TABLE OF CONTENTS

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

CHAPTER ONE

INTRODUCTION

The integrated services digital network (ISDN) is one of the hottest buzzwords today, just as the micro-chip was about fifteen years ago. A lot of development has taken place since then, in both the computing and the communications worlds. Imagine a computer sitting on your desk with which you could ring anywhere in the world and carry on a telephone conversation while at the same time accessing a remote database from within another window on your screen and using the same wall socket: your introduction to the world of ISDN. Imagine yourself then making a call to your local estate agent and scanning through their home catalogues on your screen while discussing the details with them over the phone connected to the same twisted pair cable. Imagine yourself in the office conducting multi-party, multimedia conferencing across continents. Well, all of this technology is currently available, thanks to ISDN and the developments in data and telecommunications techniques and standards.

Local area computer networks (LANs) have become as ubiquitous in today's offices as personal computers are at home. Formerly, the interconnection of LANs to each other left a lot be desired in that they mostly used the old analogue telephone lines and modems operating at the now abysmally low transmission rates of 9600 bit per second bps. Packet-switched data networks (PSDNs) have alleviated some of these problems, but the technology and applications in the LAN world have also outgrown their original bounds. What is currently needed is a method of transparent interconnection of LANs and of porting the applications developed for the relatively fast and error-free environments found on the local area networks across long distances over the ISDN. These service should be available to home workers, office interconnects and global LAN interconnection. Dynamically varying network topology's and virtual networks are now possible thanks again to the ISDN. Faster speeds, a very good bit error rate (BER), end-to-end digital connectivity, support of multiple services and types of services standardized access methods and standardized hardware and software are now a reality. For voice, data and video services supported concurrently through one global network and ubiquitous access all around the globe, the ISDN is the way forward; and, what is more, the technology is available today.

1.1 EMERGENCE OF ISDN

The invention of the transistor and the subsequent development of microchip technology have resulted in variety of useful applications in almost every aspect of life; from home to office, from industry to commerce, from government to health care to education. This is currently processing in two mains fronts; the computer and communications technologies. When intelligent functionalities were increase in the old telecommunications networks, it came in the form of the stored program control (SPC). This was the use of some computing function running special software to control the exchanges in the network. A multitude of services then became available-call logging, itemized bills, transfer of calls; conferencing, ring-back-when-free, etc. On the other hand, it became possible for computers with simple interface cards and modems to access the telecommunications services through their communications ports and software. Slowly, a bit of the functionality of each of these technologies became a part of the other. Today they are inseparable in many respects. Coupled with this background development are the user requirements of accessibility, flexibility, speed, cost effectiveness and new services on the one hand, and the desire of the service provides to increase serviceability and to ease the management problem on the other. Hence the technological and market forces have required the integration of existing networks and services already provided on separate networks. This had led to the concept of an integrated services digital network, a network that would provide solution to most of the problems in the communications world.

There are two main aspects of ISDN: the network evolution and the services provided. As far as the users are concerned, ISDN provide all the necessary communications services or access to the services provided by other networks in a transparent fashion. As far as the service providers are concerned, one homogeneous or several heterogeneous interworking networks may provide the services. This does not change the view of the ISDN from the outside: that of a uniform service provider through standard interfaces.

The integrated services digital network (ISDN) is evolving from the integrated digital network (IDN) concept and is taking shape worldwide. The IDN provides the integration of the switching and transmission facilities and extents it to the subscriber loop by digitisation in the network. It also provides for the common channel signalling which is based on the transmission of the control and signalling messages on a packet-switching network design for this purpose and is part of the public switched telecommunications network (PSTN). One of the fundamental concepts in the creation of ISDN is the provision of a multitude of switched and non-switched services to the users with in the circuit, packet of the frame modes of access, through the use of a small set of the standard user-network interfaces (UNIs). Apart from fast switching, which is possible because of the end-to-end digital connectivity, the ISDN utilizes common channel signalling (CCS), allowing the selection of different services through the use of a standard signalling protocol at the UNI. Furthermore, the ISDN provides multiple channels to the user at the UNI. The main multiplexing technology used in the digital telephony world is time division multiplexing (TDM). Assigning one or more time slots (TSs) with in a TDM frame to a channel forms multiple channels at the ISDN user-network interface. The TDM technology lends itself easily to circuit switching (CS). Hence, circuit switching is *inherent* to ISDN, and it is one of the earlier services available in ISDNs.

More recent technological developments taking place in the ISDN world are the frame mode services, based on the more efficient use of the ISDN technology, and the broadband ISDN services, based on the asynchronous transfer mode (ATM) and the fibre optic technology. This will bring a drastic improvement in the way most current services are provided. For example, the frame mode services are purposed as the main method of LAN-LAN interconnection in the very near future. The ATM technique is proposed as the main technology for multi-megabit services including high-definition TV and video conferencing.

Some of the factors affecting the development of the ISDNs are:

- *Basic technology* Among developments in the sophistication of electronic components, the current very large-scale integrated circuit (VLSI) chip technology allows the implementation of many more functions in hardware and firmware. For example, silicon chips are available accommodating the high-level data link control (HDLC) protocol for layer 2 of the OSI reference model. Developments in the fibre optic technology are another factor.

- *Increased use of computers* Computers are nowadays used in almost every walk of life, including government, commerce, banking, education, research, industry, and tourism and leisure. The development of sophisticated applications requiring more and more hardware capacity and communications bandwidth means that the computer and telecommunications technologies have to deliver. The availability of on-line information services and large databases, and the demand for electronic shopping and banking facilities, mean that more and more data networks will be built. Although currently the data services amount to only a very small percentage of the whole telecommunications sector, this is bound to change in the near future.

- *Increased use of communications services* The whole public and commercial life of individual countries and companies nowadays depends more than ever on the use of communications facilities. This provides instant access to information on a national and global basis. The whole trade and commerce sector is now completely dependent on the availability of telecommunications and computing services. For example, one cannot conceive of a just-in-time type stock provision without telecommunications and computing facilities.
- *More time and money to spare* With current developments in the economic and social spheres, most people will have more time and many to spend on technology and leisure. They will want more TV channels and easier access to electronic services and teleconferencing facilities.
- *Working from home* With the increase in computing and communications facilities, 'home commuters' are growing in number. This especially suits people who prefer to work at their home rather than the office environment or who have a got reason not to travel. It also suits some businesses, allowing them to cut down on office space, heating costs, etc.
- *Information technology* The current state of information technology has reached such a level that the storage of information is no longer the limiting factor; rather, it is the access and manipulation of the information. Hence faster and better communications facilities are needed.
- *Computer-aided production and support* Most industrial output is nowadays controlled by communicating computers. The move is to lock the parts manufacturers, stock suppliers, engineering design, manufacturing and servicing centres, as well as the distributors, into one work environment based on suitable standards, software, services and a global computer and communications network. An example of this is the Computer-Aided Logistics Support (CALS) programmer of the US Department of Defence.

1.2 THE DATA COMMUNICATIONS REVOLUTION

A revolution has been taking place since the early 1980s whereby the individual computer with its own domain of information and work is no longer a preferred solution a computing. Today, computers have become much more friendly and accessible. They have become tools for productivity rather than simply pieces of scientific equipment. Distributed computing based on workstations, personal computers and their networks has become the norm. This has resulted in the proliferation of the local area computers networks (LANs). Next have come the departmental LAN, connecting whole departments, and the company. In parallel, a more widespread revolution has taken place; that of the Internet. The interconnection of company or institution LANs to each other via a special wide-area computer network using special communication protocols has meant that a concatenated network of networks can be formed. This is called an *Internet* if they all support a common Internet protocol at the network layer of the Open Systems Interconnection (OSI) reference model.

Another development in the public domain has been the adoption of network access standards by world standards organizations for the use of national and international packet-switched public data networks (PSPDN). An example to this is the CCITT Recommendations on the X-series of protocols (eg. X.25 and X.75). Parallel to this, other standards organizations, such as the International Standards Organization (ISO) and the Institute of Electrical and Electronic Engineers (IEEE) in the United States, the British Standards Institute (BSI) in the UK and AFNOR in Germany, have been active in the definition and acceptance of other data communications and networking standards. One of the most important sets of standards is the ISO's Open System Interconnection (OSI) standard on computer communications. OSI standards cover only the data applications. However, work in the telecommunications world is converging to that of the computer world. Therefore, standards covering the overlapping area need to be defined. Indeed, today more than ever before, there is a need to provide standards for multi-service and multi-media networking.

Today's networking covers a wide spectrum/ it involves computer-to-computer as well as computer-to-other-digital-device communications like disc storage devices, tape drivers, printers and even industrial machinery. To cope with specific applications, Technical and Office Protocol standards (TOP) and Manufacturing Automation Protocols (MAPs) have been developed in some sectors of the industry.

Computer communication is necessitated by various needs. Some of these are:

- *To share data and information* Inaccessible data is not useful. The ability to share information and exchange data is vital in today's work environment.
- *To share resources* Expensive and vital resources can be shared between users and computing machinery. Examples are the expensive supercomputers, laser input or output devices, line printers, tapes and disks.
- *To increase reliability and extend services* A distributed system can be made to have increased reliability over a single resource as a result of the replication and extended services.
- *To control remote devices* Sometimes devices that need to be controlled by computers can be situated at long distances from the control centre. In this case a computer network may be installed with the appropriate hardware and software. Data collection and transmission is also required.

  Many computer applications are now available using computer communications and networks. These include electronic mail, remote login, file access, file transfer, remote windowing, remote database access, computer conferencing and computer-aided telephony.

1.3 GLOBAL IT VILLAGE

From the preceding sections, it is clear that the current trend in the computer communications revolution is towards global access to information and exchange of data using existing and future applications. This will soon create a 'global IT village', where the wonders of information technology (IT) will be exploited to its fullest. We will be able to use computers to access huge data banks, transfer information through networks at the speed of light, and request multi-media conferencing including voice, data and video components. We will have access to ever more information through document delivery services and will be able to do our banking, shopping, education and working from home. Access to these services from virtually anywhere will be possible using mobile voice and data as well as mobile LAN/MAN services. Storage, retrieval and processing of data, voice and video messages located on special servers will be possible. The world is going to shrink in size still further in terms of global communications and information dissemination.

All this and more heralded by the latest developments from the technological frontiers. However, first we must come down to search and investigate how we could utilize the flexibility and opportunities provided by the emerging ISDN. One of the requirements is to port existing services on to the ISDN. In the case of computer applications, an urgent need is to port a multitude of services (like LAN interconnection, remote windowing and networked file systems) that are by now so accustomed (and are taken so much for granted) by so many computer and LAN users to the ISDN.

1.4 INTERNETWORKING IN THE ISDN AREA

1.4.1 The Background

The first computer systems consisted of a mainframe forming the central hub of a star network serving user terminals and other peripheral equipment. The communications between a central computer and its peripherals formed the first data communications as such. These islands of computing systems were soon found to be wanting in many respects. The need for computer communication arose mainly from the need to share information and resources. This led to the formation of local computer networks. However, the need for inter-site, inter-city, inter-country and global data communications has resulted in the establishment of private, national and international computer networks. Examples of this are the networks of individual banks or companies (eg. IBM's VNET and DEC's Easynet), national research and educational networks (eg. ARPANET in the United States), JANET in the United Kingdom, European Academic Research Network (EARN) and RARE.

As computers entered every aspect of human, economic, production and administrative activity, so too did the industrial, commercial, financial, educational, governmental, health and even recreational applications of computing become widespread. As the number of computer networks at the local or wider areas increased the need for interconnectivity and interworking between different networks is becoming more acute.

For data communications between computers and a set of procedures, rules and conventions defining different activities needs to be established. These are called the communication *protocols.* Most networks are designed as a series of *layers* or *levels*, each one built upon its predecessor. This provides structured designed with reduced complexity. A set of layers and protocols is called the network *architecture.* The International Standards Organization has been working towards the establishment of worldwide standards collectively known as the Open System Interconnection (OSI) standards, based on the principle of layering. This principle helps the structured design of computer networks. The OSI reference model (OSI-RM), defines seven layers and is used as a guide for all network architectures conforming to the open system principles. Open Systems are systems that are open for communication with other systems confirming to the same principles.

In the OSI-RM, each system is decomposed functionally into a set of subsystems and is represented pictorially in a vertical sequence. Vertically adjacent subsystems communicate through their common interfaces, while *peer* subsystems collectively from a layer in the architecture. Each layer provides a set of well-defined services to the layer above, by adding its own functions to the services provided by the layer below. The layers of the model are partitioned as follows:

- *1: Physical layer* Achieves the transmission of row data bits over a communication channel (medium).
- *2: Data link layer* Converts the row transmission facility into a line that appears free of frames and delimiting them. This layer may also include access control to the medium, error detection and correction.
- *3: Network layer* Performs the routing and switching of data between any two systems across multiple data links and subnets.
- *4: Transport layer* Operates on an end-to-end basis achieving the necessary quality of service for the exchange of data between two end systems. May include end-to-end error recovery and flow control.
- *5: Session layer* Allows users on different machines to establish sessions between them, and hence establishes and messages communication dialogue between processes.
- *6: Presentation layer* Manages and transforms the syntax of structured data being exchanged. Is also concerned with the semantics of the information transmitted.
- 7: *Application layer* Deals with the information exchange between end-system application processes and defines the messages that may be exchanged.

The above layering was created according to the original design principles used in the construction of the ISO model. According to this:

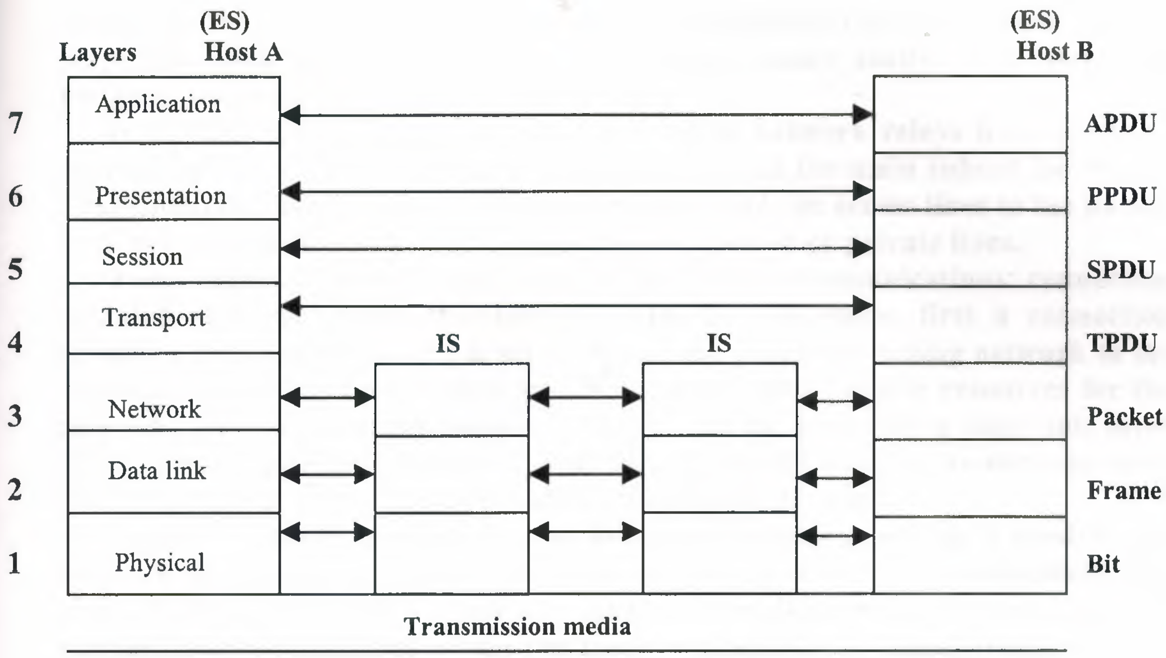


Figure 1.1 The OSI reference model.

1. Different levels of abstraction are placed in separate layers.
2. Similar functions are grouped together within a single layer with each layer performing a well defined function.
3. The function of each layer is chosen so as to be amenable to the definition of a standard protocol.
4. Minimization of information flow across interfaces in a primary goal in drawing the layer boundaries.

Although the majority of network architectures widely in use are based on the principles of layering, most do not fit the OSI model exactly in their allocation of layers and protocols used. Examples of these are the IBM's SNA (Meijer 1987), DECnet and DARPA Internet (Quarterman and Hoskins 1986), to name but a few. Conversely, some new network architectures, such as the MAP (General Motors 1988; O'Prey 1986) and TOP (Boeing 1988), have adopted the OSI-RM for their architecture and hence form 'open' networks. Open networks use internationally standardized procedures for communications rather than local or proprietary ones.

1.4.2 Technological Base for Internetworking

In computer communication three basic types of switching are used: packet, circuit or message switching. Most data networking is packet-based since, whatever the underlying switching or transmission mode used, some form of data framing is used. A *data frame* may be considered to be the smallest unit by which data transfer between networked elements is achieved, since data bits are grouped into frames before transmission. The packet size may be greater or smaller than the frame size. Packetization of data into units makes the fragmentation, transmission and reconstitution of the original data easier in that each unit may be accounted for by the communication protocol used. Hence, error detection and correction can be handled more easily. Also, flow and congestion control can be more manageable.

In packet-switched (PS) networking, hosts or network relays have single or multiple-access ports connected to packet switches in the main subnet (or WAN). Transmission links between the packet switches and the access lines to the packet switches are usually permanently connected via leased or private lines.

Two modes of operation are possible in PS data communications: connection oriented (CO) and connectionless (CL). In the CO mode, first a connection termed a virtual circuit (VC) is set up across the packet-switching network to the destination. All the intermediate switching modes must reverse resources for the new VC. Before the establishment of a VC can be achieved, a data link layer (DLL) connection must be established. This is usually done at the start-up time. Network layer (NL) packets are then transferred in DL frames.

In the CL mode of operation, no previous connection set-up is needed and network layer packets are individually routed to any one of the transmission lines over which a data link connection exists to the destination. Transmission of network packets next hop along the route is achieved by transporting them in DLL frames.

Table 1.1 Access Methods

Transmission	Access	Mode
Analogue	Acoustic coupler	Switched
	Modem	Switched/leased line
Digital	Direct	Switched/leased line
	Integrated access	Switched/permanent/semi-permanent

In circuit-switched (CS) networking, network access is achieved only after a circuit has been set up to a destination, through the intermediate network. If a packet service is required, then, once the physical layer connection is achieved, a data link connection needs to be established between the two ends. The intermediate network has to knowledge of the framing or packetization, acting solely as a 'bit pipe'. Network packets can then be forwarded between the two communicating ends. In the CS mode of networking, the delays in the establishment of physical circuits and DLL links can be time-consuming, preventing fast, on-demand set-up and removal of connections. Tables 1.1 and 1.2 show the access modes in the analogue and digital networks and the differences in the packet and circuit-switched networking, respectively.

Table 1.2 Comparison of packet-switching and circuit-switching networking

PS	CS
No 'circuit' needs to be set up (CL) VC must be set up (CO)	Long circuit set-up delays
Point-to-point, multicast, broadcast	Point-to-point
Longer queuing delays may be encountered and packet-switching delays at every switching node *en route* suffered	Once the circuit is set-up, transmission and propagation delays are the most important delays encountered
Multiplexing is inherent	Multiplexing is more convoluted

In the ISDN era, fast circuit switching will be possible with end-to-end circuit set-up times of less than 1 s within a country and using terrestrial links. (Longer distances may take up to 4.5 s on average.) Hence, with the ISDN technology, on-demand circuit establishment and removal will be possible. This will lead to a novel way of a operation using the circuit mode bearer services where the number of channels established to a destination can be varied dynamically as a function of traffic. If traffic to a destination increases, causing the number of packets in a transmission queue to build up, additional channels can be established to the same destination in order to reduce packet delays. Similarly, if the traffic to a destination decreases one or more channels to that destination can be disconnected, minimizing costs.

Table 1.3 Comparison of existing networks and ISDN

Features	Existing networks	ISDN
Access	Separate network access to telephone, CSPDN, PSPDN, telex and teletext network	Integrated digital access to bearer and teleservices
Signalling	Different in-band signalling for each network	Out-band common channel signal for access to all ISDN services
Channels	CSPDN: separate 'lines' needed PSPDN: up to 1024 VCs on each 'line' is available	Multiple channel each of which can be used for access different service & destination concurrently
User protocols	CSPDN: usually X.21 at the PL PSPDN: X.25, LAP-B	CS-BS: ISDN protocol I.431 (PL); PS-BS: X.25, LAP-B; APM-BS: frame relay/switching; (LAP-D)
Switching Speed (s)	PSTN: up to 30s CSPDN: < 1 s PSPDN: fast VC set-up times	<1s (typical 500 ms) end-end (terrestrial short distances); worst-case terrestrial: 4.5 s (mean)
Transmission Speed (s)	Switched: up to 64 kbps Non-switched: 64 kbps or higher	Switched or non-switched: 64 kbps or higher
Availability	Separate subscription to services is needed, necessitating multiple sockets	Can use existing telephone network wiring giving widest coverage single access point to all services

Another incentive for disconnecting unused connections is the tariff. Current pricing policy of postal, telephone and telegraph companies (PTTs) mean that ISDN CS calls cost the same as telephone calls (or a simple multiple). This means that, after the basic connection charges (if any), the charging for usage will be based on distance, duration and the time of day. By contrast, packet-switched networks charge on the volume of data sent in each direction, plus connect time. Therefore, ISDN CS service will cost money even if the line is not fully utilized, as the capacity must be allocated to circuits even if they do not carry traffic.

1.4.3 Using the ISDN

ISDN services At the user-network interface (UNI), the ISDN provides access to both *bearer services and teleservices.* Bearer services provide only the basic communications services (bit transmission), whereas the teleservices provide access to value added services and other network services like telephony, e-mail, voice-mail and telex. Any of these services can be requested through the use of ISDN signalling protocols on the D-channel. The bearer services further

subdivide into circuit, packet and frame mode bearer services, providing circuit-, packet- and frame-based channels to the user respectively.

Access types **Two types of access are defined in the ISDN: basic-rate access (BRA) and primary-rate access (PRA). The basic-rate access provides two user channels at 64 kbps and a signalling channel at 16 kbps on a twisted pair to the user. The primary-rate access provides 23 or 30 user (B-) channels and 1 signalling (D-) channel, each with a transmission rate of 64 kbps. Higher-rate channels, eg. H0 channel, are also available. A coaxial cable is used for each direction of transmission.**

Impact of ISDN services **Two types of interworking are envisaged between data and telecommunications equipment and the ISDN:**

1. **The use of ISDN as a transparent network to interconnect existing data networks by using only the bearer services of ISDN: this leads to the data-communications-oriented (DCO) interworking.**
2. **The use of ISDN teleservices and innate applications: this results in the telecommunications-oriented (TCO) interworking.**

The differences between these two types of interworking result in different complexity requirements from the network attachment units. In DCO interworking, only the bearer service access and control parts of signalling protocols are needed. On the user channels, two options exist: the use of ISDN data mode access protocols (currently only X.25 is supported), or of user-defined protocol stacks over CCITT defined physical layer protocols. In the TCO interworking, additional signalling protocol features are needed for access and control of teleservice sessions. On the user channels, CCITT-defined protocols must be used.

Data communications over ISDN **The CCITT Recommendations (Recs.) of the I-series describe four types of data services: circuit, packet and frame modes and the support of data terminal equipment (DTE). The CS data service provides only a 'bit pipe' down which data bytes can be sent, and imposes no user protocol restrictions apart from the ISDN Physical Layer Protocol (CCITT 1988b, 1988c). The packet mode data service specifies three types of services: user signalling, connectionless and virtual circuit services. The frame mode service (also called ~~the~~ 'additional packet mode bearer service' APMPS) supports frame relaying, frame switching and X.25 protocol operation. Support of DTEs provides access to the X-and V-series DTEs and packet mode DTEs (eg. X.25 and ISDN-compatible terminals). Packet mode DTEs can provide access to PSPDN services and can utilize the ISDN VC services. While the public switched packet data network (PSPDN) access defines the lower three layers of protocols, the APMBS defines only the lower two layers of protocols when frame mode services are used and offers the lower three layers when X.25-based services are based.**

1.4.4 The LAN-ISDN Interconnection

Users connected to the existing data communication networks may want to interconnect their local networks or individual workstations through the ISDN, or may want to interwork with the ISDN. The first case implies that the users want to use only the bearer service of the ISDN as a 'raw carrier'; the second case implies that the users want to access the teleservices as well as the bearer services. These two options present different problems for interconnection and necessitate the DCO and TCO interworking, respectively.

Most of today's local data networks can be modelled as multiple hosts, workstations, personal computers and terminals interconnected through a local area network. The need to interconnect these LANs with other LANs at remote sites has already been discussed. In the ISDN era, DCO or TCO interworking through public or private ISDNs is possible.

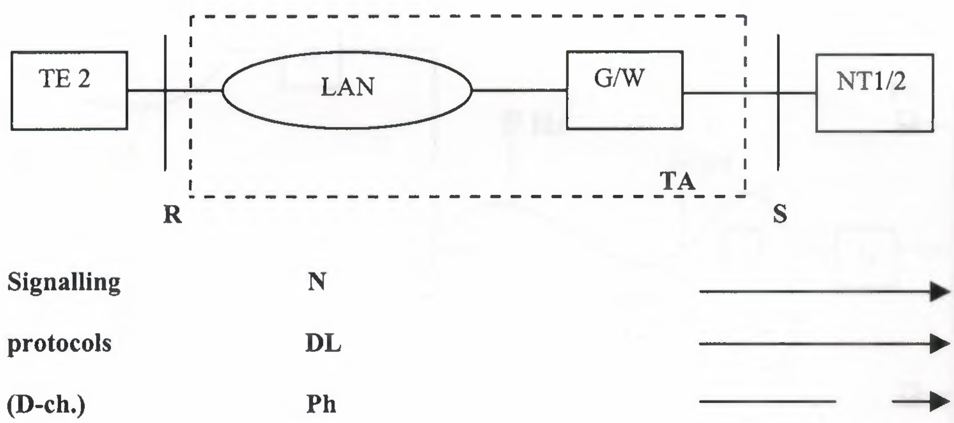


Figure 1.2 DCO internetworking: LAN and gateway (G/W) as TA

A private ISDN network may be a local integrated services private branch exchange (ISPBX) or multiple ISPBXs interconnected by leased lines to which multiple LANs can be attached. Hence, the two scenarios of interworking needs to be investigated. Figure 1.2 shows the DCO interworking scenario, while Fig. 1.3 shows the TCO interworking scenario for LAN-ISDN interconnection. The main difference between the two is in the scope of the signalling protocols. In the DCO interworking, the ISDN is made transparent to the LAN hosts. It is this mode that applies to the LAN-ISDN-LAN working since traditional LANs are data-communications-oriented. In the TCO interworking, the signalling protocols need to be terminated in the LAN hosts, making the LAN transparent to ISDN-compatible terminals and ISDN services.

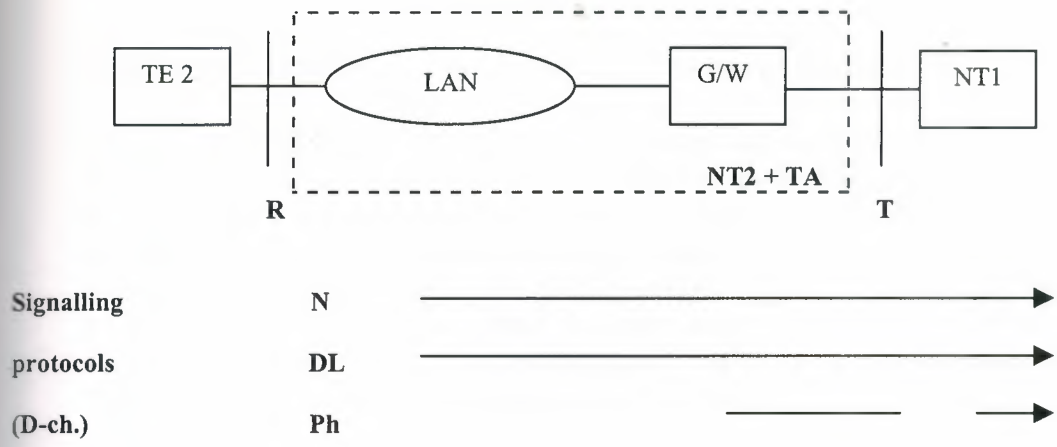
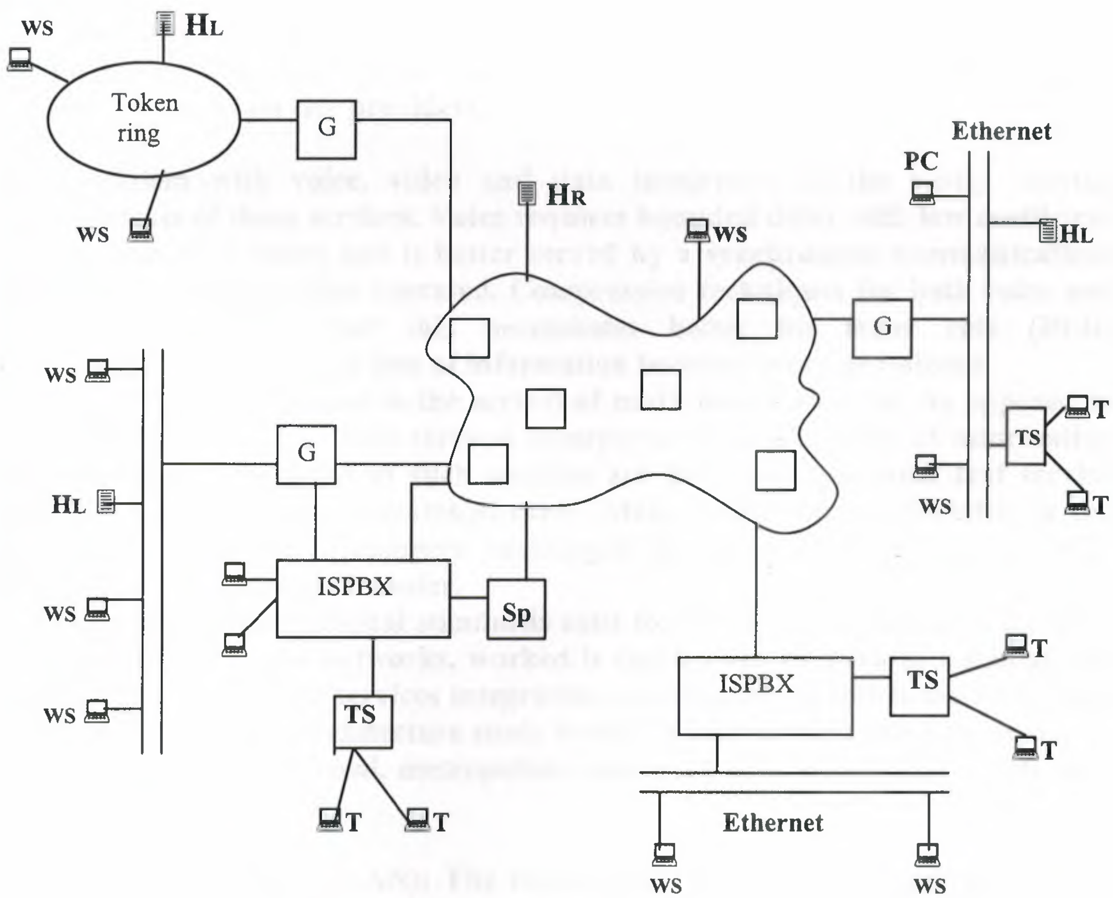


Figure 1.3 TCO interworking: LAN and gateway (G/W) as a NT1 + TA



Sp: primary-rate server; T: terminal; TS: terminal server; WS: workstation
PC: personal computer; HL: local host; HR: remote host

Figure 1.4 An interworking reference configuration.

