TRAFFIC CHARACTERISATION AND CONNECTION ADMISSION CONTROL IN ATM NETWORKS

Kalevi Kilkki

Kalevi Kilkki: Traffic Characterisation and Connection Admission Control in ATM Networks

ABSTRACT

The greatly variable requirements of different applications, particularly those of video and data sources, make high demands on the development of traffic control in ATM networks. A control scheme in ATM networks should take into account the limited knowledge of traffic behaviour, the additional expenses due to complexity and the profit achieved by high utilisation. The purpose of this study was to enhance the knowledge of the traffic process and by that means to develop efficient methods for Connection Admission Control in ATM networks. The main tools in the development of traffic models were in two divisions. Firstly, traffic variations can be classified into three time scales: cell scale with short term fluctuations, burst scale with intermediate fluctuations, and rate-variation scale with long term fluctuations that cannot be buffered in ATM nodes. Secondly, it is possible to separate the traffic models in homogeneous cases from those of heterogeneous cases. As regards the heterogeneous approximations it transpired that each type of traffic variation has a corresponding simple approximate model. There are good mathematical reasons to apply the effective bandwidth model at cell scale and a model using the variance of bit rate distribution at rate-variation scale. Burst scale processes, especially cases with fluctuations at several time scales, are much more difficult to model. A combination of effective bandwidth and effective variance (EBV-model) gives a simple and efficient solution to this problem. It is possible to use these three models as a basis of CAC procedure by introducing regulating parameters by which the required Quality of Service can be achieved. The implementation of EBV provides the opportunity for creating a very flexible scheme for Connection Admission Control in ATM networks.

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to Prof. Kauko Rahko for his invaluable support during the last twelve years. His lively views on traffic theory has greatly influenced the models and methods used in this thesis.

I became acquainted with ATM in the Telecommunications Switching Laboratory at Helsinki University of Technology during 1988-90 in a research project funded by Nokia Research Centre. I would like to thank Tapio Erke and Reijo Juvonen for introducing me to the inspiring area of traffic problems in ATM networks.

The main part of the thesis has been carried out at the Telecom Research Centre at Telecom Finland during the years 1990-94. I would like to express my thanks to Kari Kolu, Kari Nyman, Tapio Vaarnamo and other colleagues at Telecom Finland for giving me the change to do such interesting research work in a pleasant environment.

International projects have been of great importance in my work. EURESCOM P105 project has deepened my understanding of many aspects of ATM thanks to the project leader Pierre Adam and other distinguished experts at ATM. Special thanks are reserved for Anne Mette Møller and Richard Wade for several discussions which clarified the properties of Connection Admission Control methods.

The project chairman of COST 242 project, Jim Roberts, has created a stimulating environment for discussing the various topics of broadband networks. Many contributions, especially those by Karl Lindberger and Ilkka Norros, have promoted the development of the traffic models used in the thesis. I am indebted to Jorma Virtamo for his critical comments and valuable advice, in particular, concerning the presentation of effective bandwidth models.

Finally, the patience and support of my family, Eija, Olli and Juho, has been vital in accomplishing this work.

Espoo, June 1994

Kalevi Kilkki

CONTENTS

LIST OF ABBREVIATIONS

ADPCM	Adaptive Differential Pulse Code Modulation
ATM	Asynchronous Transfer Mode
BECN	Backward Explicit Congestion Notification
CAC	Connection Admission Control
CBR	Constant Bit Rate
CLP	Cell Loss Priority
COST	European Cooperation in the Field of Scientific and Technical Research
EB	Effective Bandwidth
EBV	A combination of Effective Bandwidth and Effective Variance
EFCI	Explicit Forward Congestion Indication
EV	Effective Variance
FBM	Fractional Brownian Motion
FIFO	First in First out
FRM	Fast Resource Management
FRP	Fast Reservation Protocol
FRP/DT	FRP, Delayed Transmission
FRP/IT	FRP, Immediate Transmission
GD	Gaussian Distribution
HDTV	High Definition Television
ISDN	Integrated Services Digital Network
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication Standardization Sector
KF	Kelly's Formula
LAN	Local Area Networks
LD	Large Deviation
LF	Lindberger's Formula
LM	Link Metric
MMDP	Markov Modulated Deterministic Process
MMPP	Markov Modulated Poisson Process
NPC	Network Parameter Control
NRM	Network Resource Management
PCM	Pulse Code Modulation
PR	Peak Rate
QoS	Quality of Service
TD	Traffic Descriptor
UPC	Usage Parameter Control
VBR	Variable Bit Rate
VP	Virtual Path

The notations of source and network parameters are presented at the beginnings of Chapters 3 and 4.

1 INTRODUCTION

Asynchronous Transfer Mode (ATM) is the basis for future high-speed telecommunication networks. The principle of ATM has proved usable in a wide range of networks from small local specialised networks to huge global integrated networks. The strength of ATM lies in its superior flexibility which enables a wide variety of services and applications to be efficiently integrated in one network.

At an early stage of development ATM was called Asynchronous Time-Division. This name clarifies a basic principle of ATM: all services or connections can share network resources in an asynchronous manner without any fixed reservation. Each connection can use the capacity of links, switches and buffers exactly when needed, and if for a while there is no information to be transferred, all capacity is left to other connections. On the other hand, when a number of applications compete for the same resources, the competition needs fair and efficient rules.

From the customer point of view, the main aspects for assessing telecommunication networks are Quality of Service (QoS) and price. A network operator may attempt to meet these two targets at the same time, although they are opposed. A low price can be achieved by a high exploitation of network resources whereas a low utilisation usually means high Quality of Service to the customers. A suitable traffic control strategy is the means by which the operator can satisfy both targets.

There are three extreme strategies for controlling a telecommunication network. The first one is to use the simplest possible control method and to keep the network utilisation so low that the probability of contention between different connections is very small. This strategy is typical in Local Area Networks (LANs) both at the lowest level (the network capacity is shared by competing packet flows) and at the highest level (new terminals, servers and printers are added until some user complains about poor QoS).

The other extremity is to regulate all connections so strictly that no conflict can occur during the connection. In a way, this is the principle of telephone networks since a telephone call reserves a permanent amount of resources during the call and competition occurs only when a customer tries to establish a new call. However, these two approaches are somewhat inconsistent with the principles of ATM networks, in which the operator attempts to maximise the utilisation by taking into account statistical behaviour of traffic streams. As a third strategy an operator may attempt to obtain maximum utilisation without a deteriorating Quality of Service by using an extremely complicated control architecture.

The optimum strategy in real networks is situated somewhere between these extremities; it takes into account the restriction of knowledge of traffic behaviour, the additional expenses due to complexity, and the profit achieved by a high utilisation. A solid knowledge of traffic characteristics and traffic behaviour inside the network is needed in order to find the optimum solution. The main purpose of this study is to satisfy this need.

The underlying structure of this study is depicted in Figure 1.1. The study rests on two bases: the knowledge of the characteristics of real traffic offered to ATM networks and the mathematical tools that can be used in analysing the behaviour of aggregated traffic process. A huge amount of research work has been done in the area of traffic and

queuing theory and many excellent textbooks exist, such as (Kleinrock 1975). Traffic and queuing theories are certainly useful in the analysis of ATM traffic but there is a strong demand for developing and evaluating special models of ATM traffic because of the special properties of ATM networks. A good example of this research work is COST 224 project *Performance evaluation and design of multiservice networks* (Roberts 1992a). In this study, Chapter 3 provides an insight into the particular tools for ATM traffic evaluation including a simulation program.

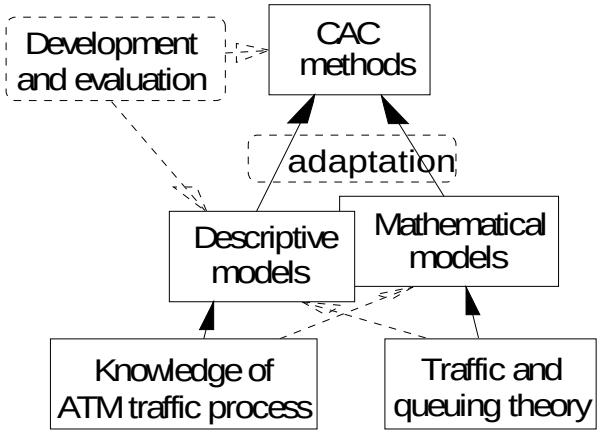

Figure 1.1. A structure for developing Connection Admission Control (CAC) methods.

The other half of the basis of evaluation is in many ways less certain because until now there has been very little experience of real traffic in ATM networks. Virtually all knowledge of traffic characteristics has been acquired from more or less separate cases, for example by measuring the characteristics of video sources or traffic in Local Area Networks. Although this type of information is helpful, combining all these in one solid model forms a real challenge for research work.

The first step in research work is to represent the results of measurement and other data in a compact form by means of descriptive models. The difference between descriptive models and mathematical models is to some degree indefinite. The main difference is that by mathematical models we attempt to achieve accuracy even at the expense of simplicity and comprehensibility whereas with descriptive models these properties are of great importance. The viewpoint of this study is mostly that of descriptive models.

Descriptive models can be divided into two groups, direct and derived models, depending on whether information about the underlying transmission and switching network is necessary. An example of a parameter used in a direct model is the variance of cell rate distribution while all models that take into account link capacity or buffer size belong to the other group. A review on the direct models and parameters are presented in Section 4.1.

The Connection Admission Control (CAC) is a set of actions taken by the network at the call set up phase in order to establish whether a connection can be accepted or rejected. Both mathematical and descriptive models are needed in the development of CAC procedures. The main requirements for the traffic models used in CAC procedures are accuracy, simplicity and general applicability. Several models meet two of these requirements but there has been hardly any that meets all of them. Section 4.2 tries to fill this gap with the aid of three models: effective bandwidth, effective variance and a combination of each.

It turns out that each of these models is especially suitable for describing a certain type of traffic variation. On the basis of this regularity two new parameters, the utilisation factor and the multiplexing factor, are introduced. On one hand, these parameters can be used as a concise way to characterise traffic sources but on the other hand they have a practical application since they determine which traffic model and which CAC method is practicable with a given traffic process.

Although a wide variety of CAC methods can be found in literature, no single method has reached a general agreement among the ATM traffic experts. In order to make a fair comparison of the methods, an adequate framework is needed. The framework presented in Section 5.1 is founded on a separation between the determination of source parameters (derived mainly from homogeneous case) and the combination of different source types (approximation of heterogeneous cases). The focus in this study is on the latter problem. The basic approaches are the same as those used to describe the ATM traffic process, namely, effective bandwidth, effective variance and a combination of both. The formulation of each model allows the use of any analytical or approximate method, or even simulation results in the determination of source parameters.

The performance evaluation of CAC methods should cover various traffic processes in order to find out the overall behaviour and weaknesses of approximations . An extensive comparison is presented in Section 5.4. Another significant issue to be considered is the simplicity of implementation. Although there is no obvious way to weigh the importance of various aspects, such as high utilisation and different parts of implementation, Section 5.5 endeavours to offer practical directions for the selection of an efficient CAC method.

Unfortunately, the reality is more complicated than the underlying model that has been used in the formal evaluation of CAC methods. Section 5.6 deals with two aspects which are essential for the realisation of the CAC method: the uncertainty of traffic parameters and the relationship between CAC and other traffic control functions. Finally, two important special cases, interconnection of Local Area Networks and Variable Bit Rate (VBR) video, have been analysed with respect to traffic control in ATM networks.

2 ASYNCHRONOUS TRANSFER MODE

2.1 ATM principles

The first article about the principles of ATM appeared eleven years ago (Coudreuse 1983). According to Coudreuse (1991) the basic target of ATM was to meet the challenge that changing telecommunications needs posed to the techniques and technologies of information transport. Although at present ATM is almost unanimously accepted as the basis for broadband networks, there have been other alternatives. The two basic alternatives are packet networks and digital networks based on the synchronous transfer mode. Let us discuss briefly the weaknesses of these alternatives in order to clarify the strength of ATM in an integrated services environment.

In a typical packet network one connection reserves the whole link during the transmission of a packet. As a result all other connections must wait until the transmission is ended. This is a substantial disadvantage in a multiservice environment because some applications, especially voice, are sensitive to delay. Although priorities can be used to alleviate this problem, it is usually not possible to break off the transmission of a packet in order to send more urgent packets.

Synchronous digital networks, particularly narrow-band ISDN (Integrated Services Digital Network), offer another alternative to ATM. In the ISDN the basic unit is 64 kbit/s channel, which is without doubt suitable for vocal communications. However, the somewhat inflexible structure of ISDN has a number of disadvantages as regards other services. The channel is reserved all the time irrespective of the actual capacity needs of the transmission. In addition to 64 kbit/s ISDN sustains only a few other rates such as 2 Mbit/s (or 1.5 Mbit/s). If a service uses more than one 64 kbit/s channel, different channels are routed through the network independently and it is difficult to guarantee that all channels have an equal delay.

The fundamental difference between ISDN and ATM is that instead of fixed speed ATM network uses fixed packets, called cells. The size of an ATM cell is 53 bytes (424 bits) of which 5 bytes are used for header and 48 for user information. Although the basic unit in ATM technology is the amount of information, there are standardised bit rates for ATM interfaces, namely, 155 and 622 Mbit/s. These bit rates and the cell size determine the time units of ATM networks, 2.73 μs and 0.68 μs, which are the transmission times of one cell at 155 Mbit/s and at 622 Mbit/s, respectively. Nevertheless, it should be noted that ATM principle can be adapted for any other bit rate.

Flexibility was given priority at an early stage of ATM development partly because of inability to predict service demand. Flexibility is achieved by an intrinsic property of ATM: all types of information (voice, data, video and still picture) are presented in the same form using equal-sized cells. There has been, on the other hand, considerable suspicion as to the viability of the integration of services with very different characteristics: Voice and video are intrinsically analogue signals; A typical data source produces variable length packets according to almost an unpredictable process; Still pictures, such as X-ray pictures, may have a huge amount of information that should be delivered through the network in a couple of seconds.

The application of a relatively small information unit results in excessive segmentations and re-assemblies, especially for data flows, and consequently additional headers and computation and possibly an increased probability of information loss. These are the main disadvantages of combining in one network voice and video services with data services. Despite these problems, the integration principle seems to be better than using separate networks for different applications, and ATM seems to be the most efficient technique to combine various traffic flows into the same network (Figure 2.1).

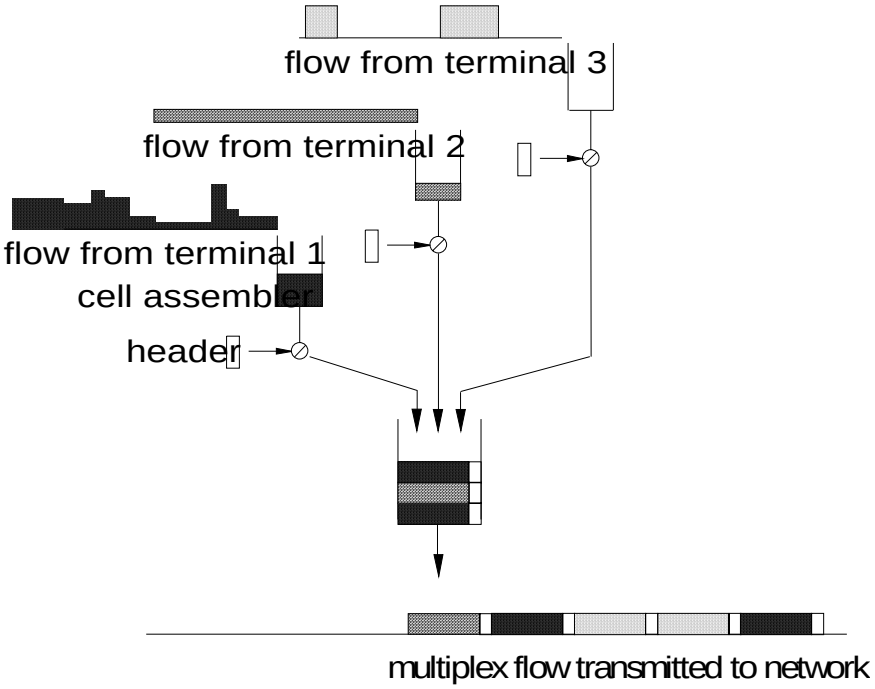

Figure 2.1. The multiplex principle of ATM networks.

In ATM networks both segmentation and buffering have the ability to alter the properties of the traffic stream. If the original traffic stream has any correlation with previous events or with other connections, the segmentation in ATM interface does not change or changes only slightly the characteristics of correlation. If all data packets are split into cells and then sent to the network one after another, the buffer requirement inside the ATM network is in principle the same as in a pure data network without segmentation. Nevertheless, there is a considerable change because the segmentation offers the *opportunity* to alter traffic flows more precisely. This is perhaps the main advantage of the ATM network as compared with packet networks. Although the ATM cell is a kind of packet, the cell size is so small that the additional delay due to transmission of one cell is negligible and even the emptying time of a full buffer is usually short in comparison with the propagation delay.

The effect of buffering depends on the buffer capacity. In this study the basic assumption is that the actual buffer capacity in ATM networks is relatively small, typically 100 cells, and if bigger buffers are necessary, they are situated outside the core network. As regards traffic analysis, the alteration effect of large buffers can be taken into account in the incoming traffic process. Because of the small buffers, the delay at network nodes does not have a substantial effect on most applications and, in consequence, the cell loss probability is usually the factor that sets a limit on the network utilisation.

2.2 Time resolution

The overall traffic process in ATM networks will be extremely complicated. One way to facilitate the analysing of ATM traffic is to divide the traffic process into several levels each of which has its typical time scale and typical traffic characteristics. In this study a resolution of four time scales is applied:

- In *call scale* traffic variations are caused by the call process. These variations are usually managed by means of Connection Admission Control.
- *Rate-variation scale* includes the variations induced by the changes in required cell rate, for example in a Variable Bit Rate video or audio connection. This scale covers typically the region from 20 ms to minutes.
- In *burst scale* the inherent phenomenon is the arrival process of bursts, such as arrivals of packets from Local Area Networks. This time scale covers typically the region from 0.1 ms to 100 ms.
- In *cell scale* each connection can be supposed to be deterministic (the interarrival time between successive cells is constant) and thus the variations in aggregated traffic process are due to randomness of phases of different connections. The time scale of these variations is approximately from 1 μs to 1 ms.

