VOICE TRANSPORT, CAPACITY AND SIGNALING IMPROVEMENTS FOR INTEGRATED WIRELESS PERSONAL COMMUNICATIONS OVER METROPOLITAN AREA NETWORKS

by

CHRISTOPHER M. COBBOLD

B.Sc. (Electrical Engineering), Queen's University, Canada, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF ELECTRICAL ENGINEERING

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 1995

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Electrical Engineering

The University of British Columbia
Vancouver, Canada

Date 9/27/95

DE-6 (2/88)

Abstract

Networks for Personal Communication Services (PCS) are currently being deployed in various parts of the world. A wide range of wireless voice and data service offerings is promised, and consumer participation is expected to be high. The IEEE 802.6 Metropolitan Area Network (MAN) has previously been proposed to provide for the distributed control of such networks.

This work continues investigations in call transport on the MAN using pre-arbitrated (PA) bandwidth, and for signaling using queue arbitrated (QA) bandwidth. A scheme for the transport of low bit-rate encoded voice, which is characteristic of mobile voice coders, is proposed that attempts to balance packetization delay with network complexity. Capacity gains are realized with PA access by locating a gateway, which provides interconnection to the Public Switched Telephone Network (PSTN), at the Head of Bus (HOB) in a closed-bus MAN. A two-dimensional Markov Chain and resulting blocking probabilities are derived which result in the determination of the Erlang and subscriber capacity of a Personal Communications Network (PCN) based on a single MAN. This provides the basis for extracting reductions in signaling delays on the MAN by applying the Preemptive Priority Mechanism (PPM), a recently proposed improvement to the MAN media access control protocol. In summary, improvements to the PA and QA transport mechanisms, as applied to PCNs, are developed and assessed.

Architectures of single and multiple MAN-based PCNs are presented in the context of the Intelligent Network (IN). The IEEE 802.6 MAN is viewed as being an evolutionary step towards the introduction of the Broadband Integrated Services Digital Network (B-ISDN). Issues arising from the migration of such architectures towards Asynchronous Transfer Mode (ATM) networks are considered in depth.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I wish to thank my research supervisor, Dr. Robert W. Donaldson, for his direction and encouragement throughout my stay here at U.B.C. This work was supported by the Canadian Institute of Telecommunications Research through a Research Assistantship provided by Dr. Donaldson. Oliver Yu deserves my gratitude for many engaging discussions on various topics in wireless networking. I offer my sincere appreciation to Dr. V. C. M. Leung for procuring the state-of-the-art in simulation software, Opnet, and to Mil 3, Inc. who provided easier access to it through their initiative with universities. Lastly, I would like to thank my family and close friends for their unwavering support for me to come to U.B.C. to pursue a goal.

Chapter 1. Introduction

When cellular communications services were offered initially to the public about 11 years ago, a forecast by AT&T predicted that there would be approximately 900,000 subscribers for such services in the United States by the year 2000. These estimates turned out to be grossly conservative. Statistics from 1994 indicate that 13 times this number were using cellular services, with another 20 million in other countries around the world [1].

Continued growth in wireless communications is anticipated as the industry looks forward to a new generation of service offerings known as Wireless Personal Communications [2] or Personal Communications Services (PCS). The scope of potential services is wide ranging. Personal Communications Networks (PCNs) will likely carry integrated wireless traffic for services such as voice, e-mail, fax, file transfers, and varying forms of messaging and telemetry, while providing for various degrees of mobility by using a combination of picocells, microcells, and macrocells. Recent Federal Communications Commission (FCC) auctions in the U.S. for slices of radio spectrum, earmarked for PCS in the 1.9 GHz range, yielded billions of dollars in government revenues and an array of corporate alliances involving firms wanting a piece of this potentially profitable sector of the communications market. Companies and shareholders are anticipating high penetration rates among the public to obtain a return on their investment in license rights to provide services.

Beyond PCS, network designers envision a Future Public Land Mobile Telecommunications System (FPLMTS) that aims to unify existing systems of varying diversity into a seamless radio infrastructure. The network seeks to deliver current and future wireline services, such as voice, video, and data communications to users situated anywhere in the world [3]. Future wireless networks will also seek integration with the Broadband Integrated Services Digital Network (B-ISDN), likely to be based on Asynchronous Transfer Mode (ATM) transport.

1.1. Motivation

Analog cellular systems, such as the Advanced Mobile Phone Service (AMPS), have a network infrastructure with centralized intelligence to handle call processing and mobility management functions. There is a growing trend towards decentralized network control. For example, in the digital GSM (Global Standard for Mobiles) system [4], radio link measurements are made by the Mobile Station (MS) instead of the Base Station (BS) and some handoffs do not require intervention by the Mobile Switching Center (MSC). Processing power and control are increasingly being pushed towards the edges of wireless networks.

Proposed wireless networks are expected to use microcells to facilitate frequency reuse, resulting in capacity enhancements. Microcells have the effect of generating a large flow of signaling traffic relating to location updates by mobiles as cell boundary crossings become more frequent. Such increases in signaling will be exacerbated by the anticipated large volume of consumer participation in PCS. These increased traffic flows will place considerable pressure on centralized network controllers to expedite call setup and handoffs. The speed of the latter will be critical to the Quality of Service (QoS) that a network operator can provide. Rapid drops in radio link quality will be more commonplace in microcellular environments, especially at higher frequencies, and congestion of the signaling network will result in a greater likelihood of the network 'dropping' a call. It is useful, therefore, to consider architectures where signaling is distributed among network elements.

1.2. Previous Work

A Metropolitan Area Network (MAN) based on the IEEE 802.6 standard [5] has been proposed as novel means of distributed network control [6,7] for a PCN. Based on a Distributed Queue Dual Bus (DQDB) media access control (MAC) protocol, the MAN in Figure 1.1 can interconnect BSs and Base Station Controllers (BSCs) to a mobile control centre and other network nodes such as bridges to Local Area Networks (LANs) and gateways to the Public Switched Telephone Network (PSTN) and to public data networks (PDNs).

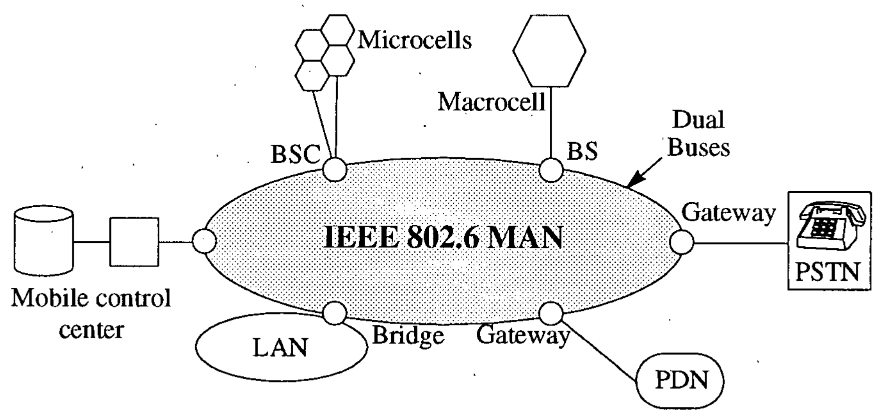

Figure 1.1. PCN architecture based on the IEEE 802.6 MAN

Subsequent work has proposed and analyzed network protocols [8,9] for PCS, and voice transport for pre-arbitrated (PA) [9] and queue arbitrated (QA) [10] media access. A hierarchical, distributed database design is also presented in [7]. Interworking the MAN-based PCN with the Digital European Cordless Telephone (DECT) standard has been investigated [11]. The reservation-arbitrated (RA) media access protocol [12] is being proposed to transport isochronous (real-time) traffic with a constant delay while enabling full statistical multiplexing among calls.

1.3. Objectives

This thesis attempts to further establish the IEEE 802.6 MAN, using PA access for voice, as a viable network architecture to support PCS. It is important to clarify the use of PA access as an alternative to other forms of transport on the IEEE 802.6 MAN, the ultimate choice of which will depend on the network operator's criteria for voice quality, fixed network capacity, flexibility, and ease of implementation. The objectives of the thesis work are:

i. to consider a modified PA voice transport scheme for low-bit rate encoding (LBE) mobile speech coders (e.g. at 8 kbps);

ii. to determine the enhanced Erlang and subscriber capacity of the MAN with modified PA voice transport and an advantageous placement of a gateway to the PSTN;

iii. to reduce PCN signaling delays on the MAN in the presence of low priority packet data using decentralized location updating and the Preemptive Priority Mechanism (PPM) [13]; and

iv. to consider a PCN network architecture based on multiple MANs and to develop its evolution towards ATM/B-ISDN.

This work differs from previous investigations in the following important ways:

- a lower coding rate (8 kbps) for mobiles is assumed
- mobile-to-mobile, mobile-to-fixed, and fixed-to-mobile calls are considered
- capacity gains from interconnection of the MAN with the PSTN are realized
- the performance of each type of signaling (e.g. call setup, location update, call hand-off) is influenced by packet data and other signaling flows
- the MAN-based PCN is presented in the context of the Intelligent Network (IN) and ATM

1.4. Outline of the Thesis

In chapter 2, relevant background information is provided on the IEEE 802.6 MAN, network modeling, and Intelligent Network concepts. Chapter 3 expands on priorities in DQDB and their relevance to a PCN, and investigates the Preemptive Priority Mechanism. Aspects of the model development of the MAN-based PCN are given in Chapter 4, with specific treatment given to voice transport, network configuration and capacity improvements. Chapter 5 considers a single-MAN PCN architecture to determine the Erlang and subscriber capacity, and to generate signaling traffic to determine if the application of the PPM to a MAN-based PCN is useful. Chapter 6 examines a PCN architecture based on multiple MANs and interworking with ATM. Chapter 7 summarizes the findings and makes suggestions for future work.

Chapter 2. Metropolitan Area Networks and the Intelligent Network

This chapter provides a brief technical description of the IEEE 802.6 MAN, specifically its architecture and protocols, network modeling, and an introduction to concepts of the Intelligent Network.

2.1. Metropolitan Area Networks

Metropolitan Area Networks (MANs), as their name suggests, seek to interconnect network nodes over a potentially large metropolitan area. There are two varieties of MANs that have received much attention - the Fiber Distributed Data Interface (FDDI) network and the Distributed Queue Dual Bus (DQDB) network as specified in the IEEE 802.6 standard [5].

FDDI and its enhanced version, FDDI-II, are based on counter-directional optical fiber rings with data rates of 100 Mbps and token-passing for media access. These standards have been adopted by industry and there are many commercial implementations, often as high-speed LANs or campus area networks.

The IEEE 802.6 MAN uses two counter-directional slotted buses to interconnect nodes which send asynchronous traffic using a distributed queueing algorithm. The standard also supports circuit-switching for isochronous (real-time) services. Implementations of the technology for private networks are rare, if any, but companies such as Alcatel, Siemens, and QPSX are distributing public networking equipment based on DQDB. In academia, interest in DQDB continues and the standard is viewed by many as an evolutionary step towards the ATM/B-ISDN [14,15]. The decision to make the Switched Multimegabit Data Service (SMDS) DQDB-based is early evidence of this evolutionary trend.

2.1.1. The IEEE 802.6 MAN Architecture

Figure 2.1 illustrates the basic open-bus architecture for the IEEE 802.6 MAN. Slots of fifty-three bytes are generated by the Head of Bus (HOB) nodes and sent downstream on each bus. Nodes have read and write access to the slots on both buses, permitting full duplex communications between any two nodes.

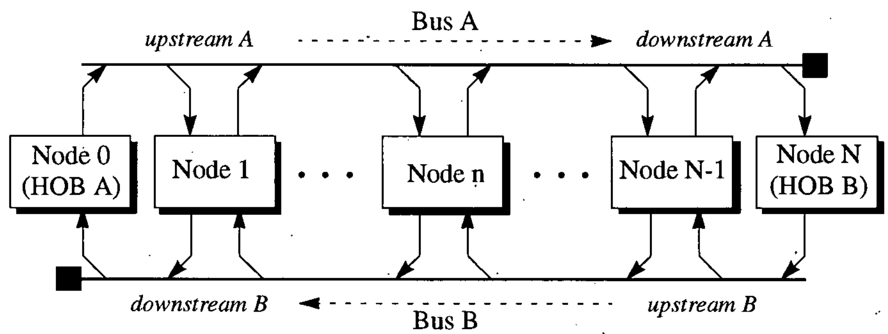

Figure 2.1. IEEE 802.6 MAN open bus topology

A closed-bus architecture, which collocates the two HOBs, improves fault tolerance. If there is a bus fault, the MAN can reconfigure the HOB functions to nodes adjacent to the fault so that an open bus MAN is effectively created.

A 125 μsec clock controls the generation and framing of slots at the HOBs. Within a frame, a number of slots are created and are designated for either asynchronous or isochronous services. Both slot types share the same header format, shown in Figure 2.2. The slot is composed of an Access Control Field (ACF), Segment Header (SH), and Segment Payload. Details of all fields can be found in the standard [5].

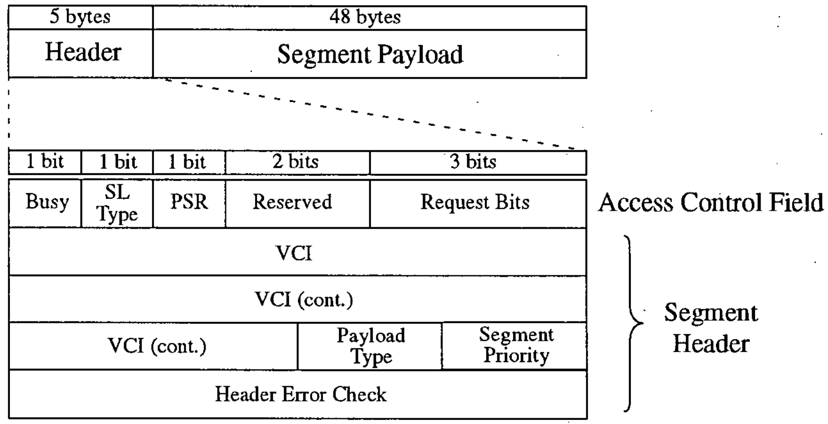

Figure 2.2. IEEE 802.6 MAN slot format

2.1.2. Scope of the IEEE 802.6 Standard

The IEEE 802.6 standard specifies the physical and media access control (MAC) layers of a MAN. There are several Physical Layer Convergence Procedures (PLCP) provided for physical layers such as DS3 at 44.736 Mbps, and Synchronous Digital Hierarchy (SDH)-based systems at 155.52 Mbps. This is consistent with the spirit of ATM, which also allows for various transmission speeds.

The MAC protocol layer is commonly known as the DQDB layer and refers not only to asynchronous media access by the distributed queueing algorithm, but also to isochronous access. The DQDB layer is organized into three levels, as is shown in Figure 2.3. The Common Functions block has the responsibility of relaying slot octets and management information octets between the two service access points (SAP) to the Physical Layer. It also allows the Arbitrated Functions block to gain read and write access to the slot octets as they are received. There are two Arbitrated Functions that are responsible for controlling access to the buses; the queue arbitrated (QA) Function controls asynchronous transmissions, while the pre-arbitrated (PA) Function controls isochronous transmissions. Above this layer is the Convergence Functions block whose responsibility is to map services from higher layers to the 48-byte segment payload. There may be several convergence functions that have varying degrees of functionality. For instance, a long message from higher layers may require the assembly of an Initial MAC Protocol Data Unit (IMPDU), which is subsequently broken into more manageable pieces by a Segmentation and Reassembly (SAR) function. Isochronous services may require functions such as buffering to smooth rate differences at higher layers.

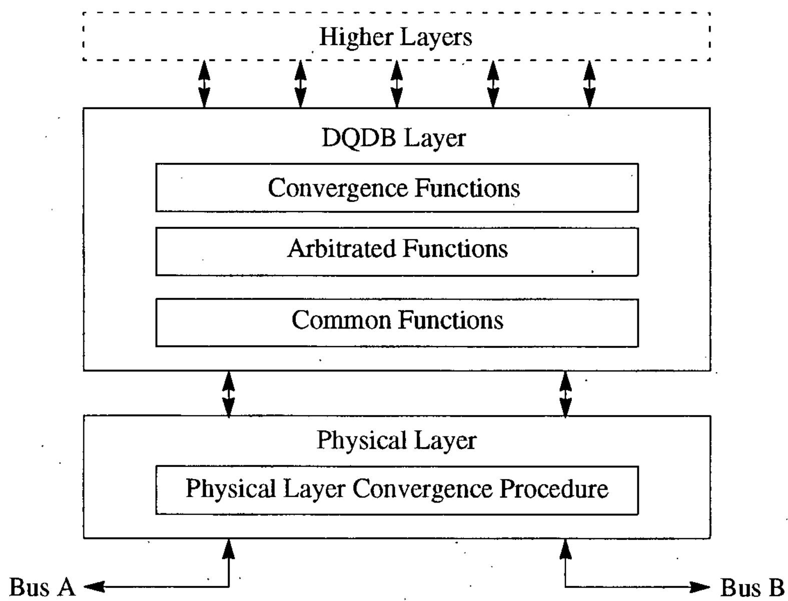

Figure 2.3. Scope of the IEEE 802.6 standard

2.1.3. Isochronous Transport

Services that have fixed bandwidth and delay requirements can use the PA Function for consistent media access. A connection setup phase (abstracted from CCITT Recommendation Q.931 in the standard) between two nodes will establish a dedicated full-duplex circuit by assigning a Virtual Channel Identifier (VCI) to a fixed number of pre-arbitrated slots in each frame. The

HOB function, having knowledge of the call, writes the VCI in the headers of the designated slots before sending them downstream. The two nodes involved, recognizing the VCIs that have been assigned to their call, can receive and write to the correct slots. In the standard, nodes may read or write to sub-slot payloads as small as 8 bits. With the basic frame length of 125 μsec, this enables real-time 64 kbps communications. Nodes, however, must be aware not only of the VCIs, but also the positional offsets in the slot that specify individual circuits. These are negotiated during the connection setup phase.

2.1.4. Asynchronous Transport

Services that use asynchronous transport, which may be connection-oriented or connectionless, contend for slots that are not designated as PA. The distributed queueing algorithm provides decentralized media access control to QA slots at three levels of priority. A VCI, which may specify a destination node or a pre-established connection [8], is written to the slot header by the node that gains access to it. The standard [5] provides a full description and state diagrams of the QA Function.

The fairness of the queueing algorithm is susceptible to long propagation delays, as nodes will sometimes have inaccurate impressions about loading in the rest of the network. Many studies and recommendations have resulted regarding how to remedy two of the problems that arise in DQDB networks: bandwidth imbalance among nodes [16] and an ineffective priority mechanism [17]. A solution to the former, Bandwidth Balancing (BWB) [16], has been included in the standard, but it has been shown to work incorrectly in certain situations for multiple, preemptive priorities [18].

2.1.5 Network Modeling

2.1.5.1. Development Environment

Simulations were developed with MIL 3, Inc.'s Opnet general communications network simulator, release 2.5A, on a Sun SPARCstation. The Opnet environment allows for hierarchical model development at three main levels of abstraction: process, node, and (sub)network. Open System Interconnection (OSI) protocol layering is readily accommodated. The process level uses a state diagram framework with user-embedded C code at desired states to invoke discrete time interrupts. This permits the flexibility to develop and alter network protocols with relative ease, while maintaining excellent readability.

2.1.5.2. Implementation of the IEEE 802.6 MAC Protocol

The MAC protocol of the IEEE 802.6 standard was implemented at the process level within Opnet, using an International Organization for Standardization (ISO) document [5] as the main reference. Two different processes were required to allow for the different actions at the HOB and non-HOB nodes. The complexity required to accommodate three QA priority levels suggested that code, rather than a graphical state machine representation was the best means to achieve flexibility and ease of alteration. As a result, few explicit states were required.

Bandwidth Balancing, although part of the standard, was not implemented due to reported inconsistencies in its action with respect to priorities and with the difficulties of choosing an appropriate BWB modulus [19]. The Preemptive Priority Mechanism [13], to be introduced in Chapter 3, was added to the original protocol and is easily switched on and off by a simulation-level parameter.

Many studies have been performed on eraser nodes [20] and other slot reuse techniques to improve capacity and media access delay. These will be exempted from this study for simplicity.

To achieve a balance between processing economies and modeling accuracy, simulations incorporate QA access only, with three levels of priority. PA access is not included in the model. As such, the amount of bandwidth available for QA traffic is fixed. This is contrary to common sense, since PA bandwidth that is laying unused would be assigned to QA traffic by the HOB in a MAN implementation. Consequently, delay figures obtained for signaling in Chapter 5 will be higher than for simulations that dynamically assign unused PA slots for QA traffic. This shortcoming should not pose a problem, however, as the purpose of the simulations, in general, is to obtain reductions in signaling delays.

Since a large percentage of bandwidth is used for speech, simulation times can be drastically reduced by excluding PA access from the simulation models. Adjustments are made to segment and signaling delay figures to properly accommodate this processing efficiency. The approximate time for a 50-node MAN to process 2,020,000 QA slots is 33 hours on a Sun SPARC 5. If QA slots consume 25% of the bandwidth, the inclusion of PA slots in the simulations would require 4-5 days to produce the commensurate level of signaling traffic.

2.2. The Intelligent Network

The Intelligent Network (IN) separates the specification, implementation, and control of telecommunications services from the physical switching network [21]. It allows network operators to rapidly develop and deploy new services without drastic alterations to the network's switching fabric. The elements of the IN include [22]:

- Service Switching Point (SSP) - stored-program control switch to intercept calls that require special handling, and to query databases for call information.
- Service Control Point (SCP) - database that provides call-handling information

- Signal Transfer Point (STP) - packet switches to route signaling messages between nodes such as SSPs and SCPs. These are often found in matched pairs for redundancy.
- Service Management System (SMS) - provides operations support. The SMS may incorporate a Service Creation Environment (SCE) [23].

The Advanced Intelligent Network (AIN) specification from Bellcore includes the definition of new elements such as the Intelligent Peripheral (IP), the Service Node (SN), and the Adjunct.

To offer personal mobility and advanced services, a PCN will incorporate Intelligent Network elements in a logical layering that separates intelligence, transport, and network access as depicted in Figure 2.4 [24].

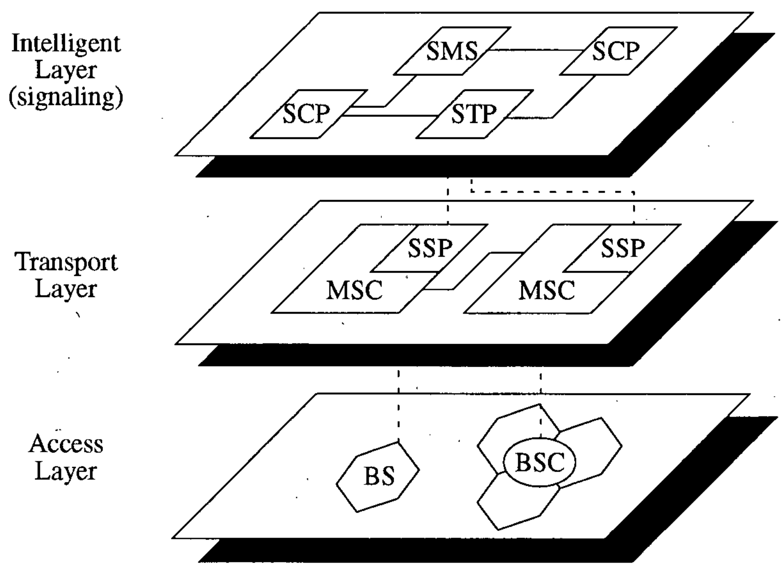

Figure 2.4. Intelligent network elements in a PCN

The underlying foundation for the Intelligent Network is Common Channel Signaling (CCS). CCS can be defined as the system that enables SSPs, SCPs, and other network nodes to exchange: i) messages relating to call and connection control; ii) information needed for distributed application processing; and iii) network management information [25].