1.4.5 Future Trends in Network Services and Networking

The common underlying trend in the network services prevision, whether at the local, metropolitan or wide area, the public or private domain, is the integration of the services. This implies the provision of voice, data and video services within an all-encompassing network resulting in the integration of services as seen by users at the access node. In earlier stages, different networks may do the provision of actual services, but the user need not be aware of this. ISDN is following this route as it id an evolutionary rather than a revolutionary network. The concept of service integration is driven by two main objectives:

1. Provision of a simplified access to a multitude of services through a single interface integration of services from the user perspective.
2. Provision of network integration such that implementation and operational costs can be reduced owing to economies of scale. For example, it is expected that revenues from the telecommunications users, making the provision of data services within the ISDN cost-effective for PTTs will meet the cost of provision of data services in ISDN. This is the integration of services from the perspective of service providers.

The problem with voice, video and data integration is the vastly varying requirements of these services. Voice requires bounded delay with low coefficient of variation (low jitter) and is better served by a synchronous communications channel. Some loss is also tolerated. Compression techniques for both voice and video are available, but this necessitates better bit error rate (BER) communications, since any loss of information becomes more prominent.

A parallel development is the arrival of multi-media services. As opposed to the integrated services, these services incorporate multiple forms of information representation. Examples of such services are the voice-annotated text service and text with bit map diagrams/pictures. Many more will be available in the future. Most of the documents exchanged by these services require huge bandwidth capacities for transfer.

Although no international standards exist for services integration at the local and metropolitan area networks, worked is carried out by various national and regional institutions for services integration. An example of this is the integrated voice and data (IVD) architecture study by the IEEE. Services integration can be further studied at the local, metropolitan and wide area networking levels as it affects all three of them.

Local area networking (LAN); The controversy of LAN versus private branch exchange (PABX or PBX) in the provision of local area networking has been going on for years. It seems that both are going through an evolution and will coexist together, filling similar roles and interworking. Integrated-services LANs (ISLANs or ISLNs), as well as integrated-services PABXs (ISPBXs), have so far concentrated on the integration of voice and data. Traditionally, LANs have been better at dealing with data, while the PABXs are better at dealing with voice.

Metropolitan area networking (MAN); Integrated services MAN is now becoming feasible. An example is the distributed queue dual bus (DQDB), which is a draft standard in the form of IEEE 802.6. These will provide access to synchronous and asynchronous services and hence provide voice, video and data integration at the metropolitan area.

Wide area networking (WAN); ISDN is a new concept that is providing a useful framework for the development of future telecommunications network and services. The natural extension to ISDN is the broadband ISDN (B-ISDN). CCITT Rec. I.121 specifies the asynchronous transfer mode (ATM) as the technology on which the B-ISDN will be based.

ATM technique will provide synchronous as well as asynchronous services. Access to ATM networks will be through fibre optic links and so far 150 Mbps and 600 Mbps access rates have been defined.

1.4.6 New Opportunities

The ISDN standardization has come of age; the first CCITT Recommendations relating to ISDN, the Recommendations of the I-series, were released in 1984 as a result of the first study period 1980-4 by the Study Committee XVIII. The second study period of 1984-8 culminated in the updating and extension of I-series Recommendations.

The approach adopted here recognizes the need for transferring current data applications to ISDN. Furthermore, new features of ISDN are attractive in that they promise much more flexible usage of resources. There is a gap in the existing networking and the defined futures of ISDN; hence I identify the following needs in data applications of circuit-switched ISDN:

- Managing the multiple channel resource effectively,
- Converting from traditional data in-band signalling to ISDN out-band signalling,
- Defining a software interface between data applications and the ISDN signalling protocols.

1.5 DYNAMIC CHANNEL MANAGEMENT

1.5.1 Definition

Resource management is a well-known concept in data communications and computer networks. It enables users or applications to manage the scarce and expensive resources in communications, such as the processing capacity of nodes, the bandwidth capacity of transmission lines and buffers at the nodes. All these resources and their management are also valid in communications over the ISDN.

As pointed out above, one of the most important features of ISDN is the provision of multiple channels at the user-network access interface (UNI) and its flexible control by the use of common channel signalling (CCS). It is these two features that are concentrated. The following distinctions are made between them in the ISDN context:

- ***Bandwidth allocation*** **Describes *how* the total bandwidth of communications facility is distributed (allocated) among different contending users. It is assumed that the bandwidth applies to the whole communications interface.**
- ***Bandwidth management*** **Describes the means (or procedures) by which the bandwidth of a communications facility is managed. It works in collaboration with and within the confines of, a bandwidth allocation policy. At the microscopic level, it is assumed to apply to the bandwidth of a single channel of variable capacity.**
- ***Channel assignment and transmission control*** **Describes the procedures for channel assignment to incoming requests and de-assignment for cost effectiveness, transmission control for data packet forwarding, and receiving and initiating signalling for link establishment and removal.**
- ***Channel management*** **Applies to multiple distinct channels a communications facility and describes their management. A channel management strategy comprises policies regarding bandwidth allocation, bandwidth management and channel assignment and transmission control operations.**

A bandwidth allocation policy provides access control to new customers (e.g. users, packets, protocol detail units (PDUs) for sharing the interface bandwidth. It also regulates changes in the bandwidth allocation to existing channels. For example, a class of users may not be allowed to access more than a certain share in the link bandwidth.

A bandwidth management policy specifies the method and the algorithm by which the bandwidth of a channel is to be increased or decreased. For example, for a given facility, a hysteresis threshold control may be designed for the service capacity (or bandwidth) control of its channels.

A channel assignment and transmission control policy specifies the conditions, i.e. when to set up a new channel and when to remove an existing one. For example, in the case of a connectionless network layer packet, the policy may specify the setting up to a new channel to a hitherto unconnected destination. Similarly, a channel may be totally removed if it is not used for τ seconds. It will deal with the packet forwarding to channel queues. Initiation of D-channel signalling activity and initiation and ending of packet transmission on any channel mapping and free-time slot hunting are also carried out under this policy directive.

Channel management encompasses policies for the above activities. Hence, it includes policies on how a channel or a group of channels is to be used, e.g. separately or aggregated together to form a larger capacity channel (i.e. a *superchannel*); when to set up new links and remove existing links; *when* and *how* to increase/decrease the existing channel capacity to a destination. It may also include a scheduling (routing) policy if multiple channels with individual queue exist to a given destination. A congestion control policy may also be implemented (e.g. throw packets away if the transmission queue gets longer than a certain level). Finally, it may involve traffic monitoring and statistics gathering operations.

1.5.2 Framework

I propose an architecture for the implementation of channel management and discuss how the functionality's described above are met, I specify the short of management information needed in order to implement these functionality's in the case of a network layer relay. To this effect, I specify the parameters of interest, the decision metrics and their measurement and the decision mechanisms involved. Queuing and simulation models of this architecture are also developed. It is shown that after suitable simplifications the resulting nodal queuing system of the network layer relay can be solved analytically using and approximate technique, and that it can be used in the validation of the simulation model.

The channel management architecture (CMA) is extended to include superchannel. This architecture is then compared with the single-channel CMA. Also, the superchannel formation in the primary-rate ISDN is investigated and several modifications and extensions to the existing CCITT protocols for the D-channel are proposed. In the section on policy performance, a simulation model is then used to compare the performance of different dynamic bandwidth management policies.

Hence, by the same token, the need for a bandwidth allocation policy that is used in performance models is reduced to: *assign a new bandwidth unit a bandwidth management request as long as free capacity is available.* This is acceptable, as no connection is assumed to exist in the simplified model used. Therefore, the interaction of multiple channels at the interface using same or different bandwidth management policies is not studied.

1.5.3 Dynamic Channel Management Architecture (DCMA)

In order to achieve dynamic channel management, a generalized approach is needed to identify the necessary modules and units of software architecture such that it could be applied to different types of ISDN user-network interfaces. Such architecture would also include events and messages that will be encountered at a UNI.

An access node interconnected to ISDN will have a physical interface as well as a software interface. The distribution of functionality within the access node can be various for different nodal architectures. In general, this interface can be represented as shown in Fig. 1.5. The dynamic channel management (DCM) module is shown to have interaction with the interface access and control process. This module needs to have regular access to the status of the particular ISDN interface it is supposed to manage. Figure 1.6 shows the relationship between the DCM and associated parameters that affect it.

The DCM monitors the ISDN interface status continuously by a report sent to it by each relevant event across the interface as well as within the ISDN interface. These events include:

- *Packet sent* A packet has been sent to the interface to be queued to a channel queue.
- *Packet received* A packet has arrived on any one of the channels.
- *Set-up channel* Set up a new channel to the destination or set-up requested by remote destination.

- *Disconnect channel* **Disconnect a channel to the destination or disconnect requested by remote destination.**
- *Vary channel bandwidth* **Increase/decrease channel bandwidth to a given destination.**

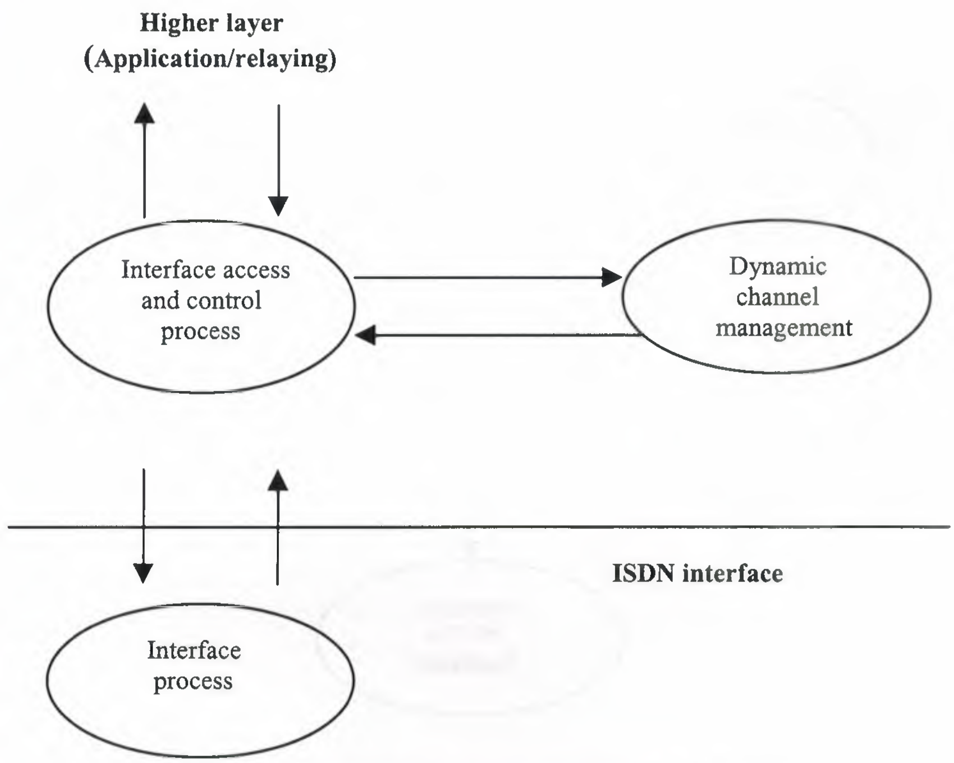


Figure 1.5 A simple schematic of the DCMA

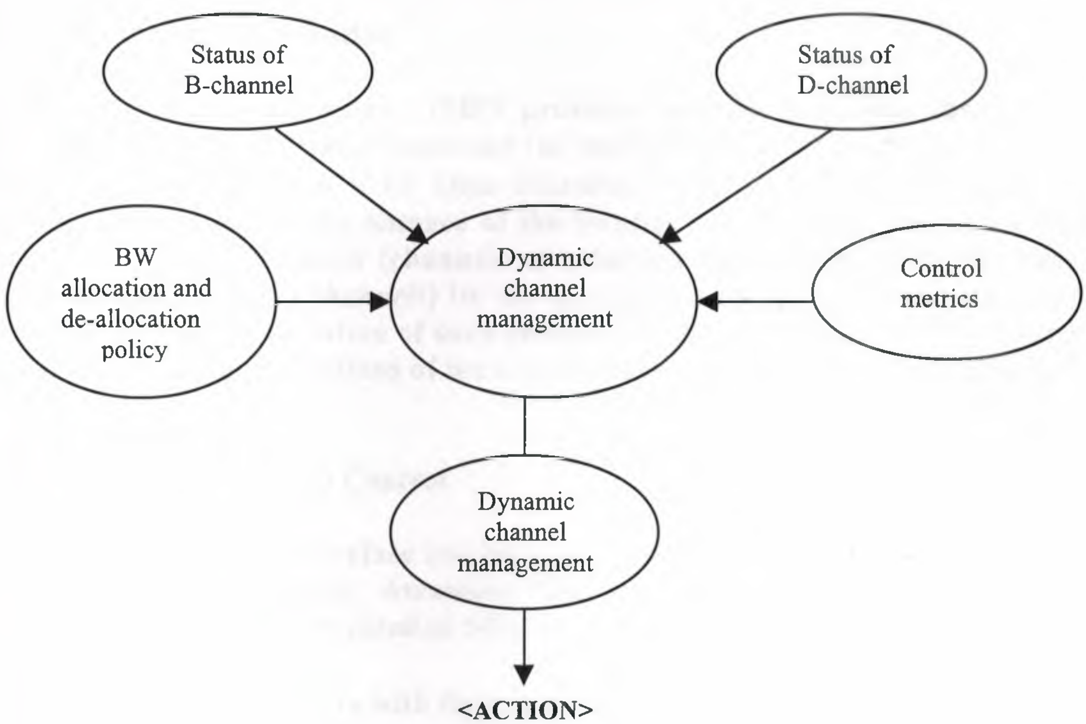


Figure 1.6 Relationship between DCM and associated parameters.

Each of these events generates a change in the status of the interface and hence causes and update of status information by the DCM module. Figure 1.7 shows the relationship of the DCM and event sequence within the architecture. As a result of this event, an action command may be generated by the DCM module and sent to the interface process.

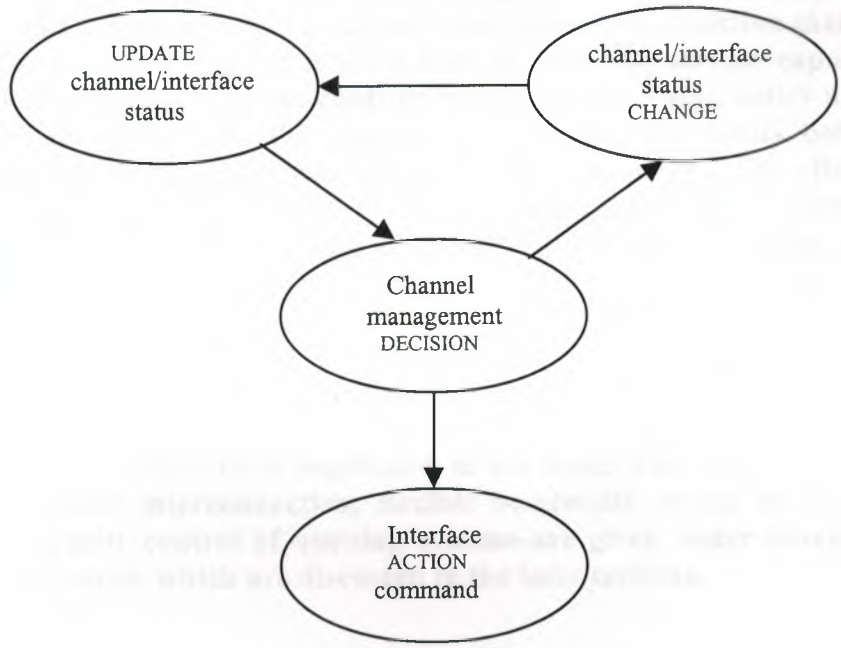


Figure 1.7 DCM events sequence.

1.5.4 Superchannel Formation

At the user-network interface, ISDN provides multiple fixed-size channels of possibly differing bandwidth capacities (in multiplies of 64 kbps). Two problems exist: the provision of *n* × 64 kbps channels, and the dynamic variation of channel bandwidth. In the absence of the former, an alternative method is the formation of superchannels (channels of arbitrary bandwidth) using the basic channel types (*physical channels*) by the *aggregation* method. Identification and dynamic bandwidth variation of such channels at the UNI needs to be provided. The CCITT Recommendations of the I-series have no provisions for the above.

1.5.5 Dynamic Bandwidth Control

A multi-channel ISDN interface can be organized in a variety of ways, leading to different queuing models. Assuming that no quality-of-service queuing is incorporated, these can be listed as follows:

1. **Multiple parallel servers with their individual queues: $n \times A/B/1$**
2. **Multiple parallel server groups, each with a single queue: $m \times A/B/c$**
3. **Single server groups, each with a single queue but variable service rate:**

$k \times$ A/B/1

4. A mixture of the above

In queuing models 1 and 2, the variation in the logical channel (a channel formed by the association of several physical channels) bandwidth to a given destination is achieved by the addition or delation of server and their queues, or just servers, respectively. The type 3 queuing model applied to single channel assumes the existence of a Superchannel structure. One immediate question that arises from the formation of a superchannel is how to vary the service capacity (rate or bandwidth). Service capacity control, based on a threshold policy applied to the number of customers in the queue or system, has previously been proposed. Here, the type of the threshold policy becomes important. The effects of policy decisions need to be evaluated. Queue-based multi-level and bi-level hysteresis threshold control policies, as well as queue- and rate-based policies, are compare and contrasted.

1.6 REWIEV OF RELATED WORK

A wide area or research is implicated in the topic. The areas of direct concern the LAN-ISDN interconnection, flexible bandwidth access in ISDNs and the service capacity control of queuing systems-are given wider coverage then the peripheral issues, which are discussed in the later sections.

1.6.1 LAN-ISDN Interconnection

The issue of LAN-ISDN interconnection is relatively new, as the emergence of ISDN standards began with CCITT Recommendations of the I-series in 1984. Here, a technical as well as historical review of the LAN-ISDN interconnection issue is presented through the published work of interest researchers.

One of the earliest investigations into the position of LANs in an ISDN world was carried out by Davies (1985), who saw the absence of structured data transmission services in the then emergent ISDN as one of the greatest weaknesses for LAN-ISDN interconnection. The major issues in this interconnection have been identified as data delimiting, message size and flow control. The synergies between the LAN access modes and the passive bus in the basic-rate ISDN are pointed out. In the case of integrated services local networks (ISLNs), offering data, voice and video services operating at very high bit rates. Two major issues in LAN-ISDN internetworking are the necessity for higher bandwidth in the wide area network, and the need for the ISDN to be able to transmit (and possibly switch) the packetized voice.

The main reasons for wanting a LAN-ISDN interconnection are listed by DePrycker (1985). The major problems when connecting a LAN to an ISDN have been identified as address inconsistency, frame structure incompatibility, and sub-addressing within a LAN. The address inconsistency is attributed to the use of layer 2 addresses by LANs and layer 3 addresses by the ISDN. Three solutions are offered for address inconsistency. The first involves the level 3 address generation within the LAN by mapping between level 2 and 3 addresses. The second solution involves the use of an adaptor between the LAN and the S-interface to the external mapping. One way of solving the frame structure incompatibility is to use a gateway to remote all frame structure elements. The solution proposed for the sub-addressing within the LAN is to have the LAN gateway interpret the first in-slot-received data and direct the following data to an appropriate address within the LAN. DePrycker also considers the use of X.25 PLP (packet-level protocol) over circuit-switched B-channels and the LAN, where the LAN-ISDN gateway performs the flow control, connection set-up and release. A mapping is to be performed between the X.25 virtual circuit (VC) and the B-channel circuit connection. The disadvantage of this method include the time consuming VC set-up and the X.25 layer 3 error correction mechanism, considered and overkill when used over a LAN which has good bit error rate (BER). When connectionless protocols are used over the LAN, the use of X.25 VC 'tunnel' is advocated, where the datagrams are ported using X.25 data packets.

The author has studied interconnection of LAN-ISLAN and ISDN in general terms. Here, the problems of interconnecting a packet-switched sub-network to the circuit-switched ISDN are investigated and the CO and CL protocol working over the ISDN B-channels are studied. The circuit set-up and removal operation is defined in both cases. When a CL protocol is used, the circuit is to be established with the first datagram to a given ISDN destination and removed after a connection *time-out* mechanism when no further packets arrive to the gateway for that destination. A further proposal includes the use of multi-link procedures over multiple B-channels giving point-to-point connections of larger bandwidth's. The possibility of using the *fast-select* facility of X.25 for transporting datagrams over semi-permanent ISDN lines is also suggested.

The definition of new ISDN packet modes using full out-of-band signalling and level 2 multiplexing are reported. The use of a terminal adaptor (TA) acting as a gateway between the LAN and ISDN for the first two solutions is advocated. The issue of connecting an ISDN terminal to a LAN is also raised. The suggested solution uses only existing data terminals over LANs and uses a PABX for ISDN-type terminal connection.

The issue of LAN-ISDN interconnection, and the type of protocols to be used over these connections, the implications for gateway design has been raised in the UCL Annual Report of research activities in October (1986). In this report, the possibility of using integrated services and multi-media application types, both at the local and wide area network domains, and the impact of this possibility on communication protocols were raised. Building of specific connection-oriented and connectionless mode network layer relays between LANs and ISDN are the topic of papers by Deniz (1987a, 1987b) and *Knight et al.* (1987). In the first of these the 'call management', and in the latter the 'circuit management' phrases are coined Deniz and Matthewson |(1986).

Tennenhouse *et al.* (1987) reported that the Unison Project would study the practical aspects of multi-media services and distributed computing across multiple LAN sites interconnected using the standard (G.732) 2.048 Mbps PCM links. The pilot network provided circuit switching through ISDN switches and, by then, supported the CCITT layer 2 (Q.921) and layer 3 (Q.931)-signalling protocols.

Berera and Jardin (1987) describe a method of interconnecting an Ethernet LAN to a PABX using the PRISDN interface given by the ECMA 104 standard (ECMA 1985a) (based on 1.431/G.703/G.732 with additional features such as maintenance loops). The first version of their equipment is an Ethernet-based terminal server. It is actually a LAN-PABX gateway that supports the basic functionality required for the support of low-speed asynchronous terminal equipment (TE) connections. It also provides 64 kbps clear connectivity. The D-channel protocol versions of ECMA 105 (1990a)/Q/921 (layer 2) and ECMA 106 (ECMA 1985b) (layer 3) are developed for signalling purposes. Cooper (1987) discusses the influence of ISDNs on local area networks and shows how the LAN must adapt to the network management requirements of the ISDN. It further considers the merging of LANs and PABXs developed to ISDN standards. The LAN-ISDN gateway is said to provide NT2 functions, to correct electrical termination to the ISDN, to transmit and receive information at ISDN rates and with ISDN protocols, and to provide network management functions. The second requirement implies that (asynchronous) packet-oriented to (synchronous) circuit-oriented conversion and buffering facilities must be provided at the gateway. The third requirement implies that the LAN or the gateway is equipped with a comprehensive network manager capable of monitoring the LAN traffic, determining the acceptance of incoming calls, establishing call priorities and registering details of the properties of all devices connected to the LAN. Henriet (1987) considers the integration of LAN, PABX and the ISDN equipment and services and its benefits to users, but offers no technical proposals.

A study by Kirstein *et al.* (1989) pointed out the possible impact of ISDN on the Joint Academic Network (JANET) in the UK. According to this, the benefits of the ISDN in an academic network environment for the period investigated will be the use of ISDN lines as catastrophe backup, on-demand extra capacity, direct capacity and remote terminal access lines.

Deniz and Knight (1989a, 1989b) define the channel management is LAN-ISDN gateways (relays) and propose a *status table* approach to channel management. The use of instantaneous or averaged queue length and/or traffic arrival rate as the control metric for channel service capacity is also proposed. The factors affecting the choice of queuing strategies in LAN-ISDN gateways are also studied. A design for a simple Ethernet-ISDN connectionless network layer relay incorporating channel management is presented. In the case of CO protocols (e.g. X.25) over ISDN, multiple virtual circuits over a single B-channel and multi-link procedure over multiple B-channels are proposed. For the case of CL protocols over ISDN, the effect of the transport protocol time-outs is also studied.

A bandwidth management system is proposed in Harita and Leslie (1989), and is used to aggregate time slots in a PRISDN interface to give instantaneous bandwidth of size $n \times 64$ kbps. An algorithm running on transmuters and providing an aggregated wide-band circuit (using $n \times$ B-channels of a PRISDN) is presented in a paper by Burren (1989). A fixed-size cell structure is superimposed over this variable bandwidth circuit-switched channel, giving an asynchronous transfer mode (ATM) type performance. Tennenhouse and Leslie (1989) report on the multi-service protocol suite used over this ATM-type network. The protocol suite used in based on the OSI layering principles but does not conform to the OSI in the distribution of layer functions. Two important aspects of this protocol suite are reported: first, the multiplexing is done at the link layer, making application-specific quality-of-service information, available to scheduling components, both in the network and the end systems; second, parallel associations are used to support the out-of-band implementation of many layer functions.

Fan (1989) gives two designs for CO and CL Ethernet-IDA gateways in the absence of an ISDN in the United Kingdom conforming to the CCITT Recommendations during the lifetime of that development. The CO gateway uses X.25 over a single B-channel, where multiple virtual circuits can be supported. The LAN-ISDN internetworking is achieved at reference point R. both of this gateways used simple set-up and removal operations for the single 64 kbps channel available at the BR-IDA interface based on the discussions presented in Deniz and Matthewson (1986), Deniz (1987b) and Knight *et al.* (1987).

INCA (Integrated Network Communications Architecture), a European collaboration under the EC-sponsored ESPRIT-1 (European Strategic Progrmme of Research in Information Technology) programme (1985-8), has looked into the LAN and WAN interconnection issues. However, it has done little work on the LAN-ISDN interconnection issue (Knight and Kirstein 1987). PROOF (Primary Rate ISDN OSI Office), an ESPRIT-2 project started in 1989, intends to bring together the PRISDN and the OSI-based data communications technology to the development of applications in distributed office systems (Knight 1989). The project will develop a high-performance LAN-PRISDN relay conforming to OSI principles and will provide an environment for the porting of X.400 mail, X.500 directory services and multi-media services over a network of concatenated LANs, private ISDNs and PABXs.

The IEEE 802.9 IVD-LAN Interface Working Group considered the LAN-ISDN interconnection via frame relay (IEEE 1988a). the method described is based upon the extended-bridge approach (MAC layer interconnection) and provides transparency while necessitating only minimal modifications to be existing LLC (link layer control protocol used over LANs) and LAP-D (link access protocol over the D-channel) protocols. Walton *et al.* (1991) describes a design for the Ethernet-ISDN connectionless gateway. The ECMA Technical Group 10 (ECMA 1990b) is currently preparing a Technical Report on the LAN-ISDN interworking which will consider the LAN-ISDN as well as LAN-ISDN-LAN interconnection issues using the ISDN bearer services.

1.6.2 Flexible Bandwidth Management in ISDNs

The issue of flexible bandwidth management in ISDNs is a relatively new topic of interest in the research community. The asynchronous transfer mode (ATM) (Minzer 1989) in broadband ISDNs has this feature as a potential facility. However, in the narrowband ISDN it has been an open issue. Provision of $n \times 64$ kbps service facility in ISDNs will of course solve the problem partially; the actual mechanism of dynamic bandwidth variation and the control doctrine to be used also need to be defined. At present, the availability of this service in the immediate future is still doubtful.

The need for channel management in LAN-ISDN interconnections, when the interworking is achieved at the network layer using connectionless mode protocols as to when to open (set-up) and close (disconnect) channels and how to vary the link bandwidth, has been raised in Deniz and Matthewson (1986), Deniz (1987b) and Knight *et al.* (1987). The need for a variable bandwidth communication link in LAN-LAN interconnection is also pointed out by

Tennenhouse *et al.* (1987), when the integration of bulk data transfer, single transection type traffic based on remote operations, voice and other multi-media services are to be integrated over a wide area connectivity. It is assumed that any two LAN sites should establish a link 'at start of day', and keep it until end of the day, varying the link bandwidth during the day as the traffic load between the two sites necessitates. It is also assumed that all the traffic between a pair of sites will be aggregated together at the source site and transported along a shared channel to the destination site.

