Flow Control for the Available Bit Rate Service in Asynchronous Transfer Mode Networks

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
genehmigte Dissertation

vorgelegt von
Diplom-Informatiker
Christian Virgil Cseh
aus Wiesbaden-Sonnenberg

Berichter: Universitätsprofessor Dr. rer. nat. Otto Spaniol
Universitätsprofessor Dr. rer. nat. Peter Martini

Tag der mündlichen Prüfung: 31. Oktober 2001

Acknowledgments

This dissertation was written during my time as a researcher with the Chair of Computer Science 4 at the University of Technology Aachen, Germany.

I would like to thank my advisor, Professor Otto Spaniol, for giving me the opportunity and support to choose and carry out my research work. I thank Professor Peter Martini for acting as co-advisor for this thesis and for providing valuable and helpful comments. I would also like to express my gratitude for the generous support of the Deutsche Forschungsgemeinschaft (DFG), during my research in the framework of the "Graduiertenkolleg Informatik und Technik".

I am very grateful that during my years at the Chair of Computer Science 4, I had the opportunity to participate in a number of scientific research projects, meet many interesting people, and work together with my colleagues in an inspiring and pleasant environment. Especially, I would like to thank my colleague Dirk Trossen for his support and personal company during many years.

It is a pleasure for me to acknowledge the important contributions of the diploma students Jens Baalmann and Yu Jiao to my work and this thesis.

Furthermore, I am very grateful for the helpful comments and corrections of Gundula Folkerts, Wolfgang Krause, and Dirk Trossen, who spent much time on carefully proof-reading this thesis. Special thanks go to Andreas Ropers for numerous interactions and his valuable comments and remarks.

I would like to express my very personal and heartfelt thanks to my partner in life, Gundula Folkerts, for her great patience and understanding as well as for her invaluable encouragement and love.

Table of Contents

CHAPTER 1

Introduction

In recent years, the convergence of classical telephone and data networks towards a single integrated multi-service network has become more and more important. New multimedia applications requiring real-time video and voice transmission generated the demand for *Broadband Integrated Service Digital Networks* (B-ISDN). The *Asynchronous Transfer Mode* (ATM) was developed as a new transmission technique for B-ISDN [I.311]. ATM is a key enabling technology for integrated networks, combining different traffic streams, like video, voice, and data, and provides end-to-end *Quality of Service* (QoS) guarantees for individual connections. ATM is a connection-oriented multiplexing and switching technique, where data is sent in packets of fixed size, called cells. In order to support different applications, the ATM Forum (an industry driven standardization group) defined several *service categories* for high and low priority traffic in ATM networks [TM4.1].

The service category *Available Bit Rate* (ABR) belongs to the low priority class and uses link bandwidth not used by high priority traffic. For this, a flow control mechanism is deployed, in order to adapt the current transmission rate of ABR sources to the varying bandwidth left unused by high priority traffic. The ABR sources gauge the current network condition and collect congestion information of the network nodes (switches) on the path. As a result, an end-to-end feedback control loop is established, providing information on the current network condition and the available bandwidth. The ABR service provides a certain minimum transmission rate per connection and guarantees a low cell loss rate.

There exist two other low priority service categories, namely, *Unspecified Bit Rate* (UBR) and *Guaranteed Frame Rate* (GFR). UBR provides a best effort service with an optional minimum transmission rate, that needs not to be supported by the network. In case of congestion situations, excess UBR traffic is simply discarded by ATM switches, which may have a deteriorating effect on the throughput of higher layer protocols. The GFR service category was developed to carry frame-based data traffic, with a main focus on the *Internet Protocol* (IP). As a consequence, the GFR service category requires a traffic descriptor specifying a maximum

frame length and a maximum burst size. Excess traffic that can not be transmitted due to congestion may be discarded by ATM switches on a per-frame basis.

The main advantage of the ABR service in contrast to UBR and GFR is the guarantee of a minimum cell rate together with a low cell loss rate without the need for a predefined traffic specification. Therefore, the ABR traffic category is best suited for aggregated network traffic or bursty applications that are tolerant to transmission delay. On the one hand, the network operator gets the possibility to use otherwise wasted link bandwidth for ABR data transport. On the other hand, the user may obtain cheap transmission capacities with a guaranteed minimum transmission rate. This makes ABR an attractive service category and an important feature of ATM networks.

1.1 Objectives and Goals

The end-to-end flow control loop for ABR traffic may lead to long feedback delays, i.e., the time delay until the rate control action is experienced at the switch, especially for wide-area connections. In order to alleviate this problem, the option of a *virtual source / virtual destination* (VS/VD) switch exists. The VS/VD switch splits the otherwise end-to-end feedback control loop into two separate segments, acting as an ABR destination for the upstream loop and as an ABR source for the downstream loop. The VS/VD switch offers the advantage of shortening the feedback delay and improving the responsiveness of the flow control mechanism. In addition, the VS/VD technique may be used as an ABR gateway at the border between fixed and wireless networks or for the control of individual ABR connections at administrative subnet boundaries. As a consequence, the VS/VD mechanism plays an important role for the performance and the traffic management of the ABR service class.

The goal of this thesis is the development and performance analysis of an ABR switch algorithm especially designed for a VS/VD switch. Since the ABR flow control mechanism describes a classical closed-loop feedback control system, *linear control theory* will be used as an analytical design method. This allows to target design and performance objectives already during the development phase and to assess the results analytically. Whereas linear control theory has already been used to analyze classical ABR switch algorithms, existing research works dealing with the design of VS/VD algorithms use only heuristics and simulation-based approaches. Therefore, this thesis meets the challenge of designing and modeling a novel VS/VD switch algorithm with the help of linear control theory and mathematical methods.

The major performance objectives are fair bandwidth allocation for ABR connections, together with an effective control of switch queue levels. Fairness is an important property of ABR switch algorithms, because each ABR connection should receive the maximum possible bandwidth allocation for optimal throughput. Small queue levels lead to small buffer requirements at the switch and hence to a good scalability in the number of supported ABR connections. In addition, a small and steady buffer length positively affects the end-to-end delay and jitter for ABR connections. Furthermore, the newly developed VS/VD algorithm should guarantee a stable and robust operation, have a low complexity, and be compliant to the ABR flow control

scheme defined by the ATM Forum. Due to the coupling of two separate ABR control loops at the VS/VD switch and the strong interaction between the different control actions taken, it is by far not obvious how these performance objectives could be reached. The analytical approach of this thesis offers the advantage of a modular and structured design, taking into account the special layout and various design options of a VS/VD switch.

In addition to the analytical performance evaluation, simulation experiments will be performed in order to validate the control model derived and to assess the actual behavior of the newly designed VS/VD algorithm under various network configurations. The use of cell-based computer simulations offers the advantage of an extensive and detailed investigation. By monitoring important performance metrics, the VS/VD switch algorithm can be analyzed under many different network topologies, source configurations, and load conditions.

In summary, the combination of linear control theory as an analytical design method and computer simulations as a flexible performance evaluation tool will be applied in this thesis to develop a new ABR flow control algorithm tailored for the implementation in a VS/VD switch.

1.2 Layout

After this introduction, the second chapter provides an overview of the Asynchronous Transfer Mode (ATM) technology. The standardization groups working on ATM, the ATM reference model with its different layers, and the ATM service classes are introduced. Furthermore, the main features and characteristics of ATM networks are discussed. In addition, the various ATM traffic management functions and their application are addressed.

Chapter 3 offers an introduction to the area of flow and congestion control. After that, an in-depth explanation of the flow control mechanism used for ABR traffic is given. The switch and end system behavior as well as the ABR connection parameters are highlighted. Furthermore, the special design of a virtual source / virtual destination (VS/VD) switch is discussed.

In the following fourth chapter, the main concepts and terms of linear control theory are introduced. The special group of closed-loop feedback control systems, relevant for the modeling of the ABR flow control scheme, is presented. Mathematical methods for the assessment of linear control systems are explained and continuous as well as discrete time control models are introduced.

Chapter 5 focuses on the development of the new VS/VD switch algorithm. First, an overview of the related work in the area of ABR flow control is given, putting special emphasis on the use of control theory for the design and on existing VS/VD switch algorithms. In the following, the design and performance objectives for the new algorithm are outlined. Based on these objectives and a traffic model describing the queue dynamics at the VS/VD switch, a corresponding control model is developed. From this control model, the novel controller is derived and its parameters are determined analytically with the help of linear control theory. Finally, implementation details are discussed.

A performance evaluation of the newly developed VS/VD switch algorithm using mathematical methods as well as simulation experiments is presented in chapter 6. After the introduction of the simulation environment and the different traffic models used, the results for the various simulation scenarios are presented. The VS/VD switch algorithm is assessed and evaluated with respect to the performance and design objectives defined in chapter 5.

The final chapter concludes the thesis with a summary of the work performed and the results obtained. A brief outlook on future work and other research areas that are related to the field of ABR flow control is given.

CHAPTER 2

Asynchronous Transfer Mode

This chapter introduces the basic concepts, key elements, and features of the *Asynchronous Transfer Mode* (ATM). The notion of logical connections, the structure of an ATM cell, and the ATM protocol layers are presented. Special attention is given to the traffic management aspects and the different ATM service classes.

2.1 History of ATM

During the 1980s the first standardization documents for an *Integrated Services Digital Network* (ISDN) were issued by the *International Telegraph and Telephone Consultative Committee* (CCITT), now named the *International Telecommunications Union - Telecommunication Standardization Sector* (ITU-T) [G.705]. During the 1990s the desire to integrate telephone and data networks into a single network emerged. Furthermore, the need for large bandwidth capacities and stringent real-time requirements (low latency and jitter) for the transmission of data and multimedia streams, like video and audio, became evident. This led to the development of *Broadband-ISDN* (B-ISDN) [I.121] by the ITU-T, who has chosen ATM as the underlying transmission technique [I.311].

In addition to the efforts of the national standardization bodies organized in the ITU, the ATM Forum was founded in 1991. As a non-profit organization, it aims at promoting the deployment of ATM by providing interoperability specifications. The ATM Forum enabled a strong industry co-operation and the agreement on numerous implementation standards. Beside major ATM equipment manufacturers, several universities and research institutions are among the over 600 members [Dobr98]. Due to the great significance of ATM Forum specifications for the theoretical background and practical application of ATM technology, this thesis will relate in great parts to these documents. When appropriate, references to the ITU-T standards will also be included throughout this work.

2.2 Key Features

ATM is a connection-oriented multiplexing and switching technique operating on packets of a fixed size (*cells*), enabling the transmission of data with a constant or variable bit rate.

The main advantages of the ATM technology are the following.

- ATM enables the *integration* of different traffic types, like data, video and voice, into a single physical network. For this reason, ATM supports the convergence of different existing networks dedicated to special services to a common B-ISDN.
- A key benefit of ATM is its ability to provide *Quality-of-Service* (QoS) guarantees to applications on a per-connection, end-to-end basis. Individual QoS parameters may be negotiated between end systems and the network, which are then guaranteed for the duration of the connection.
- ATM may operate over a wide range of transmission rates and therefore *scales in bandwidth*. This offers the opportunity to support various types of applications from low speed voice connections up to large bandwidth trunk lines with a capacity of several Gigabits per second.
- ATM allows a *seamless transition* from local to wide area networks without changing the network technology. This eliminates the need of special gateways at network boundaries or the implementation of performance degrading protocol translations. Local ATM networks may therefore flexibly be expanded to cover larger areas or be interconnected in a simple way.

ATM is based on the multiplexing of fix-sized cells, which is required by network nodes (*ATM switches*) to determine packet (cell) transmission times and queueing delays [Part94]. This in turn enables the ATM network to provide the desired QoS parameters for each connection. Each ATM cell is 53 bytes long, consisting of a 5 byte header and 48 byte payload (user data). This size resulted out of a compromise between the requirements of data communications (large cells lead to a small overhead, i.e., a small ratio between the size of the header and the size of the payload field) and the needs of voice transmission (small cells lead to short packetizing delays and fast switching times) [Blac95a], [DePr95]. Long user data packets are segmented into ATM cells, which are sent consecutively as a stream. The ATM network guarantees that cells within a stream are delivered in the same order as they were transmitted, called *in-sequence delivery* [Stal98].

As mentioned above, ATM is connection-oriented, which means that before a data transmission between a source and a destination can start, a logical connection has to be established. This is realized by a *signaling protocol* (see section 2.4). The connection is called logical or virtual, since cells from different connections may be carried over the same physical link or transmission path. The cells of different connections are multiplexed in an asynchronous manner, in the sense that no fixed time slots are reserved for each connection, like in a *Time Division Multiple Access* (TDMA) system [Blac95a]. The concept of virtual connections within ATM is explained in the following section.

2.3 Virtual Paths and Virtual Channels

ATM uses the principle of *Virtual Channels* (VC) and *Virtual Paths* (VP) [I.311], each of which are uniquely labeled by a numerical identifier, the *Virtual Channel Identifier* (VCI) and the *Virtual Path Identifier* (VPI), respectively [McSp94]. Virtual Channels are grouped to Virtual Paths, where the VCIs are unambiguous within each VP. A *Virtual Channel Link* (VCL) refers to a unidirectional transfer path, whose VCI does not change [Stal98].

In analogy to the VCL, a unidirectional transfer path of a group of VCs, who are associated by a common VPI, is called a *Virtual Path Link* (VPL). At ATM switches along the network path, VCI and/or VPI values may be changed (translated). A *Virtual Path Connection* (VPC) is composed of VPLs and extends between the points where the VCI values of the contained VCs are changed, assigned, or terminated [Stal98].

A *Virtual Channel Connection* (VCC) consists of a concatenation of Virtual Channel Links, which forms a unidirectional logical ATM connection between two communicating entities. Possible ATM end points, where a VPI and VCI are assigned or removed, are an ATM end user (source or destination) or a network node. User data is transported by user-to-user VCCs between two ATM end systems. Signaling information may be exchanged between two ATM end systems, using also user-to-user VCCs, or through a user-to-network VCC, if the end point is located inside an ATM network. Network nodes may use network-to-network VCCs in order to exchange information [Stal98].

Figure 2-1 shows a VCC from a source "S" to a destination "D", which crosses two ATM switches on its path. The example illustrates the relation between the different terms introduced above. A VCC can be routed across several VPCs, which exist independently of VCCs and VCs and allow an easy routing inside an ATM backbone, see section 2.5. VCCs and VPCs may be either point-to-point, point-to-multipoint, multipoint-to-point, or multipoint-to-multipoint.

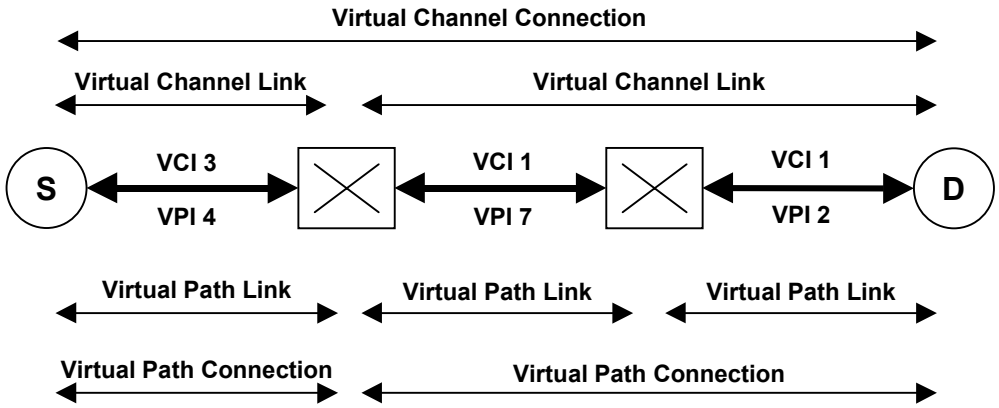

Figure 2-1: *Virtual Channel and Virtual Path Concept*

It should be noted that a *Virtual Channel* is used as a generic term for a unidirectional means of transport for ATM cells [I.311]. On a user-to-user level it is often simply referred to as a *Virtual*

Connection or *Virtual Circuit*, also called a VC [BeWa98], [BWBM98], [ChWa97], [Goya97], [LoHR00].

2.4 Signaling

A central network management instance may be used for the provision of VCCs and VPCs for a longer time period, which are then called *Permanent Virtual Connections* (PVCs). VCCs may also be dynamically set up by end systems upon request for the duration of the data transfer and are then called *Switched Virtual Connections* (SVCs) [McSp94]. In order to provide this functionality a signaling procedure is necessary at the *User-Network Interface* (UNI). The connection set-up request of the source is sent through the network to the receiver via a separate signaling channel. According to the requirements of the sender this set-up message contains traffic parameters, like the bandwidth needed, as well as QoS parameters, like the tolerated delay [UNI4.0]. Each network node passed on the way from the sender to the receiver checks if the required resources for establishing the new connection are available. If they are, the request is forwarded to the next network node via the *Network-Network Interface* (NNI) [PNNI1.0]. If the connection can not be established, an error message is returned to the previous node. This network switch may decide to propagate this error message back to the sender or to try an alternate route to the destination [Blac95b].

The ATM Forum and the ITU-T have defined different signaling standards for the UNI and NNI respectively, see table 2-1. Whereas the ITU-T is concerned with public networks, the ATM Forum specifications are used for private networks.

Signaling Interface	ATM Forum	ITU-T
User-Network Interface UNI	UNI ver. 3.0, 3.1, 4.0	DSS2 (Q.2931)
Network-Network Interface NNI	PNNI ver. 1.0, AINI	SS7 (B-ISUP)
Inter Carrier Interface ICI	B-ICI ver. 2.1	SS7 (B-ISUP)

Table 2-1: *ATM Signaling Standards*

The latest UNI standard version 4.0, dated July 1996 [UNI4.0], is based on ITU-T specification Q.2931 [Q.2931], which in turn was developed out of Q.931 [Q.931], the signaling standard for narrow-band ISDN [McSp94]. The ITU-T uses the *Digital Subscriber Signalling System 2* (DSS2) at the UNI [Q.2931] and the *Signalling System 7* at the NNI [Q.700]. The ATM Forum defined the *Private Network-to-Network Interface* (PNNI) protocol, which is based on the UNI signaling and was augmented by routing functions [PNNI1.0]. Strongly related to the PNNI is the *ATM Inter-Network Interface* (AINI) specification, that shall facilitate the inter-networking between PNNI and the *B-ISDN User Part* (B-ISUP) [AINI]. Also based on PNNI is the *B-ISDN*

Inter Carrier Interface (B-ICI), which is deployed at the interface between public ATM networks [BICI2.0], [BICI2.1].

2.5 Switching

During the set-up of a connection the signaling protocol builds up routing tables inside the ATM network nodes, that contain for each incoming VCI/VPI pair on an input port a corresponding outgoing pair and output port [ChLi95]. During the call duration the forwarding (switching) of ATM cells inside the network node is based on the VPI/VCI information contained in the cell header, see also section 2.6. Each cell belonging to one VCC takes the same route through the network and cell ordering is preserved [Stal98].

If the cell forwarding is based on the VPI only, the ATM node is called a *Virtual Path Switch* or an *ATM VP Cross-Connect* [McSp94]. VPIs are valid only locally on a link between two nodes and may change on the path from the source to the destination. Since VPs are provisioned on a longer time scale greater than the average call duration, the routing table in a VP switch is in most cases statically defined by the network management. In general it is necessary to switch cells between different VCs, which is realized by a *Virtual Channel Switch* [I.311], see figure 2-2.

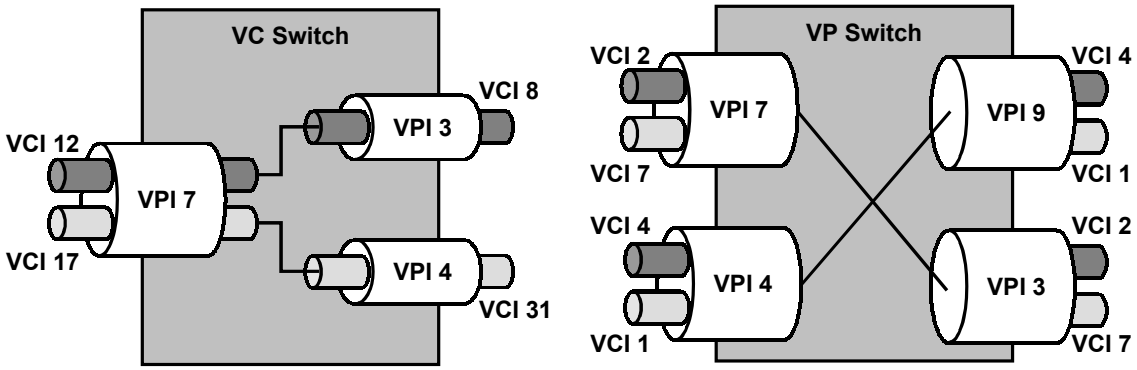

Figure 2-2: *VC / VP Switching*

In the example of figure 2-1, the left (or the first switch on the path from the source to the destination) is a VC switch, whereas the right switch (or the second switch on the path) is a VP switch.

2.6 ATM Cell Structure

Figure 2-3 presents the structure of an ATM cell. The 5 byte header is composed of a 4 byte control field and a 1 byte *Header Error Control* (HEC). The remaining 48 bytes of the ATM cell contain the payload and are called information field [DePr95]. The format of the ATM cell

header depends on the interface (UNI or NNI) that the cell is crossing. The difference between the two formats is the use of the first 4 bits as *Generic Flow Control* (GFC) at the UNI, whereas these bits are used at the NNI for the VPI. The GFC was intended to control the input of terminals attached to an ATM multiplexer via the UNI [McSp94]. This option is not supported at the moment, so that the GFC bits are always set to 0, indicating an uncontrolled source [Blac95a].

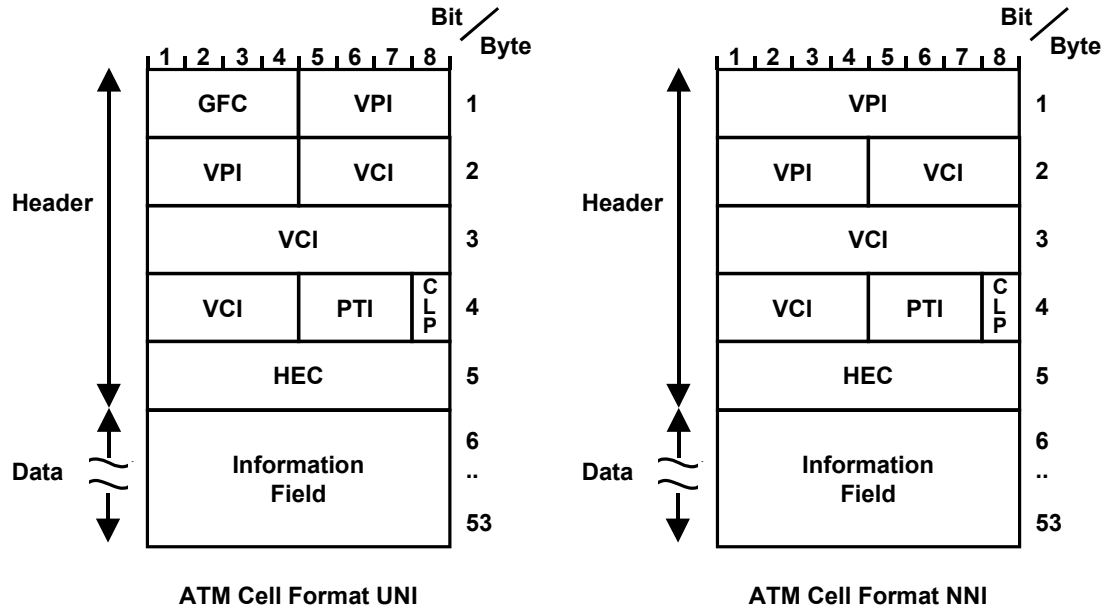

Figure 2-3: *ATM Cell Format*

The various fields of the ATM cell header have the following meaning:

- **GFC**: The *Generic Flow Control* field at the UNI is currently not used, see explanation above.
- **VPI**: The *Virtual Path Identifier* defines a label for the VP the cell is associated with.
- **VCI**: The *Virtual Channel Identifier* indicates the VC the corresponding cell belongs to.
- **PTI**: The 3 bit *Payload Type Identifier* field is used for the classification of cell types. Bit one distinguishes between data and other cells, like for *Operation, Administration, and Maintenance* (OAM). In case of a user data cell, bit two is used as an *Explicit Forward Congestion Indication* (EFCI), which is used to carry congestion information, see section 3.2.2. If the cell carries user data, then the third bit marks the last cell that belongs to a user *Protocol Data Unit* (PDU), see section 2.7.2. The value "111" (bin) of the PTI is reserved for future use [McSp94].
- **CLP**: The *Cell Loss Priority* bit is set in low priority ATM cells, which may be discarded by ATM switches in the presence of network congestion or to assure a low cell loss rate for high priority cells. An ATM cell may be tagged, i.e., the CLP bit set to 1, by the application at the source or by network nodes on the path from the sender to the destination [McSp94].

- **HEC**: The *Header Error Control* field contains an 8 bit *Cyclic Redundancy Check* (CRC) based on the first 4 bytes of the header. The checksum allows the correction of single bit errors and the detection of multiple bit errors with a certain probability [DePr95]. It has to be stressed that the checksum of the HEC covers only the cell header and not the user payload. The HEC is also used by the physical layer to detect cell boundaries in a continuous byte stream, see section 2.7.4.1.

2.7 B-ISDN Reference Model

The B-ISDN reference model of the ITU-T defines four layers and three planes [I.321], as depicted in figure 2-4. The *user plane* consists of B-ISDN applications, that use the services of the ATM Layer for data transport. The *control plane* hosts the signaling application, which uses the primitives of the signaling protocol to perform tasks like connection set-up or tear-down. The management plane implements the *Operation, Administration, and Maintenance* (OAM) functionality. It relies on the layer management for tasks, that affect only one layer and on the plane management for inter-layer coordination and problems that extend over several layers.

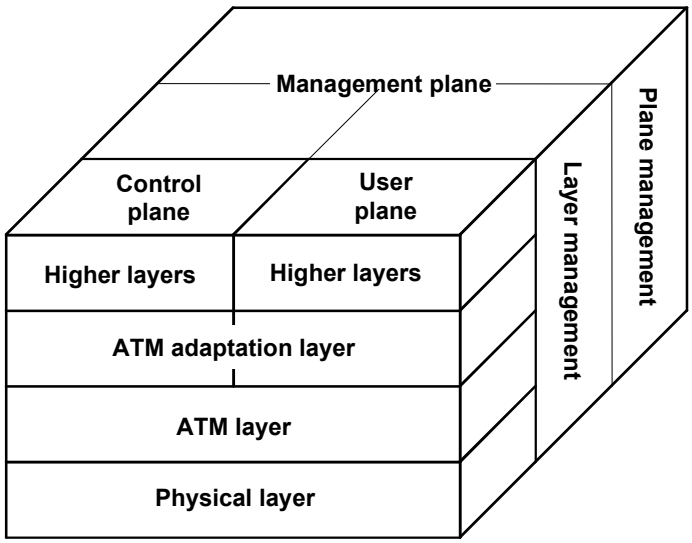

Figure 2-4: *B-ISDN Protocol Reference Model [I.321]*

The four layers defined by the ITU-T are presented in the following.

- The **Higher Layer** is hosting B-ISDN applications, that use the ATM services for data transmission.
- The **ATM Adaptation Layer** (AAL) provides functions like segmentation and reassembly to convert *Protocol Data Units* (PDUs) of higher layers to ATM cells.
- The **ATM Layer** performs the multiplexing and routing of ATM cells.
- At the **Physical Layer** the ATM cells are prepared for the transport over different physical media.

It has to be noted, that there is no clear mapping between the layers of B-ISDN reference model and the *Open System Interconnection* (OSI) basic reference model of ISO [X.200]. The ATM Layer may be compared to the OSI layers 2 and 3, whereas the ATM Adaptation Layer implements functionalities of the OSI layers 4 and 5 [Blac95a]. In the following sections the tasks and functions of the four layers will be presented in greater detail.

2.7.1 Application Layer

As this section focuses on the user part of the Higher Layer, which contains B-ISDN applications like video transport, the name *Application Layer* will be used as a synonym for this Higher Layer. Although the ITU-T discontinued its Recommendation I.362, which introduced four application classes, this approach is used in the following for illustrative reasons. Applications may be divided into the four classes (A to D) [McSp94] according to the categories:

- **Synchronization**: Indicates if a time synchronization between the sender and the receiver is required. An uncompressed live video stream would need a common time basis to determine the rate at which the received video frames have to be displayed.
- **Bit Rate**: One can distinguish between application transmitting data at a fixed or variable bit rate. A compressed video stream generates a variable bit rate, whereas a circuit emulation, e.g., for a 64 kbit/s line, will occupy a fixed amount of bandwidth.
- **Connection Mode**: Although ATM is a connection-oriented transmission technique, a B-ISDN application might be connection-oriented, like an FTP application, or operate in a connectionless mode, like email transfer.

Table 2-2 presents the correspondence of the four application classes A to D with the characteristics introduced above [Blac95a].

Parameter / Class	A	B	C	D
Synchronization	required		not required	
Bit Rate	constant	variable		
Connection Mode	connection-oriented			conn.-less

Table 2-2: *B-ISDN Application Classes*

Examples for applications of class A (isochronous, connection-oriented transfer with constant bit rate) are ISDN voice services with 64 kbit/s. Compressed video streams, like MPEG-2, have a variable bit rate and belong to class B. Programs like telnet need a connection between sender and receiver, have a variable bit rate, but no synchronization is necessary, so that they are representatives of class C. Connectionless services, like news transfer, fall into class D.

2.7.2 ATM Adaptation Layer (AAL)

The *ATM Adaptation Layer* (AAL) is responsible for making the ATM data transmission transparent to the application. It is composed of the *Convergence Sublayer* (CS) and the *Segmentation and Reassembly* (SAR) *Sublayer* [I.363].

- **Convergence Sublayer**: The CS represents the interface at which the ATM transmission starts and ends. The application hands over the user *Protocol Data Unit* (PDU), which becomes the *AAL Service Data Unit* (SDU) in the CS. The application receives the user PDU back from the CS at the destination. Underlying layers are not accessed by the application directly. The tasks of the CS depend on the AAL type and may include clock recovery at the receiver or the check if a packet was received completely. The CS may be further divided into a *Service Specific Convergence Sublayer* (SSCS) and the *Common Part Convergence Sublayer* (CPCS). The former implements the functionality common for all services of the corresponding AAL, whereas the latter may be empty or may provide special AAL services [Blac95a]. These sublayers optionally add a header and trailer, forming an SSCS-PDU or CPCS-PDU, respectively.
- The **SAR Sublayer** fragments the CS-PDU into SAR-PDU payloads, possibly adding an SAR-PDU header and trailer. These parts form a complete SAR-PDU, which has a size of 48 bytes and is transported in the payload field of an ATM cell. At the receiver side, the cells are reassembled to CS-PDUs and detected bit errors are corrected, if possible. Furthermore, the SAR layer may provide sequence numbering and may mark the end or the non-fragmentation of a CS-PDU inside the *Payload Type Identifier* (PTI) field of the ATM cell header [Stal98].

Figure 2-5 depicts the dependencies between the different data units inside the AAL. The SAR Sublayer and the CPCS may also be addressed as the *AAL Common Part* (CP).

The AAL provides five different AAL types with corresponding *Service Access Points* (AAL-SAPs) for the four service classes A to D [McSp94]. The AALs have been designed with the four applications classes in mind and therefore there is a strong correlation between these two [McSp94], see table 2-3.

<table>
<tr><td>Application Class</td><td>A</td><td>B</td><td>C</td><td>D</td></tr>
<tr><td rowspan="2">AAL Type</td><td rowspan="2">AAL 1</td><td rowspan="2">AAL 2</td><td>AAL 3</td><td>AAL 4</td></tr>
<tr><td colspan="2">AAL 5</td></tr>
</table>

Table 2-3: *Mapping of AALs and B-ISDN Application Classes*

Classes A to D map to the corresponding AALs 1 to 4 and AAL 5, as a successor of AAL 3/4, relates to both application classes C and D. However, this mapping is not mandatory, in the sense that in general the applications will use the corresponding AAL type, but in special cases may choose another AAL service [Blac95a].

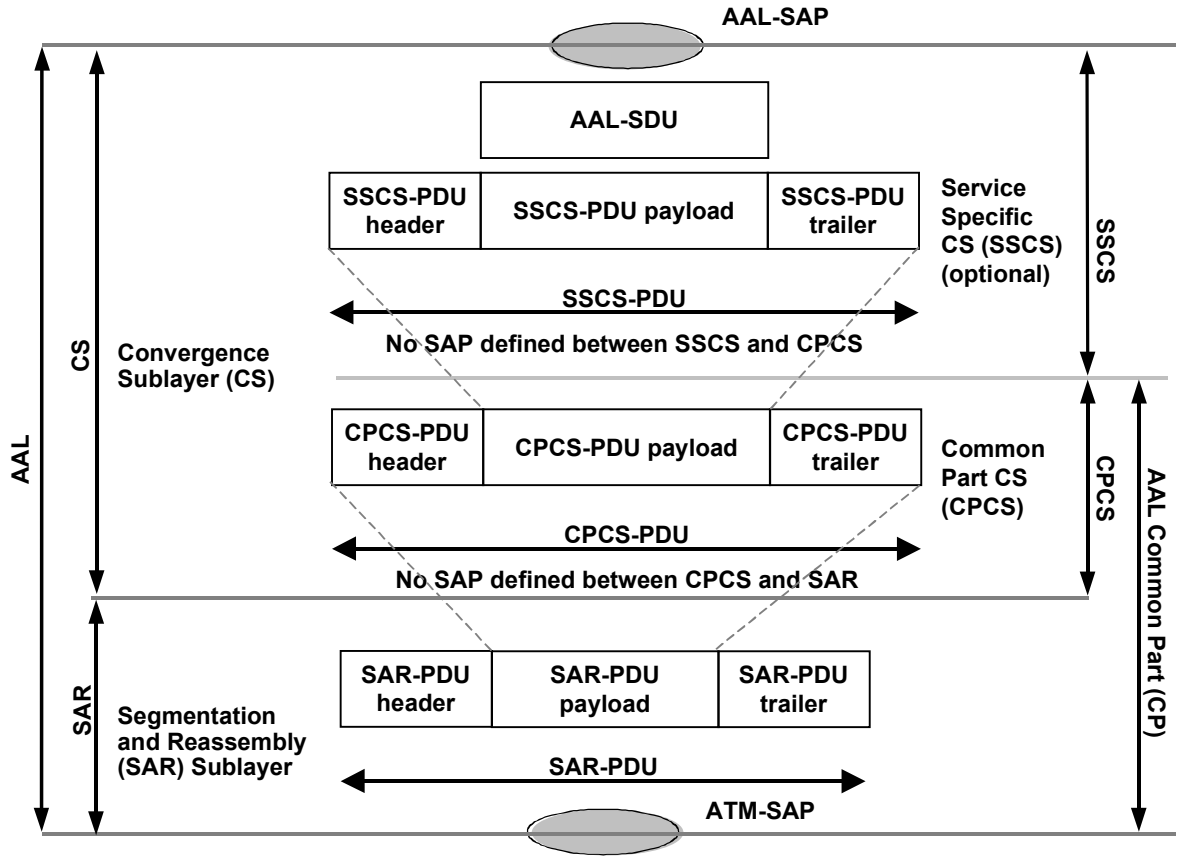

Figure 2-5: *AAL Sublayers and Data Units*

2.7.2.1 AAL 1

The *ATM Adaptation Layer Type 1* (AAL 1) is specified in ITU-T Recommendation I.363.1 [I.363.1]. It was designed for the use by constant bit rate applications, that possibly need timing information at the receiver, like circuit emulation and uncompressed video or audio streams [Blac95a].

- On the **CS**, AAL 1 provides cell buffering, detection of lost or misinserted cells and clock synchronization, so that cells can be delivered and processed at a constant rate. Furthermore, it is possible to use *Forward Error Correction* (FEC) for delay-sensitive audio and video data. In addition AAL 1 supports a bit-oriented (*Unstructured Data Transfer*, UDT) and a byte-oriented (*Structured Data Transfer*, SDT) transmission mode [I.363.1]. For the UDT, AAL 1 has no special frame format on the CS, the data is passed to the SAR Sublayer in chunks of 47 bytes. The SDT uses, for every other CS-PDU, a pointer field in the first byte, followed by 46 bytes of user data. The pointer field contains a parity bit followed by a 7 bit offset field, indicating the structure boundaries [ChLi95].
- On the **SAR Sublayer**, each ATM cell belonging to an AAL 1 stream consists of the ATM cell header, followed by a 1 byte SAR-PDU header and 47 bytes of SAR-PDU payload [McSp94].

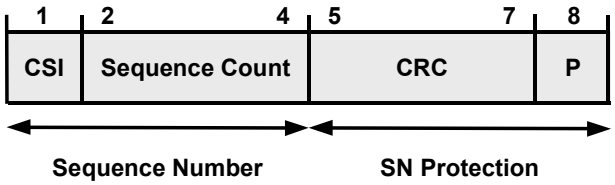

Figure 2-6: *AAL 1 SAR-PDU Header*

The AAL 1 header, see figure 2-6, consists of a *Sequence Number* (SN) and an *SN Protection* (SNP) field, each of which is 4 bits long. The SN field is composed of the *Convergence Sublayer Indication* (CSI) bit, set by the CS sublayer, and a *Sequence Count* (SC), which allows the detection of lost cells at the CS sublayer. The CSI bit is used by the CS for its services, e.g., FEC, clock recovery (in even numbered SAR-PDUs), or SDT (in odd numbered SAR-PDUs). The SN field is protected by the following 4 bit *Sequence Number Protection* (SNP) field. It consists of a 3 bit CRC checksum and a parity bit (P), which is calculated based on the complete AAL 1 header [ChLi95].

2.7.2.2 AAL 2

AAL 2 was designed to carry traffic with a variable, possibly low, bit rate, with short and variable length packets for delay sensitive applications. Furthermore, it is possible to multiplex several AAL 2 channels over one AAL 2 connection [I.363.2]. The ITU-T AAL 2 standard distinguishes between the *Service Specific Convergence Sublayer* (SSCS) and the *Common Part Sublayer* (CPS), whereas only the latter is addressed in the standard [I.363.2].

- On the **CPCS** level AAL 2 operates with *CPS-Packets*, which correspond to CPCS-PDUs. These CPS-Packets transfer AAL-SDUs of up to 45 (default) or 64 bytes.

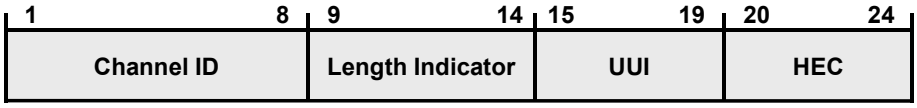

Figure 2-7: *AAL 2 CPS-Packet Header*

Figure 2-7 presents the 3 byte header preceding the AAL-SDU. The first header byte contains a *Channel ID* (CID) identifying the channel to which the payload belongs to. Thereafter follows a 5 bit *Length Indicator* (LI) that specifies the number of data bytes in the payload section. The *User-to-User Indication* (UUI) field of 5 bits transparently transports CPS layer information and the *Header Error Correction* (HEC) protects the CPS-Packet header [Bald97].

- On the **SAR** level, these CPS-Packets are then transported by AAL 2 CPS-PDUs of 48 bytes, which correspond to SAR-PDUs. They consist of a 1 byte CPS-PDU header, or *Start Field* (STF), and 47 bytes of payload [Blac95a]. This CPS-PDU payload may contain zero, one, or more (complete or partial) CPS-Packets, where unused payload is filled with padding

bytes with a zero value [I.363.2]. It has to be noted, that even if a default maximum AAL-SDU size of 45 bytes is chosen, the resulting CPS-Packet may have a size of 48 bytes and will overlap into two AAL 2 CPS-PDUs on the SAR layer.

Figure 2-8: *AAL 2 CPS-PDU Header*

The CPS-PDU header starts with a 6 bit *Offset Field* (OSF), that indicates the beginning of the next CPS-Packet or, if not applicable, the start of the padding bytes. The next field contains a 1 bit *Sequence Number* (SN) (modulo 2) of the CPS-PDUs. Finally, a parity bit (P) is used to protect the STF [Bald97].

2.7.2.3 AAL 3/4

Originally it was planned to develop two different AAL types for service classes C and D. During the standardization process is was noted that there were many similarities between connection-oriented and connectionless data transfer with no synchronization and variable bit rate. Hence, it was decided to merge the two types into a single AAL 3/4 [Stal98]. It provides a stream and a message mode, which is able to carry PDUs of variable size up to 64 kbyte. The transport is non-assured, but correct sequence and error check are guaranteed. For connection-oriented services, AAL 3/4 supports in addition the multiplexing of up to 1024 single user connections over one VC [I.363.3].

As mentioned above, the CS may be divided into the *Common Part Convergence Sublayer* (CPCS) and a *Service Specific Convergence Sublayer* (SSCS). The SSCS provides special user services and may be empty.

- On the **CPCS** AAL 3/4 uses a special PDU structure, which comprises a 4 byte header, a payload field of up to 65,535 byte, aligned on a 4 byte boundary, and a 4 byte trailer [ChLi95].

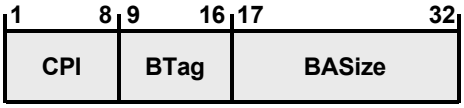

Figure 2-9: *AAL 3/4 CPCS-PDU Header*

In the CPCS-PDU header, see figure 2-9, the 1 byte *Common Part Indicator* (CPI) determines the semantics of the *BASize* and *Length* fields in the header and trailer, respectively. Currently only a null value is specified, indicating that the size and length fields relate to byte values [I.363.3]. The *Beginning Tag* (BTag) field of 1 byte carries the same bit pattern

as the *End Tag* (ETag) in the CPCS-PDU trailer and is changed for every new CPCS-PDU. Because of this mechanism the erroneous reassembly of an AAL 3/4 packet in the case of cell loss is avoided. The *Buffer Allocation Size* (BASize) field of 2 bytes indicates to the receiver AAL the size of the buffer needed to store the message. The units are bytes, as defined by the CPI field [McSp94].

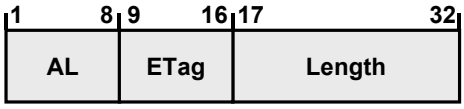

Figure 2-10: *AAL 3/4 CPCS-PDU Trailer*

The trailer of an AAL 3/4 CPCS-PDU starts with a zero filled *Alignment* (AL) byte, for alignment on a 4 byte boundary. Next, follows the *End Tag* (ETag), which contains the same pattern as the *BTag* of the header and serves as identification of the current CPCS-PDU. The size of the AAL 3/4 CPCS-PDU is noted in the 2 byte *Length* field of the trailer. As for the *BASize* field in the header, the units are bytes [Blac95a].

- Such an AAL 3/4 CPCS-PDU is segmented for transmission by the **SAR Sublayer** into corresponding SAR-PDUs. These SAR-PDUs have a 2 byte header and trailer, which enclose a 44 byte SAR payload [McSp94].

Figure 2-11: *AAL 3/4 SAR-PDU Header*

In the SAR-PDU header, see figure 2-11, the 2 bit *Segment Type* (ST) field marks the beginning, the continuation, or the end of a message as well as the presence of an unfragmented message, that fits into the 44 byte payload. The following 4 bit *Sequence Number* (SN) allows the detection of lost cells belonging to one CPCS-PDU. The *Multiplexing Identification* (MID) field allows the multiplexing of several user connections over one ATM connection on a user-to-user basis. If this service is not used, the MID is set to zero [I.363.3].

Figure 2-12: *AAL 3/4 SAR-PDU Trailer*

The SAR-PDU trailer consists of a 6 bit *Length Indicator* (LI) and a 10 bit CRC checksum based on the content of the complete SAR-PDU up to this position. The LI field indicates the number of used bytes in the payload field of the SAR-PDU [ChLi95].

2.7.2.4 AAL 5

Due to the large overhead of 9 bytes (53 byte cell size and 44 byte payload) for AAL 3/4 cells, AAL 5 was defined as a more simple and efficient adaptation layer [Stal98]. In contrast to AAL 3/4 there exists only an AAL 5 CPCS-PDU format on the Common Part Sublayer (CPCS) of the CS and no overhead due to a special cell format on the SAR Sublayer. Like AAL 3/4, type 5 provides a stream and a message mode, which is able to carry PDUs of variable size up to 64 kbyte. The transport is non-assured, but correct sequence and error check are guaranteed [I.363.5]. In contrast to AAL 3/4 the CPCS supports no multiplexing and information on the required receiver buffer size is missing [Blac95a].

- The CPCS-PDU for AAL type 5 consists of an AAL-PDU payload of 1 to 65,535 byte, an optional padding field and an AAL 5 trailer. The payload is extended by 0 to 47 padding bytes, in order to achieve alignment to a 48 byte cell boundary [Blac95a].

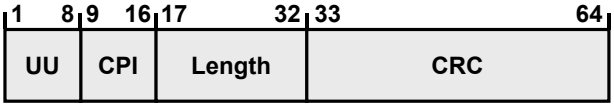

Figure 2-13: *AAL 5 CPCS-PDU Trailer*

  The 8 byte AAL 5 CPCS-PDU trailer, see figure 2-13, contains a *User-to-User* (UU) byte, that can be used to transparently convey CPCS user information, e.g., SSCS sequence numbers or multiplexing IDs, to the receiver. A *Common Part Indicator* (CPI) field of 1 byte is used for 64 bit alignment and its content is reserved for future use. The 2 byte *Length* field indicates the payload size, excluding any padding bytes. The CRC checksum in the last 4 bytes of the trailer is based on the complete AAL 5 packet [I.363.5].

- Since there is no special SAR-PDU format and CPCS-PDUs are aligned on a 48 byte boundary, the **SAR Sublayer** needs just to segment/reassemble the CPCS-PDU into/from SAR-PDUs of 48 bytes length. With respect to the segmentation, the end of an AAL 5 CPCS-PDU is indicated by setting a bit inside the *Payload Type Identifier* (PTI) field of the ATM cell header [ChLi95], see section 2.6.

2.7.3 ATM Layer

The *ATM Layer* is responsible for the transport of ATM cells between two ATM Layer instances. The important notions for the ATM Layer, namely, the Virtual Channel/Virtual Path concept and the structure of an ATM cell have already been introduced in sections 2.3 and 2.6. The ATM Layer accepts at its *Service Access Point* (SAP) ATM-PDUs of 48 bytes from the ATM Adaptation Layer and creates the ATM cell header. Furthermore, the ATM Layer determines the connection on which the cell is sent and enters the VPI/VCI values into the cell header [I.321]. The header checksum (HEC) is calculated and inserted by the physical layer, see section 2.7.4.1. The ATM Layer manages also a routing table, which contains for each incoming VPI/VCI pair a corresponding outgoing pair and link [Blac95a]. The routing table of a VP switch consists only of VPI values, see section 2.5. The entries of the routing table can be

statically configured by the network management (PVCs) or dynamically generated by the signaling protocol (SVCs). Because VPI/VCI pairs have only local significance and are valid only on the current path segment, they change at each network node along the path from the sender to the destination. Based on this path label, cells are switched on the ATM Layer between the input and output ports of ATM network nodes (switches) [SaMe96]. Therefore, queueing of ATM cells at the input and/or output ports might be necessary. The selective discarding of ATM cells with the *Cell Loss Priority* bit set to 1 (CLP = 1) in the case of an overload situation is performed at the ATM Layer [McSp94], see section 2.8.4.5. Other tasks include the asynchronous (de)multiplexing of cells from different connections, the monitoring of traffic characteristics (see section 2.8.4.2), the handling of flow control information (see section 3.2.2) and the support of *Quality of Service* (QoS) guarantees, e.g., through the implementation of certain cell scheduling or buffer management strategies [McSp94], [GiGa98], see section 2.8.4.7.

2.7.4 Physical Layer

This layer deals with the issue of the physical media used for transmission. It can be divided into two sublayers. The *Physical Medium Dependent* (PMD) sublayer, which depends on the physical media used, and the *Transmission Convergence* (TC) sublayer, that converts the cell stream into transport units that can be transmitted by the PMD [McSp94].

2.7.4.1 Transmission Convergence Sublayer

The task of the TC sublayer is to receive a data unit from, or, to send it to the PMD. This data unit may be an ATM cell or a frame containing several cells, like it is used by the *Synchronous Digital Hierarchy* (SDH) [I.432]. In case of such a frame-oriented transmission, the TC sublayer provides functions for frame generation and for cell extraction on the receiver side [I.432].

The header checksum (HEC) is generated and checked upon reception by the TC sublayer. The CRC calculation is also used in order to determine the cell boundaries (*cell delineation*) [Stal99]. If 8 bits contain a valid CRC for the preceding 4 bytes, then a cell header is assumed and a pre-sync state is entered. For a cell-based transmission with no frame structure (unsynchronized byte stream), the next 8 cells are probed. If this procedure is successful, cell synchronization is achieved and the cells are passed to the ATM Layer. If the procedure does not succeed or an error is detected during 7 consecutive cells, the procedure is restarted [I.432].

The TC sublayer processes information for *Operation, Administration, and Maintenance* (OAM) provided by the physical media [DePr95]. Another task is the insertion of idle cells into the cell stream for *cell rate decoupling* if no data cells are to be sent [McSp94]. Some media expect a continuous cell stream, in order to maintain their own cell rate, independent of the current transmission rate. The idle cells inserted have a special cell header (VPI/VCI = 0, PTI = 0, CLP = 1) and are not passed to the ATM Layer [I.432].

2.7.4.2 Physical Medium Dependent Sublayer

This sublayer is responsible for sending and receiving the bit stream. Clock recovery is performed on the receiving stream, which is then used for demodulation, synchronization of the decoding process, and for the sending process [Blac95a]. The coding may be bit-oriented or block-oriented, based on certain bit or signal patterns [G.703]. Block codes offer the advantage that the fixed bit patterns allow a simple clock recovery at the receiver and provide additional symbols for synchronization purposes, like frame end or begin. The disadvantage is the additional overhead introduced, for instance if 4 bits are coded into 5.

2.8 Traffic Management

As mentioned earlier in this chapter, one major goal of ATM is the integration of different traffic streams and the guarantee of QoS parameters for each connection. Hence, an ATM network has to accommodate heterogeneous traffic profiles and different QoS requirements. In order to specify the traffic characteristics and QoS parameters of a connection a *Traffic Contract* is used, which is described in the following section. An ATM network provides different *Service Categories*, which are presented in section 2.8.2. The *Traffic Management* defined by the ITU-T is explained in section 2.8.3. Finally, the various *Traffic Management Functions* operating in different areas and on different levels are introduced in section 2.8.4.

2.8.1 Traffic Contract

In order to guarantee a certain QoS for a connection, a *Traffic Contract* is negotiated between the sender, the network nodes and the receiver upon connection set up [GiGa98]. The traffic contract contains [TM4.1]:

- a **Source Traffic Descriptor**, indicating the traffic characteristics,
- a **QoS Specification**, indicating the desired QoS,
- the **Cell Delay Variation Tolerance** (CDVT), and
- a **Conformance Definition**, in order to identify non compliant cells.

The traffic characteristics specified by the *Source Traffic Descriptor* may be altered by the transport through the network. For example, the multiplexing inside the network nodes leads to deviations from the original cell rate. In addition cells may be delayed through the insertion of *Operation, Administration, and Maintenance* (OAM) cells or of frame information on the physical layer. The cell delay variation introduced by these events is bounded by the CDVT [Blac95a]. The CDVT varies in different parts of the network and may not be negotiated by the user [McSp94].

2.8.1.1 Source Traffic Descriptor

The *Source Traffic Descriptor* contains the following traffic parameters, which specify the characteristics of the traffic flow [GiGa98].

- The **Peak Cell Rate** (PCR) is defined as the reciprocal of the minimum inter-arrival time between two cells. Hence, it determines the maximum transmission rate of the sender [TM4.1].
- The **Sustainable Cell Rate** (SCR) of a connection specifies the required average cell rate, determined over a long time period [GiGa98].
- The **Minimum Cell Rate** (MCR) denotes the minimum transmission rate, that needs to be provided for the connection. MCR may be zero, so that the network does not need to guarantee a minimum transmission rate.
- The **Maximum Burst Size** (MBS) determines the maximum number of cells, that may be sent continuously at PCR [McSp94]. As a consequence, there must follow at least one empty cell slot after sending a burst of length MBS.
- The **Maximum Frame Size** (MFS) specifies the maximum size of a user frame at the AAL in ATM cells [GiGa98].

The rate parameters *PCR*, *SCR*, and *MCR* are associated with corresponding CDVT values, so that these can be taken into account during the conformance testing of the different cell rates [TM4.1].

2.8.1.2 Quality of Service Specification

In order to describe the QoS of a connection, the *Traffic Management Specification* of the ATM Forum [TM4.1] defines a number of QoS parameters, whose values are negotiated at connection set-up.

- Two end-to-end delay parameters are negotiated, namely, the **peak-to-peak Cell Delay Variation** (ptp CDV) and the **Maximum Cell Transfer Delay** (maxCTD). Figure 2-14 illustrates the correlation between the two delay parameters.

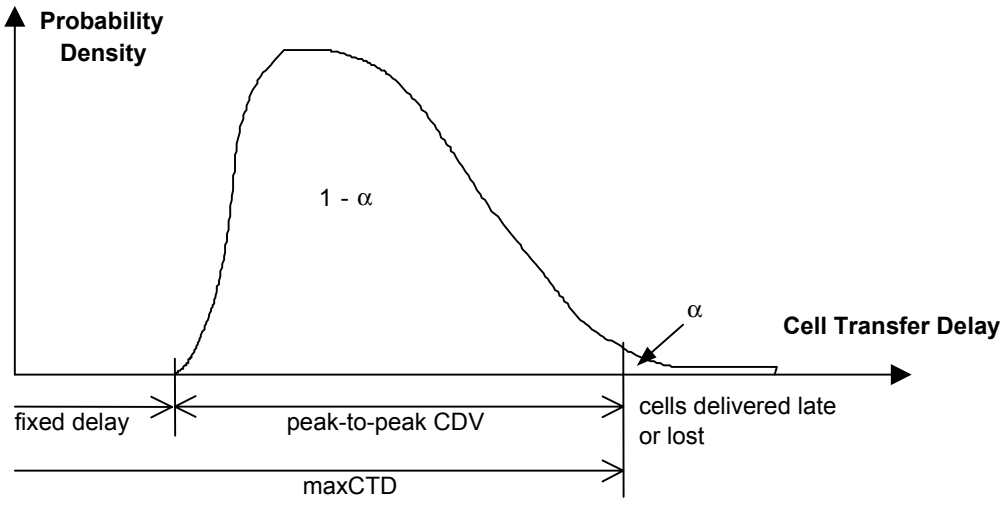

Figure 2-14: *Cell Transfer Delay Probability Density Model [TM4.1]*

The maxCTD determines the (1 - α) quantile of the cell transfer delay. The peak-to-peak CDV defines the variance of the cell transfer delay between the fixed minimum delay and the value of maxCTD. It therefore bounds the cell jitter of a connection [McDy00].

- The **Cell Loss Ratio** (CLR) is defined as the quotient of the number of lost cells and the total number of transmitted cells [Blac95a].

2.8.2 ATM Service Categories

In order to categorize the different traffic streams inside an ATM network, the ITU-T [I.356], [I.371], [I.371.1] as well as the ATM Forum [TM4.1] defined specifications for ATM traffic management. Each specifies various service classes, different traffic, and QoS parameters. Although the specifications are similar, differences in the terms used and technical details exist. Since the *Traffic Management Specification* of the ATM Forum gained a wider acceptance and therefore importance [Blac95a], [GiGa98], [McDy00], [SaMe96], it will be used as a basis for the ensuing discussion of Service Categories. However, the major differences to the ITU-T standards will be mentioned briefly in section 2.8.3.

The ATM Forum *Traffic Management Specification* defines the following six Service Categories or classes, which differ in the supported traffic and QoS parameters that were introduced above [TM4.1].

- The **Constant Bit Rate** (CBR) class is used for real-time applications, like audio/video or circuit emulation with a constant bit rate and stringent QoS requirements (small delay and jitter). The traffic characteristics are defined solely by the *Peak Cell Rate* (PCR), since the transmission uses a fixed bandwidth. The QoS is specified by the *Cell Loss Ratio* (CLR), a *maximum Cell Transfer Delay* (maxCTD) and its variation, the *peak-to-peak Cell Delay Variation* (ptp CDV). The guarantee of the delay parameters makes CBR suitable for real-time applications [TM4.1].
- The Service Category **Variable Bit Rate** (VBR) is applied for bursty traffic with fluctuating bandwidth requirements. VBR traffic is characterized by the parameters PCR, *Sustainable Cell Rate* (SCR), and *Maximum Burst Size* (MBS) [SaMe96]. This traffic description enables the network nodes to allocate corresponding resources for a VBR connection. The VBR source may transmit at PCR for a duration of MBS cells, but on average must send at the SCR, measured over a longer time period [GiGa98]. The VBR service class offers a statistical performance guarantee in the sense that the SCR is provided at all times, whereas the PCR is available with a high probability. This property together with the burstiness of the VBR traffic allows the use of statistical multiplexing for several VBR traffic flows [SaMe96]. The multiplexing gain results in a better resource utilization, because of a lower bandwidth allocation than the accumulated PCRs (*peak rate allocation*) [GiGa98].

  Regarding the QoS guarantee, the VBR category can be divided into **real-time VBR** (rt-VBR), e.g., for compressed video, and **non real-time VBR** (nrt-VBR), e.g., for video distribution or transaction traffic with low response times. Whereas nrt-VBR offers only CLR as a

QoS parameter, rt-VBR defines in addition the delay parameters maxCTD and peak-to-peak CDV [McDy00].

- Initially **Unspecified Bit Rate** (UBR) was designed as a best effort service with no QoS guarantees, i.e., when or if transmitted data will arrive at its destination. UBR is allowed to dynamically use bandwidth not claimed by other connections with a *high priority*, namely, CBR and VBR traffic. PCR may be specified as a traffic parameter, but it is not enforced by the network. All UBR traffic is accepted and if there is bandwidth left over from the high priority service classes, the UBR cells are also transmitted [SaMe96]. However, in the case of congestion, UBR cells will be discarded without feedback to the sender. This cell loss results in incomplete PDUs on the AAL, that must be recovered by higher layer protocols through retransmissions. As a consequence, even small cell loss rates on ATM level result in a large number of corrupted packets on the AAL, if segmentation had to be performed. This causes low throughput for example for TCP connections running over UBR [GiGa98].

  One possible solution is the use of intelligent buffer strategies for UBR switch queues [Stal98]. When *Early Packet Discard* (EPD) is used, complete AAL PDUs, based on the frame information in the Payload Type Identifier (PTI) field of the ATM cell header (see section 2.6), are discarded, if a congestion situation is anticipated [TzSi98]. *Partial Packet Discard* (PPD) is based on the fact, that if cells have to be dropped due to actual congestion, it is more advisable to drop cells belonging to one AAL PDU rather than randomly selecting cells. Hence, the remaining cells of the actually transmitted AAL PDU (a partial packet) are discarded by the ATM switch [RoFl95].

  As another approach to alleviate the problem of a high packet loss rate on the AAL and in order to provide a minimum throughput per connection, a *Minimum Desired Cell Rate* (MDCR) was introduced. The MDCR option is not mandatory, but may be supported for UBR connections by the network. The specification of the MCDR parameter neither enforces nor precludes a service commitment of the network. For this reason, the MCDR parameter is not a traffic parameter, but an additional connection attribute [ADTM].

- The **Available Bit Rate** (ABR) Service Category, like UBR, makes use of the available bandwidth left over by CBR or VBR traffic on a transmission link. In the presence of such high priority background traffic, the amount of bandwidth available to ABR connections varies with the burstiness of the background connections. In contrast to UBR, the ABR source specifies the *Minimum Cell Rate* (MCR) and *Peak Cell Rate* (PCR) traffic parameters for a connection. The network guarantees a minimum bandwidth equal to MCR during the duration of the connection and provides up to PCR when possible. No other QoS parameters except for MCR are guaranteed [SmAT96]. ABR is the only Service Category which uses network feedback to adjust the *Allowed Cell Rate* (ACR) of the sender. This feedback is provided by *Resource Management* (RM) cells, which are inserted into the user data stream. When a congestion situation is indicated to the source, its ACR is reduced. When congestion is absent, the ACR may be increased [BoFe95]. As a consequence, the CLR of ABR connections is low, but no quantitative value must be defined by the network. Typical examples of applications that could use the ABR service class, are LAN interconnection or simple data

transfer, like file transfer [McDy00]. ABR traffic should be able to cope with fluctuating bandwidths and be robust against cell loss or delay variation. A detailed description of the ABR flow control mechanism is presented in chapter 3 of this thesis.

- **Guaranteed Frame Rate** (GFR) supports frame-based traffic flows, like IP, by guaranteeing a *Minimum Cell Rate* (MCR) [GiGa98]. The rationale for the introduction of this Service Category was the desire to improve concepts like UBR + MDCR by adding a traffic descriptor. Since the application has to provide the *Maximum Frame Size* (MFS) and the *Maximum Burst Size* (MBS) as traffic characteristics, the network may use this information for the reservation of resources. In case of congestion the network may discard traffic exceeding the MCR on a per frame rather than on a per cell basis [TM4.1]. The user frames are transported as AAL 5 PDUs, hence indicating frame boundaries on the ATM Layer via the PTI field in the cell header (see section 2.7.2.4). The GFR service does not implement a flow control mechanism, like ABR, and is suited for bursty, frame-based applications that are able to specify their traffic characteristics with respect to PCR, MFS, and MBS [McDy00].

Table 2-4 summarizes the characteristics of the six service categories defined by the ATM Forum. *High priority* traffic, like applications in the area of multi media and distributed systems with strict QoS requirements will use the CBR or VBR service classes. The remaining bandwidth is available to the *low priority* ABR, UBR, and GFR categories. For this reason, these service classes provide only limited (ABR, GFR, UBR + MDCR) or no (UBR) QoS guarantees and are suited for data applications, like file, mail, or news transfer. Regarding the application classes and AAL types introduced in sections 2.7.1 and 2.7.2 respectively, the CBR class is mainly used for class A applications with AAL 1, VBR-rt for class B in conjunction with AAL 2, whereas class C and D applications use AAL 5 running over VBR-nrt, UBR, ABR, or GFR connections.

Service characteristic	CBR	rt-VBR	nrt-VBR	ABR	UBR	GFR
Description	Constant Bit Rate	Variable Bit Rate: real-time	Variable Bit Rate: non real-time	Available Bit Rate	Unspecified Bit Rate	Guaranteed Frame Rate
Bandwidth Requirement	constant	variable	variable	variable	variable	variable
Traffic Parameters	PCR	SCR, MBS, PCR	SCR, MBS, PCR	MCR, PCR	PCR	MCR, PCR, MBS, MFS
QoS Parameters	CLR, maxCTD, pt.-to-pt. CDV	CLR, maxCTD, pt.-to-pt. CDV	CLR	low CLR	-	low CLR

Table 2-4: *Characteristics of the ATM Forum Service Categories*

Service characteristic	CBR	rt-VBR	nrt-VBR	ABR	UBR	GFR
Minimum Bandwidth	yes	yes	yes	yes	optional (MDCR)	yes
Suitable for real-time traffic	yes	yes	no	no	no	no
Priority Class	high	high	high	low	low	low
Suitable for bursty traffic	no	yes	yes	yes	yes	yes
Congestion Feedback	no	no	no	yes	no	no

Table 2-4: *Characteristics of the ATM Forum Service Categories*

2.8.3 ITU-T Traffic Management

In contrast to the ATM Forum, where the Service Categories have corresponding traffic and QoS parameters, the ITU-T offers two separate documents specifying *QoS Classes* [I.356] and *ATM Transfer Capabilities* (ATCs) with traffic parameters [I.371]. Hence, a combination of an ATC and a QoS Class corresponds to an ATM Forum Service Category.

The ITU-T Recommendation I.356 specifies a total of four QoS Classes (stringent, tolerant, bi-level, and unbounded), with associated QoS parameters [I.356]. In contrast to the QoS parameters of the ATM Forum, the *average Cell Transfer Delay* (CTD) (not the maximum) and the CLR for both the aggregated cell stream with CLP = 1 and CLP = 0 and the high priority cell stream with CLP = 0 (not only for the aggregated stream) are defined by the ITU-T standard. Instead of the peak-to-peak CDV a *2-point CDV* is defined in the ITU-T standard. This parameter measures the jitter in CTD between two measurement points. In addition the following QoS performance measures are introduced [I.356].

- The **Severely Errored Cell Block Ratio** (SECBR) is defined as the ratio of severely errored cell blocks to total number of cell blocks transmitted. A cell block corresponds to a sequence of N cells (N = PCR / 25, where N is then rounded to the next larger power of 2) transmitted consecutively. A severely errored cell block occurs when more than M = N / 32 errored, lost, or misinserted cells are recorded in that block [GiGa98]. This notion was introduced in order to avoid the consideration of burst errors in the calculation of other QoS parameters, like the ones following below.
- The **Cell Misinsertion Rate** (CMR) is determined by the quotient of the total number of misinserted cells observed during a time period divided by the time period duration. Misinserted cells or time periods belonging to a *Severely Errored Cell Block* (SECB) are not

taken into account for the calculation of the CMR [GiGa98]. A misinserted cell does not belong to the corresponding connection and is generated out of a physical layer idle cell or of a cell of another connection. Since a misinserted cell has no corresponding cell transmission event, the MCR parameter is defined as a rate rather than a ratio [I.356].

- The **Cell Error Ratio** (CER) is defined as the number of errored cells divided by the number of received cells, excluding lost, misinserted, and cells contained in SECBs. [I.356]

Table 2-5 contains the different provisional QoS parameter values for the four QoS Classes of the ITU-T for public B-ISDNs, where "U" stands for unspecified [I.356].