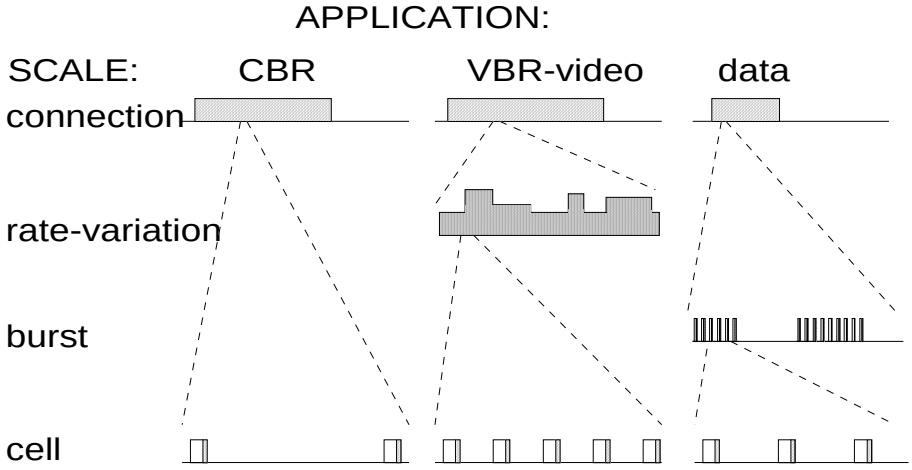

Figure 2.2. Time resolution of ATM traffic process.

A time resolution with three, four or even more scales has been applied widely (see Aagesen 1993; Castelli, Cavallero & Tonietti 1991; Heegaard & Helvik 1993; Hui, Gursoy, Moayeri & Yates 1991). There are no problems to determine and name the connection and cell scales whereas the situation is much more difficult with the intermediate scales. The names used in this study try to depict the inherent characteristics of traffic process at each scale.

A burst is interpreted as a block of information which has a certain size but not necessary a tight requirement for the peak rate during bursts (except that the transmission of a burst should end before the arrival of the next burst). In contrast, in rate-variation scale there is typically no definite amount of information but a required level for the average cell rate. This scheme differs to some extent from those in most

references, for example in (Hui 1991 et al.) and (Roberts 1992a) all traffic processes in burst and rate-variation scales are covered by one (burst) scale. However, this unification of all traffic processes between cell and call scales is unsuitable for use in this study.

The time resolution proposed by Aagesen (1993) is similar to the resolution in Figure 2.2; the most important difference is that the frame scale (from 1 ms to 1 s) and average scale (from 1 s to 1000 s) of Aagesen are united in the rate-variation scale in this study. The reason for this is that if the buffer size is relatively small, as expected in this study, the same traffic models can be applied to the whole region from 10 ms to minutes.

The traffic analysis in this study covers mainly the three lowest scales (cell, burst and rate-variation) but not the variations induced by the call process. The call process is the prime phenomenon as regards the network dimensioning, and in some cases it may be very difficult to distinguish the call scale process from the processes of other scales. A typical situation is when a Virtual Path (VP) cross connect network is used to transmit LAN traffic. Virtual Paths are typically permanent and the traffic tends to vary considerably during the holding time of a VP connection because virtual connections are established and released. As the virtual connections are transparent for a VP network, it is possible that a CAC procedure has to deal with traffic variations of a relatively long time scale. In traffic analysis these variations can be included in rate-variation scale models.

The aggregated process including cell, burst and rate-variation scale fluctuations is extremely difficult to analyse mathematically. However, a regular behaviour can be found independent of the actual traffic parameters in each scale (see Figure 2.3). Several studies (e.g., Kröner 1991; Norros, Roberts, Simonian & Virtamo 1991; Rasmussen, Sørensen, Kvols & Jacobsen 1991) have shown that, when VBR connections are aggregated on a multiplexer, the queue length distribution is composed of two distinct components. In this study a model with three components has been applied. The additional rate-variation scale component (horizontal line in Figure 2.3) arises where the input rate is permanently greater than the output rate. The burst scale component (the middle component in Figure 2.3) is due to relatively short bursts which can be partly buffered even by the small buffers of ATM nodes. The cell scale component (the steepest line in Figure 2.3) reflects the small queues which occur due to the asynchronous arrival of cells from distinct connections.

The queue components are not necessarily as clear with real traffic processes. There are regions in which the two scales are effective at the same time (rounded edges). Furthermore, since the overload periods are infinitely long only in theoretical models not in practical situations, the rate-variation scale component is horizontal only in theory and the boundary between the burst and rate-variation scales may entirely vanish.

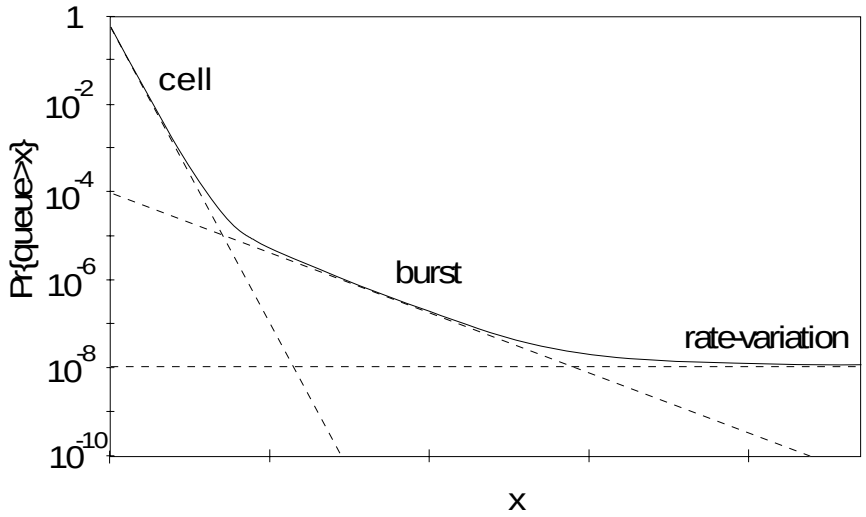

Figure 2.3. Cell, burst and rate-variation scale components of queue length distribution.

2.3 Traffic control and congestion control

2.3.1 The challenge of traffic control

The principle of ATM itself guarantees neither high utilisation nor high Quality of Service without traffic control. Congestion in its various forms is the basic problem of traffic control in the ordinary telephone network, in packet network as well as in the ATM network. Congestion occurs when the demand is greater than the available resources. According to Jain (1990) congestion is caused in packet networks:

- by a shortage of a buffer space,
- by slow links or
- by slow processors, and

this may lead to a belief that, when some or all of these problems are solved by technical development (cheap memory, high speed links and processors), the congestion problem goes away. Contrary to this belief, without proper protocol redesign, technical development may lead to more congestion and thus reduce performance (Jain 1990). This is indisputably the situation in ATM networks as well, and all over the world there is a vast effort to develop proper control methods for ATM. To quote Gilbert, Aboul-Magd and Phung (1991): the challenge is to design simple and efficient controls while still achieving reasonable bandwidth utilisation through a statistical multiplexing.

Recently there have been some proposals for complicated control architecture (see e.g., Hyman, Lazar & Pacifici 1993; Roberts 1993b; Sriram 1993). The basic idea in those proposals is that several classes of traffic with different QoS requirements are considered explicitly at every level of system design, both at the edge and at the core of the network. Therefore the network should be able to allocate the buffer capacity according to the actual requirement of each connection, not on the First in First out (FIFO) basis as in traditional control scheme of ATM networks.

This study follows the main line in standardisation and supposes that the separation of different services, if applied, is done by higher protocol layers and no parallel buffers at

the core of ATM network are used (with the possible exception of separate buffers for high and low priority flows).

2.3.2 Definitions

This section depicts the role of traffic control and congestion control as they have been defined in recommendation I.371 of International Telecommunication Union, Telecommunication Standardization Sector (ITU-T 1993a). The *primary* role of traffic control and congestion control is to protect the network and the user in order to achieve network performance objectives. An *additional* role is to optimise the use of network resources.

Traffic control refers to the set of actions taken by the network to avoid congested conditions. Congestion control refers to the actions taken by the network to minimise the intensity, spread and duration of congestion. Congestion is defined as a state of network elements in which the network is not able to meet the network performance objectives. It is to be distinguished from the state where buffer overflow is causing cell losses, but still meets the negotiated Quality of Service.

Traffic control functions are (ITU-T 1993a):

- Network Resource Management (NRM): Allocation of network resources in order to separate traffic flows according to service characteristics.
- Connection Admission Control (CAC): A set of actions taken by the network during the call set up phase in order to establish whether a virtual channel (or path) request can be accepted or rejected.
- Usage/Network Parameter Control (UPC/NPC): A set of actions taken by the network to monitor and control traffic, in terms of traffic offered and validity of the ATM connection. The main purpose of UPC is to protect network resources from malicious as well as unintentional misuse.
- Priority control: the user may generate different priority traffic flows by using the Cell Loss Priority bit. A congested network element may selectively discard cells with low priority if necessary.
- Traffic shaping is a mechanism that alters the traffic characteristics of a stream of cells to achieve a desired modification of those characteristics. Examples of traffic shaping are peak cell rate reduction and burst length limiting.
- Fast Resource Management (FRM): A typical FRM function allows a network to allocate capacity for the duration of a burst in response to a user request.

Congestion control functions are:

- Selective cell discarding: A congested network element may selectively discard cells identified as belonging to a non-compliant ATM connection and/or those cells with lower Cell Loss Priority.
- Explicit Forward Congestion Indication (EFCI) may be used to assist the network in avoidance of and recovery from a congested state. A network element in a congested state may set an Explicit Forward Congestion

Indication in the cell header so that this indication may be examined by the destination customer equipment.

As regards NRM, the service separation has an important effect on CAC because the multiplexing process is more regular if the traffic characteristics of aggregated streams, such as peak rate and burst length, resemble each other. Furthermore, it might be possible to use simpler CAC methods if sources are grouped into few service classes. These service classes may have various cell loss requirements in which case the network utilisation can be improved provided that different services use different links. However, the main reason to introduce multiple QoS classes is that they can be used to protect higher priority flows against cell loss during periods of short term traffic overflow.

According to Eckberg, Lucantoni and Prasanna (1991) there are two issues that must be addressed with respect to the Cell Loss Priority (CLP) indicator:

- the QoS given to the CLP=0 [higher priority] traffic must be only insignificantly affected by the CLP=1 traffic and
- some utility from the CLP=1 traffic must be derivable by the end-terminals.

In the first condition the analysis of CLP=0 traffic, which is the chief concern of this study, is almost independent of the CLP=1 traffic flow. Similar procedures that are used with CLP=0 traffic can be applied for the CLP=1 traffic (or for the combined traffic) using different QoS parameters.

The principle of ATM makes it possible to change the traffic stream before multiplexing mainly in order to increase the utilisation of network links, in particular when the burstiness of offered traffic is very high (e.g., Roberts 1993a). The usefulness of this approach depends on the time-scale of variations and on the delay requirements of application. As regards the performance evaluation the effect of traffic shaping can be included in the offered traffic process.

The idea of FRM is to increase the multiplexing efficiency by implementing admission control at the burst scale (or at the rate-variation scale) in addition to the connection scale. When a new burst is to be sent it is necessary to obtain a new resource allocation by means of a rapid in-band signalling exchange between user and successive network nodes. The bandwidth used by a connection is relinquished at the end of a burst. According to Roberts (1992b) the limitations of this approach are the time needed to obtain a new resource allocation which reduces efficiency particularly for short bursts, the need to implement a sophisticated protocol and the low network utilisation realisable when the connection peak rate is high. If a network node rejects the burst, it can be either buffered at the network interface or discarded depending on the application. Buffering is unavoidable if the application is file transfer whereas with real time applications, such as voice and video, it is not sensible to buffer bursts for re-transmission. In both cases the rejecting probability should be reasonably low to avoid enormous buffers or degradation of QoS.

There are two Fast Reservation Protocols (FRP) for the realisation of FRM:

- *Delayed Transmission* (FRP/DT) is based on the prior negotiation and reservation of a peak rate value on each switching node along the connection using special management cells (Tranchier, Boyer, Rouaud & Mazeas 1992);

- *Immediate Transmission* (FRP/IT) supposes that link capacity can be reserved "on the fly" by the first cells of a burst when it arrives in each switching node and on each link of its path (Roberts 1993a).

In the case of blocking, special procedures would be necessary to inform the user to allow him to make a new attempt.

Congestion indication can be send backwards by using Backward Explicit Congestion Notification (BECN). When a queue in an ATM switch exceeds a certain threshold it sends BECN cells back to the sources of virtual channels currently submitting traffic to it (Newman 1993). On receipt of a BECN cell to a particular virtual channel, a source must reduce its transmission rate for the indicated channel. If no BECN cells are received for a certain period of time, a source may gradually restore its transmission rate. According to Newman BECN could be applicable for high-burst sources without specifying traffic characteristic for every individual data source when the transmission delay is limited, as in LAN, but considerable problems might arise if the network's size is large (a good performance level might be extended to a transmission delay of several hundred kilometres).

2.3.3 Preventive vs. reactive control

There are two basic approaches for controlling broadband networks: preventive and reactive. The preventive approach relies mainly on traffic control functions while the reactive approach utilises primarily congestion control functions. The basic idea of reactive or feedback control is that the network allows the offered traffic increase until the capacity of a link is exceeded or, in a more advanced case, until some network element detects that an overload situation is probable.

To quote Blaabjerg (1991): In Europe the trend has been towards a simple and preventive strategy, based on the ideas from traditional telecommunications community whereas in the US a trend towards a more dynamic strategy based on ideas from computer communication community is seen. A good compromise, as Ramamurthy and Dighe (1991) have proposed, would be an aggressive congestion avoidance strategy that uses network resources optimally, with reactive control mechanisms as backups to relieve congestion in the unlikely event of the network experiencing congestion.

Feedback control has been proved to be useful in data networks where sources are suitable for cell rate re-allocation, buffers in network nodes are typically large and bit rates are not very high. Unfortunately, the situation is almost the reverse in a typical ATM network because it will be very hard to re-allocate most sources, buffers are small and bit rates are very high. An illustrative description of these fundamental problems of high speed networks can be found in Kleinrock (1992).

ATM networks, particularly in large areas, are dependent on the capabilities of preventive control methods, but feedback control functions can still be useful in minimising the intensity and duration of congestion. In addition, reactive functions may have an important role when exploiting the free capacity in ATM networks. Because of the statistical properties of traffic in ATM networks the mean load of high priority traffic may remain low, even less than 0.1. Network operators may attempt to utilise the remaining capacity by offering it to customers who have a large amount of data to be transferred but can tolerate occasional long delays (see Section 5.6.2). In addition, traffic sources should be able to reduce the bit rate because even a large buffer will overflow quite soon if the link is fully reserved by high priority flows.

2.3.4 Response times

The response time defines how quickly the controls react. Figure 2.4 shows a typical classification of control functions according to the response time. A clear similarity can be seen between Figure 2.2 (Time resolution of ATM traffic process) and Figure 2.4: the *connection scale* and *cell scale* in Figure 2.2 correspond respectively to *connection duration* and *cell time* in Figure 2.4.

There is an obvious relation between *rate-variation* and *burst scales* in Figure 2.2 and *round trip propagation time* in Figure 2.4 although those time levels are based on two different phenomena, the properties of traffic offered and the properties of the network. The round-trip delay in a wide area ATM-network may vary from 1 ms to 100 ms. This is the typical time scale of the arrival process of bursts and it also partly covers the rate-variation scale fluctuations. These similarities in time scales are important but they do not entirely explain the complicated relationship between offered traffic and feedback control functions (FRM, EFCI, BECN); we should take into account many other aspects, such as delay and delay variation requirements, upper layer protocols, and limited buffer capacity.

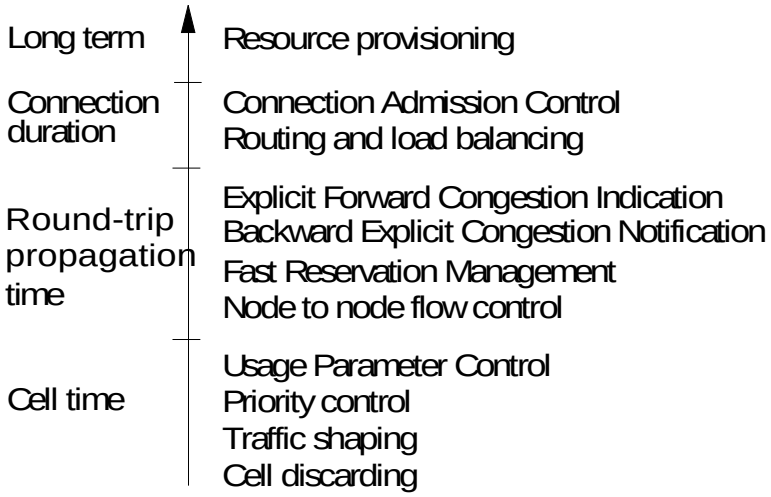

Figure 2.4. Control response times (ITU-T 1993a; Gilbert et al. 1991).

2.4 Service types and requirements

This section describes the basic properties of various service types in broadband networks. Services can be classified into five main groups: circuit emulation, voice, video, data and multimedia. Each service type has its inherent requirements for ATM networks.

2.4.1 Circuit emulation

The basic idea of circuit emulation is to hide the ATM nodes and links so that the flexible ATM technology can be brought inside the present telecommunication infrastructure with as few changes as possible. A typical situation is an operator who wants to offer switched N*64 kbit/s connections for business customers. In the present telephone network, managing these connections is a difficult task whereas in ATM networks the operator can control connections flexibly and offer transmission capacity immediate by means of ATM crossconnects. From the ATM network point of view,

circuit emulation connection is a Constant Bit Rate (CBR) source with strict requirement for cell delay variation.

