses its route depending upon the network topology and the network load. The Internet Protocol (IP), sponsored by DARPA, is the most widely used datagram protocol. With datagram service, the network does not guarantee the reliable transmission of the user data. Packets may be lost, duplicated, or get out of sequence. By building a reliable transport protocol on top of a datagram service, one can provide the end-user with reliable transport services.

#### **3.2 Virtual Circuit**

With the virtual circuit approach, a route, called a virtual circuit, is established before any data transmission can occur. The links on the route are not dedicated to the connection, they are shared by multiple virtual circuits. Signaling protocols are used to set up and tear down virtual circuits. CCITT has defined a whole set of protocols for the packet switching networks [ROB78]. Figure 7 depicts the typical applicable protocols in a packet switching network.

X.3 defines the functions of a Packet Assembler/Disassembler (PAD) which, along with X.28, works as an interface to a character originated data terminal and a packet network; X.25 is the protocol between a Data Terminal Equipment and a Data Circuit terminating Equipment. It is a protocol both for data transmission and user-network signaling; X.29 is a protocol between PADs; and X.75 is a gateway protocol between different networks.

Among the protocols introduced, X.25 is well known and commonly used in packet switching networks as the user network interface protocol. In the rest of this section, we will discuss the signaling part of X.25.

X.25 specifies three layers of protocols: physical layer, data link layer, and network layer. The data link layer uses LAPB (Link Access Protocol Balanced), which provides reliable data transfer over physical links. The network layer, or Packet Layer as defined in X.25, is responsible for virtual circuit set up, tear down, reliable end-to-end transmission, and for error and flow control.

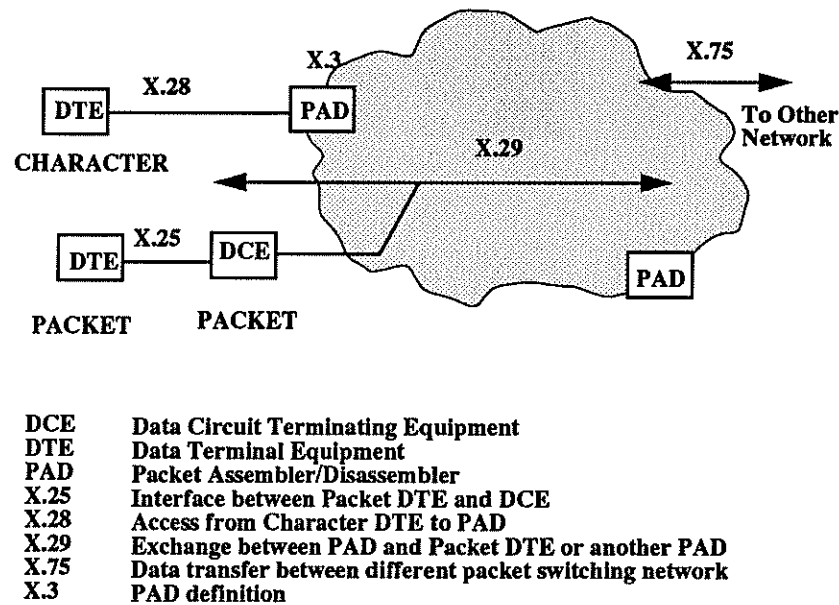


FIGURE 7. The protocol family in a Packet Switching Network

Figure 8 shows an example of a complete communication session. The source DTE sends a call-request packet to the DCE containing, in addition to the required destination DTE network address, a reference number called the Logical Circuit Identifier(LCI). The packet is forwarded through the network to the destination DCE. At the destination DCE, a second LCI is assigned to the call-request packet before it is forwarded to the destination DTE by an incoming-call request. The destination DTE sends a call-accepted message to the destination DCE. At this time, the destination DTE can start sending data packets on the virtual circuit. When the source DCE receives the message indicating the call has been accepted, it sends a call-connected message to the calling DTE, that ends the virtual circuit establishment phase.

X.25 defines the protocols between the user and the network. It does not restrict the network internal protocols. At the network node level, a network designer can use any protocols either datagram or virtual circuit.

Packet switching has a number of advantages over circuit switching: 1) since packets can be buffered at each node, the line efficiency can be higher; 2) a packet switching network can perform data rate conversion so that the data equipment in the network can work at different speeds; 3) priorities can be used to make important messages go faster. Some disadvantages are: 1) since all the packets have to be processed by software at each node on the route, the delay will be longer; 2) control information in each packet can be big overhead.