Signaling System No.7 (SS7) is an International Telecommunications Union (ITU) CCS-based layered signaling protocol that has been widely implemented by telephone network operators. The ISDN User Part (ISUP) provides signaling functions to support the control of telephone calls, non-voice calls, and advanced ISDN and IN services. The Mobile Application Part (MAP) supports mobility for voice and data services. Both can be employed in an Intelligent Network to provide a wide range of services to mobile subscribers.

Chapter 3. Priorities in the DQDB MAN

3.1. Introduction to Priorities

The IEEE 802.6 standard specifies three levels of priority for QA traffic. The implication of priority levels is that some data segments must be delivered more quickly than others. A priority classification scheme appears in [26]. A priority scheme can be considered to be implementing one of the following:

- *Pseudo-priorities* - all that can be guaranteed is that high priority traffic with an offered load greater than that for a lower priority will receive more bandwidth.
- *Weak Priorities* - like pseudo-priorities, but with the condition that higher priority throughput is lowered if the offered load from lower priority stations increases.
- *Strong Priorities* - high priority throughput has no dependence on the offered load of lower priorities.
- *Preemptive Priorities* - lower priority traffic is not admitted when the offered load of higher priority traffic exceeds the available bandwidth.

The total delay of a segment within a DQDB network is comprised partially of media access delay. Propagation delay can be significant, but is deterministic between specific nodes. An effective preemptive priority mechanism for the DQDB MAN will work in the MAC layer[1] to achieve the following stated objectives [13]:

1. The average media access delay of high priority traffic must be substantially less than that of lower priority traffic under heavy traffic loading.
2. The access delay characteristics of any given and fixed distribution of high priority

1. It would be possible for priority schemes to work above the MAC layer.

traffic should be (almost) independent of the low priority traffic.

3.2. Applicability of Preemptive Priorities to a MAN-based PCN

A network designer has the option of using PA or QA cells for the transport of voice packets on the IEEE 802.6 MAN. In a personal communications network (PCN), voice will naturally command a large proportion of the total bus bandwidth. A MAN-based PCN utilizing, say, QA-1 slots for voice transport (level 1 is the middle level of priority) and QA-2 slots for signaling (level 2 is the highest) may experience a degradation in the perceived voice quality under high signalling loads since voice packets that experience excessive access delays will be dropped from voice segment queues. This may be acceptable depending on the quality of service that is required of the network. It may, however, be contrary to a PCS goal of achieving wireline-like voice quality (i.e. 64 kbps PCM).

A preemptive priority mechanism is consistent when isochronous PA cells are used for high-quality voice transport in the (wired) network. Voice transport then has guaranteed access to the buses with deterministic delay and, subsequently, no packet loss at the MAC layer of the MAN. A high signaling load will not affect voice transport as the two methods of access are separated in the MAC layer.

After voice has been assigned to PA cells, one may choose to consider the assignment of traffic to QA priorities as shown below in Table 3.1.

QA-0	QA-1	QA-2
Wireless Packet Data	Call Processing and Location Updates	Handoffs

Table 3.1. Potential QA priority assignments in a MAN-based PCN

Handoffs should experience the least delay, since excessive delay may result in dropped calls, a very negative service implication for the PCN operator. From their point of view, some incremental rise in delays for call setup, call release, and location updates can be tolerated to reduce handoff delays. Any perceptible increase in call setup times under high handoff loads should be small. The argument for handoffs gaining priority over call processing and location updates becomes clear. This separation of signaling into different levels of priority should be preferred over the case for QA voice transport where all signaling occurs at the same level of priority (if packet data traffic is allocated to QA-0).

3.3. The Preemptive Priority Mechanism

3.3.1. Introduction

Previous studies have shown the 802.6 QA MAC (media access control) protocol to be ineffective in implementing preemptive priorities with respect to the stated objectives. An imperfect knowledge of the rest of the network due to propagation delays means that a node can only be guaranteed to obey priorities locally and not globally (i.e. with respect to the rest of the network). Furthermore, it has been shown that the current IEEE 802.6 standard implements *pseudo-priorities* [26].

To the end of achieving preemptive priorities for the DQDB MAN, a preemptive priority mechanism (PPM) mechanism has been suggested [13]. The mechanism is an enhancement to the existing IEEE 802.6 standard. Additionally, it conforms to the MAN ACF format by utilizing two bits that have been designated as Reserved. Other schemes such as that using global priority information [16] have attempted to remedy the problems of ineffective priorities and bandwidth balancing, but simulations in [26] have demonstrated long transient periods following network load changes, adversely affecting the access delay time of higher priority traffic. DQDB +/- [26] has been proposed to overcome the aforementioned problems, but the new protocol violates the ACF format.

We choose to use a reliable mechanism that requires little modification. The issue of bandwidth balancing remains. However, this problem is of serious concern only for nodes transmitting continuous streams of traffic, with a resulting domination of bandwidth. For a PCN, domination will not occur at priority levels 1 and 2 with the proposed allocation of signaling, and is unlikely at level 0 since there will always be local preemption from higher priority segments.

Studies have demonstrated that QA access delays at one node are dependent on the loading of lower priority segments at other nodes. This situation is not acceptable in many circumstances. The PPM will rectify this, subject to network propagation delays. Because nodes may become heavily loaded with QA traffic of lower priorities, and considering the time-critical nature of handoffs, the use of a preemptive priority mechanism in the 802.6 MAN is desirable.

3.3.2. Changes to the DQDB MAC Protocol

The PPM makes use of two Reserved bits in the slot ACF. These bits are used on the reverse bus to continuously indicate the highest priority of traffic awaiting transmission downstream. A node indicates to upstream nodes the value of the highest priority segment that is waiting to be transmitted downstream. This value will be referred to as the HPT, or the highest priority traffic. The two newly utilized bits compose the priority field or PF. Letting xy_2 denote the binary contents of the PF, xy_2 may take on the following magnitudes:

$$|xy_2| = \begin{cases} 0, \text{ if } xy_2 = 00_2 \\ 1, \text{ if } xy_2 = 01_2 \\ 2, \text{ if } \quad x_2 = 1_2 \end{cases}$$

Each node OR-writes either 00_2, or 01_2, or 10_2 to the PF field corresponding to an HPT of 0, 1 or 2, respectively. If there is no traffic present, the default is 00_2 since writing this value will have no effect on the queues of any priority. OR-writing of the HPT is flexible in that a *preemption distance*, *d*, can be specified. This refers to the minimum number of segments that must exist at the HPT in order for an attempt at OR-writing to take place. The default is $d = 2$. The reason for this is that, if a lone segment is waiting at a node, the request that is issued through normal DQDB operation will ensure its proper treatment. It is only when the queue grows past one segment that the PPM mechanism should be enabled. Standard DQDB operation corresponds to the case where $d = \infty$ and the PPM is deactivated so that no OR-writing of the PF field occurs.

After reading the PF bits just received on the reverse bus, a node will make a decision as to whether it should transmit a segment of its own. Upstream nodes will stop transmitting packets of lower priority when the PF bits they read on the opposite bus indicate that a downstream node has higher priority traffic. This is done by stopping the request counters and countdown counters of lower priority traffic from being decremented. It is then impossible for a lower priority segment to move up in the virtual queue of that priority.

3.3.3. Previous Analysis of the PPM

As was found in [13], there is some bandwidth wastage in the protocol when switching from high to low loads as segments waiting at a lower priority could have been transmitted in passing slots. This waste is not quantifiable and depends on the size and traffic patterns of the network. In underload situations the wastage will be minimal. In times of overload, though, the throughput of lower priorities will be sacrificed somewhat to ensure that high priority traffic may access the network with independence. The amount of wasted bandwidth decreases with increasing *d*, as nodes with lower priority traffic are less inclined to cease transmissions. The trade-off here is that high priority traffic will become increasingly dependent on lower priority segments. The choice of *d*, then, depends on the criteria of the network designer.

The size of the network affects the time taken by the PPM to take full effect. This is due to the nature of distributed networks in that each node has old information about the rest of the network due to link propagation and node processing delays. It takes time for nodes to receive new PF values as they propagate upstream, but the PPM limits the time to preempt lower priority traffic at a particular node to a maximum of one full round-trip delay.

3.3.4. A Simple Test of the PPM

Using a simple network configuration consisting of three nodes with each offering segments of a unique priority, the effectiveness of the PPM can be illustrated. Consider the MAN configuration shown in Figure 3.1. Three nodes produce segments for transmission on Bus A. The most upstream node sends packets of the lowest priority (level 0), while the most downstream node sends packets of the highest priority (level 2) and the node situated in the middle of the bus sends packets of medium priority (level 1). Times at which nodes start to produce segments for transmission will be staggered such that lower priority traffic has full access to the bus before higher priority traffic enters the network. The delay between nodes is equal to one slot length in duration (one slot length is also equivalent to a distance of about 550 meters with an SDH rate of

155.52 Mbps). Segment interarrival times are assumed to be independent and exponentially distributed with an arrival rate of λ segments per passing MAN slot. Under preemptive priority conditions, lower priority traffic should yield to high priority traffic for significantly lower access delays.

Two cases are simulated and compared: normal DQDB operation (i.e. with the PPM disabled), and DQDB with the PPM enabled and a preemption distance of $d = 2$. The chosen configuration represents a simple worst-case scenario for the former since upstream stations heavily loaded with low priority traffic are known to affect the access delay for higher priority segments that are awaiting transmission downstream.

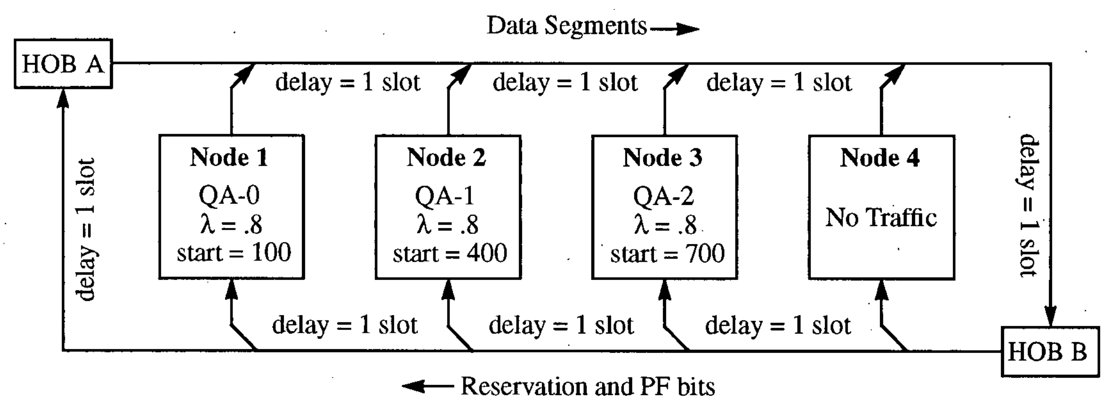

Figure 3.1. MAN configuration to illustrate the effectiveness of the PPM

Figure 3.2 shows bus access delays for each node's packets as a function of the simulation time in the case of normal DQDB operation. The window shown is small but illustrative, and is representative of the network behaviour beyond the window. The original DQDB mechanism, corresponding to a deactivated PPM (preemption distance, $d = \infty$), exhibits undesirable behaviour as nodes with higher priority traffic begin to generate segments. While only nodes 1 and 2 are on, the access delay of packets from node 2 is less than for packets from node 1 but is dependent on the presence of lower priority traffic. For this particular simulation seed, node 3, with the highest priority traffic, exhibits access delays that begin to exceed those of node 2 towards the end of the window. All three queues grow without bound in the long-run. Preemptive priorities have clearly not been achieved.

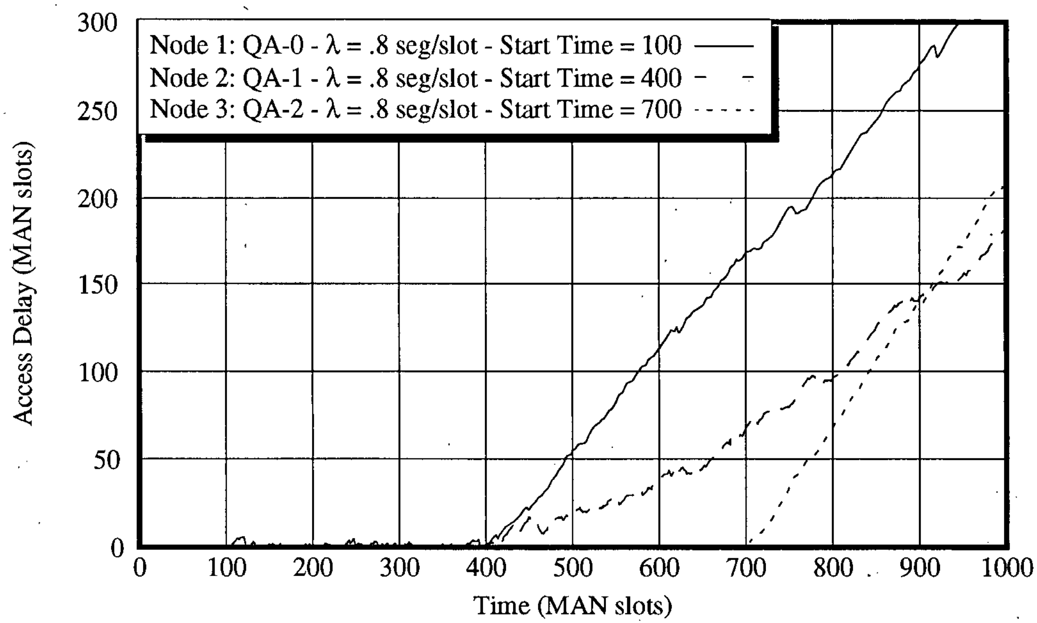

Figure 3.2. Ineffective priorities under normal DQDB operation ($d = \infty$)

In the same scenario, but with PPM activated and results shown in Figure 3.3, node access delay is consistent with the stated objectives of preemptive priorities. As nodes with higher priority traffic transmit, nodes with lower priority traffic reduce their transmissions in response to the PF bits that are set downstream. At *time* = 400 slots, node 1 cuts its transmissions significantly and the access delays for node 2 segments are bounded and independent of the offered load at node 1. Similiarly, node 3 access delays are bounded and independent of what happens at nodes 1 and 2. It is worthwhile noting that node 1 ceases transmitting altogether as soon as all nodes begin to offer traffic. The total offered load of higher priorities is $\lambda = 1.6$ segments/slot, giving node 1 no opportunity to transmit its segments.

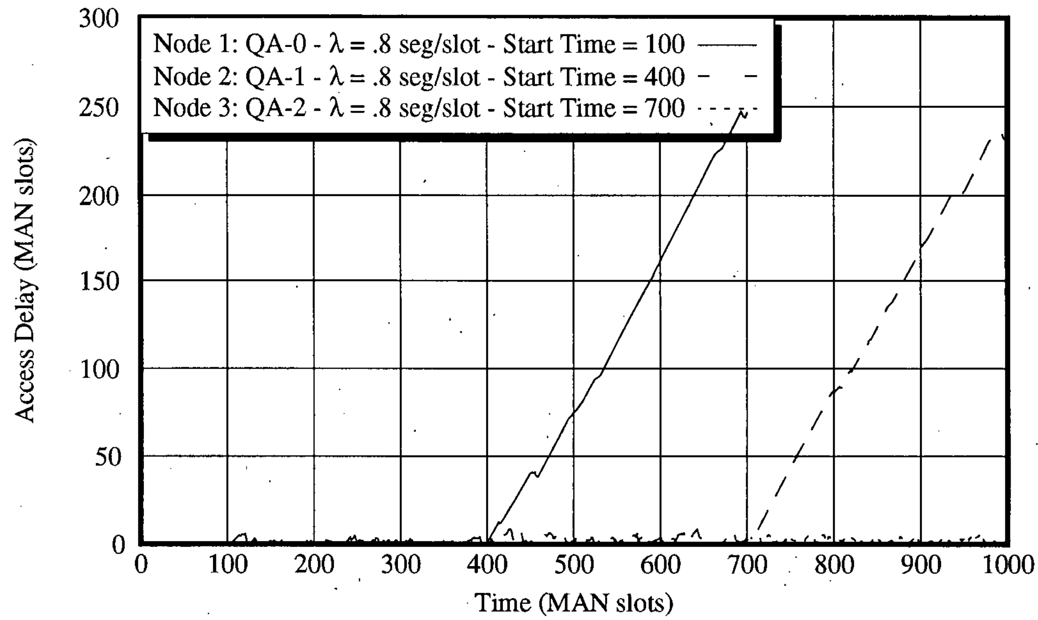

Figure 3.3. Effective priorities using the PPM (d = 2)

3.4. Further Analyses of the PPM

3.4.1. Uniformly Distributed Load

Having illustrated the effectiveness of the PPM with a simple network configuration, the long-term effectiveness of the mechanism must be shown. Essentially, improvements in the bus access delay for high priority traffic are sought. A five node, open-bus network configuration, similiar to that shown in Figure 3.1, is employed. Once again, transmissions on Bus A are considered. Each node offers the *same* amount of traffic to Bus A for a cumulative offered load of 2.4 data segments/MAN slot. Two priorities of QA traffic, QA-1 and QA-2, are offered in varying proportions but consistently throughout all nodes. The proportions are such that total QA-1 traffic exceeds the bus capacity (overloaded) while total QA-2 traffic is less than capacity (underloaded).

The delay between nodes is extended to the equivalent of five slot lengths.

Three instances of both DQDB without the PPM and DQDB with the PPM activated are simulated with a duration corresponding to 1 million MAN slots (2.73 seconds). Statistics are collected after a period of 10,000 slots to allow for a generous transient period. Figure 3.4 indicates the average access delays of QA-2 traffic at each node along the bus for all simulations. As expected, due to the unfairness in the DQDB protocol, the values increase with distance from the Head of Bus A (node 0 in this case). What is most noticeable is the dramatic improvement in access delays when the PPM is activated, especially at higher indexed nodes. With an 80/20 split between QA-1 and QA-2 traffic the access delays at node 4 are reduced by over 75%.

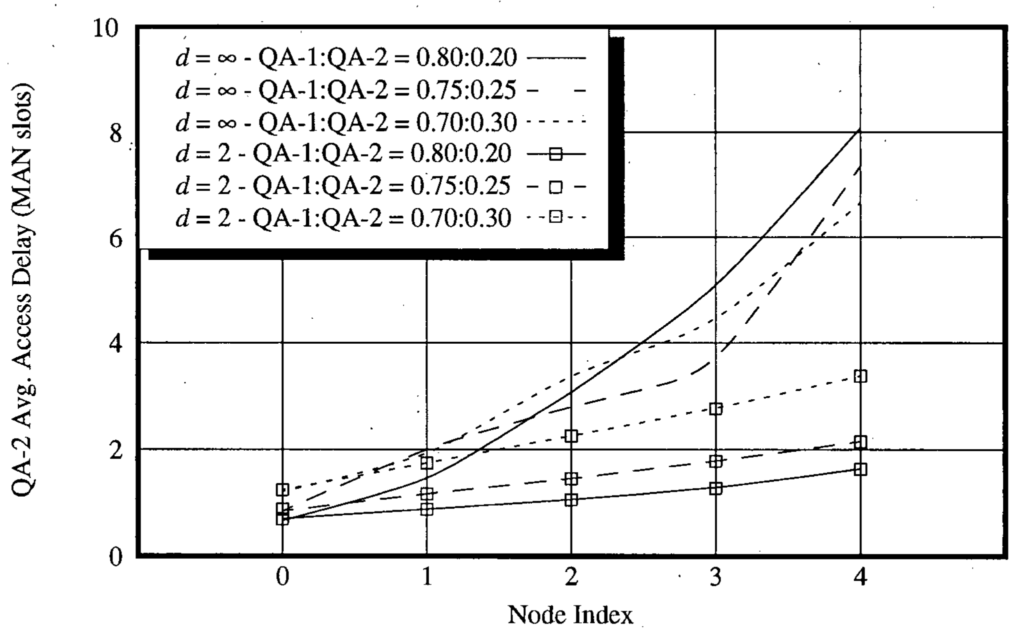

Figure 3.4. Improvements in QA-2 access delay - uniform distributed load

The PPM mechanism also appears to bring about some consistency with respect to delays resulting from differing loads. For example, witness the relative delay values without the PPM at node 4. The average access delay actually increases as the QA-2 load decreases. This problem appears at other nodes also. With the PPM activated, however, the average access delay increases at all nodes as the load increases.

3.4.2. Linearly Distributed Load

In this case, the offered load per node will decrease linearly with the distance from the HOB. This situation corresponds to a uniform distribution of destination addresses for segments being delivered to both buses. The total offered load will remain at 2.4 segments/slot/bus, as will the chosen proportions of QA-1 and QA-2 traffic at each node.

Contrary to results in [13], the noticeable change in average access delays for QA-2 traffic, indicated by Figure 3.5, is very small, but still favours the PPM. This discrepancy in results may be due to detailed timing considerations not made available in the original work. However, these results can be partially explained. The increased (compared to uniform loading) QA-2 load on more upstream nodes will mean that QA-1 traffic in these nodes will have fewer opportunities to influence behaviour in the rest of the network since their transmissions are being preempted locally in the standard DQDB protocol. Downstream, where loads have been decreased relatively, stations will have fewer occasions to preempt. The combined effect is to significantly reduce the effect of using the PPM *for this particular loading case*.

There are too many potential loading combinations to consider. These few examples indicate that the PPM will provide either no improvement to higher priority access delays or significant improvement, depending on the loading distributions. Many instances will exist, however, where delay statistics can be reduced. Considering the simplicity of its operation and implementation, the PPM is recommended as an effective technique for reducing signaling delays in a MAN-based PCN.

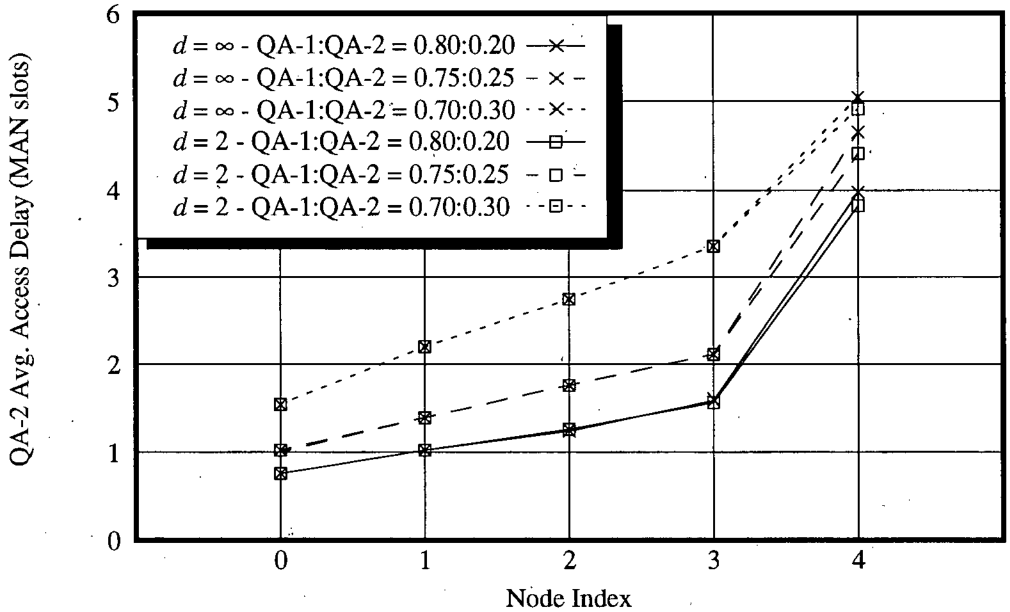

Figure 3.5. Improvements in QA-2 access delay - linearly distributed load

Chapter 4. PCS Voice Transport, MAN Configuration and Erlang Capacity

In this chapter, several key issues are explored that will impact the work that follows in subsequent chapters. Voice transport has implications for delay, speech quality, and overall network capacity. The configuration of signaling elements, databases, and gateways on the IEEE 802.6 MAN-based PCN impacts capacity and signaling. Finally, the related issues of blocking probability and Erlang capacity are presented in the context of the IEEE 802.6 MAN.

