INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Error Correction Techniques for ATM Communications

by

ABDULAZIZ S. ALMULHEM
M.A.Sc, King Fahd University of Petroleum and Minerals, 1994
B.Sc, King Fahd University of Petroleum and Minerals, 1990

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Electrical and Computer Engineering

We accept this dissertation as conforming
to the required standard

Prof. Fayez El-Guibaly, Supervisor, Dept. of Elect. & Comp. Eng.

Dr. Kin F. Li, Member, Dept. of Elect. & Comp. Eng.

Dr. Panajotis Agathoklis, Member, Dept. of Elect. & Comp. Eng.

Dr. Gholamali Shoja, Outside Member, Dept. of Computer Science

Dr. Hussein Alnuweiri, External Examiner
University of British Columbia

Supervisor: Prof. Fayez El-Guibaly

ABSTRACT

Congestion in ATM communications is a significant issue as it can have a dramatic effect on critical or real-time data. Forward Error Correction (FEC) codes are one class of protocols to decrease this effect. Conventional FEC techniques have a uniform or constant error correction rate, which can result in poor bandwidth utilization. Therefore adaptive techniques are sought. The rationale is to have better bandwidth utilization when congestion occurs. In this thesis, we investigate the related work on FEC in ATM networks. Then we propose an adaptive FEC scheme based on RS codes. This proposed scheme is then studied in different types of environments, wireline and wireless. Simulations are also conducted to measure different performance issues concerning network resources and quality of service.

Another crucial issue in ATM communications is security. The proposed FEC scheme has an added feature of being security ready. Moreover it has been shown that the security scheme is computationally secure.

Such FEC scheme has significant impact on ATM network resources and switch capacity. This has been investigated further in this work. Switch architectures utilizing FEC schemes are also studied.

Examiners:

Prof. Fayez El-Guibaly, Supervisor, Dept. of Elect. & Comp. Eng.

Dr. Kin F. Li, Member, Dept. of Elect. & Comp. Eng.

Dr. Panajotis Agathoklis, Member, Dept. of Elect. & Comp. Eng.

Dr. Gholamali Shoja, Outside Member, Dept. of Computer Science

Dr. Hussein Alnuwairi, External Examiner
University of British Columbia

Table of Contents

List of Figures

List of Tables

Notation

AAL	ATM Adaptation Layer
ABR	Available Bit Rate
AFEC	Adaptive Forward Error Correction
ARQ	Automatic Repeat reQuest
ATM	Asynchronous Transfer Mode
B-ISDN	Broadband ISDN
CAC	Connection Admission Control
CAN	Campus Area Network
CBR	Constant Bit Rate
CCITT	Consultative Committee International Telegraphy and Telephony (now known as ITU-T)
CLR	Cell Loss Ratio
CS	Convergence Sublayer
CTD	Cell Transfer Delay
EFCI	Explicit Forward Congestion Indication
EOM	End of Message
FEC	Forward Error Correction
GAN	Global Area Network
HDTV	High Definition TV
ISDN	Integrated Services Digital Network
ISO	International Standard Organization
ITU-T	International Telecommunication Union-Telecommunication Standardization Section
LAN	Local Area Network
MAN	Metropolitan Area Network
NISDN	Narrowband Integrated Services Digital Network
nrt-VBR	non real-time Variable Bit Rate

OAM	Operation and Management
OSI	Open System Interconnection
QoS	Quality of Service
RS	Reed-Solomon Codes
rt-VBR	real-time Variable Bit Rate
SAR	Segmentation And Reassembly
SRP	Selective Repeat Protocol
UBR	Unspecified Bit Rate
UPC	Usage Parameter Control
VP	Virtual Path
VC	Virtual Connection
WAN	Wide Area Network
WATM	Wireless ATM

Acknowledgement

I am indebted to my supervisor Prof. Fayez ElGuibally, who helped me at various stages of my work. I had the good fortune to be under his supervision. His support and assistance have been invaluable.

I also extend my sincere thanks to my thesis committee members, Dr. K. Li, Dr. P. Agathoklis, and Dr. G.A. Shoja, for their valuable comments and constructive criticism. In addition I would like to thank Dr. T.A. Gulliver who added to this work with his discussions.

This work would not have been possible without the generous support from King Fahd University of Petroleum and Minerals (KFUPM) in Dhahran (Saudi Arabia) , whom I undergo my Ph.D. studies under their sponsorship, and Micronet Canada.

Many people are special and have impacted my academic life, without their moral support and encouragement, I would not pursue my post graduate life. Special thanks go to Dr. M.S. Benten, Dr. S. Sait, and Dr. Y. Habib from (KFUPM).

Many thanks go to my colleagues in University of Victoria where I pursed my Ph.D. studies and in Nortel, Ottawa, where I conducted 18 month of internship.

Last but not the least, my family are always the source of enjoyment and encouragement to me. Without their care, love, and devotion, my life would not be the way it is now. Special thanks go to my wife, Haifa, for her patience and support. She has greatly contributed to this work although she is always asking about the nature of my work.

To the souls of
my father Sultan and my mother Shaikha

To my beloved wife Haifa and
my son Sultan and daughter Shaikha

Chapter 1

Introduction

A significant turning point in the history of telecommunications technology is the introduction of digital communications. It has opened the doors wide to integrate different types of information into a unified carrier and to utilize the advantages afforded by microelectronics and computer technologies. The main advances in digital communications are: powerful computing machines and connectivity.

Powerful computing machines are now affordable by individuals. These machines usually feature built-in signal and image processing capabilities with powerful graphic interfaces. This has lead to the development of more sophisticated multimedia applications, which can be supported by these machines.

Connectivity is defined as the ability of an individual getting access to local or global machines, databases, etc. which are assumed to be served by a digital communication network. Networks, depending on the geographical zoning, support different services. As a matter of fact, a network setup within a small geographical area, such as a LAN, CAN or MAN, would have more services and resources, such as bandwidth, than one with broader borders such as WAN or GAN. The rapid progression in connectivity has made practicable new work group applications, e.g. teleconferencing. This has definitely altered the way connectivity is looked at. For example, networks must provide one-to-many and many-to-many services besides the traditional one-to-one services.

LANs tend to expand in size; data traffic is growing at a rate close to 30 percent a year [6]. Technological advances in LAN internetworking have made it possible for private networks to connect to public WAN services. This has created an accelerating bandwidth demand.

Internet is a widely recognized general infrastructure for data communications.

Internet supports a wide range of computer-based applications. The type of services Internet provides have very elastic requirements [7]. Examples of these services include: electronic mail (e-mail), remote terminal (telnet), file transfer (FTP), etc. These Internet applications can back off when congestion is experienced, and retransmit packets even when packet loss is present [7].

Real-time applications (e.g. audio and video), on the other hand, have very rigid delay requirements. These applications cannot tolerate long delays and/or long congestion periods.

It is obvious from the above that user information varies in requirements and characteristics. The idea of *Integrated Services Digital Network* (ISDN) was adopted in the early 1970s [6]. ISDN mixes audio, video, and data into a common data carrier.

1.1 B-ISDN and ATM

The only deficiency with ISDN (sometimes called narrow ISDN (NISDN)) is its narrow bandwidth. Basic NISDN provides two 64-kbps Bearer (B) Channels and one 16-kbps Data (D) Channel. This interface is commonly referred to as 2B+D [6]. The bandwidth allocated by NISDN is not sufficient to transport different information services with varying bandwidth requirements. For example video information, such as digital TV and digital HDTV, requires 150 Mbps channel bandwidth [8].

Applications with high bandwidth requirements will not be supported by ISDN. An alternative proposal is to use Broadband ISDN (B-ISDN). Broadband in the sense that greater bandwidth is allocated in order to absorb different information types with high bandwidth requirements such as video.

Possible services supported by B-ISDN include: digital TV, HDTV, digital hi-fi, teleconferencing, multimedia terminals, and interconnection of LANs [8].

Asynchronous Transfer Mode (ATM) is a cell-based switching and multiplexing technology designed to be a general-purpose, connection-oriented transfer mode for a wide range of services [6]. ATM is also known as cell relay [9].

ATM differs from other technologies such as packet switching and frame relay in providing simple routing, guaranteed switching delays, light protocols, and guaranteed data sequence.

ATM is, therefore, the switching technique recommended by International Consultative Committee for Telecommunications and Telegraphy (CCITT) (now known as International Telecommunication Union- Telecommunication Standardization Section (ITU-T)), to carry B-ISDN traffic [6, 10, 11, 12]. The word asynchronous in ATM comes from the fact that time slots are not assigned to specific users as it is the case with synchronous transfer mode. Instead time slots are available to any user who is ready to transmit data. To ensure proper operation user data are prefixed with header that identifies the virtual channel [6].

ATM makes use of fixed-size cells, consisting of a 5-octet header and a 48-octet information field [6, 9, 11]. Switching or routing is based on the header information which primarily contains address information [6, 10].

ATM is a circuit-switched technology based on two generic connection concepts: virtual channel and virtual path [6, 9, 10]. A virtual channel (VC) is a generic term used to describe unidirectional transport of ATM cells associated with a common unique identifier value [9]. A bundle of VC links having the same endpoints is referred to as a virtual path (VP). VP and VC values are the elements on which ATM performs switching functions.

Although B-ISDN/ATM does not follow the International Standard Organization/Open System Interconnection (ISO/OSI) model, it makes extensive use of the OSI concepts of layering and sublayering [6]. The B-ISDN layers corresponding to the layers in the OSI model are depicted in Figure 1.1. The ATM layer in the B-ISDN model performs routing and switching functions. On the OSI model, these functions are usually part of the network layer, the layer above the data link layer (Fig. 1.1.)

1.1.1 ATM characteristics

ATM has basic characteristics that distinguish it from other switching techniques [10, 12].

- *No error protection or flow control on a link-by-link basis:*
  Since optical links of an ATM network have a very low bit error rate, no action is taken when an error occurs during transmission. Moreover no flow control functions are being considered and end-to-end protocols are employed to correct transmission errors.

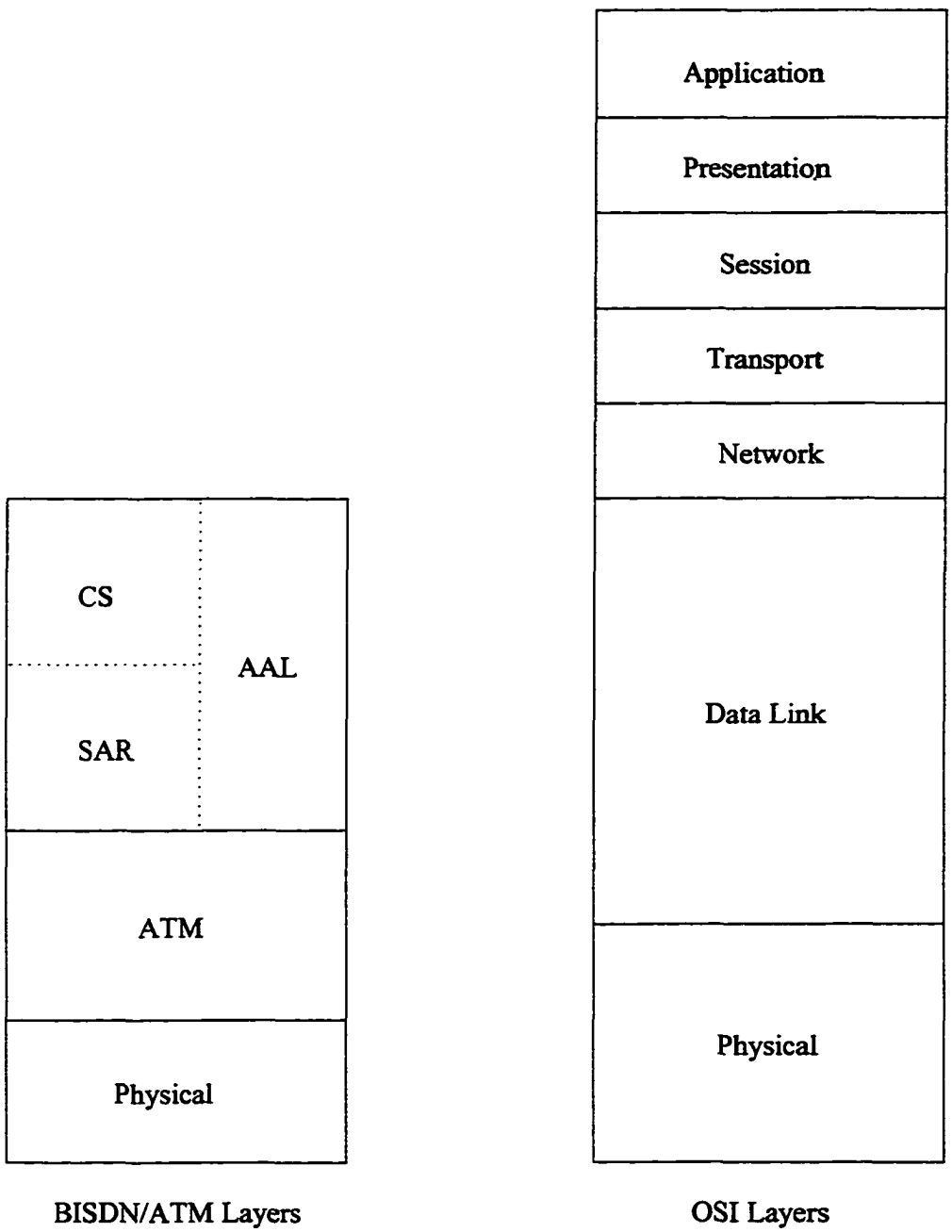

Figure 1.1. *Comparison between the lower two OSI layers and B-ISDN layers.*

- *Connection oriented mode of operation:*
  A connection setup is needed before any information transfer. Resources are reserved during the setup phase and the connection is rejected if no sufficient resources are available. Upon call completion, network resources are released. With this mode of operation minimal cell loss rate is guaranteed. This mainly simplifies cell routing and minimizes cell header information.
- *Reduced header functionality:*
  The information contained in the cell header is very limited. Basically it contains the identification of a connection and its route. This guarantees fast processing and consequently short delays between communicating parties.
- *Relatively small information field:*
  The small fixed-size information field (48 octets) has the advantage of reducing the buffering requirements and hence reducing the queuing delays.

1.2 BISDN Traffic Description

Applications and services that ATM networks are required to transport vary in bit rate from low bit rate ($\leq$ 64 Kbps; e.g. voice and low speed networks) to high bit rate ($\geq$ 100Mbps; e.g. video.)

ATM network design essentially requires understanding of the different types of traffic the network is transporting [13]. A recent classification after the ATM Forum Traffic Management working group is found in [13], where traffic is classified into five service categories:

- *Constant Bit Rate (CBR):*
  CBR services is defined to offer very simple, reliable, and guaranteed communication channels. CBR is intended for real-time applications (audio and video). A fixed bandwidth is assigned to an application source. The application is assumed to generate a continuous cell stream during the connection lifetime.
- *Unspecified Bit Rate (UBR):*
  Applications in this category are alternatively classified as unit-oriented [13]. The object is to transfer a fixed quantity of bits rather than to provide a continuous flow. Thus applications send discrete units of information without any

explicit rate parameter. This behavior is exactly the service model found in the Internet [13].

- *Real-time Variable Bit Rate (rt-VBR):*
  This serves real-time applications where the source rate is allowed to vary [13]. An obvious consequence is that bandwidth is efficiently utilized since real-time application sources are statistically multiplexed. rt-VBR can further be divided into: peak-allocated VBR (PVBR) and statistically multiplexed VBR (SMVBR) [13]. In PVBR, network resources are allocated according to the peak rate but sources generate traffic at variable rates. For SMVBR, network resources are allocated less than the peak rate for each source.
- *Non-real-time Variable Bit Rate (nrt-VBR):*
  The bursty traffic generated in this service category is specified by peak rate, sustainable rate, and loss rate parameters. These parameters are used to allocate resources to each connection. The statistical multiplexing loss rate should conform with the loss rate specified by the connection.
- *Available Bit Rate (ABR):*
  This is a best effort traffic which benefits from the extra bandwidth but has no resource allocation. ABR protocol identifies an optional minimum cell rate which is useful for allocating minimum resources to the connection [13]. Cell loss and delay are therefore improved. Moreover, ABR services use rate-based feedback flow control mechanisms to minimize loss and provide fairness.

The spectrum of applications differ in their requirements. These application can be differentiated in their requirements based on cell loss and cell delay. Figure 1.2 shows the requirements of some of the existing services.

Network resources such as channel bandwidth and switching node buffers are shared among different applications. Different applications enforce different requirements; network resources are allocated accordingly.

It is important to know the traffic requirements before establishing a connection. A connection is accepted or rejected based on a set of traffic descriptors (e.g. peak bit rate, average burst length, etc.) Another traffic descriptor which may alter traffic characteristics is cell delay variation (CDV). It is defined as the variable cell interarrival delay which is different from cell intergeneration time due to variable waiting

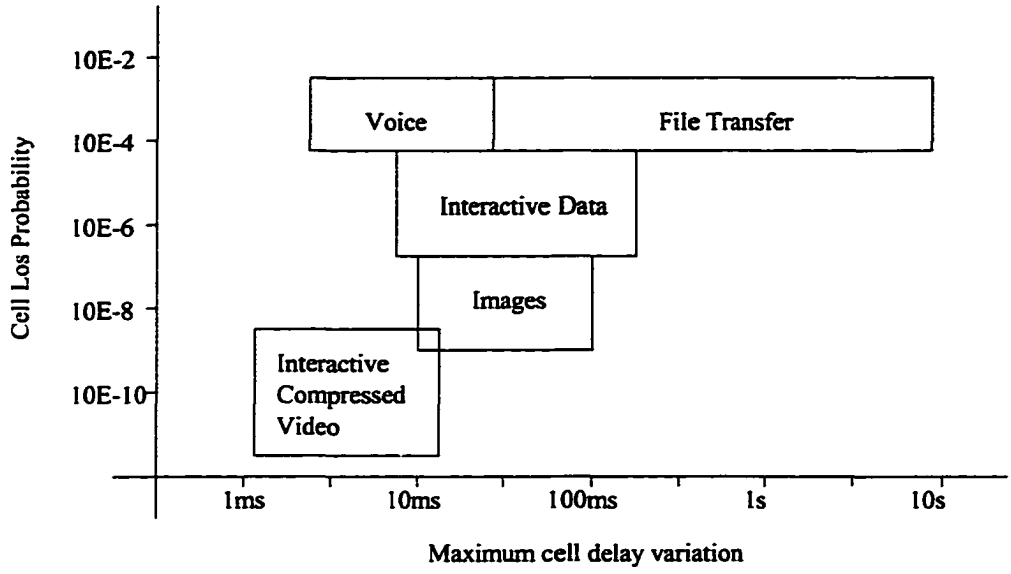

Figure 1.2. *Depending on the application, different requirements are enforced.*

time in buffers [14]. The network uses these traffic descriptors to judge whether resources are sufficient to accept the connection or not. Not complying with this may affect the network performance and ultimately the Quality of Service (QoS) provided.

1.3 Data loss in ATM network

Unlike conventional networks, ATM networks are characterized by low data loss ratios. This is attributable to the fact that low bit error rate characterizes ATM physical media.

However this does not prevent data which is carried in the form of ATM cells to be lost and/or corrupted. There are three factors that significantly contribute to cell loss:

1. Network limited resources. When network resources such as buffers are consumed, incoming traffic is rejected in the form of cell loss.
2. Quality of service non-conformance. Cells that do not satisfy certain QoS parameters such as CDV are tagged and/or dropped. For example, when real-time data cells get delayed beyond the recommended CDV values, they are tagged since user is less interested in receiving them.

3. Traffic contract violation. If a source of data starts sending data at a higher rate than what was negotiated, network policers will respond by tagging and/or dropping excess rate data cells.

Among the above factors, the first is the major cause of cell loss. Overloading a network beyond its capacity is inevitable when traffic with different requirements shares network resources. This is investigated in the next section.

1.4 Congestion in ATM

Network congestion, in general, is the condition where network resources are exceeded by the accumulation of demand [6]. A symptom of congestion is when the number of cells within the network causes the performance (e.g. throughput) to fall off dramatically [9]. Specific to ATM, congestion occurs when offered load from the user to the network approaches or exceeds the network design limits for guaranteeing the QoS agreed upon [6].

The sources of congestion are limited resources such as buffer size, and inherent characteristics of switching nodes such as multiplexing delay (switch service delay).

Generally in a congested switch, low priority cells are rejected once the switch buffer is full. Priority cells are rejected when there are no low priority cells to discard from the queue (buffer). Priority cells may contain critical data such as control information. Losing such cells could affect the QoS provided.

1.4.1 Congestion Control Techniques

ATM networks are expected to carry diverse traffic types ranging from pure data to complex multimedia signals. This traffic diversity forces the network to manage its resources wisely and apply preventative actions such that demand does not exceed available resources (i.e. network congestion). In order to ensure this, congestion control techniques have been proposed. Congestion control is defined as the set of actions reducing the spread and duration of congestion [14]. These techniques can either be reactive or preventive.

Reactive mechanisms, which are sometimes referred to as congestion management mechanisms, attempt to ensure that congestion is never experienced [6]. The general

mechanism is to balance or limit the traffic admitted to the network so as to virtually eliminate congestion.

Preventive mechanisms, on the other hand, referred to as congestion avoidance mechanisms, attempt to avoid severe congestion while admitting more traffic to the network.

In the following section, several congestion avoidance techniques are discussed briefly.

Connection Admission Control (CAC)

When a new connection setup request is received by the network, certain action are taken to accept or deny this connection. The decision is based on the connection anticipated traffic characteristics, the required QoS, and the current network load [14]. While the current network load is estimated from traffic descriptor values of the existing connections, anticipated traffic characteristics are estimated from the traffic descriptor and CDV values.

The objective of CAC is to maintain QoS for existing connections while guaranteeing QoS for the upcoming ones.

Once the connection is accepted sufficient resources are allocated. In any case, the VC connection is accepted if and only if the VC connection is accepted at each VP along the route [14].

CAC provides a protection measure against congestion by minimizing the probability of network congestion [15]. This would mainly contribute to rejecting new connections when low network congestion probabilities are desired. Equivalently, call blocking probability is increased. By call blocking, it is meant that a call cannot be established due to connection rejection or denial. Thus it is obvious that the number of connections admitted are a tradeoff with the level of network congestion expected. In Chapter 2 we show that more connections either CBR or VBR can be admitted into the network while maintaining low congestion levels.

User Parameter Control (UPC) and Network Parameter Control (NPC)

Users are interfaced to ATM networks by traffic contracts. A traffic contract basically consist of traffic descriptors. Violation of this contract may cause network congestion. Therefore, mechanisms to ensure compliance with traffic contract and to enforce the contract terms are needed. UPC and NPC are used for this purpose.

UPC and NPC have similar functions but performed at different interfaces; UPC is performed at the user-network interface (UNI) and NPC is performed at the network-network interface (NNI) [14]. Their functions include: monitoring cell flow, checking traffic descriptors conformity, and taking necessary actions when violation is detected [14].

At UNI(NNI), UPC(NPC) monitors the traffic. If the monitored traffic does not comply with the anticipated traffic (i.e. violation), certain enforcement actions are exercised on violating cells. These actions may either be cell discard or cell tagging.

In cell tagging the CLP field of the ATM cell header is changed to low priority. That is CLP=0 is changed to CLP=1. Tagged cells as well as originally set low priority cells will be discarded at UNI and NNI once congestion is experienced.

Priority Control

The heterogeneous QoS requirements supported by ATM networks require some priority schemes to satisfy this varying requirements. Having CAC being performed at connection setup time, priority controls performed during cell transmission [13].

Connection-level priority and cell-level priority are two priority control schemes [15]. In connection-level priority, connections are prioritized implicitly by their VPI/VCI at setup time [14, 15]. Such priority schemes can be performed as time priority or space priority [14]. In the former case, high priority cells experience shorter queuing delay. In the latter case, when high priority cells are to be queued and buffer is full, low priority cells are discarded. This is to cope with QoS required.

In the cell-level priority scheme, different priority classes are assigned to different cells coming in the same VC connection. These services are distinguished and cells are accordingly buffered at that particular class queue.

Priority mechanisms in a switching node can be performed at three different levels: input port (priority cell assignment), service class buffers (priority cell discard) [14], and output ports (cell scheduling) [13]. These different levels are shown in Figure 1.3. Priority cell assignment mainly assigns prioritized cells to their designated service class. Priority cell discard mechanism composes of discarding cells of the same class when the designated buffers are full and discarding low priority cells of the class if high priority cells are to be pushed into the full buffer.

Scheduling is employed at the output ports of a switch (see Figure 1.3) and de-

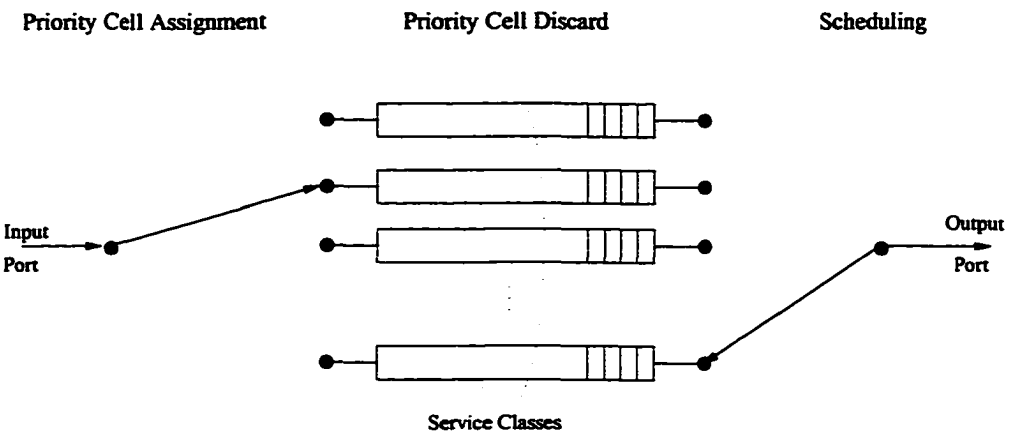

Figure 1.3. *Priority control is done at different levels in a switching node.*

termines which set of waiting cells is to be served next [13]. To provide fairness and attain QoS, scheduling algorithms have been proposed to balance network resources allocation to different users. One such algorithm is weighted fair queuing (WFQ) [13]. Although different variations are being used, the objective is to provide fair queuing implementation.

Explicit Congestion Notification (ECN)

For non-real-time applications, it is difficult to provision and engineer a network with sufficient confidence to ensure a very low probability of congestion, and simultaneously maintaining high network efficiency [15]. Therefore, a control mechanism to reduce the probability of retransmission due to cell loss, thereby preventing network congestion, is required.

ECN in conjunction with flow control mechanisms would modulate the traffic admitted into the network [15]. It has been shown that ECN is efficient in reducing information loss during congestion periods that are at least an order of magnitude larger than the round-trip network propagation delay [15].

Switching nodes are capable of monitoring the traffic flowing through and accordingly a state of congestion can be detected. Upon detecting a congestion, all cells passing through these congested switches are set to indicate a congestion state is being experienced. Explicit forward congestion indication (EFCI) is one such protocol where connection end points are informed of any congestion. This is usually done by setting the payload type identifier (PTI) in the cell header to 010 or 011 [14].

1.5 Using FEC in ATM

It has been shown by Biersack [16, 17, 18], and Zhang *et al.* [19] that FEC coding is efficient for video transmission (i.e. real-time data) over ATM. FEC trades bandwidth for latency to reduce the cell loss rate.

ATM is characterized by protocol simplicity which should be preserved. To ensure that ATM protocol stack is not being effected, FEC can be deployed above or below the ATM stack.

In the former case, an end-to-end error correction schemes residing over the AAL layer are implemented. The objective here is to compensate for data lost due to network errors or congestion. Non-critical real-time data such as video, is an example of this case. Frame control information is of high importance and need to be delivered correctly. Usually control information is short and frequent. Cell loss compensation schemes such as FEC, protect real-time data from loss and hence improve the network performance. Moreover a compensation scheme that can adapt to the QoS requested and the level of congestion is even better. In this thesis we are proposing a novel adaptive FEC scheme for this purpose.

In the latter case (i.e. where FEC is deployed below ATM stack), a hop-by-hop error correction scheme is used. This is feasible in cases where wireless or satellite channels connect ATM switches.

Using FEC in ATM networks is advantageous due to the observation that less network resources are required and less QoS is required by applications [20]. This way QoS is either retained with the option of admitting more connections to the switch, or improved. The impact of using FEC on ATM network resources and QoS is discussed at different spots of this thesis.

An attractive feature of ATM is multicasting, and FEC has been shown to be more appropriate for multicasting than ARQ [21]. Data recipients can compensate for lost cells, This saves the data source from retransmitting dropped cells. Without FEC, sources need to keep track of which recipient did not receive what. Certainly such practice would require sources to store unacknowledged data which means huge memories, and implement sophisticated management strategies.

1.6 Thesis Contribution

In this thesis we are proposing a low cost adaptive FEC scheme based on Reed-Solomon (RS) codes to counteract the impact of CLR on traffic with stringent QoS requirements. Moreover the applicability of this scheme in the wireline ATM networks and the wireless ATM networks is studied.

The performance of AFEC is studied in term of throughput, delay and overhead, and compared to standard ARQ, if applicable.

The proposed AFEC has a unique feature of providing security as well as reliable communications. Network security is an issue specially when traffic is transported over open media. Interestingly enough, the security and the coding operations are done in a single stage as opposed to traditional multistage operation. Moreover the security algorithm proposed in this thesis is shown to be computationally secure.

Multicasting is problematic to ATM communications. Network resources and switch architecture may contribute to availability and/or limitation of the multicasting services especially when ARQ schemes are used in the upper layers. In this thesis we investigate deploying AFEC in an ATM network with multicasting capabilities. The performance of the AFEC in this case is compared to selective repeat protocol (SRP), which is a well-known ARQ scheme.

Also the thesis discusses the possible impact of deploying FEC schemes in ATM network in general and AFEC schemes in particular. Possible switch architectures with FEC capabilities are investigated. In addition, the effects of applying such techniques on resource allocation and management are investigated.

1.7 Thesis outline

This section briefly outlines the remaining chapters. Throughout this thesis, we focus on investigating the applicability of error correcting techniques in ATM networks.

Chapter 2 discusses the different FEC schemes being adopted for ATM. An introduction to RS codes and their error correction properties are also found in this chapter. The chapter concludes with a simple study on the impact of FEC on ATM network resources and QoS. The emphasis is to study buffer requirements as well as the network capacity when FEC coding is deployed.

Chapter 3 presents the new proposed adaptive FEC scheme, which is based on RS codes. The scheme parameters are presented and their impact on network performance is discussed. The chapter then discusses the mathematics of the encoder and the decoder incorporating the new adaptive parameters. An example over GF(2^3) is given to show the operation. A framework for utilizing this scheme is also proposed. Moreover the security issue is discussed. That section starts with a brief introduction on cryptography and the different classification of cryptosystems. The security scheme that to be supported by the proposed FEC coding scheme falls into the category of private key cryptosystems. It has also been shown that the proposed security scheme is computationally secure.

Chapter 4 shows the performance modeling of the proposed scheme. In this chapter we investigate the throughput and end-to-end delay of the proposed FEC when compared with the counterpart ARQ schemes. The analysis is conducted assuming an aggregate traffic load on the network. Moreover monitoring process and the overhead of the proposed FEC is discussed.

Chapter 5 describes a new approach of deploying the proposed adaptive scheme into wireless ATM networks.

The chapter starts with a brief treatment on the directions, issues and challenges of wireless communications in general and wireless ATM in particular. Next the framework of deploying the proposed scheme into the wireless environment is discussed. The chapter then concludes by studying the wireless framework into Rayleigh fading channels.