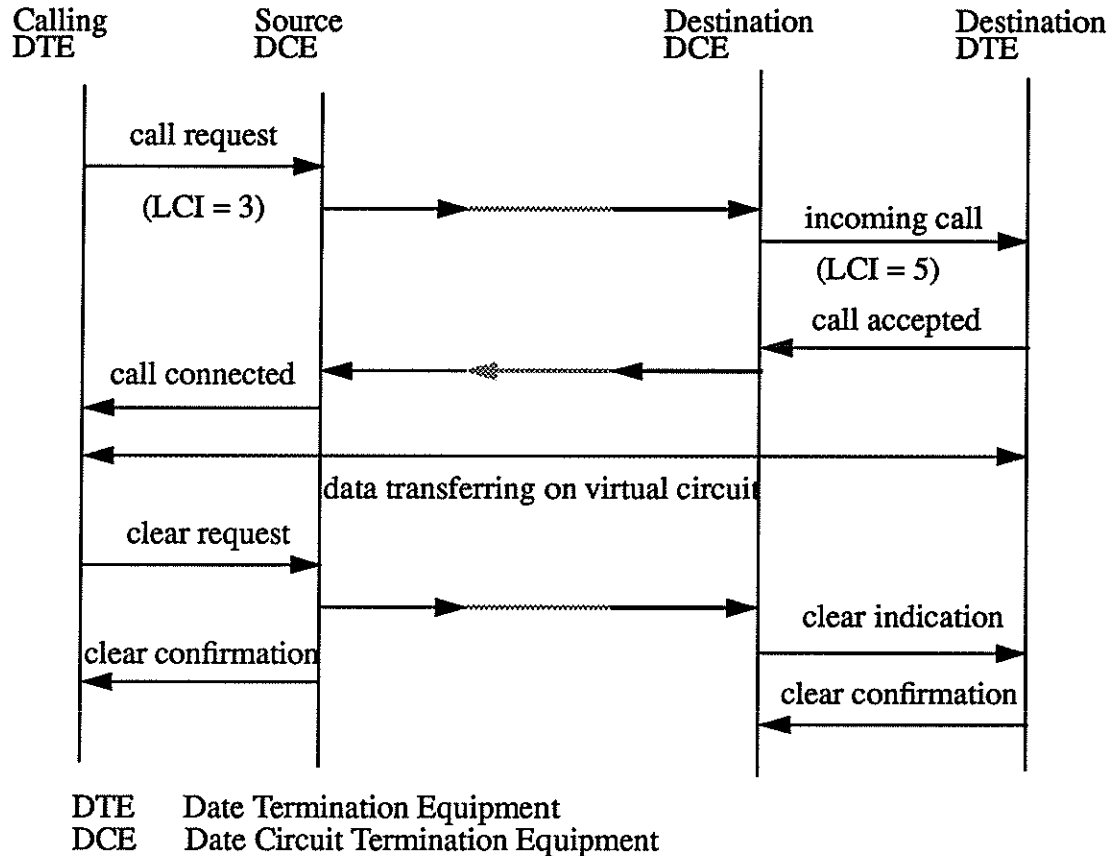


FIGURE 8. Sequence of Events: X.25 Protocol

The design of fast packet switching networks has been motivated in part by the desire to combine the flexibility of packet switching with the relative simplicity of circuit switching.

#### 4.0 Signaling in Modern High Speed Network

Two main factors are driving the telecommunication towards the fast packet networking [TURN86]. The first driving force is the diverse new communication service requirements, such as multipoint and multimedia communications, high quality video and audio communications. To meet these new requirements, a network has to have some primary capabilities such as multipoint signaling, dynamic allocation of network resources, synchronization among multimedia information streams, etc.



The second driving force is the development of new technologies. The dropping cost and increasing capacity of fiber optics makes high speed communication, up to gigabits per second, realistic. The development of integrated circuit and computer technologies provides the possibility to produce fast packet switching system.

In section 4.1, we will briefly introduce the fast packet switching technology. Since ATM, Asynchronous Transfer Mode, has been designated as the standard for Broadband ISDN by CCITT, in the following sections we concentrate on signaling in an ATM network. Section 4.2 discusses signaling protocols at the user-network interface. Section 4.3 introduces the signaling protocols at the network node level; Section 4.4 discusses the routing problem which is closely related to the signaling problem.