4.1. Supporting PCS Voice Transport on the IEEE 802.6 MAN

4.1.1. Voice Quality

There is a strong desire to provide high toll-quality, wireline-like, speech services in future PCNs [2]. This is required to draw potential subscribers to a service that will compete for access with the conventional local telephony loop. New speech coders are achieving better quality at lower bit rates, thereby reducing the amount of radio frequency spectrum required per call at the air interface.

There is a growing trend towards variable-rate speech coders as is evidenced in the IS-96 voice coding standard for Code Division Multiple Access (CDMA) [27]. One of four different bit rates is used to code a 20 msec frame of speech, depending on the speech energy level of the frame. For example, instead of using a full bit rate of 8 kbps to code portions of silence, a low rate of 800 bps is used with little or no loss of quality. The corresponding reduction in transmitted bits over the air-interface allows more subscriber transmissions over the same RF link.

Network designers are considering speech interpolation techniques that cease transmission during periods of silence. This extreme of variable-rate coding, in which the speech coder moves between rates of 0 bps and the full-rate, is also known as Voice Activity Detection (VAD). It allows channels on the air interface to be torn down and reused by voice channels which enter a period of speech activity, as is the case with Hughes' Extended-TDMA [28] and the Packet Reservation Multiple Access (PRMA) scheme [29]. Capacity increases on the order of two or more can be realized. In order to reduce co-channel interference and save on battery life, GSM has a Discontinuous Transmission (DTx) option that works on the same principal but maintains the same radio channel throughout a voice call.

Some voice quality issues arise as a result of VAD schemes:

- Voice clipping results from delays in switching from one coding rate to the other (this effect will become less apparent with more coding levels). It is often difficult to detect the onset of speech, especially in a noisy environment, such as in an automobile [2].
- Voice clipping can occur due to an unavailability of channels to be assigned when speech activity resumes.
- The absence of background noise during speech silence may be disconcerting to listeners. Some 'comfort' noise may have to be artificially generated at the receiver to remove the perception of abrupt transitions between silence and speech [28].

These factors improve the case for multi-rate vocoders or constant rate vocoders without VAD in a PCN network where high voice quality is desired.

4.1.2. Media Access Alternatives for Voice Traffic

As discussed in Chapter 3, either QA or PA slots may be used for voice transport on the DQDB MAN. For air interfaces like those of E-TDMA and PRMA, QA transport is an efficient fit due to its asynchronous packet nature. Care will have to be taken not to overload the MAN network such that excessive delays cause MAN voice slots to be dropped, with a subsequent loss of voice quality.

For air interface speech coding schemes not using VAD, QA transport is possible but the overhead for QA slots is 80% higher than for PA slots. The problem of slot delay variability must also be weighed against any potential benefits.

QA transport with VAD at the MAN/PSTN interface is possible, but there is a risk of degrading speech quality in order to use less wired network capacity and thereby increase the potential MAN coverage area. Any capacity bottleneck should be designed to occur at the air interface and not in the wired network. Network operators may not wish to sacrifice voice quality to reduce the landline capacity required per call.

PA transport on the MAN is intended for isochronous users communicating at 64 kbps. The standard MAN frame length of 125 μsec may be extended to permit the lower bit rates of speech coders used in mobile communications. PA transport guarantees zero delay variability in the slot access times, so there will be no loss of voice quality due to segment dropping. VAD may be employed with a variation of PA access known as Bi-State PA [9], although this recalls the issue of speech quality. PA is well suited to constant rate coding. In the case of multi-rate coders, there will be some wastage of MAN bandwidth when PA slots accommodate all but the highest rate of coding. Efficient MAN transport of multi-rate coding schemes is not considered here, although it is anticipated that this would increase the network complexity and introduce some voice clipping. Table 4.1 summarizes the speech quality and transport issues.

Air I/F Coding	Const. Rate w/ VAD		Multi-Rate		Constant Rate	
Example(s)	E-TDMA, PRMA		QCDMA (IS-95)		IS-54, GSM	
MAN Transport	QA	PA	QA w/ VAD	PA	QA w/ VAD	PA
Lost speech information?	Yes	Yes	Yes	Some	Yes	No
Relative MAN BW Efficiency	High	Low	High	Low	High	Low

Table 4.1. Speech quality and MAN transport for mobile speech coding schemes

4.1.3. Transcoding

Transcoding for speech refers to the translation of digitized voice at a particular bit rate to another rate. This is a common operation in digital cellular where speech that is coded at, say, 8, 13, 32,or 64 kbps must be transcoded to a rate suitable for transport in a different part of the network. Some loss of speech quality may result from transcoding. The frequency of transcoding stages should, therefore, be minimized.

For a MAN-based PCN architecture, transcoding is required to accommodate translation between low bit-rate encoding (LBE) and coding in the PSTN and, potentially, between different LBE rates if the PCN is used to accommodate different air interfaces. In the simulations that follow, a single LBE rate for mobile vocoders is assumed. Speech can be transported at this rate on the MAN and not at some higher, common rate such as PCM (64 kbps). This will allow greater calling loads to be offered to the MAN. The transcoding operation is needed only for calls involving the fixed network (i.e. mobile-to-fixed and fixed-to-mobile) and can be located either at the MAN gateway or between the gateway and the fixed network, as shown in Figure 4.1.

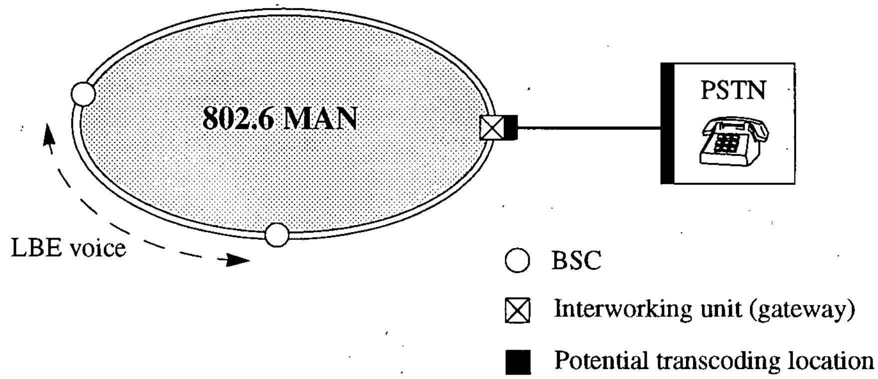

Figure 4.1. Location of transcoding operation

4.1.4. Delay and Call Packing

Excessive end-to-end delay of speech, as in a satellite communications link, is more a source of irritability than a factor of speech playback quality. However, delay affects overall conversation quality. Delays for cordless telephony standards are usually less than 50 msec, with delays for CT-2 being less than 10 msec [2]. Simple air-interface standards that require relatively little signal processing minimize delays. For cellular, the harsher radio link environment requires advanced signal processing for operations such as low bit-rate encoding, channel coding and decoding, bit interleaving and equalization. The end-to-end delay is often in the range of 200 msec [2].

Using the IEEE 802.6 MAN as a PCN infrastructure can cause delay since some packetization[1] into MAN slots is required. At the basic frame length of 125 μsec, a single octet (8 bits) within each frame provides transport to a 64 kbps source. Each PA slot can then accommodate 48 PCM (pulse-code modulation) voice sources. The additional end-to-end delay resulting from packetization - in this case 8 bits / 64,000 bps = 125 μsec - corresponds to the MAN frame

1. Packetization delay results from accumulating enough source bits to fit a target payload.

length. This format of voice transport is illustrated in Figure 4.2.

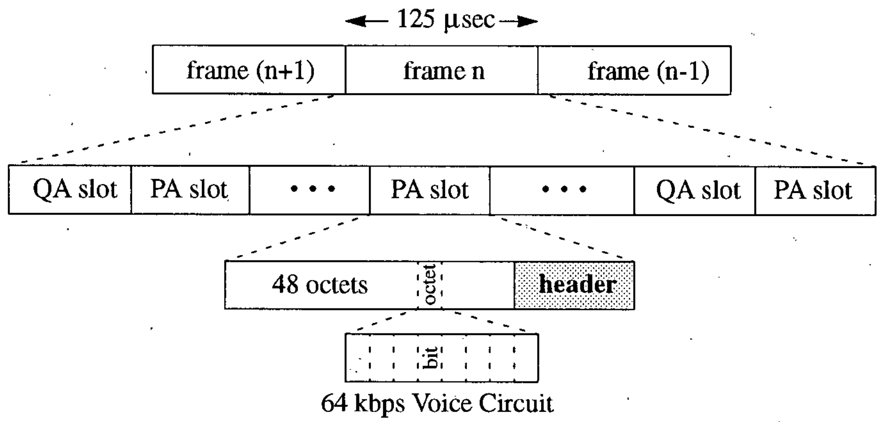

Figure 4.2. IEEE 802.6 standard voice (isochronous) transport

In QA access and Bi-state PA access, the full data payload is used (44 bytes for QA, 48 for PA) by a single voice source. At low bit rates, packetization delay can be excessive. For a 8 kbps source, 48 msec is required to fill the 48-byte segment payload of a PA slot. This figure doubles to 96 msec when the coding rate drops to 4 kbps. To mitigate this problem, multiple calls per slot may be used, but with the constraint that a simplex connection is not spread across multiple slots within a frame. It is advantageous to maintain a small number of circuits per cell since there is: i) reduced nodal processing, and ii) more efficient routing of slots to other parts of the PCN (e.g. in an ATM network).

When considering the frame and slot format for voice transport, it is useful to consider the speech coder frame length used at the mobile and base station. Many speech coders used or proposed for digital cellular have 20 msec frame lengths. That is, 20 msec of a speech waveform is digitally encoded at a time to generate a bit stream. To reduce hardware complexity at the BSC, a MAN frame length that matches that of the speech coder can be used.

In the simulations that are performed, two assumptions are made that affect voice transport and the capacity of the MAN:

1. Speech coders for the air interface use 20 msec frame lengths.

2. A maximum speech coding rate of 9.6 kbps is transported on the MAN.

The second assumption implies that channel coding and decoding for the mobile environment takes place somewhere between the BS and the BSC and not at a centralized location on the MAN.

Based on the above assumptions, a MAN transport format that uses a frame length of 20 msec (this is also the packetization delay) and fits two simplex connections per PA slot is proposed. The PA slot payload is partitioned to fit 192 bits (9.6 kbps $\times$ 20 msec) per connection as shown in Figure 4.3. Some wastage will inevitably occur if rates less then 9.6 kbps are carried on the MAN. Such is the case with the standards IS-54 and IS-136 (8 kbps) and with IS-96 (8.6 kbps peak rate [27]). 16.7% of the payload is wasted with 8 kbps coding.

A similiar transport scheme has also been proposed for an ATM-based PCS architecture [30] and requires a modified ATM Adaptation Layer (AAL). Similiar call transport schemes in the MAN and in ATM will facilitate migration of the MAN-based PCN towards ATM/B-ISDN.

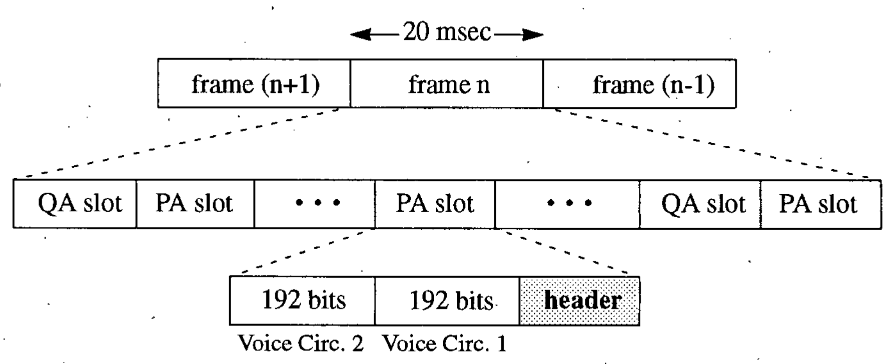

Figure 4.3. Proposed voice transport for low bit rate encoded speech with PA slots

Alternative voice transport schemes with varying frame lengths and calls per slot may require consideration for other popular coding rates such as the one used by GSM at 13 kbps and Qualcomm Inc.'s Pure Voice vocoder, also at 13 kbps. In this case, 20 msec frame lengths will leave much waste as only one call at 260 bits per frame can fit into the 384-bit PA slot payload without overflowing into other slots. Smaller frame lengths must be considered.

4.2. The IEEE 802.6 MAN Configuration

4.2.1. MAN Signaling Architecture

The effects of various signaling architectures on call setup delays has been investigated [9]. The largest delay reductions were found with the scheme having the following features:

- the Signaling Termination (ST) function distributed among all nodes;
- a closed-bus (loop) configuration; and
- a single Bandwidth Manager (BWM) function co-located with the two HOBs.

This configuration only will be considered for simulations. These features take advantage of the substantial signaling that occurs between the BSCs and the ST function, and between the BWM and the HOBs to reduce the cumulative access delay that is experienced by messages during call setup. The configuration used has positive implications for handoffs that may require the establishment of a new duplex voice channel on another MAN. The use of the closed-bus configuration may require longer cabling runs but the added fault tolerance is desirable.

4.2.2. Network Databases

PCN databases store static mobile subscriber information such as service profiles as well as dynamic information to support user mobility. In GSM [4], a mobile subscriber is permanently registered at a Home Location Register (HLR). The HLR stores service profiles of its mobile subscribers and a pointer to another database, the Visitor Location Register (VLR). The VLR is specific to a particular coverage area and tracks visiting subscribers. As subscribers roam from the current coverage area into a new one, their service profiles must be made available at the new VLR database; as well, the VLR pointer at the subscriber's HLR must be updated.

The HLR/VLR concepts have been applied to an IEEE 802.6 MAN-based PCN with a hierarchical, distributed database design [7]. Database partitions exist at the node (BSC) level as well as the MAN level to reduce subscriber location search times. Location area boundary crossings by mobile subscribers trigger location updates that may involve the transfer, copying, and/or deletion of a subscriber's service profile to or from another database. The required operation is determined by the type of boundary crossing and the depth to which service profiles are made available to the network.

An advantage to having service profiles available at the BSC level is the speed with which caller and callee authentication can be made. The obvious disadvantage is that, under aggressive location updating, the bandwidth required to move service profiles between BSCs could be prohibitive. An alternative is to eliminate BSC level partitions and store service profiles at a MAN level VLR. However, the signaling traffic load placed on the node where the database is located, although manageable, may lower the throughput of lower priority data traffic. These database arrangements are illustrated in Figures 4.4a and 4.4b.

A compromise, proposed in [7], is that BSC level database partitions list the subscribers currently within its coverage area, while their service profiles are stored at a VLR located above the BSCs in the MAN, as shown in Figure 4.4c. To locate a subscriber at the BSC level, a broadcast query, facilitated by the MAN architecture, can be made to all BSCs located below the subscriber's current VLR. Bandwidth is saved by not transporting service profiles between BSCs, and throughput for data traffic will be higher at the node where the VLR resides. This scheme, and the case where no BSC database partitions exist will be considered for their effects on signaling behaviour in Chapter 5.

From a signaling perspective, it is most efficient to locate the HLR and VLR, and, if required, an Equipment Identity Register (EIR) and Authentication Center (AC), such that they can be accessed through the gateway to the PSTN. This arrangement will lower the number of messages sent over the MAN during a call setup involving a subscriber connected directly to the fixed network.

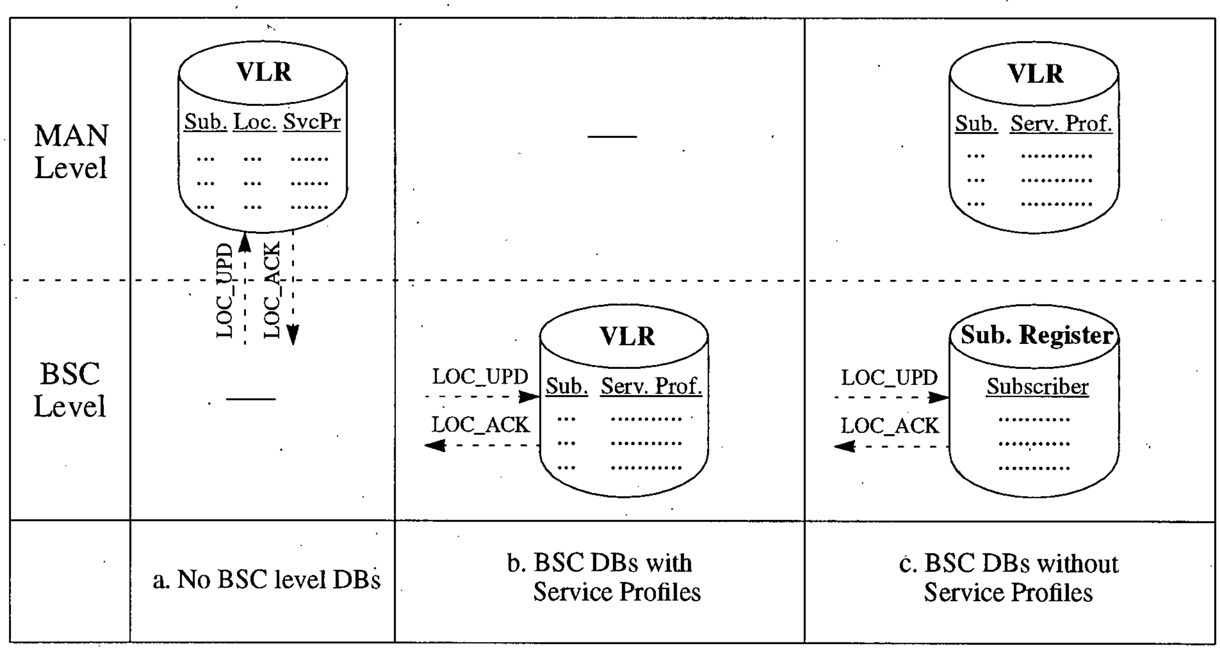

Figure 4.4. VLR configurations with intra-MAN location update message flows

4.2.3. Connectivity to the Fixed Network

As implied in earlier sections, PCNs must provide connectivity to the local fixed network which may be a Public Switched Telephone System (PSTN), an Integrated Services Digital Network (ISDN) or an ATM/B-ISDN. For example, in today's GSM networks, each Mobile Switching Center (MSC) that coordinates the activities of a group of BSs is connected to the fixed network. For the MAN-based PCN, this connectivity may be provided through a gateway that allows the PCN and the fixed network (herein referred to as the PSTN) to 'talk' to each other by incorporating user transport and signaling protocol conversion.

An OSI model for the router that provides layer 3 conversion for voice transport between the MAN and Time Division Multiplexed (TDM)-based DS3 trunks is shown in Figure 4.5. At layer 3, individual voice streams are found in their raw format, stripped of all frame and slot overhead. Banks of transcoders may be utilized here to provide conversion between different vocoder rates, that is, 64 kbps PCM and a LBE rate. In the future, trunks that provide the physical link for user transport may be ATM-based, in which case the transcoding operation can be moved out of the gateway and closer to the PSTN. A dynamic routing table is required to switch from a combination VCI/sub-frame offset to an outgoing circuit and vice-versa.

CCS-based SS7 can support call-connection and mobility management functions outside of the MAN. Separate links (e.g. 56 kbps, 1.544 Mbps) or associated signaling (e.g. ATM) may be used. A signaling protocol converter will be considered in Chapter 5.

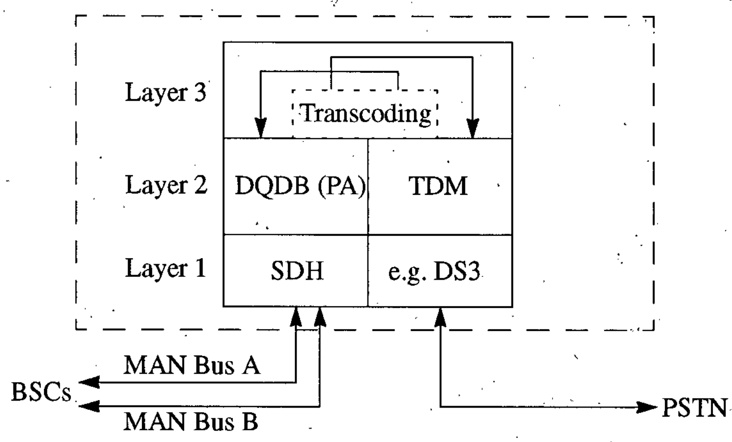

Figure 4.5. Router for handling of voice between MAN and PSTN

The issue of where on the MAN to locate the gateway is not trivial. Locating it at the HOB in a closed-bus MAN can increase the PA capacity of the MAN, as will be described. In standard DQDB operation two isochronous channels (PA) are established in opposing directions to allow full-duplex communications between two nodes. If one of the nodes is the HOB, a closed-bus configuration allows one established circuit to provide the same service since the packet that was created by the HOB will end up at the HOB after traversing either of the buses.

The other node can reuse the circuit after it has removed the voice payload destined for itself. To accomplish this removal, the node must have erasure capability on a sub-slot level. After copying and erasing the contents of the original payload, the node can then write its own payload and send the packet along the same bus to its next destination, the HOB. This operation is illustrated by the diagrams in Figure 4.6.

The improvement to the call capacity of the MAN will be significant considering the amount of calls that will likely involve a PSTN party. For example, if 75% of the calls involve only one mobile party, then the capacity of the MAN is increased by 60%. In general, a proportion, f, of mobile-to-fixed and fixed-to-mobile calls relates to a capacity increase of

$100\% \times f/(2-f)$. It is assumed that there is no BSC functionality at the HOB. All calls involving two mobiles, then, will still require two opposite-direction circuits.

