INTEROPERABILITY AMONG RATE-BASED CONGESTION CONTROL ALGORITHMS FOR ABR SERVICE IN ATM NETWORKS

by

MAHMOUD KAYALI

B.Sc. Computer Engineering, KFUPM, Dhahran, Saudi Arabia, 1995
B.Sc. Computer Science, KFUPM, Dhahran, Saudi Arabia, 1995

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF APPLIED SCIENCE
in
THE FACULTY OF GRADUATE STUDIES
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

THE UNIVERSITY OF BRITISH COLUMBIA

August 1997

© Mahmoud Kayali, 1997
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Electrical and Computer Engineering

The University of British Columbia
Vancouver, Canada

Date August 20, 1997

DE-6 (2/88)
Abstract

The ATM Forum has recently adopted rate-based schemes as the standard congestion control mechanisms for Available Bit Rate (ABR) service. Two types of rate-based schemes are supported by the ATM Forum specification, namely binary feedback and explicit rate. Explicit rate feedback schemes by themselves can be divided into exact and approximate. Since the implementation of one of these schemes in ATM switches is left to switch manufacturers, it is expected to have these types working on the same network. The compatibility of switches each running a different scheme was considered in the ATM Forum specification. However, since each type of switches has its own merits and demerits, the interoperability of these switch types and its impact on the performance of the network deserves extensive studies. This thesis investigates two types of interoperability in multi-vendor (heterogeneous) ATM networks: the interoperability between binary feedback and explicit rate schemes and the interoperability among different explicit rate schemes. For the first type of interoperability, we look at the steady-state performance of ABR service in networks consisting of both binary feedback and explicit rate switches. In order to find the best locations for each switch type, several cases of switch type locations are considered. Moreover, the impact of ABR sources parameter setting on the performance of heterogeneous networks is studied. For the second type of interoperability (i.e. the interoperability among explicit rate schemes), we investigate potential unfairness problems resulting from situations whereby different ABR sources receive network feedback from different subsets of switches along a given network path. Three types of unfairness problems that arise in such networks are identified. One type of unfairness appears while the sources are increasing their rates and the second type appears while they are decreasing their rates. The third type is a new cause of unfairness generated by the presence of highly bursty Variable Bit Rate (VBR) traffic which can cause
unfairness not only in the steady-state periods but also in the transient-state periods on a network link. In addition to identifying the causes of unfairness, our results provide quantitative evaluation of the level of unfairness as a function of various network and source parameters.
Table of Contents

Abstract ii
List of Figures vii
List of Tables ix
Acknowledgment xi

Chapter 1 Introduction 1

1.1 Background 1

1.1.1 ATM Technology 1
1.1.2 ATM Service Categories 2
1.1.3 ATM Reference Model 3

1.2 Congestion Control for ABR Service Category 5

1.3 Thesis Objectives 6

1.4 Thesis Overview 7

Chapter 2 Congestion Control for the ABR Service Category in ATM Networks 9

2.1 Quality of Service Parameters 9

2.2 Congestion Control in ATM Networks 10

2.2.1 Preventive Control 10
2.2.2 Reactive Control 11

2.3 Congestion Control for the ABR Service Category 13

2.3.1 Credit-Based Congestion Control 13
2.3.2 Rate-Based Congestion Control 15

2.4 ATM Forum Specification 19

2.4.1 ABR Traffic Parameters 20
Chapter 3  Interoperability Issues 31

3.1 Types of Interoperability.......................................................31
  3.1.1 Interoperability Between Binary Feedback and ER Switches ....32
  3.1.2 Interoperability Between ER switches..................................32

3.2 Issues Affecting the Study of Interoperability...........................33
  3.2.1 Switch Location.............................................................33
  3.2.2 ABR Source Parameters Setting......................................35
  3.2.3 Switch Algorithm Parameters.........................................37
  3.2.4 Mixed Traffic Categories..............................................37

3.3 Previous Work......................................................................38

Chapter 4  Interoperability between Binary Feedback and Explicit Rate Algorithms 41

4.1 Simulation Model ..................................................................41
  4.1.1 Network Topology............................................................42
  4.1.2 Switch Mechanisms.........................................................44
  4.1.3 Source Parameters..........................................................45

4.2 Simulation Results.................................................................46
  4.2.1 Impact of RIF and RDF....................................................48
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>ATM Reference Model</td>
<td>4</td>
</tr>
<tr>
<td>2.1</td>
<td>Credit-based congestion control</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>Rate-based congestion control</td>
<td>15</td>
</tr>
<tr>
<td>2.3</td>
<td>Segmented rate-based end-to-end feedback loop using a virtual source/destination switch</td>
<td>16</td>
</tr>
<tr>
<td>2.4</td>
<td>The ATM Forum Specification</td>
<td>22</td>
</tr>
<tr>
<td>2.5</td>
<td>Summary of congestion control techniques</td>
<td>30</td>
</tr>
<tr>
<td>3.1</td>
<td>An example of a typical ATM network configuration</td>
<td>34</td>
</tr>
<tr>
<td>4.1</td>
<td>A simplified ABR traffic source model</td>
<td>43</td>
</tr>
<tr>
<td>4.2</td>
<td>Simulated network topology</td>
<td>44</td>
</tr>
<tr>
<td>4.3</td>
<td>ACR dynamics for All-binary feedback switches case (RIF = 1/512 RDF = 1/512)</td>
<td>50</td>
</tr>
<tr>
<td>5.1</td>
<td>Simple two ABR sources network</td>
<td>59</td>
</tr>
<tr>
<td>5.2</td>
<td>ABR source’s ACR dynamics</td>
<td>60</td>
</tr>
<tr>
<td>5.3</td>
<td>Heterogeneous network topology</td>
<td>64</td>
</tr>
<tr>
<td>5.4</td>
<td>The rate increase problem</td>
<td>65</td>
</tr>
<tr>
<td>5.5</td>
<td>The rate decrease problem</td>
<td>68</td>
</tr>
<tr>
<td>5.6</td>
<td>Heterogeneous network topology with “deterministic” VBR source</td>
<td>73</td>
</tr>
<tr>
<td>5.7</td>
<td>ACR dynamics of the simulated ER algorithms</td>
<td>76</td>
</tr>
<tr>
<td>5.8</td>
<td>ACR dynamics in a heterogeneous network (switch1 algorithm = ERICA, switch2 algorithm = EPRCA)</td>
<td>79</td>
</tr>
<tr>
<td>5.9</td>
<td>ACR dynamics in a heterogeneous network (switch1 algorithm = EPRCA-Bounded, switch2 algorithm = EPRCA)</td>
<td>80</td>
</tr>
</tbody>
</table>
Figure 5.10  ACR dynamics in a heterogeneous network (switch1 algorithm = EPRCA, switch2 algorithm = EPRCA-Binary) ................................................................. 81

Figure 5.11  ACR dynamics in a heterogeneous network with VBR source present (switch1 algorithm = EPRCA-Bounded, switch2 algorithm = EPRCA) ........................................ 84

Figure 5.12  ACR dynamics in a heterogeneous network with VBR source present (switch1 algorithm = ERICA, switch2 algorithm = EPRCA-Binary) ........................................ 85
List of Tables

Table 1.1 Characteristics of ATM service categories ............................................. 3
Table 2.1 Traffic contract attributes .................................................................... 10
Table 2.2 ABR Source Parameters ........................................................................ 20
Table 4.1 EPRCA parameter settings ................................................................... 45
Table 4.2 Source parameters .................................................................................. 46
Table 4.3 Switch location criticality ...................................................................... 47
Table 4.4 RIF and RDF setting for multi-link VCs ................................................ 48
Table 4.5 Percentage throughput and fairness for the first simulation set (with ERICA) (RIF=1/512 RDF=1/512) ................................................................. 49
Table 4.6 Percentage throughput and fairness for the second simulation set (with ERICA) (RIF=1/128 RDF=1/512) .............................................................. 51
Table 4.7 Percentage throughput and fairness for the third simulation set (with ERICA) (RIF=1/128 RDF=1/32) ............................................................... 52
Table 4.8 Percentage throughput and fairness for the fourth simulation set (with ERICA) (RIF=1/512 RDF=1/32) ............................................................... 53
Table 4.9 Percentage throughput and fairness for the first simulation set (with EPRCA) (RIF=1/512 RDF=1/512) ................................................................. 54
Table 4.10 Percentage throughput and fairness for the second simulation set (with EPRCA) (RIF=1/128 RDF=1/512) .............................................................. 55
Table 4.11 Percentage throughput and fairness for the third simulation set (with EPRCA) (RIF=1/128 RDF=1/32) ............................................................... 55
Table 4.12 Percentage throughput and fairness for the fourth simulation set (with EPRCA) (RIF=1/512 RDF=1/32) ............................................................... 56
Table 4.13 Percentage throughput and fairness when PCR=100Mbps (with EPRCA) (RIF=1/512 RDF=1/32) ................................................................. 57
Table 5.1 EPRCA tuning parameters setting ......................................................... 74
| Table 5.2 | Source parameters setting | 75 |
| Table 5.3 | Simulation results for heterogeneous network topology | 77 |
| Table 5.4 | Simulation results for heterogeneous network topology with VBR source | 82 |
Acknowledgment

I would like to express my gratitude to my research supervisors, Dr. H. M. Alnuweiri and Dr. V. C. M. Leung, for their guidance and support throughout my study. This work was supported by a research grant from the Canadian Institute of Telecommunications Research (CITR), by a University Graduate Fellowship (UGF) from the University of British Columbia, and by Mil3 Inc. through OPNET licences provided by the University Consortium program. The most important acknowledgment goes to my great parents. Without their support and guidance, this work would not have been possible. I would like to thank my wife for her patience and continuous encouragement. Special thanks go to my brothers, my sister, and my true friends for their sincere support to finish this work. Lastly, I thank God for blessing me with knowledge and energy to accomplish this goal, and I thank Him for making all this possible.
Chapter 1  Introduction

1.1  Background

1.1.1  ATM Technology

Since the introduction of the Internet in the last decade as a public worldwide information super-highway, the number of Internet users is growing fast. Not only the population of the Internet user community is increasing, but with the advances in computer technology, even the number and the nature of applications running over the Internet are rapidly changing. In addition to the traditional applications like file transfer and electronic mail, multimedia applications such as real-time video conferencing, multimedia email, and remote interactive simulations are finding their way quickly into the Internet. Considering the high diversity of applications expected to run over the Internet and their wide range of requirements in terms of delay and bandwidth, the necessity for a single integrated network technology capable of supporting all these types of applications efficiently has emerged.

Asynchronous Transfer Mode (ATM) has been proposed by the International Telecommunication Union-Telecommunication Sector (ITU-T) as the solution for the future broadband telecommunication network [17]. The most valuable advantage of ATM is its ability to carry voice, video, and data over the same infrastructure while providing acceptable Quality of Service (QoS) in an efficient and simple way [29][51]. ATM is based on a fixed-size virtual circuit-oriented packet-switching technology. In ATM, traffic, regardless of its type, is carried in 53-byte cells (packets) consisting of a 5-byte header field and a 48-byte payload field.
1.1.2 ATM Service Categories

The ATM Forum has classified traffic carried by ATM network into five main service categories [2]. Each category has its own QoS requirements in terms of delay, bandwidth guarantees, and cell loss tolerance [15][50]. These traffic categories are:

**Constant Bit Rate (CBR):** the CBR category is intended for real-time ATM applications that require very strict delay and cell delay variation requirements. These applications need a static amount of bandwidth for their connection lifetime. Telephone connections and video conferencing are two typical CBR applications.

**Real-Time Variable Bit Rate (rt-VBR):** this service category is intended for real-time applications requiring tightly constrained delay and delay variation. As opposed to CBR applications which require a static bandwidth, rt-VBR bandwidth requirements vary strongly with time. An example of rt-VBR is interactive compressed video.

**Non-Real-Time Variable Bit Rate (nrt-VBR):** nrt-VBR category is meant for non-real-time applications that have varying bandwidth requirements, but are less sensitive to delay variations than rt-VBR applications. Multimedia email is an example of an nrt-VBR application.

**Unspecified Bit Rate (UBR):** The UBR category is intended for non-real-time applications not requiring tightly constrained delay and delay variation. Traditional data applications such as file transfer and email are typical UBR applications. No bandwidth guarantees are promised by the network for UBR applications. Moreover, UBR applications, as opposed to CBR and VBR applications, are not sensitive to cell loss.

**Available Bit Rate (ABR):** like UBR, this category, is designed for data applications such as file
transfer and email, but with the difference that ABR applications respond to congestion feedback messages from the network. It is desirable for the ATM network components (e.g. switches) to try to reduce the ABR traffic delay and cell loss to the minimum although the specification does not require such guarantees.

Since this thesis is devoted to the ABR service category, more details are provided about ABR traffic in next chapter. For more details about the other service categories, the reader is referred to [2][15][17]. Table 1.1 summarizes the general characteristics of ATM service categories [50].

<table>
<thead>
<tr>
<th>Service characteristic</th>
<th>CBR</th>
<th>rt-VBR</th>
<th>nrt-VBR</th>
<th>ABR</th>
<th>UBR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth guarantee</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Optional</td>
<td>No</td>
</tr>
<tr>
<td>Suitable for real-time traffic</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Suitable for bursty traffic</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Feedback about congestion</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

1.1.3 ATM Reference Model

Figure 1.1 shows the ATM reference model [20][41]. It consists of three planes: a user plane, a control plane, and a management plane. The user plane is responsible for transporting user information whereas the control plane is in charge of call and connection signalling. The management plane performs layer management functions, related to each layer separately, and plane management functions, to provide coordination between all planes.

The user plane and the control plane are further divided into three main layers: the physical layer, the ATM layer and the ATM Adaptation Layer (AAL).
Chapter 1 Introduction

Figure 1.1 ATM Reference Model

The physical layer deals directly with the physical medium. It converts the ATM cell stream into bits to be transmitted over the physical medium dealing with voltages, bit timing, and various other issues related to the physical medium.

The ATM layer deals with cells and their transportation. It defines cells layout and the meaning of each field in the cell header. Also, it controls establishment, release, and multiplexing of virtual circuits. Congestion control is also a job of the ATM layer.

The ATM Adaptation Layer (AAL) is divided into two sublayers, the Segmentation And Reassembly (SAR) at the bottom, and the Convergence Sublayer (CS) at the top. The SAR sublayer is responsible for breaking the upper layer packets into ATM cells at the transmitter end and then putting them back together at the receiver end. The CS provides the interface between the ATM network and the upper layers by performing different application-dependent functions.
1.2 Congestion Control for ABR Service Category

Congestion occurs in ATM network when multiple input ports of a switch carry data that is destined for a single output port. If the sum of the bandwidths of the input ports is higher than the bandwidth of the output port, congestion occurs since the output link of the switch is receiving more data than it can process.

As mentioned earlier, ABR traffic is the only ATM traffic type that supports congestion control. After considering several proposals, the ATM Forum has chosen the rate-based feedback scheme to be the standard congestion control mechanism for ABR service category. In rate-based control, ABR sources are requested to adjust their rates according to the feedback messages coming from the network (mainly ATM switches). These feedback messages indicate explicitly or implicitly the rate by which each ABR source can send data as computed by the ATM switches [6].

Rate-based schemes, implemented in ATM switches, fall into two main categories: binary feedback and explicit rate [2]. Binary feedback schemes ask ABR sources to decrease (or increase) their rates based on the congestion state of the network without providing the source explicitly with a specific rate value. The feedback is carried using a special type of control cells, known as Resource Management (RM) cells. On the other hand, explicit rate switches (i.e. running explicit rate algorithms) compute the fair rate for each ABR connection and then provide the source with the computed value through the RM cells.

 Explicit rate feedback schemes, by themselves, can be divided into two main classes, \textit{approximate} and \textit{exact}. From their names, approximate algorithms tend to simplify the rate computation by approximating the explicit rate value and the exact algorithms attempt to compute
the explicit rate value exactly by performing more accurate rate calculation. In general, exact algorithms have better performance figures in terms of fairness and responsiveness. Approximate algorithms, however, are much simpler to implement.

The ATM Forum's specification gave the ATM switch manufacturers a great level of flexibility by leaving the implementation details of the rate-based scheme running in a switch to be specified by the manufacturers themselves. Moreover, the specification allowed ATM switches running binary feedback or explicit rate algorithms to coexist in the same network.[6]

Many ATM switch vendors have already deployed switches with binary feedback or explicit rate capabilities and it is likely that different vendors will implement different forms of algorithms. This brings into question the issue of interoperability between different switches in multi-vendor (or heterogeneous) networks. In particular, the issue of fair bandwidth allocation becomes crucial in this type of networks.

1.3 Thesis Objectives

The main goal of this thesis is to evaluate the performance of multi-vendor ATM networks where several rate-based algorithms coexist in the same network. To achieve this goal, the thesis will:

1. Provide a detailed characterization of the interoperability issue by separately investigating two types of interoperability: the interoperability between binary feedback and explicit rate algorithms, and the interoperability among explicit rate algorithms.
2. Identify the best places in the network to locate binary feedback and explicit rate switches to enhance the performance of ABR service in multi-vendor ATM networks. The importance of this identification becomes very apparent at the network design stage where the best locations of each switch type need to be determined to get the best network performance.