## 4.1 Switching Technology

Switches are the key elements in a fast packet switching network. To keep up with the line speed, a switch has to route packets in hardware. There are two main techniques to route a packet in a high speed switching network: source routing, or label swapping.

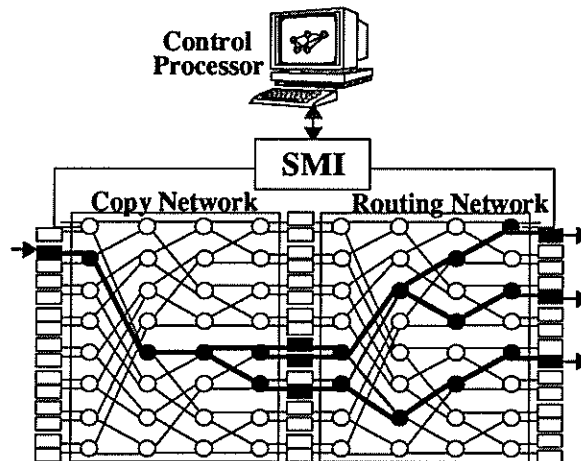
With source routing, the packet sender has to know the whole route the packet will take. It attaches the whole route, as a sequence of link identifiers, as a part of the packet header. When the packet is received at a switch, the switch chops the first link identifier off, and forwards the packet along the link indicated by the first link identifier. This procedure continues until the destination is reached. Source routing is efficient in the case that the message size is long. The advantages are: the user can send a whole message in one packet that simplifies the intermediate layer processing, for example fragmentation and reassembly are not necessary; switch functions are very simple since switches do not have to hold information for routing; connection set-up is not necessary. The disadvantages are: for short packets, the routing information can be big overhead and the source has to know the global network status which is impractical in a large network. As a result, source routing fits well in a small private network, but it can not work well in large public telecommunication environments which may involve thousands of switches.

Another routing technique is called label swapping routing. A user sends a packet with a label as part of the packet header. When the packet is received at a switch, the switch looks up information in a table with the label as the index to find out the forward link and the next label. Then the packet is sent along the forward link with the new label. Two problems have to be solved before the technique can be implemented: a switch has to have some memory to store the tables and a mechanism to access the table quickly; before the data transmission can take place, the tables along the route have to be filled with consistent values. The first problem has been solved by using VLSI. The second problem implies that we have to have a connection set-up phase to signal the switches to update their tables. In contrast with the source routing technique, label swapping scales very well. CCITT has assigned ATM, Asynchronous Transfer Mode, which uses the label swapping technique, as the standard for Broadband ISDN (BISDN).

In an ATM network, data are divided into small fixed size packets, called cells. Each cell has a label, a virtual circuit identifier (VCI), associated with it. Cells are routed through the network by label swapping as described.

ATM networks are designed to meet multipoint, multimedia and other emerging service requirements, so that the switches in an ATM network have to be intelligent enough to support various service requirements.

Figure 9 shows the architecture of one ATM switch, Turner's *Broadcast Packet Switch*. This switch contains a *Copy Network* (CN) concatenated with a *Routing Network* (RN). ATM cells enter the switch on the left. Multicast cells, destined for several locations, are replicated by the CN, then routed to the appropriate destinations by the RN. Point-to-point cells follow an arbitrary path through the CN, then are routed by the RN. Cells leave the switch on the right, where they traverse fiber optic links to other switches or clients.



**FIGURE 9. Architecture of Turner's Broadcast Packet Switch Demonstrating Multicast Routing.**

The switch is controlled by a *Control Processor* (CP) connected to the switch via a *Switch Module Interface* (SMI). The CP configures the switch hardware to route incoming cells to the appropriate outgoing links by modifying tables within the switch, thus establishing connections. The switch routes signaling cells from clients or other nodes (as distinguished by the header) to the CP via port 0. The CP manages the switching resources, and it is responsible for setting up or tearing down connections.

## 4.2 User-Network Signaling

As mentioned previously, network signaling can be classified into user-network signaling, network-node signaling, and user-user signaling. Since ATM is a new technology, all the protocols, standards, and testbeds are under development. Neither user-network signaling protocols nor network-node signaling protocols have been standardized. In the rest of this section, we briefly introduce Q.93B which is a user-network signaling protocol under development by CCITT Study Group XVIII. Then we introduce some concepts of the

Connection Management Access Protocol (CMAP) proposed by Applied Research Lab of Washington University (ARL). CMAP can be viewed as an extension of Q.93B. In the next section, we will introduce the Connection Management Network Protocol (CMNP) which is a network node protocol proposed by ARL.