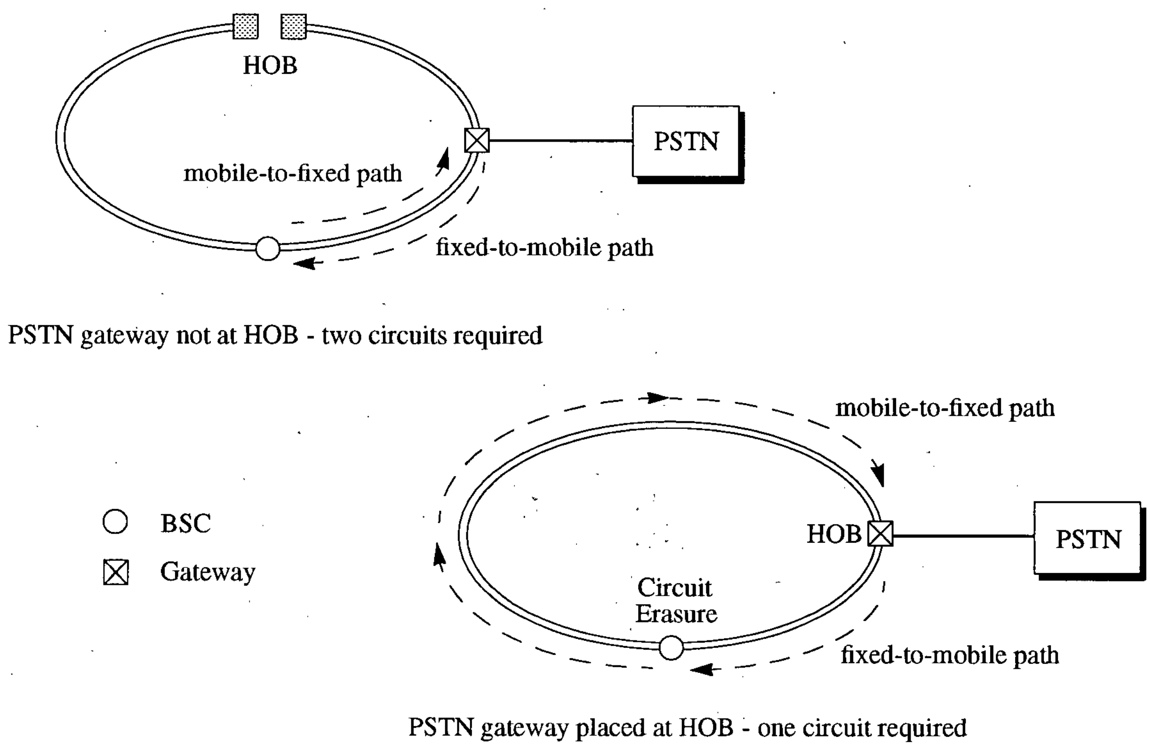

Figure 4.6. Voice circuits required for call involving a PSTN party

4.3. Erlang Capacity of the IEEE 802.6 MAN-based PCN

4.3.1. PCN Call Blocking Probability

A metric often used in telephony is the call blocking probability. It is the likelihood of not obtaining an end-to-end connection due to a lack of available resources somewhere in the network. This probability is the sum of individual blocking probabilities in each section of the network through which a conversation will pass. In a cellular network, it is often equated to the Grade of Service (GoS). Table 4.2 captures typical probabilities, P_B, for a mobile-to-PSTN connection in an AMPS cellular network.

Section of Network	Typical P_B
Mobile to Cell Site (radio)	0.020
Cell Site to Mobile Exchange (trunks)	0.005
Mobile Exchange Switch	0.001
Mobile Exchange to PSTN (trunks)	0.005
TOTAL	0.031

Table 4.2. Typical blocking probabilities in an AMPS cellular network

Blocking probabilities can be applied to the design of a MAN-based PCN. GoS goals for a PCN are likely higher than for current cellular systems. With the extensive use of microcells a blocking probability of 0.01 for the radio link is reasonable. Trunks are required to connect BSCs to individual BSs and to connect the MAN to the PSTN. The connection from BS to BSC should be capable of achieving reduced blocking since bandwidth is inexpensive. Due to the volume of traffic that must be handled, a similiar decrease will not be as easy for the trunks that provide PSTN access. The target blocking probability on the MAN is assumed to be equivalent to that for trunks between cell sites and the mobile exchange. Table 4.3 suggests potential design blocking probabilities.

Section of Network	Suggested P_B
Mobile to Cell Site (radio)	0.010
Cell Site to BSC (trunks)	0.001
MAN	0.005
MAN to PSTN (trunks)	0.005
TOTAL	0.021

Table 4.3. Suggested blocking probabilities in a MAN-based PCN

4.3.2. Determining the MAN Erlang Capacity for Two Session Classes

The blocking probability, P_B, for a trunk group that can be modeled as an M/M/m/m queue is easily found with the Erlang-B formula [31]

$$P_B = \frac{\rho_{MAN}{}^{m}/m!}{\sum_{n=0}^{m} \rho_{MAN}{}^{n}/n!} \tag{4.1}$$

where m is the maximum allowable number of two-way calls and ρ_{MAN} is the traffic load, in Erlangs[2], offered to the MAN portion of the PCN. For example, with 75% of the 311 Mbps MAN bandwidth reserved for voice transport (including overhead) and using the multiple call per cell transport scheme, a desired blocking probability of 0.005 allows a traffic load of approximately 10510 Erlangs, assuming that each call requires two isochronous circuits. Note that this figure represents the offered load to the MAN and not to the air interface.

The analysis becomes more complex when we take advantage of a PSTN interconnection

2. The Erlang is a dimensionless unit measuring traffic load, and is equal to the mean call arrival rate multiplied by the average call duration.

that is located at the HOB, as detailed in 4.2.3. Recall that there are now two session classes:

1. fixed-to-mobile and mobile-to-fixed calls that can utilize one circuit
2. mobile-to-mobile calls that require two circuits

Although the average call duration of the two classes will be considered identical, the two sessions types are not indistinguishable and the whole system can not be represented by the one dimensional Markov Chain implied by (4.1). This is because the second session type consumes twice as many resources than the first.

A two-dimensional Markov Chain must be constructed to determine the call blocking probability of the different session types. The value m (which is even) now refers to the total number of circuits available in the MAN, and not the maximum allowable number of two-way calls. Let the number of calls of the first and second session class occupying the network be denoted by n_1 and n_2, respectively. The current state of the network, defined by the number of active calls of each type, is denoted (n_1, n_2). The possible system states, shown in Figure 4.7, are constrained according to the condition $n_1 + 2n_2 \leq m$.

The shaded states represent potential blocking cases given that there is a call arrival. The probability of call blocking for the first session class is given by

$$P_{B_1} = \sum_{n_2 = 0}^{m/2} P((m - 2n_2), n_2) \tag{4.2}$$

and for the second session class by

$$P_{B_2} = \sum_{n_1 = 0}^{m} P\left(n_1, \left\lfloor \frac{(m - n_1)}{2} \right\rfloor\right), \tag{4.3}$$

where $\lfloor x \rfloor$ is the largest integer less than or equal to x and $P(x, y)$ is the probability that the network will be in state (x, y).

The steady state probabilities found in (4.2) are a subset of those in (4.3) and so the blocking probability for the second session class is greater than that for the first. This is intuitive as one circuit will be found more easily than two when the system is close to capacity.

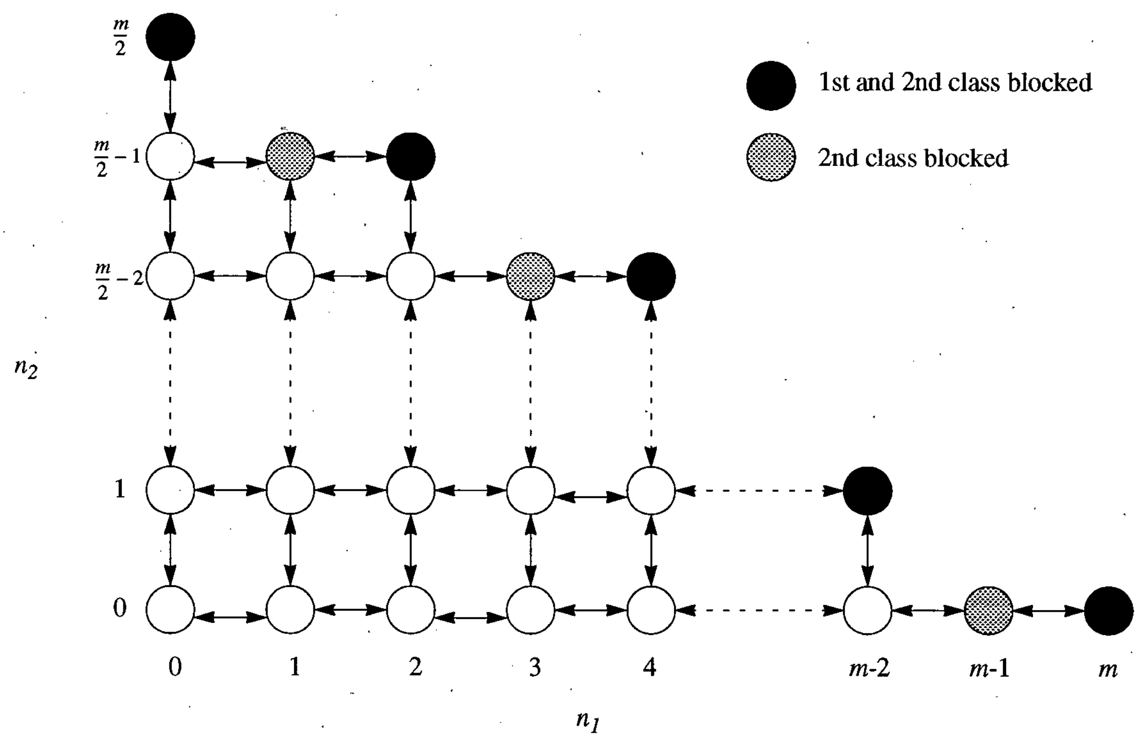

Figure 4.7. Markov chain for the two session class MAN

The Markov Chain in Figure 4.7 represents a truncation of a system consisting of two independent *M/M/∞* queues. The stationary probability distribution, $P(x_1, x_2, \ldots, x_K)$, of a system of K independent *M/M/∞* queues is given by [32]

$$P(n_1, n_2, \ldots, n_K) = \frac{\frac{\rho_1^{n_1}}{n_1!} \cdot \frac{\rho_2^{n_2}}{n_2!} \cdots \frac{\rho_K^{n_K}}{n_K!}}{G} \tag{4.4}$$

where G is a normalization constant,

$$G = \sum_{(n_1, n_2, \ldots, n_K) \in S} \frac{\rho_1^{n_1}}{n_1!} \cdot \frac{\rho_2^{n_2}}{n_2!} \cdots \frac{\rho_K^{n_K}}{n_K!}, \tag{4.5}$$

ρ_k is the Erlang load of the kth queue, and S is the set of states of the truncated Markov Chain.

Using this product form solution, the blocking probabilities of the respective service classes are

$$P_{B_1} = \frac{\displaystyle\sum_{n_2=0}^{m/2} \frac{\rho_1^{(m-2n_2)}}{(m-2n_2)!} \cdot \frac{\rho_2^{n_2}}{n_2!}}{\displaystyle\sum_{n_2=0}^{m/2} \sum_{n_1=0}^{(m-2n_2)} \frac{\rho_1^{n_1}}{n_1!} \cdot \frac{\rho_2^{n_2}}{n_2!}} \tag{4.6}$$

and

$$P_{B_2} = \frac{\displaystyle\sum_{n_1=0}^{m} \frac{\rho_1^{n_1}}{n_1!} \cdot \frac{\rho_2^{\lfloor (m-n_1)/2 \rfloor}}{\lfloor (m-n_1)/2 \rfloor !}}{\displaystyle\sum_{n_2=0}^{m/2} \sum_{n_1=0}^{(m-2n_2)} \frac{\rho_1^{n_1}}{n_1!} \cdot \frac{\rho_2^{n_2}}{n_2!}}, \tag{4.7}$$

where ρ_1 and ρ_2 are the Erlang loads of the respective session types offered to the MAN and $\rho_1 + \rho_2 = \rho_{MAN}$.

To evaluate (4.6) and (4.7) requires much computational power for large values of *m*. Additionally, the offered loads of the two sessions classes must be determined by trial and error since P_{B_1} and P_{B_2} are design parameters. P_{B_2} will be the constraining value as P_{B_1} is guaranteed to be lower. To simplify matters, the MAN portion of the PCN can be designed such that each session class has a maximum blocking probability of 0.005. This will ensure a cumulative blocking probability of less than 0.005.

4.3.3. 'Soft' Capacity Scheme for Handoffs

The Erlang capacity of the MAN has implications not just for call setup but for call handoffs as well. There are several instances in a MAN-based PCN where a call may be dropped during a handoff because of unavailable channels on the MAN:

- one of two mobiles involved in an intra-BSC call (no MAN voice circuits required) roams into a new BSC coverage area
- a mobile in conversation outside of a particular MAN roams into that MAN's coverage area
- a combination of the above requires a new channel on each of two different MANs

Call dropping represents a greater threat to the GoS for network operators than blocked calls. In current cellular networks, call dropping will occur either because of poor radio link quality or because of unavailable channels in a target cell. The use of microcells should help to lower the probability of the latter event occurring.

To avoid call dropping due to unavailable MAN channels, a flexible PA bandwidth allocation scheme should be implemented that will allow for temporary increases in the number of available voice channels. This 'soft' capacity scheme would allocate extra PA channels when the original PA capacity of m circuits is reached and handoffs require the establishment of new MAN channels. This implies a momentary reduction in the amount of available QA bandwidth. Signaling delays, however, should hardly be affected since data traffic will assume the lowest QA priority level. When the current PA consumption equals or exceeds m, all calls are blocked until there are enough call releases to bring the PA activity down below m. During this period handoffs are given preferential treatment over call setups to effectively nullify the likelihood of call dropping on the MAN, at the expense of reducing data traffic throughput temporarily. Figure 4.8 presents a flow diagram to illustrate the scheme.

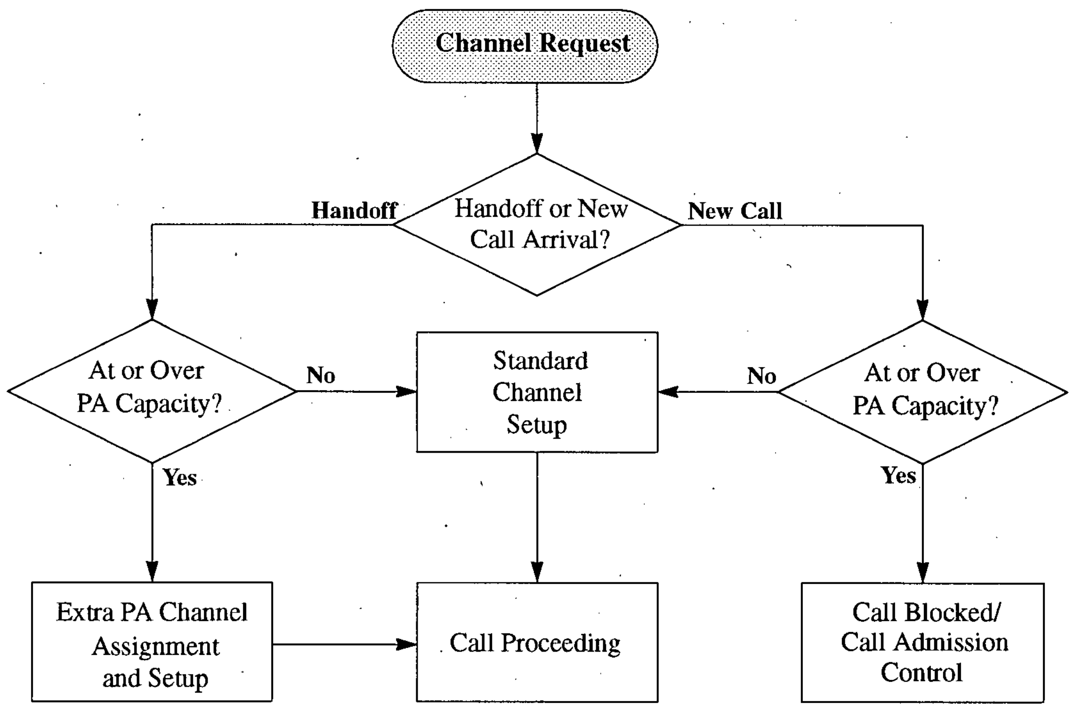

Figure 4.8. 'Soft' capacity scheme to avoid call dropping during inter-MAN handoffs

4.3.4. Implications of CDMA Soft Handoff on MAN Capacity

In a CDMA-based cellular network, handoffs are performed in which a mobile communicates with two different base stations simultaneously. By using a Rake receiver, the mobile can isolate both signals and select the best speech frame of the two radio paths based on speech quality indicators [32]. A frame selector is also required in the fixed portion of the cellular network to provide the same benefits for the reverse link. In this way a 'soft', high-quality handoff is ensured. The soft handoff concept can be extended to provide macrodiversity for mobiles in areas where reception from multiple base stations is possible.

In a MAN-based PCN utilizing a CDMA air interface, frame selection in the PSTN must be decentralized to the node level (including the gateway to the PSTN) to accommodate mobile-to-mobile connectivity. A benefit is that macrodiversity involving base stations of the same BSC (intra-BSC macrodiversity) will not require additional MAN capacity. However, several cases arise where additional MAN channels are needed over and above the number required for a non-CDMA conversation. For example, a mobile-to-fixed call in which the mobile communicates with two base stations under the control of different BSCs (inter-BSC macrodiversity) will require two channels rather than the one channel normally required under the PSTN/HOB collocation scheme. A mobile-to-mobile call in which both mobiles communicate with two different BSCs will require four channels. The former example is illustrated in Figure 4.9. The number of possibilities quickly escalates as the number of potential connections, or the degree of macrodiversity, increases. These factors must be detailed to enable accurate estimates of the MAN Erlang capacity and call blocking probabilities. Simulations reported in subsequent chapters assume a Time Division Multiple Access (TDMA) air interface with no macrodiversity.

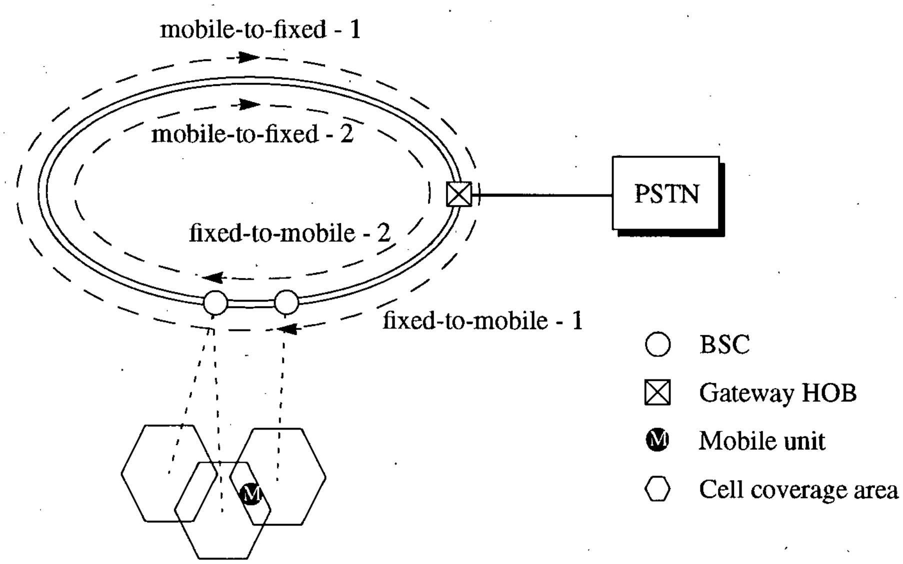

Figure 4.9. Mobile-to-fixed call with inter-BSC macrodiversity

Chapter 5. Analysis of a PCN Based on an IEEE 802.6 MAN

Using a network architecture based on a single IEEE 802.6 MAN, a PCN that provides wireless voice and data services to an urban population is investigated. Capacity and signaling improvements are sought, using ideas previously discussed.

5.1. Services, Architecture and Signaling

5.1.1. Services

A PCN based on a single IEEE 802.6 MAN can provide voice and wireless packet data transport to a city, or a portion thereof. It will connect to the PSTN as well as to public data networks (PDNs). Current terrestrial wireless data services, such as the offerings from Ardis and RAM Mobile Data, do not carry voice. In current cellular networks, data is transported on a circuit-switched basis, although the introduction of Cellular Digital Packet Data (CDPD) will change this to some extent. Wireless packet data traffic will likely see considerable growth, due in part to the ever-growing popularity of the Internet. The trend in PCNs is moving towards the transport of integrated traffic, and networks with highly flexible media access control can help to deliver network architecture solutions.

Voice traffic on the MAN portion of the PCN can be separated from signaling and data traffic by using PA media access. This separation implies some bandwidth planning on the part of the network designers. It is difficult to predict the relative future growth patterns of wireless voice and data usage. The degree to which voice will dominate the MAN bandwidth is highly debatable. It is instructive to choose a high PA-to-QA traffic ratio and proceed with the analysis. Predicting the impact of altering this ratio is straightforward. For example, increased voice traffic will increase signaling delays and lower the packet data throughput. Emphasis is placed on the analysis and improvements to performance, rather than on the initial conditions since the number of possibilities is endless.

5.1.2. Architecture

The PCN can provide coverage to an entire city area. Consistent with other analyses [19], a 50-node (including collocated HOBs) MAN network is considered. The physical spacing between successive BSCs has been enlarged to four slot lengths (about 2 km), giving a total effective separation (including a slot transmission delay) of five slots.

The MAN network elements such as the ST, BWM and databases are arranged as discussed in 4.2 for capacity gains and signaling efficiency on the MAN. The router required for voice transport between the MAN and the PSTN was discussed in 4.2.3. The type of trunks employed for user transport between the gateway and the PSTN may be TDM or ATM-based DS3 or SDH at 155.52 Mbps.

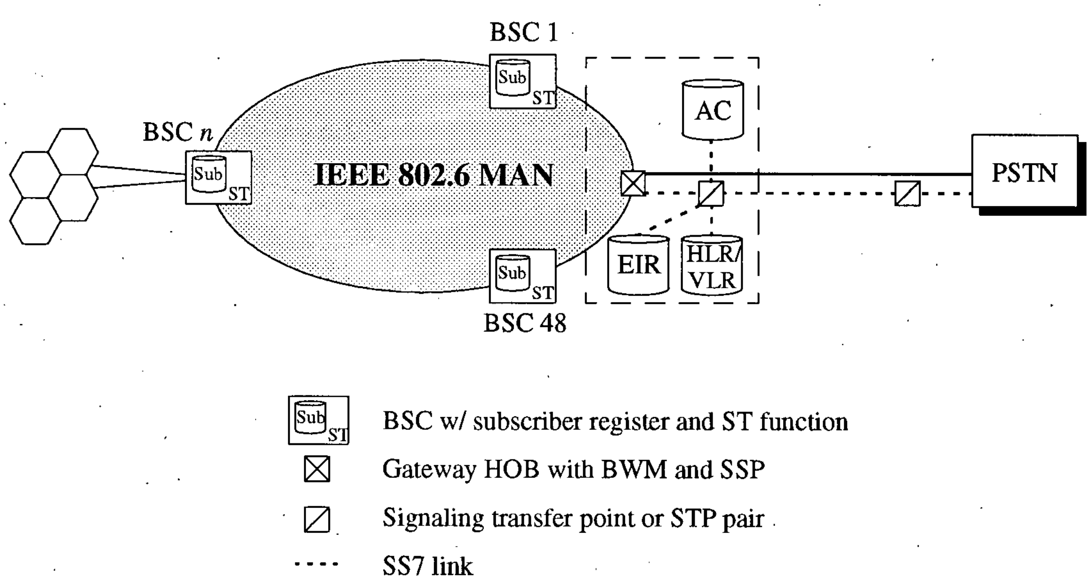

Figure 5.1. PCN architecture based on a single IEEE 802.6 MAN

The architecture, shown in Figure 5.1, includes IN elements such as a Service Switching Point (SSP), Signal Transfer Points (STPs), and Service Control Points (SCPs). These will be discussed further in the context of PCN signaling.

5.1.3. Signaling

It is assumed that the PCN communicates with the PSTN using Common Channel Signaling based on Signaling System No. 7. The connectionless nature of the protocol facilitates the real-time mobility applications required in a PCN [23]. Any call processing that requires the establishment of a PCN-PSTN trunk connection will use the ISUP (ISDN User Part) portion of the SS7 protocol to provide basic bearer services. The gateway HOB incorporates a Service Switching Point (SSP) module that sends ISUP messages to the PSTN via the local Signal Transfer Point (STP).

Signaling between the various PCN databases (Service Control Points) and the gateway HOB function will also employ SS7, with messages being routed through the local STP. The MAP (Mobile Application Part) of the SS7 protocol will be called upon to support functions such as location updates, service profile retrieval and authentication. For efficiency, proprietary signaling, instead of SS7, could be used for location updates at the BSC level since these databases are not adjacent to the PSTN.