3. Evaluate the impact of some important ABR source parameters on the performance of multi-vendor ATM networks especially to solve potential unfairness problems that were identified for some types of rate-based algorithms. This is considered a very practical issue to evaluate because the impact of some ABR source parameters is really high and can noticeably affect the performance of the whole network if not carefully set.

4. Investigate and analyze some potential unfairness problems that can appear in multi-vendor ATM networks. This issue is of great importance for telephone companies and service providers since they need always to ensure that their network equipment supplied by different vendors can provide their customers with a fair service.

5. Identify new directions that can be explored to better understand the behavior of ABR traffic since a full implementation of the ATM Forum's specification for this service category is still at an early stage.

This work complements and expands on previous results reported in [8][9][40][52]. Specific comparison with previous research is pointed out in the respective chapters of this thesis.

1.4 Thesis Overview

This thesis is organized as follows. Chapter 2 provides an introduction to congestion control for ABR service in ATM networks. It outlines the different feedback schemes suggested
to solve congestion problem for the ABR service category with more emphasis on the rate-based schemes. Chapter 3 introduces the issue of interoperability in multi-vendor ATM networks. It presents the goals of studying interoperability along with the factors that must be considered when addressing the issue of interoperability.

Chapter 4 is devoted to the interoperability between binary feedback and explicit rate algorithms. The model used in our simulations is described and the performance results are presented. Chapter 5 addresses the interoperability among explicit rate algorithms. It analyzes the dynamic behavior of ABR sources while responding to the feedback messages and describes three types of unfairness problems that can exist in multi-vendor ATM networks. Some simulation results are presented. Chapter 6 concludes the work presented in this thesis and provides some suggestions for future work.
Chapter 2  Congestion Control for the ABR Service Category in ATM Networks

One of the primary goals of ATM is to support voice, data, and video traffic integration with guaranteed Quality of Service (QoS) for each traffic type. Any factor that results in failure of supporting the expected QoS guarantees must be contained and removed as soon as possible. Congestion is considered one of the major threats that can weaken the ability of ATM networks to guarantee the required QoS requirements.

This chapter is dedicated to explain congestion control for the ABR service category in ATM networks. First, the QoS parameters specified for each ATM connection are briefly outlined. Second, several congestion control techniques for ATM networks are presented. Then, the main two classes of congestion control approaches for ABR traffic: the credit-based and the rate-based, are discussed thoroughly. The ATM Forum specification for the ABR service category is then briefly reviewed. Finally, some examples of the proposed rate-based congestion control algorithms are given.

2.1 Quality of Service Parameters

For each application running over ATM, there is a specific set of QoS parameters that need to be guaranteed by the ATM network for that application. These parameters along with some other traffic parameters are negotiated at connection set-up time and must be agreed upon by both the application and the network. This agreement is sometimes referred to by the “traffic contract” [50].

The ATM Forum has specified a list of attributes that contains the traffic parameters and
QoS parameters that are to be maintained in the traffic contract [2][50]. Table 2.1 below outlines these attributes with a brief description of each. Since not all the attributes are specified for each service category, Table 2.1 shows for each attribute the applicable service categories.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Acron.</th>
<th>Meaning</th>
<th>Service Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic Parameters:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peak Cell Rate</td>
<td>PCR</td>
<td>The maximum rate at which cells will be sent</td>
<td>All categories</td>
</tr>
<tr>
<td>Sustained Cell Rate</td>
<td>SCR</td>
<td>The long-term average cell rate</td>
<td>VBR</td>
</tr>
<tr>
<td>Minimum Cell Rate</td>
<td>MCR</td>
<td>The minimum acceptable cell rate</td>
<td>ABR</td>
</tr>
<tr>
<td>Maximum Burst Size</td>
<td>MBS</td>
<td>The maximum number of cells that can be sent at the peak rate</td>
<td>VBR</td>
</tr>
<tr>
<td>Cell Delay Variation Tolerance</td>
<td>CDVT</td>
<td>The maximum acceptable cell jitter</td>
<td>All categories</td>
</tr>
<tr>
<td>QoS Parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell Delay Variation</td>
<td>CDV</td>
<td>The variance in cell delivery times</td>
<td>CBR and rt-VBR</td>
</tr>
<tr>
<td>Cell Transfer Delay</td>
<td>CTD</td>
<td>The maximum and the mean delivery times</td>
<td>CBR, and rt-VBR</td>
</tr>
<tr>
<td>Cell Loss Ratio</td>
<td>CLR</td>
<td>Fraction of cells lost or delivered too late</td>
<td>CBR, VBR and ABR (optionally)</td>
</tr>
<tr>
<td>Other Attributes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback support</td>
<td></td>
<td>The ability to respond to feedback messages from the network</td>
<td>ABR</td>
</tr>
</tbody>
</table>

### 2.2 Congestion Control in ATM Networks

Congestion happens when the input rate at some node in the network (e.g. a switch) is more than the available output link capacity [24]. ATM networks solve the congestion problem using two strategies: preventive control and reactive control. The use of one strategy or the other depends mainly on the traffic service category. Preventive control and reactive control are explained in the following sections.

#### 2.2.1 Preventive Control

Preventive congestion control (sometimes called ‘open-loop’ control) aims to avoid the congestion by ensuring that each connection remains within the limits that were used to allocate network resources during connection set-up time [12]. Preventive control consists of two main
functions: connection admission control and usage parameter control.

Connection Admission Control (CAC) is defined as the set of actions taken by the network during the connection set-up phase in order to determine whether the connection request can be accepted or not [48]. A connection request is accepted only when sufficient resources such as bandwidth and buffer space are available along the path of the connection. The decision is mainly based on the service category, desired Quality of Service (QoS) and the state of the network.

Usage Parameter Control (UPC) is defined as the set of actions taken by the network to ensure that user's traffic conforms to the QoS parameters specified at connection set-up time [48]. UPC protects network resources from misbehaving users who can affect the QoS of other connections. If one of the connections is discovered to be violating its requested QoS parameters, then the UPC function can possibly discard the non-conforming cells or alternatively can tag them as low priority cells by setting the Cell Loss Priority (CLP) Bit in the ATM cell header [38]. Cells with the CLP bit set are discarded before other cells if congestion is detected inside the network. Several implementations were suggested for the UPC function [42]. Although the ATM Forum specifies a general algorithm that can be used to test the conformance of the cells called the Generic Cell Rate Algorithm (GCRA), no specific implementation is suggested as a standard [2].

2.2.2 Reactive Control

As the name suggests, reactive control allows congestion to occur then it reacts to the congestion by informing the involved traffic sources to lower their rates in order to relieve the network from overload. The applicability of reactive control mechanisms in ATM networks was originally suspected mainly because of the slow reaction to congestion in real-time traffic such as VBR [12]. However, when ABR traffic was introduced, the necessity for reactive control became
very clear. ABR traffic utilizes the available bandwidth that is left after giving CBR and VBR traffics their bandwidth requirements. Since the amount of the available bandwidth changes widely over time, the need for a feedback mechanism (i.e. reactive control) becomes a must. Reactive control is sometimes referred to as “closed-loop” control since there is always a loop between the traffic sources and the network through feedback messages.

There are several design criteria to be considered when designing a feedback mechanism [1][24]. As the number of feedback mechanisms suggested for controlling ABR traffic is rather large, these criteria can be used as a basis for comparing the performance of different mechanisms. The following is a summary of these design criteria:

1. **Scalability**: the feedback scheme should not be limited to a specific range of speed, distance, number of switches, or number of connections.

2. **Fairness**: the feedback scheme should ensure a fair rate allocation for all connections and at the same time should not allow for bandwidth underutilization. Fair allocation is not as simple as dividing the link bandwidth over the number of connections. Several fairness policies were suggested to compute the fair allocation for a specific connection [2]. Among these different policies, the “**max-min fairness criterion**” is by far the most commonly used [1][23].

3. **Robustness**: the scheme should not be very dependent on the tuning parameters. It should tolerate slight parameters mistuning and loss of control messages. Moreover, it must be able to protect the network from misbehaving users.

4. **Convergence**: the feedback scheme must ensure that at the steady-state, connections’ rates will converge to their fair rate allocations from any initial conditions. Large rate oscillations at
steady-state should be avoided since it will degrade the performance of the whole network.

5. **Responsiveness:** the scheme must react as fast as possible to sudden traffic changes in the network with minimum cell loss and with minimum bandwidth underutilization.

6. **Interoperability:** as network components are manufactured by different vendors each with his own feedback scheme, effective operation of feedback schemes on a multivendor network environment is highly needed.

7. **Implementability:** The feedback scheme should not be too complex to implement. Both time and space complexities must be considered when designing any feedback mechanism.

### 2.3 Congestion Control for the ABR Service Category

In addition to preventive congestion control which is applied in all traffic categories, ABR is the only service category that uses feedback control (i.e. reactive control). Since this thesis is concerned with congestion control for ABR traffic, the rest of this chapter is dedicated to the issue of feedback congestion control in ABR service.

To solve the problem of congestion in ABR service, two main schemes were suggested: the credit-based and the rate-based [6][13][24][32]. After a long debate in the ATM Forum, the rate-based scheme was chosen as the standard mechanism. In the following sections, we look at these two scheme in details.

#### 2.3.1 Credit-Based Congestion Control

The credit-based scheme was the first mechanism to be introduced for controlling congestion in ABR traffic. As shown in Figure 2.1, the credit-based scheme is a window-based, link-by-
Chapter 2 Congestion Control for the ABR Service Category in ATM Networks

link, per-VC flow control. Along the path from the VC source to the VC destination, there is a feedback loop between each two consecutive nodes (ATM switches or end systems). In every feedback loop, the destination node maintains a separate buffer for each VC. The source node in the feedback loop cannot send any data cells on any VC before receiving a “credit” from the feedback loop destination. The credit represents the number of cells that can be transmitted on that VC without causing any buffer overflows at the destination [24][32].

Credit-based congestion control can be either static or adaptive depending on how the buffer space is allocated at the destination node of a feedback loop. In the static version, the destination reserves a fixed amount of buffer space for each VC at the connection set-up time. Regardless of the behavior of the source or the congestion status of the network, the same amount of buffer space is maintained during the lifetime of the VC. On the other hand, the adaptive control version changes the amount of reserved buffer space as the activity of the VC changes [24][32].

The adaptive version of the credit-based control utilizes the buffer space more efficiently than the static version since it is wasteful of resources to reserve a large amount of buffer for the whole VC’s lifetime given that the VC does not always have data to send. However, having always a large buffer space reserved for each VC in the static version, allows a suddenly-active
VC to transmit data immediately without the need to wait for a larger amount of buffer to be allocated for it as in the case of adaptive control [24].

2.3.2 Rate-Based Congestion Control

In contrast to the credit-based, rate-based congestion control is an end-to-end flow control scheme. This means that the feedback loop in rate-based control is from the source of the connection to the end destination as depicted in Figure 2.2. Within a feedback loop, the source responds to the feedback messages coming from the network (ATM switches and/or the destination) by changing its transmission rate accordingly. The feedback messages are carried by special control cells called Resource Management (RM) cells. [6]

![Figure 2.2 Rate-based congestion control](image)

In some networks where the feedback delay can be long, as in the case of Wide Area Networks (WANs), using end-to-end flow control becomes ineffective. In order to overcome this problem, the end-to-end feedback loop is segmented into smaller loops by having some intermediate switches to act as "virtual sources/destinations", see Figure 2.3. The virtual sources interact with the destination as if they are real sources and similarly the virtual destinations interact with the source as if they are real destinations. By segmenting the end-to-end feedback loop, misbehav-
ing users in one feedback loop are isolated from other users in other loops. Of course, having the switches to act as virtual sources/destinations adds up to the complexity of the intermediate switches. [6][24]

The ATM Forum specification supports two main types of algorithms for implementing the rate-based congestion control scheme: binary feedback and explicit rate (ER) feedback. In the following sections, the basic operation of each is presented. From now on, the term “binary feedback switch” refers to an ATM switch running a binary feedback algorithm and the term “explicit rate switch” or “ER switch” refers to a switch running an explicit rate algorithm.

2.3.2 Binary Feedback Algorithms

Binary feedback algorithms use the Explicit Forward Congestion Indication (EFCI) bit in the ATM cell header to indicate whether one or more of the network elements (intermediate switches or the destination) along the connection path from the source to the destination is congested. The term “binary” is used to indicate that these algorithms are just capable of inform-
The basic binary feedback algorithm works as follows. The source sends an RM cell every \(N_{rm}\) data cells sent to the destination. The data cells are sent with the EFCI bit cleared. Any switch along the path can set the EFCI bit in the data cells if it detects congestion. When an RM cell is received by the destination with the last received data cell having the EFCI bit set, then the destination informs the source, through a feedback message carried by the RM cell, that there is a congestion in the network. Upon receiving this RM cell, the source decreases its rate by an amount proportional to its current rate. If the EFCI bit in the last data cell received by the destination was not set, then the RM cell is sent to the source without indicating any congestion and hence the source can increase its rate safely. Several variations of the binary feedback algorithms were considered by the ATM Forum before reaching the final version of the specification (for examples check [3][21][35]).

One way to classify the binary feedback algorithms is by the “polarity” of the feedback message they send to the source [24]. If the binary feedback scheme sends just only feedback messages telling the source to increase its rate then the feedback is called a “positive” feedback. On the other hand, if the feedback is sent to tell the source to reduce its rate then it is called “negative” feedback. Moreover, some binary feedback algorithms use both “positive” and “negative” feedbacks, and they are classified as “bipolar” algorithms.

### 2.3.2.2 Explicit Rate Algorithms

The fact that switches in connection-oriented networks such as ATM have a lot of connection-related information is not used in the binary feedback algorithms [13]. ER algorithms were introduced to utilize the extra information available in the switches. In ER algorithms, switches
along the connection path play more active role than in the case of binary feedback.

The basic operation of the ER algorithms can be explained as follows. As in the case of binary feedback algorithms, the source sends an RM every \( N_{rm} \) data cells. For every connection passing through a switch, the switch computes a fairshare of the bandwidth. When the RM cell passes through the switch, the ER field value of the RM cell is changed to the minimum of the current ER value and the connection’s fairshare as computed by the switch. When the destination receives an RM cell it returns the RM cell back to the source. Again, the switches can modify the ER field in the RM cell while it is in its way back to the source. Then upon receiving the RM cell, the source will change its rate according to the ER value carried by that RM cell.

Since the bandwidth fairshare for a connection can be computed in different ways and with different accuracies, various ER switch algorithms were introduced [4][18][26][27][34][44]. Two types of ER algorithms can be identified: approximate and exact [1]. Approximate algorithms try, using the buffer occupancy levels and some information carried by the RM cells, to approximate the fairshare for all connections. To estimate the fairshare in an efficient way and without keeping Per-connection information, they mainly depend on using running averages. On the other hand, exact algorithms use per-connection state information to compute the exact fairshare for every connection.

In terms of performance, exact ER algorithms are generally more robust, more responsive, and more fair than approximate algorithms [1]. However, approximate algorithms are much simpler to implement in hardware as switch algorithms since no per-connection state information need to be kept. Examples for the two types of ER algorithms are given in section 2.5.
Each type of rate-based algorithms has its own merits [1][13]. Looking at the performance issue, ER algorithms outperforms binary feedback algorithms in several aspects. First, ER algorithms allow for immediate rate changes with minimum oscillations. Second, they show robustness against RM cells loss. If an RM cell is lost, the source rate will be corrected by the next coming RM cell. Third, it was shown that binary feedback algorithms has a serious fairness problem. VCs passing through more switches have higher chances of having the EFCI bit set than those traveling across less number of switches. This problem is better known as the “beat-down” problem [5]. If we assume that \( p \) is the probability of having the EFCI bit set in one binary feedback switch, then the probability of having the EFCI bit set for a VC traveling \( n \) switches is \( 1 - (1 - p)^n \) or \( np \) [24]. Therefore, VCs with longer paths have fewer chances to increase their rates and are beaten down more often than VCs with shorter paths. On the other hand, from a hardware complexity point of view binary feedback switches are certainly much simpler to implement than ER switches.

2.4 ATM Forum Specification

The ATM Forum has chosen rate-based control to be the standard congestion control scheme for ABR traffic. The choice of rate-based control over credit-based was mainly taken because of the high hardware complexity of the switches if the credit-based scheme was to be implemented. The complexity comes from the fact that credit-based control uses Per-VC queueing and it is widely known that Per-VC queueing is not feasible when a large number of VCs is expected as in the case for Wide Area Networks (WANs). [13][24]

The ATM Forum specification has defined a general framework for the rate-based congestion control scheme. This framework was designed to allow for a wide range of possible
implementations for rate-based algorithms. The following three points have highly affected the development of this specification [6]:

- In addition to supporting an end-to-end flow control, the specification must optionally allow for a segmented flow control loop by supporting the idea of "virtual" source/destination switches. (see Section 2.3.2)

- The specification should support both binary feedback and ER congestion control algorithms.

- Both binary feedback and ER algorithms should be allowed to coexist in the same feedback loop without incompatibility problems.

The following sections highlight the major parts of the ATM Forum specification that are necessary for the course of this thesis. For the specification’s full details, the reader is referred to the specification document [2].