Chapter 6 shows the results of a simulation study of investigating the advantages of deploying the proposed FEC into ATM networks with multicasting capabilities. The simulation is done using OPNET which is a network simulation tool. The study compared the proposed FEC with ARQ schemes. The performance measures of interest are network resources usage and end-to-end delay. For real-time applications it is crucial to meet the stringent delay requirements for these applications specially when multicasting or broadcasting type of service is required.

Chapter 7 investigates possible ATM switch designs when FEC techniques are used. In addition, resource allocation problem is investigated under FEC environment. One observation is that the number of connections (CBR and VBR) is increased

when FEC schemes are used, maintaining the QoS.

Chapter 8 provides a summary and some conclusive remarks. Suggestions for future work and further investigation are also given.

Chapter 2

FEC Techniques and ATM Networks

ATM is the technology by which diverse user information is to be transported. From the user prospective, the QoS agreed upon at the time of connection setup should be fulfilled during the duration of the connection. For service integrity purposes, sufficient resources (e.g. bandwidth) are allocated for the connection. Thus these resources are to be utilized wisely.

Usually the incoming traffic is multiplexed into one channel. This is essentially the statistical multiplexing defined in ATM networks. More connections, and ultimately more users, are served even if the total bandwidth is more than what the network can handle. Simply, this is due to the fact that users will utilize fraction of the bandwidth offered. In other words, the total average input load from all users is less than or equal to what the networks offers. Users, however at some time periods, may require to input more load than the expected average. As a result, the network is overloaded with user data. Therefore, policies are required either to manage the incoming traffic flows or to compensate for partial traffic loss.

Congestion control schemes have been discussed in the introduction chapter. This chapter discusses cell compensation schemes being proposed for BISDN networks. These schemes mainly utilize FEC codes. So the chapter gives with a brief discussion on Reed-Solomon (RS) codes and their properties. RS codes are chosen to serve two aspects. First, they are widely known for their correction capabilities and relative simplicity. Second, the scheme this thesis is proposing is based on RS codes.

The gains of deploying FEC schemes to ATM networks and QoS are also investigated. This investigation represents the usefulness of FEC to ATM networks. Further

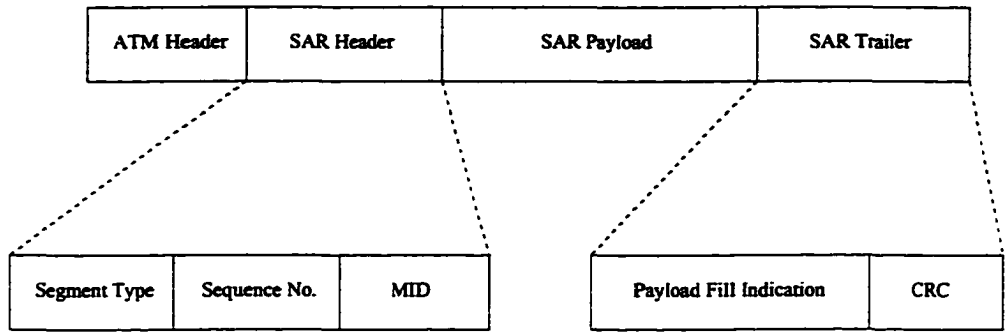

Figure 2.1. *Segmentation and reassembly (SAR) PDU showing MID field.*

studies are found in later chapters of this thesis.

2.1 FEC in B-ISDN

Error correcting codes have been proposed and applied since the emergence of digital communications. One popular class of error correcting codes is Hamming codes [8, 22]. The theory behind Hamming codes, and error correcting codes in general, is to separate information symbols by adding redundant bits to these symbols. The goal is to make the resulting symbols or codes as separably distant as possible.

Since FEC algorithms are expensive to implement in silicon, they are only used where retransmission is costly or impractical such as in satellite communications, remote sensing and navigation, deep-space, and CD players.

Usually FEC schemes are of constant error correcting rate due to high cost in design and manufacturing. Such schemes do not properly utilize the transmission media as the channel condition is changing with time. Adaptive FEC is more appropriate [23, 24].

Dravida and Damodaram [25] are among the first researchers who investigated the application of error correction alternatives for BISDN. The main motivation is that CRC at the ATM cell header is not sufficient to correct misrouted cells. They are not interested in correcting data being lost as much as maximizing correct routing [25]. Therefore, they substantially explored the different possible ways to accomplish their goal by focusing on the MID field of SAR_PDU as shown in Figure 2.1. MID field associates all cells belonging to a given higher layer PDU.

A major conclusion is that cell misdelivery levels are significantly reduced when per-cell AAL CRC is applied [25]. Moreover, regardless of the type of data service supported, cell misdelivery is sufficiently improved.

2.1.1 FEC for ATM communication

Conventional FEC methods generate redundant information to recover lost data at the destination without the need for retransmission. On the other hand, the use of these conventional FECs in ATM communications can result in added processing overhead and reduced throughput which might be undesirable. In this section we review several FEC approaches used for ATM communications.

McAuley [21] proposed an RS erasure technique to guarantee maximum end-to-end delay required by real-time applications and to decrease the effective cell loss ratio. The technique has a constant error correcting capability which increases redundancy and reduces throughput. Although it was mentioned that variable block size and redundancy are desirable, this strategy was not examined.

Cell loss compensation schemes have been introduced by Kitami and Tokizawa [26]. Data streams of different applications are input to the transmitting node which encodes cells according to their virtual path indicator (VPI). Two coding schemes are proposed: one dimensional and two dimensional [26]. In the one dimensional scheme, a single data stream is coded sequentially and then interleaved with other data streams. In the two dimensional scheme, cells of the data stream is arranged in a matrix form. Dedicated cells at the bottom of each column are used to hold the parity information for error control purposes. The parity is generated by the modulo-2 operation.

At the receiving node the streams of cells are decoded. Lost cells are recovered by FEC decoders. On transmission, data streams are interleaved to distribute cell loss over the different traffic streams. This is shown to produce satisfactory performance results of error correction and throughput [26].

Ohta and Kitami [1, 27] proposed node-to-node and end-to-end FEC approaches. In the node-to-node approach, losses were detected at virtual path (VP) terminating nodes. The issue of cell detection was also addressed. In the end-to-end approach, the ATM adaptation layer (AAL) sequence number (SN) is used to detect lost cells.

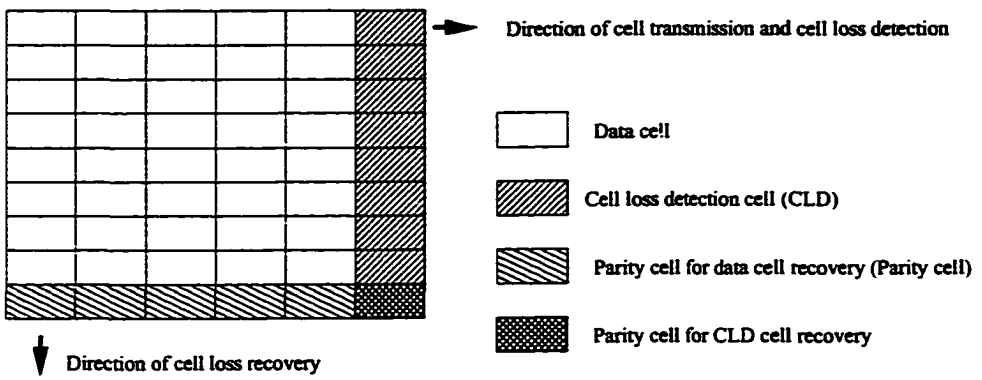

Figure 2.2. *Coding matrix structure used in [1].*

Using SN in the other scheme limits the block of data to be coded to less than 16 cells. This decreases the effective throughput since the ratio of redundant information to data is relatively high.

In the node-to-node approach, the data cells are gathered in a matrix and each row is terminated with cell loss detection cell (CLD) as shown in Figure 2.2 [1].

The purpose of CLD is to housekeep the data cells in a row. Each column in the matrix is terminated with a parity cell. The parity coding used is similar to the one used by Kitami and Tokizawa (i.e. modulo-2 addition). Single errors can be detected and corrected. Therefore a burst error of length at most equal to row length is allowed in order to maintain the integrity of the service. Longer burst errors produce double errors in a column which is not detected by the parity cell and, thus, cannot be corrected.

The maximum coding efficiency is achieved when row size is 17 [27]. Correction power, however, may be increased by adding more parity cells at the end of each column [27]. By doing so, the allowable number of cells being lost in each column is increased. This can lead to inefficient utilization of bandwidth and network resources since coding overhead is increasing accordingly. To efficiently utilize the network resources, suitable error coding schemes are needed. These codes generate minimal overhead for maximum correcting power and bandwidth utilization. This is shown in Chapter 3.

The drawback of this scheme is that it requires huge memory to store data cells in order to calculate CLD and parity cells for node-to-node scheme. Moreover, the tech-

nique, with both approaches, does not consider the conditions at which the network is operating (i.e. congestion and noise).

Ayanoglu *et al.* [28, 29, 30] proposed a two-level FEC scheme. Their approach is very similar to the approach devised by Ohta and Kitami except they used RS codes instead on parity coding. This approach although shows high tolerance to noise and cell loss, it is computationally complex and costly.

Shacham [31] addressed the problem of buffer management. He suggested that the switch (multiplexing node) need to be used to evenly distribute cell rejection over the data blocks passing through it. Then fixed-rate FEC coding can be designed to correct the average number of erasures. Interleaving is shown to equivalently produce the same effect [31].

2.2 Reed-Solomon Codes

(n, k) Q-ary Reed-Solomon (RS) block codes are symbol error correcting codes with symbol size $q = \log_2 Q$ bits and block length $n = Q - 1$[32]. Each group of k information symbols input to the encoder is transformed into an n-symbol codeword. RS codes are optimal (i.e. maximum distance separable or MDS code), and so have minimum distance $d_{min} = n - k + 1$. An RS code can correct up to

$$t = \lfloor (d_{min} - 1)/2 \rfloor = \lfloor (n - k)/2 \rfloor$$

symbol errors, where $\lfloor x \rfloor$ is defined as the largest integer less than or equal to x. In addition, up to e errors and s erasures can be corrected where $2e + s = n - k$. This means that with erasures only correction, $n - k$ errors can be corrected, if the error locations are known. An important property of RS codes is that shortening, puncturing and extending (within limits) results in another MDS code.

Bounded distance decoding of the RS codes is assumed in this thesis. This can be achieved with the Berlekamp-Massey or Euclidean algorithms[33]. In this case, decoding is successful only if the received word lies within the decoding sphere of a valid codeword. Otherwise, the decoder reports a decoding failure. Both algorithms employ an iterative technique to first find the error locations, then the values of the code word symbols at these locations. For ATM communications, erasures-only decoding is employed, so that only the symbol values at the erased positions need

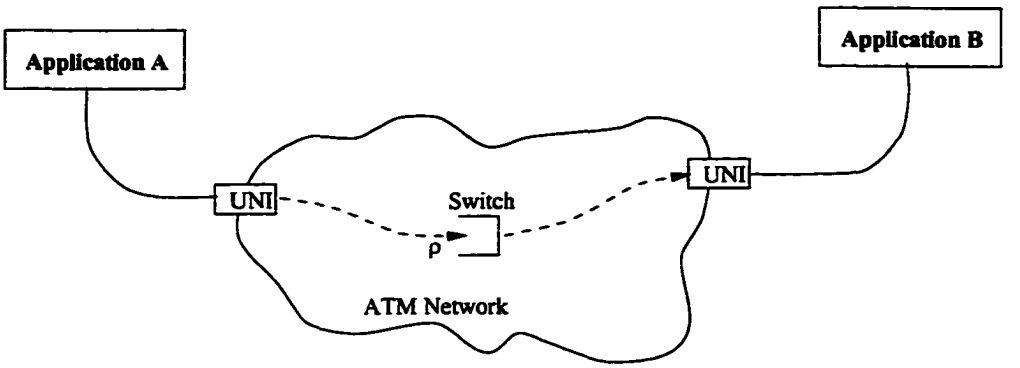

Figure 2.3. *Configuration assumed to study impact of FEC schemes on ATM networks.*

be computed. Erasures are more suitable to ATM since data loss is mainly due to congestion. ATM transmission media, mainly fibers, are generally characterized by a very low bit error rate, which does not contribute significantly to data loss.

2.3 Impact of Error Correction Codes on ATM Network Resources and Quality of Service

Studies such as [17, 18] show that FEC is advantageous when deployed in ATM networks. However none has shown the impact of using such schemes on ATM network resources and quality of service. In this section, a simple approach to analytically study the impact of Reed-Solomon (RS) codes on ATM network resources is attempted.

2.3.1 Preliminaries and Modeling

Using FEC schemes in ATM networks has potential impact on network resources. In order to study the gains to ATM networks in general and ATM switch design in particular due to FEC deployment, the configuration in Figure 2.3 is assumed.

A connection is established between points A and B which may traverse one or more switches. In order to study the impact of traffic that is FEC encoded on ATM network resources, we consider one switch (refer to Fig. 2.3) with buffer size B. The

switch receives a load, called offered load (ρ), which is defined as the average fraction of an input link bandwidth that is used.

For FEC schemes we assume RS erasure codes which are different from the classical RS codes in that the location of errors are known and hence no attempt to find them is required [21]. Let us assume an (n,k) RS erasure only coding scheme in GF(2^8). This implies that the symbol length is 8 bits or an octet. In this case a code word can still be recovered if $n-k$ or less symbols are lost. Since data in ATM networks are lost in cells or bursts of 48 octets, n and k are chosen such that $n-k$ is an integer multiple of 48 octets. This guarantees that lost cells can be recovered.

To transmit the higher layer data (i.e. application data), we assume a stream of data to be delivered continuously to the encoder. The encoder then gathers the data into blocks of size k. In the case where no FEC is used, each data block will require $K = \lceil k/48 \rceil$ cells and will have a block error probability ($BEP_{uncoded}$) of:

$$BEP_{uncoded} = 1 - (1 - CLR)^K. \tag{2.1}$$

On the other hand, if FEC is used then each of the data blocks of size k is encoded with the RS encoder. The coded block is now of length n and will require N ATM cells where $N = \lceil n/48 \rceil$. The probability of the encoded block being in error upon reception (BEP_{coded}) is defined as the probability of losing $R = \lceil \frac{n-k}{48} \rceil + 1$ or more cells:

$$BEP_{coded} = \sum_{i=R}^{N} \binom{N}{i} CLR^i (1 - CLR)^{N-i}. \tag{2.2}$$

where CLR is cell loss ratio.

For the purpose of this study, a (255,207) RS FEC scheme is assumed. This means that redundant symbols occupy one cell only. The corresponding probability functions as described in Eq. (2.2) and Eq. (2.1), are plotted in Figure 2.4. The figure depicts that a lower error probability is attained when FEC is deployed. Alternatively, a block of data cells with actual CLR is delivered at a much lower apparent CLR when FEC is used. This plot serves as a reference throughout this chapter.

In what follows we investigate the major effects of using FEC schemes on ATM switch resources using the above assumptions. Using FEC schemes potentially reduces

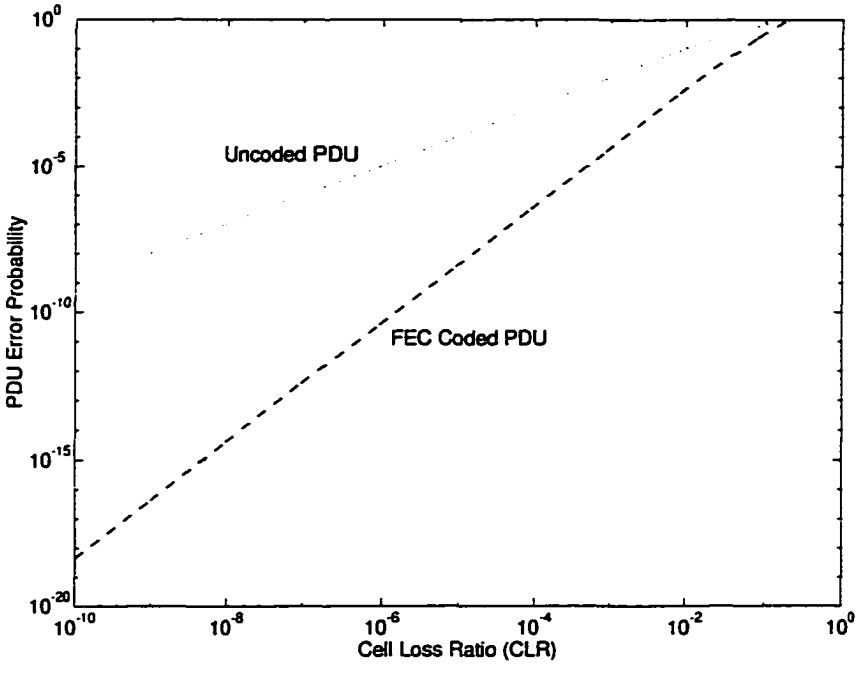

Figure 2.4. *Block error probabilities for uncoded and FEC coded blocks. FEC coded blocks have lower error rates than uncoded blocks.*

ATM switch resources requirements. The interest is in envisioning the impact of deploying FEC on switch buffer size and capacity. Capacity is defined in terms of the maximum number of connections that can be established in a switch.

2.3.2 Resource allocation requirements

It is crucial to study the impact of deploying FEC on switch resources. One important switch resource is the buffer size. Regardless of the buffering strategy followed, limited buffer sizes ultimately lead to congestion and cell loss. One way to overcome this congestion problem is to use large buffers.

One way to study the impact of FEC on switch buffer requirements is to assume the configuration of Figure 2.3. In addition we assume that a connection generates a traffic load (ρ) and requests a desired CLR. Then buffer requirements, B, in the switch are approximated by

$$B = D \times \frac{\log(CLR)}{\log(\rho)} \tag{2.3}$$

where D is the PDU burst size in cells [6]. D equals $\lceil \frac{n}{48} \rceil$ for FEC, otherwise it equals $\lceil \frac{k}{48} \rceil$.

The objective CLR is derived from Figure 2.4. For a specific BEP (y axis), which corresponds to a *desired* CLR, the corresponding CLR is derived (x axis) that we will refer to as *actual* CLR. When establishing a connection, actual CLR value is negotiated as extended QoS parameter [34]. In this case, acceptable forward and backward CLR parameters of the calling user are specified. CLR is expressed as an order of magnitude m, where CLR takes the value 10^{-m} [34].

Using the mapping strategy described above and Eq. (2.3), buffer size requirements are shown in Figures 2.5 when desired CLR equals 10^{-10}. In this case actual CLR are 10^{-10} for uncoded PDUs and 10^{-5} for FEC encoded PDUs. Similarly when the desired CLR is about 10^{-5}, buffer requirements are shown in Figure 2.6. Actual CLR for uncoded PDU is 10^{-5} and that for FEC encoded is 10^{-3}. The two figures indicate that for a given offered load, buffer requirement generally increases as the objective CLR decreases. This is expected since traffic with lower CLR guarantee (i.e. higher m) imposes more stringent requirements and hence more buffer spaces are required to store the incoming cells.

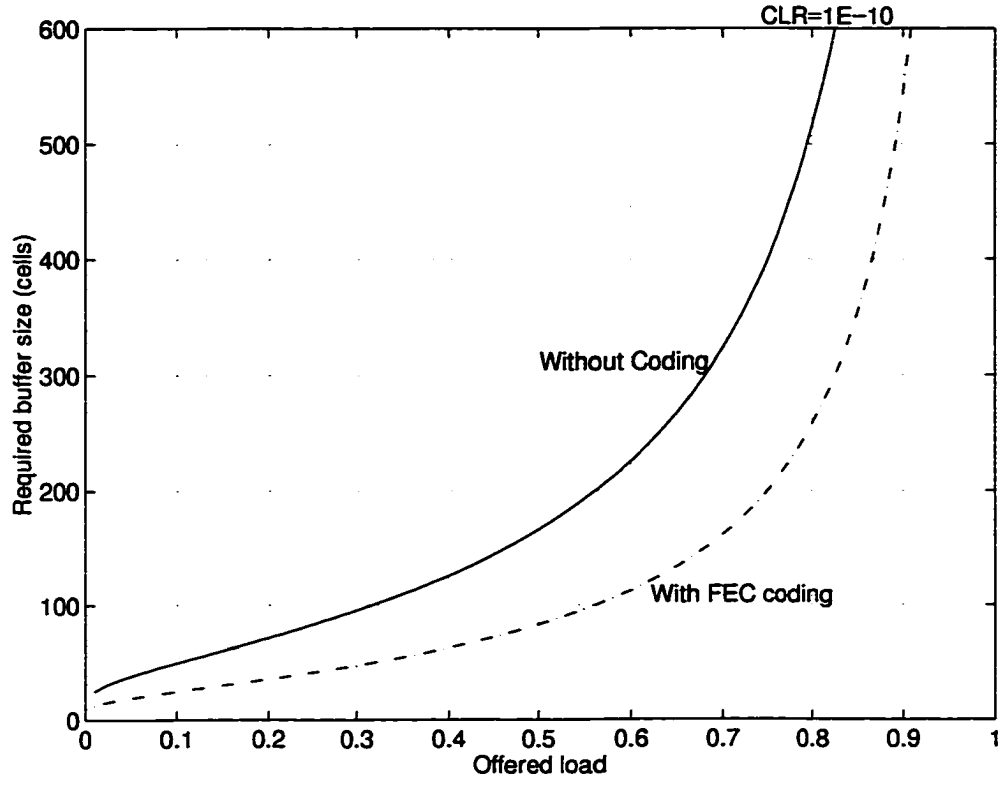

Figure 2.5. *The impact of using FEC on switch buffer size at cell loss ratio (CLR)* 1×10^{-10}.

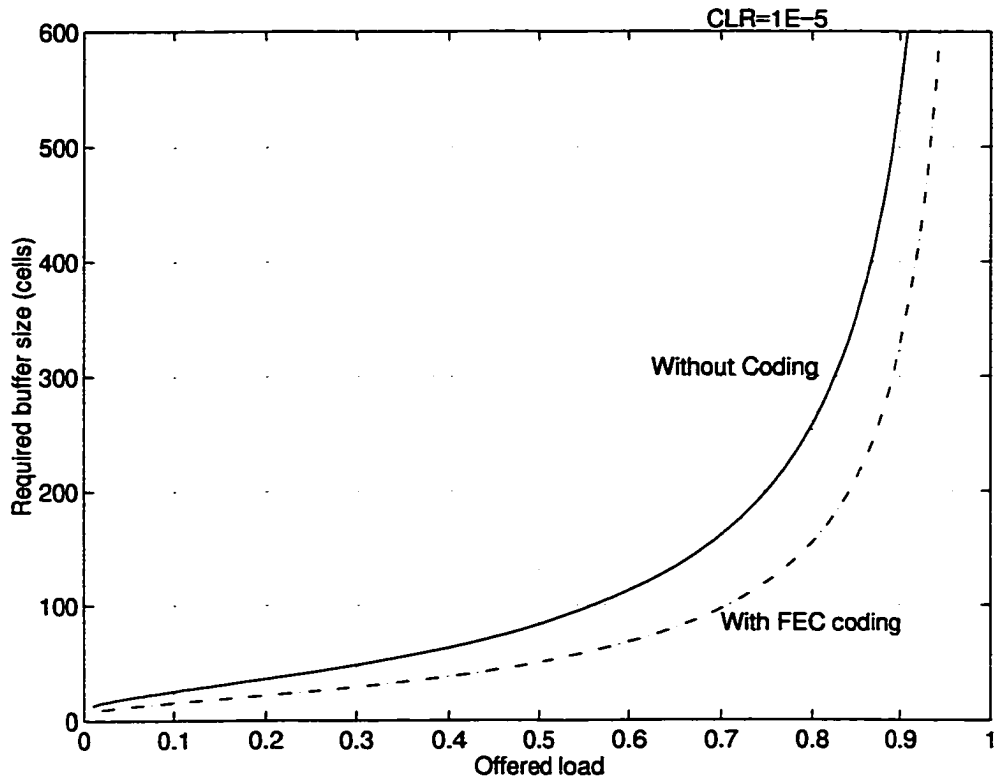

Figure 2.6. *The impact of using FEC on switch buffer size at cell loss ratio (CLR)* 1×10^{-5}.

When using FEC schemes, buffer spaces required by ATM switches at a given offered load can be made limited. The sizes of these buffers are less than or equal to the sizes needed by the worst CLR that might be encountered. This is a valid argument since FEC schemes are capable of regenerating a limited number of lost cells.

As a conclusion, deployment of FEC schemes relaxes CLR requirements and hence smaller buffering spaces could be engineered. This effect has double benefit when compared to a switch of similar buffer size without FEC. First, more buffer space is allocated to other connections to improve their QoS. Second, CLR is improved so that connection admission control (CAC) mechanism may admit more connections (see Section 2.3.3).

To study the saving in buffer sizes due to the deployment of FEC schemes, let us assume B' to be the buffer size corresponding to the case where FEC traffic with actual CLR' is entering the switch and B is the buffer size when no FEC is used with actual CLR. D and ρ are the same in both cases in Eq. (2.3), thus we obtain:

$$\frac{B}{B'} = \frac{\log CLR}{\log CLR'}. \tag{2.4}$$

Essentially buffer size required by FEC traffic is less than non-FEC traffic. Buffer saving, which is calculated as $\frac{B-B'}{B}$, is shown in Figure 2.7. When there is no coding more buffer is required to maintain QoS as CLR increases. However we can save upto 80% of buffer sizes when FEC is deployed and still maintain the same QoS for all established connections.

2.3.3 Capacity

ATM switches generally are characterized by limited resources mainly buffers. Incoming connection requests to the switch are accepted under the condition that there is no QoS degradation to existing connections and the QoS of those incoming connections is guaranteed. For example, when a new connection request is received, CLR is extrapolated [14]. The new CLR value is compared against an objective CLR. A connection is accepted if the new CLR is less than or equal to the objective CLR, else the connection is rejected.

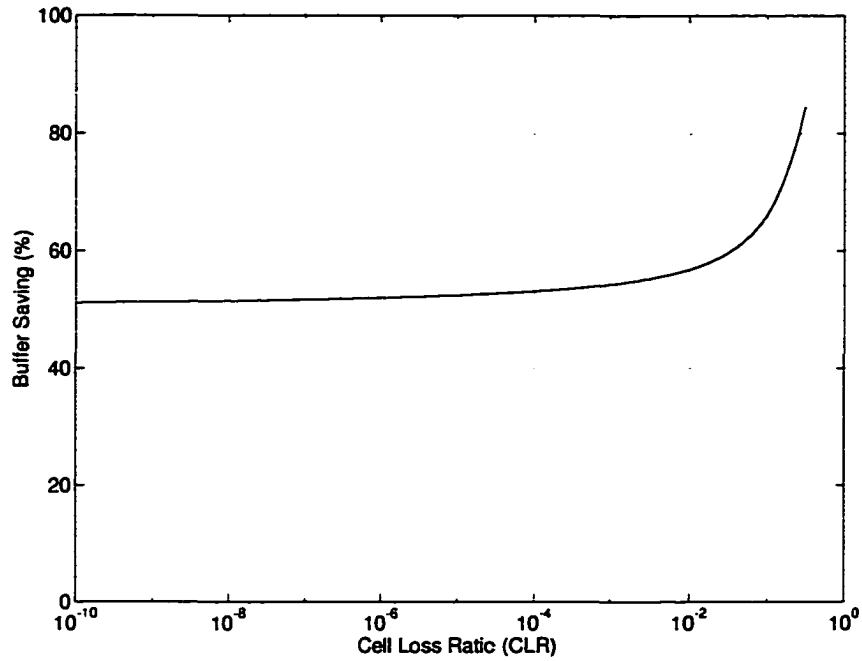

Figure 2.7. *Percentage buffer sizes saved when FEC is used at different CLR values.*

In the previous section we showed that a connection with specific load will require less buffer spaces when FEC is used. It is expected then that FEC schemes might add the advantage of accepting more connections without exceeding the switch CLR objective. To validate this, two cases with two different connection types namely; constant bit rate (CBR) connections and variable bit rate (VBR) connection are studied analytically. The emphasis of the study is to show how many connections of the above types separately the switch can handle based on the switch buffer space available.

Case 1: CBR connections: CLR for a switch with buffer size B and C CBR connections is approximated by:

$$CLR \approx e^{[-2B(1-\rho)-\frac{2B^2}{C}]} \tag{2.5}$$

where ρ is the offered load [6, 35].

Solving Eq. (2.5) for the number of CBR connections, C, that can be admitted given an objective CLR, gives:

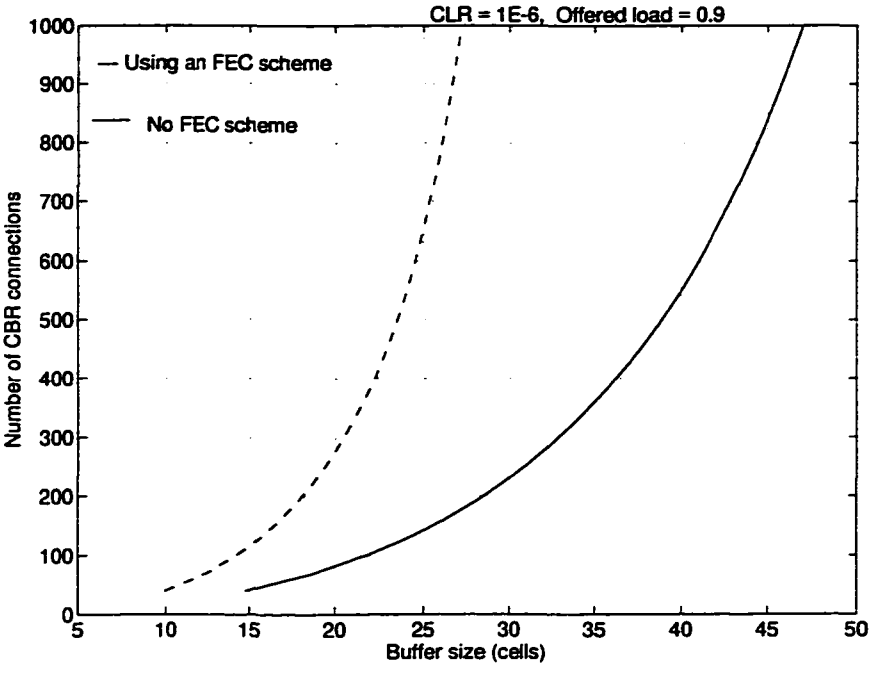

Figure 2.8. *Number of CBR connections under different buffer sizes with 90% offered load at CLR of* 1×10^{-6}.

$$C \approx \frac{2B^2}{-\ln CLR - 2B(1-\rho)} \tag{2.6}$$

Figures 2.8 and 2.9 plot the average number of CBR connections admitted into a switch with different buffer sizes B at two values of offered load 0.9 and 0.7, respectively. CLR value is picked such that it reflects voice traffic characteristics; the value used is 10^{-6}.

The above figures show that more CBR connection would be admitted into a switch if FEC schemes are used given a fixed buffer size.