There is some disparity between the signaling that takes place on the MAN and the SS7 protocol. This must be accommodated through a signaling protocol converter at the gateway that maps one application level to the other. This is shown in Figure 5.2. MAN signaling messages must be converted above the application layer to one of the MAP or ISUP aspects of SS7, depending on the functionality required.

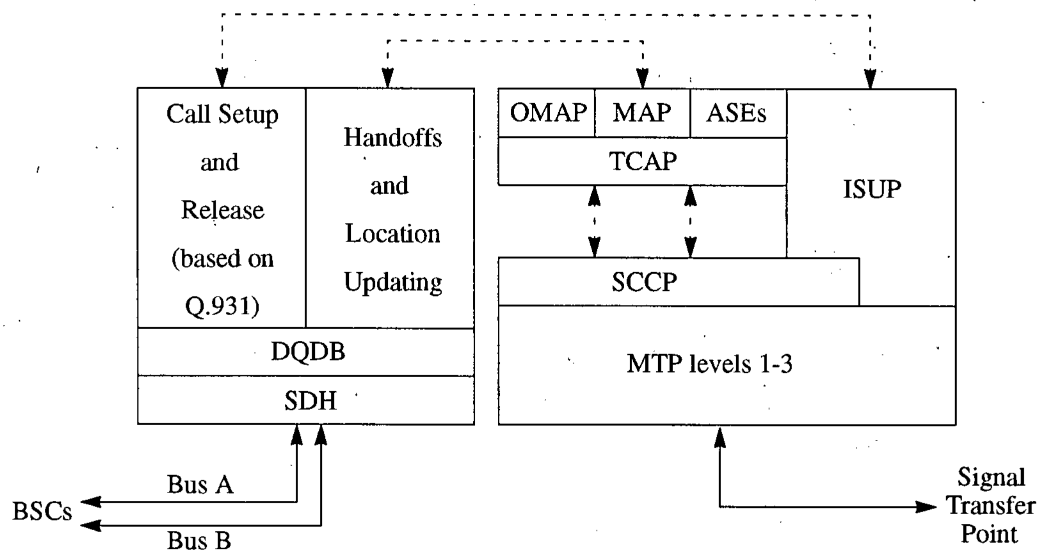

Figure 5.2. Signaling protocol converter at MAN gateway

Signaling on the MAN that is used to support PCS in the simulations that follow is based on methods proposed in [7, 9]. Call setup procedures are based on the ISDN Q.931 protocol. Signaling message abbreviations can be found in Appendix B.

The distributed database scheme with BSC level databases tracking subscriber locations has implications for location update and call-setup signaling message flows. Intra-MAN location update messages will move between BSCs only and will not place a heavy burden on any single node, unlike a centralized location update process that will involve message flows to and from the VLR. As a result of the proposed database scheme, the signaling required for intra-MAN location updates is kept to a minimum (since service profiles are not transferred between BSCs) and is highly decentralized (if authentication is not required for intra-MAN boundary crossings).

An intra-MAN call setup involving two mobiles is illustrated in a modification of the signaling diagram borrowed from [9] and shown in Figure 5.3. Bus access and processing delays are not implied. Arrows indicate signaling messages sent on the MAN only, recalling the signaling efficiencies gained by using: i) a distributed Signaling Termination element; and ii) a BWM and VCI server located at the HOB in a closed-bus MAN. Call setup involving a fixed party differs only slightly on the MAN portion of the PCN and is easily accommodated.

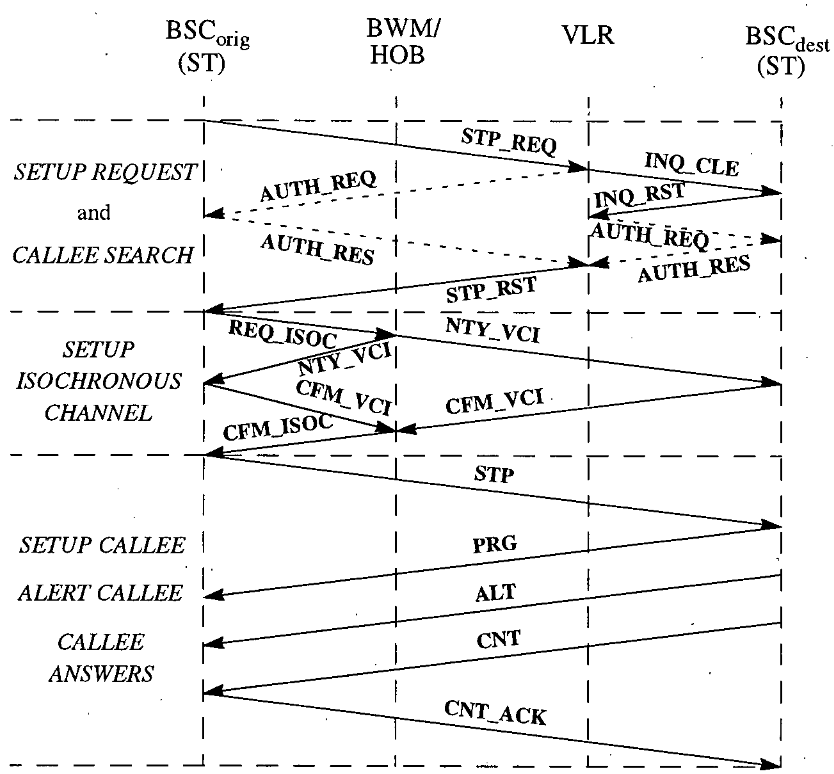

Figure 5.3. MAN signaling for a modified intra-MAN call setup between two mobiles

The originating BSC (which has ST functionality) requests service from the VLR for call setup using the STP_REQ message. The database must locate the service profile for the calling party to determine if service can be provided. Depending on the extent of network security, the database may initiate an authentication request [8] from the calling party. This step may begin in parallel with the database partition search for the callee (INQ_CLE) to reduce the overall setup time. As well, authentication of the callee may be required in which case a similiar exchange takes place between the VLR and the destination BSC. Authentication message flows are indicated by dashed lines and are omitted from simulations. If required, their inclusion will lengthen the initial phases of call setup and only slightly increase the amount of bandwidth consumed by signaling.

5.2. Erlang and Subscriber Capacity

5.2.1. Erlang Capacity

In [7], the Erlang load carried per MAN in a PCN was found by applying the Erlang-B formula in (4.1). This quantity was then used to determine the city block coverage based on the offered load per block and a given proportion of intra-MAN traffic. In [33], the number of cells per MAN is calculated, and other capacity measures are derivable from this.

For the simulations that follow, the Erlangs carried per MAN will be determined based on bandwidth allocation and proportions of various call types. This approach attempts to be non-specific with regards to the cell size or offered load per cell, since these quantities may vary dramatically between and within cellular networks.

PA traffic for voice transport is assigned to 75% of the available MAN bandwidth (including slot and SDH frame overhead), with the remaining 25% to be used for the transport of multiplexed signaling and wireless packet data traffic in QA slots. This generous allocation in favour of voice is based on the belief that voice will dominate PCS traffic. Using the two circuit per cell transport scheme and a single bus data rate of 155.52 Mbps (149.76 without SDH frame overhead), the number of available one-way circuits is 21120.

With a fixed network gateway located at the HOB in closed-bus MAN, the various percentages of call types will have a great impact on capacity. Three two-way call types are considered: fixed-to-mobile, mobile-to-mobile and mobile-to-fixed, with proportions *fm*, *mm*, and *mf*, respectively. Currently, fixed origin calls are a small percentage (less than 15%) of the total calls in a cellular network. There is reason to believe that the proportion of fixed origin calls will increase with the introduction of PCNs, as personal communications services become commonplace. Destinations of mobile origin calls are heavily dominated by fixed network parties, as opposed to mobile terminations. This will change as the mobile subscriber population grows, increasing the percentage of mobile-to-mobile calls.

An allocation of call types for the purpose of capacity planning is proposed in Table 5.1. Values are based on two factors: 1) fixed-to-mobile calls represent 25% of the total calls on the network; and 2) 25% of mobile originated calls have mobile destinations. On average, then, 81.25% of all calls on the MAN consume only one circuit, leading to an expected capacity gain of approximately 68%.

Call Type	Fraction of Total
Fixed-to-mobile	$fm = 0.25$
Mobile-to-mobile	$mm = 0.1875$
Mobile-to-fixed	$mf = 0.5625$

Table 5.1. Distribution of PCN calls by type

Given a desired probability of call blocking of 0.005 for the second session type (involving two circuits), and the ratio of the traffic loads, ρ_1 and ρ_2, the maximum sustainable offered load, ρ_{MAN}, can be determined. Unfortunately, the amount of computation required to evaluate (4.6) and (4.7) is prohibitive for very large values of *m*. A crude approximation is made by focusing on that area in the two-dimensional Markov Chain that surrounds the location of the expected system state. Even further approximation is necessary and so points in the area of interest are sampled in a grid-like fashion. The error of this method is difficult to predict without actual values. The problem requires more rigorous treatment than is possible here. More details on the method of approximation are found in Appendix D.

Based on the given proportions of call types, an offered load of $\rho_{MAN} = 17680$ Erlangs has been determined numerically to give the following blocking probabilities: $P_{B_1} \cong 0.0025$ and $P_{B_2} \cong 0.0050$. As expected, the enhanced capacity represents an increase of 68% over the original figure of 10510 Erlangs found in 4.3.2. Additionally, simulations of the Markov Chain in Figure 4.7 confirm these quantities.

5.2.2. Subscriber Capacity

The offered load to the PCN that is mobile originated, ρ_{mob}, will give an indication of the maximum sustainable number of mobile subscribers. This is given by

$$\rho_{mob} = \left[\frac{mm}{\left(1 - P_{B_{radio}}\right)^2} + \frac{mf}{\left(1 - P_{B_{radio}}\right)}\right]\rho_{MAN} \tag{5.1}$$

where $P_{B_{radio}}$ is the blocking probability on the air interface and the exponent of two in the first term accounts for the probability of blocking in two different cells. Also,

$$\begin{aligned} \rho_{mob} &= \rho N \\ &= \lambda X N \end{aligned} \tag{5.2}$$

where ρ is the Erlangs generated per mobile subscriber, N is the maximum sustainable number of subscribers to the PCN, λ is the call generation rate per mobile subscriber during the busy hour, and X is the mean call duration. From (5.1) and (5.2),

$$N = \left[\frac{mm}{\left(1-P_{B_{radio}}\right)^2} + \frac{mf}{\left(1-P_{B_{radio}}\right)}\right]\frac{\rho_{MAN}}{\lambda X}. \tag{5.3}$$

The busy hour refers to that hour in the week when mobile users collectively generate the most voice traffic, normally late on Friday afternoons. For the calculations and simulations that follow, the values $\lambda = 0.2$ arrivals/hour and $X = 90$ seconds are chosen to represent realistic busy hour statistics. Using this data, the maximum number of subscribers to the MAN-based PCN is $N \cong 266{,}000$. If a market penetration rate of 20% is assumed, the single MAN-based PCN may effectively serve a metropolitan area with a population of approximately 1.3 million people.

5.2.3. Effect of Subscriber Capacity on Location Updates

The number of subscribers to a PCN greatly affects the amount of signaling generated to support mobility management. In particular, it will be most significant for location update traffic. The subscriber rate of location updates, L, in the busy hour can be given by [33]:

$$L = (1 - P_{active})\lambda_{LA} \tag{5.4}$$

where P_{active} is the probability that the mobile terminal is in conversation, and λ_{LA} is the rate of location area boundary crossings for a mobile. Additionally,

$$P_{active} = \left[\left(1-P_{B_{radio}}\right)\left(1-P_{B_1}\right)\left(\frac{mf}{mf+mm}\right) + \left(1-P_{B_{radio}}\right)^2\left(1-P_{B_2}\right)\left(\frac{mm}{mf+mm}\right)\right]\rho$$
$$+\left[\left(1-P_{B_1}\right)fm + \left(1-P_{B_2}\right)mm\right]\frac{\rho_{MAN}}{N} \tag{5.5}$$

where the first and second terms represent activity from being the caller and callee, respectively. Recall that the Erlangs per mobile subscriber, ρ, are subject to blocking probabilities on the radio interface, while ρ_{MAN} is not. Finally, the global rate of location updates is given by

$$L_{global} = L \times N. \tag{5.6}$$

5.3. Improvements in PCN Signaling Over the IEEE 802.6 MAN

Results from Chapter 3 indicate that the Preemptive Priority Mechanism (PPM) should be implemented to achieve proper behaviour for multi-priority QA access. Two sample loading scenarios have been investigated; however, these traffic distributions may not necessarily be representative of the signaling load in a PCN. The only way to truly test the PPM's effectiveness for a MAN-based PCN is to integrate the signaling processes required to support PCS voice services with previous simulations, which exclude PCN signaling traffic considerations.

Additionally, comparisons are made between the effects on signaling using two possible intra-MAN location updating methods based on different MAN database schemes discussed in section 4.2.2:

- Centralized location updates - the MAN-level VLR is updated with the most recent location area of mobiles under its coverage area. There are no BSC-level databases.
- Decentralized location updates - BSC-level databases list subscribers in their location area, but service profiles are kept at the MAN-level VLR.

The latter scheme is desirable since congestion can be reduced significantly at the node which enables VLR access.

Handoff signaling assumes a higher priority level of QA access than signaling for call processing and location updates. In a subsequent section, this priority assignment is compared with a scheme where all signaling accesses the MAN at one priority level only. The loading of packet data traffic at QA-0 is assumed to be persistent at all nodes, such that a QA-0 slot is always ready for transmission after one has left. Another QA-0 loading arrangement, in which the sum of the loads at all BSCs equals the load at the fixed network gateway, is considered later.

5.3.1. Setup

The following assumptions are consistent throughout the simulations that follow.

- The physical distance between nodes is four slots - the effective distance is five slots.
- The total dual bus MAN bandwidth is 311.04 Mbps, including SDH overhead
- The ratio of PA bandwidth to QA bandwidth (including overhead) is 3:1.
- Simulations are run for a period of 1 million QA slots, equivalent to approximately 10.9 seconds in real network time, given the PA-to-QA ratio. A generous transient period is allowed for, after which statistics are collected.
- The distribution of call types is given by Table 5.1.
- 20 msec MAN frames are used and the speech coding rate on the air interface is less than or equal to 9.6 kbps.
- There are two simplex voice circuits per PA cell.
- Packet data have exponentially distributed interarrival times and use QA-0 slots.
- All signaling messages use one QA slot, except for LOC_UPD and SPY_INFO (handoffs) which use two, and for the STP_REQ and STP messages which use five slots.
- All location updates, handoffs, and mobile-to-mobile call setups are intra-MAN and inter-BSC.
- The occurrence of signaling events are modeled as Poisson processes with exponentially distributed interarrival times.
- The mean call duration is 90 seconds. The subscriber call attempt rate is 2 calls/hour.
- Aggressive location update and handoff rates of 1 and 2 events/minute/subscriber,

respectively, are assumed. Each mobile subscriber is assumed to be in a position where either a handoff or a location update may occur, but not both.

- A constant processing delay of 100 μsec occurs between the receipt of a signaling message and the subsequent transmission of a new one in response. This should improve the variance of delay statistics over the case where delay is exponentially distributed without sacrificing much of the modeling accuracy.
- No signaling for wireless packet data services at QA-1 or QA-2 is assumed.

The parameters ρ_{MAN}, P_{B_1}, and P_{B_2}, along with the distribution of call types are supplied to the simulation at run time to create an 'active calls' variable that tracks the PA bandwidth usage based on the occurrence of call setups and call arrivals. The initial value of this variable is given by the expected number of calls at any given time:

$$E\{\text{calls}\} = \left[\left(1-P_{B_1}\right)(mf+fm) + \left(1-P_{B_2}\right)mm\right]\rho_{MAN}. \tag{5.7}$$

Some definitions of the various signaling delays are in order. These are specific to the MAN-based PCN and are useful as a benchmark for potential improvements.

- *call setup delay* - time elapsed from the receipt of the STP_RST message to the receipt of the PRG message at the originating node, including processing delays. This period does not include database searches and potential authentication requirements.
- *call clear delay* - time elapsed from sending the REL_ISOC message until receipt of its acknowledgment at the initiating node
- *location update delay* - time elapsed from sending the LOC_UPD message until receipt of its acknowledgment at the initiating node
- *handoff delay* - time elapsed from sending the HO_REQ message at the target node

until receipt of the CFM_HO message at the previous node

5.3.2. Simulation Results

The positive effect of an enabled PPM ($d = 2$) is readily seen in the *nominal* reduction in average signaling delays. Table 5.2 details these results.

The *real* improvement due to the PPM is found by removing the processing delay of 100 μsec for messages, as well as the transmission and propagation delay components from the above statistics. The average delay figures are then due only to message segments waiting for media access. In general, the nominal improvement will depend on two factors:

1. The size of the cumulative processing delay experienced in the signaling exchange.
2. The size of expected propagation delays.

An increase in either of these two quantities will diminish the noticeable gains of the PPM. Table 5.3 illustrates the real improvement that is gained by using the PPM. Average access delays, measured in slot lengths, indicate higher real improvements. For example, when decentralized location updating is used, the PPM improves the delay of QA-2 segments by 26%, a significant difference from the nominal improvements to handoffs seen in Table 5.2.

Signaling Exchange	Signaling Exchange Delay (msec)					
	Centralized loc. updates			Decent. loc. updates		
	$d = \infty$	$d = 2$	**%dec**	$d = \infty$	$d = 2$	**%dec**
Call Setup (no fixed-to-mob.)	3.07	1.97	**35.8**	2.43	2.07	**14.8**
Call Clear (mobile initiated)	0.76	0.51	**32.9**	0.60	0.55	**8.3**
Location Update	1.02	0.57	**44.1**	0.64	0.62	**3.1**
Handoff	1.79	1.56	**12.8**	1.67	1.60	**4.2**

Table 5.2. Nominal improvements in MAN signaling delays

QA Priority Level	Bus Access Delay (MAN slots)					
	Centralized loc. updates			Decent. loc. updates		
	$d = \infty$	$d = 2$	**%dec**	$d = \infty$	$d = 2$	**%dec**
QA-1 (BSCs only)	164.66	31.48	**80.9**	28.38	17.34	**38.9**
QA-2 (handoff)	24.04	8.04	**66.5**	15.30	11.22	**26.7**

Table 5.3. Real improvements in MAN segment delays at BSC nodes

With the PPM disabled ($d = \infty$), the decentralized location update scheme lowers signaling delays compared with the centralized scheme. This result is due to congestion relief at the HOB, since the signaling burden there is reduced by using intra-MAN location updates that do not require centralized control. With a lighter offered load, the bus access advantage at the HOB will be less pronounced with a heavy traffic load. This is made apparent through the bus access delays for QA-1 segments at the BSCs.

When the PPM is activated, a decentralized location update scheme appears to increase signaling delays very slightly. QA-1 segment delay figures indicate a decrease in the average bus access delay, although this decrease is distinctly less than for the case where the PPM is disabled. A factor that increases the location update signaling delays is greater bus propagation delay. Larger distances must be traversed in decentralized location update signaling exchanges, since the HOB, which is no more than half of the bus length away from any node, is no longer used. Instead, some messages may now have to traverse almost the entire bus length to reach its destination. This small degradation, which will diminish with decreasing inter-BSC delays, must be weighed against the benefits of decentralized location updates, one of which is the reduced traffic bottleneck at the HOB.

It is interesting to examine the signaling delays for the given scenarios on a BSC-by-BSC basis. A surprising result is that delays for decentralized signaling exchanges (i.e. not involving the HOB) depend inversely on the distance from the HOB to the involved BSC. This is the situation for intra-MAN location updates under the decentralized location update scheme, and for all intra-MAN handoffs as shown in Figures 5.4 and 5.5. At first, this may not seem to be appropriate, as the DQDB protocol is known to favour nodes nearer to the HOB. Upon consideration of the propagation delays, however, the results can be explained. The expected distance between a given BSC and any other becomes greater as it moves closer towards the HOB. Depending on the distance between successive nodes and the total length of the bus, the effect of propagation may manifest itself significantly in higher delays for decentralized signaling exchanges at nodes nearer the HOB, thereby diminishing any access advantage that these nodes have over nodes located further in on the buses.

Delay values for centralized signaling exchanges, such those for call setup shown in Figure 5.6, show the expected nodal favouritism, due in part to the propagation advantage made possible since the HOB is accessible by both buses from any BSC.