2.4.1 ABR Traffic Parameters

The ATM Forum specification defines a long list of parameters that are used to describe each ABR connection. Since not all the parameters presented in the specification have the same importance for our later discussions, Table 2.2 shows a list of the important parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Acron.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak Cell Rate</td>
<td>PCR</td>
<td>The maximum rate that can be used by the ABR source</td>
</tr>
<tr>
<td>Minimum Cell Rate</td>
<td>MCR</td>
<td>The minimum cell rate guaranteed by the network for that ABR connection</td>
</tr>
<tr>
<td>Initial Cell Rate</td>
<td>ICR</td>
<td>The start-up rate of the source initially or after an idle period</td>
</tr>
<tr>
<td>Rate Increase Factor</td>
<td>RIF</td>
<td>Controls the maximum rate increase permitted by the ABR</td>
</tr>
</tbody>
</table>
Table 2.2 ABR Source Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Acron.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate Decrease Factor</td>
<td>RDF</td>
<td>Controls the decrease in the ABR source rate upon receiving an RM cell</td>
</tr>
<tr>
<td>Number of data cells/ RM cell</td>
<td>Nrm</td>
<td>Controls the frequency of the RM cells sent with the data cells</td>
</tr>
<tr>
<td>Allowed Cell Rate</td>
<td>ACR</td>
<td>The maximum current rate that can be used by the ABR source</td>
</tr>
</tbody>
</table>

All the above mentioned parameters are negotiated at the connection set-up time except for the case of ACR which changes during ABR’s connection lifetime and is set to the PCR initially by default.

2.4.2 RM Cell Structure

RM cells represent the media by which the network feedback is carried to ABR sources. As specified by the ATM Forum, the header of the RM cell is the same as the header of the data cells. However, the payload field of the RM cell has a different structure. The following is a list of a subset of the RM cell fields that is used in our later discussions.

1. **Congestion Indication (CI) bit**: is set to ‘1’ to indicate congestion in the network.

2. **No Increase (NI) bit**: is set to ‘1’ to stop the source from increasing its ACR. It signals mild congestion situation in the network.

3. **Explicit Rate (ER)**: is used to limit the source’s ACR to a specific value.

4. **Current Cell Rate (CCR)**: is used to indicate the current cell rate of the source.

5. **Minimum Cell Rate (MCR)**: is the MCR parameter assigned to the source.
2.4.3 Basic Operation

As shown in Figure 2.4, the basic operation of the ATM Forum specification can be summarized in the following steps:

1. Every $N_{rm}$ data cells, the source sends a forward RM cell. The fields of the RM cell are set as follows: ER = PCR (default), CI = 0, and NI = 0 (default).

2. When the forward RM cell passes through a congested switch along the connection path, two possible actions can take place:

   • If the switch is a binary feedback switch, then it cannot modify any field in the RM cell, but the congestion is indicated by setting the EFCI bit in the header of the data cells passing through the switch.

   • If the switch is an ER switch, then the value of the ER field can be reduced depending on the ER algorithm running on the switch. Also, the CI bit can be set by the switch. If the congestion is mild, then the switch can set the NI to stop the ABR source from increasing its ACR.

---

Figure 2.4  The ATM Forum Specification
3. At the destination, the value of the ER field can be reduced, the CI bit can be set, or the NI bit can be set. The destination plays another role by setting the CI bit in the RM cell if in the last data cell received, the EFCI bit was set. The destination returns the RM cell back to the source.

4. While the RM cell in its way back to the source (i.e. backward RM cell), the switches can do the same as in step 2 above.

5. When the RM cell is received by the source, one of the following actions will be taken by the source:

   - If CI = 0 and NI = 0, then ACR is increased to the minimum of \( ACR + PCR \cdot RIF \) and ER.
   - If CI = 1, then ACR is reduced to the minimum of \( ACR - ACR \cdot RDF \) and ER.
   - If CI = 0 and NI = 1 (i.e. there is no congestion but the source is not allowed to increase its rate), then ACR is set to the minimum of ACR and ER.

In addition to the above mentioned steps, the following points should be noted:

   - Active sources decrease their rates proportional to ACR after sending a specific number of cells without receiving any backward RM cells. This action acts as a guard against lost RM cells.
   - Both the destination and the intermediate switches can optionally generate backward RM cells to reduce the source rate at anytime instead of waiting until an RM cell passes through them.
It should be clear from the previous description that the ATM Forum specification was written in a very flexible way. It does not impose any specific congestion detection method, any cell scheduling algorithm, or even any switch type or architecture. In the specification, only few examples of possible ER switch algorithms, that can be implemented by switch vendors, are provided. The vendors are free to implement one of these algorithms or introduce their own.

2.5 Examples of ER Algorithms

As mentioned earlier in section 2.3.2.2, a large number of ER algorithms were proposed. Wide ranges of performance figures and implementation complexity are shown by these different algorithms. In this section, we shed the light on two of these algorithms, namely the Enhanced Proportional Rate Control Algorithm (EPRCA) and the Explicit Rate Indication for Congestion Avoidance (ERICA) algorithm. These two algorithms were among the first ones introduced and their performance was widely studied.

2.5.1 Enhanced Proportional Rate Control Algorithm (EPRCA)

EPRCA is one of the early explicit rate algorithms proposed to overcome the shortcomings of the binary-feedback algorithms [44][47]. The key advantage of this algorithm is to allow each congested switch to compute an estimate of the optimal cell rate for each VC with small processing and very low memory requirements. EPRCA maintains just one variable for each output queue and does not require per-VC accounting or per-VC queueing. EPRCA is an approximate ER algorithm.

The main difference between EPRCA and binary-feedback algorithms is that EPRCA performs intelligent marking when congestion is detected in the switch [6]. In EPRCA, only those
VCs, that are really causing the congestion, are requested to reduce their rates instead of blindly forcing all the VCs passing through the congested switch to reduce their rates as in the case of binary feedback algorithms.

An EPRCA switch can be in three states: non-congested, congested, and severely congested [47]. If the buffer occupancy level for an output port is higher than some threshold value, $QT$, then the switch is said to be congested. The switch will reach the severely congested state when its buffer occupancy level reaches another threshold value, $DQT$. If the buffer level is below $QT$ then the switch is not congested.

In order to approximate the optimum VC’s fairshare, the switch computes the Mean Allowed Cell Rate (MACR) for every output port. MACR is computed as an exponential running average by the following equation:

$$MACR = (1 - AV) \cdot MACR + AV \cdot CCR$$  

(EQ 1)

Where $CCR$ is copied from the forward RM cell and $AV$ is the exponential averaging factor set to $1/16$ by default. EPRCA’s basic operation as described in [44] is presented below. For other versions of EPRCA reader can refer to [7][18][30][34][46][47].

Depending on the direction of the RM cell passing through the switch, EPRCA performs different operations as follows:

1. If the received cell is a forward RM cell and the switch is non-congested, then MACR is updated according to (EQ 1) above only if CCR is higher than $MACR \times VCS$, where $VCS$ is the VC’s Separator. $VCS$ is set to $7/8$ by default.
2. If the received cell is a forward RM cell and the switch is congested or severely-congested, then $MACR$ is updated as in (EQ 1) only if the $CCR$ is less than $MACR$

3. If the received cell is a backward RM cell and the switch is non-congested then the RM cell is sent intact (i.e. not marked). This allows the source to increase its rate.

4. If the received cell is a backward RM cell and the switch is congested then the ER field in the RM cell is replaced by $\min(ER, MACR \times ERF)$ only if the $CCR$ is higher than $MACR \times DPF$. Where $ERF$ is called Explicit Reduction Factor and is set to $15/16$ by default. $DPF$ is the Down Pressure Factor and it is by default set to $7/8$.

5. If the received cell is a backward RM cell and the switch is severely-congested then the ER field in the RM cell is replaced by $\min(ER, MACR \times MRF)$, where $MRF$ is called Major Reduction factor and is set to $1/4$ by default.

Ideally, $MACR$ should be equal to the fair share according to the max-min fairness criterion [34]. Therefore, it is used to selectively mark the VCs. During congestion, those VCs whose rate is higher than $MACR$ will have their rates lowered to $MACR$ and those VCs whose rate is below $MACR$ are let to increase their rates. It must be noted here that the marking can be either binary by setting the CI bit in the RM cells or explicit by changing the ER field in the RM cells [44].

The most attractive feature of EPRCA is its low computation complexity and memory requirement as it keeps track of just only one variable per output port, namely $MACR$ [1]. Despite this key advantage over the other algorithms, it was suggested to increase the complexity of the
computation [47] or to use per-VC accounting [7] in order to increase the accuracy of the fair share calculation and hence the performance of the algorithm.

Although EPRCA maintains just only $MACR$ per output link, the number of tuning parameters ($VCS, AV, MRF, DPF$, and $ERF$) for EPRCA is really high. If these parameters are not carefully set, EPRCA will experience severe oscillations and $MACR$ will not converge to the correct fair share value. Moreover, the performance of EPRCA is highly affected by the values of source parameters (e.g. $RDF$ and $RIF$) [1].

The severely congested state of the EPRCA switch was shown that it may cause a serious fairness problem since while the switch in this state, it marks all the RM cells without any selectivity [46]. As a solution to this problem, the ‘severe congested’ state can be removed. However, this may lead to high buffer occupancy levels in some cases although the fairness is restored. To lower the buffer levels, buffer growth-based congestion detection method was suggested. So, the switch will record the buffer level every $N$ cells received and then will use this recorded value to determine whether the buffer size is increasing or decreasing, hence the switch is congested or not.

2.5.2 Explicit Rate Indication for Congestion Avoidance (ERICA)

The ERICA algorithm proposed in [26] is an exact ER algorithm that keeps Per-VC state information to calculate VCs fairshares accurately. ERICA proves to be a very robust and responsive algorithm [1]. Moreover, it introduces almost no oscillations in bandwidth allocations in the steady-state. Compared to EPRCA, ERICA has very few parameters which can be easily tuned. However, ERICA is a much more complex algorithm than EPRCA and it needs to keep a lot of information about each connection.
The basic ERICA algorithm works as follows. The switch calculates the load on each output link and determines a load factor. The load factor is updated periodically depending on the setting of the monitoring period duration. The load factor, $z$, is calculated according to the following ratio:

$$z = \frac{ABRInputRate}{ABRCapacity}$$

(EQ 2)

Where $ABR Input Rate$ is the rate of the ABR traffic during a monitoring interval and $ABR Capacity$ is equal to: $TargetUtilization \times LinkCapacity$. The value of the $Target Utilization$ parameter is close to, but less than 100%.

The load factor, $z$, indicates how congested is the link. A value of $z > 1$ indicates that the link is congested and a value of $z < 1$ means that the link is underutilized. Neither state is desirable and a value of $z = 1$, which is the optimal, must be maintained.

In addition to the load factor value, the switch calculates the $Fairshare$ value depending on the number of the active VCs during the last monitoring interval. $Fairshare$ is computed as follows:

$$Fairshare = \frac{ABRCapacity}{NumberofActiveVCs}$$

(EQ 3)

For each VC, the switch keeps the CCR value. This value gets updated whenever a forward RM cell is received by the switch. The CCR value of a specific VC is used to compute the
VCShare whenever a backward RM cell is received by the switch. VCShare is computed as follows:

\[
\text{VCShare} = \frac{CCR}{z}
\]  

(EQ 4)

Depending on the VCShare and the Fairshare values, the ER field in the backward RM cell is set to:

\[
ER = \min\{ER, \max(\text{VCShare}, \text{Fairshare})\}
\]  

(EQ 5)

The VCShare and the Fairshare values complement each other. Setting the ER field of the backward RM to the VCShare value aims at reaching the optimal operating point (i.e. \(z = 1\)). If all VCs change their rates to their VCShare values, then in the next monitoring interval, the load factor, \(z\), will be equal to unity. However, setting the ER field of the backward RM to the VCShare value does not guarantee fairness in bandwidth allocation among active VCs. On the other hand, setting the ER field to the Fairshare value aims at ensuring fairness and not necessarily optimal operation. For this reason, the two values are used to update the ER value in (EQ 5) above.

Several extensions to the basic ERICA algorithm were suggested [26]. Although these extensions were proposed to enhance the performance of the basic algorithm and to make it more robust, they resulted in a more complex algorithm [1].

2.6 Chapter Summary

This chapter introduced two types of congestion control in ATM networks, namely
preventive and reactive. Some general preventive techniques that solve the congestion problem were summarized. Reactive control, which is used only for ABR service category, was explained in more details by looking at different types of reactive schemes, comparing their performance and lastly giving examples of some of the algorithms that were suggested for implementing these schemes. Figure 2.5 below summarizes the congestion control mechanisms presented in this chapter.

Figure 2.5 Summary of congestion control techniques
Chapter 3  Interoperability Issues

The ATM Forum specification for ABR service does not recommend a specific switch algorithm to be implemented in switches involved in congestion control. ATM switch vendors are free to implement any type of algorithm in their switches as long as they adhere to the general guidelines given in the specification. Typically, as customers tend to buy their switches from different vendors, it is expected to have more than one switch type to coexist in the same network.

In this chapter, several issues of the interoperability between switch algorithms in multi-vendor networks are reviewed. First, the two main types of interoperability are presented along with the objectives of studying each type. Then, the factors that should be taken into account while evaluating the performance of heterogeneous networks are identified. Finally, a brief look at the previous work that was done in the area of interoperability is presented.

3.1 Types of Interoperability

Studying the interoperability of switch algorithms in ATM networks focuses mainly on looking at the performance of the whole network when several switch algorithms work together within the same feedback loop. Generally, switches can employ either binary feedback or explicit rate (ER) congestion control algorithms. Furthermore, ER switch algorithms are divided into either approximate or exact. Therefore, addressing the interoperability problem among different switch types involves investigating: 1) interoperability between binary feedback switches and ER switches, and 2) interoperability among ER switches of different types (e.g. exact and approximate).
3.1.1 Interoperability Between Binary Feedback and ER Switches

The low implementation cost of the binary feedback switches is by far the most valuable advantage of this type of switches. However, binary feedback switches were found to be generally unfair due to the well-known "beat-down" problem [24]. On the other hand, in spite of their better performance in terms of fairness and throughput, ER switches are much more complex to implement because they require per-VC (flow) computations.

Given the wide range of trade-offs between switch implementation cost and performance, studying the interoperability between binary feedback and ER switches aims at answering the following questions:

1. What is the impact of having a specific switch type at a specific location in the network (e.g. in the backbone or at the network edges)?
2. Given the high overhead complexity of implementing ER algorithms in switches, is it possible to alleviate the "beat-down" problem by strategically replacing a small subset of binary feedback switches by ER switches?
3. Considering the two types of ER switches: the approximate and the exact, which type performs better (i.e. more compatible) with binary feedback switches?
4. How does ABR source parameter setting affect the performance of heterogeneous networks in general?

3.1.2 Interoperability Between ER switches

The difference between the performance of approximate and exact ER switches is not as radical as the difference between the performance of binary feedback and ER switches. Generally
speaking, approximate algorithms tend to generate some oscillations in the traffic due to the simple mechanisms they use to compute the fairshare of each VC and, in some few cases, they may fail to compute the fairshare correctly. At the expense of higher implementation complexity, exact algorithms usually compute the VC fairshare accurately and, hence, generate an oscillations-free bandwidth allocation. For more details about the differences between ER algorithms see [1], and Chapter 5 of this thesis.

As in the case of the interoperability between binary feedback and ER switches, studying the interoperability among different ER-type switches is motivated by a similar range of trade-offs between switch implementation cost and performance. The goal of examining the interoperability among different ER-type switches is to answer the following questions:

1. What is the expected performance of networks consisting of a mix of approximate and exact ER switches? Is it comparable to the performance of a network employing one type of ER switches?
2. What is the impact of having a specific ER switch type at a specific location in the network?

3.2 Issues Affecting the Study of Interoperability

3.2.1 Switch Location

Although the ATM Forum specification allows for compatible operation between different switch types, performance of heterogeneous networks is highly affected by the location of each switch type. Moreover, for any specific VC, certain switches along its path will be more crucial than others [9].

In typical ATM networks, switches can be generally classified into: access switches and
backbone switches [36]. An access switch, which is usually deployed at the customer premises or service provider locations, adapts a user or a Local Area Network (LAN) traffic to ATM traffic. Access switches are referred to as “edge” switches since they reside at the ATM network edges. A backbone switch deals with ATM traffic aggregating from access switches or from other backbone switches. Since backbone switches reside at the heart of ATM network they are sometimes called “core” switches. Backbone switches are much larger in capacity than access switches as more traffic is expected at the heart of the network. Figure 3.1 shows an example of a typical ATM network configuration [36].

![A diagram of an ATM network configuration showing backbone and access switches](image)

**Figure 3.1** An example of a typical ATM network configuration

Switches along a VC path fall into two categories based on their impact on the fairshare computation for the VC [9]. The first category consists of all the switches controlling the most bottlenecked links along the VC’s path. In other words, according to the *max-min fairness* criterion [23], these switches give the least amount of bandwidth to the VC. These switches are
called “critical switches”. All other switches fall into the second category, the “non-critical” switches. Each VC has its own set of critical and non-critical switches. This set is not static and it changes during the lifetime of a VC because the available bandwidth for each VC over different links may change widely over time. Having a critical switch as a binary feedback or as an ER has a remarkably different impact on the performance of bandwidth allocation for a VC [9].

Although both access and backbone switches can be critical switches, we should expect to find more VCs having backbone switches as their critical switches since the traffic load is higher at the core of the network where a large number of VCs share the same links.