According to many authors (see e.g., Decina & Toniatti 1990) a VBR connection should be interpreted as a CBR source determined by the peak rate if the ratio of peak rate to link rate is greater than 1/15 or 1/20. Therefore the actual bit rate from a source determined as a CBR connection is not necessarily always the same as the declared peak rate.

2.4.2 Voice

Although voice communication is frequently considered an insignificant service for broadband networks, it should not be totally ignored. A typical telephone conversation generates more than 10 Mbit information in both directions and, for example, the amount of information transferred by the Finnish long distance telephone network is roughly 10 Gbit per year and inhabitant. Some data applications, such as remote use of supercomputers, can generate perhaps 1000 times as much information during a year but, on the other hand, applications of this type will only be exploited by a few specialists.

In fact, voice communication has been taken into account in the standardisation of ATM. The cell size is partly determined by the requirements of voice connections because the larger cell size, the longer it takes to gather up a whole cell from the bit stream. This delay is 6 ms for a 64 kbit/s connection and if there are several ATM parts in the path of the connection, these delays, together with the propagation delay, may have a disturbing effect on a telephone connection. For the same reason, large buffers at network nodes are not recommendable. On the other hand, voice connection is usually less sensitive to cell losses than video or data applications.

There is much knowledge of the general behaviour of traffic in telephone networks. However, the situation in ATM network differs from that of the ordinary telephone network since it is possible to adapt to the varying bit rate demands during conversation. Nowadays a telephone call uses a constant 64 kbit/s channel, but this is not what is really needed since both talkers are seldom talking at the same time and, in addition, there are clear pauses between successive words and sentences. Depending on how accurately the silence periods are detected, the proportion of active periods varies from 0.35 to 0.5 (Brady 1969; Sriram & Whitt 1986). If we take into account that the 64 kbit/s PCM (Pulse Code Modulation) coding can be replaced by the 32 kbit/s ADPCM (Adaptive Differential PCM) coding without deterioration in speech quality, the average needed bandwidth of a voice connection can be reduced to 12 kbit/s in ATM networks. Consequently, an ATM link with a capacity of 622 Mbit/s may transfer roughly 40 000 telephone calls simultaneously.

2.4.3 Video

In the long term, the most important type of service of broadband networks is presumably video communications (e.g., Lyons, Jensen & Hawker 1993). Video communications consist of a wide variety of services from slow rate videophones to High Definition Television (HDTV) and the required bit rate may vary from tens of kilobits to hundreds of megabits per second. In order to utilise network resources efficiently layered coding schemes have been suggested. The idea of layered coding is, according to Ramamurthy and Sengupta (1990):

Applications like broadcast video that require large bandwidth, may use layered coding and mark packets as essential and enhancement packets. Essential packets help to reproduce the basic picture at the receiver and keep the session intact, and hence have to be carried without loss. Enhancement packets enhance the quality of the picture, and can be dropped in the event of congestion in the network without disrupting the session.

Typical properties of video sources with VBR coding are:

- a sharp peak occurs when the scene changes but variations are relatively slight for the same scene (Bae & Suda 1991; Roberts, Guibert & Simonian 1991);
- the form of stationary distribution depends on the type of sequence (videophone, videoconference, entertainment) (Roberts et al.) and on the coding method (Kawashima & Saito 1990);
- the autocorrelation function decreases rapidly over the first few frames but the rate of decrease then slows down (Ramamurthy & Sengupta 1990; Roberts et al.);
- if burst scale traffic variations are buffered the necessary buffer capacity might become very large (Alparone, Argenti, Capriotti & Benelli 1992; Ramamurthy & Sengupta; Roberts et al.);
- if complicated coding methods are used, cell loss rates for the important data shall be very low, in the order of 10^{-10} , whereas it may be possible to tolerate a greater cell loss rate for the remainder (Roberts et al.);
- video phones and video conferences will require all their packets to be delivered without delay (Ramamurthy & Sengupta);
- a video connection is seldom used without voice and other service components; this can bring about correlation between different connections and complicate the traffic control.

Kawashima and Saito (1990) have presented a summary of video source models with three bit rate parameters: mean (m), standard deviation (σ) and maximum (h). Since these models are concerned with a wide range of sources from videophones to studio television, it is not reasonable to take direct averages from these figures, instead we can use parameters such as ratios of standard deviation to mean and mean to peak. From Table 3 in (Kawashima & Saito) we can obtain the following average values for these parameters:

$$\frac{1}{M}\sum_{i=1}^{M}\frac{\sigma_i}{m_i} = 0.36,$$

$$\frac{1}{M}\sum_{i=1}^{M}\frac{m_i}{h_i} = 0.41.$$

These figures can be obtained by the following bit rate (λ) distribution (the mean bit rate is 1 Mbit/s):

$\Pr\{\lambda = 0.86 \text{ Mbit/s}\} = 0.60,$
$\Pr\{\lambda = 1.00 \text{ Mbit/s}\} = 0.34,$
$\Pr\{\lambda = 2.41 \text{ Mbit/s}\} = 0.06.$

Although the real bit rate distribution may be much more complicated, these figures contain the essence of the video source: there is relatively stable behaviour most of the time (0.88 and 1 Mbit/s in the example) and intermittent periods with a substantially higher bit rate requirement (2.41 Mbit/s).

2.4.4 Data

One definition for data communication is that it consists of all possible applications which use a computer as terminal equipment, in fact, everything that is not voice or video is data (Roberts, Bensaou & Canetti 1992). Up to now it has been possible to distinguish data applications from voice and video, but the recent development of telecommunication services has blurred the edges between different service types. Let us take for example a video art library. A fraction of a video movie has been saved on a computer disk and then manipulated by means of a sophisticated program that changes essentially the original content of the video. Then the edited video is sent automatically to the network after a request from a customer who uses it as a part of a multimedia application. It is not at all clear whether the result is data, video or some other type of connection.

In this study we are dealing with ATM traffic and its characteristics (there are many other viewpoints but they are not considered here). Consequently, the primary issue is what requirement an application or user has, in particular, whether or not the bandwidth requirement at any given time is determined by the source. The prime issue in the previous example is therefore whether the video tape is played immediately or saved on disk and played afterwards. In the latter case the used bit rate during transmission may be low or high depending on the charging policy and network load at the time, and it may vary independently of the actual content of the original source. A connection with these properties should be classified as a data connection rather than a video one.

Similarly, the properties of a data connection may approach those of video. For example, if a designer utilises computer aided animation remotely by aid of a supercomputer, user requirements are similar to a typical video connection even though no real video camera has been used. Thus the demands an application makes on the network are more important than the type of terminal. This must be taken into account in the source description: it is not possible to classify all sources to pre-defined groups according to the terminal type and other permanent information because the user application is in many cases more important than the terminal type.

In consequence, there is no typical data connection and no single feature appropriate to every data connection but, nevertheless, there are some typical characteristics for most data traffic:

- the bit rate needed can be very high but usually the peak rate is used only over a small fraction of time and thus the mean rate is much lower than peak rate (Doshi, Dravida, Johri & Ramamurthy 1991; Roberts 1992b);
- long bursts of information are interspersed with short messages (e.g., acknowledgements) (Doshi et al.; Roberts);
- unknown and unpredictable on- and off-period statistics (Roberts);

- loss tolerance depends on the coding scheme (Doshi et al.);
- sources are controllable in the sense that they can be slowed down, if needed, without affecting the viability of service (Doshi et al.).

There are two fundamentally different types of situation with respect to the traffic control in ATM networks:

- *individual connections*, which means that the original properties of connections are visible to the ATM network;
- *aggregated process*, typically between Local Area Networks, when the ATM network have little, if any, knowledge of the actual connections used by different applications.

According to Doshi et al. (1991) the former group can be divided into three different data types:

- *Relatively smooth data* comes from sources for which the cell arrival process is not as periodic as for CBR sources but the ratio of mean rate to peak rate is relatively high, say ≥ 0.1, and the bursts at peak rate are relatively short and nearly constant.
- *Bursty interactive traffic* and short intermittent file transfer are characterised by a relatively small value of the mean rate to peak rate requirement (could be < 0.01), and data bursts at peak rate ranging from a few bytes to a few hundreds of kilobytes.
- *Bursty long file transfers* correspond to long infrequent file transfer. Typically, the idle periods between such file transfer are much longer than the time to transmit file, implying a small ratio of mean to peak bandwidth requirement. These sources are delay tolerant.

The packet length distribution depends on the application but typically it has clear peaks at the minimum and maximum packet sizes (e.g., Drakopoulos 1993). According to Falaki and Sørensen (1992) the best fit to the interarrival time distribution on a local area computer network is provided by a hyperexponential distribution with two contributing terms: 68% of the traffic has a mean interarrival interval of 25.2 ms and the remaining 32% has a significantly larger mean interarrival interval of 235.2 ms. A more complex but basically similar model with hyperexponential distribution has been presented by Heegaard and Helvik (1993).

Data connection models may be complicated but much more difficult is to determine general LAN interconnection traffic because essential information (what the main properties of connections are, when they start and end and so forth) is either uncertain or unknown. This problem of modelling LAN traffic is dealt with further in Section 4.1.7.

2.4.5 Multimedia

A multimedia call may consist of audio, video and data components, and traffic control can treat these components as separate connections. However, problems may arise because of the interdependence of separate connections inside a multimedia call, for example between audio and video components. Unfortunately, almost every model for the aggregate traffic process relies on the assumption that different connections are independent of each other. This intricate problem needs further study but because we do

not have enough knowledge of real multimedia traffic and its interdependencies, the basic assumption in this study is that different connections are independent of each other.

2.4.6 Requirements for traffic models

From the results in previous sections we can infer that a wide range of models is needed for a proper analysis of ATM traffic. Applying the time resolution presented in Section 2.2 we can require that the following models are included:

- cell scale: deterministic sources for circuit emulation and other CBR sources, bit rates from 10 kbit/s to 100 Mbit/s;
- burst scale: periodic, bursty sources as models for worst case traffic and Markov models for uncontrolled data sources;
- rate-variation scale: models with three bit rate levels;
- combination: rate-variation scale modulation of burst scale models.

In the rest of this study these theoretical models are applied instead of specified models of video, voice or data sources. At the same time we now leave *the knowledge of ATM traffic process* box in Figure 1.1 and move onto *the mathematical models* box.

3 TOOLS FOR QOS EVALUATION

In this chapter we describe the available tools for the determination of cell loss probability and other QoS parameters by using the separation of time scales. We attempt to answer the following questions: how random is each time scale and what are the typical traffic models and their solution techniques. This set of tools is the starting point for the introduction of new description models and CAC methods presented in Chapters 4 and 5. The following notations have been applied (all terms are measured in cells, cells per second or time slots):

c = link capacity,

K = buffer size,

P_{loss} = cell loss probability,

P_{sat} = saturation probability,

h = peak rate (in cell and burst scales),

h_{rv} = peak rate in rate-variation scales
$= h\,p_{burst}$,

p_{burst} = *on* probability in the burst scale
$= L\,D_{cell}/D_{burst}$,

p_{rv} = *on* probability in the rate-variation scale,

L = mean burst size,

D (or D_{cell}) = distance between two consecutive cells during a burst
$= c/h$,

D_{burst} = distance between two consecutive bursts,

m = mean rate of a source
$= h\,p_{burst}\,p_{rv}$,

v = variance of cell rate distribution of a source,

λ_j = cell rate level j,

ρ = average load
$= \frac{1}{c}\sum_{i}^{N} m_i$,

N = the number of sources.

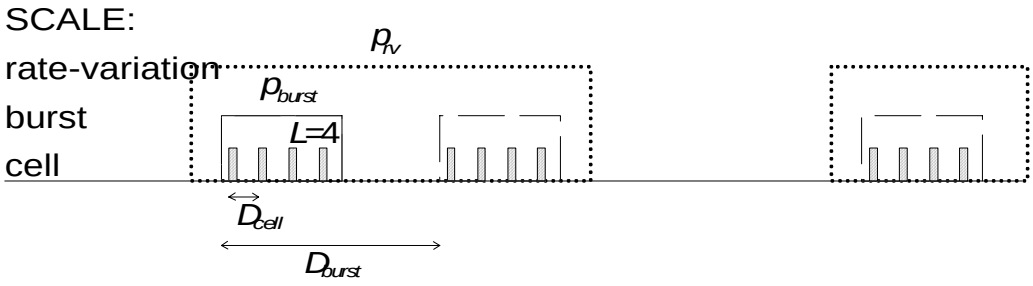

Figure 3.1. Definition of source parameters.

3.1 Cell scale

3.1.1 Models

In cell scale models we suppose that each connection is periodic (or deterministic) with respect to interarrival time of consecutive cells. Let us define the arrival time of a random cell of connection *i* by T_i, the period of source *j* by D_j and the arrival time of the first cell of connection *j* after T_i by T_j. The random variable $T_{i,j} = T_j - T_i$ is evenly distributed between 0 and D_j provided that connections *i* and *j* are independent of each other. The point is that all randomness in this system ($\Sigma D_i/D/1/K$) lies in the distribution of $T_{i,j}$. It can be said that this system is both realistic and possible to present in a pure mathematical form and, consequently, the result of analysis is both exact and applicable for practical purposes.

The independence requirement is valid evidently at the first multiplexing stage because the effect of dependency between connections is significant only if it occurs at time scales below 1 ms and this is quite unlikely (note that long term dependencies belong either to burst or rate-variation scale). A typical reason for dependency at the later multiplexing stages is that several connections to a network node come from the same link. However, this phenomenon has only a small effect which is even positive in so far as it decreases the probability of contentions of cells.

In homogeneous case we obtain a discrete time $N*D/D/1/K$ system and as a limit system when *D* grows to infinity and *N*/*D* remains constant, we obtain $M/D/1/K$ system. In this case the number of cells arriving in one time slot is Poisson distributed. In addition, we can take into account the limited number of input links since only one cell can arrive at each time slot from one input (if the link rates are equal for all input and output links). In this case we should replace the Poisson distribution by a binomial one (Geo/*D*/1/K system). Finally, if the traffic process consists of periodic input streams with different periods, we obtain a $\Sigma D_i/D/1/K$ system.

3.1.2 Solutions

An efficient technique to solve the above-mentioned problems is to use the Beneš formula (Beneš 1963; Roberts 1992a Section 5.3). The Beneš formula makes it possible in many cases to calculate the complementary distribution function of virtual waiting time, and what is perhaps more important, it makes it possible to derive near approximation even if the exact equation is very difficult to solve. The exact formulae for virtual waiting time for $M/D/1$, Geo/*D*/1 and $N*D/D/1$ systems can be found in Roberts (1992a). Because of the regular behaviour of these systems it is possible to obtain the queue length distribution for a finite buffer system and by that means the cell loss probability and other QoS parameters. An approximation for $\Sigma D_i/D/1/K$ system has been presented by Virtamo and Roberts (1989).

3.2 Burst scale

3.2.1 Requirements for traffic models

Burst scale models depict the behaviour of the traffic process that arises when variable length packets from data networks arrive at an ATM network. These packets should be

split into several ATM cells since the typical size of data packet is much larger than the size of an ATM cell. A very important issue is the speed at which the cells are delivered to the ATM network. Cells can be delivered one after another, or the cell stream can be smoothed by the aid of buffers. The first approach is usually advantageous for the implementation of packet/cell and cell/packet converters because this strategy minimises the need for buffers at the interfaces but at same time it maximises the buffer requirement inside the ATM network. The second approach has exactly the opposite effect on the buffer requirements. In a practical situation we have had to compromise between these two approaches but, unfortunately, this compromise has lead to the most laborious model as far as the performance evaluation is concerned.

The real traffic processes of data connections tend to be very complicated and many different approaches have been applied to catch the essence of traffic behaviour. We have to determine at least the packet length (or burst size) distribution and the interarrival time distribution of packets. However, this is not an adequate description of the real traffic process because in addition to this packet process there are usually long-term variations in the arrival process. In this study, these variations are matters of rate-variation scale and they are managed by the characteristic tools of that scale.

We must keep in mind that the traffic models should be solvable and, what is an even harder requirement, the parameters applied should be suitable for measuring and controlling in real implementations. Thus the models should be simple and preferably give upper bounds for cell loss probability.

The most common models applied are the deterministic and the Markov. In the deterministic model both burst size and interarrival time are constant. This model seems to be unfit for a description of data traffic but in ATM networks we should take into account the effect of traffic control (UPC and NPC) and the worst case traffic. The worst case traffic pattern that can go through a control device is typically a deterministic on/off process with a constant burst size and constant cell rate during a burst. As Aarstad (1993) and Doshi (1993) have shown, this assumption is not exactly true but for practical evaluation it is obviously an acceptable assumption because the worst patterns are more complicated and they are likely to appear only if some customers use them intentionally.

The original traffic behaviour of a typical data connection can be better reached by a Markov model than by deterministic one. In the simplest Markov model both burst size and interarrival time distribution are geometrically distributed. Although this model does not correspond precisely to the measurement results (see Section 2.4.4), it gives a more realistic picture of the real traffic process than the deterministic process, at least before a controlling unit (UPC). Furthermore, it is possible to use more complicated models in order to achieve more accurate results at the expense of the simplicity of solution.

3.2.2 Approximate models

The most common approximations for traffic in the burst scale are:

- Markov Modulated Poisson Process (MMPP);
- Markov Modulated Deterministic Process (MMDP);
- fluid flow models;
- diffusion approximation (Gaussian).

3.2.2.1 MMPP and MMDP

An MMPP is a Poisson process with an instantaneous arrival rate that varies according to the state of the continuous time Markov chain. Heffes and Lucantoni (1986) have used a 2-state MMPP to match four parameters of the superposition process: the mean arrival rate, the variance-to-mean ratio of the number of arrivals in $(0,t_1)$, the long term variance-to-mean ratio of number of arrivals, and the third moment of the number of arrivals in $(0,t_2)$. Slightly different approaches have been presented by Baiocchi, Melazzi, Listanti, Roveri and Winkler (1991), Okuda, Akimaru and Nagai (1992), and Sykas, Vlakos and Anerousis (1991). A more complicated, 4-state MMPP model has been employed by Yegenoglu and Jabbari (1993).