Case 2: VBR connections: The analytical model that relates the average number of VBR connections, n, to buffer sizes, is derived from [6]. The model is derived from the fact that VBR traffic distribution can be modeled using a normal approximation to the binomial distribution for V VBR sources with an average activity per source of ρ [6]. The required buffer size is approximated by:

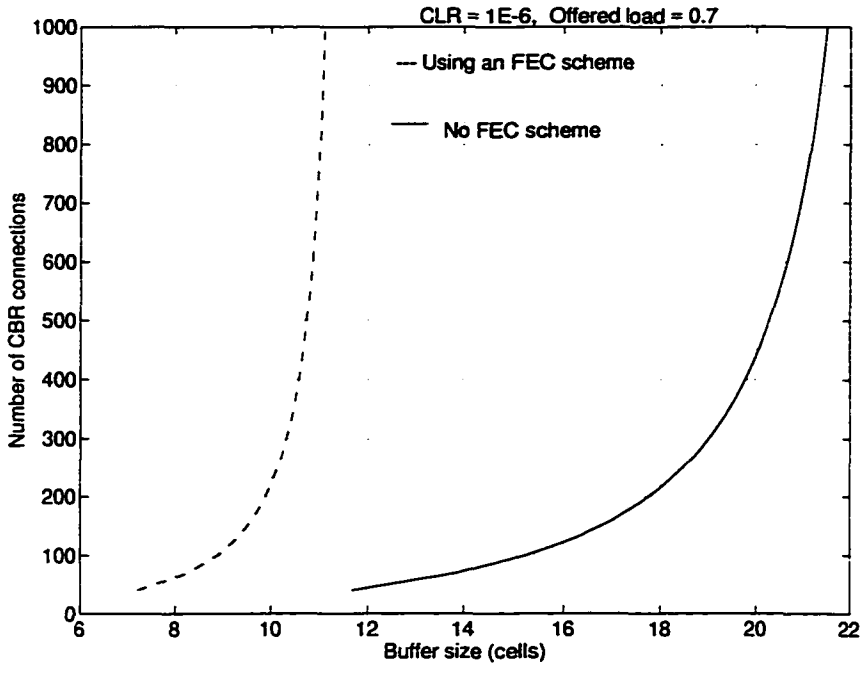

Figure 2.9. *Number of CBR connections under different buffer sizes with 70% offered load at CLR of* 1×10^{-6}.

$$B \approx V\rho + \alpha\sqrt{V\rho(1-\rho)} \tag{2.7}$$

where

$$\alpha = \sqrt{-2\ln(2 \times CLR)}. \tag{2.8}$$

Solving Eq. (2.7) for V gives:

$$V \approx \frac{\gamma - 2B + \alpha^2(1-\rho)}{2\rho} \tag{2.9}$$

where

$$\gamma = \sqrt{4B^2 + \alpha^4(1-\rho)^2 + 4B\alpha^2(1-\rho) - 4}. \tag{2.10}$$

The number of VBR connections to be admitted is solely based on buffer spaces available given that CLR is fixed. Figure 2.10 shows the number of VBR connections handled when FEC schemes are used for different buffer sizes and source activities ρ. A specified CLR objective is again picked to reflect data type traffic characteristics.

It is needless to mention that using FEC schemes potentially increase the number of VBR connections to be admitted into an ATM switch without requiring more buffer sizes. Equivalently FEC guarantees VBR traffic QoS for a fraction of the buffer size.

2.4 Concluding Remarks

Different FEC schemes had been proposed in the literature. The only drawback with these schemes is their constant correction rate. Due to the nature of errors in ATM communications, it is essential to have adaptive correction schemes especially to be used with real-time applications where delay is crucial.

Moreover the impact of deploying FEC schemes on ATM network resources has been studied. The analysis conducted in this chapter shows that it is advantageous to consider FEC, and the key benefits are:

1. Significant reduction in buffer sizes even at low CLR. Almost 50% buffer space saving is achieved when FEC is used at a CLR range of 10^{-10} to 10^{-4}. At higher CLR values greater savings are observed.

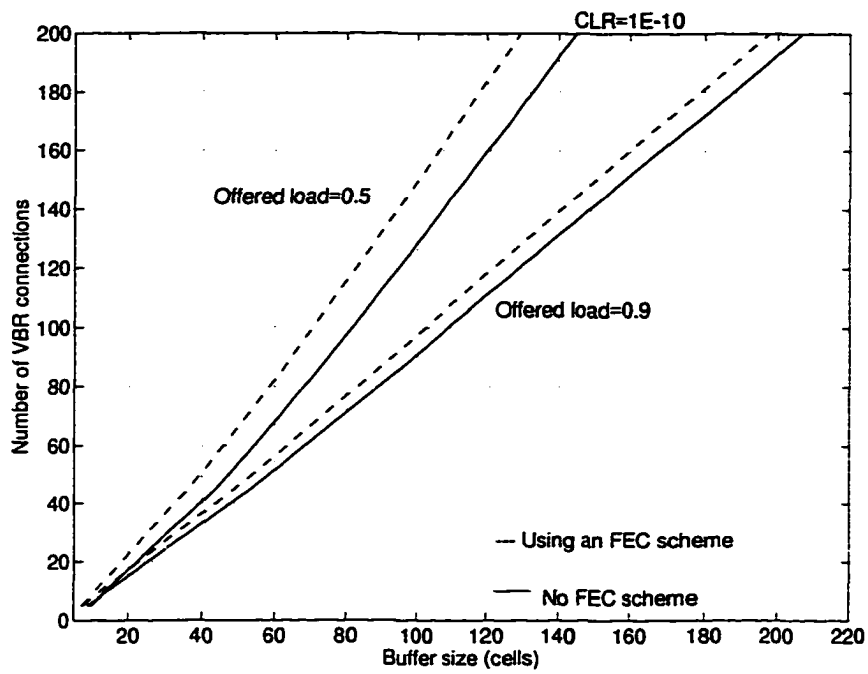

Figure 2.10. *Number of VBR connections that a switch can handle under different buffer sizes at CLR of* 1×10^{-10}.

2. More connections to be admitted into the switch without deterioration in the QoS. For CBR traffic it has been observed that for the same number of connections half the buffer space is needed when FEC is deployed. However not as much enhancement is to be achieved when VBR traffic is assumed. One possible reason is that the behavior of VBR traffic is not as predictable as CBR and thus the formulas used to describe it is much of an estimation.
3. FEC can be applied at different points in the ATM communication infrastructure. Some services, such as real-time video over VBR, would require FEC at the UNIs. Other services such as CBR can utilize the power of FEC at any point in the network.

Chapter 3

An Adaptive FEC based on RS Codes

Chapter 1 listed some of the congestion control mechanisms used in ATM networks. Chapter 2 showed some of the effort invested in applying FEC codes and techniques to ATM communication. Different approaches had been proposed at different levels of the BISDN/ATM model [25, 26, 1, 27].

In this chapter, we propose a novel adaptive FEC scheme based on RS codes. We further investigate its applicability and operation. Adaptive FEC codes have been used intensively in satellite communications. References [23, 24] have more details in this regard. This chapter however contains a new proposal for a framework utilizing adaptive FEC codes.

The adaptive FEC scheme proposed in this chapter is based on RS codes introduced in Chapter 2. The detailed scheme is explained in Section 3.1. The scheme theory is given in Section 3.2. In addition the encoding and decoding procedures are described mathematically in GF(2^m). Based on the mathematical results obtained, an application framework is proposed in Section 3.3. The framework consists of a protocol that has to be executed by communicating entities. A key feature of the proposed scheme is security. This feature is thoroughly discussed in Section 3.4.

3.1 The Novel Adaptive Scheme

In the classical (n, k) RS codes, $n-k$ redundant symbols are introduced to the original k symbols. These $n-k$ symbols protect the data from loss and the RS CODECs will be capable of recovering up to $n-k$ symbols in the resulting code word of length n.

Therefore the correction rate is constant. These $n - k$ symbols are always appended to the data symbols even if there are no errors. This means that channel bandwidth is sometimes wasted.

In ATM networks, cell loss can happen due to several factors as described in Chapter 2. Moreover the state at which the network is causing this cell loss is changing. Therefore it is more suitable to have an adaptive scheme after which only lost cells are regenerated. The motivation here is to achieve higher bandwidth utilization by reducing the redundancy in the code word. One possible scheme is to have the flexibility of varying k. Although this is theoretically possible it might be practically not feasible. The design of such RS decoders may be very costly.

Alternatively we introduce an additional coding parameter, l, which denotes the number of information symbols punctured (i.e. set to zero or any other value agreed upon) from the original (n, k) RS code. The main function of l is to increase the error correcting rate by zeroing l of the k information symbols. The parameter l can also be used for data security as will be discussed in Section 3.4.

For ATM communication deployment, RS encoding and decoding are done on cells. Figure 3.1 shows the symbol distribution of the new coding technique. $k - l$ cells are to be encoded and sent together. One byte of the cell payload is used for sequencing since the RS CODEC assumes knowledge of erasure location. The symbols are interleaved over the cell length.

This approach has two advantages. First, it creates a wider or more versatile range of error correcting capabilities. A conventional (n, k) RS code has the ability to correct up to $n - k$ erasures. Zeroing l of the k information symbols increases the error correcting capability. For example, with a (255,201) RS code over GF(2^8), 54 symbol erasures out of 255 symbols (i.e. 20%) can be corrected. If $l = 100$ (i.e. 100 symbols are set to zero), then 54 erasures out of 155 symbols (i.e. 30%) can be corrected. Figure 3.2 shows the percentage of erasures that can occur in a (255,223,l) RS code and still be recovered. Conventional techniques have a constant rate, while correction capability of the proposed technique is variable. Hence the proposed scheme has more flexibility over the conventional schemes.

Second, using a conventional RS code requires that n cells be transmitted across the communication channel. This significantly contributes to the ATM congestion

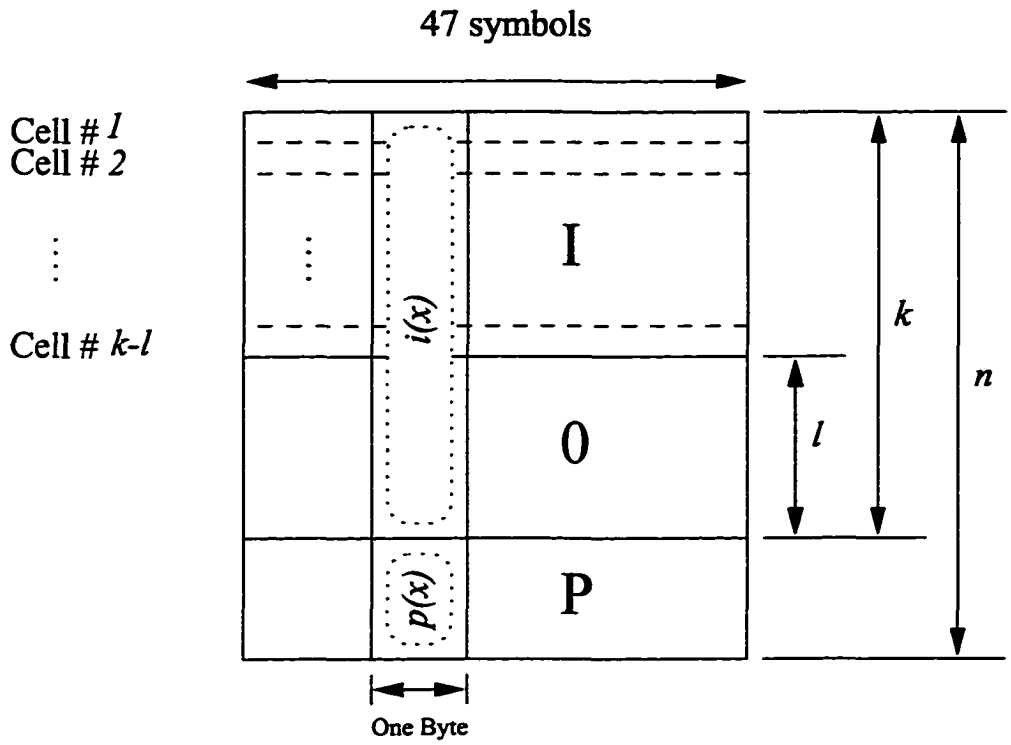

Figure 3.1. *The adaptive (n,k,l) RS code over GF(2^8) in ATM communications. The first byte of each cell is used as a sequence number. $(k-l) \times 47$ data bytes (**I**) and $l \times 47$ zero bytes (**0**) are used to generate $(n-k) \times 47$ parity bytes (**P**). The resultant $n-l$ cells of size 53 bytes are interleaved and transmitted.*

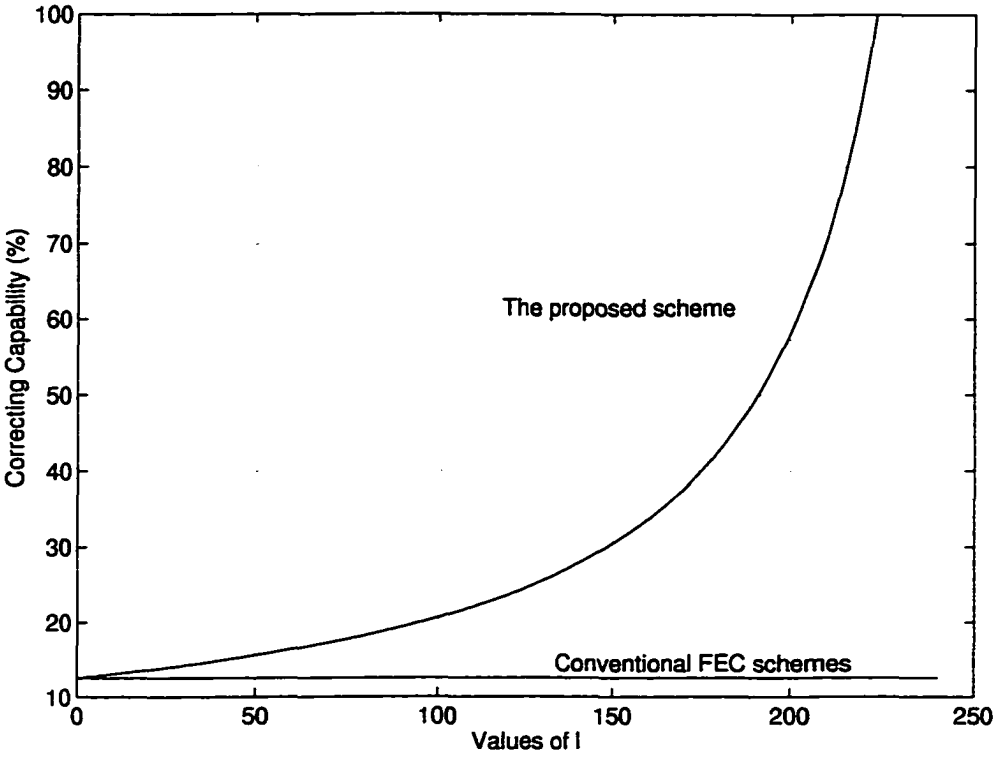

Figure 3.2. *Comparison of the erasure correction capability of a conventional (255,223) RS code and a versatile (255,223,l) RS code.*

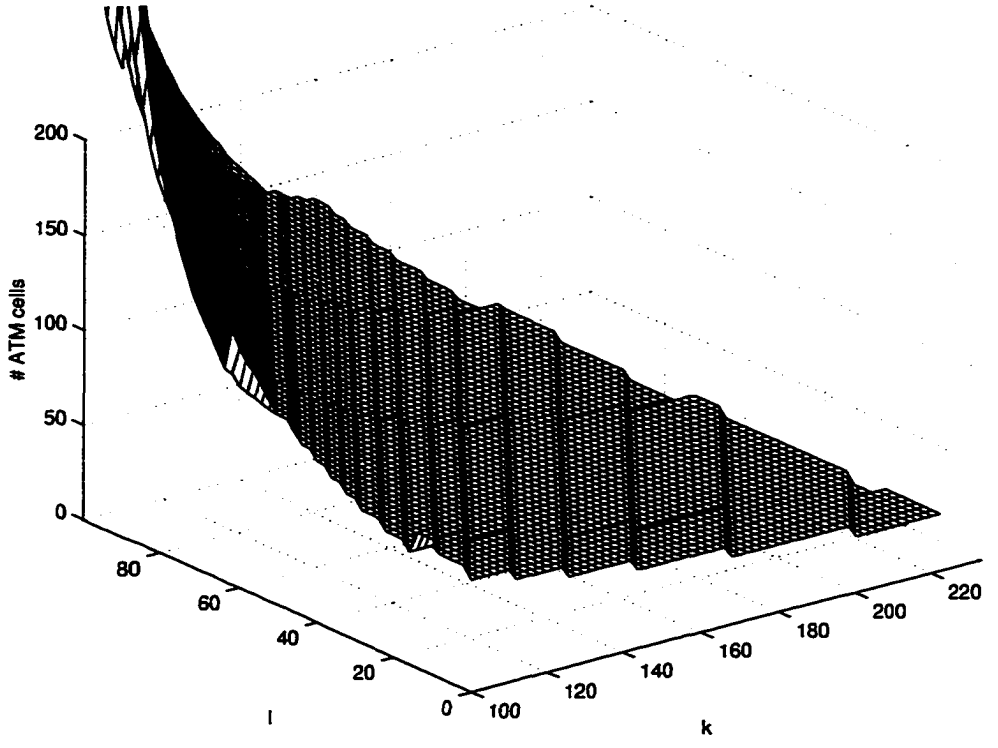

Figure 3.3. *The number of ATM cells needed to carry a PDU of size 1000 octets using AFEC (255,k,l) RS code, where* $100 \leq k \leq 233$, *and* $0 \leq l < k$.

problem. On the other hand, the new technique requires that only $n - l$ cells be transmitted, since the zeroed symbols can simply be recreated at the receiver. A PDU of size D will require a number of ATM cells to carry it. The number of ATM cells in this case, assuming one byte of the cell payload is used for sequencing, can be expressed by:

$$\#\ \text{ATM cells} = \lceil \frac{D}{k-l} \rceil \lceil \frac{n-l}{47} \rceil$$

Figure 3.3 depicts the number of ATM cells needed to transfer a PDU of size 1000 octets using AFEC (255, k, l) where $100 \leq k \leq 233$, and $0 \leq l < k$. The figure shows that there exists a range of values for k and l that satisfies the criterion of carrying a PDU over certain number of ATM cells. It can easily be argued that for a specific quality of service and network congestion, k and l can be chosen from that range to satisfy the network conditions.

For a given value of l, Figure 3.4 shows the number of cells needed to carry a

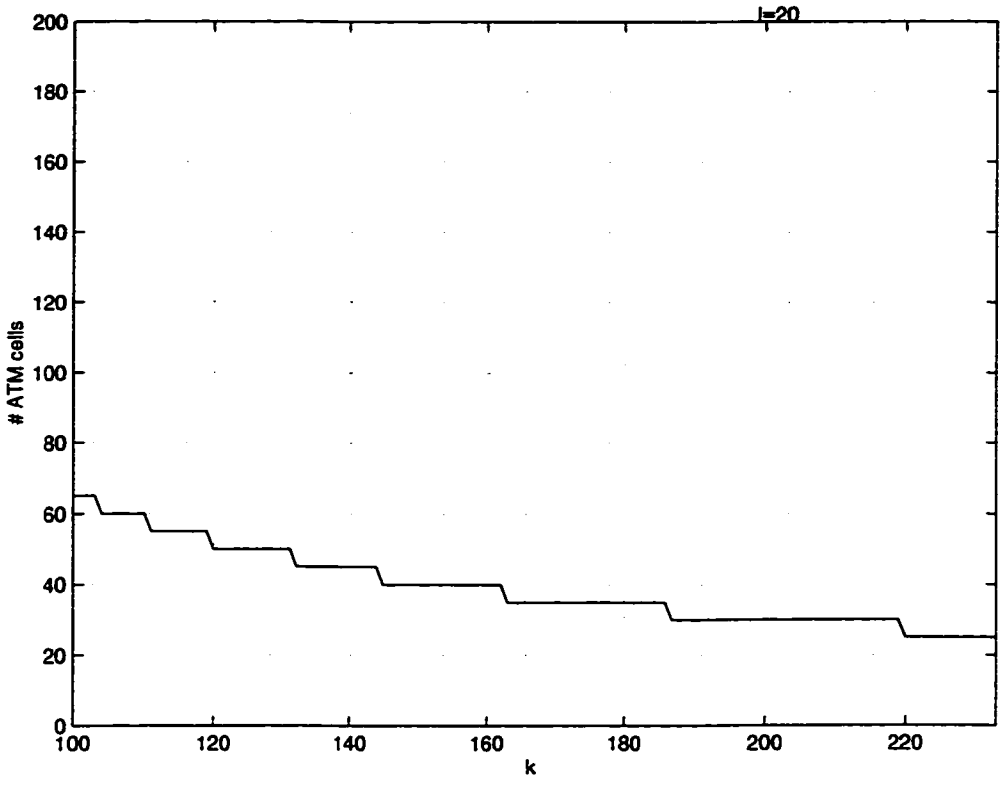

Figure 3.4. *The impact of varying k on the number of ATM cells to carry a PDU of size 1000 when* $l = 20$.

PDU of size 1000 octets when k varies. The stair shaped curve is due to the ceiling function effect of dividing the PDU over ATM cells. The last cell may carry padding information

On the other hand, for a given value of k, Figure 3.5 shows the number of cells needed to carry a PDU of size 1000 octets when l varies. One could observe that as l reaches k the number of cells increases quickly. This behavior may not be desired after certain limit since coding efficiency may be low and therefore low throughput may occur. This suggests that an upper bound on the value of l could be used.

Finally, the end-to-end delay associated with transmitting a group of cells is reduced, depending on the level of quality and reliability requested. Further discussion will be given in Chapter 4.

Before we proceed any further let us explore the gain of using AFEC over classical FEC schemes. The lemma below gives quantitatively the increase in correction rate

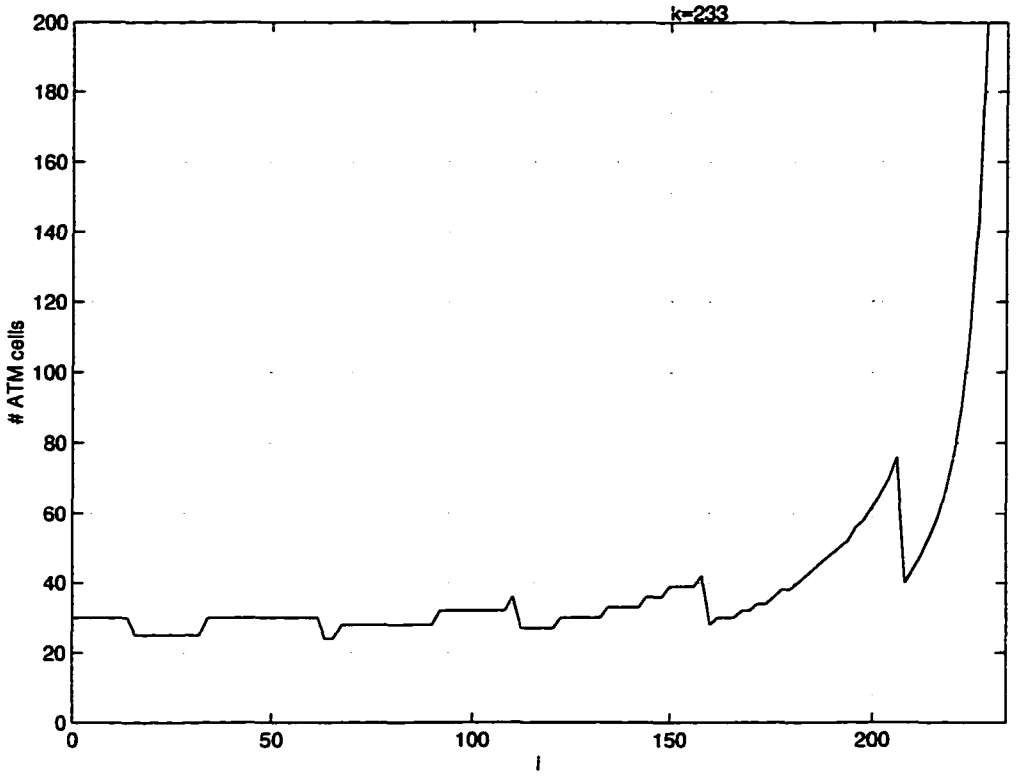

Figure 3.5. *The impact of varying l on the number of ATM cells to carry a PDU of size 1000 when $k = 233$.*

due to AFEC deployment on a traffic stream.

Lemma 1 *The adaptive (n, k, l) RS scheme increases the correction rate by a factor of $l/(n-l)$ when compared to the classical RS FEC.*

Proof. In the classical RS scheme the correction rate (R) is defined as:

$$R = \frac{n-k}{n}. \tag{3.1}$$

Now when the adaptive RS is used, the number of symbols is reduced by l. The correction rate (R') is given by:

$$R' = \frac{n-k}{n-l}. \tag{3.2}$$

The increase in correction rate (δ) is found to be:

$$\delta = \frac{n-k}{n-l} - \frac{n-k}{n}. \tag{3.3}$$

After some manipulation, we have the increase in correction rate to be equal to:

$$\delta = \frac{l}{n-l}\frac{n-k}{n}. \tag{3.4}$$

Hence the correction rate is increased by a factor of $l/(n-l)$.

■

The corollary of the above lemma is that a traffic stream that uses AFEC can tolerate to losing up to $l/(n-l)$ percent of its traffic, when compared to the same traffic but using FEC, without losing the level of QoS.

In the rest of this thesis, the versatile or adaptive RS code is denoted by the parameters (n,k,l). In this context sometimes we will refer to this scheme by AFEC standing for adaptive FEC scheme. The parameters k and l are part of the QoS negotiated between entities[1]. Section 3.3 describes how these parameters are negotiated.

[1]The parameter n is chosen to be constant for simplicity.

3.2 Adaptive (n,k,l) RS Erasure Correcting Code

One advantage of this adaptive coding technique is its low design and operational complexities. This is essentially due to the fact that the sender and receiver have comparable hardware complexity. The coding parameters, k and l, are set according to the QoS negotiated. Although RS decoding is a complex algebraic operation, using only erasure correction greatly simplifies the decoding algorithm. It is also assumed that the code word length n is fixed, so that the same encoder and decoder can be used throughout the network. In this work a similar approach to the one described in [21] is followed.

In the general case, consider the $k-l$ information symbols to be transmitted. These can be represented by a polynomial of degree $k-l-1$

$$i(x) = i_0 + i_1 x + \ldots + i_{k-l-2}x^{k-l-2} + i_{k-l-1}x^{k-l-1}$$

The code word is constructed in a systematic form so that it can be represented by the following polynomial of degree $n-1$

$$c(x) = \underbrace{c_0 + c_1 x + \ldots + c_{k-l-1}x^{k-l-1}}_{I} + \underbrace{c_{k-l}x^{k-l} + \ldots + c_{k-1}x^{k-1}}_{O} + \underbrace{c_k x^k + \ldots + c_{n-1}x^{n-1}}_{P} \tag{3.5}$$

where $c_i \in \mathrm{GF}(2^m)$ and I, O and P are as defined in Figure 3.1. Note that the l information symbols which are set to 0 will not be transmitted.

The $n-k$ parity symbols are chosen such that the code word is divisible by the generator polynomial. The generator polynomial, $g(x)$, has degree $n-k$ and is of the form

$$g(x) = (x-\alpha^1)(x-\alpha^2)\ldots(x-\alpha^{n-k}) \tag{3.6}$$

where the α^i are $n-k$ consecutive powers of $\alpha \in \mathrm{GF}(2^m)$ [36].

The systematic code word is constructed by dividing $x^{n-k}i(x)$ by $g(x)$ to obtain the parity symbols, $p(x)$ (Refer to Fig. 3.1). Then the code word is

$$c(x) = i(x) + x^k p(x)$$

and is clearly divisible by $g(x)$. Note that the l most significant symbols of $i(x)$ are zero. Next we need to show how the encoder and the decoder work.

Encoder

Encoding means to find the parity polynomial $p(x)$ which is of length $n-k$. Moreover the parity polynomial symbols are chosen such that $g(x)$ divides $c(x)$. From Eq. 3.6, we know that:

$$c(x) = 0 \text{ for } x = \alpha^1, \alpha^2, \ldots, x = \alpha^{n-k} \tag{3.7}$$

Hence to ensure that $g(x)$ divides $c(x)$, the following $n-k$ equations must be true:

$$0 = i_0\alpha^{1(0)} + \ldots + i_{k-1}\alpha^{1(k-1)} + c_k\alpha^{1k} + \ldots + c_{n-1}\alpha^{1(n-1)} \tag{3.8}$$

$$0 = i_0\alpha^{2(0)} + \ldots + i_{k-1}\alpha^{2(k-1)} + c_k\alpha^{2k} + \ldots + c_{n-1}\alpha^{2(n-1)} \tag{3.9}$$

$$\vdots$$

$$0 = i_0\alpha^{(n-k)(0)} + \ldots + i_{k-1}\alpha^{(n-k)(k-1)} + c_k\alpha^{(n-k)k} + \ldots + c_{n-1}\alpha^{(n-k)(n-1)} \tag{3.10}$$

The set of $n-k$ simultaneous equations, with n terms and $n-k$ unknowns, can be uniquely solved for the $n-k$ unknowns which form $p(x)$. One way to solve this system of equations is with matrix manipulation. The above simultaneous equations are represented by a matrix as shown below.

$$\begin{bmatrix} \alpha^{1k} & \cdots & \alpha^{1(n-1)} \\ \vdots & & \vdots \\ \alpha^{(n-k)k} & \cdots & \alpha^{(n-k)(n-1)} \end{bmatrix} \begin{bmatrix} c_k \\ \vdots \\ c_{n-1} \end{bmatrix} = \begin{bmatrix} i_0\alpha^{1(0)} + \ldots + i_{k-1}\alpha^{1(k-1)} \\ \vdots \\ i_0\alpha^{(n-k)(0)} + \ldots + i_{k-1}\alpha^{(n-k)(k-1)} \end{bmatrix} \tag{3.11}$$

Decoder

For upto $n-k$ errors or erasures in a transmitted codeword, can be corrected if their locations are known; in our case, the location of errors is known from the missing cell sequence number. The decoding algorithm in this case is identical to the encoding scheme described above except the unknowns are in different positions. For example, assume that i_o to i_{u-1} are lost and that $(n-k) > u$, then using the same technique as in the encoder we can express the u simultaneous equations with n terms and u unknowns as:

$$0 = i_0\alpha^{1(0)} + \ldots + i_{u-1}\alpha^{1(u-1)} + \ldots + c_k\alpha^{1k} + \ldots + c_{n-1}\alpha^{1(n-1)} \tag{3.12}$$

$$0 = i_0\alpha^{2(0)} + \ldots + i_{u-1}\alpha^{2(u-1)} + \ldots + c_k\alpha^{2k} + \ldots + c_{n-1}\alpha^{2(n-1)} \tag{3.13}$$

$$\vdots$$

$$0 = i_0\alpha^{u(0)} + \ldots + i_{u-1}\alpha^{u(u-1)} + \ldots + c_k\alpha^{uk} + \ldots + c_{n-1}\alpha^{u(n-1)} \quad (3.14)$$

The above system of equations can be solved by matrix manipulation and be represented by:

$$\begin{bmatrix} \alpha^{1(0)} & \cdots & \alpha^{1(u-1)} \\ \vdots & & \vdots \\ \alpha^{u(0)} & \cdots & \alpha^{u(u-1)} \end{bmatrix} \begin{bmatrix} i_0 \\ \vdots \\ i_{u-1} \end{bmatrix} = \begin{bmatrix} i_u\alpha^{1(u)} + \ldots + c_{n-1}\alpha^{1(n-1)} \\ \vdots \\ i_u\alpha^{u(u)} + \ldots + i_{n-1}\alpha^{u(n-1)} \end{bmatrix} \quad (3.15)$$

Solving the matrix equation above will reveal the lost information symbols.

The above explains simply the mathematics behind encoding and decoding operation for RS codes. There are many VLSI implementations for RS CODEC. References [37], [38], [39] and [40] are some examples of RS CODECS which can be used in ATM communications.

An AFEC example

Here we will show the encoding and decoding operations as described above. We will be using AFEC (7,5,1) over GF(2^3). The symbol length is 3 bits and AFEC can correct upto 2 missing symbols. The field elements are then defined using GF(2^3) primitive polynomial, $x^3 + x + 1$, where α is a root so that $\alpha^3 = \alpha + 1$ [36]. Table 3.1 shows the power representation of the field elements.