### 3.2.2 ABR Source Parameters Setting

Among the ABR source parameters (see chapter 2), rate increase factor (RIF) and rate decrease factor (RDF) have the highest impact on network performance since they control the amount of bandwidth by which the source can increase its rate if there is no congestion in the network or decrease its rate when congestion is detected [1]. It is very important to realize that the more the congestion control scheme is dependent on these parameters for changing source rate, the less robust it is.

The binary feedback scheme depends solely on RIF and RDF to change the source rate. Therefore, one should expect the impact of these parameters setting to be most prevalent in the case of binary feedback schemes. Most approximate ER schemes are also sensitive to the choice of RIF and RDF values but to a much lesser degree than binary feedback switches [1]. On the other hand, these parameters do not dictate the performance of exact ER schemes at all.

Several studies were devoted to determine the optimal values for RIF and RDF that can
enhance network performance in terms of link utilization and buffer requirements [22][39][49]. These studies provide general guidelines for setting up these parameters in homogenous networks consisting of binary feedback switches only. Unfortunately, the guidelines offered by these studies cannot be used for setting up RIF and RDF values for heterogeneous networks as network conditions are different from those used to derive the equations that control the behavior of homogenous networks. Moreover, most of these studies were based on highly simplified network conditions in order to ease the analysis process. Since it is expected that network conditions will be much more complex in real ATM networks, especially in heterogeneous ones, following these guidelines becomes less important. It must be noted here that although these guidelines are generally not accurate enough to be used for optimizing the performance of heterogeneous networks, they do help in understanding the general impact of certain values of RIF and RDF on network performance.

Given the fact that RIF and RDF values are negotiated at connection setup time [2][45], it is possible to have their values not set to the values requested by the source. Since different VCs may have their paths set up through different subsets of switches, it is possible not to have these parameters set to the same values for different VCs even for sources asking for the same setting. Therefore, in heterogeneous networks where different switch types are co-existing such scenario will be more likely to happen, and finding the best locations for each switch type regardless of the RIF/RDF parameters setting will be of great interest.

Another ABR source parameter that is worth investigation is the Peak Cell Rate (PCR). This parameter limits the maximum rate that can be used by a specific ABR source. Despite its importance, most of the studies that deal with ABR traffic sets PCR value to the link capacity to
avoid any underutilization of the bandwidth. It is interesting to evaluate the impact of PCR value setting on the performance of heterogeneous networks as it may lower the unfairness levels experienced in some network configurations.

### 3.2.3 Switch Algorithm Parameters

The various proposed types of ER algorithms differ mainly in the number and type of tuning parameters that need to be controlled by each algorithm. This is due to the various number of ways used by different algorithms to compute the fairshare. Approximate ER algorithms use quite a large number of parameters compared to exact ER algorithms because approximate algorithms tend to estimate the fairshare using a small number of computations and without the need for per-connection information [1]. Additionally, incorrect settings of these parameters can result in large oscillations in rates and potential unfairness. Consequently, the performance of heterogeneous networks can be affected by such parameters in general. However, switch parameters do not have the same level of impact on the performance of heterogeneous networks as the source parameters. This is due to the fact that source parameter settings can affect not only the performance of ER switches, but also the performance of binary feedback switches.

### 3.2.4 Mixed Traffic Categories

In real world environments, ABR traffic will not be the only type of traffic running over network links. Other traffic categories such as Variable Bit Rate (VBR) and Constant Bit Rate (CBR) will always co-exist and share the same links with ABR traffic. Although the influence of other traffic types on ABR is not directly related to the issue of interoperability, it is important to investigate how ABR congestion control mechanisms will perform in the presence of other traffic types in heterogeneous networks. Two questions that suggest themselves here are:
• How are the results obtained by evaluating the performance of heterogeneous networks with ABR traffic only applicable to the heterogeneous networks supporting other traffic types (especially VBR) besides ABR?

• How do VBR traffic dynamics impact the performance of control mechanisms for ABR?

3.3 Previous Work

The behavior of congestion control mechanisms for ABR in homogenous networks has been studied extensively by many researchers [10][11][19][26][28][37][43]. However, despite its extreme importance especially from a practical point of view, the issue of interoperability between different rate-based congestion control schemes in heterogeneous networks was hardly addressed in the literature. Unlike the work reported in this thesis (see Chapter 5), most of the previous studies focused on one type of interoperability, namely the interoperability between binary feedback and explicit rate switch mechanisms [8][9][52]. In the following, we review briefly some previous investigations.

Chang et al. [8][9] studied the performance of ATM networks consisting of binary feedback switches and switches running the US National Institute of Standards and Technology (NIST) ER scheme. Several network configurations were considered in the study. It was concluded that replacing one or more binary feedback switches with explicit rate switches may or may not enhance the performance of networks consisting of binary feedback switches only. Whether the explicit rate switch will improve the performance or not depends on the location of that switch in the network. Moreover, by looking at one parameter, namely RIF, this study pointed out that the selection of ABR source parameters can highly affect the performance of heteroge-
In another study of interoperability, it was reported that replacing the network edge switches by binary feedback switches in networks using ER mechanisms in all switches may result in performance degradation in terms of convergence time, throughput, and buffer occupancy [52]. This study considered several simple network configurations and used the NIST ER scheme too. In order to look at the performance of the network in the transient state, VBR traffic was also introduced in all network configurations simulated.

In all the previous studies of the interoperability between binary feedback and ER algorithms only one ER algorithm was considered by each study. Given the fact that there are approximate and exact ER algorithms, a more detailed look at the issue of interoperability is needed. Furthermore, as the binary feedback scheme is highly dependent on source parameter setting, different settings for these parameters must be considered.

The only study we found dealing with interoperability between ER algorithms among themselves is the one conducted by Plotkin et. al. [40]. In this study, a “rate mismatch problem” between different VCs in heterogeneous networks was identified. This problem occurs when two VCs having the same critical switch (according to the max-min fairness criterion) experience the same rate decrease during congestion periods but increase their rates differently when there is no congestion. The rate mismatch problem causes unfairness in rate allocations for VCs in some heterogeneous networks depending on the type of the ER switches used. The well-known “parking-lot” network topology was used in this study [24].

In the next two chapters, the two interoperability problems outlined above will be investi-
gated in much more detail. The aim is to provide a much more accurate characterization of such interoperability issues.
Chapter 4 Interoperability between Binary Feedback and Explicit Rate Algorithms

Despite the fact that explicit rate (ER) algorithms outperform binary feedback algorithms in several aspects, namely fairness, robustness, and responsiveness, the interest in binary feedback algorithms still exists. Binary feedback algorithms possess a low implementation complexity which makes this type of rate-based algorithms a quick solution for supporting ABR service in ATM switches, at least until more efficient and less complex ER algorithms are introduced. Therefore, until a full deployment of ER algorithms, it is very likely to have both binary feedback and ER algorithms work together in the same ATM network. For more details about the main differences between binary feedback and ER algorithms reader is referred to Chapter 2 in this thesis.

This chapter evaluates the performance of heterogeneous networks consisting of both binary feedback and ER algorithms. First, the simulation model used in the performance evaluation is described. The simulated network topology along with the source and switch algorithms parameters are introduced. Then, the simulation results, where we look at two ER algorithms and different settings for source parameters, are presented.

4.1 Simulation Model

For our simulation purposes, we use OPNET™, a discrete event simulation package. OPNET™ features a sophisticated Graphical User Interface (GUI) and a wide range of built-in communications protocols models. The built-in ATM model does not support the ATM Forum specification for the ABR service, so extensive modifications were applied to the provided model in order to implement the ATM Forum specification.
The main components of the ABR source model used in our simulations are depicted in Figure 4.1. The traffic source sends packets continuously at a rate of 155.52 Mbps. When a packet is received by the AAL (specifically, AAL 5) module, it is inserted into a small input buffer of size 20 packets. The AAL module takes one packet at a time from the input buffer and segments it into ATM data cells and then sends the cells to the ATM layer module. Moreover, the AAL module sends an RM cell every $N_{rm}$ data cells sent to the ATM layer module. The Segmentation And Reassembly (SAR) rate of the AAL module, which indicates the maximum number of cells that can be produced or reassembled at the AAL per second, is varied according to the feedback carried by the backward RM cells. Although, according to the ATM reference model the congestion control function is part of the ATM layer, we chose to implement it in the AAL module because it is difficult to control the ABR source rate from the ATM layer module in the built-in ATM model provided by OPNET™.

OPNET™ provides a built-in output-buffered ATM switch model. For each output port, four separate output queues are implemented; one for each service category (CBR, VBR, ABR and UBR). VBR and CBR traffics have higher priority than ABR traffic. In other words, all VBR and CBR cells waiting in their output queues will be serviced before sending any cell from the ABR output queue.

### 4.1.1 Network Topology

Figure 4.2 shows the network topology used in our simulations. It consists of five switches and seven ABR VCs. Each VC $i$ has a source $S_i$ and a destination $D_i$. Two VCs (1 and 2) traverse all five switches while two other VCs (4 and 5) traverse only three switches. All other VCs (3, 6 and 7) pass through one switch to create some cross traffic in the switches. The source-to-switch
delay is higher for VCs 2 and 5 in order to study the effect of having different link delays. All the links have the same capacity of 155.52 Mbps. Link delay is assumed to be 5 μsec/km.

This topology was carefully designed to cover many different possible congestion scenarios found in real ATM networks. Switches SW1 and SW5 act as access switches. Switch SW1 combines the traffic coming from sources S1, S2, and S3 into one output link. Switch SW5 splits the combined traffic coming from sources S1, S2, S4, S5, and S7 into five output links each.
Figure 4.2  Simulated network topology

connected to the corresponding destination. In this topology, switches SW2, SW3, and SW4 can be viewed as backbone switches, where most of the traffic is concentrated. It should be emphasized here that sources S4, S5, S6, and S7 are meant to create additional traffic coming to the backbone switches from other sections of the ATM network (e.g. traffic coming from other backbone switches) and not necessarily traffic coming from single sources as may be understood from Figure 4.2.

4.1.2 Switch Mechanisms

For binary feedback switches, buffer occupancy level is used as the basis for detecting congestion. As long as the buffer length is below a predetermined threshold value (200 cells in our model), the switch is considered not congested. The switch becomes congested if the buffer length exceeds the threshold value.

For explicit rate (ER) switches, two congestion control algorithms are considered. One is an approximate ER algorithm, Enhanced Proportional Rate Control Algorithm (EPRCA) [44][47], and the other is an exact algorithm, Explicit Rate Indication for Congestion Avoidance
Chapter 4 Interoperability between Binary Feedback and Explicit Rate Algorithms

(ERICA) [26]. For more details about these two algorithms see Chapter 2. The motivation behind investigating two different types of ER algorithms is to ensure that our results are valid for both types of ER algorithms. Table 4.1 shows the parameter settings used for the EPRCA algorithm. These parameters have been defined in section 2.5.1. Setting \( DQT \) to such a high value compared to \( DQ \) lowers the possibility of driving the EPRCA switch into the “severely-congested” state. It was shown that this state causes unfairness because, when in this state, the switch marks all VCs without any selectivity [46]. For the other parameters, the default values are used [44][47]. In contrary to EPRCA, ERICA maintains a small number of tuning parameters. These parameters are the \textit{averaging interval length} and the switch \textit{target utilization} which are set to \( 0.5 \text{ms} \) and \( 95\% \) respectively. A switching delay of \( 1^{-10} \text{sec.} \) and infinite buffer sizes are assumed for all switches.

Table 4.1 EPRCA parameter settings

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQ</td>
<td>100</td>
</tr>
<tr>
<td>DQT</td>
<td>5000</td>
</tr>
<tr>
<td>VCS</td>
<td>( 7/8 )</td>
</tr>
<tr>
<td>MRF</td>
<td>( 1/4 )</td>
</tr>
<tr>
<td>DPF</td>
<td>( 7/8 )</td>
</tr>
<tr>
<td>ERF</td>
<td>( 15/16 )</td>
</tr>
<tr>
<td>AV</td>
<td>( 1/16 )</td>
</tr>
</tbody>
</table>

4.1.3 Source Parameters

All ABR traffic sources used in our simulations are greedy (i.e. persistent). The use of greedy sources allows for testing the network under worst case conditions where all the sources send traffic at their ACR at the same time. Table 4.2 shows the source parameter settings used in our simulations. For those parameters that take more than one value, namely PCR, RIF, and RDF,
the value used in the simulation will be specified before presenting the corresponding results.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>155.52 or 100 Mbps</td>
</tr>
<tr>
<td>MCR</td>
<td>1 Mbps</td>
</tr>
<tr>
<td>ICR</td>
<td>20 Mbps</td>
</tr>
<tr>
<td>Nrm</td>
<td>32</td>
</tr>
<tr>
<td>RIF</td>
<td>1/128 or 1/512</td>
</tr>
<tr>
<td>RDF</td>
<td>1/32 or 1/512</td>
</tr>
</tbody>
</table>

4.2 Simulation Results

In order to study the impact of switch type location on the performance of the network, six different cases of switch locations are considered:

1. All switches are ER.
2. All switches are binary feedback.
3. Switches SW2,3,4,5 are binary feedback and switch SW1 is ER.
4. Switches SW1,2,4,5 are binary feedback and switch SW3 is ER.
5. Switches SW1,2,5 are binary feedback and switches SW3,4 are ER.
6. Switches SW2,4,5 are binary feedback and switches SW1,3 are ER.

Cases 1 and 2 are used to determine baseline performance provided by a homogeneous network that employs a single switch type. Cases 3 and 4 are designed to evaluate the impact of replacing one binary feedback switch with an ER switch. In case 3, access switch SW1 is replaced by an ER switch where SW1 is a non-critical switch for VCs 1 and 2. In case 4, SW3 which is a critical switch for VCs 1, 2, 4, 5 and 6 is replaced by an ER switch. In Cases 5 and 6 two binary feedback switches are replaced by ER switches. In case 5, the backbone switches are ER switches.
and in case 6 an edge switch and a backbone switch are replaced by ER switches. It must be noted here that SW5 type has no impact at all on the performance of the network as no congestion is experienced at its output links.

Since the switch location affects the performance of each VC, it is important to show which switch types are critical for each VC. Such information is supplied by Table 4.3 below. For those VCs not traversing a specific switch, “N/A” is used to mean not applicable.

**Table 4.3 Switch location criticality**

<table>
<thead>
<tr>
<th>VC</th>
<th>SW1</th>
<th>SW2</th>
<th>SW3</th>
<th>SW4</th>
<th>SW5</th>
</tr>
</thead>
<tbody>
<tr>
<td>VC1</td>
<td>non-critical</td>
<td>non-critical</td>
<td>critical</td>
<td>critical</td>
<td>non-critical</td>
</tr>
<tr>
<td>VC2</td>
<td>non-critical</td>
<td>non-critical</td>
<td>critical</td>
<td>critical</td>
<td>non-critical</td>
</tr>
<tr>
<td>VC3</td>
<td>critical</td>
<td>non-critical</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>VC4</td>
<td>N/A</td>
<td>N/A</td>
<td>critical</td>
<td>critical</td>
<td>non-critical</td>
</tr>
<tr>
<td>VC5</td>
<td>N/A</td>
<td>N/A</td>
<td>critical</td>
<td>critical</td>
<td>non-critical</td>
</tr>
<tr>
<td>VC6</td>
<td>N/A</td>
<td>N/A</td>
<td>critical</td>
<td>non-critical</td>
<td>N/A</td>
</tr>
<tr>
<td>VC7</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>critical</td>
<td>non-critical</td>
</tr>
</tbody>
</table>

To evaluate the performance, the average throughput for each VC divided by its fair rate allocation, according to the *max-min fairness criterion*, as well as the overall fairness are used. The average throughput is computed as the average rate of bits received at the destination during the period from time = 50ms to time = 1000ms of the simulation time, where all simulations start as time zero. The first 50ms are not included in the average throughput calculation to eliminate false results due to any transient behavior that may show up during the ramp-up time.

Rate allocation fairness is computed using the *fairness index* proposed in [25] (see Appendix A). Fairness is bounded between 0 to 100% where a value of 100% would be ideal. According to the *max-min fairness* criterion, the fair rate allocation for each VC in the simulated
network topology is: 31.1 Mbps for VCs 1,2,4,5,6,7 and 93.3 Mbps for VC3.

In the following sections, the performance of heterogeneous networks consisting of both binary feedback and ER switches is analyzed by presenting the results of our simulations. The impact of the source parameters: RIF, RDF and PCR on the performance of heterogeneous networks is evaluated by varying the values for these parameters.

4.2.1 Impact of RIF and RDF

As mentioned earlier, two ER algorithms, namely ERICA and EPRCA, are considered in our simulations. For each algorithm, we report the results for four sets of simulations. The sets differ in the values of the parameters RIF and RDF for the multi-link VCs (i.e. VCs 1,2,4 and 5) only. For single-link VCs (i.e. VCs 3,6, and 7), RDF and RIF are both given a value of 1/512 in all simulation sets. Table 4.4 shows the values used for each simulation set for the case of multi-link VCs.