The characteristic of services requiring multi-slot connections and their impact of ISDN design are described in a paper by Roberts and Van (1987). Roberts and Liang (1988) study the performance of the X.25 multi-link protocol for $n \times 64$ kbps data transmission in ISDN, while Altarah and Motard (1989) and Boltz *et al.* (1989) study the dynamic usage of bandwidth and narrowband ISDN. The latter two papers propose methods for solving the time slot sequence integrity (TSSI) problem in ISDNs.

Deniz and Knight (1989a, 1989b) provide simple bandwidth management architecture for dynamic channel management. Harita and Leslie (1989) present analytic and experimental results for dynamic bandwidth management using threshold control. Barren (1989) describes an algorithm for solving the TSSI problem. The protocol issues in channel management have been the subjects of the paper by Deniz and Knight (1990). The problem of channel management and dynamic bandwidth management formed on queue-based multi-level threshold control policies and their performances have been the subject of a study by Deniz (1991). A recent study by Harita (1991) also investigates dynamic bandwidth management in an ATM overlay over ISDN. A paper by Altarah and Seret (1991) propose a closed-loop control method for rate-based bandwidth management. The rate-based bandwidth control has also been the subject of study by Du and Knight (1991).

1.6.3 Service Capacity Control of Queuing Systems

The classical queuing theory deals mainly with *descriptive* models, where, for specified characteristics of a queuing process, the state probabilities and (excepted value) measures of effectiveness describing the system are obtained. In contrast, the control of queuing systems results in a *prescriptive* model. The prescriptive queuing models also fall into a category of queuing theory referred to as the 'design and control of queues' and they prescribe the optimal course of action to follow (Gross and Harris 1985). The design and control models relate to static and dynamic models, respectively. Static models prescribe a combination of control parameter values, which optimize certain cost functions. The control models are usually involved in determining the optimal control policy. Queuing processes can be characterized by the following six features (Gross and Harris 1985):

1. Arrival pattern
2. Service pattern
3. Queue discipline
4. System capacity
5. Number of service channels

6. **Number of service stages**

The control of a queuing process can therefore be exercised by controlling one or more of these characteristics.

Instantaneous Switching: **The earliest research in this field is attributed to Romani (1957) and Moder and Phillips (1962) and Gross and Harris (1985). Both of this works study the M/M/c queue where the number of servers in the variable under control. In Romain's control doctrine, if the queue builds up to a certain critical value (*threshold*) *N*, additional servers are added as new arrivals come. The maximum number of servers available is assumed to be infinite. Hence the queue never exceeds the value *N*. servers are removed when then finish service and the queue length is zero.**

In the Moder-Phillips (1962) study, a minimum number of servers σ ($\sigma \geq 1$), is always available and the maximum number of servers, *S*, is fixed. The switching-on of additional servers is done in a similar fashion to Romain's model until the limit *S* is reached. The additional servers are removed when they finish serving and the queue falls below another critical value v. These works provides closed-form expressions for the mean queue length and mean number in the system, as well as idle time, mean waiting time and the number of interruptions (server starts).

Yadin and Naor (1967) have generalized the work in this field. In their model, customer arrivals are assumed to be Poisson and service exponential. The service station is capable of providing service at different controllable rates, so the queuing model is essentially M/M/1 with state-dependent service. Alternatively, the number of active servers within the station may be varied in some fashion, producing a variable-capacity service station. It is noted that, if the service capacity increases (set-ups) incur costs, it is desirable to reduce the number of service capacity changes per unit time to a minimum. In such a situation, the optimal doctrine may provide for delay of setting up new (or removing active) servers capacity, leading to a hysteresis phenomenon. The following class of policies is considered. For a sequence of feasible service capacities $\mu = (\mu o, \mu 1, \ldots\ldots\ldots, \mu k, \ldots\ldots\ldots)$ where $\mu(k+1) > \mu k$ and $\mu o = 0$, the two integer sequences R = (*R1, R2,……, Rk,……*), S = (*So, S1,……, Sk,…..*) such that $R(k+1) > Rk$, $S(k+1) > Sk$, $R(k+1) > Sk$, $So = 0$, represent the management doctrine. The policy is stated as follows: 'Whenever system size reaches a value *Rk* (from below) and service capacity equals $\mu(k\text{-}1)$, it is increased to μk; whenever system size drops to *Sk* (from above) and service capacity is $\mu(k+1)$, it is decreased to μk.'

Gebhard (1967) has studied an M/M/1 queuing process using a single hysteresis with bi-level service capacity control. The policies described by Gebhard are special cases of the Yadin-Naor (1967) policies. The control doctrine used prescribes the average service rate depending upon the queue length. A single-threshold model is also analysed for completeness. The terminology used in Genhard's paper to describe the control doctrines is somewhat different from that adopted. In the single- (point) threshold model, the service rate is $\mu 1$, for queue lengths less or equal to $\mu 1$ and $\mu 2$ ($\mu 2 > \mu 1$) for queue lengths greater than *N1*. In the case of the bi-level hysteretic control, the service rate is increased from $\mu 1$ to $\mu 2$ whenever the queue length increases to a

value equal to or greater than and upper control level *N2*. The service rate is reduced from $\mu 2$ to $\mu 1$ whenever the queue length drops to a value *N1* (*N1*< *N2*). For both models, closed-form expressions are obtained for the expected value and variance of the number of customers in the queue. In both cases the steady state is assumed to exist provided the higher of the two service rates, $\rho 1$, is less than unity. Two *figures of merit* for measure of performance of the bi-level hysteretic control doctrine are derived: the rate of switching between service rates, and the proportion of time that the queuing system operates at the higher service rate.

Several cost formulae are superimposed on the queuing systems considered for a comparison of costs when the control doctrine is varied. Although optimum solutions are not derived, it is concluded that the *optimum* control doctrine depends on the cost formula. The total cost formula is assumed to have two components: cost *Cs* associated with service (including switching costs) and cost *Cq* associated with queuing. The single-level control is found not be best for any cost combination, while the hysteretic control is found to be best under certain cost combinations. The results obtained exemplify that cost allocations do not always make the best control option apparent.

Delayed Switching: Yadin and Naor (1963) study an M/G/1 queuing system with a removable service station and introduce delays in both the set-up and close-down operations. Their work considers a single server which could be either present or absent (on/off control). The distribution of the switching delays as well as the service times are assumed to be general. They state that, in the absence of any financial (or other) penalties for the establishment and dismantling of a server, the trivially optimal doctrine is one that eliminates the idle fraction of the server by closing down the server and reopening it at suitable times. However, ordinarily the dismantling of a server leads to penalties arising from: an increase in average queue length and queuing time, set-up/close-down times and costs. Hence, their study looks at the effect of an (0, *R*) doctrine ('this mantle the service station, if 0 customers are in the queue; re-establish the service station, if the queue has accumulated to size *R*') on the above queuing system. Their results include the average queue length and queuing time. A cost structure is also superimposed on the system, and optimization procedures are outlined.

King and Shacham (1986, 1989) before the second server becomes available introduce an M/M/2 system with delay. King (1990) also studies the system with delayed second server. In their study, King and Shacham (1986) look at a system with a single queue and two parallel identical servers were one is available permanently and the other is switched in after a random switching delay (exponentially distributed with mean $1/\varsigma$) whenever the queue length reaches *K*. The second server is switched off (instantaneously) only when the queue length falls back to zero. That is, the control doctrine used is a single hysteresis threshold with bi-level capacity control. A cost structure composed of two factors-cost of waiting time (queuing time plus service time) and cost of dial-up (channel set-up) – is superimposed on the system. According to this, at light system load the switching cost is dominant, and at loads close to capacity, waiting is the dominant factor in cost. The optimum-sold is found to be decreasing function of the offered traffic to the queuing system. The cost function used has a low gradient in the region of the optimum-sold. Therefor, it is concluded that choosing a threshold in the vicinity but not exactly at the optimum threshold will not increase the operating costs excessively. It is also found that for fixed threshold values *K*, the cost increases with load for large *K* but has a minimum for small *K*.

A second model introduced by King and Shacham (1986) deals with two queues with a dial-up channel operating in full duplex mode. Each node that has its buffer occupancy above the threshold dials up a channel. When the first dial-up is complete, the other buffer aborts any ongoing dial-up process and uses the established channel. The channel is released when both buffers are empty. No closed-form results are obtained, but a numerical solution procedure is outlined. Again, it is found that the choice of threshold is not critical as long as the operating threshold is close to the optimal.

Generalization of the above to an M/M/1 system with delayed service capacity increase is carried out by Harita and Leslie (1989). The control doctrines used are single (point) threshold with bi-level capacity control and single hysteresis threshold with bi-level capacity control. The channel switching-on delay is assumed to be exponentially distributed with parameter γ. The switching-off is assumed to be instantaneous. Expressions in Z-transform for the mean and variance of the number of customers in the queue are obtained.

1.6.4 Other Issues

Standardization bodies active in ISDN: The work on ISDN can be reviewed under three groups: the national and international organizations responsible from standards development, the service providers (e.g. PTTs) and manufactures and the users. ISDN is such a wide field that is almost impossible to review all the relevant work in this field. Therefor, here I present the work of national and international organizations, while the other related work, especially research aspects, are referred to in the rest of the book where necessary.

The CCITT (International Telegraph and Telephone Consultative Committee) has been working on the definition of the ISDN concept as well as on interfaces and signalling procedures for the ISDN since before 1980. Its study periods are separated into groups of four years. The first ISDN study period, 1980-4, has culminated in the I-series Recommendations known as the Red Book, while the study period 1984-8 has produced by Blue Book version of the recommendations. The ISO (International Standard Organization) prepares standards for data communication. CCITT and ISO co-operate in some of these areas. ISO standards are relevant for data communications using the ISDN. Indeed, the layering principle of the Open Systems Interconnection (OSI) is adopted by the CCITT in developing the I-series Recommendations.

In Europe, the European Computer Manufactures' Association (ECMA) deals with various aspects of data communications between data processing systems, including their interconnection to ISDN PABXs. ECMA has a series of Technical Standards relating to ISDN systems. Also, a new ISDN users' forum, the European ISDN Users' Forum, has been establish to co-ordinate the activities of the service providers, manufacturers and users on ISDN services and equipment.

The European Telecommunications Standards Institute (ETSI) was formed by the CEPT (European PTTs) in November 1988 to carry out the standardization activities for the European countries. This standardization activity also applies to the ISDN in selecting common options and services. An additional objective is the co-ordination of the installation planning activities. A group has also been formed within CEPT to study the use of ISDN to provide the OSI network service.

QUAD is joint venture between the German, French and Italian PTTs and British Telecom, which are co-operating within this venture to produce documents setting out how the ISDNs will be implemented. Its aim was to achieve terminal portability through technical harmonization by 1992 (Trudgett 1988).

The Standard Promotion and Application Group (SPAG) is a European industry grouping formed to promote functional standards and a common conformance testing facility. SPAG produces a Guide to the Use of Standards (GUS) which contains profiles as well as a description of their use and purpose.

In the United Kingdom, the UK ISDN Standards Forum (UISF) comprises UK service providers (e.g. British Telecom) and users; it co-ordinates the submission of information and views of the UK organizations to the CCITT on relevant standards. The Department of Trade and Industry has also undertaken work to co-ordinate ISDN standards, implementations and usage in the UK. Another forum that is co-ordinated by the British Standard Institute (BSI) is the TCT/8/-/4 Committee, looking at the LAN-ISDN and LAN-ISPBX interconnection issues. The work of this forum is also passed on to the relevant committees of ECMA.

In the United States, the Exchange Carriers Standards Committee (ECSA) sponsors the Standards Committee T1-Telecommunications. The membership of the committee is open to all interested parties. T1 develops national standards, which are promulgated by the American National Standards Institute (ANSI). It also co-ordinates the US contributions to CCITT. The T1S1 subcommittee T1 deals with the ISDN standards in the United States. The North American ISDN Users' Forum (NIU-Forum), which is set up by the National Institute of Standards and Technology (NIST) to look into ways of using ISDN and developing new applications and services, supports two sets of committees: a user set and an implementors' set.

Another organisation working on standards in the United States is the Institute of Electrical and Electronic Engineers (IEEE). The relevant technical committees of this institute have been working on the use of frame relaying techniques in LAN-ISDN-LAN interconnection (IEEE 1988a) as well as B-ISDN and interworking with metropolitan area networks (MANs).

<u>Integrated switching</u>: The requirements of different types of traffic sources have traditionally been addressed by the circuit-switched (CS) and packet-switched (PS) communication system. For example, circuit and packet switching better serve the voice and data services, respectively. In order to improve bandwidth utilization and reduce the multiplicity of access and network arrangements necessitated, transmission of circuit-switched and packet-switched traffic sources on the same medium can be achieved by using time division multiplexing (TDM). A time division multiplex frame can be separated into two parts: one for the circuit-switched applications (blockable traffic type, class 1) and one for the packet-switched applications (queueable traffic type, class 2). The boundary between the two can be either fixed or moveable. Kummerle (1974) and Zafiropulo (1974) first proposed the moveable-boundary strategy. The method is

shown to offer a better performance advantage for class-2 users over a fixed-boundary scheme, in which the same reserved slot allocations are made. The blocking performance of class-1 calls remains the same in either case. The data packets achieve a reduction in their queuing delays under the movable-boundary strategy (Schwartz 1987). Weinstein *et al.* (1980) have shown that a dynamically movable boundary is not very effective in allowing class-2 traffic (e.g. data packets) to utilize free capacity arising from momentary fluctuations in class-1 traffic (e.g. voice) unless an appropriate flow control mechanism is used. The analysis and design of hybrid switching networks is carried out by Gitman *et al.* (1981), among others. Maglaris and Schwartz (1982) proved that the movable boundary provides optimal or near optimal system performance for the single-user case. The multiple-user case has been analysed by Krainmeche (1984), who showed that for the balanced traffic case, and for the range of parameters used, the movable boundary mechanism provided an improvement in packet traffic delay.

A refined hybrid switching method, the slotted envelope network (SENET) concept is proposed by Coviello and Vena (1975). Here, pseudo-synchronous 'envelopes' are used to convey multiplexed traffic through the network using a wideband slotted transmission format, such that the traditional characteristics of both circuit and packet switching's are simultaneously provided. The concept can provide flexibility for a varying mix of traffic and the capability to provide variable channel bandwidth's to match different applications.

Access control problem in ISDNs: Narrowband ISDN can be viewed as a time division multiplex (TDM) facility where each time slot in the TDM frame can be accessed separately. When larger bandwidth links are required, multiple slots are accessed. A mixture of traffic comprising *K* types of customers is assumed to share this facility. Such a situation occurs when more than one class of traffic (e.g. voice and data) is allowed to access the facility.

For a hybrid system with packet- and circuit-switched traffic types, the access control problem can then be defined as: 'dynamic assignment of bandwidth units (slots, bits, etc.) to incoming requests, as a function of both circuit (CS) and packet (PS) switched load, in order to optimise the performance of PS traffic to achieve minimum delay given an upper bound of acceptable blocking probability for the CS traffic' (Kraimeche 1984). A considerable amount of interest has been shown in the performance of the above model in the last eight years. Maglaris and Schwartz (1979, 1982), Li and Mark (1984), and Yum and Schwartz (1987) have studied hybrid switching in TDM systems. A taxonomy for hybrid switching is given by Goeldner (1985). Modelling and simulation techniques are used to obtain message, packet or bit-level performances of the models described. However, only a few have concentrated on the message-level analysis (Daigle and Whitehead 1984). A good survey of previous work in this field is given in Fletcher *et al.* (1986).

Channel access control strategies are needed for integrated communication system with heterogeneous users (data, voice, and video) who have differing service requirements, so that it can handle its traffic demands. Close control of access and switching at the input node is necessary for high efficiency and flexibility (Kraimeche and Schwartz 1984). The same applies if ISDN is used in a data-only environment, for example LAN-LAN interconnection using the ISDN. The difference is that, although the real-time requirements of voice and video do

not exist, the requirements of various data services still need to be addressed (Deniz 1987b). For example, a suitable quality of service must be provided to users, requested requirements may be given higher priorities over those with less stringent requirements (e.g. file transfer and electronic mail).

Analytic models for ISDN access: Queuing systems for ISDN access have generally been modelled as multi-server queuing systems with *F* identical, parallel servers, and where customers upon arrival demand a variable (random) number of servers (basic bandwidth units-BBUs) for the duration of their service times. The random number of servers required represents customers bit rate requirements. The most important property of this class of queues is that a given customer can enter service only when all of this his or her requested servers are available.

Different customers are identified by their bit rate (bandwidth) requirement. Other attributes characterizing customer traffic futures are the average service (or holding) time, and their generation rate and pattern. A customer arriving at the ISDN access node makes an *access request.* Access requests for bandwidth service are assumed, in general, to be of *K* types. A type *i* access is identified by its arrival rate λi in requests/time, its mean service time $1/\mu i$ in units of time, and its information bit rate *bi* in BBUs. Two variants of this model exist. In the first, the servers allocated to the same customer are all released simultaneously. This is the model used by Brill and Green (1984) and by Kraimeche (1984). In the second model, the servers allocated to the same customer free up independently. This model has been studied by Green (1980, 1981).

In hybrid switching models, with both the CS and PS traffic access, the CS traffic is usually assumed to operate on the 'blocked calls cleared' principle and the PS traffic is assumed to be queuable. The same principle is extended to analytic models where *K* traffic streams compete for the bandwidth resource so that two variants of the model are studied: the case where all *K* traffic streams are assumed to be blockable or queuable, respectively. A special case of this model is when the traffic types are categorized into two classes: wide-band (WB) and narrow-band (NB) traffic types. In this case either one or both are assumed blockable or queuable.

Channel assignment techniques in hybrid switching are analysed by Schwartz and Kraimeche (1982). The performance of blockable WB and queuable NB traffic types is studied by Kraimeche and Schwartz (1985b) as a mixed blocked/queued traffic. The case of blocked calls cleared for NB and WB traffic with trunk reservation is studied by Liao *et al.* (1989). Serres and Mason (1988) study a multi-server queuing system with blockable NB and queuable WB customers under a wide-band restricted-access strategy. Kraimeche and Schwartz (1984) study the case of K heterogeneous CS traffic types in the 'blocked calls cleared' model. The performance of complete sharing and complete partitioning is investigated. A class of restricted access policies is also proposed with improved performance. Kraimeche and Schwartz (1985a, 1986) consider the case of all queuable K streams of traffic. A channel access structure for wide-band ISDN is studied by Kraimeche and Schwartz (1987a, 1987b), where the interface is ordered hierarchically in channel capacity and *K* queuable traffic types are allowed to contend for the bandwidth with channel overflow facility. The case of *K* traffic types grouped into clusters with blocked calls cleared and Kashper (1987) studies a restricted access policy.

Queuing model for CL mode packed access to CS ISDN: For the CL packed access to the circuit-switched ISDN, the general queuing models mentioned above are no longer valid. In the first place, the concept of customers *demanding* a (random) number of servers is not true for this case. This is because the customers arrive as disjoint packets (datagrams) and have no files in their headers to indicate the bandwidth that will be required for a complete session (ISO 1987b; Postel 1981a). In fact, they do not have a means of specifying the bandwidth required for that particular packet, either. Second, each customer can enter service when at least one BBU (channel) is available. However, it will not be efficient to assign a new channel (e.g. a B-channel of 64 kbps capacity) to every packet source with a different Internet address, since some of the Internet address will inevitably be on the some subnetwork (e.g. hosts on a LAN) and hence will be served by the some ISDN gateway. In such cases, the traffic to the same subnet can be multiplexed over one or more channels or a superchannel of variable bandwidth.

Multiplexing all the different data traffic sources (e.g. file transfer, e-mail and terminal access) to the same ISDN destination (e.g. a gateway) may produce problems in the provision of quality of service (QoS). It is necessary to study the queuing systems for QoS in CL interworking in order to see the effect of such multiplexing and bandwidth control algorithms on the performance of different traffic types. Prue and Postel (1988) suggest separate queues for different classes of data packets and apportion the server capacity accordingly between the separate queues (single server, multiple queues). A further increase in bandwidth of the channel is assumed to occur under the control of a channel or bandwidth management unit. This unit increases bandwidth of a channel by adding further BBUs to the channel capacity on a need basis. Furthermore, the release of allocated BBUs for a CL customer is done independently rather than simultaneously.

Hence it is necessary to clarify the notation of a 'customer' and the method of bandwidth allocation and release mechanisms within such an environment. In real computer networks, users communicate via processes. Processes communicate with each other to provide services. A single host or application can have multiple processes, which are active at one time. Each user process uses a transport level process to provide end-to-end communications on an internetwork or intra-network domain. On a local network multiplicity of hosts, each supporting a number of such processes, will want to communicate with peer entities at remote sites. At the ISDN access node (e.g. a network layer relay), all communications directed to a specific destination could be viewed as 'one' customer, despite the fact that their source host or process could be different. The difference in this case is that we now have an 'aggregate customer' whose features of transmission in terms of packet arrival rate, mean service time and information bit rate are changing over time (Deniz and Knight 1989b).

CHAPTER TWO

THE INTEGRATED SERVICES DIGITAL NETWORK

This chapter highlights the most relevant features of the integrated services digital network (ISDN) in its application to data communications and the interconnection of local area networks using the ISDN. It also provides a general background on the ISDN and its current development status.

The ISDN has evolved from a digital network concept, and this evolution is expected to continue for years to come (one or more decades). In this respect, the features of ISDN can be discussed meaningfully only with reference to the global principles, concepts and features of the network and its services, or to its *current* consolidated and standardized features, of proposals for the forthcoming CCITT study period of four years. Beyond that time scale, concrete evolution can not be made easily owing to the evolutionary process; only migratory paths could be indicated. Here, we shall concentrate mainly on the first two cases, with some reference to the immediate proposals expected for the next study period. Some possible future trends will be indicated.

Since its conception, the ISDN has diversified into two main streams: ISDN (also termed narrowband or N-ISDN) and broadband ISDN (B-ISDN). N-ISDN refers to network access at bit rates up to in including 2 megabits per second (Mbps), while B-ISDN refers to network access at bit rates (up to and) greater then 2 Mbps, usually in tens or hundred of megabits per second. It also concentrates on the circuit mode bearer service, because of its wider and earlier availability in Europe. However, a summary looks into other aspects of N-ISDN as well as the current and expected developments in B-ISDN are also presented here for completeness. Furthermore, this chapter refers mainly to the most recent CCITT Recommendations of the I-series (CCITT 1988a), and its predecessor known as the CCITT 1984. Designated work area for the third study period of 1988-92 are also indicated where necessary.

2.1 THE ISDN ERA

The CCITT is the international from that is leading the way in establishing a world-wide definition and standardization of ISDN. Its studies on the ISDN are published Recommendations of I-series. The I-series Recommendations coverall aspects of ISDN under the following sub-sections; 1.100-series, General Concept and Structure; 1.200-series, Service aspects; 1.300-series, Network aspects; 1.400-series, User-network interface aspects; 1.500-series, Internetwork interfaces; 1.600-series, Maintenance principles. These and other relevant recommendations are referred to the following sections.

2.1.1 Principles of ISDN

Its three essential futures can characterize an ISDN:

- End-to-end digital connectivity
- Multi-service capability (voice, data, video, and multi-media)
- Standard terminal interfaces

The ISDN concept includes the support of a wide range of voice and non-voice application in the same network. A range of services using a limited set of connection types and multi-purpose user-network interface arrangements will provide the service integration for an ISDN. ISDNs will support a variety of applications including both switched (circuit or packet-switched) and non-switched connections. They will contain intelligent for the provision of service features, maintenance and network management functions. The specification of access to an ISDN should be by layered protocol structure. ISDN may be implemented in a variety of configurations to be determined by the national service providers. Hence the details of the underlying network will not be important for the users, since they will see a common service across ISDNs (save or optional services and features).

2.1.2 Evolution of ISDN

ISDNs will be based on the concepts developed for the ISDN and will progressively in corporate additional functions and features at present provided by separate networks, as well as other new ISDN-specific features and services. During the transition period, arrangements will be developed for the interworking of services on ISDNs and services on other networks.

2.2 TECHNICAL FEATURES

Three key ISDN communication capabilities could be used to define the user's perception of an ISDN service: the provision of end-to-end digital connectivity between end users, the provision of multiple-access channels; and intelligent out-band (common channel) signalling and user control. The end-to-end digital connectivity will provide larger bandwidth, better signal qualities and better error rates than most of the present networks. Multiple access channels will provide more flexibility in the use of the total bandwidth available at the user-network interfaces; to this end the CCITT has defined different channel types and interface structures. The new common channel signalling scheme has been chosen so as to allow the user to control both the network connections and the delivery of user services and applications independently of the user information paths in an effective manner.

The basic architectural model of an ISDN as described in Rec. I.324, the ISDN Network Architecture (CCITT 1988a), includes the following main switching and signalling functional capabilities: ISDN local functional capabilities (e.g. user-user and user-network signalling, charging); 64 kbps circuit-switched and non-switched capabilities; rates greater than 64 kbps switched and non-switched capabilities; packet-switched and non-switched functional capabilities; and common channel signalling (user-exchange and inter-exchange) functional capabilities. For the circuit-switched connections, user-bit rates less than 64 kbps are rate-adapted as described in Rec. I.460 before any switching can take place in the ISDN. Multiple information streams from a given user may be multiplexed together in the same B-channel, but for circuit switching an entire B-channel will be switched to a single user-network interface. Packet-switching capabilities are formed by two functional groupings; packet handling functional grouping and interworking functional grouping. The first grouping contains functions relating to the handling of packet calls within

the ISDN, while the latter ensures interworking between ISDN and packed-switched data networks (PSDNs). Another features of the ISDNs will be its fast switching ability, made possible by digital switching techniques. Although the performance issues relating to this have not been fully defined, average national circuit switching speeds are expected to be below 1 s while for international calls the aim is for a speed below 5 s.

2.2.1 Service Aspects

Services provided by ISDNs have been divided into two broad categories: bearer services and teleservices. The bearer service is a type of telecommunications service that provides the capability for the transmission of signals between user-network interfaces and as such conveys information between users in real-time and without alteration to the information 'message'. The teleservice is a type of telecommunication service that provides the complete capability, including terminal equipment functions, for communication between users according to protocols established by agreement between administrations.

Bearer Service: Two types of bearer services have been defined: circuit and packet modes. The circuit mode bearer service has been defined for 64 kbps, 384 kbps, 1536 kbps and 1920 kbps rates (Rec. I.231, CCITT 1988a). In circuit mode, the 'unrestricted information transfer' with ' 8 kHz structural integrity' capability is of interest for data applications since it provides switched 'bit pipes' over which transfer of user information can be achieved without alteration (within the quality-of-service (QoS) limitations).

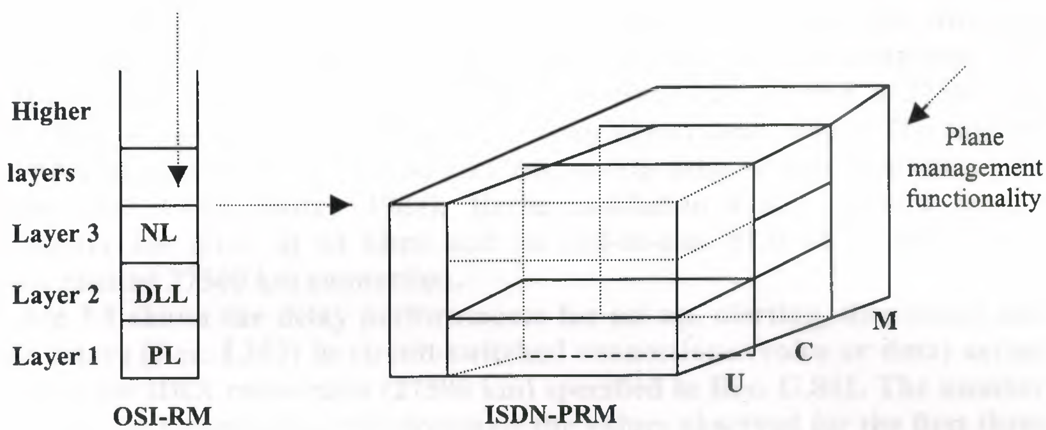
The packet mode bearer service involves packet-handling functions. The following service categories identified in Rec. I.230 (CCITT 1988a) are described in Rec. I.232: virtual call and permanent virtual circuit (connection-oriented (CO) mode), connectionless (CL) and user signalling packet mode bearer services. In the CO mode of operation, 'unrestricted information transfer capability' and 'service data unit structural integrity' are supported. User information is conveyed over a virtual circuit (VC) within B- or D-channel. Signalling may be provided via D-channel and/or VC within B-channel. Access protocols Rec. I.440, I.450/1, I.462 (X.31) and X.25 (layers 2 and 3) (CCITT 1988a).

Teleservices: Teleservices use bearer services to move information around, and in addition employ 'higher-layer functions' corresponding to layers 4-7 of OSI-RM. The mixed-mode teleservice provides combined text and facsimile communication.

2.2.2 Network Aspects

These cover the network functional principles, the ISDN Protocol Reference Model (PRM), numbering, addressing and routing principles, connection types and performance objectives. The ISDN-PRM, which is based on the principles of OSI-PRM, allocates functions in a modular fashion that facilitates the definition of telecommunications protocols and standards. The ISDN-PRM extends the OSI

model beyond the traditional point-to-point, user-network-user, in-band signalling, data-only communications. This extension includes the separation of signalling and management operations from the flow of application information within a piece of equipment; definition of communication contexts which may operate independently from each other; and the application of the above to internal network components. This separation leads to the definition of independent communications contexts. Each context can be modelled individually and use independent protocols.