According to Norros et al. (1991) MMPP models can be criticised on two points: they do not accurately represent short-term correlation effects, and performance evaluation remains complex. As the approach in this study is to separate the calculation of homogeneous cases (or more generally, the calculation of traffic parameters of a single source) and heterogeneous cases, the question is whether MMPP is suitable for either cases. The problem in homogeneous cases is that MMPP as such is not very well suited to describe any typical source because of the Poisson process assumption in the cell scale. As regards heterogeneous approximations the MMPP model is too complicated for practical CAC implementations. For these reasons MMPP is not used in this study.

The difference between MMPP and MMDP is that the cell scale process of MMDP is deterministic. A deterministic model in the cell scale is usually a better approximation for a single source while the combined process of several sources can be modelled better by means of MMPP, in particular if the number of sources is great.

A closed-form solution can be obtained in one special case, namely when both burst and idle periods are exponentially distributed the necessary bandwidth of a single separate source is (Guérin, Ahmadi & Naghshineh 1991):

$$k_i = \frac{h_i}{2}\left(1 - x + \sqrt{(1 - x)^2 + 4 p_{burst,i} x}\right), \tag{3.1}$$

where: $$x = \frac{K}{-L(1 - p_{burst,i})\ln P_{loss}}.$$

3.2.2.2 Fluid flow

In fluid flow models the arrival rate fluctuations are accurately represented but the work to be accomplished by the server is assumed to arrive in a continuous flow rather than in discrete units (Bensaou, Guibert & Roberts 1990). Norros et al. (1991) have pointed out that the fluid flow model may be viewed as a way to calculate exactly the burst scale component of the real queue (see Figure 2.3). Akar and Arikan (1993) have presented an approximation that also captures the short term fluctuations of the queue length in the cell scale. However, the use of fluid models usually leads to a negligible overestimation of the load allowed (Roberts 1992c), and thus the short term (i.e., cell scale) fluctuations can be omitted.

A typical approach is to model the arrival rate as a Markov process. In this case we can write and solve equations for the stationary distribution for both homogeneous and heterogeneous cases (e.g., Blaabjerg 1991). Using the Beneš result, a more common

approach can be obtained (see Bensaou et al. 1990); however, the solution needs much computational effort due to numerical integration and, therefore, it is virtually unsuitable for real time implementation.

It can be easily seen that when the burst size substantially exceeds the buffer capacity, a fluid flow model is reduced to the bufferless model used in the rate-variation scale (e.g., Castelli et al. 1991). Although there is no exact boundary between burst and rate-variation scales, fluid flow (and other burst scale) models are needed only if the burst size is at most four times as large as the buffer size (this phenomenon is studied in Section 4.3).

3.2.2.3 Gaussian

Addie and Zukerman (1993) have modelled burst scale traffic streams using a stationary Gaussian process. The results of this model are based on three parameters of the aggregated input process: the mean, the variance and the autocovariance sum. A special property of this model is that only the peak rate is specified by the user and consequently there are strict requirements for the real time calculation of other parameters and this may be very difficult with the autocovariance sum. Another problem is the validity of Gaussian distribution as an input process approximation particularly when the number of connections is small.

3.3 Rate-variation scale

The intrinsic phenomenon in the rate-variation scale is the variation of needed bit rate. A typical example is a VBR video source in which the instant bit rate depends on the scene and on the motion of the camera. There are two important points with respect to traffic modelling:

- the variations are slow (one bit rate level can remain for several seconds) as compared with variations in the cell and burst scales;
- the variations can be predicted using general knowledge of statistical properties of various source types whereas the possibility of predicting the behaviour of an individual connection is limited (this holds even for the mean bit rate which can be estimated only approximately).

In the rate-variation scale the only important issue is whether at a certain instant there is enough link capacity for all sources. The effect of buffering can be ignored provided that the buffers are sufficiently large to cope with the cell scale variations; this seldom becomes a problem since the number of VBR sources is usually limited. The situation is much more complex if a VBR source has also considerable variations in the burst scale because it is not always evident which one of the variations, those of burst or rate-variation scale, has the larger influence on the allowable load. We return to this issue in Sections 3.4 and 4.3.4.

3.3.1 Exact solution

The rate-variation scale evaluation is easy in respect that the cell loss probability can be obtained by a simple formula (e.g., Roberts 1992a p. 150):

$$P_{loss} = \frac{\sum_{j:\lambda_j > c} \Pr\{\lambda = \lambda_j\}(\lambda_j - c)}{\rho c}, \qquad (3.2)$$

where $\Pr\{\lambda=\lambda_j\}$ is the probability that the aggregated process needs cell rate λ_j, c is the link capacity offered to the connections and ρ is the mean load. The main problem is to calculate or approximate the cell rate distribution $\Pr\{\lambda=\lambda_j\}$.

3.3.2 Approximations

Although (3.2) is exact if rate-variation scale assumptions are valid and $\Pr\{\lambda=\lambda_j\}$ distribution is calculable, the calculation becomes numerically difficult when the number of source types and the number of cell rate levels grow. The main approach to solving this problem are:

- to keep $\Pr\{\lambda=\lambda_j\}$ distribution as simple as possible by using a coarse granularity;
- to replace $\Pr\{\lambda=\lambda_j\}$ distribution by a simpler one;
- a large deviation approximation of $\Pr\{\lambda=\lambda_j\}$.

3.3.2.1 Convolution with limited granularity

The main benefits of the use of convolution are:

- the accuracy of results (provided that the calculation is based on exact source description), and
- the decentralised calculation of P_{loss}.

To achieve a reasonable implementation of convolution procedure it is inevitable to use a coarse granularity because the number of states of cell rate distribution is inversely proportional to the granularity unit. The peak cell rate may be defined as an integer variable of 3 octets (ITU-T 1993b Item 9) and consequently the granularity unit can be so small that the number of possible states becomes far too large for real time implementations.

Though it is possible to use a simple convolution method based on a limited observation period (see Section 5.2.4), it may be difficult to develop a convolution method that is both easy to implement and accurate for heterogeneous traffic. Moreover, the utilisation gain to be achieved by the convolution method in comparison with the most advanced CAC methods is so small that the use of convolution does not seem to be useful in practical implementations (see the results in Section 5.4.5).

3.3.2.2 Distribution approximations

The mean and variance of cell rate distribution can be calculated easily if sources are independent of each other. This leads to the idea of replacing the original distribution by a simpler one with the same parameters. The main candidates are *Gaussian, Poisson* and *binomial distributions*. Note that these distributions have been applied in the traditional teletraffic theory when calculating call and time congestion (see Rahko 1976; Rahko 1983). The rate-variation scale evaluation has many other points in common with teletraffic theory and the long-term knowledge acquired in that area can be utilised

in ATM traffic evaluation. On the other hand, there are differences as well. The main difference lies in desired blocking probabilities since classical teletraffic theory is dealing with probabilities in the order of 10^{-2} or 10^{-3} whereas in ATM networks the cell loss probabilities are typically in the order of 10^{-9}.

The benefit of Gaussian distribution is that it is wholly determined by mean and variance and these parameters are additive if the independence requirement is satisfied. A direct use of Gaussian distribution in (3.2) leads to a numerical calculation of Gaussian distribution. Poisson and binomial distributions can be applied in the same way (Uose, Shioda & Mase 1990).

A further approximation is to apply the saturation probability as a QoS requirement instead of cell loss probability. With Gaussian distribution this leads into a simple formula because the saturation probability depends only on a safety factor $\left(c - \sum m_i\right) \Big/ \sqrt{\sum v_i}$. The connection admission can then be obtained by the following formula:

$$\sum_i m_i + \kappa \sqrt{\sum_i v_i} \le c\,, \tag{3.3}$$

where κ depends only on the acceptable cell loss ratio. A feasible approximation for κ is (Guérin et al. 1991):

$$\kappa = \sqrt{-2 \ln P_{loss} - \ln(2\pi)}\,. \tag{3.4}$$

With a typical cell loss requirement 10^{-9} parameter κ is roughly 6.3.

Lindberger (1991) has developed an approximation for the bandwidth needed for rate-variation scale sources. The underlying idea of the approximation is that the original cell rate distribution can be replaced by a process composed of equivalent Poisson bursts. After some rearrangings and approximations Lindberger obtained a formula for the necessary bandwidth of a single source *i*:

$$k_i = a\left(m_i + b\frac{v_i}{c}\right), \tag{3.5}$$

where in most cases *a* and *b* depend only on P_{loss}:

$$a = 1 - \frac{\log P_{loss}}{50},$$

$$b = -6 \log P_{loss}.$$

In the complete formula *a* and *b* depend in some special cases on m_i, v_i and *c* (see Tidblom 1992). Formula (3.5) can be also used as a CAC method (see Section 5.2.1).

3.3.2.3 Large deviation approximation

The basic problem in the previous approximations is that the interesting region in the aggregated distribution is far from the mean and the relative error in approximation grows rapidly when the approximated probability decreases. The large deviation theory offers a excellent solution to this problem (see Bean 1993; Griffiths 1990; Hui 1990;

Kelly 1991). Firstly, it can be used to obtain an accurate estimate for small saturation probabilities:

$$P_{sat} = \Pr\{\lambda_t > c\},$$

when λ_t is composed of a number of independent streams. The idea is to shift the most accurate point of estimation from the region of original mean value to the interesting region of very small saturation probabilities. The shifted distribution can be approximated accurately by a Gaussian distribution around its own mean. The result is (Hui 1990 p. 206):

$$P_{sat} \approx \frac{1}{\sqrt{2\pi}\beta^*\sigma(\beta^*)} e^{-\beta^* c + \mu(\beta^*)}, \tag{3.6}$$

where $\mu(\beta) = \ln \mathrm{E}\{e^{\beta\lambda_t}\}$, $\sigma^2(\beta) = \mu''(\beta)$ and β is a free parameter by which we can ascertain in which region the approximation is best. Now we are interested in the region near the link capacity *c* and therefore the optimum value (β^*) can be obtained from the equation:

$$m(\beta^*) = c, \tag{3.7}$$

where *m(β)* is the expectation of the shifted distribution. The function *m(β)* has simple expressions for many distributions such as Poisson, binomial, exponential and Gaussian (see e.g., Roberts 1992a p. 109).

The same technique can be used to approximate the cell loss probability instead of saturation probability (Roberts 1992a p. 154):

$$P_{loss} \approx \frac{1}{\sqrt{2\pi} m\beta^{*2}\sigma(\beta^*)} e^{-\beta^* c + \mu(\beta^*)}, \tag{3.8}$$

which differs from (3.6) only by the appearance of an extra factor $m\beta^*$ in the denominator. This formula gives an excellent approximation for the cell loss probability as can be seen from the results presented in Section 5.4.1. The factor $m\beta^*$ is typically of the order 100 which means that P_{sat} criterion is roughly two orders of magnitude tighter than P_{loss} one (Roberts 1992a p. 154).

Let us take the simplest homogeneous case with on/off sources. Then we obtain from the binomial distribution:

$$\mu(\beta) = N \ln(1 - p_{rv} + p_{rv} e^{h\beta}), \tag{3.9}$$

where p_{rv} is the on-probability and *N* is the number of sources. Applying the basic formulae:

$$\begin{aligned} m(\beta) &= \mu'(\beta), \\ \sigma^2(\beta) &= \mu''(\beta), \end{aligned} \tag{3.10}$$

and combining (3.8) and (3.9), we obtain the following approximation for P_{loss}:

$$P_{loss} = \frac{\left(Nhp_{rv}/c\right)^{N-1} e^{\beta^*(Nh-c)}}{c^2 \beta^{*2} \sqrt{2\pi\left(h/c - 1/N\right)}}, \tag{3.11}$$

where $\beta^* = \frac{1}{h} \ln \frac{1 - p_{rv}}{\left(Nh/c - 1\right) p_{rv}}$.

In this simple example we can obtain a closed-form solution which is, however, rather complicated in comparison with (3.3), (3.4) and (3.5). Moreover, it is difficult to find any simpler approximation based on (3.11) because the first term in nominator is very small (e.g., 10^{-67}) while the second term is very large (e.g., 10^{60}). With more complicated traffic processes iterations are needed in order to obtain an optimum value for β^*.

Using the saturation probability as a QoS criterion and Chernoff's bound, Kelly has developed the following formula (Kelly 1991):

$$\frac{1}{\beta^* + \frac{\ln P_{loss}}{c}} \sum_i \mu_i\left(\beta^*\right) \le c. \tag{3.12}$$

Notwithstanding the linear form of (3.12), the optimum value of β^* depends on the traffic combination and so does $\mu_i\left(\beta^*\right)$. As Kelly has pointed out, β^* can be fixed according to a typical traffic mix but since β^* also fixes the maximum load with CBR sources, it can be applied in the same way as ρ_{max} in some other CAC formulae (see Section 5.3.2.5).

A substantial benefit of (3.12) is that it invariably guarantees the required cell loss probability. However, it should be noted that this proposition is true only statistically: (3.12) guarantees that the long term average of cell loss probability is smaller than the required value if all source parameters are exactly known but, unfortunately, the uncertainty of source parameters may cause larger errors in the traffic evaluation than any other reason (see Section 5.6.1).

3.4 Combination of different time scales

The real traffic process in ATM networks contains simultaneously properties from all time scales and thus it is necessary to combine the results attained in the previous sections. In practice, it is unlikely that the effect of cell scale fluctuations is equal to those of the rate-variation scale whereas the situation is not so clear when burst scale and rate-variation scale processes are concerned. We have three alternatives for the combination; they can be named the modulation, addition and separation approaches.

The basic idea of the modulation approach is that a deterministic process is modulated by rate-variation scale process (see e.g., Fuhrmann & Le Boudec 1991; Hübner & Tran-Gia 1991). Then the cell loss probability can be calculated separately for each cell rate level λ_j (or more generally for each combination of number of active sources) by methods presented in Section 3.1 and these probabilities can be added up:

$$P_{loss} = \frac{\sum_j \lambda_j \Pr\{\lambda = \lambda_j\} P_{loss}(\lambda_j)}{\sum_j \lambda_j \Pr\{\lambda = \lambda_j\}}. \quad (3.13)$$

In principle this is an easy and accurate technique to deal with the combination problem but for a real time implementation it is not suitable because the number of states in $P\{\lambda=\lambda_j\}$ distribution may be huge and the calculation of each single case may be quite difficult.

The modulation principle can be also applied to the combination of cell and burst scales. If a fluid flow model is used for the burst scale process, the cell scale component behaves like a $\Sigma D_i/D/1$ queue when the burst scale component of cell loss probability is zero, and it constitutes a small positive bias when the latter is positive. According to Norros et al. (1991) the expected value of this bias is approximately equal to the mean of a $\Sigma D_i/D/1$ queue with load equal to 1.

The second alternative is to calculate the cell loss probabilities separately for cell and rate-variation scales and add up these probabilities. To calculate the rate-variation scale process we can use the methods and approximations presented in Section 3.3 whereas the cell scale is more problematic because it is not clear which traffic model should be used. A typical choice is to calculate the saturation probability of the cell scale queue ($P_{sat,c}$) by $M/D/1/K$ system, and to calculate also a saturation probability for the rate-variation scale process ($P_{sat,rv}$). Then the upper bound for the combined saturation probability is according to Rasmussen et al. (1991):

$$P_{sat} = \alpha P_{sat,c} + P_{sat,rv}, \quad (3.14)$$

where α is a constant smaller than 1.65 for sources with peak to mean ratio larger than 2.

The last approach is based on complete separation of time scales:

- the buffer capacity is determined according to the cell scale fluctuations;
- the allowable load is determined according to rate-variation scale fluctuations.

In this case the allowable cell loss probability can be divided into cell scale and rate-variation scale parts by a constant α', $0 < \alpha' < 1$. The required cell loss probabilities are $\alpha' P_{req}$ and $(1-\alpha')P_{req}$ respectively for cell scale and rate-variation scale fluctuations (Miyao 1993).

3.5 General models

The main problem common to all approaches in the previous section is the vagueness between cell, burst and rate-variation scales in reality. In some cases it is very difficult to split the traffic process into two (or three) parts and analyse them separately, and by that means obtain satisfactory results for the combined traffic process.

In some recent studies different approaches have been applied in which all time scales have combined inside a model without any discrete boundaries between cell, burst and rate-variation scales. One of the main origin of these models is the measurement made in Local Area Networks, particularly those by Fowler and Leland (1991). The main

conclusion to be drawn from the measurements is that the traffic process has fluctuations at time scales from milliseconds to months and the properties of these fluctuations are similar at all time scales.

One possible approach for modelling traffic of this type is to use a Fractional Brownian Motion (FBM), $Z(t)$. $Z(t)$ has the following self-similarity property (Norros 1993):

> $Z(\alpha t),\ t \geq 0$, is identical in distribution to $\alpha^H Z(t),\ t \geq 0$, for every $\alpha > 0$.

If $H > 1/2$, the process is said to possess long-range dependence. We can use FBM in ATM traffic analysis for modelling the arrival process. The number of cells entering the multiplexer within the time interval (0,t] is (Norros 1993):

$$A(t) = mt + \sqrt{am}Z(t)\cdot \tag{3.15}$$

However, the traffic process in ATM networks may differ in many respects from those measured in LANs because the sophisticated methods for traffic control in ATM networks have the capability to restrict the traffic fluctuations at every time scale. A possible application is to use a control scheme with real-time traffic measurement if a ATM network connects LANs and there is no efficient CAC and UPC capabilities in LAN-ATM interface (see Section 5.2.5).

3.6 Tools used for analysis

3.6.1 Mathematical models

Most of the foregoing mathematical models have been implemented during this study, and though some of them are not used directly in the following sections, all presented models form the basis of the analysis in the following sections. The main traffic models applied in this study are:

- Cell scale: *M*/*D*/1/*K*;
  Geo/*D*/1/*K*;
  *N***D*/*D*/1/*K*.
- Burst scale: Fluid flow approximation,
  - heterogeneous case with Markov modulated on/off sources.
- Rate-variation scale: exact formula for cell loss probability (3.2);
  - sources with three different cell rate levels; three different sources simultaneously;

  Gaussian distribution approximation (3.3), (3.4);
  Lindberger's approximation (3.5);
  large deviation approximation (3.8),
  - homogeneous case, three different cell rate levels;

  Kelly's approximation (3.12).