Using the elements of Table 3.1 we then define the addition and multiplication tables as shown in Tables 3.2 and 3.3, respectively.

Let us assume that the information to be transmitted is:

$$I = (i_0 = \alpha^0, i_1 = \alpha^5, i_2 = \alpha^2, i_3 = \alpha^4).$$

Using Eq. 3.6, the generator polynomial is defined as:

$$g(x) = (x - \alpha^1)(x - \alpha^2)$$

and from 3.7, we get:

$$c(x) = 0 \text{ for } x = \alpha^1 \text{ and } \alpha^2$$

Table 3.1. *A representation of GF(2^3) generated from $\alpha^3 = \alpha + 1$.*

Zero and powers of α	Binary value
0	000
1	001
α^1	010
α^2	100
α^3	011
α^4	110
α^5	111
α^6	101

Table 3.2. *The addition table for GF(2^3).*

+	0	1	α	α^2	α^3	α^4	α^5	α^6
0	0	1	α	α^2	α^3	α^4	α^5	α^6
1	1	0	α^3	α^6	α	α^5	α^4	α^2
α	α	α^3	0	α^4	1	α^2	α^6	α^5
α^2	α^2	α^6	α^4	0	α^5	α	α^3	1
α^3	α^3	α	1	α^5	0	α^6	α^2	α^4
α^4	α^4	α^5	α^2	α	α^6	0	1	α^3
α^5	α^5	α^4	α^6	α^3	α^2	1	0	α
α^6	α^6	α^2	α^5	1	α^4	α^3	α	0

Table 3.3. *The multiplication table for GF(2^3).*

•	0	1	α	α^2	α^3	α^4	α^5	α^6
0	0	0	0	0	0	0	0	0
1	0	1	α	α^2	α^3	α^4	α^5	α^6
α	0	α	α^2	α^3	α^4	α^5	α^6	1
α^2	0	α^2	α^3	α^4	α^5	α^6	1	α
α^3	0	α^3	α^4	α^5	α^6	1	α	α^2
α^4	0	α^4	α^5	α^6	1	α	α^2	α^3
α^5	0	α^5	α^6	1	α	α^2	α^3	α^4
α^6	0	α^6	1	α	α^2	α^3	α^4	α^5

In the encoding operation, the unknown parity symbols, c_5 and c_6, need to be found. Using equations 3.8 through 3.10, we know that:

$$0 = i_0\alpha^{1(0)} + i_1\alpha^{1(1)} + i_2\alpha^{1(2)} + i_3\alpha^{1(3)} + i_4\alpha^{1(4)} + c_5\alpha^{1(5)} + c_6\alpha^{1(6)} \quad (3.16)$$

$$0 = i_0\alpha^{2(0)} + i_1\alpha^{2(1)} + i_2\alpha^{2(2)} + i_3\alpha^{2(3)} + i_4\alpha^{2(4)} + c_5\alpha^{2(5)} + c_6\alpha^{2(6)} \quad (3.17)$$

where i_4 is the zero symbol. Substituting the information symbols with their respective values, and after some manipulation we derive the following matrix representation.

$$\begin{bmatrix} \alpha^5 & \alpha^6 \\ \alpha^3 & \alpha^5 \end{bmatrix} \begin{bmatrix} c_5 \\ c_6 \end{bmatrix} = \begin{bmatrix} \alpha^3 \\ \alpha^4 \end{bmatrix} \quad (3.18)$$

The unique solution to the above matrix is:

$$c_5 = \alpha^2 \quad (3.19)$$

$$c_6 = \alpha^2 \quad (3.20)$$

Therefore the codeword to be transmitted is:

$$I = (i_0 = \alpha^0, i_1 = \alpha^5, i_2 = \alpha^2, i_3 = \alpha^4, i_4 = 0, c_5 = \alpha^2, c_6 = \alpha^2).$$

Please note that when this codeword is sent out, the zero symbol (in this case i_4) is not sent out. However this zero symbol is recreated by the receiver since its location is known to the receiver.

Let us assume that two errors occurred to the codeword, the location of a symbol in error will be represented by e. The received codeword after adding the zero symbol to it, will look like this:

$$I = (i_0 = \alpha^0, i_1 = \alpha^5, i_2 = \alpha^2, i_3 = e, i_4 = 0, c_5 = e, c_6 = \alpha^2).$$

The receiver applies the decoding algorithm to recover the erasures. Again the decoder generates the two simultaneous equations as described in Eq. 3.12 through Eq. 3.14. This produces the following equations:

$$0 = \alpha^0\alpha^{1(0)} + \alpha^5\alpha^{1(1)} + \alpha^2\alpha^{1(2)} + i_3\alpha^{1(3)} + 0\alpha^{1(4)} + c_5\alpha^{1(5)} + \alpha^2\alpha^{1(6)} \quad (3.21)$$

$$0 = \alpha^0\alpha^{2(0)} + \alpha^5\alpha^{2(1)} + \alpha^2\alpha^{2(2)} + i_3\alpha^{2(3)} + 0\alpha^{2(4)} + c_5\alpha^{2(5)} + \alpha^2\alpha^{2(6)} \quad (3.22)$$

After some manipulations, the following matrix is generated:

$$\begin{bmatrix} \alpha^3 & \alpha^5 \\ \alpha^6 & \alpha^3 \end{bmatrix} \begin{bmatrix} i_3 \\ c_5 \end{bmatrix} = \begin{bmatrix} 0 \\ \alpha^2 \end{bmatrix} \quad (3.23)$$

The unique solution of the above matrix is:

$$i_3 = \alpha^4 \quad (3.24)$$

$$c_5 = \alpha^2 \quad (3.25)$$

Thus the decoder output is:

$$I = (i_0 = \alpha^0, i_1 = \alpha^5, i_2 = \alpha^2, i_3 = \alpha^4, i_4 = 0, c_5 = \alpha^2, c_6 = \alpha^2).$$

which is the codeword generated by the sender's encoder.

In the next section, the negotiation of the coding parameters is discussed.

3.3 A Protocol Framework Deploying AFEC

FEC schemes can be deployed either above the ATM protocol stack or below it (i.e. over the physical media). In this section we are interested in investigating scenarios for how the communication entities will negotiate the coding parameters based on the network congestion sensed.

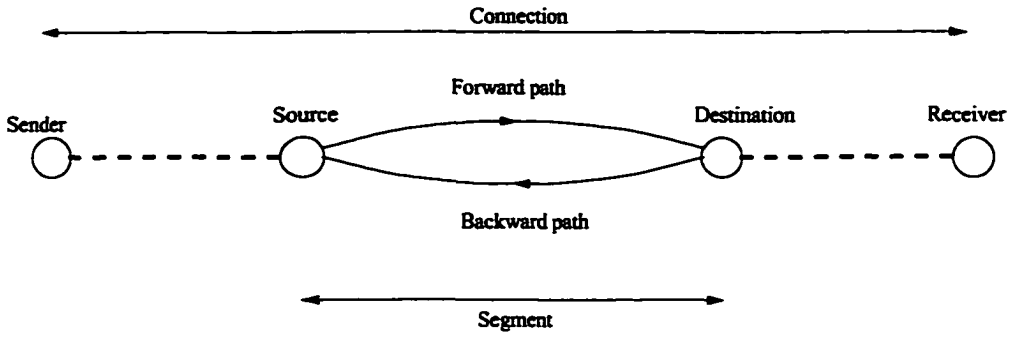

Figure 3.6. *Protocol entities.*

For simplicity and without losing generality, a unicast communication session is considered. The protocol can be applied on a segment basis or on a connection basis. A segment is either a VCC or VPC communication path between two VC or VP switching elements or nodes, and no other similar switch exists between them. A connection is two or more segments that establish an end-to-end communication path, as shown in Fig. 3.6. For simplicity the two nodes of a segment are referred to as the source and destination. Usually ATM channels are bidirectional. The source transmits data in one direction (forward direction). The bandwidth in the other direction (reverse direction) is set to zero. However periodic control messages are transmitted over the reverse channel.

The protocol under consideration is running over connections. A similar approach can be derived for segments with the assumption that the source and destination are able to store and manipulate data cells.

The sender begins by initiating cell transmission in the forward path. At the receiver, the number of missing cells is determined. This number indicates that either network congestion occurred along the transmission path or the path is noisy. If the physical medium is characterized by a high error rate, such as a satellite transmission channel, then noise may be the major source of errors. Based on this estimate of cell loss, the receiver determines the most suitable FEC parameters k and l. A Management Cell (MC) is sent through the reverse path (Fig.3.6). The sender receives the MC and updates the FEC parameters associated with this connection.

A lookup table is used to store the FEC parameters associated with each destination address. Initially (i.e. upon connection setup) k and l are set to zero (no coding)

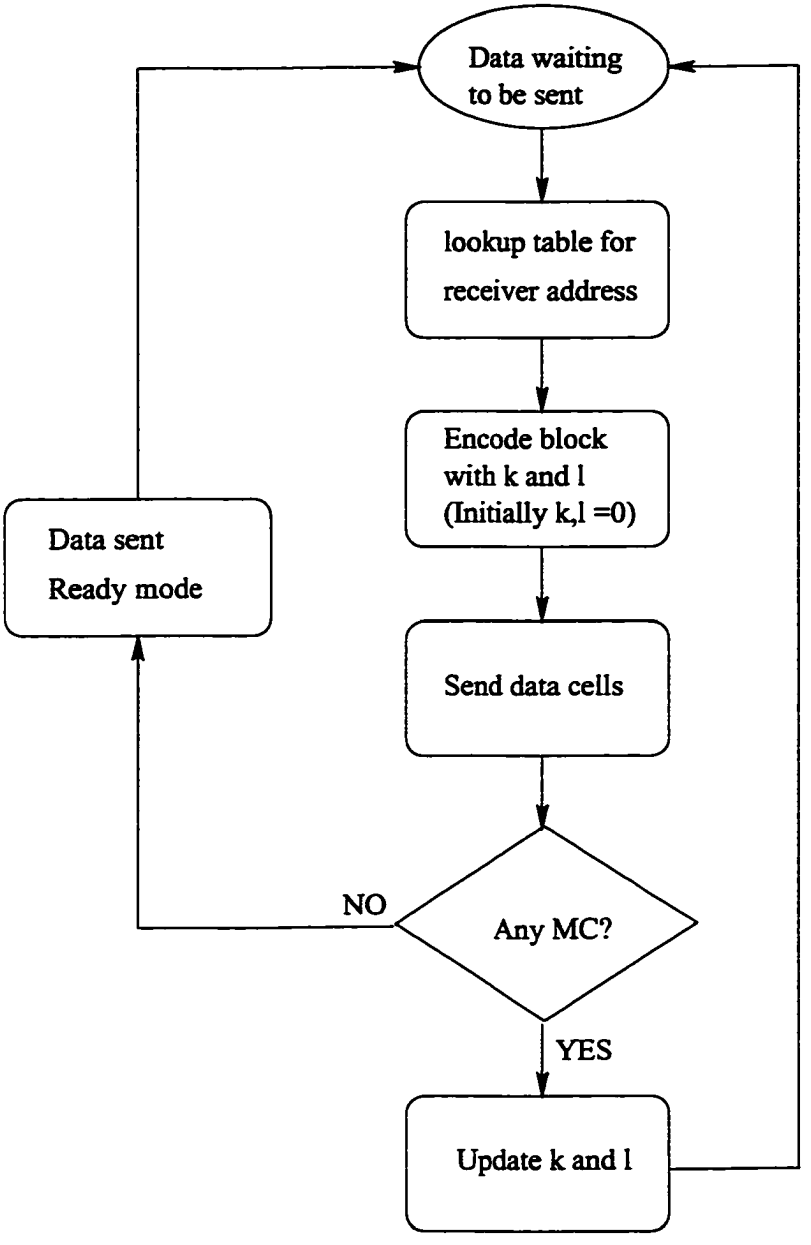

Figure 3.7. *Flowchart of actions taken by sender.*

or to some value that corresponds to low error rate. The receiver updates k and l according to the cell loss encountered. These parameters are negotiated depending on congestion and noise in the path.

As data loss is detected by the receiver, the values of k and l are estimated. Then a coding request is initiated and sent back to the sender. Figures 3.7 and 3.8 show flowcharts of the actions taken by the sender and receiver, respectively.

Non zero values in the FEC parameter values in the lookup table indicate that cells are collected as a block of data and passed to the (k,n,l) RS encoder which produces a coded block of length $n - l$, as shown in Fig. 3.1. At the receiving side, same k and l values are used for decoding.

The signaling in this proposed protocol is assumed to be in the forward path, which means that only the receivers determine the FEC parameters. This is more effective because the receiver can easily determine the number of lost cells. The traffic can then be reshaped accordingly. In this protocol, usage of the reverse connection is minimized. Only management cells (MC) with the FEC parameters are sent back to the source. Once the sender receives these MCs, it resets the cell sequence number. The receiver then expects coded data from the next cell with sequence number 0.

3.4 Security Feature

Among the three main operations carried out in information systems namely, processing, storage and transmission, transmission carries the greatest security risks [41]. Cryptography is one way to boost the security measures. Cryptography is the art of hiding information in a secret way so as to preserve its confidentiality [41]. Two operations are usually associated with cryptography: encipherment or to produce secure information and decipherment or to retrieve the original information.

Cryptosystems, which are pieces of software or hardware to conduct encipherment and decipherment, have two main components to function properly. These components are the cipher algorithm and the cipher key [42]. Communicating entities must agree on those components prior to any data transfer. For an intruder to reveal the encrypted message, both components need to be obtained. The strength of a cryptosystems is measured by the difficulty of obtaining such information.

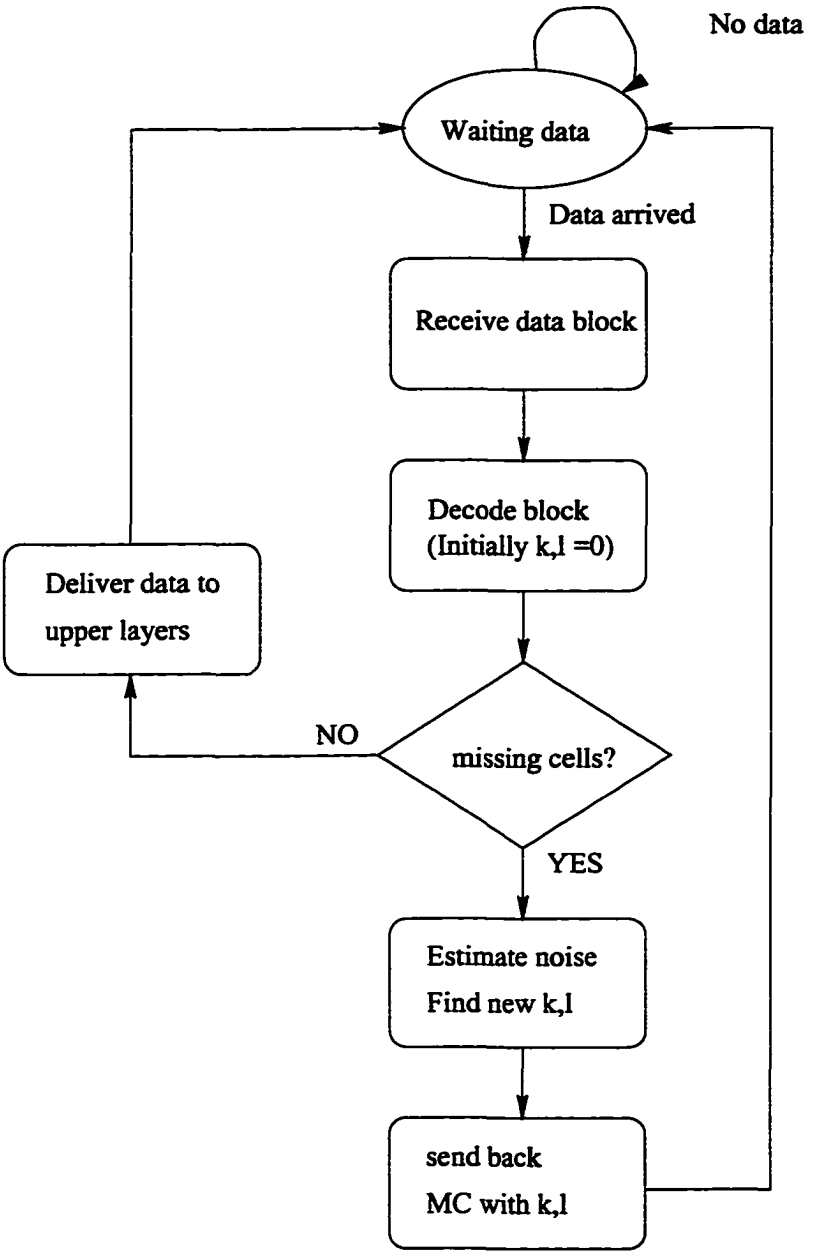

Figure 3.8. *Flow chart of actions taken by receiver.*

Shannon's theory of secrecy systems had contributed greatly to the theory of encipherment [41, 42]. Shannon identified two significant classes of encipherment systems: Unconditionally secure and computationally secure. Unconditionally secure cryptosystems cannot be broken even with infinite computational resources.

Computationally secure cryptosystems, on the other hand, are breakable, however, the algorithm used to break it requires a number of operations that is very large. This large number of operations is impossible to complete even with the fastest conceivable computational power. For example if a machine is capable of testing 10^6 cipher keys simultaneously, it would take it 107 days to exhaustively search for a key of length 64 bits if each search takes one microsecond. Whereas more than one hundred years are needed if a search takes one millisecond on the same machine [41].

Many cryptosystems have been invented. One example is the data encryption standard (DES) which is a widely used encipherment scheme. This cryptosystem is sometimes referred to as private-key where the same key is used in encipherment and decipherment [42]. The drawback of this cryptosystem is that the cipher key needs to be communicated between entities prior to any data transfer assuming secure channels [42]. Another popular encipherment scheme is the Rivest, Shamir, and Adleman (RSA) cryptosystem or sometimes called the public-key systems. The basic idea behind RSA is that encipher keys are made public whereas decipherment is done using privately owned decipher keys. Further treatment on the subject of cryptography and different cryptosystems can be found in [41, 42].

After the brief introduction of cryptography, now we need to look into the security aspect of the AFEC. Usually encipherment algorithm is handled independent of error correction algorithm. The two algorithms use different fields or spaces with different operations. This means that two stages of operation or layering are required. Here we are proposing a new schema to add security to the operation along with the correction capability of the code. Unlike classical approaches, this is done in one stage. The advantages are cut in operational costs and ease of implementation since no different fields and operations are required.

The technique we are proposing is to use some predefined polynomial instead of the string of the zero symbols. This key is used by both the encoder and encipherer as shown in Figure 3.9. The crypto algorithm proposed here is a simple GF() addition

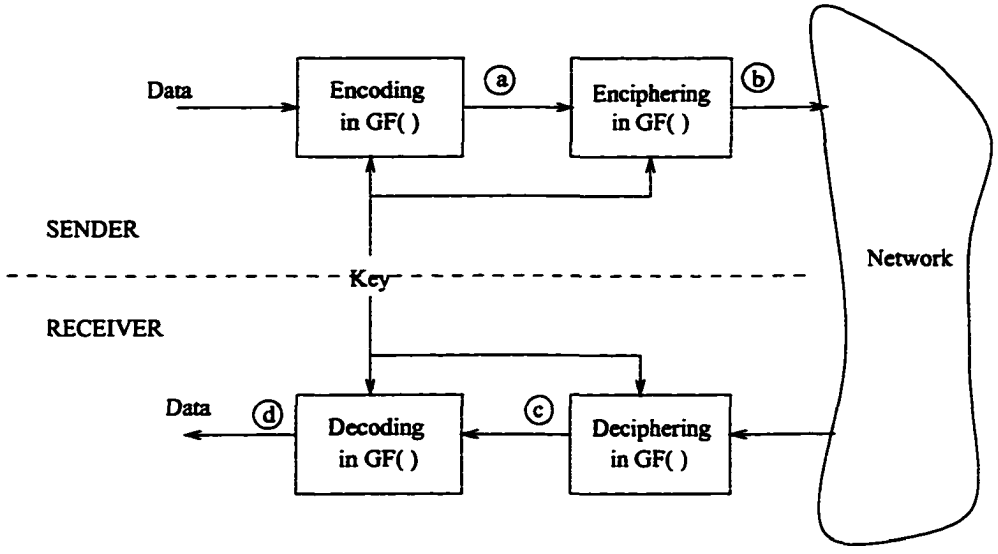

Figure 3.9. *Enhancemnet on AFEC to incorporate security. The same key is used for both AFEC CODEC and the cryptosystem.*

operation. The combined encryption algorithm (i.e. encoding and encryption) is outlined in the following steps:

1. The data is encoded using AFEC with the key, κ. The codeword is then passed for enciphering (point a in Figure 3.9).
2. The key, κ, is replicated to form a new polynomial, γ, of length n, which is the length of the codeword.
3. The codeword is then added to γ to generate the ciphertext for transmission (point b in Figure 3.9).

The combined decryption algorithm (i.e. decryption and decoding) is given in the following steps:

1. The ciphertext is received with erasures and deciphered using γ (point c in Figure 3.9). In this step if there is no error or the error is happening in the parity part of the codeword then there is no need for decoding.
2. If error or erasures are sensed in the data part of the codeword, then the key, κ, is used for decoding. Error location is always known due to sequencing. The original data transmitted is then recovered (point d of Figure 3.9).

The advantages of such scheme is that it does not add cost due to the security operation. Existing tables and manipulation for AFEC CODEC are used. It is needless to mention that a simple addition is only needed. This operation is done on a symbol-by-symbol basis in both enciphering and deciphering which makes it simple. In addition the security in this scheme comes from the length of the key polynomial, κ, and the choice of its terms.

In a related issue the key is always changing either due to congestion or to add more sophistication to the security. This scheme is not a one-time pad cryptosystem. However it can be used as a minimum security for communications.

In Eq. 3.5, the O symbols, namely c_{k-l} through c_{k-1}, are set to the zero element. When these symbols are set to non-zero, we establish the security in the code. For example if the security key (κ), which is of length l, is given by:

$$\kappa(x) = s_0 + s_1 x + \ldots + s_{l-1} x^{l-1}$$

then Eq. 3.5 can be rewritten with the above key as follows:

$$c(x) = c_0 + c_1 x + \ldots + c_{k-l-1} x^{k-l-1} + \underbrace{s_0 x^{k-l} + \ldots + s_{l-1} x^{k-1}}_{\text{security key } (\kappa)} + c_k x^k + \ldots + c_{n-1} x^{n-1}.$$

Now this codeword is encrypted using the concatenated version of κ, which is denoted by γ, as described above. Adding $\gamma(x)$ to $c(x)$ produces the encrypted codeword.

The security key (κ) is not transmitted with the message. However the destination, which uses the same key, adds the key to the received codeword.

These security keys are basically the polynomials needed to encipher and decipher the message as well as encoding and decoding. The security technique, we are proposing, is a private-key type of cryptosystems. The GF() addition operation is considered as the cipher algorithm. Although this operation is known and simple, wrong keys will always result in incorrect deciphering and decoding even if there is no error. The receiver and the transmitter should agree on a polynomial before transmission based on noise and security level required.

With such scheme, two levels of communication security are defined. One is related to the adaptive parameters of the AFEC scheme and the other is to the key polynomial. An intruder would not be able to retrieve the original message since the key is unknown to him/her. Moreover the AFEC parameter values used are also

unknown. More sophistication can be added to the security scheme by considering the location of the key within the information symbols. Different locations of the key within the information symbols result in different parities and different encryption.

This scheme is well suited for common access media where many users share the same media whether wireless or wireline. Moreover if minimum security is required in a wireline network, then AFEC and its security scheme can be applied above the ATM stack. More sophisticated security schemes, such as RSA, can be applied on data before AFEC CODEC operations and the key polynomial is either set to zero symbols for no encryption or to some key, κ, for another level of security.

A possible framework using this modified secure scheme is governed by the following assumptions:

1. Security keys are used to substitute the zero symbols as defined in the proposed AFEC.
2. Keys with the same polynomial length (l) are clustered into the key set, Υ_l.
3. Each key set, Υ_l , contains many keys from which a key, κ, is to be picked up.
4. The process of choosing a key from a set is agreed upon among the communicating nodes.
5. The receiver is responsible for choosing a key since coding parameters are updated by the receiver too. The transmitter is then informed of any changes.
6. Network security does not enforce any AFEC parameter change. This means that due to security issues no key longer than the current value of l is needed unless a parameter change takes place.
7. Keys can be changed, as necessary, but within the same key set. This is one level of sophistication for added security.

A crucial measure of a cryptosystem performance is the space of possible keys to be used. A good cryptosystem should provide huge space of keys which makes it computationally secure. To estimate the number of keys to be offered by the proposed AFEC, it is essential that the security key is not divisible by the generator polynomial, $g(x)$ (refer to Eq. 3.6). One way to guarantee this is to restrict the key polynomial length to be less than the generator polynomial length (i.e. $l < n-k$). Now l symbols in GF(2^m) will have 2^{ml} combinations. Therefore the key set, Υ_l, has $2^{ml} - 1$ keys,

where the key of all zero symbols is left out. Since at least one non-zero symbol is considered as a key, the total number of keys (X) in the system is given by:

$$X = \sum_{i=1}^{l} \Upsilon_i = \sum_{i=1}^{l} (2^{mi} - 1). \tag{3.26}$$

After some manipulation we derive the following expression for the total number of keys in the system:

$$X = \frac{2^{m(l+1)} - 2^{m+1} + l}{2^m - 1}.$$

For example if we use AFEC(255, 233, l), where $0 \leq l \leq 21$, then there are more than 3×10^{50} keys!

The above discussion assumes that keys are picked to exactly fit the zero polynomial. However if a key is picked such that its length is less than or equal to the required value of l then the number of keys are much greater than what has been given in Eq. 3.26. This is due to the fact that key position does give another level of security and therefore Eq. 3.26 becomes:

$$X == \sum_{i=1}^{l} \binom{l}{i} 2^{mi} - l.$$

Now the operation with this framework is summarized in the following steps:

1. A connection is established between the transmitter and the receiver. The AFEC parameters used initially reflect minimal channel noise. For a specific value l, a key, s, is pulled from the set of keys, Υ_l. The key should fit into the rest of the message to be transmitted.
2. If due to data loss the AFEC parameters need to be changed, the key will also be changed.
3. The transmitter is updated with the new coding parameters and the new key.

3.5 Concluding Remarks

The proposed adaptive FEC addresses reliable communication without significantly overloading the bandwidth available. Only enough redundancy is added as a reactive measure to sensed noises over the channels.

The introduction of the adaptive puncturing parameter l has made it possible to control the destructive nature of noise and or congestion on user data.

Although both parameters, k and l, can be made variable, it might be beneficial to fix k and vary l. Usually the cost associated with designing and development of RS codes with variable k is high especially when n is large. The l parameter can be varied while fixing the parameters n and k. In this case the CODEC cost are made lower since only one RS CODEC, RS (n,k), need to be designed. The adaptation can be then perceived with varying l which does not add cost the design of RS CODEC.

On another track, the parameter l might be used for data security. This is done by setting its value to non-zero, and in this case it is called a key. When l polynomial is added to the data polynomial, then resulting code word cannot be retrieved unless the key and the proper RS parameters are known. Data encryption comes from the Galoy field algebra. Although private-key type of cryptosystem is proposed in this chapter, adapting public-key cryptosystem can further be investigated.

The added security feature of the proposed scheme does not increase operation. As a matter of fact both operations (i.e. error correction and data encryption) are done in one stage. A strategy to take advantage of the security feature of the proposed FEC is also outlined.

Chapter 4

Performance Modeling

The adaptive FEC code introduced in the previous chapter is evaluated analytically here. The motivation is to attempt to study the relevant parameters that closely affect system performance. In addition analytical solutions can be used reasonably quickly.

The modeling in this chapter assumes an ATM network running over fiber optical links. The intention is to study the overall performance issues when AFEC is used. First we will be looking at the system throughput when AFEC is used. Then we compare this throughput with throughputs achieved by classical FEC, Go-back-N and selective repeat protocols. This is described in Section 4.2.

We compared the end-to-end delay for AFEC against that of the classical FEC. The comparison is given in Section 4.3. The comparison of end-to-end delay of AFEC to ARQ schemes is left for the simulations of Chapter 6. The rationale is to express the significant gain of applying both schemes in a fairly complicated environment such as broadcasting.

An important issue when using AFEC is the amount of overhead to be expressed as the frequency of adaptation of coding parameters. This issue has been studied in Section 4.4.

4.1 Preliminaries and Notation

In this section the different parameters used in this chapter are defined.

C = channel capacity (bps),

CLR = cell loss ratio,

D = PDU size (octets),
n, k, l = FEC coding parameters,
p' = cell loss probability (CLR) with FEC encoding,
p'' = cell loss probability (CLR) with AFEC encoding,
t_{dec} = FEC decoder delay (s),
t_{enc} = FEC encoder delay (s),
t_{trans} = PDU transmission time (s) ,
t_{prop} = channel propagation delay (s),
β' = codeword error probability with FEC encoding,
β'' = codeword error probability with AFEC encoding,
ϱ' = PDU error probability with FEC encoding,
ϱ'' = PDU error probability with AFEC encoding,
η' = throughput with FEC encoding,
η'' = throughput with AFEC encoding,
τ' = delay (s) with FEC encoding,
τ'' = delay (s) with AFEC encoding,
Δ = cell transmission time (s).

Timing parameters:

PDU transmission time is the time needed to transmit a user PDU between two nodes. PDU transmission time (t_{trans}) can then be defined as follows:

$$t_{trans} = \frac{D \times 8}{C}. \tag{4.1}$$

This transmission time is not generally an integer multiple of cell transmission time (Δ). Since we assume ATM networks to be the transportation media then t_{trans} is essentially an integer multiple of Δ.

CLR with FEC coding:

To transmit an average higher layer Protocol Data Unit (PDU) of length D octets, it is broken down into $\lceil D/k \rceil$ data blocks. The data block which is of size k is encoded

into a codeword of length n. The probability of a codeword be in error is given by:

$$\beta' = \sum_{i=R}^{N} \binom{N}{i} CLR^i (1 - CLR)^{N-i} \tag{4.2}$$

where $N = \lceil \frac{n}{48} \rceil$ and $R = \lceil \frac{n-k}{48} \rceil + 1$, assuming space for sequencing is negligible (≤ 3 bits).

There are $\lceil \frac{D}{k} \rceil$ codewords that form a PDU. The probability of receiving an erroneous coded PDU using fixed error correction FEC can then be expressed by:

$$\varrho' = \sum_{j=1}^{\lceil \frac{D}{k} \rceil} \binom{\lceil \frac{D}{k} \rceil}{j} (\beta')^i (1 - \beta')^{\lceil \frac{D}{k} \rceil - j}.$$

Since cell loss is identical and independent and the medium that carries the cells is assumed memoryless, then the new CLR (p') as a result of applying FEC is defined as:

$$p' = \frac{\varrho'}{\lceil \frac{D}{k} \rceil \lceil \frac{n}{48} \rceil}. \tag{4.3}$$

In the case where AFEC is used, the parameter k and l are changing and hence the PDU is broken down into $\lceil \frac{D}{k-l} \rceil$ data blocks each of which is encoded into a codeword of length $n - l$ symbols. The codeword is further broken down into $\lceil \frac{n-l}{48} \rceil$ ATM cells. The probability of a PDU in error is expressed by:

$$\varrho'' = \sum_{i=1}^{\lceil \frac{D}{k-l} \rceil} \binom{\lceil \frac{D}{k-l} \rceil}{i} (\beta'')^i (1 - \beta'')^{\lceil \frac{D}{k-l} \rceil - i} \tag{4.4}$$

where β'' is defined as:

$$\beta'' = \sum_{j=1}^{N} \binom{N}{j} CLR^j (1 - CLR)^{N-j}$$

where $N = \lceil \frac{n-l}{48} \rceil$ and $R = \lceil \frac{n-k}{48} \rceil + 1$, assuming space for sequencing is negligible (≤ 3 bits).