#### 4.2.1 Q.93B

Q.93B is a user-network signaling protocol designed for ATM networks [ATM93]. Q.93B supports both point-to-point and one-to-many multipoint communications. From the user's point of view, the basic communication object is a call. Calls can have different parameters associated with them, such as bandwidth, Quality of Service (QOS), etc. Within each call, there is only one connection. This excludes the possibility to synchronize multiple media by the network.

A one-to-many call can be modeled as a tree. The root is responsible for setting up the tree by sending ADD\_PARTY messages to all the participants. Only the root can send messages to the users participating in a one-to-many call.

Q.93B defines the signaling messages, the states at both the user and the network sides, and the procedures to process messages. All the signaling message types are listed in Table 1. Messages can be classified into four categories: 1) call establishment messages are for setting up a new point-to-point call; 2) one-to-many messages are for manipulating a multipoint connection; 3) call clearing messages are involved in the disconnection routines; and 4) miscellaneous messages are related to the status inquiries.

TABLE 1. Q.93B Message Types

Message Name	Description
Call establishment messages:	
SET_UP	Calling party to network or network to called party to request to set up a point-to-point connection
CALL_PROCEEDING	Network to calling party or called party to network to indicate the SET_UP message has been received
CONNECT	Network to calling party or called party to network to indicate a connection has set up
One-to-many messages:	
ADD_PARTY	The root to the network, or the network to called party to request to add an end-point
ADD_PARTY_ACK	Called party to network or network to the root to indicate a new end-point has been added
ADD_PARTY_REJECT	Network to calling party or called party to network to reject an ADD_PARTY request
DROP_PARTY	The root to the network, or the network to called party to request to drop an end-point
DROP_PARTY_ACK	Called party to network or network to the root to indicate a new end-point has been dropped

**TABLE 1. Q.93B Message Types**

<b>Message Name</b>	<b>Description</b>
<b>Call clearing messages</b>	
DISCONNECT	Either calling or called party to network to request to clear a connection
RELEASE	The sender intends to release the resources
RELEASE_COMPLETE	Indicating the resources such as VCI can be reused
<b>Miscellaneous messages:</b>	
STATUS_INQUIRY	Request the status information
STATUS	Status information

A possible one-to-many call set up sequence is given in Figure 10. One user, the root, starts to set up a point-to-point connection by sending a SET\_UP message. When accepted, the network node sends a CALL\_PROCEEDING back, and the SET\_UP message is forwarded to the called party. The called party accepts the call by sending a CONNECT message, which is forwarded back to the root. After the point-to-point connection has been set up, the root can issue multiple ADD\_PARTY requests to add all the participants. An ADD\_PARTY request may be rejected by the network due to insufficient resources, or rejected by the called party by an ADD\_PARTY\_REJECT message. When accepted, the called part sends an ADD\_PARTY\_ACK message, which is forwarded to the root.

Q.93B also defines the timers and error processing procedures.

Comparing with Q.931, a user-network access signaling protocol for ISDN, there are several points worth noticing:

1. Q.931 supports only point-to-point calls while Q.91B supports one-to-many calls as well as point-to-point calls.
2. In the SET\_UP message and ADD\_PARTY message, the user can define required services such as bandwidth and QOS, so that network resources can be more efficiently used.

- ATM networks require service capabilities that are not yet addressed in Q.93B, meaning that future extensions will be required.