All these models are implemented in a personal computer using the Pascal programming language. The numerical accuracy is sufficient to calculate cell loss probabilities of order 10^{-10}. A typical calculation time is a few seconds for cell scale

models, several minutes for fluid flow approximation and ten seconds for the exact formula of rate-variation scale and less than one second for other rate-variation scale models.

3.6.2 The simulation program and its accuracy

Although we have practicable mathematical tools for each separate time scale, there remains the difficult problem of the aggregated traffic process. The mathematical tools for burst scale are often numerically complicated and hence a straightforward combination of burst and rate-variation scale models quickly becomes unusable. One solution to these problems is to employ simulation tools, which have been in use since the sixties (see Rahko 1976). The main properties of the simulation tool used in this study are presented in Appendix B.

Simulation programs provide opportunities for analysing every traffic process that can be presented in a suitable form and, in addition, they are indispensable for the validation of mathematical models. The primary difficulty in applying simulation in performance evaluation of ATM traffic is that the important events, namely cell losses, should be very rare in ATM networks. Probabilities of the order of 10^{-9} are almost impossible to simulate with reasonable accuracy, and because we are interested in very complicated aggregated processes, it is difficult to use any analytical method to arrive at these probabilities.

In this study we have chosen a cell loss probability level of 10^{-4} for all simulations. We can suppose that the underlying phenomenon in the traffic process is similar to cases with smaller probabilities and, in addition, it is possible to attain sufficient accuracy for analysing purposes. In study the basic target is to obtain an error ratio of less than 10% for cell loss probabilities. However, it is not easy to conclude how long a simulation time is needed to reach this value because we should know the exact traffic process in order to determine the simulation accuracy. An important fact is that cell losses come in bursts (e.g., Virtamo & Norros 1991), and, accordingly, the basic unit in terms of the accuracy of simulation is not a single cell but a burst of lost cells. Therefore the main parameters of the cell loss process are the burst size and the interarrival time of bursts. In addition, the distribution of these parameters effects the accuracy of simulation.

In this study a conservative definition for a burst of lost cells has been applied: a burst of lost cells consists of all cells that have been lost during a traffic generation period on one output link. In this case we have every reason to believe that bursts of lost cells are independent of each other. The generation period of the program used in this study is typically from 1000 to 16000 time slots depending on the properties of the incoming traffic process (see Appendix B).

The other problem concerning burst size distribution is more problematic but according to the simulation results the standard deviation of the distribution usually equals the mean, which leads to an assumption of geometrical distribution. The only distinct exception is when:

- there are fluctuations of both the burst and rate-variation scale,
- these cause roughly the same amount of cell loss, and
- the average number of lost cells due to rate-variation scale fluctuations is substantially larger than that due to burst scale fluctuations.

In these situations the standard deviation to mean ratio is sometimes as high as two. Fortunately, we have accurate analytical approximations for rate-variation scale models and by combining analytical and simulation results it is possible to obtain sufficient accuracy during a reasonable simulation time.

Thus the basic question about the accuracy of simulation results is analogous to the accuracy of a traffic measurement in which the incoming traffic process is Poisson and call duration is exponentially distributed. The variance of measured mean traffic caused by the limited measuring period is Riordan (1951; also in Rahko & Hertzberg 1988):

$$\sigma^2\{A\} = \frac{2At_c}{T} + \frac{2At_c^2}{T^2}\left(e^{-T/t_c} - 1\right), \tag{3.16}$$

where: A = theoretical offered traffic
= number of lost cells in a time unit
= ρP_{loss},

T = the length of measurement (in time slots),

t_c = average holding time of calls
= average number of lost cells during a loss period.

Omitting the second term in (3.16), we obtain an approximation for the simulation error:

$$\sigma^2\{P_{loss}\} \approx \frac{2P_{loss}^2}{N_{lost}},$$

where: N_{lost} = the number of lost bursts
= $TP_{loss}\rho/t_c$.

In most simulations made in this study the number of loss bursts is at least 500. Then the standard deviation of P_{loss} is at most:

$$\sigma\{P_{loss}\} \approx 0.063P_{loss}.$$

If we suppose that the simulation error is normally distributed, the probability that the error of P_{loss} is larger than 10% is about 11%. This accuracy is necessary for the evaluation of some parameters characterising source behaviour, especially the multiplexing factor (presented in Section 4.2.4) is sensitive to the inaccuracy of simulation results. If $P_{loss} < 5\ 10^{-5}$, a looser requirement has been applied: the number of lost bursts should be more than $10^7 \cdot P_{loss}$. Although the error is then bigger in relation to P_{loss}, the absolute error is smaller than in the original case.

4 TRAFFIC CHARACTERISATION

As Figure 1.1 depicts, measurement results and other general knowledge of the behaviour of potential traffic in ATM networks form the basis of traffic analysis. This knowledge should be transformed into relatively simple mathematical models with a limited number of parameters in order to describe the inherent traffic behaviour. There has been an obvious lack of descriptive models both appropriate to all types of traffic process and capable of capturing the essence of traffic behaviour. Typically, a traffic model is intended to form a basis for developing mathematical formulae and it may lead to a practical solution, but only within the limits of the underlying traffic model. An example is large deviation approximation which is very useful for evaluating ATM traffic but only with traffic models at rate-variation scale. On the other hand, we have models based on index of dispersion and correlation which are suitable for all traffic processes but which are complicated and difficult to apply in performance evaluation.

As yet there has not been a simple way to describe ATM traffic sources with a few parameters from the performance evaluation point of view. Below we introduce two approaches. The first one, using the concepts *effective bandwidth* and *effective variance*, is suitable for traffic evaluation of a wide variety of traffic processes, and the second one is appropriate to describe any traffic source with the aid of two simple parameters, *the utilisation factor* and *the multiplexing factor*.

Before determining of these new concepts we review other descriptive models. These models can be grouped into two types: direct models that are independent of any network model, and derived models that require some information on network properties, such as link capacity and buffer size. The primary idea of derived models is that they attempt to depict the source behaviour in a typical traffic situation.

The main notations used in this chapter are (see also the beginning of Chapter 3):

$N_{c,i}$ = the allowed number of sources of type i in homogeneous case,

k_i = effective bandwidth of source i (EB_1 methods),

k_i^* = effective bandwidth of source i (EB_2 methods),

v_i^* = effective variance of source i,

$\rho_{hom,i}$ = the allowed load in homogeneous case

$$= \frac{m_i N_{c,i}}{c},$$

N_i = the number of sources of type i (in a certain traffic case),

$$\psi_i = \frac{N_i}{N_{c,i}},$$

$\psi_{max,i}$ = the maximum allowed ψ_i with EB_2 type of methods,

ρ_{cbr} = the load induced by CBR sources,

ρ_{max} = the maximum allowed load,

ε_u = utilisation factor,

ε_m = multiplexing factor.

4.1 Direct models and parameters

4.1.1 Source classes

Traffic characteristic may be declared directly or by means of predefined classes. According to Appleton (1991) the main reason for the use of predefined classes is that it makes the declaration of traffic characteristics user friendly, since the customer need only specify a class such as "video telephony" or "high speed data". The number of source types is the basic difficulty of this approach:

- on the one hand, if the traffic classes are used in Connection Admission Control without any numerical method, the number of classes should be very limited in order to keep the admission decision rule feasible (for instance, with ten traffic classes the number of possible combinations is enormous), and
- on the other hand, if there are only a few classes, a traffic class (e.g., "high speed data") may contain a wide variety of connections with various properties and there will unquestionably be connections that do not fall into by any predefined class.

Even though the network provides predefined traffic, in order to retain full service integration it must be capable of providing service to customers with exceptional traffic characteristics that do not fall into any traffic class (Appleton 1991). In this study we suppose that a numerical specification is always used either directly or through predefined traffic classes.

4.1.2 Controllable parameters

One of the most important requirements of traffic parameters is controllability, that is, the possibility of controlling traffic parameters efficiently. A good example of this approach can be found in (ITU-T 1993b Item 10) which defines in addition to peak cell rate two optional traffic parameters:

- sustainable cell rate and
- intrinsic burst tolerance.

These parameters can be used to determine needed bandwidth in a CAC method with a statistical multiplexing scheme or in a CAC procedure with peak rate allocation and FRM (Roberts 1993c). A further approach is to develop a resource allocation in which both bandwidth and buffer space are determined directly by these traffic parameters: a bandwidth equal to the sustainable cell rate is allocated on each link and an amount of memory based on burst tolerance is reserved in each multiplexing buffer (Roberts). This resource allocation guarantees a service without cell loss but at the same time it requires large buffer space and network nodes to operate complicated queue scheduling algorithms.

4.1.3 Rate-variation scale parameters

The traffic behaviour at rate-variation scale can be described by the cell rate distribution $\Pr\{\lambda = \lambda_j\}$. Cell rate distribution can describe both a single source and the combined traffic process, particularly if the sources are independent of each other. For practical purposes, such as a real implementation of CAC, a complete distribution is too

complicated, and consequently a simpler expression is needed. A typical choice for traffic parameters is:

- peak rate: $\mathrm{Max}\{\lambda_j\}$;
- mean rate: $\mathrm{E}\{\lambda_j\}$;
- variance: $\mathrm{Var}\{\lambda_j\}$.

Higher moments are usually avoided because they are not additive. When these three parameters are given, an on/off model is a good approximation to worst case traffic. The basic problem of controlling statistical parameters, in particular variance, is very difficult although some methods have been proposed (see Andrade & Villen 1993).

Cell rate distribution does not give any information related to the burst length and therefore it is not suitable for burst scale description. The traffic models that also take into account cell or burst scale behaviour produce an additional difficulty because there are not any permanent cell rate levels but a process with continuously varying cell rates. If we, after all, want to use parameters related to cell rate distribution we should determine a basic time unit (or observation period) for the calculations. Two extreme cases can be distinguished. First, if each observation takes only one time slot, all higher moments are determined by the mean cell rate—this is clearly an inadequate method. Second, if the period is equal to connection duration, it is not possible to measure the variance of one connection (though it is possible to include in the variance the uncertainty of mean cell rate). The optimum value for the observation period is between these approaches but, unfortunately, it is hard to find any simple method to obtain the optimum value since it depends both on the traffic process and buffer capacity. One proposal for this value is 1 ms (Lindberger 1991).

4.1.4 Index of dispersion

To solve the above-mentioned problem of optimum observation period one alternative is to present variance as a function of the observation period. The *index of dispersion for count* at time t is the variance of the number of arrivals in an interval length t divided by the mean number of arrivals in t (Gusella 1991):

$$I(t) = \frac{\mathrm{Var}\{N(t)\}}{\mathrm{E}\{N(t)\}}. \tag{4.1}$$

$I(t)$ has been used mainly in describing and evaluating the properties of traffic models; in particular, how well the models fit the real data from existing networks in respect of $I(t)$. However, it is very difficult to find any method to calculate QoS parameters, such as cell loss probability, directly from the index of dispersion.

4.1.5 Burstiness and peakedness

Many approaches to capture the intrinsic behaviour of a connection using a one traffic parameter have been presented; *burstiness* is perhaps the most popular. Bae and Suda (1991) have presented six definitions for burstiness. The definition most commonly used is the ratio of the peak cell rate to the mean cell rate (h/m).

As has been shown by Iversen and Bohn Nielsen (1992) this definition for burstiness is not a practical parameter for describing the behaviour of a aggregated traffic stream because it does not take into account the number of sources. If one thousand low cell rate sources with burstiness two is multiplexed, the result is very distinct from a case in

which ten high cell rate sources with the same burstiness are multiplexed. In this respect a better parameter is the ratio of the variance of cell rate to the mean cell rate or *peakedness* $z = v/m$.

Other definitions for burstiness have their weaknesses and obviously there is no one definition appropriate to all cases, at least, it is difficult to use burstiness as a direct parameter (i.e., without presupposing any underlying network properties). In this study we use burstiness as a name for the *h/m* ratio, mainly for the sake of simplicity, but we do not use it as an intrinsic source parameter.

4.1.6 Correlation

Some source types have inherent correlations in the arrival process. This is, in fact, a well-known phenomenon in telephone and data traffic (see Rahko 1967). A typical example is VBR video, in which successive frames have strong correlation because the required number of bits of one frame depends on the type of scene, and there may be correlation over the whole duration of the connection because of the constancy in the presentation of the motif. There is a fixed relation between index of dispersion and serial correlation: $I(\tau)$, $\tau \leq t$ provides the correlation structure related to intervals within distance of t (Helvik, Hokstad & Stol 1991).

Doshi et al. (1991) have proposed the following way to model correlation. The behaviour of a virtual circuit during a call holding time is given by $\{(\Delta_n, I_n); n \geq 1\}$, where Δ_n and I_n are respectively the number of cells transmitted in the n^{th} data burst and the length of the n^{th} idle period. The sequence of random vectors, $\{(\Delta_n, I_n); n \geq 1\}$, could be serially correlated. Doshi et al. have presented typical behaviour of these random vectors for different traffic types, for example, long bursty file transfer is characterised by large values of Δ_n and I_n.

Gropp (1993) has presented models for VBR video sources using autoregressive processes. The simplest model uses first-order autoregressive processes but it has the major disadvantage that it can only match the short term correlation. When complicated traffic models, such as the autoregressive moving average process, are used, the queuing performance will be usually obtained only by simulation as in Grünenfelder, Cosmas, Manthorpe and Odinma-Okafor (1991). Therefore, although the correlation vector can be used for purposes of analysing, the negotiating parameters should be simpler. One alternative is to adjust the original correlation curve to a well-known traffic model and then use this model as a substitute in QoS evaluation.

4.1.7 Fractional Brownian Motion

The foregoing models are suitable primarily for cases where:

- it is possible to obtain sufficient information on each connection, and
- connections are independent of each other.

If these assumptions are not valid, we must use a different approach based on traffic measurement. This demand is especially true with traffic between LANs, which has the property that the index of dispersion monotonically increases throughout a time span of 6 orders of magnitude from 1 ms to hours (Fowler & Leland 1991). This phenomenon can be explained only if the traffic contains strong and complicated correlation effects, which are very difficult to model by a tractable traffic model. A promising approach is to use Fractional Brownian Motion which can be fitted into the index of dispersion

curve by using self-similarity parameter *H* (see Section 3.5). Then the FBM model can be used for obtaining approximations for queue length distribution (Norros 1993).

4.2 Derived models and parameters

The target of this section is to develop models and parameters for the description of ATM traffic. The basic assumption is that the intrinsic traffic parameters (Traffic Descriptor), network parameters (link capacity and buffer size), and QoS parameters (cell loss probability) are known and from these parameters we calculate derived traffic parameters that describe in a simple way the main characteristic of each traffic source. The emphasis is both to assist the development of CAC methods and to improve the knowledge of ATM traffic process and by that means to help the selection of a proper CAC method.

We introduce three models for ATM traffic: effective bandwidth, effective variance and a combination of both, EBV. Both effective bandwidth models and models based on variance have been applied by many authors (see Section 5.2). The main strength of the following presentation is that the determination of source parameters is based either on a homogeneous case or on a very simple heterogeneous case and, moreover, the same solution to the homogeneous case can be applied to every CAC method. In short, the approach used in this study provides a very flexible way to develop efficient CAC procedures.

Despite the efficiency of the presented models, they do not fully satisfy the comprehension aspect. In order to fulfil this need two new parameters are introduced in Section 4.2.4.

4.2.1 Effective bandwidth

The basic problem in developing an efficient CAC method is to find a suitable approximation for heterogeneous traffic cases. Although it is possible to solve this problem exactly in some cases, solutions are usually too complicated for real-time purposes. The simplest approach is to suppose that the bandwidth required by a source is independent of other traffic components, or in other words, that the acceptance region is approximately linearly constrained (Bean 1993). Various terms, such as equivalent bandwidth, virtual bandwidth and effective bandwidth, have been applied to the needed bandwidth. The last term is used in this study.

Although there are numerous CAC methods based on the concept of effective bandwidth, the presentation is usually restricted to certain traffic cases (see Section 5.2.1). These methods use almost invariably rate-variation scale models in spite of the fact that the effective bandwidth is most accurate when the burst size is much smaller than the buffer size (see Section 4.3). In this study we present an effective bandwidth concept that is independent of the underlying traffic process and time scale. Let us approach the problem from the basis of a homogeneous case and denote the allowed number of sources of type *i* by $N_{c,i}$ when the link capacity is *c* ($N_{c,i}$ can be obtained by any exact or approximate formula). Then the *effective bandwidth* of the source *i* can be defined as:

$$k_i = c / N_{c,i} . \tag{4.2}$$

The acceptance rule can then be expressed by the formula:

$$\sum_i k_i \le c. \tag{4.3}$$

We denote this formula as the first effective bandwidth model (EB_1). The basic principle of EB_1 models is that the determination of effective bandwidth is based purely on the homogeneous case and a factor common to all sources is used in regulating the value of the allowed load (see Section 5.3.2.1).

In order to obtain a clear view of the main characteristic of EB_1 model let us take a simple example in which sources of type *i* are aggregated with CBR load. According to the effective bandwidth model the allowed number of sources of type *i* depends linearly on the CBR load (ρ_{cbr}):

$$k_i N_i + \rho_{cbr} c \le c. \tag{4.4}$$

Consequently, the value of effective bandwidth of a source is supposed to be valid for all link capacities less than *c*. This is an important property because in reality the link capacity between two network nodes be may divided by several semi-permanent Virtual Paths and therefore the capacity shared by a group of virtual connections may be anything less than the link capacity.

Formula (4.4) can be further simplified by defining factor ψ_i as the allowed number of sources of type *i* divided by $N_{c,i}$ (see Figure 4.1):

$$\begin{aligned} \psi_i &= \frac{N_i}{N_{c,i}} \\ &\le 1 - \rho_{cbr}. \end{aligned} \tag{4.5}$$

However, (4.2) is not necessary the optimum determination of the effective bandwidth because the traffic process usually consists of different types of sources and the needed bandwidth of a source may depend on the actual traffic mix. We denote all models that apply the effective bandwidth concept and in which the determination of effective bandwidth is based on any heterogeneous model by EB_2.