NL: network layer; DLL: data link layer; PL: physical layer
U: user plane; C: control plane; M: management plane

Figure 2.1 OSI reference model (OSI-RM) and ISDN protocol reference model (ISDN-PRM)

The ISDN-PRM refers to two distinct logical planes: user (U) and control (C) planes. The layering principles apply to each of these planes; each one can potentially accommodate a seven-layer protocol stack (CCITT 1988a; Duc and Chew 1985; Potter 1985). A plane management function is required to allow co-ordination between the activities in different planes. Further study is needed to specify the types of layer services required to describe a telecommunications service, data flow modelling and ISDN management. The relationship between the OSI-RM and the ISDN-PRM is shown in Fig. 2.1

Numbering principles within the ISDN (Rec I.330) will be based on the expansion of the international PSTN numbering plan (Rec. I.331/E.164). An ISDN address is composed of an international ISDN number and an ISDN subaddress of up to 15 decimal digits and 40 decimal digits (20 octets) respectively (CCITT 1988a). the ISDN address may be of variable length, and all ISDNs will be capable of conveying the ISDN subaddress transparently. This subaddress may be used by private communications facilities such as LANs and PABXs or by other private networking arrangements. In other to identify different OSI network layer service access points (NSAPs), methods have been defined in the Rec. I.330 (CCITT 1988a) to relate the ISDN number to the OSI network layer address.

The performance objectives regarding propagation delays, availability and errors were not dealt with in full during the 1980-4 period. However, the specification is still by no means complete. Currently, provisional performance values are quoted, while the actual target values are left for further study. The general network performance (QoS) measures are defined in Rec. G.106, while ISDN performance objectives are defined in Rec. I.350. The network performance objectives for connection processing delays in an ISDN are given in Rec. I.352. These are further discussed below. The network performance objectives regarding error and slip rates that apply at reference point T of an ISDN are given in Rec. G.821/2.

The CCITT G-series Recommendations define hypothetical reference connections (HRX) in order to develop performance objectives for different HRXs. An HRX presents a typical worst-case end-to-end digital connection. The longest (international) terrestrial connection envisaged by CCITT is 27500 km. (Rec. G.801). This represents a *propagation delay* of less than 100 ms. If a satellite link is present, this would easily add up a round-trip delay of approximately 520-560 ms (Peel 1989; Sastry 1984). Recommendation G.821 defines overall performance for HRX at 64 kbps and an end-to-end BER of 0.000001 on a circuit-switched 27500 km connection.

Table 2.1 shows the delay performances for set-up, alerting, disconnect and release delays (Rec. I.352) in circuit-switched connections (voice or data) across the worst-case HRX connection (27500 km) specified in Rec. G.801. The number of exchanges in a connection will dominate the values observed for the first three parameters. For shorter-length connections the observed values will be lower. Delays are specified for a nominal busy hour and the delays that are dependent upon a user equipment (or network) and user response time is not included. Additional delays may be incurred at signalling message queues. Furthermore, the specified performances are for connections exclusively over ISDNs (i.e. no interworking).

Table 2.1 Objectives for connection processing delays in an ISDN (worst-case HRX connection)

Statistic	Delay (ms)			
	Set-up	Alerting	Disconnect	Release
Mean	4500	4500	2700	300
95%	8350	8350	4700	850

2.2.3 User-Network Interfaces

An important aspects of service integration for an ISDN is the provision of a limited set of standard multi-purpose user-network interfaces (UNIs). An ISDN is recognized by the service characteristics available through user-network interfaces, rather than by its internal architecture, configuration or technology. Hence the CCITT has put most of its initial effort into the specification of this interface to cover both the existing terminal/user base and the future ISDN terminals/users.

The I.400 series of Recommendations (CCITT 1988a) describe the interface between user equipment and ISDNs. They include a reference configuration, giving the functional model for the subscriber, a definition of interface structures and access capabilities, and detailed specifications of layers 1-3 for the basic and primary-rate interface structures. Several other Recommendations in the series provide standardized methods for rate-adapting classical X.1 and V.5 data rates into B-channels, for submultiplexing B-channels and corresponding circuit-mode bearer services, and for interworking with digital network facilities providing only restricted 64 kbps information transfer rates. Furthermore, a terminal adaptor for converting X.21 circuit-switched interfaces into the ISDN basic user-network interface is defined in Rec. I.461/X.30. Recommendation I.462/X.31 defines how existing packet-switched public data network (PSPDN) services can be provided in or through an ISDN to a packet mode user terminal. Packet mode user terminals, which use the X.25 packet-layer protocol but attach to an ISDN user-network interface, are also covered.

Recommendation Q.940 (CCITT 1988a) defines the ISDN UNI Protocol for Management (General Aspects). The management functions so far identified include: fault, configuration, accounting, performance and security aspects (CCITT 1988a; Ishii 1988).

Reference Configurations: The user-network interface lies within the customer premises. The reference configurations include reference points which are the conceptual points dividing functional groups. These functional groups include the termination (network termination-NT1), distribution and switching (NT2), terminal equipment with ISDN user-network interfaces (TE1), and terminal equipment with other interfaces (TE2) and associated terminal adaptors (TA). Figure 2.2 shows the reference configurations for the ISDN UNI (Rec. I.411).

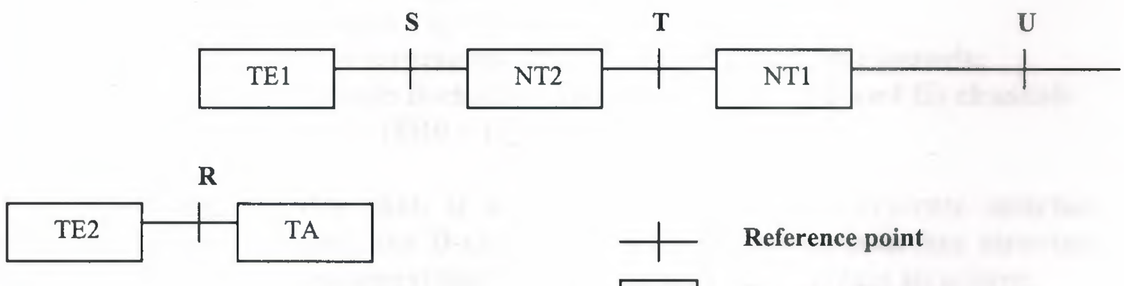


Figure 2.2 Reference configurations for ISDN user-network Interfaces

Channel Types: A limited set of channel types is defined, including B, D and H channels (Rec. I.412). The B-channel (64 kbps) carries user information. The D-channel (16 or 64 kbps for basic and primary rate, respectively) carries packetized signalling (s-data) information to control the set-up, modification and clearing of calls and services. Surplus bandwidth capacity on this channel may also be used for packet mode user data (p-data) and telemetry (t-data). The H-channels–384(H_0), 1536(H_{11}) or 1920(H_{12}) kbps—carry circuit or packet mode user information. New channel types for B-ISDN are also defined (CCITT 1988a).

Interface Structures: Only two generals (narrowband) ISDN user-network interfaces are described. These are the basic rate (Rec. I.420) and the primary rate (Rec. I.421) interface structure.

2.2.4 Basic-Rate Access (BRA)

This interface has an information-carrying capacity of 144 kbps. It is organized as (2B + D) channel types. With the framing, synchronization and D-channel echo bits, the aggregate transmission rate is 192 kbps. This interface is mainly intended for the connection of ISDN terminals or other individual terminals/workstations which will not need very high bandwidth. At this interface both the point-to-point and bus (S-interface bus, or S-bus) distribution types are available. In the S-bus configuration, up to eight terminals can be connected to the NT2, all sharing the same D-channel. Recommendation I.420 provides a connection resolution procedure for B-channel allocation since only two terminals at a time can access these.

2.2.5 Primary-Rate Access (PRA)

This interface has two subdivisions with differing information-carrying capacities. These are 1536 and 1984 kbps, and they represent requirements of US and European networks. Their aggregate transmission rates are 1544 and 2048 kbps, respectively. This interface can be organized in several ways:

1. Primary-rate B-channel interface structures (23B + D) and (30B + D)
2. Primary-rate H-channel interface structures:
   (a) H_0 channel structures, e.g.: $4H_0$, ($3H_0$ + D), ($5H_0$ +D)
   (b) H_1 channel structures, e.g.: H_{11}, (H_{12} + D)
3. Primary-rate interface structures for mixtures of B and H_0 channels:
   (a) Can consist of a single D-channel and any mixture of B and H_0 channels
   (b) Examples structure: (5H0 + D) and (2H0 + 11B + D)

It is interesting to note that, if a D-channel is one primary-rate interface structure is not activated, the D-channel of a neighbouring interface structure may be used for carrying signalling information for that interface structure.

The physical layer of a primary-rate interface conforms to CCITT Recs. G.703/4/6. Figure 2.3 shows a time division multiplex (TDM) frame structure (Rec. I.431/G.704) with associated time slots (TSs).

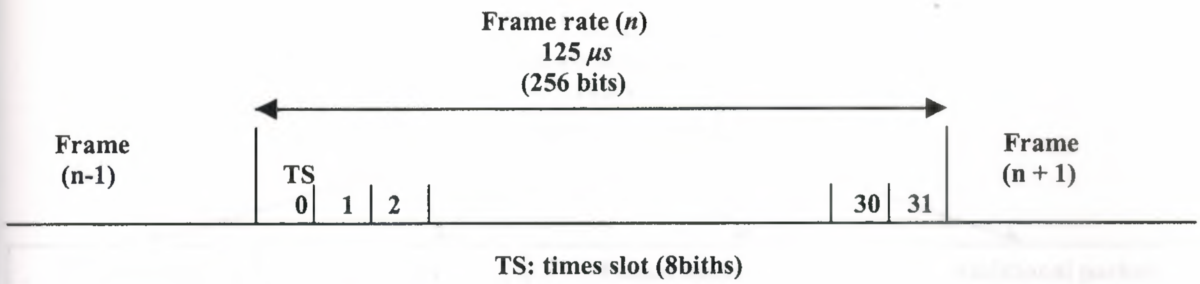
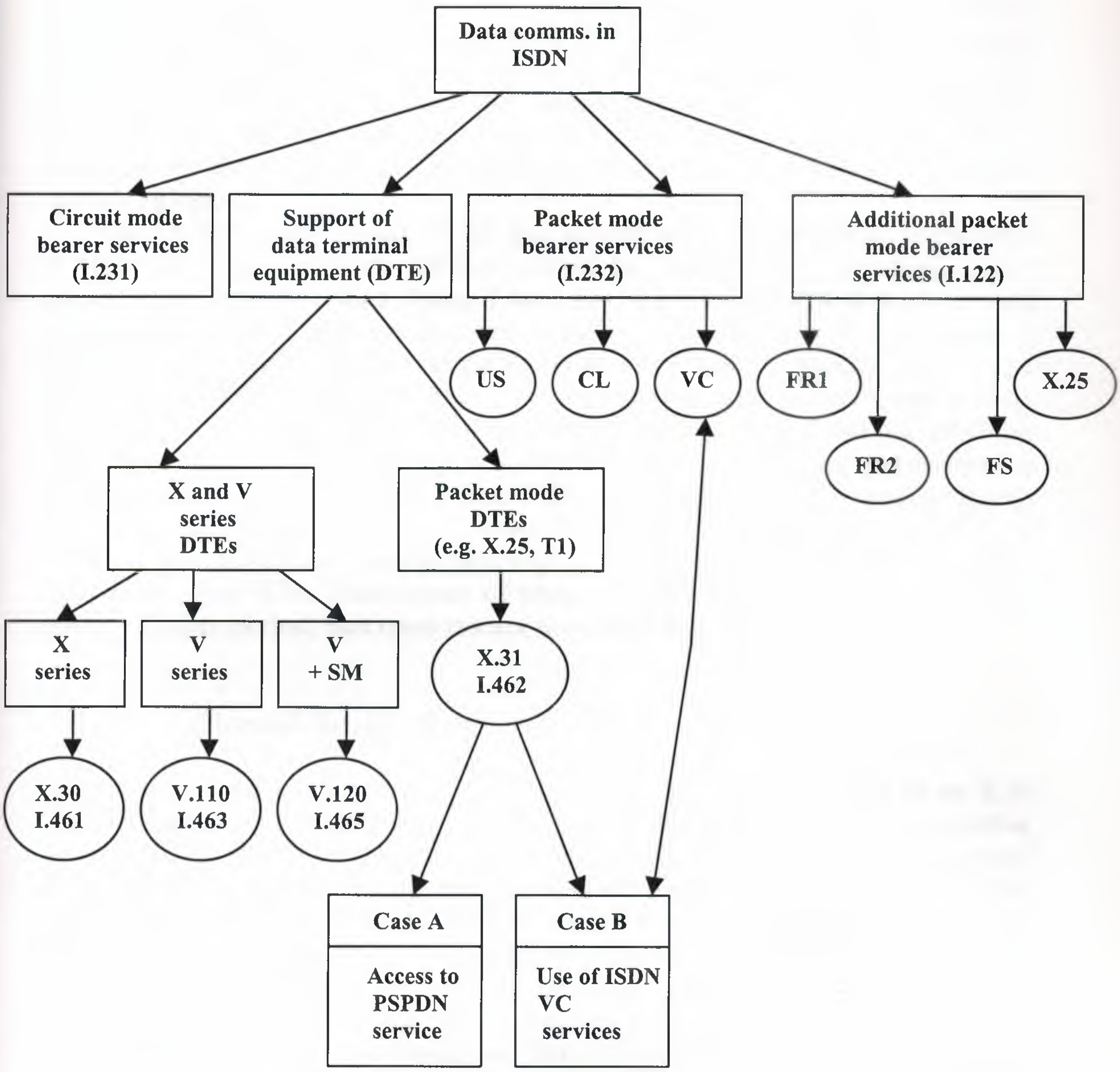


Figure 2.3 TDM frame structure used at PRA physical layer (Europe)

2.3 DATA COMMUNICATION IN ISDN

Data communication in ISDN can be achieved by using either the circuit or packet-switched and/or non-switched bearer services. The current alternatives for data communication in ISDN are summarized in Fig. 2.4. The circuit mode in ISDN is described in section 2.2.1. This service provides a digital bit pipe where the user channels are presented to the users at the physical layer. A channel occupies an integer number of time slots (TSs) and the same time slot positions in every frame. A circuit formed by the allocation of TSs to a channel in consecutive TDM frames available at the interface and connecting this to the destination by similar allocation of TSs in TDM trunks *en route*. The type of protocols to be used is left to the users. This provides a challenge to users as to how best to utilize this flexibility. The good BER of ISDNs means that light-weight protocols could be used at the data link and the network layers, hence reducing communications overhead. The packet modes in ISDN are described in the following section.



US: user signalling; CL: connectionless; VC: virtual circuit
FR 1/2: frame relay 1/2; FS: frame switching; SM: statistical multiplexing

Figure 2.4 Summary of data communication alternatives in ISDN.

2.4 PACKET MODES IN ISDN

As defined by the CCITT (CCITT 1984), the ISDN is principally a circuit-switched network which can be used for packet network interworking. The definition of a framework for providing additional packet mode services (I.122) has improved the data communications aspects of ISDNs. CCITT Rec. X.31 describes how an ISDN may offer access to an existing PSPDN from X.25 terminals connected to the ISDN via a terminal adaptor. It also describes how an X.25-like service may be supported wholly within an ISDN to offer a virtual

circuit bearer service; however, this requires a two-stage call set-up and as such is awkward. The recent developments and draft proposals indicate that a move is being made to incorporate packet network access in a much more tightly coupled manner than access to some distinct packet-switching function or packet handler as was previously defined (Atkins and Cooper 1989). This will be a closer step to the prevision of packet mode services wholly within the ISDN, which will be achieved by the expansion of the D-channel layer 3 protocols to cover packet-switching cases.

The CCITT Study Group XVIII has agreed on an evolutionary framework within which packet mode services would be considered. Accordingly, four phases have been described. Phase 1 incorporates enhancement of X.31 services described. Phase 2 includes the extension to X.31 service based on layer 2 multiplexing in the B-channel for an access arrangement to a packet handler and the use of X.25 PLP at layer 3. Phase 3 cover the additional packet mode services based on full out-band call control and layer 2 multiplexing (I.441—LAP-D) on the B-channel. Phase 4 covers broadband services including asynchronous transfer mode (ATM) and synchronous transfer mode (STM) services. Although Rec. X.31 (1984) describes a possible packet network internal to ISDN, it concentrates on the provision of interworking to existing PSPDNs. In addition to revisions to the X.31, foundations of phase 3 and 4 have been laid during the 1984—8 study period, and these too are described below.

2.4.1 Packet Network Interworking

Recommendation I.462/X.31 describes two different access scenarios to an X.25 based packet network (PSPDN): minimum and maximum integration scenarios. In the CCITT (CCITT 1988a) these are classified as 'Access to PSPDN Services' (case A) and 'ISDN Virtual Circuit Service' (case B) respectively. In the minimum integration scenario, PSPDN access is via access units (AUs), which are situated within the PSPDN. In this case, the ISDN provides transparent, or rate-adapted connection, for an X.25 terminal to/from a PSPDN port (AU). Therefore, as such is offers access to non-ISDN packet mode services. This access could be semi-permanent or switched. For switched access, a two-stage call set-up procedure is needed. In the first stage, ISDN D-channel signalling procedures (I.451) are used to establish a connection to/from the AU. Then X.25 call set-up procedures are used over the circuit-switched connection to complete an X.25 call over the PSPDN. The AU will be allocated an ISDN address. The user data access will necessarily be over the B-channel.

2.4.2 ISDN Virtual Circuit Bearer Service

The service is defined in Rec. I.232 (CCITT 1988a). It was named the 'maximum integration scenario' in Rec. I.462/X.31. This provides an X.25-like service within an ISDN from the user's point of view. Packet terminals using the ISDN numbering plan will have their protocols terminated by an ISDN packet handler (PH) within the ISDN rather than the PSPDN access unit. Access to the PH may be via the B- or D-channel.

In access via the B-channel, a two-stage call set-up is needed. The user will not be aware of this because of the initial 'packet transfer mode' requested using the I.451 signalling protocol. The ISDN will then establish a circuit-switched connection to a PH. for access using the D-channel, LAP-D frames identifying packet data are interleaved with the signalling information. The distinction is done using a service access point identifier (SAFI) within the frame header. The network will route packet frames to/from the PH. Signalling frames are given priority over packet frames. An X.25 terminal uses and terminal adaptor (TA) to interface to the network termination. Throughput of data over D-channel is limited by the 16 kbps data rate, and additional delays may be incurred in the TA. However, it can be used in parallel with the circuit-switched B-channels.

2.4.3 Additional Packet Modes

The signalling scheme defined for the D-channel is for controlling circuit-switched user traffic channels and non-connection-related applications. The new Rec. I.122 (CCITT 1988a) aims to use the enriched call control technique is controlling packet-switched virtual circuits. Recommendation I.122 describes only the architectural framework and service descriptions of the associated bearer services. It will apply to data communications at bit rates up to 2 Mbps. The reasoning behind the additional packet modes include the following: the provision of packet mode bearer services, which would be accessible using clearly separate control and user planes employing the same signalling protocols as other ISDN bearer services; and fast data transfer (with reduced delay and high throughput), by avoidance of full three-layer termination at each node in the network.

The method involves the use of outslot signalling with single-step call establishment in accordance with I.451, rather than the X.25 PLP call control procedure. Also, the multiplexing and switching of virtual connections takes place in layer 2 (LAP-D over B-channels) rather than the X.25 PLP. In the control plane, information elements specific to packet switching have to be added to existing I.451 signalling messages. In the user plane, the LAP-D functionality which will be terminated in every exchange may be restricted to a core subset including the frame delimiting, alignment and transparency; frame multiplexing and demultiplexing using the address field; and detection of transmission errors. These core function result in a low-cost service, independent frame-relaying mechanism on a link-by-link basis, which will result in low overheads and could be used in various applications.

Recommendation I.122 (CCITT 1988a) specifies four potential bearer services: frame relaying types 1 and 2, frame switching and X.25-based additional packet mode.

- Frame relaying involves establishing a transparent network path for a call at call set-up time and then conveying subsequent frames along that path checking only the validity of frame format. Frame order is preserved, but frame acknowledgement within the network is not undertaken. Additional network functions to monitor and enforce throughput and achieve congestion control as well as frame error detection and QoS parameters are left for future study. The difference between the types of frame relaying is that type 1 implements only the *core* functions of Rec. I.441 (layer 2), whereas type 2 implements all the extended I.441 functions within the user terminals. In both *cases, the network implements only the core layers 2 functions.*

- Frame switching is similar except that the network passes on frames, which correctly observe the layer 2 protocols. Additionally, it provides frame acknowledgement, error detection and recovery, and provides control. It necessitates the full implementation of layer 2 protocol (I.441 with extensions if necessary) both in the user terminal and in the network.
- X.25-based service adds to this the network support for X.25 PLP data transfer part.

Many important issues in the provision of additional packet modes still remain for features studies, including throughput enforcement, flow and congestion control (due to lack of layer 3 congestion and flow control mechanisms at layer 2), signalling enhancements to I.451, application categories (to determine type of bearer service), and charging (frame types need to be distinguished to enable charging).

2.5 PROTOCOL CONSIDERATIONS AND STANDARDS

With the introduction of distinct user and control planes in the ISDN reference model, functionally separate interactions associated with the control and signalling functions and transferring data are divorced, leading to separate and parallel developments of their functionalities and protocols. Within the CCITT, the protocols for the D-channel have been defined for the layers 1—3. For the B-channel, in the case of circuit mode services only layer 1 has been defined, while for the packet mode DTEs layers 1—3 based on the X.25 model are proposed. Additional packet modes incorporate layers 1—2 and 1—3 depending on the type of service, with LAP-D in layer 2 and X.25 PLP in layer 3 (when used). In the circuit-switched ISDN case, the user is left free to decide which particular protocol stacks to use on the B-channels for data communications over the links established. The D-channel signalling protocol is message/frame-structured and layered according to the OSI seven-layer model. Basic and primary rate interface layer 1 specifications are necessarily different, but are common for layers 2 and 3. Table 2.2 shows protocol stacks on the D-channel for both BRA and PRA.

Table 2.2 protocol stacks for the D-channel

Protocol	Signalling protocols	
layer	BRA	PRA
Layer 3	I.450 / I.451	I.450/1 (Q.930/1)
Layer 2	I.440 / I.441	I.440/1 (Q.920.1)
Layer 1	I. 430	I/431

Recommendations I.450/1 (alias Q.930/1) define the network layer features and protocol. This has 'broad' similarities to the X.25 PLP functionality (e.g. frame formats), but is much more complex. It has a sophisticated message set in order to achieve call control for different types of services (bearer and teleservices, e.g. circuit- and packet-switched; additional packet mode services) and telephony functions. However, the protocol is easily expandable to support more complex calls, standardized supplementary services and network specific services. Furthermore, this signalling procedure may also be used within private networking arrangements (e.g. between two PABXs).

Recommendations I.440/1 (Q.920/1) define the data link layer procedures for the ISDN D-channel (LAP-D). This protocol extends the X.25 LAP-B protocol. The main difference is that it allows multiple data links by virtue of an extended address field (two octets). It also uses several additional HDLC commands and responses.

Recommendations I.430 and I.431 define the physical characteristics of the basic and primary-rate-user-network interfaces, respectively. The basic rate interface supports point-to-point as well as passive bus (S-bus) arrangements (including collision detection and contention resolution) and it has a four-wire 'S' interface. The primary-rate interface specifies coaxial wires for receive and transmit paths, and only the point-to-point configuration is supported according to CCITT Rec. G.703 (CCITT 1988a).

2.6 BROADBAND ISDN

Recommendation I.121 on 'Broadband Aspects of ISDN' describes the basic concepts, service aspects, user-network interfaces and architecture models of B-ISDN. Asynchronous transfer mode (ATM) is chosen, as the target transfer mode for B-ISDN, while synchronous transfer mode (STM) will also be supported in the interim. Unlike the narrowband ISDN (based on 64 kbps), it employs optical fibres for customer access. It proposes that all access to broadband and 64 kbps-based services will be via a single user-network interface S_B. Control of all services will be by means of CCS in line with the existing ISDN principles.

2.6.1 B-ISDN Services

Two broad service categories are identified; interactive and distribution services. The interactive services are subdivided into conversational services (bidirectional end-to-end information transfer), messaging services (user-to-user communication via store and forward or mailbox type functions) and retrieval services (interactive videotex, video, tex and graphics). The distribution services are subdivided into services with or without user individual presentation control.

2.6.2 B-ISDN User-Network Interfaces

Two types of user-network interfaces are proposed:

1. 150 Mbps interface: mainly for interactive services; symmetrical and providing an H_4 channel plus some B- and D-channel capacity.
2. 600 Mbps interface: for interactive and distributive services; may be asymmetrically.

It is expected that the initial emphasis will be on the 150 Mbps interface as this will be preferred for fast data transfer because of its reduced transmission costs compared with the higher bit rate transmissions.

The following additional interface structures are proposed (through these channel rates will not appear physically at the S_B-T_B interface owing to the ATM principle of variable bit rates):

- H_{21} and H_{22} channels (32.768 and 43—45 Mbps respectively)
- H4 channel (132—138.240 Mbps)

In addition to these, the SB-TB interface will have B, H0, H11 and H12 narrowdand channels converted to ATM and separate virtual signalling channels. Carrying the corresponding physical channels as virtual channels over the broadband access (Frantzen 1987) can provide the narrowband S-T interface.

2.6.3 Asynchronous Transfer Mode (ATM)

ATM is a new multiplexing and switching concept, which is capable of handling both the bursty and stream type of traffic. Its most important feature is the flexible bandwidth allocation scheme, which decouples the interface from the network design. ATM is based on virtual channels (CO), with varying bit rates determined by service needs, and employs packet-switching technology. Fixed-size cells consisting of a header and an information field are used to transport data, reducing complexity and variation in delay. The basic ISDN protocol architecture regarding the separation of user, control and management planes is preserved. It is chosen by both the CCITT and CEPT as the target solution for B-ISDN.

The parties involved in its definition including CEPT (NA5), CCITT (Study Group XVIII) and the United States (T1/S1). On the technical side, some problems exist regarding the provision of time transparency (e.g. smoothing of propagation delay variations) and semantic transparency (e.g. error processing) in ATM networks. The cell and header sizes, call delineation, synchronization, flow control, congestion handling and adaptation between ATM and non-ATM parts of ISDN require further study.

2.7 CURRENT STATUS AND FUTURE TRENDS

2.7.1 Current Status

The Recommendations developed during the CCITT study period 1980—4 concentrated on general principles and on a definition of ISDN services as well as the definition of D-channel protocol specifications. During the CCITT study period 1984—8 much effort was put into further extensions. These included the extension of D-channel signalling protocols, the development of further network standards for ISDNs (numbering and addressing, network and terminal identification, network interworking and routing), the ISDN protocol reference model and consolidation with the OSI-RM, definition of additional packet modes, and the work on the definition of the B-ISDN principles.

One of the work areas that has progressed during the second study period has been the Q.931 expansion for layer 3 specification. This includes the essential features, procedures and messages required for call control (CS, user-user, PS) in the D-channel. However, some procedural details have not yet been specified and have been left for further study (e.g. transport or other message-based information flows, and the alignment of the functions and protocol with those of OSI NL). Further developments have been achieved in the CCITT Study Group regarding the subaddressing and NSAP addressing issues.

During the period 1984—8, the CCITT (1984) provided enough impetus and reasonably firm standards for integrated circuit manufacturers to start work on new ISDN VLSI chips. Several manufacturers have developed and introduced ISDN chip families, board-level products and test instruments for the basic and primary-rate interface structures. It can be expected that CCITT (1988a) recommendations will further proliferate product and service implementations for the narrowband ISDN as these recommendations become even more stables.

Further enhancements and clarity were brought to the existing Rec. I. 462/X.31. This is expected to increase the development of terminal adaptors for X.31 access to ISDNs as well as carrier provisions for such a service. However, it seems that X.31 (case B i.e. the maximum integration) is the only recommendation on which the ISDN packet services can now be based. It is also to be used as the basis of early European functional standards for ISDN. There seems to be little interest in the UK for the provision of a service based on X.31 (case B).

Recommendation I.122, providing a framework for the additional packet modes (APMBS), has appear in the CCITT. Recommendation I.232 for packet mode bearer services details three further bearer service categories: virtual call and PVC, connectionless, and user signalling. (Deletion of the last two services has been proposed to the CCITT Study Group XVIII.) these will add to the existing X.31 options for data transmissions within ISDNs and should provide that the common channel signalling on the D-channel can be used for all types of connections within the ISDNs. Its implications for hardware and software need to be evaluated. Several studies have so far been carried out for the use of APMBS in LAN interconnection. These studies include a Temporary Document from British Telecom (TD 17 : BT to CEPT/NA1-WP3 Report) and the IEEE IVD-LAN Interface Working Group (IEEE 1988a).

Recommendation I.121 on the broadband aspects of ISDN, which appears in the CCITT, is intended to provide a framework for the service definition, but there still is a lot of confusion regarding the technical implementation details (Minzer 1989) and B-ISDN applications despite the European Commission's aim to hold field trials during the 1991-5 period. However, we cannot expect to see much proliferation in the equipment or services available to customers before 1995. During the third CCITT study period (1988-92) the consolidation of the technical aspects of ATM and B-ISDN are expected. Without these, it is difficult to see manufacturers committing themselves to interim standards.

The CCITT Study Group XVIII prepared a new list of issues to be pursued during the 1989-92 period. These included the ATM; performance aspects; the interworking of ISDN and PSTN, PSPDN, CSPDN, ISDN; and private networks including LANs and user-network interfaces; layer 1 characteristics of 64 kbps ISDN including the updating/completion of Rec. I.430/431.

The earliest option available for data transmission over (narrowband) ISDN in Europe will be the circuit-switched access (as this is 'inherent' to the system). The second option will be packet-switched (X.31-based) access (through this will depend on the carries implementation policies). It seems that, for the X.25-based applications, the X.31-based implementations will be the norm for the immediate future. This protocol is now quite mature, although further evolution may be expected.