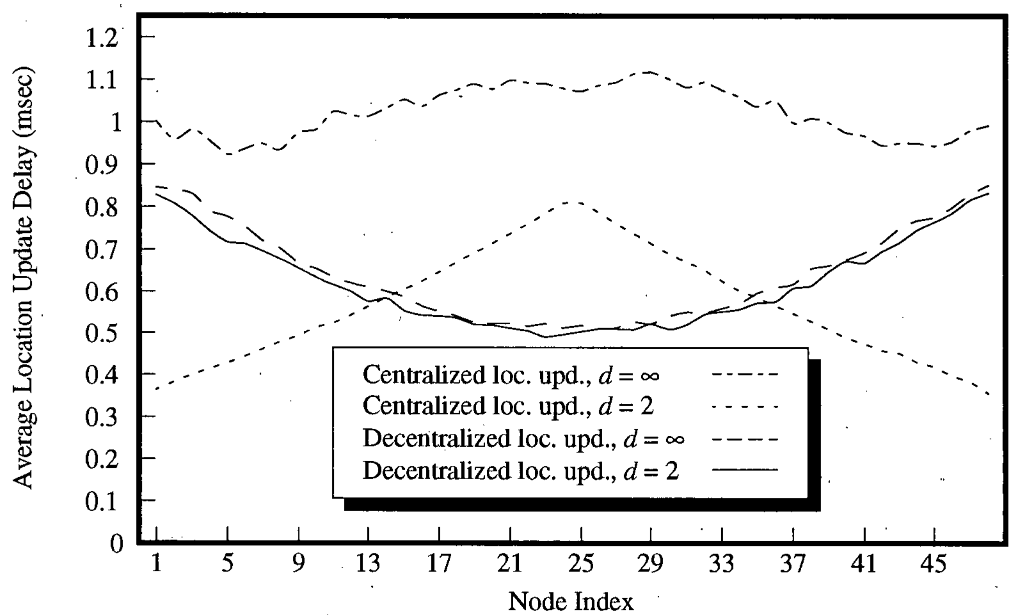

Figure 5.4. BSC-by-BSC location update delays

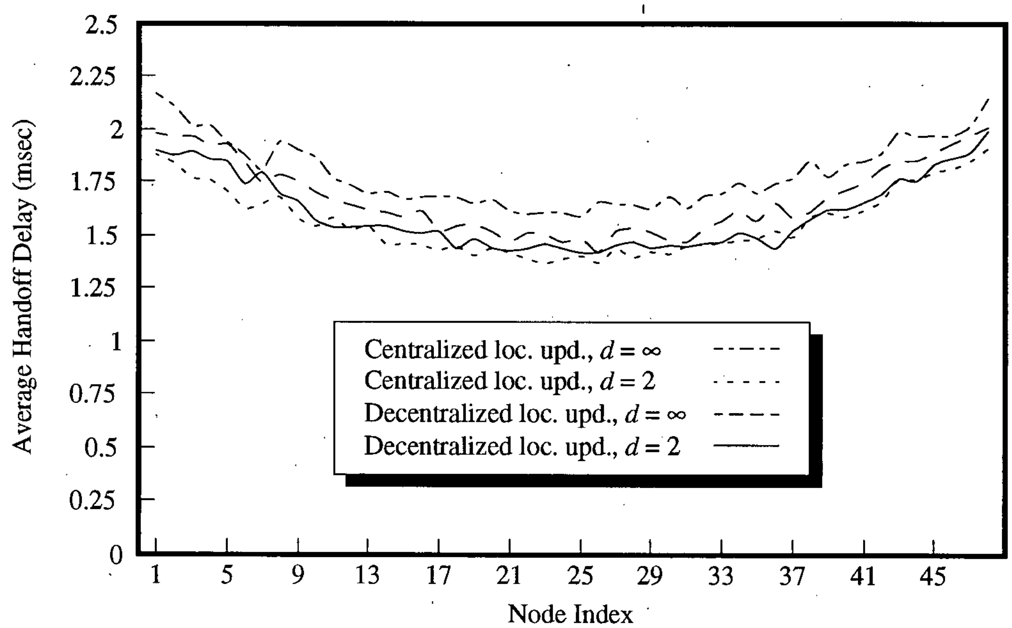

Figure 5.5. BSC-by-BSC handoff delays

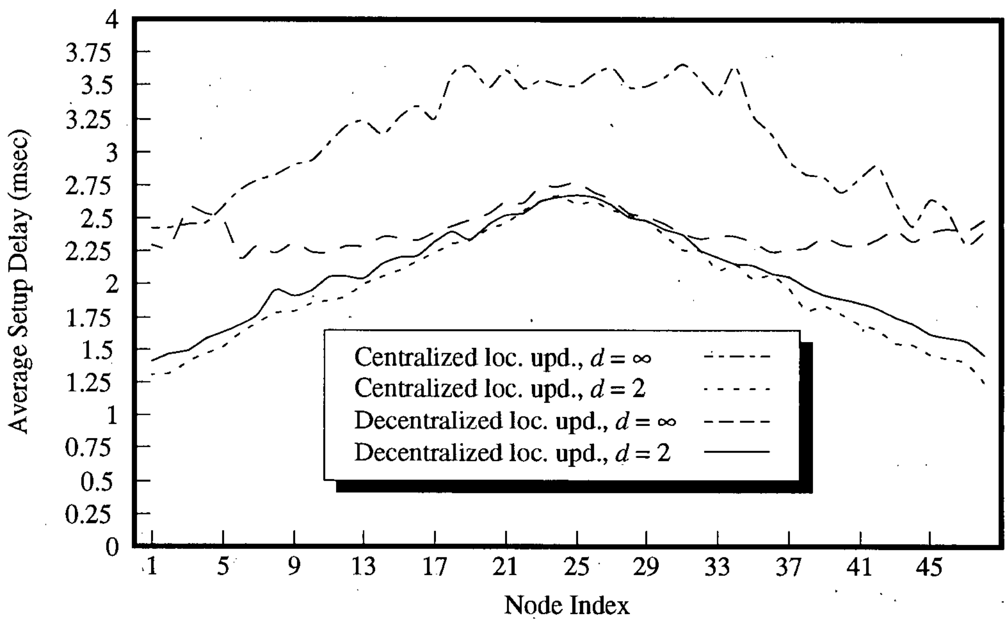

Figure 5.6. BSC-by-BSC call setup delays

To determine the effect of propagation delay on signaling delays, the effective inter-BSC delay has been reduced from five slots to one slot for the case where the PPM is enabled and decentralized location updating is employed. The BSC-by-BSC signaling delays are shown beside values from previous plots in Figure 5.7. Indeed, the lower propagation delay not only reduces signaling delays, but also reduces the spread of average delays among BSCs.

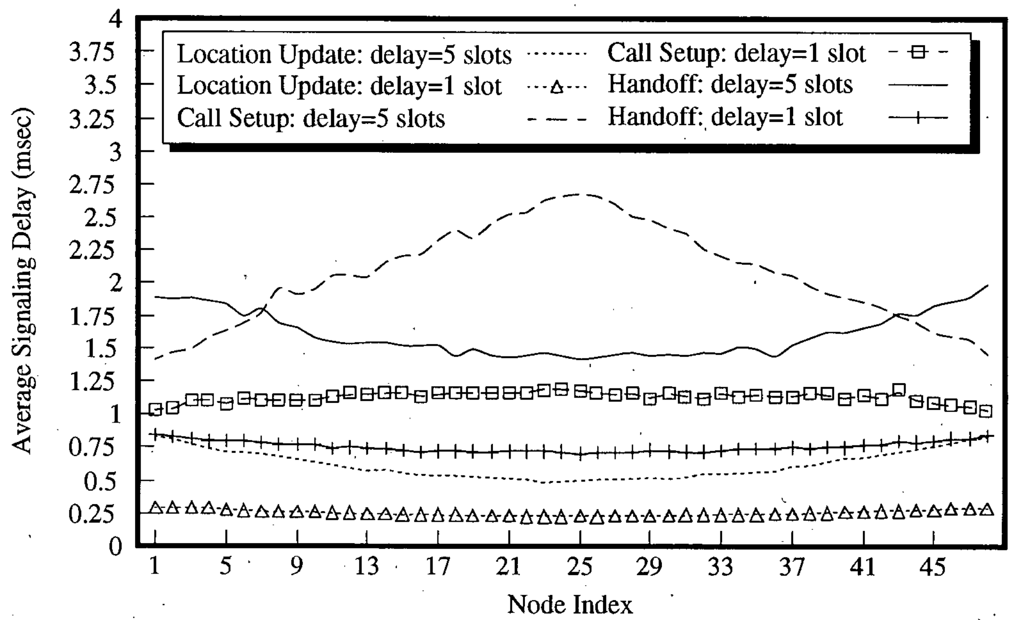

Figure 5.7. Effect of inter-BSC propagation on signaling delays

5.3.3. Comparison Using Signaling at One Priority

The assignment of signaling traffic to QA priorities in the previous section places greater importance on handoff traffic than on all other signaling. This assignment may be preferred over the case where all signaling is assigned to one priority level, but verification through simulation is required. As expected, the handoff delay is improved, although only very slightly, according to Table 5.4. The exception is the case with centralized location updating and a deactivated PPM. These improvements will be more substantial as the amount of available QA bandwidth decreases, since preemption will occur more often. Directing handoffs to QA-1 access has increased the average access delay for all signaling segments, producing a slight increase in call setup delays, also.

Signaling Exchange	Signaling Exchange Delay (msec)							
	HO → QA-1				HO → QA-2			
	Cent. loc. upd.		Decent. loc. upd.		Cent. loc. upd.		Decent. loc. upd.	
	$d = \infty$	$d = 2$	$d = \infty$	$d = 2$	$d = \infty$	$d = 2$	$d = \infty$	$d = 2$
Call Setup (excl. *fm*)	2.98	1.98	2.42	2.05	3.07	1.97	2.43	2.07
Handoff	2.00	1.58	1.69	1.62	1.79	1.56	1.67	1.60

Table 5.4. Signaling delays for different handoff priority levels

5.3.4. Packet Data Throughput

To best demonstrate the potential benefits of the PPM, a heavy load of QA-0 (packet data) traffic was introduced. This was similiar to Chapter 3, where the offered load of lower priority traffic pushed the cumulative load past bus capacity in order to show the effectiveness of the PPM. Now that benefits for PCN signaling have been shown, it is useful to examine the throughput of QA-0 traffic from the previous simulations. Figure 5.8 shows that the throughput of nodes is greatly affected by both the use of decentralized location updating and the PPM. With centralized location updating and the PPM disabled, QA-0 throughput is higher as one moves closer to the HOB. This is consistent with DQDB behaviour. The exceptional node is the HOB, whose QA-0 throughput is severely limited by the signaling traffic at higher priorities.

The throughput curve changes dramatically as decentralized location updating and the PPM are introduced. BSCs located closer to the HOB, no longer seem to have the same access advantage and nodes closer to the middle of the closed-bus have achieved higher, nearly equivalent throughputs. Although not apparent from Figure 5.8, the HOB throughput has improved with decentralized location updating. The media access for low priority data at the HOB is constrained by signaling traffic due to call setup and clearing, and by the heavy packet data loading at other nodes.

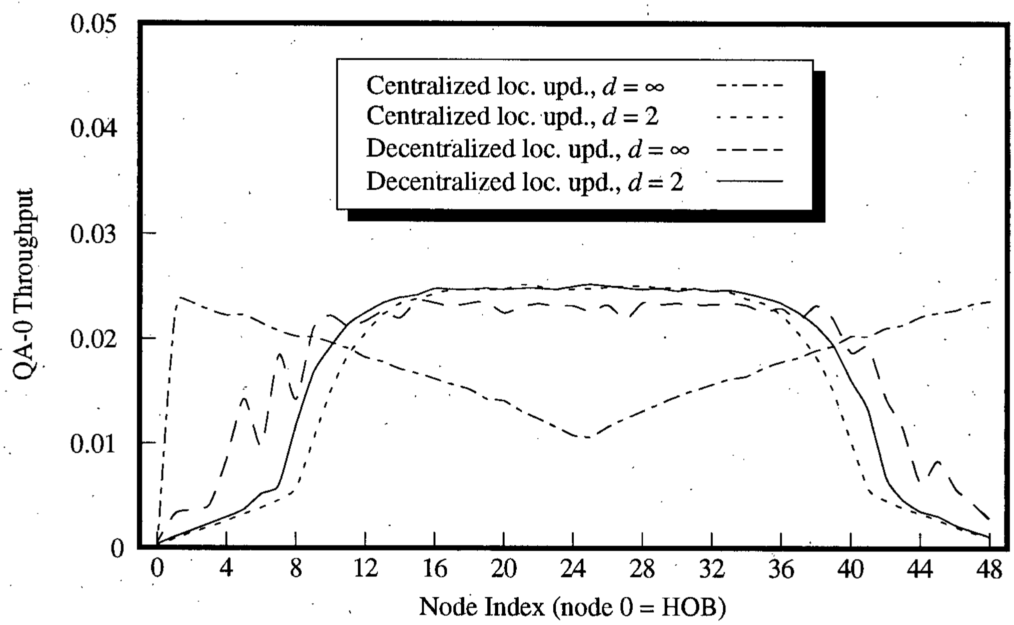

Figure 5.8. QA-0 throughput of BSCs under heavy loading

The high throughput potential of the MAN under heavy traffic loading is evident from the results of the case where decentralized locating updating is used and the PPM is disabled. The QA throughput averages 98.9% of capacity for the simulation duration. When the PPM is enabled, the throughput of lower priority segments is sacrificed, and the total throughput average is 93.4%.

The previous QA-0 loading scenarios may be unrealistic in that the load at the BSCs is assumed to be the same as that at the HOB. If a gateway to a public data network is located at the HOB, the loading will likely be asymmetrical. To model this approximately, the total QA-0 load offered to the MAN is divided such that the sum of QA-0 loads at all BSCs equals the offered load at the HOB. The total QA-0 load on the MAN is chosen such that it is slightly less than the QA capacity after signaling is considered. In this scenario, the load at the BSCs is manageable and delay characteristics are typical for DQDB, as shown in Figure 5.9. Interestingly, access delays for QA-0 segments are improved with the PPM enabled.

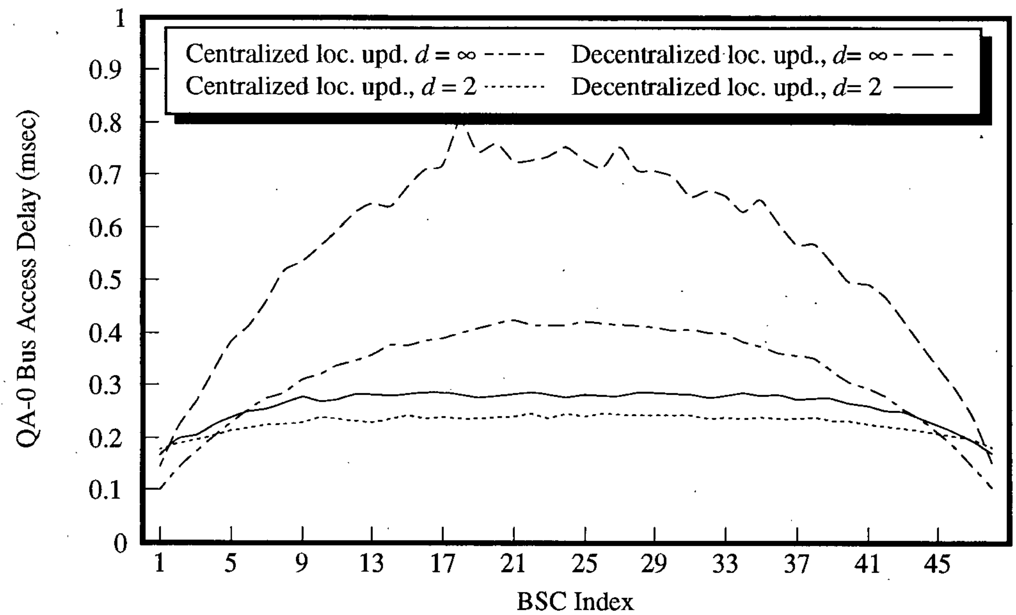

Figure 5.9. QA-0 segment delays at BSCs using asymmetrical traffic loading

HOB throughputs for the various schemes are shown in Table 5.5. In the cases where the PPM is disabled, the normalized offered load at the HOB, $\lambda_{HOB} = 0.44$, is carried successfully. Decentralized location updates provide the best HOB throughput but, as before, use of the PPM lowers throughput. The trade-off between lower signaling delays and higher packet data throughput must be weighed by the network designer. To help alleviate the traffic congestion at the HOB, Public Data Network gateways (there could be more than one per MAN) may be located at nodes other than the HOB. To increase the potential QA capacity, erasure nodes may also be considered.

Scenario	Centralized Location Updates		Decentralized Location Updates	
	$d = \infty$	$d = 2$	$d = \infty$	$d = 2$
QA-0 Throughput	.44	.31	.44	.38

Table 5.5. Packet data throughput at the Head of Bus - $\lambda_{HOB} = 0.44$

Chapter 6. A Multiple-MAN PCN and Migration Towards ATM

This chapter extends the architecture for a PCN based on a single IEEE 802.6 MAN to that of a PCN based on multiple IEEE 802.6 MANs in the context of the Intelligent Network (IN). One of the salient features of the IEEE 802.6 MAN standard is the fifty-three byte slot structure that reflects the ATM cell layout. This is not mere coincidence, as the standard is expected to be an evolutionary step towards broadband ATM networks [15]. As such, it is useful to consider the evolution of a MAN-based PCN towards ATM/B-ISDN.

6.1. PSTN Interconnection to Multiple-MAN PCNs

6.1.1. The Backbone Interconnection Scheme

Work has been performed in the area of QA voice access for multiple interconnected IEEE 802.6 MANs [34]. A backbone MAN provides connection between non-adjacent access MANs, as well as redundant connection between adjacent ones. Voice packet delays and throughputs have been tabulated for various levels of intra-MAN traffic.

In keeping with the spirit of the previously presented PCN architecture based on a single MAN, the multiple-MAN interconnection scheme is considered with an interconnection to the PSTN. The six-MAN network with a PSTN gateway located on the backbone MAN is illustrated in Figure 6.1.

Considering the proportion of calls involving a fixed party, it quickly becomes apparent that the traffic carried per access MAN is limited by the capacity of the backbone MAN, which is assumed to be the same as that for the access MANs. This is true regardless of the bus access mode (QA or PA). Higher capacity backbones may solve the problem, but bus speeds past 155.52

Mbps are not part of the ISO IEEE 802.6 standard [5]. For the interconnection scheme shown, the proportion of calls in an access MAN involving the PSTN would be limited to less than 17%. This is not consistent with current cellular network calling patterns and is unlikely to be so for PCNs. If the number of access MANs decreases, backbone interconnection becomes increasingly possible, but the expense of installing a backbone MAN commands a proportionally higher burden on capital expenditures.

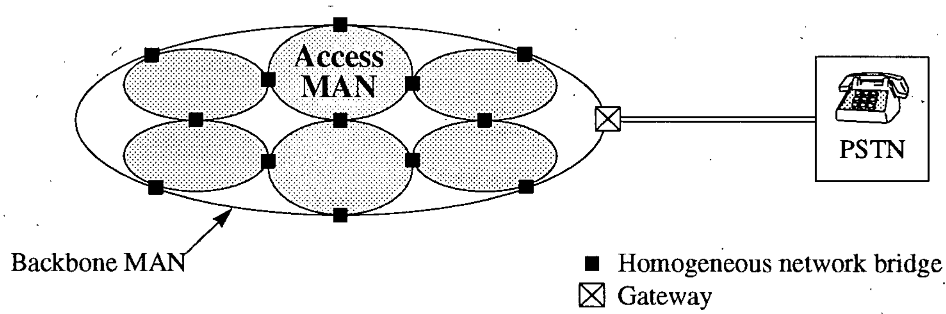

Figure 6.1. Six-MAN PCN with interconnection to the PSTN

6.1.2. The MAN Gateway-PSTN Interconnection Scheme

A PCN network architecture based on multiple MANs is proposed that involves some combination of HOB-PSTN interconnection and inter-HOB connectivity, depending on factors such as the intended coverage area of each MAN and the physical location of PSTN switches. Interconnected HOBs will allow for some trunking efficiencies for user transport to and from the PSTN. GSM provides similiar inter-MSC connectivity so that a conversation involving two mobiles need not switch through the PSTN [4].

Figure 6.2 presents a variation on this idea by showing the connectivity that may exist for two physically separated PCNs that require the PSTN as an intermediate carrier. PCN-1 features three MANs that have interconnected HOB functions, one of which, named the gateway HOB, provides connectivity to the PSTN. Three of the MANs share two point-to-point homogeneous bridges between MANs that provide VCI conversion and can support inter-MAN voice transport and signaling exchanges such as call setup and handoff. These procedures are described in [19,35]. Bridges are not essential as the aforementioned functions can be handled by the trunks between HOBs which already exist. Handoffs are probably best handled by the bridges, though, as only one additional full duplex circuit is required (on the new MAN), as opposed to two new ones (on the new MAN and additionally on the inter-HOB trunks) when the bridge isn't employed. This should reduce inter-MAN handoff delays, thus reducing the likelihood of call dropping. As in GSM, the problem of executing an inter-MAN handoff between disjoint networks remains.

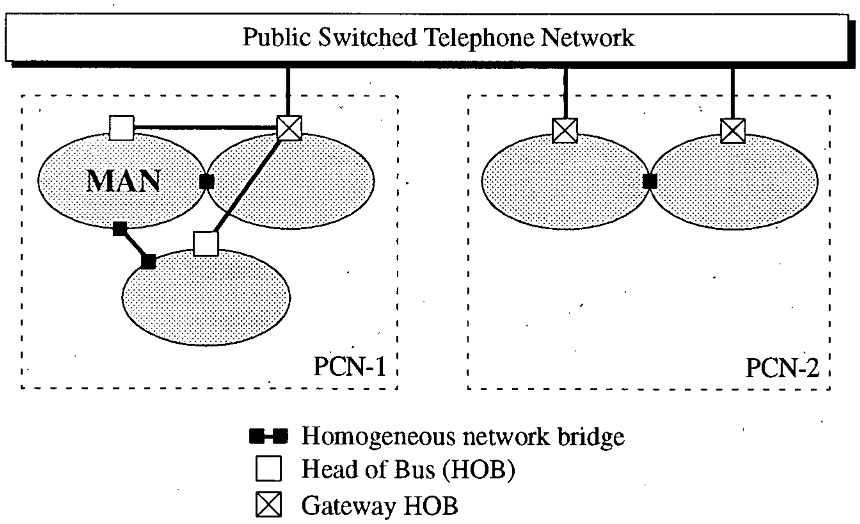

Figure 6.2. Variation on multiple-MAN PCN user transport architecture

6.1.3. Network Architecture to Support a Multiple-MAN PCN

A potential network architecture to support the proposed MAN interconnection scheme is suggested in Figure 6.3. It is adapted from [23] and places PCN-1 from Figure 6.2 in the context of the Intelligent Network. On the user transport plane, the point of interconnection at the PSTN is a tandem office (or toll office or class 4 switch), since an end office (class 5 switch) may not have the means to support 911 calls and operator services [36]. The subscriber registers at the BSC level are not assumed to be a part of the signaling network since proprietary protocols to support database transactions may be more efficient than using the signaling links employed to support the PCN-PSTN interconnection.

In a variation on the database (SCP) locations, the HLR may be considered to be a part of the fixed network [8]. To allow for quicker database transactions, the VLRs may be Adjuncts [23], as opposed to SCPs. Alternatively, each HOB may collocate an HLR function with a VLR if it is warranted by the size of the subscriber base per MAN.

Common Channel Signaling based on Signaling System No.7 is proposed as the signaling transport mechanism to support call control and mobility management outside of the MANs. Signaling for a PCS network interconnected to the PSTN has been investigated [36]. The IS-41 digital cellular intersystem operations protocol is implemented in the MAP portion of SS7 for mobility management, while the ISUP is used for call setup and release.

In SS7, A-links are used for the signaling exchanges between the HOB and the an STP pair. D-links provide interconnection between STP pairs. If ATM transport is used to interconnect HOBs, the STP can be implemented in the HOB to allow for associated signaling. This will effectively integrate the STP and SSP functions.

Since several network operators may be involved, it is not assumed that the PSTN and PCNs are controlled by the same Intelligent Network. Operators of different PCNs can facilitate interworking by using the same MAP application in their signaling protocols.

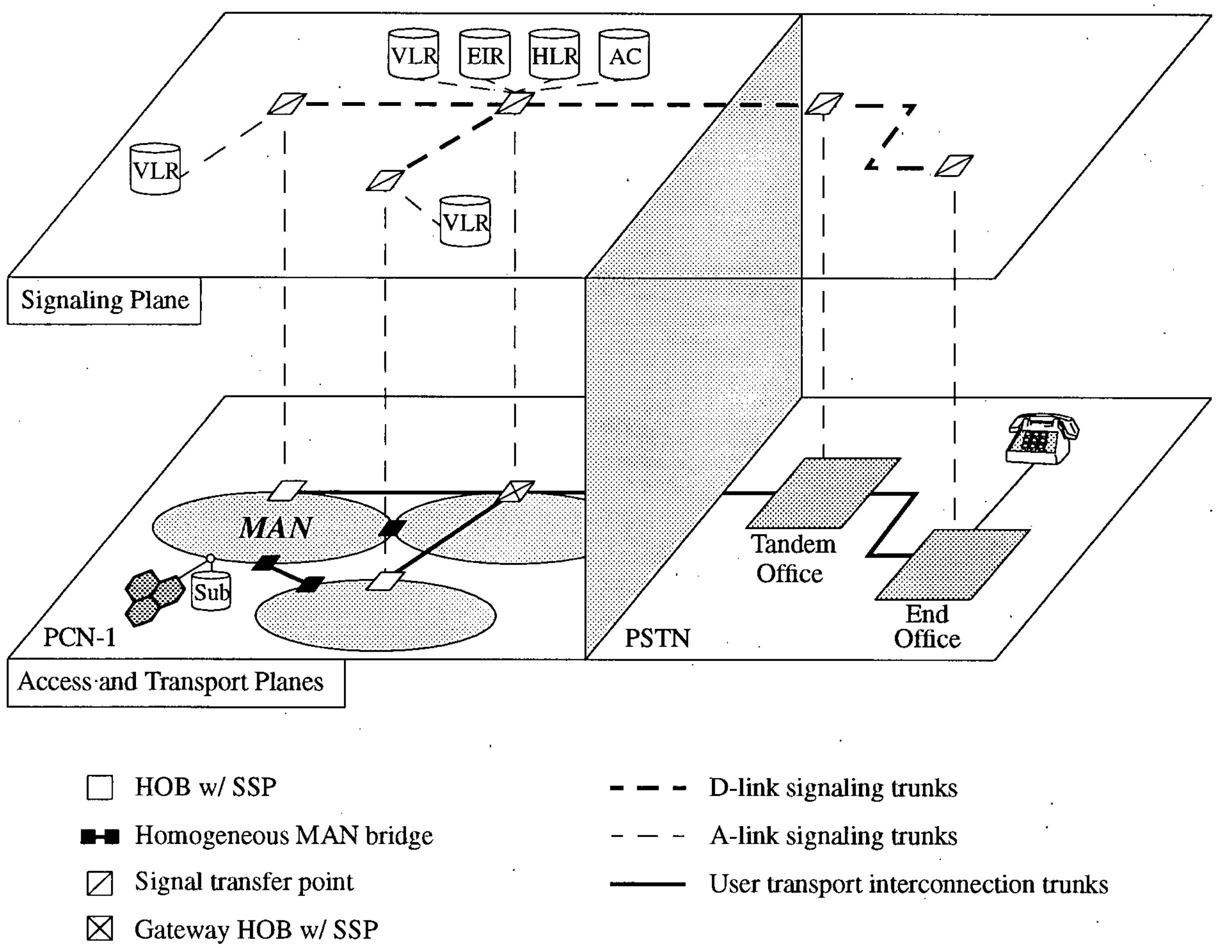

Figure 6.3. Potential network architecture for a multiple-MAN PCN

6.2. Migration Towards Asynchronous Transfer Mode

Prior to standardization, the IEEE 802.6 MAN slot format was aligned to that of the ATM cell. The purpose of this alignment is to have MANs fill an evolutionary role towards ATM/B-ISDN. Before an ATM network infrastructure is prevalent, MANs can serve to interconnect various data networks and, of course, provide a fixed network architecture for the distributed control of a PCN. As ATM is introduced into private and public networks, MANs will be able to merge with proposed broadband architectures.