The simplest technique to realise the EB_2-principle is to aggregate the sources under study with a CBR load. In order to obtain a unequivocal determination for the effective bandwidth the following definitions are applied throughout this study:

- factor ρ_{max} is defined as the maximum attainable load among all possible traffic cases, particularly with a homogeneous CBR load;
- the effective bandwidth of a CBR source is equal to peak rate.

The main reason to introduce factor ρ_{max} is that by means of it the maximum load can be limited in case of approximation errors (see Section 5.3.2.1). The second definition fixes the unit of measure for an effective bandwidth. Using these two definitions we can determine an other effective bandwidth k_i^* in the following way:

$$k_i^* = \mathrm{Max}\left\{ \frac{\rho_{max} c}{\psi_{max,i} N_{c,i}}, m_i \right\}, \tag{4.6}$$

where $\psi_{max,i}$ is the maximum value that satisfies the condition:

- for all values of CBR load ($0 \le \rho_{cbr} \le 1$) the straight line between points (0, $\psi_{max,i}$) and (ρ_{max}, 0) is in the acceptance region (see illustration in Figure 4.1).

The acceptance region (i.e., the allowed number of sources *i* for different values of CBR load) can be calculated by means of any exact or approximate method. The formula for acceptance decision is similar to the EB_1 model except for the additional factor ρ_{max}. In a general traffic case the acceptance rule of EB_2 is:

$$\sum_i k_i^* \le \rho_{max} c. \tag{4.7}$$

Note that the second alternative in (4.6) (i.e., $k_i^* = m_i$) is caused by the definition of ρ_{max} together with (4.7).

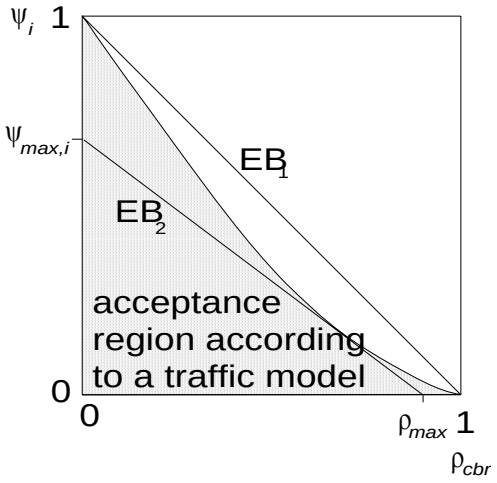

Figure 4.1. The principles of two effective bandwidth models, EB_1 and EB_2; ρ_{cbr} = constant bit rate load, ψ_i = the allowed number of sources of type *i* divided by $N_{c,i}$.

In the case of the superposition with CBR load we obtain:

$$k_i^* N_i + \rho_{cbr} c \le \rho_{max} c. \tag{4.8}$$

From this formula we obtain (EB_2 in Figure 4.1):

$$\psi_i \le \left(1 - \frac{\rho_{cbr}}{\rho_{max}}\right) \psi_{max,i}, \tag{4.9}$$

if the first part in (4.6) is valid whereas if the second part is valid, the allowed ψ_i is lower.

The same principle with a tangent plane approximation has been applied for example by Kelly (1991) and Miyao (1993), although the method for determining k_i^* varies considerably.

4.2.2 Effective variance

There are various ways of applying the variance of cell rate distribution in Connection Admission Control (see Section 5.2.2). The underlying idea is that the Gaussian

distribution can be used for modelling rate-variation scale fluctuations (see Section 3.3.2.2). In that case the admittance function can be written in the following form:

$$\sum_i m_i + \sqrt{\sum_i \kappa^2 v_i} \le c, \tag{4.10}$$

where Σv_i is the variance of aggregated cell rate distribution and κ depends only on the cell loss requirement.

Because the starting point of this approach is the cell rate distribution, (4.10) can be applied directly with rate-variation scale models but not with other traffic models. However, we can use v_i as an effective parameter instead of a real parameter that can be measured from traffic flow and controlled directly. If we including factor κ in the variances v_i we obtain the *effective variance* model (Kilkki 1992):

$$\sum_i m_i + \sqrt{\sum_i v_i^*} \le c, \tag{4.11}$$

where v_i^* can be obtained by the following formula when $N_{c,i}$ is known:

$$v_i^* = \frac{\left(c - m_i N_{c,i}\right)^2}{N_{c,i}} \cdot \tag{4.12}$$

The main usefulness of this formulation is that the application of v_i^* is independent of traffic model, time scale and the approximation method used in homogeneous cases. Furthermore, it can be deduced from (4.2) and (4.12) that the effective variance model is always applicable when the mean cell rate and the effective bandwidth of each source are known:

$$v_i^* = \frac{c}{k_i}\left(k_i - m_i\right)^2 . \tag{4.13}$$

4.2.3 Combination of effective bandwidth and effective variance

As can be seen from the above-mentioned formulae, the behaviour of effective bandwidth and effective variance are essentially different, and the same difference can be observed in real traffic situations. As is shown in Sections 4.4.1 and 4.4.3, the effective bandwidth and effective variance models can estimate accurately only the two extreme positions, and the estimation of anything between those two limits (with burst size roughly equal to buffer size) is less accurate. A plausible approach is to combine (4.3) and (4.11). The following demands can be made for the combined formula:

- effective variance and effective bandwidth should be special cases of the combined formula;
- the formula should be mathematically as simple as possible;
- it should be suitable to all types of traffic process.

A possible approach is to calculate an acceptance region using both effective bandwidth and effective variance formulae and then select either of them. Various selection rules can be applied (e.g., select smaller or larger) but because of the complicated nature of ATM traffic there is no simple rule that can be applied to all cases (see Section 5.2.3).

Another approach is to start from the derived source parameters of the effective bandwidth and effective variance models. The first two requirements can be realised by the following combined formula:

$$\sum_i m_i + \left(\left(\sum_i \sigma_i^{**} \right)^{2\gamma} + \left(\sum_i v_i^{**} \right)^{\gamma} \right)^{1/2\gamma} \le c, \tag{4.14}$$

where: σ_i^{**} = an effective standard deviation representing cell scale fluctuations and partly the burst scale fluctuations, and

v_i^{**} = an effective variance representing mainly the rate-variation scale fluctuations.

It can be easily seen that effective bandwidth (EB_1) and effective variance formulae are special cases of (4.14):

$k_i = m_i + \sigma_i^{**}$ in (4.3) when $v_i^{**} = 0$;

$v_i^{*} = v_i^{**}$ in (4.11) when $\sigma_i^{**} = 0$.

Factor γ can be chosen so that the last requirement is fulfilled. According to simulation results the choice $\gamma = 1$ seems to be most practical, see Section 4.4.2. Then we obtain the following EBV formula (Kilkki 1992):

$$\sum_i m_i + \sqrt{\left(\left(\sum_i \sigma_i^{**} \right)^2 + \sum_i v_i^{**} \right)^{+}} \le c. \tag{4.15}$$

Since (4.15) has two free parameters for each source, two points are needed for the determination of σ_i^{**} and v_i^{**}. The homogeneous case is the first one and for the other we can use a case in which a half of the link capacity is reserved by CBR traffic. Let us define the allowed number of sources under consideration by $N_{c/2,i}$. Then we obtain:

$$v_i^{**} = \frac{N_{c,i}\left(c/2 - N_{c/2,i} m_i\right)^2}{N_{c/2,i}\left(N_{c,i} - N_{c/2,i}\right)} - \frac{N_{c/2,i}\left(c - N_{c,i} m_i\right)^2}{N_{c,i}\left(N_{c,i} - N_{c/2,i}\right)}, \tag{4.16}$$

$$\sigma_i^{**} = \frac{\sqrt{\left(c - N_{c,i} m_i\right)^2 - N_{c,i} v_i^{**}}}{N_{c,i}} \cdot \tag{4.17}$$

We can deduce from (4.16) and (4.17) that the application of the EBV model is independent of the source model; the only parameters needed are m_i, $N_{c,i}$ and $N_{c/2,i}$ for each source.

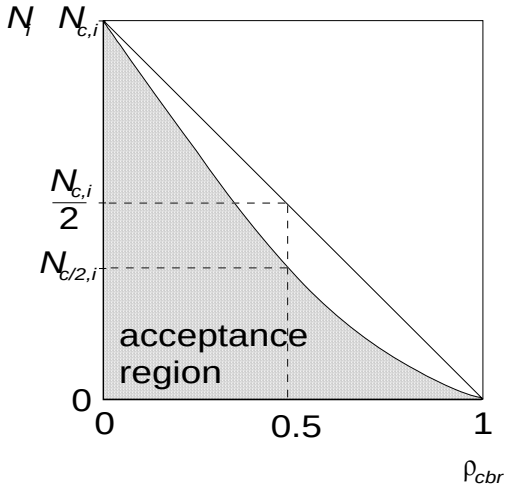

Figure 4.2. Determination of $N_{c,i}$ and $N_{c/2,i}$.

It should be noted that although the determination of $N_{c/2,i}$ is based on a superposition of a CBR load, the difficulty of solving these cases is roughly equal to that of homogeneous case. This statement is obviously true with rate-variation scale models, and, in addition, most traffic models at cell and burst scales can be modified easily in order to apply a homogeneous model with link capacity c/2 (see Section 4.4.1).

Parameter v_i^{**} obtained from (4.16) is negative if $N_{c/2,i} > N_{c,i}/2$. In practical implementation it is better not to use negative values because this property causes problems in some cases (see Sections 4.4.2 and 4.4.4). In the rest of this chapter, however, the basic rule is that negative values are accepted.

4.2.4 Scale factors

In the previous sections we presented four models for description of the ATM traffic process. The parameters of these models (k_i, k_i^*, v_i^*, σ_i^{**} and v_i^{**}) are useful so far as traffic modelling is concerned but they are not very understandable if we attempt to describe the general behaviour of a source. In this section we introduce two parameters that meet this demand. By these two parameters we can describe the main characteristic of the source and by that means develop efficient rules for the selection of analysing methods and CAC procedures.

The starting point is that we determine two standard traffic models: the first one relating to the short term fluctuations at cell scale and the other relating to the long-term fluctuations at rate-variation scale. First of all, we shall determine precisely the time scales by defining the intrinsic traffic process of each scale. An obvious candidate for the cell scale model is the arrival process of independent cells (i.e., *M*/*D*/1/*K* queuing system, see Section 3.1). Correspondingly, at rate-variation scale a typical model is a VBR source with infinite duration of each bit rate level. We can use these two traffic models as standards of comparison.

In addition, we should define in which respect the comparison is made. The most important criteria are:

- the allowable load and
- the behaviour of multiplexing process.

The first criterion is needed when appraising the suitability of different traffic models for homogeneous cases and the second one is suitable for the assessment of heterogeneous approximations. The application of the first criterion is simple because the solutions of both standard models are known (see Sections 3.1 and 3.3). The second criterion needs some knowledge of the multiplexing process at cell scale and at rate-variation scale. This problem is studied thoroughly in Section 4.4. The main result is that at cell scale, particularly with *M*/*D*/1/*K* model, effective bandwidth is a very accurate approximation whereas at rate-variation scale effective variance is a suitable approximation.

Using these two standard models and two criteria, we denote scale factors based on utilisation (ε_u) and the multiplexing process (ε_m) in the following way:

- $\varepsilon_u = 0$ if the allowable load in homogeneous case is the same as that for *M*/*D*/1/*K* model;
- $\varepsilon_u = 1$ if the allowable load in homogeneous case is the same as that for corresponding VBR model;
- $\varepsilon_m = 0$ if the effective bandwidth model is exact when sources are multiplexed with CBR sources;
- $\varepsilon_m = 1$ if the effective variance model is exact when sources are multiplexed with CBR sources.

The corresponding VBR source means in this study a source which has the same parameters as the original source except that the duration of each state is supposed to be infinite. Then we can neglect the buffer capacity and use the formulae for cell loss probability presented in Section 3.3.

Let us define the allowable load according to *M*/*D*/1/*K* model by ρ_0 and the allowed number of sources according to corresponding VBR model by $N_{c,i}\{\text{VBR}\}$. Then the utilisation factor of source *i* can be defined as:

$$\varepsilon_{u,i} = \frac{N_{c,i}\{M/D/1/K\} - N_{c,i}}{N_{c,i}\{M/D/1/K\} - N_{c,i}\{\text{VBR}\}}, \tag{4.18}$$

where $N_{c,i}\{M/D/1/K\} = c\rho_0 / m_i$.

The scale factor of source *i* is defined respectively:

$$\varepsilon_{m,i} = \frac{\frac{1}{2}N_{c,i} - N_{c/2,i}}{\frac{1}{2}N_{c,i} - N_{c/2,i}\{\text{EV}\}}, \tag{4.19}$$

where $N_{c/2,i}\{\text{EV}\}$ is the allowed number of sources when a half of the link capacity is reserved for CBR traffic. $N_{c/2,i}\{\text{EV}\}$ can be obtained by means of the effective variance model:

$$N_{c/2,i}\{\text{EV}\} = \frac{cm_i + v_i^* - \sqrt{v_i^*\left(2cm_i + v_i^*\right)}}{2m_i^2}, \tag{4.20}$$

where v_i^* is obtained from (4.12).

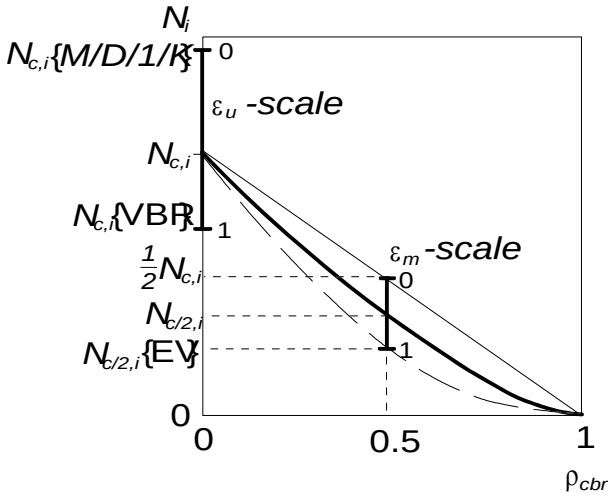

Figure 4.3. Determination of ε_u and ε_m scales.

4.3 Description of burst scale sources by scale factors

The target of this section is to provide an insight into the traffic behaviour between typical cell scale and rate-variation scale sources, and by that means to give relatively simple rules for the selection of suitable traffic models at burst scale. The main tools used in the examination are the scale factors ε_u and ε_m defined in the previous section. For the examination we need methods to solve burst scale models. In some special cases we have appropriate analytical methods (such as fluid flow approximation) but in most cases the only suitable method of evaluation is to use simulation tools. Throughout the examination buffer size is 100 cells, acceptable cell loss probability is 10^{-4} and link capacity is used as a unit for cell rates (i.e., we denote $c = 1$).

Firstly, we should determine the accuracy of simulation results so as to make sure of the validity of the inferences. The standard deviation of cell loss probability due to inaccuracy of simulation results is roughly 6% (see Section 3.6.2). Using this information we can firstly determine the accuracy of determining the allowed number of sources and, secondly, the accuracy of factors ε_u and ε_m (see Appendix C). We can obtain the standard deviations of error caused by the inaccuracy of simulation results as follows:

- allowed load: from 0.001 to 0.005;
- ε_u: from 0.001 to 0.006;
- ε_m: from 0.015 to 0.06.

The inaccuracy of ε_u is almost discernible in the following figures whereas the inaccuracy of ε_m is perceivable for example in Figure 4.6, but it has no effect on the general conclusion to be drawn from the results.

4.3.1 From cell scale through burst scale into rate-variation scale

Let us examine some typical traffic cases in order to provide a further insight into the boundaries between cell and burst scales, and between burst and VBR scales. In the first example presented in Figure 4.4 we have deterministic sources with the following parameters:

- peak rate, $h = 0.1$;
- *on* probability in burst scale, $p_{burst} = 0.2$.

Since the corresponding rate-variation scale model depends only on h and p_{burst} but not on burst size, the allowable load is always the same (= 0.38) for the limit model of rate-variation scale.

If the burst size is less than 10, both ε_u and ε_m are negative and hence these sources can be clearly classified as cell scale sources while all sources with burst size larger than 400 cells behave as rate-variation scale sources. Furthermore, if the burst size equals the buffer size, the utilisation factor is as high as 0.8, which means that the error in omitting the buffer capacity (i.e., application of rate-variation scale approximation) is relatively small even if the burst size is of the same order as the buffer size. In this case the region of burst scale lies roughly from 10 to 400 cells measured in burst size. This result is quite general (only if p_{burst} is larger than 0.3 is the situation somehow different as we can see later) and important because it reveals the difficulty of buffering burst scale fluctuations using typical buffers in ATM networks.

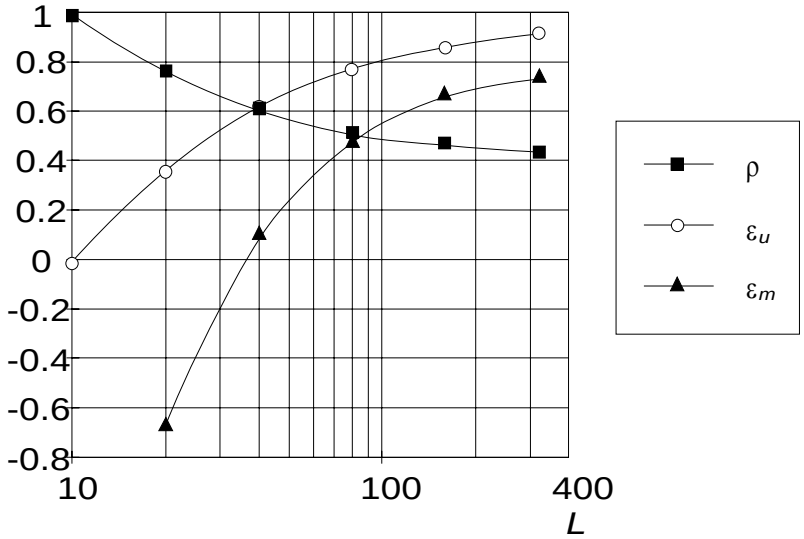

Figure 4.4. Allowable load (ρ) and scale factors (ε_u and ε_m) as a function of burst size L; $h = 1/10$, $m = 0.02$, $p_{burst} = 0.2$, $K = 100$, $P_{loss} = 10^{-4}$.