2.7.2 Future Trends

Customer applications will probably determine and justify the selection of bearer services, in which case some of the existing Recommendations may or may not be implemented by all carries. The degree of acceptance of Rec. I.462/X.31 will depend on the success of the additional packet modes, since the frame-relaying/switching may be preferred for some applications and the X.25-based approach for others where a common call control procedure will be advantageous to the X.31 methods. The X.31 methods provide an early solution to the packet modes in ISDNs, while the use of X.25 over highly reliable (good BER) ISDN lines seem wasteful. Hence, light-weight protocols for data transfers are needed. However, it is still by no means certain that all the bearer services proposed in the additional packet modes will mature to be implemented. X.31 is still evolving, and ISDNs capable of supporting X.31 are currently being implemented.

The future of circuit-switched services within the ISDNs seems to be guaranteed (essential for telephony, fax and video), and they may be incorporated into the ATM-based ISDNs as well. However, it is interesting to note that the connectionless packet mode of working within the ISDN is not available. The alternative seems to be a move towards an all CO mode of communications, both at the local networks and at the public networks or the provision of a CL mode of operation on top of circuit-switched or semi-permanent lines or within some frame-relaying/switching or X.25-based 'tunnel'. Another solution may be the full CL and CO interworking arrangement, as indicated with the recent work by the ISO (1989).

Another issue is the recent and future developments regarding the ATM and B-ISDN, Co-ordinate B-ISDN field trials in Europe are expected during the 1991—5 period; an earlier start apparently could not be made because of the difficultly of deciding on the right multiplexing technology to used. No one yet has a clear idea of how broadband will be used in the near future; the only existing demand is for high-speed data (and to some extent for video). One of the main problems the B-ISDN activists have been facing is finding real applications for B-ISDNs (e.g. motion video, medical applications, etc.). The ATM is a statistical technique and as such its capability to provide circuit-switched (CS) services are questioned. It is expected that it will eventually be able to replace CS services, even for voice. However, the costs involved are prohibitive for now. Another problem of the ATM technique is not so much the speed, but the handling of a large number of basic channels like the 64 kbps narrowband ISDN channels.

Value-added services (VASs) depend on the provision of higher-layer functions associated with the OSI layers 4—7 and are provided on top of basic services by value added carriers. The VAS service modules or servers can be accessed from a number of different networks, including the ISDNs. These services will include database access, store-and-forward communications, communications support services and compatibility services. The enhanced signalling capabilities of an ISDN is seen as a useful method for allowing signalling messages to be exchanged between a VAS user and server parellel with, or in the absence of an established network connection. It is thought that ISDNs will provide an optimum framework for VAS. Hence, widespread implementations of VASs are expected in the ISDNs in future.

2.8 RELEVANT FEATURES OF ISDN IN LAN-ISDN INTERCONNECTION

In the design of data-oriented LAN-ISDN relays, the following features of the ISDN are most relevant:

- Multiple channel access structures
- Common channel signalling (CCS) using the D-channel
- Switched and non-switched capabilities
- Circuit and packet-switching capabilities

Multiple channels within one interface, which are user-controlled with regards to the way they are used and the type of ISDN services selected, are available in the ISDN era through the use of a CCS system. This is a completely new development in the field of data communications, where most networks still use leased lines for interconnection. Packet data is by nature bursty and the bandwidth requirements of data networks connected to ISDNs may vary during a given time duration. Hence one could conclude that the channels at the UNI could best be utilized by dynamically controlling their total bandwidth. This leads us to the idea of superchannels and dynamic channel (bandwidth) management, issues that are studied in Chapters 4 through 10.

Naturally, the integration of services within the ISDN also has implications for the integrated services LANs (ISLANs) and integrated services PABXs. However, these issues are not further discussed. Since here we are focusing on the data applications across the ISDN itself.

2.9 SUPERCHANNELS IN ISDN

In this case, channels of arbitrary bandwidth size (in units of 64 kbps) are termed 'superchannels'. Current ISDN recommendations do not specify such a facility, although channel at predetermined higher bandwidth values are defined. However, there is a need in many applications to have channel capacities that are not limited by the basic channel types. Hence formation of superchannels by suitable aggregation of basic channel types is desirable. However, identification of such channels on an end-to-end basis, as well as provision of dynamic variation of their allocated bandwidth, is not yet available within the CCITT recommendation.

2.10 TARIFFS

This issue has not been completely addressed so far by the parties involved. However, the early indications show that, for data communications using the circuit-switched ISDN, a tariff structure similar to that of telephony will be applied. This means that each call is charged according to the charge band applicable (depending on the time of day) and distance. Each unit is charged to the customer at the beginning of the charge period. In some countries the charge period (or the charge applied to each charge period) is not linear, the first charge period being shorter (or the first charge unit being higher), so that a type of 'connection cost' is incurred by the users.

CHAPTER THREE

INTERCONNECTING

LANs AND ISDN

Developments in the local and wide area carrier technologies, and the need for new applications with large bandwidth requirements, running on very fast processors, mean that the methods of interconnection of these devices will be the crucial issue in their collective performance. Today, in the local area, the workstation/server model with fast local area network (LAN) interconnection has become the norm in data communications and local networking. In the wider area, private and public networks with ever increasing capabilities are providing better facilities for the interconnection of local networks, servers and workstations.

In data communications, the interconnection of LANs using ISDN has become an issue of great importance because of the ISDN capabilities described in Chapter 2. The ISDN, apart from integrating services and providing common user interfaces, will facilitate fast switching via end-to-end digital connectivity as well as multiple channels at various transmission capacities. The use of common channel signalling (CCS) will mean that control of these channels can be made in an outband mode concurrent with the user data transmission over the user channels. To date, little has been reported about how these facilities may be used in LAN-LAN interconnections using the ISDN. Hence our interest in LAN-ISDN-LAN interconnection arises from two ambitions: to investigate the issues involved in the LAN-ISDN interconnection, and to study the channel management issue at the boundary of the packet switching—circuit switching network interconnection. The LAN-ISDN interconnection will be common source of PS—CS network traffic.

3.1 PACKET NETWORK INTERCONNECTION ISSUES

For the multitude of reasons, which have already been discussed above, computer networks need to be interconnecting. The common objective of all interconnection methods is to allow all subscribers a transparent means of accessing a host or service on any of the interconnected networks.

To achieve this objective, data produced at a source in one network must be able to delivered and correctly interpreted at the destination(s) in another network. This necessitates the provision of inter-process communication across the network boundaries (Cerf and Kirstein 1978). In order to achieve this, some sort of 'commonality' is required across the interconnected networks. This can be achieved by two general approaches:

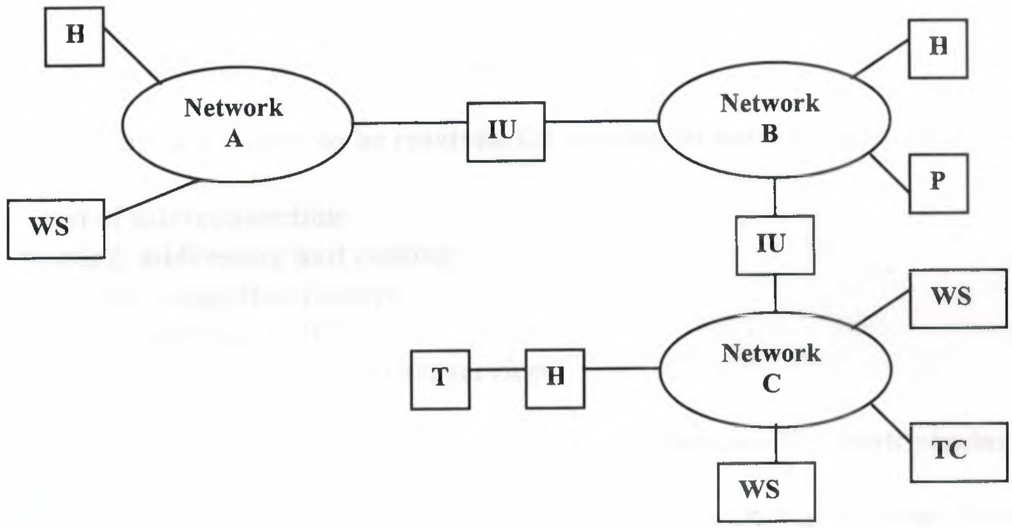
1. Translation of one protocol into another, when different protocols operate and separate networks.
2. Using protocols common among the communicating parties.

Also, a strategy combining the two approaches may be employed (e.g. using some common protocol at a given layer and translating protocols at one or more layers to the interworking protocols). In order to provide a common model for data communications, ISO has developed the Reference Model for Open Systems Interconnection (OSI-RM) (CCITT 1988a; ISO 1984). This model provides the basis for development of common protocols as well as for interworking between heterogeneous environments (Burg and Iorio 1989).

3.1.1 Network Interconnection Approaches

From the point of view of the desired functionality, three general approaches to network interconnection can be described (Schneidewind 1983):

1. *Network access* Achieves physical access and protocol compatibility in the lower layers of the OSI model (e.g. physical, data link and network layers of X.25). This method is used when the networks (e.g. LANs and the long-distance network) already exist with their own established protocol structures.



IU: interconnecting unit (e.g. gateway); H: host/server
WS: workstation; T: terminal concentrator; P: pad

Figure 3.1 Multiple network interconnection.

2. *Network services* Obtains the use of specific types of services (e.g. virtual circuit service) to satisfy user needs. This method is used for resource sharing or values added networks. It stresses obtaining network services for the user (e.g. provisions of remote interactive processing involving session control and sequenced error-free message delivery).
3. *Protocol functions* Matches the number and types of protocols on each type of network. Protocol conversion is necessary for this method, since different sets of protocol functions are assumed at each end.

Two similar terms are usually used in describing the interaction between two or more networks:

- Network interconnection
- Interworking (or internetworking)

Interworking issues are defined as those issues relating to the provision of services and facilities available on one network to the users on another network or to single users accessing that network using isolated terminals or workstations.

The networks access approach falls into the interconnection classification, while the network services falls into the interworking classification described above. In layered communications architecture, the protocols used in each layer provide a service to the layer above. Hence the protocol functions approach falls somewhere between the two. We are interested in network access as far as the LAN-ISDN interconnection is concerned. For the LAN-ISDN-LAN interconnection, however, we are interested in the LAN-LAN interworking. Hence the use of ISDN to interconnect LANs and isolated individual hosts/workstations is investigated from this viewpoint in Section 3.3.

3.1.2 Technical Issues in Interconnection

The following issues need to be resolved for a coherent network interconnection:

- **Level of interconnection**
- **Naming, addressing and routing**
- **Flow and congestion control**
- **Access control (security)**
- **Common services (e.g. Internet services)**

Other issues such as buffering, fragmentation and re-assembly, multiplexing and error control must also be considered.

Different networks can be interconnected by 'mediating' systems that are (generically) called *gateways*, corresponding to *relays* in the OSI terminology. The fundamental role of a gateway is to terminate the internal protocols of each network to which it is attached, while at the same time providing a 'ground' across which data from one network can pass into another. Gateways can be used in the following strategies (Cerf and Kirstein 1978):

- **Packet-level interconnection (common subnet technology)**
- **Common network access interface (datagram and virtual circuit)**
- **General host gateways**
- **Protocol translation gateways**

In general, when different protocols operate on separate networks, two approaches can be achieved for interconnection:

1. ***Media conversion* A media conversion gateway bridges the gap between differing data link and physical layer protocols; messages from one network are read by unwrapping their network packaging, the necessary routing is computed, and messages are sent to another network by wrapping them into that networks packaging. Examples of these are repeaters and bridges operating at the physical and data link layers, respectively.**

2. Protocol conversion A protocol conversion gateway bridges the gap between differing network and higher-layer protocols; messages received from one network are replaced by different messages with the same protocol semantics and sent into another network. Examples of these are network or transport layer relays.

In the OSI approach, transport gateways are an anathema, since they are not supposed to be needed because network interconnection is expected to be at layer 3 or below. Gateways are not discussed in any detail, since the OSI model of interconnection described below is adopted.

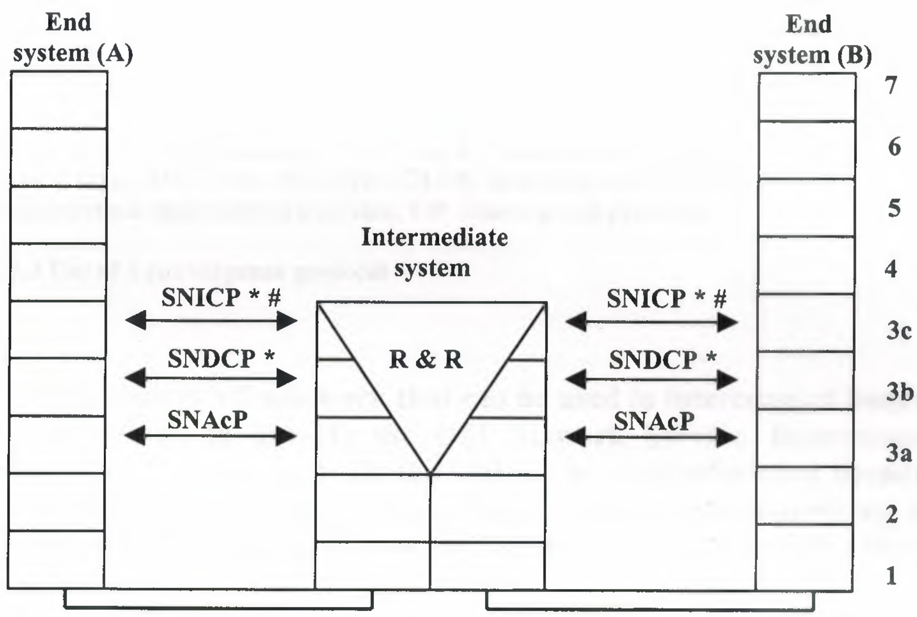
3.1.3 The OSI Model for Interconnection

The OSI-RM (CCITT 1988a; ISO 1984) refers to a seven-layer architecture. The top three layers (layers 5—7) are responsible for the processing of information, while the bottom four layers (layers 1—4) are responsible for moving information units. According to the definition in Section 3.1.2, a co-operation between heterogeneous networks at layers 1—4 providers interconnection, while the full 1—7 layer co-operation provides interworking between end systems and processes. It is the heterogeneity of networking technologies at both the local and wide areas that causes the interconnection problems, since information units must be transported access these networks securely. Interworking necessitates interconnection, but not vice versa. Intermediaries can interconnect heterogeneous networks. A *relay* described by the OSI-RM is such an intermediary. It's the function is to facilitate communications between peers in the protocol hierarchy. A relay implements a set of procedures by which data from one system is forwarded to another. This is called a *relaying function.* A relay sharing a common protocol at layer n with other systems but not participating in a layer $n + 1$ protocol in implementing its relaying function is called a *layer n relay* in OSI terminology.

In the OSI-RM, layers 4 and above operate on an end-to-end basis. Furthermore, OSI principles state that interconnection of data networks (*subnetworks* in OSI terminology) must be achieved at layer 3, the network layer, since this layer provides global network addressing and provides routing and switching as well as relaying functions, among others. Network interconnection can also be achieved at layer 1 and 2; however, as these result in *extended subnetworks*, they do not contravene the OSI principles. These issues are further dealt within Section 3.2.

The OSI-RM specifies that layer $n + 1$ makes use of the layer n service in performing its own service. This leads to the *n-layer service interface* and the *service relay* concept for network interconnection. This assumes that the services operating on either side of a relay are identical (e.g. connection mode network service—CONS). In this case, the relaying function consist of a simple mapping of indications and confirmations arriving through one service interface onto requests and responses sent through the other. Provision of different services on either side causes incompatibilities. In practice, the services offered by real subnetworks (e.g. LANs and X.25 PSPDNs) are rarely identical and usually they do not provide a true OSI network service (CCITT 1998d; ISO 1987a). A further complication is that the OSI Network Service (OSI-NS) can have two guises: the connection mode (CO) and connectionless (CL) network services. If these are operated on either side of a relay, a CO/CL interworking arrangement is needed (ISO 1989). This is often the case when LANs (usually CL) and WANs (usually CO) are interconnected.

According to the OSI principles, service incompatibilities on either side of a relay can be remedied by modifying the service on one or both sides such that a common service is achieved (Lenzini 1984). This aim is made easier to achieve since the OSI network layer functions are divided into three sublayers (Burg and Iorio 1989; ISO 1987c): two convergence protocols and a subnetworking access protocol (SNAcP). The convergence protocols are referred to as the subnetwork independent and dependent convergence protocols, (SNICP and SNDCP) respectively.

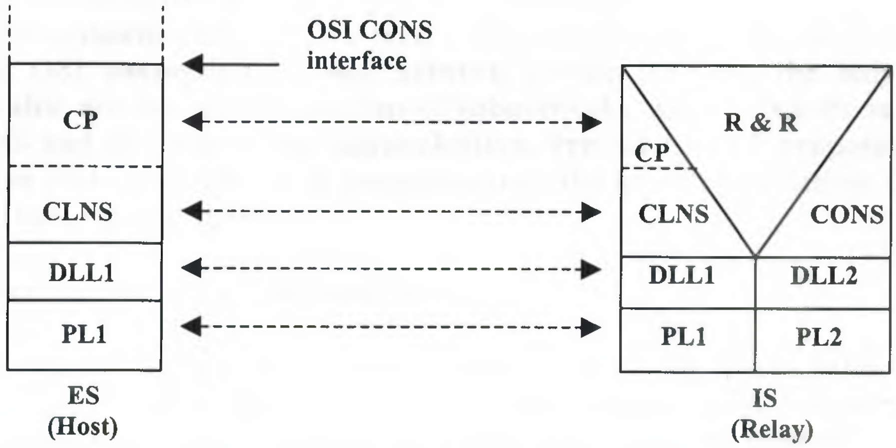


R&R: routing and relaying; SNAcP: subnetwork access protocol
SNICP: subnetwork independent convergence protocol
SNDCP: subnetwork dependent convergence protocol

* May not be needed in all instances.
# These protocols may be identical.

Figure 3.2 Partitioning of network layer

The roles assigned to different sublayers can be provided by a single protocol (e.g. the X.25 Packet Level Protocol (PLP)—as well as by a set of protocols. The role of the convergence protocols is to enhance or de-enhance an existing service to provide the OSI Network Service at the service interface to the transport layer. This extra layer of protocol needs to be added to every network node wishing to communicate with a NL relay configured in this fashion. A practical way to avoid the large number of modifications the above scheme implies is to implement relaying at the transport layer where a commonality can be achieved (using either CO or CL network service modes). However, this compromises the end-to-end nature of the transport service and hence violates the OSI-RM.



ES: end system; IS: intermediate system; R & R: routing and relaying
PL: physical layer; DLL: data link layer; CLNS: connectionless network service
CONS: connection mode network service; CP: convergence protocol

Figure 3.3 Use of a convergence protocol

ISO work providers a framework that can be used to interconnect both OSI and non-OSI networks to provide the OSI Network Service. Interconnection of networks whose SNAcPs provide the OSI-NS is straightforward because of the service compatibility across all participants. When interconnecting networks that include non-OSI networks, two approaches are available under the common service concept:

1. Hop-by-hop harmonization
2. Internetworking protocol

The hop-by-hop harmonization approach prescribes the 'harmonization' of the subnetwork service available on each non-OSI network to that of the OSI-NS. The harmonization functions reside in the SNICP sublayer and possibly, when (de-) enhancement is necessary, the SNDCP sublayer. Routing and relaying functions are employed to interconnect the harmonized subnetwork services to provide the OSI-NS end-to-end. With this approach, all end and intermediate systems attached to a given subnetwork have to support the same sublayer protocol(s) to enable communication between all end system in this subnetwork.

In the internetworking protocol approach, an Internet protocol (operating in the SNICP role) runs over the interconnected series of networks, which may be of different types, providing the OSI-NS across these networks and intermediate systems. An internetworking protocol is based on a predefined set of capabilities over which it is to operate. If a network participating in the interconnection cannot provide a subnetwork service adequate for the capabilities required by the internetworking protocol, a sublayer SNDCP protocol has to be used to enhance its subnetwork service to the level of that required by the internetworking protocol. When this approach is used for network interconnection, the same internetworking protocol has to be used in all end-systems and relays to enable communication between these systems to take place.

Interestingly, the internetworking protocol international standard is the connectionless network protocol (CLNP) (ISO 1988), resulting in the provision of the OSI connectionless network service (CLNS). Furthermore, for the provision of the OSI connection-oriented network service (CONS), the only methods applicable are the interconnection of subnetworks whose SNAcPs support the OSI-NS and the hop-by-hop harnonization. Present CCITT proposals support only the OSI CONS; hence it recognizes only the latter two methods (Burg and Iorio 1989; CCITT 1988f).

3.1.4 Interconnection at Different Layers

As mentioned earlier, interconnection of networks at layers other than the network layer is possible. Here, these are mentioned briefly. It is interesting to note that session or presentation layer relays have so far not been developed.

Data link layer relay: The MAC bridge (IEEE 1988b) is a widely used example of such interconnection. It achieves relating within layer 2 and interconnects LANs. Local and remote bridging techniques are available. Mote than two LANs can be connected via a single bridge. However, MAC layer bridges present problems in areas of address management, security broadcast control, resilience, load balancing and upgradability. Further problems arise in (global) addressing and protocol compatibility when LAN services are to be accessed by remote terminals or workstations.

Transport layer relay: The CO/CL network service interworking without modifications at the hosts, can be established by using a transport layer relay. An ISO Technical Report (ISO 1989) proposes passive and active types of transport layer relays (TLRs). A TLR is called an interworking functional unit (IFU) and achieves interworking between CO and CL network services by relaying and/or conversion of protocol data units (PDUs) from one network type to another. This method assumes the use of a connection oriented transport service on both networks in order for them to be interconnected. However, these modes are outside the OSI architecture, and hence this solution is not an International Standard but a Technical Report.

Application layer relay: Application layer relays are used for mapping largely incompatible (and often proprietary) protocol stacks, masking incompatibilities at lower layers. Examples are terminal and mail gateways. They may also be used for administrative reasons, e.g. to restrict inter-domain traffic to certain applications. Their major disadvantages are the need to develop new gateway modules for each application and set of protocol stacks, and slower performance.

3.2 INTERCONNECTING WITH ISDN

Two issues determine the level of interworking between an ISDN and an external user:

- **Type of user interface (or user interface capabilities)**
- **ISDN services needed, ie. Bearer and/or teleservices**

The type of user interface determines the level of services that can be obtained from the ISDN. For example, a terminal or host supporting only traditional data interfaces, which are non-ISDN-compatible, needs to be connected through a terminal adaptor to the ISDN. This means that it cannot request the full range of services offered through an ISDN-compatible interface. Since teleservices require layers 4—7 compatibility with ISDN, these services are offered only to equipment supporting ISDN interfaces (compatible with S or T reference interworking). When interconnecting LANs and terminal or workstations to ISDN, for cases can be described:

1. **Provision of all ISDN services to LAN users**
2. **Provision of LAN services to ISDN terminals/workstations**
3. **Provision of LAN services to non-ISDN terminals/workstations**
4. **LAN-LAN interworking across ISDN**

The first case necessitates the provision of ISDN-compatible interfaces to the LAN users capable of interworking at the S reference point. This necessitates the support of the S-interface over the LAN and cannot be easily implemented over the existing data-only LANs. With such a functionality the LAN becomes more like an ISPBX. We shall call such a LAN *an ISDN-compatible* LAN (IcLAN). Also, a LAN may be modified to support integrated services, in which case it is called an integrated services LAN (ISLAN). However, an ISLAN does not necessarily need to support the S-interface.

The second case can easily be supported if the LAN is connected to the ISDN via a network layer relay. The third case necessitates the use of either the circuit or packet mode bearer services by both the LAN and the non-ISDN terminal/workstation for data transfer. Finally, the fourth case necessitates the access of ISDN switched or non-switched bearer services for interconnection and can be implemented to operate at the R, S or T reference point. Table 3.1 summarizes the ISDN services needed and or useable in each case.

3.3 INTERCONNECTION OF LANs AND ISDN

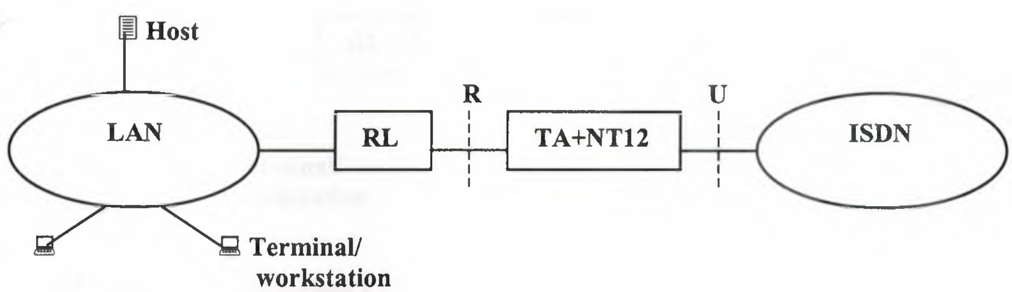
The choice of reference points with regard to the usage scenarios described in the previous section is given in the following subsections. In all of these scenarios, we assume that the LAN is connected to the ISDN via a relay, as opposed to individual LAN hosts having direct access to the ISDN.

Table 3.1 Services in LAN/workstation-ISDN interconnection

Interconnection type	ISDN services: Bearer service	ISDN services: Teleservices
T/WS-ISDN-Host/Server	√	–
T/WS-ISDN-LAN	√	–
LAN-ISDN-LAN	√	–
LAN-ISDN	√	–
IcLAN-ISDN	√	√
IWS-ISDN	√	√
IWS-ISDN-IcLAN	√	√

T/WS: terminal/workstation; IWS: ISDN workstation;
IcLAN: LAN modified/designed to support S-interface.

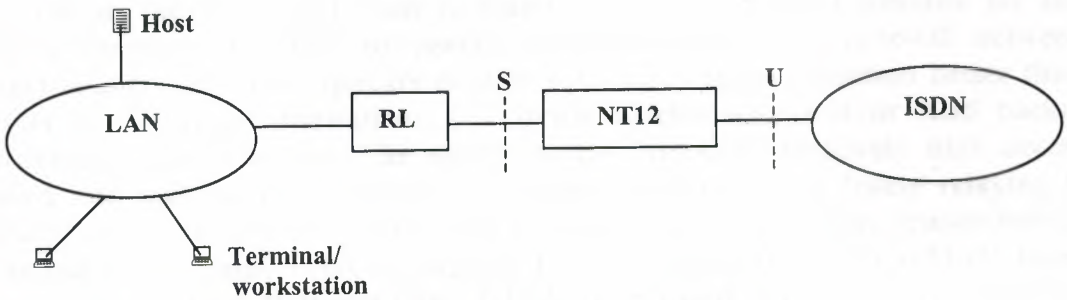
Interconnection at reference point R: The LAN-ISDN relay can have two types of interfaces: circuit or packet mode, hence acting as a circuit or packet mode DTE, respectively. In either case, a compatible terminal adaptor is used to interwork with the S-interface. Thus, usage cases 2, 3 and 4 can be supported with this configuration. Ordinarily TAs will support only one B- or H-channel at the ISDN side, thereby limiting the Internet traffic that can be handled. However intelligent TAs, which can aggregate B- or H-channels in ISDN, are also possible, and these could support larger-capacity connections. Figure 3.4 shows the LAN-ISDN interconnection at reference point R.



RL: relay; TA: terminal adaptor; NT12: network termination 1 and 2

Figure 3.4 LAN-ISDN interconnection at reference point R.

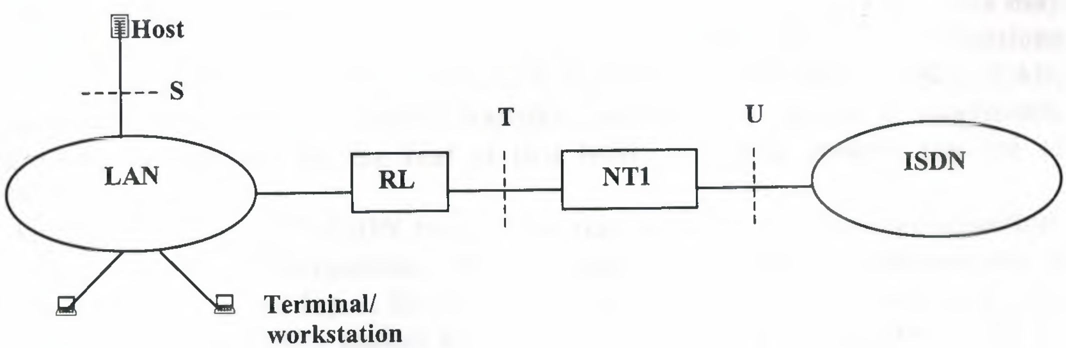
Interconnection at reference point S: This configuration, shown in Fig. 3.5, can be used to support usage cases 2, 3 and 4. The advantage of this configuration is that it provides direct access to the S-interface by the LAN-ISDN relay, which can use the flexibility of the control of that interface by ISDN signalling procedures. In this mode, LAN-relay combination acts as an intelligent terminal adaptor. This is the preferred interworking point for the purposes.



RL: relay; NT12: network terminatio 1 and 2

Figure 3.5 LAN-ISDN interconnection at reference point S.

Interconnection at reference point T: This is the only configuration where usage case 1 can be supported in addition to cases 2, 3 and 4. The LAN needs to be interconnected at reference point T to the ISDN. The LAN acts as an NT2 and provides the multiplexing and distribution functions as well as the S-interface to LAN users. However, a LAN supporting the S-interface needs to be designed carefully in order to benefit from this configuration, as none of the current LANs provide this facility, nor can normal LAN hosts exploit it.



RL: relay; NT1: network termination 1

Figure 3.6 LAN-ISDN interconnection at reference point T.

3.3.1 Using the ISDN Bearer Services

In chapter 2, the data communications in ISDN were classified under the circuit- and packet-switched bearer services and the additional packet mode bearer services (APMBS). The packet-switched bearer services are currently available only using the X.25 protocol suite, and the APMBS is not expected to be available in Europe in the immediate future.