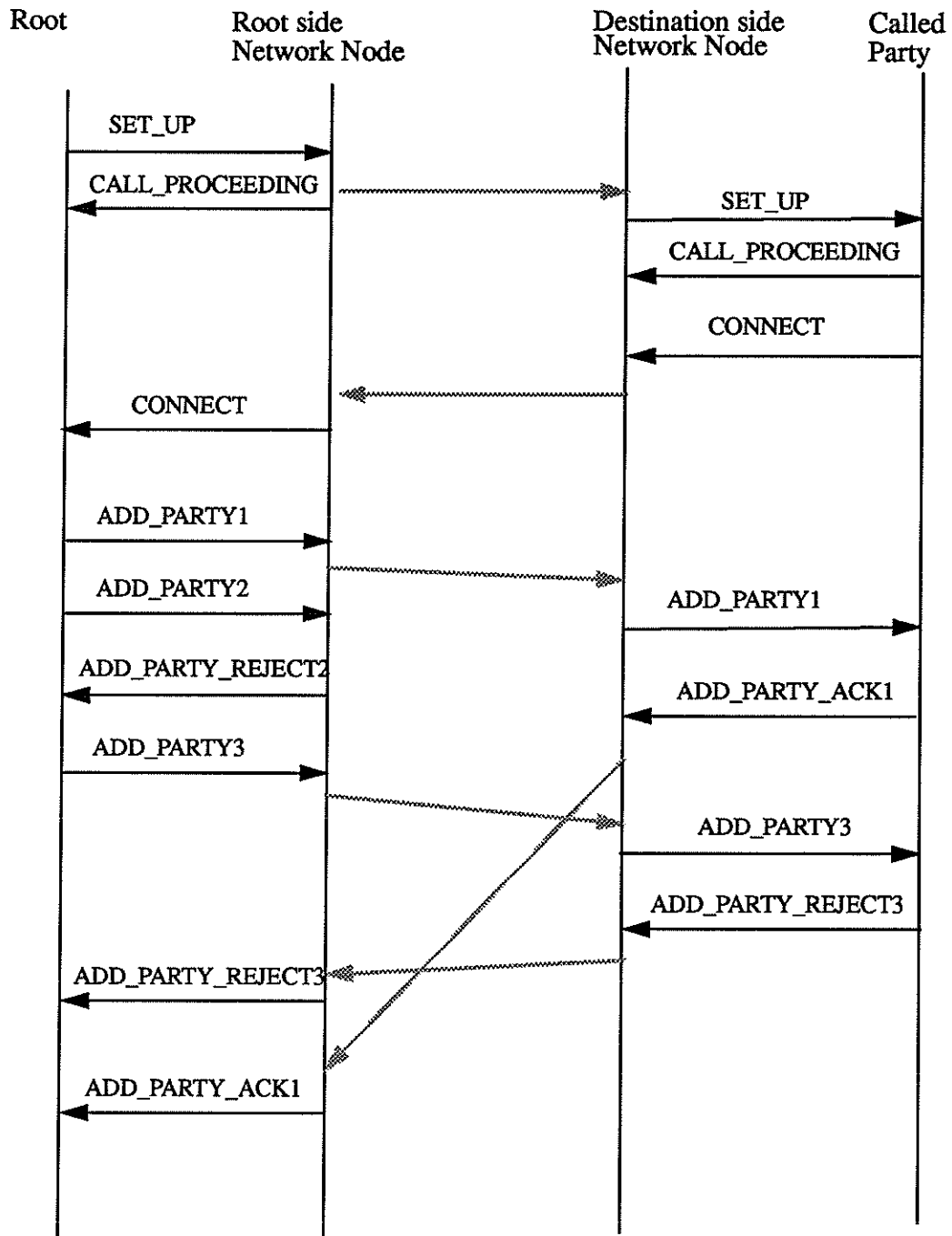


FIGURE 10. A one-to-many call set up

### 4.2.2 CMAP

Many new applications require more services than Q.93B can support [DEH92]. For example, a conference call may need many-to-many communications and multimedia synchronization. A distributed data collection system may need many-to-one services. The Applied Research Lab at Washington University (ARL) proposed a new network access protocol for ATM networks named Connection Management Access Protocol (CMAP).

CMAP can be viewed as an extension of Q.93B. CMAP allows users to create, manipulate, and delete multipoint, multimedia communication channels, which are termed calls. A multipoint multimedia call contains a set of connections and a set of end-points. Each multipoint connection is a communication channel between two or more clients (end-points) of the network, where all the data sent by one client is received by all other clients who have the receive attribute set. A point-to-point, or a one-to-many connection is a special case of a multipoint multiconnection call.

When a call is created, one or more connections are established between the network and the client who creates the call. This client is designated the owner of the call. Additional clients can be added by: 1) invitation from the owner, where the invited party has the option of refusing the invitation, 2) request from a client where the owner has the option of denying the request, or 3) request from a third party, who is not necessarily in the call, where both the owner and the client being added have the option to refuse. Once a call has been created, additional connections and clients can be added to, or deleted from, the call.