Similar to the case where classical FEC is used, the new CLR due to deployment of AFEC (p'') can be expressed as follows:

$$p'' = \frac{\varrho''}{\lceil \frac{D}{k-l} \rceil \lceil \frac{n-l}{48} \rceil}. \tag{4.5}$$

4.2 Effective throughput

Sophisticated applications produce huge amounts of data to be carried out by the network. FEC schemes should provide a high throughput with low delays. Assuming the network shown in Fig. 4.3, a MATLAB program was written to determine the theoretical effective throughput. Simulations were performed to compare the throughput performance of the conventional and proposed FEC schemes with two different ARQ schemes. The two ARQ schemes considered are Selective Repeat Protocol (SRP) and Go Back N (GBN).

ARQ schemes

The impact of cell loss on user application throughput using ARQ scheme is presented here. The closed-form formula derivations for the ARQ schemes are reproduced from [6]. The number of cells in the retransmission window W is determined by the transmission rate C, the user PDU D (in octets), and the propagation delay t_{prop} as follows:

$$W = \left\lceil \frac{2Ct_{prop}}{D \times 8} \right\rceil . \tag{4.6}$$

A PDU is considered in error once one of its cells is in error. The probability that an individual PDU is in error due to a random cell loss probability π derived from CLR is approximated by:

$$\pi \approx \left\lceil \frac{D}{48} \right\rceil \times CLR. \tag{4.7}$$

In the GBN strategy, if a single PDU is in error, then the entire window is retransmitted. For the GBN retransmission strategy, the usable throughput $\eta\ (GBN)$ is approximated by the inverse of the average number of times the entire window must be sent, which is approximately:

$$\eta\ (GBN) \approx \frac{1-\pi}{1+\pi W}. \tag{4.8}$$

In the SRP, only PDUs in error are retransmitted. The usable throughput $\eta\ (SRP)$, in this case, is approximately the inverse of the average number of times any individual PDU is sent, which is:

$$\eta\,(SRP) \approx 1 - \pi\ . \tag{4.9}$$

FEC schemes

The usable throughput of both conventional and adaptive FEC schemes are approximated as the percentage of the number of bytes in the corresponding coded block that represent user data. For the conventional FEC, usable throughput $\eta\ (FEC)$ is defined as follows:

$$\eta' = \left(\frac{k}{n}\right)(1 - p') \tag{4.10}$$

where p' is the CLR when using classical FEC as defined by Eq. (4.3).

The usable throughput for adaptive FEC $\eta\ (AFEC)$ is given by the following expression:

$$\eta'' = \left(\frac{k-l}{n-l}\right)(1 - p'') \tag{4.11}$$

where p'' is the CLR when AFEC is used as given by Eq. (4.5).

Equations (4.10) and (4.11) model the behavior of the corresponding FEC code given certain CLR. For adaptive FEC, the code adapts its coding parameters according to the level of cell loss experienced. This is expressed in Eq. (4.11) as the probability of an individual coded block is in error due to CLR.

Equations (4.8) to (4.11) are simulated with MATLAB. The Cell Loss Ratio (CLR) is varied from 1×10^{-10} to 1×10^{-1}. Figure 4.1 plots the throughput of the various schemes described above.

Fig. 4.1 shows that adaptive FEC has an effective throughput comparable to ARQ for low CLR ($\leq 1 \times 10^{-5}$). At a high CLR, the effective throughput tends to decrease slowly (CLR $> 1 \times 10^{-3}$). Generally the effective throughput of adaptive FEC is higher than that of conventional FEC. The stair-case pattern found in the figure for AFEC is due to the rounding of the number of redundancy cells to the nearest integer.

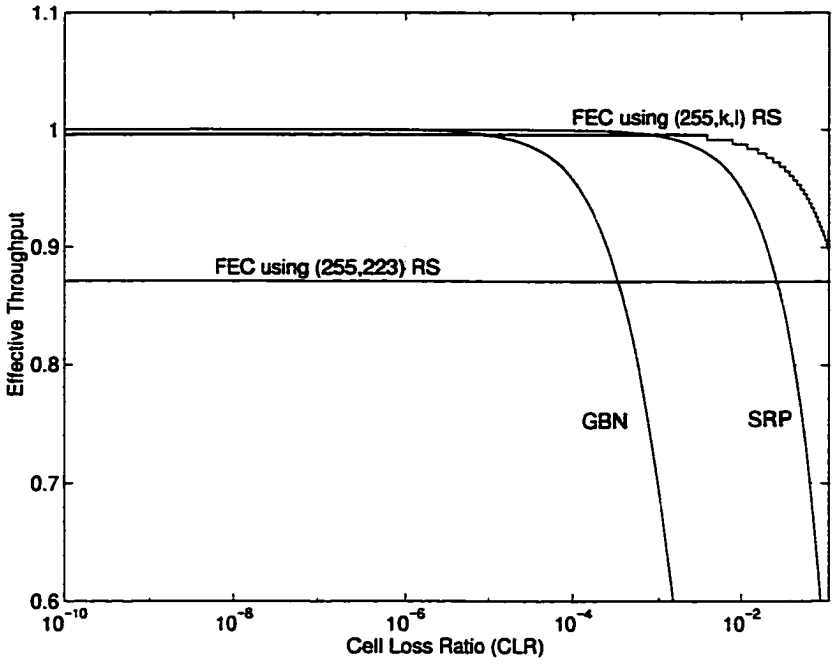

Figure 4.1. *Dependence of throughput on Cell Loss Ratio for several error control techniques.*

4.3 End-to-end delay

Communication delays when FEC schemes are deployed, are very deterministic. End-to-end delay of adaptive FEC is analyzed, simulated and compared to that of conventional FEC.

FEC CODECs have deterministic delays, t_{enc} and t_{dec}, associated with encoders and decoders, respectively. These delays are inherent characteristics of CODECs.

Usually encoding delays are smaller than decoding delays. This is due to the fact that decoders cannot deliver PDUs to upper layers until the whole coded block is received, of course with some loss.

Figure 4.2 shows the delays experienced when a block of cells are transmitted between a transmitter and a receiver.

The total one-way delay (T_{total}) for a coded block with no loss, derived from Figure 4.2, is as follows:

$$T_{total} = t_{trans} + t_{prop} + t_{enc} + t_{dec} + \delta \tag{4.12}$$

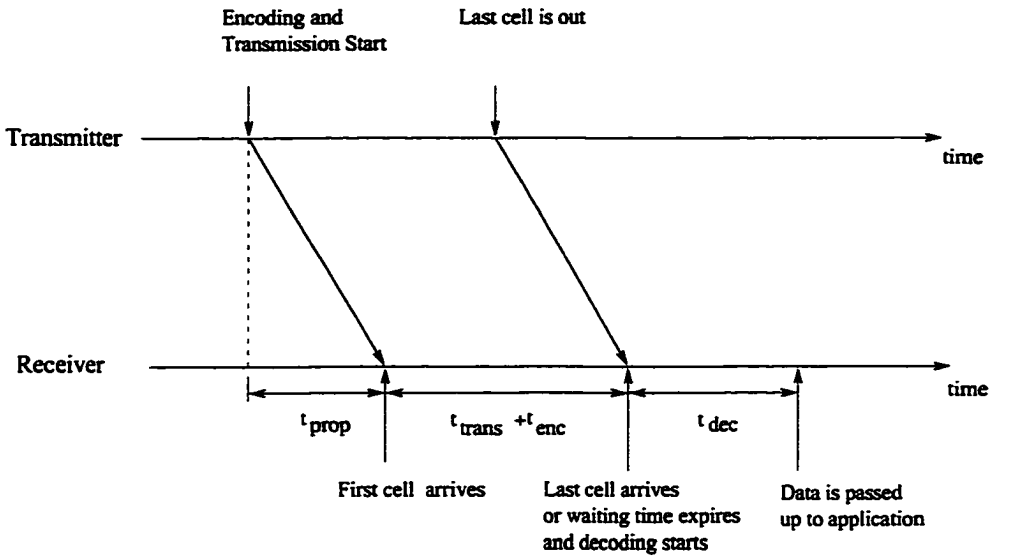

Figure 4.2. *Timing diagram showing the delay components associated with transmitting a block of cells.*

where δ is a delay consisting of different components such as packetization delay, which we assumed to be negligible. The encoding delay (t_{enc}) can be deleted from the above equation since a copy of the cell can be kept. Redundancy cells will follow the last data cell.

Since AFEC uses two adaptive parameters, k and l, it is therefore important to study the impact of parameter l on the end-to-end delay. The network employed is illustrated in Figure 4.3. A (255,223) RS CODEC is assumed for the conventional FEC and a (255,223,l) RS CODEC for adaptive FEC. Cells are delivered to the

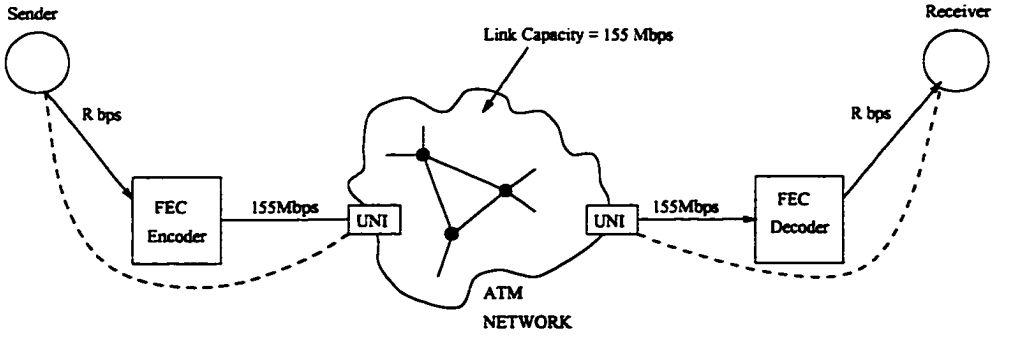

Figure 4.3. *The network model with FEC used for analysis.*

encoders and leave the decoders at a rate of 160 cells/sec. The application modeled here is an interactive audio session. Higher speeds, as required for video applications over ATM, produce comparable results.

The study is conducted under the following assumptions:

1. Conventional RS encoders wait till k cell payloads are delivered from the application. AFEC encoders wait till $(k-l)$ cell payloads are ready for transmission.
2. For the AFEC CODEC, the time to encode or decode $k-l$ cell payloads is constant. The validity of this assumption is due to the fact that fixed n in FEC coding is assumed through out the network.
3. The maximum delay is a one way communication delay. This is true since FEC schemes regenerate lost cells within certain CLR.
4. ATM network switching and propagation delays are negligible.
5. User data is delivered to the FEC CODEC at the rate of r octet/sec. The total time required to deliver user data is

$$\frac{D}{r}\text{sec.} \tag{4.13}$$

A user PDU of D octets is broken down into segments to fit into the data part of the FEC encoder. For the conventional FEC, the number of segments (S_{FEC}) required is:

$$S_{FEC} = \left\lceil \frac{D}{k} \right\rceil . \tag{4.14}$$

The number of segments (S_{AFEC}) needed in AFEC is:

$$S_{AFEC} = \left\lceil \frac{D}{k-l} \right\rceil . \tag{4.15}$$

It is clear that the number of segments needed when adaptive FEC is used, increases as the value of l increases. And the actual data sent in each coded PDU of the AFEC is less when compared to conventional FEC. $n-l$ bytes are sent using AFEC. While n bytes are transmitted when conventional FEC is used.

Each coded PDU in both cases is further divided into 48 byte segments forming ATM cell payloads. The number of ATM cells required for conventional FEC and AFEC are defined by equations (4.16) and (4.17), respectively.

$$\text{ATM cells for FEC} = \left\lceil \frac{n}{48} \right\rceil \tag{4.16}$$

$$\text{ATM cells for AFEC} = \left\lceil \frac{n-l}{48} \right\rceil . \tag{4.17}$$

One way delay through ATM network is composed of the number of segments as defined in equations 4.14 and 4.15, the number of ATM cells as defined in equations 4.16 and 4.17, and cell transmission time. Therefore the delay experienced by conventional FEC (τ')is expressed by the following:

$$\tau' = \left\lceil \frac{D}{k} \right\rceil \left\lceil \frac{n}{48} \right\rceil \Delta \tag{4.18}$$

where Δ is a cell transmission delay.

On the other hand, the delay encountered by AFEC (τ'')is formulated as follows:

$$\tau'' = \left\lceil \frac{D}{k-l} \right\rceil \left\lceil \frac{n-l}{48} \right\rceil \Delta. \tag{4.19}$$

Therefore, the maximum end-to-end delay encountered by user application is defined as the delay given by Eq. (4.13) added to the transmission delay in Eq. (4.18) for conventional FEC, and Eq. (4.19) for adaptive FEC. Based on the above assumptions and the closed form formulae derived, simulation is conducted with MATLAB. A user data PDU of 500 cells (D=500) is assumed.

The end-to-end delay encountered is depicted in Figure 4.4. With conventional FEC, there is a fixed delay. On the other hand, with adaptive FEC the same block of cells encounters a smaller delay, depending on the value of l. One observation can be derived from Figure 4.4. As the number of data units in a PDU decreases due to the increase in value of l, the end-to-end delay decreases. An added advantage is that the correcting capability of the scheme increases (refer to Fig. 3.2.) The oscillations in Figure 4.4 are due to dividing the PDU by the length of data portion in AFEC scheme (i.e. $k-l$). The instants where $k-l$ divides D with no remainder give the lowest end-to-end delays.

In general, low end-to-end delays are achieved with AFEC, and the worst delay that may be encountered is bounded by the delay of the conventional FEC. The delay improvement will positively impact QoS of real-time application.

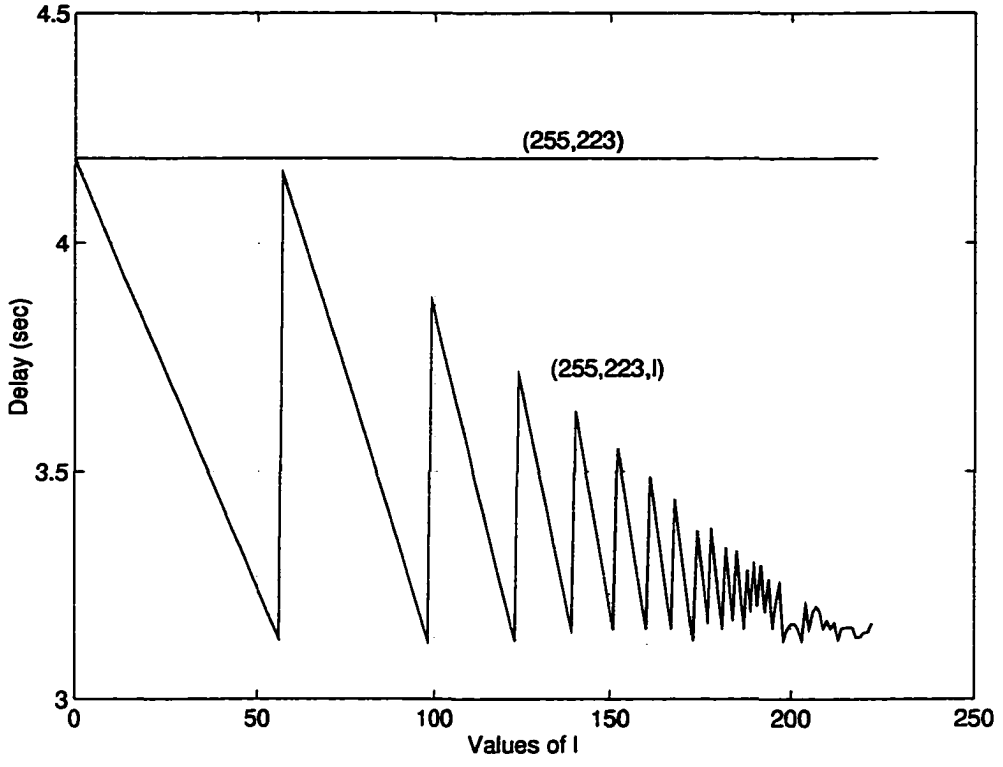

Figure 4.4. *The delay encountered by a block of 500 cells using conventional FEC with an RS (255,223) code, and adaptive FEC with a (255,223,l) RS code.*

4.4 Parameter Update Technique

AFEC is an adaptive coding technique where the parameters k and l are changed. To counteract any channel deficiency or data loss, the coding parameters should be changed by increasing the correction rate of the code. On the other hand, if the network recovers from a congestion state then coding parameters should be changed to reflect the current network state. The objective is to minimize the protocol overhead and yet achieve higher throughput.

The channel is best monitored by the destination. As stated in Chapter 3, destinations are capable of quantifying the loss. This is basically done with the help of sequencing. At most one control message is generated when the coding parameter need to be changed. This significantly reduces the operation costs associated with AFEC since network states (i.e. congestion or no congestion) tend to exist for a fair

period of time.

Another interesting issue is the incremental change that the coding parameters should be updated to. When a congestion is sensed then the coding parameters should be changed to eliminate the effect of such congestion on the data. In other word the values of the coding parameters should somehow be increased. However if the network is recovering from congestion then less data will be lost and thus it is important to change the coding parameters (i.e. their values should be decreased).

The problem in the latter case is how fast should change or decrease be. There are two solutions to this problem. Either the coding parameter changes with the instantaneous data loss or the destination collects statistics over the past readings and predicts suitable parameter values. This has some analogy in the field of digital filters. Therefore to solve such problem, we can use digital filters that take the number of cell lost, $x(n)$, at time n as the input function and output the function $y(n)$ which describes the values for the parameters k and l.

Different approaches can be used to find $y(n)$ given the above constraints. Finite input response (FIR), infinite input response (IIR) or Kalman prediction [43] are possible candidates. For simplicity we assume that parameter change follows the instantaneous reading of the data loss.

4.5 Concluding Remarks

In this chapter, the performance of the proposed AFEC scheme is analytically compared to some other schemes. Among the known schemes GBN and SRP are chosen to represent ARQ strategies.

Table 4.1 summarizes some of the parameters and variables of FEC and AFEC for a PDU of size D octets. FEC schemes are assumed in the GF(2^8). Note that $0 \leq l < k$ and $k < n$.

It is found that the proposed scheme improves end-to-end delay experienced by real-time application such as Audio. This is mainly due to the deterministic delay associated with operation. Moreover the l parameter is found to enhance the end-to-end delay and at the same time increase data reliability. This enhancement is achieved when l is increased.

Table 4.1. *Comparison of conventional and adaptive FEC as opposed to no coding.*

	No Coding	FEC	AFEC
Data fraction	1	$\frac{k}{n}$	$\frac{k-l}{n-l}$
Cells sent	$\frac{D}{48}$	$\lceil\frac{D}{k}\rceil n$	$\lceil\frac{N}{k-l}\rceil(n-l)$
Maximum delay	$\frac{D}{48}\Delta$	$\lceil\frac{D}{k}\rceil n\Delta$	$\lceil\frac{D}{k-l}\rceil(n-l)\Delta$
Error Correcting	0	$\frac{n-k}{n}$	$\frac{n-k}{n-l}$

On another track the goodput (i.e. effective throughput) of the proposed scheme is comparable to ARQ schemes at low CLR and is by far much better at relatively high CLR even when compared to conventional FEC. Generally, AFEC is found to outperform other coding schemes in throughput. It is also found that AFEC maintains high throughput even at high CLR values.

Not as in ARQ schemes, the number of control messages generated by AFEC is manageable. AFEC generates a parameter update request every time the network congestion state changes. To improve this, prediction techniques such as Kalman prediction, can be used to change the coding parameters based on the readings of the previous data losses. The destination is most suitable to monitor the network congestion.

Chapter 5

Wireless ATM

The diverse attributes of ATM such as simplicity, teletraffic engineering, quality of services, and the definition of different services, have made it attractive to other technologies. Wireless communications is one field where ATM has found great interest in. Wireless ATM is basically an extension of the existing ATM technology. Therefore wireless ATM need to preserve the integrity of ATM communications.

ATM mainly operates over optical fiber channels where bit error rate is low ($\leq 10^{-10}$). This has great impact on the simplification of the ATM protocol stack where no data link layer (DLC) functionality is needed. However when ATM is thought of to support wireless, new set of protocols and layers need to be considered.

In principle, wireless communications has three stages of folded operation [2]. First stage is the source encoding/decoding. In source encoding raw data are processed, e.g. compressed, to be replaced by shorter but equivalent one. Source decoders do the reverse job. In the second stage there are the channel encoders/decoders. Channel encoders add redundancy to protect data from transmission noise. Channel decoders, on the other hand, reconstructs the encoded message and compensates for any loss. At this stage, communication reliability is an issue.

The final stage is the modulation/demodulation. At this stage data are prepared to be transported over wireless channels using some form of coding and modulation techniques such as quadrature amplitude modulation (QAM) [44]. The demodulators, on the other hand, transforms the radio message back into some form of signal that is readable by the channel decoders. Figure 5.1 shows the principle operation of wireless communication systems.

Based on the above description of wireless systems, below are some areas where to improve wireless communications [2]:

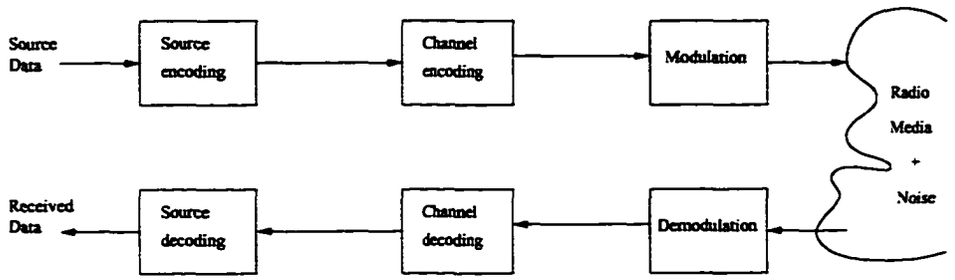

Figure 5.1. *General description of transmission in wireless channel as shown in [2].*

1. increasing source coding efficiency by using data compression techniques that gives more compression ratios.
2. using efficient modulation techniques that is characterized by high noise immunity.
3. using efficient channel coding technique that overcomes the problems of wireless communications. Such techniques include FEC. This point is mainly emphasized on in this chapter.

Boutillon *et al* [2] define an efficient channel coding as the coding scheme,which for an existing noise level, minimizes the signal power to achieve the desired error rate. This is our motive to study the bit error rate versus signal power for the AFEC.

In this chapter a brief treatment on wireless ATM is given in Section 5.1. This is intended to describe the issues and limitations confronting radio ATM. More details can be found in the references listed in that section.

The chapter then continues with the deployment of the proposed AFEC scheme into ATM radio networks. This is described in Section 5.2. The adaptive component in the proposed scheme fits well with the different types of noise encountered in Wireless media. Obviously more redundancy is needed when more noise is encountered. The security feature of AFEC is also discussed. CLR performance of deploying AFEC in Rayleigh fading type of media is studied in Section 5.3.

5.1 Wireless ATM: A Review

Wireless ATM is a collection of technologies, techniques, architectures and devices that are to carry traffic and services, as been described and defined by ATM and

ISDN standard bodies, over wireless media.

In general there is distinction between wireline and wireless communication. For example wireless communications offer limited quality of service. The areas where wireless communications differ from wireline communications are listed below [45]:

1. Wireless links has a limited radio spectrum which makes the available capacity for service limited.
2. Wireless communication has time-varying impairments due to the rapidly changing surrounding environment. Examples of time-varying impairments include multipath propagation, shadow fading, co-channel interference. These impairments has a great impact on the bit error rate performance.
3. Mobile users tend to change their position with respect to time. This means that access point is also changing throughout the duration of a connection. The implication of such behavior is significant on the quality of service to be provided.

The rationale of using wireless ATM is attributable to the following characteristics that distinguish ATM technology [46]:

1. Fixed and small-sized cells, and
2. QoS-specifiable VC.

Wireless bandwidth is expensive and therefore it is wasteful of resources to carry data in the format that wireline ATM uses (i.e. 48 octets). One solution is to combine two or more cells into one radio frame. As such wireless channel throughput is improved.

From the QoS-specifiable VC perspective, ATM intelligently provide a schema for resource allocation and admission control upon VC establishment. Moreover ATM provides a traffic flow ID that make traffic engineering possible. This ID can be extended to the wireless part of ATM. The usefulness of this ID is in the allocation and scheduling of the wireless channel resources [46]. This suggests that the usual media access control or multi-access (MAC) protocol need to be altered to accommodate channel resource scheduling.

From the above, when wireless communications is to get integrated transparently into the ATM networks, it is necessary for one or more of the following modifications

to take place [47, 3, 4]:

1. Design of physical (PHY) layers that provide both the required BER and transmission rates that support multimedia services. The objective is to counteract the channel impairment described previously. High speed radio links is one example. In addition radio modems must support burst operation which may involve innovation in modulation techniques. One modulation candidate is the equalized quadrature phase shift keying (QPSK)/QAM [48]. This scheme requires complex equalizer. Another viable option is direct sequence code division multiplexing (DS/CDMA) [49].
2. Definition and integration of MAC and data link layer (DLC) protocols that support dynamic bandwidth allocation with traffic policing functions. Specific techniques which have been considered for wireless ATM as MAC protocols include packet reservation multiple access (PRMA) [50] and multiservice dynamic reservation (MDR-TDMA) [51]. Recently, the work by Passas *et al.* [52] proposes a MAC protocol which is oriented to the QoS to be provided for an application. Channel allocation is assigned to an application based on the type of service to be provided and the service class priority (i.e. CBR is higher priority than rtVBR, and so on).

   On the other hand, DLC protocols are needed to mitigate the effect of radio channels on data. Possible protocols to be deployed in such layers include ARQ and FEC.
3. Modification of the signaling protocols of the user-network interface (UNI) [34] and private network-network interface (PNNI) [53] to support mobility management.

From the above discussion, a possible wireless ATM architecture and protocol stack [3, 4] is shown in Figure 5.2. In this figure, ATM base stations support wireline and wireless communications simultaneously. In the case of wireline ATM service, normal ATM operation takes place. However when wireless connection is to be established, the connection request processing and data traffic flow in the wireless plane (shaded areas in Figure 5.2). Further description can be found in [3, 4].

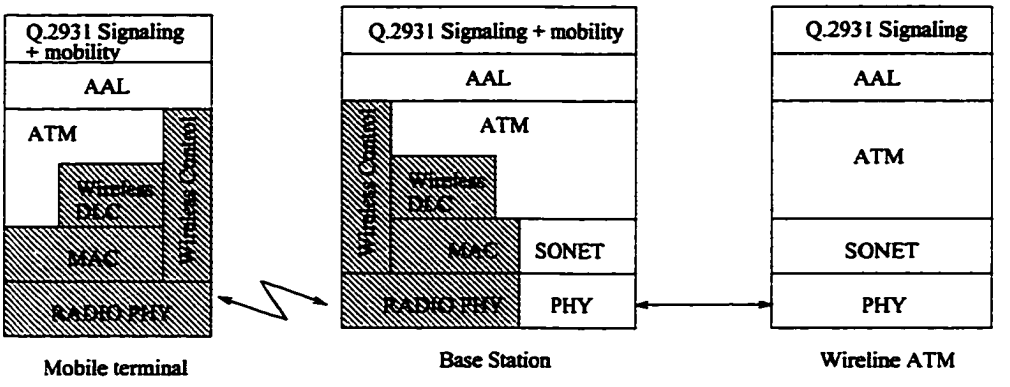

Figure 5.2. *Protocol stacks to integrate wireless ATM users to a wireline ATM network as illustrated in [3, 4].*

5.2 Integration of Proposed AFEC into Wireless ATM

The proposed AFEC described in Chapter 3 is more suited for wireless ATM networks since noise levels are high.

ATM cells generated by the user are packaged into radio frames. A frame could contain one or more ATM cells. These frames are then transmitted over radio channels to a base station as shown in Figure 5.3. The base station checks and converts back the radio frames to ATM cells in order to be handed to the wireline ATM network.

In order to overcome the errors due to channel noise, different schemes have been proposed to incorporate FEC into the radio frame. A detailed study of such schemes is found in [54]. The main drawbacks of those schemes are:

1. The correction rate of all the schemes used is constant which makes the schemes not suitable to variable noise channels.
2. Some proposals use multiple FEC schemes. Such approaches tolerate noise. However it is very costly since a radio frame is chopped into different segments where a different FEC is applied on every segment.
3. None of the schemes supports secure communication. In order to incorporate security, another level of operation is required.

To overcome the above drawbacks, the proposed AFEC can be used. The AFEC is used under the following assumptions:

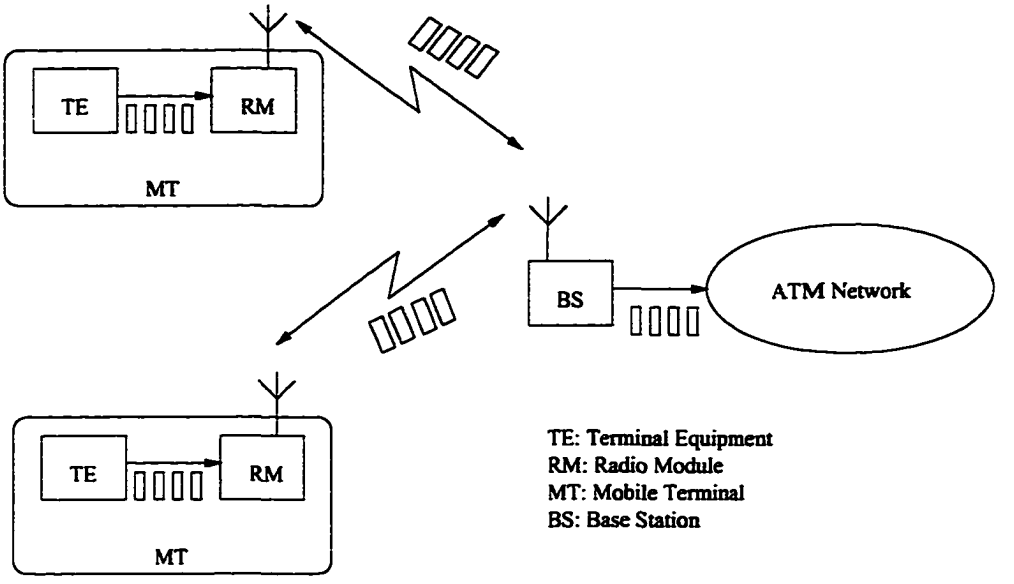

Figure 5.3. *The construction of wireless ATM.*

1. A control message is periodically sent from the central station to the radio modules. This message contains information about AFEC coding parameters and security keys.
2. Control messages are short and need not be of an ATM cell length. So for simplicity these messages are assumed to be reliably transmitted using ARQ schemes. Noise variation is assumed to take longer time compared to the time required to communicate a control message.
3. For simplicity security key length is assumed to be fixed. This choice still guarantees a computaionally secure cryptosystem.