QoS Parameter	Default	Class 1 stringent	Class 2 tolerant	Class 3 bi-level	U Class
mean CTD	none	400 ms	U	U	U
2-pt. CDV	none	3 ms	U	U	U
CLR (CLP 0 + 1)	none	$3 \cdot 10^{-7}$	10^{-5}	U	U
CLR (CLP 0)	none	none	none	10^{-5}	U
CER	$4 \cdot 10^{-6}$	default	default	default	U
CMR	1 / day	default	default	default	U
SECBR	10^{-4}	default	default	default	U

Table 2-5: *Provisional QoS Class Parameters [I.356]*

Similar to the Service Categories of the ATM Forum, but not including any QoS parameters, the ITU-T defined *ATM Transfer Capabilities* (ATCs) [I.371]. These ATCs have associated traffic parameters only, namely, PCR, SCR, and an *Intrinsic Burst Tolerance* (IBT). The IBT is related to the *Maximum Burst Size* (MBS), which is the parameter used for signaling, by

$$\tau_{IBT} = \left\lceil (MBS - 1)\left(\frac{1}{SCR} - \frac{1}{PCR}\right)\right\rceil \tag{2.1}$$

where SCR and PCR are specified in cells / s and $\lceil X \rceil$ is the first value above X out of a generic list specified for time intervals [I.371]. The following ATCs were defined [I.371].

- **Deterministic Bit Rate** (DBR) corresponds to the CBR Service Category and provides a fixed bandwidth for the duration of the connection.
- **Statistical Bit Rate** (SBR) is related to the VBR service class for connections with fluctuating bandwidth requirements. According to the specified traffic characteristics, the SBR ATC may be further divided into three subclasses [LSMA98].

- **SBR1** with a PCR specification and an SCR specification that both apply to the aggregate cell stream (CLP 0 + 1).
- **SBR2** with a PCR specification and an SCR specification that apply to the aggregate cell stream (CLP 0 + 1) and to cells with CLP = 0 only, respectively.
- **SBR3** is identical to SBR2, except that tagging is applied to cells non-confirming to the SCR specification.

• The **ATM Block Transfer** (ABT) capability allows the unidirectional, point-to-point transmission of a group of ATM cells as a *block*. Such a block is enclosed by two RM cells and its size needs not be related to the CPCS-PDU of the AAL [I.371]. ABT is based on the reservation of bandwidth on a per-block basis, at the time the block is ready for transmission at the ABT source. For this reason, it was formerly known as *Fast Resource Reservation Protocol* [McSp94]. For each block, the network allocates bandwidth on demand for a transmission according to a *Block Cell Rate* (BCR) indicated in the leading RM cell. After the cell block is transmitted, the allocated bandwidth is released. ABT is hence suited for bursty applications, that are able to adapt to variable network bandwidth [LSMA98]. At connection setup, the traffic parameters PCR, PCR for RM cells (limiting the maximum rate of BCR negotiations), and SCR (possibly zero) with a corresponding IBT, are defined. One can distinguish two modes for the ABT capability [I.371.1].

  - When **ABT with Delayed Transmission** (ABT/DT) is used, the first RM cell carries the desired BCR. This RM cell travels along the network path from the sender through the network nodes and is returned by the destination. In case of a successful reservation (ACK) all switches on the path to the receiver have allocated the desired bandwidth and the block can be transmitted by the source [GiGa98]. Since the source has to wait for one *Round-Trip Time* (RTT) before it can transmit the block, this mode is called DT. The delayed transmission limits the possible throughput, especially for high capacity links in combination with a small block size and large RTT [AtMM98].
  - **ABT with Immediate Transmission** (ABT/IT) was designed to overcome this problem. Operating in the IT mode, the source sends the cell block (user data) immediately after the leading RM cell without waiting for an acknowledgment. If the required bandwidth is available at a network switch, it is reserved and the block is forwarded to the next hop. If the resources are not available, the block is discarded [I.371]. Hence, the source does not know if the block will actually reach the destination at the time of transmission.

  Both transfer modes allow the use of a *rigid/elastic* bit, which indicates that the Block Cell Rate (BCR) value in RM cells may be decreased by switches along the network path for ABT/DT, or the block may be forwarded to the next hop with a lower BCR for ABT/IT, respectively. Although ABT uses RM cells and offers a frame-based transmission, it is not directly related to the ABR or GFR Service Categories of the ATM Forum.

• The **Available Bit Rate** (ABR) transfer capability corresponds, apart from minor differences, e.g., in the RM cell format and the detailed specification of the end system behavior, to the ABR service class of the ATM Forum.

In contrast to the ATM Forum, where Service Categories imply traffic as well as QoS parameters, the ITU-T chose to define *QoS Classes* [I.356] and *ATCs* independently [I.371]. Table 2-6 suggests a mapping of ITU-T ATCs to QoS Classes and the ATM Forum Service Categories, where "U" stands for unspecified and "n.a." for non applicable [LSMA98], [TM4.1]. Because there is a flexible assignment of ATCs and QoS Classes, it is possible to combine the unspecified class with every ATC, if desired [I.356].

ITU-T ATC	ITU-T QoS Class				ATM Forum Service Category
	1	2	3	U	
DBR	x	(x)		(x)	CBR
SBR 1	x	(x)		(x)	rt-VBR
SBR 2		x	(x)	(x)	nrt-VBR
SBR 3		x	(x)	(x)	nrt-VBR
ABT (DT/IT)	x	(x)		(x)	n.a.
ABR			x	(x)	ABR
n.a.				x	UBR
n.a.			x	(x)	GFR

Table 2-6: *Mapping of ATCs, QoS Classes, and Service Categories*

It has to be noted, that the ITU-T is in the process of revising and updating its standards for ATM Layer cell transfer performance and B-ISDN traffic and congestion control.

2.8.4 Traffic Management Functions

The task of ATM traffic management is to avoid and resolve congestion situations, to assure certain performance and QoS guarantees for connections, and to provide an efficient use of network resources. The primary means to accomplish this task is the use of *Traffic Management Functions* to control and monitor traffic flows and congestion [I.371]. The most important of these functions are described in the following sections.

2.8.4.1 Connection Admission Control (CAC)

During connection setup the network nodes on the path from the source to the destination have to decide whether they have sufficient resources according to the requirements of the traffic contract of the new connection [McSp94]. The new connection is established only, if the requested QoS guarantees for the new connection can be provided and the ones of already existing connections are not violated [McDy00]. This decision is therefore based on the traffic

and QoS parameters of the new connection, its service class, and the available resources. The CAC may use additional information, like the measurement of the current network load, to make its decision [GiGa98]. In case the new connection is admitted, the required resources with respect to bandwidth, memory, CPU load, etc. have to be reserved by the network nodes. The implementation of CAC algorithms is not standardized neither by the ATM Forum nor the ITU-T and is hence vendor and network specific [GiGa98]. The CAC operation is correct, if it prevents the violation of QoS guarantees for all established connections.

2.8.4.2 Usage Parameter Control (UPC)

This function decides if a cell stream is compliant to the traffic characteristics specified in the traffic contract and may *discard* or *tag* ATM cells in the presence of congestion. Tagging of cells means that their *Cell Loss Priority* (CLP) is set from high to low (CLP = 1). These cells are then marked or *tagged* for discard. This function is essential for the control and monitoring of ATM traffic, because it prevents uncompliant flows from affecting the QoS of other connections [TM4.1]. It is located at the UNI (UPC) or at the NNI (*Network Parameter Control*, NPC). Whereas the ITU-T recommends the use of a UPC and defines the NPC as optional [I.371], the ATM Forum specifies the UPC/NPC as optional [TM4.1]. The actual implementation of the UPC/NPC functions has not been standardized neither by the ATM Forum nor the ITU-T. The conformance test defined at the UNI to determine if a cell stream is compliant to the traffic contract may also be used as a UPC/NPC function [GiGa98], [McDy00]. The *Generic Cell Rate Algorithm* (GCRA) is part of the traffic contract and is defined in the ITU-T and ATM Forum standards [TM4.1], [I.371].

The GCRA operates with two parameters, the *Increment* I, which corresponds to the average cell inter arrival time, and the *Limit* L, which determines the tolerance which a conformant cell may deviate from its estimated arrival time. Figure 2-15 shows the definition of the GCRA as a *Virtual Scheduling* and a *Continuous State Leaky Bucket* algorithm. Both models define the same behavior [TM4.1], [I.371].

The *Virtual Scheduling* algorithm calculates for each cell a *Theoretical Arrival Time* (TAT), which corresponds to the estimated arrival time of a conformant cell. If the cell arrives after this TAT, it is classified as compliant and a new TAT is calculated based on the current arrival time and the increment I. If the cell arrives prior to the TAT, but within an interval determined by the limit L, the cell is also defined as compliant. The new TAT is the old TAT plus the increment I. In all other cases the cell is not compliant and the TAT remains unchanged [Blac95a].

The *Continuous State Leaky Bucket* algorithm uses a buffer (queue) of size L + I, which is drained with a constant rate (leaky bucket). If the actual queue length is smaller than L upon arrival of a new cell, the cell is conformant and the queue length is increased by I units. If the queue length is greater than L, the cell is not compliant and may be discarded or tagged [Blac95a].

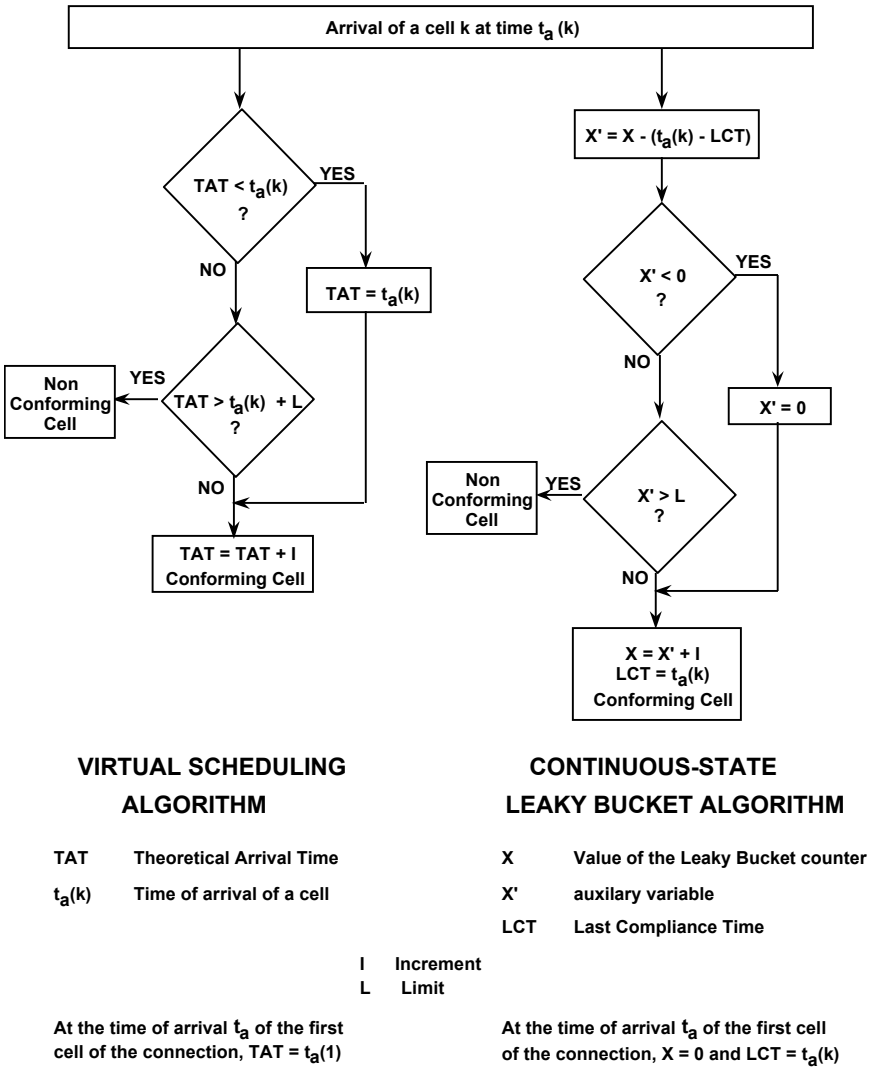

Figure 2-15: *Generic Cell Rate Algorithm [TM4.1]*

It is possible to combine two instances of the GCRA, called a *Dual Leaky Bucket Algorithm*, where non conforming cells of the first instance enter the second algorithm [GiGa98]. This method allows to control a VBR connection with respect to SCR and PCR. The cells enter at first a GCRA for the SCR, where the bucket depth corresponds to the Maximum Burst Size (MBS). Non-conforming cells are passed to the second GCRA, checking the PCR constraint, where the bucket depth relates to the Cell Delay Variation Tolerance (CDVT), see section 2.8.1. Furthermore, it is possible to use this model to control the PCR parameters of a connection for high priority cells at the first GCRA and the aggregated cell stream (CLP 0 + 1) at the second GCRA [McSp94].

In addition to this function, which is also called *traffic policing*, the UPC/NPC checks the VPI and VCI fields in the ATM cell header. If the values are invalid with respect to the corresponding connection, the cell is discarded [TM4.1].

2.8.4.3 Traffic Shaping

The ATM traffic management may use functions to deliberately change the traffic characteristics so that the flow conforms to the traffic contract and suits the actual network needs. This is called *traffic shaping* and possible actions, among others, are cell buffering, cell spacing, peak rate reduction, or limiting burst durations [McSp94]. The GCRA introduced above as a UPC/NPC function may also serve as a traffic shaping function. If the GCRA is used as a conformance test or as a UPC/NPC function, the cells inspected for conformance or policed do not actually enter a leaky bucket, but are rather checked against this theoretical model. Hence, the cell stream is not altered in this case. When the GCRA is used for traffic shaping, the cells are actually queued and this buffer is drained at a fixed rate [McDy00]. Thereby peaks in the traffic flow are smoothed at the expense of increasing the mean cell transfer delay. In accordance to the philosophy of the ATM Forum and the ITU-T, the actual implementation of this function is vendor specific [TM4.1], [I.371].

2.8.4.4 Network Resource Management

The use of *Virtual Paths* (VPs) can facilitate the management and allocation of network resources, e.g., bandwidth. The grouping of several *Virtual Channels* (VCs) with the same service class or QoS requirements into a single VP simplifies the traffic control, e.g., CAC or UPC [McSp94].

2.8.4.5 Selective Cell Discard

In case of a congestion situation a switch may selectively discard cells that have their CLP bit set and/or non-conforming cells with respect to the traffic parameters of the corresponding connection [TM4.1]. When tagged cells, that conform to the traffic contract of the connection, are discarded, the CLR specified for that connection has to be met. The idea of the selective cell discard, which is also called *priority control*, is to protect high priority (CLP = 0) cells or compliant connections during congestion situations [I.371]. The procedure of cell discard may be extended to include AAL information, like AAL PDU boundaries indicated in the cell header PTI field (e.g., EPD or PPD). For the Guaranteed Frame Rate (GFR) Service Category of the ATM Forum (see section 2.8.2), this feature is mandatory [TM4.1].

2.8.4.6 ABR Flow Control

As mentioned in section 2.8.2, the ABR Service Category needs a feedback mechanism, that allows the network switches to signal congestion situations to the sources, which adapt their transmission rate to the current network load. This ABR flow control mechanism is described in detail in chapter 3 of this thesis.

2.8.4.7 Other Functions

In addition to the functions mentioned above, the network management can take additional actions to manage and control traffic flows. In case of equipment and/or UPC/NPC faults, the affected connections should be isolated and/or released by the network management [I.371].

The intelligent choice of routing decisions may support the transfer performance of the network and lead to a more balanced network load, avoiding congestion. By the use of sophisticated cell buffering and scheduling algorithms at network switches (e.g., priority queueing), meeting of QoS requirements is facilitated [GiGa98].

2.9 Summary

This chapter provided an introduction to the Asynchronous Transfer Mode (ATM) as a switching and multiplexing technology. Its key features and area of applicability in Broadband Integrated Digital Networks (B-ISDN) were explained. The key players in ATM standardization, the ATM Forum and the ITU-T, and their standardization efforts were introduced. The connection-oriented transfer of fixed sized ATM cells lead to the concept of Virtual Channels (VCs) and Virtual Paths (VPs) as well as to the notion of cell switching inside the network nodes. The B-ISDN layered reference model was highlighted to present a structured view on the different tasks inside an ATM-based network. Starting at the Application Layer, the user data is processed at the ATM Adaptation Layer (AAL) leading to a cell stream at the ATM Layer. This cell stream may then be transported on the Physical Layer by different transmission media.

Since ATM was designed to support integrated networks with per connection, end-to-end QoS guarantees, ATM networks have to cope with heterogeneous traffic profiles. Hence, traffic management is of great importance for fulfilling the transfer performance requirements. The concept of Service Categories and QoS Classes was introduced, juxtaposing the traffic management specifications of the ATM Forum and the ITU-T. The different traffic management functions and their operational areas were introduced.

One task of the ATM traffic management is the use of congestion control mechanisms to avoid and relief overload situations in the network. This area and the specific problem of providing feedback information to control the traffic flow of Available Bit Rate (ABR) sources are covered in the following chapter.

CHAPTER 3

ABR Flow Control

In this chapter a more detailed description of the flow and congestion control mechanism for the Available Bit Rate Service Category of ATM networks will be given. First, an introduction to the area of congestion control for data networks will be presented. The goals and different techniques of congestion control will be highlighted. Next, an in-depth explanation of the ABR flow control mechanism as defined by the ATM Forum will be given, including the means for giving feedback from the network back to the ABR traffic source. The different functional specifications for the end systems and the switch behavior as well as the parameters of an ABR connection are presented. The interaction between the very common *Transmission Control Protocol* (TCP) and the ABR flow control mechanism is discussed. Furthermore, the design option of the virtual source / virtual destination technique will be explained. Finally, a short summary will conclude this chapter.

3.1 Congestion Control

A data network element is said to be in a congested state, if its traffic load is exceeding its available resources, i.e., the traffic demand is higher than the processing capability of the network element [BeGa92]. Therefore, congestion manifests itself in a shortage of network resources. These resources may be link capacity, buffer space, memory, processing power, or any other means that is necessary for handling, processing, or transmitting data [Kala97]. In this thesis we will focus on the usage of link capacity and buffer requirements of network switching nodes as the main resources to be controlled. The reasons for this choice are the general importance of these parameters for the characterization and dimensioning of network nodes and their independence of implementation details. Furthermore, this approach does not limit the analysis of network congestion, since other resources, like for example processing power, may be mapped to link capacity or buffer space of network nodes.

Overload situations may be categorized into *short term* or transient congestion and *long term* or persistent congestion. The average duration of a data connection is used as a time scale for this characterization. Depending on the duration of congestion, different traffic management mechanisms for dealing with these situations can be distinguished. For example, call admission control is an appropriate function to control long term congestion, due to overbooking of available bandwidth [Jain96a]. The congestion control mechanism of the ABR service class introduced in the next section handles short-time overload situations that are cleared during a time scale of a few tens of a second.

Another feature of congestion control algorithms is the goal to *avoid* or *relief* overload situations. For congestion avoidance, preventive actions are taken, if a resource shortage is foreseen. This is done for example by call admission control, if the establishment of a new connection, possibly leading to overload, is rejected. Congestion recovery aims at bringing the network element back to an uncongested state after an overload has occurred. However, most of the intelligent congestion control algorithms operating on network nodes, like the one introduced in chapter 5, incorporate both goals of avoidance and recovery and try to detect congestion as early as possible [Kaly00], [AfMO00], [RoBO95].

There are different modes of operation that can be distinguished for congestion control mechanisms, namely, *open-loop* and *closed-loop* [McDy00],[YaRe95], [KoRa97a]. With open-loop control the traffic source specifies its traffic characteristics and quality of service requirements before the beginning of the data transmission, like for example in a traffic contract as described in chapter 2.8.1. This approach is especially suitable for deterministic service classes, like ATM constant bit rate or variable bit rate [Kaly97]. These service classes have a constant traffic profile over the duration of the connection, which they are able to define. Network nodes may then reserve the necessary resources for the duration of the connection. Therefore, open-loop congestion control is mainly concerned with long-term overload situations and congestion avoidance. Closed-loop mechanisms on the other hand indicate the current network situation to the traffic sources, which must be able to adapt their transmission rates to the received feedback. This is an appropriate technique for low priority traffic classes, that are not able to fully describe their traffic and quality of service profile, but tolerate fluctuations in transfer capacity. The available bit rate service class in ATM networks is a typical candidate for this kind of closed-loop congestion control, which is elaborated in detail in the next section. For this reason, we will concentrate on closed-loop congestion control for the remainder of this chapter.

Closed-loop congestion control mechanisms may be executed on an *end-to-end* or on a *link-to-link* basis [YaRe95]. Whereas in the end-to-end case the network feedback is sent all the way back to the traffic source, in the link-to-link approach the network node indicates its current load situation to the preceding node on the connection path. If end-to-end congestion control is used, the source adjusts its traffic input to the network. With link-to-link control all network nodes must have the ability to control the traffic flow sent downstream.

Furthermore, an important design criterion for a closed-loop congestion control mechanism is what kind of feedback and how it is given. One can differentiate between *window-based* and *rate-based* feedback, which may be given *implicitly* or *explicitly* [Jain96a]. In the case of win-

dow- or credit-based feedback, the network status is indicated by an amount of data which will be accepted by the network node as input. This may be a "window" worth of bytes or "credits" corresponding to a number of data packets [MiKT96]. The other possibility is to give a rate-based feedback, which influences the rate at which new data is received at the input of the network element [Krish97]. The feedback information, either window- or rate-based, may be conveyed explicitly as a congestion information or may be implicitly derived from other network events, like for example from delayed acknowledgments [Kala97]. The use of an explicit congestion indication allows to convey the network status more accurately, but also requires the cooperation of the network nodes and leads to the use of additional overhead bandwidth.

Finally, one distinguishes between *positive* and *negative* closed-loop feedback control. When positive feedback control is used, the source is allowed to increase its transmission rate or the amount of data injected into the network only if it has received or implicitly detected that the network has indicated no congestion. If no information on the network status is available, the source has to reduce or at least is not allowed to increase its traffic load. On the contrary, if negative feedback control is employed, the source is forced to reduce its traffic load only if it detects a congestion situation. In the lack of any information the source may increase its traffic. Whereas negative feedback reduces the control overhead in the case of explicit congestion messages, positive feedback leads to a more conservative, less aggressive and therefore more robust control performance.

The taxonomy of congestion control introduced above will be illustrated for the special case of the ABR service class in section 3.2 of this chapter.

3.1.1 Performance Goals

There are several desired performance features of closed-loop congestion control algorithms, which will be introduced in the following [Kala97], [Kaly97], [Kara98], [VJGF99].

- **Efficiency**: The congestion control mechanism should lead to a high link utilization, where the available bandwidth is completely allocated to the connections. Furthermore, a low average and maximum buffer occupation at the intermediate network nodes is desirable, in order to avoid long queueing delays and large delay variance (jitter). The network node may however keep a "pocket full of cells" in its queues, to guarantee the use of short peaks in the available bandwidth. Therefore, there exists a trade-off between efficient use of available bandwidth and a low buffer occupancy [Jaff81], [SuVW00], [RoBe97], [TiSu90].
- **Fairness**: In general there are several connections controlled by the congestion control algorithm that share one or more common transmission links. Therefore, the need for a fair distribution of available bandwidth to competing connections arises. There have been developed several definitions and possibilities of a fair bandwidth allocation [AbKu97], [TM4.1], [BeGa92], [HoTP97a], [HoTP97b], [HTPK98b], [LoHR00], [Simc94], [Vand00]. A detailed discussion will be presented in section 5.3 in connection with the performance goals of the newly designed congestion control algorithm.

- **Stability**: Network conditions may vary in time, for example when new connections are admitted, existing connections are terminated or sources change their traffic load. Such events may result in short-term overload and it is the task of congestion control to drive the network back to a desired operating point. The congestion control algorithm is said to be stable, if this steady state is eventually reached from an arbitrary starting condition [AiHo99], [TKCT98]. Stability is needed to protect the network from permanently staying in an unwanted overload condition.
- **Responsiveness**: The time a congestion control algorithm needs to resolve an overload situation and to restore a normal network condition, should be as small as possible. A fast relief of congestion leads to improved network performance, since the duration of overload conditions is minimized.
- **Scalability**: The congestion control algorithm must scale with the number of controlled connections. Especially in wide area networks a network node must be able to handle a large number of connections simultaneously. This may be a problem for algorithms with explicit congestion indications that operate on a per-connection basis and maintain extensive state information for each traffic flow [ChRL96], [LoHR00].
- **Robustness**: The congestion control algorithm should operate correctly, even in the presence of misbehaving network elements or lost congestion information. The possibilities to isolate single non-conforming traffic flows depend also on the design of the network node, for example if per-connection queueing is used. Also the chances for the congestion control algorithm to protect the network from the effects of a severe malfunction are somewhat limited. Especially for congestion control mechanisms using explicit feedback, robustness also implies that the algorithm is able to cope with lost indications for a limited time period and is insensitive to slight mistuning of any control parameters. Furthermore, the presence of time-varying or unknown feedback delays may be a problem for the design of robust congestion control algorithms [AlBS98], [OKKI99], [Atas00], [ZhYM97a].
- **Conformity**: If there exists an already standardized framework in which the congestion control algorithm is to be executed, the algorithm has to comply to these standards to be interoperable with existing network elements. If for example there is assumed a certain end system behavior, an end-to-end based congestion control algorithm has to take this into account [TM4.1], [MoKl99].
- **Complexity**: An algorithm for avoiding and/or relieving overload situations has to be executed on network elements and possibly on the end systems involved. Therefore, the computational complexity and memory requirements of the algorithm have to match the available resources of the executing device [MaVG98].

After this introduction to congestion control mechanisms, we will focus on the special techniques used in ATM networks for the available bit rate service class. Hence, the network nodes will be ATM switches and the end systems will be the source and destination of an ABR connection (virtual circuit).

3.2 ABR Congestion Control

As already explained in chapter 2.8.2, ABR traffic has to use, together with the other best-effort traffic UBR and GFR, the remaining link bandwidth left over by the CBR and VBR service classes. As a consequence, ABR traffic sources have to cope with fluctuating capacities available for the ABR service class [ChLS96]. Therefore, a congestion control mechanism has to be used to adapt current ABR traffic loads to available link capacities. Figure 3-1 illustrates the bandwidth usage of the different ATM service classes, leading to a heterogeneous traffic profile.

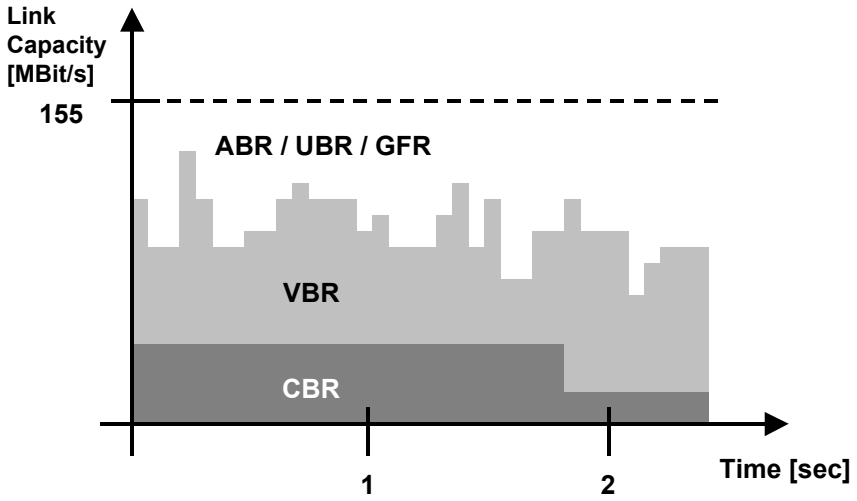

Figure 3-1: *Traffic Profile on an ATM Link*

Since the ABR traffic flow injected into the network has to be controlled, it has become common to speak of ABR flow control as a special case of congestion control [BeGa92]. Strictly speaking the term flow control refers to the fact that a traffic source may be throttled by the receiving side, in order not to cause an overload at the destination. Since the ABR source is throttled by the bottleneck link on the transmission path, the ABR congestion control mechanism may be interpreted as a kind of flow control to avoid network overloads. For these reasons, in the remainder of this thesis, we will use the two terms congestion and flow control interchangeably in the context of the ABR service class.

As mentioned in chapter 2.1, the ITU and the ATM Forum are the two bodies working on ATM standardization. Due to the slow progress of the ITU activities for the definition of an ABR congestion control mechanism and the greater practical importance of the ATM Forum standards in the area of the ABR traffic class, we will focus in this thesis on the traffic management specification of the ATM Forum [TM4.1]. However, there exists a liaison between the two bodies and it can be expected that the ITU standardization for the ABR service class follows closely the suggestions of the ATM Forum [LSMA98].

Inside the traffic management working group of the ATM Forum, which is responsible for drafting the ABR flow control, two different approaches for congestion control were discussed

[SiRo94]. One group advocated a window-based approach on a link-by-link basis, which is controlling the number of cells sent in one network hop. This approach was called *credit-based*, because each ATM switch indicates the amount of free buffer space (credits) for each ABR virtual circuit (VC) to its predecessor on the ABR connection path [CFKS97]. The upstream switch is then allowed to send at maximum a number of ABR cells equal to the number of credits received for the corresponding VC. At the time when the different possibilities for ABR flow control were discussed inside the ATM Forum in 1994, per-VC accounting of credits and per-VC queueing of ABR connections was considered a too complex implementation task for an ATM switch [SiTz94]. Another problem of the credit-based scheme is, that special synchronization procedures have to be employed in order to recover from lost credit update messages [KuBC94], [KuMo95]. Without these inter-switch synchronization protocols lost credit update messages result in a diminishing number of circulating credits and may eventually lead to a starvation of the traffic flow [OzSV94], [OzSV95]. On the one hand the credit-based technique yields fast response times in local networks, due to the short control loop between adjacent switches [RaNe95]. On the other hand, if a single ABR connection should be able to utilize the total link capacity, then the per-VC buffer size has to equal the bandwidth-delay product of the link [MoKu95]. This fact leads to high memory usage in high-speed, wide-area networks, where in addition the number of passing VCs is high. There were made several attempts to alleviate this problem by intelligent buffer management strategies, where many VCs share a common buffer pool [KuBC94], [KuMo95]. Finally, the traffic management group of the ATM Forum did not adopt the credit-based flow control approach, due to its complex implementation on the switch side and its limited performance in the field of wide area networks [SiTz94], [SiJa95].

Therefore, the ATM Forum agreed 1994 on a second approach for ABR flow control, an end-to-end, *rate-based* scheme with positive feedback [Fend96], [SmAT96]. The *Allowed Cell Rate* (ACR) of the ABR source is determined by a control loop between sender and receiver, passing through all ATM switches on the ABR connection path. The ABR source is allowed to increase its transmission rate only if it receives positive feedback from the network. If no or negative feedback is received, the ACR has to be reduced. For the duration of the ABR connection, a certain *Minimum Cell Rate* (MCR) is guaranteed to the source [BoFe95]. This *rate-based* algorithm is defined in the traffic management specification of the ATM Forum's traffic management group and is the de-facto standard for ABR flow control [TM4.1].

In the framework of the classification of congestion control algorithms given in the previous section, the two closed-loop approaches discussed by the ATM Forum can be characterized according to table 3-1.

ABR Flow Control	short term / long term	end-to-end / link-to-link	window-based / rate-based	implicit / explicit	positive / negative
credit-based	short-term	link-to-link	window-based	explicit	positive
rate-based	short-term	end-to-end	rate-based	explicit	positive

Table 3-1: *Classification of ABR Congestion Control*

A complete description of the ABR rate based flow control scheme is included in the traffic management specification of the ATM Forum [TM4.1]. The document fully specifies the ABR service class, by describing the RM cell structure, the switch and end system behavior, as well as the connection parameters [LRMS97], [LeRM98]. The following sections follow closely this standard definition and will give a detailed overview of the ABR flow control mechanism. In addition, the characteristics of transporting TCP traffic over the ABR service class are discussed.

3.2.1 Basic Operation

As mentioned above, the ABR flow control scheme of the ATM Forum operates as an end-to-end, rate-based algorithm, where the source receives feedback from the network through *Resource Management* (RM) cells. The source inserts these RM cells in the forward direction (also called *Forward Resource Management cells*, FRM) into the data cell stream of an ABR connection. The RM cells travel along the path through the ATM network and are received by the destination. The destination returns the RM cells (also called *Backward Resource Management Cells*, BRM) and sends them back to the sender [GiGa98]. Through this mechanism a feedback loop is established, reaching from the sender to the destination of an ABR connection and back, as illustrated in figure 3-2. In related works, this principle of ABR flow has been analyzed analytically, proving its correct operation [LRMS97], [LeRM98], [MoKl99].

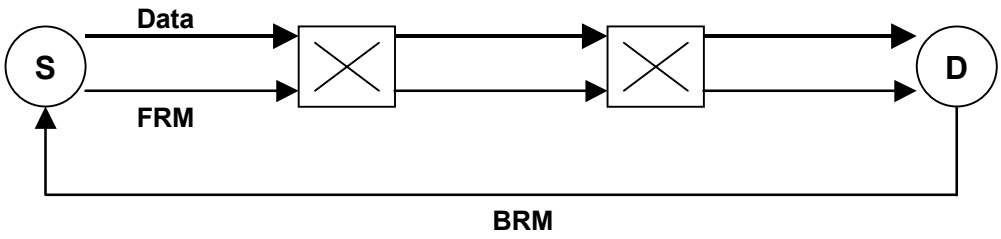

Figure 3-2: *ABR Closed-Loop Feedback Control*

RM cells are special ATM cells, which are recognized by a payload identifier (PTI) value in the ATM cell header, see chapter 2.6. The body of an RM cell contains different fields, as indicated in table 3-2, [TM4.1].

Field	Length	Description
Header	5 bytes	ATM header, PTI = "110" (bin)
ID	1 byte	Protocol Identifier: 1 for ABR service
DIR	1 bit	Direction of the RM-cell: 0 for forward RM-cells and 1 for backward RM-cells
BN	1 bit	Backward Notification: 1 for switch generated (BECN) RM-cells and 0 for source generated RM-cells
CI	1 bit	Congestion Indication: 1 for congestion and 0 otherwise
NI	1 bit	No Increase: 1 indicates no additive increase of ACR and 0 otherwise
RA	1 bit	Request/Acknowledge: used by [I371] or set to 0
Reserved	3 bit	For future use, set to 0
ER	2 bytes	Explicit Rate: limits the ACR of the source to a specific value
CCR	2 bytes	Current Cell Rate: indicates current cell rate of the source
MCR	2 bytes	Minimum Cell Rate: minimum cell rate that is desired by the source
QL	4 bytes	Queue Length: used by [I.371] or set to 0
SN	4 bytes	Sequence Number: used by [I.371] or set to 0
Reserved	30 bytes	For future use, set to "6A" (hex)
Reserved	6 bit	For future use, set to 0
CRC-10	10 bit	Cyclic Redundancy Check: checksum over the cell content

Table 3-2: *RM Cell Structure*

A short description of the various fields is given in the following [TM4.1].

- **Header**: The PTI value is set to "110" (bin), in order to mark the ATM cell as an RM cell. The VCI and VPI fields are set according to the ABR virtual channel connection (VCC) the RM cells belongs to. It is also possible to define ABR VPs; in this case a fixed VCI value of 6 is used. RM cells may be sent *in-rate*, which means that they are included in the ACR of the source and the CLP bit in the RM cell header is set to 0. On the other hand, RM cells could be sent *out-of-rate* at a fixed maximum rate, where the CLP bit is set to 1. Please refer to section 3.2.4 for details.

- **ID**: In accordance with the ITU, a protocol ID of 1 identifies the ABR service.
- **Message Type Field**: The second byte of the RM cell payload containing the fields DIR, BN, CI, NI, RA and Reserved is also called the Message Type Field.
- **DIR**: The DIR field indicates the direction in which the RM cell is sent along the network path. A forward RM cell is marked with DIR = 0 and travels with data cells from the source to the destination. When the RM cell is turned around by the destination, DIR is changed to 1. Hence, backward RM cells, travelling from the destination back to the source, are identified by DIR = 1.
- **BN**: In addition to the ABR sources it is also possible, that network switches or the destination generate an RM cell. The *Backward Notification* bit indicates whether the RM cell is a *Backward Explicit Congestion Notification* (BECN) cell, i.e., non-source generated (BN = 1) or source generated (BN = 0).
- **CI**: If a network switch detects that there is congestion in the network, it may set the *Explicit Forward Congestion Indication* (EFCI) bit in passing data cells. When an ABR destination turns around an RM cell it sets the *Congestion Indication* (CI) bit according to the EFCI state of the previous received data cell. The ABR source is analyzing the CI bit of a received BRM cell and hence gets information about the network condition. A CI value of 1 indicates congestion and the source is forced to lower its ACR, while a value of 0 means no congestion and the ABR source may increase its ACR.
- **NI**: A network element is allowed to set the *No Increase* (NI) bit to 1, in case of an expected overload, in order to prevent the ABR source from increasing its ACR. In general the source will initialize NI to 0, but it may set NI to 1, if it does not need to increase its current ACR.
- **RA**: The *Request/Acknowledge* (RA) bit is used in accordance with the ITU Standard I.371 for Broadband ISDN. The field is set to 0 if the ABR service of the ATM Forum is used.
- **ER**: The *Explicit Rate* (ER) may be set by the switches on the ABR network path from the source to the destination or by the destination itself to limit the ACR of the source. Each network element is only allowed to reduce the current ER value, so that the minimum supported rate on the network path is reported to the source. In each FRM cell the ER value is initialized by the source to a requested rate, smaller or equal to its *Peak Cell Rate* (PCR).
- **CCR**: The ABR source indicates its ACR in the *Current Cell Rate* (CCR) field of an FRM cell. This information may be used by ATM switches along the network path to the destination. For non-source generated RM (BECN) cells, the CCR value is set to 0.
- **MCR**: The *Minimum Cell Rate* (MCR) of the ABR connection is indicated in the MCR field of an RM cell. This information may be used by network nodes on the path to the receiver. As for the CCR field, it is set to 0 for BECN cells.
- **QL**: The *Queue Length* (QL) field is used for the ITU ABR service according to I.371. When the ABR service of the ATM Forum is used, this field is set to 0.

- **SN**: As for the other ITU specific fields, like RA and QL, the *Sequence Number* (SN) field is used only for the ITU ABR service. In case of the ATM Forum ABR service its value is set to 0.
- **CRC-10**: The *Cyclic Redundancy Check* (CRC) field contains a 10 bit checksum covering the RM-cell payload excluding the CRC field (374 bits), which is placed with the least significant bit right justified in the CRC field. Each bit of the payload is considered as a binary coefficient of a polynomial of degree 373 using the first bit as the coefficient of the highest order term. The CRC-10 generating polynomial is: $1 + x + x^4 + x^5 + x^9 + x^{10}$.

It should be noted that bi-directional ABR connections are also possible, which result in FRM and BRM cells travelling on the same network path.

3.2.2 Switch Behavior

From the above description it becomes clear, that the feedback given to the ABR source is *binary* through the CI and NI bits and/or *explicit* through the ER field of BRM cells [Stal99]. Consequently, for an ATM switch there are two major options of indicating a congestion situation to the source of an ABR connection, which are introduced in the following.

- **Binary feedback**: The switch may set the Explicit Forward Congestion Indication (EFCI) bit in data cells, when an overload situation is detected. In general a switch enters a congestion state, if a certain high threshold of the ABR output queue level is exceeded. The status is reset to normal, when the queue occupancy falls below a low threshold [McDy00]. The destination will observe the EFCI status in received data cells and in turn will set the CI bit in turned-around BRM cells [Jain96b]. This mode of operation is called *EFCI marking*. The advantage of this approach is its simple implementation at the involved ATM switches. It is not necessary for the switch to process any RM cells either in the forward or backward direction. The main drawback of EFCI marking is the long feedback delay of the congestion indication until it reaches the source [KoRa97b].

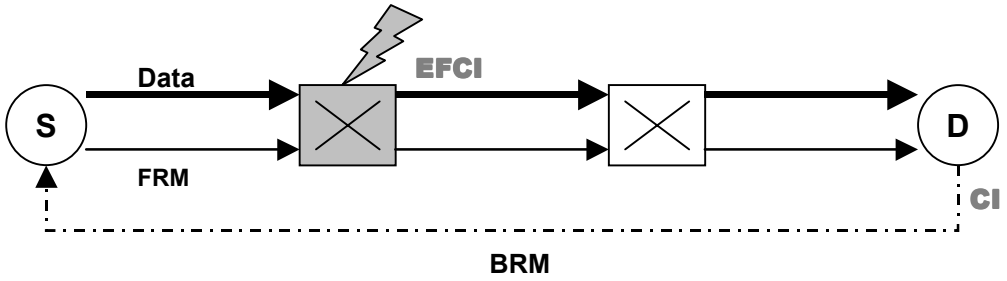

Figure 3-3: *EFCI Marking*

This effect is illustrated in figure 3-3, where the grey colored ATM switch experiences congestion and sets the EFCI bit. The overload information has to travel all the way from the point of congestion to the ABR destination and then via BRM cells back to the source. This feedback delay may become unacceptably large especially when the distance between

source and destination is large and the congestion occurs near the source, at the beginning of the network path. As a worst case, the feedback delay may almost reach the round trip time (RTT) of the ABR connection.

A way to improve the performance of the binary feedback scheme is to set the CI or NI bits directly in forward and/or backward RM cells. This method is called *Relative Rate Marking* and shortens the feedback loop, if BRM cells of the connection are marked by the overloaded switch [ABLM96]. It is also possible to additionally mark FRM cells to inform down stream switches and the receiver about the occurrence of congestion on the network path. The option of marking only FRM cells seems not desirable, since it offers no benefits compared to EFCI marking.

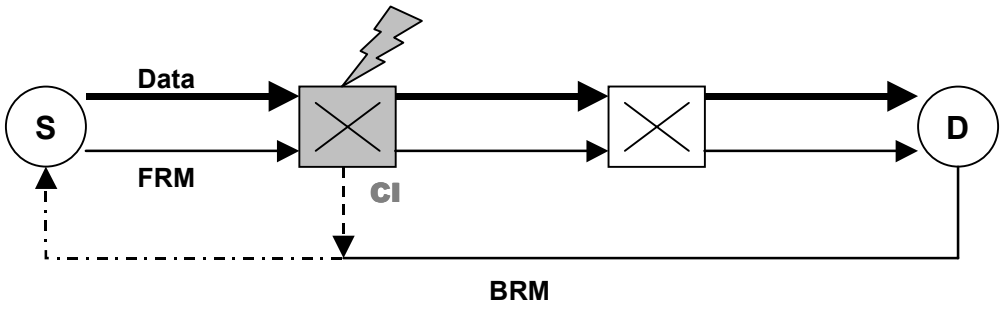

Figure 3-4: *Relative Rate Marking*

Furthermore, *Binary Enhanced Switches* (BES) can process the path information (VPI/VCI) of BRM cells to select certain ABR connections that will receive a congestion indication [Ohsa95d]. By this mechanism it is possible to throttle only specific connections with a current cell rate above the actual mean ACR of all connections, in the case of network overload. If nevertheless the ABR queue level reaches a very high threshold, a severe congestion is detected and BRM cells of all ABR connections are marked with CI = 1. Binary Enhanced Switches offer also a possibility to avoid the so-called *"beat-down problem"*. It leads to unfairly low rates for long ABR connections travelling through several EFCI switches in comparison to shorter VCs [KoRa97b]. The advantage of lowering the feedback delay and selectively tagging certain BRM cells comes along with the drawback of a more complex switch implementation than in the case of EFCI marking [Ohsa95a], [Ohsa95b].

- **Explicit rate**: In contrast to the parsimonious feedback information of the binary feedback scheme, the *Explicit Rate Marking* allows the switch to indicate a desired rate for the ABR source in the ER field of forward and/or backward RM cells [Stal99]. As with Relative Rate Marking, setting the ER field in BRM cells shortens the feedback loop and using the ER field in FRM cells may provide down stream switches and the destination with congestion information. Setting the ER field only in FRM cells is not an optimal strategy, because the feedback delay is then maximized.

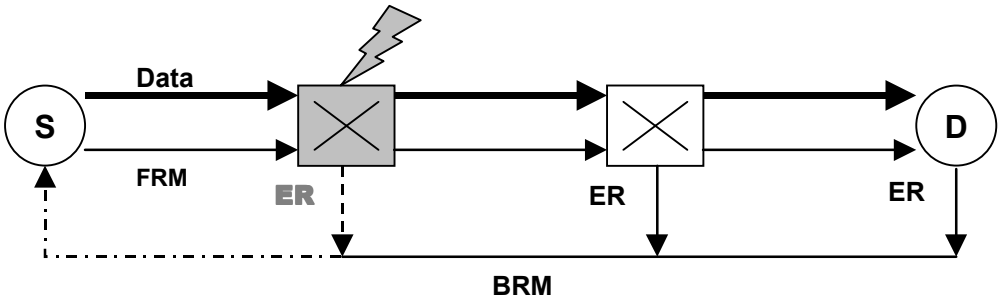

Figure 3-5: *Explicit Rate Marking*

As mentioned above, in the description of the RM fields, the switches along the ABR path are only allowed to reduce the value of the ER field [TM4.1]. Thus, the source of an ABR connection receives the minimum supported rate on the network path to the destination. The explicit rate mechanism is illustrated in figure 3-5, where the grey colored switch experiences a congestion and sets the ER field of BRM cells to the bottleneck rate. The major benefit of using Explicit Rate Marking is that the ABR source can be set to a specific transmission rate in one control step with the duration of the feedback delay. Whereas in case of binary feedback the source is only informed about the presence of an overload situation and has to adjust its transmission rates by possibly several reduction steps, the source may be slowed to the desired rate directly by using explicit rate feedback. This results in an improved performance for adjusting the source rates due to fewer oscillations, than for EFCI switches [CFKS97], [KoRa97b], [RaQi95], [Ritt97b]. On the other hand an explicit rate switch is even more complex than a relative rate marking switch, because in addition to the processing capabilities of RM cells, it needs to calculate a desired rate for each ABR connection traversing the switch.

Beside these two general schemes of indicating congestion by an ATM switch to an ABR source, the ATM Forum specified the option of segmenting the otherwise end-to-end control loop by special *Virtual Source / Virtual Destination* switches [Kaly98a], which are introduced in section 3.3.

According to the ATM Forum traffic management specification the ATM switch is allowed to generate BRM cells itself with a rate of up to 10 cells/ second. The switch has to set the CI and/ or NI fields of the RM cell to 1, so that this mechanism is to be used as an exceptional congestion indication. The other fields of the BRM cell have to be set according to table 3-2, e.g., BN = 1 [TM4.1]. The switch may issue such a non-source generated BRM cell immediately when congestion occurs, without waiting for the next regular BRM cell in the flow of an ABR connection.

Furthermore, the ATM switch is allowed to transmit RM cells with priority over ABR data cells, only the sequence of RM cells has to be maintained [TM4.1]. This means that if the ABR source emits RM cells with a fixed spacing in between data cells, see also section 3.2.3, this regular pattern might not be preserved after the first switch.

Another problem that can arise when using the ABR service, is the fact that a source might not be able to fully utilize its assigned ACR. This occurs when the source rate is limited by an external constraint, for example data I/O. If this constraint is revoked, the ABR source may suddenly reclaim its higher ACR, which would lead to a sudden load increase at the switch. For this reason, the switch is allowed to implement a so called "*use-it-or-loose-it*" policy to reduce the ACR of a connection whose actual transmission rate is below its assigned ACR [Kaly98b].

It should be noted that the ATM Forum standardized only the principle of ABR flow control and did not define any implementation specific details or algorithms. Especially it is not specified how or when an ATM switch will detect a congestion situation or how an explicit rate for an individual source is to be calculated. The ATM Forum traffic management specification guarantees that ATM switches using different ABR flow control algorithms are inter-operable [LRMS97], [LeRM98], although certain configurations may lead to slightly unfair rate allocations [PlSy97], [GoSS98].

3.2.3 End System Behavior

In addition to the switch operation, the ATM Forum standardized the behavior of ABR end systems [Jain96b]. This guarantees the interoperability between different end stations using the ABR service and allows the ATM switches on the path to expect a deterministic reaction of the ABR sources, e.g., to a congestion indication. The traffic management specification of the ATM Forum defines 13 source and 6 destination rules, that an ABR compliant end-system has to obey [TM4.1]. The ABR connection parameters used in the following end system rules are summarized and explained in more detail in section 3.2.4.

- **Source End System (SES)**: As mentioned in section 3.2.1, the term in-rate RM cell refers to RM cells with the cell loss priority (CLP) bit set to 0, whereas out-of-rate RM cells have CLP = 1. Furthermore, in-rate RM cells are included in the *Allowed Cell Rate* (ACR) of the source, out-of-rate RM cells are not. ABR data cells should never be tagged with CLP = 1 by the source.
  - **Source Rule 1**: *"The value of ACR shall never exceed PCR, nor shall it ever be less than MCR. The source shall never send in-rate cells at a rate exceeding ACR. The source may always send in-rate cells at a rate less than or equal to ACR."* [TM4.1]

    This defines the following relationships between CCR, MCR, ACR and PCR:

    $$(MCR \leq ACR \leq PCR) \wedge (CCR \leq ACR) \quad (3.1)$$

  - **Source Rule 2**: *"Before a source sends the first cell after connection setup, it shall set ACR to at most ICR. The first in-rate cell sent shall be a forward RM-cell."* [TM4.1]

    This starts the initial open-loop phase of the ABR flow control until the first BRM cell is received by the source [GiGa98]. The duration of this phase is minimized by starting the transmission with an FRM cell. In order to avoid congestion at the switches during this initial open loop phase, the source is allowed to send only with its *Initial Cell Rate* (ICR), which has been advertized to the switches at connection set-up.

- **Source Rule 3**: *"After the first in-rate forward RM-cell, in-rate cells shall be sent in the following order:*

  a) The next in-rate cell shall be a forward RM-cell if and only if, since the last in-rate forward RM-cell was sent, either:

  i) at least Mrm in-rate cells have been sent and at least Trm time has elapsed, or

  ii) Nrm-1 in-rate cells have been sent.

  b) The next in-rate cell shall be a backward RM-cell if condition (a) above is not met, if a backward RM-cell is waiting for transmission, and if either:

  i) no in-rate backward RM-cell has been sent since the last in-rate forward RM-cell, or

  ii) no data cell is waiting for transmission.

  c) The next in-rate cell sent shall be a data cell if neither condition (a) nor condition (b) above is met, and if a data cell is waiting for transmission." [TM4.1]

  This rule defines the order in which the three possible types of in-rate cells are sent by an ABR source. These three possible cell types are FRM, BRM and data, as mentioned in subsections a, b, and c, respectively. It should be noted, that the turn-around of an FRM cell is realized by emitting a BRM cell from the source according to the above rule. Therefore, every ABR destination has to implement in addition to the receiver side also the source end system rules.

  Part a of the source end system rule 2 defines the frequency at which FRM cells will be sent. In general the source inserts an FRM cell every *Nrm* - 1 data cells (ii). The default value for *Nrm* is 32, which guarantees a fixed control overhead due to FRM cells of 1 ⁄ 32 or approximately 3 % of the ABR bandwidth. If the BRM cells on the return path are also taken into account, the total overhead for RM cells is about 6 %. This overhead is independent of the number of ABR connections.

  Because of this mechanism for sending FRM cells, sources with a low ACR suffer from large inter FRM cell intervals and hence a low update rate for their feedback information. For this reason, part (i) of source rule 2a defines with *Trm* an upper bound for the time between two consecutive FRM cells. The default value for *Trm* is 100 milliseconds. This poses the new problem, that connections with a very low traffic rate (under 10 cells ⁄ second) will be dominated by RM cells and will not be able to send any data cells. Therefore, in addition to the time-out value *Trm* the parameter *Mrm* has been defined. It guarantees that at least *Mrm* (default value = 2) data cells are sent between two consecutive FRM cells. As indicated in the previous section 3.2.2, the spacing of RM cells between data cells might be altered by switches implementing a priority mechanism for the transmission of RM cells, see section 3.2.2.

- **Source Rule 4**: *"Cells sent in accordance with source behaviors #1, #2, and #3 shall have CLP = 0."* [TM4.1]

  ABR data cells are always sent as in-rate cells, only RM cells may be tagged by CLP = 1 as out-of-rate.

- **Source Rule 5**: *"Before sending a forward in-rate RM-cell, if ACR > ICR and the time T that has elapsed since the last in-rate forward RM-cell was sent is greater than ADTF, then ACR shall be reduced to ICR."* [TM4.1]

  This rule implements a use-it-or-loose-it policy at the ABR source. If the source is idle and the time between two consecutive FRM cells exceeds the *ACR Decrease Time Factor* (ADTF), its ACR is reduced to ICR. The default value for the ADTF is 0.5 sec.

- **Source Rule 6**: *"Before sending an in-rate forward RM-cell, and after following behavior #5 above, if at least CRM in-rate forward RM-cells have been sent since the last backward RM-cell with BN = 0 was received, then ACR shall be reduced by at least ACR · CDF, unless that reduction would result in a rate below MCR, in which case ACR shall be set to MCR."* [TM4.1]

  In case there is no feedback information received from the network, a congestion situation is assumed. The *Missing RM-cell Count* (CRM) defines the number of FRM cells that may be sent out without an acknowledgment (a BRM cell returned by the destination), before the ACR has to be reduced. The *Cutoff Decrease Factor* (CDF) determines the amount of the rate reduction, while the MCR is maintained. Hence, source rule no. 6 implements the feature of a *positive* congestion control.

- **Source Rule 7**: *"After following behaviors #5 and #6 above, the ACR value shall be placed in the CCR field of the outgoing forward RM-cell, but only in-rate cells sent after the outgoing forward RM-cell need to follow the new rate."* [TM4.1]

  The CCR field of FRM cells carries the actual ACR of the ABR source. It is however possible that the cell rate of an ABR connection varies throughout the network path, due to queueing and/or congestion. In this case the CCR experienced by the switches along the path for an individual ABR VC may differ from the value of the CCR field in FRM cells [Kaly97].

- **Source Rule 8**: *"When a backward RM-cell (in-rate or out-of-rate) is received with CI = 1, then ACR shall be reduced by at least ACR · RDF, unless that reduction would result in a rate below MCR, in which case ACR shall be set to MCR. If the backward RM-cell has both CI = 0 and NI = 0, then the ACR may be increased by no more than RIF · PCR, to a rate not greater than PCR. If the backward RM-cell has NI = 1, the ACR shall not be increased."* [TM4.1]

  This rule defines the source reaction to *positive* (CI = 0 and NI = 0, no congestion detected) or *negative* (CI = 1, congestion detected) binary feedback from the network. In the absence of congestion, the source is allowed to increase its rate by an additive amount, determined by the *Rate Increase Factor* (RIF) and the PCR of the connection. If an overload condition is signaled to the source, the *Rate Decrease Factor* (RDF) defines a multiplicative quantity by which the ACR is decreased. When the switches along the ABR network path have set the NI bit, the source will not raise its current transmission rate. In accordance to source rule 1, the ACR is adjusted to remain within the limits of MCR and PCR respectively.

It was shown that a multiplicative decrease and an additive increase yields a fast convergence near the optimal ACR [ChJa89]. It has to be noted however, that the RIF and RDF in general are chosen to obtain a good transient behavior. That means that the ACR is reduced and retained fast in a few update operations. As a result, binary feedback leads to oscillations of the ACR in steady state, since the desired optimal ACR may not be reached exactly [KoRa97b], [Ritt97a].

- **Source Rule 9**: *"When a backward RM-cell (in-rate or out-of-rate) is received, and after ACR is adjusted according to source behavior #8, ACR is set to at most the minimum of ACR as computed in source behavior #8, and the ER field, but no lower than MCR."* [TM4.1]

  Switches along the ABR network path may use the Explicit Rate field of BRM cells to convey the minimum supported rate to the source, as explained in section 3.2.2. The ABR source has therefore to take into account this bottleneck rate by limiting its ACR to ER. Since the ER field of FRM cells is initialized by the source to its desired rate (see section 3.2.1), source rule 9 will impose no constraints on the ACR in the absence of explicit rate marking switches.

- **Source Rule 10**: *"When generating a forward RM-cell, the source shall assign values to the various RM-cell fields as specified for source-generated cells [...]."* [TM4.1]

  The fields of an FRM cell generated by the ABR source have to be initialized according to table 3-2.

- **Source Rule 11**: *"Forward RM-cells may be sent out-of-rate (i.e., not conforming to the current ACR). Out-of-rate forward RM-cells shall not be sent at a rate greater than TCR."* [TM4.1]

  An ABR source may emit FRM cells at arbitrary time instants. They will however not be accounted for the current ACR of the source and are therefore tagged as out-or-rate FRM cells (CLP = 1). The parameter *Trm*, which was defined in source rule 3, applies also to these out-of-rate FRM cells and limits the rate with which they are sent out by the ABR source.

- **Source Rule 12**: *"A source shall reset EFCI on every data cell it sends."* [TM4.1]

  As described in section 3.2.2 on switch behavior, binary marking switches will set the EFCI bit in data cells in case of a congestion situation. Therefore, the source has to initialize the EFCI value to its default value of 0 (no congestion).

- **Source Rule 13**: *"The source may implement a use-it-or-lose-it policy to reduce its ACR to a value that approximates the actual cell transmission rate."* [TM4.1]

  In addition to source rule 5, which possibly triggers a reduction of ACR to ICR at the time an FRM cell is sent, the source may optionally adjust its ACR to its actual transmission rate. The algorithms for these source-based use-it-or-lose-it mechanisms are implementation-dependent [Kaly98b]. The goal of this policy is to prevent ACR retention, as described in section 3.2.2.

Besides these source rules the traffic management specification of the ATM Forum [TM4.1] defines 6 rules for ABR destinations, which are listed below.

- **Destination End System (DES)**: In accordance to the source end system rules, cells with a CLP of 0 are referred to as in-rate, whereas cells with a CLP of 1 are referred to as out-of-rate.
  - **Destination Rule 1**: *"When a data cell is received, its EFCI indicator is saved as the EFCI state of the connection."* [TM4.1]

    In binary mode operation, the EFCI bit inside the header of data cells is set by the congested switches along the forward path. In order to convey this information back to the source, the destination has to store the EFCI status to correctly set the CI bit in BRM cells, see destination rule 2.
  - **Destination Rule 2**: *"On receiving a forward RM-cell, the destination shall turn around the cell to return to the source. The DIR bit in the RM-cell shall be changed from "forward" to "backward", BN shall be set to zero, and the CCR, MCR, ER, CI, and NI fields in the RM-cell shall be unchanged except:*

    a) If the saved EFCI state is set, then the destination shall set CI=1 in the RM-cell, and the saved EFCI state shall be reset. It is preferred that this step is performed as close to the transmission time as possible;

    b) The destination (having internal congestion) may reduce ER to whatever rate it can support and/or set CI=1 or NI=1. A destination shall either set the QL and SN fields to zero, preserve these fields, or set them in accordance with ITU-T Recommendation I.371. [...]" [TM4.1]

    Received FRM cells are returned to the source as BRM cells, i.e., the DIR bit is changed from 0 to 1. The CI field is set according to the EFCI status of the ABR connection, see destination rule 1. The destination end system has the possibility to throttle the ABR source, either by binary feedback (setting the CI and/or NI bits) or by requesting an explicit rate (setting the ER field). Hence, the BRM cells of the feedback loop serve both as a congestion control mechanism and as a "true" flow control scheme for controlling the source by the destination.
  - **Destination Rule 3**: *"If a forward RM-cell is received by the destination while another turned-around RM-cell (on the same connection) is scheduled for in-rate transmission:*

    a) It is recommended that the contents of the old cell are overwritten by the contents of the new cell;

    b) It is recommended that the old cell (after possibly having been over-written) shall be sent out-of-rate; alternatively the old cell may be discarded or remain scheduled for in-rate transmission;

    c) It is required that the new cell be scheduled for in-rate transmission." [TM4.1]

    In case of several RM cells waiting at the destination for turn-around, the most recent information should be sent out first. This is done as an in-rate cell, to avoid a possible discarding of the cell.

- **Destination Rule 4**: *"Regardless of the alternatives chosen in destination behavior #3 above, the contents of an older cell shall not be transmitted after the contents of a newer cell have been transmitted."* [TM4.1]

  Recent information returned by BRM cells should not be replaced by out-dated information sent thereafter.

- **Destination Rule 5**: *"A destination may generate a backward RM-cell without having received a forward RM-cell. The rate of these backward RM-cells (including both in-rate and out-of-rate) shall be limited to 10 cells / second, per connection. When a destination generates an RM-cell it shall set either CI=1 or NI=1, shall set BN=1, and shall set the direction to backward. The destination shall assign values to the various RM-cell fields as specified for destination generated cells [...]."* [TM4.1]

  Like ATM switches, see section 3.2.2, ABR destination end systems are allowed to generate BRM cells in order to signal an exceptional congestion situation. The rate of these end system generated cells is limited to 10 cells / second to limit their load on the network.

- **Destination Rule 6**: *"When a forward RM-cell with CLP=1 is turned around it may be sent in-rate (with CLP=0) or out-of-rate (with CLP=1)."* [TM4.1]

  Out-of-rate FRM cells may change their status to in-rate.

The rules introduced above completely define the behavior of an ABR end system. The closed-loop feedback flow control mechanism defined by the ATM Forum relies on the fact, that participating end systems are compliant to these rules and behave as expected. Any misbehaving ABR sources may result in persistent congestion at network switches or even performance degradation of other ABR connections. This deteriorating effect on other ABR connections can be limited by the use of per-VC queueing at the intermediate ATM switches, since traffic of different connections is buffered in separate queues.

3.2.4 ABR Connection Parameters

At connection set-up time, when a new ABR VC is established, the ATM signaling protocol carries a set-up message from the source to the destination and back, along the network path [PNNI1.0], [UNI4.0], [NNIABR], [UNIABR]. This message contains ABR connection parameters, which in part must or may be negotiated between the source and the network elements, i.e., ATM switches and the destination end system [TM4.1]. Table 3-3 presents a summary of all ABR connection parameters of the flow control mechanism used by the end systems, the switches, and the signaling protocol.

Parameter	Description	Value	Default Value
PCR	Peak Cell Rate	0 .. 4 billion cells/s	-
MCR	Minimum Cell Rate	0 .. 4 billion cells/s	0
ACR	Allowed Cell Rate	0 .. 4 billion cells/s	-
ICR	Initial Cell Rate	0 .. 4 billion cells/s	PCR
TCR	Tagged Cell Rate (CLP = 1)	constant	10 cells/s
Nrm	Number of data cells between FRM cells	2 .. 256	32
Mrm	Controls bandwidth allocation between FRM, BRM, and data cells	constant	2
Trm	Upper bound on inter-FRM time	$100 \cdot (2^{-7})$ .. $(100 \cdot 2^{0})$ ms	100 ms
RIF	Rate Increase Factor, determines the maximum allowed increase of the sending rate, in one step	2^{-15} .. 1	1
RDF	Rate Decrease Factor, determines the decrease of source rate, if the network is congested	2^{-15} .. 1	1/32768
ADTF	ACR Decrease Time Factor, sets the timeout interval between FRM cells	0.01 .. 10.23 s with granularity of 10 ms	500 ms
TBE	Transient Buffer Exposure, determines the maximum number of cells that may be sent before the source receives the first BRM cell	0 .. 16,777,215 cells	16,777,215
CRM	Missing RM cell Count, determines the maximum number of FRM without receiving a BRM cell	integer	-
CDF	Cutoff Decrease Factor, defines the reduction of ACR with respect to CRM	0 or 2^{-6} .. 1	1/16
FRTT	Fixed Round-Trip Time, minimum RTT	0 .. 16,777,215 µs	-

Table 3-3: *ABR Connection Parameters*

In the following, each connection parameter will be introduced briefly [TM4.1].

- **PCR**: This rate may never be exceeded by the source. The source sets this value to its maximum transmission rate. PCR is negotiated downwards during connection set-up, to reflect the minimum value supported along the path.
- **MCR**: Minimum transmission rate that will be guaranteed for the duration of the ABR connection by the network and at which the source is always allowed to send. If no minimum rate is needed, the default value is set to zero. If in addition to MCR a tolerable absolute minimum value (MCRmin) is signaled by the source, then MCR is negotiable down to MCRmin, otherwise it is not [TM4.1]. Switches along the path will use the MCR value in their call admission control (CAC) procedure, because resources for the MCR of an ABR VC must always be provided.
- **ACR**: Rate at which the source is allowed to transmit cells. ACR is computed by the source in response to congestion information conveyed through RM cells. This parameter is not used by the signaling protocol.
- **ICR**: Rate at which the source may send after the successful connection set-up, before the first RM cell has returned (open-loop phase), and after an idle period. The default value is the PCR of the connection and ICR is negotiated downwards during connection set-up to reflect the minimum value supported along the path. ICR is updated by the source after the call setup is complete, in accordance with TBE (see below) as [TM4.1]:

$$ICR = max\left(MCR, min\left(ICR, \frac{TBE}{FRTT}\right)\right) \tag{3.2}$$

- **TCR**: The Tagged Cell Rate defines the rate at which a source may sent out-of-rate RM cells, with a CLP of 1. Its value is a constant of 10 cells / sec and is not changed by signaling.
- **Nrm**: Defines the maximum number of cells a source may send for each forward RM cell. This value may be optionally specified by the source during call setup, but is not negotiated. If not specified, the default value of 32 is used.
- **Mrm**: Assures that sources with a low transmission rate (< 10 cells / sec, see also parameter Trm) send at least 2 data cells between consecutive FRM cells, as described in section 3.2.3 (source rule no. 3).
- **Trm**: Defines an upper bound on the time between two consecutive FRM cells for an active source. This value may be optionally specified by the source during call setup, but is not negotiated. If not specified, the default value of 100 ms is used.
- **RIF**: Determines an additive increase ($RIF \cdot PCR$) of the allowed transmission rate at the source, after a BRM cell is received, see source rule no. 8 (section 3.2.3). RIF is negotiated downwards during connection set-up to reflect the minimum value supported along the path.
- **RDF**: Defines a multiplicative decrease ($RDF \cdot ACR$) of the ACR at the source, see source rule no. 8 (section 3.2.3). During negotiation the value of RDF may be altered, taking into account that the ratio RDF / RIF shall not be decreased. Consequently, if RIF is decreased by a factor k, RDF may be decreased by at most a factor k, or it may be increased [TM4.1].