Figure 11 shows an example of a multimedia call at a client. Calls have a number of parameters that describe how clients may access the call. The main parameters are listed in Table 2.

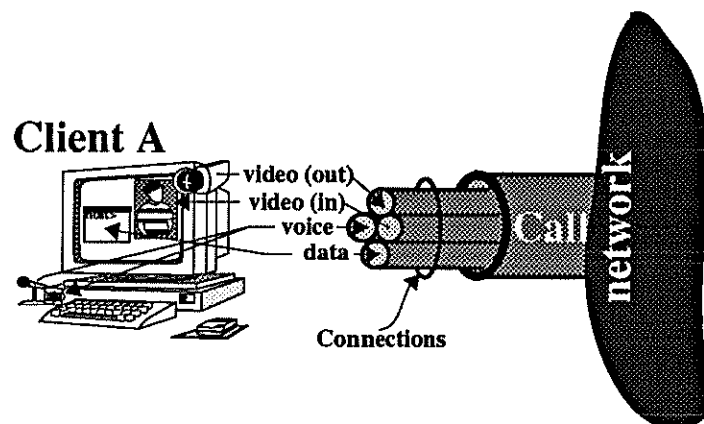


FIGURE 11. Example Multimedia Call at the UNI.

TABLE 2. Call Parameters

Parameter	Description
Call_id	a globally unique identifier for the call
Accessibility	Freedom (or lack thereof) with which clients can join the call
Monitoring	Level of notification of client joins and drops
Modifiability	Whether a nonowner client may add connections
Connection List	Current connections in call

The Call\_id is a globally unique identifier for the call. The accessibility parameter defines who has the right to add a new end-point into the call. The monitoring parameter defines who will be notified when an end-point joins or drops the call. The modifiability parameter controls whether a non-owner end-point has the permission to add connections to the call. The connection list is a list of connection identifiers that define the different connections under the control of the call.

A connection corresponds to a multipoint channel. Table 3 lists the main parameters for a connection.

TABLE 3. Connection Parameters.

Parameter	Description
Connection_id	a unique channel identifier within a call
Type	Three tuple consisting of VP/VC, dynamic/static bandwidth and quality of service
Bandwidth	The reserved bandwidth for the connection ( <i>best-effort</i> connections are supported)
Permission	Restrictions on the transmit/receive access of endpoints
Mapping	The current transmit/receive access mapping of each endpoint

The Connection\_id uniquely identifies a communication channel within a call. The connection type parameter defines the channel type, either a VC or a VP, the bandwidth type, either static or dynamic, and the quality of service. The bandwidth parameter defines the bandwidth requirements including the average bandwidth, peak bandwidth and peak burst length. The permission parameter restricts the way that a client can access the communication channel. The connection mapping parameter switches the transmission, receiving, and echo ability on or off to realize one-to-many, many-to-one, or many-to-many multipoint communications.

One of the most important contributions of CMAP is the separation of connections from a call. By manipulating the call and connections within the call, users can flexibly construct communication channels for their special needs.

### 4.3 Network Node Signaling

Each network node consists of one or more switches. The function of the network node signaling is to make the nodes in a network work together to establish, modify, and release

connections consistently and efficiently [CMNP93]. The functional part that performs network node signaling is called the Connection Management Functional Area (CMFA).

There are several important functions that a CMFA should provide: it should provide a simple yet powerful model of abstract network connectivity to its clients that flexibly supports a wide range of bandwidth, QOS, and network topologies and it has to efficiently manage network resources.

For practical reasons of scale, processing load, geography, administrative partitioning, flexibility, reliability, etc., the CMFA has to be implemented as a distributed collection of connection management functional instances called Connection Managers (CM). Each CM will exclusively control the resources at a node. Each CM has to support three kinds of interfaces:

1. An interface to the clients of the CM. Since different users may use different user-network signaling protocols to access the communication network, the CM has to provide a powerful and general interface to support different access protocols.
2. An interface to the network resources under its control. Through this interface, the CM can change the underlying switch configurations to realize the signaling functions.
3. An interface among the CMs through which CMs can communicate with each other to manage network resources consistently.

ARL has designed a Connection Management Network Protocol (CMNP). CMNP specifies message formats used to pass the control information among network nodes to create, modify and delete multipoint multimedia connections named Connection Groups (CG). CMNP also defines the actions at network nodes when messages are received.