The deployment of this coding scheme is illustrated in Figure 5.4. ATM cells are generated by the user equipment or coming from the wireline ATM network. These cells include HEC (Header Error Check) which is the standard error control scheme used in ATM cell headers. In the mobile terminal two actions take place. First the HEC field is removed from the cell (point 1 in Figure 5.4). HEC is removed since the function of this field is substituted by a more powerful technique. Second the AFEC parity is added to the block of cells (point 2). For proper operation of AFEC, a sequence number is added to the cell in place of HEC field. The added parity already holds the security information since the parity is calculated using a predefined security

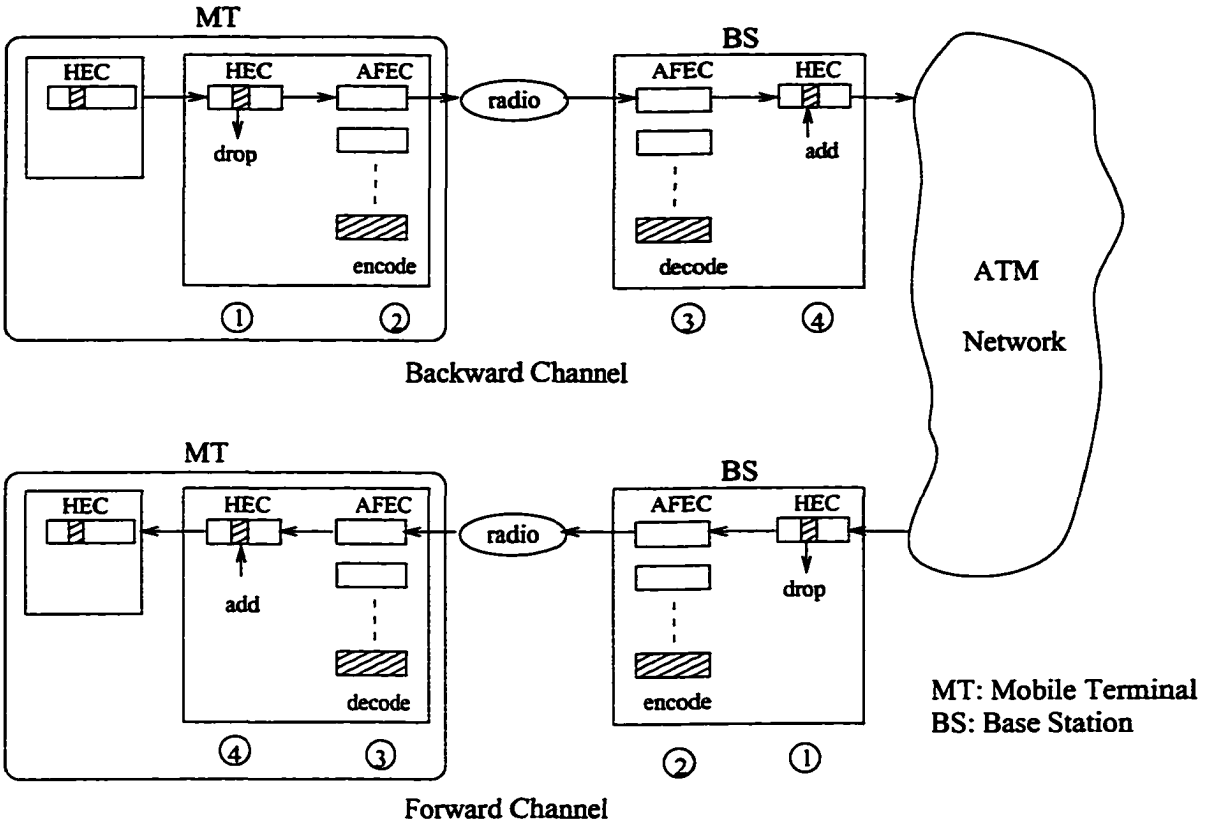

Figure 5.4. *The operation of the proposed AFEC scheme into radio ATM networks.*

key. The radio frame is transmitted over the wireless media.

Once the radio frame is received by the base station in the backward channel or the mobile terminal in the forward channel, it is decoded as shown at point 3 in Figure 5.4. The original frame payload will be retrieved if errors are within permitted range of the AFEC code and the correct security key is used. Otherwise, the frame is dropped. This action is recorded and the proper statistics are collected. The new coding parameters are periodically propagated back to the mobile terminal in the backward channel or the base station in the forward channel.

To fully recover the original ATM cell, HEC is calculated and added to the cell as shown at point 4 in the figure. In the backward direction, ATM cells are then forwarded to the wireline ATM network for further transmission.

Next we will study the CLR performance of this scheme.

5.3 CLR performance in Rayleigh fading

In this section the CLR performance of the AFEC scheme is evaluated in the environment described in the previous section. Boutillon *et al.* [2] define an efficient channel coding scheme as the one, which for an existing noise level, minimizes the signal power to achieve the desired error rate. We use this definition to study the bit error rate versus signal power for the AFEC.

A (15,k,l) is used over the block of cells. The radio frame is assumed to be composed of a maximum of 15 cells of which $k - l$ are data cells. Finally the channel rates are assumed to be 64 Kbps.

The signal is assumed to travel in a Rayleigh fading channel. In such an environment, there is no nonfading component and dominant components of the signal envelope tend to fade away [44, 55]. Moreover we assume that there is no direct path between the sender and receiver. Receiver will receive reflected or refracted echoes of the signal. The received signal power envelope then corresponds to a Rayleigh distribution [55]. The *pdf* of this distribution is given by [44]:

$$p(x) = \frac{x}{\sigma^2} \exp -\frac{x^2}{2\sigma^2} \tag{5.1}$$

where x is the receiving voltage and σ^2 is the variance. The cumulative distribution function (CDF) of the power envelope is:

$$P(X) = \text{Prob}(x \leq X) = \int_0^X p(x)dx = 1 - \exp \frac{-X^2}{2\sigma^2} \tag{5.2}$$

The CLR here is calculated with the assumption that QPSK modulation technique with noncoherent detection in additive white Gaussian noise (AWGN) channel. Therefore bit error rate (p_c) is given by [44]:

$$p_c = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \tag{5.3}$$

where $Q()$ is the Q-Function as defined in [44, 56], E_b is the energy per bit and N_0 is the white noise power spectral density.

In this scheme an ATM cell is lost if it is misrouted. This occurs when one or more bits in the header (excluding the HEC field) is erroneous. The case where errors

occur in the payload does not result in cell loss but it may deteriorate the quality of service.

The cell loss ratio (CLR) is improved due to the AFEC. Since there are 32 bits in the cell header which may contribute to the cell loss based on the above assumption, CLR can then be expressed in term of p_c as follows:

$$CLR = \sum_{i=1}^{32} \binom{32}{i} p_c^i (1 - p_c)^{32-i}. \tag{5.4}$$

CLR after AFEC is depicted in Figure 5.5. The correction capability of the AFEC is changing with channel noise and BER. Lower CLR is achieved with the use of AFEC. The staircase shape in the graph is due to parameter change in the AFEC code.

Another aspect is the CLR in response to the signal level. Figure 5.6 plots CLR versus signal to noise ratio for channels with Rayleigh fading. For the same signal to noise ratio, CLR is significantly improved when compared to no FEC.

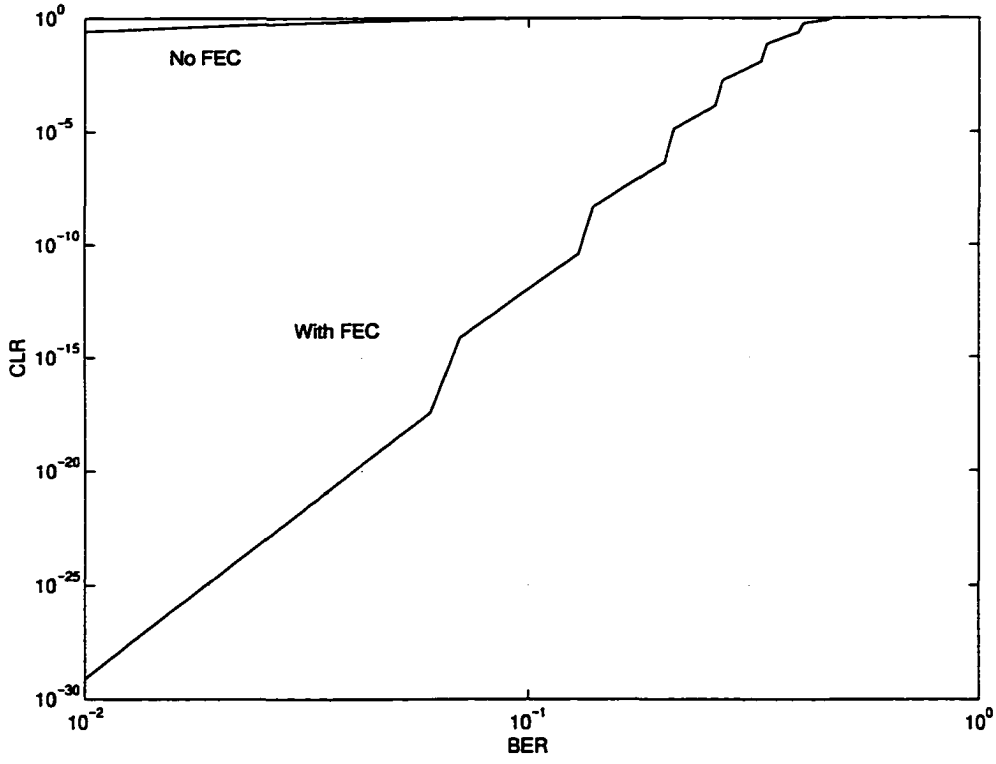

Figure 5.5. *CLR before and after application of AFEC.*

The correction capability of the AFEC is changing with channel noise and CLR. Lower CLR is achieved with the use of AFEC.

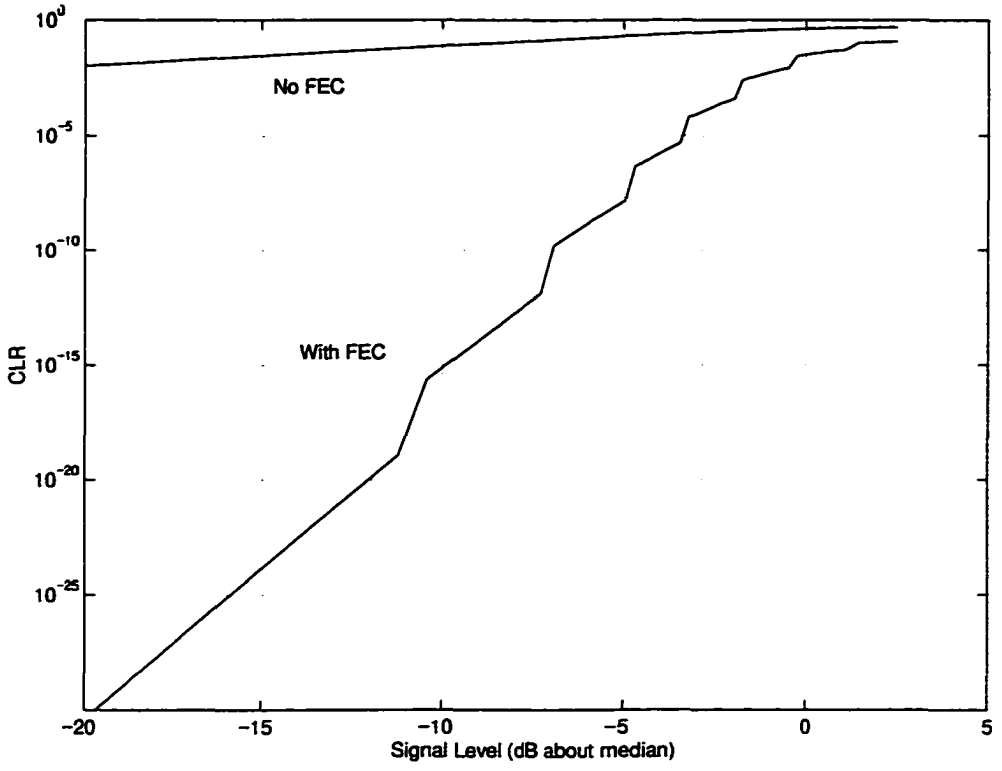

Figure 5.6. *AFEC improves CLR in Rayleigh fading media. A 64 Kbps channel is assumed with noncoherent QPSK modulation.*

5.4 Cocncluding Remarks

In this chapter wireless ATM has been explored and the principles of wireless communications were briefly discussed. For wireless communication to support ATM it needs to work in harmony with ATM infrastructure. This can be accomplished by exploiting the advantages of ATM and introducing new type of DLC and MAC protocols to deal with the wireless media limitations and challenges.

The proposed AFEC is then adapted to work in a wireless ATM environment. The AFEC has also been analyzed in Rayleigh fading channels. It has been shown that AFEC is an efficient channel coding scheme.

Besides AFEC tolerances to noisy channels, the framework has also take advantage of the security feature of the scheme.

Chapter 6

A Simulation Study: rt-VBR

In this chapter we simulate the proposed adaptive RS scheme in an ATM network supporting multicasting with real-time traffic.

In addition for ATM to carry Internet traffic, ATM has to provide multicast and broadcast services. Multicasting is the ability to send one message to one or more destinations in one operation. Diverse applications will make use of multicasting such as distance learning.

Different approaches are used for multicasting in an ATM switch. Examples of ATM switches with multicasting capabilities, found in the literature, are [57], [58] and [59]. The last reference gives a thorough review of multicasting engines in ATM communications. The basic idea behind these schemes is to be able to duplicate an incoming ATM cell to as many outputs as the set of destinations forming a multicast group. A copy network (CN) is mostly used in ATM switch fabrics that are not bus based [60]. Associated with an ATM multicasting switch is a factor called fanout. Fanout, which is defined as the maximum number of copies that a CN can do in one operation, is a critical design issue [60]. This is mainly due to technology limitations such as bus width as in the Knockout switch [60].

To overcome the technological limitations due to multicasting of real-time traffic, error control mechanisms are used at a layer above the ATM protocol suite to ensure proper functionality and guaranteed quality of service (QoS). Examples of some of the proposed schemes include SMART [61], RMTP [62], MESH [63], and destination set splitting [64]. These schemes, in general, utilize window-based flow control mechanisms. An apparent disadvantage is when the network is congested. In this case one or more receiver will starve or even worse they will not be guaranteed the negotiated QoS.

A remedy for this case is to use FEC schemes which can be used as an end-to-end protocol to regenerate lost cells at the receiver side of the network. Since each destination experiences a different congestion state, adaptive FEC schemes are advised as discussed in Chapter 3 and Ref. [65].

In this chapter we compare window-based flow control schemes represented by SRP and forward error correction schemes represented by an AFEC mechanism. Both mechanisms are simulated to deliver real time VBR (rtVBVR)traffic through ATM networks with multicasting. The choice of rtVBR is because video traffic such as MPEG, is variable in nature [66].

Traffic sources used in this study are both artificial and real world sources. These sources, as well as simulation setup, are discussed in Section 6.1. Throughout the sets of simulation conducted, multicasting is considered as a standard feature to be supported. Moreover the switches are assumed to be fanout limited and that cell duplication is not done in one operation. This is discussed in Section 6.1. In Section 6.2, ATM switch resources, mainly buffer spaces, are discussed. To lower the load on ATM switches, buffering may be handled at the edge of the network. This is also discussed in Section 6.2.

An important aspect of ATM handling real time traffic is cell transfer delay (CTD). Real-time traffic imposes stringent requirements mainly on delay. If cells exceed a certain delay value, they become useless to the destination application. This is discussed in details in Section 6.3. Conclusions and recommendations are found in Section 6.4.

6.1 Simulation Setup and Modeling Preliminaries

To conduct our study, The network configuration shown in Figure 6.1 is used throughout the simulations. There are two basic components in the above network: traffic generators and sinks, and switches. Each will be looked into closely. In order to ensure proper protocol behavior, protocol bodies are embedded in traffic generators and sinks. These protocol bodies will also be discussed. All links are assumed to be of OC-3 (i.e. 155 Mbps) rate with $1us$ propagation delay unless otherwise stated.

In the above network more assumptions are found in the following sections.

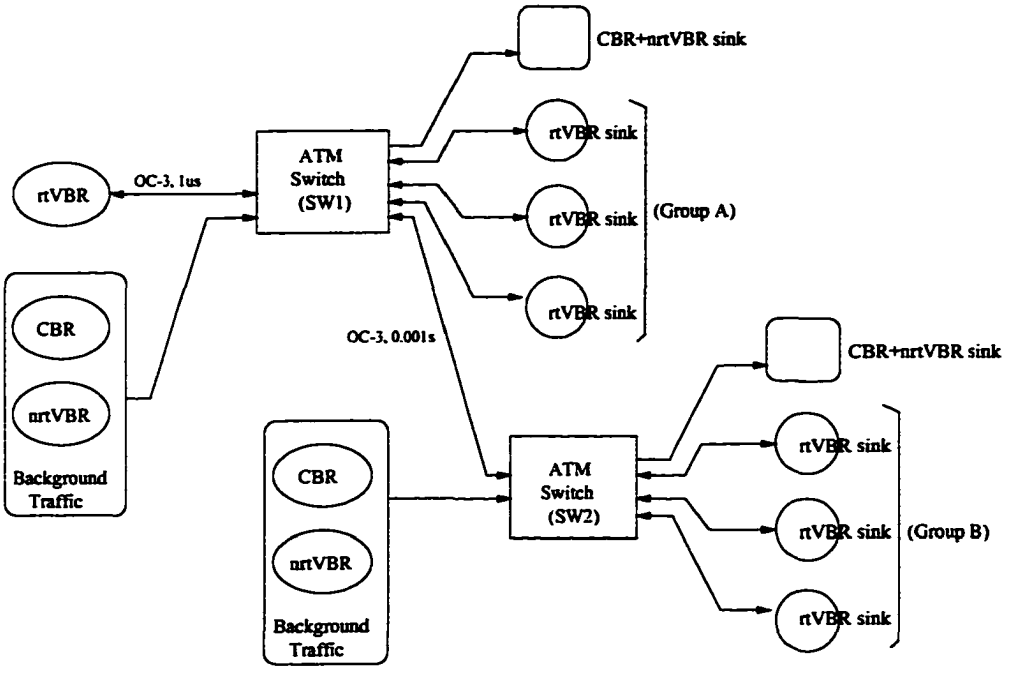

Figure 6.1. *Configuration assumed to study error recovery schemes on ATM networks with multicasting capabilities.*

6.1.1 Traffic sources

In this simulation we are primarily investigating:

1. The impact of error recovery schemes on real-time VBR traffic, and
2. the impact of real-time traffic utilizing those schemes on ATM switches with multicasting.

For this purpose, real time traffic are generated either from on-off sources or from real traffic traces. Each is described below.

On-off sources:

On-off sources are widely used since they are easily coded. In the literature, references [67, 68] discuss how such models can be used to generate video traffic. Two types of on-off sources are used in this study: rtVBR on-off sources and nrtVBR on-off sources. For rtVBR traffic 20 on-off sources corresponding to video sources are used with mean bit rate of 768 Kbps and peak bit rate of 4.608 Mbps [69].

For non-rt VBR traffic, an on-off model with source activity of 0.077 and peak rate of 85 Mbps is used. This is intended to resemble a general low activity VBR

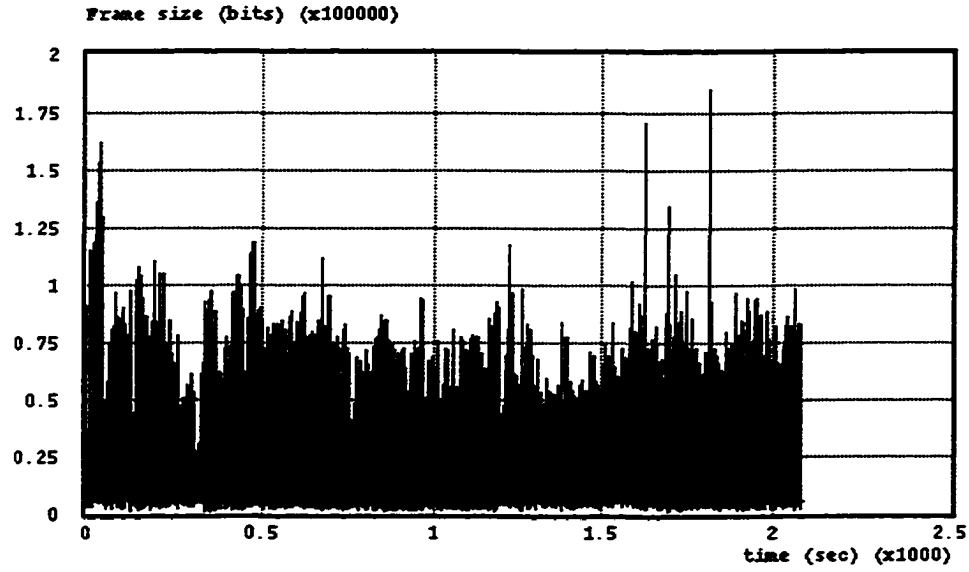

Figure 6.2. *Traces from "Star War" movies as found in [5].*

source with no particular application.

Traffic traces:

Real traffic traces are of importance since they are collected from real world video sources. In addition traces are more accurate and provide more reliable results. For this reason, an MPEG traffic trace taken from "Star War" movies as described in [70] is used. The traffic trace is shown in Figure 6.2.

Beside VBR traffic sources, constant bit rate (CBR) traffic is injected as a background traffic. CBR traffic is generated such that it occupies approximately 40% of the total link bandwidth.

6.1.2 ATM switch

ATM switches that are modeled here are general ones and do not imply any technology or methodology. Each switch has input ports that accepts traffic of different classes of service. Traffic is buffered into the corresponding class of service.

The switch is an output buffer type. It has four classes of service queues: CBR, rtVBR, VBR and UBR. Since we are not interested in UBR traffic, this class of service is not used although it does exist in our switch model for completeness purposes. The

servicing of these queues is in a round robin fashion with weight associated to each class. The higher the class weight the more frequently it is serviced; CBR has the highest weight, then rt-VBR, and so on. Details on weighted round robin are found in [13].

In this study multicasting is implemented for rtVBR traffic only. In Figure 6.1, a sink represents a set of destinations. Each sink corresponds to the maximum fanout that a switch can multicast a cell into at any time. The assumption here is at each cell time a multicast cell is transmitted to one sink only. This means that a cell is transmitted to D sinks would require D cell time slots. Sinks are grouped into Group A and Group B. Group A corresponds to one hop delay while Group B corresponds to two hops delay with large propagation delay. The intention with such scenario is to study the worst case a real time traffic could experience. Therefore in studying the cell transfer delay (CTD), Group B delay is observed.

6.1.3 Error recovery protocol bodies

These bodies are implemented to be part of the ATM adaptation layer or above it and their functions are to properly implement a protocol and to guarantee correct protocol data units (PDU) delivery to higher layer applications.

As mentioned earlier, two error control schemes are used in this work: SRP and AFEC. In SRP when a window is sent out, a timer is activated. The window slides as an acknowledgment (ACK) or negative acknowledgment (NACK) is received. However if the timer expires before receiving an ACK, this means that the cell is not received correctly. In this case the same cell is sent again and the timer is reset. Ref. [9] has more details on such protocols. In this thesis, the notation SRP/n is used to represent SRP protocol with window size of n cells. The timeout value is set to 3ms which is 3 times the worst-case propagation delay.

For AFEC schemes we assume Reed-Solomon (RS) erasure codes which are different from the classical RS codes in that the location of errors are known and hence no attempt to find them is required [21]. Moreover we are assuming these codes are adaptive as been described in Ref. [65] and Chapter 3. AFEC schemes are denoted by (n,k,l); where n is the length of the code word, k is the length of data word and l is the number of zero symbols. AFEC is applied on cells; this means that n, k and

l are expressed in units of cells. In the event of cell loss, destinations feedback traffic sources with their new coding parameters since this cell loss is due to probable network congestion and only sensed by destinations. This is intended to protect future data from loss.

To accomplish correct protocol behavior sources and destinations communicate their control information, such as acknowledgments in RSP and new coding parameters in AFEC, in a reliable backward channel. This control information is prioritized and for simulation purposes could be sent over a different channel that does not go through the switch but has same channel characteristics like propagation delays.

6.1.4 Simulation setup

The network shown in Figure 6.1 has been coded using OPNET. OPNET is a network simulation tool that uses C-like programming as its modeling language [71]. The different components described in the above sections are also coded in OPNET. The simulations were conducted on a SUN Ultra 1 machine.

In the remaining of this chapter we show the results of the simulations conducted.

6.2 Switch Resources

In order to study the switch buffering required, switches are set such that there is no cell loss due to limited buffer sizes.

When using SRP/256 and AFEC (255,207,l), both schemes require almost the same queue sizes. In fact AFEC occupies more spaces due to redundant cells. Figures 6.3 and 6.4 depict the queue sizes for switches SW1 and SW2, respectively, when SRP/256 is used. Queue size in SW1 reaches its peak at around 0.5 sec. This is due to the sources activity specially the bursty ones at around this time. The queue size in SW2 (Figure 6.4) does not have any queue growth as experienced in SW1. As such it is deduced that congestion in the switches of this simulation study is not correlated among adjacent switches. This is to say that if congestion occurs at a specific switch, it is not necessary that other switches along the path will also be congested.

Since SRP and AFEC are end-to-end mechanisms, it is necessary to monitor queue buildups at the edges of the network. In the case of AFEC there is no need to buffer

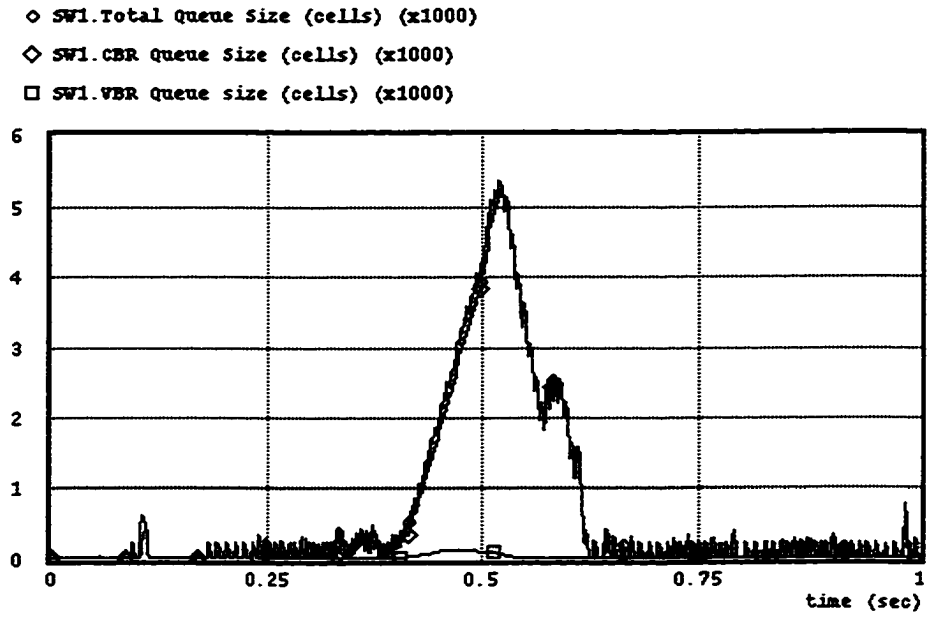

Figure 6.3. *Total queue occupancy in SW1.*

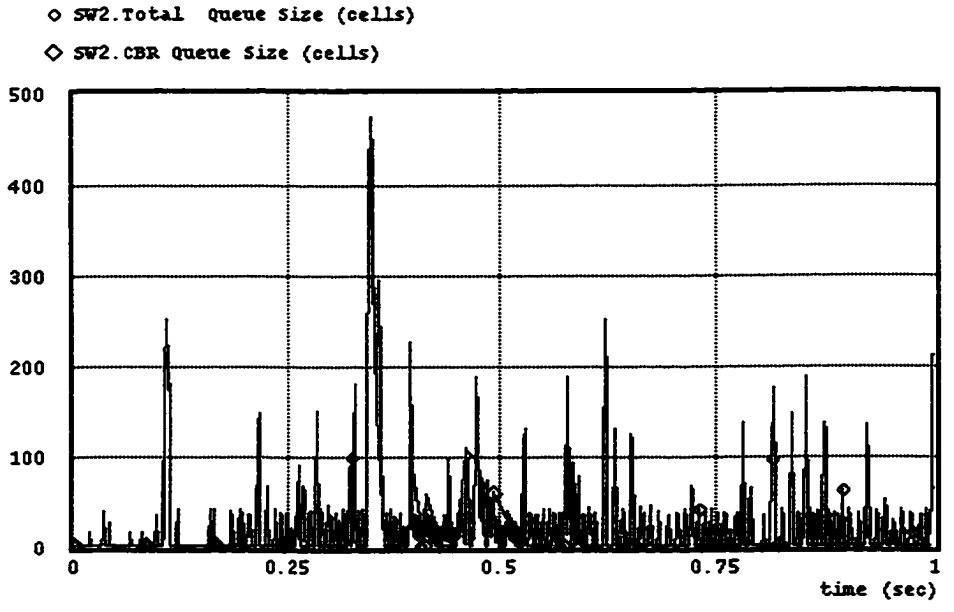

Figure 6.4. *Total queue occupancy in SW2.*

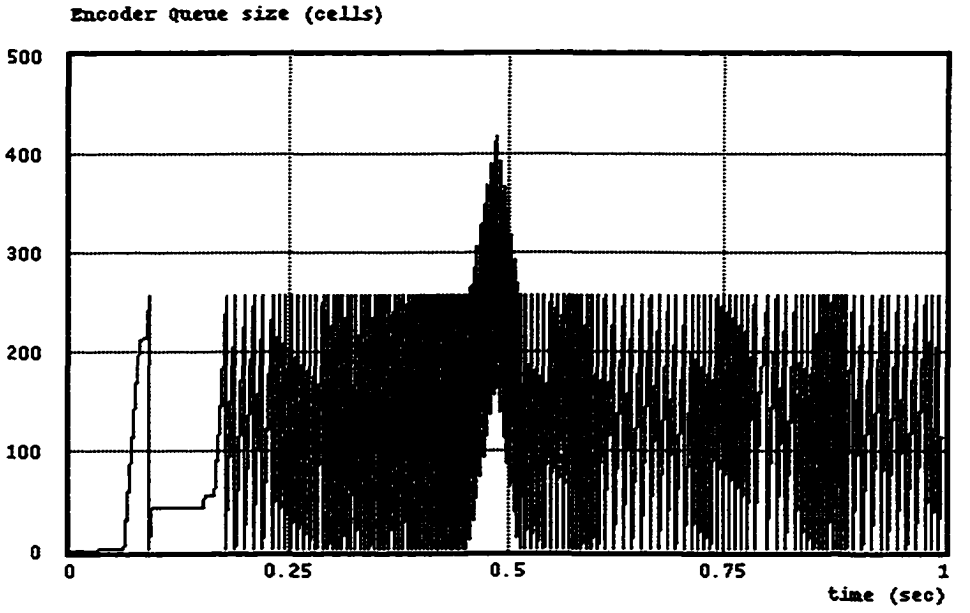

Figure 6.5. *Buffer spaces occupied at the network edge for SRP/256.*

cells. Cells are transmitted downstream. A cell copy is only needed for the purpose of generating redundant cells.

On the other hand, SRP needs a queuing mechanism since traffic sources will not send out the next window till a correct delivery of the current window is received. Figure 6.5 shows the buffer spaces in cells at the edge of the network when using SRP/256. In Figure 6.5 traffic sources may generate cell streams that are larger than the window size. The excess traffic is therefore queued for the next windows. This is shown in Figure 6.5 around time 0.5 second.

It is interesting to observe that switch buffer sizes are reduced when SRP with lower window sizes are envisioned. Switch buffer sizes in the case of SRP/64 are depicted in Figure 6.6. This is mainly due to the fact that SRP/64 controls the flow of the traffic to be injected into the network in blocks of 64 cells. When compared to Figure 6.3, significant reduction in buffer sizes is observed.

The drawback, however, is that the edge queue will grow large. This is shown in Figure 6.7. Although it seems that SRP/64 is more advantageous, SRP/64 transferred the queuing problem to the edge of an ATM network. This queuing introduces notable delay to real-time traffic and need to be considered.