- **ADTF**: Time-out value between two FRM cells, after which ACR is reset to ICR, see source rule no. 5 (section 3.2.3). This value may be optionally specified by the source during call setup and is negotiated downwards. If it is not specified, a default value of 500 ms is assumed.
- **TBE**: Limits the number of cells injected into the network during the open-loop phase after a successful connection set-up and the return of the first BRM cell at the source. TBE is negotiated downwards during connection set-up to reflect the minimum value supported along the path.
- **CRM**: Limits the number of FRM cells that may be sent by the source without receiving any BRM cells. CRM is an integer value not used by signaling and is calculated by the source after connection set-up as: CRM = $\left\lceil \frac{TBE}{Nrm} \right\rceil$, where $\lceil X \rceil$ is the smallest integer greater than or equal to X.
- **CDF**: Defines a decrease of ACR by $ACR \cdot CDF$, if CRM FRM cells have been sent without receiving a BRM cell, see source rule no. 6 (section 3.2.3). This value may be optionally specified by the source during call setup and is negotiated upwards. If it is not specified, a default value of 1/16 is assumed.
- **FRTT**: The Fixed Round-Trip Time is set by the source to its fixed delay. FRTT is then accumulated as a signaling parameter during the call setup and is the sum of all fixed delays plus propagation delays in the round trip call path experienced by transmitted ATM cells. This value does not include any possible queueing delays during the actual transmission.

In case of a mixed environment of ER and EFCI switches, rate in- or decrease is controlled by the RIF and RDF parameters, constrained by the ER value of BRM cells. In a pure ER environment, i.e., only ER switches are present along the network path, it is possible to adjust the ACR of the source directly by the ER value of BRM cells. For that purpose, it is necessary that the source sets RIF and RDF to their default values, namely, 1 and 1/32768, respectively.

3.2.5 Interaction with TCP

A very common transport protocol is the *Transmission Control Protocol* (TCP) [Jaco88]. In combination with the *Internet Protocol* (IP) it is widely deployed and supports a broad range of Internet applications. TCP as a transport protocol provides a connection-oriented, reliable, byte stream data transport to the user [Stev94]. For this, it uses a flow control mechanism which uses, according to the taxonomy introduced in section 3.1, a short-term, closed-loop, end-to-end, window-based, implicit, and positive algorithm.

The TCP sender forwards application data in TCP segments to the receiving station and is at every point in time allowed to send a certain number of bytes. This number is determined by the minimum of two window sizes, namely the congestion window (*cwnd*), maintained by the sender, and the advertized window (*wrcvr*) indicated by the receiver [Jaco88]. Whereas the advertized window flow controls the sender to avoid an overflow of the receiver, the congestion window limits the amount of traffic injected into the network to avoid congestion. A network congestion is inferred from segment loss, which is detected by the sender with the help of a

retransmission timer. When the retransmission timer expires without the reception of an acknowledgment, the sender times out and retransmits the corresponding segment [Stev94].

The algorithm for the adjustment of the congestion window operates in two phases *slow start* and *congestion avoidance*, determined by a threshold value (*ssthresh*) [Jaco88]. During the slow start phase the congestion window *cwnd* is set to one and is increased by one segment for each acknowledgment received. Despite its name slow start, the congestion window grows exponentially during this phase. Whenever the retransmission timer expires, the threshold variable is set to half of the current window size, or more exactly

$$ssthresh = \max \{2, \min \{cwnd / 2, wrcvr\}\}, \tag{3.3}$$

and the congestion window *cwnd* is set to one. The sender enters slow start until the window size reaches the threshold variable *ssthresh*. At this point the sender enters the congestion avoidance phase and the congestion window is increased only by 1 / *cwnd* for every received acknowledgment. That means that the sender increases its congestion window linearly by one packet each time it receives acknowledgments for a full congestion window of packets [Jaco88].

This basic algorithm was improved and enhanced by newer TCP implementations adding features like:

- fast retransmit (TCP Tahoe), i.e. retransmission is triggered by the reception of three or more *duplicate acknowledgments* [RFC2001], indicating that a segment was received out of order,
- fast recovery (TCP Reno), after the fast retransmit was performed, the sender enters directly the congestion avoidance phase rather than slow start [RFC2001],
- selective acknowledgments (SACK TCP) [FaFl96],
- new retransmission algorithm, modified congestion avoidance and slow start algorithms (TCP Vegas) [BrPe95].

Since TCP/IP is a very common and widely used protocol combination, its transport over ATM networks has been intensively studied. The link utilization of TCP connections over high speed, low delay ATM networks may be low, due to two implementation details, namely the coarse granularity of the retransmission timer (100 - 500 ms) and the limited window size of 64 kbyte. These problems may be avoided by using a finer grained clock (0.1 ms) [RoFl95] and the window scale option, increasing the maximum window size in correspondence to the bandwidth-delay product of the link [Stev94].

Furthermore, the choice of the ABR Service Category as defined in chapter 2.8.2 influences the transport of TCP traffic over ATM networks. The use of the Unspecified Bit Rate (UBR) category results in a poor effective throughput on TCP level, as mentioned in chapter 2.8.2. Adding special buffer discard mechanisms [RoFl95] and queueing techniques [TzSi98] leads to an improved performance. The Guaranteed Frame Rate (GFR) class with its frame-oriented transport and guaranteed minimum cell rate offers a simple and well-suited ATM Service Category for TCP/IP traffic. The required traffic parameters like the maximum frame and burst sizes may be defined for single, end-to-end TCP connections.

When using the Available Bit Rate (ABR) service for TCP, two separate flow control algorithms are operating in parallel: the TCP window-based flow control on the transport layer and the ABR rate-based flow control on the ATM layer. Since both algorithms are not coupled and ABR rate information is not conveyed to the TCP layer, their interaction is critical and could have unforeseen effects on TCP throughput. Related works show that a good performance of TCP over ABR is possible for carefully tuned ABR connection parameters and persistent (or greedy) TCP sources [Kaly96], [KJFG98]. When non-persistent sources and interfering high priority traffic are assumed, throughput on TCP level might be negatively influenced [HoSu99].

For these reasons, the performance evaluation of the newly developed ABR flow control algorithm, presented in chapter 6, will take into account the transport of aggregated TCP/IP traffic over ABR using the ABR self-similar source model. The ABR Service Category is then used for certain parts of the TCP network path, rather than for individual, end-to-end TCP connections, which are better served by the GFR Service Category. As a consequence, the aggregated TCP connections will experience the ABR rate adjustments, e.g., on an access link to the Internet, as fluctuations of the network capacity, common to a best effort service.

3.3 Flow Control using Virtual Source / Virtual Destination

As mentioned in section 3.2.2, the ATM Forum traffic management specification [TM4.1] offers an optional flow control feature for an ATM switch, namely, the virtual source / virtual destination (VS/VD) design. In this section the operation and properties of the VS/VD switch will be introduced in more detail, since this design is assumed for the flow control algorithm developed in this thesis.

3.3.1 Mode of Operation

The VS/VD switch splits the otherwise end-to-end control loop of a single ABR connection into two separate parts, the *previous* or *upstream* loop and the *next* or *downstream* loop. The VS/VD switch therefore acts as a virtual destination for the previous loop and as a virtual source for the next loop [BoFe95]. Received FRM cells from the upstream loop are turned around as BRM cells by the VS/VD switch, which in turn generates new FRM cells for the downstream loop and receives them as BRM cells. Data cells are forwarded to the downstream loop according to the network path [TM4.1].

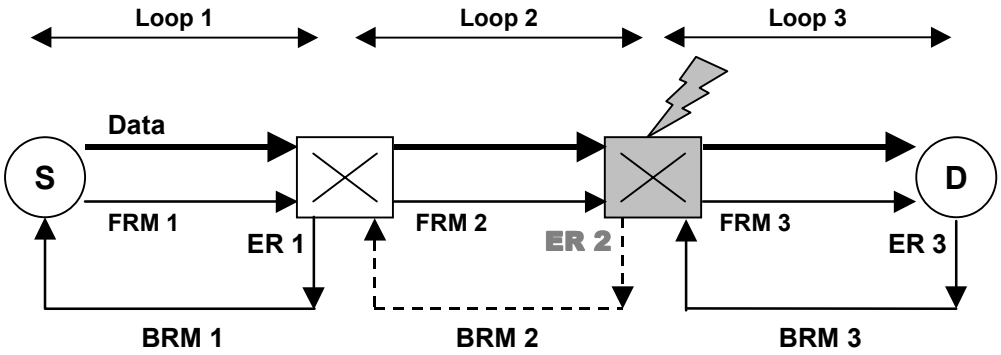

Figure 3-6: *VS/VD Control Loops*

Figure 3-6 depicts an ABR connection with 2 VS/VD switches and 3 control loops. Each of the 3 feedback loops has its own RM cell flow, with FRM cells generated by the source (loop 1) or the virtual source of a VS/VD switch (loops 2 and 3). BRM cells are turned around by the destination (loop 3) or the virtual destination of a VS/VD switch (loops 1 and 2). In case of a congestion, e.g., at the second switch marked with grey color, a reduced explicit rate (ER 2) is requested from the virtual source of loop 2 at the first VS/VD switch.

The ATM Forum traffic management specification defines, that a VS/VD switch has to implement the ABR end system rules, as described in section 3.2.3. Each ABR control segment, except the first, is fed by a virtual source, which behaves like an ABR source end system. BRM cells received by a virtual source are used to calculate the ACR and are then discarded. Each ABR control segment, except the last, is terminated by a virtual destination, which behaves like an ABR destination end system. Received FRM cells are turned around as BRM cells according to destination rule no. 2 and are not forwarded to the next segment. The MCR value of the connection shall not be changed and conveyed to the following segment. Other ABR connection parameters may be altered by a VS/VD switch [TM4.1]. Any other details, for example how the two feedback loops are coupled, i.e., when or if any congestion information of the next loop is conveyed to the previous loop, are implementation-specific and are not defined. Each individual ABR connection (VC) passing through a VS/VD switch may operate at a different sending rate, according to the ABR control scheme. In addition each VC may experience a different current cell rate (CCR) in the previous and the next loop. Therefore, per-VC queueing at the VS/VD switch is required for cell buffering and to enforce individual sending rates for each VC [Goya97].

3.3.2 Areas of Application

The main advantages of the VS/VD switch are the reduction of the feedback delay by shortening the control loop and the control of individual ABR VCs at administrative network boundaries [BoFe95]. VS/VD switches may be deployed at the network-network-interface when operators do not want to rely on the correct behavior of ABR sources outside their administrative domain. Since the ABR flow injected into the downstream network segment is controlled by the virtual source of the VS/VD switch, the network operator is able to directly control this

traffic load. Any misbehaving ABR sources will affect only the corresponding per-VC queue of the first VS/VD switch along the path [Jain96a], [McDy00], [Sait96]. Furthermore, VS/VD switches may be used at the boundaries of sub-networks with special transmission characteristics, like radio- or digital subscriber line (DSL)-access networks [NuMa98], [SmAT96], [Kwok99]. In addition, VS/VD switches could be used to transport non-ABR, best effort traffic over an ABR segment along the network path [KKON99], [RaNe95]. In the case of wide area networks VS/VD switches are able to divide the otherwise long end-to-end feedback loop into smaller segments with shorter feedback delay, improving bandwidth usage and lowering peak buffer occupancy [BoFe95], [PeJa99]. The ATM Forum addressed in its traffic management specification also the point of extending the point-to-point oriented ABR service to a point-to-multipoint approach [TM4.1]. For ABR multicast traffic, the problem of consolidating ABR feedback information (BRM cells) at branch points of the multicast tree arises [TzSi97]. One possible solution is the use of VS/VD switches at these locations [ChCh00].

Besides these operational advantages of the VS/VD option, the need of implementing individual end system behavior and buffering for each VC leads to a complex switch design and implementation. This can be considered a major drawback of the VS/VD option [Kaly98a].

Since the VS/VD switch behaves as an ordinary ABR end system and in no way alters the ABR flow control scheme, it may be used in conjunction with other ABR switches (explicit rate or EFCI) along the path. The virtual destination will turn around all FRM cells received and the virtual source reacts on all congestion indications of the downstream loop. Any other ABR switches on the path do not detect the presence of a VS/VD switch, because the ABR VC seems to end / or start at a regular ABR end system. This seamless interoperation allows to integrate complex VS/VD switches at sub-network ingress or egress boundaries alongside with simpler ER or EFCI switches along the network path [SmAT96]. This allows a flexible partition of the ABR control loop, which may be as short as hop-by-hop (using adjacent VS/VD switches) or stretch over several network hops (using regular ABR switches in between). It must be noted however, that under certain combinations of flow control algorithms unfair rate allocations may occur, especially when algorithms with a large oscillation of the ACR operate at the bottleneck switch of the VC [GoSS98], [PlSy97].

Although the ATM Forum adopted a rate-based flow control scheme, the VS/VD option allows to incorporate some desired features of the credit-based approach into this framework. The control of the ABR source is still rate-based, but the advantage of a short feedback loop can be exploited, if the burden of a more complex implementation of the VS/VD switch is accepted.

3.4 Summary

In this chapter a detailed introduction into congestion control algorithms and their features was given. A taxonomy for the classification of such algorithms was presented and the performance goals to be achieved by a congestion control algorithm were defined. Next, a special application of congestion control, namely, the ABR flow control scheme defined by the ATM Forum was introduced. Its operation on the side of end systems and network components was

explained in detail. Furthermore, the interaction between the flow control algorithms of ABR and TCP were discussed. Special attention was given to the virtual source / virtual destination (VS/VD) implementation option for ATM switches, since the ABR flow control algorithm presented in this thesis will be based on this design. The operation of a VS/VD switch and its possible impacts on the performance of ABR flow control were discussed.

In order to provide a means for the analytical analysis and design of a VS/VD flow control algorithm, linear control theory will be used. The following chapter will give a concise introduction into this subject area.

CHAPTER 4

Linear Control Theory

The ABR flow control mechanism described in the previous chapter controls the transmission rate of an ABR source through a closed-loop feedback system. In this chapter a brief introduction to linear control theory will be given, which follows closely the works of [Dorf89] and [PhHa91]. This methodology will then be used to design and analyze the flow control algorithm presented in the following chapter 5. The analogy of the ABR control flow technique and a linear, time-invariant feedback control system is explained. Control system components are identified and their representation in a block diagram is introduced. Furthermore, different mathematical approaches to derive performance metrics for continuous and discrete time linear control systems are presented.

4.1 Feedback Control Systems

A *control system* in general consists of two components, a physical system or *process* to be controlled, which possesses input and output variables (or *signals*), and a *controller* injecting input to this process. The controller receives as input a desired output response of the process, from which it determines a corresponding input signal for the process [Dorf89]. The input variables of the process, and hence the output of the controller, influence its output variables.

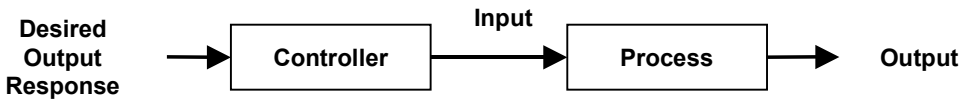

Figure 4-1: *Simple Control System [Dorf89]*

For reasons of simplicity, only systems with a single input/output variable, called the *control variable*, will be investigated in the following. Figure 4-1 illustrates the structure of a simple control system [Dorf89].

In the case where the input signal of the process depends on its output signal, the system is called a *closed-loop feedback control system*. Then, the output variable of the process is measured by a *sensor* and is compared to the desired output response. The resulting *error signal*, as the difference between the target and the actual value, is fed into the controller. There the error signal is processed, e.g., amplified, and enters the process, which is also called the *plant* [PhHa91]. Sometimes the output signal of the process is influenced by other factors, that can not be controlled and which are accumulated into a *disturbance signal* [Dorf89]. Figure 4-2 depicts the components of a closed-loop feedback control system.

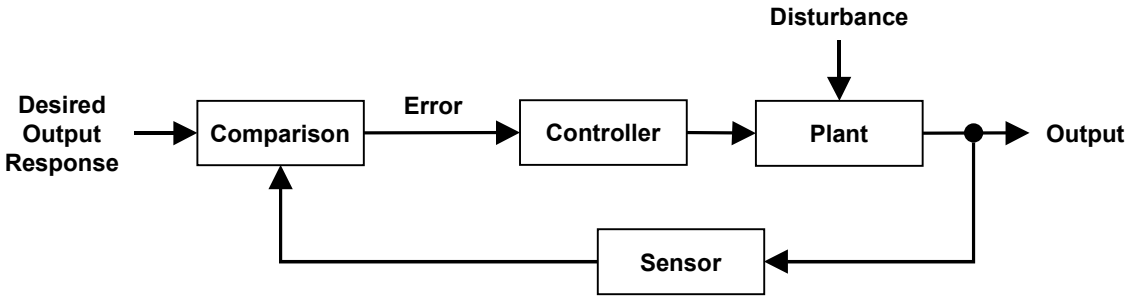

Figure 4-2: *Closed-Loop Feedback Control System*

As mentioned in the previous chapter, the ABR flow control mechanism in ATM networks is based on such a closed-loop feedback control mechanism. An ATM switch along the path from the source to the destination provides either its desired transmission rate as an *Explicit Rate* (ER) or indicates its *congestion state* (via a congestion bit). With respect to the *Allowed Cell Rate* (ACR) of the ABR source, one can identify one or more *bottlenecked* switches along the transmission path. A bottleneck switch controls the ACR of the ABR source by setting the lowest ER or by indicating an overload situation for the longest time period of all traversed switches.

When taking the view of such a bottleneck switch, the process of rate adaption can be modeled by a closed-loop feedback control system. In general an ATM switch determines the desired explicit rate or its congestion state by comparing a switch parameter with a target value, e.g., a desired queue length or traffic load. The resulting error signal, as the difference between the desired and actual value, is then fed into the controller, which implements the ABR switch algorithm. Depending on the type of the algorithm, the controller calculates an ER value for the corresponding ABR connection or, in the presence of congestion, sets the congestion bits in data or backward RM cells. The new traffic load according to the given feedback is experienced only after a delay. This time period is called *feedback* or *action delay* [AlBS97], [Srik99], because the result of the control action is perceived only after this time delay. The plant accounts for this transport of the control information to the ABR source, the adaptation of the ACR, and the impact on the output variable, e.g., a change in the queue level. Fluctuations of

the available bandwidth, due to high priority background traffic or changing number of ABR sources, can be seen as external disturbances, affecting the output signal.

4.2 Analytical Model

In order to describe a control system and its performance characteristics in a structured way and to fully understand the interaction between different system variables, it is necessary to obtain a mathematical model of the control system. This analytical model relates the output or *response* of a system to its input or *excitation* by the use of mathematical functions.

A system is defined as *linear*, if for two excitations $r_1(t)$, $r_2(t)$ and corresponding responses $c_1(t)$, $c_2(t)$ the following property holds for any two constants a_1, a_2

$$\text{input: } a_1 \cdot r_1(t) + a_2 \cdot r_2(t), \text{ output: } a_1 \cdot c_1(t) + a_2 \cdot c_2(t) \text{ [PhHa91].} \tag{4.1}$$

As control systems are dynamic, their transient behavior can be described by *differential equations*. For *linear*, *time-invariant* (the coefficients of the equations are constant) differential equations, it is possible to derive quantitative analytical results. A common technique for solving these linear differential equations is their conversion to algebraic equations by the use of the *Laplace transform* [PhHa91]. This corresponds to a shift from the time to the complex frequency domain.

The Laplace transform of a function of time $f(t)$ is defined as

$$F(s) = \mathfrak{L}[f(t)] = \int_0^\infty f(t) \cdot e^{-st} dt \,, \tag{4.2}$$

where $\mathfrak{L}$ denotes the Laplace transformation. The resulting Laplace transform of $f(t)$ is a function of the complex variable s [Doet61]. The *inverse Laplace transform* is given by

$$f(t) = \mathfrak{L}^{-1}[F(s)] = \int_{\sigma - j\infty}^{\sigma + j\infty} F(s) \cdot e^{st} ds \,, \tag{4.3}$$

where $\mathfrak{L}^{-1}$ denotes the inverse Laplace transformation and $j = \sqrt{-1}$. The value of σ in eq. 4.3 is determined by the singularities of $F(s)$ [Doet61]. The technique of *partial fraction expansion* is often used to find the inverse Laplace transform, but in general this is not a trivial task [PhHa91].

With the help of the Laplace transform, linear and time invariant differential equations may be represented by a *transfer function*, ignoring the initial conditions. When a system has an input function $e(t)$ and an output function $c(t)$ with corresponding Laplace transforms $E(s)$ and $C(s)$, then the transfer function $G(s)$ is defined by [Ebel87]

$$G(s) = \frac{C(s)}{E(s)} \,. \tag{4.4}$$

A control system is composed of different *functional blocks*, like the controller and the plant, which in turn can be described by individual transfer functions. Therefore, it is possible to represent the system model by a *block diagram* [Dorf89]. It contains operational blocks of the sys-

tem, which are characterized by the three functions $E(s)$, $C(s)$, and $G(s)$. The transfer function $G(s)$ multiplied by the error function $E(s)$ yields the output $C(s)$. In a block diagram, a comparison of signals is depicted by a circle, which represents the summation of all input functions. Positive and negative terms are marked by a plus or minus sign, respectively [PhHa91]. The closed-loop feedback control system of figure 4-2 (ignoring the disturbance signal) may be represented as a block diagram as shown in figure 4-3 [Dorf89], where $G_c(s)$, $G_p(s)$, $R(s)$, and $H(s)$ denote the transfer functions of the controller, the plant, the input, and the sensor, respectively.

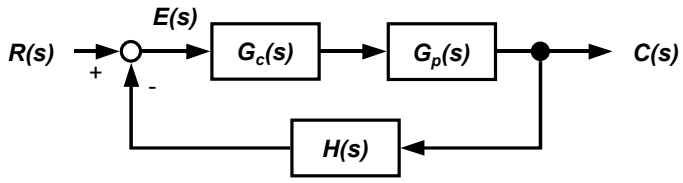

Figure 4-3: *Block Diagram of Closed-Loop Feedback Control System [Dorf89]*

The concatenated block of controller and plant may be expressed by the joint transfer function $G(s)$. Hence, the following equations hold for a closed-loop feedback control system [PhHa91]:

$$E(s) = R(s) - H(s) \cdot C(s) \,, \tag{4.5}$$

$$C(s) = G(s) \cdot E(s) \,. \tag{4.6}$$

When substituting the first equation into the second and solving for $C(s)$ one obtains the *system transfer function* $T(s)$ [PhHa91]:

$$T(s) = \frac{G(s)}{1 + G(s) \cdot H(s)} \,. \tag{4.7}$$

The transfer function is of great importance for the analysis of the system, because the *transient response* of the control system is determined by the location of the transfer functions's poles and zeros [PhHa91]. In the case of a closed-loop feedback control system, the poles of $T(s)$ correspond to the roots of $1 + G(s) \cdot H(s)$. This results in the so called *characteristic equation*

$$1 + G(s) \cdot H(s) = 0 \,, \tag{4.8}$$

where $G(s)H(s)$ is defined as the *open-loop transfer function* [PhHa91]. Hence, the transient response of a closed-loop feedback system depends on the root locus of its open-loop transfer function [Dorf89].

4.3 Performance Characteristics

The design of the controller for a closed-loop feedback system is in general based on certain performance characteristics that may include, but are not limited to, the following criteria [PhHa91].

- **Disturbance Rejection** aims at minimizing any unwanted effects on the output variable from uncontrollable, external disturbances.
- The **Steady State Error** is defined as the limit of the error signal, when time approaches infinity.
- The **Transient Response** is defined as the response of a system as a function of time. It is used to describe the transient behavior of a system after a change in the input signal until the system reaches a steady state again.
- **Stability** is defined as the fact that the control system will converge to a steady state after a change in the input signal.

These characteristics are introduced in the following sections. Especially the last point is of major importance for the effectiveness of the control system, since a secure operation is desired under all possible circumstances.

Other performance parameters [PhHa91], that will not be addressed in detail are the following.

- **Sensitivity** characterizes the effects of small parameter changes in the plant.
- **Robustness** means that the system shall be insensitive to small inaccuracies of the system model.

4.3.1 Disturbance Rejection

Feedback control systems may suffer from *external disturbances* that can not be controlled but influence the output signal. In the block diagram these disturbances are merged into a *disturbance signal* $D(s)$, that serves as an additional input to the plant [PhHa91]. Its direct effect on the output signal $C(s)$ can be denoted through a transfer function $G_d(s)$, see figure 4-4. Sometimes however it is possible to measure or quantify the disturbance signal. As a consequence, it is possible to *feedforward* the disturbance signal through a suited compensator, whose transfer function is $G_{cd}(s)$ to the point where the error signal $R(s)$ is calculated [PhHa91]. This approach is shown in figure 4-4, where the disturbance signal $D(s)$ is taken into account already at the calculation of the error signal and hence the input to the controller.

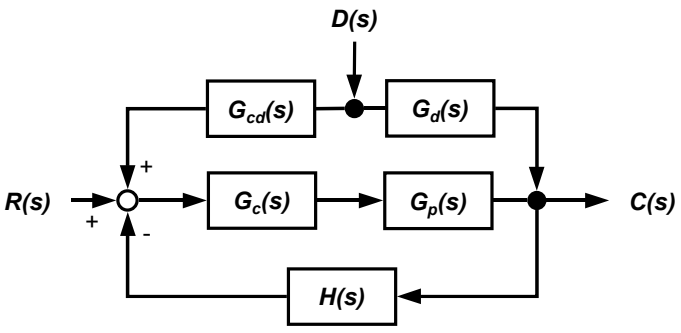

Figure 4-4: *Feedforward Compensation [PhHa91]*

4.3.2 Steady State System Error

In order to assess the accuracy of a control system, one can determine the steady state system error resulting from a change of the system input $r(t)$. Using the final limit theorem of the Laplace transform [Doet61], the steady state system error e_{ss} is defined as

$$e_{ss} = \lim_{t \to \infty} e(t) = \lim_{s \to 0} s \cdot E(s), \tag{4.9}$$

provided that the limit exists [Dorf89]. For the change of the system input, usually two standard functions are examined, namely, the step input and the ramp input [Ebel87]. Whereas the step input describes the instantaneous change from one constant input value T_a to another constant value T_b at a certain time instant (e.g., $t = 0$), the ramp input consists of a constantly increasing signal with slope of a [Dorf89].

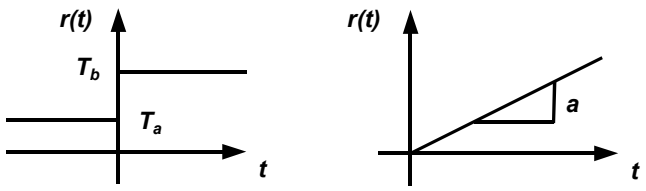

Figure 4-5: *Standard Input Functions [Dorf89]*

Figure 4-5 illustrates the two input functions.

4.3.3 Transient Response

The *transient response* of a system describes the behavior of its output signal as a function of time. It is used to analyze the system output signal between a change of the input signal and the point where the system reaches a steady state. The performance specifications are described in the time domain referring to certain points of the function $c(t)$. Figure 4-6 illustrates the different performance parameters, using a typical damped response to a step input, and assuming a steady state value c_{ss} of 1 for $c(t)$ [PhHa91].

The *rise time* T_r denotes the time period in which the output signal rises from 10 % to 90 % of the final value c_{ss} [PhHa91].

The *peak time* T_p is defined as the duration the output function needs to reach its peak value M_{pt}. This peak value is often called *percent overshoot* and is then expressed in percent, based on the final value of $c(t)$ [PhHa91].

The *settling time* T_s specifies the point in time since which the output function remains inside the interval $[c_{ss} - d , c_{ss} + d]$, where the value of d is usually chosen between 2 and 5 % of c_{ss} [PhHa91].

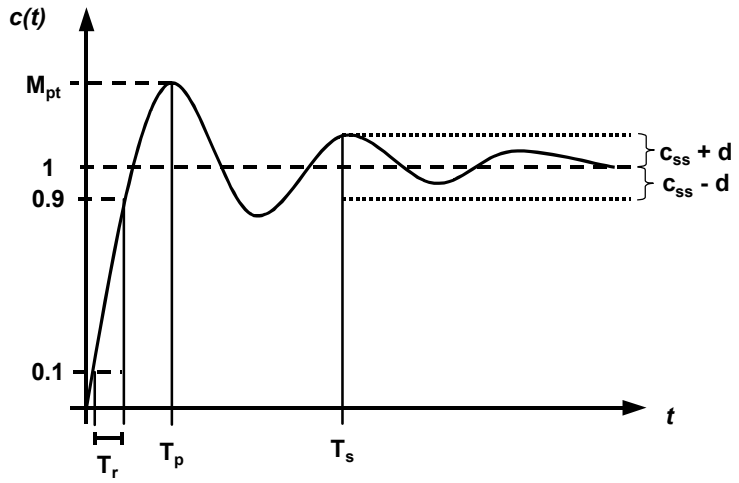

Figure 4-6: *Typical Transient Response to a Step Input [PhHa91]*

4.3.4 Stability Analysis

Stability of a control system means that the system generates a bounded response to a bounded input or disturbance [Meyr94]. As mentioned in the previous section, the locations of the roots and poles of the system transfer function define the transient behavior and stability of the feedback system. In a stable system the roots of the characteristic equation are required to be located in the left-hand side of the *s*-plane [PhHa91]. In order to determine the root locations, there exist different approaches. If for instance the characteristic equation consists of a polynomial, the Routh-Hurwitz criterion may be used [Ebel87]. Other techniques involve the plotting of the root locus, when some system parameters are varied [Dorf89]. A common method for system and stability analysis is the use of the frequency response method, which is introduced in the next section. For this method, not only one form of the characteristic equation, like for the Routh-Hurwitz criterion, or a very accurate system model, like for the root locus technique, is required [PhHa91].

4.4 Frequency Response Analysis

The main idea of the *frequency response analysis* is to evaluate a control system based on its response to a sinusoidal input. For this, steady-state system response is observed as the frequency of a sinusoidal input signal is varied [Dorf89]. When assuming a linear, closed-loop feedback control system, like depicted in figure 4-3, the system transfer function $T(s)$ (see eq. 4.7) is a function of the complex variable s. For the frequency response analysis, the variable s of $T(s)$ is replaced by $j\omega$ ($j = \sqrt{-1}$), which corresponds to applying a *Fourier transformation* $\mathfrak{F}$

$$F(j\omega) = \mathfrak{F}[f(t)] = \int_{-\infty}^{\infty} f(t) \cdot e^{-j\omega \cdot t} dt \qquad (4.10)$$

instead of a Laplace transformation (see eq. 4.2), if $f(t)$ is defined only for $t \geq 0$ [Dorf89]. For a general transfer function $G(s)$, s is substituted by $j\omega$ and the resulting system response may be described either by *magnitude* and *angle* as:

$$G(j\omega) = \left|G(j\omega)\right| \cdot e^{\phi(j\omega)} = \left|G(j\omega)\right| \angle \phi(j\omega), \tag{4.11}$$

or by the *real* and *imaginary* parts as:

$$G(j\omega) = \Re\left(G(j\omega)\right) + j \cdot \Im\left(G(j\omega)\right) \text{ [Dorf89].} \tag{4.12}$$

4.4.1 Nyquist Diagrams

The latter representation can be used to draw a *polar plot* of the real and imaginary parts of the frequency response [Dorf89]. Figure 4-7 shows the polar plot of the transfer function

$$G(j\omega) = \frac{5}{(1+s)^3}, \tag{4.13}$$

for $\omega = [\,0\,,\,\pi\,]$.

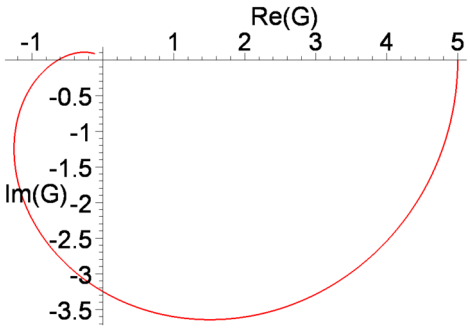

Figure 4-7: *Polar Plot of Transfer Function $G(j\omega)$*

For a closed-loop feedback control system, the characteristic equation is defined as in eq. 4.8. A polar plot of the open-loop transfer function $G(s) \cdot H(s)$ is also called a *Nyquist diagram* [PhHa91].

4.4.2 Bode Diagrams

An alternative for graphically displaying the system response is the use of logarithmic plots, called *Bode diagrams* [Dorf89]. These diagrams contain:

- a plot of the *logarithm of the magnitude* of $G(j\omega)$ against the *logarithm of the frequency* ω on one set of axes and
- a plot of the *phase* of $G(j\omega)$ against the *logarithm of the frequency* ω against a second set of axes.

The logarithm of the magnitude of $G(j\omega)$, also called the *logarithmic gain*, is plotted in units of *decibel* (dB), where

$$\text{logarithmic gain in dB} = 20 \log (| G(j\omega) |) \text{ [Dorf89].} \qquad (4.14)$$

Figure 4-8 presents the Bode diagram of the transfer function $G(j\omega)$ of eq. 4.13 for $\omega = [\ 0.01\ ,\ 10\ ]$. It has to be noted, that angles smaller than -180 degrees are expressed as a positive value. Therefore, the phase plot shows a peak at the -180 deg. crossing point.

The main advantage of Bode diagrams is the fact that the plot of a transfer function $G(j\omega)$ with different factors can be constructed by pointwise adding of the Bode diagrams for the single factors. Many control systems consist of a concatenation of different blocks, which results in different factors for $G(j\omega)$. This multiplicative composition of $G(j\omega)$ is transformed to a more simpler additive operation in a Bode diagram [Ebel87].

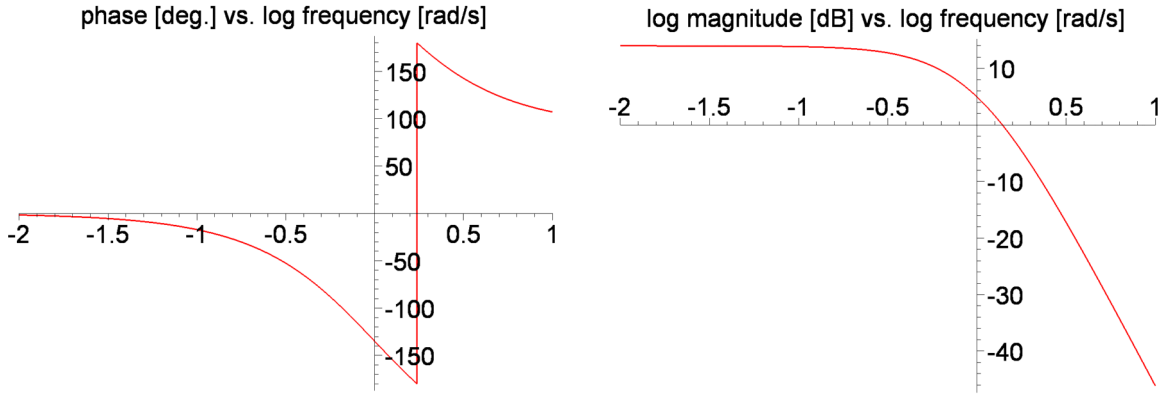

Figure 4-8: *Bode Diagram of Transfer Function $G(j\omega)$*

4.4.3 Nyquist Criterion

As mentioned in section 4.3.4, *stability* is a major design goal for control systems, since it allows a predictable and safe operation for all possible bounded input signals. Another important performance measure is the *relative stability* of a control system. This term quantifies the degree of stability, if certain system parameters are varied or if the system model is not accurate [PhHa91]. A system which is only marginally stable poses the danger to become unstable under certain conditions, since it operates near the border of stability. For a closed-loop feedback system, the characteristic equation has the form of eq. 4.8 and the system stability may be determined from the frequency response of the open loop transfer function $G(j\omega)H(j\omega)$ with the help of the *Nyquist criterion* [PhHa91]. It is based on a rational mapping function $F(s)$ from the complex s-plane to the $F(s)$-plane. Furthermore, the frequency response analysis allows to assess the relative stability of the control system.

The Nyquist criterion uses the *theorem of Cauchy*, also called the *principle of the argument*, which states:

Let $F(s)$ be a ratio of two polynomials in s. Let the contour Γ_s in the s-plane be mapped into the complex plane as Γ_F through the mapping $F(s)$ as the traversal is in the clockwise direction along the contour. If $F(s)$ is analytic within and on Γ_s, except at a finite number of poles, and if

$F(s)$ does not pass through any poles or zeros of $F(s)$ as the traversal is in clockwise direction along the contour Γ_s, then

$$N = Z - P, \tag{4.15}$$

where Z is the number of zeros of $F(s)$ in Γ_s, P is the number of poles of $F(s)$ in Γ_s, and N is the number of encirclements of Γ_F of the origin of the $F(s)$-plane in the clockwise direction [Dorf89], [PhHa91].

Taking into account the stability criterion noted in section 4.3.4, that the zeros of the characteristic equation must lie in the left-hand s-plane, one defines (according to eq. 4.8)

$$F(s) = 1 + G(s) \cdot H(s) = 0, \text{ or alternatively } G(s) \cdot H(s) = F(s) - 1 = P(s). \tag{4.16}$$

Taking a contour Γ that encloses the right-hand s-plane, Cauchy's theorem can be used to determine the number Z of zeros of $P(s)$ that occur inside this contour [Dorf89]. By plotting Γ_P in the $P(s)$-plane and counting the number N of clockwise encirclements of the (-1,0) point, the following equation holds:

$$Z = N + P \text{ [Dorf89]}. \tag{4.17}$$

If P, the number of poles of $P(s)$ in Γ, equals zero, then the number of encirclements of the (-1,0) point N corresponds to the number of unstable zeros of $P(s)$ or the open-loop transfer function. The encirclements have to be analyzed with respect to the (-1,0) point instead of the origin in the $P(s)$-plane, since the rational mapping function is shifted by one unit to the left [PhHa91]. A path which follows the imaginary axis from $-j\infty$ to $+j\infty$ and encloses the right-hand s-plane with an arc of infinite radius in clockwise direction is called a *Nyquist path* [PhHa91] or *Nyquist contour* [Dorf89], see figure 4-9.

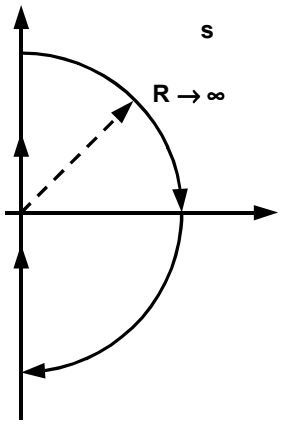

Figure 4-9: *Nyquist Path*

In the case where $P(s)$ has single poles or zeros on the Nyquist path, e.g., at the origin, these locations may be excluded from the contour by making an infinitesimal small detour around these points. The result of Cauchy's theorem still remains valid in these cases [PhHa91].

The Nyquist criterion follows exactly the approach introduced above [Dorf89], [Ebel87]:

A feedback control system is stable, if and only if the mapping of the Nyquist contour Γ_P in the $P(s)$-plane does not encircle the (-1,0) point, when the number of poles of $P(s)$ in the right-hand s-plane is zero ($P = 0$).

When the number of poles of $P(s)$ in the Nyquist contour is greater than zero, then the Nyquist criterion becomes:

A feedback control system is stable, if and only if for the mapping of the Nyquist contour Γ_P in the $P(s)$-plane, the number of clockwise encirclements of the (-1,0) point is equal to the number of poles of $P(s)$ in the right-hand s-plane ($N = -P$).

Figure 4-10 illustrates the Nyquist diagram for the open-loop transfer function $G(s)$ of eq. 4.13 for a closed-loop, unity feedback control system, where $H(s) = 1$. Using the Nyquist criterion and observing that

- $G(s)$ has no poles in the right-hand s-plane ($P = 0$) and that
- the mapping shows no encirclements of the (-1,0) point ($N = 0$),

one can conclude that the number of zeros of the open-loop transfer function in the right-hand s-plane is zero ($Z = N + P = 0$) and hence that the closed-loop control system is stable.

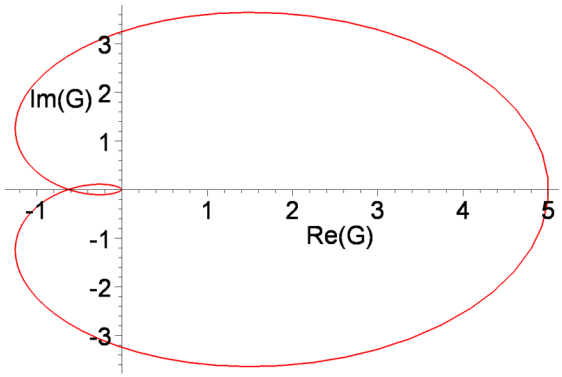

Figure 4-10: *Nyquist Diagram of Open-Loop Transfer Function $G(j\omega)$*

4.4.4 Relative Stability

As mentioned in the previous section, *relative stability* is another crucial performance metric in controller design. Beside the fact that the system operation is stable, the *safety margin* of this stability plays an important role. On the one hand an adequate safety margin prevents the system from drifting into an unstable operation, even if certain system parameters have not been modeled accurately. On the other hand there exists a trade-off between a large safety margin and a fast transient response of the system [Dorf89], [PhHa91]. Therefore, it is not possible to achieve both goals at the same time and controller parameters have to be chosen to put more emphasis on one of the two system characteristics, depending on the actual application context.

In order to assess the relative stability of a closed-loop feedback control system quantitatively the Nyquist diagram of the frequency response analysis can be used. Derived from the Nyquist

criterion, the notion of the *phase margin* defines the distance of the Nyquist diagram to the (-1,0) point and hence a measure of the relative stability of the system [PhHa91]. The phase margin is expressed in degrees and is defined as the magnitude of the minimum angle by which the Nyquist diagram has to be rotated to intersect the (-1,0) point for a stable closed-loop system [Ebel87]:

$$\text{phase margin } \phi_m \text{ in degrees} = \angle\,(\,G(j\omega_1)H(j\omega_1)\,)\,-180°, \tag{4.18}$$

where ω_1 denotes the frequency at which the magnitude of $G(j\omega_1)H(j\omega_1)$ equals 1.

Another measure for relative stability of a closed-loop control system is the *gain margin*, which is defined as the reciprocal of the magnitude of $G(j\omega)H(j\omega)$ on the -180 degrees crossover point of the Nyquist diagram. If there are several points where the Nyquist diagram intersects the negative real axis, the point resulting in the smallest gain margin is chosen [PhHa91], [Ebel87]. The gain margin is usually expressed in decibels (dB) and hence:

$$\text{gain margin } \alpha \text{ in dB} = -20 \log (\,|\,G(j\omega_0)H(j\omega_0)\,|\,) \text{ dB}, \tag{4.19}$$

where ω_0 denotes the frequency at which the phase angle of $G(j\omega_0)H(j\omega_0)$ equals -180 degrees. Taking the closed-loop unity feedback system of figure 4-10 ($H(j\omega) = 1$), the phase and gain margins are depicted in figure 4-11.

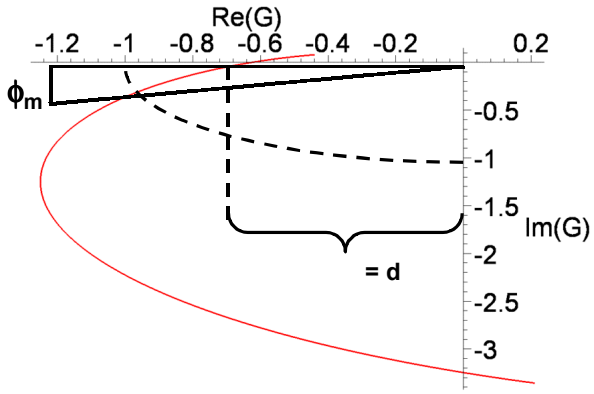

Figure 4-11: *Phase and Gain Margins in a Nyquist Diagram*

The gain margin α corresponds to 20 log ($1 / d$) or -20 log (d) dB. When considering the Bode diagram of the open-loop transfer function $G(j\omega)H(j\omega)$, the phase margin ϕ_m equals the difference between the -180 deg. line and the phase plot, taken at the frequency ω_1, where the magnitude curve crosses the ω-axis. The gain margin equals the difference of the magnitude plot and the ω-axis, taken at the frequency ω_0, where the phase curve intersects the 180 deg. line [PhHa91].

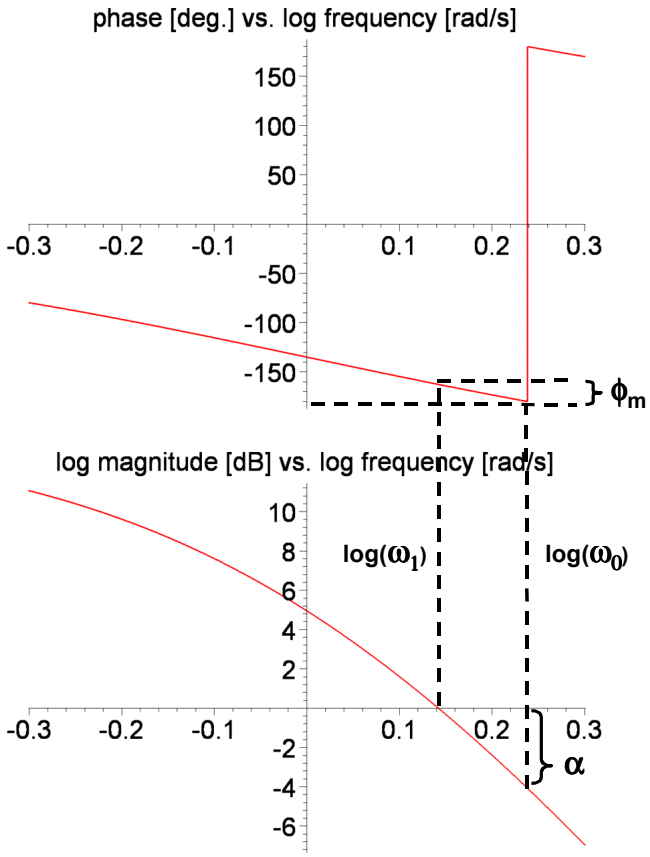

Figure 4-12: *Phase and Gain Margins in a Bode Diagram*

Figure 4-12 shows the phase margin ϕ_m and the gain margin α in the Bode diagram of the closed-loop unity feedback system of figure 4-10 ($H(j\omega) = 1$).

4.4.5 Ideal Time Delay

In many control systems an *ideal time delay* occurs, which means that the input signal is experienced as an output signal of the delay instance after a fixed time period t_0, without any attenuation of the signal [Dorf89]. Mathematically the output signal can be described as

$$c(t) = e(t - t_0) \cdot u(t - t_0), \tag{4.20}$$

where $u(t)$ denotes the unit step function, which has the value of 1 for $t \geq 0$ and of 0 for $t < 0$, and $e(t)$ is the input function [PhHa91]. Applying the Laplace transform yields

$$C(s) = e^{-t_0 \cdot s} \cdot E(s), \tag{4.21}$$

where $C(s)$ and $E(s)$ are the Laplace transforms of $c(t)$ and $e(t)$, respectively [PhHa91]. Hence, the transfer function of the ideal time delay is

$$G(s) = e^{-t_0 \cdot s} \text{ [Ebel87].} \tag{4.22}$$

When looking at the frequency response of an ideal time delay, the variable s has to be replaced by $j\omega$ in eq. 4.22, which gives

$$G(j\omega) = e^{-t_0 \cdot j\omega} = 1 \angle -\omega t_0 \,. \tag{4.23}$$

The magnitude of the transfer function is unity and therefore an ideal time delay does not alter the shape of the input signal. It introduces only a phase shift or lag, which becomes larger with increasing frequency [Dorf89]. The Nyquist criterion defined in section 4.4.3 assumes a transfer function that is a ratio of two polynomials. However, the criterion remains valid, if an ideal time delay is added to the system, since its transfer function does not add any poles or zeros to the right-hand *s*-plane [Dorf89]. The added phase lag of the ideal time delay has a destabilizing effect on the control system, because the Nyquist diagram is rotated towards the (-1,0) point [PhHa91].

Revisiting the closed-loop, unity feedback system with the open-loop transfer function $G(j\omega)$ of eq. 4.13 and adding an ideal time delay of 0.8 sec, the new transfer function becomes

$$G(j\omega) = \frac{5 \cdot e^{-0.8 \cdot s}}{(1+s)^3} \,. \tag{4.24}$$

When using the Nyquist criterion to investigate the system stability, the Nyquist diagram, shown in figure 4-13, proves that the control system is now unstable. The additional phase lag of the ideal time delay has caused an encirclement of the (-1,0) point. This finding is confirmed by the Bode diagram of the modified transfer function of eq. 4.24. The phase margin is negative and the gain margin positive, as depicted in figure 4-14. The ideal time delay of 0.8 sec results in a phase lag of -0.8 rad for $\omega = 1$ rad/sec. This corresponds to a phase shift of about -45 degrees in the Bode diagram at the point where the phase plot crosses the *y*-axis. The phase shift can be observed when comparing the phase plots of figure 4-12 and figure 4-14. However, the magnitude plots of these Bode diagrams are identical, because the magnitude of the ideal time delay transfer function is equal to unity, which results in an additive component of 0 dB.

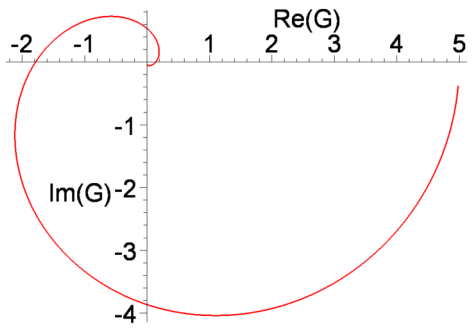

Figure 4-13: *Nyquist Diagram of an Unstable Feedback Control System*

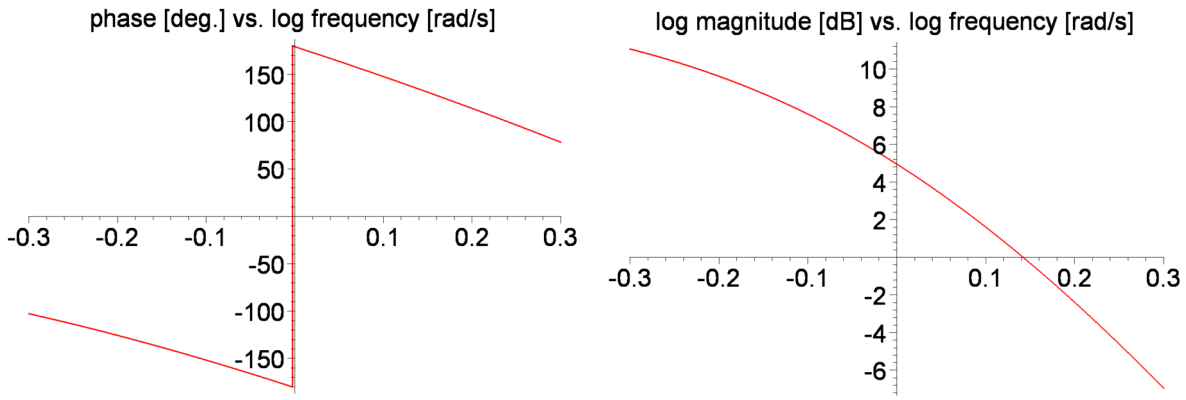

Figure 4-14: *Bode Diagram of an Unstable Feedback Control System*

Hence, it can be concluded that the frequency response method, using the Nyquist criterion for stability analysis and the Nyquist and Bode diagrams for visualization, is well suited for the study of closed-loop feedback systems with ideal time delays.

4.5 Discrete Time Systems

So far the properties of continuous time or analog control systems were discussed in this chapter. In many cases digital computer systems are used as controllers, which process input data only at fixed time instances, called the *sample time* [Dorf89]. With respect to the application in the area of ABR flow control, the fact that feedback information is given only at distinct time intervals when RM cells are received, determines the discrete time characteristic of the closed-loop feedback control system. In addition, any measurements performed by the switch, e.g., queue levels or available bandwidth, are executed at sample intervals. As a consequence, a more accurate and detailed analysis of the ABR flow control scheme would lead to a discrete time closed-loop feedback system.

Whereas system dynamics for continuous time control systems are described by *differential equations*, these are replaced by *difference equations* for a discrete time system [PhHa91]. A special problem in discrete time control systems is the sampling of continuous time signals. The value read at a sampling instant remains valid for the duration of the sampling interval. For this reason, a quantization component in a control system is called a *sampler and zero-order hold* [Oppe89]. The output of the sampler component can be defined as

$$r^*(t) = \sum_{k=0}^{\infty} r(kT) \cdot \delta(t - kT) \ , \tag{4.25}$$

for $t > 0$, where $r(kT)$ denotes a series of sample values, T represents the length of the sample interval, kT is the k^{th} sample interval, and $\delta(t)$ is the impulse function

$$\delta(t) = \begin{cases} 1 \text{ for } t = 0 \\ 0 \text{ for } t = kT, k \neq 0 \end{cases} \text{ [Dorf89].} \tag{4.26}$$

The effects of quantization may be neglected, if the sampling intervals are sufficiently small compared to the changes of the sampled signal [Dorf89]. Applying the Laplace transformation $\mathfrak{L}$ to eq. 4.25 yields

$$R^*(s) = \mathcal{L}\left[r^*(t)\right] = \sum_{k=0}^{\infty} r(kT) \cdot e^{-kTs} \text{ [Dorf89].} \tag{4.27}$$

Substituting $e^{Ts} = z$ results in the (single-sided) *z-Transformation* $\mathcal{Z}$, which is defined as

$$R(z) = \mathcal{Z}\left[r(t)\right] = \mathcal{Z}\left[r^*(t)\right] = \sum_{k=0}^{\infty} r(kT) \cdot z^{-k}, \tag{4.28}$$

where the *z-Transform* $R(z)$ of $r(t)$ is a function of the complex variable z. In the analysis of the stability and transient response of a discrete time control system the z-Transform replaces the Laplace transform for analog systems [PhHa91].

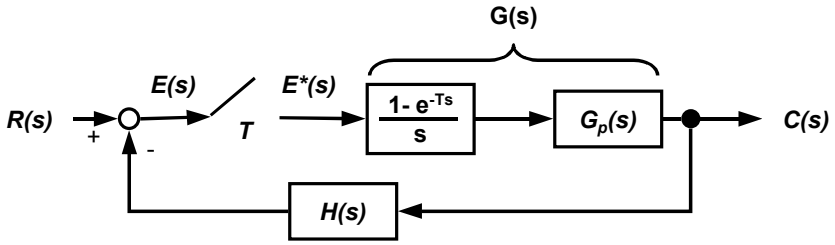

Figure 4-15: *Closed-Loop, Sampled-Data Feedback Control System*

Figure 4-15 presents a closed-loop, sampled-data feedback control system. It has to be noted, that it is not possible to determine a Laplace transform of the transfer function of an ideal sampler and that eq. 4.27 represents a Laplace transform of the output sequence only [PhHa91]. The zero-order hold component however, can be described by the transfer function

$$G_{ho}(s) = \frac{1 - e^{-Ts}}{s} \text{ [PhHa91].} \tag{4.29}$$

This component may be combined with the plant transfer function $G_p(s)$ to a common transfer function $G(s)$. Then, the closed-loop, sampled-data feedback control system has a system transfer function

$$T^*(s) = \frac{C^*(s)}{R^*(s)} = \frac{G^*(s)}{1 + \overline{GH}^*(s)}, \text{ with } \overline{GH}^*(s) = \mathcal{L}\left[g^*(t) \cdot h^*(t)\right] \text{ [PhHa91].} \tag{4.30}$$

Using the substitution $e^{Ts} = z$, results in the characteristic equation of the closed-loop, sampled-data control system

$$1 + \overline{GH}(z) = 0, \text{ with } \overline{GH}(z) = \mathcal{Z}\left[g(t) \cdot h(t)\right]. \tag{4.31}$$

The substitution $e^{Ts} = z$ defines a mapping of the s-plane into the z-plane, where for instance the imaginary axis of the s-plane is transformed into the unit circle in the z-plane [PhHa91]. As mentioned in section 4.3.4, a closed-loop feedback system is stable, if the open-loop transfer function has no poles in the right-hand s-plane. Correspondingly a sampled-data system is stable, if all the roots of the characteristic equation (eq. 4.31) are located inside the unit circle of

the *z*-plane [Dorf89]. The Nyquist criterion introduced in section 4.4.3 can be adapted to the case of discrete time systems. The Nyquist path in the *z*-plane is the unit circle passed in counter-clockwise direction. The Nyquist criterion results in

$$Z = N + P, \tag{4.32}$$

where Z and P are the number of the zeros and poles of the characteristic equation outside the unit circle, respectively, and N is the number of clockwise encirclements of the (-1,0) point in the *z*-plane. As for the analog system, if $P = 0$, then N has to be 0 for a stable sampled-data, closed-loop feedback control system. Phase and gain margins are defined as for analog control systems [PhHa91]. The zero-order hold component of the sampler introduces a phase lag to the system that has a destabilizing effect. If this effect is taken into consideration when designing the controller and the sampling period T is small in relation to the time constants of the system, it is possible to analyze a discrete time system using a continuous time model.

4.6 Summary

This chapter presented a short introduction to linear control theory. Since the ABR flow control mechanism operates as a closed-loop feedback control system, emphasis was put on the analysis of these kind of systems. An analytical model using the Laplace transform was presented and the performance characteristics of disturbance rejection, steady state error, transient response, and stability were introduced.

Stability is a major concern in controller design, because it assures a reliable operation of the system. In order to assess the stability of a control system the common technique of frequency response analysis was explained. It uses the visualization of the system output to sinusoidal inputs with the help of Nyquist and Bode diagrams. A closed-loop feedback control system is stable, if all pole locations of its transfer function are in the left-hand *s*-plane. In order to determine the pole locations, the locus of the roots of the characteristic equation is analyzed. For this, the Nyquist criterion uses the open-loop transfer function to map the Nyquist path into a Nyquist diagram. This Nyquist diagram or alternatively a Bode diagram then provides information on the relative stability of the control system. This approach remains valid, if an ideal delay is added to the closed-loop feedback system, which is the case in many applications.

For a discrete-time control system, that uses sampled data, the methods presented may be adapted, e.g., in applying the *z*-Transform, so that an analysis is also possible. In the following chapter, the theory of closed-loop feedback control systems will be used to develop an ABR flow control algorithm for a virtual source / virtual destination switch. For this task, the analytical methods of linear control theory can guarantee to meet desired performance goals.

CHAPTER 5

Flow Control Algorithm Design

This chapter will focus on the design of a flow control algorithm for a virtual source / virtual destination (VS/VD) switch. First, an overview of the related work available in the research area of ABR flow control is presented. In the following, the VS/VD switch layout and the internal cell flows are explained. Next, the design goals of the newly developed flow control algorithm are outlined. Then, a model for the traffic flow and the resulting equations for the queue dynamics at the VS/VD switch are introduced. A corresponding control model using a closed-loop feedback control system is presented. Next, the values of the controller parameters for the two feedback loops are derived. Finally, the complete controller design and implementation details of the ABR flow control algorithm for a VS/VD switch are highlighted.

5.1 Related Work

The development and performance analysis of ABR flow control algorithms is a very active research area, with a great number of publications and proposals, e.g., [AfMO00], [AAEE99], [ArCA96], [CFKS97], [Char94], [ChWa97], [FGFM97], [GoSS98], [HoTP97a], [KaVa95], [Kaly00], [KuMo95], [MoFd97], [Muth00], [Ohsa95c], [TsHK98], [ZhLi96]. The various works may be categorized with respect to the following criteria.

- **EFCI** (Explicit Forward Congestion Indication): The binary scheme for congestion indication was introduced in section 3.2.2, where possible enhancements and its pros and cons were discussed. The oscillatory behavior of the *Allowed Cell Rate* (ACR) of the source is caused by the alternating congestion indications of the EFCI switch. This effect was modeled in several works assessing the different phases of the rate changes and the performance of the algorithm [BoSh90], [LaLi98], [PaAg95], [PaAg97], [RaQi95], [Ritt97a], [Yin95]. It can be concluded that queue levels increase linearly with the number of ABR sources and that EFCI switches, with no further enhancements [KoRa95], are suited only for LAN environments, where the feedback delay is low. The ATM Forum presented in its Traffic Man-

agement Specification [TM4.1] an example for an EFCI algorithm, based on two queue threshold values.

- **Comparison of ER and EFCI**: Several existing works deal with the comparison of simple EFCI and more sophisticated *Explicit Rate* (ER) switch algorithms [ArCA96], [CFKS97], [KoRa97b], [Ohsa95a], [Ohsa95b], [Ohsa95d], [Ritt97b]. The superior performance of ER over binary algorithms with respect to fairness and buffer occupancy is confirmed, especially under extreme load conditions, large feedback delays and varying number of ABR sources. Furthermore, it was proven, that the two schemes may inter-operate [LRMS97], [LeRM98].
- **Fair Distribution of Bandwidth**: The goal of distributing the available capacity in a fair manner among competing ABR sources gained a lot of attention [ArCA96], [HTPK98b], [Jaff81], [PlSy97]. The proposed algorithms may be subdivided into a class that maintains a state for each connection [Char94], [ChCJ95], [ChRL96], [GhMa97a], [HoTP97b], [HTPK98b], [Kala97], [KaVa95], [LaTs99], [PrKB98], [Vand99], [Vand00] and a class with a stateless operation [AfMO00], [FuLL97], [HoTP97a], [HTPK98a], [Kaly00], [KoRa97a]. Some of the works derive a distributed asynchronous algorithm from a centralized approach [AbKu97], [Char94] or prove the convergence of the procedure to a fair allocation [Char94], [Kala95].
- **Background Traffic**: There are existing studies that focus on the efficient use of ABR link capacities in the presence of high priority background traffic, like e.g., VBR video transmissions [DaSt98], [HKSO98], [Ohsa95c]. A common solution to deal with high frequency changes in the ABR capacity, due to the highly variant nature of the background traffic, is the use of low pass filters for the measurement of the available bandwidth [FuLL97], [ZhLS97], [ZhLi96].

Beside these characteristics, there exist numerous proposed algorithms, which are not compatible to the ATM Forum Traffic Management Specification. They either

- presume a non-standard behavior of the end system [ChCC98], [GhMa97a], [KoRa99], [NaSi96], [RoBO95], [ZhYM97a], or
- of the switches [HuLP95],
- modify the structure of the *Resource Management* (RM) cells [AkLN97], [AnAA98], [NaSi97], or
- use a credit-based approach [KuBC94], [KuMo95], [KuWa98], [MoKu95], [OzSV94], [OzSV95].

As a consequence, these schemes can operate only in a homogeneous environment and are not compatible with other switches or end systems that rely on the ATM Forum standard.

Another aspect of ABR flow control is its usage for multicast connections, where the problem of providing feedback information at branch points is investigated [ChLL97], [ChCh00], [TzSi97].

The different methods and techniques to develop and assess switch algorithms for the control of ABR sources presented in related works, range from artificial intelligence [GaBo96], [LiDo97], fuzzy logic [DoDe95], [PiSR95], [PiSR97], prediction [Woo98], statistical physics [AiHo99], pricing models [LoLa99], to stochastic processes [AlBB94], [KaMM95].

After this overview of related work in the area of ABR flow control, the next sections will introduce common ER switch algorithms, that have been studied in the literature. Next, two aspects that play an important role in the framework of this thesis are presented in greater detail, namely, the group of VS/VD algorithms and the use of control theory.

5.1.1 Common ER Algorithms

In order to compare the performance of ER with VS/VD algorithms in chapter 6, several well-known and already analyzed ER algorithms [ArCA96], [GoSS98], [Jain96a], [Ohsa95c], will be introduced in the following. These algorithms will in addition be used in chapter 6 to demonstrate the inter-operability of VS/VD and ER switches in the ABR control loop.

5.1.1.1 EPRCA

The *Enhanced Proportional Rate Control Algorithm* (EPRCA) [Robe94] is mentioned in the ATM Forum Traffic Management Specification [TM4.1] as an example of a simple ER algorithm. The switch maintains a running exponential average of the current cell rates of all passing connections. Whenever a forward RM (FRM) cell is received, this *Mean Allowed Cell Rate* (MACR) is calculated as

$$MACR = (1 - \alpha) \cdot MACR + \alpha \cdot CCR, \tag{5.1}$$

where the exponential averaging factor α is generally chosen to be $1/16$ and the *Current Cell Rate* (CCR) value is the one indicated in the FRM cell. The switch determines its load by monitoring the queue level of a common output queue, that is shared by all connections. There exist two congestion states: the normal congestion is reached, if a first threshold HT is exceeded, whereas a severe congestion state is declared, if a second, higher threshold DQT is reached. A common *FairShare* for all connections is computed as

$$FairShare = MACR \cdot DPF, \tag{5.2}$$

where the *Down Pressure Factor* (DPF) has a default value of $7/8$. During normal congestion ($HT \leq$ queue level $<$ DQT), the ER field of backward RM (BRM) cells is reduced to the *FairShare* for connections which have a CCR higher than this *FairShare*. Other connections are allowed to increase their rates. In case of severe congestion (queue level $>$ DQT), the ER field of all BRM cells is reduced to the *FairShare*, regardless of the CCR of the connection. The congestion state is cleared, if the buffer level falls below a low threshold LT.