The basic manageable objects of CMNP are CGs. A CG contains a set of connections as well as a set of attributes. Table 4 lists the main parameters for a CG.

**TABLE 4. Connection Group Parameters**

<b>Parameter</b>	<b>Description</b>
<b>CG_ID</b>	Connection Group's identifier.
<b>CM_Ack_Flag</b>	indicating whether the owner of the CG has to be informed when a new end-point is to be added into the CG.
<b>CG_correlation</b>	whether all the connections in the CG have to be routed on the same path.
<b>Connection_list</b>	Current connections in the CG

A CG is initially created by receiving a CREATE\_CG request from a session manager when the session manager has received an OPEN\_CALL request from a user-network interface. The call-id of the call becomes the connection group id of the CG. The CM\_Ack\_Flag is a boolean variable that indicates whether the owner of the CG has to approve before a new end-point can be added into the CG. CG\_correlation is another boolean variable to indicate whether all the connections within the CG have to be routed on the same path. Connection\_list defines multiple channels within the CG.



Table 5 lists all the main parameters of a connection in a CG.

**TABLE 5. Connection Parameters.**

Parameter	Description
Connection_id	a unique id within the CG
Con_type	VC or VP
rx_bw	receive bandwidth
tx_bw	transmission bandwidth
qos	quality of service

A connection\_id uniquely identifies a connection within a CG. Con\_type indicates the type of the connection, either VC or VP. Rx\_bw and tx\_bw give the receiving and transmission bandwidth requirements. QOS defines the quality of service.

Table 6 lists the major messages defined in CMNP.

**TABLE 6. Messages in CMNP**

Message	Description
Create_CG	from a session manager to CM to request to create a CG
Join_CG	from a session manager to a CM or from one CM to another to request to join an existing CG
Drop_CG	from a session manager to a CM, or from one CM to another to drop nodes from a CG.
Destroy_CG	issued by the owner of a CG, then passed from CM to CM to destroy a CG
Add_Connection	from a session manager to a CM or from a CM to some others to request to add a new connection into an existing CG
Drop_Connection	from a session manager to a CM, or from a CM to another to drop nodes from a connection within a CG
Destroy_Connection	issued by the owner of a CG, then passed from CM to CM to destroy a connection within a CG

Each message has two types, a request type and a response type. The response type of a message works as an ACK or NACK. In the case of NACK, a cause field in the message gives the reason of the failure. A general multipoint CG creation scenario is given as follows.

When a session manager receives an OPEN\_CALL request from a user-network interface, the session manager sends a CREATE\_CG request to the local CM. The CM creates a CG and the session manager is assigned as the owner of the CG. When another end-point wants to join an existing CG, the session manager serving that end-point sends a JOIN\_CG request to the local CM. The CM checks its own resources. If the resources allow, the CM will reserve the resources, and then after calling a routing subroutine to find the next node towards the owner, the CM will send another JOIN\_CG request to the next node. The CMs pass messages in this way until a node at which the CG has already existed is reached. Then an ACK response is sent backwards. In this way, a multipoint call can be constructed by multiple JOIN\_CG requests.

CMNP also defines a transaction mechanism. The user of the CMNP can group his requests into a transaction that will be executed as an atomic operation.

CMNP is still under development. Details can be found in [CMNP93].

#### 4.4 Routing Algorithm

In a network signaling algorithm, routing plays an important role. The performance of a routing algorithm is crucial to the efficiency of a network system. A network can be modeled as a bidirectional graph with a weight associated with each edge. For point-to-point connections, a distributed shortest path algorithm can be used. Maruyama [Maru83] gave a good overview of routing algorithms for session based communications. For a multipoint connection, the ideal is to find a minimum spanning tree to cover all the terminals involved in a multipoint connection. Unfortunately this has been proved intractable in the theoretical sense. Waxman and Imase, [IMAS91, WAXM93], define the routing problem as a dynamic Steiner tree problem and Waxman [WAXM93] gives a greedy algorithm for this problem and shows, by simulation, that the greedy algorithm performs well in the sense of efficiently allocating network resources. Further discussion of routing algorithms is beyond the scope of this paper.

#### 5.0 Conclusion

Signaling is an important component of communication networks. A signaling protocol has great impact on the efficient use of network resources, and the services a network can provide to its users. ATM networks have the potential capability to provide flexible services. Signaling protocols are responsible to make full use of these capabilities and to provide users all the services an ATM network can provide. ATM networking is still a relative new area, and so is signaling for ATM networks. Most of the protocols are not standardized. Some existing protocols need to be extended. Some new and potential services have to be added.