<table>
<thead>
<tr>
<th>Set No.</th>
<th>RIF</th>
<th>RDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/512</td>
<td>1/512</td>
</tr>
<tr>
<td>2</td>
<td>1/128</td>
<td>1/512</td>
</tr>
<tr>
<td>3</td>
<td>1/128</td>
<td>1/32</td>
</tr>
<tr>
<td>4</td>
<td>1/512</td>
<td>1/32</td>
</tr>
</tbody>
</table>

4.2.1.1 ERICA and Binary feedback Algorithms Interoperability

In the first set of simulations, we consider a conservative value of 1/512 for both RIF and RDF for all VCs. Table 4.5 summarizes the simulation results for this set. As should be expected, the best performance in terms of fairness and throughput is achieved by using the ER scheme in all switches (case 1). The throughput of all VCs is about 95% since the target utilization of
ERICA was set to 95%. In case 2, where all the switches were replaced by binary feedback switches, we notice a large drop in multi-link VCs throughput and the overall fairness due to the well-known "beat-down" problem experienced by the multi-link VCs [5]. Furthermore, VCs 1 and 2 experience more degradation in their performance than VCs 4 and 5 since they traverse more switches and therefore have a higher chance to be beaten down (see Figure 4.3 below).

For cases 3 and 4, where only one switch is an ER and the rest are binary feedback, a clear performance improvement over the all-binary switches case (case 2) is noticed since the number of binary feedback switches is decreased, and the chance of having the multi-link VCs beaten down drops. In case 3, SW1, which is a non-critical switch for VCs 1 and 2, is an ER switch and according to the max-min fairness criterion, SW1 gives VCs 1,2 and 3 a rate allocation of 51.84 Mbps (higher than the rate given by the critical switches which is 31.1 Mbps). Both VCs 1 and 2 are given a higher chance to increase their rates because SW1 tries to limit the bandwidth of VC 3. Moreover, since switch SW1 is no longer a binary feedback switch, VCs 1 and 2 have similar chances to be beaten down as VCs 4 and 5. Consequently, the throughputs of these four VCs is almost the same. In case 4, SW3 is a critical switch for VCs 1 and 2, so their throughputs is expected to be lower than it was in case 3. VCs 4 and 5 experience less "beat-down" than VCs 1

### Table 4.5 Percentage throughput and fairness for the first simulation set (with ERICA) (RIF=1/512 RDF=1/512)

<table>
<thead>
<tr>
<th>No.</th>
<th>Case</th>
<th>VC1</th>
<th>VC2</th>
<th>VC3</th>
<th>VC4</th>
<th>VC5</th>
<th>VC6</th>
<th>VC7</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All ER</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>All Binary</td>
<td>44</td>
<td>44</td>
<td>136</td>
<td>63</td>
<td>62</td>
<td>281</td>
<td>280</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>All Binary, SW1 = ER</td>
<td>71</td>
<td>70</td>
<td>111</td>
<td>72</td>
<td>70</td>
<td>209</td>
<td>209</td>
<td>78</td>
</tr>
<tr>
<td>4</td>
<td>All Binary, SW3 = ER</td>
<td>47</td>
<td>47</td>
<td>133</td>
<td>92</td>
<td>92</td>
<td>194</td>
<td>220</td>
<td>77</td>
</tr>
<tr>
<td>5</td>
<td>All Binary, SW3,4 = ER</td>
<td>94</td>
<td>94</td>
<td>103</td>
<td>95</td>
<td>95</td>
<td>96</td>
<td>96</td>
<td>99</td>
</tr>
<tr>
<td>6</td>
<td>All Binary, SW1,3 = ER</td>
<td>77</td>
<td>77</td>
<td>103</td>
<td>78</td>
<td>77</td>
<td>160</td>
<td>163</td>
<td>89</td>
</tr>
</tbody>
</table>
and 2 as they traverse less binary feedback switches.

In cases 5 and 6, more performance improvement is achieved because the number of binary feedback switches is decreased further. Case 5 shows higher performance improvement than case 6 since for all VCs, the critical switches (namely SW3 and SW4) are ER switches. Having SW3 and SW4 as ER switches means that VCs 6 and 7 cannot cause any beat-down problems for the other VCs. It is interesting to observe that cases 1 and 5 have almost the same performance figures, even though the network of case 5 employs only two ER switches. This makes the network setup of case 5 a very attractive choice.

Figure 4.3  ACR dynamics for All-binary feedback switches case (RIF = 1/512 RDF = 1/512)
In the second simulation set, RIF of the multi-link VCs is set to a more aggressive value of 1/128 and the RDF remains the same. By setting RIF to this value, multi-link VCs have an advantage over the single-link VCs during the rate increase periods in the sense that they can increase their rates faster. This acts as some type of compensation for the "beat-down" problem that affects the multi-link VCs. Table 4.6 summarizes the results for this simulation set. Generally speaking, the performance of all cases is improved by giving the multi-link VCs a more aggressive RIF. However, the amount of improvement in the performance is not equal for all cases. The reason is that although the beat-down problem can happen when using binary feedback switches, there is no guarantee that it will always show up. Moreover, even if it shows up, the severity of the beat-down problem may not be equal for all cases. Nevertheless, it is very obvious that the performance of this set (with RIF = 1/128) is much better than it was in the previous set (with RIF = 1/512).

<table>
<thead>
<tr>
<th>No.</th>
<th>Case</th>
<th>VC1</th>
<th>VC2</th>
<th>VC3</th>
<th>VC4</th>
<th>VC5</th>
<th>VC6</th>
<th>VC7</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All ER</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>All Binary</td>
<td>78</td>
<td>72</td>
<td>116</td>
<td>138</td>
<td>126</td>
<td>73</td>
<td>81</td>
<td>93</td>
</tr>
<tr>
<td>3</td>
<td>All Binary, SW1 = ER</td>
<td>109</td>
<td>101</td>
<td>87</td>
<td>110</td>
<td>104</td>
<td>64</td>
<td>72</td>
<td>97</td>
</tr>
<tr>
<td>4</td>
<td>All Binary, SW3 = ER</td>
<td>66</td>
<td>65</td>
<td>120</td>
<td>93</td>
<td>91</td>
<td>156</td>
<td>182</td>
<td>88</td>
</tr>
<tr>
<td>5</td>
<td>All Binary, SW3,4 = ER</td>
<td>93</td>
<td>94</td>
<td>103</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>All Binary, SW1,3 = ER</td>
<td>93</td>
<td>93</td>
<td>95</td>
<td>93</td>
<td>93</td>
<td>102</td>
<td>127</td>
<td>99</td>
</tr>
</tbody>
</table>

In the third simulation set, the multi-link VCs are endowed with an aggressive RDF value of 1/32 while keeping the RIF value as it was in the second set. Setting RDF to this value makes the effect of the "beat-down" problem worse since the ACR of the multi-link VCs is lowered more whenever an RM cell is received with the CI bit set. The results of this set are shown in Table 4.7.
Multi-link VCs throughput is generally lower than it is in the first simulation set due to the worse "beat-down" effect. However, the general picture of the results for this set is very comparable to that of the first set. Case 5 is still showing a high performance figures compared to the other cases.

The results for the last set of simulations we consider are shown in Table 4.8. In this set, we keep the aggressive value of RDF = 1/32 for the multi-link VCs, but we change the RIF value back to 1/512. For the multi-link VCs, this parameter setting is considered the worst among the settings we consider. Multi-link VCs are more affected by the beat-down problem since their RDF is aggressive, while their RIF is small. The main purpose for considering this parameter setting is to test whether the network of case 5 can still keep its good performance. As shown in Table 4.8, the throughput of all multi-link VCs in mixed switch types is very low. An exception of this is case 5 whereby VCs 4 and 5 show a very high throughput although VCs 1 and 2 experience a low throughput. Moreover, case 5 is still showing an excellent fairness figure compared to the other cases.

<table>
<thead>
<tr>
<th>No.</th>
<th>Case</th>
<th>VC1</th>
<th>VC2</th>
<th>VC3</th>
<th>VC4</th>
<th>VC5</th>
<th>VC6</th>
<th>VC7</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All ER</td>
<td>95</td>
<td>95</td>
<td>94</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>All Binary</td>
<td>33</td>
<td>31</td>
<td>143</td>
<td>63</td>
<td>57</td>
<td>300</td>
<td>306</td>
<td>58</td>
</tr>
<tr>
<td>3</td>
<td>All Binary, SW1 = ER</td>
<td>55</td>
<td>50</td>
<td>121</td>
<td>56</td>
<td>51</td>
<td>270</td>
<td>279</td>
<td>63</td>
</tr>
<tr>
<td>4</td>
<td>All Binary, SW3 = ER</td>
<td>60</td>
<td>54</td>
<td>124</td>
<td>62</td>
<td>59</td>
<td>209</td>
<td>258</td>
<td>70</td>
</tr>
<tr>
<td>5</td>
<td>All Binary, SW3,4 = ER</td>
<td>70</td>
<td>69</td>
<td>118</td>
<td>109</td>
<td>114</td>
<td>111</td>
<td>111</td>
<td>96</td>
</tr>
<tr>
<td>6</td>
<td>All Binary, SW1,3 = ER</td>
<td>61</td>
<td>57</td>
<td>117</td>
<td>62</td>
<td>58</td>
<td>180</td>
<td>255</td>
<td>71</td>
</tr>
</tbody>
</table>

As can be seen from Tables 4.5 through 4.8, the performance of VCs 1 and 2, and VCs 4...
Table 4.8 Percentage throughput and fairness for the fourth simulation set (with ERICA) (RIF=1/512 RDF=1/32)

<table>
<thead>
<tr>
<th>No.</th>
<th>Case</th>
<th>VC1</th>
<th>VC2</th>
<th>VC3</th>
<th>VC4</th>
<th>VC5</th>
<th>VC6</th>
<th>VC7</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All ER</td>
<td>94</td>
<td>93</td>
<td>93</td>
<td>95</td>
<td>94</td>
<td>97</td>
<td>97</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>All Binary</td>
<td>17</td>
<td>22</td>
<td>139</td>
<td>22</td>
<td>25</td>
<td>361</td>
<td>357</td>
<td>46</td>
</tr>
<tr>
<td>3</td>
<td>All Binary, SW1 = ER</td>
<td>22</td>
<td>30</td>
<td>138</td>
<td>19</td>
<td>23</td>
<td>354</td>
<td>351</td>
<td>47</td>
</tr>
<tr>
<td>4</td>
<td>All Binary, SW3 = ER</td>
<td>17</td>
<td>17</td>
<td>154</td>
<td>31</td>
<td>30</td>
<td>376</td>
<td>402</td>
<td>46</td>
</tr>
<tr>
<td>5</td>
<td>All Binary, SW3,4 = ER</td>
<td>33</td>
<td>32</td>
<td>143</td>
<td>137</td>
<td>127</td>
<td>143</td>
<td>143</td>
<td>84</td>
</tr>
<tr>
<td>6</td>
<td>All Binary, SW1,3 = ER</td>
<td>26</td>
<td>25</td>
<td>129</td>
<td>26</td>
<td>29</td>
<td>334</td>
<td>316</td>
<td>51</td>
</tr>
</tbody>
</table>

and 5 is almost the same for most of the cases. Having a shorter source-to-switch delay does not result in better performance and vice versa. This conclusion is based on our simulations and matches with some earlier results reported in [13]. However, we cannot generalize this result since for some types of links, such as satellite links, the delay is very high and it may affect the performance of the network differently especially for short-lived connections. This issue will not be further discussed here as it is out of the scope of this thesis.

4.2.1.2 EPRCA and Binary Feedback Algorithms Interoperability

Except for slight differences in some cases, the simulation results for EPRCA are very consistent with the results shown above for the case of ERICA. Tables 4.9 through 4.12 summarize the simulation results for EPRCA.

Slight differences in the performance of heterogeneous networks between the case of ERICA and the case of EPRCA should be expected due to the different mechanisms employed by each algorithm. ERICA is a congestion avoidance algorithm so it tries to keep the output link utilization slightly below 100%. In our case, the target link utilization was set to 95%. The effect of this can be clearly seen by comparing the first row of Table 4.5 with the first row of Table 4.9,
where all the switches are ER. One can notice that under ERICA, the maximum throughput that was achieved by any VC was 95%. However, under EPRCA, throughput of each VC is approximately 100%.

Table 4.9 Percentage throughput and fairness for the first simulation set (with EPRCA) (RIF=1/512 RDF=1/512)

<table>
<thead>
<tr>
<th>No.</th>
<th>Case</th>
<th>VC1</th>
<th>VC2</th>
<th>VC3</th>
<th>VC4</th>
<th>VC5</th>
<th>VC6</th>
<th>VC7</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All ER</td>
<td>98</td>
<td>98</td>
<td>99</td>
<td>98</td>
<td>98</td>
<td>100</td>
<td>103</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>All Binary</td>
<td>44</td>
<td>44</td>
<td>136</td>
<td>63</td>
<td>62</td>
<td>281</td>
<td>280</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>All Binary, SW1 = ER</td>
<td>71</td>
<td>71</td>
<td>118</td>
<td>72</td>
<td>70</td>
<td>207</td>
<td>207</td>
<td>79</td>
</tr>
<tr>
<td>4</td>
<td>All Binary, SW3 = ER</td>
<td>64</td>
<td>63</td>
<td>123</td>
<td>106</td>
<td>106</td>
<td>154</td>
<td>157</td>
<td>91</td>
</tr>
<tr>
<td>5</td>
<td>All Binary, SW3,4 = ER</td>
<td>96</td>
<td>96</td>
<td>101</td>
<td>101</td>
<td>101</td>
<td>104</td>
<td>104</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>All Binary, SW1,3 = ER</td>
<td>89</td>
<td>89</td>
<td>106</td>
<td>89</td>
<td>88</td>
<td>136</td>
<td>143</td>
<td>96</td>
</tr>
</tbody>
</table>

More importantly, since EPRCA causes oscillations in the VC rate as opposed to ERICA, the effect of the beat down problem can be either higher or lower than in the case of ERICA. To explain this, we use the term “ACR cycle” to refer to the period during which a VC source increases its rate (i.e. the ACR) until congestion is detected in the switch and then start decreasing its rate until the congestion is relieved. This cycle is repeated again and again during a VC’s lifetime. ABR traffic controlled by switches that can cause rate oscillations, such as binary feedback and EPRCA switches, experiences this “ACR cycle”. However, such behavior does not show up in traffic controlled by a switch algorithm like ERICA since this algorithm generates oscillation-free traffic in the steady-state. Now, if the ACR cycle caused by one switch algorithm was in phase with the cycle caused by the other algorithm we end up having the VCs traversing both switches affected by one cycle only. On the other hand, if the two cycles are not in phase, then the VCs traversing both switches will be affected by both cycles and this is the worst case. If we apply this observation to the network of Figure 4.2, one can observe that when the binary
feedback switches ask the sources to reduce their rates while the sources are already reducing their rates in response to feedback from the EPRCA switches, then the beat-down problem will not show up at all. By the same token, if the binary feedback (respectively, EPRCA) switches are asking the sources to reduce their rates while these sources are trying to increase their rates according to the feedback coming from EPRCA (respectively, binary) switches then the net result will be to reduce the rate. This observation is a general one and can be applied to any two or more algorithms causing oscillations in the traffic.