The algorithm represents a class of simple ER algorithms, where known problems include the convergence of the procedure to the fair share and possible unfairness for connections bottlenecked elsewhere on the path from sender to receiver [ArCA96], [GoSS98]. The running exponential average raises the problem of quickly utilizing all available bandwidth in the presence

of highly variant background traffic [HKSO98], [Ohsa95c]. Although there were suggested various improvements of the EPRCA [ArCA96], [ChWa97], [HKSO98], [Ohsa95c], the original version introduced above [Robe94] will be used in chapter 6 as an example of a simple ER algorithm. Table 5-1 lists the configuration parameters used for the simulations in presented in chapter 6.

Parameter Name	Parameter Value
Exponential Averaging Factor	0.9
Down Pressure Factor	0.875
Low Threshold	50 cells
High Threshold	70 cells
DQT	100 cells

Table 5-1: *EPRCA Configuration Parameters*

5.1.1.2 DERA

The *Distributed Explicit Rate Allocation* (DERA) algorithm maintains a state for each ABR VCs passing the switch and divides the connections into two sets: locally bottlenecked (*U*) and restricted elsewhere (*R*) [Char94]. At regular time intervals, the switch calculates a *Local Fair Share* (LFS) for each connection and loops through all passing ABR VCs. If the current transmission rate of the VC is lower than the LFS, the connection is marked as restricted and its current transmission rate is substracted from the capacity available to locally bottlenecked VCs. This procedure is repeated until all restricted connections have been identified and the remaining available bandwidth is determined. This remaining bandwidth is then evenly distributed among the locally bottlenecked connections [ChCJ95]. The transmission rate of a VC is updated whenever a forward or backward RM cell is received by the switch. When the ER field in a backward RM cell is lower than the current transmission rate recorded, it is updated with the ER value and hence VCs restricted downstream are detected. In order to identify a connection that is restricted upstream, the switches need to convey the LFS in forward RM cells [ChRL96]. Since this behavior is not stipulated by the ATM Forum traffic management standard, the DERA algorithm needs to operate in a homogeneous environment, where all switches along the network path from the sender to the receiver implement the DERA scheme. The implementation of the DERA algorithm used in chapter 6 incorporates some minor modifications in accordance to [GoSS98] to obey the ATM Forum Traffic Management Specification [TM4.1]. These consist of the adoption of the standardized end system behavior and the use of the ER field instead of the CCR field to convey the LFS in forward RM cells.

5.1.1.3 ERICA+

The *Explicit Rate Indication for Congestion Avoidance* (ERICA+) algorithm is a flow control scheme developed out of the OSU and ERICA algorithms [Kaly97]. This switch algorithm targets an optimal link utilization of 100 %, fairness, and a desired queue length for ABR traffic. The switch implementing ERICA+ will keep a "pocket" full of cells in its ABR queue, in order to fill in any short-term ABR capacities on the link [Kaly97]. This ABR queue has a First-In-First-Out (FIFO) scheduling strategy and is shared by all passing ABR connections through the corresponding link. The switch will perform measurements of the traffic load (ABR, CBR and VBR) and the ABR queue at regular intervals.

At first, ERICA+ needs to calculate the available bandwidth for ABR traffic (*TotalABRCapacity*), which is defined as

$$TotalABRCapacity = LinkBandwidth - CBRLoad - VBRLoad. \tag{5.3}$$

This value is then modified according to the current queue length (which depends on the current load), so that a certain amount of bandwidth is reserved for draining the queue. The target queue length *Q0* is determined dynamically by the *TotalABRCapacity* and a desired queueing delay *T0*, as

$$Q0 = TotalABRCapacity \cdot T0. \tag{5.4}$$

ERICA+ calculates a *TargetABRCapacity* based on the *TotalABRCapacity* and the current queue length or the corresponding queueing delay T_q as

$$TargetABRCapacity = TotalABRCapacity \cdot f(T_q), \tag{5.5}$$

where *f* is a queue control function [Kaly97]. There exist different options for an appropriate function *f*, that drives the queue towards the desired operating point *Q0*, but in general a hyperbolic function is recommended [VJGF98], [VJGF99]. The queue control function *f* should have a value greater or equal to one, when the queue level is below *Q0* and a value smaller than one, when the queue length exceeds *Q0*. This leads to two curves with parameters *a* and *b*. In order to avoid that too much of the available bandwidth is reserved for draining the ABR queue, the value of the queue control function is limited to a *Queue Drain Limit Factor* (QDLF), with a default value of 0.5 for WAN and 0.8 for LAN environments. The queue control function for ERICA+ is defined as

$$f(T_q) = \frac{b \cdot Q0}{(b-1) \cdot q + Q0} \text{ , for } 0 \le q \le Q0 \text{ and} \tag{5.6}$$

$$f(T_q) = \max\left(QDLF, \frac{a \cdot Q0}{(a-1) \cdot q + Q0}\right), \text{ for } q > Q0. \tag{5.7}$$

The recommended values, determined by simulations, for *a* and *b* are 1.15 and 1.05, respectively [Kaly97], where larger values result in a more intense reaction to queue errors.

Based on the *TargetABRCapacity* and the measured total *ABRInputRate* of all ABR connections, the *LoadFactor* is determined as

$$LoadFactor = \frac{ABRInputRate}{TargetABRCapacity} \text{ [Kaly00].} \quad (5.8)$$

Because highly variant background traffic can lead to fluctuating ABR capacities, an exponential averaging factor is used for the calculation of the *LoadFactor*. With the help of this *LoadFactor* a *VCShare* is computed for each VC, based on its Current Cell Rate (CCR)

$$VCShare = \frac{CCR}{LoadFactor} \text{ [Kaly97].} \quad (5.9)$$

Next, the *FairShare* of an ABR connection is determined as the ratio of the *TargetABRCapacity* and the number of active ABR sources *n*

$$FairShare = \frac{TargetABRCapacity}{n} \text{ [Kaly97].} \quad (5.10)$$

For the calculation of the number of active ABR sources, a decay factor may be applied to each VC, so that connections that stop sending data still contribute to the number of active sources in the following measurement intervals. This accounts for bursty ABR sources that will resume their transmission after an idle period [Kaly97].

The *Explicit Rate* (ER) calculated by the ERICA+ switch for an ABR connection is the maximum of the *VCShare* and the *FairShare*, not exceeding the *TargetABRCapacity*

$$ER = \min(\max(VCshare, FairShare), TargetABRCapacity) . \quad (5.11)$$

Whereas the calculations of the *ABRCapacity*, the *FairShare* and the *LoadFactor* are performed at each measurement interval for each link, the *VCShare* is determined whenever a backward RM cell is received. The *FairShare* guarantees a fair distribution of the available bandwidth, whereas the *VCShare* guarantees an optimal load [Fahm99].

A modification has been suggested regarding the calculation of the *Explicit Rate* in order to guarantee a fair distribution under all traffic configurations. The changes include an accounting of the maximum rate allocated in the current and previous averaging interval for each link and a max-min fairness parameter δ, which defines the steady state operating region. For a detailed description, please refer to [Kaly00].

These modifications are included in the ERICA+ implementation used in chapter 6. For the discrete event simulations presented in chapter 6, the configuration parameters of table 5-2 were used.

Parameter Name	Parameter Value
Averaging Interval	5 ms
T0	3 ms
Factor a	1.15

Table 5-2: *ERICA+ Configuration Parameters*

Parameter Name	Parameter Value
Factor b	1.05
Queue Drain Limit Factor	0.5
Fairness Parameter Delta	0.1
Decay Factor	0.9
Exponential Averaging Factor	0.8

Table 5-2: *ERICA+ Configuration Parameters*

5.1.2 VS/VD Technique

In the literature the use of the virtual source / virtual destination (VS/VD) technique has been suggested

- to shorten the ABR control loop [AlBH99], [Deat96], [Soum99],
- to isolate misbehaving ABR sources [Sait96],
- to generate feedback information at branch points of ABR multicast trees [ChCh00],
- as a gateway between
  - Guaranteed Frame Rate (GFR) and ABR connections [KKON99],
  - fixed networks and wireless ATM links [Goya98], [NuMa98],
  - credit- and rate-based ABR flow control schemes [RaNe95],
  - two ATM subnets [SmAT96].

[Soum99] proposed a dual VS/VD structure for ATM switch design, where one VS/VD pair is used to control the external ABR traffic and one internal VS/VD pair is used to control the ABR and UBR traffic inside the switch, from input to output port.

However, none of the works mentioned above defines an actual flow control algorithm for the implementation in a VS/VD switch. There exist only few contributions in which a VS/VD flow control algorithm is described [Baal99], [BaCs99], [CsKR97], [Goya97]. The most detailed description and analysis is provided for an extension of the ERICA+ scheme for VS/VD switches [Goya97], which will be compared in chapter 6 to the algorithm developed in this thesis.

The enhancements of the ERICA+ algorithm for VS/VD switches, called VERICA in the following, are based on the special layout of the VS/VD switch, which uses per-VC queueing, see chapter 3.3. VERICA operates with three *Explicit Rate* (ER) values for each VC (see figure 5-1):

- *ERfeedback*, that is sent to the source in the previous loop via backward RM (BRM1) cells, i.e., turned around forward RM cells (FRM1),

- *ERinternal*, that is used to control the transmission rate of the VS from the per-VC queue into the ABR class queue, and
- *ERexternal*, that is conveyed by the BRM2 cells received by the VS in the next loop [Goya97].

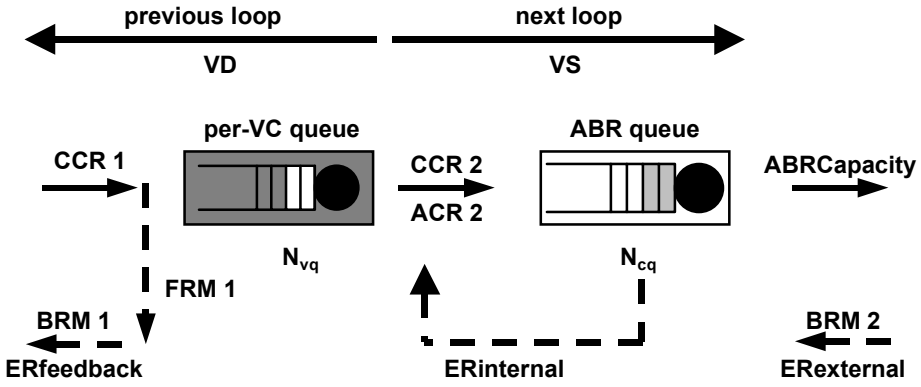

Figure 5-1: *ERICA+ for VS/VD*

Whenever a BRM cell is received in the next loop by the VS, its value is copied to *ERexternal*. When a forward RM cell is received in the previous loop by the VD, it is returned to the source of the previous loop as a BRM cell with an ER value of *ERfeedback*. The algorithm executes the following steps at each measurement interval [Goya98].

$$OverLoad = \frac{\sum_{\text{VCs to link}} CCR2}{f\left(N_{cq}\right) \cdot ABRCapacity} \tag{5.12}$$

$$ERinternal = \min\left(\max\left(\frac{CCR2}{Overload}, \frac{f\left(N_{cq}\right) \cdot ABRCapacity}{N}\right), ERexternal\right) \tag{5.13}$$

$$ACR2 = fn(ERinternal, \text{ end system rules}) \tag{5.14}$$

$$ERfeedback = f\left(N_{vq}\right) \cdot ACR2 \tag{5.15}$$

The variables used in figure 5-1 and in the equations above are defined as:

- *CCR1*, the *Current Cell Rate* in the previous loop that feeds the per-VC queue of the VD,
- *CCR2*, the measured *Current Cell Rate* in the next loop from the per-VC queue to the ABR queue,
- *ACR2*, the *Allowed Cell Rate* in the next loop,
- *ABRCapacity*, the cell rate at which the ABR queue can transmit,
- N_{vq}, the level of the per-VC queue,
- N_{cq}, the occupancy of the ABR queue, and
- *N*, the number of ABR connections.

The *ACR2* is determined by a function *fn* of the ATM Forum end system rules and the value of *ERinternal*. The *CCR1* of the previous loop is defined by *ERinternal* of the previous feedback cycle [Goya97]. The queue control function *f* is a hyperbolic function, as defined for ERICA+ (see eq. 5.6 and 5.7).

For the discrete event simulations presented in chapter 6, the configuration parameters listed in table 5-2 were used.

Parameter Name	Parameter Value
Averaging Interval	5 ms
T0 (ABR class queue)	3 ms
Factor a (ABR class queue)	1.15
Factor b (ABR class queue)	1.05
Queue Drain Limit Factor (ABR class queue)	0.5
T0 (per-VC queue)	0.3 ms
Factor a (per-VC queue)	1.05
Factor b (per-VC queue)	1.0
Queue Drain Limit Factor (per-VC queue)	0.5
Fairness Parameter Delta	0.1
Decay Factor	0.9
Exponential Averaging Factor	0.8

Table 5-3: *ERICA+ for VS/VD Configuration Parameters*

5.1.3 Use of Control Theory

Since the ABR flow control scheme establishes a closed-loop feedback control system, there have been several proposals to use classical control theory for system analysis [Shen90], [Kesh91], [TKCT98].

In one early work [BeMe93] suggested a control algorithm that is composed of a *proportional-derivative* (PD) controller, based on a queue threshold, and a controller of order *d*, based on recent transmission rates. The algorithm depends on the last *d* advertized rates and the last two queue levels, where *d* depends on the longest feedback delay of all connections. The model has therefore $d + 2$ gain values or parameters and the poles of the system transfer function have to

be placed appropriately to obtain a stable system. Hence, the resulting algorithm is fairly complex and difficult to implement. When the number of active ABR sources changes, the controller has either to be adaptive (new calculation of gain values) or robust, in the sense that it will perform in an acceptable manner for a different number of sources. For this reason, [KoRa97a] uses two PD controller of dimension $d + 2$, one for high (large number of connections) and one for low (small number of connections) load conditions. The corresponding gain values are calculated off-line [KoRa97c]. In later improvements, [BeWa98] proposed a table driven approach with gain values that are calculated off-line and incorporating per-VC queueing. [BeWa98], as well as [KiTh98a] and [KiTh98b], solve a *linear quadratic* (LQ) control problem instead of performing a pole placement. The optimal control problem is called LQ, because the cost function to be minimized has a quadratic form and the resulting optimal controller is a linear function of the system state [BeWa98]. Furthermore, the application of standard controller types, like PD, for binary feedback algorithms has been proposed [ABLM96], [RoBO95].

A very common approach is the use of a simple *first-order proportional* controller [ChNW98a], [ChNW98b]. It will determine an explicit rate that is proportional to the deviation of the ABR queue level with respect to a target length [HaBA99], [LeKa00], [PeJa99], [ZhYM97a]. This idea has also been used for explicit rate algorithms that operate hop-by-hop [WeHW98], [MiKT96], [ZhYM00] instead of end-to-end. On the one hand, this approach offers the major advantage that only one gain parameter has to be selected as a controller parameter. On the other hand, the steady state error of the system is not equal to zero, resulting in rate oscillations [HaSA99], [RoBe97]. The combination of a first-order proportional controller and a Smith predictor, to eliminate the feedback delay in the control loop, has been suggested by [CaMG96], [Masc97b], [Masc97a], [Masc99]. The number of cells in flight (between source and switch) for each connection is predicted, which requires per-VC queueing [MaCG97].

In order to combine the idea of rate and queue control, the use of two controllers was proposed by [NaSi96]. There exists a *rate reference* controller matching the source rate to the available bandwidth and a *queue reference* controller driving the queue to a desired operating point. The latter controller operates on an effective (or virtual) queue with no delay [NaSi97]. Alternatively, an observer that estimates the queue size (Smith predictor or a per-VC accounting model) may be used [Narv97]. The algorithm assumes a special source behavior, i.e., the source stops its transmission after a specific time period conveyed through backward RM cells [Narv97], which makes it incompatible to the ATM Forum traffic management standard [TM4.1]. In [SuVW00] an algorithm was proposed, that incorporates rate control proportional to queue length, available bandwidth, and number of active sources. An ABR source is marked as active, if its transmission rate exceeds a certain threshold. This fact is conveyed to the switch by setting a proprietary bit in forward RM cells [SuVW00]. In addition it is assumed, that an active source may increase its transmission rate only linearly [SuVW00]. These assumptions are not consistent with the definitions of the ATM Forum Traffic Management Specification [TM4.1].

Some of the related works deal with the analysis of ABR algorithms using *delay differential equations*. The developed algorithms are designed either for binary [BoSh90], [BoSh92],

[BoMS95], [RoBO95], or explicit rate end-to-end control [Elwa95], [Izma96]. The same idea has been also used for hop-by-hop algorithms [ZhYM97b].

In the area of *optimal control theory*, stochastic equations are used to define a centralized control problem with action delays [Srik99]. It is shown that this problem, i.e., the switch has to assign transmission rates that are experienced at the switch only after an action delay, is equivalent to a decentralized team problem with information delays [AlBS98], [AlBS97]. That means that each team member (ABR source) has access to differently delayed information (explicit rates), but the members pursue a common goal, which is characterized by a linear quadratic Gaussian (LQG) cost function [ImBa99a], [ImBa99b]. The cost function may define certain performance goals, like good tracking of the available capacity or matching a desired queue level [AlBS99]. To find a solution for this cost function, a discrete-time algebraic Riccati equation has to be solved. Because the available capacity for ABR traffic is modeled as an autoregressive (AR) process, the dimension of the Riccati equation is determined by the magnitude of the largest system delay and the order of the AR process [AlBS99]. The advantage of this approach is, that an optimal controller with respect to a defined cost function may be derived. The major drawback is the complexity involved in solving the system state matrices and the fact that the parameters of the AR process have to be known.

Instead of an LQG cost function, an H^{∞}-controller may also be used to solve the optimal control problem [AlBH99], [Atas00], [OzKI98], [OKKI99], [ShMa98]. Alternatively H^2-optimal control theory can be applied, when the optimization problem consists of driving the low frequency, unused ABR capacity to zero [ZhLS97].

[PiLa97] suggests a *predictive adaptive scheme* with feedforward of the measured disturbance and feedback of the output, but relies on the incorporation of the Call Admission Control (CAC) procedure.

For the choice of an appropriate control model, the frequency of feedback information has to be taken into account. The feedback or action delay for the ABR control loop consists of the inter-RM cell interval plus the backward and forward propagation delays, as explained in section 5.4. Whenever a backward RM (BRM) cell is received by the switch, it is forwarded to the source on the return path. The calculated ER value is placed in the ER field of the RM cell, if it is lower than the current value contained in this field. For this reason, a possible control model would be a discrete time model with a sampling period equal to the inter-RM cell interval. Because of the fact that the frequency of RM cells depends on the Current Cell Rate (CCR) of the ABR source (see chapter 3.2.3), this inter-RM cell interval however is variable. Possible models for dealing with varying feedback intervals, that have been suggested in the literature, are:

- the assumption of *periodic feedback* information, ignoring the fact of variable RM intervals [AlBH99],
- a *discrete time model* operating with a fixed sample interval, where missing feedback information in one time period is estimated [MaCG97]

- incorporating the fact of a *varying sample time* into a discrete time model, but assuming non-standard source behavior in the stability analysis [NaSi97], or
- a *continuous time modeling* [ChNW98a],[ChNW98b], [NaSi96], [SuVW00].

After this overview of related works in the area of ABR flow control algorithms, the use of the VS/VD technique, and the application of classical control theory, the following sections focus on the newly proposed approach.

5.2 VS/VD Switch Layout

As outlined in chapter 3.3.1, VS/VD switches must have the capability to control the *Allowed Cell Rates* (ACRs) of ABR connections individually and therefore have to implement per-VC queueing. In this work the following design of the VS/VD switch is assumed, see figure 5-2, which is also used by [Goya97], [Goya98], [Kaly98a].

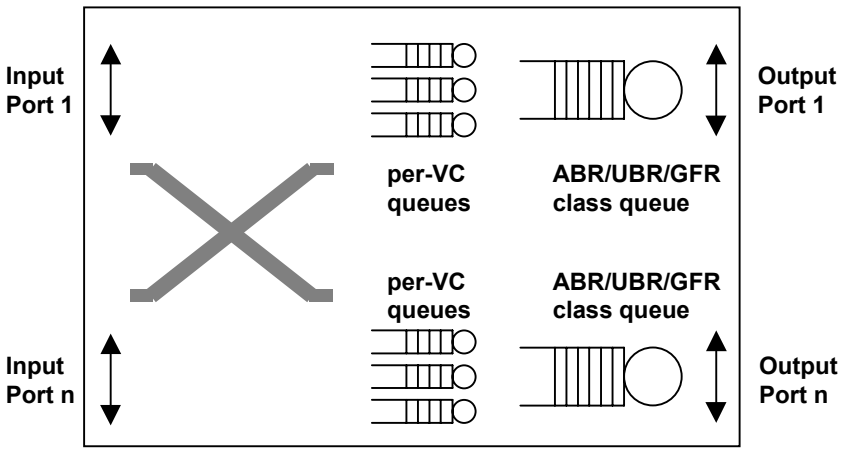

Figure 5-2: *VS/VD Switch Architecture*

The cells of an ABR VC are switched from the input port to an output port, where they enter a per-VC queue. Next, the cells are sent to a common low priority class queue for UBR, GFR, and ABR traffic of the corresponding output port. At each output port a high priority class queue for CBR/VBR traffic is present as well, which is not shown in figure 5-2 for conciseness. The high priority queue is served with strict priority over the low priority (ABR/UBR/GFR) class queue, so that ABR traffic is sent only, when no high priority cells are waiting for transmission.

Figure 5-3 illustrates the cell flow of an ABR VC passing through a VS/VD switch according to the above layout. The source feeds data into its own VC-queue from where it is sent out with the Current Cell Rate of the previous loop (*CCR1*). After reception at the corresponding VC queue of the virtual destination (VD) inside the VS/VD switch, data cells are forwarded with a rate of *CCR2* to the low priority class queue at the virtual source (VS). The *CCR2* may be lower than the Allowed Cell Rate for the next loop (*ACR2*), it is however equal to *ACR2* at the bottle-

neck switch along the path. A switch represents the bottleneck for a VC, if its allocated rate for the connection is the minimum on the path from the sender to the receiver. Forward RM cells of the first segment (*FRM1*) are returned to the source as backward RM cells (*BRM1*) by the VD, giving direct feedback without waiting for backward RM cells in the next loop. At the VS the data cells and forward RM cells for the second segment (*FRM2*), generated by the switch, are transmitted to the destination. These *FRM2* cells are then turned around by the destination as *BRM2* cells and are interpreted by the VS to determine the Allowed Cell Rate in the next loop (*ACR2*).

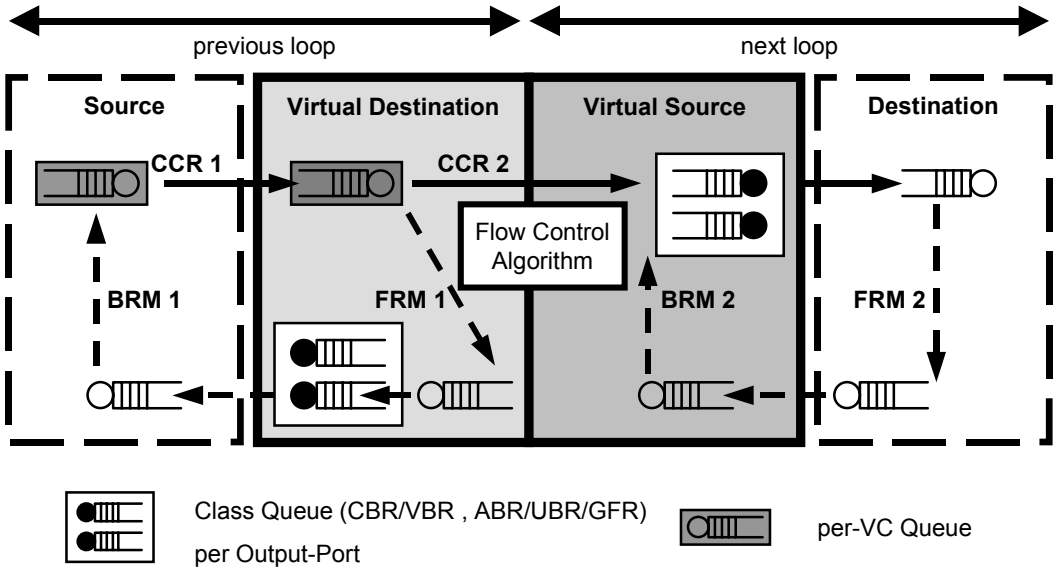

Figure 5-3: *VS/VD Data and RM Cell Flow*

It has to be noted that this is a conceptual model only, which does not stipulate any special implementation in a practical system. For example, it might be reasonable to implement a separate UBR queue, since UBR traffic can be admitted to the network without restrictions. Hence, UBR VCs may possibly dominate a common low priority queue, since ABR connections reduce their rate in the presence of congestion, leaving more space for UBR traffic. In the following we will assume that no other low priority traffic, i.e., UBR or GFR, is present and will refer to the low priority queue as ABR class queue.

5.3 Design Goals

The major contribution of this work consists in developing an ABR flow control algorithm, analytically designed with the help of linear control theory and tailored for the special layout of a VS/VD switch. The reason for using linear control theory for the design of the VS/VD flow control algorithm is to ensure certain performance properties of the algorithm. Revisiting the performance goals for ABR flow control schemes (see chapter 3.1.1) and closed-loop feedback

control systems (see chapter 4.3), the following objectives can be identified and will be taken into account for the design.

- **Efficiency:** Whereas in general a low buffer occupancy is desirable from the viewpoint of delay and jitter, a certain amount of buffering is advantageous when short-term ABR bandwidth capacities shall be exploited. Therefore, the goal of efficiency combines the two conflicting aims of low buffer levels and high link utilization, which will be addressed separately.
  - **Buffer occupancy**: Like in the ERICA+ concept [Kaly97], the ABR class queue may be filled to a certain threshold to keep a reservoir of cells, which can be used to fill short-term gaps, without notifying the source to increase its rate. Since the ABR class queue is fed by several VCs, each source needs to contribute only in part in driving the class queue back to its target value. Large queue levels however would lead to long queueing delays and are a possible source of jitter. The maximum class queue length should be sufficiently larger than the target value, in order to allow an additional buffering to avoid cell loss in case of buffer overshoots.

    The per-VC queue has to be allocated for each ABR connection passing through the VS/VD switch and hence its maximum queue length is determined by the available buffer space at the VS/VD switch. Furthermore, the scalability of the algorithm with respect to the number of supported ABR connections depends on the buffer requirement for the per-VC queues. On the one hand, these reasons suggest that the maximum per-VC buffer capacity should be kept low. On the other hand, the per-VC queue will have to buffer transient overloads of a single ABR connection. Especially in the case of a large bandwidth delay product in the previous loop, this might lead to a large buffer requirement. Therefore, it can be concluded that there exists a trade-off between scalability and low cell loss at the per-VC queues. This problem may be alleviated by the use of buffer pool techniques, that are offered by modern ATM switches [BWBM98], [EAYN99], [Soum99]. This mechanism allows an on-demand assignment of buffer space from a common buffer pool to individual connections. It has to be noted however, that an overload situation for a link affects all passing ABR VCs and that all VCs with a large feedback delay in the previous loop may require substantial additional buffer space at the same time.

    Since the corresponding per-VC queue has to be passed by all ABR cells in addition to the ABR class queue, its target value shall be kept low. This design decision is supported by the fact, that the per-VC queue is fed by a single connection only and large queue deviations from the target level could directly affect the calculated explicit rate for the ABR source.

    A discussion of actual target levels for the per-VC and ABR class queues depending on the link rate for the new VS/VD switch algorithm will be presented in section 5.6.1.

  - **Link utilization**: A high link utilization is desirable to obtain a high throughput for low priority traffic and to fully use the available link bandwidth. In order to adapt quickly to

the desired cell rate and to avoid rate oscillations, the newly developed VS/VD algorithm will use an explicit rate indication and not a binary scheme.

Short-term capacities with a duration less than the feedback delay pose a special problem to ABR flow control. The available bandwidth can not be used by a rate increase notification to the source, because of the action delay of the control loop. The increased traffic load would be experienced at the VS/VD switch only after the capacity was available. As mentioned above, for this reason a certain reservoir of cells buffered at the ABR class queue is desirable.

Another problem for bandwidth usage is the static reservation of a certain amount of unused link capacity. Some of the existing ABR flow control algorithms target a link utilization below 100 % at steady state, in order to reserve bandwidth for new ABR connections [AfMO00] or to drain ABR queues, that have built-up during periods of transient overloads [Kaly00]. As the available bandwidth should be used completely for an efficient operation, the newly designed VS/VD algorithm tries to achieve a 100 % link utilization. Bandwidth for draining the ABR class queue is allocated dynamically when needed.

The method for detecting unused bandwidth available for ABR traffic also affects the link utilization. Whereas some formerly proposed algorithms, e.g., [KoRa97a], [KoRa97b], use information on bandwidth allocation to high priority traffic (CBR/VBR) from the Call Admission Control (CAC), this approach does not allow to access reserved bandwidth currently not used by CBR or VBR traffic. For this reason, the VS/VD algorithm proposed in this work measures the actual high priority traffic to determine the ABR capacity.

- **Fairness**: An ABR flow control algorithm should distribute the available bandwidth in a fair manner among the competing sources. From the view point of the VS/VD switch, the passing ABR connections may be either locally bottlenecked by the corresponding switch or locally unconstrained, i.e., constrained elsewhere on the network path, see the DERA in section 5.1.1.2. Since the transmission rate of the locally unconstrained VCs may not be controlled by the VS/VD switch, the ABR capacity minus the bandwidth used by these unconstrained VCs should be divided among the bottlenecked connections. The ATM Forum presents in its Traffic Management Specification a total of six fairness definitions, using the following variables [TM4.1].

  ABRCapacity = total available bandwidth for all ABR connections on a given link.

  CCR_U = sum of bandwidth of ABR connections bottlenecked elsewhere. This includes source-bottlenecked connections, that are limited by the Peak Cell Rate of the sender.

  TotalFairShare = *ABRCapacity* - CCR_U, bandwidth to be shared by ABR connections bottlenecked on this link.

  n = total number of active ABR connections.

  u = number of active ABR connections bottlenecked elsewhere.

$r = n - u$, number of active ABR connections bottlenecked on this link.

MCR_r = sum of MCRs of active ABR connections within r.

$FairShare_i$ = fair allocation for ABR connection i.

MCR_i = Minimum Cell Rate of ABR connection i.

w_i = pre-assigned weight of ABR connection i.

a. **MCR Plus Weighted Share**: This fairness definition is the most general of the six terms and includes four of the others (b. to e.) as special cases.

$$FairShare_i = MCR_i + (TotalFairShare - MCR_r) \cdot w_i \Big/ \sum_{j=1}^{r} w_j \qquad (5.16)$$

The fair allocation to connection i is defined as its MCR plus a weighted share of the *TotalFairShare* reduced by the MCRs of the active bottlenecked connections [Vand99], [Vand00].

b. **Max-Min Fairness**: This very common fairness definition has been used in many of the already proposed ABR flow control algorithms [AbKu97], [Char94], [Kala95], [Kaly00], [KoRa99], [LaTs99], [PrKB98], [TzSi97].

$$FairShare_i = \frac{TotalFairShare}{r} \qquad (5.17)$$

The *TotalFairShare* is equally shared by the r connections bottlenecked at the link. An intuitive, but informal definition states, that a bandwidth allocation at a switch is max-min fair, if none of the bandwidth assignments may be increased without decreasing one of the smaller rates, i.e., *maximize* the *minimum* rate among all sessions [HoTP97a]. Max-min fairness corresponds to "MCR plus weighted share", where all MCRs are zero and all VCs have an equal weight.

c. **MCR Plus Equal Share**: The fair allocation for a connection i is defined as its MCR plus an equal share of the *TotalFairShare* reduced by the MCRs of the active bottlenecked connections.

$$FairShare_i = MCR_i + \frac{TotalFairShare - MCR_r}{r} \qquad (5.18)$$

This fairness criterion corresponds to "MCR plus weighted share", where all VCs are equally weighted. Furthermore, when all MCRs are zero, it corresponds to max-min fairness.

d. **Proportional to MCR**: The fair allocation for a connection i is weighted according to its MCR.

$$FairShare_i = TotalFairShare \cdot \frac{MCR_i}{MCR_r} \qquad (5.19)$$

This criterion applies only, when all MCRs are greater than zero. The fairness definition corresponds to "MCR plus weighted share", if the MCRs are chosen as weights and the MCRs are set to zero during the calculation.

e. **Weighted Allocation**: The fair allocation for a connection *i* is proportional to its assigned weight.

$$FairShare_i = TotalFairShare \cdot w_i \Big/ \sum_{j=1}^{r} w_j \tag{5.20}$$

The criterion corresponds to "MCR plus weighted share", where the MCRs are set to zero. When all weights are equal, this fairness definition corresponds to max-min fairness.

f. **Maximum of MCR or Max-Min Share**: The fair allocation for a connection *i* is the maximum of its MCR and its max-min fair share. Again, if all MCRs are zero, this criterion corresponds to max-min fairness.

Since the novel VS/VD algorithm developed in this thesis should support MCR guarantees and no other external information on user VCs, e.g., preassigned weights, are provided, the "MCR plus equal share" criterion is chosen. It will take into account MCR constraints and distributes the remaining bandwidth equally among all bottlenecked connections, not favoring VCs with large MCRs as "Proportional to MCR" or penalizing them as "Maximum of MCR or Max-Min Share". When comparing the new VS/VD algorithm developed in this thesis with existing ER (see section 5.1.1) or VS/VD (see section 5.1.2) algorithms in chapter 6, that support only max-min fairness, the MCRs will be adapted accordingly.

It has to be noted that ABR flow control algorithms that implement per-VC queueing may support arbitrary fairness definitions by the use of an appropriate cell scheduling discipline [BeWa98], [ChRL96], [ChWa97], [LoHR00], [Masc97a]. Because all VS/VD switch algorithms must implement per-VC queueing, the task of a fair bandwidth allocation could be assigned to the cell scheduler. In this work however, no special cell scheduling algorithm will be assumed, allowing for simple implementations like strict priority scheduling of class queues. As a consequence, the newly designed VS/VD algorithm will explicitly take into account the fairness criterion "MCR plus equal share" as defined above.

- **Complexity**: As mentioned in chapter 3.3, the complex implementation of a VS/VD switch is the major drawback of this technique. Since in the design of a VS/VD switch (presented in section 5.2) a simple priority queueing discipline is assumed, the processing power provided may mainly be used to implement the ABR end system rules (see chapter 3.2.3) and the VS/VD switch algorithm. The source and destination behavior has to be executed for each ABR connection passing the switch, requiring per-VC state information and queueing. This implementation complexity is an inherent part of the VS/VD concept and can not be avoided without violating the ATM Forum Traffic Management Specification [TM4.1].

  It was already stated above, that for reasons of link utilization an Explicit Rate (ER) algorithm shall be used. Since the VS/VD switch carries already the burden of implementing the ABR end system behavior, the ER algorithm shall not impose a large additional computing load. As a consequence, the developed VS/VD algorithm should have a simple structure and a small number of parameters, that may be determined without a costly computation. Espe-

cially the use of optimal control theory, where large state space matrices have to be solved to compute an optimal solution, is not advisable.

- **Scalability**: Due to the fact that modern ATM switches offer a modular design [BWBM98], [EAYN99], [Soum99], [Kou99] scalability may be addressed in the context of line cards that service a single connection link. The most important factor is the number of concurrent ABR connections supported. As already mentioned when discussing the complexity in the previous paragraph, state information for each ABR VC has to be stored and processed. The main problem for scalability and hence the limiting factor for the number of ABR connections supported, is the allocation of per-VC buffer space. In defining an upper limit for the number of concurrent ABR VCs, it can be assumed that about 2/3 of the link capacity will be allocated to high priority traffic and that a bandwidth of at least 64 kbit/s per ABR VC is desirable. A detailed discussion of implementation issues is presented in section 5.6.
- **Conformity**: A major goal for the VS/VD algorithm is conformity to the ATM Forum Traffic Management Specification [TM4.1]. The key idea of the VS/VD technique is its transparent operation, that does not require any changes to existing ABR end systems or switches. The other network elements, i.e., end systems and switches, are not aware of the fact, that a VS/VD switch is in operation along the network path from the sender to the receiver. The VS/VD switch needs to inter-operate with a standard ABR source end system in the previous loop and a standard destination end system in the next loop. In order to be inter-operable with other network elements, the use of proprietary formats for RM cells or the assumption of any non-standard behavior is strictly prohibited. Furthermore, the VS/VD algorithm shall not require a certain network element as the next hop in the previous or next loop, but has to allow for another VS/VD switch, a standard ATM switch or an end system as a legal next neighbor.
- **Stability**: The use of linear control theory for the design of the VS/VD flow control algorithm enables a mathematical assessment of its stability. The closed-loop transfer function of the VS/VD algorithm may be analyzed with respect to the Nyquist criterion (see chapter 4.4.3) guaranteeing a stable operation. With the help of stability analysis, controller parameters can be selected that result in a stable operation. In contrast to choosing controller parameters based on heuristics or certain simulation scenarios, stability analysis offers the key advantage of analytically proven performance characteristics under all possible conditions. As mentioned in chapter 4.4.3, stability is of crucial importance for a control system, since an unstable operation could lead to unbounded buffer occupancies with a deteriorating effect on ABR performance.
- **Robustness**: In order to develop a robust algorithm, that is tolerant against small inaccuracies in the system model, a sufficient relative stability (defined by the phase and gain margin, see chapter 4.4.4) will be targeted during the controller design. On the one hand a large phase margin offers the advantage of a very robust system, but on the other hand the swiftness of the system response to changes of the input variable is reduced [PhHa91]. Therefore, there exists a trade-off between the relative stability and the transient behavior (see next point) of a control system [Dorf89]. In order to assure a good responsiveness, a phase mar-

gin between 40 and 45 degrees and a gain margin of at least 5 dB are targeted. These values are slightly smaller than the recommended standard values of 45 degrees and 8 dB [PhHa91], [Ebel87], which put an emphasis on stability.

- **Transient Behavior**: The transient behavior of the VS/VD flow control algorithm can be described by the performance metrics introduced in chapter 4.3.3. In general it is not possible to obtain a short rise and settling time as well as a small overshoot [PhHa91]. Following the discussion of buffer occupancy above, it can be assumed that the controlled variables are the buffer levels of the per-VC and the ABR class queue. The settling time of these two queues depends on the feedback or action delay in the previous loop, delaying the impact of the control action taken. Therefore, newly developed VS/VD algorithm should achieve a settling time of 4 times the feedback delay. The overshoot of the class queue should not exceed a factor of 3 times (or 200 % of) the target value, in order to avoid large queue levels (see discussion of buffer occupancy above). Since the per-VC queues will have to handle the transient overload for an individual VC and the average queue level is small, an overshoot of a factor of 4 to 5 of the target value may be accepted. It has to be noted that at the initial start-up of the system, rise and settling times depend on the *Initial Cell Rates* (ICRs) of the ABR connections and the network topology. The fair rates have to be propagated through the exchange of RM cells in the different control loops and the per-VC and ABR class queues have to be filled to the desired levels. Therefore, the aforementioned rise and settling times as well as the buffer overshoots are valid only for a step change of the input signal fed into the controller, starting at steady state.
- **Steady State Error**: In order to avoid unnecessary fluctuation of the controlled variable, a steady state error of zero for a step change in the input signal of the controller should be obtained. For a ramp error, a bounded steady state error is acceptable.
- **Disturbance Rejection**: The rates at which the per-VC or the ABR class queue are drained, namely, the Current Cell Rate of the next loop (*CCR2*) or the current ABR capacity, respectively, may be modeled as disturbance signals. This allows the controller to directly react on rate changes without waiting for a queue error to built up. Besides keeping the queue lengths of the per-VC and the ABR class queue at the target level, the tracking of the rate at which these queues are drained, is an important task of the controller. Both rates are known at the VS/VD switch, because the *CCR2* is determined as part of the ABR source rules of the VS and the current ABR capacity may be measured as defined in eq. 5.3. These information can be taken into account by the controller as a feedforward signal for the calculation of the desired input rates to the per-VC and ABR class queues, respectively.

After this analysis of the performance goals, it can be concluded, that the newly developed, explicit rate VS/VD flow control algorithm must control the queue lengths of the per-VC and the ABR class queue, considering the Current Cell Rate of the VS (*CCR2*) and the measured ABR capacity. Furthermore, it has to guarantee "MCR plus equal share" fairness. This has to be achieved by a simple control law, whose parameters must assure a stable operation of the control system. In order to derive a linear differential equation that describes the queue dynamics as a basis for a control model, an appropriate traffic model has to be defined first.

5.4 Traffic Model

Since the design of the new VS/VD switch algorithm is based on linear control theory, a traffic model is needed, that will define the temporal behavior of the controlled variable, i.e., the buffer level. A common approach presented in many of the related works consists of a fluid flow model describing the traffic flow at a queueing point [AlBS99], [ChNW98a], [GhMa97b], [HaSA99], [Izma96], [KoRa99], [WeHW98], [MaCG97], [MiKT96], [NaSi96], [PaAg95], [RoBO95], [RoMa99], [SuVW00], [WoBo98], [ZhYM00][WeHW98]. The fluid flow model uses a continuous-space deterministic process as an approximation of an underlying stochastic process describing the traffic flow at a buffer [BeMe93], [BoSh92]. The model is based on first-order delay-differential equations [BoMS95], where no cell boundaries are taken into consideration (fluid flow) [Elwa95]. The queue level N_q at a certain instant in time depends on the previous buffer occupancy and the difference between the number of cells that enter and that depart from the queue during this time interval.

$$N_q(t+\Delta t) = Sat_{Nqmax}\left\{N_q(t) + \Delta t \cdot \left(In(t) - Out(t)\right)\right\}, \text{ where} \tag{5.21}$$

$$Sat_x(t) = \begin{cases} x \text{ if } t \geq x \\ \text{t else} \\ 0 \text{ if } t \leq 0 \end{cases}$$

assuming that the input rate $In(t)$ and the output rate $Out(t)$ remain constant over the interval Δt and are measured in the unit of cells/ s. The saturation function accounts for the fact, that a buffer level can not be negative or exceed a certain maximum value. In order to derive a model for a linear control system, the saturation boundaries in eq. 5.21 have to be removed. As a consequence, the model is valid only when the buffer level is close to the desired target value and does not reach the lower or upper boundary [KoRa99]. Calculating the derivation leads to

$$N_q'(t) = In(t) - Out(t)\ . \tag{5.22}$$

Hence, the change in the queue level corresponds to the difference between its input and output rate.

When applying this model in the context of Explicit Rate ABR flow control, the ABR queue of the bottleneck switch is of interest. This switch controls the remote source, in the sense that its advertized Explicit Rate is the minimum on the path from the sender to the receiver. From the viewpoint of the other switches along the network path, the VC is constrained elsewhere and its rate can not be influenced. Taking the position of the bottleneck switch, the input rate to the ABR queue is determined by the source rate and hence by the Explicit Rate calculated by the switch. Due to the feedback delay between the switch and the remote source, the advertized Explicit Rate is experienced as an input to the ABR queue only after this time delay. This feedback or action delay τ_{action} is composed of:

- the RM cell delay τ_{RM} until a returned RM cell passes the switch on its way back to the source,

- the backward propagation delay τ_{br} until the BRM cell reaches the source, and
- the forward delay τ_{fr} until the next data cell reaches the switch.

$$\tau_{action} = \tau_{RM} + \tau_{br} + \tau_{fr} \tag{5.23}$$

As a consequence, the input rate $In(t)$ in eq. 5.22 may be substituted by the Explicit Rate calculated by the switch at the time t - τ_{action}, i.e., $ER(t - \tau_{action})$. In order to obtain a transfer function for the ABR buffer model at the switch, the Laplace transform is applied to the modified eq. 5.22, describing the queue dynamics. This results in

$$N_q(s) = \frac{1}{s}\left(e^{-\tau_{\text{action}} \cdot s} \cdot ER(s) - Out(s)\right), \tag{5.24}$$

where $Nq(s)$, $ER(s)$, and $Out(s)$ are the Laplace transforms of $Nq(t)$, $ER(t)$, and $Out(t)$, respectively. The corresponding block diagram is depicted in figure 5-4.

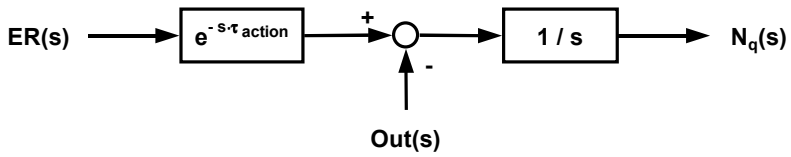

Figure 5-4: *Block Diagram of Queue Model*

The queue has an integrative transfer function, that converts an incoming rate signal (in cells/ s) to the number of cells in the queue. Based on this queue model and the differential equation defining the queue dynamics, a control model for the novel VS/VD flow control algorithm may be designed.

5.5 Control Model

In this section the newly designed ABR flow control algorithm for a VS/VD switch will be presented. Taking into account the VS/VD switch layout, the design goals, and the traffic model described in the previous sections, an ABR flow control algorithm will be developed using control theory.

As concluded in section 5.3, the VS/VD flow control algorithm is responsible for keeping the per-VC queues, existing for each ABR connection, and the common ABR class queue, shared by all ABR VCs, filled at the desired levels. Whereas the per-VC queues are fed by the ABR sources transmitting at the Current Cell Rate of the previous loop (*CCR1*), the ABR class queue is fed by all ABR connections in parallel, each sending at the Current Cell Rate of the next loop (*CCR2*). Furthermore, the fair distribution of the ABR capacity, which drains the ABR class queue, is also a task affecting the rates of all ABR VCs. Hence, it is a straightforward and reasonable design decision to develop two controllers, that are responsible for the previous and the next loop, respectively. This approach reflects also the layout and architecture of a VS/VD switch as presented in section 5.2.

- The controller for the previous loop will control the buffer level of the per-VC queue by calculating an Explicit Rate for the previous loop (*ER1*) for each connection.
- The controller for the downstream loop is responsible for the fair allocation of the available ABR capacity to the bottlenecked connections and for the control of the ABR class queue buffer level. It calculates for each ABR VC an Explicit Rate for the next loop (*ER2*), that is used by the corresponding virtual source (VS).

Since the per-VC queue is drained into the ABR class queue by *CCR2*, the two controllers are coupled by this variable. When taking the position of the bottleneck VS/VD switch that determines the rate of the ABR source for the corresponding connection, the controller of the next loop determines *ER2* and hence *CCR2*. *CCR2* in turn is the rate that empties the corresponding per-VC queue and therefore has an influence on *ER1*.

The dual controller structure is analogous to the VERICA+ concept, see figure 5-1. The rate calculation of VERICA+ however uses function parameters derived by heuristics and simulation [Goya98]. Other related works have also proposed the use of a dual controller for a standard ABR switch, operating as the combination of a rate and a queue driven controller [AiHo99], [NaSi96], [NaSi97], [Narv97]. Whereas the algorithm of [AiHo99] is based on thermodynamics and statistical physics, [NaSi97] use a proportional controller operating on a virtual queue plus rate measurements. Both proposals combine the output of the two controllers to a single Explicit Rate indication.

In the following the design of the two controllers for the previous (upstream) and the next (downstream) loop as well as the joint control structure of the VS/VD switch is introduced.

5.5.1 Upstream Loop

The first control model describes the controller operating on the per-VC queue of a single ABR VC and calculating the Explicit Rate for the previous or upstream loop (*ER1*). In the following the viewpoint of the bottleneck VS/VD switch for the corresponding VCs is assumed. That is, the VS/VD switch determines the rate of the ABR source by indicating the lowest Explicit Rate of all switches along the network path from the sender to the receiver. Furthermore, a greedy source is assumed, that is able to transmit data at the corresponding Explicit Rate, i.e., the connection is not bottlenecked or constrained at the source. Therefore, the Explicit Rate calculated by the VS/VD switch for the previous loop (*ER1*) equals the Allowed Cell Rate of the source (*ACR1*) and its Current Cell Rate (*CCR1*).

For the modeling of the previous control loop, a continuous time feedback system is assumed, in order to apply standard techniques, like the Nyquist criterion, for the stability analysis. The feedback delay introduced by the periodic transmission of RM cells is incorporated in the action delay $\tau_{1action}$ of the traffic model, see figure 5-4.

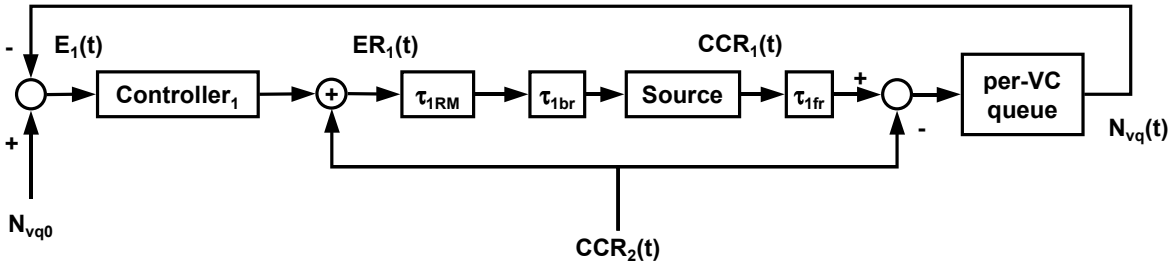

Figure 5-5: *Feedback Control System of Upstream Loop*

For a single ABR connection at the bottleneck VS/VD switch, the previous loop control system is described by figure 5-5, where

- $CCR_i(t)$ = Current Cell Rate of source in loop i (1 = previous loop, 2 = next loop), i.e., CCR_1 = Current Cell Rate of upstream source, CCR_2 = Current Cell Rate of virtual source,
- $ER_1(t)$ = Explicit Rate signaled to the upstream source,
- $E_1(t)$ = error signal (difference between desired and actual per-VC queue level),
- τ_{RM} = delay until next backward RM cell, *br* = backward path, *fr* = forward path),
- τ_{br} = propagation delay on backward path,
- τ_{fr} = propagation delay on forward path,
- N_{vq0} = desired queue level for per-VC queue, and
- $N_{vq}(t)$ = queue length of per-VC queue.

It has to be noted that the per-VC queue is drained with CCR_2, the Current Cell Rate of the next loop (see figure 5-3), which is modeled as an external disturbance. Since CCR_2 is known to the VS/VD switch (it is controlled by the VS), this signal can be feedforward to the controller of the previous loop (*Controller1*). The components defining the delays, the upstream source, and the per-VC queue may be described as the process or plant, which has the output signal $N_{vq}(t)$, the input signal $ER_1(t)$, and the disturbance signal $CCR_2(t)$. The plant input $ER_1(t)$ equals the output of *Controller1*, whose input variable is the error signal $E_1(t)$.

Next, the controller process for the previous loop has to be defined. Since the per-VC queue has the effect of an integrator and the feedback delay adds phase lag to the system [PhHa91], a proportional-derivative (PD) controller is a preferable choice. It compensates the phase lag and hence leads to a more stable control system. Together with the integral (I) queue transfer function a PID system is composed, which is already well known and has been intensively studied [Dorf89], [PhHa91]. Furthermore, PD-controllers have already been identified as a promising approach for standard ER switches [BeMe93], [BeWa98], [KoRa97a], [KiTh98a], [KiTh98b]. For a PD-controller, only two gains K_{1p} and K_{1d} have to be chosen for the proportional and the derivative component, respectively. Therefore, the requirement of a simple controller structure is met. The PD-controller process can be described as

$$ER_1(t) = K_{1p} \cdot E_1(t) + K_{1d} \cdot E_1'(t) + CCR_2(t)\ , \tag{5.25}$$

where the error signal $E_1(t) = N_{vq0} - N_{vq}(t)$ and the disturbance CCR_2 is feedforward to the PD-controller. Applying the Laplace transform to eq. 5.25 yields

$$ER_1(s) = K_{1p} \cdot E_1(s) + K_{1d} \cdot s \cdot E_1(s) + CCR_2(s)\,, \tag{5.26}$$

where $ER_1(s)$, $E_1(s)$ and $CCR_2(s)$ are the Laplace transforms of $ER_1(t)$, $E_1(t)$ and $CCR_2(t)$, respectively. Adapting eq. 5.24, describing the queue dynamics, to the case of the per-VC queue of the previous loop yields

$$N_{vq}(s) = \frac{1}{s}\left(e^{-\tau_{1\text{action}} \cdot s} \cdot ER_1(s) - CCR_2(s)\right), \tag{5.27}$$

where the action delay of the previous loop is defined as

$$\tau_{1action} = \tau_{1RM} + \tau_{1br} + \tau_{1fr}\,. \tag{5.28}$$

Combining eq. 5.26 and eq. 5.27 results in the complete controller equation relating the input signal $E_1(s)$ to the output signal $N_{vq}(s)$

$$N_{vq}(s) = E_1(s)\frac{\left[K_{1p} + K_{1d} \cdot s\right] \cdot e^{-s \cdot \tau_{1action}}}{s} + \frac{CCR_2(s) \cdot e^{-s \cdot \tau_{1action}}}{s} - \frac{CCR_2(s)}{s}\,. \tag{5.29}$$

The first term defines the transfer function of the controller process (PD-controller) concatenated with the plant process (action delay and per-VC queue). The second term accounts for the feedforward component entering the plant process. The last term describes the disturbance signal draining the per-VC queue. Rewriting eq. 5.29 lead to

$$N_{vq}(s) = \left[\left(E_1(s)\left(K_{1p} + K_{1d} \cdot s\right) + CCR_2(s)\right) \cdot e^{-s \cdot \tau_{1action}} - CCR_2(s)\right] \cdot \frac{1}{s}\,, \tag{5.30}$$

which results in the block diagram for the feedback PD-control system of the upstream loop, see figure 5-6.

One can identify the transfer functions $G_{1c}(s)$ of the control process, consisting of the PD-controller, and $G_{1p}(s)$ of the plant process, comprising the action delay and the per-VC queue.

$$G_{1c}(s) = K_{1p} + K_{1d} \cdot s \tag{5.31}$$

$$G_{1p}(s) = \frac{e^{-s \cdot \tau_{1action}}}{s} \tag{5.32}$$

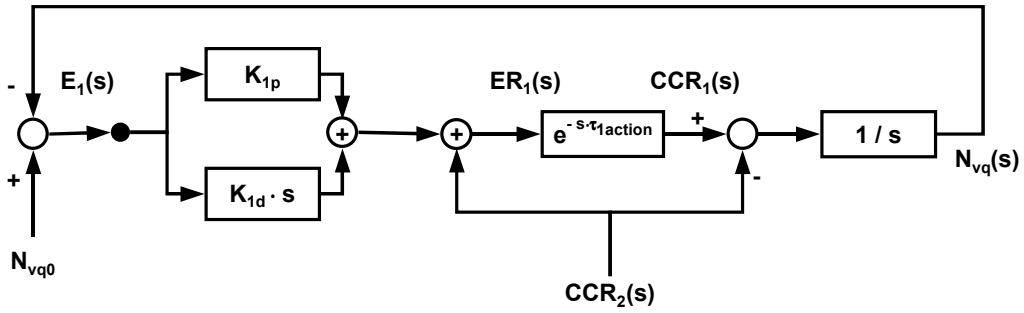

Figure 5-6: *Block Diagram of Upstream Feedback Control System*

The open-loop transfer function of the previous loop control system $G_1(s)$, with respect to the input signal $E_1(s)$ and the output signal $N_{vq}(s)$ is hence defined as

$$G_1(s) = G_{1c}(s) \cdot G_{1p}(s) = \left(K_{1p} + K_{1d} \cdot s\right) \cdot \frac{e^{-s \cdot \tau_{1\,action}}}{s} \quad . \tag{5.33}$$

Since the per-VC queue level can be measured directly at the VS/VD switch, the transfer function of the sensor is equal to one, and hence the previous loop is a unity, closed-loop feedback control system with a closed-loop transfer function

$$T_1(s) = \frac{1}{1 + G_1(s)} \quad . \tag{5.34}$$

5.5.1.1 Controller Process

The frequency response of the PD-controller can be described by a Nyquist diagram of the transfer function $G_{1c}(j\omega)$, replacing s by $j\omega$ in eq. 5.31.

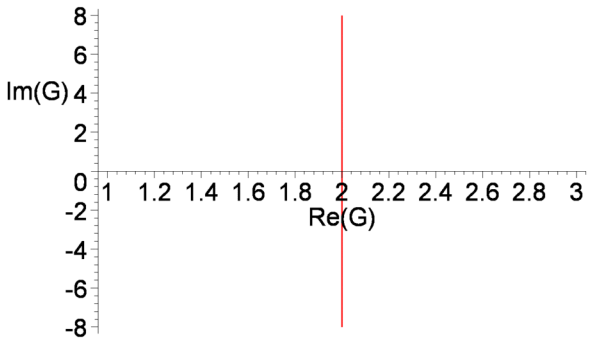

Figure 5-7: *Nyquist Diagram of Controller Transfer Function*

Figure 5-7 depicts the polar plot for the PD-controller transfer function for two sample parameter values of $K_{1p} = 2$ and $K_{1d} = 0.2$, which consists of a vertical line, parallel to the imaginary axis at a position equal to K_{1p}.

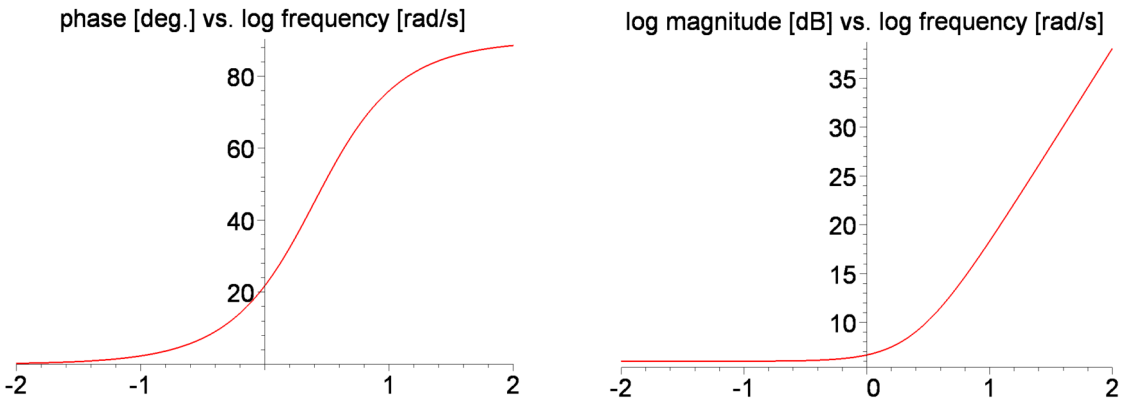

Figure 5-8: *Bode Diagram of Controller Transfer Function*

Figure 5-8 shows the Bode diagram of the transfer function $G_{1c}(j\omega)$ of a PD-controller for $\omega = [\ 0.01\ ,\ 100\ ]$, using the same controller parameters as above. From the phase plot it can be seen, that the PD-controller adds a positive phase angle to the system, i.e., the curve has positive phase values for all frequencies. The plot of the magnitude indicates that the controller gain, corresponding to the magnitude of the controller response, grows with no bounds with increasing frequency.

5.5.1.2 Plant Process

For the frequency response analysis of the plant, s is substituted by $j\omega$ in its transfer function $G_{1p}(s)$ (eq. 5.32), resulting in

$$G_{1p}(j\omega) = -\frac{j \cdot e^{-j\omega \cdot \tau_{1action}}}{\omega} = -\frac{\sin(\omega \cdot \tau_{1action})}{\omega} - \frac{j \cdot \cos(\omega \cdot \tau_{1action})}{\omega} \tag{5.35}$$

The polar plot of $G_{1p}(j\omega)$ for $\omega = [\ \pi/4\ ,\ 10 \cdot \pi\ ]$ is presented in figure 5-9, assuming a sample feedback delay parameter $\tau_{1action}$ of 0.5 s.

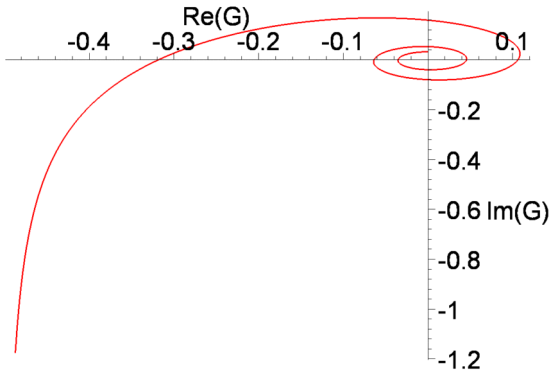

Figure 5-9: *Nyquist Diagram of Plant Transfer Function*

The real part of $G_{1p}(j\omega)$ converges to $-\tau_{1action}$ for ω reaching zero. The frequency ω_0 at which $G_{1p}(j\omega)$ intersects the real axis for the first time is defined as

$$\omega_0 = \inf\left\{\omega \middle| \Im\left(G_{1p}(j\omega)\right) = 0 \wedge \omega > 0\right\} = \frac{\pi}{2 \cdot \tau_{1action}} \tag{5.36}$$

At this frequency, the transfer function of the plant has a magnitude of

$$\left|G_{1p}(j\omega_0)\right| = \frac{2 \cdot \tau_{1action}}{\pi}\ , \tag{5.37}$$

which corresponds to the reciprocal of ω_0.

Figure 5-10 presents the Bode diagram of the plant transfer function $G_{1p}(j\omega)$ for $\omega = [\ 0.1\ ,\ 10\ ]$ and the sample feedback delay $\tau_{1action}$ of 0.5 s. The transfer function reaches a phase angle of -180 deg. at $\log(\omega_0) = 0.497$ with a magnitude of -9.9 dB in accordance with eq. 5.36 and eq. 5.37.

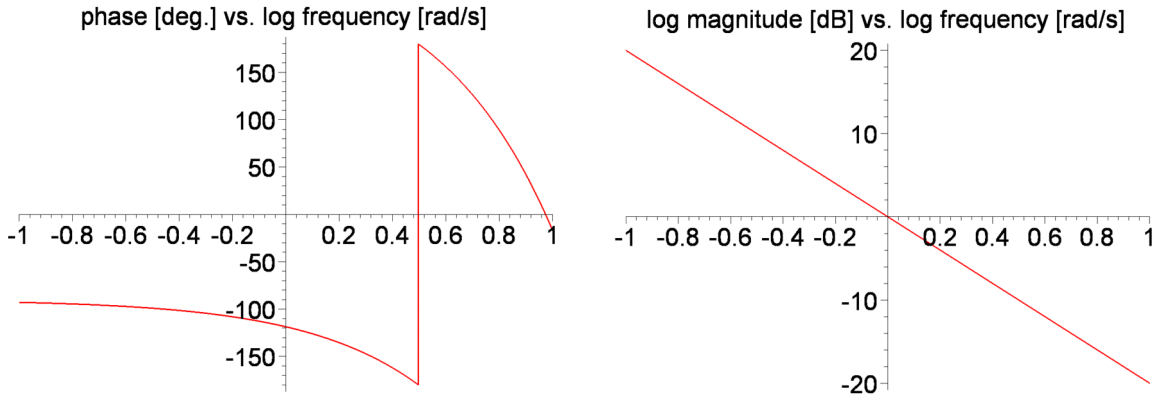

Figure 5-10: *Bode Diagram of Plant Transfer Function*

5.5.1.3 Choice of Controller Parameters

As a next step, the two parameters of the PD-controller, (K_{1p} and K_{1d}), have to be selected appropriately, in order to guarantee a stable operation and the desired phase margin. The Nyquist criterion, introduced in chapter 4.4.3, will be used to derive the two parameters needed. The open-loop transfer function $G_1(s)$ (eq. 5.33) is a ratio of two polynomials in s, including an ideal time delay of $\tau_{1action}$ and hence the Nyquist criterion may be applied. $G_1(s)$ has a zero at the origin and a pole for s approaching zero. The conclusion that the control system is stable, when $G_1(s)$ has no encirclements of the (-1,0) point is valid only, if $G_1(s)$ has no poles in the right-hand s-plane, see chapter 4.4.3. In order to avoid the zero of $G_1(s)$ along the Nyquist path (figure 4-9) at the origin, the path may be modified to include an infinitesimal small detour around the origin. The location of the pole will be determined by the use of the frequency response analysis. Therefore, the variable s of the open-loop transfer function $G_1(s)$ is replaced by $j\omega$.

$$G_1(j\omega) = -\frac{j\left(K_{1p} + K_{1d} \cdot j\omega\right) \cdot e^{-\omega \cdot \tau_{1action}}}{\omega} \tag{5.38}$$

Separating eq. 5.38 into its real and imaginary parts yields

$$\Re\left(G_1(j\omega)\right) = \frac{K_{1d} \cdot \omega \cdot \cos(\omega\tau_{1action}) - K_{1p} \cdot \sin(\omega\tau_{1action})}{\omega} \tag{5.39}$$

$$\Im\left(G_1(j\omega)\right) = \frac{-K_{1p} \cdot \cos(\omega\tau_{1action}) - K_{1d} \cdot \omega \cdot \sin(\omega\tau_{1action})}{\omega} \; . \tag{5.40}$$

Taking the limit as ω approaches zero gives

$$\lim_{\omega \to +0} \Re\left(G_1(j\omega)\right) = K_{1d} - K_{1p} \cdot \tau_{1action} \tag{5.41}$$

$$\lim_{\omega \to +0} \Im\left(G_1(j\omega)\right) = -\infty \text{ for } K_{1p} > 0 \; . \tag{5.42}$$

As a result, it can be concluded, that the pole of the open-loop transfer function of the previous loop feedback control system is located in the left-hand s-plane, if $K_{1d} < K_{1p} \cdot \tau_{1action}$.