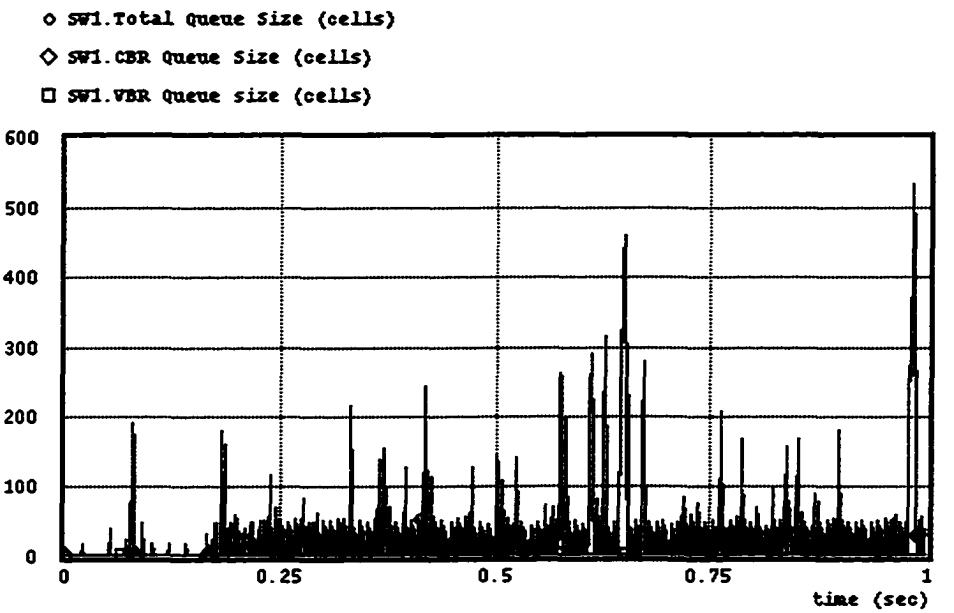

Figure 6.6. *Smaller buffers at switch SW1 are required for SRP with lower window sizes (SRP/64).*

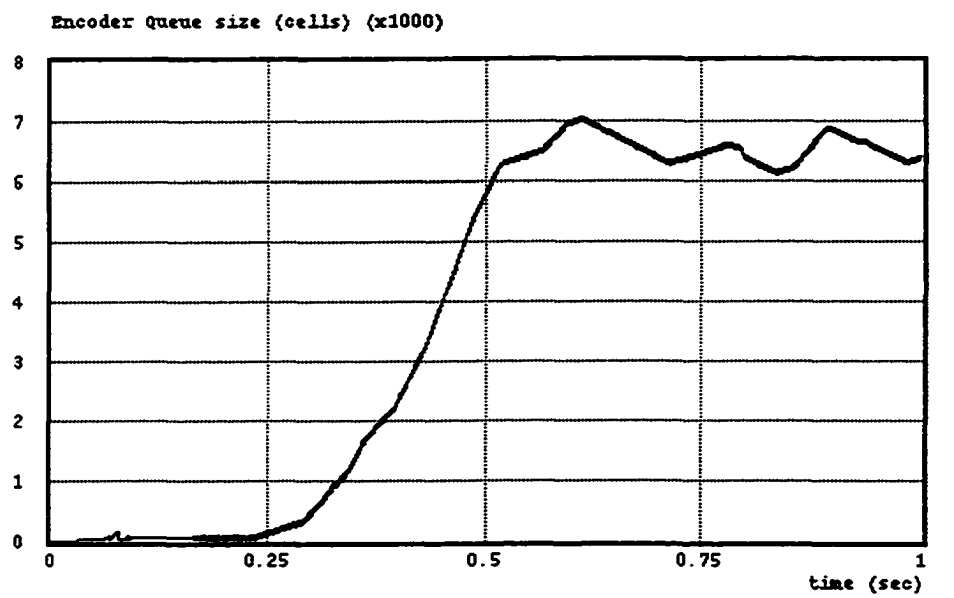

Figure 6.7. *Queue sizes at the edge of the network when SRP/64 is used.*

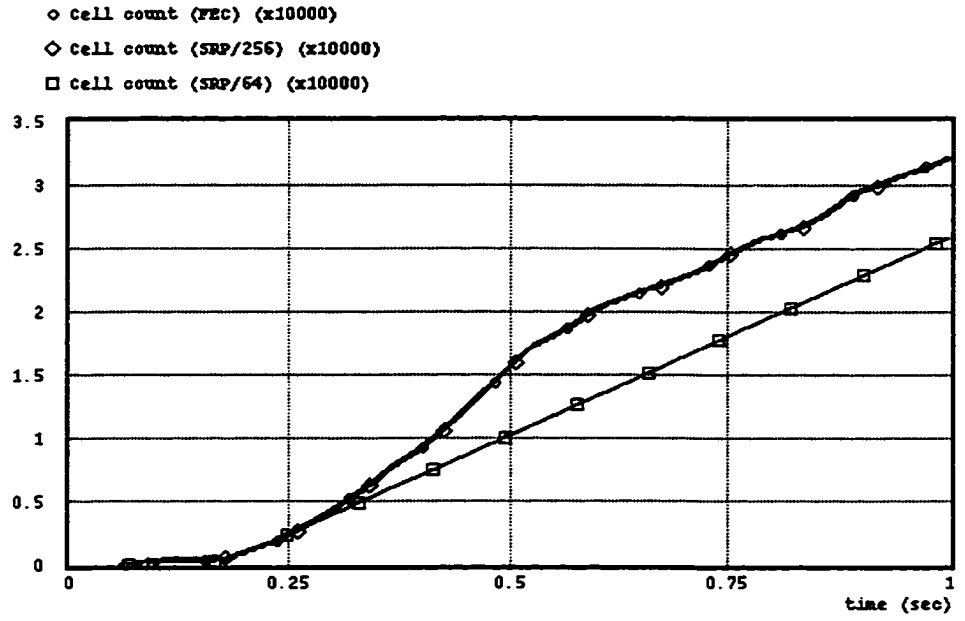

Figure 6.8. *The maximum number of correct cells received during a single simulation run for SRP/256, SRP/64 and AFEC.*

Another observation worth mentioning is the maximum cell count that has been received correctly by the destination application. Figure 6.8 shows cell count for SRP/256, SRP/64 and AFEC. The cell count is being calculated as the number of correct cells delivered. The figure shows that SRP/256 and AFEC are comparable in terms of goodput. Whereas SRP/64 has a lower goodput. One reason to such degradation in goodput is the window size.

6.3 Delay Issues

In this section delay issues are investigated. The delay of interest is CTD which is measured from the source switch entry point to a destination.

To exercise SRP and AFEC protocols, limited switch buffers are assumed. Initially switch queue thresholds are set to 500 cells. When queue threshold is reached cells are dropped except for CBR and end of message (EOM) cells which are queued. In this work, EOM is primarily used in AFEC though it can be avoided.

Before we observe cell delay, let us observe the queue dynamics. In the case

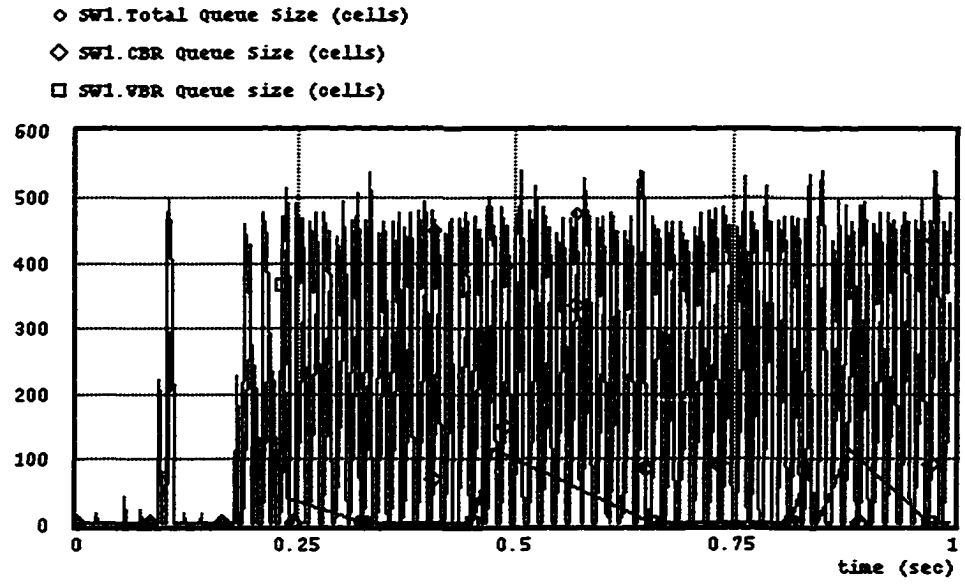

Figure 6.9. *Queue sizes in switch SW1 with threshold of 500 cells using SRP/256. Occasionally queue sizes exceed their threshold for reasons explained in text.*

of SRP/256, switches (SW1 and SW2) queues are shown in Figures 6.9 and 6.10. Both figures verify the proper functionality of the switches by not exceeding their thresholds. Queue sizes may occasionally exceed their threshold (see Figure 6.9). This is due to the fact that CBR cells are not dropped. The corresponding queue sizes in the case of AFEC are shown in Figures 6.11 and 6.12. When comparing Figure 6.11 with Figure 6.9, one could observe that lower buffer occupancies are achieved with AFEC. This verifies the results shown in [20] and Chapter 2.

CTD is measured for cells directed to Group B. The motive is to observe the worst delay experienced by the traveling cells to reach their destinations. Figures 6.13 and 6.14 show CTD for SRP/256 and AFEC, respectively. These figures show that when AFEC is used, significantly low CTD is attained. CTD has been reduced by almost an order of magnitude. This is despite the fact that no queues at the network edge are needed. In the case where SRP is used, queues are needed and they grow large due to the dynamics of SRP and the rate at which an application generates its data. Queue size at the network edge in this case is shown in Figure 6.15.

To study the limitations of SRP, queue thresholds are reduced to a value lower than the window size, say to 200 cells. In this case SRP/256 will starve since most of

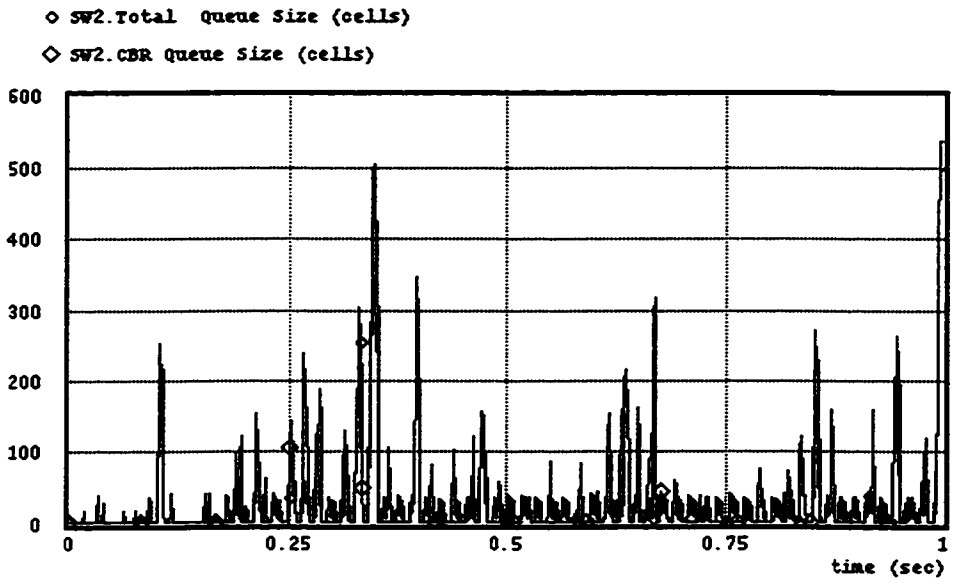

Figure 6.10. *Queue sizes in switch SW2 with threshold of 500 cells using SRP/256. Occasionally queue sizes exceed their threshold for reasons explained in text.*

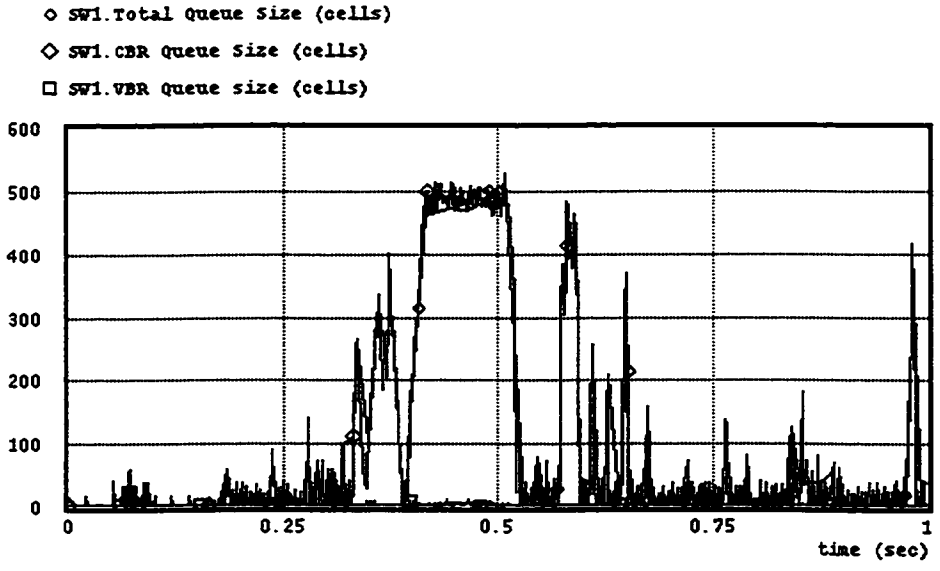

Figure 6.11. *Queue sizes in switch SW1 with threshold of 500 cells using AFEC. Occasionally queue sizes exceed their threshold for reasons explained in text.*

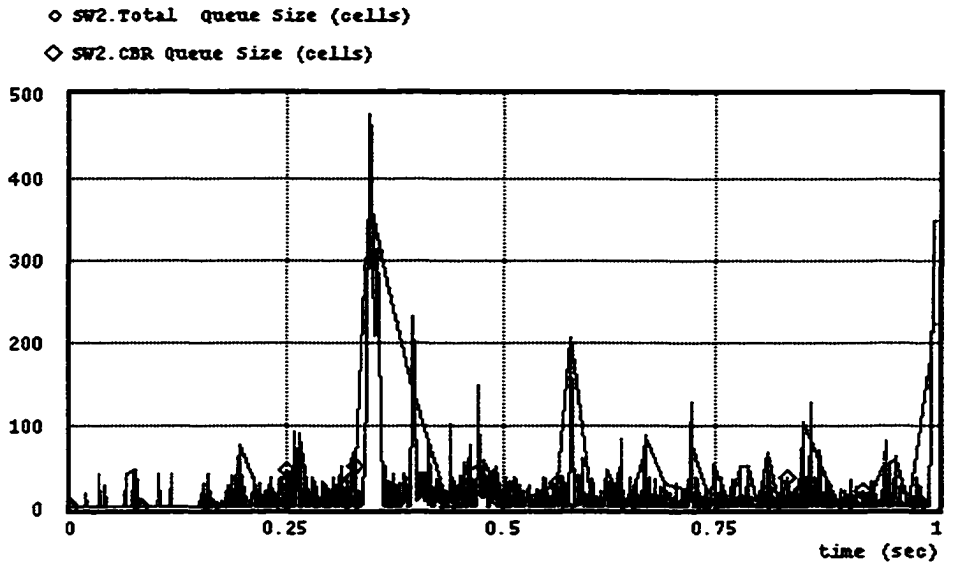

Figure 6.12. *Queue sizes in switch SW2 with threshold of 500 cells using AFEC. Occasionally queue sizes exceed their threshold for reasons explained in text.*

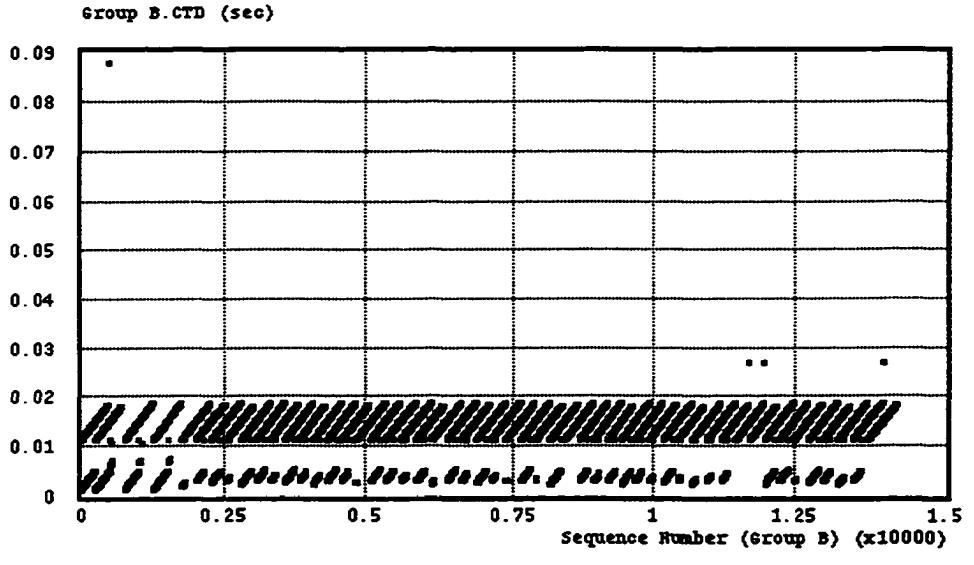

Figure 6.13. *CTD measured for Group B when SRP/256 is used with switch buffer size of 500 cells.*

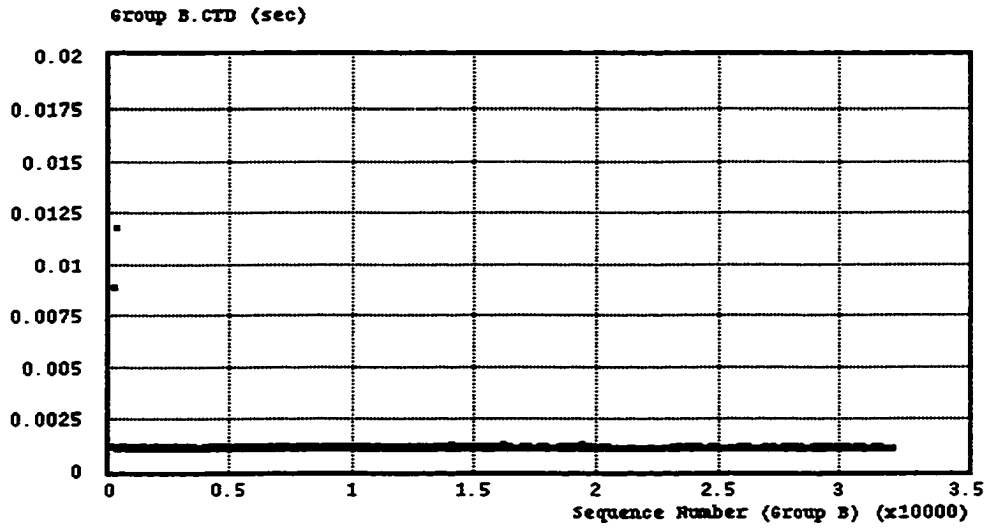

Figure 6.14. *CTD measured for Group B when AFEC is used with switch buffer size of 500 cells.*

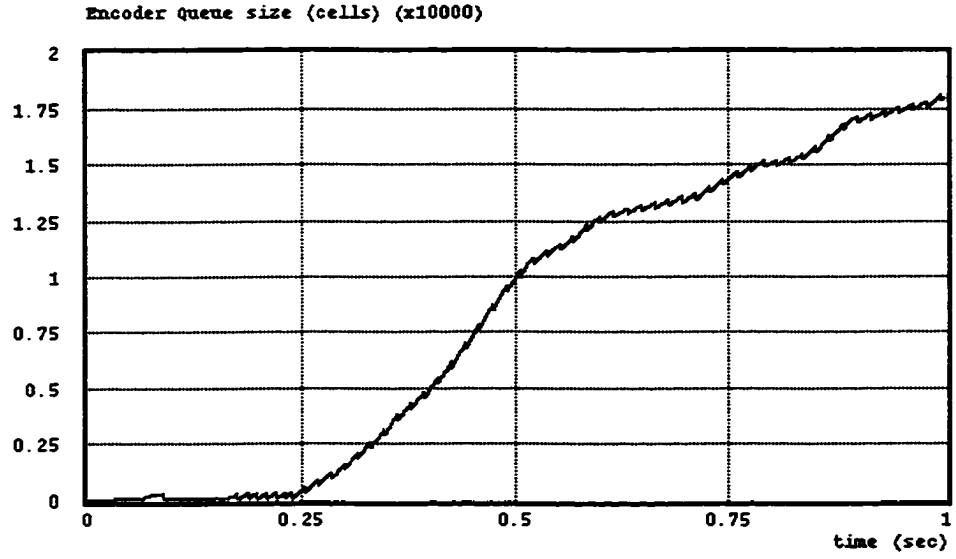

Figure 6.15. *Queue sizes as observed at the edge of the ATM network when SRP/256 is used.*

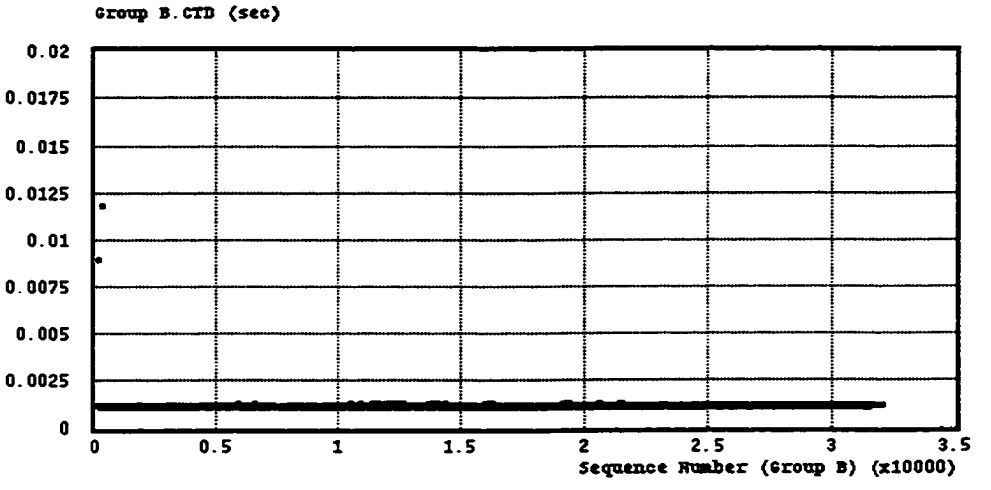

Figure 6.16. *CTD when AFEC is used over a switch with queue threshold set at 200 cells. CTD measurement does not count AFEC decoding delay.*

the cells in a window are dropped. This will activate the retransmission mechanism many times. The worst case that could happen is when multiple timeouts occur.

On the other hand, when AFEC is used there has not been any significant increase in CTD. In fact, CTD is observed to be reasonably low taking into account the fact that when AFEC is used, decoding delay at destination may be significant. Figure 6.16 shows CTD in the case of AFEC, for simplicity no decoding delay is assumed. When Figure 6.16 is compared to Figure 6.14, no change in CTD is observed. The justification for this is that as soon as some cells are lost AFEC changes its parameters to reflect the new network state. The redundancy does decrease the throughput; however CTD is not affected. The conclusion derived from this is that AFEC has a very deterministic delay on traffic.

In the case of SRP, lower window size is a possible remedy for the cases of switches with small queues. A possible configuration is using SRP/128. Low CTD is achieved in this case but still much greater than the case using AFEC. This is depicted in Figure 6.17. In general when window-based ARQ schemes are used, CTD changes as the window sizes and threshold values change.

MPEG traffic

To study the impact on real traffic, traces of MPEG video as described in Section 6.1.1 is used. A 100 second video segment of the whole video is extracted and

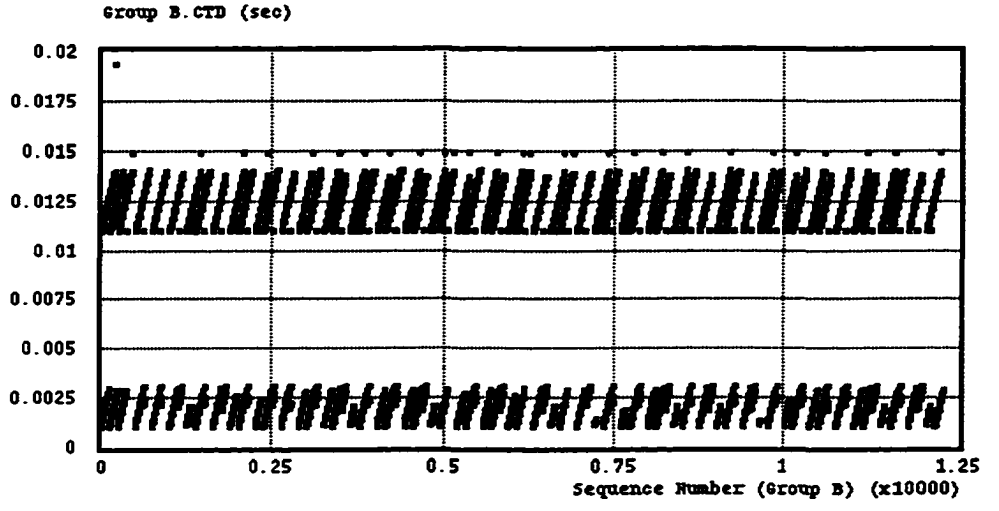

Figure 6.17. *low CTD is achieved with SRP/128 when switch threshold is set to 200 cells.*

used for this simulation.

Buffer requirements at switch (SW1) and at the edge of the network when SRP/256 are shown in Figures 6.18 and 6.19, respectively. Buffer occupancy in the case of AFEC is shown in Figure 6.20. From Figures 6.18 and 6.20, one can observe that when AFEC is deployed buffers occupancy is less than the case where SRP/256 is used.

In studying CTD, Figures 6.21 and 6.22 show the delay experienced by cells during simulation time. From looking at the graphs it is obvious that AFEC accomplishes lower CTD values. In fact CTD when SRP/256 is used is more than ten times greater than when AFEC is used. The throughput of both schemes, on the other hand, is comparable.

6.4 Concluding Remarks

In this chapter we studied AFEC schemes over ATM multicasting networks. The study is conducted using OPNET network simulation tool.

Several observations are derived from the study:

1. The proposed AFEC scheme outperforms ARQ schemes when multicasting is

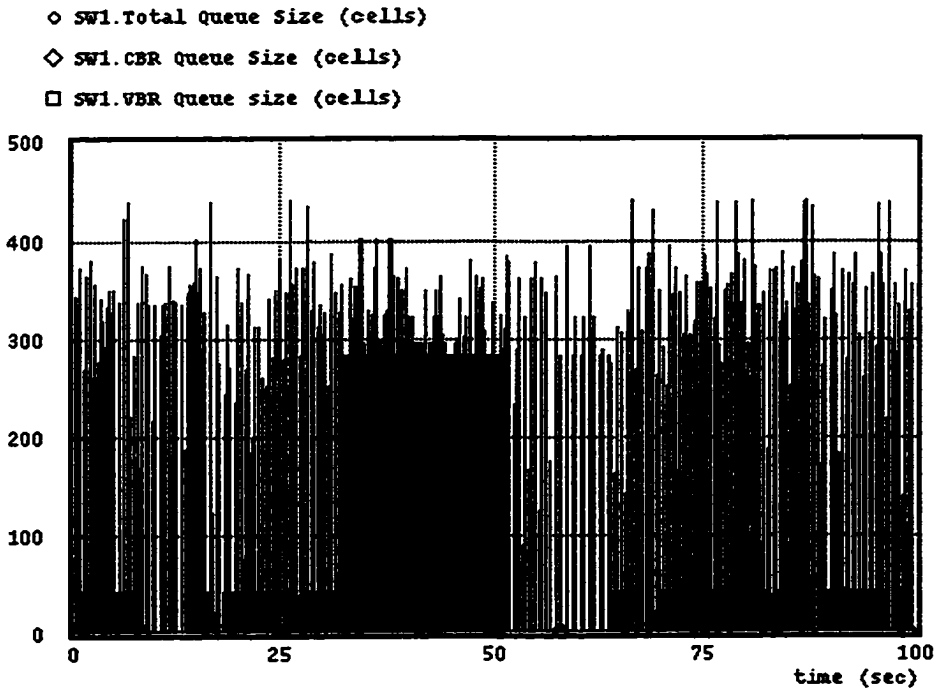

Figure 6.18. *Queue sizes for SW1 with threshold value of 400 cells when SRP/256 is used. Application is MPEG video traces.*

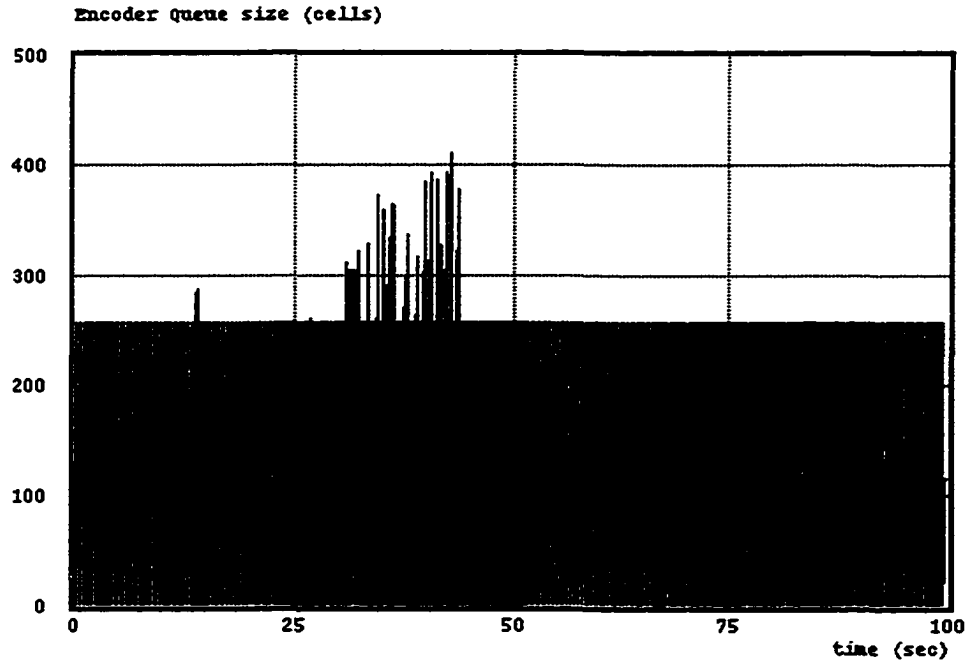

Figure 6.19. *Queue sizes at the edge of the network when SRP/256 is used. Application is MPEG video traces.*