The frame-relaying type 1 (Lai 1989a, 1989b, 1989c) implies a core layer 2 service on the B-channels and is found to offer significant benefits for the interconnection of LANs by *remote bridges*. Indeed, its trade-off between functionality and speed appears to suite LAN-LAN interconnection better than either of the other alternatives, i.e. circuit-switched channels or X.25 packet switching. This is because the simple access interfaces relatively high access speed and statistical multiplexing capability (in layer 2) of frame relaying is ideally suited to LAN interconnection (Lamount *et al.* 1989). The remote bridge solution using frame relaying suggest the encapsulation of LAN MAC layer frames in the LAP-D frame (Rec. I.441). High-speed operation is possible by transparent operation because of the non-termination of higher-layer protocols and the use of core layer 2 functions only (e.g. no frame acknowledgment). Owing to layer 2 multiplexing, high-speed access is more cost-effective, especially to multiple destinations. This facility is similar to the multiple virtual circuits in X.25 networks but at reduced protocol overhead and associated speed/cost penalty of layer 3 multiplexing.

However, because of the additional requirement of enabling isolated workstations to be able to access the LAN services and the earlier availability and ubiquity of the circuit-switched service, it is this ISDN bearer service that will be referred to in the rest of this book. Therefore, LAN interconnection using either the packet mode bearer service or the APM bearer service is not discussed further in this chapter.

A further requirement in LAN-LAN interconnections is adequate bandwidth provision. In cases where off-site traffic is light enough, a basic-rate access may suffice. However, when off-site traffic volumes are high and LAN applications best served by high-bandwidth links are to be used (e.g fast file transfer, CAD, bit-map and multi-media document transfer), primary rate access arrangements need to be considered. In the rest of this book, we shall assume the use of PRISDN).

Another issue in LAN-ISDN interconnection is the layer of interconnection. Following the earlier discussions, we shall assume that the interconnection is achieved at the network layer. Hence the network service is to be provided over the circuit-switched ISDN digital bit pipes. The protocol stacks that could be used over these bit pipes are discussed in Section 3.3.5.

3.3.2 Application Running on LANs

Although it is difficult to predict future communication needs and patterns of various user groups, all existing data communication applications could be supported over the ISDN. Furthermore, it is expected that remote terminal, file and database access, fast file transfer, distributed databases, CAD applications, electronic mail, multi-media document service and telematic services application will be prominent in the immediate future. It is obvious that some of this applications will not be able to use the full 64 kbps channel capacity when offered, for example, a remote terminal access session. Several such sessions can therefore be multiplexed onto a single channel if they pass through the same pair of ISDN stations (access nodes). Also, interactive applications such as terminal access and file access need 'real-time' responses in order to be useable. Delay becomes more important than throughput in these cases.

3.3.3 Interconnection Scenarios

The following LAN-ISDN-LAN interconnection scenarios can be visualized:

- *Scenario 1* A fixed number LAN of sites communicates with each other on a regular basis (e.g. branches of a corporation) and has regularly distributed traffic.
- *Scenario 2* Users on a large number of LANs want to communicate with each other on a irregular basis.

In scenario 1, a leased-line or semi-permanent line solution may be more cost-effective than a circuit-switched solution, as the advantage of ISDN will not be evident if the tariffing structure is not favourable to switched connections. However, the ISDN could provide added bandwidth on demand or act as an emergency backup facility. In scenario 2, a circuit-switched solution will be more cost-effective. Hence a judicious management of the circuit-switched connections under different tariff structures, user demand (traffic volume) and service/user priority constraints will bring about cost-effective utilization of the local ISDN bandwidth. It seems more likely that an advanced automated office environment will need the flexibility of scenario 2 for its applications.

3.3.4 The LAN-ISDN Relay

The LAN-ISDN relay is a specialized relay, which must overcome several technical problems. It differs from most other relays is that:

- It is multichannel.
- It must mange many switched connections simultaneously.
- It has common channel signalling for outband control of circuits.
- It connects to a network (the ISDN) which imposes no restrictions on the protocols to be used the B-channels appearing merely as bit pipes (assuming circuit-switched bearer service).

These properties present to following problems:

1. Intelligent management of multiple channels for switched connections; circuit set-up and removal actions as well as dynamic bandwidth control.
2. Efficient queue management for channels.
3. Conversion of inband data signalling (e.g. addressing, routing and other packet information) on the LAN to outband control signals for circuit-

switched connection set-up on the ISDN. This is basically a way of protocol conversion from the single-plane format of the OSI-RM to the two-plane (user and control planes) format of the ISDN-PRM.

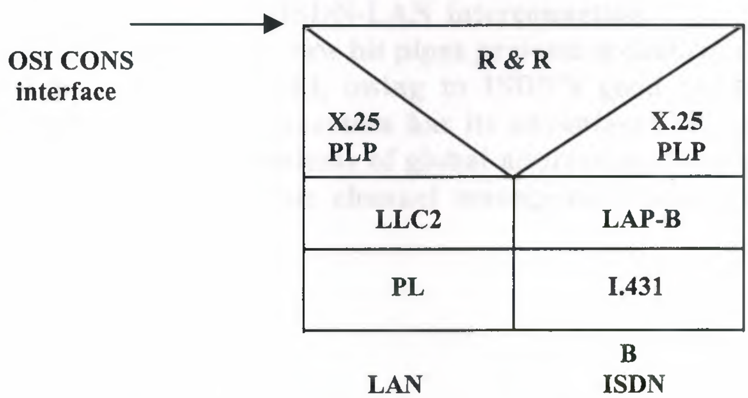
4. Choice of protocol stacks to be used over bit pipes.

In the local area, the workstation/server/LAN system model has become the most popular environment, and this must be interfaced to other LANs via the ISDN. A LAN-ISDN relay with a LAN interface on one side and a PRISDN interface on the other is needed. The capabilities that required from this relay must reflect service requirements as well as the underlying communications requirements. The LAN-ISDN relay must be able to manage multiple channels, which may be switched individually or in bundles, depending on user/network requirements. Procedures need to be defined to manage a variable number of B-channel connections depending on system status and traffic variations.

3.3.5 Protocols Over the ISDN Bit Pipes

Another issue in the LAN-ISDN relay is the type of protocols to be supported. Both the CO and CL network services can be provided over the digital bit pipes provided by the ISDN circuit-switched bearer service. This leads to different protocol stacks in the LAN-ISDN network layer relay. For simplicity, the same type of network service is assumed in each of the scenarios described below. Here, we note that other possible scenarios exist (e.g. X.25 tunnelling, where connectionless packets could be carried within connection mode packets, etc.). However, there are not discussed, as they are not relevant for the rest of the book. In both cases, only the B-channels are shown.

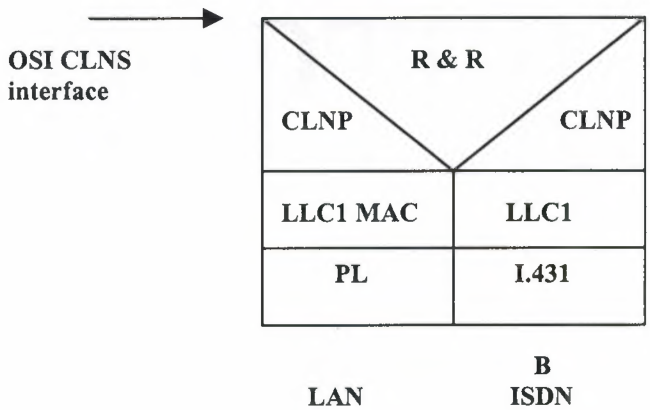
<u>CONS across ISDN</u>: Figure 3.7 show a NL relay protocol stack where the CONS are provided across both the LAN and the ISDN. At the network layer both the LAN and the ISDN are assumed to have the X.25 packet layer protocol (PLP). The lower layers are necessarily different. When a new B-channel is set up, the data link must be established over it.



PLP: packet level protocol (IS 8208); LLC2: logical link control type 2 (IS 8802-2)
LAP-B: link access protocol-balanced (IS 7776); PL: physical layer;
R & R: rouiting and relaying

Figure 3.7 A connection oriented LAN-ISDN network layer relay.

CLNS across ISDN Figure 3.8 show a NL relay protocol stack where the CLNS is provided across both the LAN and the ISDN. At the network layer both the LAN and ISDN are assumed to have the CL network protocol (ISO Internet protocol). At the data link layer, the ISDN need not have the MAC-level protocol. Indeed, HDLC core protocol could also be used instead of the LLC1. If LC1 is used, not data link needs to be estavlished after a circuit set-up since this is a mode layer 2 protocol. If HDLC is used, however, the data link needs to be established anew with every new circuit set-up. The interaction of CLNP and the channel management are studied further in Chapter 4.



R & R: rouitng and relayink; PL: physical layer;
CLNP: connectionless network protocol (IS 8473); LLC1: logical link control type 1 (IS 8802-2)
MAC: medium access control protocol; I.431: ISDN physical layer

Figure 3.8 A connectionless LAN-ISDN network layer relay.

3.4 SUMMARY

Circuit-switched ISDN, providing fast switching, multiple channels, common channel signalling and digital bit pipes, presents a flexible wide area network for LAN-ISDN and LAN-ISDN-LAN interconnetion. The user flexibility to choose suitable protocols over raw bit pipes presents a challenge. These protocols do not need to be heavy-weight, owing to ISDN's good bit error rate performance. Network layer interconnection has its advantages in conforming with the OSI-RM and eases the problems of global addressing. Multiple user channels at the UNI precipitate dynamic channel management issues, which are dealt with in Chapter 4.

CHAPTER FOUR

CHANNEL MANAGEMENT

The OSI Reference Model (ISO 1984) specifies that relaying should be done at the network layer (NL or below. (NL is favoured for interconnecting on a very large scale.) In the traditional sense, the NL relay has fixed transmissions lines (e.g. leased lines as opposed to switched), linking it to one or more destinations (e.g. other relays, LANs or packed – switched network inter working units). However, a NL relay operating in packed mode over the circuit – switched ISDN has multiple channels that can be assigned on demand. This gives an added flexibility to relay controlling the ISDN access interface so that can assign bandwidth to requests dynamically. This chapter investigates the implications of this flexibility in terms of channel and bandwidth management and protocols.

Packet operations can be carried out in connectionless (CL) or connection – oriented (CO) modes, as described in Chapter 3. Assuming, for simplicity, CO transport and data link services, two sets of protocol stacks emerge: one with a CO network service (CONS) and one with the CL network service (CLNS). The mechanism of channel management is necessarily different in the two types of network service. In either case, end-to-end associations are established by transport layer entities wishing to communicate with other peer entities. Furthermore, in both cases the establishment of data link service is necessitated once new physical channels are switched in. This is in contrast to the permanent link case where the data link connection is established once at the start-up time and is assumed to exist permanently.

From the channel and bandwidth management aspects, two questions arise:

- When should connections be set up and removed?
- How can the bandwidth of existing connections be varied?

The first question is not as trivial as it sounds. In the first instance, a certain amount of bandwidth must be allocated from the bandwidth pool to a new request. This may be subject to some form of access control. Furthermore, a network layer relay has no knowledge of to end-to-end associations (the duty of the transport layer protocol), and hence it needs to have a strategy for setting up and removing circuit-switched connections. The second questions raise the issue of the type of policy to be used to determine the instantaneous bandwidth of a connection. In both cases, issues such as the use of signalling protocols for channel management, practical mechanisms involved in bandwidth variation (e.g. channel aggregation) and provision of quality of service also arise.

The burst nature of packet data lends itself to statistical multiplexing. This is an inadvertent feature of the CL mode of communication, and the same can achieved by the use of multiple virtual circuits (VCs) over the same physical link is usually assume to be fixed. The idea of statistical multiplexing can naturally be extended to the LAN – ISDN interconnection. An added complication here is that the bandwidth of the underlying physical link between two entities may vary. Note that this form a model similar to the one considered by Harima and Leslie (1979), where variable bandwidth circuit-switched ISDN links using ATM-type packet structures are utilized in interconnecting LANs over a wide area.

In a multi-channel NL relay, there may exist multiple channels (physical or logical) to the same or different destination currently. The NL relay dispatches to their destinations according to some specified routing strategy. As mentioned in chapter 1 and 2, multiple parallel channels or a single super channel to a given destination may be created. When multiple parallel channels a given destination exist, the question of packet scheduling to these channel queues arises. As the traffic load (i.e. the aggregate bandwidth requirement) of a set of multiplexed traffic streams may vary over a given period of time, so will the channel (i.e. connection or link) capacity needed to that destination vary. This can depend upon the changes in the number of users or in the types of activities they partake in during the course of an end-to-end communication. It would be wasteful if a large amount of capacity were allocated to the communication needs of group of users to a given destination when, as a result of changes in their traffic activity, they no longer needed the same high capacity. It would also be unfair to other users who might want to use the same resources. It pays, then to adjust the allocated capacity (bandwidth) of a channel according to the traffic activity, some specified QoS parameter and predefined cost structures. We call this collection of management activities channel management, as it involves the interaction of several related but distinct disciplines.

The channel management problem is described in sections 4.1 and 4.2 Sections 4.3 and 4.4 present the bandwidth allocation and management issues. Queuing systems and queuing strategies relevant in channel management in topic of Section 4.5. Section 4.6 presents a short summary of the possible scheduling policies that could be used in multi-channel interfaces. The effects of CO and CL protocols used over the ISDN B-channels on channel management in a network layer relay are examined in Section 4.7. Section 4.8 presents a short discussion on the management information needed for effective channel management in network layer relays. Cost factors are briefly studied in Section 4.9. Finally, a summary of the chapter is given in Section 4.10.

4.1 THE CHANNEL MANAGEMENT PROBLEM

Frequently in computer network analysis, issues related to each other are, out are necessity, studied and analyzed separately for case of approach. Examples are bandwidth allocation, bandwidth management, scheduling, design and control of queues, channel assignment and transmission control. However, these issues seldom occur in isolation, and they need a unified approached the achieve the best practical understanding. Channel management is such a problem, as it encompasses several distinct but interrelated sub-problems. Furthermore, the decisions at policy level usually effect the performance at different levels of communications hierarchy by varying amounts, and their collective effect also needs to be investigated. Some definitions for the above issues are as follows:

- *Bandwidth Allocation* (BA) This relates to the problem of partitioning the bandwidth resource at a communication facility among a group of homogeneous users. Usually, it is assumed that each user has a static requirement, specified at the start of its communication activity.
- *Bandwidth Management* (BM) This concern the problem of managing the bandwidth of an individual user a group of users as its/their requirement change over time during a connection. Hence it is assumed that a bandwidth allocation strategy already exist and that some amount of bandwidth has been assigned to this/these application. The bandwidth management policy than operates within the confines of the BA strategy such that the bandwidth of a given channel is varied (increased-decreased) accordingly.

- *Scheduling* This describes the type of policy that is used for dispatching packets addressed to a given destination when there exist multiple parallel channels with individual queues. This problem does not exist if only one packet (user/customer) queue exists per destination.
- *Design and control of queuing systems* Classical queuing theory deals with queuing systems in a descriptive way. Dynamic control of queue systems necessitates a prescriptive approach. Most real-life situations are not static, and the control of queuing systems as a discipline is of prime importance in channel and bandwidth management.
- *Channel assignment and transmission control* Channel assignment describes the produces for assigning channels to incoming requests. It includes channel hunting and channel de-assignment for cost minimization. Transmission control involves the use of the signalling procedures for link establishment and removal and data packet handling (transmission and reception).

In this chapter, we concentrate on the dynamic channel management problem in network layer relays in general and in the connectionless NL relay in particular. Some of the issues also apply to equipment directly interfacing the ISDN (rather than through a relay). The control of queuing systems comes into play in the design of bandwidth control strategies. Some aspects of the allocation and dynamic bandwidth management issues are also considered in relation to the channel management problem. Scheduling issues are discussed in relation to the management of multiple parallel channels between two communication end-points. Channel assignment and transmission control is also relevant to the discussion; however, these issues are discussed in Chapter 5.

4.2 DEFINITION OF CHANNEL MANAGEMENT

In designing NL relays between LANs and the ISDN, it is necessary to manage the ISDN resource in the most effective way in order the meet several conflicting criteria. Bandwidth utilization user-perceived response, delay, throughput and real communication costs must all be considered. Therefore, the main objective of channel management maybe defined as the optimization of bandwidth utilization (hence the reduction of costs) for a given set of users with a given quality-of-service (QoS) distribution. This applies the multiple distinct channels (or a single channel with controllable multiple basic bandwidth units) and describes their management.

A collection of policies regarding bandwidth allocation, bandwidth management, scheduling, channel assignment and transmission control forms a channel management (CM) strategy. A CM policy can be implemented by having controls at two distinct levels. First, the bandwidth allocation policy level determines how the available bandwidth is distributed among the different requests and hence provides an admission (access) control policy. Second, the bandwidth management policy level determines how the bandwidth of an existing channel should be varied as traffic characteristics vary in time. In the case where multiple parallel channels with individual queues exist to given destination, a third level of control is available. This is scheduling policy level, which determines how the traffic is “routed” to different parallel channels established to the same destination. The scheduling policy in turn provides a handle on the queuing policy to be adopted for the outgoing links. The queuing policy may be QoS-associated.

Thus, channel management encompasses the bandwidth allocation, bandwidth management and scheduling policies. The BA policy is needed to prescribe from the outset the “rules” of the game, that is, how much bandwidth each user class or group may have, while the BM policy determines how the bandwidth is varied (increased or decreased) from the initially allocated amount within the limits prescribed by the BA policy. The scheduling policy determines how the packets are routed to different channels connected to the same destination (and forming part of the same link) in order to achieve some QoS or performance criteria.

In the rest of this book, we will concentrate on the channel management problem, in particular the channel set-up/disconnect and the dynamic bandwidth management issues. But first, discussions on bandwidth allocation, control of queuing systems and scheduling are presented in order to highlight their relevance.

4.3 BANDWIDTH ALLOCATION

Bandwidth allocation policies determine how the total available bandwidth of a communication channel is allocated (partitioned) to multiple user (customer) requests. It is also sometimes referred as circuit-channel access control policy since the total bandwidth is partitioned into ‘channels’ (Kraimeche and Schwartz 1984; Schwartz and Kraimeche 1982). Most BA schemas are based on a *prescriptive* method. This means that each user (user process, or *protocol data unit*) requests certain amount of bandwidth at the start of each session, and is content with that amount of bandwidth for the duration of a session. Also, there are limits to the amount of bandwidth a class of users is allowed to occupy in a situation of bandwidth connection with other classes of users.

An important design aspect from the BA policy point of view is the composition of the users. Generally, they are classified into two types: blockable and queueable (class 1 and class 2, respectively). These two classes are sometimes referred to as circuit and packet types of user. However strictly speaking this is not true, since packetized voice would still be in class 1 despite its packetization. This terminology merely indicates the type of service required by a class of users. Hence the circuit users are assumed to be operating on a blocked-call-lost principle. Different BA policies can be devised on the assumptions of homogeneous or heterogeneous block able/queue able users. Traditionally, the circuit and packet traffic has been treated separately. Each type of traffic would use different types of networks and hence switching. In the ISDN era this no longer true. Switching systems handling both the circuit and packet types of users are sometimes called *hybrid switching* systems.

One way of dealing with mixed traffic types is to separate the available bandwidth into the ‘circuit’ and ‘packet’ groups. In a TDM frame, partitioning available time slots, N, into two groups, N_1 and $N_2=N-N_1$ slots. A boundary is thus formed between the class 1 and class 2 traffic types where N_1 slots are assigned to the block able class while N_2 slots are reserved for the queue able class of users. This boundary can be fixed or movable boundary scheme, class 2 packets may occupy any of the unused N_1 class 1 slots, with a proviso that any arriving class 1 call can pre-empt a class 2 packet occupying these slots. The movable-boundary hybrid-switching scheme is found to be superior to the fixed-boundary scheme in its improved time delay performance for the queued class of users. However, care must be with the movable strategy. Any attempt to increase the class 2-user throughput above which is normally available by N_2 slots reserved Per frame, by the utilization of class 1 user slots released on the average, may lead to exceptionally long packet queues (Schwartz 1987; Schwartz and Kraimeche 1982).

In either case, parts of the total bandwidth may be reserved as a general shared pool, which could be equally accessible by all user classes or could have some form of restricted access superimposed. The prescriptive nature of these policies guarantees a certain QoS, or resource utilization, and fairness to contending users (Schwartz and Kraimeche 1982).

In the case where only blockable (circuit-switched) type traffic is available, bandwidth allocation can be done on the homogeneity/heterogeneity basis. Some of the previously studied strategies for BA in circuit-switched systems include: complete sharing (CS), complete partitioning (CP), restricted access with priority (RA-P). In complete sharing, no explicit control is exerted on the access of the K circuit streams and an arriving message is rejected only if it requires a bandwidth greater than the available amount. In complete partitioning, the total bandwidth is divided into K separate portions such that each circuit stream has its own decided bandwidth. In the restricted access type of policies, the users are grouped into K user types where each type is characterized by its bandwidth requirement; RA policy then restricts the resource occupancy of users from each group, hence reserving part of the total bandwidth to each group. An RA policy with priority permits users of lower bit rate to access the bandwidth reserved for users of higher bit rate which have a right to pre-empt lower bit rate users.

It is known that, for homogeneous circuit traffic (i.e. same bit rate and holding time), the CS policy yields best throughput (Kraimeche and Schwartz 1984). For non-homogeneous traffic with a large spread in its bandwidth requirements (i.e. bit rates), a hybrid CP policy, where the bandwidth dedicated to given group is completely shared by user types in that group, outperforms any other simple policy. It is also known that a RA policy augmented by a priority scheme provides improved performance over other policies described (Kraimeche and Schwartz 1984).

4.3.1 Packet Data over Circuit-Switched Connections

Bandwidth allocation in the above sense can be achieved only if the bit rate requirements of users/application are known in advance. That is, when such users arrive at a resource facility, they request a certain amount of bandwidth. They could then be allocated their requested bandwidth according to the BA policy implemented by the resource manager.

This form of BA is suitable only in the CO mode of inter working as the QoS parameters (indicating bandwidth or data rate required as well as the acceptable delay and throughput parameters) for each CO session are agreed upon between the user and the resource managers at the intermediate nodes of communications link at the start of an association during the set-up procedure.

In the case of CL working, the QoS is not negotiated at the start of communication but is determined in an a priori fashion between the users and the service providers. Also, as there is no connection set-up phase in this mode operation, the bandwidth requirement of an application cannot be determined. Furthermore, the QoS parameters of the connectionless network service do not include bandwidth. In CL network layer relays, traffic multiplexing is an inherent feature, although this may be specifically prevented in order to guarantee a QoS. However, the CL networks operate on the best-effort principle and do not guarantee a QoS. When several traffic sources are multiplexed onto a physical channel and the occurrence and duration of these sources vary stochastically, the bandwidth requirement of the channel also varies in a stochastic manner. Under these conditions, the BA policy may actually limit only the upper bound of bandwidth allowable for each channel. It is then left to the interaction of the user traffic and the bandwidth management policy to determine the amount of instantaneous bandwidth allocate to such a channel. In the rest of this study, this form of BA control will be assumed.

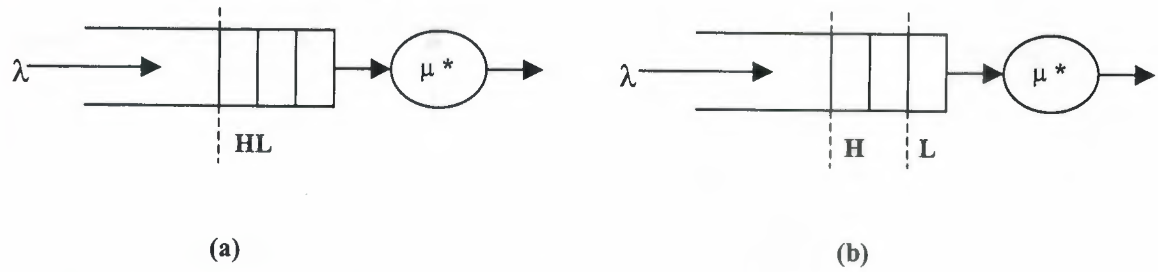
4.4 BANDWIDTH MANAGEMENT

This is the resource management function associated with the adaptation of user bandwidth to the needs of user traffic, which may vary during the lifetime of a connection. We have already considered scenarios where such variations in bandwidth requirements may occur.

The bandwidth of a TDM access structure is divided into channels, each of which may be mapped to one or more time slots. Each TS can be considered to be basic bandwidth unit. As the traffic load of a stream (or a multiplexed stream) increases over time, the bandwidth allocated to this user should be increased in order to provide a constant QoS performance. This bandwidth increase can be put into effect by associating more TSs to this channel.

At the core of the bandwidth management problem is the question of how and using which type of control this increase and decrease in channel capacity (bandwidth) should be achieved. Service capacity (rate) control of queuing systems using threshold policies based on the system state is well known (e.g. Gebhard 1967; Moder and Phillips 1962; Romani 1957; Yadin and Naor 1963). These policies could be based on single or multiple thresholds and point or hysteresis forms. Figure 4.1 shows a schematic of point and hysteresis threshold forms when the system state used is the instantaneous queue length. A *point threshold* policy implies that decision to increase the service capacity of the server is made as the system state value 'crosses' the threshold value form below; a decision to decrease of bandwidth is made as the system state value crosses the threshold from above. A *hysteresis threshold* policy has two such levels. A decision to increase the bandwidth is made when the system state value exceeds the high threshold. This capacity is held until the state drops back again to the low threshold value, whence a decision is made to decrease the bandwidth. This is could a two-phase hysteresis. More complex schemas with a multiple threshold pairs that may or may not overlap can be devised. Also the level of control could be based on either bi-level service rate control could be based on either bi-level or multi-level types. In the bi-level service rates, μ_1 or μ_2. Whichever is selected depends on the state of the control variable. In multi-level service rate control, the system can have different service rates selected from a set of $\mu(\mu_1,\ldots,\mu_N)$.

Point threshold control is a special case of the hysteresis control in that the two threshold levels are coincident. This form of control is better at conserving resources but can produce a very high number of switching transients when the system is operating around the threshold value. Another special case of the hysteresis control is one where the lower threshold is set at zero level of the control variable (Gebhard 1967).



H: high threshold; L: low threshold; λ and μ: mean packet arrival and service rates, respectively. HL: point (high-low) threshold; μ *: variable capacity server

Figure 4.1 An example of (a) point and (b) hysteresis thresholds.

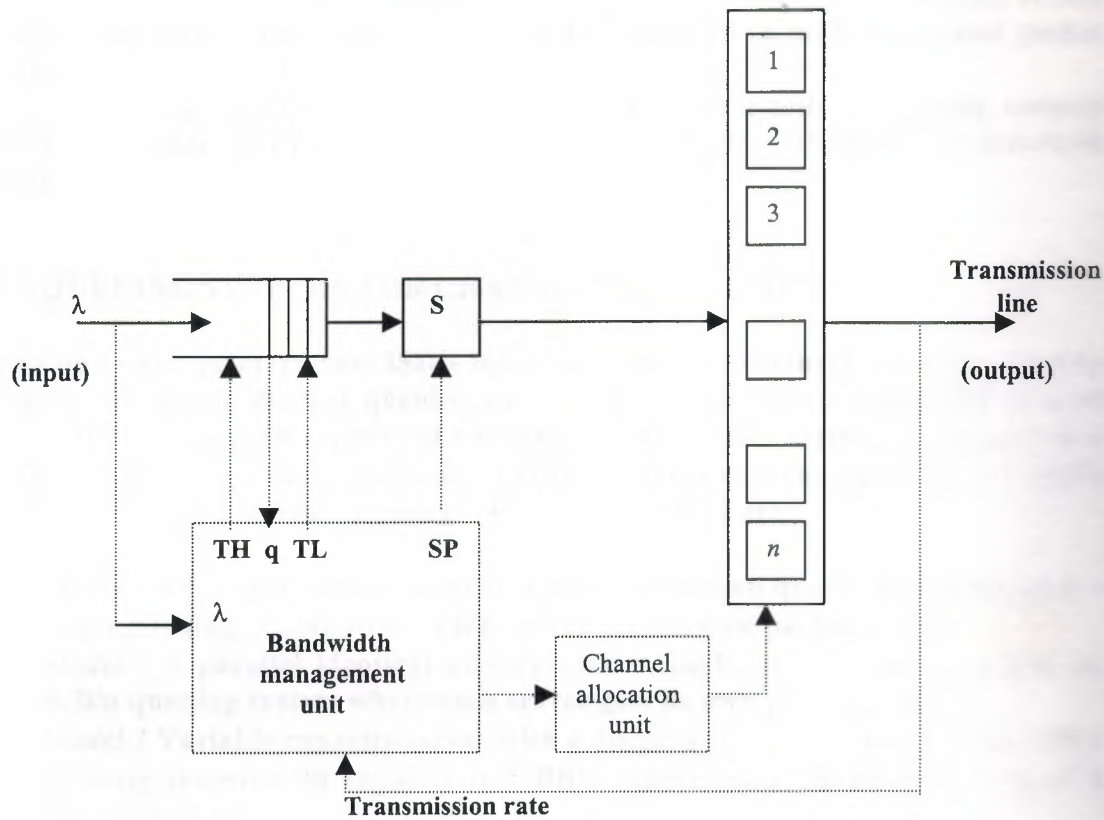
4.4.1 Control Variables and Metrics for BW Management

Various control variables and metrics are available for service rate control in queuing systems. These include the instantaneous or averaged values of the number of customers in the queue or the system, the amount of unfinished work in the system (i.e. the actual number of bits to be transmitted rather than the number of packets in the system) and the packet arrival rate (or the tarffic bit rate).

Figure 4.2 shows the control variables and modules for dynamic bandwidth management in a multi-channel packet communications facility as applied to an aggregate packet stream.