In order to guarantee the relative stability of the system, a phase margin of 40 to 45 deg. is targeted, see section 5.3., which will be reached at a frequency ω_1. At this frequency the magnitude of the open-loop transfer function $G_1(j\omega)$ is equal to one, see figure 4-11.

$$\left|G_1(j\omega_1)\right| = \frac{\sqrt{K_{1p}^2 + K_{1d}^2\omega_1^2}}{\omega_1} = 1 \tag{5.43}$$

Solving for ω_1 gives

$$\omega_1 = \frac{\sqrt{1-K_{1d}^2}\cdot K_{1p}}{1-K_{1d}^2} \tag{5.44}$$

and hence $K_{1d} < 1$.

Respecting the requirement $K_{1d} < K_{1p} \cdot \tau_{1action}$, the choice $K_{1p} = 1 / \tau_{1action}$ combines both conditions. The value of the open-loop transfer function at the frequency ω_1 is then defined as

$$G_1(j\omega_1) = \left(-j\sqrt{1-K_{1d}^2} + K_{1d}\right)\cdot e^{\frac{-j}{\sqrt{1-K_{1d}^2}}} . \tag{5.45}$$

The phase margin ϕ_m may be expressed as

$$\phi_m = \angle G_1(j\omega_1) = -\arctan\left(\frac{\sqrt{1-K_{1d}^2}}{K_{1d}}\right) - \frac{1}{\sqrt{1-K_{1d}^2}} . \tag{5.46}$$

Figure 5-11 shows a plot of the phase margin ϕ_m for different values of K_{1d} between 0 and 1. A parameter value of 0.2 yields a phase margin of about 42 deg. The other possible value of 0.84 would result in a less robust controller, since the phase margin shows greater variance in the vicinity of this point.

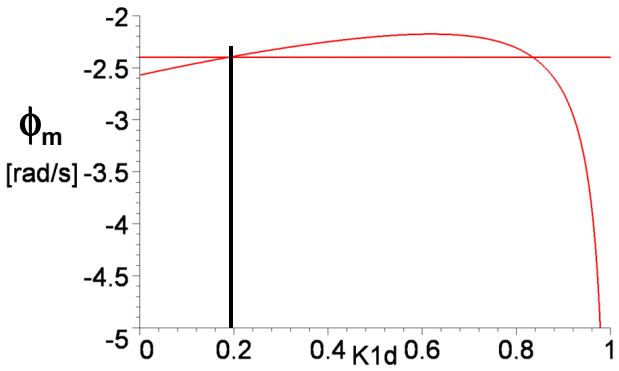

Figure 5-11: *Phase Margin*

Therefore, the two PD-controller parameters are

$$K_{1p} = \frac{1}{\tau_{1action}} \text{ and} \tag{5.47}$$

$$K_{1d} = 0.2 . \tag{5.48}$$

5.5.1.4 Stability

In order to prove that the designed PD-controller is stable, the Nyquist diagram of the open-loop transfer function must not encircle the (-1,0) point. The gain margin of the controller is determined by the leftmost intersection of the open-loop transfer function of $G_1(j\omega)$ with the real axis. Since $G_1(j\omega)$ has a pole at (-0.8 , -∞) for ω approaching zero and a zero at the origin, the critical point of (-1,0) is not encircled, if this intersection point is located to the right of the critical point. The angle of the open-loop transfer function $G_1(j\omega)$ at the intersection with the real axis is -180 deg., occurring at a frequency ω_0. Hence, for a stable system the magnitude of $G_1(j\omega_0)$ has to be smaller than one, avoiding the critical point (-1,0).

$$\Im\left(G_1\left(j\omega_0\right)\right)=0 \Rightarrow -\frac{K_{1p}}{K_{1d}}=\omega_0 \cdot \tan(\omega_0 \cdot \tau_{1action}) \tag{5.49}$$

Inserting the chosen parameters according to eq. 5.47 and eq. 5.48 leads to

$$\omega_0 = \frac{1.93722}{\tau_{1action}} \quad , \tag{5.50}$$

and hence

$$\Re\left(G_1\left(j\omega_0\right)\right) = -0.55359 > -1 \quad . \tag{5.51}$$

Thus, it can be concluded that the closed-loop feedback system for the previous loop of a VS/ VD switch using a PD-controller with parameters $K_{1d} = 0.2$ and $K_{1p} = 1 / \tau_{1action}$ is stable with a phase margin of 42 deg. and a gain margin of 5.1 dB.

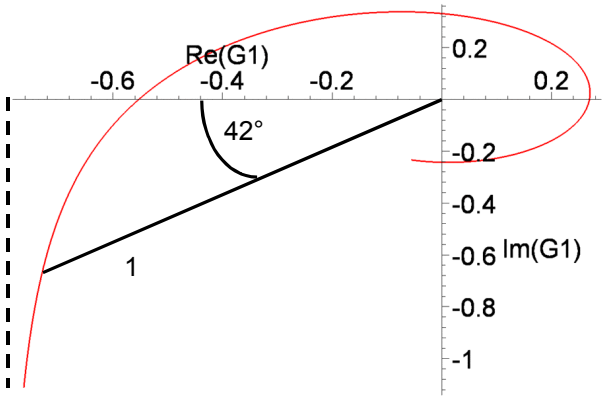

Figure 5-12: *Nyquist Diagram of Feedback Control System for Previous Loop*

Figure 5-12 shows a Nyquist diagram of the open-loop transfer function of the previous loop for a feedback delay $\tau_{1action}$ of 5 ms, resulting in the PD-controller parameters $K_{1d} = 0.2$ and $K_{1p} = 200$, for $\omega = [\ 150\ ,\ 1500\ ]$. The plot illustrates the pole at (-0.8 , -∞), the phase margin ϕ_m of 42 deg. and the leftmost intersection point with the real axis at -0.55.

Figure 5-13 presents the corresponding Bode diagram for $\omega = [\ 0.5\ ,\ 700\ ]$. The phase margin in degrees and the gain margin in dB can be read from the diagram. The logarithm of the two frequencies ω_0 and ω_1 may also be determined from the plot.

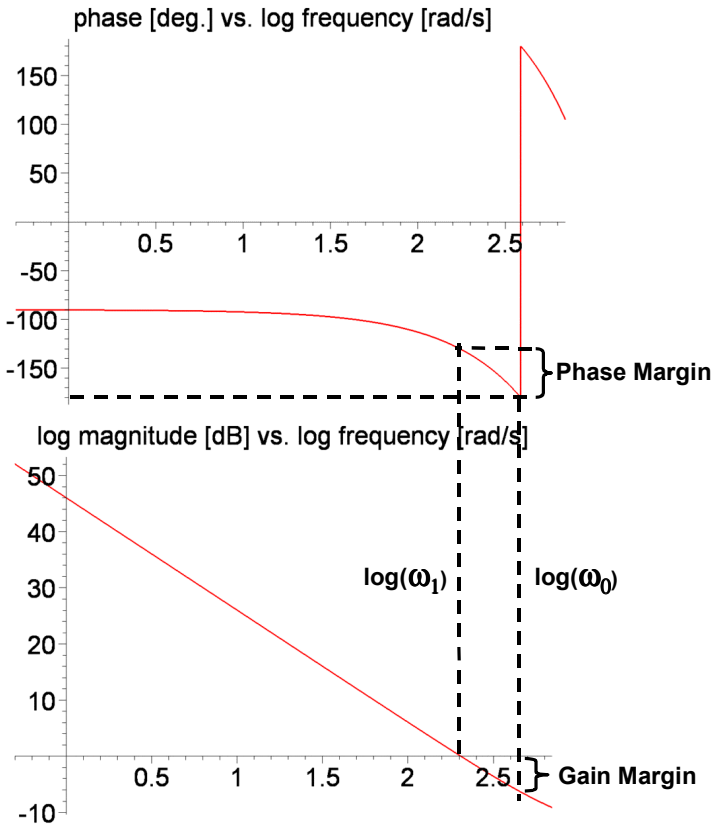

Figure 5-13: *Bode Diagram of Feedback Control System for Previous Loop*

5.5.1.5 Steady State System Error

The steady state error of the PD-controller is defined with respect to the disturbance of the input signal. Common input functions are the step and the ramp input, see chapter 4.3.2.

For a step change, the input function is defined as follows.

$$r_1(t) = A \cdot u(t) \ , \tag{5.52}$$

where A is constant specifying the magnitude of the step change and $u(t)$ denotes the unit step function, which has the value of 1 for $t \geq 0$ and of 0 for $t < 0$, see chapter 4.4.5. Thus, the steady state system error, with respect to the error signal $E_1(t)$ and the unity, closed-loop feedback system of the VS/VD previous loop (see figure 5-6), can be calculated as

$$\lim_{t \to \infty} E_1(t) = \lim_{s \to 0} \frac{s \cdot (A/s)}{1 + G_1(s)} = 0 \ . \tag{5.53}$$

For a ramp input, the input function can be described as

$$r_1(t) = A \cdot t \ , \tag{5.54}$$

where A is a constant defining the slope of the straight line. For the steady state system error of the closed-loop control system (see figure 5-6)

$$\lim_{t\to\infty} E_1(t) = \lim_{s\to 0} s \cdot \frac{A/s^2}{1+G_1(s)} = \frac{A}{K_{1p}} = A \cdot \tau_{1\,action} \tag{5.55}$$

holds, i.e., the error is bounded by the feedback delay times the slope of the ramp increase.

After the design of the PD-controller for the previous loop of the VS/VD switch, the controller for the next or downstream loop is developed in the following section.

5.5.2 Downstream Loop

The controller of the downstream loop calculates for each ABR connection an Explicit Rate for the next loop (*ER2*), that is used by the virtual source (VS) to determine its Allowed Cell Rate (*ACR2*). The controller has the two tasks of

- allocating the ABR capacity to the ABR VCs according to the "MCR plus equal share" fairness criterion, see section 5.3, and of
- keeping the ABR class queue at a desired level.

5.5.2.1 Fairness

The VS/VD switch divides the ABR connections into two sets:

- A set of locally restricted (*R*) or bottlenecked connections, whose Current Cell Rate (*CCR1*) of the ABR source is determined by this VS/VD switch. That means, that the Explicit Rate allocated by this VS/VD switch is the minimum rate along the network path from the sender to the receiver.
- A set of locally unconstrained (*U*) connections, that are bottlenecked elsewhere along the network path from source to destination.

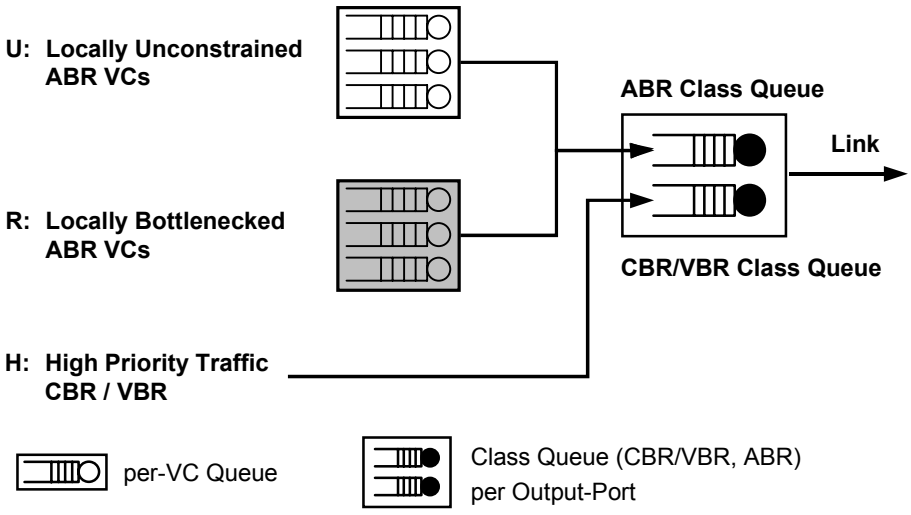

Figure 5-14: *Connection Sets*

In addition, there is a set H of high priority VCs, carrying CBR and VBR traffic, see figure 5-14.

The ABR capacity ($ABR_{Capacity}$) is calculated as the link bandwidth C reduced by the bandwidth used by high priority traffic (VBR/CBR).

$$ABR_{Capacity}(t) = C - \sum_{h \in H(t)} CCR^h(t) \tag{5.56}$$

$ABR_{Traffic}$ is defined as the sum of locally bottlenecked and unconstrained ABR connections.

$$ABR_{Traffic}(t) = \sum_{r \in R(t)} CCR_2^r(t) - \sum_{u \in U(t)} CCR_2^u(t) \tag{5.57}$$

Since the corresponding VS/VD switch is able to control only the transmission rates of the bottlenecked connections, the *TotalFairShare* minus the Minimum Cell Rates (*MCRs*) of the bottlenecked connections is divided among these VCs. This rate is called the *FairShare* of connection i. The *TotalFairShare* is defined as the available bandwidth for ABR traffic ($ABR_{Capacity}$) reduced by the Current Cell Rates of the locally unconstrained ABR VCs, belonging to set U.

$$TotalFairShare(t) = ABR_{Capacity} - \sum_{u \in U(t)} CCR_2^u(t) \tag{5.58}$$

$$FairShare^i(t) = \frac{TotalFairShare(t) - \sum_{r \in R(t)} MCR^r}{|R(t)|} + MCR^i \tag{5.59}$$

The allocation of the link capacity C according to the different terms and groups defined above is illustrated in figure 5-15.

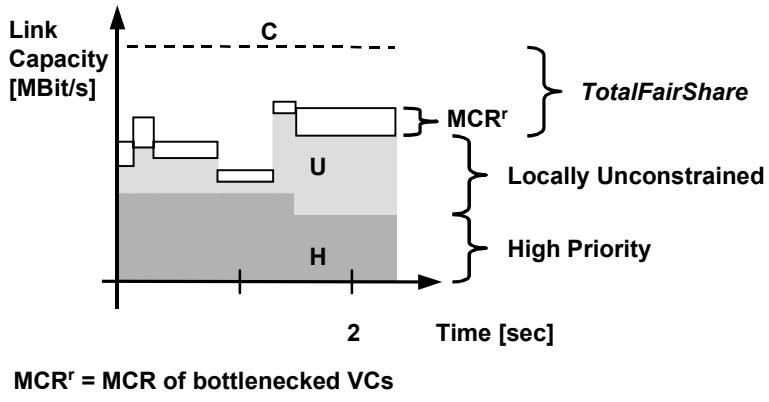

Figure 5-15: *Link Bandwidth Allocation*

The fair share calculated for each bottlenecked ABR connection as specified in eq. 5.59 implements the fairness criterion "MCR plus equal share" according to eq. 5.18.

For a connection j that is locally unrestricted, i.e., bottlenecked elsewhere on the network path, the *FairShare* is calculated as if all connections would be bottlenecked.

$$FairShare^{j}(t) = \frac{TotalFairShare(t) - \sum_{r \in R(t)} MCR^{r} - \sum_{u \in U(t)} MCR^{u}}{|R(t)| + |U(t)|} + MCR^{i} \qquad (5.60)$$

On the one hand, this allocation is more conservative than calculating the fair share as if only the single connection *j* would change its status from unconstrained to bottlenecked. On the other hand, it prevents large temporal overloads at the VS/VD switch, when more than one unconstrained VC becomes bottlenecked at the same time.

5.5.2.2 Controlling the ABR Class Queue

The second task of the controller of the next loop at the VS/VD switch is keeping the ABR class queue filled at its target level N_{cq0}. The controller model is similar to the approach developed for the controller of the previous loop. The per-VC queue of the previous loop is replaced by the ABR class queue in the case of the next loop. The queue is filled by the $ABR_{Traffic}$ and is drained by the $ABR_{Capacity}$. In contrast to the previous loop, where the per-VC queue was fed only by a single ABR connection, the ABR class queue is fed by all ABR VCs passing through the VS/VD switch. Because the buffer level of the ABR class queue can be controlled by the virtual source (VS) only for the locally constrained connections, the viewpoint of such a bottlenecked VC is assumed in the ensuing discussion. The main idea is, that each bottlenecked VC contributes by its Current Cell Rate in the next loop (*CCR2*) to the ABR class queue level. For this reason, the queue error signal $E_2(t)$ is divided by the number of bottlenecked VCs *r* ($r = |R(t)|$).

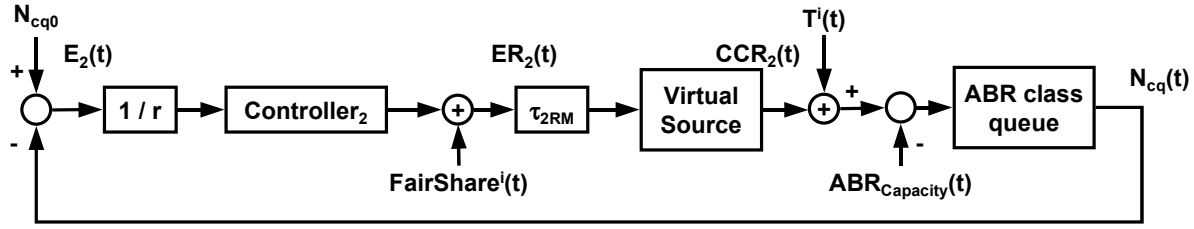

Figure 5-16: *Control Model for Downstream Loop*

Figure 5-16 illustrates the control model for a bottlenecked ABR VC *i* of the downstream loop. The ABR class queue is drained by the available capacity for ABR traffic ($ABR_{Capacity}$) and is filled by the individual VC ($CCR_2(t)$) and the remaining traffic $T^i(t)$, where

$$T^i(t) = ABR_{Traffic} - CCR_2(t). \qquad (5.61)$$

The class queue error signal $E_2(t)$ divided by *r* is fed into the controller for the next or downstream loop. The calculated $FairShare^i$ is used as a feedforward signal for the connection *i* to determine the Explicit Rate of the next loop ($ER_2(t)$). The action delay until this rate becomes effective equals the time until the next backward RM cell of the downstream loop (BRM2) for that VC is received (τ_{2RM}). Since the VC is bottlenecked at the corresponding VS/VD switch, the Current Cell Rate of the virtual source ($CCR_2(t)$) equals the explicit rate calculated ($ER_2(t)$). It becomes clear, that the structure of the control model for the downstream loop is analogous to

the one of the model for the previous loop. As a consequence, the same type of PD-controller is used to derive the explicit rate value for the next loop ($ER_2(t)$). The analysis of the closed-loop feedback control system of the previous loop can be adopted for this analogous system. The reduction factor r for the error signal of the class queue may be included in the two PD-controller parameters. Thus, the parameters K_{2p} and K_{2d} are defined as follows.

$$K_{2\,p} = \frac{1}{r \cdot \tau_{2\,action}} \text{ and} \tag{5.62}$$

$$K_{2\,d} = 0.2/r \ , \tag{5.63}$$

where the feedback delay $\tau_{2action}$ equals the time until the next BRM2 cell for this VC is received at the VS/VD switch (τ_{2RM}).

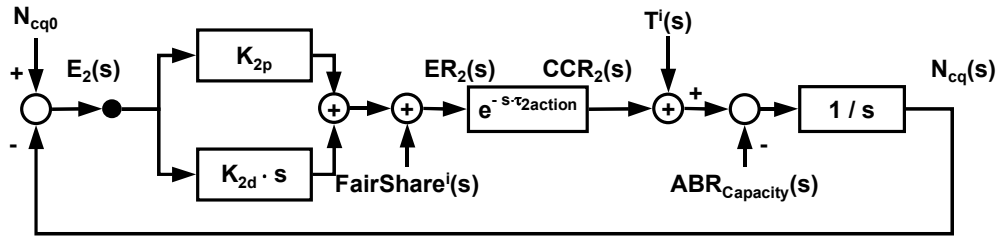

Figure 5-17: *Block Diagram of Downstream Feedback Control System*

Figure 5-17 presents the resulting block diagram of the downstream feedback control system, where the $E_2(s)$, $FairShare^i(s)$, $ER_2(s)$, $CCR_2(s)$, $T^i(s)$, $ABR_{Capacity}(s)$, and $N_{cq}(s)$ are the Laplace transforms of the corresponding functions $E_2(t)$, $FairShare^i(t)$, $ER_2(t)$, $CCR_2(t)$, $T^i(t)$, $ABR_{Capacity}(t)$, and $N_{cq}(t)$, respectively.

The open-loop transfer function of the previous loop as defined in eq. 5.33, corresponds to the open-loop transfer function for the next loop, adapting the feedback delay parameter and the queue error input signal as described above. Therefore, the results regarding the stability and the steady state system error derived for the feedback control system of the previous loop, can be adopted accordingly.

5.5.2.3 Sampled Operation

As described in the previous section, the VS/VD switch needs to maintain a connection state for each ABR VC, indicating whether the connection is locally bottlenecked or unconstrained. Since the explicit rate calculated for the next loop (*ER2*) depends on the fair share and hence on the number of bottlenecked connections, this state information has to be updated at regular intervals. For this reason, a measurement interval is introduced, after which the VS/VD switch will loop through the ABR VCs to determine the connection state. Then, the control action for the next loop is executed, calculating *ER2* for each of the connections. Due to this fixed measurement interval, an alternative discrete time control model for the downstream loop may be developed.

Using a discrete time model, the feedback delay introduced by the BRM cells of the next loop is included in the sampling interval *T*. If a BRM2 cell has been received during the sampling period, the calculated value *ER2* is used to determine the Allowed Cell Rate for the downstream loop (*ACR2*). If no BRM2 cell has been received by the VS/VD switch, a new *ER2* value is calculated by the controller for the ABR class queue based on updated measurements, which remains valid for the upcoming sample period. Therefore, in the digital control model the feedback delay $\tau_{2action}$ is removed. The sampled time system requires a digital PD-controller, where the differential operation is performed on consecutive sample values.

$$m\left((k+1)T\right) = \frac{e\left((k+1)T\right) - e\left(kT\right)}{T} \quad , \tag{5.64}$$

where $m(kT)$ defines the output and $e(kT)$ the error series of sample values. Assuming all initial conditions to be zero and applying the *z*-Transform yields

$$M\left(z\right) = E(z) \cdot \frac{z-1}{Tz} \quad , \tag{5.65}$$

where $M(z)$ and $E(z)$ are the *z*-Transforms of $m(kt)$ and $e(kt)$, respectively [PhHa91]. Hence, the transfer function of a digital PD-controller is defined as

$$K_p + \frac{K_d \cdot (z-1)}{Tz} \text{ [PhHa91].} \tag{5.66}$$

The resulting discrete time control system for the next loop consists of the sampler, the digital PD-controller, the zero-order hold component, and the ABR class queue, as depicted in figure 5-18.

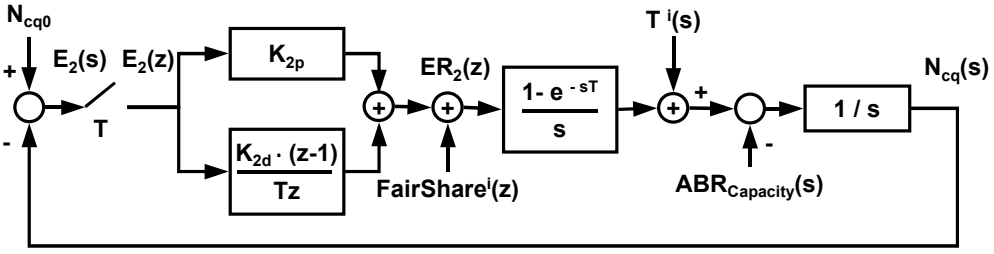

Figure 5-18: *Discrete Time Control Model for Downstream Loop*

The functions of the complex variable *z* denote the *z*-Transforms of the corresponding Laplace transform functions. Combining the zero-order hold component and the ABR class queue leads to

$$\left(1 - e^{-sT}\right)\Big|_{z=e^{sT}} \cdot \mathcal{Z}\left[\frac{1}{s^2}\right] = \frac{T}{z-1} \quad . \tag{5.67}$$

This results in the open-loop transfer function

$$G_2(z) = \frac{K_{2d}}{z} + \frac{K_{2p} \cdot \mathrm{T}}{(z-1)} \tag{5.68}$$

for the digital control model of the downstream VS/VD loop.

Following the design of the analog control model, the two controller parameters are chosen as

$$K_{2p} = \frac{1}{r \cdot \mathrm{T}} \text{ and} \tag{5.69}$$

$$K_{2d} = 0.2/r \ , \tag{5.70}$$

where r denotes the number of bottlenecked connections and T equals the measurement interval. The closed-loop transfer function for the discrete time control system of the downstream loop and its characteristic equation (the denominator) are presented in eq. 5.71.

$$T_2(z) = \frac{G_2(z)}{1 + G_2(z)} \tag{5.71}$$

The control system is stable, if the roots of the characteristic equation are located inside the unit circle, see chapter 4.5. Choosing the controller parameters according to eq. 5.69 and eq. 5.70, leads to two roots at r_1 = -0.558 and r_2 = 0.358, located inside the unit disc of the complex z-plane.

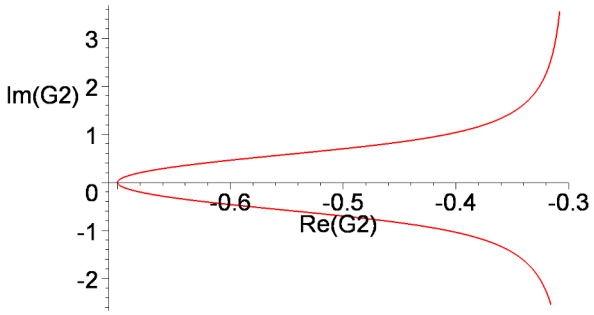

Figure 5-19: *Nyquist Diagram of Discrete Time Control System for Next Loop*

Another possibility to determine the stability is the use of the Nyquist criterion. Figure 5-19 shows the Nyquist diagram of the open-loop transfer function $G_2(z)$ of the downstream loop for ω = [0.4 , 6], using the substitution $z = e^{-j\omega T}$. The Nyquist path used for the mapping consists of the unit circle, making a small detour around the point (1,0), because the open-loop transfer function $G_2(z)$ has a pole at this location. The Nyquist diagram does not encircle the critical (-1,0) point, confirming the stability of the discrete time control system.

A simple method to quantify the phase and gain margins of the digital feedback control system is the use of the Bode diagram. Thus, the Bode diagram of the open-loop transfer function $G_2(z)$ for ω = [0.8 , 3.5], using the substitution $z = e^{-j\omega T}$, is presented in figure 5-20. From the diagram the gain margin is determined to about 3 dB and the phase margin to about 65 degrees.

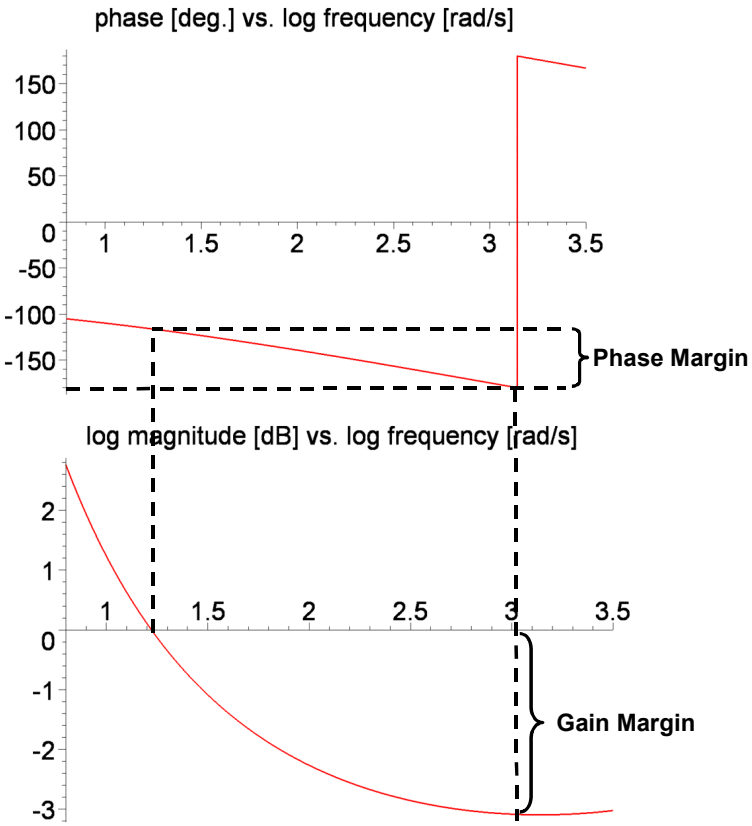

Figure 5-20: *Bode Diagram of Discrete Time Control System for Next Loop*

5.5.2.4 Steady State System Error

The steady state system error for a discrete time control system is defined as

$$\lim_{k\to\infty} e(kT) = \lim_{z\to 1} (z-1)E(z) \quad \text{[PhHa91].} \tag{5.72}$$

When considering a step input with magnitude A for the discrete time control system of the downstream loop presented in the previous section, the steady state error becomes

$$\lim_{z\to 1} \frac{Az}{1+G_2(z)} = 0 \ . \tag{5.73}$$

This result is in accordance with the fact, that a unity ($H(z) = 1$), discrete time control system, whose open-loop transfer function has a pole at $z = 1$, has a steady state system error of zero for a step input [PhHa91].

For a ramp input, i.e., input function $r(t) = A \cdot t$, the steady state error is calculated as

$$\lim_{z\to 1} \frac{A\cdot T}{(z-1)\cdot G_2(z)} = A\cdot T \ . \tag{5.74}$$

Hence, the steady state error of the closed-loop feedback control system of the VS/VD next loop for a ramp input, is bounded by the slope of the input function times the sampling interval.

5.5.3 Coupling of Control Loops

In the previous sections two closed-loop feedback control models for the previous (upstream) and the next (downstream) loop of a VS/VD switch were developed. The two controllers are coupled by the Current Cell Rate of the next loop (*CCR2*). In the case where the VS/VD switch is the bottleneck for the corresponding ABR connection and the source has data to send, this rate equals the Explicit Rate (*ER2*) and the Allowed Cell Rate (*ACR2*) for the next loop. The per-VC queue of the virtual destination (VD) is drained by *CCR2* and the data is fed into the ABR class queue, see figure 5-3. Hence, the controlled variable of the downstream loop acts as a disturbance and feedforward signal of the previous loop.

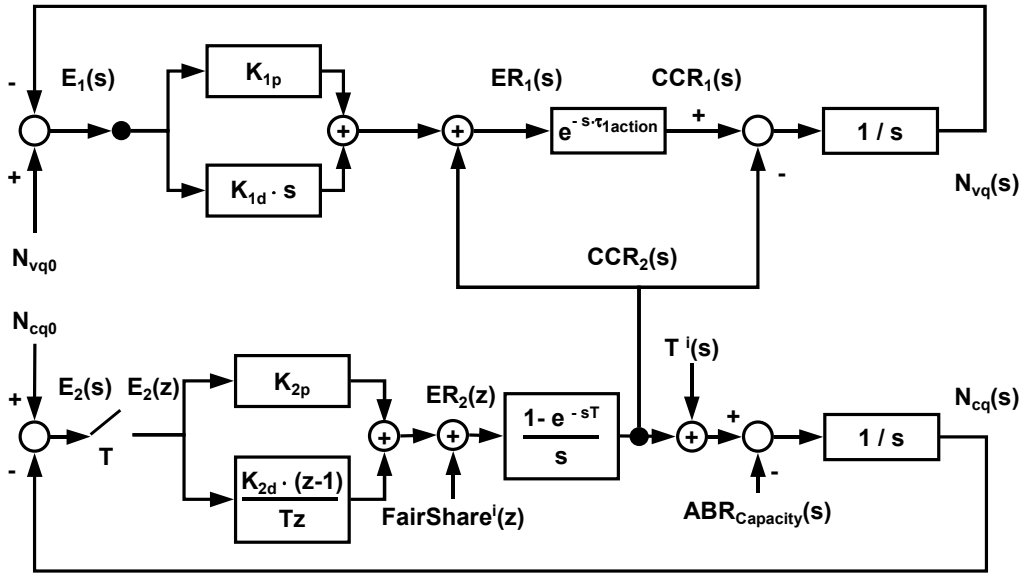

Figure 5-21: *Block Diagram of Dual Controller Model for VS/VD Switch*

Figure 5-21 presents the complete block diagram for the dual PD-controller model for the upstream and downstream loop of the VS/VD switch. The upper part of figure 5-21 represents the analog model for the previous loop, whereas the lower part depicts the control model for the next loop using a discrete time PD-controller. Choosing the PD-controller parameters as indicated in the previous sections, leads to stable closed-loop feedback systems, with desired phase and gain margins as well as steady state system errors, as defined in section 5.3. The PD-controller of the VS/VD downstream feedback control loop drives the ABR class queue to its target level and divides the remaining ABR bandwidth in a fair manner among the bottlenecked ABR connections, taking into account their minimum cell rates. The PD-controller of the previous loop controls for each connection the buffer level of its per-VC queue.

It has to be noted that the PD-controller of the previous loop follows in its *ER1* calculations the disturbance signal *CCR2* of the next loop. The PD-controller of the previous loop may not be able to track the disturbance signal accurately, if its action delay $\tau_{1action}$ is larger than the sample interval of the PD-controller in the next loop. For this reason, the controller parameter K_{2p} should be chosen as the reciprocal of the maximum of $\tau_{1action}$ and the sample interval T.

5.6 Implementation

In the previous section of this chapter a dual PD-controller model for the two feedback control loops of a VS/VD switch was designed. In the following some implementation issues, regarding the actual choice of threshold parameters and the realization of required operations are discussed. This allows an actual implementation of the VS/VD flow control algorithm, which is evaluated in chapter 6 of this work.

5.6.1 Buffer Threshold

Each of the two PD-controllers uses an input signal, that represents the desired buffer level of the per-VC and the ABR class queue, respectively. Therefore, an appropriate setting per output port and per-VC is required. For reasons of simplicity, a common per-VC threshold for all ABR connections sharing an output port will be used, depending on the downstream link capacity.

A common transport link bandwidth for ATM networks is OC-3c using a *Synchronous Optical Network* (SONET) on the physical layer, offering a nominal bandwidth of 149.76 MBit/s (after accounting for the SONET overhead) [UNI3.1]. A target buffer occupancy of the ABR class queue of 300 cells would lead to a queueing delay of about 0.85 ms. Larger values would lead to longer queueing delays and are a possible source of jitter. A maximum class queue length of 1000 cells would allow an additional buffering of up to 2 ms at OC-3c link speed before cell loss occurs. Furthermore, a queue length overshoot of 3 times the average value could be accommodated, as stated in the requirements for the transient behavior (see section 5.3).

The per-VC queue has to be allocated for each ABR connection passing through the VS/VD switch and hence the maximum per-VC queue length is limited by the maximum number of ABR connections supported. Therefore, the scalability of the algorithm with respect to the number of supported ABR connections influences the maximum buffer capacity for the per-VC queues. As mentioned in section 5.3, it can be assumed that on average 2/3 of the link bandwidth is allocated to high priority traffic. When targeting an MCR of 64 kbit/s per ABR connection, a maximum supported number of 800 ABR VCs on a OC-3c link is sufficient. Modern ATM switches offer a buffer space of up to 200,000 cells per port [BWBM98] and hence a maximum per-VC buffer level of 250 cells is possible. The ABR class queue is fed by several ABR VCs, and as a consequence, a few cells of each connection are sufficient to influence its length. Furthermore, the ABR traffic is delayed in both the per-VC queue and the ABR class queue. For these reasons, a low target level of 50 cells for the per-VC queue is reasonable. A per-VC queue of 250 cells can hence accommodate about 5 times of its average occupancy, fulfilling the requirements defined in section 5.3.

5.6.2 Measurements

As mentioned in section 5.5.2.3, the VS/VD switch has to perform measurements of the high priority traffic and to determine the fair share for each ABR connection in regular intervals. The transient response of the feedback control system depends on the length of this measurement or sampling interval. On the one hand a short sampling period leads to an improved transient

response and fast adaptation to changing bandwidth conditions. On the other hand, a short measurement interval imposes a large computational load on the VS/VD switch. In related works, that define measurement-based ABR flow control algorithms, sampling intervals between 0.5 ms [HTPK98b] and 5 ms [Vand00] are used. In order to find a good balance between complexity and fast response, in this work a sampling interval of 2 ms is assumed. This corresponds to about 700 cell times for an OC-3c link. That means, that even if the maximum possible ABR bandwidth of 149.76 MBit/s is withdrawn for a complete sample interval, the resulting overload may be buffered at the ABR class queue with no cell loss.

In order to determine the current ABR capacity, the input rate of high priority traffic is measured. Since the PD-controller leads to unbounded gains at high frequencies, see section 5.5.1.1, an exponential averaging is used for the measurements. To avoid a slow adaption to larger changes in the available bandwidth, e.g., due to starting or terminating CBR connections, an averaging threshold of 10 MBit/s is introduced. If the difference between the last and the current value exceeds this threshold, the current value is chosen as the input rate of the high priority traffic.

In addition to the transmission rates, the buffer levels of the per-VC queue and the ABR class queue are measured by the VS/VD switch. It has to be taken into account, that the previous and next loop need not to be synchronized with respect to their clocks defining an ATM cell slot. As a result, the per-VC queue may experience small oscillations, even if the input and output rate are the same. Furthermore, the Current Cell Rate in the downstream loop (*CCR2*) may assume arbitrary bit rates, whereas the ABR class queue is served at fixed cell slot times, determined by the downstream link rate. Again, this may lead to small oscillations of the class queue level at steady state. In addition the time instant at which the queue measurement is performed may fall exactly in a short time interval between the insertion and the deletion of a cell at a queue. For these reasons, a dead-band with a tolerance of 2 cells for the per-VC queues and 4 cells for the ABR class queue around the target value is introduced in the measurement of the queue levels.

5.6.3 Connection Status

For the fair distribution of the available bandwidth, the VS/VD switch has to maintain a connection status for each ABR VC, as described in section 5.5.2.1. The VS/VD switch has to determine, if an ABR connection is locally bottlenecked or unconstrained. If the ABR connection is locally unconstrained, then the VC is bottlenecked elsewhere on the network path from the sender to the receiver. This could happen either in the upstream loop ($CCR1 < ER1$) or in the downstream loop ($CCR2 < ER2$). The latter case is easily detected, because the *CCR2* is determined by the ABR source rules, based on the external Explicit Rate, received in backward RM (BRM2) cells of the next loop. Whenever a BRM2 cell is received by the VS/VD switch, its external Explicit Rate carried in the ER field is compared to the Explicit Rate calculated for the downstream loop (*ER2*). If the external value is smaller than *ER2*, the VC is bottlenecked downstream, i.e., locally unconstrained at the corresponding VS/VD switch.

The detection of an upstream bottleneck is not that simple, because the *CCR1* value carried in the forward RM cell of the previous loop (FRM1) indicates the source rate at a previous time

instant, i.e., the time of the reception of the FRM1 cell minus the forward propagation delay (τ_{1fr}), see figure 5-22. The VS/VD switch returns a BRM1 cell at time instant t_0 to the upstream source, containing the calculated Explicit Rate for the previous loop (ER1). On its way to the ABR source the upstream bottleneck switch reduces the *ER1* value carried in the BRM1 cell to a smaller value *ER1´*. This smaller value becomes the new ABR transmission rate of the source (*CCR1*) at time t_1. The source sends at least after *Nrm* (default value = 32) data cells the next FRM1 cell. This FRM1 cell reaches the VS/VD switch at time t_3 and carries the source rate *CCR1*, see figure 5-22

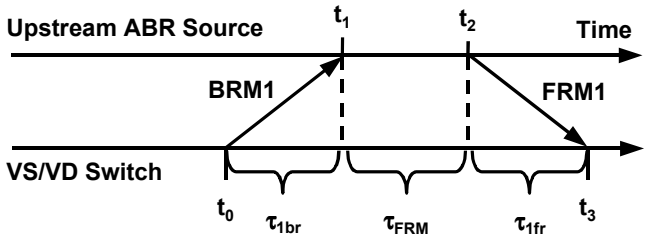

Figure 5-22: *Time Delays in Previous Loop*

In turn, the connection is locally bottlenecked, if this *CCR1* equals the delayed explicit rate for the previous loop (*ER1*) calculated by the VS/VD switch. The time delay until the Explicit Rate *ER1* is seen by the VS/VD switch in a FRM1 cell equals the backward propagation delay of the previous loop (τ_{1br}), plus the time period until the source sends its next FRM1 cell (τ_{FRM}), plus the forward propagation delay (τ_{1fr}). It has to be noted that the propagation delays correspond to the variables of the upstream control model, depicted in figure 5-5. However, τ_{FRM} defines the time at the source between a BRM1 cell is received and the next FRM1 cell is sent, whereas the variable τ_{1RM} in the control model specifies the delay between the execution of the control action and the transmission of a BRM1 cell at the VS/VD switch.

One possible method for a VS/VD switch to detect an upstream bottleneck of an ABR connection would be to store the *ER1* values calculated and to compare these to the *CCR1* values received in FRM1 cells. Taking into account only the propagation delay of 5 µs/km and assuming a path length of 1000 km, the *Round-Trip Time* (RTT) for the previous loop evaluates to 10 ms. At a source rate of 149.76 MBit/s, the VS/VD switch would have to store about 100 *ER1* values before the first of these rates would be seen in the CCR field of a FRM1 cell. Observing the number of 800 parallel ABR connections supported, this approach is considered too complex.

Instead, the VS/VD switch stores the *CCR1* rates and reception times of the last two FRM1 cells, t_{-1} and t_{-2}. If the RTT of the previous loop is smaller than the last inter-FRM1 cell period (t_{-1} - t_{-2}), then the Explicit Rate calculated at t_{-2} is compared to the *CCR1* value received at t_{-1}. If $CCR_1(t_{-1})$ is equal to $ER_1(t_{-2})$, then the connection is locally bottlenecked, else it is locally unconstrained. If the RTT of the previous loop is larger than the last inter-FRM1 period, the VS/VD switch calculates the *CCR1* of the ABR source based on the number of cells received.

$$CCR_1(now) = \frac{n + Nrm}{now - t_{-2}} \quad , \tag{5.75}$$

where *n* corresponds to the number of cells received since the last FRM1 cell and *now* denotes the current time instant. The ABR VC is assumed to be bottlenecked upstream, if

1. this *CCR1* value is significantly smaller that the Current Cell Rate of the next loop (*CCR2*) and
2. the per-VC queue is nearly empty for a long time period.

Significantly smaller in this context means, that the difference delta between the two rates is greater than 5 % of *CCR2* or at least 2 MBit/ s. For the second condition to be true, an exponential average of the per-VC queue level is maintained by the VS/VD switch, that has to be below a threshold of 2 cells.

In general the change of the connection status at the end of a measurement interval leads to a change of the fair share and hence of *ER2* of the VC. Therefore, this new rate has to become effective before another change of status is allowed. Hence, the VC status may be changed again by the VS/VD switch, only after a BRM2 cell has been received (*CCR2* is updated), a BRM1 cell was sent (*ER1* is updated) and the RTT has elapsed (*CCR1* is updated).

5.6.4 Delay Estimates

The VS/VD switch algorithm is based on the knowledge of the RTT of the previous loop. The fixed delays resulting from the signal propagation and possible switching delays of upstream switches are included in the *Fixed Round Trip Time* (FRTT) parameter received during connection setup, see chapter 3.2.4. When receiving the setup request for an ABR VC, this value corresponds to the forward propagation delay (τ_{fr}). For the return path, a symmetric connection is assumed, having the same value as the backward delay (τ_{br}). The RTT then corresponds to twice the forward or backward delay. The actual delay experienced by ATM cells may be longer, since queueing delays in upstream switches add to this RTT. For RM cells, which carry the feedback control information, the ATM Forum allowed the option of express queueing (see chapter 3.2.2). The RM cells may bypass the queued data cells, resulting in a feedback delay equal to the RTT. For the VS/VD switch implementation, it is assumed that upstream switches located in the previous loop implement this option and hence the RTT reflects the actual feedback delay.

5.7 Summary

In this chapter an analytic design based on linear control theory for a VS/VD flow control algorithm was developed. At first, a review of the related work, putting an emphasis on VS/VD switch algorithms and the use of control theory was presented. This included the introduction of common Explicit Rate ABR flow control algorithms for regular and VS/VD switches, which will be used for a performance comparison in the next chapter.

In the following, the VS/VD switch layout and the cell flow between the two control loops was explained. Each ABR connection passes through a per-VC queue of the virtual destination and at the virtual source all ABR VCs share a common ABR class queue.

Next, performance and design goals for the novel VS/VD switch algorithm were established.

A traffic model using a deterministic fluid flow approximation was used to derive linear differential equations describing the queue dynamics.

As a major result, a dual proportional-derivative (PD) control model for the VS/VD switch was designed. For an ABR connection, the PD-controller of the upstream loop controls the buffer level of the per-VC queue and is coupled to the PD-controller of the downstream loop by the VC's transmission rate in the next loop. The PD-controller of the downstream loop is responsible for a fair allocation of the available bandwidth and the buffer occupancy of the ABR class queue. A continuous time control model for the upstream PD-controller was developed and the key design goal of stability was proven by the use of the Nyquist criterion. Based on this continuous time control model, a discrete time model for the PD-controller was derived. The stable operation was also proven for this feedback control system. It was shown, that both controllers have a sufficient relative stability and fulfill the desired conditions for the steady state error.

Finally, some implementation issues were discussed, that have been taken into account for the realization of the dual PD-controller model for a VS/VD switch. Table 5-4 summarizes the configuration parameters of the dual PD-controller algorithm determined during the design process.

Parameter Name	Parameter Value
Measurement Interval	2 ms
Target ABR Class Queue Level	300 cells
ABR Class Queue Capacity	1000 cells
Target ABR per-VC Queue Level	50 cells
ABR per-VC Queue Capacity	250 cells
Measurement Dead-band ABR Queue	4 cells
Measurement Dead-band per-VC Queue	2 cells
Exponential Averaging Factor	0.95
Averaging Threshold	10 MBit/s
Bottlenecked Delta	max(0.05 · CCR2, 2 MBit/s)

Table 5-4: *Dual PD-Controller Configuration Parameters*

In the following chapter a performance evaluation of the newly developed VS/VD switch algorithm is presented. In addition to a summary of the analytic results derived in this chapter, discrete event simulations are used to assess the performance objectives of the novel flow control algorithm.

CHAPTER 6

Performance Evaluation

In the previous chapter, a novel ABR flow control algorithm for a virtual source / virtual destination (VS/VD) switch was developed, using linear control theory. In this chapter, an evaluation of the proposed algorithm will be presented, taking into account the design goals defined in chapter 5.3. The analytic results derived by linear control theory allowed to target certain performance characteristics already during the design phase. These goals will be addressed in the following section. Nevertheless, the remaining performance objectives can not be investigated analytically, and for this reason discrete event simulations are used. The simulation tool, the network components, and the different source models used for the performance evaluation of the VS/VD flow control algorithm are introduced. Next, the simulation scenarios and the results obtained are presented.

6.1 Analytical Results

Revisiting the performance goals established in chapter 5.3, the following objectives have already been met during the design phase.

- **Complexity**: The proposed dual PD-controller operates on the upstream and downstream control loops, using for each loop a set of two gain parameters K_p and K_d, for the proportional and the derivative component, respectively. K_d is fixed for both loops and K_p is dynamically adjusted by the VS/VD switch, as described in chapter 5.5. For this, the action delay of the previous loop is estimated using the upstream Current Cell Rate (CCR1) and the fixed round trip delay, which is conveyed during the connection set-up for the ABR VC. The target queue levels may be set automatically with respect to the output port link bandwidth. Hence, no user-adjusted or -defined parameters are required. The computational complexity of the PD-controller is restricted to simple operations to determine the queue error and growth. The measurements of ABR and high priority traffic are implemented by simple counters, which are evaluated at fixed averaging intervals, e.g., every 2 ms. This operation

may be performed as a background process on the VS/VD switch and leads to a constant overhead.

- **Scalability**: The newly developed VS/VD switch algorithm should scale with respect to the number of supported ABR connections. As discussed in chapter 5.6, buffer allocation is the major factor limiting the maximum number of parallel ABR VCs. The chosen maximum per-VC and ABR queue lengths will allow a sufficiently large number of 800 parallel ABR VCs, assuming a modern VS/VD switch [BWBM98], see chapter 5.6. The low computational complexity of the algorithm, as mentioned above, also supports its scalability.
- **Conformity**: The newly designed VS/VD switch algorithm operates in full accordance with the ATM Forum Traffic Management Specification [TM4.1]. The VS/VD switch implements the standard end system behavior, see chapter 3.2.3, and may be connected to other VS/VD switches, regular switches, or ABR end systems. The suggested VS/VD flow control algorithm does not require any special RM cell format or proprietary extensions to the standard ABR flow control mechanism.
- **Stability**: The stability, i.e., a bounded input signal will always lead to a bounded output signal, of the proposed dual PD-controller is analytically proven by the fulfillment of the Nyquist criterion. The controller parameters were determined in such a way that the criterion can be applied and a stable operation is guaranteed. The use of linear control theory allows to describe the ABR closed-loop feedback control system by mathematical functions and to define a control model. From this model performance parameters, like system stability, have been derived.
- **Robustness**: The relative stability of the control system, i.e., the distance of its stable operating point from the unstable operating region, is determined with the help of Nyquist and Bode plots. Again, the defined control model yields quantitative results regarding the relative stability of the system. The continuous time model for the upstream control loop of the VS/VD switch has a phase margin of 42 deg. and a gain margin of 5 dB, see chapter 5.5.1.4. For the discrete time model of the downstream loop, a phase margin of 65 deg. and a gain margin of 3 dB were determined, see chapter 5.5.2.3. Thus, the discrete time model shows a larger phase margin and a lower gain margin, with respect to the targeted values of 40-45 deg. and 5 dB. The impact on the transient response of the control system will be analyzed with the help of a system simulation in the following sections.
- **Steady State Error**: The steady state error of a control system is defined as the limit of the error signal, i.e., the difference between the actual output signal and its desired value, when time approaches infinity. For a ramp input signal, a bounded steady state error and for a step input a zero steady state error were targeted. The mathematical analysis for both the continuous time control model, see chapter 5.5.1.5, and the discrete time control model, see chapter 5.5.2.4, confirm the desired steady state errors.

The performance objectives mentioned above can be analyzed analytically with the help of the control model and the Laplace and *z*-Transforms in the complex *s*- or *z*-plane, respectively. The remaining goals of efficiency, i.e., buffer occupancy and link utilization, fairness, transient behavior, and disturbance rejection would require a mathematical analysis in the time domain,

assessing the actual temporal system behavior. In order to obtain quantitative results in the time domain for a controller model, a re-transformation of Laplace or *z*-Transforms is required. However, this is analytically not tractable for the transfer functions derived from the dual PD-controller model in the previous chapter.

As an alternative, the system behavior may be studied either by observing an actual implementation of the proposed flow control algorithm on a real VS/VD switch or by simulation. Not only is the first approach outside the scope of this thesis, but it has also the disadvantage of being costly and time consuming. Furthermore, the possible evaluation scenarios are limited by the physical restrictions of the testing environment. Hence, in this thesis a computer simulation is used to evaluate the performance of the developed VS/VD flow control algorithm and to check the correctness of the control model, defined in the previous chapter.

6.2 Discrete Event Simulation

Computer simulations offer the possibility to investigate complex systems without actually implementing them. The real system is represented by a *simulation model* that reflects its key properties. As a result, performance metrics of the modeled system are obtained, that allow to draw conclusions regarding the behavior of the actual system. The introduction to the terms and concepts of computer simulations presented in this section follows closely the works of [BrFS87] and [SpHo95].

The time period covered by the simulation is called the *simulation time* and differs in general from the actual time the simulation program needs for its execution on a computer system. Like the real system to be evaluated, the simulation model will contain certain system components, which are called *entities*. These entities are related by interactions, depending on the simulation time. An entity is characterized by its current state and a set of transitions, that modify the entity's state during the simulation time. These state changes are also called *events* and the time instants at which a system enters a state are called *event epochs*. Since an event-oriented simulation is used for the analysis of the VS/VD switch algorithm, this type of simulation is addressed in the remainder of this section.

In an event-oriented simulation, the simulation time skips from one event epoch to the next and no actions are performed in between. For each event, a subroutine is defined, e.g., in a high level programming language, that defines the actions to be executed and the next events triggered together with a corresponding activation time. The *simulation environment* maintains a *simulation clock*, indicating the simulation time, together with an *event queue*. This event queue stores all pending events ordered by their activation dates. An *event scheduler* processes this event queue, adjusting the simulation clock, and invoking the subroutine of the current event, see figure 6-1. Different events having the same activation time may be executed according to a certain scheduling strategy, e.g, first come, first served, or in an arbitrary order.

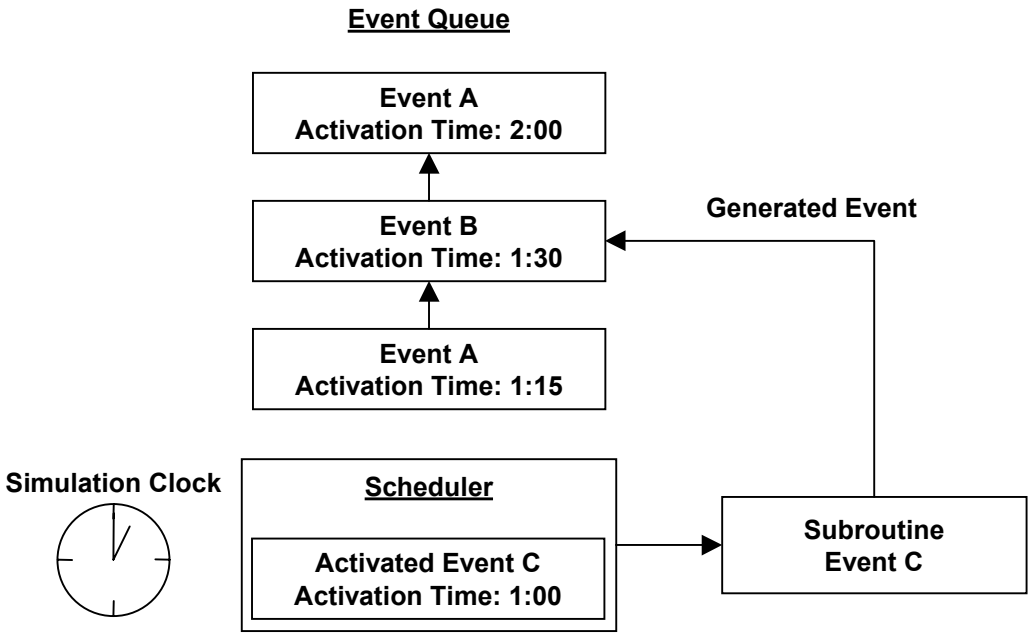

Figure 6-1: *Event-Oriented Simulation Environment*

State changes in *discrete time* systems occur only at certain points in time, like in the event-driven systems described above. These time instants need not to be equally spaced and may even be random. Due to its simple structure and general approach, discrete-time, event-driven simulations, also called *discrete event simulations*, are a very common simulation technique. They are applied in a wide area of application domains, like queuing systems, data networks, or road traffic [SpHo95].

The state change of an entity may also evolve *continuously* in time in contrast to discrete time systems described above. In a continuous time simulation, differential equations, that describe the simulation model, are solved numerically. This kind of simulation is often used for physical systems, where the system behavior may be described by differential equations and the system state changes continuously over time.

In case of ABR flow control, it would be possible to define a continuous time simulation model, based on the traffic model introduced in chapter 5.4. Due to the deterministic fluid flow approximation of a stochastic process, this simulation model would allow only a coarse modeling of traffic ATM flows, not taking into account individual cell boundaries. However, for the analysis of the newly designed VS/VD flow control algorithm a detailed simulation of the ATM cell flow and the single Resource Management (RM) cells, carrying the feedback information is necessary. Furthermore, the analytical results using linear control theory show that the resulting algebraic equations in the complex frequency domain, expressed by Laplace and *z*-Transforms, can in general not be re-transformed to the time domain. For these reasons, the evaluation of the proposed VS/VD flow control algorithm, implementing the dual PD-controller, will be performed using a discrete event simulation.

Using a *cell-based* simulation model allows a detailed investigation of the temporal behavior of the system, taking into account the effects of individual RM cells. The implementation of the ABR end system rules and the ABR flow control mechanism, as defined by the ATM Forum Traffic Management Specification [TM4.1], leads to a realistic simulation model. The sending and receiving of individual ATM cells at the different simulation entities is treated as an event. This approach models the temporal flow of single ATM cells from the sender to the receiver along the network path and its related effects on the states of the simulation entities. A drawback of this cell-based approach is the large number of events that have to processed by the simulation tool. A single ABR connection at 100 MBit/s passing through two ATM switches along the path from source to destination generates a total of approximately 3.5 million events for a simulation time of 1 second. As a consequence, an efficient implementation of the simulation software is crucial.

In the following section, a detailed description of the actual simulation environment used for the performance evaluation of the VS/VD flow control algorithm is presented.

6.3 Simulation Environment

The *NIST ATM Simulator* is a discrete event simulation software tool developed by the National Institute of Standards and Technology (NIST) [Golm98]. The software is based on the Network Simulator tool of the Massachusetts Institute of Technology and offers predefined components for common ATM network elements like switches and end systems. The software is written in the C programming language, runs on UNIX platforms, and provides a graphical user interface under the X Window System. This graphical interface allows the display of simulation parameters, such as bar graphs or meters, and a user-friendly input of network topologies. The software has been maintained and supported by NIST for several years, and a user manual as well as a programmer's guide are available [Golm98]. The simulation scenario, specifying the network elements and topology, is stored in a configuration file that can be loaded at the start of the simulation run.

The software has a modular structure with a core library that implements the simulator kernel functions, like for event handling and queue management, and a component library, that may be modified according to the special needs of the user.

6.3.1 Component Classes

There exist four *classes* of components [Golm98], which are introduced in the following and may contain several types of network elements.

6.3.1.1 Switch

The *switch* class defines ATM switches, that are able to route cells from one input to an output port, may buffer cells in queues, and support strict priority queueing for different service classes (high/low priority traffic). A switch introduces a switching delay and operates at an

internal processing rate, when serving the queues. The switch class originally consisted of three types, two offering credit-based flow control protocols, not compatible to the ATM Forum Traffic Management Specification [TM4.1], and one rate-based model providing the standard ABR flow control mechanism [Golm98]. For this thesis, the switch class was extended by a new fourth type, implementing the VS/VD switch layout introduced in chapter 5.2. Regarding the ABR flow control mechanism, the VS/VD switch type offers the ERICA+ scheme for VS/VD switches (see chapter 5.1.2) and the novel dual PD-controller developed in the previous chapter.

The Explicit Rate switch components used in the discrete event simulations use the following default configuration and input parameters.

Parameter Name	Parameter Value
Switching Delay	10 µs
Internal Processing Rate	155 MBit/s
Output Queue Size	7000 cells

Table 6-1: *Explicit Rate Switch Parameters*

The same switching delay and internal processing rate is also used for VS/VD switches.

6.3.1.2 Broadband Terminal Equipment

The *Broadband Terminal Equipment* (BTE) class specifies, together with the ATM applications (see section 6.3.1.4) the end systems connected to the ATM network. On the sending side of an ATM connection, the BTE forwards the source traffic generated by the ATM application to the next ATM switch on the network path. For this, the BTE provides input and output queues for cell buffering. On the destination side, the BTE counts the cells received from the last switch on the network path, calculates the transmission delay, and discards the cells. In case of an ABR connection, the BTE implements the ABR end system behavior, according to the ATM Forum Traffic Management Specification [TM4.1], see also chapter 3.2.3. Thus, it controls the *Current Cell Rate* (CCR) on the sending side of the ABR *Virtual Connection* (VC). The BTE class provides two credit-based and two rate-based types (one regular BTE and one for hybrid fiber coax networks) [Golm98].

If not otherwise stated, the following configuration parameters (see table 6-2) are used for a BTE in the simulations presented in section 6.4. The ABR connection parameters apply only, if an ABR application is attached to the corresponding BTE. Furthermore, if any VS/VD switches are present along the network path from the sender to the receiver, these ABR connection parameters are also used for the configuration of their virtual sources (VS). For a detailed explanation of the ABR connections parameters, please refer to chapter 3.2.4.

Parameter Name	Parameter Value
BTE Input Queue Size	∞
BTE Output Queue Size	∞
Peak Cell Rate	149.76 MBit/s
Minimum Cell Rate	1.49 MBit/s
Initial Cell Rate	7.49 MBit/s
Tagged Cell Rate	0.00424 MBit/s
Nrm	32 cells
Trm	0.1 s
Rate Increase Factor	1
Rate Decrease Factor	1/32768
ACR Decrease Time Factor	0.5 s
CRM	32
Cutoff Decrease Factor	0.0625

Table 6-2: *BTE and ABR Connection Parameters*

6.3.1.3 Links

The *link* class implements the bi-directional physical connection between switches or between BTEs and switches. There exists only a single link type, with the two input parameters speed and distance. The link accepts one ATM cell per cell slot and delivers the cell at its other end after a propagation delay corresponding to 5 µs/km. A common physical layer also used for the transport of ATM cells is defined by the *Synchronous Optical Network* (SONET) [McSp94]. The link speed is adjusted to a cell rate supported by SONET, taking into account the SONET frame overhead [UNI3.1]. For this work, the link output parameter utilization was modified to represent the percentage of link bandwidth used per measurement interval, rather than a cell rate in MBit/s. For the simulation studies presented in section 6.4, a link speed of 149.76 MBit/s is assumed, if no other information is given.

6.3.1.4 ATM Applications

This class consists of the *source traffic generators*, that produce a certain number of ATM cells per time period, according to the source model selected. ATM applications belong to one of the ATM Forum service categories [TM4.1], i.e., *Constant Bit Rate* (CBR), *Variable Bit Rate*

(VBR), *Unspecified Bit Rate* (UBR), or *Available Bit Rate* (ABR), see also chapter 2.8.2. The recently specified service class *Guaranteed Frame Rate* (GFR) [TM4.1] is not yet supported by the NIST ATM Simulator tool. Several ATM applications belonging to the same service category may be attached to a single BTE, multiplexing their cell streams at the sender side and demultiplexing it at the receiving side, respectively. The NIST ATM Simulator tool provides 12 different types of ATM applications, and in the following section the ones used for the performance evaluation of the newly designed VS/VD switch algorithm will be discussed in greater detail.

6.3.2 Source Models

The various traffic sources connected to the simulated ATM network will be responsible for the traffic load experienced at the switches along the connection path. Hence, it is desirable to generate different traffic streams, in order to investigate the network behavior under several conditions. The following traffic generators are used in the simulation scenarios, presented in section 6.4. Beside the source models for the ABR service category, traffic models for the high priority classes *Constant Bit Rate* (CBR) and *Variable Bit Rate* (VBR) are used to generate background traffic that affects the ABR capacity.

6.3.2.1 Available Bit Rate - Persistent

ABR persistent sources always have data to send and thus fully utilize the available bandwidth indicated to the source. For this reason, these sources are also called *greedy*. As a consequence, the CCR of the source equals the *Allowed Cell Rate* (ACR), which is determined by the ABR end system rules and the current network condition. Persistent ABR sources are used to determine the efficiency of the flow control algorithm, in the sense that the maximum possible link and buffer usage is reached.

6.3.2.2 Available Bit Rate - On/Off

As described in chapter 2.8.2, the ABR service category is used to transport bursty traffic of interactive applications or aggregated LAN traffic. The former may be modeled by an on/off traffic stream, where the source is in one of two alternating states, either *active* (on) or *idle* (off) [Jiao00]. During the active phase, a constant bit rate cell stream is injected into the network. After sending the burst, the source enters an idle period, where it does not send any data. The durations of the active and idle periods are drawn from an exponential distribution. Such an ABR on/off traffic source is provided by the NIST ATM Simulator tool, where the mean values for the active and idle times as well as the transmission rate during the on period are defined as input parameters in [Golm98]. The bursty nature of ABR on/off sources allows a multiplexing of different connections, where an effective bandwidth for each source may be calculated taking into account the average idle and active times as well as the transmission rate. The sum of the effective bandwidths of the multiplexed connections should not exceed the link capacity, in order to allow transient queues to drain. Since the ABR source is not persistent, a mismatch between the ACR and the ABR source rate may occur. Either the ACR exceeds the source rate,

then the VC is bottlenecked at the source, or the ACR is lower than the source rate, then the associated BTE component has to buffer the surplus cells.

6.3.2.3 Available Bit Rate - Self-Similar

As mentioned above, the ABR service category is suited for the transport of aggregated traffic, either as an interconnection between two LANs or as an access link to the internet, carrying WAN traffic. Since the mid-1990s, research has demonstrated that the traffic characteristics of aggregated IP-traffic are not accurately modeled by classical Poisson models. As an improvement, self-similar models were developed that better reflect the long-range dependence of LAN [LTWW94] or WAN [PaFl95] traffic.

A stationary process with an autocorrelation function $r(k)$ that is not summable, i.e.,

$$\sum_{k=1}^{\infty} r(k) = \infty , \tag{6.1}$$

is *long-range dependent* (LRD); if it is summable, it is *short-range dependent* (SRD) [Jiao00]. Self-similar processes have hyperbolically-decaying autocorrelation functions and are LRD. Their degree of long-range dependence is characterized by the *Hurst parameter H*, which takes values of $0.5 < H < 1$, while larger values indicate a stronger long-range dependence. The most common self-similar processes are *fractional Gaussian noise* (FGN) and *fractional autoregressive integrated moving-average* (F-ARIMA) [FrMe94], see also section 6.3.2.7.