Beyond today's choices for PCN network infrastructures lies the possibility of ATM-based transport for wireless services. Future generations of PCNs will likely internetwork through the ATM/B-ISDN [7,9]. It is, therefore, of interest to consider how a MAN-based PCN will fit to ATM.

6.2.1. Asynchronous Transfer Mode

Much excitement has been generated by the technology that is widely viewed as having the capabilities to deliver the Broadband Integrated Services Digital Network (B-ISDN). Asynchronous Transfer Mode (ATM) will provide the transmission, multiplexing, and switching facilities to support a rich mixture of constant and variable bandwidth traffic such as voice, video, and high speed data [14]. In the past, networks have been optimized for particular applications such as voice in the PSTN and packet data in X.25 networks.

ATM is based on fast packet switching techniques that route information at network nodes by means of an address contained in the packet header. The address is a combination of two hierarchical addresses: the Virtual Path Indicator (VPI) (highest) and Virtual Channel Indicator (VCI). The packet, referred to as a cell, has a fixed length of fifty-three bytes, five of which are for the header. High switching speeds are achieved by using a network with limited functionality, as evidenced by the relatively small header. Although bit error rates for optical fibre transport are very low, a reliable end-to-end bit pipe must be provided by higher layers.

The ATM protocol stack is shown in Figure 6.4. Several physical layers have been defined by the ATM Forum at data rates up to 155.52 Mbps. The ATM layer performs multiplexing, address translation, cell header generation and removal, and flow control (at the user-network interface only). The ATM Adaptation Layer (AAL) provides an interface between service requirements of higher layers and the 48 byte payload sent to the ATM layer. There are four variations:

AAL types 1, 2, 3/4, and 5. Types 1 and 2 are used for, respectively, constant isochronous bit rate services and variable isochronous bit rate services. The others provide for connection-oriented and connectionless data services, with Type 5 utilizing a simplified Segmentation and Reassembly (SAR) sub-layer for more efficient bandwidth utilization. New AALs have been proposed specifically for the transport of wireless voice traffic [30].

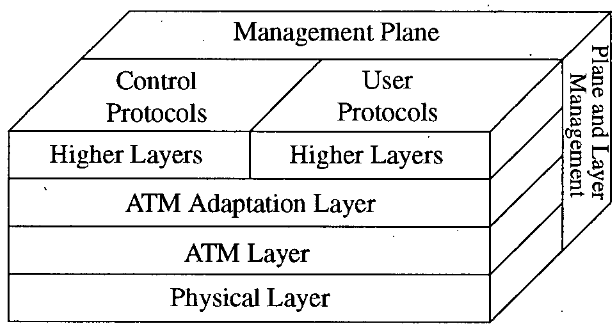

Figure 6.4. ATM protocol stack

Different traffic types are multiplexed on the same physical link by switches that have some knowledge of the negotiated Quality of Service (QoS) of a particular session. Cell headers do not indicate the AAL type being used so switches must associate VPI/VCI addresses with the QoS to ensure proper cell handling.

6.2.2. Architecture for a Hybrid IEEE 802.6 MAN-ATM PCN

A proposed ATM-transport network employing a hierarchical arrangement of ATM switching elements to interconnect microcells has been suggested [37]. This arrangement may be stand-alone and owned by a PCN operator, or be a subset of a large ATM network, so that resources are shared by other services and operators. In either case, an ATM transport network may be traversed to reach other networks such as the PSTN or other PCNs.

ATM can also be used to incrementally increase the capacity of a MAN-based PCN if the addition of another MAN is not economical. If consumer demand for wireless services grows unexpectedly, shifting base station interconnections and adding small ATM switches and multiplexers with dedicated links to new or existing microcells may alleviate capacity problems. A hybrid MAN-ATM PCN would then evolve. The concept of distributed control is partially lost, though. Figure 6.5 illustrates the potential architecture evolving from the previous example of two PCNs. Inter-HOB connections are replaced with dedicated ATM links to small ATM switches whose traffic is gathered by a larger switch. ATM multiplexers concentrate the user transport and signaling information from base stations. Connections between the smaller switches should exist to facilitate mobile-to-mobile calls and handoffs within the confines of the PCN. Investigations to support fast handoffs in an ATM architecture are currently being conducted [38].

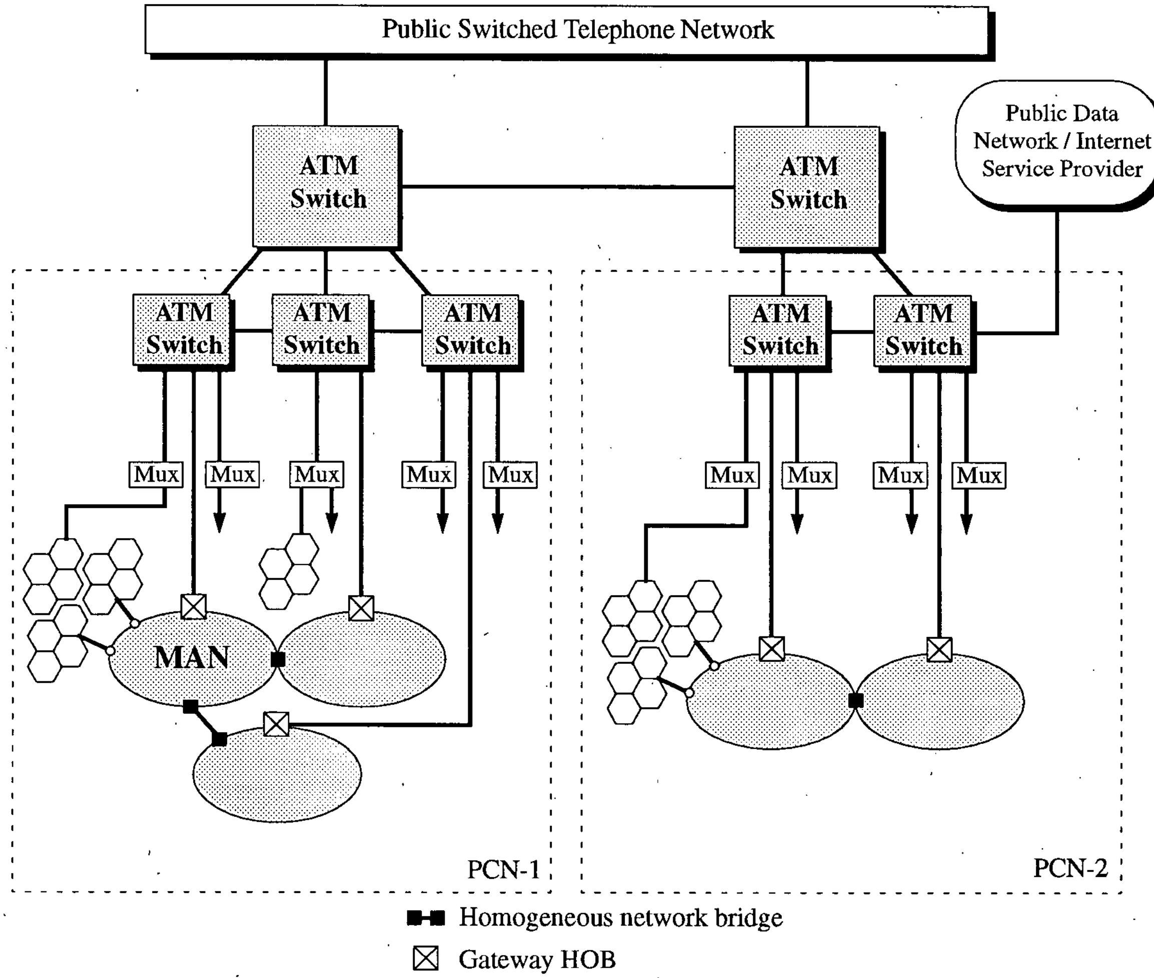

Figure 6.5. Hybrid MAN-ATM PCN user transport architecture

6.2.3. Interworking Between the IEEE 802.6 MAN and ATM

Although the ATM cell and MAN slot share the same length and common fields, interworking between their respective networks is not trivial. There are differences in voice transport, signaling and addressing, to name a few areas, that require consideration.

6.2.3.1. Voice Transport

The PCN architectures proposed will use PA access for voice on the MAN portion of the network. PA transport is partially characterized by its delivery of slots at a destination with constant interarrival times. This reduces the requirements for playback buffers to overcome jitter caused by asynchronous transmission. Since all types of traffic are multiplexed in ATM, no direct equivalent of isochronous service is available. The best that can be done currently is circuit emulation via AAL-1 and AAL-2. If a constant bit rate (CBR) for voice is assumed, AAL-1 transport can be employed to carry voice in the ATM network. Voice travelling through any ATM portion of the PCN would require buffering at the destination BS if it is served by an ATM link or by the HOB if the destination BS is served by the MAN. In the latter case, the Isochronous Convergence Function (ICF) is employed to smooth rate differences.

To avoid costly packetization delays it is suggested that ATM incorporate the proposed MAN voice transport scheme of 2 calls/slot. The Convergence Sublayer (CS) of AAL-1 would require modification to handle two source streams to form the Segmentation and Reassembly (SAR) Protocol Data Unit (PDU) payload. In AAL-1, an octet of overhead within the SAR-PDU is required for functions such as sequencing and source clock recovery with the result that the actual payload is reduced to 47 bytes. For vocoder rates at 9.6 kbps, this poses a problem. However, these functions may not be required and so a simplified (or non-existent) AAL-1 SAR layer can be used which has the benefit of reduced processing requirements.

6.2.3.2. Signaling

A signaling gateway to map call control and mobility management functions from the MAN to the ATM network is required. The requirements for B-ISDN signaling will be complex, in part because development of intelligent network and mobile-customer services are occurring at the same time that B-ISDN is being implemented [25]. For CCS/SS7, features required for connection control in B-ISDN will be reflected in the Broadband ISDN User Part (BISUP), based on ISUP.

In cases of high voice traffic load at ATM switches, a trade-off must be made between voice packet loss and signaling delay for handoffs. As on the MAN, it is important that packet data does not adversely affect the queueing delays of signaling traffic. The Cell Loss Priority (CLP) bit in the ATM header may be set for packet data to prevent excessive signaling delays, especially for handoffs. QA traffic from the MAN can be mapped to AAL-3/4 or AAL-5 depending on the service requirements.

6.2.3.3. Addressing

A notable difference between ATM and the DQDB MAN is the space allocated for addressing; 20 bits are allocated in the MAN slot header as opposed to 24 bits in the ATM cell header. Address translation at the HOB gateway should not be problematic since the MAN provides only one route for slots. Routing connectionless data from the MAN will be more challenging, since these slots are identified by a series of 1's in the VCI field of the slot header.

Multiplexing two calls per ATM cell implies an additional level of addressing beyond the VPI/VCI combination. It has been recommended that identifiers in the ATM header define the voice source payload boundaries and also identify the connection for each source [30]. The header may not have room for such flexibility, however. Instead, the position that a voice call assumes in the cell can be assigned during call setup, as is the case for isochronous service in the IEEE 802.6 standard. Calls are then identified and routed at an ATM switch by their VPI/VCI/offset combination. Two entries per VPI/VCI at a switch's routing table are required- one for each voice circuit.

The additional level of addressing requires additional processing in intermediate switches, specifically at the AAL layer. Here the benefit of using a small number of calls per cell is seen. It is most efficient in terms of processing at the switches to assign circuits such that calls share a common path in the MAN-ATM PCN for as long as possible. For example, if two calls that share a cell take diverging paths at the next switch, the ATM cell payload must be processed by the AAL. If they share the same path after the switch, AAL processing can be avoided and a single entry in the routing table to translate the incoming VPI/VCI combination is sufficient. As the number of calls carried per cell increases, the likelihood of having to process at the AAL becomes greater as all calls within a cell are less likely to have similiar paths. Carrying one call per cell obviates the need for any AAL processing at intermediate switches. This situation, however, must be compromised to lower the cell packetization delay. Figure 6.6 shows two simplex circuits originating at a MAN. They share a MAN slot and ATM cell until their paths diverge at a switch that separates the voice payloads in the AAL and multiplexes them with calls from another stream that share similiar destinations.

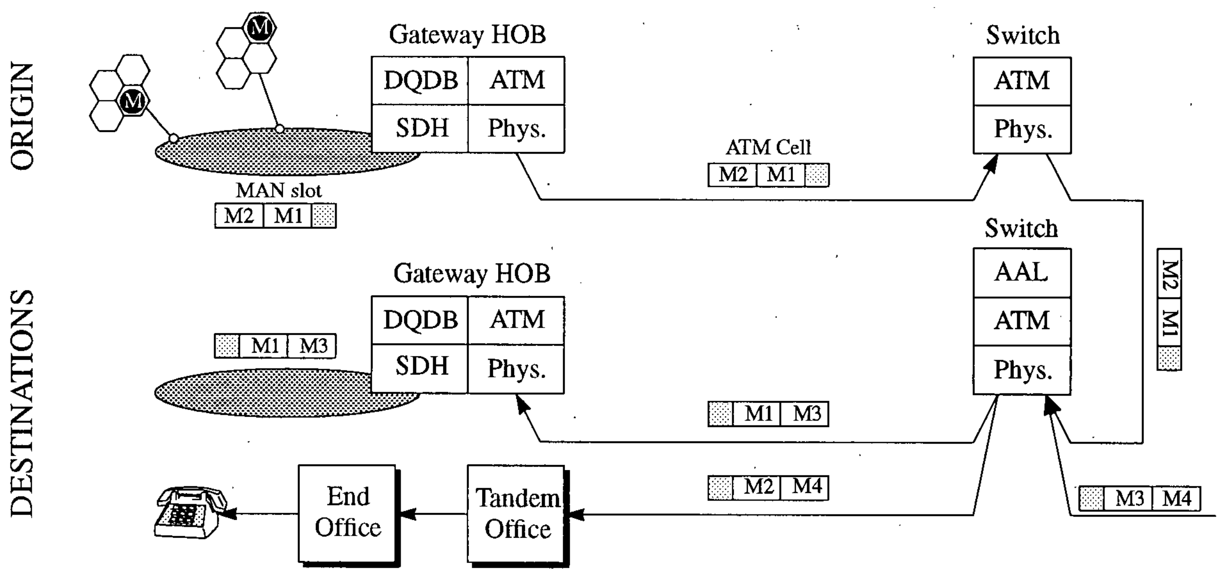

Figure 6.6. Conversations following diverging paths in an MAN-ATM PCN

Multicasting cells to outgoing links may be used as an alternative. The processing at intermediate nodes will be reduced as the AAL is avoided. If bandwidth can be spared, this may be a more attractive routing option.

Chapter 7. Summary and Conclusions

The interest in the IEEE 802.6 MAN for Personal Communication Networks is motivated by the need for distributed network control to handle an expected rise in signaling traffic caused by the widespread deployment of microcells and picocells. Previous work has demonstrated the viability of the technology, and this thesis attempts to make improvements to aspects such as voice transport, signaling, and capacity, while examining an alternative multiple-MAN architecture and its evolution towards ATM/B-ISDN.

7.1. Summary of Findings

The use of the Preemptive Priority Mechanism has been investigated as an improvement to the DQDB protocol. Under different loading conditions, the mechanism exhibited varying degrees of improvement. It is important to recall that the implementation of the PPM will forego any future use of the Reserved bits in the Access Control Field of the MAN slot. It was suggested, however, that the mechanism be adopted to improve the signaling delays for a MAN-based PCN due to its simplicity and effectiveness.

On the topic of voice transport for PCS over a MAN, an extension of the MAN frame length to 20 msec was proposed to accommodate two low bit-rate encoded voice calls per PA slot. The scheme represents a trade-off between two factors: i) reducing packetization delay; and ii) minimizing the number of calls per cell. Transporting voice on the MAN that is coded at lower rates has obvious capacity advantages.

The configuration of MAN databases has implications for signaling traffic. BSC-level subscriber registers allow for decentralized location updates, and maintaining subscriber profiles at a

MAN-level VLR reduces the signaling load. A logical procedure for an intra-MAN mobile-to-mobile call setup using this database arrangement illustrated its benefits.

It was shown that the placement of a gateway at the HOB of a closed-bus MAN will provide significant capacity benefits if the proportion of calls to the fixed network is high. These calls can use one MAN circuit with circuit erasure capabilities at the BSC to provide full-duplex communications. To determine the Erlang capacity of the MAN under this assumption, a two-dimensional Markov Chain was derived to model two session classes competing for MAN resources. Expressions for the blocking probabilities of each session class were derived so that the sustainable traffic load of each class could be determined. For a single MAN transporting two 9.6 kbps calls per cell, a QA:PA bandwidth split of 1:3, and a maximum blocking probability of 0.005, the sustainable traffic load during a busy hour was found to be 17680 Erlangs. Using derived expressions, the number of potential voice subscribers was determined to be approximately 266,000. If a market penetration rate of 20% is assumed for PCS, a PCN based on a single MAN could potentially cover a metropolitan population of around 1.3 million people.

With these quantities and a derived expression for location updating, PCN signaling on a MAN was simulated to determine the potential reduction in signaling delays by using the PPM. Results indicate significant improvements for call setup, although the throughput of QA-0 traffic suffers somewhat as a result of the PPM. Decentralized location updates afforded improvements in QA-0 throughput at the HOB.

Single and multiple-MAN network architectures for PCNs have been suggested in the context of the Intelligent Network (IN). A hybrid MAN-ATM architecture has been proposed. The migration from a MAN-based PCN to ATM has been considered in terms of voice transport, signaling, addressing, and routing.

7.2. Topics for Future Investigation

Subject to the continued interest in using IEEE 802.6 MANs for PCNs, further research may be pursued in the following topics:

- call transport to accommodate 13 kbps source coding
- MAN signaling requirements for wireless data services
- capacity determination for a MAN-based PCN using CDMA

Some of the topics discussed in Chapter 6 with regards to wireless networking with ATM will require more analysis.

- routing of calls that share the same ATM cell
- expediting signaling traffic at ATM switches under heavy loading due to voice and packet data

References

[1] "The end of the line: A survey of telecommunications", *The Economist*, October 23, 1993.

[2] D. C. Cox, "Wireless personal communications: What is it?", *IEEE Personal Communications*, vol. 2, no. 2, pp. 20-35, April 1995.

[3] M. H. Callender, "Future public land mobile telecommunications systems", *IEEE Personal Communications*, vol. 1, no. 4, pp. 18-22, 4th Qtr. 1994.

[4] M. Rahnema, "Overview of the GSM system and protocol architecture," *IEEE Communications Magazine*, vol. 31, no. 4, pp. 92-100, April 1993.

[5] International Standard ISO/IEC 8802-6, *Distributed Queue Dual Bus (DQDB) access method and physical layer specifications*, First Edition. New York: IEEE, 1994.

[6] D. J. Goodman, "Cellular Packet Communications," *IEEE Transactions on Communications*, vol. 38, no. 8, pp. 1272-1280, August 1990.

[7] A. D. Malyan, L. J. Ng, V. C. M. Leung, R. W. Donaldson, "Network architecture and signaling for wireless personal communications," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 6, pp. 830-840, August 1993.

[8] K. S. Meier-Hellstern, G. P. Pollini, D.J. Goodman, "Network protocols for the cellular packet switch," *IEEE Transactions on Communications*, vol. 42, no. 2/3/4, pp.1235-1244, February/March/April 1994.

[9] V. C. M. Leung, N. Qian, A. D. Malyan, R. W. Donaldson, "Call control and traffic transport for connection-oriented high speed wireless personal communications over metropolitan area networks," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 8, pp. 1376-1388, October 1994.

[10] W. Y. L. Wong, V. C. M. Leung, R. W. Donaldson, "Call transport alternatives for interconnected wireless communication networks based on the IEEE 802.6 MAN," *IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing*, pp. 235-240, Victoria, B.C., May 1995.

[11] S. Rao, D. J. Goodman, G. P. Pollini, K. S. Meier-Hellstern, "Interworking between Digital European Cordless Telecommunications and a distributed packet switch," *Wireless Networks*, vol. 1, no. 1, pp. 83-93, February 1995.

[12] H. C. B. Chan, V. C. M. Leung, "Reservation-arbitrated access for isochronous voice transport over dual-bus metropolitan area networks," *Proc. IEEE International Conference on Communications*, pp. 1467-1471, Seattle, WA, June 1995.

[13] C. Bisdikian, A. N. Tantawy, "A mechanism for implementing preemptive priorities in DQDB subnetworks," *IEEE Transactions on Communications*, vol. 42, no. 2/3/4, pp. 834-839, February/March/April 1994.

[14] C. Bisdikian, B. Patel, F. Schaffe, M. Willebeek-LeMair, "Approaching B-ISDN: An overview of ATM and DQDB," *Asynchronous Transfer Mode Networks*, Y. Viniotis and R. O. Onvural, Eds., New York, NY: Plenum Press, 1993.

[15] W. R. Byrne, G. Clapp, H. J. Kafka, G. W. R. Luderer, B. L. Nelson, "Evolution of metropolitan area networks to broadband ISDN," *IEEE Communications Magazine*, vol. 29, no. 1, pp. 69-82, January 1991.

[16] E. L. Hahne, A. K. Choudhury and N. F. Maxemchuk, "DQDB networks with and without bandwidth balancing," *IEEE Transactions on Communications*, vol. 40, no. 7, pp. 1192-1204, July 1992.

[17] H. R. van As, J. W. Wong, P. Zafiropulo, "Fairness, priority and predictability of the DQDB

MAC protocol under heavy load," *Proc. 1990 International Seminar on Digital Communications*, Zurich, Switzerland, pp. 410-417, March 1990.

[18] V. Catania, L. Mazzola, A. Puliafito, L. Vita, "Throughput analysis of DQDB in overload conditions," *Proc. IEEE International Conference on Communications*, pp. 741-747, June 1991.

[19] N. Qian, "Call control signaling for personal communications over interconnected metropolitan area networks," University of British Columbia, Department of Electrical Engineering, M.A.Sc. thesis, April 1993.