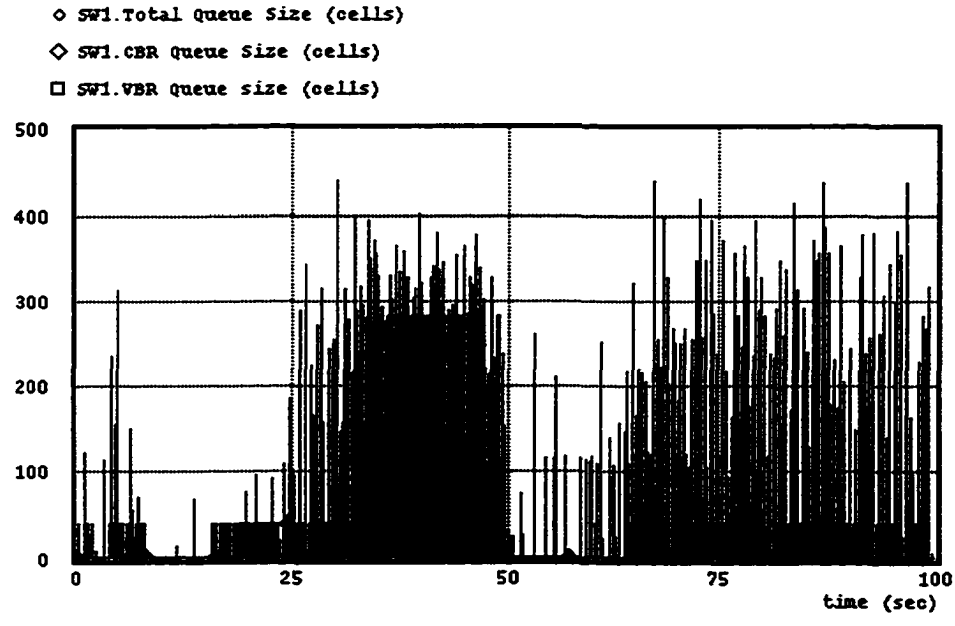

Figure 6.20. *Queue sizes for SW1 with threshold value of 400 cells when AFEC is used. Application is MPEG video traces.*

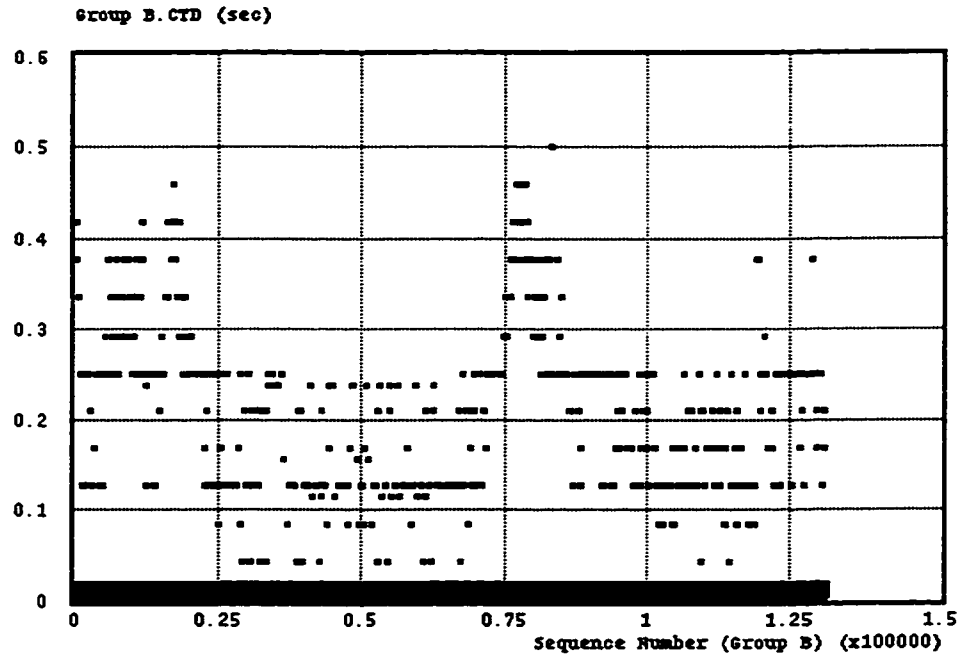

Figure 6.21. *CTD observed by destination for the MPEG traces when SRP/256 is used.*

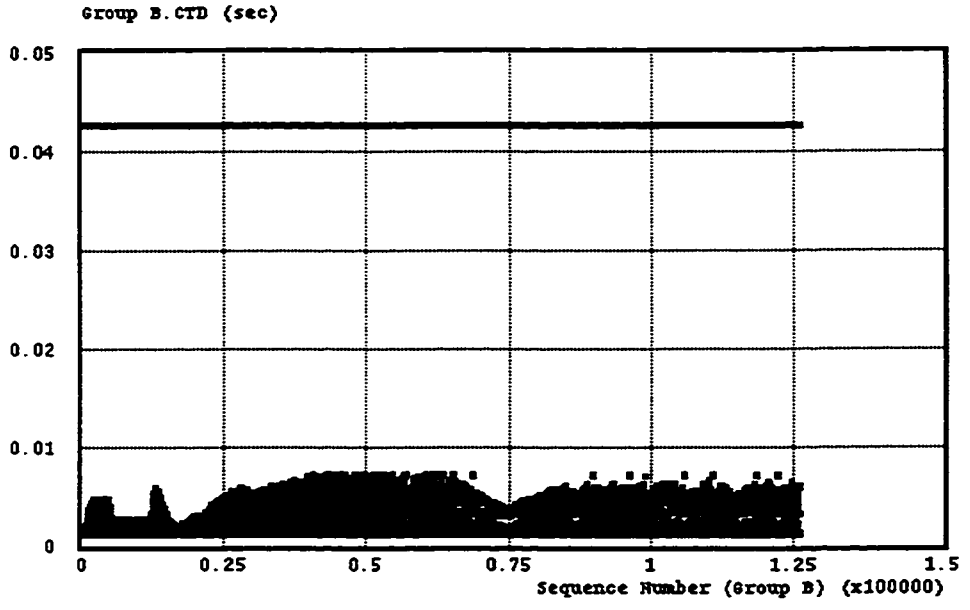

Figure 6.22. *CTD observed by destination for the MPEG traces when AFEC is used.*

considered. With multicasting each destination will have different pattern of the cells it losses. This suggests that time-out events for SRP protocols are more probable to occur in multicasting.

2. For ARQ scheme to perform well in multicasting environments, adaptive window schemes are needed such as the ones described in [72] and [73].
3. Less buffering, and sometimes none, is needed at the UNI points of the network when AFEC scheme is used.
4. More traffic is carried when AFEC is used. This means that real-time multicasting application will be able to achieve higher throughput.
5. Cell transfer delay (CTD) is improved when AFEC is deployed over multicasting networks. It is needless to mention that CDT is upper bounded in the case of AFEC; whereas CDT is changing when SRP protocols are used as a function of the window size and the threshold values of the switches in the network.

Chapter 7

Switch Design Requirements Under FEC Environment

In Chapter 2 we showed that ATM network resources are improved when FEC is deployed. This apparent advantage of using FEC implies potentially simple ATM switch design since smaller buffer sizes and yet simpler switch management mechanisms are required. Moreover it has been shown in Chapters 4 and 6 that end-to-end delay is improved by upper bounding its CTD. Other advantages include cost effectiveness and simple implementations of multicast services. Deterministic cell delay and higher throughput are two major enhancements to be seen by an application running over ATM networks.

In this chapter we will investigate the potential requirements of switch design under the influence of FEC deployment. First a list of the general requirements of an ATM switch design is given in Section 7.1. Based on the impacts of using FEC codes in ATM switch design which are discussed in Chapter 2, two switch approaches are given in Section 7.2. The tradeoffs between these approaches is then are discussed.

7.1 General switch design requirements

Designing ATM switches involves many engineering decisions. These decisions affect the switch performance and cost. Different switch designs implies different design requirements. Generally, in designing an ATM switch certain factors are to be considered. These factors are:

1. Delays: Delay encountered by ATM cells is an implementation issue. An ATM switch should carry cells from an input port to specific output port(s) in reason-

able delays. The major delay components in ATM switches are: segmentation and reassembly delays, and switching and queuing delays.

2. Strategies: Two types of strategies are considered when designing ATM switches: switching and queuing strategies. Examples of switching strategies which also called switching fabrics, include: crossbar, multi-stage interconnection networks, bus, etc.

   Queuing strategies involves decisions on whether single or multiple queues (buffers) are appropriate. Moreover queuing strategies may specify whether to use input, output, or shared buffers. For example single, shared buffering strategy has lower cell loss characteristics, but greater average delays.

3. Management and control: An ATM switch requires management and connection admission control mechanisms to maintain proper functionality of the switch and manage the switch resources such as buffers. These mechanisms can either be global or distributed [12].
4. Switch scalability: A good switch design is scalable in terms of size and speed.
5. Switch area: The silicon area of an ATM switch, in particular single chip switch design, is of concern to designers. Limitations are imposed by the technology used. Switch area usually limits the availability of resources and hence limit the scalability of the switch.
6. Bus width: The size of the switch internal bus, if any is used, is an important factor to count. It affects the switch speed as it regulates the usage of its resources.
7. Buffer size: Unlimited buffer sizes guarantee congestion-free switch design. However, this is not possible and impractical. A switch has always limited buffer size.

7.2 ATM switch design under FEC

Having considered buffer size and switch capacity as discussed in Section 2.3.1, two possible approaches in designing ATM switches can be investigated [20, 74]. In the first approach, type A, parts of the FEC coding scheme are implemented in the switch. The emphasis in this approach is to overcome the channel noise problem which may

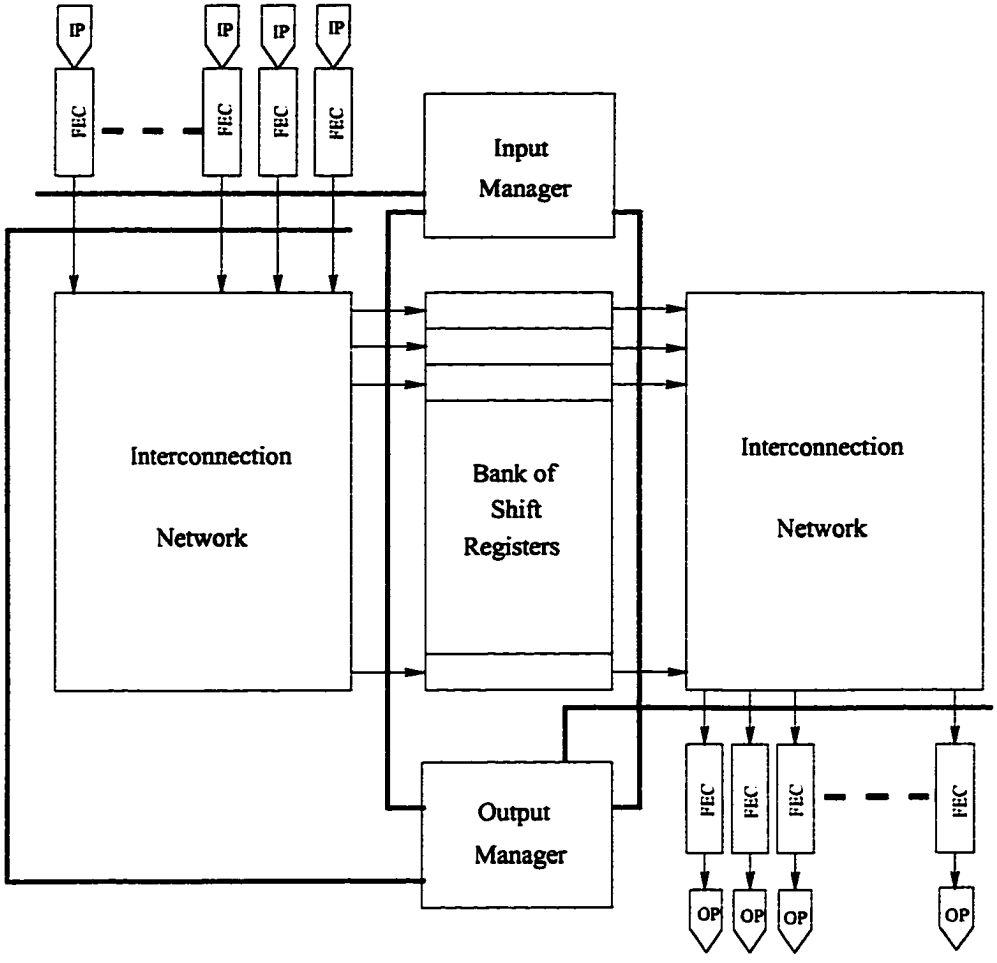

Figure 7.1. *Functional blocks of an ATM switch of type A. FEC decoders are implemented at the input (IP) ports of the switch. Whereas FEC encoders are implemented at the output (OP) ports.*

result from satellite channels, for example. This is depicted in Figure 7.1.

Type A switches assume noisy channels or paths. FEC schemes are used to compensate for cell lost due to channel noise. Such switch configurations are candidate for ATM communications using satellite channels or to support wireless ATM as discussed in Chapter 5.

Another approach, type B, is to design the switch with limited resources. Unlike the previous approach, FEC schemes are applied end-to-end. FEC encoders and decoders are only located at sources and destinations of data (i.e. at the UNIs). The assumption here is that cell loss will be due to either congestion or channel noise.

With such configuration, switch resources such as buffers are reduced conditional to maximum CLR. This maximum CLR guarantees that lost cells will be compensated by the FEC schemes.

In type A switches, FEC schemes are used to decrease CLR which is primarily due to channel noise. This does not guarantee small buffer sizes. Instead it may require larger buffer sizes in order to accommodate incoming cells. On the other hand, type B switch design can acquire buffer sizes that are lower than what is required normally. This is due to the fact that FEC schemes will compensate for lost cells.

Type A switch designs will be slower than type B switch designs. This is mainly because that FEC scheme encoding/decoding delays are on link-by-link basis and this delay is accumulative. Type B, on the other hand, produces lower delays because of two factors: FEC is applied end-to-end, and smaller queue size is required. Table 7.1 compares both switch design approaches. The row, entitled "Ranges to span", means whether the switch design can be used in an end-to-end communication or the FEC should be used on segment-by-segment basis.

7.3 Concluding Remarks

In this chapter, general ATM switch design requirements are outlined. These requirements are controlled when using FEC schemes. Using FEC schemes is advantageous since they are capable of regenerating lost cells. This implies that less resources are required. For example, buffer spaces can be reduced. We also showed that FEC schemes are used to reduce the effect of congestion while using reasonable buffer sizes.

QoS is indeed improved when FEC schemes are applied. More connections (CBR and VBR) can be accepted by an ATM switch at no extra buffer size and no QoS degradation.

Two ATM switch design proposals, namely type A and type B, have been given. It is observed that type B switch design is more appropriate for wireline ATM networks while type A is more suitable for wireless ATM.

Table 7.1. *Comparison between Type A and Type B switch designs.*

	Type A	Type B
Source of cell loss	Channel noises	network congestion as well as channel noises
Environment to be used in	Noisy channels such as satellite	Wireline or less noisy channels such as twisted pairs or Coax
Ranges to span	Segment-by-segment, hop-by-hop, VCI or VPI	End-to-end, entry-to-exit
Capacity	Switch will not accept more connections than recommended	More CBR and VBR connections can be accepted
Delay	delay is mainly due to CODECs	No extra delay to be experienced
Buffering	No extra buffering is needed since FEC is applied on PHY	Less buffer spaces are needed since FEC will regenerate lost cells (refer to Ch. 2)
Design Complexity	High design complexity	Not complex, in fact, less buffering is needed
Cost	High cost due to FEC CODEC support at each port	No cost addition to regular switch design

Chapter 8

Summary and Future Work

In this thesis we investigated the applicability of forward error correction (FEC) coding in ATM networks. FEC is not a new subject to ATM. Many researchers have studied different FEC schemes over ATM communications. Chapter 2 has more details on the different schemes that have been proposed in the literature. The rationale behind this is that FEC has been found to be more viable to delay sensitive applications or real-time traffic such as real-time audio and video applications. Although ATM uses low BER media for transmission, data can be lost due to network congestion. Since the level of network congestion varies, it is more appropriate to have sort of adaptive coding scheme that will alter its coding parameters based on the network congestion sensed.

8.1 Thesis Contributions

In this work we presented a new adaptive FEC (AFEC) scheme. The new scheme devised is based on RS codes. In general, FEC coding can be used into two different places in ATM networks: above ATM protocol stack as part of the transport layer or application, or in the physical layer. Both cases have been investigated in this thesis.

In the case where FEC is deployed above ATM stack, performance modeling is conducted to study its throughput compared to ARQ schemes and classical FEC. AFEC delivers higher throughput even at high CLR. AFEC is also observed to accomplish lower end-to-end delay than that of classical FEC.

Since AFEC is an adaptive scheme, it is essential to study the overhead associated with the parameter change via estimating the control message that need to be generated requesting such parameter update. At the worst case, AFEC will gener-

ate a parameter update control message every time the destination records a new value for the data loss. However this can be improved with the use of digital filtering techniques and prediction algorithms.

Another contribution is the security feature of AFEC. The security operation is conducted via the use of keys which are embeded into the parameters of AFEC. Security keys are not exchanges between communicating entities but they are referred to by some indices. Further, it has been shown that AFEC is a computationally secure cryptosystem.

Another interesting field, where AFEC is applicable, is multicasting environment. A simulation study is conducted using OPNET network simulation tool to study AFEC performance against ARQ/SRP coding schemes. Two performance measures where observed: cell delay and switch resources, mainly buffer requirements. The results of the study are briefly stated below.

1. Low buffer requirements in ATM networks (i.e. switch buffers) and at the UNI interfaces.
2. AFEC has a deterministic delay associated with its operation which makes it very reliable when real-time traffic is carried by ATM networks especially for broadcast and multicast environments. It has been observed that AFEC achieve CTD an order of magnitude less than the delay when SRP protocols where used.
3. In the case of SRP schemes, CTD depends mainly on window-size and switch thresholds. It might be beneficial to have some sort of adaptive window scheme to adapt well in multicasting environments.

On the other hand, AFEC is also studied in wireless environment. This falls in the class of application where FEC can be used in the physical. A framework of integrating AFEC with a wireless ATM network is given and analyzed with respect to CLR. The study in Chapter 5 which is done in Rayleigh type of wireless environment, shows that AFEC is more viable and withstands well against channel impairments

Using FEC has made ATM switch design potentially simpler. This issue has been explored and two architectures have been devised. Based on the type of service to be provided and the media that the switch should be operating into decided which approach of switch design is needed.

In summary, the key features of this scheme are:

1. Higher correction rates due to the puncturing.
2. Higher throughput than ARQ schemes over noisy channels.
3. Higher throughput is achieved when AFEC is used over multicasting networks.
4. Upper bounded delay components is associated to this scheme.
5. AFEC has potential impact on switch design. It has been shown that switch resources can be reduced without deterioration in quality of service offered.
6. Secure communication is an added feature.

8.2 Future Work

Although we attempted to study AFEC in different environments and situations, there is always room for exploring new modifications and areas to improve the code and its adaptive components.

Future direction with this research can be seen in the following points:

1. Studying the code on wireless media with different channel characterization other than Rayleigh. One possibility is exploring the benefits of deploying AFEC in CDMA.
2. Studying the viability of AFEC for Internet applications. The motive in here is that significant interest has been shown by researchers in Internet and in improving its services.
3. Investigating the possibilities of increasing the security measures of this code. One way could be to incorporate public key as in RSA cryptosystems.
4. Investigating possible VLSI realizations for AFEC. This is of great interest since the cost of developing RS CODECs grows with their coding length.
5. Modifying ARQ schemes to incorporate adaptive window sizes. This is to counteract the delay and buffer issues in multicasting and broadcasting.
6. Studying the impact of multicasting and coding parameters specially when different sources request parameter changes that scans broader range than what has been studied in this thesis. Possibly optimization techniques may be of interest to answer these issues.

7. Channel monitoring is a crucial issue in adaptive coding schemes. Further investigation of using digital filter techniques and prediction algorithms need to be addressed.

Bibliography

[1] H. Ohta and T. Kitami. "A Technique to Detect and Compensate Consecutive Cell Loss in ATM Networks". In *Proc. INFOCOM*, pages 781–790, 1991.

[2] E. Boutillon and J. Urunuela. " A VLSI Decoder For a New Type of Constellations Adapted to the Rayleigh Fading Channel". *Journal of Wireless Networks*, (1):17–26, 1991.

[3] Dipankar Raychaudhuri. " Wireless ATM Network: Architecture, System, and Prototyping". *IEEE Personal Communications*, 3(4):42–49, August 1996.

[4] O.Kubar and H. Mouftah. " Multiple Access Control Protocols for Wireless ATM: Problems Definition and Design Objectives". *IEEE Communications*, 35(11):93–99, November 1997.

[5] Bellcore. "Star Wars MPEG Traces". Available at ftp://ftp.bellcore.com/pub/vbr.video.trace.

[6] D. McDysan and D. Spohn. *ATM: Theory and application.* McGraw-Hill, NY, 1995.

[7] T. Nishida and K. Taniguchi. "QOS Controls and Services Models in the Internet". *IEICE Trans. Commun.*, E78-B(4):447–457, April 1995.

[8] J. Walrand. *Communication Networks: A First Course.* Aksen Associates, Boston, Massachusets, 1991.

[9] W. Stallings. *Data and Computer Communications.* Macmillan Publishing, NY, fourth edition, 1994.

[10] M. de Prycker. *Asynchronous Transfer Mode: Solution for Broadband ISDN.* Ellis Horwood, NY, second edition, 1993.

[11] J. Le Boudec. "The Asynchronous Transfer Mode: a tutorial". *Computer Networks and ISDN Systems*, 24:279–309, 1992.

[12] T. Chen and S. Liu. *ATM Switching Systems.* Artech House, Boston, MA, 1995.

[13] M. Garrett. "A Service Architecture for ATM: From Application to Scheduling". *IEEE Network*, pages 6–14, May 1996.

[14] H. Saito. *Teletraffic Technologies in ATM Networks.* Artech House, Boston, MA, 1994.

[15] M. Wernik, O. Abdul-Magd, and H. Gilbert. "Traffic Management for B-ISDN Services". *IEEE Network*, pages 10–19, September 1992.

[16] E. Biersack. "Performance Evaluation of Forward Error Correction in ATM Networks". In *Proc. Sigcomm*, pages 248–257, Philadelphia, Pennsylvania, Sept. 1992.

[17] E. Biersack. "Performance Evaluation of Forward Error Correction in an ATM Environment". *IEEE JSAC*, 11(4):631–640, May 1993.

[18] E. Biersack. "A Simulation Study of Forward Error Correction in ATM Networks". *ACM Sigcomm*, 22(1):36–47, January 1992.

[19] Y. Zhang, W. Wu, K. Kim, R. Pickholtz, and J. Ramasastry. "Variable Bit Rate Video Transmission in the Broadband ISDN Environment". In *Proc.* 15^{th} *Conf. on LANs*, pages 213–221, 1990.

[20] A. Almulhem and F. El-Guibally. "ATM Switch Design Requirements Under FEC Environment". In *Proc. IEEE Can. Conf. on Elec. And Comp. Eng.*, pages 810–813, Calgary, Canada, May 1996.

[21] A. McAuley. "Reliable Broadband Communication Using a Burst Erasure Correcting Code". In *Proc. Sigcomm*, pages 297–305, 1990.

[22] S. Lin and D. Castello. *Error Control Coding: Fundamentals and Applications.* Prentice-Hall, Englewood Cliffs, NJ, 1993.

[23] S. Wicker and V.K. Bhargava. *Reed-Solomon codes and their applications.* IEEE Press, Piscataways, NJ, 1994.

[24] M.H. Khan, T. Le-Ngoc, and V. K. Bhargava. "Adaptive Forward Error Control for Digital Satellite Systems". *IEEE Trans. on Aerospace and Electronic Systems*, 21(4):547–558, July 1985.

[25] S. Dravida and R Damodaram. "Error Detection and Correction Options for Data Services in B-ISDN". *IEEE JSAC*, 9(9):1484–1495, December 1991.

[26] T. Kitami and I. Tokizawa. "Cell Loss Compensation Schemes in Asynchronous Broadband ISDN". In *Proc. INFOCOM*, pages 116–123, 1990.

[27] H. Ohta and T. Kitami. "A Cell Recovery Method Using FEC in ATM Networks". *IEEE JSAC*, 9(9):1471–1483, December 1991.

[28] E. Ayanoglu *et al.* "Forward Error Control for MPEG-2 Video Transport in a Wireless ATM LAN". In *Proc. Internation Conf. on Image Processing*, pages 833–36, Lausanne, Switzerland, September 1996.

[29] E. Ayanoglu, K. Eng, and M. Karol. " Wireless ATM: Limits, Challenges, and Proposals". *IEEE Personal Communications*, 3(4):18–34, August 1996.

[30] N. Oguz and E. Ayanoglu. "Performance Analysis of Two-Level Forward Error

Correction for Lost Cell Recovery in ATM Networks". In *Proc. IEEE INFOCOM '95*, pages 723–37, Boston, Massachusetts, April 1995.

[31] N. Shacham and P. McKenney. "Packet Recovery in High-Speed Networks Using Coding and Buffer Management". In *Proc. INFOCOM*, pages 124–131, 1990.

[32] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes.* North-Holland Publishing, NY, 1977.

[33] G.C. Clark Jr. and J.B. Cain. *Error-Correction Coding for Digital Communications.* Plenum Press, NY, 1981.

[34] ATM Forum. "ATM User-Network Interface (UNI) Signalling Specification". Technical report, Ver. 4.0, July 1996.

[35] L. Dron, G. Ramamurthy, and B. Sengupta. "Delay Analysis of Continuous Bit Rate Traffic Over an ATM Network". *IEEE JSAC*, 9(3):402–407, April 1991.

[36] A. Michelson and A. Levesque. *Error-Control Tecgniques For Digital Communication.* John Wiley & Sons, New York, 1985.

[37] K. Iwamura, Y. Dohi, and H Imai. "A Design of Reed-Solomon Decoder with Systolic-Array Structure". *IEEE Trans. on Computers*, 44(1):118–122, January 1995.

[38] S. Wei and C. Wei. "High-Speed Decoder of Reed-Solomon Codes". *IEEE Trans. on Communications*, 41(11):1588–1593, November 1993.

[39] H. Shao, T. Troung, L Deutch, J. Yuen, and I. Reed. "A VLSI Design of a Reed-Solomon Decoder". *IEEE Trans. on Computers*, 37(10):1273–1280, October 1988.

[40] I. Hsu, I. Reed, T. Truong, K. Wang, C. Yeh, and L. Deutsch. "The VLSI Implementation of A Reed-Solomon Encoder Using Berlekamp's Bit-Serial Multiplier Algorithm". *IEEE Trans. on Computers*, 33(10):906–911, October 1984.

[41] D. Davies and W. Price. *Security for Computer Netwroks.* John Wiley & Sons, New York, second edition, 1989.

[42] Douglas R. Stinson. *Cryptography: Theory and Practice.* CRC Press, Boca Raton, Florida, 1995.

[43] E. Ifeachor and B. Jervis. *Digital Signal Processing: A Practical Approach.* Addison-Wesley, Reading, Massachusetts, 1993.

[44] Theodore Rappaport. *Wireless Communications: Principles and Practice.* Prentice Hall, Upper Saddle River, NJ, 1996.

[45] Anthony Acampora. " Wireless ATM: A Perspective on Issues and Prospects". *IEEE Personal Communications*, 3(4):8–17, August 1996.

[46] P. Agrawal *et al.* " SWAN: A Mobile Multimedia Wireless Network". *IEEE Personal Communications*, 3(2):18–33, April 1996.

[47] G. Sfikas, C. Apostolas, and R. Tafazolli. " The U.K. Link-PCP Approach to the Wireless ATM System". *IEEE Communications*, 35(11):60–70, November 1997.

[48] R. Valenzuela. " Performance of Quadrature Amplitude Modulation for indoor Radio Communications". *IEEE Transactions on Communications*, 35(11):1236–38, November 1987.

[49] Frank Amorosa. " Use of DS/SS Signalling to Mitigate Rayleigh Fading in A Dense Scatterer Environment". *IEEE Personal Communications*, 3(2):52–61, April 1996.

[50] D. Goodman *et al.* " Packet Reservation Multiple Access for Local Wireless Communications". *IEEE Transactions on Communications*, 37(8):885–90, August 1989.

[51] N. Wilson *et al.* " Packet CDMA Versus Dynamic TDMA for Multiple Access in an Integrated Voice/Data PCN". *IEEE JSAC*, 10(6):870–83, August 1993.

[52] N. Passas, S. Paskalis, D. Vali, and L. Merakos. "Quality-of-Service-Oriented Medium Access Control for Wireless ATM Networks". *IEEE Communications*, 35(11):42–50, November 1997.

[53] ATM Forum. "PNNI Specification". Technical report, Ver. 1.0, March 1996.

[54] S. Aikawa, Y. Motoyama, and M. Umehira. " Error Correction and Error Detection Techniques for Wireless ATM". *Journal of Wireless Networks*, (3):285–290, 1997.

[55] William C.Y. Lee. *Mobile Communications Engineering.* McGraw-Hill, New York, second edition, 1998.

[56] William C.Y. Lee. *Mobile Communications Design Fundamentals.* John Wiley & Sons, New York, second edition, 1993.

[57] H. chao and B. Cheo. "Design and Analysis of a LArge-Scale Multicast Output Buffered ATM Switch". *IEEE/ACM Trans. Networking*, 3(2):126–1383, April 1995.

[58] P. Huang and Y. Tanaka. "Multicast Routing Based on Predicted Traffic Statistics". *IEICE Trans on Commun.*, E77-B(10):1188–1193, October 1994.

[59] X. Liu and H. Mouftah. "High Performance Copy Network Design for Multicast ATM Switching". *Computer Communications*, 20:89–96, 1997.

[60] M. de Prycker. *Asynchronous Transfer Mode: Solution for Broadband ISDN.* Prentice-Hall International, UK, third edition, 1995.

[61] E. Gauthier, J. Le Boudec, and Ph. Oechslin. " SMART: A Many-to-Many Multicast Protocol for ATM". Technical report, Swiss Federal Institute of Technology, Lausanne, Switzerland, August 1996.

[62] J. Lin and S. Paul. "RMTP: A Reliable Multicast Transport Protocol". In *Proc. IEEE INFOCOM*, pages 1414–1424, San Francisco, CA, 1996.

[63] M. Lucas, B. Dempsey, and A. Weaver. " MESH: Distributed Error Recovery for Multimedia Streams in Wide-Area Multicasting Networks". In *Proc. IEEE ICC97*, Montreal, Canada, 1997.

[64] M. Ammar and L. Wu. " Improving he Throughput of Point-to-Multipoint ARQ Protocols Through Destination Set Splitting". In *Proc. INFOCOM*, pages 262–271, Florence, Italy, 1992.

[65] A. Almulhem, T.A. Gulliver, and F. El-Guibally. "Adaptive Error Correction for ATM Communications using Reed-Solomon Codes". In *Proc. IEEE Southeastcon96*, pages 230–233, Tampa, Florida, 1996.

[66] Reinhard Posch (Editor). *Communications and Multimdia Security.* Chapman and Hall, London, UK, 1995.

[67] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, and J. Robbins. "Performance Models of Statistical Multiplexing in Packet Video Communications". *IEEE Trans. on Comm.*, 36(7):834–843, July 1988.

[68] D. Lucantoni, M. Netus, and A. Reibman. "Methods for Performance Evaluation of VBR Video". *IEEE/ACM trans. Network*, 2(2):176–180, April 1994.

[69] J. Pitts and J. Schormans. *Introduction to ATM: Design and Performance.* John Wiley and Sons, West Sussex, England, 1996.

[70] T. Stock and X. Garcia. " On the Potential of Forward Error Correction Mechanisms applied to Real-Time Services caried over B-ISDN". In *International Zurich Seminar*, Zurich, Switzerland, February 1996.

[71] Mil3 Inc. *OPNET User Manual.* Washington, DC, 1996.

[72] D. Mitra and J. Seery. "Dynamic Adaptive Windows For High Speed Data Networks with Multiple Paths and Propagation Delays". In *Proc. IEEE INFOCOM '91*, pages 39–48, Boston, Massachusetts, April 1991.

[73] F. Bonomi, D. Mitra, and J. Seery. "Adaptive Algorithms for Feedback-Based Flow Control in High Speed, Wide Area ATM Networks". *IEEE JSAC*, 13(7):1267–1283, September 1995.

[74] A. Almulhem and F. El-Guibally. "Design Requirements of an ATM Switch with FEC Capabilities". In *Proc. IEEE ISITA96*, pages 230–233, Victoria, Canada, 1996.