The NIST ATM Simulator tool provides a self-similar ABR source model, based on a fast, approximative synthesis of FGN [Paxs97]. The algorithm synthesizes sample paths, which have a power spectrum that approximates that of FGN. The main idea of the approach can be summarized as follows. Let $f(\lambda, H)$ be the power spectrum of an FGN process. Then, one can generate a sequence of complex numbers z_i, which is called a frequency domain sample path, corresponding to this power spectrum. With the help of the inverse *discrete time Fourier transform* (DTFT), the corresponding x_i in the time domain can be calculated. Because of the properties of the DTFT, the generated x_i have the autocorrelation function and the power spectrum of FGN. Because DTFT and inverse DTFT can be computed efficiently by the *fast Fourier transform* (FFT) algorithm, this method has a small computational complexity compared to other approaches [Paxs97].

The main difficulty is to accurately compute the FGN power spectrum $f(\lambda, H)$, which is defined as [Paxs97]:

$$f(\lambda, H) = \mathcal{A}(\lambda, H)\left[|\lambda|^{-2H-1} + \mathcal{B}(\lambda, H)\right], \tag{6.2}$$

for $0 < H < 1$ and $-\pi \le \lambda \le \pi$, where

$$\mathcal{A}(\lambda, H) = 2 \cdot \sin(\lambda H) \cdot \Gamma(2H+1)(1-\cos\lambda) \tag{6.3}$$

$$\mathcal{B}(\lambda, H) = \sum_{j=1}^{\infty}\left[(2\pi j + \lambda)^{-2H-1} + (2\pi j - \lambda)^{-2H-1}\right]. \tag{6.4}$$

Since no closed form for the infinite summation in eq. 6.4 is known, the following approximation is used in the calculations.

$$\mathcal{B}(\lambda, H) \approx a_1^d + b_1^d + a_2^d + b_2^d + a_3^d + b_3^d + \frac{a_3^{d'} + b_3^{d'} + a_4^{d'} + b_4^{d'}}{8H\pi}, \tag{6.5}$$

with

$$d = -2H - 1,\ d' = -2H,\ a_k = 2k\pi + \lambda,\ b_k = 2k\pi - \lambda \ \text{[Paxs97]}. \tag{6.6}$$

The power spectrum computed with this approximation method is defined as $\tilde{f}(\lambda, H)$.

The algorithm to generate the approximated FGN sample path needs two inputs: the desired Hurst parameter *H* and the desired even number *n* of observations in the synthesized sample path. The fast approximation method is described by five steps [Paxs97].

1. Generate a sequence $\{f_1, ..., f_{n/2}\}$, with

$$f_j = \tilde{f}(\frac{2\pi j}{n}, H), \tag{6.7}$$

   corresponding to an FGN power spectrum, for frequencies from $2\pi/n$ to π.

2. Alter each $\{f_i\}$ by multiplying it by an independent exponential random variable with a mean value of 1, obtaining the new sequence $\{f'_i\}$.

3. Generate a sequence of complex values $\{z_1, ..., z_{n/2}\}$ with

$$|z_i| = \sqrt{f'_i}, \tag{6.8}$$

   such that the phase of z_i is uniformly distributed in the interval $[0, 2\pi]$. This random phase technique preserves the power spectrum and the autocorrelation of the $\{f'_i\}$ and ensures the independence of different generated sample paths. At the same time, the marginal distribution of the final result is normal, which is a property of FGN.

4. Construct $\{z'_0, ..., z'_{n-1}\}$ as

$$z'_j = \begin{cases} 0 & \text{if } j = 0 \\ z_j & \text{if } 0 < j \le n/2 \\ \overline{z}_{n-j} & \text{if } n/2 < j < n \end{cases} \tag{6.9}$$

   The $\{z'_j\}$ preserve the power spectrum of the $\{z_i\}$ and are symmetric about $z'_{n/2}$, so that they correspond to the Fourier transform of a real valued signal.

5. Apply the inverse Fourier transform to the $\{z'_j\}$, in order to obtain the approximate FGN sample path $\{x_i\}$.

In the ABR self-similar model implementation of the NIST ATM Simulator, the number of generated sample points is 4096 [Golm98]. The FGN sample path has a zero mean and variance σ^2 and is scaled by

$$A(i) = m + m\frac{x_i}{2\sigma}, \tag{6.10}$$

where m is the mean bit rate of the source, defined by the user. After this transformation, values greater than $2 \cdot m$ are set to $2 \cdot m$ and negative values to zero. Since $A(i)$ has a normal Gaussian distribution, 95 % of the values will belong to the interval $[0, 2m]$ [Golm98]. The values of $A(i)$ define the transmission rate for a certain time interval (bin) of length *Timegran* µs. Based on this interval, the number of cells per bin is calculated, which are evenly spaced over the interval. In order to avoid empty bins, *Timegran* should be larger than the smallest possible cell inter-arrival time, corresponding to the transmission rate of $2m$. Thus, the ABR self similar model generates cell transmission for a duration of $4096 \cdot$ *Timegran* µs [Golm98].

For the simulations presented in section 6.4, the following input parameter for the ABR self-similar model were used, see table 6-3.

Input Parameter Name	Parameter Value
Mean Bit Rate	70 MBit/ s
Timegran	300 µs
Hurst Parameter	0.8

Table 6-3: *Required Parameters for the ABR Self-Similar Model*

The Hurst parameter of 0.8 was chosen as a typical value for aggregated WAN traffic [Jiao00].

6.3.2.4 Constant Bit Rate

CBR connections represent high priority traffic with a constant transmission rate during the lifetime of the connection. The input parameters for this source type are cell rate, start time and amount of data to send [Golm98]. CBR sources are used to study the network behavior in the presence of infrequently changing load conditions. For the analysis of ABR flow control algorithms, starting or ending CBR traffic streams lead to sudden changes in the ABR capacity. Hence, a changing number of CBR sources allows to investigate the transient behavior of the flow control algorithm in the presence of congestion.

6.3.2.5 Variable Bit Rate - On/Off

This VBR source model is similar to the ABR model introduced in section 6.3.2.2. The only difference is that the service category of the traffic stream is VBR, and the on and off periods are not drawn from an exponential distribution, but are fixed. The model generates an on/off traffic stream, as it would originate from a bursty application. As for the ABR model, the source enters in turn two states: active (on) or idle (off), where during the on period a constant bit rate cell stream is generated. The input parameters for the model are the values for the active and idle times, and the transmission rate during the active period. VBR on/off traffic leads to a frequently changing ABR capacity, where congestion situations at the beginning of on periods alternate with underload conditions at the beginning of off periods.

6.3.2.6 Variable Bit Rate - MPEG-2 Video Real-Time

In order to investigate the impact of highly variant VBR background traffic on the behavior of the VS/VD switch algorithm, the transmission of MPEG-2 coded video sequences is modeled. MPEG-2 is a standardized video compression technique that allows to reduce the data rate required to transmit a full resolution (NTSC), full motion (30 frames/s) digital video from over 100 MBit/s to about 3 to 4 MBit/s [I13818-2]. Temporal redundancies are exploited by using the following three frame types.

- **Intra-Frames** (I-frames): Using intra-frame coding, a frame is divided into blocks of 8 by 8 pixels and each block is coded using a *Discrete Cosine Transformation* (DCT). The DCT coefficients are quantized in such a way that the low frequency components are represented with a higher accuracy than the high frequency components.
- **Predicted-Frames** (P-frames): P-frames use a coding algorithm similar to that of I-frames. Motion detection is used to predict blocks in the current frame from previous I- or P-frames, which results in a smaller frame size than for an I-frame.
- **Bidirectional-Frames** (B-frames): The coding of B-frames is similar to that of P-frames. The main difference is, that motion compensation is performed with respect to the previous and the next I- or P-frames, again resulting in a smaller frame size than for I-frames [LeGa91].

These frames are typically grouped together in a *group of pictures* (GOP), whose first frame is an I-frame which is then followed by a pattern of P- and B-frames. The frame pattern has a cyclic (*N*, *M*) format, where *N* is the spacing between successive I-frames and *M* is the distance between successive I- or P-frames, also called anchor frames [FrNg00].

The NIST ATM Simulator tool offers a source model for real-time VBR traffic carrying MPEG-2 video sequences [Golm98]. Due to the real-time streaming of the content, no large buffer delays are allowed. As a consequence, the model takes into account only the short range dependence of the transmitted frame sequence using a *Gamma-Beta autoregressive* (GBAR) model [Golm98].

The GBAR model has two important features, namely, the marginal distribution is Gamma distributed and the autocorrelation function is geometric [Heym97].

The probability density function of a Gamma distribution $Ga(\beta, \lambda)$ is defined as

$$f_{\Gamma}(x) = e^{-\lambda x} \frac{\lambda(\lambda x)^{\beta-1}}{\Gamma(\beta)} \quad , \tag{6.11}$$

with shape parameter β and scale parameter λ, where $\Gamma(x)$ denotes the Gamma function [GaWi94].

A Beta distributed random variable $Be(p, q)$ has the probability density function

$$f_{Be}(x) = \frac{\Gamma(p+q)}{\Gamma(p) \cdot \Gamma(q)} x^{p-1} (1-x)^{q-1} \quad \text{[Fish78].} \tag{6.12}$$

In order to define the GBAR model, let $Ga(\beta, \lambda)$ be a Gamma distributed random variable with shape parameter β and scale parameter λ. Let $Be(p, q)$ be a Beta distributed random variable with parameters p and q. The GBAR model uses the two properties:

- the sum of independent $Ga(\alpha, \lambda)$ and $Ga(\beta, \lambda)$ random variables is a $Ga(\alpha + \beta, \lambda)$ random variable, and
- the product of independent $Be(\alpha, \beta - \alpha)$ and $Ga(\beta, \lambda)$ random variables is a $Ga(\alpha, \lambda)$ random variable [Heym97].

Therefore it follows, that if X_{n-1} is $Ga(\beta, \lambda)$, B_n is $Be(\alpha, \beta - \alpha)$, and W_n is $Ga(\beta - \alpha, \lambda)$, and these three are mutually independent, then

$$Z_n = B_n \cdot Z_{n-1} + W_n \tag{6.13}$$

defines a stationary GBAR process $\{Z_n\}$, that has a marginal $Ga(\beta, \lambda)$ distribution [Heym97]. It is an autoregressive process of order one, because the current value depends on its predecessor and hence is also called a GBAR(1) process.

For the generation of successive frame sizes, the GOP GBAR source model of the NIST Simulator tool uses three independent stationary GBAR processes $\{Z_{ik}, k = 0,1,2,..\} \sim \text{GBAR}(a_i, r_i)$ for $i = 1, 2, 3$ [FrNg00]. The size of X_k of the k^{th} frame, starting with an I-frame as a first frame and using a GOP pattern of (N, M) is defined as [Golm98]:

$$X_k = \begin{cases} \lambda_1 Z_{1k} + \lambda_2 Z_{2k} + \lambda_3 Z_{3k} & \text{if } k = 1 \bmod N \\ \lambda_1 Z_{1k} + \lambda_2 Z_{2k} & \text{if } k \neq 1 \bmod N, \text{ but } k = 1 \bmod M \\ \lambda_1 Z_{1k} & \text{otherwise} \end{cases} . \tag{6.14}$$

The GOP frame sequence $\{X_k\}$ has a geometric autocorrelation function, composed of the geometric autocorrelations of the three component GBAR processes [FrNg00].

The parameters a_i, l_i, r_i $(i = 1, 2, 3)$ of the GBAR model can be estimated from the user-specified input parameters mean m_j, variance v_j, and the first-lag correlation r_j, for $j = I, P, B$, and a GOP pattern of (N, M) as [Golm98]:

$$\begin{aligned} \lambda_1 &= \frac{v_B}{m_B} \quad \lambda_2 = \frac{v_P - v_B}{m_P - m_B} \quad \lambda_3 = \frac{v_I - v_P}{m_I - m_P} \\ a_1 &= \frac{m_B^2}{v_B} \quad a_2 = \frac{(m_P - m_B)^2}{v_P - v_B} \quad a_3 = \frac{(m_I - m_P)^2}{v_I - v_P} \end{aligned} \tag{6.15}$$

$$\begin{aligned} \rho_1 &= \rho_B \\ \rho_2 &= \left(\frac{v_P \rho_P^M - v_B \rho_B^M}{v_P - v_B} \right)^{1/M} \\ \rho_3 &= \left(\frac{v_I \rho_I^N - (v_P - v_B) \rho_P^N - v_B \rho_B^N}{v_I - v_P} \right)^{1/N} \end{aligned} \tag{6.16}$$

The sizes of the k^{th} frame X_k is then calculated recursively according to eq. 6.14 using the values of Z_{ik}, $i = 1, 2, 3$. For this, three independent Gamma variables $W_{ik} \sim Ga(\alpha_i(1 - \rho_i), 1)$, and three independent Beta variables $B_{ik} \sim Be(\alpha_i \cdot \rho_i, (1 - \rho_i) \cdot \alpha_i)$, for $i = 1, 2, 3$ are generated. The beta variables are obtained using

$$B = \frac{X}{X + Y} \sim Be(x, y), \tag{6.17}$$

where the variable B is generated from two independent Gamma variables $X \sim Ga(x, 1)$ and $Y \sim Ga(y, 1)$ [Golm98]. According to eq. 6.13, the Z_{ik} are calculated as

$$Z_{ik} = B_{ik} \cdot Z_{i,k-1} + W_{ik}, \text{ for } i = 1, 2, 3. \tag{6.18}$$

As mentioned above, the GBAR(1) is a short-range dependent process, taking into consideration only the first-lag correlation. This is sufficient, because long-term correlations do not have a significant impact on the cell loss ratio of real-time VBR video transmissions with a low bandwidth utilization and small buffer sizes [RyEl96].

The required input parameters for the GBAR VBR real-time model have been determined by measurements and statistical analysis of MPEG-2 coded video sequences and are listed in table 6-4 [Jiao00].

Input Parameter Name	Parameter Value
Frame Rate	30 Hz
GOP Frame Spacing (N, M)	(12, 3)
Average I-Frame Size (in kBits)	92.38
Average P-Frame Size (in kBits)	50.99
Average B-Frame Size (in kBits)	24.62
Standard Deviation of I-Frames (in kBits)	24.46
Standard Deviation of P-Frames (in kBits)	10.03
Standard Deviation of B-Frames (in kBits)	5.02
Interframe Correlation of I-Frames	0.85
Interframe Correlation of P-Frames	0.34
Interframe Correlation of B-Frames	0.44

Table 6-4: *Required Parameters for the VBR MPEG-2 GBAR(1)Video Model*

For the VBR real-time traffic model, successive frame sizes for the corresponding GOP frame types are generated using the GBAR(1) model. The values are rounded to the nearest integer value and are translated to an according number of ATM cells, taking into account the overhead involved in an AAL5 transmission, see chapter 2.7.2.4. The ATM cells generated are transmitted by the VBR source at evenly spaced time instants throughout the inter-frame interval of $1/30$ s.

6.3.2.7 Variable Bit Rate - MPEG-2 Video Non Real-Time

In contrast to the previous model, where real-time streaming of MPEG-2 coded video sequences is covered, this source model deals with the *non real-time* transmission of MPEG-2 video frames. Possible application areas are video distribution and buffered playback, e.g., for set-top boxes. There exist relaxed delay and jitter constraints for the transmission, because the video material is received and buffered before playback. Therefore, the non real-time VBR (nrt-VBR) service category is used for this type of traffic, see chapter 2.8.2. Since the NIST ATM Simulator does not offer a non real-time MPEG-2 traffic generator, the tool was enhanced by an appropriate model [Jiao00].

The model introduced is based on [GaWi94] and takes self-similarity and long-range dependence into consideration. This is required, because for non real-time video transmissions larger buffer delays and a higher bandwidth utilization than for real-time streaming are possible. Since a self-similar process has observable bursts on all time scales, it shows a long-range dependence property, that means, values at any time are correlated with all future values. The autocorrelation function decays hyperbolically rather than exponentially.

The source model presented in this section captures two important aspects of MPEG-2 coded video, namely, a precise marginal distribution and an autocorrelation function with a long-range dependence [GaWi94]. The main idea is to generate a normal distributed random value using a fractional autoregressive integrated moving-average (F-ARIMA) process with Hurst parameter H and the average frame size of the corresponding MPEG-2 frame type. This process is responsible for the long-range dependence property of the model. The actual frame size distribution for the different MPEG-2 frame types is approximated by a Gamma marginal distribution with shape parameter k and scale parameter b. The Gamma distribution has a mean value of $b \cdot k$ and a variance of $b^2 \cdot k$. The size for the respective frame is calculated by applying the inverse of the Gamma distributed random variable to the generated normal distributed random value [Jiao00].

In a fractal ARIMA(p, d, q) process, the orders p and q are the classical autoregressive and moving-average parameters. The fractional difference parameter d is related to the Hurst parameter H by $H = d + 0.5$. The F-ARIMA(p, d, q) process $\{X_k\}$ ($k = 0, 1, 2, ...$) is represented as

$$\Phi(B)\nabla^d X_k = \Theta(B)\varepsilon_k \text{, where} \tag{6.19}$$

$$\Phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p, \;\; \Theta(B) = 1 - \theta_1 B - \ldots - \theta_p B^p \tag{6.20}$$

are polynomials in the backshift operator $BX_k = X_{k-1}$, and ∇^d is the fractional differencing operator defined as

$$\nabla^d = (1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \text{, with} \tag{6.21}$$

$$\binom{d}{k}(-1)^k = \frac{\Gamma(-d+k)}{\Gamma(-d)\Gamma(k+1)}, \tag{6.22}$$

and ε_k being a white noise process [Jiao00].

In order to calculate successive frame sizes for each of the three MPEG-2 frame types, fractional noise is generated, using a fractional ARIMA(0, d, 0) process $\{X_k\}$. The process has Gaussian marginals with zero mean and variance v_0. Its autocorrelation function has an asymptotically hyperbolic shape and is defined as

$$\rho_k = \frac{d(1+k)\ldots(k-1+d)}{(1-d)(2-d)\ldots(k-d)} \text{ [GaWi94].} \tag{6.23}$$

The values $\{X_k\}$ of the F-ARIMA(0, d, 0) process are generated in the following way. X_0 is a value from the Normal distribution $N(0, v_0)$. Setting $N_0 = 0$ and $D_0 = 1$, one can generate the next n points by iterating the following calculations for $k = 1, 2, 3, ..., n$ [GaWi94].

$$N_k = \rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j} \cdot \rho_{k-j} \tag{6.24}$$

$$D_k = D_{k-1} - \frac{N_{k-1}^2}{D_{k-1}} \tag{6.25}$$

$$\phi_{kk} = \frac{N_k}{D_k} \tag{6.26}$$

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j} \text{ for } j = 1,2,\ldots,(k-1) \tag{6.27}$$

$$m_k = \sum_{j=1}^{k} \phi_{k,j} \cdot X_{k-j} \tag{6.28}$$

$$v_k = (1-\phi_{kk}^2) \cdot v_{k-1} \tag{6.29}$$

Each X_k is then chosen from the Normal distribution $N(m_k, v_k)$ [GaWi94].

A realization of the fractional ARIMA(0, d, 0) process $\{X_k\}$ has a Gaussian distribution and cannot match arbitrary empirical distributions. Hence, inversion methods need to be used to transform the Gaussian marginal to the desired Gamma marginal distribution by

$$Y_k = F_\Gamma^{-1}\left(F_N(X_k)\right) \text{ for } k > 0, \tag{6.30}$$

where F_N is the cumulative probability function of the Normal distribution and F_Γ^{-1} is the inverse cumulative probability function of the Gamma distribution [GaWi94]. The Gamma distribution scale parameter *b* and the shape parameter *k* for each MPEG-2 frame type can be calculated from the average frame size and standard deviation. These values have been determined from statistical analysis of MPEG-2 coded video frames [Jiao00]. The sequence of generated frame types is defined by the GOP pattern.

Thus, the proposed non real-time MPEG-2 video traffic model requires a total of four source parameters for each frame type, namely, the average frame size, the Gamma shape parameter *k* and scale parameter *b*, and the Hurst parameter *H*. The values of the Hurst parameter *H* are derived from measurements of MPEG-2 coded video sequences [Jiao00], using variance-time plots and the rescaled adjusted range statistics (R/S) analysis [LTWW94].

Input Parameter Name	Parameter Value
Frame Rate	30 Hz
GOP Frame Spacing (*N*, *M*)	(12, 3)
Average I-Frame Size (in kBits)	92.38
Average P-Frame Size (in kBits)	50.99
Average B-Frame Size (in kBits)	24.62
Gamma Parameter *b* of I-Frame	1.81
Gamma Parameter *b* of P-Frame	1.38
Gamma Parameter *b* of B-Frame	4.08
Gamma Parameter *k* of I-Frame	10.12
Gamma Parameter *k* of P-Frame	12.28
Gamma Parameter *k* of B-Frame	12.36
Hurst Parameter *H* of I-Frame	0.90
Hurst Parameter *H* of P-Frame	0.80
Hurst Parameter *H* of B-Frame	0.80

Table 6-5: *Required Parameters for the VBR MPEG-2 Self-Similar Video Model*

Although this model is designed for non real-time usage, the parameter frame rate describes at what intervals frames are generated at the source.

Frame sizes are generated separately for each frame type, following the two steps:

1. Generate one normal distributed random value by using the F-ARIMA process with Hurst parameter H and the average frame size for the corresponding frame type;
2. Invert it to a Gamma distributed random value, using Gamma parameters b and k; this determines the frame size in bits.

This frame size in bits is converted to the corresponding number of ATM cells, taking into account the AAL 5 overhead, see chapter 2.7.2.4. The generated cells are then evenly distributed during the inter-frame interval.

It has to be noted that for the implementation of this model no closed-form for the inverse *Gamma cumulative density function* (CDFGamma) could be found. Taking the normal distributed random value $y(t)$ of the F-ARIMA process and applying the inverse of the cumulative Gamma distributed density function $CDFGamma^{-1}$, yields a corresponding value x. In order find the roots of

$$CDFGamma(x) - y = 0 \tag{6.31}$$

the bisection method was used. This method locates such a root by repeatedly narrowing the interval around the solution [Zach96]. The computation time of this method can be so long that it may affect the performance of the traffic generator. In this case, the frame sizes will be generated and stored at the beginning of the simulation run, so that the source model can provide successive frame sizes, whenever needed.

6.3.3 Performance Metrics

Since the discrete event simulations will be used for the performance evaluation of the newly developed dual PD-controller, performance metrics to be monitored have to be defined.

For the comparison of the novel VS/VD switch algorithm and already existing Explicit Rate algorithms, introduced in chapter 5.1.1, the following data will be investigated.

- **ACR at the Source**: The *Allowed Cell Rate* (ACR) at the ABR source indicates if and how fast a fair bandwidth allocation is reached and if the rate control leads to oscillations.
- **Queue Levels**: The buffer occupancy of the ABR class queue and, if present, of the per-VC queues at the bottleneck switch is investigated. Maximum values are reported in order to check if the provided buffer capacities are sufficient. Furthermore, the temporal behavior of the queue length indicates the transient behavior of the system and if a steady state is reached. In addition, the queue lengths and their variance determine the end-to-end delay and jitter for an ABR connection.
- **Delay**: The end-to-end delay of an ABR cell is calculated as the difference of the receive and send time at the corresponding BTEs. According to the QoS parameters of the ATM Forum the *Maximum Cell Transfer Delay* and the *peak-to-peak Cell Delay Variation*, see chapter 2.8.1.2, are investigated. Although no restrictions for these values are defined for ABR traffic, the goal of an ABR flow control algorithm is to provide as low values as possible.

- **Link Utilization**: The utilization of the bottleneck link is presented to check, if all bandwidth capacity is claimed by ABR traffic, since a utilization of 100 % is targeted.
- **Throughput**: The number of ABR cells transmitted per connection and per inspection interval is used to calculate the throughput value.

For the assessment of the performance of the suggested VS/VD switch algorithm, the metrics ACR at the source, queue levels, and link utilization, as defined above, are used. In addition, the ACR for the downstream loop at the virtual source of the bottleneck VS/VD switch is analyzed, in order to investigate the behavior of the downstream controller. This ACR value indicates how the virtual source is able to track the changing ABR capacity, which is influenced by high priority background traffic. In the case of non-persistent ABR sources, the buffer occupancy at the BTE component is also recorded. This queue is used if the generated source traffic exceeds the current ABR capacity of the connection. For the comparison of the new dual PD-controller with the existing ERICA+ for VS/VD algorithm, delay and throughput values as introduced above will be observed.

For the calculation of the delay and throughput values, all received cells are logged at the receiving BTE. The link utilization is determined every 350 cells times, which corresponds to approximately 1 ms for a link speed of 155 MBit/s. The queue lengths are determined whenever a queue is served, i.e., a cell is dequeued. ACRs are evaluated if a backward RM cell is received. The queue length and ACR values are recorded only, if the sample interval between two successive sampling instants is at least 10 µs, or approximately 3 cell times assuming a cell processing speed of 155 MBit/s. This resolution is sufficient to capture all dynamic and temporal effects and allows to reduce the amount of data to be stored and processed.

6.4 Simulation Scenarios

This section presents the different simulation scenarios used for the performance evaluation of the investigated ABR flow control algorithms. The network topology and the traffic sources involved are chosen to represent a large number of load conditions, that allow an insight to the performance of the flow control algorithm and a generalization of the results.

In addition to the analytical results presented in section 6.1, the performance goals of efficiency, fairness, transient behavior, and disturbance rejection have to be analyzed using discrete event simulations. The objective of efficiency is represented by the buffer occupancy and link utilization at the bottleneck switch.

The default configuration and input parameters related to the different Explicit Rate ABR flow control algorithms investigated have been introduced in chapter 5.1.1. The parameter set for the ERICA+ for VS/VD algorithm is listed in chapter 5.1.2 and for the dual PD-controller in chapter 5.7, respectively. The configurations of the simulator components have been introduced in the previous section 6.3. If not otherwise stated, these default settings are used for the simulation experiments.

In general, each ATM application component (source or destination) is associated with a single dedicated BTE. For reasons of simplicity, the BTE is omitted in the figures illustrating the network topology of the different simulation scenarios. If two or more ATM application components are connected to a single BTE, this will be explicitly noted. In the remainder of this chapter, ATM source application names start with the letter "S", whereas destination names start with a "D". ABR sources and destinations are depicted by white circles, VBR or CBR applications as shaded circles, respectively. All ABR connections are uni-directional and are identified by a VC number, which is indicated in the names of the ATM applications at the start and the end of the VC. For instance, ABR VC no. 3 starts at the ABR application "S3" (source) and ends at "D3" (destination).

In the following sections, the various simulation scenarios and the observed performance metrics are presented.

6.4.1 Comparison of Explicit Rate and VS/VD Switches

In order to compare the performance of standard Explicit Rate (ER) and VS/VD switches, the first scenario is illustrated in figure 6-2. Two persistent ABR VCs 1 and 2 share a network path across four switches and at the third switch, two CBR connections are joined. The links interconnecting the switches are 500 km long, all end systems are connected by 100 km links. The CBR connection S3-D3 starts a 15 MBit/s transmission after 100 ms, in addition CBR VC 4 starts to send at the same rate after 200 ms, whereas the two persistent ABR connections start immediately. As a consequence, the link between switch 1 and 2 is the bottleneck link and switch 1 is the bottleneck switch for the first 100 ms of the simulation. After that, switch 3 becomes the bottleneck switch, and its link to switch 4 becomes the bottleneck link. Since the two ABR VCs are symmetric and greedy sources are used, the two ABR connections show exactly the same behavior. For this reason, only the results for the first VC are presented. The last switch serves the destinations on dedicated lines and therefore never experiences any congestion. Hence, performance metrics for the fourth switch are omitted.

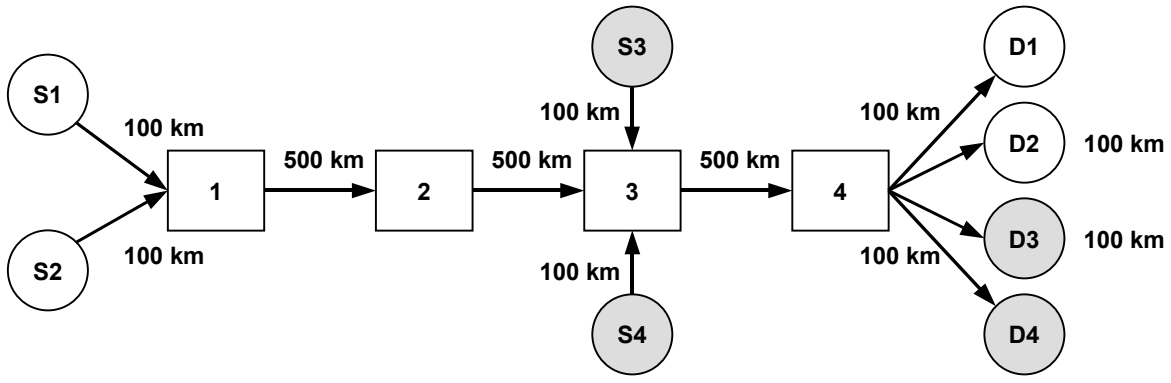

Figure 6-2: *Scenario 1: Comparison of ER and VS/VD Switches (CBR Sources)*

In three simulation runs each of the standard ER algorithms EPRCA, DERA, and ERICA+, introduced in chapter 5.1.1, is executed on all four switches. In a fourth experiment, the four switches are of VS/VD type, and the novel PD-Controller algorithm is deployed.

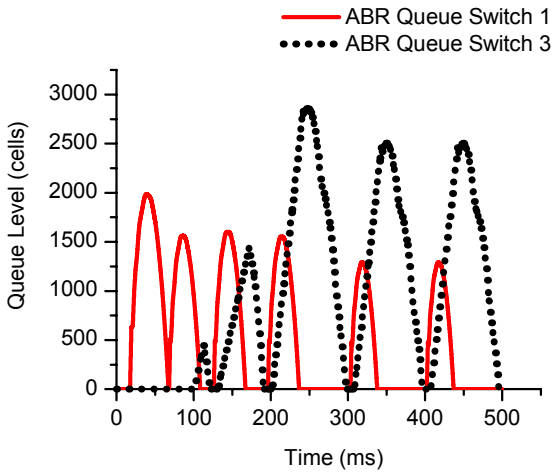

Figure 6-3.a: *ABR Class Queues EPRCA*

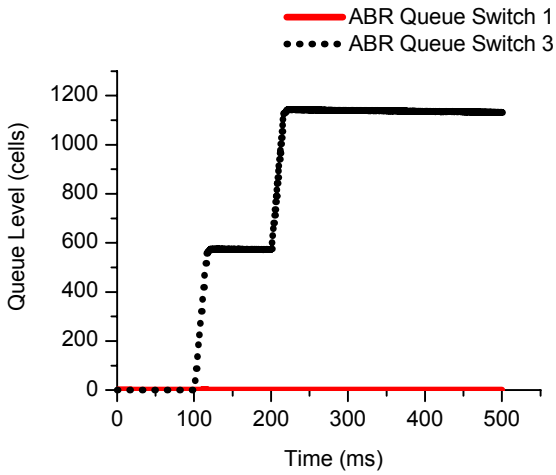

Figure 6-3.b: *ABR Class Queues DERA*

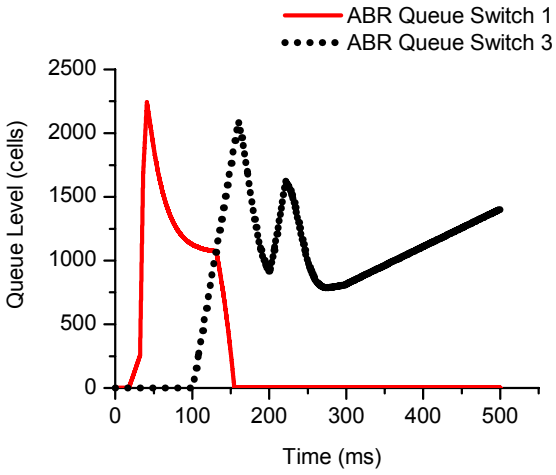

Figure 6-3.c: *ABR Class Queues ERICA+*

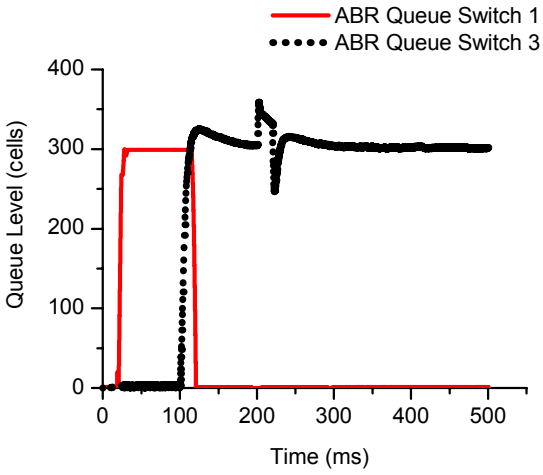

Figure 6-3.d: *ABR Class Queues VS/VD*

The ABR class queue levels of the EPRCA, DERA, ERICA+, and the VS/VD flow control algorithms at the first and third switch (the bottleneck switches) are depicted in figures 6-3.a to 6-3.d, respectively. For the VS/VD algorithm, the per-VC queues at the switches 1, 2, and 3 are shown in figure 6-3.e.

The simple EPRCA offers only a coarse queue control with its three congestion states. As a consequence, the buffer lengths exhibit large oscillations, even at the first switch after the congestion situation is revoked. This behavior is confirmed by the large oscillations of the source rate, illustrated in figure 6-4.a.

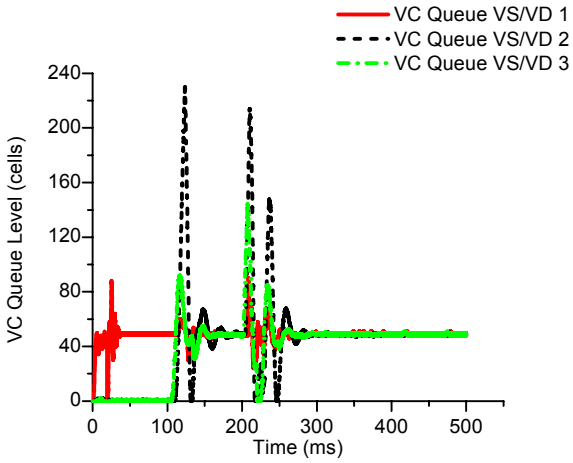

Figure 6-3.e: *Per-VC Queues VS/VD*

The DERA algorithm tracks exactly the fair share of the ABR connections, which is shown in figure 6-4.a. This leads to a zero buffer occupancy at the first switch, because the correct rate is advertized right from the start-up of the connection. However, when overload conditions occur at the third switch at 100 and 200 ms, the resulting queue backlogs are not drained, as shown in figure 6-3.b. The reason for this is, that the DERA algorithm takes into account only the fair share and not the queue level at the switch. As a result, switch queue levels may grow without bounds, if no underload conditions occur during which built-up queues may be cleared.

The ERICA+ algorithm tries to reach the target buffer delay T_0 of 3 ms, which corresponds to a queue length of about 1000 cells. The minimal recommended value for T_0 is $1/8$ of the round trip time [Kaly97] and hence 2.1 ms for this scenario. ERICA+ leads to a relatively high buffer overshoot of approximately 220 % during the first congestion situation at each switch and shows a long settlement time, see figure 6-3.c. This is due to the combined impact of queue level and traffic load on the explicit rate calculated, depicted in figure 6-4.a.

The newly developed VS/VD algorithm drives quickly both the ABR class queues (figure 6-3.d) and the per-VC queues (figure 6-3.e) to the desired levels of 300 and 50 cells, respectively. The per-VC queues are used to buffer the transient overload at 100 and 200 ms. When the bottleneck switch changes from VS/VD switch 1 to 3 at 100 ms, the ABR class queue at the first switch is drained and the one at the third switch is filled, see figure 6-3.d. When VS/VD switch 2 becomes bottlenecked downstream after 100 ms, its upstream control loop may increase the ACR of VS/VD switch 1 (figure 6-4.b) to fill its per-VC queue (figure 6-3.e). The downstream controller however has to follow the bottleneck rate of VS/VD switch 3 and hence the ABR class queue at the second VS/VD switch remains empty (not shown in figure 6-3.d). The PD-controller tracks the fair share accurately and uses small rate changes to adapt the queue levels when possible, see figure 6-4.b.

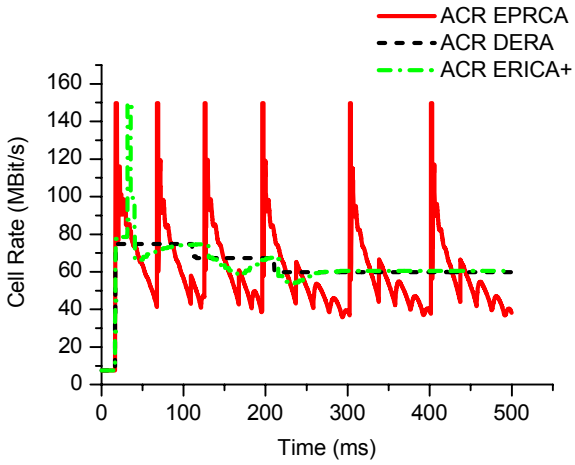

Figure 6-4.a: *Source ACRs (ER Algorithms)*

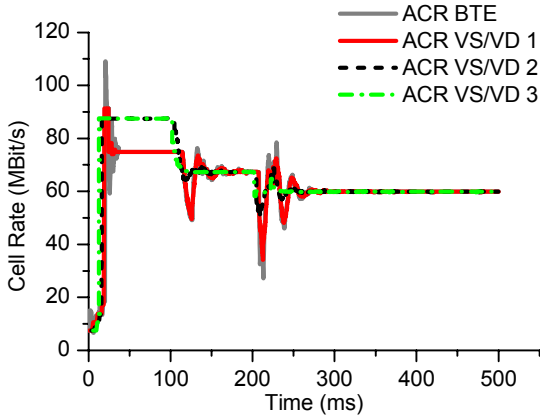

Figure 6-4.b: *Source and VS/VD ACRs*

Figures 6-5.a and 6-5.b show the utilization of the bottleneck link between switches 3 and 4 for the ER and the VS/VD algorithms, respectively. Except for the simple EPRCA with its large rate oscillations, all other flow control algorithms do not suffer from empty ABR class queues throughout the duration of the ABR connections. Therefore, they are able to fully utilize the available bandwidth not claimed by the high priority CBR sources.

The high buffer occupancy and their large variation at the ER switches result in a fluctuating end-to-end delay on a high average level, see figure 6-6.a. The VS/VD switches along the network path have more steady queue lengths than the ER switches, which leads to a smooth end-to-end delay curve, see figure 6-6.b.

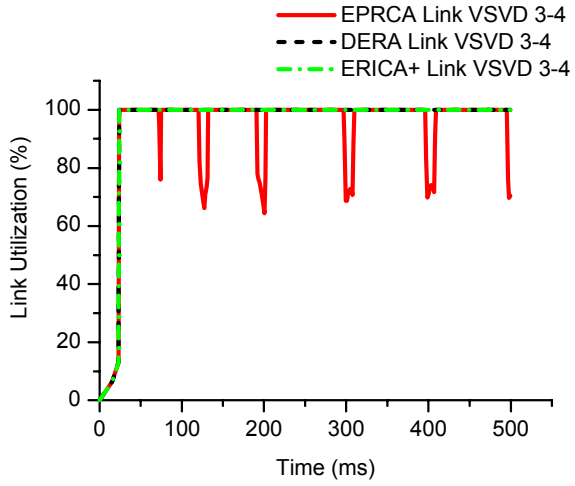

Figure 6-5.a: *Link Utilization (ER Algorithms)*

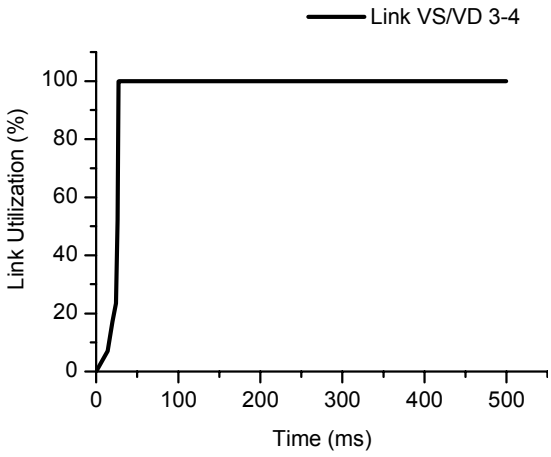

Figure 6-5.b: *Link Utilization (VS/VD Algorithm)*

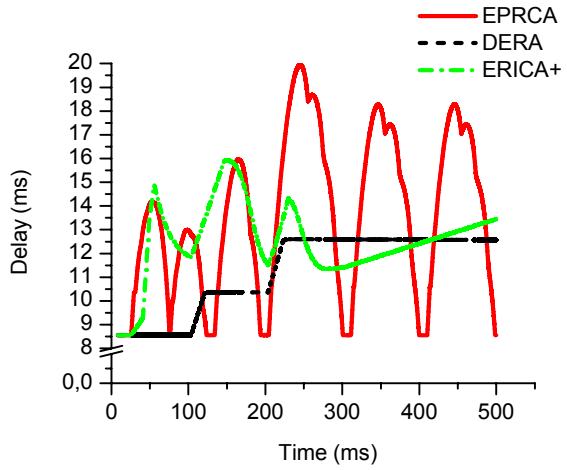

Figure 6-6.a: *Delay (ER Algorithms)*

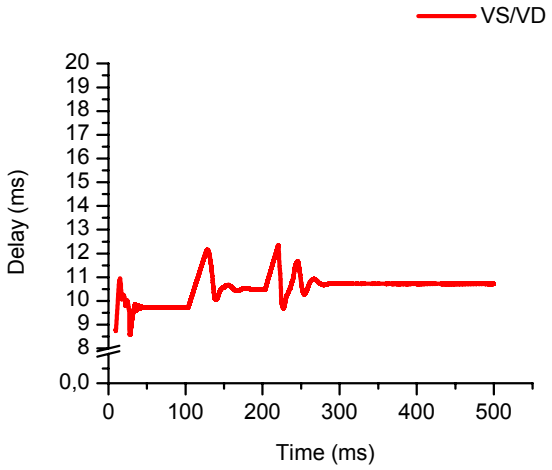

Figure 6-6.b: *Delay (VS/VD Algorithm)*

The values for the QoS parameters *Maximum Cell Transfer Delay* (maxCTD) and *peak-to-peak Cell Delay Variation* (ptp CDV), as defined by the ATM Forum for an alpha value of 1 % (see chapter 2.8.1.2), are noted in table 6-6. The dual PD-controller of the VS/VD switches is able to achieve smaller average, maximum and variance values, than all tested ER algorithms.

Flow Control Algorithm	Average Delay	Standard Deviation	maxCTD	ptp CDV
EPRCA	13.8 ms	3.2 ms	19.9 ms	11.3 ms
DERA	11.2 ms	1.6 ms	12.6 ms	4.1 ms
ERICA+	12.7 ms	1.4 ms	15.9 ms	7.4 ms
Dual PD VS/VD	10.5 ms	0.5 ms	12.1 ms	3.5 ms

Table 6-6: *Delay Values ER and VS/VD Algorithms*

The throughput values for the start-up period (from 0 to 100 ms) and the following transmission phase with congestion periods (from 100 to 500 ms) are listed in table 6-7. The VS/VD algorithm suffers from a slightly smaller throughput during the start-up phase. The reason for this is that the feedback information to raise the Initial Cell Rate (ICR) to the fair allocation needs to be propagated through five control loops, instead of being conveyed by the first RM cell returned. During the following transmission phase, the VS/VD algorithm achieves the same throughput as the two best ER algorithms (DERA and ERICA+). It has to be noted that the VS/VD switches need a smaller maximum buffer length to deliver this performance, than the ER switches, see figures 6-3.b, 6-3.c, and 6-3.d. Whereas the DERA algorithm has no explicit queue bounds, ERICA+ needs about 2500 cells of buffer space. When this resource is not available, cell loss at the ER switches will occur, having a negative impact on the through-

put. This effect could not be seen in this simulation scenario, because the ER switch buffer capacities were chosen generously (7000 cells) to determine the peak value needed.

Flow Control Algorithm	Throughput (0 - 100 ms)	Throughput (100 - 500 ms)
EPRCA	56.6 MBit/s	59.5 MBit/s
DERA	56.9 MBit/s	61.9 MBit/s
ERICA+	56.9 MBit/s	61.9 MBit/s
Dual PD VS/VD	56.1 MBit/s	61.9 MBit/s

Table 6-7: *Throughput Values ER and VS/VD Algorithms*

In order to investigate the behavior of the different ABR flow control algorithms under dynamic load conditions, the previous network scenario (shown in figure 6-2) is modified. The CBR VC 4 together with source S4 and its corresponding destination D4 are deleted and CBR connection VC 3 is changed to a VBR connection (figure 6-7). The sender S3 becomes a VBR on/off source, with equal on/off periods of 100 ms each and a transmission rate of 15 MBit/s during the active period. The source starts its transmission at 50 ms with an idle phase.

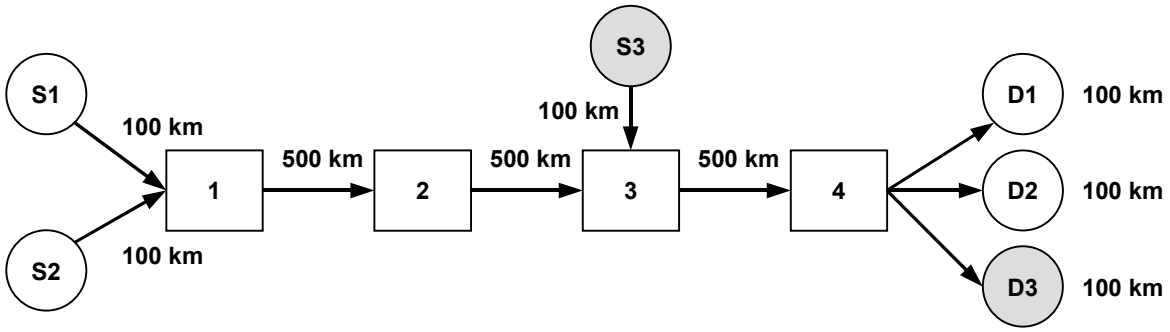

Figure 6-7: *Scenario 2: Comparison of ER and VS/VD Switches (VBR Source)*

As in the case of two CBR sources, the EPRCA leads to strong fluctuations of the ACR at the ABR source, see figure 6-9.a, which result in oscillating queue lengths, see figure 6-8.a. The DERA algorithm shows alternating queue built-ups at the two bottleneck switches, i.e., switch 1 during the off periods of VBR source S3 and switch 3 during the on periods. The resulting queue lengths at the switches get larger with growing distance to the source or feedback delay, but can be drained completely during the underload conditions (figure 6-8.b).

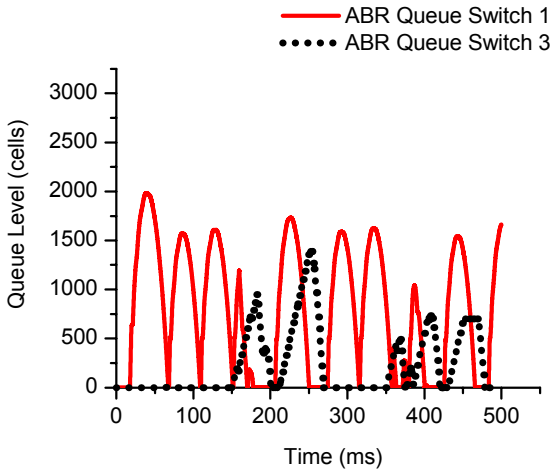

Figure 6-8.a: *ABR Class Queues EPRCA*

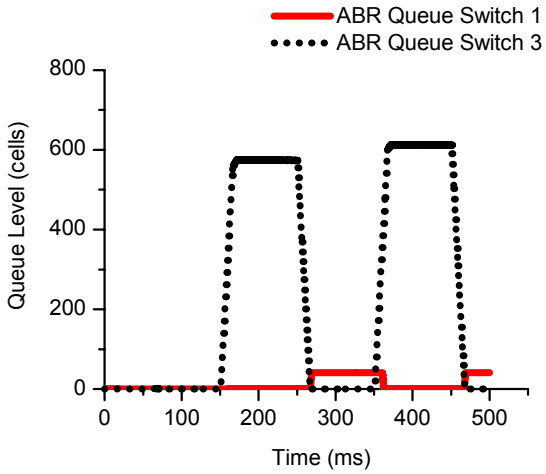

Figure 6-8.b: *ABR Class Queues DERA*

Again, the DERA algorithm is able to track the available bandwidth exactly, see figure 6-9.a, but is not able to effectively control the maximum buffer occupancy.

ERICA+ allocates fair rates including small changes to control the queue levels (figure 6-9.a). However, ERICA+ is not able to drive the ABR class queues to the desired level of about 1000 cells quickly enough during the short periods of constant traffic load, see figure 6-8.c. The newly designed PD-controller in the VS/VD switches shows a fast response and reaches the target buffer level of 300 cells for the ABR class queue at the bottleneck switch quickly (figure 6-8.d). The per-VC queues are kept at the desired length of 50 cells (figure 6-8.e). The allocated rate corresponds to the fair share plus the Minimum Cell Rate (MCR) of the connection, with some fluctuations after changes of the load condition to correct the queue lengths, see figure 6-9.b.

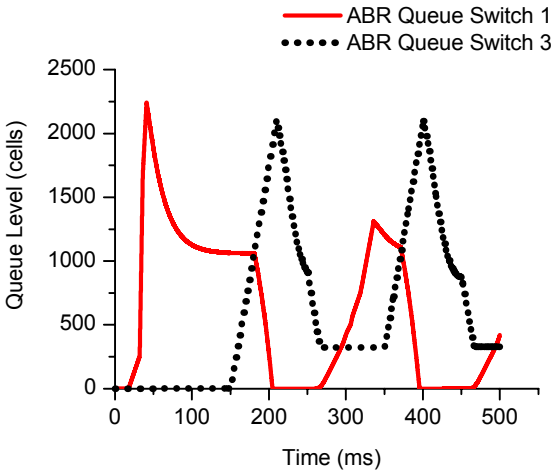

Figure 6-8.c: *ABR Class Queues ERICA+*

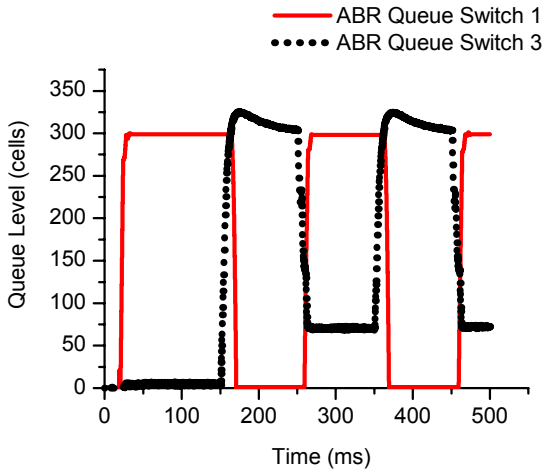

Figure 6-8.d: *ABR Class Queues VS/VD*

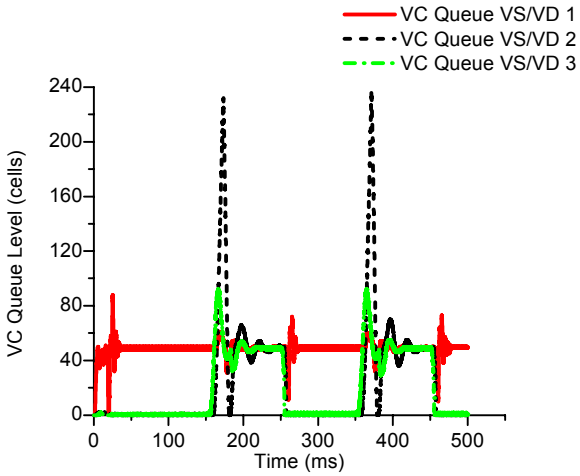

Figure 6-8.e: *Per-VC Queues VS/VD*

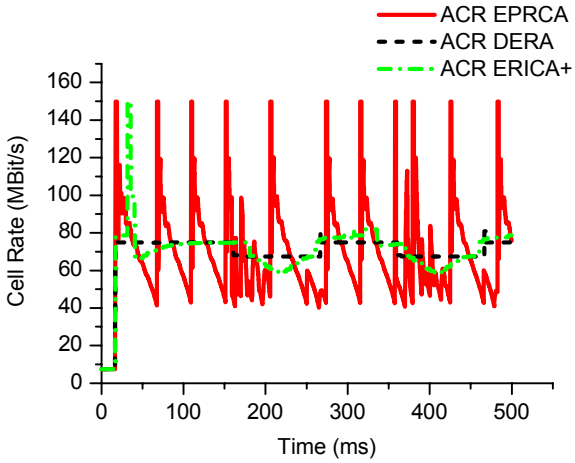

Figure 6-9.a: *Source ACRs (ER Algorithms)*

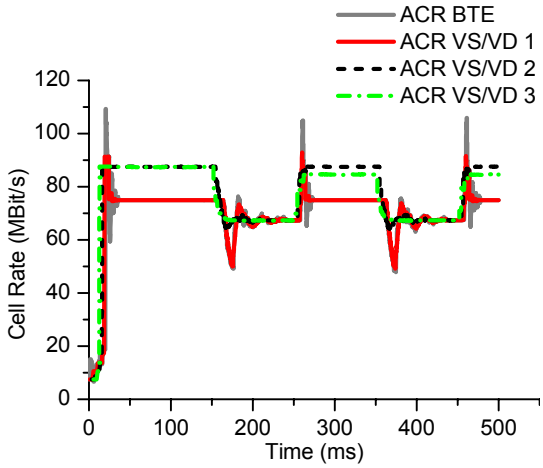

Figure 6-9.b: *Source and VS/VD ACRs*

The utilization of the link between switch 3 and 4 shows a similar result as for the previous CBR scenario, presented in figures 6-5.a and 6-5.b. The main difference is, that the DERA algorithm leads to a slight under-utilization of 90 % for a short time period after the VBR source changes from active to idle.

The end-to-end delays for the ABR connections using an ER algorithm, depicted in figure 6-10.a, reflect the temporal evolution of the switch buffer lengths along the network path. The DERA algorithm shows a very advantageous behavior, but it has to be noted that it lacks any queue control mechanism. The proposed dual PD VS/VD switch algorithm offers a steady end-to-end delay with small disturbances. Detailed statistical values are presented in table 6-8.

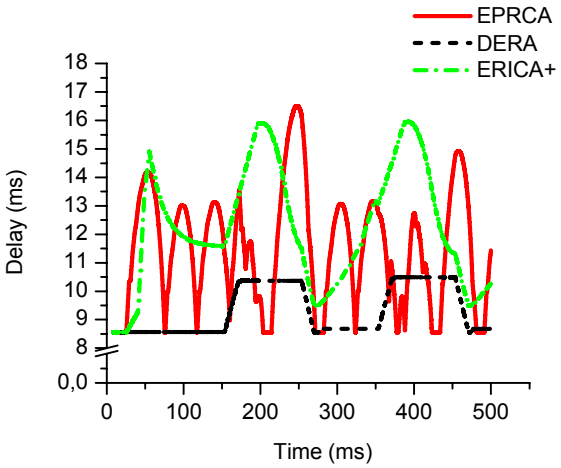

Figure 6-10.a: *Delay (ER Algorithms)*

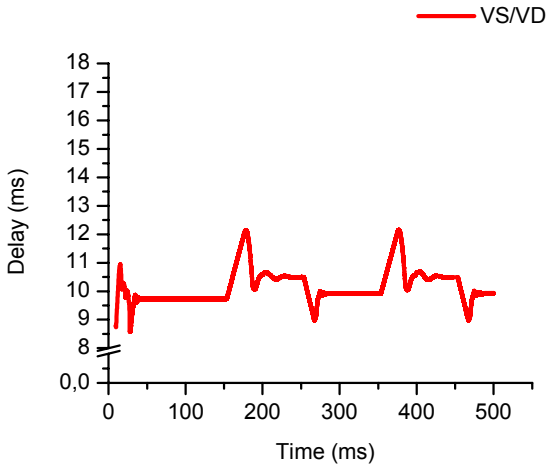

Figure 6-10.b: *Delay (VS/VD Algorithm)*

Flow Control Algorithm	Average Delay	Standard Deviation	maxCTD	ptp CDV
EPRCA	11.7 ms	2.0 ms	16.5 ms	7.9 ms
DERA	9.3 ms	0.8 ms	10.5 ms	1.9 ms
ERICA+	12.3 ms	2.0 ms	15.9 ms	7.4 ms
Dual PD VS/VD	10.1 ms	0.6 ms	12.1 ms	3.5 ms

Table 6-8: *Delay Values ER and VS/VD Algorithms*

The newly developed VS/VD switch algorithm outperforms the ERICA+ algorithm with combined queue and rate control and simple algorithms like EPRCA. DERA yields the best delay values for this scenario, but has the major disadvantage that it does not bound the buffer space needed.

The throughput values obtained for the dynamic environment are similar to the ones presented in table 6-7 for the previous CBR scenario and show the same effects of a smaller throughput during the start-up of the connection for the VS/VD algorithm.

It can be concluded that the VS/VD technique achieves stable queue lengths at the switches and leads to a steady end-to-end delay with small jitter. Due to the small buffer space needed, cell loss is unlikely, which affects the possible throughput positively. After this comparison of existing ER and the new VS/VD switch algorithm, the following section will focus on the performance analysis of the dual PD-controller.

6.4.2 Transient Response

The transient response of the dual PD-controller is determined by the temporal behavior of the controlled variable after a step change in the input variable. Whereas the PD-controller of the upstream loop controls the per-VC queue level, the PD-controller of the downstream loop controls the occupancy of the ABR class queue at the VS/VD switch. A step change of the input variable would correspond to a change in the target buffer length for the two queues. Since this effect can not be investigated under normal operation of the ABR flow control mechanism (the target values remain fixed), the VS/VD switch operation was modified for this experiment. The target ABR class queue value is changed from 300 to 400 cells at 50 ms and the per-VC queue value is raised from 50 to 100 cells at 100 ms.

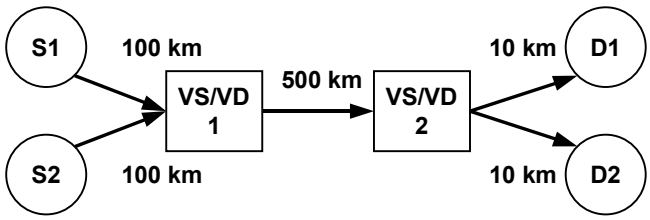

Figure 6-11: *Scenario 3: Transient Behavior (Step Change)*

The scenario illustrated in figure 6-11 is used to evaluate the temporal evolution of the queue lengths at VS/VD switch 1. Two persistent ABR sources share the bottlenecked link between VS/VD switches 1 and 2. The results obtained are analyzed with the help of the performance metrics for the transient response of a control system introduced in chapter 4.3.3.

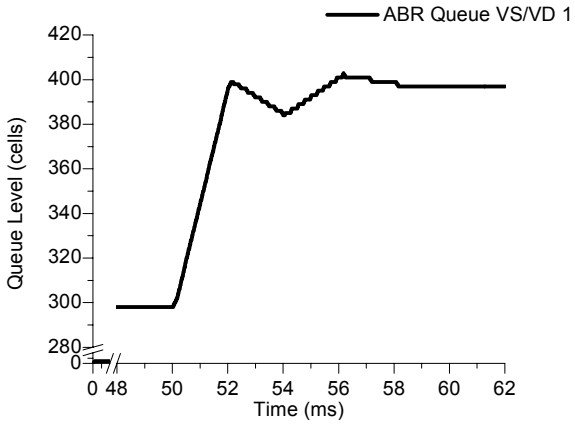

Figure 6-12.a: *ABR Class Queue Length*

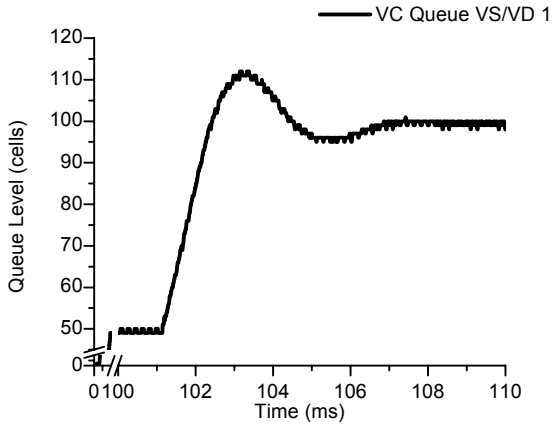

Figure 6-12.b: *Per-VC Queue Length*

Figure 6-12.a presents the buffer level of the ABR class queue after the step change at 50 ms. The peak value of 403 cells, which corresponds to an overshoot of 0.75 %, is reached after a *peak time* of $T_p = 6.175$ ms. Taking into account the action delay of the downstream control loop T_2, that equals the measurement interval of 2 ms, T_p corresponds to approximately 3 times T_2. The *rise time* T_r evaluates to 1.574 ms and the *settling time* T_s for a 2 % interval to

4.986 ms. Hence, the settling time corresponds to approximately 2.5 times the action delay T_2. It has to be noted however that the peak value of 403 cells lies already inside the measurement dead-band of 4 cells around the target value and also inside the 2 % settlement interval around the final value of 400 cells.

The evolution of the per-VC queue for the step change of 50 to 100 cells at 100 ms is shown in figure 6-12.b. After a *peak time* of T_p = 3.144 ms, the peak value of 112 cells (12 % overshoot) is reached. The rise time T_r equals 0.985 ms, and the settling time T_s for a 5 % interval evaluates to 4.073 ms. The action delay for the previous loop consists of the round trip time of 1 ms and the inter-RM cell interval of 0.18 ms (32 cell times at 74.88 MBit/ s). Therefore, the settling time of the per-VC queue corresponds to about 4 times the action delay of the previous loop.

These values confirm that the dual PD-controller meets the design objectives defined in chapter 5.3 and achieves the desired transient behavior.

Although the previous simulation scenario indicates that the newly designed VS/VD control algorithm offers a quick and swift transient response, a step change of the input variable will not occur during the normal operation of the ABR flow control algorithm. A more realistic case is the change of the error variable due to a change of the ABR capacity (external disturbance). In order to investigate this case, the previous network configuration is modified by adding a third CBR connection (S3-D3), see figure 6-13.

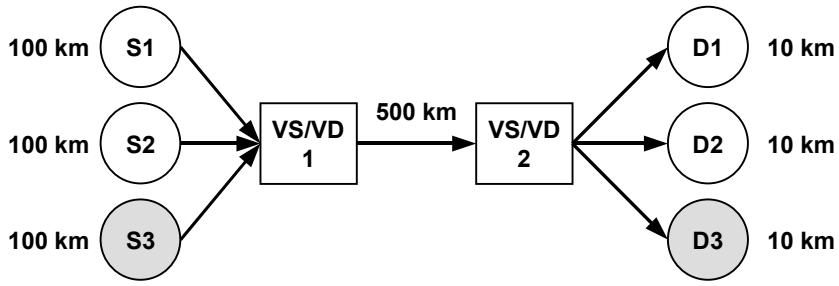

Figure 6-13: *Scenario 4: Transient Behavior (CBR Source)*

The CBR source S3 starts a 15 MBit/ s transmission at 50 ms and stops at 75 ms. Figure 6-14.a presents the buffer lengths of the per-VC queue and the ABR class queue at the bottleneck switch VS/VD 1. The corresponding Allowed Cell Rates (ACRs) for the first ABR VC at the source and at the VS of VS/VD switch 1 are depicted in figure 6-14.b. At the start of the CBR source (50 ms), the per-VC queue reaches a peak value of 79 cells (58 % overshoot) and a settling time of 10.147 ms (10 % settling interval). The ABR class queue shows a peak value of 353 cells (18 % overshoot) and a settling time of 11.101 ms (5 % interval). When the CBR source stops sending at 75 ms, the per-VC queue has a peak value of 80 cells (60 % overshoot) and a settling time of 12.982 ms for a 10 % interval around the steady state value of 50 cells. The overshoot of the ABR class queue at the end of the CBR transmission evaluates to 308 cells (2.7 % overshoot) and the settling time is 7.55 ms for a 2 % interval.

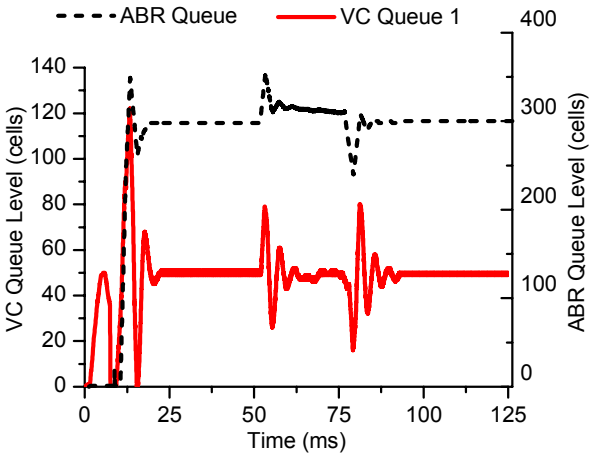

Figure 6-14.a: *VS/VD Queue Lengths*

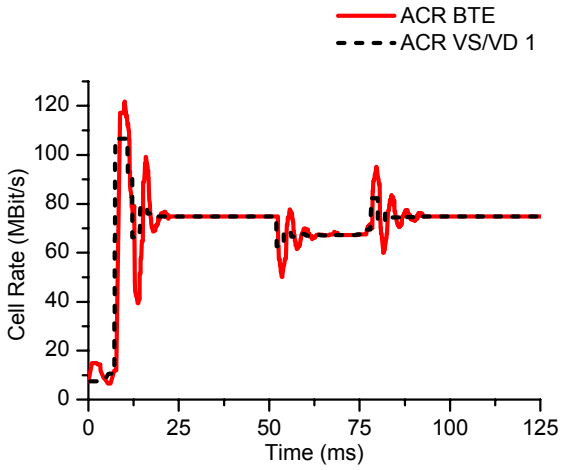

Figure 6-14.b: *Source and VS/VD ACRs*

The analysis proves that the novel dual PD VS/VD algorithm is able to control the lengths of the per-VC and ABR class queues effectively without exceeding the maximum buffer space and that it provides a fast transient response to a disturbance of the output signal, i.e., the current queue length. For both of the last two scenarios, a full utilization of the bottleneck link was achieved after a short start-up period.