Table 4.10  Percentage throughput and fairness for the second simulation set (with EPRCA)  
(RIF=1/128 RDF=1/512)

<table>
<thead>
<tr>
<th>No.</th>
<th>Case</th>
<th>VC1</th>
<th>VC2</th>
<th>VC3</th>
<th>VC4</th>
<th>VC5</th>
<th>VC6</th>
<th>VC7</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All ER</td>
<td>101</td>
<td>100</td>
<td>93</td>
<td>101</td>
<td>100</td>
<td>94</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>All Binary</td>
<td>78</td>
<td>72</td>
<td>116</td>
<td>138</td>
<td>126</td>
<td>73</td>
<td>81</td>
<td>93</td>
</tr>
<tr>
<td>3</td>
<td>All Binary, SW1 = ER</td>
<td>107</td>
<td>101</td>
<td>96</td>
<td>113</td>
<td>105</td>
<td>62</td>
<td>70</td>
<td>96</td>
</tr>
<tr>
<td>4</td>
<td>All Binary, SW3 = ER</td>
<td>90</td>
<td>90</td>
<td>102</td>
<td>91</td>
<td>91</td>
<td>112</td>
<td>135</td>
<td>98</td>
</tr>
<tr>
<td>5</td>
<td>All Binary, SW3,4 = ER</td>
<td>97</td>
<td>97</td>
<td>101</td>
<td>104</td>
<td>104</td>
<td>97</td>
<td>97</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>All Binary, SW1,3 = ER</td>
<td>91</td>
<td>91</td>
<td>103</td>
<td>91</td>
<td>91</td>
<td>105</td>
<td>133</td>
<td>98</td>
</tr>
</tbody>
</table>

Table 4.11  Percentage throughput and fairness for the third simulation set (with EPRCA)  
(RIF=1/128 RDF=1/32)

<table>
<thead>
<tr>
<th>No.</th>
<th>Case</th>
<th>VC1</th>
<th>VC2</th>
<th>VC3</th>
<th>VC4</th>
<th>VC5</th>
<th>VC6</th>
<th>VC7</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All ER</td>
<td>101</td>
<td>100</td>
<td>93</td>
<td>101</td>
<td>100</td>
<td>94</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>All Binary</td>
<td>33</td>
<td>31</td>
<td>143</td>
<td>63</td>
<td>57</td>
<td>300</td>
<td>306</td>
<td>58</td>
</tr>
<tr>
<td>3</td>
<td>All Binary, SW1 = ER</td>
<td>54</td>
<td>49</td>
<td>130</td>
<td>58</td>
<td>52</td>
<td>270</td>
<td>278</td>
<td>64</td>
</tr>
<tr>
<td>4</td>
<td>All Binary, SW3 = ER</td>
<td>53</td>
<td>49</td>
<td>128</td>
<td>65</td>
<td>59</td>
<td>221</td>
<td>262</td>
<td>68</td>
</tr>
<tr>
<td>5</td>
<td>All Binary, SW3,4 = ER</td>
<td>72</td>
<td>70</td>
<td>117</td>
<td>121</td>
<td>121</td>
<td>108</td>
<td>112</td>
<td>96</td>
</tr>
<tr>
<td>6</td>
<td>All Binary, SW1,3 = ER</td>
<td>61</td>
<td>56</td>
<td>125</td>
<td>63</td>
<td>58</td>
<td>161</td>
<td>252</td>
<td>72</td>
</tr>
</tbody>
</table>
Table 4.12 Percentage throughput and fairness for the fourth simulation set (with EPRCA) (RIF=1/512 RDF=1/32)

<table>
<thead>
<tr>
<th>No.</th>
<th>Case</th>
<th>VC1</th>
<th>VC2</th>
<th>VC3</th>
<th>VC4</th>
<th>VC5</th>
<th>VC6</th>
<th>VC7</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All ER</td>
<td>98</td>
<td>98</td>
<td>99</td>
<td>98</td>
<td>98</td>
<td>100</td>
<td>103</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>All Binary</td>
<td>17</td>
<td>22</td>
<td>139</td>
<td>22</td>
<td>25</td>
<td>361</td>
<td>357</td>
<td>46</td>
</tr>
<tr>
<td>3</td>
<td>All Binary, SW1 = ER</td>
<td>15</td>
<td>14</td>
<td>155</td>
<td>15</td>
<td>14</td>
<td>439</td>
<td>439</td>
<td>41</td>
</tr>
<tr>
<td>4</td>
<td>All Binary, SW3 = ER</td>
<td>16</td>
<td>15</td>
<td>155</td>
<td>31</td>
<td>29</td>
<td>402</td>
<td>405</td>
<td>45</td>
</tr>
<tr>
<td>5</td>
<td>All Binary, SW3,4 = ER</td>
<td>33</td>
<td>33</td>
<td>144</td>
<td>141</td>
<td>141</td>
<td>147</td>
<td>148</td>
<td>83</td>
</tr>
<tr>
<td>6</td>
<td>All Binary, SW1,3 = ER</td>
<td>29</td>
<td>29</td>
<td>146</td>
<td>30</td>
<td>28</td>
<td>377</td>
<td>380</td>
<td>48</td>
</tr>
</tbody>
</table>

4.2.2 Impact of PCR

In most studies dealing with congestion control for the ABR service, the PCR is usually set to the full link capacity. Setting PCR to the full link capacity aggravates any fairness problem since it allows VCs getting higher rates than their fairshares to acquire more bandwidth. Consequently, this means that those VCs getting lower rates than their fairshares will keep getting less and less bandwidth. Table 4.13 shows the simulation results when the PCR is set to 100 Mbps instead of 155.52 Mbps (i.e. the full link capacity). For multi-link VCs, RIF and RDF are set to 1/512 and 1/32 respectively. A value of 1/512 is used for both RIF and RDF for the single-link VCs. EPRCA is used as the ER switch algorithm.

When comparing the results shown in Table 4.13 with the results shown in Table 4.12, where the same setting is used for all parameters except for PCR which was set to 155.52 Mbps in Table 4.12, we notice a clear performance improvement due to the fact that any VC which took more than its fairshare cannot take more than 100Mbps of the link capacity. Therefore, more bandwidth is available for those VCs which took less than their fairshare in Table 4.12.

Setting the PCR to 100 Mbps is an arbitrary choice and may not be the optimal setting. It
should be expected that the performance will be enhanced more if we lower the PCR further. However, it must be pointed out that setting PCR to a very low value can cause underutilization of the bandwidth for some links. This is because some VCs cannot use the available bandwidth due to the limit imposed by the low PCR value, even when the other VCs cannot use their fairshares of the bandwidth due to other factors.

Table 4.13 Percentage throughput and fairness when PCR=100Mbps (with EPRCA)
(RIF=1/512 RDF=1/32)

<table>
<thead>
<tr>
<th>No.</th>
<th>Case</th>
<th>VC1</th>
<th>VC2</th>
<th>VC3</th>
<th>VC4</th>
<th>VC5</th>
<th>VC6</th>
<th>VC7</th>
<th>Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All ER</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>98</td>
<td>99</td>
<td>100</td>
<td>103</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>All Binary</td>
<td>50</td>
<td>47</td>
<td>105</td>
<td>50</td>
<td>47</td>
<td>301</td>
<td>302</td>
<td>57</td>
</tr>
<tr>
<td>3</td>
<td>All Binary, SW1 = ER</td>
<td>50</td>
<td>47</td>
<td>105</td>
<td>50</td>
<td>47</td>
<td>301</td>
<td>302</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>All Binary, SW3 = ER</td>
<td>50</td>
<td>48</td>
<td>104</td>
<td>51</td>
<td>48</td>
<td>289</td>
<td>298</td>
<td>59</td>
</tr>
<tr>
<td>5</td>
<td>All Binary, SW3,4 = ER</td>
<td>91</td>
<td>89</td>
<td>104</td>
<td>104</td>
<td>104</td>
<td>107</td>
<td>109</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>All Binary, SW1,3 = ER</td>
<td>50</td>
<td>48</td>
<td>104</td>
<td>51</td>
<td>48</td>
<td>289</td>
<td>298</td>
<td>59</td>
</tr>
</tbody>
</table>
Chapter 5 Interoperability Among ER Algorithms

Given their performance superiority, a fairly large number of explicit rate (ER) algorithms has been proposed in the literature (for examples see [4][14][26][34][44][47]). In general, the proposed ER algorithms can be classified into two main classes: approximate and exact. As the name implies, approximate algorithms estimate the bandwidth fairshares for different connections (VCs) using exponential running averages rather than computing exact fairshare values. As a result, approximate algorithms are relatively less responsive to network changes and can result in high oscillations in the generated bandwidth allocations. At the expense of higher implementation complexity, exact algorithms are more accurate in computing bandwidth fairshares and generate, in most cases, oscillation-free bandwidth allocations [1].

This chapter investigates the issue of interoperability among ER algorithms in heterogeneous networks. The issue of fair bandwidth allocation is considered in details. First, the dynamics of the ABR source rate controlled by an ER algorithm are analyzed. Based on the analysis, some of the problems that can cause unfairness in heterogeneous ER networks in both the transient and the steady-state cases are identified and analyzed. In order to show how these problems occur between some of the existing ER algorithms, we show some simulation results for a heterogeneous network topology experimenting with different ER algorithms.

5.1 ABR Source Rate Dynamics in Networks with ER Switches

One of the main advantages of exact ER algorithms is their ability to generate oscillation-free ABR traffic in steady-state network conditions. By contrast, approximate algorithms tend to generate oscillatory ABR traffic. How large are these oscillations depend highly on the algorithm itself and on the tuning of its parameters [1]. ABR traffic generated by a fair exact algorithm
matches with the expected fairshare value with almost no oscillations. For an oscillatory fair
approximate algorithm, although the generated traffic is oscillatory, it oscillates around the
fairshare value.

Consider the simple network topology depicted in Figure 5.1, where congestion happens
at the output link of switch 1. This network will be used to elucidate potential causes of unfairness.
There are two identical ABR VC's in the network of Figure 5.1. One connects the source node S1
to the destination node D1 and the other connects source S2 to destination D2. The function
$ACR(t)$ represents the ACR changes during a VC lifetime as perceived by the source in response
to the feedback messages carried by the RM cells. For the topology shown in Figure 5.1, $ACR(t)$ is
the same for both sources since they are identical and the end-system-to-switch delays for both
VC's are the same. Figure 5.2 shows three possible examples of the shape of the function $ACR(t)$
in the steady-state. While $ACR1(t)$ and $ACR2(t)$ are possibly generated by approximate ER
algorithms running at switch 1, $ACR3(t)$ is generated by an exact ER algorithm.

For the oscillatory algorithms (i.e. algorithms showing oscillations in $ACR(t)$), one can
divide the ACR cycle into two phases: an 'upper' phase during which $ACR(t) >$ fairshare and a
'lower' phase during which $ACR(t) <$ fairshare, where the fairshare value for a specific VC is the
fair rate allocation that is given to that VC according to a certain fairness criterion (in our case, the \textit{max-min fairness criterion} is used [23]). This applies only when dealing with $ACR(t)$ generated by fair algorithms where the ACR value always oscillates around the fairshare value. In Figure 5.2, the time period from $t_0$ to $t_2$ is considered one cycle for $ACR(t)$. As shown in the Figure, the amplitude of the oscillations during the 'upper' phase and during the 'lower' phase need not be equal. However, in order to achieve a fair operation, the average rate of a VC over its connection time should be equal to its fairshare.

If we define the throughput as the total number of bits received by the destination during the lifetime of a VC, then for a VC that has the ACR dynamics given by $ACR1(t)$ in Figure 5.2, the
throughput is calculated by:

\[ \text{Throughput} = \int_{t_0}^{t_f} ACR_1(t)dt \]  

(EQ 6)

where \( t_0 \) refers to the connection start-up time and \( t_f \) refers to the connection end time.

The average rate of a VC during a specific period of time can be computed by dividing the total throughput over the period duration. Therefore, if we assume that the VC which has the ACR dynamics given by \( ACR_1(t) \) gets its fairshare then the throughput as calculated by (EQ 6) divided by the connection duration is equal to the fairshare:

\[ \text{Fairshare} = \frac{\text{Throughput}}{t_f - t_0} \]  

(EQ 7)

Ideally, for (EQ 7) to apply, each ACR cycle should be fair by itself. In other words, the throughput during an ACR cycle divided by the cycle duration should be equal to the fairshare. By applying this to our example in Figure 5.2, we get the following equation:

\[ \text{Fairshare} = \frac{\int_{t_0}^{t_1} ACR_1(t)dt + \int_{t_1}^{t_2} ACR_1(t)dt}{t_2 - t_0} \]  

(EQ 8)

During the period from \( t_0 \) to \( t_1 \) the VC average rate (termed \( Rate_1 \)) is given by:

\[ Rate_1 = \frac{\int_{t_0}^{t_1} ACR_1(t)dt}{t_1 - t_0} \]  

(EQ 9)
And during the period from \( t_1 \) to \( t_2 \), the VC average rate (termed \( Rate_2 \)) is given by:

\[
Rate_2 = \frac{\int_{t_1}^{t_2} ACR_1(t)\,dt}{t_2-t_1}
\]  

(EQ 10)

Clearly, \( Rate_1 > fairshare > Rate_2 \) since \( ACR_1(t) > fairshare \) from \( t_0 \) to \( t_1 \) and \( ACR_1(t) < fairshare \) from \( t_1 \) to \( t_2 \). However, since (EQ 8) simply averages the rate of the VC over the entire cycle duration and since we are considering fair allocation, it should be expected that:

\[
Rate_1 - fairshare = fairshare - Rate_2
\]

(EQ 11)

or

\[
Gain = Loss
\]

(EQ 12)

(EQ 11) means that for a fair algorithm, the rate increase gained during the 'upper' phase will be nullified by the decrease in rate anticipated during the 'lower' phase for every ACR cycle. Although the above result was derived for an ideal case, fair allocation will not hold in practice unless (EQ 11) applies for a fairly large number of oscillations.

### 5.2 Unfairness problems

In heterogeneous networks a VC traverses several switches running different ER algorithms. The resulting \( ACR(t) \) function computed for one of the VCs depends on the values stored in the ER field and the CI bit of the RM cell. The final value stored in the ER field, which is the one that will be used to change the ACR at the source, is basically the minimum of all the values calculated by the ER switches along the VC's path. Therefore, one of the important
questions that remain to be answered is: If a VC traverses several switches and these switches are running fair ER algorithms, does this VC always get its correct fairshare? and if not, why?

In order to answer this question, we use the network topology shown in Figure 5.3. In this network, link 1 bandwidth is half that of link 2 and the PCR of the sources S1 and S2 is equal to the bandwidth of link 2 in order to force congestion in switches 1 and 2. According to the max-min fairness criterion, both VC1 (from S1 to D1) and VC2 (from S2 to D2) have the same fairshare which is equal to half the bandwidth of link 2. The source-to-switch delay for S2 is set to the sum of link 1 delay and the source-to-switch delay for S1 in order to remove the impact of delay (if any) on the performance of VC1. The destination-to-switch delay is the same for both VCs. Switch 1 and switch 2 are running two different oscillatory but fair ER algorithms. The algorithm running on switch 3 does not affect the performance of the network since there is no congestion at its output links.

The motive behind studying this topology is to analyze the effect of having source S1 change its ACR value according to the RM cells which are altered by the two ER algorithms running at switches 1 and 2, while having source S2 respond only to the feedback generated by switch 2. Practically, the two algorithms running at switches 1 and 2 are compatible if source S1 is responding only to the feedback coming from switch 2. This occurs if switch 1 always sets the ER field of the RM cell to values equal to or higher than those set by switch 2. As a result, the ACR of sources S1 and S2 is computed by the same algorithm.

Depending on the two algorithms running at switches 1 and 2 and on the conditions postulated in the previous section for fair operation, we anticipate several problems to arise in the network topology of Figure 5.3 (or any similar topology) where some VCs receive feedbacks
generated by multiple switches along a path while other VCs respond to feedbacks generated by one algorithm only. In the following, we identify three such problems (the rate increase, the rate decrease, and the presence of VBR traffic) and investigate each problem thoroughly.

5.2.1 The Rate Increase Problem

Certain algorithms allow the sources to increase their ACRs much faster than others. If switch 2 algorithm in Figure 5.3 allows a fast increase in the ACR value while switch 1 does not allow as fast an increase, then there is a high chance of having a fairness problem. This is because source S1 will use the final value recorded in the ER field of the RM cell which represents the minimum of the two values proposed by switches 1 and 2. The end result is that VC 2 will be granted more than its fairshare of bandwidth at the expense of VC 1 getting less than its fairshare. This unfairness problem is referred to as the rate increase problem.

Figure 5.4 illustrates a possible scenario that causes the rate increase problem. From time t 0 to t 1 both VCs 1 and 2 are increasing their ACRs at the same rate. After time t 1, the rate increase allowed by the algorithms running at switches 1 and 2 start to differ. During the period
from $t_1$ to $t_2$, switch 1 algorithm slows the rate of increase of the ACR for VC1 while switch 2 algorithm allows VC2 to increase its ACR much faster. Therefore, VC2 now has a bandwidth advantage over VC1.

After some time, Switch 2 detects congestion and asks VC2 to lower its rates since during the interval from $t_2$ to $t_3$ only VC2 is causing the congestion. At time $t_3$, switch 2 is still congested but now it will ask both VC1 and VC2 to lower their rates since they have the same rates. From now on, both VCs 1 and 2 will have the same $ACR(t)$ functions until the end of the cycle at time $t_7$.

Careful scrutiny of Figure 5.4 reveals that to attain its fairshare of bandwidth, VC1 should start increasing its rate at time $t_5$. To simplify analysis, it is assumed that the $ACR(t)$ function is
symmetric about the fairshare line. However, VC1 will not be able to achieve this performance because switch 2 is still congested at that time, and it will not allow VC1 to increase its ACR. In fact, \( ACR_1(t) \) will continue to match \( ACR_2(t) \) value. Since both VC1 and VC2 are lowering their rates, the congestion at switch 2 will be relieved earlier than it would in the case of only VC2 lowering its rate. At time \( t_6 \), both VCs will start increasing their rates since switch 2 is not detecting congestion any more. This will continue until the end of the cycle at time \( t_7 \).

It is obvious from the previous discussion that VC2 had an advantage over VC1 and (Eq 11) and (Eq 12) do not hold for both VCs. Formally, for VC1 we have:

\[
\int_{t_0}^{t_4} \frac{ACR_1(t)dt}{t_4 - t_0} - \text{Fairshare} < \int_{t_4}^{t_7} \frac{ACR_1(t)dt}{t_7 - t_4} \quad \text{(Eq 13)}
\]

and for VC2 we have:

\[
\int_{t_0}^{t_4} \frac{ACR_2(t)dt}{t_4 - t_0} - \text{Fairshare} > \int_{t_4}^{t_7} \frac{ACR_2(t)dt}{t_7 - t_4} \quad \text{(Eq 14)}
\]

Simply, (Eq 13) means that for VC1 the gain during the ‘upper’ phase is less than the loss during the ‘lower’ phase. On the other hand, (Eq 14) means that for VC2, the gain in the ‘upper’ phase is higher than the loss during the ‘lower’ phase. Therefore, VC1 average rate during this cycle is less than its fairshare and VC2 average rate is higher than its expected fairshare. Of course, for an ideal case where the output link of switch 2 is fully utilized, the decrease in the average rate anticipated by VC1 is equal to the average rate increase gained by VC2.
A related unfairness scenario can arise if switch 1 runs an exact ER algorithm that generates a non-oscillatory bandwidth allocation while switch 2 algorithm is running an approximate algorithm which generates an oscillatory bandwidth allocation. This behavior is considered a special case of the rate increase problem, where VC1 will not have a chance at all to increase its rate above the fairshare value. This is because switch 1 algorithm will always try to set VC1 rate equal to the fairshare value and at the same time switch 2 algorithm will generate oscillations that push VC1 rate below the fairshare value during $ACR_2(t)$ ‘lower’ phase. This scenario will be explained in more details in section 5.3.2.4

5.2.2 The Rate Decrease Problem

Another case which can lead to unfairness is when different ER algorithms ask ABR sources to decrease their rates at different rates. In Figure 5.3, if switch 2 algorithm asks VCs 1 and 2 to decrease their ACRs to a value that is higher than the value demanded by switch 1 for VC1, then VC1 will have its rate decreased more than VC2. This is due to the fact that VC1 will always follow the minimum of the ER values demanded by switches 1 and 2. This unfairness problem is called the rate decrease problem because it happens when there is a mismatch during the rate decrease process.