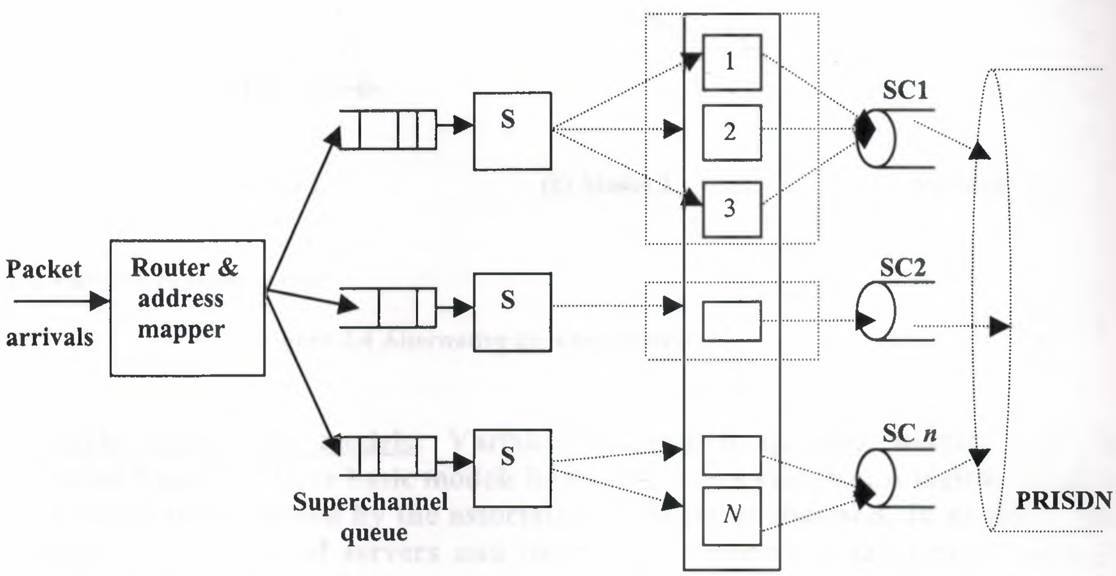
The bandwidth management unit (BMU) monitors the system state periodically or on each new packet arrival. This includes the queue length and the data input rate on aggregate traffic streams destined to the same remote location or device (e.g. a network layer relay). The BMU signals the channel allocation unit (CAU) for the implementation of any bandwidth increase/decrease operation. A packet or job schedular operates to determine which channel (indicated as 1, 2, ... , *N* BBUs) receives the next packet or byte. The schedular operates according to an assigned scheduling policy.

For multiple aggregated traffic streams, the queue and schedular components are replicated. The output transmission rates of individual aggregated channel groupings must also be input separately into the bandwidth management unit.



TH: threshold high; TL: threshold low; q: current queue size
SP: scheduling policy; S: packet scheduler; λ: packet arrival time

Figure 4.2 Components of bandwidth management in a multi-channel packet communications facility.



S: packet (byte) scheduler SC *n*: superchannel *n*

Figure 4.3 Formation and use multiple superchannels in a network layer relay.

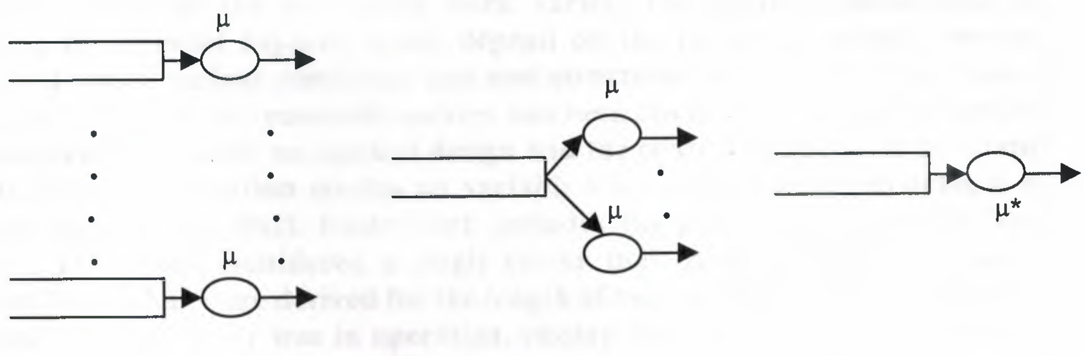
Figure 4.3 shows the formation of multiple superchannels in a packet switch or network layer relay. The BMU of Fig. 4.2 is applied to each aggregated packet stream in superchannel queues.

In queuing theory, bandwidth management corresponds to queuing systems with removable or variable capacity servers. This class of queues is described next.

4.5 QUEUING SYSTEMS FOR CHANNEL MANAGEMENT

The basic and primary-rate ISDN interface channel structures were discussed in Chapter 2. Three distinct queuing models can be developed depending on how the *n* time slots (each capable of carrying a byte of information per frame) in a TDM frame are utilized. Assuming that the interface is composed of n × BBUs (i.e. 64 kbps) and no QoS queuing is undertaken, these are:

- *Model 1 n* parallel identical servers each with its own queue. This gives an n × A/B/1 queuing system where each server gets its own packet to serve.
- *Model 2 n* parallel identical servers with a single queue. This results in an A/B/n queuing system where each server gets its own packet to serve.
- *Model 3* Variable capacity server with a dingle queue. This results in an A/B/1 queuing system with capacity n × BBU, where each TS serves a byte of a packet in turn.



(a) Model 1 (b) Model 2 (c) Model 3

μ*: Variable capacity server

Figure 4.4 Alternative queuing models

Variable bandwidth models: Variable bandwidth queuing models can be obtained from the above basic models by having n as a variable. A logical channel is assumed to be formed by the association of physical channels. In model 1, the addition or deletion of servers and their queues can vary the logical channel bandwidth. When multiple servers with individual queues to the same destination exist, the problem of packet scheduling arises. In model 2, a BBU can be associated with, or dis-associated from an existing logical channel queue, thereby providing a variable bandwidth model. In model 3, each added TS is closely integrated with the existing TSs, thereby effectively increasing the server capacity. Simply dropping some TSs from the logical channel can reduce server capacity.

In practice, models 1 and 2 present a re-sequencing problem at the destination since this form of transmission can reorder packets. Response time performance of model 3 is superior to model 2, which itself is superior to model 2, which itself is superior to model 1. However, model 3 presents a time slot sequence integrity problem. These issues and the practical ways of implementing variable bandwidth channels in an ISDN interface are described further in Chapter 5. In Chapter 5 dynamic channel management architecture is presented for models 1 and 3. This architecture can easily be modified to address model 2.

Variable bandwidth control for logical channels: For models 2 and 3, this can be achieved simply by use of threshold policies described in Section 4.4. In model 1 two options exist regarding the use of threshold policies for service capacity control: the use of a logical queue length as the control variable, and the use of separate thresholds on each queue. The latter could be combined with the scheduling policies but inevitably leads to a more complex control strategy.

4.5.1 Queuing Systems with Removable Servers

Queuing systems for channel management in ISDN form a class of models that includes add able/removable servers. In general, the performance of these systems can be studied by considering queuing systems in which the service capacity varies as the unfinished work varies. The decision mechanism for varying the channel capacity could depend on the buffer occupancy, the cost incurred under various conditions and cost structures in the model. The class of queuing systems with removable servers has been studied by various researches. A comprehensive study on optimal design and the control of queues is by Crabill et all. One of the earliest studies on variable bandwidth queuing systems is by Moder and Phillps(1962). Later work includes the paper by Yadin and Naor (1963). This study considered a single server that could be either present or absent. Formulate were derived for the length of busy periods and the proportion of time that the server was in operation, closing down, starting up or idle. In a later paper (Yadin and Naor 1967), a system in which the service capacity could vary was considered. As the number of customers waiting increased, at various threshold the server capacity would increase; the service capacity would be decreased as the number of customers decreased past different thresholds. In the latter paper, the addition of new service capacity was made instantaneously. Several other authors have dealt with variable-rate servers, but they have all assumed zero dial-up time (Bell 1980; Levy and Yechiali 1975), which is not realistic in telecommunications.

In a study by King and Shacham (1986), two models of buffers with dial-up capability are analyzed. In the first model, a buffer that has a permanent and a dial-up channel connected to its output is analyzed. The dial-up channel is added into the system forming an M/M/2 queuing system whenever the occupancy increases above some threshold. When the buffer is empty, the overflow channel is released. A second model considers a system that consists of two buffers that *use dial-up channels to communicate. Each node that has its buffer occupancy above the threshold dials up a channel. A connection, once established, is full* duplex, allowing data to flow in both directions. Thus, when the first dial-up is complete, the other buffer, if it is in the middle of a dial-up, aborts that process and uses the existing channel to send messages. The channel is released when both buffers are empty.

In this class of queuing systems it is known that a threshold policy, i.e. initiating connection establishment on the arrival of the Kth message to the buffer, is optimal (Heyman 1968). The problem then becomes one of finding the optimum threshold that will minimize the total cost owing to channel set-up and to delays incurred in the queue. This threshold depends on system configuration, rates of service and arrival and various costs.

4.5.2 Threshold Control of Server Capacity

It is known that the M/M/1 queue has a monotone hysteresis optimal service rate control policy when switching costs exist and a monotone optimal control policy without switching costs (Lu and Serfozo 1984). They hysteresis control form, where the switching costs exist, arises from the needed to minimize the number of switching and hence make it economical once an increase in service rate is achieved. Gebhard (1967) and Li (1988) have studied the hysteresis service rate control, where the switching occurs instantaneously. Harita and Leslie (1989) have studied the case where the switching takes a time that exponentially distributed. All of these studies consider only a bi-level change in the service capacity.

4.5.3 Queuing Systems with Fluctuating Parameters

Many real-life queuing systems, such as communication channels, exhibit random fluctuating in their message arrival rate. The statistical variations of the offered traffic load to such systems introduce system design requirements that cannot be accounted through the use of a conventional time-homogeneous traffic model (Zukerman and Rubin 1986). The ISDN relay described in this is no exception.

Non-homogeneous queuing systems so far analyzed in the literature can be divided into two main categories. In the first, the arrival process is assumed to be non-homogeneous Poisson with rate function $\lambda(t)$, where $\lambda(t)$ is a deterministic function of time; and in the second, the arrival process is assumed to be a doubly stochastic Poisson process; i.e., the arrival rate function, $\lambda(t)$, is itself stochastic process (Zukerman and Rubin 1986). The latter approach is relatively new, and Yechiali and Naor (1971) looking published the first work at a special case where the arrival and service rates to a single M/M/1 queue are step function stochastic processes. Under this assumption, the queuing system can be in one or two modes, 0 or 1. When the system is in mode i, the inter arrival and service times are exponentially distributed with parameters λi and μI, (i = 0,1), respectively. The time interval during which the system functions at level i (i = 0,1) is exponentially distributed with parameter γi. Further extensions to the theory have been achieved by Neuts (1971) by assuming a general number of levels and general service time. Zukerman and Rubin (1986) develop generalization of the above case more than one server.

In the case where the arrival process of a queuing system can be one of two modes these modes could represent light and heavy loads. This leads to the definition of two further study models: model I, where a single-mode arrival process is assumed, and model II, where a two mode arrival process assumed. Model I is dealt with in chapters 7 through 10 in detail, while Model II is considered beyond the scope of this book.

4.5.4 Queuing Strategies

The factors affecting the queuing strategies in a relay within a heterogeneous environment are somewhat different from those of data-only environment that we are considering. In the heterogeneous environment the relative mixes of different traffic types as well as their service requirements in terms of switching are more important. The delay on synchronous services like packet voice and video must be bounded. Error rates become more important if some form of compression has been used. In the data-only case, delay is tolerated to a certain extent but reliable delivery is necessary. The factors described below are considered to be most relevant this book.

Application types: Here the difference is between the interactive versus non-interactive applications. The problem is that the network layer has no knowledge regarding the type of higher-layer protocol data units carried within a packet. It is up to transport service to determine the QoS parameters that will be included in the network service data units. Also, different priorities may be assigned to different application types.

Quality of service The determination of the quality service for CO and CL modes of transmission is different from one another. Here we consider only the CL case. The OSI connectionless network service (CNLS) definition describes the associated QoS parameters (ISO 1987b). The DARPA IP types of service parameters (Postel 1981a) are similar. The transit delay and priority QoS parameters are most important in the context of channel allocation strategies.

The transit delay QoS parameter will indicate how that packet should be treated in terms of queuing for output within the CL network relay. For low transit delay requirements, more channels may need to be opened in order to satisfied the network service user, although the CL mode of operation only endeavors best efforts in delivering a particular packet to its destination. Note that the OSI CLNS does not specify a QoS parameter for throughput.

Prue and Postel (1988) develop a queuing algorithm for type of service TCP/IP networks, where the multiple channel queues are assigned to a server. The server spends a proportion of time serving each queue. The time spent on each queue is linked to the number of queues, the service rate and the QoS requirements. However, for simplicity we do not implement the full QoS queuing in our channel management plan. The only QoS parameter considered is the 'delay' parameter. Hence, we assume that all the packets to the same destination are multiplexed into a single queue, irrespective of their QoS requirements.

Priorities and pre-emption The use of priorities for different classes of applications has not been widely implemented in CL networks although some work has begun. However, the QoS parameter field is available for this purpose. It could possibly be used as basis for giving higher properties to packets from certain types of application. For example, an interactive application such as a file access can be given higher priority than that for an electronic mail application. Pre-emption suggests that a higher-priority packet could pre-empt a lower-priority packet in order to get speedier service.

The issues of properties and pre-emption are beyond the scope of this book and hence will not be considered any further.

Costs: opening a new circuit for a burst arrival of packets may be expensive if that channel is not going to be utilized after the burst has been cleared through the system. Therefore, ways of determining the best threshold for new circuit set-up on possibly multiple queues (e.g. model 1) to same destination need to be found.

4.6 SCHEDULING POLICIES

By 'scheduling policies' we mean the way in which the individual packets are 'assigned' to different physical channel queues of a logical channel, and such this applies to model 1. In queuing theory terminology, this entails the scheduling of a new job to a server. Some well-known scheduling policies are 'round robin', random 'and' shortest queue policies. In real data communication interfaces the size of the channel queues are bounded; therefore, scheduling policies like 'sequential' and 'alternate sequential' policies can also be used.

Under the round robin policy, each queue receiver a packet in turn within a strict order. Random policy queues packets in a random fashion to the channels. Under the shortest queue policy, an arriving packet is placed in the queue whit the shortest length. A variant of this policy, the 'shortest remaining work policy', places the arriving packet in the queue whit the least unfinished work (i.e. fewest total number of bits waiting to the transmitted, rather than smallest number of packets in the queue).

The sequential policy applied to a two-channel system would first start with channel 1 and schedule all incoming packets to its queue until it reaches the maximum queue size or a set threshold value. The second channel would be used as an overflow channel only, so that, whenever the channel 1 queue size fell below the set value, it would get the following packets as described above. The advantage of this method is that it will use the overflow channels only when it can handle no more packets. An idle-timeout mechanism can be used to remove the unused overflow channels. The alternate sequential policy behaves like the sequential policy except that it alternately fills the channel queues to their limit.

4.7 PROTOCOLS OVER THE ISDN B-CHANNEL

As was mentioned the two modes of interworking at the network layer –CO and CL – lend themselves to different protocol stacks over the ISDN B-channels. Here, we investigate the effects of the mode of operation. The choice of protocol stacks to be used over the ISDN B-channels and their behaviour are important in the type of channel management schemes to be adopted. This is discussed below with reference to CO and CL network services. It is assumed for simplicity that a CO transport service is used both cases.

The channel and bandwidth management aspects mentioned at the beginning of this chapter is relevant here. This section analyses only the protocol interaction with the connection set-up and removal operations. The question of bandwidth variation using threshold control applies to the CL relay more readily than to a CO relay, since in the former multiplexing of packets from several TP entities into one channel queue is assumed. This also relates to window-based flow control issues in transport protocols , since the bandwidth used by a TP would be controlled by this mechanism when operating over a connectionless network service. In the CO relay case , the various virtual circuits multiplexed over a single physical circuit produce a queuing model where a server is serving multiple packets queues, one Per VC.A detailed study of that model is beyond the scope of this book.

4.7.1 CO relay

A CO services data for call set-up addressed to a remote host will be recognized and the next hop along the route will be determined. Assuming that there exist no circuits to the destination, the relay will set-up in a circuit-switched path to the next hop, using its ISDN layer 3 call set-up procedures on the D-channel. This will be then be followed by the normal X.25 call set-up procedures on the B-channel. Following the set-up of a virtual circuit, the data transfer will take place. In this case X.25 calls will be served to completion, i.e. until one of the parties terminates the connection.

If other users wish to communicate to the same destination relay while this circuit is set up, the same physical circuit could be used for the new transfer by setting a new VC which will be multiplexed over the existing physical circuit. However, if the QoS parameters exchanged at the beginning of the session demand a full 64 kpbs service, then a second B-channel will be used for a new call set-up; otherwise, a disconnect packet will be returned.

An alternative mode of working is to use multiple physical ISDN circuits provide a larger bandwidth for applications requiring it. In this case, the X.25 multi-link procedure (MLP) could be used over single-link procedures (SLPs) running on each physical circuit.

The above method of usage means that the slow terminal or file transfer traffic could be multiplexed over one ISDN circuit-switched connection, whereas the rapid response requirements of a file access would be met fully by avoiding further call multiplexing on that physical circuit. In the case of fast file transfers, MLP could be used over multiple circuit-switched connections to provide even larger bandwidths.

In order the achieve an effective channel management, therefore, the relay needs to know about the parameters of each connection as well as other state-depend information-for example the delay and throughput requirements of each VC connection, the address of the destination relay and the B-channel to which it is connected. Also, a binding of VCs to 'physical' ISDN circuits is necessary.

4.7.2 CL relay

In the case of connectionless relay, the problems with the circuit set-up and removal include the following:

1. Applications running on LAN hosts may need to establish an association across the ISDN. Ideally, an ISDN circuit should be established as soon as there are no associations extant. If the relay is a network layer relay, it has no idea of the progress of the association, its duration, or the number of packets to be transferred; it needs to have a time-out mechanism which will close the connection down if no more packets arrive during that duration. There is obviously a danger of closing the circuit while the association is merely quiescent.

2. As the end-to-end connection are initiated and maintained by the transport protocol, its features become important in the above case. For the practical case of ARPANET transmission control protocol (TCP) over internet protocol (IP), it is necessary to be aware of the time-out values of the implementations. The main time-outs in TCP are the user time-out (UTO), retransmission time-out (RTO), time-wait time-out (TWTO) and the maximum segment lifetime (MSL) which indicates the lifetime any segments remaining in the network. These are discussed below.

In most implementations the hosts determine the transmission time-out dynamically. This means that the relay has no idea of its value for that connection. Furthermore, a transport entity will usually continue to retransmit a transport protocol data unit (TPDU) that requires an acknowledgement for a number of times N; hence (N*RTO) is the persistence value. A transport protocol (TP) connection is closed if the inactivity time-out value is exceeded. It usually indicates failure of the supporting network connection or the remote transport entity. In order to prevent expiration of the remote transport entity's inactivity timer when no data is being sent, the local transport entity sends acknowledgement TPDUs (ACKs) at suitable intervals. This inactivity time-out value may be used calculating an ISDN channel's no activity time-out (NTO) within the relay. Hence the circuit-switched channels could be closed soon after such an event in the absence of traffic on other TP connections. It is also, plausible that, in the absence of data TPDUs, if the 'inactivity' ACKs sent by each TE are generated at reasonably long intervals, one may close the channel while there is no real data traffic, and occasionally open the channel to transmit these ACKs and maintain the two end TP connections. A similar argument can be proposed for the sufficiently long RTU values.

To summarize, it is necessary the close unused connections, while care must be taken not to be too quick for fear of closing down and ongoing transmission. Selection of the channel NTO value is of great importance in this respect. It can be seen, therefore, that there is a need studying the effects of the channel management schemes on the higher-layer protocol operation; herein lie the trade-offs and the route to an efficient relay implementation.

4.8 CHANNEL MANAGEMENT INFORMATION

This is an important issue, as the detail of state information available to the channel management function will determine the efficiency and performance of any channel management policy. Another important issue is the timing of the availability of such information. If there is a delay in state information updating, the actions taken can be detrimental to performance.

From the foregoing discussions, it can be concluded that the following information is necessary for channel management; status of each physical channel available at the interface and the destination address connected; bandwidth associated with each channel in use; and transmission queue sizes. Additionally, if the traffic arrival rate is to be used at the state variable for control, a rate measure per logical channel needs to kept and updated. It is also necessary to keep a timer variable Per logical channel in order the decide the removal of an unused channel. Furthermore, QoS parameters associated with each logical channel need to be kept if certain performance values are to be provided. The QoS parameters indicated would update this field by the network protocol packets.

4.9 COST FUNCTIONS

The form of cost functions to be adopted in evaluating a queuing system is important in determining the relative importance of cost factors as well as performance measures. Cost factors can be separated into the telecommunications costs incurred in accessing the bandwidth, and the user-perceived response in terms of delay and throughput. In the latter case, by penalizing the increase in delay or reduction in throughput arising from the implementation of a certain management policy, a basis for comparison between different policies can be achieved. However, combining the two cost factors is not straightforward.

4.9.1 ISDN Tariffs

The tariffs in circuit-switched ISDN are discussed in section 2.10 above, which points to a tariff structure based on the same principles as for telephone calls; the cost incurred is calculated by the duration and charging index. It is assumed further that a stepwise cost structure exists, i.e. that the charging is made at the beginning of each charge period. This mean that, once a call is established, it makes more sense to keep a channel connected until the end of the current charge-period, unless other messages are waiting for the channel to be released for access. Therefore, tariff cost minimization calls for cutting a circuit connections short as possible in order the minimize tariff costs.

4.9.2 Delay Cost Factor

A cost factor depending on the waiting time in a transmission queue or the sojourn time of packets can be used the add the delay penalty into the cost function. This cost factor can be given a linear or non-linear weighting in order to see the effect on the policy performances.

4.10 SUMMARY

This chapter has presented an examination of the channel management problem. Associated problems of bandwidth allocation, bandwidth management, control of queuing systems, factors affecting queuing strategies and scheduling have also been discussed. It is shown that the multiple channels at the ISDN user-network interface can be used in a variety of ways which could be modelled by alternative queuing systems. On the protocol side, the CO and CL mode services over the ISDN B-channel were considered with particular reference to the channel management problem. The chapter ended with a description of the types of information needed for effective channel management and the cost factors affecting the performance of management policies.

CHAPTER FIVE

CHANNEL MANAGEMENT ARCHITECTURES

Dynamic channel management (DCM) is needed in order to manage dynamically the multiple channels that are available at the user-ISDN interface. An increase or decrease in the bandwidth of a given logical channel formed by the aggregation of several physical channels can be achieved according to the requested quality of service e.g. delay, and traffic patterns for providing the QoS and minimizing operational costs.

This chapter describes a dynamic channel management architecture (DCMA) based on the status table approach, whereby the information relating to the individual call, channel statutes and the ISDN interface is kept and updated within the interface by a channel management function. The model describes applies to connectionless network layer (CL-NL) LAN-ISDN relays in general and to the Ethernet-PRISDN CL-NL relay in particular. By applying suitable modifications, the architecture can be ported to connection oriented (CO) type relays as well as the user equipment operating in either the CL or CO modes and having direct access to a multi-channel user-ISDN interface. Furthermore, the architecture described refers to the packet mode usage of the circuit-switched ISDN. The interface referred to is the primary-rate ISDN (PRISDN) with 30 B-channels, although other multi-channel interfaces are also relevant.

The DCMA developed in this chapter will be shown to serve two purposes: first, as a general model for implementation within a 'real' LAN-ISDN CL-NL relay to provide dynamic channel management; second, as the basis for a simulation model for measuring the performance of different of different bandwidth management strategies. The implementation of the DCMA within a simulation model has the advantage of providing the viability of the architecture. The level of detail provided in the DCMA model presented here is necessary for the simulation model since the interaction between different decision processes as well as packet scheduling to the channel queues needs to be investigated.

5.1 GENERAL ARCHITECTURE MODELS

The main assumptions regarding the architecture design are the existence of:

- Common channel signalling (CCS) facility
- Multiple channels at the user-network interface

In chapter 2, the notion of channels formed by the individual or aggregated usage of time slots (TSs) in a TDM frame was introduced. In Chapter 4, channel management was defined and the pertinent models for different types of usage of these channels were described as queuing systems for channel management. Here, the architecture models for two types of queuing systems will be considered.

- The n x A/B/1 queuing systems (model 1)
- The A/B/1 queuing system with capacity n x C (model 3)

Model 1 represents the case where individual B-channels (time slots) are used in parallel, each with its own queue, and to which packets are 'routed' according to the scheduling discipline used. Each B-channels transmits a complete packet. Packet re-sequencing is left to the transport layer protocol operating end-to-end.

In model 3, the additional B-channels that are added to the first one share the load in such a way that each TS transmits a byte of a given packet in turn. This mode of usage necessitates the provision of time slot sequence integrity (TSSI) between the individual TSs on an end-to-end basis (at the physical layer) across the ISDN. This is assumed to exist in the rest of this chapter. The issue of TSSI provision is discussed further in Chapter 6. In this model, the underlying physical channels are transparent to the packets since the aggregated channels from a logical channel. Table 5.1 shows a comparison of the two basic models.

Model 2 results from the A/B/n type of queuing system .In this model, one transmission queue Per super channel is needed, and each server (time slot) must have a buffer space for one NL packet while transmitting it. Each server interrupts or raises a flag at the end of its busy period to indicate its idle state so that the next packet can be assigned to it for transmission. Also, the TSSI is not necessary if the transport protocol is used for re-sequencing of network layer packets.

The architectural requirements of model 2 are very similar to those of models 1 and 3. The similarity whit model 1 arises since each TS is also a physical channel, and a super channel is formed by the association of different B-channels into a larger transmission group. In contrast, model 3 may or may not be composed of individual physical channels: in the aggregation method of super channel formation each TS will correspond to a B-channel, while if the n x 64 kbps method is used several TSs will make up a super channel. However, model 2 is similar to model 3 in its use of a single super channel queue, hence resulting in a similar type of threshold control structure. This model will not be studied further for reasons of similarity with models 1 and 3.

Table 5.1 Comparison of x n A/B/1 & A/B/1 – n x C queuing models

Features	n ×A/B/1	A/B/1 – n × Ca
Scheduling	Packets are scheduled to the individual server queues	Not needed since only one queue per 'super channel' exists
TSSI	Not necessary since transport protocol is used for packed re-sequencing when CL-NP is used	Needed since byte reordering may result. Packet re-sequencing is not needed if TSSI is provided.
Response time	Inferior	Superior
BW management	Complicated	Simpler
End-to-end superchannel identification	'A priori' or as needed	'A priori' or as needed

C: capacities of one time slot (i.e. one basic bandwidth unit—BBU).
CL-NP: connectionless network protocol.
Superior response time in the A/B/1—$n \times C$ over the $n \times$ A/B/1 queuing model is achieved as a result of the *scaling effect* shown by Kleinrock (1974,1976).

B-channel, while if the $n \times 64$ kbps method is used several TSs will make up a superchannel. However, model 2 is similar to model 3 in its use of a single superchannel queue, hence resulting in a similar type of threshold control structure. This model will not be studied further for reasons of similarity with models 1 and 3.

5.2 THE CHANNEL MANGEMENT ARCHITECTURE

The main requirements of the management architecture are the provision of efficient management functions and practicability. The minimal set of channel management functions can be identified as:

- Ability to monitor and update the status of the interface under control
- Ability set up and remove channels
- Ability to vary the bandwidth of logical channels

In all the architecture models described in the previous section, a further issue needs to be considered: that of whether a priori or negotiated super channels are to be used. In the priori method, all physical channels established between two communicating ends are assumed to be part of the same super channel. This prevents the operation of multiple parallel super channels across the same user-network interface and is less flexible. The negotiated super channels method supports multiple super channels between two communicating ends. In model 1, the priori method leads to a simpler interface status table structure and necessitates a different status block for reporting current interface status. Both of these methods are dealt with in the following sections. The examples given will be mainly for model 1—*a priori*—method, and differences indicated where necessary.

The main functional units necessary for channel management and the design of the messaging structures at the management–ISDN interfaces are described in the following sections.

5.2.1 The Channel Management Functional Units

The above requirements can simply be satisfied by the provision of the following channel management functional units (CMFUs):

1. *Channel status table (CST)* Keeps the current state of the ISDN interface.
2. *Channel management routine (CMR)* Interrogates the CST at specified epochs and makes a decision to set up or remove channels or vary their bandwidth according to some performance and/or cost criteria.

The channel status table is updated every time a packet is routed to the ISDN and whenever a message is received from the ISDN interface module reporting the completion a packet transmission and the UNI status. In its simpler form, the CST can be used to monitor only the outbound traffic. The two important metrics to be updated are the instantaneous transmit queue size and the traffic arrival rate. Complexity of the CST and its operation can be increased if incoming traffic is also monitored to record the incoming traffic arrival rate and the receive queue(s).

In the design of the architecture, placement of the main functional units can be implementation-specific. However, a convenient separation of the management operations at the logical level from those operations performed at the physical interface level, including any signalling protocol implementations at layers 1-3, cases the design effort and provides a much needed level of abstraction. Furthermore, there exist no standard software interface between the user application and the ISDN D-channel protocols at layer 3. Indeed, such standardization may never be adopted. Therefore, such a software interface needs to be designed.

Two levels of decision processes are identified within the channel management function (routine): high-level and low-level.