6.4.3 Fairness

In order to assess the fairness property of the VS/VD switch algorithm, the following simulation scenario is investigated, see figure 6-15.

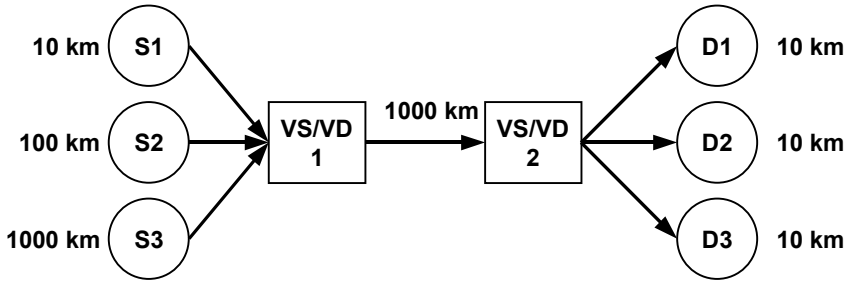

Figure 6-15: *Scenario 5: Fair Bandwidth Allocation (3 ABR Sources)*

The three ABR sources S1, S2, and S3 have different distances of 10, 100, and 1000 km to the first VS/VD switch 1 and start at time 0, 30 and 80 ms, respectively. Furthermore, the Minimal Cell Rates (MCRs) of the three ABR connections are set to 10, 20, and 30 MBit/ s, respectively.

Figures 6-16.a and 6-16.b illustrate the lengths of the ABR class and the three per-VC queues at the bottleneck VS/VD switch 1, respectively. As soon as the first two ABR connections are present after 30 ms, the corresponding two per-VC and the ABR class queues can be filled to the desired level. After the third connection (VC 3) has started its transmission, its per-VC

queue is also filled. The time to reach a steady queue length is relatively long for the third ABR connection compared to the other two, due to the long round trip delay of 10 ms.

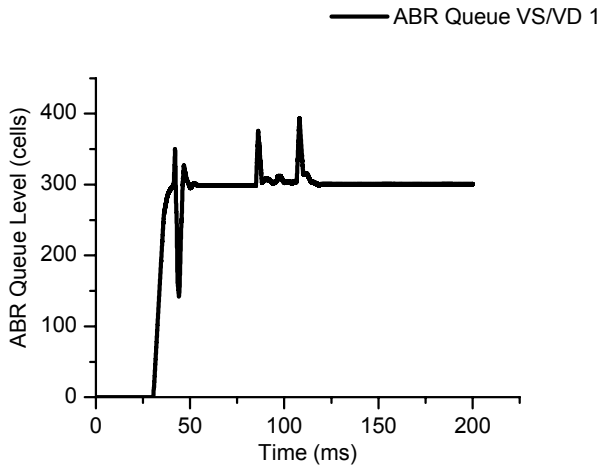

Figure 6-16.a: *VS/VD ABR Queue Length*

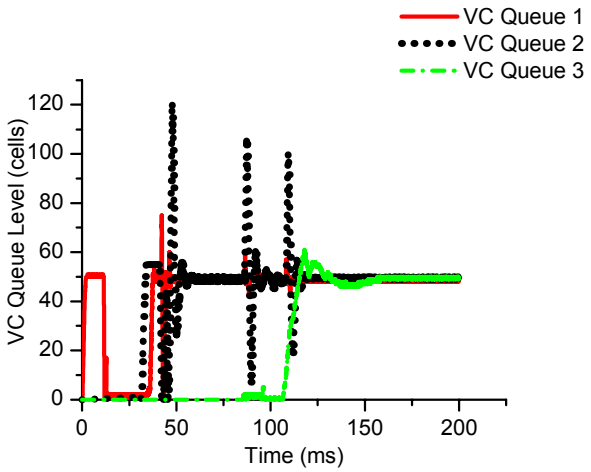

Figure 6-16.b: *VS/VD Per-VC Queue Lengths*

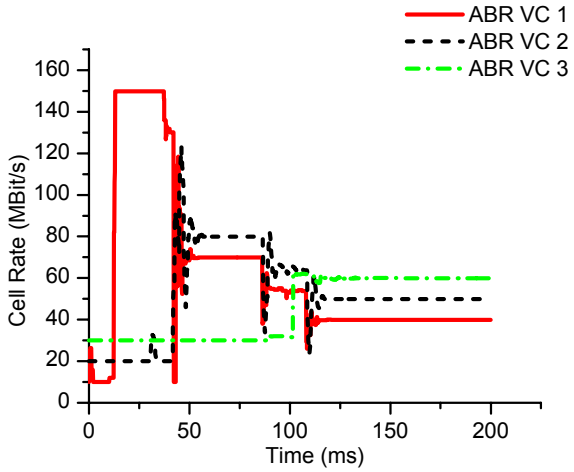

Figure 6-16.c: *Source ACRs*

The ABR source rates are presented in figure 6-16.c. During the complete simulation run, the active sources are allocated their "Equal Share plus MCR" rates. Thus, the newly developed dual PD-controller is able to guarantee the "MCR + Equal Share" fairness criterion in the presence of different MCRs, different round trip times, and a changing number of ABR sources.

The traffic management group of the ATM Forum has defined a common standard fairness test [Simc94], based on max-min fairness, see chapter 5.3. Since the "Equal Share + MCR" fairness criterion includes max-min fairness if all MCR are set to zero, this test may be applied also to the new dual PD VS/VD switch algorithm. However, an MCR of zero allows the bottleneck switch to stop the ABR source by indicating an Explicit Rate and hence an ACR of zero. The

source, which is still active, has then to probe the network condition by sending RM cells after the *Trm* time-out of 100 ms, see chapter 3.2.3. In order to avoid this effect, a small MCR of 1.49 MBit/s is defined for all sources.

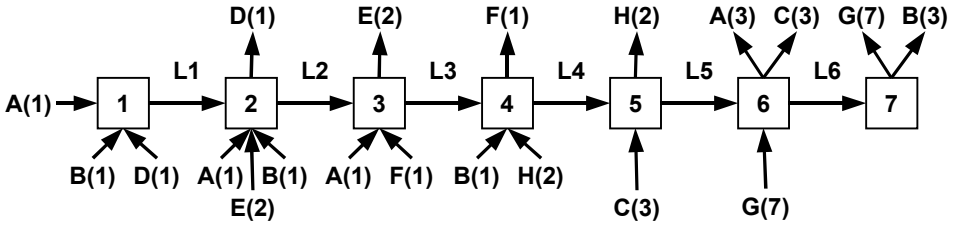

Figure 6-17: *Scenario 6: Generic Fairness Configuration 2*

The ATM Forum has specified two test cases, where the more general one, also called the *Generic Fairness Configuration 2* (GFC2) [Simc94], defines a network scenario according to figure 6-17. The configuration consists of 7 switches and 7 groups of persistent ABR connections, labeled A to G. Each group has one or more VCs, where the number of connections is indicated in brackets after the group letter. The distances of links L1 to L6 and their transmission rates are defined by the test configuration and are listed in table 6-9.

Link Name	Distance	Speed
L1	$4 \cdot D = 4000$ km	49.536 MBit/s
L2	$2 \cdot D = 2000$ km	100 MBit/s
L3	$D = 1000$ km	49.536 MBit/s
L4	$D = 1000$ km	149.76 MBit/s
L5	$D = 1000$ km	149.76 MBit/s
L6	$2 \cdot D = 2000$ km	49.536 MBit/s

Table 6-9: *Link Configuration*

The link distances are defined with respect to a standard length *D*, which is chosen to be 1000 km to address a Wide Area Network (WAN) topology. The fair bandwidth allocations (rounded to the nearest integer value) and the bottleneck link for the various groups of connections are presented in table 6-10.

Connection Group	Bottleneck Link	Fair Allocation
A	L3	11 MBit/s
B	L6	5 MBit/s
C	L5	34 MBit/s
D	L1	33 MBit/s
E	L2	34 MBit/s
F	L3	11 MBit/s
G	L6	5 MBit/s
H	L4	51 MBit/s

Table 6-10: *Fair Bandwidth Allocation*

All ABR end systems, except for the ones belonging to the two group "H" connections, are attached to the VS/VD switches by 10 km links with a capacity of 49.536 MBit/s. The ABR end systems for connections H1 and H2 are connected by 149.76 MBit/s lines, because of their fair share of 51 MBit/s.

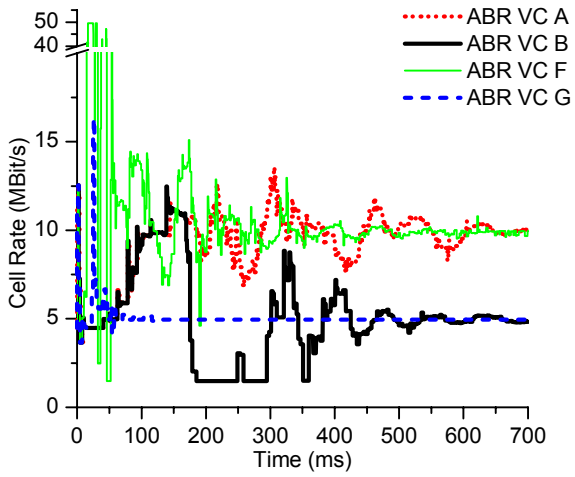

Figure 6-18.a: *ACR of Sources A, B, F, G*

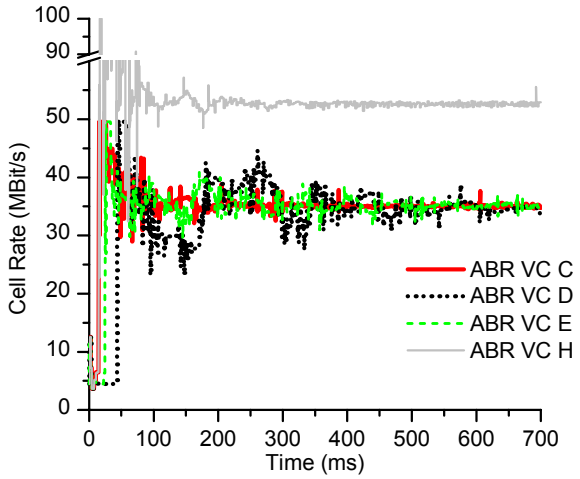

Figure 6-18.b: *ACR of Sources C, D, E, H*

Figures 6-18.a and 6-18.b depict the ACR at the source for one ABR connection of each group. If a group consists of several connections with different VC lengths, the source with the longest distance to the destination was chosen. The dual PD-controller allocates the fair rates listed in table 6-10 to the different ABR sources, passing the generic fairness test of the ATM Forum. It can be seen that the ACRs exhibit small rate fluctuations around the fair allocation. The PD-

controller changes the advertized rates to compensate for mismatches of the queue lengths at the VS/VD switches. These rate changes affect the rate allocations of other groups, which in turn influences the queue lengths. Thus, this generic fairness scenario defines a network configuration where an optimal queue length and rate allocation can not be obtained at the same time.

It can be concluded that the newly developed VS/VD switch algorithm fulfills the "MCR plus Equal Share" fairness criterion, chosen during the design process.

6.4.4 Background Traffic

In section 6.4.1, it was demonstrated that the proposed dual PD-controller is able to track large changes of ABR capacity. During the design phase of the novel algorithm, it was stated that the ABR class queue should keep a reservoir of cells to fill in short bandwidth gaps. Furthermore, an exponential averaging was used for the measurement of the high priority traffic, to filter high frequency noise. In order to test the behavior of the new VS/VD control algorithm in the presence of highly variant VBR traffic, the following scenario (figure 6-19) is used. There are three groups of ABR connections, containing two VCs each. The sources of the three groups have a distance to the first VS/VD switch of 10, 100, and 1000 km, respectively. In addition to the ABR VCs, the common link between VS/VD switch 1 and 2 is also shared by 10 VBR connections carrying real-time MPEG-2 traffic as defined in section 6.3.2.6.

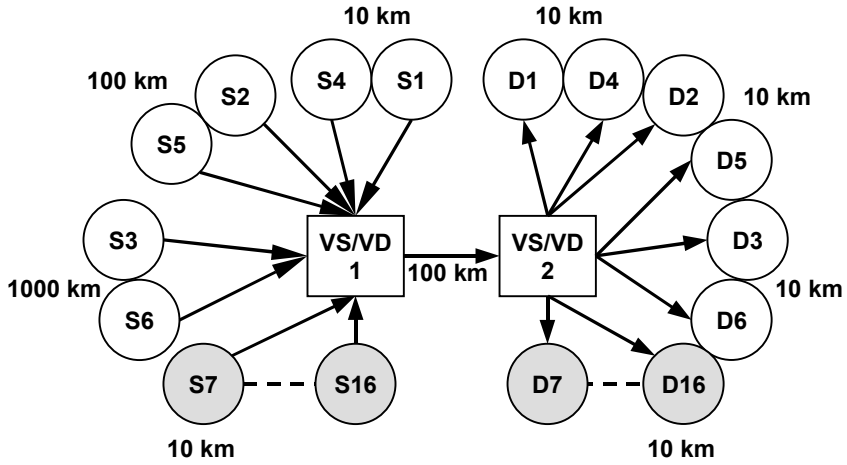

Figure 6-19: *Scenario 7: VBR Background Traffic (Real-Time MPEG-2)*

The 10 VBR real-time sources, pairs of which are attached to one associated BTE, have access links of 10 km length and start their transmission after an interval of 40 ms, in order to avoid synchronization between sources. Thus, all 10 sources are sending after 400 ms. The following figures present results for ABR VCs labeled 1, 2, and 3, where each belongs to one of the three groups with a distance of 10, 100, and 1000 km, respectively.

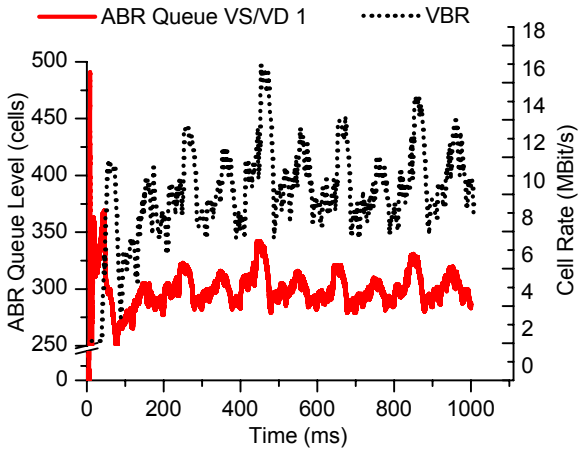

Figure 6-20.a: *VS/VD ABR Queue Length (Dual PD-Controller)*

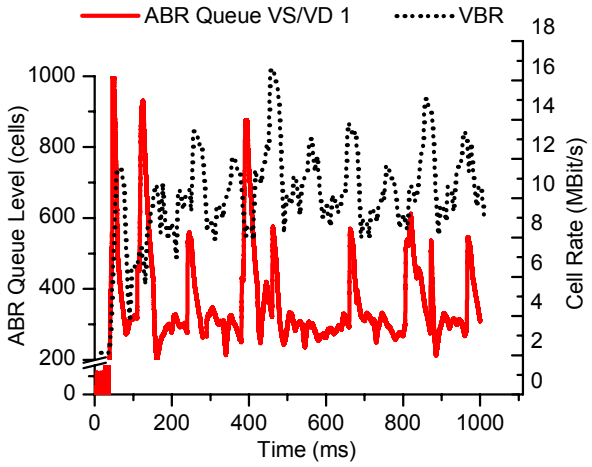

Figure 6-20.b: *VS/VD ABR Queue Length (ERICA+ for VS/VD)*

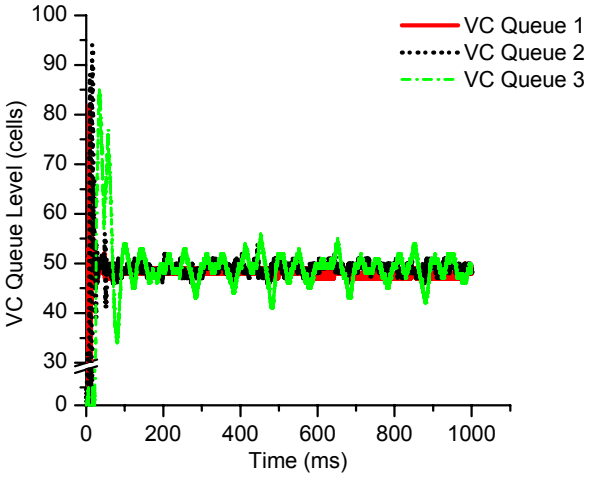

Figure 6-21.a: *VS/VD Per-VC Queue Lengths (Dual PD-Controller)*

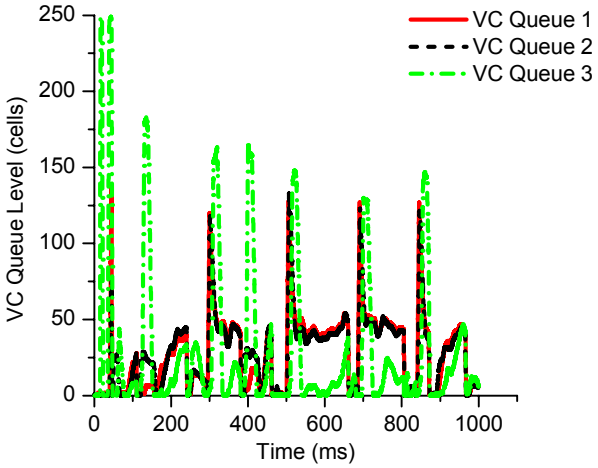

Figure 6-21.b: *VS/VD Per-VC Queue Lengths (ERICA+ for VS/VD)*

Figures 6-20.a and 6-21.a present the lengths of the ABR class queue together with the VBR traffic load and the level of per-VC queues for the dual PD-controller at the bottleneck switch VS/VD 1, respectively. The per-VC queues are kept at the desired operating point of 50 cells. The long-haul ABR VC 3 shows minor oscillations, due to the long propagation delay of 10 ms. The ABR class queue exhibits same fluctuations around the target value of 300 cells, because ABR cells are used to fill VBR bandwidth gaps. The utilization of the bottleneck link is therefore kept constantly at 100 %, after a short start-up period. The ACRs of the ABR sources and the VBR traffic load at VS/VD switch 1 are depicted in figure 6-22.a. Because of the fact, that the ABR class queue is able to compensate the disturbances of the ABR capacity caused by the VBR background traffic, the source rates need only minor corrections and show a steady behavior.

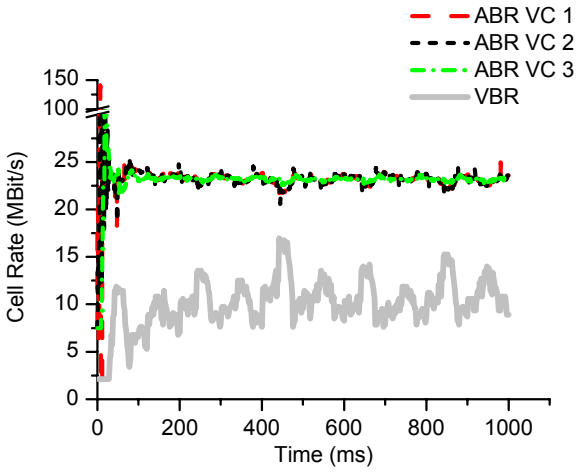

Figure 6-22.a: *Source ACRs and VBR Load (Dual PD-Controller)*

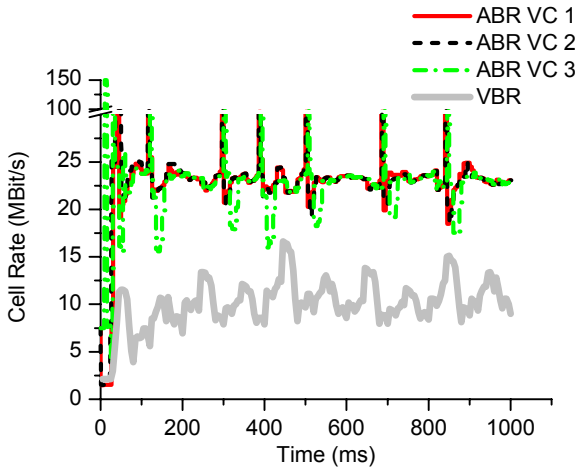

Figure 6-22.b: *Source ACRs and VBR Load (ERICA+ for VS/VD)*

In order to compare the performance of the newly designed dual PD-controller with existing VS/VD algorithms, ERICA+ for VS/VD is used, see chapter 5.1.2. The algorithm parameters listed in table 5-3 have been chosen according to scenario 7 to obtain similar results, as for the dual PD-controller algorithm. The desired delay values of 0.85 ms for the ABR class queue and 0.9 ms for the per-VC queue lead to a target buffer occupancy of approximately 300 and 50 cells, respectively. For this, an ABR capacity of 150 MBit/s and an average cell rate of 23 MBit/s per ABR VC is assumed. The slope parameters *a* and *b* for the two queue control functions have been tuned based on the default values by trial and error using simulation experiments. The suggested standard values led to strong oscillations of the queue lengths [BaCs00].

Figures 6-21.b and 6-22.b illustrate the ABR class and the per-VC queue lengths for the simulation scenario 7. Both queues demonstrate large oscillations around the target value and are heavily disturbed by the VBR background traffic. During the start-up phase at the beginning of the simulation, both queues have large overshoots, exceeding the maximum buffer occupancy and hence resulting in cell loss. The negative effect of the VBR traffic can be felt even at the ABR sources, because the ACRs have to be altered to correct the VS/VD switch queue levels, see figure 6-22.b. As an example for the resulting end-to-end delays, ABR VC 1 with a distance of 10 km of the source to the first VS/VD switch is chosen. The delay values for the dual PD-controller and ERICA+ for VS/VD for VC 1 are shown in figures 6-23.a and 6-23.b, respectively. The frequent changes in queue length at the VS/VD switch can be observed in both curves, but the ERICA+ for VS/VD leads to higher variance, due to large peaks in the buffer occupancy.

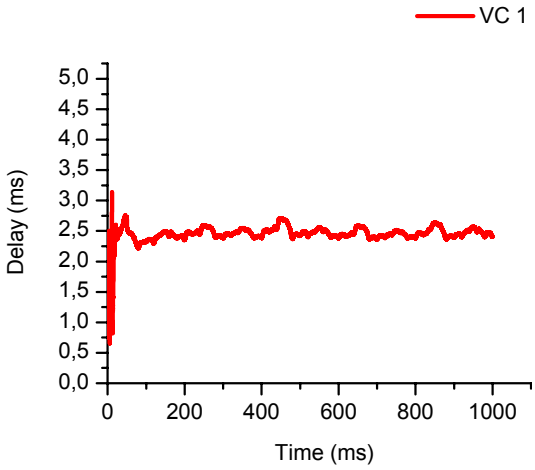

Figure 6-23.a: *End-to-End Delay (Dual PD-Controller)*

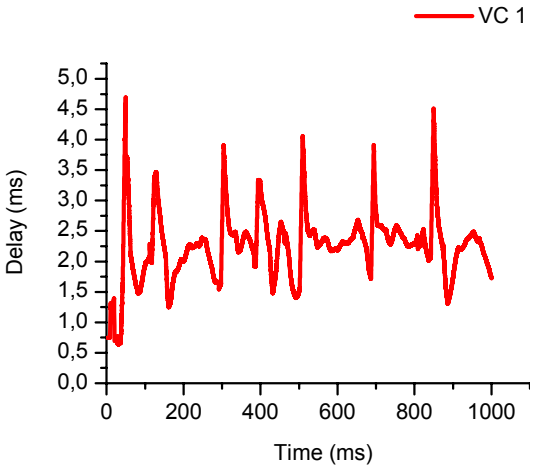

Figure 6-23.b: *End-to-End Delay (ERICA+ for VS/VD)*

Since the ABR class queue is not drained completely, both algorithms are able to reach a utilization of 100 % for the bottleneck link. As a consequence, the throughput for the dual PD-controller is better than for ERICA+ for VS/VD only during the start-up phase of the connections, because overflowing queues lead to cell loss for the latter algorithm. After that, the obtained throughput for VC 1 is nearly the same, with a slightly higher value for the adapted ERICA+ algorithm (see table 6-11).

Flow Control Algorithm	Throughput (0 - 400 ms)	Throughput (400 - 1000 ms)
ERICA+ for VS/VD	21.82 MBit/ s	23.33 MBit/ s
Dual PD VS/VD	23.69 MBit/ s	23.11 MBit/ s

Table 6-11: *Throughput Values ER and VS/VD Algorithms*

In the following scenario 8, the previous configuration (figure 6-19) is modified by replacing the 10 VBR sources by non real-time VBR sources, generating self-similar MPEG-2 video traffic as described in section 6.3.2.7. As for scenario 7, results for three ABR VCs are presented, labeled 1, 2, and 3, where each belongs to one of the three groups with a distance of 10, 100, and 1000 km from the source to VS/VD switch 1, respectively.

The results for non real-time, MPEG-2 video traffic are similar to those of the previous scenario. Figure 6-24.a shows the ABR class queue for the dual PD-controller together with the VBR traffic load at VS/VD switch 1. The length of the per-VC queues for the three ABR connections using the novel ABR algorithm at VS/VD switch 1 are depicted in figure 6-24.b. For the dual PD-controller, the buffer lengths of both the ABR class and the per-VC queues are close to the desired values (figures 6-24.a and 6-25.a). Again, the ABR class queue holds a res-

ervoir of cells, that are used to exploit small bandwidth capacities. The ABR source rates using the dual PD-controller remain steady, even though the VBR background load perturbs the ABR bandwidth, see figure 6-26.a.

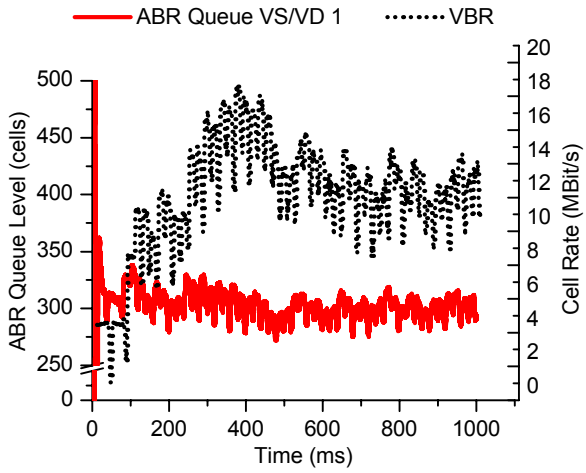

Figure 6-24.a: *VS/VD ABR Queue Length (Dual PD-Controller)*

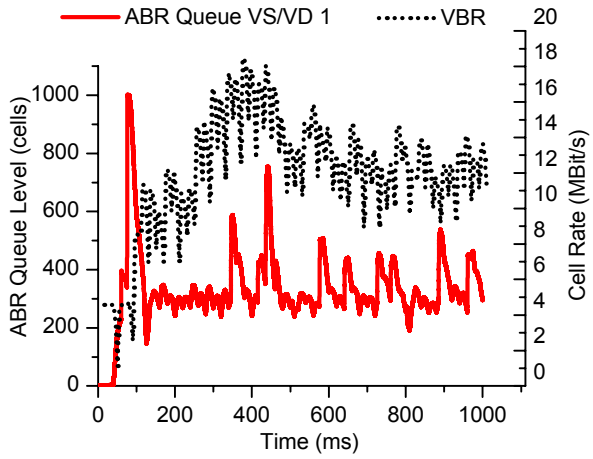

Figure 6-24.b: *VS/VD ABR Queue Length (ERICA+ for VS/VD)*

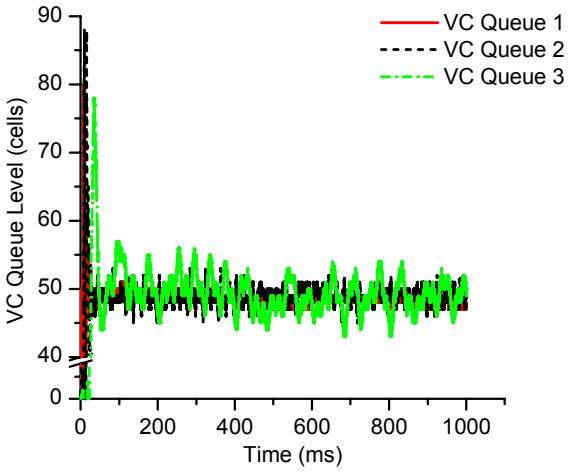

Figure 6-25.a: *VS/VD Per-VC Queue Lengths (Dual PD-Controller)*

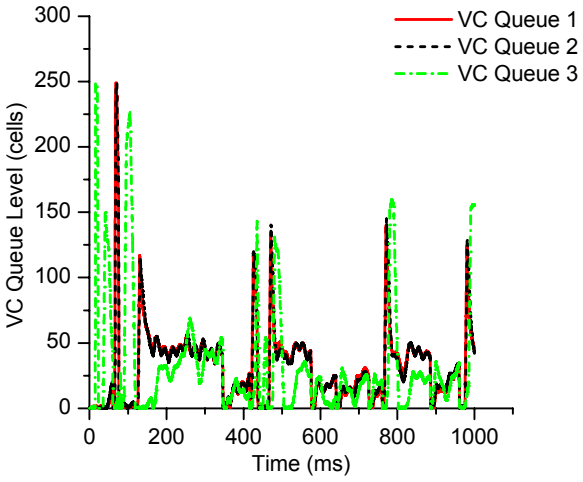

Figure 6-25.b: *VS/VD Per-VC Queue Lengths (ERICA+ for VS/VD)*

The ERICA+ for VS/VD algorithm is not able to retain a steady buffer level for the ABR class or the per-VC queues, as shown in figures 6-24.b and 6-25.b, respectively. The ACRs at the ABR sources suffer from rate changes, due to the large queue oscillations, see figure 6-26.b.

Both algorithms are able to fully utilize the bottleneck link capacity for the whole simulation time, except for a short initialization phase.

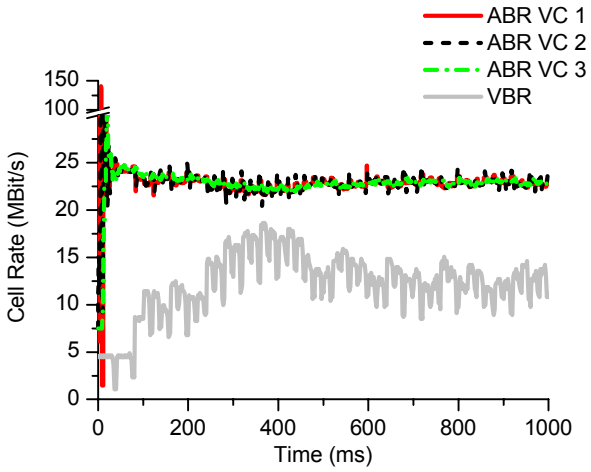

Figure 6-26.a: *Source ACRs and VBR Load (Dual PD-Controller)*

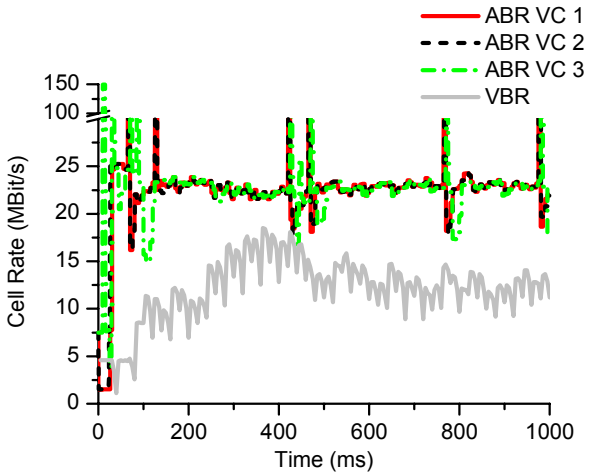

Figure 6-26.b: *Source ACRs and VBR Load (ERICA+ for VS/VD)*

As for the previous scenario, the end-to-end delays for ABR VC 1 (the source has a distance of 10 km to the first VS/VD switch) for both VS/VD algorithms are juxtaposed in figures 6-27.a and 6-27.b. Again, the ERICA+ for VS/VD algorithm results in large oscillations of the delay values and hence in large jitter.

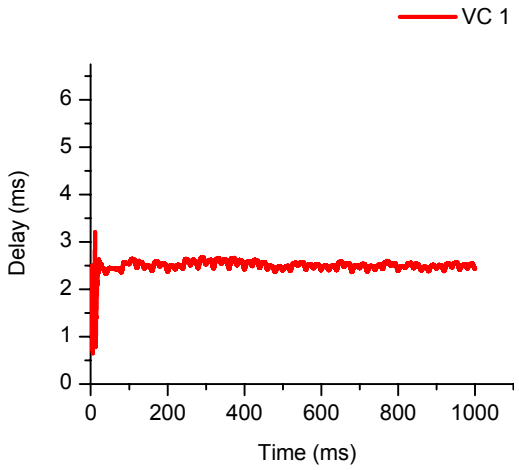

Figure 6-27.a: *End-to-End Delay (Dual PD-Controller)*

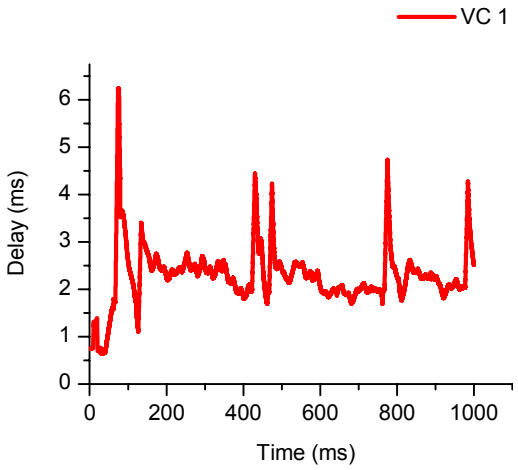

Figure 6-27.b: *End-to-End Delay (ERICA+ for VS/VD)*

The throughput values for ABR VC 1 for the two VS/VD switch algorithms are presented in table 6-12. As in the previous scenario, the cell loss caused by buffer overflows at the first VS/VD switch leads to a reduced throughput for the ERICA+ for VS/VD algorithm. Since both algorithms are able to fully utilize the bottleneck link after the start-up phase, the obtained throughput for the remaining time period is almost identical.

Flow Control Algorithm	Throughput (0 - 400 ms)	Throughput (400 - 1000 ms)
ERICA+ for VS/VD	21.32 MBit/s	22.86 MBit/s
Dual PD VS/VD	23.23 MBit/s	22.82 MBit/s

Table 6-12: *Throughput Values ER and VS/VD Algorithms*

It can be concluded that the novel VS/VD algorithm shows a better performance with regard to steady switch queue lengths, source rates, and delay values, than the ERICA+ for VS/VD algorithm. The results are confirmed for both real-time, short range dependent, and non real-time, self-similar MPEG-2 video background traffic. The proposed dual PD-controller is able to effectively compensate bandwidth fluctuations introduced by highly variant VBR background traffic and therefore provides a good disturbance rejection.

6.4.5 Non-persistent ABR Sources

In this section, the evaluation of the new dual PD-controller for VS/VD switches with respect to non-persistent ABR sources is presented. In contrast to greedy ABR sources, the ACR of the BTE component may be different from the source rate generated by the ABR traffic model. As a consequence, the utilization of the bottleneck link may drop below 100 % if the source rate is lower, or the buffer length at the BTE may grow if the source rate is larger than the ACR.

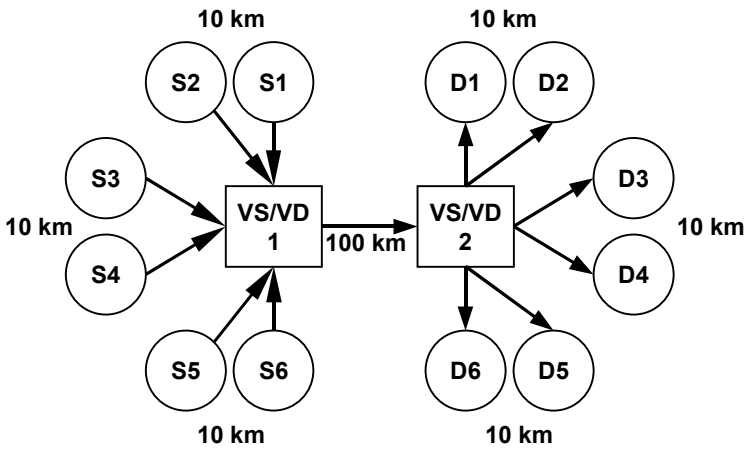

Figure 6-28: *Scenario 9: ABR On/Off Sources*

Figure 6-28 depicts scenario 9, where 6 ABR on/off sources, as described in section 6.3.2.2, share a common bottleneck link of 100 km. All 6 sources have exponential distributed on/off periods of 20 ms and a transmission rate of 50 MBit/s during their active periods. Since the 6 multiplexed sources are active for 50 % of the time on average, the bottleneck link capacity of 149.76 MBit/s is fully exploited.

The transmission rate of one of the 6 ABR on/off sources is presented in figure 6-29.a. The random durations of the successive on and off periods can be clearly observed. Figure 6-29.b depicts the corresponding ACRs of the BTE component and of the downstream loop of the VS/VD switch 1. The cell rate oscillates around the fair rate of 25 MBit/s in order to correct the deviations of the switch queue lengths due to the random source rates, see figure 6-29.c.

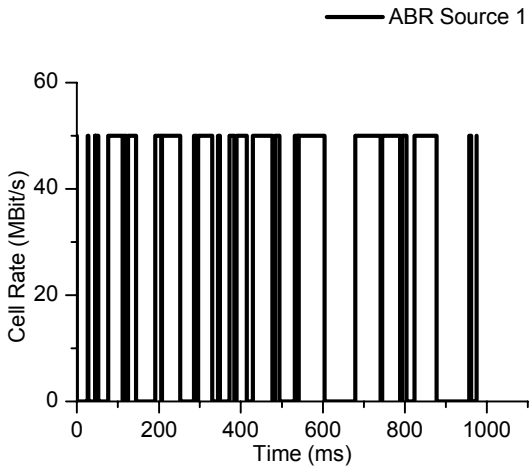

Figure 6-29.a: *Rate of ABR On/Off Source*

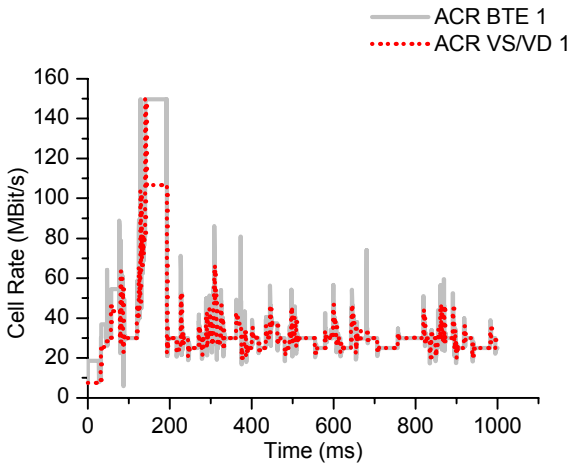

Figure 6-29.b: *ACR of BTE and VS/VD 1*

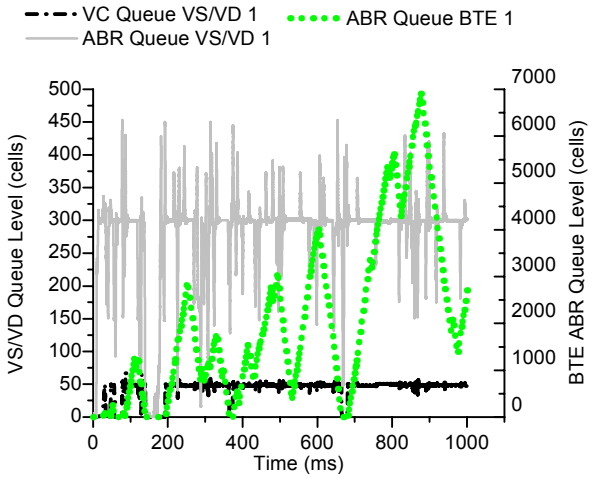

Figure 6-29.c: *VS/VD and BTE Queues*

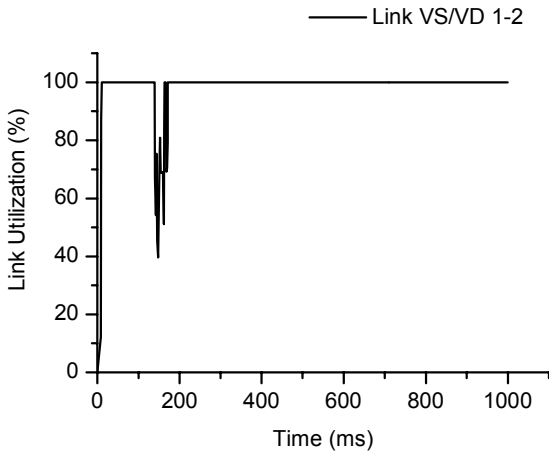

Figure 6-29.d: *Link Utilization*

The per-VC queue of the sample ABR connection at the VS/VD switch is kept closely at the desired level with small peaks. The ABR class queue shows larger oscillations around the target value of 300 cells, which is caused by the multiplexing of the random sources (figure 6-29.c). Also depicted in this figure is the BTE buffer length, which builds up whenever cells have to be queued due to a surplus of source generated traffic. The peak value reaches 7000 cells, which corresponds to a buffer space of about 400 kbyte that has to be provided by the BTE.

Figure 6-29.d illustrates the utilization of the bottleneck link between VS/VD switches 1 and 2. The utilization is kept at 100 % nearly all the time, except for short drops, which is a conse-

quence of the large degree of multiplexing. In the configuration above, the sum of the effective bandwidths of all 6 ABR on/off sources equals the link capacity.

Whereas the previous scenario uses a classical Poisson model for the ABR source traffic, the following experiment analyzes the behavior of the novel VS/VD switch algorithm in the presence of aggregated data traffic. For this, the self-similar ABR traffic model, introduced in section 6.3.2.3, is deployed in scenario 10, depicted in figure 6-30. Two self-similar ABR sources with an average cell rate of 70 MBit/s each share a common link of 100 km between VS/VD switches 1 and 2. Because the generated traffic of the single ABR source fluctuates around the average cell rate, a spare link capacity of 10 MBit/s remains to drain transient queues.

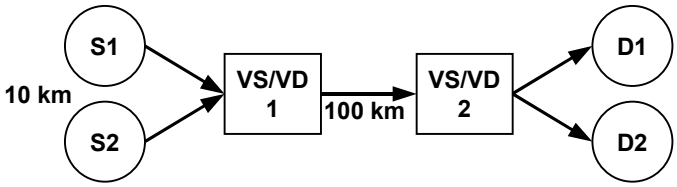

Figure 6-30: *Scenario 10: Self-Similar ABR Sources*

Figure 6-31.a illustrates the source rates for the two ABR self similar sources, which are derived from a sample path of fractional Gaussian noise, see section 6.3.2.3. The picture shows a noisy distribution of the source rates around the chosen average value of 70 MBit/s. The corresponding ACRs of the BTE component and of VS/VD switch 1 are presented in figure 6-31.b. The ACR at the VS/VD is often raised above the fair share of 75 MBit/s up to a value of approximately 110 MBit/s, in order to fill the ABR class queue with the help of the two ABR VCs. The ACR at the BTE reaches the Peak Cell Rate (PCR) of 149.76 MBit/s most of the time, because the PD-controller of the upstream loop at VS/VD switch 1 tries to keep the per-VC queue at the target level, despite the low average occupancy of the bottleneck link of 140 MBit/s.

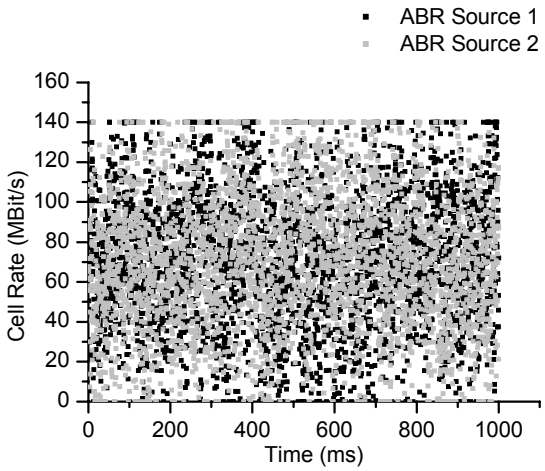

Figure 6-31.a: *Rates of ABR Self-Similar Sources*

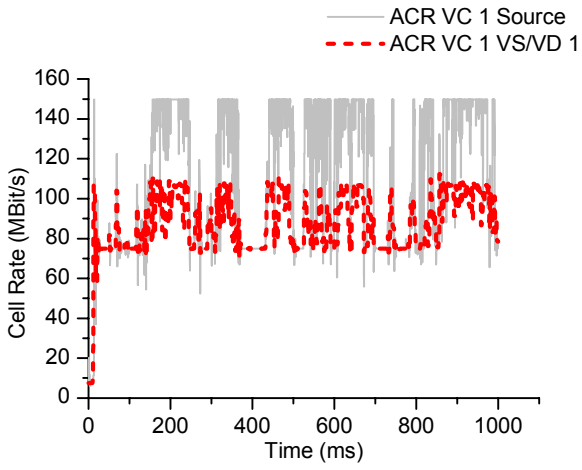

Figure 6-31.b: *ACR of BTE and VS/VD 1*

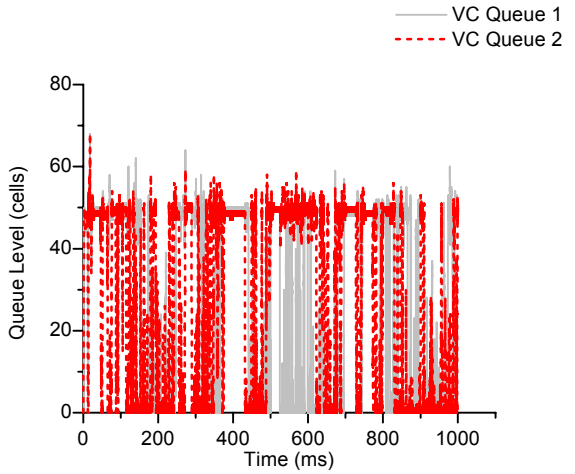

Figure 6-32.a: *Per-VC Queue Lengths at VS/VD 1*

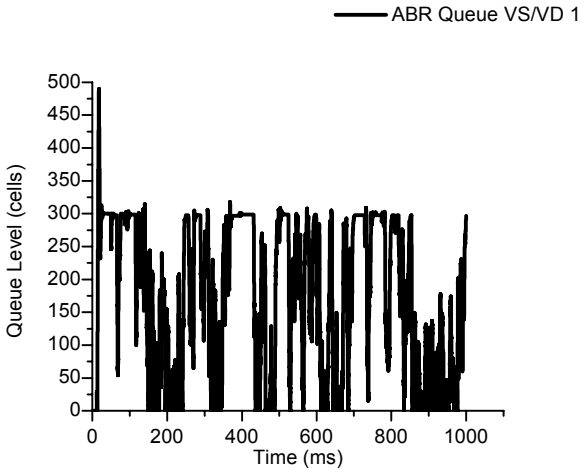

Figure 6-32.b: *ABR Class Queue Length at VS/VD 1*

This effect can also be observed at the per-VC queues and the ABR class queue of VS/VD switch 1, presented in figures 6-32.a and 6-32.b, respectively. Due to the non-persistent ABR sources and their average transmission rates below the fair share, both queue levels reach their target value only partly during the simulation time. Nevertheless, there exist times, where the generated source rates exceed the ACRs at the BTE and corresponding queues build-up at the BTE component, see figure 6-33.a. Figure 6-33.b demonstrates, that although the sources will not fully use the bottleneck link capacity on average, the buffering at the ABR class queue allows a high link utilization of 100 % for longer time periods.

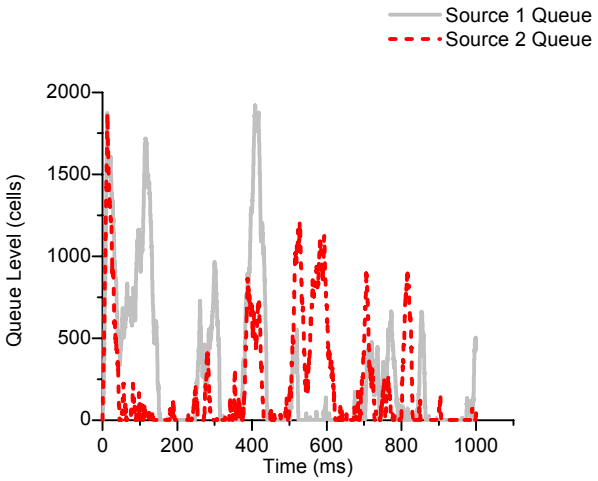

Figure 6-33.a: *BTE Queue*

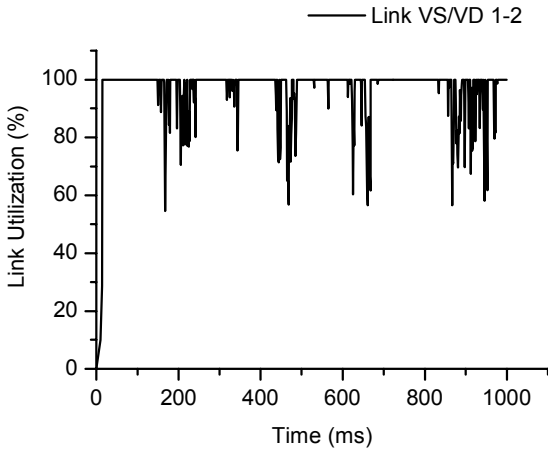

Figure 6-33.b: *Link Utilization*

It can be summarized that the developed VS/VD switch algorithm is able to control non-persistent ABR sources that carry multiplexed or aggregated LAN/WAN traffic. The queue levels and the utilization of the bottleneck link are kept close to their target levels, with temporal devia-

tions due to the random source rates. For this type of ABR traffic, the source end system (or the BTE component in the simulation model) has to provide buffer space in case the source rate exceeds the current ACR of the ABR VC.

6.4.6 VS/VD and Explicit Rate Mixed Environment

As mentioned in section 6.1, the new VS/VD algorithm developed in this thesis is fully compliant to the ATM Forum Traffic Management Specification and implements the standard ABR end system rules and ABR flow control mechanism [TM4.1]. In order to test the performance of the novel dual PD-controller VS/VD switch algorithm in combination with other Explicit Rate (ER) algorithms, scenario 1 (illustrated in figure 6-2) is adapted. A mixed VS/VD and ER configuration is defined by assigning the following flow control algorithms to switches 1 to 4: dual PD-controller for VS/VD, EPRCA, DERA, and ERICA+, see figure 6-34.

As in scenario 1, there are 2 persistent ABR sources starting at 0 ms, which are passing through the four switches. Two CBR sources are attached to the third switch and their connections share the link between the third and fourth switch with the two ABR VCs. CBR source S3 starts a transmission of 15 MBit/s at 100 ms, whereas source S4 starts with the same rate at 200 ms.

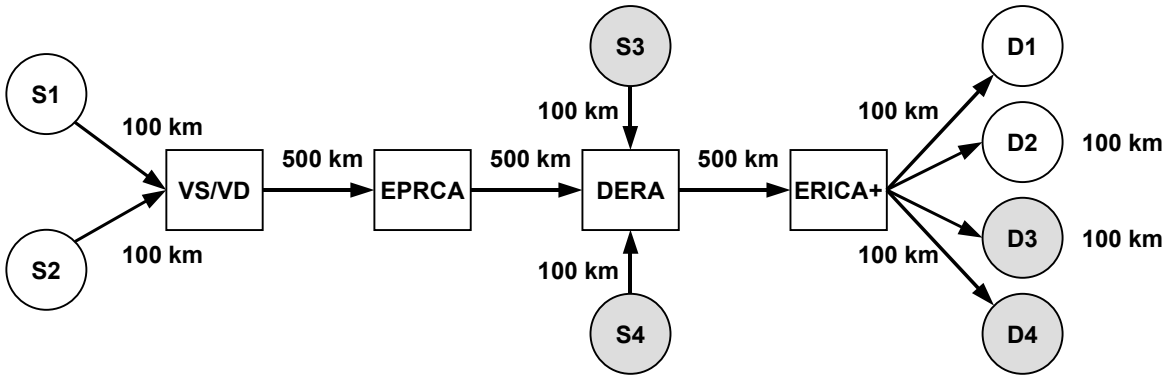

Figure 6-34: *Scenario 10: Mixed VS/VD and ER Configuration*

Since the two ABR VCs are symmetric, only the results for VC 1 are presented. Figure 6-35.a illustrates the per-VC queue of ABR VC 1 at the VS/VD switch - the other ER switches have an ABR class queue only. In the same figure, the ABR class queue at switch 3, operating with the DERA algorithm, is shown. All other switches have an empty ABR class queue, which is not depicted in figure 6-35.a. The VS/VD switch is not able to fill the ABR class queue to the desired level, because the PD-controller feeding the ABR class queue is bottlenecked downstream. The DERA switch determines the fair rate for each ABR connection, which then limits the ACR for the next loop at the VS/VD switch. The PD-controller for the previous loop however is allowed to raise its rate indicated to the BTE, in order to fill the per-VC queue. These rate adaptations can be seen in figure 6-35.b, the resulting per-VC queue level in figure 6-35.a. The ABR class queue length at the third switch raises at the two time instants where the starting CBR sources generate an overload condition. This queue is not drained, because the DERA algorithm offers no queue control feature.

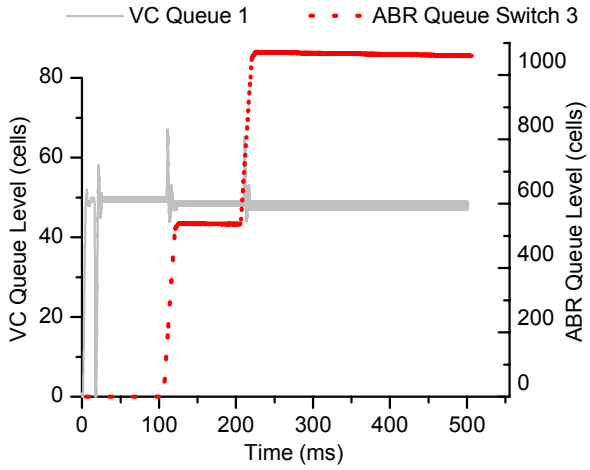

Figure 6-35.a: *Per-VC and ABR Class Queue Lengths*

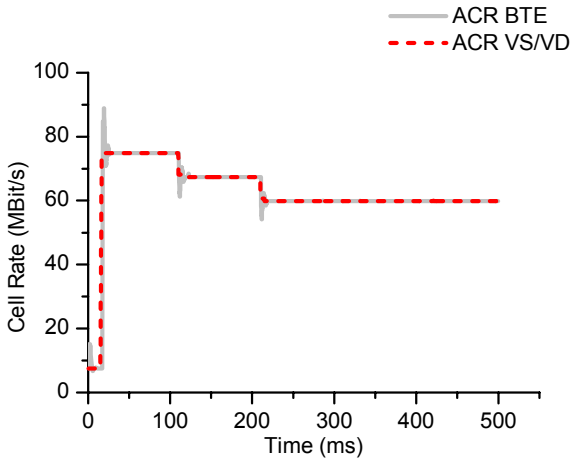

Figure 6-35.b: *ACR (BTE and VS/VD Switch)*

The simulation results for the mixed ER and VS/VD environment prove that the new dual PD-controller behaves as expected, inter-operating with the standard ER algorithms. Under such conditions, the performance of the VS/VD algorithm depends on the restrictions imposed by other coexisting flow control algorithms.

6.5 Summary

In this chapter the assessment of the design objectives defined in section 5.3 and the performance evaluation of the VS/VD switch algorithm, developed in the previous chapter, were presented.

The use of classical control theory allowed to derive analytical results, which prove that the design objectives with respect to complexity, scalability, conformity, stability, robustness, and steady state error could be fulfilled. As it was not possible to transform the VS/VD control model designed in the frequency domain with the help of Laplace and *z*-Transforms back to the time domain, a simulation model was used for the subsequent analysis. The remaining performance goals have been analyzed with the help of discrete event simulations, in order to investigate the temporal behavior of the performance metrics under various conditions.

First, the different operations of the new VS/VD and Explicit Rate (ER) algorithms were demonstrated. It has been shown that some of the existing ER algorithms target only the performance goals of fair bandwidth allocation or simple implementation. Even the most advanced ERICA+ algorithm incorporating queue and rate control performs only moderate with respect to buffer requirements and jitter. The newly designed VS/VD algorithm leads to steady queue lengths and small jitter, especially in the presence of changing ABR capacities.

The transient behavior of the dual PD-controller was assessed by a step change of the controller input signals, i.e., the target queue lengths. It was demonstrated that the new VS/VD switch algorithm meets the requirements for rise and settlement times defined during the design process. The predicted zero steady state system error derived from analytical analysis was confirmed by the simulation experiment. In addition, the more realistic step change of ABR capacity due to high priority CBR traffic was investigated. The dual PD-controller demonstrated a swift response even in the presence of a changing number of ABR connections and different propagation delays.

The fairness criterion "Minimum Cell Rate (MCR) + Equal Share" (see chapter 5.5.2.1) was successfully tested for a changing number of active ABR connections with different round trip times and MCRs. Furthermore, the novel VS/VD switch algorithm passed the generic fairness test scenario defined by the ATM Forum.

During these experiments, the dual PD-controller algorithm already showed an efficient operation, in the sense that the desired queue levels and a 100 % utilization of the bottleneck link were achieved. In the following, the most challenging environment for ABR flow control including highly variant VBR background traffic was investigated. For this, both real time and non real-time, self-similar MPEG-2 video traffic models were used. The VS/VD flow control algorithm developed in this thesis achieved a good disturbance rejection, with a full utilization of the bottleneck link and small queue length variations. In contrast, the existing ERICA+ for VS/VD algorithm showed strong queue oscillations with a large end-to-end jitter, despite a tuning of its controller parameters.

Since the ABR service category is suited for bursty applications and aggregated LAN/WAN traffic, non-persistent ABR source models were also deployed. The multiplexing of on/off sources and the use of self-similar traffic models revealed that the proposed dual PD-controller is able to keep buffer levels at the target levels as long as sufficient ABR traffic is generated. The random source rates result in fluctuations of the queue lengths, but the buffering at the ABR class queue still allows a 100 % link usage for most of the time.

Finally, it was proven that the developed VS/VD flow control algorithm is able to inter-operate with existing ER algorithms in a mixed environment. The performance of the dual PD-controller may be however restricted by the behavior of the coexisting switch algorithms.

It can be concluded that the ABR VS/VD switch algorithm presented in this work demonstrates a stable, robust, and efficient operation with fair bandwidth allocations and effective queue length control. The analytical as well as the simulation results confirm that the design objectives defined in the previous chapter are fully met. The simulation results were obtained for a variety of network configurations, realistic source models, and different load conditions and therefore offer a high level of credibility.

CHAPTER 7

Conclusion

The *Asynchronous Transfer Mode* (ATM) is a network technology that integrates different kinds of traffic streams, like video, audio, and data and therefore has to provide various service categories. Besides the categories *Constant Bit Rate* (CBR) and *Variable Bit Rate* (VBR) for high priority traffic, the lower priority class *Available Bit Rate* (ABR) attains a lot of interest, as it is expected to offer a cheaper service. Furthermore, it will provide a guaranteed *Minimum Cell Rate* and a low cell loss rate for the transmitted traffic. Whereas newly defined low priority traffic classes like *Guaranteed Frame Rate* (GFR) need a traffic description (maximum frame size, maximum burst size), ABR connections may be used for aggregated, bursty traffic flows without knowledge of specific traffic parameters. Typical application areas are Internet access using *Asymmetric Digital Subscriber Lines* (ADSL) [Kwok99] or local area networks [HaSA99]. The ABR service class will therefore play an important role in ATM networks.

As ABR traffic has to use the link bandwidth spared by CBR and VBR connections, it has to cope with fluctuating capacities available for the ABR service class. Therefore, a flow control mechanism is used to adapt current ABR traffic to available link capacities. For this, the ABR source establishes an end-to-end feedback control loop along the network path to the destination and back. If the *Virtual Source / Virtual Destination* (VS/VD) mechanism is used at the ATM switches, this control loop is split into separate segments. The advantages of this technique are the isolation of ABR traffic flows in subnets and the possibility to reduce the feedback delay by shortening the control loop.

The goal of this thesis was to develop and evaluate a novel VS/VD switch algorithm, that realizes the separation of the ABR control loop and provides guaranteed and analytically proven performance characteristics. A fair allocation of the ABR capacity to the ABR connections as well as an effective control of the switch queue lengths were targeted as major performance objectives. Existing works either do not address the special case of a VS/VD switch in their analytical analysis of switch algorithms or use only heuristics and trial and error approaches for the VS/VD switch algorithm design.

For the development of the switch algorithm, an analytical approach using linear control theory was chosen. The ABR flow control mechanism was modeled as a closed-loop feedback control system, where the controller process is located at the VS/VD switch. It determines the transmission rate of the established ABR connections for the upstream and downstream control loops. Hence, a dual control model was developed, whose structure suggested the use of a proportional-derivative (PD) controller for each of the two separate control loops. With the help of mathematical techniques like the Laplace-Transform or the Nyquist criterion, it was possible to target important performance objectives, like stability and robustness, already during the design process of the novel VS/VD switch algorithm.

Design objectives related to the temporal behavior of the system, like queue control and fairness, could not be analyzed mathematically. The reason for this is, that the control model of the dual PD-controller developed in this thesis is analytically not tractable in the time domain. As a consequence, the temporal behavior of the novel VS/VD switch algorithm was investigated by cell-based, discrete event simulations.

7.1 Achievements

The major contribution of this thesis is the development, implementation, and performance evaluation of an ABR flow control algorithm tailored for the special design of a VS/VD switch. This work provides an analytically designed VS/VD switch algorithm based on linear control theory offering mathematically proven performance characteristics. Hence, ATM switch manufactures are able to take this reliable and sound ABR flow control algorithm design as a basis for a simple and effective implementation of the VS/VD technique, improving and enhancing their products.

An important quality of a flow control algorithm is its stability, i.e., a bounded controller input results in a bounded output signal. The fulfillment of this goal was proven mathematically using the Nyquist criterion. The stable operation of the controller is maintained, even in the presence of small inaccuracies in the control model or of measurements of system variables at regular sampling intervals. This robustness was analytically assessed by the use of Nyquist and Bode plots. Therefore, network operators using the dual PD-controller developed in this thesis can rely on a safe and correct behavior of their VS/VD equipment under all network conditions.

The computational complexity of the new VS/VD algorithm is low, because no complex evaluations have to be performed. This allows a flexible implementation on different VS/VD switch platforms with no special requirements for the available hardware.

The PD-controller parameters are determined automatically and hence the developed VS/VD switch algorithm has no user-defined parameters, that need to be selected or tuned. As a consequence, the network operator is freed from any complicated and error-prone parameter configurations or adjustments, always obtaining an optimal ABR performance.

Due to the low target buffer lengths at the VS/VD switch, the developed algorithm scales well in the number of supported ABR connections. Thus, the algorithm is well suited for its main

application area, i.e., on backbone VS/VD switches, that operate at inter-carrier subnet boundaries.

The newly developed dual PD-controller is fully compliant to the ABR traffic management standard of the ATM Forum. The proposed VS/VD switch algorithm implements the standard ABR end system behavior and does not rely on any proprietary modifications of the standard ABR flow control mechanism. Therefore, it is inter-operable with other ATM switches and may easily be integrated into existing ATM networks, enhancing their functionality and performance.

Beside the analytical assessment by mathematical methods, a rigorous and extensive performance evaluation using discrete event simulations was performed in this work. The simulation experiments investigated various performance metrics, network topologies, source configurations, and load conditions.

The ABR flow control algorithm presented in this thesis shows a swift response to changing load conditions and allocates fair rates. It keeps the ABR switch queues at their target levels, avoiding cell loss, and achieves an optimal utilization of the available bandwidth, which is an important feature for ATM network operators. The small and steady ABR switch queue levels achieved by the VS/VD algorithm lead to a low end-to-end delay with a remarkably small jitter, which improves the *Quality of Service* (QoS) perceived by the end user.

Furthermore, the special features and advantages of VS/VD switches may be used not only in wide area networks, for shortening the control loop, or at administrative subnet boundaries, for controlling individual ABR connections, but also for special network topologies. Wireless access networks or satellite links benefit from VS/VD switches at their ingress or egress points, bridging fixed and wireless segments with their different characteristics [Goya98], [NuMa98]. The dual controller design developed in this thesis reflects the structure of the VS/VD principle. Therefore, the approach is very flexible and can easily be adapted to the special requirements of wireless networks. The PD-controller responsible for the wireless network segment can be tuned to support the long feedback delay and high loss rates of radio links or may even be replaced by another controller type, if desired.

Compared to existing ABR flow control algorithms for classical ATM switches, the newly developed VS/VD algorithm leads to more steady switch queue lengths and a reduced buffer capacity because of the shortened control loop. In contrast to the only other switch algorithm adapted for VS/VD that was extensively discussed in related works, namely, the ERICA+ algorithm [Goya97], the dual PD-controller offers an improved buffer occupancy with only small oscillations and the elimination of parameters, that have to be tuned by the user.

From the analytical and simulation results it could be concluded that the dual PD-controller based VS/VD switch algorithm developed in this thesis does fully meet the performance requirements, defined during the design process. An efficient and simple VS/VD switch algorithm was developed in this thesis providing a fair, stable, and robust ABR flow control, leading to a high throughput and link utilization. An implementation on a VS/VD switch would result in improved performance and Quality of Service parameters for ABR connections in ATM networks.