Figure 5.5 shows a possible scenario that leads to unfairness caused by the rate decrease problem. During the period from $t_0$ to $t_1$ both VCs increase their rates by the same amount. At time $t_1$, switch 1 and switch 2 detect congestion, and start asking the VCs to decrease their ACRs. Switch 1 asks VC1 to decrease its rate much faster than switch 2 does for VC2. After some time, when the congestion is relieved at switch 1, VC1 starts increasing its ACR at time $t_3$. This will continue until time $t_5$. But switch 2 is still congested and so it will demand both VCs to decrease
their rates. At time t6, switch 2 is no longer congested and it will allow both VCs to increase their rates. Since we assumed that the shape of the $ACR(t)$ function is symmetric about the fairshare line, VC2 should have lowered its rate to the expected value to keep its average rate equal to the fairshare. However, this does not happen because during the period from t5 to t6 VC1 is also lowering its rate in response to the feedback coming from switch 2. When inspecting Figure 5.5, we notice that VC1 was unfairly asked to lower its rate more than it should to keep its average rate equal to the fairshare. In contrast, VC2 had an advantage by not being asked to lower its rate to the expected value needed to achieve the fairshare.

![Figure 5.5 The rate decrease problem](image)

As in the case for the rate increase problem, (EQ 11) and (EQ 12) do not hold. Instead, for VC1 we have:
\[
\frac{\int_{t_0}^{t_2} ACR1(t)dt}{t_2 - t_0} - \text{Fairshare} < \frac{\int_{t_2}^{t_7} ACR1(t)dt}{t_7 - t_2}
\]  
(EQ 15)

and for VC2 we have:

\[
\frac{\int_{t_0}^{t_4} ACR2(t)dt}{t_4 - t_0} - \text{Fairshare} > \frac{\int_{t_4}^{t_7} ACR2(t)dt}{t_7 - t_4}
\]  
(EQ 16)

(EQ 15) means that for VC1, the gain during the ‘upper’ phase is less than the loss during the ‘lower’ phase. (EQ 16) means exactly the opposite for VC2, where the gain during the ‘upper’ phase is higher than the loss during the ‘lower’ phase.

5.2.3 Presence of Other Traffic Types

In real ATM networks, ABR traffic is not the only type of traffic that exists in the network. In the presence of other traffic types, especially those of bursty nature such as Variable Bit Rate (VBR), ABR traffic may experience very aggressive changes in the available bandwidth. Dealing quickly with such available bandwidth changes is one of the objectives of an efficient congestion control mechanism for ABR service [1].

Due to high changes in the available bandwidth when other traffic types exists along with ABR traffic, unfairness that arises during the available bandwidth increase or decrease periods can sometimes be very severe. For example, if the bandwidth available for ABR suddenly increases then those VCs controlled only by algorithms that allow quick ACR increases will utilize the extra bandwidth much faster than those VCs controlled by algorithms that allow a very slow increase in ACR. These are the same symptoms of the rate increase problem. Similarly, if the available
bandwidth is reduced quickly, then those VCs controlled by algorithms not capable of responding quickly to the available bandwidth reduction will have an advantage over other VCs controlled by a very responsive algorithms that reduce the ACR very fast and this is considered a form of the rate decrease problem.

The type of unfairness that occur when other traffic types exist shows up only during changes in the ABR available bandwidth (i.e. during the transient-state) and it is completely independent from the unfairness that can happen during the steady-state period which was discussed in sections 5.2.1 and 5.2.2 above. Therefore, even if two ER algorithms appear to be compatible in the steady-state (i.e. causing no unfairness), they may cause unfairness during the transient-state due to their different responses to the changes in the available bandwidth.

5.3 Simulation Results

This section demonstrates the concepts presented in previous sections by simulating four representative ER algorithms. Three algorithms are based on the well-known EPRCA [44][47] algorithm and the fourth is ERICA [26]. For more details about these algorithms, see Chapter 2. The idea behind choosing these algorithms is to illustrate how the rate increase and the rate decrease problems can lead to unfairness. Although several other ER algorithms have been proposed in the literature [4][14][34], the representative subset of algorithms used in this chapter are adequate for providing significant insight into the compatibility and interoperability among different algorithms. As for the simulations presented in Chapter 4, we use OPNET™ as our simulation tool.
5.3.1 Simulation Model

Our simulations are divided into two sets. The first set is dedicated to the steady-state case where ABR is the only traffic category existing in the network. The second set adds a bursty VBR source to simulate heterogeneous networks with mixed traffic types (ABR and VBR in our case), where the main intent is to evaluate the performance in the transient-state. Aside from the network topology, almost all source and switch algorithms are the same for the two simulation sets.

5.3.1.1 Network Topology

We will use the network topology shown in Figure 5.3 for our first simulation set. Using this simple network rather than a large one allows for better understanding of the unfairness problems. All the links have the same length of 0.1 km except the link between source S2 and switch2 which has a length of 0.2 km in order to guarantee that the delay between each source and switch2 is the same. Link delay is assumed to be 5 µsec/km. It is assumed that all links have a rate of 155.52 Mbps except link1 which has a rate of 77.76 (i.e. 155.52 / 2) Mbps in order to allow congestion to occur in switch1. Moreover, setting link1 rate to this value makes both switches 1 and 2 critical switches for VC1, where a critical switch is the switch that gives the least amount of bandwidth to a specific VC according to the max-min fairness criterion. Therefore, both switches will affect the performance of VC1. It is worth mentioning that even if only one of the two switches is critical for VC1, there is still a possibility of having unfairness problem. For more details about this scenario see [40].

In the second simulation set, we modify the network topology presented in Figure 5.3 to include a "deterministic" VBR source that traverses switches 1 and 2. The new network topology used in the second simulation set is shown in Figure 5.6. The presence of the VBR source will
change the amount of bandwidth available for ABR service traffic at different times. The VBR source is called "deterministic" because it switches from the 'active' state to the 'inactive' state and vice versa periodically in a deterministic fashion. Our VBR source is a simple ON-OFF source whose rate is equal to the PCR during the ON period and is equal to zero during the OFF period. In the simulation, the VBR source switches from one state to the another every 50 milliseconds. During the 'active' period it sends traffic with a rate of 80 Mbps (i.e. PCR = 80 Mbps). Setting the PCR for the VBR source to 80 Mbps may not represent the worst case where changes in the available bandwidth may be more aggressive. We use this moderate value to show that the rate increase and the rate decrease problems can truly affect the performance of heterogeneous networks in the transient-state and subsequently can lead to unfairness.

In addition to the VBR source, the second simulation set assumes that link1 bandwidth is equal 125 Mbps to allow switch2 to be the critical switch during the steady-state period. Thus VC1 and VC2 are both controlled mainly by switch2 algorithm. This configuration enables evaluating the impact of quick changes in the available bandwidth on the performance of ABR traffic without having any unfairness persisting during the steady-state. Based on these changes, the fairshares of the ABR VCs during the 'active' period are different from their fairshares during the 'inactive' period of the VBR source. While the VBR source is in the 'active' state, the available bandwidth on link1 becomes 125 - 80 = 45 Mbps and the available bandwidth on link2 becomes 155.52 - 80 = 75.52 Mbps. However, since link2 is shared by VCs 1 and 2, then the fairshare according to the max-min fairness criterion becomes $75.52 / 2 = 37.76$ Mbps for each VC. On the other hand, during 'inactive' period of the VBR source, links 1 and 2 available bandwidth is 125 Mbps and 155.52 Mbps, respectively, and according to the max-min fairness criterion the fairshare for each VC is $155.52 / 2 = 77.76$ Mbps.
5.3.1.2 Switch Algorithms

As mentioned earlier, we consider a number of different ER algorithms operating in different switches but within the same network. The first algorithm is the standard Enhanced Proportional Rate Control Algorithm (EPRCA) proposed in [44] and [47]. EPRCA is an approximate (an hence oscillatory) ER algorithm. The second algorithm, which will be called EPRCA-Binary, sets the CI bit in the backward RM cells to inform the sources of the presence of congestion instead of using the ER field as in the case of the standard EPRCA. Other than using the CI bit, EPRCA and EPRCA-Binary are the same. Although, EPRCA-Binary operates by setting the CI bit only and does not use the ER field of the RM cell, it is still different from the binary-feedback algorithm, because it utilizes the concept of selective marking when setting the CI bit. EPRCA-Binary was suggested in [44] as a binary-mode EPRCA.

The third algorithm which is also based on EPRCA, differs from EPRCA in the way it controls the value by which a VC increases its rate. In EPRCA, if there is no congestion, then the sources are allowed to increase their rates depending only on the RIF parameter. In this new version of EPRCA, even if there is no congestion, the algorithm will mark the ER field in the RM
cell by a value equal to $\alpha \cdot MACR$, where MACR is the fairshare running average maintained by the switch for every output port [44][47]. By doing so, the increase of the ACR of the sources will be bounded by the MACR value which usually increases slowly resulting in a slow increase in the ACR value. This version of the EPRCA will be called EPRCA-Bounded and was proposed in [40]. The fourth ER algorithm that we consider is ERICA and it was developed by Jain et. al. [26]. ERICA is an exact ER algorithm.

Table 5.1 shows the parameter settings used for EPRCA, EPRCA-Binary, and EPRCA-Bounded. For the case of EPRCA-Bounded, the parameter $\alpha$ is set to 1.015. ERICA’s averaging interval length and target utilization are set to 0.5msec and 99.5% respectively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQ</td>
<td>250</td>
</tr>
<tr>
<td>DQT</td>
<td>5000</td>
</tr>
<tr>
<td>VCS</td>
<td>7/8</td>
</tr>
<tr>
<td>MRF</td>
<td>1/4</td>
</tr>
<tr>
<td>DPF</td>
<td>0.82</td>
</tr>
<tr>
<td>ERF</td>
<td>0.65</td>
</tr>
<tr>
<td>AV</td>
<td>1/16</td>
</tr>
</tbody>
</table>

In the second simulation set, we use the same algorithms with the same parameter settings. The only exception is that for EPRCA-Bounded algorithm, the parameter $\alpha$ is set to 1.5 instead of 1.015 to allow for higher increase rates since the presence of VBR will result in higher changes in the available bandwidth.
5.3.1.3 Source Parameters

All ABR traffic sources used in our simulations are greedy (i.e. persistent). Table 5.2 shows the source parameter settings used in our simulations. This setting is used for the two simulation sets.

Table 5.2 Source parameters setting

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>155.52 Mbps</td>
</tr>
<tr>
<td>MCR</td>
<td>1 Mbps</td>
</tr>
<tr>
<td>ICR</td>
<td>77.7 Mbps</td>
</tr>
<tr>
<td>Nrm</td>
<td>32</td>
</tr>
<tr>
<td>RIF</td>
<td>1/128</td>
</tr>
<tr>
<td>RDF</td>
<td>1/256</td>
</tr>
</tbody>
</table>

5.3.2 Results

5.3.2.4 Simulation Set 1 (ABR Traffic Only)

Before presenting our simulation results for the first set, which represents the steady-state case, we compare the performance of the ER algorithms we choose for our simulations. Figure 5.7 shows the $ACR(t)$ functions generated by each algorithm for the simple network topology shown in Figure 5.3, where all link rates are 155.52 Mbps and the parameters setting summarized in Table 5.1 and Table 5.2 is used. In Figure 5.7, only the $ACR(t)$ function for one of the two VCs is shown as the other ACR function is exactly the same.

It is clear from Figure 5.7 that the approximate algorithms, namely EPRCA, EPRCA-Binary, and EPRCA-Bounded, are oscillating around the fairshare line which exactly matches with $ACR(t)$ function generated by ERICA. Figure 5.7 also shows that the relations expressed by (EQ 11) and (EQ 12) are highly applicable in the case of real algorithms. The differences between
the three oscillatory algorithms stem from their different increase and decrease rates. Looking at the increase rates, we notice that EPRCA and EPRCA-Binary have the same increase rates which is higher than the increase rate generated by EPRCA-Bounded. Specifically, EPRCA-Bounded has the same initial increase rate as EPRCA and EPRCA-Binary, but then it exhibits a slower increase rate as the rate becomes bounded by the fairshare running average (MACR). With respect to the decrease rates, we notice that EPRCA-Binary has a much slower decrease rate than EPRCA and EPRCA-Bounded.

Twelve different simulations were performed covering all the possible combinations in which switch1 and switch2 in the network of Figure 5.3 run two different algorithms. To assess how these different algorithms affect the fairness of rate-allocation, the average throughput of
VC2 is compared to that of VC1. The difference between the two throughput values is calculated as a percentage of VC1 throughput. If there is no unfairness, then the resulting value will be 0% which is the ideal case. The average throughput is computed as the average rate of bits received at the destination during the simulation interval which is 1 second. It should be noted here that in this chapter, we have chosen not to use the fairness index [25], used in Chapter 4, as a basis for our fairness comparison. The reason behind this choice is that we have only two VCs in the network topology we consider in this chapter and a simple fairness measure is adequate.

Table 5.3 summarizes simulation results for all the twelve different cases. As a general observation, there is no unfairness in the following cases:

1. All cases where the algorithm of switch2 has a slower increase rate compared to switch1 algorithm (cases: 2 and 5).

2. All cases where switch2 algorithm is faster than switch1 algorithm in its decrease rate given that there is no unfairness during the rate increase period (case 4).

3. All cases where the algorithm of switch2 is non-oscillatory (cases: 3, 6, and 9).

<table>
<thead>
<tr>
<th>No.</th>
<th>Switch 1 Algorithm</th>
<th>Switch 2 Algorithm</th>
<th>VC1’s Throughput (Mbps)</th>
<th>VC2’s Throughput (Mbps)</th>
<th>Unfairness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EPRCA</td>
<td>EPRCA-Binary</td>
<td>73.5</td>
<td>81.2</td>
<td>10.5</td>
</tr>
<tr>
<td>2</td>
<td>EPRCA</td>
<td>EPRCA-Bounded</td>
<td>77.3</td>
<td>77.4</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>EPRCA</td>
<td>ERICA</td>
<td>77.2</td>
<td>77.5</td>
<td>0.4</td>
</tr>
<tr>
<td>4</td>
<td>EPRCA-Binary</td>
<td>EPRCA</td>
<td>77.4</td>
<td>77.4</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>EPRCA-Binary</td>
<td>EPRCA-Bounded</td>
<td>77.3</td>
<td>77.4</td>
<td>0.4</td>
</tr>
<tr>
<td>6</td>
<td>EPRCA-Binary</td>
<td>ERICA</td>
<td>77.2</td>
<td>77.6</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Table 5.3 Simulation results for heterogeneous network topology

<table>
<thead>
<tr>
<th>No.</th>
<th>Switch 1 Algorithm</th>
<th>Switch 2 Algorithm</th>
<th>VC1’s Throughput (Mbps)</th>
<th>VC2’s Throughput (Mbps)</th>
<th>Unfairness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>EPRCA-Bounded</td>
<td>EPRCA</td>
<td>74.6</td>
<td>80.1</td>
<td>7.4</td>
</tr>
<tr>
<td>8</td>
<td>EPRCA-Bounded</td>
<td>EPRCA-Binary</td>
<td>70.5</td>
<td>84.3</td>
<td>19.6</td>
</tr>
<tr>
<td>9</td>
<td>EPRCA-Bounded</td>
<td>ERICA</td>
<td>77.1</td>
<td>77.6</td>
<td>0.6</td>
</tr>
<tr>
<td>10</td>
<td>ERICA</td>
<td>EPRCA</td>
<td>73.2</td>
<td>81.5</td>
<td>11.3</td>
</tr>
<tr>
<td>11</td>
<td>ERICA</td>
<td>EPRCA-Binary</td>
<td>73.7</td>
<td>81.1</td>
<td>10.0</td>
</tr>
<tr>
<td>12</td>
<td>ERICA</td>
<td>EPRCA-Bounded</td>
<td>73.4</td>
<td>81.2</td>
<td>10.6</td>
</tr>
</tbody>
</table>

For all other cases, switch 2 gives an advantage to VC2 and hence results in unfairness. These cases can be grouped into three classes:

1. All cases where the algorithm of switch 2 has a higher increase rate than that of switch 1 (cases: 7, 10, 11, 12).

2. All cases where the algorithm of switch 2 has a lower decrease rate than that of switch 1 (case 1).

3. All cases where the algorithm of switch 2 has a higher increase rate than that of switch 1 and in the same time switch 2 algorithm has a lower decrease rate than that of switch 1 (case 8).