High-level decision process (HLDP): At this level, a decision needs to be made to open/close a new/old logical channel to a new/existing destination. Once a logical channel is established, a decision is made to send or drop packet, depending on the node/channel queue congestion level or other decision criteria. Furthermore, for dynamic control of channel bandwidth, a decision needs to be according to some prescribed criteria, and using a decision metric, whether to increase or decrease the channel bandwidth and by how much. For this, a high-level primitive, CHange BandWidth (CHBW), could be used. Hence the following high-level primitives can be defined for use between the HLDP and lower-level modules:

1. *OPEN (isdn_address, bandwidth)* Low-level decision process (LLDP) module returns the assigned logical channel reference number (LCRN).
2. *CLOSE (LCRN)* Indicates the complete removal of all physical channels assigned to a logical channel (LC) with a given LCRN.
3. *SEND (LCRN, data_buffer)* LCRN value used the queue the packet the correct logical channel queue.
4. *CHBW (LCRN, bandwidth, direction)* Indicates request for change in the bandwidth of a LC by the amount 'bandwidth'. Increase/decrease is indicated by the 'direction' value.

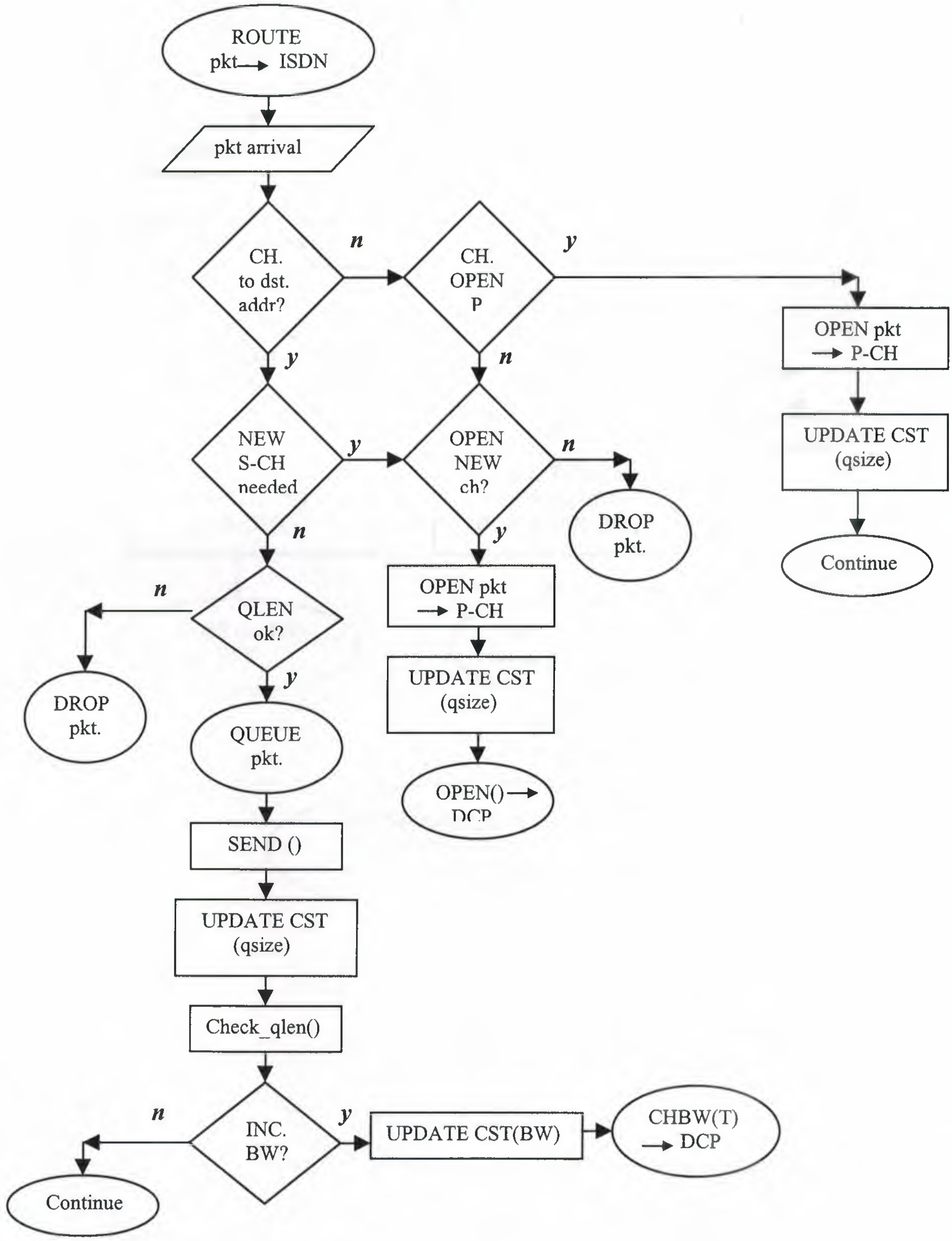
Of course, primitives are also needed to indicate the successful or non-successful completion of each command sent from the HLDP to the LLDP. These are OPENED, CLOSED, PKT_SENT and BW_CHANGED. The asynchronous type of interface suggested is necessary since the HLDP needs to be free to deal with other requests while it is waiting for, say, OPENED for a previous open command. This mode of operation is deemed to be better than, say, suspending all other operations while waiting for responses from the LLDP.

Low-level decision process (LLDP): The main duty of this decision process is to act upon the high-level primitives and interface with the B and D-channel processes. Furthermore, at this level a decision is made as to which types of physical channel (e.g. B or H-channel type) to use in order to aggregate the bandwidth to the requested logical channel bandwidth. The assignment of LCRNs and call references is also undertaken. A logical channel-to-physical

channel mapping is necessary at this point. A possible time slot mapping (TSMAP) table is shown Fig. 5.3. A decision to close an existing connection and re-established the connection using different types of physical channel structures will also need to be placed here if such functionality is supported. In the case where H-channels are used, another decision is needed to determine the TS to be assigned to each physical channel. This may be determined by the TS hunting strategies adopted. However, the dynamic reassignment of TSs to physical channels for an existing call is not supported in the current PRISDN UNI. In the case where only B-channels exist, CCITT Rec. I.450/1 specifies that each TS number also corresponds to the channel number.

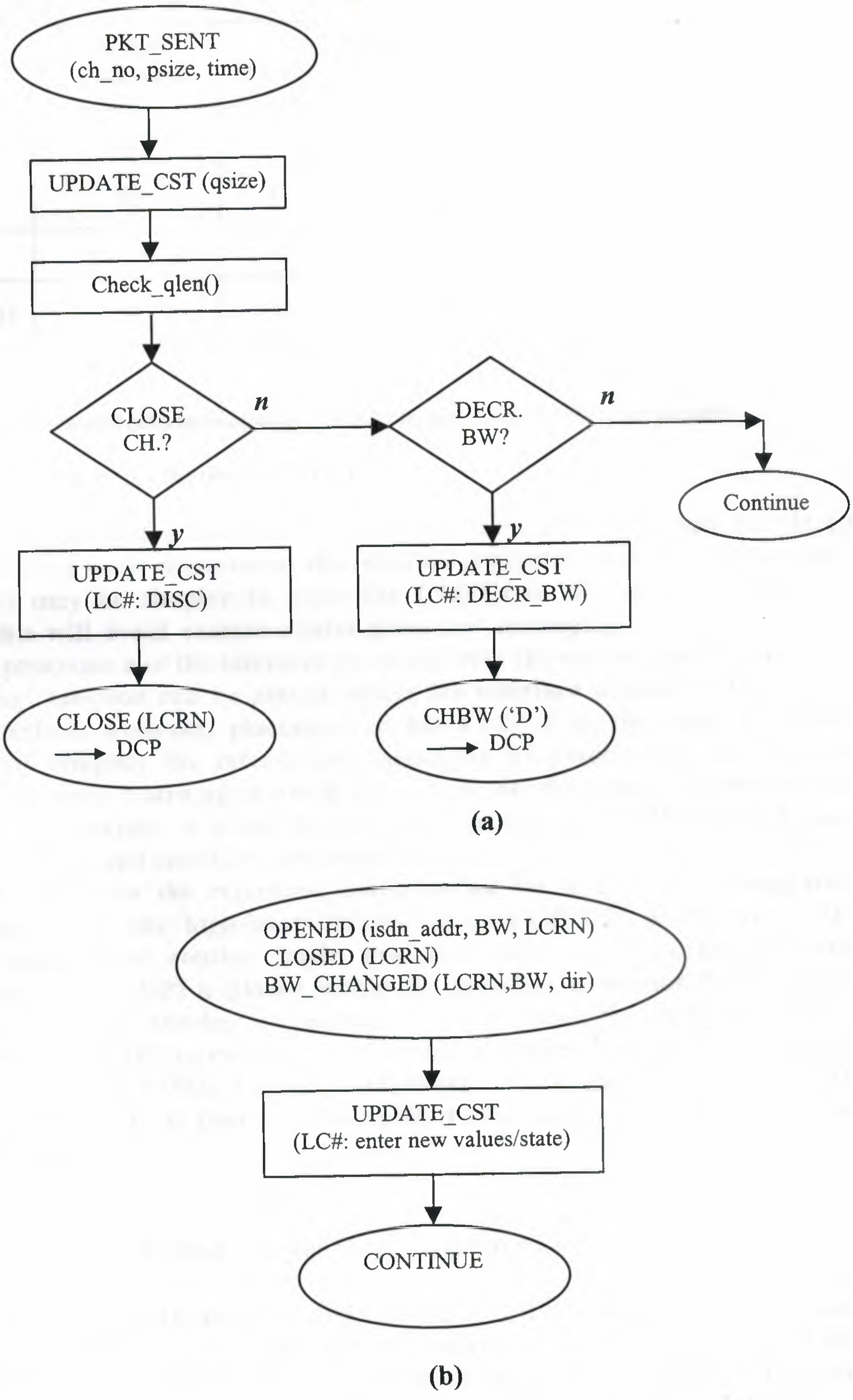
5.2.2 Placing the CMFUs

As mentioned in the previous section, this is very much dependent upon the hardware and software configurations as well as on performance issues. A broad assumption that the network layer relay is placed on a hardware configuration where a main processor board controls the access to each network type, each of which is interfaced by an interface board with on-board interface software for low-level protocol/hardware interfacing, is not unrealistic, as there are many examples of this in the existing networks. According to the above assumptions, two broad classifications can be made with regard to the ISDN interface. In the first the ISDN hardware interface could be composed of a single board (type 1) or multiple boards (type 2). The second class of interface may be necessitated for speed reasons where each interface board is given the responsibility for a limited number of B-channels as well as separating the B and D-channel boards. An example of this the PROOF LAN – ISDN relay (Knight 1989), where handling of X.25 protocols over the B-channels is also required from the interface boards.



**pkt: packet; CH (ch): channel; dst: destination; addr: address; P: pending;
P-CH: pending channel; qsize: queue size; BW: ISDN bandwidth; CST: channel status table;
DCP: D-channel process; CHBW: change bandwith function**

Figure 5.1 High-level decision process: action routine



LC: logical channel; LCRN: logical channel number; DISC: disconnect
DECR_BW: decrement bandwidth; psize: packet size; dir: direction

Figure 5.2 High-level decision process: reply routines

TS	PCRN (Call_Ref)	LCRN
0	•	•
1	25	121
•	•	•
31	57	121

PCRN: physical channel reference numder; LCRN: logical channel reference number

Figure 5.3 Details of a TSMAP table.

Thus, the channel management functional units (CMFUs) can be placed either within the *main processor* or the *interface processor* domains. In the type 1 interface it may be simpler to place the CMFUs in the interface processor domain; this will avoid excessive inter-processor messaging between the main processor processes and the interface processes over the system bus. In fact, even the 'routing' function can be placed within the interface processor domain. In type 2 interface, however, placement of the CMFUs at the main processor domain may simplify the inter-board messaging by combaining the decision process at the main board upon which the correct interface board is activated. In the rest of this chapter, a model is described whereby the CMFUs are divided between the main and interface processor domains.

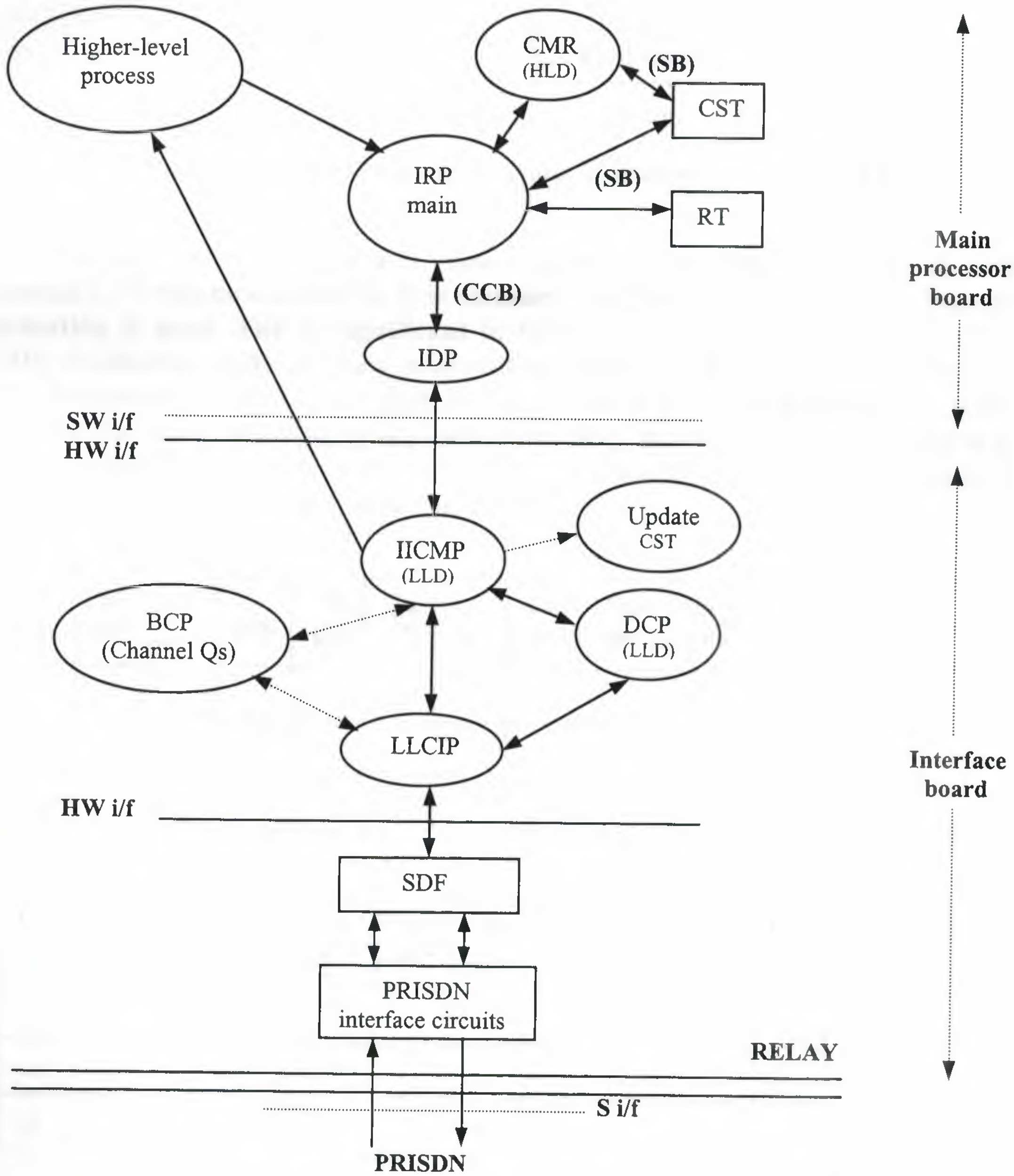
Figure 5.4 shows the reference configuration for a channel management architecture. Here, the high-level decision process (HLDP) is left within the channel management routine (main processor domain) while the low-level decision process (LLDP) is placed within the interface processor domain. This would minimize the number of message exchanges and the speed of reponse since, once the LLDP is called, it will result in either a D-channel protocol activation (OPEN, CLOSE, CHBW primitives) or a B-channel activity (SEND primitive). The LLDP is shared between the B and D-channel processes of the interface processor.

5.2.3 Designing the Channel Control Messaging Interface

From the reference configuration shown in Fig. 5.4, it can be seen that the B and D-channel interfaces can be addressed in a unified manner. This approach has been adopted in the design of the channel cotrol messaging structure. A Channel Control message Block (CCB) is design to achieve the exchange of information between the high and low-level decision processes. The structure of these message blocks is similar to the Q.931 message structure (CCITT 1988a) and is shown in Fig. 5.5.

5.2.4 The Channel Status Table (CST)

The channel status table contains all the information about the ISDN interface necessary for decision-making by the CMR. The form and structure of this table could be implementation specific. Also, depending on whether the model 1 or model 3 architecture type is used, its information content may be different. This will also affect the form of the status block used to interrogate the CST.



CMR: channel management routine (within IRP); HLD: high-level decision; SB: status block; **CST: channel status table; RT: routing table; CCB: channel control block; BCP: B-channel process; IDP: ISDN port driver process; IICMP: ISDN interface control main process; DCP: D-channel process; LLCIP: low-level control interface process; LLD: low-level decision; SDF: synchronous data formatter (HDLC).**

Figure 5.4 Reference configuration for CMFU placement.

Protocol discriminator
Call reference
Message type
Information element (Mandatory/optional)

(a)

Call_Ref
Msg_Type
Ch_Type
Ch_No/LCRN
Issrc_Add
Cause
Bandwidth

(b)

Figure 5.5 Details of (a) Q.931 message structure; (b) channel control block (CCB)

A sample CST for model 1 is shown in Fig. 5.6(a). Here, the state of each channel (TS) can be monitored. It is assumed that the a *priori* type superchannel formation is used. This is significant in that a search of the table for a given ISDN destination address (*isdst_add*) will reveal all the B-channels connected to that destination. A packet can then be scheduled to the most appropriate queue according to the scheduling policy being used (e.g. shortest queue). The status of those channels connected to the destination is returned by the status block. A sample CST for model 3 is shown Fig. 5.6(b).

Ch. no.	Call ref.	Ch. state	BW	Tx pkts	Tx Ql	Act	Call src	Time out	ipsrc add	ipdst add	isdst add
1											
.											
30											

(a)

SC no.	LCRN	SC state	BW	Tx pkts	Tx Q1	Dst LCRN	Act	Time out	ipsrc add	ipdst add	isdst add
1											
.											
30											

(b)

Ch: channel; Ref: reference; BW: bandwidth; Tx: transmit; Pkts: packets; dst: destination; Act: activity (tarffic rate); Ql: queue length; ip: Internet protocol; isdst: ISDN destination; Add: address; SC: superchannel; LCRN: logical channel reference number

Figure 5.6 Details of the channel status tables: (a) model 1 (b) model 3

5.2.5 The Channel Management Routine (CMR)

This routine can interrogate the CST by a messaging block called the *status message block* (*SB*) and a function called, *check_con* (*isdn_address, sb_p*), where 'check_con' is short for *check_connection* and *sb_p* is a pointer to the status message block.

After an interrogation, the SB contains the necessary information for the CMR to make a decision described in the HLDP, and using one of the described primitives, it activates the LLDP for the correct action across the main processor port interface. Figure 5.7 shows the SB structure and fields applicable to model 1 with a *priori* superchannel formation.

	Ch. no.	Call ref.	Ch. sta.	Qued pkts	Qlen
1					
.					
m					

	Full	No. of cons.	No. of free chs.	Free ch. no.	No.of resv. chs.	Resv. ch. no.	ISDN dst. add	Send ch. no.	C1	C2
1										

Sta: state; Cons: connections; Full: channels (not) busy; Resv: reserved; C1, C2: code fields

Figure 5.7 Details of the status message block (SB)

5.2.6 Status Updating

As a result of each instance of interface and channel activity, the CST is updated by a procedure called *update_cst* (*io_blk_p, msg*), where *io_blk_p* is the pointer to an input-output message block containing the data packet, *sb_p* is the pointer to the status block (SB) and *msg* is the message indicating the type of operation causing the update. The message types are broadly classified as those pertaining to the D-channel activity and to the B-channel activity. The first type affects mainly the channel state, while the second type affects mainly the updating of B-channel queues and activity rate fields. The messages used for the D-channel activity are: SETUP, RESV, CONN, CONN_ACK, DISC, REL, FREE and UPDATE_BW. These messages correspond to the D-channel protocol messages-states at layer 3, i.e. Setup, Reserve, Connect, Connect_Acknowledge, Disconnect Release and Free. The message UPDATE_BW results from a successful CHBW() functional call to the interface. A further message for Release Complete can also be implemented. The messages used for the B-channel activityare UPDATE_QLEN, UPDATE_ACT and UPDATE_TO. Every time a packet is

sent to the interface for channel queues and when packets are tranmitted from the interface, the UPDATE_QLEN message is used to update the queue size and number of packets in the system. The activity message UPDATE_ACT is used to measure the packet arrival rate (over a period) or the last time a channel was used. The UPDATE_TO message can be used to record/modify the time-out values set for each channel.

5.3 A DCMA FOR ETHERNET-ISDN RELAY

The main principles of the dynamic channel management architecture discussed in the previous section can be applied to the design of a DCMA for the Ethernet-ISDN relay. Assuming a rather simple-minded connectionless network layer relay, all of the DCMA functionalities described can be implemented.

In the design of the DCMA, several assumptions have been made regarding the capabilities of the ISDN interface hardware. These assumptions are based mainly on the VLSI chips available today or soon to be available. It is assumed that the ISDN interface board has enough processing power derived from a single or multiple processors to manage a PRISDN interface, including all the B-channels as well as the signalling D-channel. The existence of enough buffer space for the transmit and receive buffers for each channel is also assumed. It is further assumed that a form of framing will be present or established to implement the data link functions on the B-channels. The main processor can communicate with the interface processor(s) using global memory and passing pointers to data/message structures. A common interface for signalling and data paths between the ISDN router process (IRP) and the ISDN interface modules is designed in line with Figures 5.4 and 5.9. this interface is assumed to be capable of specifying connection set-up or removal on multiple B-channels at the same time.

5.3.1 Protocols

The relay reference protocol structure is shown in Fig. 5.8. This protocol structure has previously been suggested in Deniz and Knight (1989b). The CL network protocol (CLNP) could be replaced by the Internet protocol (IP) in networking environments using the TCP/IP suite (DARPA transport control protocol/internet protocol). The data link layer protocol on the Ethernet could be null in some implementations, in which case only the MAC frames would be used. As the same network layer protocols are used over both the Ethernet and ISDN, the CLNP relaying is rather simple; it involves the extraction of the CLNP packet from one link layer, framing and placing it in the other network's framing. Of course, issues such as address resolution, fragmentation and re-assembly need to be considered. In the case of the circuit-switched ISDN, the existence of a channel to a given destination must also be verified. A channel management routine within the relaying function acts like a decision process, determining whether a new circuit should be opened, an existing one close, the bandwidth of an existing channel changed or a network layer (NL) packet simply forwarded over an existing channel. The NL packets are then relayed over the ISDN B-channel(s) within core HDLC frames assuming that a channel to the given destination exists. For channel management simulations, only the basic D-channel signalling functions of I.451 (layer 3) and LAPD (layer 2) need to be implemented in a much simplified manner.

	RELAYING FUNCTION		
Network layer	CLNP	CLNP	I.451
Link layer	LLCI	HDLC	I.APD
Physical layer	CSMA/CD	I.431	I.431
		B	D
	Ethernet	ISDN	

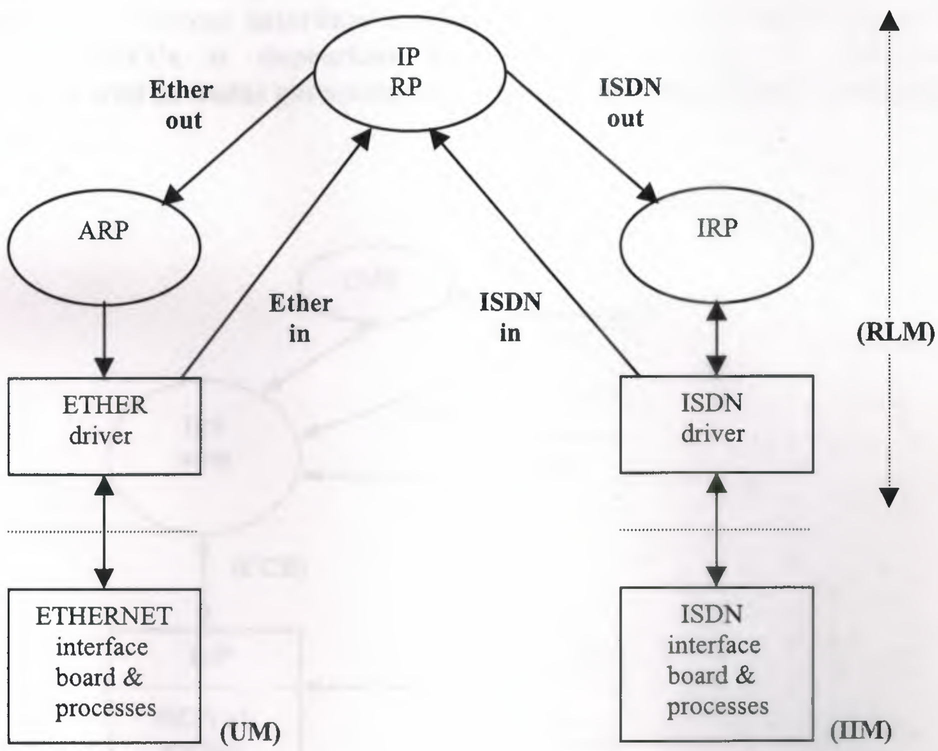
CLNP: connectionless network protocol; LLCI: connectionless logical link protocol
LAPD: link access protocol for the D-channel; HDLC: HDLC framing
B, D: ISDN user and signalling channels, respectively.

Figure 5.8 CL Ethernet/ISDN relay protocols.

5.3.2 Main Processes

The main tasks within the Ethernet/ISDN relay (EIR) are shown in Fig.5.9. The IP relaying process (IPRP) performs all the general relaying functions including buffer management, initial address resolution, fragmentation and reassembly. It calls the address resolution protocol (ARP) or the ISDN router process (IRP) for address resolution functions to Ethernet and ISDN respectively. The ISDN router process achieves the IP-ISDN address mapping via a fixed routing table and monitors the status of ISDN interface via a channel status table (CST) shown in Fig. 5.6. The IRP may not include a full routing function if only point-to-point communication is used or if only one interface type exist between the Ethernet and ISDN. The call control and message forwarding for each channel buffer is accomplished by the use of a channel control message block (CCB) used between the ISDN router (IRP) and driver (IDP) modules. The status information regarding the PRISDN interface board(s) in returned using the same CCB message structure and is used for CST updating. The CST status can be called up at any time by the IRP or the channel management routine, which itself is activated by the IRP every time a datagram arrives at the relay and is to be routed to the ISDN output interface. CMR is also activated when a packet transmission occurs and its completion is reported back by the ISDN interface. CST status is reported back using a status message block (SB). Figure 5.7 shows the SB, while Fig. 5.10 shows a detailed diagram of the IRP and IDP process of the EIR.

In Fig. 5.9, all the processes running on the relay main processor are shown within the relay module (RLM). Two further modules, the user module (UM) and the ISDN interface module (IIM). Note that these arbitrary module names are for simulation purposes only. The CCB message structure is used for communicating with both the B and D-channel processes in the IIM. This represents a unified and simpler approach to ISDN interface access. With this approach, the actual interface hardware implementation could be assumed to be on single or multiple boards using single or multiple processors. Within the ISDN interface module, a circuit control status table (CCST) can be used optionally to monitor and manage the status of individual ISDN channel, (similar



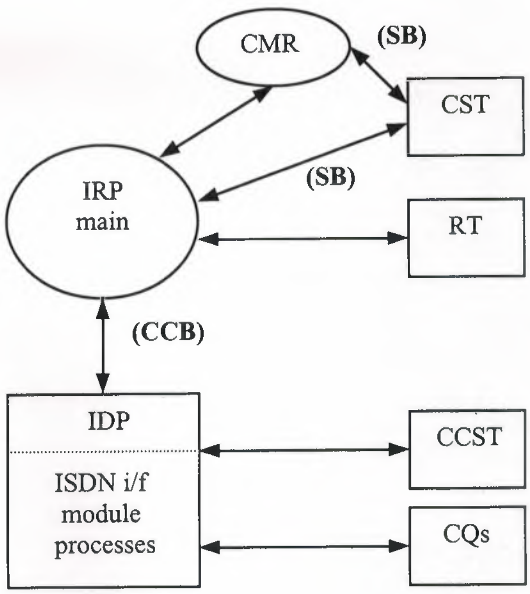
ARP: address resolution protocol; IPRP: internet protocol relaying process; UM: user module
IRP: ISDN router process; IIM: ISDN interface module; RLM: relay module

Figure 5.9 Ethernet/ISDN relay processes

to the CST) as well as to transmit and receive channel queues (CQs). This depends on whether or not be CST data structures are globally shared. If they are shared, as shown in Fig. 5.4, both the main and interface processors can use a common CST updating function. Within the EIR core functions the same message blocks are used, and their pointers are passed to processes along the way. This ensures that there is as little buffing copying as possible for operational speed. Communications between the core functions and the driver modules are established using global memory. The same is assumed between the driver module and the PRISDN interface board(s).

5.4 SUMMARY

This chapter has presented a dynamic channel management architecture design based on the status table approach. The decision process involved in the management function is classified into the high-level and low-level processes. It is shown that a high-level decision process using simple primitives can implement opening, closing and bandwidth variation of channels as well as communicate packets to the transmission interface. The main channel management functional units (CMFUs) are identified as the channel status table and the channel management routine. The importance of the placement of CMFUs is discussed with reference to different interface configurations. It is concluded that the placement of CMFUs is dependent upon the hardware and software configurations as well as nodal performance. A simple channel control messaging



CMR: channel management routine; CST: channel status table
RT: routing table; SB: status block; CCB: channel control block
CCST: circuit control status table; CQs: channel queues; IDP: ISDN (interface) driver process

Figure 5.10 Details of the ISDN relay process (IRP)

interface is designed and shown to be effective in combining D and B-channel messaging. Finally, a DCMA for Ethernet-ISDN relay is designed and its operation described.