7.2 Future Work

The VS/VD switch algorithm presented in this thesis was developed for unicast ABR connections. As a future extension, it may be modified to also support ABR multicast transmissions. For this, additional efforts are necessary to investigate the consolidation of feedback information at the branch points of the multicast tree [ChCh00].

The insight gained during the development of the novel dual PD-controller is helpful for investigating flow control mechanisms other than for ABR traffic in ATM networks. The performance results achieved by the proposed algorithm make this approach a promising option for flow and congestion control on higher network layers.

This concerns especially higher layer protocols where adequate congestion control is of major importance, such as reliable multicast [RFC2357]. Due to the possibly large geographical extension of multicast distribution trees and the possibly large number of receiving stations, effective mechanisms, like the segmentation of the control loop as investigated in this thesis, are of particular interest.

Furthermore, there are existing works, which propose a distributed, rate-based flow control algorithm on the transport layer as an enhancement for existing unicast protocols. The scheme could achieve a fair bandwidth distribution and a possible prioritizing of certain traffic streams [FrMa99].

Other approaches address directly the common *Transmission Control Protocol* (TCP) on the transport layer, which operates with a window-based flow and congestion control mechanism [Jaco88]. Whereas the standard TCP and its enhancements use lost packets and time-out mechanisms to detect network congestion (implicit congestion indication), a binary *Explicit Congestion Notification* (ECN) was also proposed [RFC2481], [Floy94]. This scheme corresponds to the binary congestion indication for ABR traffic in ATM networks. Several options for the transport of the congestion bit [Floy94] and for the reaction of the TCP source to congestion have been proposed [OMUM99], [HaWa00].

But emerging multimedia streaming applications transmitting real-time voice and video require other congestion control mechanisms than TCP, leading to a fair bandwidth allocation, a fast rate adaptation to available capacities, and omitting retransmissions of lost packets. Especially large source rate oscillations caused by the TCP congestion control have a deteriorating effect on multimedia streaming applications. Therefore, several proposals exist to replace the TCP and its window-based congestion control mechanism by a rate-based approach.

The *Rate Adaption Protocol* (RAP) uses an additive increase and multiplicative decrease algorithm for rate adaption, focusing on a TCP-friendly operation, i.e., it behaves fair with respect to competing TCP connections [ReHE99]. Another proposal leading to a more steady rate allocation than RAP is the *TCP-friendly Rate Control Protocol* (TFRC)[FHPW00].

It has to be noted however, that the original window-based operation of TCP as well as the newly proposed rate-based protocols still rely on measurements of the round trip time and the packet loss at the end systems. The source calculates a transmission rate based on these values

and hence tries to deduct or even to guess the actual congestion situation in the network. However, modern network nodes, combining routing and switching technology, provide sophisticated scheduling and queueing capabilities together with a large memory and computing power [BWBM98], [Cisc01], [Shio00], [Soum99]. This technological progress would allow today a shift from existing TCP or rate-based flow control algorithms towards an *explicit rate indication*, where the source rate is determined by the network nodes along the path. This shift could profit from the extensive research in the area of QoS and congestion control for ATM networks.

The flow control algorithm developed in this thesis may be used as a basis to define an end-to-end, explicit rate congestion control protocol on the transport layer. A step-wise migration strategy could be targeted, where the existing TCP connections are supported in parallel, but handled separately from the explicit rate connections. Hence, the network nodes rather than the sources would determine a fair allocation of available bandwidth and the ratio between TCP and explicit rate bandwidth. As a result, the explicit rate algorithm needs not to mimic the TCP behavior any more, in order to be TCP-friendly. In particular, the small and steady switch buffer levels achieved by the proposed dual PD-controller result in low end-to-end delays with small jitter. Multimedia applications for voice and video transport over the Internet would directly benefit from these features. Including a QoS aware connection admission control in an explicit rate flow control protocol, would allow connection-based QoS guarantees like for instance a minimum transmission rate.

The achievements and results of this thesis are therefore an essential building block for the development of a new explicit rate transport protocol or "next generation TCP", that would surely bring a new push for evolving multimedia Internet applications due to unprecedented performance and QoS capabilities.

Appendix A

Glossary

A

AAL . . . . . ATM Adaptation Layer 11
ABT . . . . . ATM Block Transfer 27
ABT/DT . . . . . ABT with Delayed Transmission 27
ABT/IT . . . . . ABT with Immediate Block Transmission 27
ACK . . . . . Acknowledgment 27
ACR . . . . . Allowed Cell Rate 38
ADSL . . . . . Asymmetric Digital Subscriber Line 169
ADTF . . . . . ACR Decrease Time Factor 47
AINI . . . . . ATM Inter-Network Interface 8
AL. . . . . Alignment 17
AR . . . . . Autoregressive 87
ATC . . . . . ATM Transfer Capability 25
ATM . . . . . Asynchronous Transfer Mode 5

B

BASize . . . . . Buffer Allocation Size 17
BCR . . . . . Block Cell Rate 27
BECN . . . . . Backward Explicit Congestion Notification 41
BES. . . . . Binary Enhanced Switch 43
B-Frame . . . . . Bidirectional-Frame 132
B-ISDN. . . . . Broadband-ISDN 5
B-ISUP . . . . . B-ISDN User Part 8
BN . . . . . Backward Notification 41
BRM . . . . . Backward Resource Management 39

BTag . . . . . Beginning Tag 16
BTE. . . . . Broadband Terminal Equipment 126

C

CAC . . . . . Connection Admission Control 28
CBR . . . . . Constant Bit Rate 22
CCITT. . . . . International Telegraph and Telephone Consultative Committee . 5
CCR . . . . . Current Cell Rate 41
CDF. . . . . Cutoff Decrease Factor 47
CDFGamma . . . . . Gamma Cumulative Density Function 138
CER. . . . . Cell Error Ratio 26
CI. . . . . Congestion Notification 41
CID . . . . . Channel ID 15
CLP. . . . . Cell Loss Priority 10
CLR. . . . . Cell Loss Ratio 22
CMR . . . . . Cell Misinsertion Rate 25
CP . . . . . Common Part 13
CPCS. . . . . Common Part Convergence Sublayer 13
CPI . . . . . Common Part Indicator 16
CPS . . . . . Common Part Sublayer 15
CRC . . . . . Cyclic Redundancy Check 11
CRM . . . . . Missing RM-Cell Count 47
CS . . . . . Convergence Sublayer 13
CSI . . . . . Convergence Sublayer Indication 15

D

dB . . . . . Decibel 66
DBR . . . . . Deterministic Bit Rate 26
DCT. . . . . Discrete Cosine Transformation 132
DERA . . . . . Distributed Explicit Rate Allocation 80
DIR . . . . . Direction 41
DPF. . . . . Down Pressure Factor 79
DSL. . . . . Digital Subscriber Line 57
DSS2. . . . . Digital Subscriber Signalling System 2 8
DTFT. . . . . Discrete Time Fourrier Transform 129

E

ECN. . . . . Explicit Congestion Notification 172
EFCI . . . . . Explicit Forward Congestion Notification 41

EPD . . . . . Early Packet Discard 23
EPRCA . . . . . Enhanced Proportional Rate Control Algorithm 79
ER . . . . . Explicit Rate 41
ERICA . . . . . Explicit Rate Indication for Congestion Avoidance 81
ETag . . . . . End Tag 17

F

F-ARIMA . . . . . Fractional Autoregressive Integrated Moving-Average 129
FEC . . . . . Forward Error Correction 14
FFT . . . . . Fast Fourrier Transform 129
FGN . . . . . Fractional Gaussian Noise 129
FIFO . . . . . First-In-First-Out 81
FRM . . . . . Forward Resource Management 39
FRTT . . . . . Fixed Round-Trip Time 53
FTP . . . . . File Transfer Protocol 12

G

GBAR . . . . . Gamma-Beta Autoregressive 132
GCRA . . . . . Generic Cell Rate Algorithm 29
GFC2 . . . . . Generic Fairness Configuration No. 2 153
GFR . . . . . Guaranteed Frame Rate 24
GOP . . . . . Group of Pictures 132

H

HEC . . . . . Header Error Control 11

I-K

IBT . . . . . Intrinsic Burst Tolerance 26
ICR . . . . . Initial Cell Rate 45
I-Frame . . . . . Intra-Frame 132
IP . . . . . Internet Protocol 53
ISDN . . . . . Integrated Services Digital Network 5
ITU-T . . . . . International Telecommunications Union - Telecommunication Standardization Sector 5

L

LAN . . . . . Local Area Network 23
LFS . . . . . Local Fair Share 80

M

N

O

P

PVC . . . . . Permanent Virtual Connection 8

Q

QDLF . . . . . Queue Drain Limit Factor 81
QL. . . . . Queue Length 41
QoS. . . . . Quality of Service 6

R

RA . . . . . Request/Acknowledge 41
RAP . . . . . Rate Adaption Protocol 172
RDF . . . . . Rate Decrease Factor 47
RIF . . . . . Rate Increase Factor 47
RM . . . . . Resource Management 39
RTT . . . . . Round-TripTime 27
rt-VBR . . . . . Real-Time Variable Bit Rate 22

S

SAR . . . . . Segmentation and Reassembly 13
SBR . . . . . Statistical Bit Rate 26
SC. . . . . Sequence Count 15
SCR . . . . . Sustainable Cell Rate 21
SDH . . . . . Synchronous Digital Hierarchy 19
SDT . . . . . Structured Data Transfer 14
SDU . . . . . Service Data Unit 13
SECBR. . . . . Severely Errored Cell Block Ratio 25
SN. . . . . Sequence Number 42
SNP. . . . . Sequence Number Protection 15
SONET. . . . . Synchronous Optical Network 115
SRD . . . . . Short-Range Dependent 129
SSCS. . . . . Service Specific Convergence Sublayer 13
ST . . . . . Segment Type 17
STF. . . . . Start Field 15
SVC . . . . . Switched Virtual Connection 8

T

TAT . . . . . Theoretical Arrival Time 29
TBE . . . . . Transient Buffer Exposure 53
TC. . . . . Transmission Convergence 19

U

V

W-Z

Appendix B

List of Figures

Appendix C

List of Tables

Appendix D

References

D.1 Books and Papers

[AbKu97] Abraham S. P., Kumar A.: "Max-Min Fair Rate Control of ABR Connections with Nonzero MCRs", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 97), Vol. 1, pp. 498-502, Phoenix, November 4-8, 1997

[AfMO00] Afek Y., Mansour Y., Ostfeld Z.: "Phantom: A Simple and Effective Flow Control Scheme", Computer Networks, Vol. 32, Issue 3, pp. 277-305, March 2000

[AiHo99] Aida M, Horikawa K.: "Stability Analysis for Global Performance of Flow Control in High-Speed Networks Based on Statistical Physics", IEICE Transactions on Communications, Vol. E82-B, No. 12, pp. 2095-2106, December 1999

[AAEE99] Ait-Hellal O., Altman E., Elouadghiri D., Erramdani M.: "Performance Evaluation of the Rate-Based Flow Control Mechanism for ABR Service", Telecommunication Systems, Vol. 12, No. 2-3, pp. 211-236, 1999

[ABLM96] Ajmone-Marsan M., Bianco A., Lo Cigno R., Munafo M.: "Four Standard Control Theory Approaches for the Implementation of RRM ABR Services", ATM Networks Performance Modelling and Analysis Volume 3, IFIP TC6 WG6.3/WG6.4, Fourth International Workshop on Performance Modelling and Evaluation of ATM Networks, pp. 259-279, Kouvatsos D. (Ed.), Chapman & Hall, Ilkley, July 8-10, 1996

[AkLN97] Akyildiz I. F., Liebeherr J., Nikolaidis I.: "Multi-Level Rate-Based Flow Control for ABR Traffic", Performance Evaluation, Vol. 31, Issue 1-2, pp. 107-131, November 1997

[AlBB94] Altman E., Baccelli F., Bolot J.-C.: "Discrete-Time Analysis of Adaptive Rate Control Mechanisms", High Speed Networks and their Performance, Perros H. G., Viniotis Y. (Eds.), pp. 121-140, North Holland, 1994

[AlBH99] Altman E., Basar T., Hovakimyan N.: "Worst-Case Rate-Based Flow Control with an ARMA Model of the Available Bandwidth", Advances in Dynamic Games and Applications, Filar J. A., Mizukami K., Gaitsgory V. (Eds.), Annals of the International Society of Dynamic Games, Vol. 5, Birkhauser, 1999

[AlBS97] Altman E., Basar T., Srikant R.: "Flow Control in Communication Networks with Multiple Users and Action Delays: A Team-Theoretic Approach", Proc. of the 36th IEEE Conference on Decision and Control, pp. 2387-2392, San Diego, December 10-12, 1997

[AlBS98] Altman E., Basar T., Srikant R.: "Robust Rate Control for ABR Services", Proc. of IEEE INFOCOM'98, pp. 166-173, San Francisco, March 29 - April 2, 1998

[AlBS99] Altman E., Basar T., Srikant R.: "Congestion Control as a Stochastic Control Problem with Action Delays", Automatica, Vol. 35, Issue 12, pp. 1937-1950, December, 1999

[AnAA98] An L., Ansari N., Arulambalam A.: "TCP/IP Traffic over ATM Networks with FMMRA ABR Flow and Congestion Control", Computer Networks and ISDN Systems, Vol. 29, Issue 17-18, pp. 2091-2102, February 1998

[ArCA96] Arulambalam A., Chen X., Ansari N.: "Allocating Fair Rates for Available Bit Rate Service in ATM Networks", IEEE Communications Magazine, Vol. 34, No. 11, pp. 92-100, November 1996

[AtMM98] Ata S., Murata M., Miyahara H.: "Performance Evaluation of TCP over ABT Protocols", SPIE Conference on Performance and Control of Network Systems II, Lai W. S., Cooper R. B. (Eds.), pp. 423-433, Vol. 3539, Boston, November 2-4, 1998

[Atas00] Ataslar B. et al.: "Robust Rate-Based Flow Controllers for High-Speed Networks: The Case of Uncertain Time-varying Multiple Time-Delays", Proc. of the 2000 American Control Conference (ACC 2000), pp. 2804-2808, Chicago, June 28-30, 2000

[Baal99] Baalmann J.: "Flow Control for Available Bit Rate Traffic in ATM-Networks: Evaluation of the Virtual Source/Virtual Destination-Technique", Diploma Thesis, University of Technology Aachen (RWTH Aachen), Department. of Computer Science IV, March 1999

[BaCs99] Baalmann J., Cseh C.: "Rate Indication Flow Control Algorithm Using the Virtual Source / Virtual Destination Technique for Available Bit Rate Service", Proc. of the 3rd IFIP Workshop on Traffic Management & Design of ATM Networks, p. 4/1-4/16, London, April 1999

[BaCs00] Baalmann J., Cseh C.: "Evaluation of the Virtual Source/Virtual Destination-Technique for Available Bit Rate Traffic in ATM-Networks", Proc. of the 2nd IEEE International Conference on ATM (ICATM'99), pp. 72-77, Colmar, June 21-23, 1999, extended version in Annals of Telecommunications, Vol. 55, No. 11-12, pp. 577-587, November-December 2000

[Bald97] Baldwin J. H. et al.: "AAL-2 – A New ATM Adaptation Layer for Small Packet Encapsulation and Multiplexing", Bell Labs Technical Journal, Spring 1997

[BeMe93] Benmohamed L., Meerkov S. M.: "Feedback Control of Congestion in Packet Switching Networks: The Case of a Single Congested Node", IEEE/ACM Trans. on Networking, Vol.1, No. 6, pp. 693-708, December 1993

[BeWa98] Benmohamed L., Wang Y. T.: "A Control-Theoretic ABR Explicit Rate Algorithm for ATM Switches with per-VC Queueing", Proc. of IEEE INFOCOM'98, pp. 183-191, San Francisco, March 29 - April 2, 1998

[BeGa92] Bertsekas D., Gallager R.: "Data Networks", Second Edition, Prentice Hall, 1992

[Blac95a] Black U.: "ATM, Foundation for Broadband Networks", Volume 1, Second Edition, Prentice Hall, 1995

[Blac95b] Black U.: "ATM, Signaling in Broadband Networks", Volume 2, Prentice Hall, 1995

[BoSh90] Bolot J.-C., Shankar A. U.: "Dynamical Behavior of Rate-Based Flow Control Mechanisms", ACM Computer Communication Review, Vol. 20, No. 2, pp.35-49, April 1990

[BoSh92] Bolot J.-C., Shankar A. U.: "Analysis of a Fluid Approximation to Flow Control Dynamics", Proc. of IEEE INFOCOM'92, pp. 2398-2407, Florence, May 4-8, 1992

[BoFe95] Bonomi F., Fendick K. W.: "The Rate-Based Flow Control Framework for the Available Bit Rate ATM Service", IEEE Network, Vol. 9, No. 2, pp. 25-39, March/April 1995

[BoMS95] Bonomi F., Mitra D., Seery J. B.: "Adaptive Algorithms for Feedback-Based Flow Control in High-Speed, Wide-Area ATM Networks", IEEE Journal on Selected Areas in Communications, Vol. 13, No. 7, pp. 1267-1283, September 1995

[BrPe95] Brakmo L. Peterson L.: "TCP Vegas: End to End Congestion Avoidance on a Global Internet", IEEE Journal on Selected Areas in Communications, Vol. 13, No. 8, pp. 1465-1480, October 1995

[BrFS87] Bratley P., Fox B. L., Schrage L. E.: "A Guide to Simulation", Second Edition, Springer, 1987

[BWBM98] Briem U., Wallmeier E., Beck C., Matthiesen F.: "Traffic Management for an ATM Switch with Per-VC Queueing: Concept and Implementation", IEEE Communications Magazine, Vol. 36, No. 1, pp. 88-93, January 1998

[CaMG96] Cavendish D., Mascolo S., Gerla M.: "SP-EPRCA: an ATM Rate Based Congestion Control Scheme Based on a Smith Predictor", UCLA Technical Report #960001, February 1996

[CFKS97] Chandra P., Fisher A., Kosak C., Steenkiste P.: "Experimental Evaluation of ATM Congestion Control Mechanisms", Proc. of IEEE INFOCOM'97, pp.1324-1332, Kobe, April 7-11, 1997

[Char94] Charny A.: "An Algorithm for Rate Allocation in a Packet-Switching Network with Feedback", Master Thesis, Massachusetts Institute of Technology (MIT), May 1994

[ChCJ95] Charny A., Clark D. D., Jain R.: "Congestion Control With Explicit Rate Indication", Proc. of IEEE International Conference on Communications (ICC'95), pp. 1954-1963, Seattle, June 18-22, 1995

[ChRL96] Charny A., Ramakrishnan K. K., Lauck A.: "Time Scale Analysis and Scalability Issues for Explicit Rate Allocation in ATM Networks", IEEE/ACM Transactions on Networking, pp. 569-581, Vol. 4, No. 4, August 1996

[ChCC98] Chen J.-R., Chen Y.-C., Chan C.-T.: "Performance Evaluation of an End-to-End Rate Control Approach for ABR Services", IEICE Transactions on Communications, Vol. E81-B, No. 12, pp. 2400-2412, December 1998

[ChLi95] Chen T. M., Liu S. S.: "ATM Switching Systems", Artech House, 1995

[ChLS96] Chen T. M., Liu S. S., Samalam V. K.: "The Available Bit Rate Service for Data in ATM Networks", IEEE Communications Magazine, Vol. 34, No. 5, pp. 56-71, May 1996

[ChJa89] Chiu D.-M., Jain R.: "Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks", Computer Networks and ISDN Systems, Vol. 17, pp. 1-14, June 1989

[ChWa97] Chiussi F., Wang Y. T.: "An ABR-Based Congestion Control Algorithm for ATM Switches with Per-VC Queueing", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 97), Vol. 2, pp. 771-778, Phoenix, November 4-8, 1997

[ChLi98] Cho K.-H., Lim J.-T.: "Supervisory Rate-Based Flow Control of ATM Networks for ABR Services", IEICE Transactions on Communications, Vol. E81-B, No. 6, pp. 1269-1271, June 1998

[ChLL97] Cho Y.-Z., Lee S.-M., Lee M.-Y.: "An Efficient Rate-Based Algorithm for Point-to-Multipoint ABR Service", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 97), Vol. 2, pp. 790-795, Phoenix, November 4-8, 1997

[ChNW98a] Chong S., Nagarajan R., Wang Y.-T.: "Designing Stable ABR Flow Control with Rate Feedback and Open-Loop Control: First-Order Control Case", Performance Evaluation, Vol. 34, Issue 4, pp. 189-206, December 18, 1998

[ChNW98b] Chong S., Nagarajan R., Wang Y.-T.: "First-Order Rate-Based Flow Control with Dynamic Queue Threshold for High-Speed Wide-Area ATM Networks", Computer Networks and ISDN Systems, Vol. 29, Issue 17-18, pp. 2201-2212, February 1998

[ChCh00] Chuang C.-M., Chang J.-F.: "Virtual Source / Virtual Destination (VS/VD) Congestion Control for Multicast ABR Services in ATM Networks", IEICE Transactions on Communications, Vol. E83-B, No. 3, pp. 646-658, March 2000

[Cisc01] Cisco Systems: "BPX 8600 Series — Wide-Area Switches", Cisco Product Catalog, February 2001

[CsKR97] Cseh C., Karabek R., Reichl P.: "Segmentation of the Traffic Control Loop for Available Bit Rate-Service", Proc. of the 5th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, pp. 38/1-38/10, Ilkley, July 1997

[DaSt98] Dag T., Stavrakakis I.: "Evaluation of ABR Traffic Management under Various System Time Scales", Computer Networks and ISDN Systems, Vol. 29, Issue 17-18, pp. 2103-2117, February 1998

[Deat96] Deaton G.: "Juggling ATM Traffic", Data Communications, Vol. 25, No. 5, pp. 130-138, April 1996

[DePr95] De Prycker M.: "Asynchronous Transfer Mode, Solution for Broadband ISDN", Third Edition, Prentice Hall, 1995

[Dobr98] Dobrowski G. H.: "The ATM Forum: Developing Implementation Agreements", IEEE Communications Magazine, Vol. 36, No. 9, pp. 121-125, September 1998

[Doet61] Doetsch G.: "Guide to the Applications of the Laplace and Z-Transforms", Van Nostrand Reinhold, 1961

[Dorf89] Dorf R. C.: "Modern Control Systems", Addison-Wesley, 1989

[DoDe95] Douligeris C., Develekos G.: "A Fuzzy Logic Approach to Congestion Control in ATM Networks", Proc. of IEEE International Conference on Communications (ICC'95), pp. 1969-1973, Seattle, June 18-22, 1995

[Ebel87] Ebel T.: "Regelungstechnik", Teubner 1987

[Elwa95] Elwalid A. I.: "Analysis of Adaptive Rate-based Congestion Control for High-Speed Wide-Area Networks", Proc. of IEEE International Conference on Communications (ICC'95), pp. 1948-1953, Seattle, June 18-22, 1995

[EAYN99] Endo N., Aimoto T., Yazaki T., Nishijima T.: "Large Scale Enterprise ATM Node with Per-VC Traffic Control", Computer Networks, Vol. 31, Issue 6, pp. 617-625, March 1999

[Fahm99] Fahmy S. et al.: "Quality of Service for Internet Traffic over ATM Service Categories", Computer Communications, Vol. 22, Issue 14, pp. 1307-1320, September 15, 1999

[FaFl96] Fall K., Floyd S.: "Simulation-based Comparisons of Tahoe, Reno, and SACK TCP", ACM Computer Communication Review, Vol. 26, No. 3, pp. 5-21, July 1996

[Fend96] Fendick K. W.: "Evolution of Controls for the Available Bit Rate Service", IEEE Communications Magazine, Vol. 34, No. 11, pp. 35-39, November 1996

[FGFM97] Fichou A., Galand C., Fdida S., Moret Y.: "Evaluation of the ER Algorithm ERAQLES in Different ABR Environments", Proc. of the 5th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, pp. 48/1-48/13, Ilkley, July 21-23, 1997

[Fish78] Fishman G. S.: "Principles of Discrete Event Simulation", John Wiley & Sons, 1978

[FHPW00] Floyd S., Handley M., Padhye J., Widmer J.: "Equation-based Congestion Control for Unicast Applications", Proc. of ACM SIGCOMM 2000, Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, pp. 43-56, Stockholm, August 28 - September 1, 2000

[Floy94] Floyd S.: "TCP and Explicit Congestion Notification", ACM Computer Communication Review, Vol. 24, No. 5, pp. 10-23, October 1994

[FrMa99] Frank M., Martini P.: "Eine Erweiterung der Transportebene zur Bandbreitenregelung auf Ende-zu-Ende Basis", Praxis der Informationsverarbeitung und Kommunikation, Vol. 22, No. 1, pp. 11-21, January-March 1999

[FrNg00] Frey M., Nguyen-Quang S.: "A Gamma-Based Framework for Modeling Variable-Rate MPEG Video Sources: The GOP GBAR Model", IEEE/ACM Transactions on Networking, Vol. 8, No. 6, pp. 710-719, December 2000

[FrMe94] Frost V., Melamed B.: "Traffic Modelling for Telecommunication Networks", IEEE Communications Magazine, Vol. 32, No. 3, pp. 70-80, March 1994

[FuLL97] Fulton C., Li S. Q., Lim C. S.: "An ABR Feedback Control Scheme with Tracking", Proc. of IEEE INFOCOM'97, pp. 805-814, Kobe, April 7-11, 1997

[GaBo96] Gaiti D., Boukhatem N.: "Cooperative Congestion Control Schemes in ATM Networks", IEEE Communications Magazine, Vol. 34, No. 11, pp. 102-110, November 1996

[GaWi94] Garrett M. W., Willinger W.: "Analysis, Modeling and Generation of Self-Similar Video Traffic", ACM Computer Communication Review, Vol. 24, No. 4, pp. 269-280, October 1994

[GhMa97a] Ghani N., Mark J. W.: "Measurement-Based Flow Control for ABR Services in ATM Networks", European Transactions on Telecommunications & Related Technologies, Vol. 8, No. 1, pp. 39-53, January-February 1997

[GhMa97b] Ghani N., Mark J. W.: "Queueing Analysis of a Distributed Explicit Rate Allocation Algorithm for ABR Services", Proc. of IEEE INFOCOM'97, pp. 1315-1323, Kobe, April 7-11, 1997

[GiGa98] Giroux N., Ganti S.: "Quality of Service in ATM Networks: State-of-the-Art Traffic Management", Prentice Hall, 1998

[GoSS98] Golmie N., Saintillan Y., Su D.: "ABR Switch Mechanisms: Design Issues and Performance Evaluation", Computer Networks and ISDN Systems, Vol. 30, Issue 19, pp. 1749-1761, October 1998

[Golm98] Golmie N.: "The NIST ATM/HFC Network Simulator, Operation and Programming Guide", Version 4.0, December 1998

[Goya97] Goyal R. et al.: "Per-VC Rate Allocation Techniques for ABR Feedback in VS/VD Networks", ATM Forum, Traffic Management Working Group, ATM Forum Contribution 97-1086R1, Ohio State University (OSU), February 1997

[Goya98] Goyal R. et al.: "Per-VC Rate Allocation Techniques for ABR Feedback in Virtual Source Virtual Destination Networks", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 98), pp. 1778-1785, Sydney, November, 8-12, 1998

[HaWa00] Hamann T., Walrand J.: "A New Fair Window Algorithm for ECN Capable TCP (new-ECN)", Proc. of IEEE INFOCOM 2000, pp.1528-1530, Tel Aviv, March 26-30, 2000

[HKSO98] Hashmani M. A., Kawahara K., Sunahara H., Oie Y.: "Efficient Use of ABR Bandwidth in Highly Variant VBR/CBR Background Traffic", Proc. of 6th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks (IFIP ATM'98), pp. 07/01-07/10, Ilkley, July 20-22, 1998

[HaBA99] Hassan M., Breen J., Atiquzzaman M.: "HG-RCP: A Rate-Based Flow Control Mechanism for Intranets Interconnected by ATM Networks", Computer Networks, Vol. 31, Issue 22, pp. 2361-2379, November 1999

[HaSA99] Hassan M., Sirisena H., Atiquzzaman M.: "A Congestion Control Mechanism for Enterprise Network Traffic over Asynchronous Transfer Mode Networks", Computer Communications, Vol. 22, Issue 14, pp. 1296-1306, September 15, 1999

[Heym97] Heyman D. P.: "The GBAR Source Model for VBR Videoconferences", IEEE/ACM Transactions on Networking, Vol. 5, No. 4, pp. 554-560, August 1997

[HoSu99] Hong D. P., Suda T.: "Performance of ATM Available Bit Rate for Bursty TCP Sources and Interfering Traffic", Computer Networks, Vol. 31, Issue 1-2, pp. 7-18, January 1999

[HoTP97a] Hou Y. T., Tzeng H., Panwar S. S.: "A Generalized Max-Min Network Capacity Assignment Policy with a Simple ABR Implementation for an ATM LAN", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 97), Vol. 1, pp. 503-508, Phoenix, November 4-8, 1997

[HoTP97b] Hou Y. T., Tzeng H., Panwar S. S.: "A Weighted Max-Min Fair Rate Allocation for Available Bit Rate Service", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 97), Vol. 1, pp. 492-498, Phoenix, November 4-8, 1997

[HTPK98a] Hou Y. T., Tzeng H., Panwar S. S., Kumar V. P.: "ATM ABR Traffic Control with a Generic Weight-Based Bandwidth Sharing Policy: Theory and a Simple Implementation", IEICE Transactions on Communications, Vol. E81-B, No. 5, pp. 958-972, May 1998

[HTPK98b] Hou Y. T., Tzeng H., Panwar S. S., Kumar V. P.: "On Rate Allocation Policies and Explicit Feedback Control Algorithms for Packet Networks", SPIE Conference on Performance and Control of Network Systems II, Lai W. S., Cooper R. B. (Eds.), pp. 2-20, Vol. 3539, Boston, November 2-4, 1998

[HuLP95] Hu X., Lambert J. F., Pitsillides A.: "Fast Backward Predictive Congestion Notification for ATM Networks with Significant Propagation Delays", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 95), Vol. 1, pp. 275-279, Singapore, November 13-17, 1995

[ImBa99a] Imer O. C., Basar T.: "Constructive Solutions to a Decentralized Team Problem with Application in Flow Control for Networks", Topics in Control and Its Applications, Miller D. E., Qiupp L. (Eds.), pp. 117-141, Springer, 1999

[ImBa99b] Imer O. C., Basar T.: "Optimal Solution to a Team Problem with Information Delays: An Application in Flow Control for Communication Networks", Proc. of the 38th IEEE Conference on Decision and Control, pp. 2697-2702, Phoenix, December 7-10, 1999

[Izma96] Izmailov R.: "Analysis and Optimization of Feedback Control Algorithms for Data Transfers in High-Speed Networks", SIAM Journal on Control and Optimization, Vol. 34, No. 5, pp. 1767-1780, September 1996

[Jaco88] Jacobson V.: "Congestion Avoidance and Control", Proc. of ACM SIGCOMM'88, Conference on Communications Architectures & Protocols, pp. 314-329, Stanford, August 16-19, 1988

[Jaff81] Jaffe J. M.: "Bottleneck Flow Control", IEEE Transactions on Communications, Vol. COM-29, No. 7, pp. 954-962, July 1981

[Jain96a] Jain R.: "Congestion Control and Traffic Management in ATM Networks: Recent Advances and a Survey", Computer Networks and ISDN Systems, Vol. 28, Issue 13, pp. 1723-1738, October 1996

[Jain96b] Jain R. et al.: "Source Behavior for ATM ABR Traffic Management: An Explanation", IEEE Communications Magazine, Vol. 34, No. 11, pp. 50-57, November 1996

[Jiao00] Jiao Y.: "Development and Evaluation of Traffic Models for Available Bit Rate-Traffic in ATM-Networks", Diploma Thesis, University of Technology Aachen, Department of Computer Science IV, July 2000

[KaVa95] Kalampoukas L., Varma A.: "Dynamics of an Explicit Rate Allocation Algorithm for Available Bit-Rate (ABR) service in ATM Networks", University of California Santa Cruz (UCSC), Technical Report UCSC-CRL-95-54, December 5, 1995

[Kala95] Kalampoukas L.: "An Efficient Rate Allocation Algorithm for ATM Networks", Master Thesis, University of California Santa Cruz (UCSC), June 1995

[Kala97] Kalampoukas L.: "Congestion Management in High Speed Networks", Ph.D. Thesis, University of California Santa Cruz (UCSC), September 1997

[Kaly96] Kalyanaraman S. et al.: "Performance of TCP/IP over ABR Service on ATM Networks," Proc. of IEEE Global Telecommunications Conference (GLOBECOM 96), Vol. 1, pp. 468 -475, London, November 18-22, 1996

[Kaly98a] Kalyanaraman S. et al.: "Design Considerations for the Virtual Source / Virtual Destination (VS/VD) Feature in the ABR Service of ATM Networks", Computer Networks and ISDN Systems, Vol. 30, Issue 19, pp. 1811-1824, October 1998

[Kaly98b] Kalyanaraman S. et al.: "Use-It-or-Lose-It Policies for the Available Bit Rate (ABR) Service in ATM Networks", Computer Networks and ISDN Systems, Vol. 30, Issue 24, pp. 2293-2308, December 1998

[Kaly00] Kalyanaraman S. et al: "ERICA Switch Algorithm for ABR Traffic Management in ATM Networks", IEEE/ACM Transactions on Networking, Vol. 8, No. 1, pp. 87-98, February 2000

[Kaly97] Kalyanaraman S.: "Traffic Management for the Available Bit Rate (ABR) Service in Asynchronous Transfer Mode (ATM) Networks", Ph.D. Thesis, The Ohio State University, 1997

[KJFG98] Kalyanaraman S., Jain R., Fahmy S., Goyal R.: "Performance and Buffering Requirements of Internet Protocols over ATM ABR and UBR Services", IEEE Communications Magazine, Vol. 36, No. 6, pp. 152-157, June 1998

[Kara98] Karabek M. R.: "Data Communications in ATM Networks", Dissertation, RWTH Aachen, Shaker, September 1998

[KaMM95] Kawahara K., Murata M., Miyahara H.: "Performance Analysis of Reactive Congestion Control for ATM Networks", IEEE Journal on Selected Areas in Communications, Vol. 13, No. 4, pp. 651-661, May 1995

[KKON99] Kawahara R., Kamado Y., Omotani M., Nagata S.: "Method of Implementing GFR Service in Large-Scale Networks Using ABR Control Mechanism and Its Performance Analysis", IEICE Transactions on Communications, Vol. E82-B, No. 12, pp. 2081-2094, December 1999

[Kesh91] Keshav S.: "A Control-Theoretic Approach to Flow Control", ACM Computer Communication Review, Vol. 21, No. 4, pp. 3-15, September 1991

[KiTh98a] Kim B.-K., Thompson C.: "ABR Traffic Control in ATM Networks using Optimal Control Theory", Proc. of 1st International Conference on ATM (ICATM'98), pp. 327-333, Colmar, June 22-24, 1998

[KiTh98b] Kim B.-K., Thompson C.: "Optimal Feedback Control of ABR Traffic in ATM Networks", Proc. of the IEEE International Conference on Communications (ICC'98), pp. 844-848, Atlanta, June 7-11, 1998

[KoRa95] Kolarov A., Ramamurthy G.: "End-to-end Adaptive Rate Based Congestion Control Scheme for ABR Service in Wide Area ATM Networks", Proc. of IEEE International Conference on Communications (ICC'95), pp. 138-143, Seattle, June 18-22, 1995

[KoRa97a] Kolarov A., Ramamurthy G.: "A Control Theoretic Approach to the Design of Closed Loop Rate Based Flow Control for High Speed ATM Networks", Proc. of IEEE INFOCOM'97, pp. 293-301, Kobe, April 7-11, 1997

[KoRa97b] Kolarov A., Ramamurthy G.: "Comparison of Explicit Rate and Explicit Forward Congestion Indication Flow Control Schemes for ABR Service in Wide Area Networks", Performance Evaluation, Vol. 31, Issue 1-2, pp. 89-106, November 1997

[KoRa97c] Kolarov A., Ramamurthy G.: "A Testbed Implementation of Explicit Rate Based ABR Service via a Dual Proportional-plus-Derivative Controller", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 97), Vol. 2, pp. 806-811, Phoenix, November 4-8, 1997

[KoRa99] Kolarov A., Ramamurthy G.: "A Control-Theoretic Approach to the Design of an Explicit Rate Controller for ABR Service", IEEE/ACM Transactions on Networking, Vol. 7, No. 5, pp. 741-753, October 1999

[Kou99] Kou K. Y.: "Realization of Large-Capacity ATM Switches", IEEE Communications Magazine, Vol. 37, No. 12, pp. 120-123, December 1999

[Krish97] Krishnan R.:"Rate-Based Control Schemes for ABR Traffic - Design Principles and Performance Comparison", Computer Networks and ISDN Systems, Vol. 29, Issue 5, pp. 583-593, April 1997

[KuBC94] Kung H. T., Blackwell T., Chapman A.: "Credit-Based Flow Control for ATM Networks: Credit Update Protocol, Adaptive Credit Allocation, and Statistical Multiplexing", Proc. of ACM SIGCOMM'94, Conference on Communications Architectures, Protocols and Applications, pp. 101-114, London, August 31 - September 2, 1994

[KuMo95] Kung H. T., Morris R.: "Credit-Based Flow Control for ATM Networks", IEEE Network, Vol. 9, No. 2, pp. 40-48, March/April 1995

[KuWa98] Kung H. T., Wang S. Y.: "Zero Qeueing Flow Control and Applications", Proc. of IEEE INFOCOM'98, pp. 192-200, San Francisco, March 29 - April 2, 1998

[Kwok99] Kwok T. C.: "Residential Broadband Architecture over ADSL and G.Lite (G.992.2): PPP over ATM", IEEE Communications Magazine, Vol. 37, No. 5, pp. 84-89, May 1999

[LaTs99] Lai W. K., Tsai J.-Y.: "A Flow Control Scheme on ATM Networks with Max-Min Fairness", Computer Communications, Vol. 22, Issue 6, pp. 543-555, April 1999

[LaLi98] Lai Y.-C., Lin Y.-D.: "Performance Analysis of Rate-Based Flow Control under a Variable Number of Sources", Computer Networks and ISDN Systems, Vol. 29, Issue 17-18, pp. 2139-2152, February 1998

[LRMS97] Lee D., Ramakrishnan K. K., Moh W. M., Shankar A. U.: "Performance and Correctness of the ATM ABR Rate Control Scheme", Proc. of IEEE INFOCOM'97, pp.785-794, Kobe, April 7-11, 1997

[LeRM98] Lee D., Ramakrishnan K. K., Moh W. M.: "A Formal Specification of the ATM ABR Rate Control Scheme", Computer Networks and ISDN Systems, Vol. 30, Issue 19, pp. 1735-1748, October 1998

[LeGa91] Le Gall D.: "MPEG: A Video Compression Standard for Multimedia Applications", Communications of the ACM, Vol. 34, No. 4, pp. 47-58, April 1991

[LTWW94] Leland W. E., Taqqu M. S., Willinger W., Wilson D. V.: "On the Self-Similar Nature of Ethernet Traffic (Extended Version)", IEEE/ACM Transactions on Networking, Vol. 2, No. 1, pp. 1-15, February 1994

[LeKa00] Lengliz I., Kamoun F.: "A Rate-based Flow Control Method for ABR Service in ATM Networks", Computer Networks, Vol. 34, Issue 1, pp. 129-138, July 2000

[LiDo97] Liu Y.-C., Douligeris C.: "Rate Regulation with Feedback Controller in ATM Networks - A Neural Network Approach", IEEE Journal on Selected Areas in Communications, Vol. 15, No. 2, pp. 200-208, February 1997

[LoHR00] Long Y. H., Ho T. K., Rad A. B.: "Fair Flow Control of ABR Service by per-VC Virtual Queuing", Computer Communications, Vol. 23, Issue 1, pp. 71-83, January 1, 2000

[LoLa99] Low S. H., Lapsley D. E.: "Optimization Flow Control - I: Basic Algorithm and Convergence", IEEE/ACM Trans. on Networking, Vol. 7, No. 6, pp. 861-874, December 1999

[LSMA98] Luetchford J. C., Schreinemachers M., Morita N., Arai H.: "Applications of ATM in Global Networks", IEEE Communications Magazine, Vol. 36, No. 8, pp. 104-109, August 1998

[MaVG98] Martinez J., Vidal J. R., Guijarro L.: "A Low Complexity Congestion Control Algorithm for the ABR Class of Service", Interactive Distributed Multimedia Systems and Telecommunication Services, 5th International Workshop, IDMS'98, Plagemann T., Goebel V. (Eds.), Springer, LNCS 1483, pp. 219-230, Oslo, September 8-11, 1998

[MaCG97] Mascolo S., Cavendish D., Gerla M.: "ATM Rate-Based Congestion Control Using a Smith Predictor", Performance Evaluation, Vol. 31, Issue 1-2, pp. 51-65, November 1997

[Masc97a] Mascolo S.: "Linear Control Theory for Modelling, Designing and Performance Evaluation of ATM Congestion Control Algorithms", Performance Analysis of ATM Networks, IFIP TC6 WG6.3/WG6.4, Fifth International Workshop on Performance Modelling and Evaluation of ATM Networks, pp. 309-336, Kouvatsos D. (Ed.), Kluwer, Ilkley, July 21-23, 1997

[Masc97b] Mascolo S.: "Smith's Principle for Congestion Control in High Speed Networks", Proc. of 36th IEEE Conference on Decision and Control, pp. 4595-4600, San Diego, December 10-12, 1997

[Masc99] Mascolo S.: "Congestion Control in High-Speed Communication Networks Using the Smith Principle", Automatica, Vol. 35, Issue 12, pp. 1921-1935, December 1999

[McSp94] McDysan D. E., Spohn D. L.: "ATM, Theory and Application", McGraw-Hill, 1994

[McDy00] McDysan D. E.: "QoS & Traffic Management in IP & ATM Networks", McGraw-Hill, 2000

[Meyr94] Meyr H.: "Regelungstechnik und Systemtheorie", Agustinus, 1994

[MiKT96] Mishra P. P., Kanakia H., Tripathi S. K.: "On Hop-by-Hop Rate-Based Congestion Control", IEEE/ACM Transactions on Networking, Vol. 4, No. 2, pp 224-239, April 1996

[MoKl99] Monin J.-F., Klay F.: "Correctness Proof of the Standardized Algorithm for ABR Conformance", FM '99 - Formal Methods. World Congress on Formal Methods in the Development of Computing Systems, Wing J., Woodcock J, Davies J. (Eds.), Springer, LNCS 1708, Vol. 1, pp. 662-681, Toulouse, September 20-24, 1999

[MoFd97] Moret Y., Fdida S.: "ERAQLES an Efficient Explicit Rate Algorithm for ABR", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 97), Vol. 2, pp. 801-805, Phoenix, November 4-8, 1997

[MoKu95] Morris R., Kung H. T.: "Impact of ATM Switching and Flow Control on TCP Performance: Measurements on an Experimental Switch", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 95), Vol. 2, pp. 888-892, Singapore, November 13-17, 1995

[Muth00] Muthukrishnan R. et al.: "Design, Implementation and Evaluation of an Explicit Rate Allocation Algorithm in an ATM Switch", Proc. of IEEE INFOCOM 2000, pp.1313-1322, Tel Aviv, March 26-30, 2000

[NaSi96] Narváez P., Siu K.-Y.: "Design of Feedback Control Algorithms for ABR Service in ATM", Technical Report 11-96, D'Arbeloff Laboratory for Information Systems and Technology, Massachusetts Institute of Technology (MIT), 1996

[NaSi97] Narváez P., Siu K.-Y.: "Optimal Feedback Control for ABR Service in ATM", Technical Report 1-97, D'Arbeloff Laboratory for Information Systems and Technology, Massachusetts Institute of Technology (MIT), 1997

[Narv97] Narváez Guarnieri, P. L.: "Design and Analysis of Flow Control Algorithms for Data Networks", Master Thesis, Massachusetts Institute of Technology (MIT), June 1997

[NuMa98] Nuno F., Matsumoto Y.: "VD/VS Coupling Scheme Using Feed-Forward Congestion Indication for ABR Services in Wireless ATM", IEICE Transactions on Communications, Vol. E81-B, No. 12, pp. 2392-2399, December 1998

[OzKI98] Özbay H., Kalyanaraman S., Iftar A.: "On Rate-Based Congestion Control in High Speed Networks: Design of an H-infinity Based Flow Controller for Single Bottleneck", Proc. of the American Control Conference, pp. 2376-2380, Philadelphia, June 24-26, 1998

[OKKI99] Özbay H., Kang T., Kalyanaraman S., Iftar A.: "Performance and Robustness Analysis of an H-infinity Based Flow Controller", Proc. of the 38th IEEE Conference on Decision and Control, pp. 2691-2696, Phoenix, December 7-10, 1999

[OzSV94] Özveren C. M., Simcoe R., Varghese G.: "Reliable and Efficient Hop-by-Hop Flow Control", Proc. of ACM SIGCOMM'94, Conference on Communications Architectures, Protocols and Applications, pp. 89-100, London, August 31 - September 2, 1994

[OzSV95] Özveren C. M., Simcoe R., Varghese G.: "Reliable and Efficient Hop-by-Hop Flow Control", IEEE Journal on Selected Areas in Communications, Vol. 13, No. 4, pp. 642-650, May 1995

[Ohsa95a] Ohsaki H. et al.: "Analysis of Rate-Based Congestion Control Algorithms for ATM Networks, Part 1: Steady-State Analysis", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 95), Vol. 1, pp. 296-303, Singapore, November 13-17, 1995

[Ohsa95b] Ohsaki H. et al.: "Analysis of Rate-Based Congestion Control Algorithms for ATM Networks, Part 2: Initial Transient State Analysis", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 95), Vol. 2, pp. 1095-1101, Singapore, November 13-17, 1995

[Ohsa95c] Ohsaki H. et al.: "Performance Evaluation of Rate-Based Congestion Control Algorithms in Multimedia ATM Networks", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 95), Vol. 2, pp. 1243-1248, Singapore, November 13-17, 1995

[Ohsa95d] Ohsaki H. et al.: "Rate-Based Congestion Control for ATM Networks", ACM Computer Communication Review, Vol. 25, No. 2, pp. 60-72, April 1995

[OMUM99] Ohsaki H., Murata M., Ushio T., Miyahara H.: "A Control Theoretical Approach to a window-based Flow Control Mechanism with Explicit Congestion Notification", Proc. of the 38th IEEE Conference on Decision and Control, pp. 2715-2720, Phoenix, December 7-10, 1999

[Oppe89] Oppenheim A. V.: "Discrete-Time Signal Processing", Prentice Hall, 1989

[Part94] Partridge C.: "Gigabit Networking", Addison-Wesley, 1994

[Paxs97] Paxson V.: "Fast, Approximative Synthesis of Fractional Gaussian Noise for Generating Self-Similar Network Traffic", ACM Computer Communication Review, Vol. 27, No. 5, pp. 5-18, October 1997

[PaFl95] Paxson V., Floyd S.: "Wide-Area Traffic: The Failure of Poisson Modelling", IEEE/ACM Transactions on Networking, Vol. 3, No. 3, pp. 226-244, June 1995

[PaAg95] Pazhyannur R. S., Agrawal R.: "Feedback-Based Flow Control of B-ISDN/ATM Networks", IEEE Journal on Selected Areas in Communications, Vol. 13, No. 7, pp. 1252-1266, September 1995

[PaAg97] Pazhyannur R. S., Agrawal R.: "Rate Based Flow Control with Delayed Feedback in Integrated Services Networks", Technical Report ECE-97-4, ECE Dept., University of Wisconsin - Madison, July 1997

[PeJa99] Petrovic S., Jamalipour A.: "A Novel Approach to Congestion Control in ATM Wide Area Networks", Proc. of IEEE ATM Workshop, pp. 385-390, Kochi, May 24-27, 1999

[PhHa91] Phillips C. L., Harbor R. D.: "Feedback Control Systems", Prentice Hall, 1991

[PiLa97] Pitsillides A., Lambert J.: "Adaptive Congestion Control in ATM Based Networks: Quality of Service and High Utilization", Computer Communications, Vol. 20, Issue 14, pp. 1239-1258, December 1997

[PiSR97] Pitsillides A., Sekercioglu Y. A., Ramamurthy G.: "Effective Control of Traffic Flow in ATM Networks Using Fuzzy Explicit Rate Marking (FERM)", IEEE Journal on Selected Areas in Communications, Vol. 15, No. 2, pp. 209-225, February 1997

[PiSR95] Pitsillides A., Sekercioglu Y. A., Ramamurthy G.: "Fuzzy Backward Congestion Notification (FBCN) Congestion Control in Asynchronous Transfer Mode (ATM) Networks", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 95), Vol. 1, pp. 280-285, Singapore, November 13-17, 1995

[PlSy97] Plotkin N. T., Sydie J. J.: "The Rate Mismatch Problem in Heterogeneous ABR Flow Control", Proc. of IEEE INFOCOM'97, pp. 1304-1314, Kobe, April 7-11, 1997

[PrKB98] Prasad S., Kiasaleh K., Balsara P.: "LAPLUS: An Efficient, Effective and Stable Switch Algorithm for Flow Control of the Available Bit Rate ATM Service", Proc. of IEEE INFOCOM'98, pp. 174-182, San Francisco, March 29 - April 2, 1998

[RaNe95] Ramakrishnan K. K., Newman P.: "Integration of Rate and Credit Schemes for ATM Flow Control", IEEE Network, Vol. 9, No. 2, pp. 49-56, March/April 1995

[RaQi95] Ramamurthy G., Qiang R.: "Analysis of the Adaptive Rate Control for ABR Service in ATM Networks", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 95), Vol. 2, pp. 1083-1088, Singapore, November 13-17, 1995

[ReHE99] Rejaie R., Handley M., Estrin D.: "RAP: An End-to-end Rate-based Congestion Control Mechanism for Realtime Streams in the Internet", Proc. of IEEE INFOCOM'99, pp. 1337-45, New York, March 21-25, 1999

[Ritt97a] Ritter M.: "Analysis of a Queueing Model with Delayed Feedback and Its Application to the ABR Flow Control", Technical Report No. 164, University of Würzburg, Institute of Computer Science, Research Report Series, January 1997

[Ritt97b] Ritter M.: "Congestion Detection Methods and their Impact on the Performance of the ABR Flow Control", Proc. of the 15th International Teletraffic Congress, pp. 1107-1118, Washington, June 23-27, 1997

[Robe94] Roberts L.: "Enhanced PRCA (Proportional Rate-Control Algorithm)", ATM Forum Contribution, 94-0735R1, ATM Systems, August 11, 1994

[RoBO95] Rohrs C. E., Berry R. A., O'Halek S. J.: "A Control Engineer's Look at ATM Congestion Avoidance", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 95),Vol. 2, pp. 1089-1094, Singapore, November 13-17, 1995

[RoBe97] Rohrs C. E., Berry R. A.: "A Linear Control Approach to Explicit Rate Feedback in ATM Networks", Proc. of IEEE INFOCOM'97, pp. 277-282, Kobe, April 7-11, 1997

[RoFl95] Romanow A., Floyd S.: "Dynamics of TCP Traffic over ATM Networks", IEEE Journal on Selected Areas in Communications, Vol. 13, No. 4, pp. 633-641, May 1995

[RoMa99] Ros D., Marie R.: "Loss Characterization in High-Speed Networks Through Simulation of Fluid-Models", Proc. of the 1999 Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS 99, pp. 251-260, Chicago, July 11-15, 1999

[RyEl96] Ryu B. K., Elwalid A.: "The Importance of Long-Range Dependence of VBR Video Traffic in ATM Traffic Engineering: Myths and Realities", Computer Communication Review, Vol. 26, No. 4, pp. 3-14, October 1996

[SaMe96] Sackett G. C., Metz C. Y.: "ATM and Multiprotocol Networking", McGraw-Hill, 1996

[Sait96] Saito H. et al: "Performance Issues in Public ABR Service", IEEE Communications Magazine, Vol. 34, No. 11, pp. 40-48, November 1996

[ShMa98] Shen X., Mark J. W.: "Explicit Rate Controller Design for ABR Services: An H-infinity Optimization Approach", Proc. of the IEEE 48th Annual International Vehicular Technology Conference, VTC'98, pp. 2416-2420, Ottawa, May 18-21, 1998

[Shen90] Shenker S.: "A Theoretical Analysis of Feedback Flow Control", Proc. of ACM SIGCOMM'90, Symposium on Communications Architectures & Protocols, pp. 156-165, Philadelphia, September 26-28, 1990

[Shio00] Shiomoto K. et al.: "Scalable Multi-QoS IP + ATM Switch Router Architecture", IEEE Communications Magazine, Vol. 38, No. 12, pp. 86-92, December 2000

[SiRo94] Simcoe R., Roberts L.: "The Great Debate over ATM Congestion Control", Data Communications, pp.76-80, September 21, 1994

[Simc94] Simcoe R.: "Test Configurations for Fairness and Other Tests", ATM Forum Contribution, TM 94-0557, Digital Equipment Corporation, July 1994

[SiJa95] Siu K.-Y., Jain R.: "A Brief Overview of ATM: Protocol Layers, LAN Emulation, and Traffic Management", ACM Computer Communication Review, Vol. 25, No. 2, pp. 6-20, April 1995

[SiTz94] Siu K.-Y., Tzeng H.-Y.: "Intelligent Congestion Control for ABR Service in ATM Networks", ACM Computer Communication Review, Vol. 24, No. 5, pp. 81-106, October 1994

[SmAT96] Smith A., Adams J., Tagg G.: "Available Bit Rate - a New Service for ATM", Computer Networks and ISDN Systems, Vol. 28, Issue 5, pp. 635-640, March 1996

[Soum99] Soumiya T. et al.: "The Large Capacity ATM Backbone Switch "Fetex-150 ESP", Computer Networks, Vol. 31, Issue 6, pp. 603-615, March 1999

[SpHo95] Spaniol O., Hoff S.: "Ereignis-orientierte Simulation", Thomson Publishing, 1995

[Srik99] Srikant R.: "Control of Communications Networks", Perspectives in Control Engineering - Technologies, Applications, New Directions, Samad T. (Ed.), IEEE Press, 1999

[Stal98] Stallings W.: "High Speed Networks: TCP/IP and ATM Design Principles", Prentice Hall, 1998

[Stal99] Stallings W.: "Data & Computer Communications", Sixth Edition, Prentice Hall, 1999

[Stev94] Stevens W. R.: "TCP/IP Illustrated, Vol. 1, The Protocols", Addison-Wesley, 1994

[SuVW00] Su C.F., de Veciana G., Walrand J.: "Explicit Rate Flow Control for ABR Services in ATM Networks", IEEE /ACM Transactions on Networking, Vol. 8, No. 3, pp. 350-361, June 2000

[TiSu90] Tipper D., Sundareshan M. K.: "Numerical Methods for Modeling Computer Networks under Nonstationary Conditions", IEEE Journal on Selected Areas in Communications, Vol. 8, No. 9, pp. 1682-1695, December 1990

[TsHK98] Tsai W. K., Hu L. C., Kim Y.: "A Temporal-Spatial Flow Control Protocol for ABR in Integrated Networks", Interactive Distributed Multimedia Systems and Telecommunication Services, 5th International Workshop, IDMS'98, Plagemann T., Goebel V. (Eds.), Springer, LNCS 1483, pp. 207-218, Oslo, September 8-11, 1998

[TKCT98] Tsai W. K., Kim Y., Chiussi F., Toh C.-K.: "Stability and Sensitivity of ABR Flow Control Protocols", Proc. of SPIE Conference on Performance and Control of Network Systems, pp. 43-54, Boston, November 1-6, 1998

[TzSi97] Tzeng H.-Y., Siu K.-Y.: "On Max-Min Fair Congestion Control for Multicast ABR Service in ATM", IEEE Journal on Selected Areas in Communications, Vol. 15, No. 3, pp. 545-556, April 1997

[TzSi98] Tzeng H.-Y., Siu K.-Y.: "Performance of TCP over UBR in ATM with EPD and virtual queuing techniques", Computer Communications, Vol. 21, Issue 12, pp. 1070-1079, August 1998

[Vand99] Vandalore B. et al.: "Overload Based Explicit Rate Switch Schemes with MCR Guarantees", Proc. of IEEE International Conference on Computer and Communications Networks (IC3N'99), p. 655-661, Boston, October 11-13, 1999

[Vand00] Vandalore B. et al.: "General Weighted Fairness and Its Support in Explicit Rate Switch Algorithms", Computer Communications, Vol. 23, Issue 2, pp. 149-161, January 2000

[VJGF98] Vandalore B., Jain R., Goyal R., Fahmy S.: "Design and Analysis of Queue Control Functions for Explicit Rate Switch Schemes", Proc. of IEEE International Conference on Computer and Communications Networks (IC3N'98), pp. 780-786, Lafayette, October 12-15, 1998

[VJGF99] Vandalore B., Jain R., Goyal R., Fahmy S.: "Dynamic Queue Control Functions for ATM ABR Switch Schemes: Design and Analysis", Computer Networks, Vol. 31, Issue 18, pp. 1935-1949, August 1999

[WeHW98] Wei K. L., Hwang C. C., Wen J. H.: "A Proportional Feedback Scheme for ATM Networks", Information Sciences, Vol. 110, Issue 3-4, pp. 237-253, October 1998

[WoBo98] Wong M. K., Bonomi F.: "A Novel Explicit Rate Congestion Control Algorithm", Proc. of the SPIE International Symposium Voice, Video & Data Communications, Conference on Performance and Control of Network Systems II, pp. 31-42, Boston, November 2-4, 1998

[Woo98] Woo T.-K.: "A Predictive Approach for ABR Traffic Management in ATM Networks", Computer Communications, Vol. 21, Issue 1, pp. 81-87, February 1998

[YaRe95] Yang C.-Q., Reddy A. V. S.: "A Taxonomy for Congestion Control Algorithms in Packet Switch Networks", IEEE Network, Vol. 9, No. 4, pp. 34-45, July/August 1995

[Yin95] Yin N.: "Analysis of a Rate-Based Traffic Management Mechanism for ABR Service", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 95), Vol. 2, pp. 1076-1082, Singapore, November 13-17, 1995

[Zach96] Zachary J. L.: "Introduction to Scientific Programming: Computational Problem Solving Using Maple and C", Springer 1996

[ZhYM97a] Zhang H., Yang O.W., Mouftah H.: "Design of Robust Congestion Controllers for ATM Networks", Proc. of IEEE INFOCOM'97, pp. 302-309, Kobe, April 7-11, 1997

[ZhYM97b] Zhang H., Yang O.W., Mouftah H.: "The Hop-by-Hop Flow Controller for High Speed Networks: Single VC Case", Proc. of IEEE Global Telecommunications Conference (GLOBECOM 97), Vol. 2, pp. 785-789, Phoenix, November 4-8, 1997

[ZhYM00] Zhang H., Yang O.W., Mouftah H.: "A Hop-by-Hop Flow Controller for a Virtual Path", Computer Networks, Vol. 32, Issue 1, pp. 99-119, January 2000

[ZhLS97] Zhao Y., Li S. Q., Sigarto S.: "A Linear Dynamic Model for Design of Stable Explicit-Rate ABR Control Schemes", Proc. of IEEE INFOCOM'97, pp. 283-292, Kobe, April 7-11, 1997

[ZhLi96] Zhao Y., Li S. Q.: "Feedback Control of Multiloop ABR Traffic in Presence of CBR/VBR Traffic Transmission", Proc. of IEEE International Conference on Communications (ICC'96), pp. 1717-1721, Dallas, June 23-27, 1996

D.2 Standards and Recommendations

[BICI2.0] ATM Forum, Technical Committee: "B-ISDN Inter Carrier Interface (B-ICI) Specification", Version 2.0, AF-BICI-0013.003, December 1995

[BICI2.1] ATM Forum, Technical Committee: "Addendum to B-ISDN Inter Carrier Interface (B-ICI) Specification, (B-ICI Specification, v2.1)", AF-BICI-0068.000, November 1996

[PNNI1.0] ATM Forum, Technical Committee: "Private Network-Network Interface Specification", Version 1.0, AF-PNNI-0055.000, March 1996

[UNI3.1] ATM Forum, Technical Committee: "ATM User-Network Interface Specification", Version 3.1, September, 1994

[UNI4.0] ATM Forum, Technical Committee: "ATM User-Network Interface (UNI) Signalling Specification", Version 4.0, AF-SIG-0061.000, July 1996

[NNIABR] ATM Forum, Technical Committee: "Addendum to PNNI V1.0 for ABR Parameter Negotiation", AF-PNNI-0075.000, January 1997

[UNIABR] ATM Forum, Technical Committee: "Addendum to UNI Signalling V4.0 for ABR Parameter Negotiation", AF-SIG-0076.000, January 1997

[TM4.1] ATM Forum, Technical Committee: "Traffic Management Specification", Version 4.1, AF-TM-0121.000, March 1999

[ADTM] ATM Forum, Technical Committee: "Addendum to Traffic Management V4.1 for an Optional Minimum Desired Cell Rate Indication for UBR", AF-TM-0150.000, July 2000

[AINI] ATM Forum, Technical Committee: "ATM Inter-Network Interface (AINI) Specification", AF-CS-0125.000, July 1999

[G.703] CCITT: "Recommendation G.703: Physical/Electrical Characteristics of Hierarchical Digital Interfaces", Geneva, April 1991

[G.705] CCITT: "Recommendation G.705: Characteristics Required to Terminate Digital Links on a Digital Exchange", Geneva, November 1988, withdrawn October 1996

[I.121] CCITT: "Recommendation I.121: Broadband Aspects of ISDN", Geneva, April 1991

[I13818-2] ISO/IEC 13818-2:2000: "Information Technology - Generic Coding of Moving Pictures and Associated Audio Information: Video", Geneva, December 2000

[I.311] ITU-T: "Recommendation I.311: B-ISDN General Network Aspects", Geneva, August 1996

[I.321] CCITT: "Recommendation I.321: B-ISDN Protocol Reference Model and Its Application", Geneva, April 1991

[I.356] ITU-T: "Recommendation I.356: B-ISDN ATM Layer Cell Transfer Performance", Geneva, October 1996

[I.362] ITU-T: "Recommendation I.362: B-ISDN ATM Adaptation Layer (AAL) Functional Description", Geneva, cancelled June 1997

[I.363] ITU-T: "Recommendation I.363: B-ISDN ATM Adaptation Layer (AAL) Specification", Geneva, March 1993

[I.363.1] ITU-T: "Recommendation I.363.1: B-ISDN ATM Adaptation Layer Specification: Type 1 AAL", Geneva, August 1996

[I.363.2] ITU-T: "Recommendation I.363.2: B-ISDN ATM Adaptation Layer Specification: Type 2 AAL", Geneva, September 1997

[I.363.3] ITU-T: "Recommendation I.363.3: B-ISDN ATM Adaptation Layer Specification: Type 3/4 AAL", Geneva, August 1996

[I.363.5] ITU-T: "Recommendation I.363.5: B-ISDN ATM Adaptation Layer Specification: Type 5 AAL", Geneva, August 1996

[I.371] ITU-T: "Recommendation I.371: Traffic Control and Congestion Control in B-ISDN", Geneva, March 2000

[I.371.1] ITU-T: "Recommendation I.371.1: Traffic Control and Congestion Control in B-ISDN: Conformance Definitions for ABT and ABR", Geneva, June 1997

[I.432] ITU-T: "Recommendation I.432: B-ISDN User-Network Interface - Physical Layer Specification", Geneva, March 1993

[Q.700] ITU-T: "Recommendation Q.700: Introduction to CCITT Signalling System No. 7", Geneva, March 1993

[Q.931] ITU-T: "Recommendation Q.931 (05/98) - ISDN User-Network Interface Layer 3 Specification for Basic Call Control", Geneva, May 1998

[Q.2931] ITU-T: "Recommendation Q.2931: Broadband Integrated Services Digital Network (B-ISDN) - Digital Subscriber Signalling System No. 2 (DSS 2) - User-Network Interface (UNI) - Layer 3 Specification for Basic Call/Connection Control", Geneva, February 1995

[X.200] ITU-T: "Recommendation X.200: Data Networks and Open System Communications, Open Systems Interconnection - Model and Notation, Information Technology - Open Systems Interconnection - Basic Reference Model: The Basic Model", Geneva, July 1994

[RFC2001] Stevens W.: "TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms", IETF, Network Working Group, RFC 2001, January 1997

[RFC2357] Mankin A., Romanow A., Bradner S., Paxson V.: "IETF Criteria for Evaluating Reliable Multicast Transport and Application Protocols", IETF, Network Working Group, RFC 2357, June 1998

[RFC2481] Ramakrishnan K. K., Floyd S.: "A Proposal to Add Explicit Congestion Notification (ECN) to IP", IETF, Network Working Group, RFC 2481, January 1999

Curriculum Vitae

Christian Virgil Cseh

May 26, 1967	born in Wiesbaden-Sonnenberg
1973 - 1977	Geschwister Scholl primary school in Wiesbaden
1977 - 1980	Leibniz secondary school in Wiesbaden
1980 - 1986	Schiller secondary school in Cologne
June 1986	Abitur
1986 - 1988	Social service at Kuratorium für Dialyse und Nierentransplantation in Cologne
1988 - 1994	Studies in Computer Science at the University of Technology Aachen
October 1994	Diploma in Computer Science
1995	Research Assistant with the German Aerospace Center (DLR) in Göttingen
1995 - 2001	Research Assistant with the Department of Computer Science (Informatik 4) at the University of Technology Aachen and scholar of the Graduiertenkolleg Informatik und Technik
since September 2001	Researcher with DaimlerChrysler AG, Research and Technology in Ulm