[20] M. Zukerman, P. G. Potter, "A protocol for eraser node implementation within the DQDB framework," *Proc. IEEE International Conference on Communications*, pp. 1400-1404, December 1990.

[21] J. Homa, S. Harris, "Intelligent network requirements for personal communications services," *IEEE Communications Magazine*, vol. 30, no. 2, pp. 70-76, February 1992.

[22] R. B. Robrock, "The intelligent network - Changing the face of telecommunications," *Proceedings of the IEEE*, vol. 79, no. 1, January 1991.

[23] B. Jabbari, "Intelligent network concepts in mobile communications," *IEEE Communications Magazine*, vol. 30, no. 2, pp. 64-69, February 1992.

[24] A. D. Malyan, V. C. M. Leung, "Network architecture and signaling for urban personal communications," *Proc. IEEE International Conference on Selected Topics in Wireless Communications*, Vancouver, B.C., pp. 305-309, June 1992.

[25] P. J. Kuhn, C. D. Pack, R. A. Skoog, "Common channel signaling networks: Past, present, future," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 3, pp. 383-394, April 1994.

[26] J. Liebeherr, I.F. Akyildiz, A.N. Tantawi, "Dual bus MANs with multiple-priority traffic," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 8, pp. 1202-1213, October 1993.

[27] R. Padovani, "Reverse link performance of IS-95 based cellular systems", *IEEE Personal Communications*, vol. 1, no. 3, pp. 28-34, 3rd Quarter 1994.

[28] S. Fist, "GSM - Better late than never?" *Australian Communications*, pp. 73-78, June 1993.

[29] D. J. Goodman, R. A. Valenzuela, K. T. Gayliard, B. Ramamurthi, "Packet reservation multiple access for local wireless communications," *IEEE Transactions on Communications*, vol. 37, no. 8, pp. 478-485, August 1989.

[30] B. T. Doshi, A. Sawkar, "An ATM based PCS/cellular architecture," *Proc. of the 5th WINLAB Workshop*, pp. 23-35, East Brunswick, NJ, April 1995.

[31] D. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ: Prentice Hall, 1991.

[32] A. J. Viterbi, A. M. Viterbi, K. S. Gilhousen, E. Zehavi, "Soft handoff extends CDMA cell coverage and increases reverse link capacity," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 8, pp. 1281-1288, October 1994.

[33] G. P. Pollini, "Capacity of an IEEE 802.6 based cellular packet switch," *Proc. IEEE International Conference on Communications*, Geneva, Switzerland, 1993.

[34] W. Y. L. Wong, "Delay throughput analysis of inter- and intra-MAN voice and data integrated traffic in IEEE 802.6 MAN based PCN," University of British Columbia, Department of Electrical Engineering, M.A.Sc. thesis, February 1995.

[35] L. J. Ng, "Distributed architectures and databases for personal communications networks and services," University of British Columbia, Department of Electrical Engineering, M.A.Sc. thesis, March 1993.

[36] Y. Lin, S. K. DeVries, "PCS network signaling using SS7," *IEEE Personal Communications*, vol. 2, no. 3, pp. 92-100, April 1993.

[37] D. Raychaudhuri, N. D. Wilson, "ATM-based transport architecture for multiservices wireless personal communications networks," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 8, pp. 1401-1414, October 1992.

[38] O. T. W. Yu, V. C. M. Leung, "Extending B-ISDN to support user terminal mobility over an ATM-based personal communications network," *To appear in IEEE GLOBECOM '95*, Singapore, November 1995.

Appendix A. List of Abbreviations and Acronyms

ACF	Access Control Field
AIN	Advanced Intelligent Network
AMPS	Advanced Mobile Phone Service
ATM	Asynchronous Transfer Mode
B-ISDN	Broadband Integrated Services Digital Network
BS	Base Station
BSC	Base Station Controller
BWB	Bandwidth Balancing
CBR	Constant Bit Rate
CCS	Common Channel Signaling
CDPD	Cellular Digital Packet Data
DECT	Digital European Cordless Telephone
DQDB	Distributed Queue Dual Bus
FPLMTS	Future Public Land Mobile Telephone System
GoS	Grade of Service
GSM	Global System for Mobiles
HOB	Head of Bus
IMPDU	Initial MAC Protocol Data Unit
IN	Intelligent Network
IP	Intelligent Peripheral
ISDN	Integrated Services Digital Network

ISO	International Organization for Standardization
ISUP	ISDN User Part
ITU	International Telecommunications Union
LAN	Local Area Network
LBE	Low Bit-Rate Encoding
MAC	Media Access Control
MAN	Metropolitan Area Network
MAP	Mobile Application Part
MSC	Mobile Switching Center
OSI	Open System Interconnection
PA	Pre-Arbitrated
PCM	Pulse-Code Modulation
PCN	Personal Communications Network
PCS	Personal Communications Services
PDN	Public Data Network
PDU	Protocol Data Unit
PLCP	Physical Layer Convergence Procedure
PPM	Preemptive Priority Mechanism
PSTN	Public Switched Telephone Network
QA	Queue Arbitrated
QoS	Quality of Service
RA	Reservation Arbitrated
SAP	Service Access Point

SAR	Segmentation and Reassembly
SCE	Service Creation Environment
SCP	Service Control Point
SH	Segment Header
SMDS	Switched Multimegabit Data Service
SMS	Service Management System
SN	Service Node
SS7	Signaling System No. 7
SSP	Service Switching Point
STP	Signal Transfer Point
VCI	Virtual Circuit/Channel Indicator
VPI	Virtual Path Indicator

Appendix B. MAN Signaling Message Abbreviations [19]

Call Setup

ALT	Alerting (callee)
AUTH_REQ	Authentication Request
AUTH_RST	Authentication Result
CNT	Connect
CNT_ACK	Connect Acknowledge
CPD	Call Proceeding
INF	Information
INQ_CLE	Inquiring Callee
INQ_RST	Inquiry Result
PRG	Progress (of call setup)
STP	Setup
STP_ACK	Setup Acknowledge
STP_REQ	Setup Request
STP_RST	Setup Result

Call Clear

DISC	Disconnect
REL	Release
RCP	Release Complete

Handoff

CFM_HO	Confirm Handoff
EXG_INF	Exchange Call Information
HO_PD	HandoffProceeding
HO_REQ	Handoff Request
REQ_INF	Request Call Information
SPY_INF	Supply Call Information

Location Update

LOC_ACK	Location Update Acknowledgment
LOC_UPD	Location Update Request

MAN Facility Setup/Release

CFM_ISOC	Confirm Isochronous Channel
CFM_VCI	Confirm VCI
NTY_VCI	Notify VCI
RCP_ISOC	Release Complete Isochronous Channel
RCP_VCI	Release Complete VCI
REL_ISOC	Release Isochronous Channel
REL_VCI	Release VCI
REQ_ISOC	Request Isochronous Channel
RES_VCI	Reserve VCI

Appendix C. Selections from Opnet Simulation Models

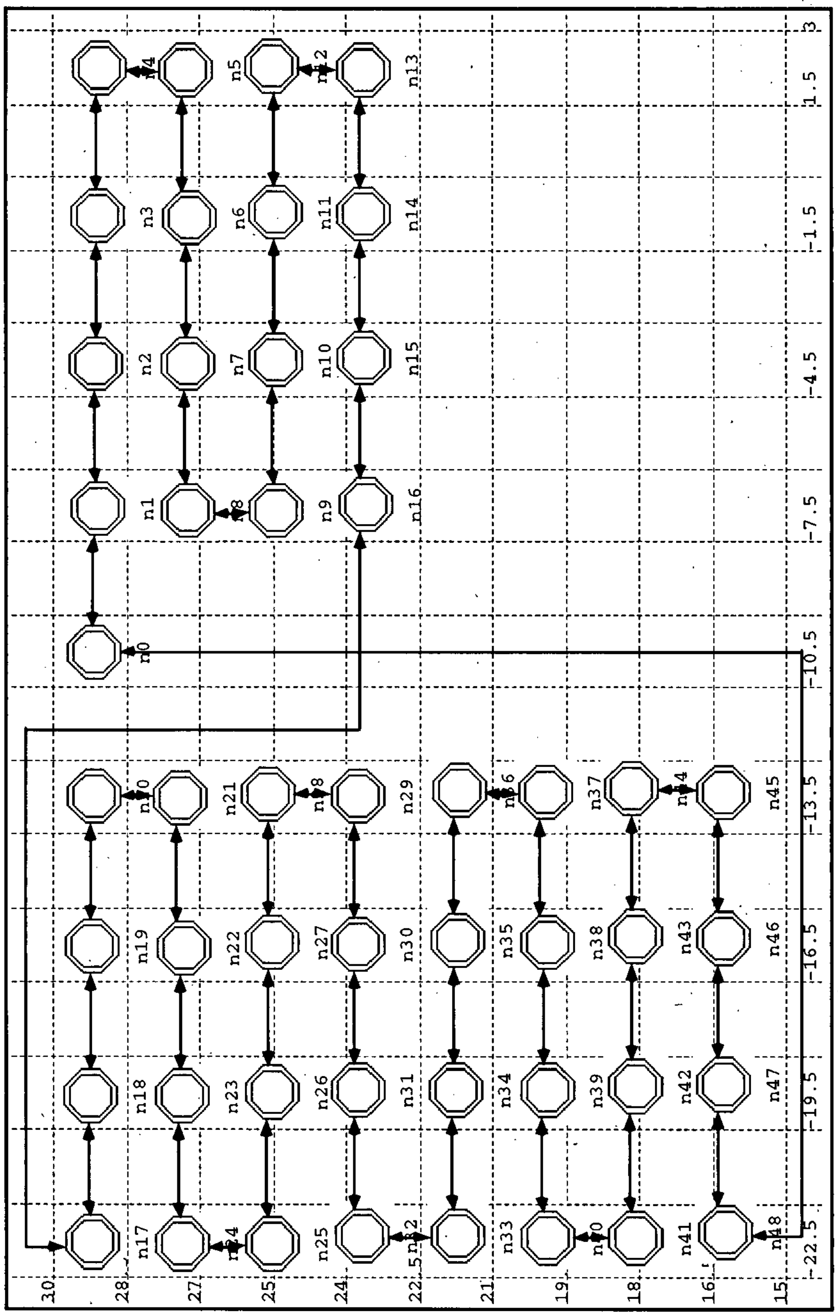

Figure C.1. 50-node MAN network model

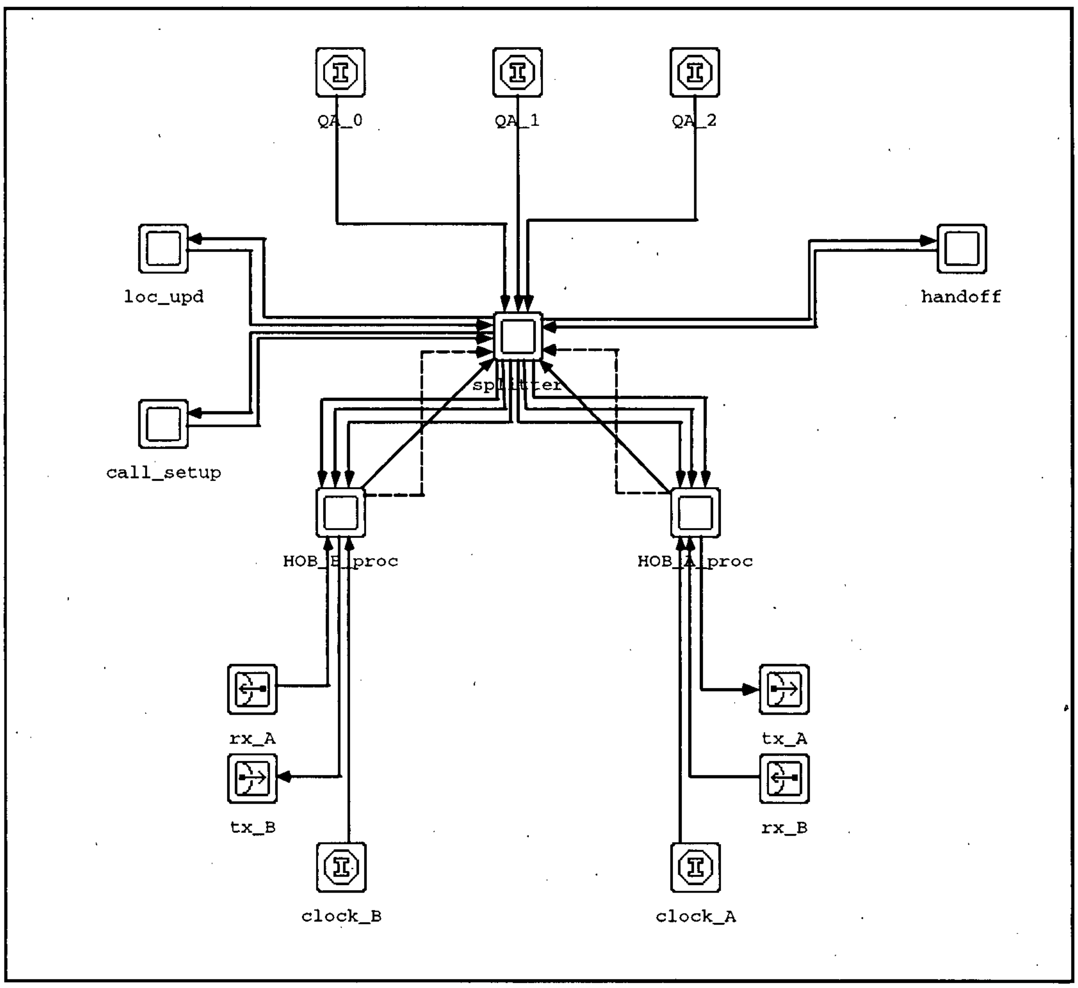

Figure C.2. Head of Bus node model

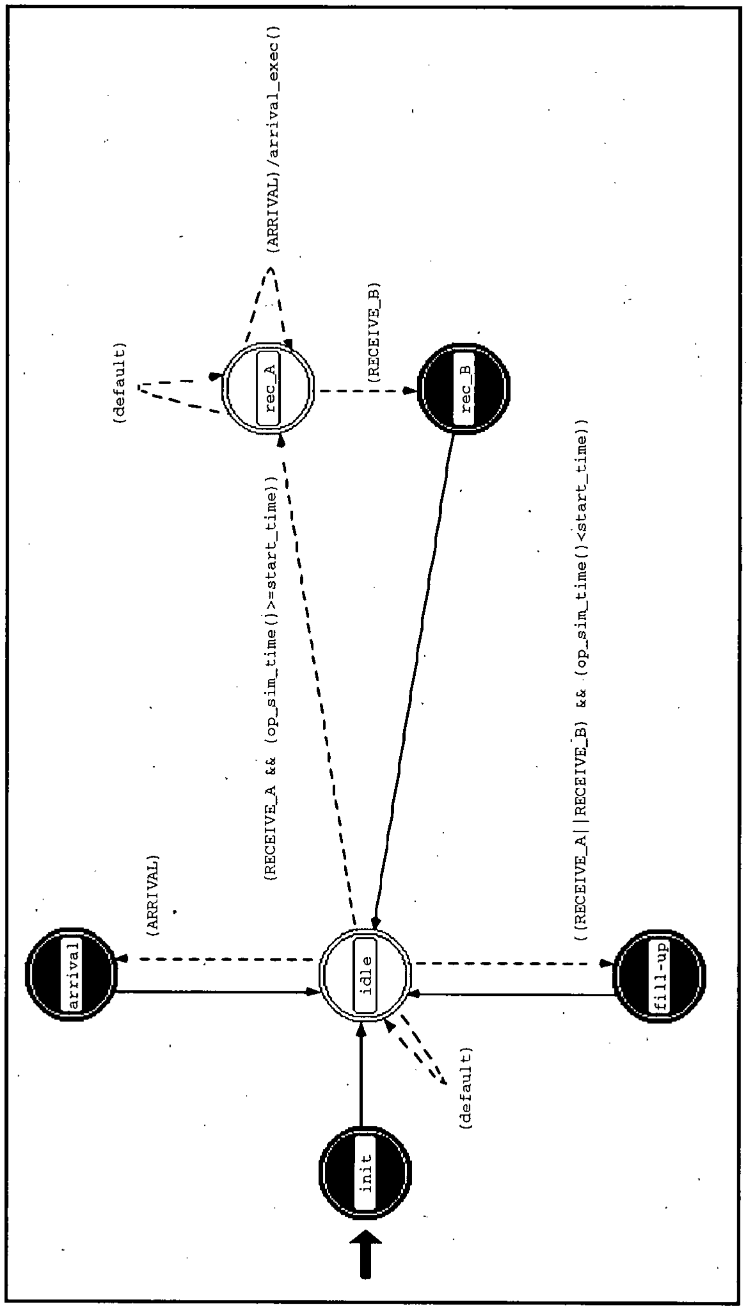

Figure C.3. DQDB process model

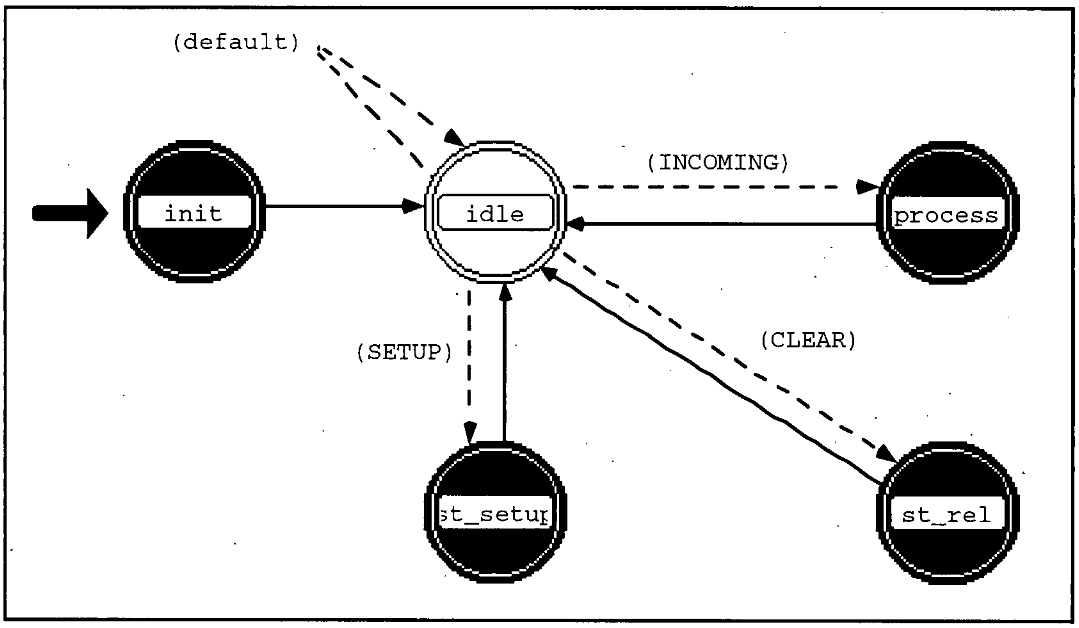

Figure C.4. Call setup and clear process model

Appendix D. 2-dimensional Markov Chain Approximation

Using the first session class of the two-dimensional Markov Chain representation described in Chapter 4, the blocking probability, P_{B_1}, is given by:

$$P_{B_1} = \frac{\displaystyle\sum_{n_2=0}^{m/2} \frac{\rho_1^{(m-2n_2)}}{(m-2n_2)!} \cdot \frac{\rho_2^{n_2}}{n_2!}}{\displaystyle\sum_{n_2=0}^{m/2} \sum_{n_1=0}^{(m-2n_2)} \frac{\rho_1^{n_1}}{n_1!} \cdot \frac{\rho_2^{n_2}}{n_2!}} \tag{D.1}$$

$$= \sum_{n_2=0}^{m/2} \frac{t_{n_2}}{G} \tag{D.2}$$

where G represents the set of potential system states reflected in the denominator of (D.1) and can be expressed as

$$G = G_{0,0} + G_{1,0} + \ldots + G_{N_1,N_2} + \ldots + G_{0,m/2} \tag{D.3}$$

where N_1, N_2 is a potential system state in the Markov Chain. The terms t_{n_2} represent blocked system states in the numerator of (D.1) and

$$t_{n_2} = \frac{\rho_1^{(m-2n_2)}}{(m-2n_2)!} \cdot \frac{\rho_2^{n_2}}{n_2!}.$$

A quantity must be calculated for each potential blocked state-system state pair:

$$\frac{G_{N_1,N_2}}{t_{n_2}} = \frac{\rho^{N_1}\rho^{N_2}(m-2n_2)!n_2!}{\rho^{m-2n_2}\rho^{n_2}N_1!N_2!}. \tag{D.4}$$

For each value of n_2, the terms of (D.4) are summed. Each sum is inverted and all the resulting

n_2 terms are added, as is implied by (D.2), to determine the blocking probability. A similiar procedure is used to determine P_{B_2}. In calculating each term, processing efficiencies can be realized by reducing exponents and factorials where allowable, and by aborting calculations when terms begin to exceed a threshold value.

The factorials and exponents in (D.4) are prohibitive in determining the blocking probability for a large value of *m*. To shorten the computation time, a subset of the Markov Chain is used, as is illustrated in Figure D.1. This subset includes the expected system state, given by (5.7), which will lie approximately on the line with the slope given by the ratio of traffic loads for each session type, ρ_1 and ρ_2.

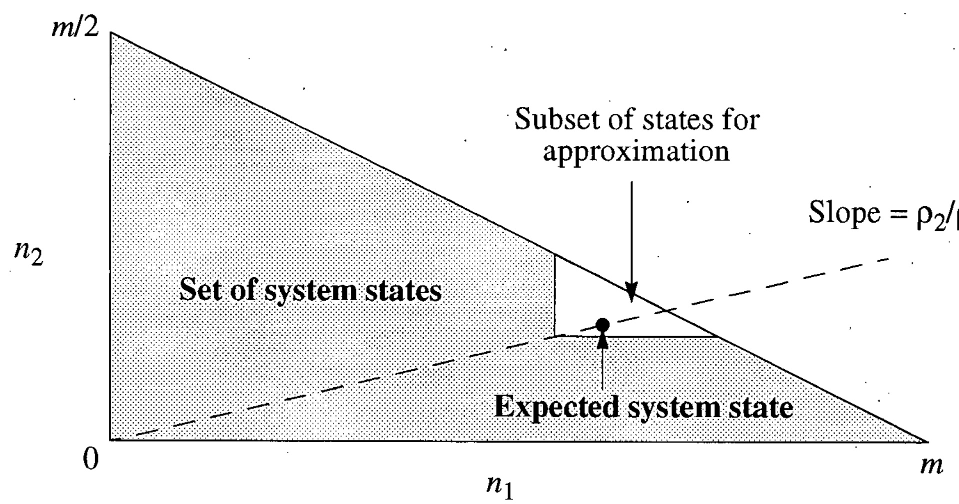

Figure D.1. 2-dimensional Markov Chain approximation

Even with a small subset of the original Markov Chain, computation levels are still very high for large *m* (in this thesis, $m = 21{,}120$). To further reduce calculations, system states are sampled in a grid-like fashion so that each gridpoint term (given by (D.4)) approximates many other terms in its physical vicinity. For example, a step size of 20 and 10 used for n_1 and n_2, respectively, means that 199 terms for every 200 are approximated. The choice of the grid step size and the size of the subset depends on a desired trade-off between accuracy and processing time. Smaller step sizes will mean greater accuracy for a given subset, while a larger subset provides a more optimistic estimate of the blocking probabilities (i.e. values are smaller), as was found through trial and error. The MAN capacity enhancement determined through this method, then, is conservative. Validation of the estimated sustainable traffic loads and resulting blocking probabilities must be provided through simulations.