As can be seen from Table 5.3, the highest unfairness among the cases we considered is in case 8 where the rate increase problem and the rate decrease problem are both existing at the same time.
Figures 5.8 through 5.10 show the ACR dynamics for VC1 and VC2 in three different representative cases. In Figure 5.8, switch1 employs ERICA and switch2 employs EPRCA. The resulting unfairness is caused by the rate increase problem where VC1 ACR does not exceed the fairshare value of 77.7 Mbps computed by ERICA, whereas VC2 ACR is not limited by ERICA at all. Figure 5.9 shows a similar situation that results when switch1 algorithm is replaced by the EPRCA-Bounded algorithm which has a very slow increase rate compared to that of EPRCA algorithm running in switch2. Although EPRCA-Bounded allows VC1 to increase its rate little above the fairshare value, it still does not match the rate increase allowed for VC2 by the EPRCA algorithm running in switch2.

Figure 5.10 shows the case when switch1 algorithm is EPRCA and switch2 algorithm is
EPRCA-Binary. As can be seen from Figure 5.10, the problem happens during the rate decrease period because VC1 decreases its rate faster than VC2. What aggravates the problem here is that VC1 does not get the chance to increase its rate to the rate of VC2 after it has been forced to decrease its rate faster than VC2 due to the nature of the EPRCA-Binary algorithm. For example, at time $t = 0.12$ seconds, VC1 tries to increase its rate to attain the same rate as VC2, but this does not happen because the EPRCA-Binary algorithm running in switch2 will force VC1 to decrease its rate when it comes close to the running average (MACR) before it can reach VC2 rate. This shows how the rate decrease and increase problems can be aggravated, in some cases, by other factors caused by the nature of the running algorithm. Such problems are more likely to occur in the cases of approximate ER algorithms since they depend heavily on running averages and this
can sometimes lead to inaccurate estimates of the fairshares.

Figure 5.10  ACR dynamics in a heterogeneous network (switch1 algorithm = EPRCA, switch2 algorithm = EPRCA-Binary)

5.3.2.5 Simulation Set 2 (Presence of VBR Traffic)

Table 5.4 below shows the simulation results for the twelve cases considered previously, but with the inclusion of a VBR source. The cases that show unfairness can be grouped into three groups:

1. When the algorithm of switch1 responds much faster than that of switch2 to the reduction of the available bandwidth caused by having the VBR source change from an ‘inactive’ state to an ‘active’ state. More specifically, switch1 algorithm tries to reduce the rate of VC1 to 125 - 80 = 45 Mbps while switch2 algorithm tries to reduce the rates of VC1 and VC2 to (155.52 -
80) / 2 = 37.76 Mbps each. Due to the fact that switch1 algorithm responds much faster to the bandwidth reduction, it will reduce VC1 rate to 45 Mbps, but VC2 rate will be controlled by switch2 algorithm which has a slower reduction rate. Eventually, the two VCs will respond only to the algorithm of switch2 and both will have their rates reduced to 37.76Mbps. Cases 1 and 11 illustrate this behavior.

2. When the algorithm deployed in switch1 responds much faster than that of switch2 to the increase in the available bandwidth caused by having the VBR source changing from an ‘active’ state to an ‘inactive’ state. More Specifically, switch2 algorithm tries to increase the rates of VC1 and VC2 to $\frac{155.52}{2} = 77.76$ Mbps. However, because switch1 algorithm does not allow the same increase rate allowed by switch2, the increase rate for VC1 will be limited and hence VC2 gains a bandwidth advantage over VC1. Cases 7 and 9 are good examples of this behavior.

3. When the rate increase and decrease problems outlined in 1 and 2 above occur simultaneously between the two algorithms at switches 1 and 2, then a severe unfairness is noticed. Case 8 shows how severe this combined effect can be.

Table 5.4 Simulation results for heterogeneous network topology with VBR source

<table>
<thead>
<tr>
<th>No.</th>
<th>Switch 1 Algorithm</th>
<th>Switch 2 Algorithm</th>
<th>VC1's Throughput (Mbps)</th>
<th>VC2's Throughput (Mbps)</th>
<th>Unfairness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EPRCA</td>
<td>EPRCA-Binary</td>
<td>53.5</td>
<td>61.1</td>
<td>14.2</td>
</tr>
<tr>
<td>2</td>
<td>EPRCA</td>
<td>EPRCA-Bounded</td>
<td>55.5</td>
<td>55.6</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>EPRCA</td>
<td>ERICA</td>
<td>57.5</td>
<td>57.6</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>EPRCA-Binary</td>
<td>EPRCA</td>
<td>56.4</td>
<td>57.1</td>
<td>1.2</td>
</tr>
<tr>
<td>5</td>
<td>EPRCA-Binary</td>
<td>EPRCA-Bounded</td>
<td>54.4</td>
<td>54.6</td>
<td>0.4</td>
</tr>
<tr>
<td>6</td>
<td>EPRCA-Binary</td>
<td>ERICA</td>
<td>57.5</td>
<td>57.7</td>
<td>0.3</td>
</tr>
<tr>
<td>7</td>
<td>EPRCA-Bounded</td>
<td>EPRCA</td>
<td>47.8</td>
<td>64.2</td>
<td>34.3</td>
</tr>
<tr>
<td>8</td>
<td>EPRCA-Bounded</td>
<td>EPRCA-Binary</td>
<td>39.6</td>
<td>69.3</td>
<td>75.0</td>
</tr>
</tbody>
</table>
Table 5.4  Simulation results for heterogeneous network topology with VBR source

<table>
<thead>
<tr>
<th>No.</th>
<th>Switch 1 Algorithm</th>
<th>Switch 2 Algorithm</th>
<th>VC1's Throughput (Mbps)</th>
<th>VC2's Throughput (Mbps)</th>
<th>Unfairness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>EPRCA-Bounded</td>
<td>ERICA</td>
<td>51.3</td>
<td>62.3</td>
<td>21.4</td>
</tr>
<tr>
<td>10</td>
<td>ERICA</td>
<td>EPRCA</td>
<td>56.9</td>
<td>57.1</td>
<td>0.4</td>
</tr>
<tr>
<td>11</td>
<td>ERICA</td>
<td>EPRCA-Binary</td>
<td>49.2</td>
<td>62.6</td>
<td>27.2</td>
</tr>
<tr>
<td>12</td>
<td>ERICA</td>
<td>EPRCA-Bounded</td>
<td>54.4</td>
<td>54.9</td>
<td>0.9</td>
</tr>
</tbody>
</table>

The results summarized in Table 5.4 do not represent the worst case unfairness, but by comparing Table 5.3 and Table 5.4, one observes that two ER algorithms which are compatible (i.e. they do not cause unfairness) in the steady-state, can become severely incompatible in the transient-state (when large changes in the available bandwidth are expected). For example, compare case 9 in Tables 5.3 and 5.4.

The unfairness exhibited by some algorithms in Table 5.4 is due mainly to the substantial changes in the available bandwidth, caused by the activity of the VBR source. This type of unfairness is quite different from the unfairness that results in the steady-state when there are no changes in the available bandwidth. In other words, this unfairness is due to the mismatch among the different algorithms in response to transients in the available bandwidth.

To further elucidate the effect of VBR on the fairness of ER algorithms, we will consider in detail cases 7 and 11 of Table 5.4. Figure 5.11 shows the ACR dynamics for case 7 where the algorithm of switch1 is EPRCA-Bounded and switch2 algorithm is EPRCA. When the VBR source switches from an ‘active’ state to an ‘inactive’ state (at times = 0.05, 0.15, 0.25, 0.35, 0.45, and 0.55 seconds in Figure 5.11), the EPRCA-Bounded algorithm in switch1 responds very slowly to the increase in the available bandwidth compared to EPRCA at switch2, which results in a large difference in the throughput of the two VCs.
Chapter 5 Interoperability Among ER Algorithms

Figure 5.11 ACR dynamics in a heterogeneous network with VBR source present (switch1 algorithm = EPRCA-Bounded, switch2 algorithm = EPRCA)

Figure 5.12 below shows another way by which the unfairness could happen in a heterogeneous network (case 11 of Table 5.4). Here, switch1 algorithm is ERICA and switch2 algorithm is EPRCA-Binary. At times when the VBR source switches from an ‘inactive’ state to an ‘active’ state (at times = 0.0, 0.1, 0.2, 0.3, 0.4, and 0.5 seconds), the ERICA algorithm forces VC1 to drop its rate almost immediately in response to the reduction of the available bandwidth, whereas the EPRCA-Binary algorithm reduces VC2 rate very slowly. It should be emphasized that this mismatch in the rate of reduction is the major cause of the unfairness. The EPRCA-Binary algorithm, forces VC1 to reduce its rate even after the large reduction experienced by VC1 due to ERICA’s feedback. This reduction occurs because VC1 rate is very close to the running average (MACR) maintained by EPRCA-Binary. Therefore, VC1 is asked to reduce its rate according to
Figure 5.12  ACR dynamics in a heterogeneous network with VBR source present (switch1 algorithm = ERICA, switch2 algorithm = EPRCA-Binary)
Chapter 6  Conclusions and Future Work

6.1 Conclusions

The performance of ABR service in heterogeneous ATM networks was investigated. The various congestion control mechanisms used for ABR service were reviewed. The issue of interoperability between different rate-based congestion control algorithms running at different switches in the network was defined. The two types of interoperability: 1) interoperability between binary feedback and explicit rate algorithms and 2) interoperability among explicit rate algorithms were analyzed and simulated.

Our simulation results, presented in chapter 4, show that when one or more of the binary feedback switches (i.e. switches running the binary feedback algorithm) is replaced by an explicit rate switch (a switch running an explicit rate algorithm), the performance of ABR service varies widely depending on the location of each switch type and on the ABR source parameters setting. Since the multi-link VCs (i.e. VCs traversing several switches) are vulnerable to being beaten-down when binary feedback switches are used, giving these VCs higher Rate Increase Factor (RIF) and lower Rate Decrease Factor (RDF) enhances their performance. Moreover, setting the Peak Cell Rate (PCR) for all VCs to a value less than the total link bandwidth can noticeably improve the fairness of bandwidth allocation because it prevents misbehaving VCs from acquiring a large portion of the link bandwidth at the expense of other VCs.

If certain optimal parameter setting is not possible and if explicit rate switches cannot be used everywhere in the network (e.g. due to their high cost), our results show then that the best location for the binary feedback switches is at network edges. Edge switches are much less critical to VCs than backbone switches. Replacing backbone switches, rather than edge switches, by
binary feedback ones may degrade the performance of a large number of VCs and hence will have a higher impact on the network performance. Moreover, as shown by the simulations, using binary feedback switches at the network edges does not impact the performance of the backbone switches.

Another advantage of having binary feedback switches at network edges is that usually the last switch in a VC's path suffers little congestion (if any) and for bidirectional VCs this last switch in one direction is the first switch in the other direction. Therefore, it would be more economical to have this switch as a binary feedback switch since it will deal with congestion in one direction only most of the time. Having this switch as an explicit rate switch is considered wasteful given its high implementation complexity.

Chapter 5 presented an in-depth study of interoperability among various explicit rate algorithms in ATM networks. It also presented simulation results demonstrating various types of unfairness problems that can arise when several explicit rate (ER) algorithms run at different switches in the network. We have identified two distinct types of unfairness problems, one that persists during the steady-state allocation over a link, and the other arises during transient-state allocation (i.e. changing bandwidth) over a link.

In the steady-state, unfairness can appear in heterogeneous networks while ABR sources are increasing their rates (the rate increase problem) or while they are decreasing their rates (the rate decrease problem). As our analysis and simulations show, unfairness is mainly caused by the different shapes of oscillations in bandwidth allocations generated by different approximate ER algorithms. Since exact ER algorithms generate oscillation-free bandwidth allocations, it should be expected that heterogeneous networks employing only exact algorithms will not show any
unfairness problems in the steady-state.

In the transient-state case, unfairness is due to the different levels of responsiveness to the high changes in the available bandwidth demonstrated by different ER algorithms. Generally speaking, exact algorithms are more responsive than approximate algorithms and hence we recommend using exact algorithms in heterogeneous networks to reduce potential unfairness in bandwidth allocations when other traffic categories coexist with ABR traffic.

6.2 Future Work

Transmission Control Protocol (TCP) is the *de facto* standard for data transfers in Wide Area Networks (WANs). TCP implements its own congestion control schemes. Since the ABR service category was mainly designed to carry data traffic, it is interesting to see how TCP, with its congestion control schemes, performs over ABR service in heterogeneous ATM networks. More specifically, finding optimal parameter settings for TCP congestion control schemes when working over heterogeneous networks is of great importance.

Most of the rate-based algorithms suggested in the literature are based on the *max-min fairness criterion*. It is widely known that this fairness criterion is not applicable in the cases where the Minimum Cell Rate (MCR) parameter of the ABR source is set to a value greater than zero [6]. Therefore, it would be important to investigate how the rate-based algorithms, which are based on other fairness criteria, behave in heterogeneous networks when the MCR parameter is set to a non-zero value.

A third issue that is worth investigation is the impact of long link delays on the performance of heterogeneous networks. One practical question that needs to be answered is how the
fairness of rate-based algorithms is affected if some parts of the network are connected by satellite links that have delays of approximately 250ms.
Bibliography


Appendix A. Fairness Index

Given any fairness criterion (e.g. max-min fairness criterion), the fair bandwidth allocation for all VCs in the network can be determined. This is called the optimal allocation. If a rate-based congestion control algorithm (binary feedback or explicit rate) gives a bandwidth allocation that is different from the optimal one, then the unfairness can be numerically quantified using what is called the fairness index as follows [Jain].

For an n-VC network, assume that the congestion control algorithm allocates the bandwidth as follows \( \{ \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n \} \) instead of the optimal allocation \( \{ x_1, x_2, \ldots, x_n \} \). Then, we can calculate the normalized allocations \( x_i = \hat{x}_i / \hat{x}_i \) for each VC where \( i = 1 \ldots n \). The fairness index can be computed as follows:

\[
\text{fairnessIndex} = \frac{\left( \sum_{i} x_i \right)^2}{n \sum_{i} x_i^2}
\]

The fairness index gives a value between 0 and 100% where 100% is ideal. For example, if for a specific network topology that consists of two VCs, the optimal bandwidth allocation is \{50 Mbps, 50 Mbps\} and a certain rate-based congestion control algorithm gives a bandwidth allocation of \{20 Mbps, 80 Mbps\}, then the fairness of this algorithm given by the fairness index can be computed by the above equation and it gives a value of 73.5%. 
## Glossary

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAL</td>
<td>ATM Adaptation Layer</td>
</tr>
<tr>
<td>ABR</td>
<td>Available Bit Rate</td>
</tr>
<tr>
<td>ATM</td>
<td>Asynchronous Transfer Mode</td>
</tr>
<tr>
<td>CAC</td>
<td>Connection Admission Control</td>
</tr>
<tr>
<td>CBR</td>
<td>Constant Bit Rate</td>
</tr>
<tr>
<td>CCR</td>
<td>Current Cell Rate</td>
</tr>
<tr>
<td>CI</td>
<td>Congestion Indication</td>
</tr>
<tr>
<td>CLP</td>
<td>Cell Loss Priority</td>
</tr>
<tr>
<td>CS</td>
<td>Convergence Sublayer</td>
</tr>
<tr>
<td>EFCI</td>
<td>Explicit Forward Congestion Indication</td>
</tr>
<tr>
<td>EPRCA</td>
<td>Enhanced Proportional Rate Control Algorithm</td>
</tr>
<tr>
<td>ER</td>
<td>Explicit Rate</td>
</tr>
<tr>
<td>ERICA</td>
<td>Explicit Rate Indication for Congestion Avoidance</td>
</tr>
<tr>
<td>GCRA</td>
<td>Generic Cell Rate Algorithm</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>ITU-T</td>
<td>International Telecommunication Union-Telecommunication Sector</td>
</tr>
<tr>
<td>LAN</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>MACR</td>
<td>Mean Allowed Cell Rate</td>
</tr>
<tr>
<td>MCR</td>
<td>Minimum Cell Rate</td>
</tr>
<tr>
<td>NI</td>
<td>No Increase</td>
</tr>
<tr>
<td>NIST</td>
<td>US National Institute of Standards and Technology</td>
</tr>
<tr>
<td>nrt-VBR</td>
<td>Non-Real-Time Variable Bit Rate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>PCR</td>
<td>Peak Cell Rate</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RDF</td>
<td>Rate Decrease Factor</td>
</tr>
<tr>
<td>RIF</td>
<td>Rate Increase Factor</td>
</tr>
<tr>
<td>RM</td>
<td>Resource Management</td>
</tr>
<tr>
<td>rt-VBR</td>
<td>Real-Time Variable Bit Rate</td>
</tr>
<tr>
<td>SAR</td>
<td>Segmentation And Reassembly</td>
</tr>
<tr>
<td>TCP</td>
<td>Transmission Control Protocol</td>
</tr>
<tr>
<td>UBR</td>
<td>Unspecified Bit Rate</td>
</tr>
<tr>
<td>UPC</td>
<td>Usage Parameter Control</td>
</tr>
<tr>
<td>VC</td>
<td>Virtual Circuit</td>
</tr>
<tr>
<td>WAN</td>
<td>Wide Area Network</td>
</tr>
</tbody>
</table>