Figures 4.5 and 4.6 show the effect of peak rate on the factors ε_u and ε_m. Now the mean rate is constant, the burst scale period D_{burst} is constant for each burst size L, and the peak rate is a variable. This type of situation occurs at LAN/ATM interfaces where packets are segmented into ATM cells: the packet size and the interarrival time of packets are fixed whereas it is possible to adapt the peak cell rate.

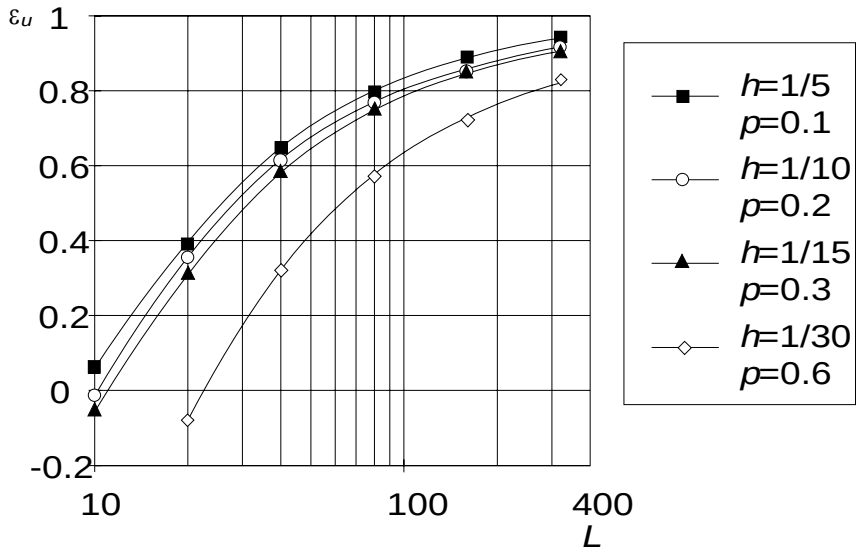

Figure 4.5. Utilisation factor ε_u as a function of burst size L for different peak rates h; m = 1/50, K = 100, $P_{loss} = 10^{-4}$.

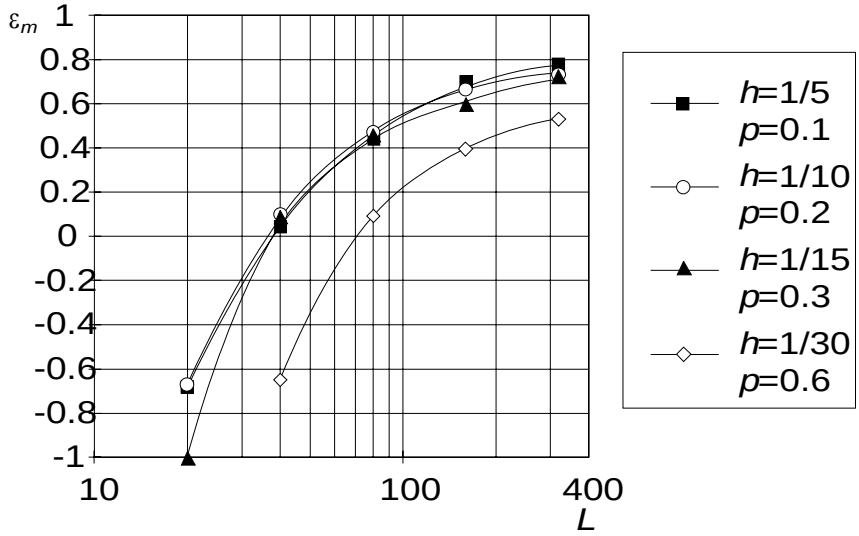

Figure 4.6. Multiplexing factor ε_m as a function of burst size L for different peak rates h; m = 1/50, K = 100, $P_{loss} = 10^{-4}$.

Both utilisation and multiplexing factors are almost independent of the cell scale period if p_{burst} is less than 0.3. This means that the general behaviour of burst scale sources depends only slightly on the peak rate provided that the source burstiness is high. However, it should be stressed that this independence between peak rate and the utilisation factor does not mean that the allowable load is independent of peak rate because the allowable load of limit case (rate-variation scale model) depends strongly on the peak rate. Furthermore, Figure 4.6 shows that the effective bandwidth scheme is valid if the buffer size is considerable larger than the burst size. This result is in line with theoretical studies concerning the applicability of effective bandwidth (see e.g., Elwalid & Mitra 1993).

As we might infer from the foregoing figures, factor p_{burst} may have a substantial influence on the scale factors if it exceeds 0.3. Figures 4.7 and 4.8 further illustrate this phenomenon. Next we keep peak rate constant and vary mean rate, or in other words, equal bursts (determined by h and L) are arriving at a network node at different speeds

($1/D_{burst}$). In this case the effect of p_{burst} is more distinct particularly in respect of the multiplexing factor.

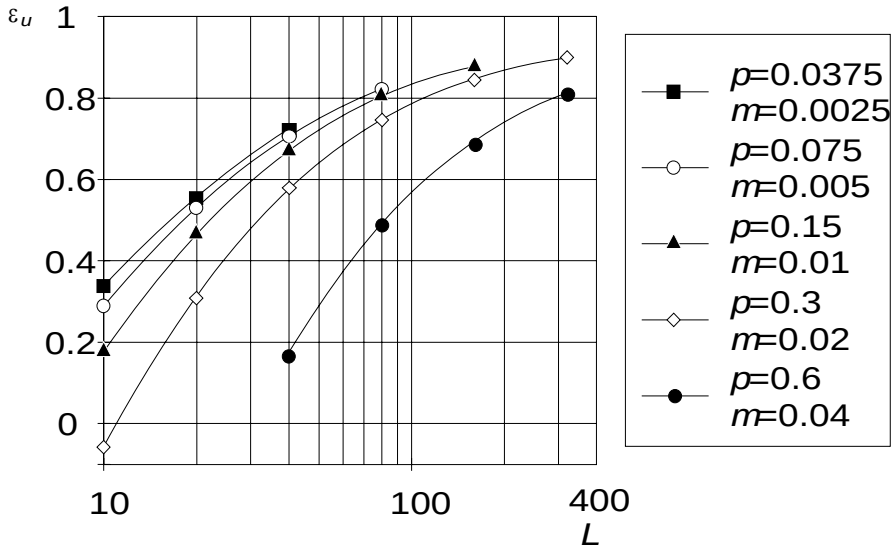

Figure 4.7. Utilisation factor ε_u as a function of burst size L for different values of p_{burst}; h = 1/15, K = 100, $P_{loss} = 10^{-4}$.

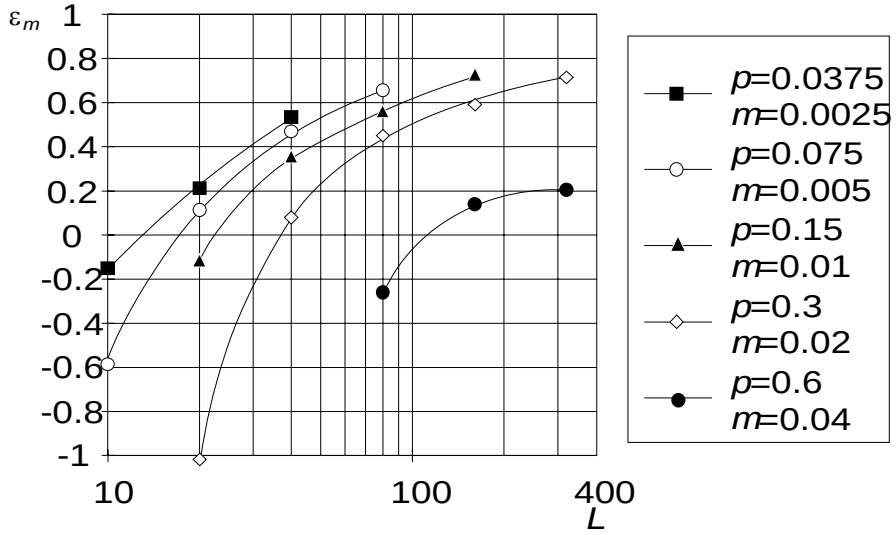

Figure 4.8. Multiplexing factor ε_m as a function of burst size L for different values of p_{burst}; h = 1/15, K = 100, $P_{loss} = 10^{-4}$.

One way to illustrate the characteristic of a source is to use a plane determined by the utilisation and multiplexing factors. The positions of sources with burst lengths from 10 to 320 cells and burstiness from 1.6 to 160 are shown in Figures 4.9, 4.10 and 4.11 for three peak rate values, 1/5, 1/15 and 1/60, respectively. The inferences are much the same as earlier:

- burst size is the most important parameter for the classification;
- burstiness ($=1/p_{burst}$) has a minor effect on the utilisation and multiplexing factors when burst length is constant;
- sources can be classified explicitly as a cell scale source only if the burst size is very small;

- if the burst size is at least four times as large as the buffer size, the source can be classified as a rate-variation scale source.

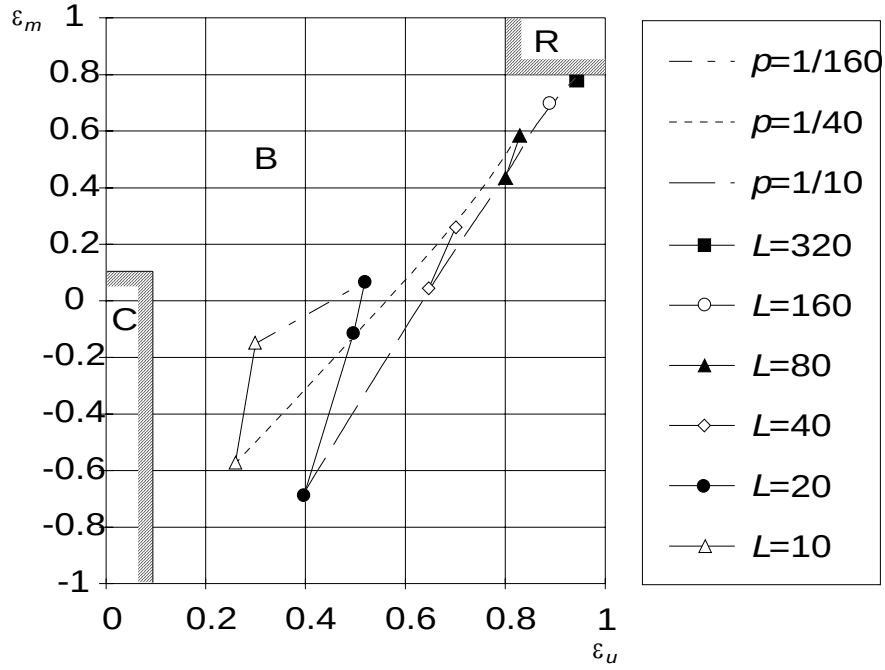

Figure 4.9. Source position on the scale factor plane as a function of p_{burst} and burst size L; h = 1/5, K = 100, P_{loss} = 10^{-4};
C: cell scale, B: burst scale, R: rate-variation scale.

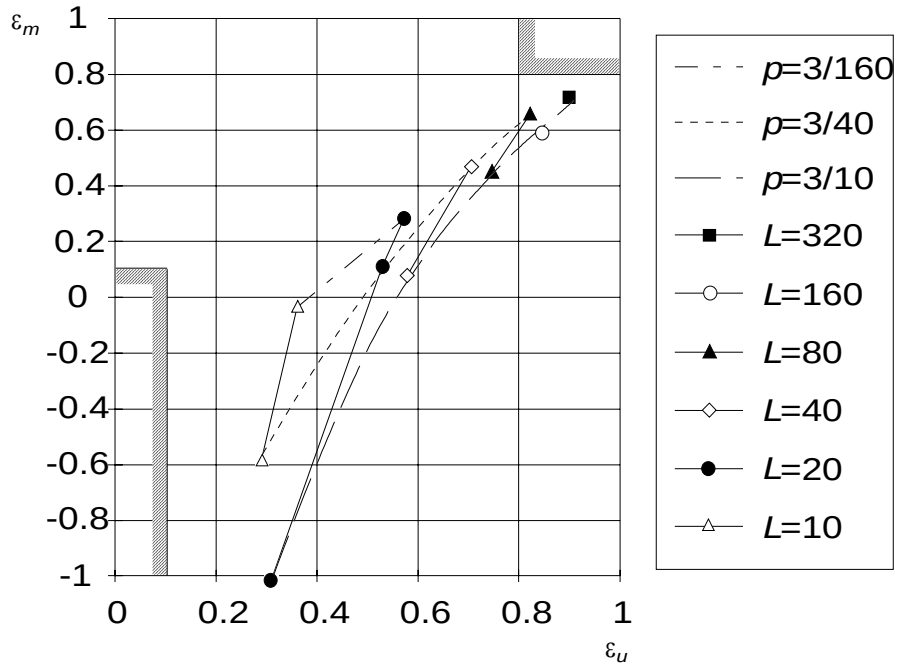

Figure 4.10. Source position on scale factor plane as a function of p_{burst} and burst size L; h = 1/15, K = 100, P_{loss} = 10^{-4}.

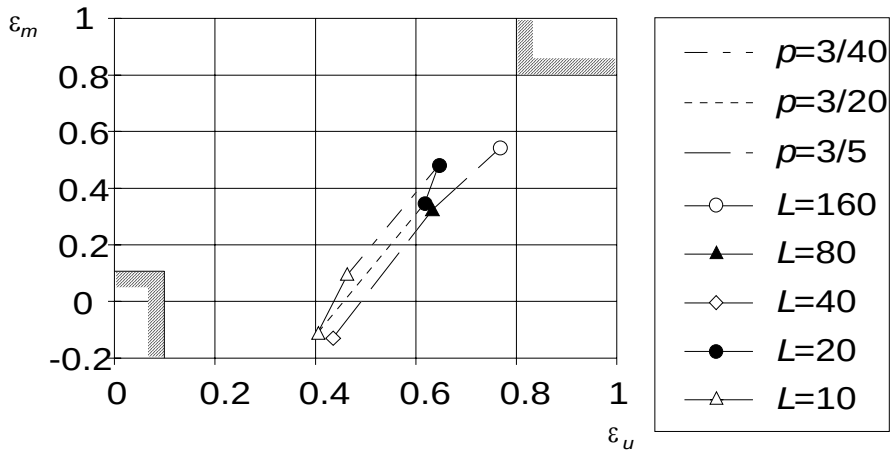

Figure 4.11. Source position on the scale factor plane as a function of p_{burst} and burst size L; h = 1/60, K = 100, P_{loss} = 10^{-4}.

4.3.2 Deterministic vs. the Markov process

All results in the previous section were based on the deterministic process (both burst size and interarrival time of bursts were constant). However, a typical data source can be better illustrated by the Markov process. For the comparison of deterministic and Markov processes we can apply the scale factor plane and examine how the position of a source shifts while the type of process changes but the average values of burst scale parameters (L and p_{burst}) remain unchanged. The following results are based on a fluid flow approximation with geometrically distributed on- and off-periods.

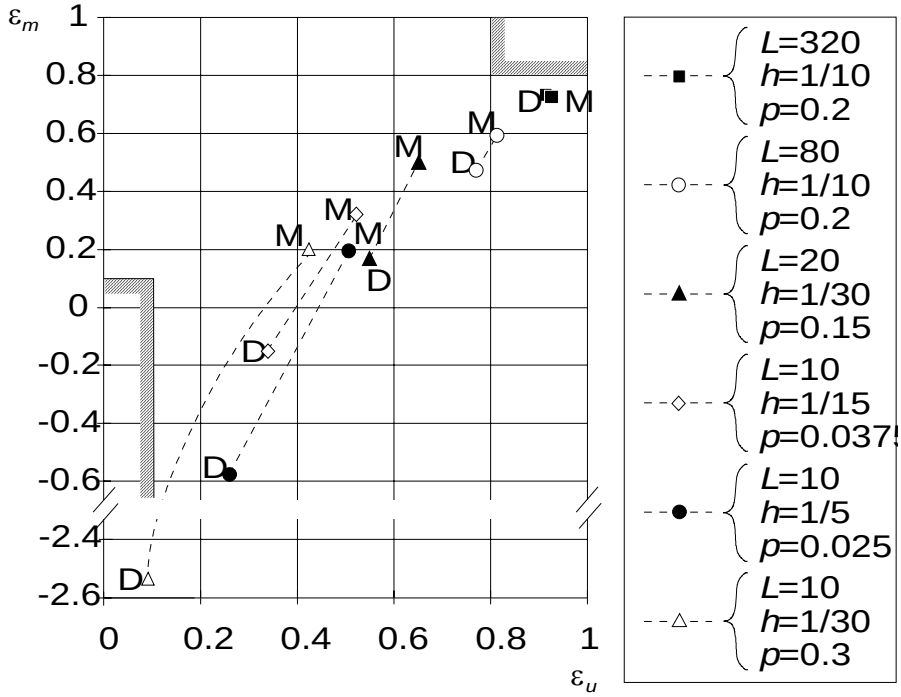

Figure 4.12. The effect of process type on the source classification; M = Markov model, D = deterministic model, K = 100, P_{loss} = 10^{-4}.

Figure 4.12 shows the shift for six different sources. In all cases Markov sources have larger scale factors (except for the case with the largest burst size, in which the accuracy of simulation does not allow a discoverable difference between the models). The

difference is noticeable with a small burst size: when L = 10, h = 1/5 and p_{burst} = 2000, the difference in the utilisation factor is about 0.25 and in the multiplexing factor as high as 0.8. A rough approximation is that a Markov source has the same scale factors as a deterministic source with twice as large a burst size as the Markov source.

4.3.3 Effect of cell loss probability standard on scale factors

Another weakness of the previous examination is the cell loss probability level since 10^{-4} is not sufficient for most applications in ATM networks. Let us see what the difference is when the P_{loss} standard changes from 10^{-4} to 10^{-9}. From Figure 4.13 we can deduce a rule (though the analysis is brief): the tougher the cell loss requirement the greater scale factors. The difference is again more distinct when the burst size is small. An interesting observation is that the sources with constant peak rate (h) and burstiness ($1/p_{burst}$) form a nearly straight line on the scale factor plane; a simple approximation can perhaps be developed if this phenomenon is applied to the burst scale traffic.