In this paper, we have reviewed the history and evolution of network signaling. We pay most of our attention to signaling in emerging ATM networks. At the time this paper is written, the ATM Forum has worked out the "ATM User-Network Interface Specification (v3.1)", which contains the new version of Q.93B. ARL at Washington University in St.Louis has implemented the core CMAP and CMNP. One trend of the signaling system is to support multipoint, multimedia applications. CMAP and CMNP is quite flexible and successful in this direction. Some of the concepts built in CMAP and CMNP have been accepted by the standard bodies. It is the author's perception that some technique that supports multimedia and multipoint communications will appear in future signaling standards.

There are a lot of research topics in the network signaling field. Some basic ones are: how to model the signaling system; how to define the efficiency of a signaling system and to evaluate it; how to make the signaling system scalable and reliable. For practical reasons, software that implements signaling protocols has to be designed as a distributed system.

this introduces more problems such as synchronization among processes, information sharing, communication reliability, etc. Routing remains a problem. In a distributed system, how nodes cooperate and update their routing information, how to find an efficient route are still hot research topics.

## **Acronym List**

The following acronyms are used within this paper:

**ACM** — Address Complete Message

**ATM** — Asynchronous Transfer Mode

**B-channel** — Data channel for ISDN

**BISDN** — Broadband Integrated Services Digital Network

**BW** — Bandwidth

**CCITT** — International Telegraph and Telephone Consultative Committee

**CM** — Connection Manager

**CMAF** — Connection Management Access Protocol

**CMNP** — Connection Management Network Protocol

**DCE** — Data Circuit termination Equipment

**D-channel** — Signaling channel for ISDN

**DTE** — Data Termination Equipment

**IAM** — Initial Address Message

**ISDN** — Integrated Service Digital Network

**ISDN-UP** — ISDN User Part

**QOS** — Quality of Service

**NNI** — Network Node Interface

**SS7** — Signaling System No. 7

**STP** — Signaling Transfer Point

**UNI** — User Network Interface

**VC** — Virtual Channel

**VCI** — Virtual Channel Identifier (ATM header field)

**VP** — Virtual Path

## References

- APP86 Appenzeller, H. R., "Signaling System No. 7 ISDN User Part," IEEE Journal on Selected Areas in Communications, May 1986
- ATM93 ATM Forum, "ATM UNI Specification (v2.1)," May 1993
- COM91 Comer D., Stevens D., "Internetworking with TCP/IP," Prentice Hall, Englewood Cliffs, NJ, 1991
- CMNP93 Dehart, J., Wu, D., "CMNP specification," Computer Science Department, Washington University in St. Louis, Technical Report (in preparation)
- DEH92 Dehart, J., Gaddis, M., Bubenik, R., "Connection Management Access Protocol (CMAP) Specification," Computer Science Department, Washington University in St. Louis, Technical Report WUCS-92-01, 1992
- IMAS91 Imase, M., Waxman, B., "Dynamic Steiner Tree Problem," SIAM J. Disc. Math., Aug. 1991
- MAR83 Maruyama, K., Shorter D., "Dynamic Route Selection Algorithms for Session Based Communication Networks," Computer Communication Review, April 1983
- MIN89 Minzer, S., Spears, D., "New Directions in Signaling for Broadband ISDN," IEEE Communications Magazine, Feb. 1989
- MOD90 Modarressi, A., Skoog R., "Signaling System No. 7: A Tutorial," IEEE Communication Magazine, July 1990
- ROB78 Roberts, L., "The evolution of Packet Switching," Proceedings of IEEE, v. 66, Nov. 1978
- REY83 Rey, R., editor, "Engineering and Operations in the Bell System," Murray Hill, NJ: AT&T Bell Lab, 1993
- STAL89 Stallings, W. "ISDN an Introduction," Macmillan Publishing Company, 1989
- TURN86 Turner, J. S., "New Directions in Communications," IEEE Communications Magazine, Oct. 1986
- VERM90 Verma. P., "ISDN System Architecture, Technology, and Applications," Prentice-Hall Inc., 1990
- WAX93 Waxman, B., "Performance Evaluation of Multipoint Routine Algorithms," INFOCOM 1993