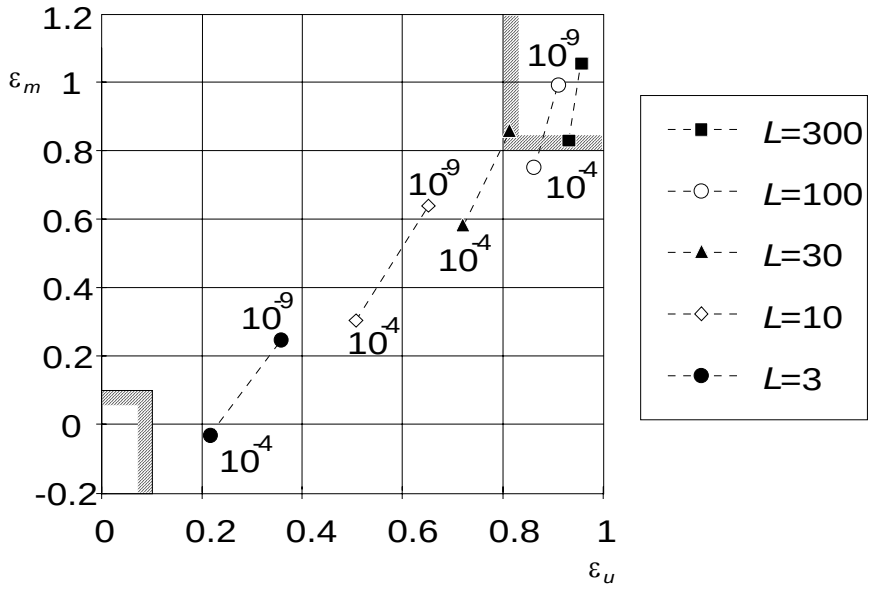

Figure 4.13. Scale factors change from $P_{loss} = 10^{-4}$ to $P_{loss} = 10^{-9}$ for different burst sizes; MMDP model, h = 1/20, p_{burst} = 0.1, K = 100.

4.3.4 Combination of rate-variation and burst scales

In this section we attempt to clarify some phenomena arising when rate-variation and burst scale processes are combined in one source. Several traffic types may lead to this type of model (see Section 2.4). In order to achieve an efficient traffic control it is important to know whether both rate-variation and burst scale fluctuations are effective at the same time and which one of the variations is dominant.

Let us take a source model in which the traffic process is of the on/off type both at the burst scale and at the rate-variation scale, burst size is 20 cells and $D_{burst} = 16000 p_{rv}$. This means that the mean rate of every source is 1/800, and the form of a burst is unchangeable for a given peak rate. The differences between sources are related to the time scale of fluctuations. If there are no rate-variation scale fluctuations (p_{rv} =1), the interarrival time of the bursts is large (16000 time slots), which indicates large burst scale fluctuations. In contrast, if p_{rv} is small, the traffic process is smooth in the burst scale (in an extreme case $D_{burst} = L/h$) while rate-variation scale burstiness is high.

In Figures 4.14 and 4.15 we can see how scale factors change from typical burst scale values ($\varepsilon_u \approx 0.5$, $\varepsilon_m \approx 0.2$) to typical rate-variation scale values ($\varepsilon_u > 0.8$, $\varepsilon_m > 0.9$). When rate-variation scale fluctuations increase (or p_{rv} decreases) there are at first no perceptible changes but later the change of scale factors is rapid. When the peak rate increases, burst scale fluctuations are larger in relation to rate-variation scale fluctuations and for this reason the turning point occurs at the smaller value of p_{rv}. In any event, the general behaviour is independent of the peak rate.

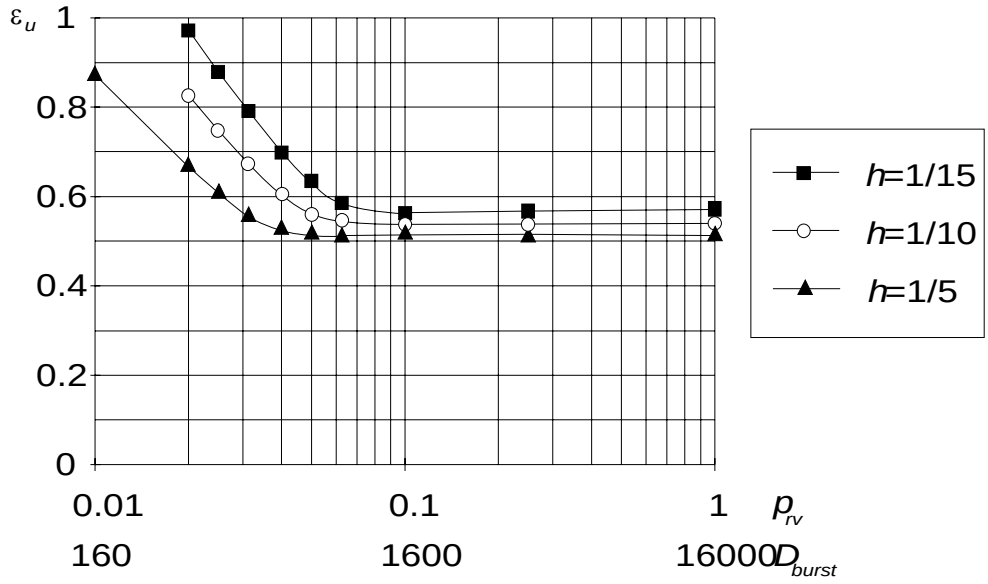

Figure 4.14. Utilisation factor (ε_u) as a function of p_{rv} for three peak rate values; $L = 20$, $m = 1/800$, $K = 100$, $P_{loss} = 10^{-4}$.

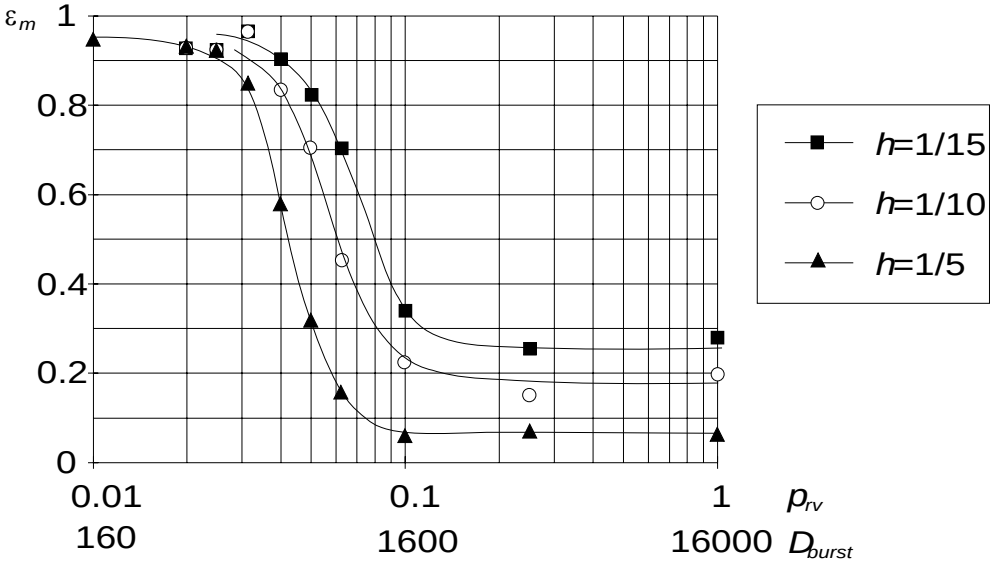

Figure 4.15. Multiplexing factor (ε_m) as a function of p_{rv} for three peak rate values; $L = 20$, $m = 1/800$, $K = 100$, $P_{loss} = 10^{-4}$.

Usually either the burst scale or rate-variation scale process is dominant with respect to the source classification. This phenomenon is even clearer if we examine the allowable load. Figure 4.16 shows the allowable load curve for three peak rate values (1/5, 1/10 and 1/15) and, in addition, for a situation in which peak rate has the smallest possible value (= $1/800p_{rv}$ in this example). The behaviour of the allocation curve has two limits:

the allowable load according to a pure rate-variation scale model, and the allowable load based on Poisson bursts. The allowable load of the combined source is always near one or the other of them. There is a small rounding in the allocation curve only on the narrow range where both the limit cases lead to roughly equal allowable loads. The actual values for allowable load with peak rate 1/10 are presented in Table 4.1.

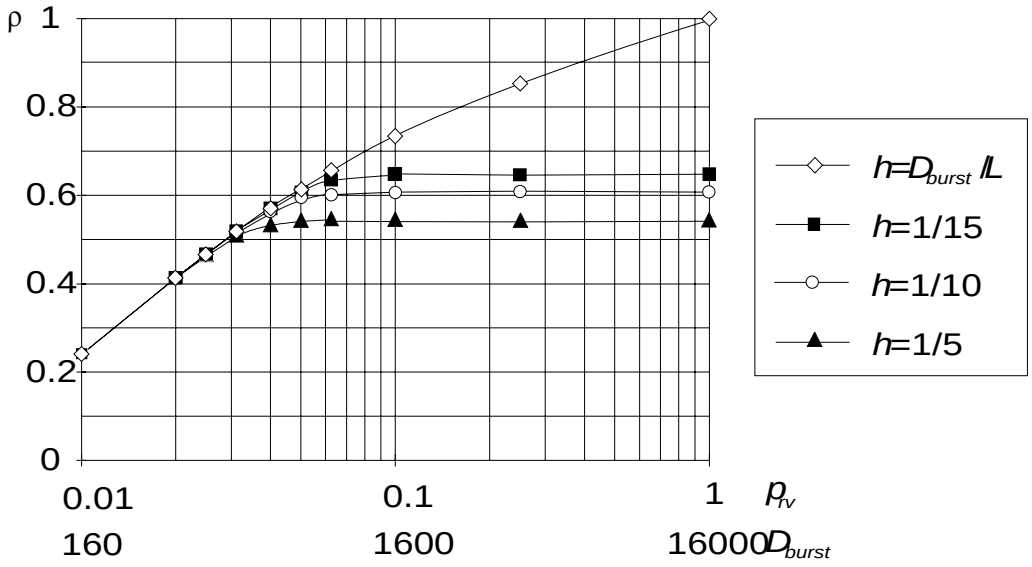

Figure 4.16. Allowable load (ρ) as a function of p_{rv} for three peak rate values and for maximum peak rate ($h=D_{burst}/L$); L = 20, K = 100, $P_{loss} = 10^{-4}$.

Table 4.1. Allowable load of source with both burst scale and rate-variation scale fluctuations, L = 20, h = 1/10, K = 100, $P_{loss} = 10^{-4}$

p_{rv}	D_{burst}	allowable load of		
		Poisson bursts	rate-variation scale	combined source
0.03125	500	0.607	0.517	0.517
0.04	640	0.607	0.570	0.562
0.05	800	0.607	0.615	0.594
0.0625	1000	0.607	0.656	0.602
0.1	1600	0.607	0.734	0.607

The fluctuations of both burst scale and rate-variation scale have a considerable effect on the allowable load in some rare cases and where this happens, the allowable load is only slightly lower than the extreme model with the lower allowable load. This result implies that a combination method similar to (3.14) is valid for the combining of burst scale and rate-variation scale processes. The main difference is that the Poisson arriving process now concerns bursts of cells instead of individual cells (as in $M/D/1/K$ system).

4.3.5 General remarks on burst scale sources

The target of Section 4.3 has been to elucidate the complicated traffic process in ATM networks. Two extreme cases, cell scale model and rate-variation scale model, are relatively simple to solve and they give accurate results provided that the underlying assumptions are valid. Therefore it is preferable to use these models instead of

complicated traffic models characteristic of burst scale process, if the accuracy of either extreme model is sufficient.

The boundary between cell and burst scales is very critical because if a source is classified erroneously as cell scale source, the allowable load will be overestimated either in the homogeneous case (utilisation criterion) or in the multiplexing case (multiplexing criterion). Therefore we should determine this boundary with very small values for ε_m and ε_u, preferably zero. If a larger value (such as 0.1 in Figures from 4.9 to 4.13) is used, a larger reserve for approximation errors in CAC formulae should be used (the reserve for approximation errors is managed by factor ρ_{max}, see Section 5.3.2.1).

The boundary between burst and rate-variation scales is less critical because if a burst scale source is classified as a rate-variation scale source, the allowable load will be underestimated. In addition, the multiplexing factor ε_m of a real VBR source is often not precisely 1 because the approximation used in effective variance formulae is exact only in some special cases. The boundary between burst and VBR scales may be determined by a value of 0.8.

The burst size is the most important parameter in respect of source classification. According to the previous examination the utilisation factor ε_u is almost independent of the link rate to peak rate ratio if the mean to peak rate ratio is less than 0.3. As small a burst as five cells has a considerable influence on the allowable load even though the source model is deterministic and, moreover, in the case of MMDP nearly all sources should be classified as either burst or VBR sources. Thus the *M*/*D*/1/*K* model as such is valid only if the maximum burst size is very small (i.e., two or three cells). If the burst size is at least four times as large as the buffer size, the source can be classified as a rate-variation scale source.

4.4 Traffic models for different time scales

In this section we show either by analytical methods or by simulation results that each traffic model presented in Section 4.2 is valid on a certain time scale: effective bandwidth at cell scale, effective variance at rate-variation scale and EBV at burst scale. In addition, general traffic cases with sources of various types are analysed in Section 4.4.4.

In spite of the suitability of each model in certain traffic cases, each model has its weak points in other cases. These weaknesses are identifiable and ways to overcome the problems which arise are proposed. A common source of error for all models is that the performance evaluation has been based on the average cell loss probability although the cell loss probability obtained by a source may vary considerably depending on the characteristic of each source and on the traffic mix. This issue is examined in Section 4.4.5.

4.4.1 Cell scale and effective bandwidth

Let us first examine the mixing of the Poisson and deterministic processes, and by that means the suitability of effective bandwidth for approximating cell scale processes. The Poisson arrival process can be determined as follows:

- the number of cells arriving during a time slot is independent of all preceding events and is Poisson distributed with mean ρ.

The solution of this model is known (see Section 3.1.2). If we suppose that a deterministic source produces a cell every second time slot and the mean number of cells from the Poisson process is ρ/2, we obtain system S1 with the following arrival process:

- in $(2i)^{th}$ time slot one deterministic cell arrives and Poisson distributed cells with mean ρ/2;
- in $(2i+1)^{th}$ time slot Poisson distributed cells arrive with mean ρ/2.

We can suppose that all arrivals at the buffer occur in the first half of the time slot, all leavings take place in the second half of the time slot and the rejections due to buffer overflow take place in the middle of the time slot.

Now we can modify the original system by shifting all Poisson cells from the $(2i)^{th}$ time slot to the $(2i+1)^{th}$ time slot. Thus we obtain system S2 where:

- in the $(2i)^{th}$ time slot one deterministic cell arrives;
- in the $(2i+1)^{th}$ time slot Poisson distributed cells with mean ρ arrive.

The shifting process of Poisson cells has an effect on the cell loss probability because it is possible that one cell that has been lost in system S1 in time slot $2i$ is shifted to time slot $2i+1$ in S2 and not lost there. The reverse (i.e., that the number of cells lost during time slots $2i$ and $2i+1$ is higher in S2 than in S1 when the state of S2 is equal to that of S1 at the beginning of time slot $2i$) is impossible (see Table 4.2). We can suppose that the shifted and rescued cells have a lower priority than the other cells without affecting the average cell loss probability. Therefore there will be a difference in the number of lost cells only if the rescued cell can find an empty buffer before the next buffer overflow situation and this probability is usually much lower than 1.

Table 4.2. The number of lost cells in systems S1 and S2 during two consecutive time slots; x = the number of empty buffer places at the beginning of time slot $2i$, n_1 = the number of Poisson cells in $(2i)^{th}$ time slot, n_2 = the number of Poisson cells in $(2i+1)^{th}$ time slot

	the number of lost cells in S1	the number of lost cells in S2
$n_1 < x,\ n_2 \le x - n_1$	0	0
$n_1 < x,\ n_2 > x - n_1$	$n_1 + n_2 - x$	$n_1 + n_2 - x$
$n_1 \ge x,\ n_2 = 0$	$n_1 + 1 - x$	$n_1 - x$
$n_1 \ge x,\ n_2 > 0$	$n_1 + n_2 - x$	$n_1 + n_2 - x$

In system S2 in time slot $2i$ one cell arrives and one leaves the buffer, and thus these time slots do not affect the number of lost cells. The number of lost cells during T time slots can be obtained by the aid of $M/D/1/K$ queuing system:

$$N_{lost}\{S2,T\} = \frac{\rho T}{2} P_{loss}\{\rho, M/D/1/K\}, \tag{4.21}$$

and the cell loss probability for S2 is:

$$P_{loss}\{S2\} = \frac{\rho}{1+\rho} P_{loss}\{\rho, M/D/1/K\}. \quad (4.22)$$

Finally we obtain the following formula for the cell loss probability of the original system S1:

$$P_{loss}\{S1\} \approx \frac{\rho}{1+\rho} P_{loss}\{\rho, M/D/1/K\}. \quad (4.23)$$

The point is that $P_{loss}\{S1\}$ differs from that of $M/D/1/K$ system roughly by a factor of 2. Moreover, it should be noted that (4.23) gives an average cell loss probability and in this case the individual cell loss probability of the Poisson stream is substantially higher than that of the CBR stream. If we suppose that the admission decision is based on the individual cell loss probabilities rather than the average one, the admittance function is nearly linear. Therefore the multiplexing factor ε_m for Poisson process of cells is slightly less than 0. Accordingly, the application of effective bandwidth results in a small underestimation of the allowable load when cells with Poisson arrivals are combined with a CBR stream.

4.4.2 Burst scale and EBV model

There are well-grounded reasons to apply effective bandwidth and effective variance at cell scale and at rate-variation scale, respectively, whereas the EBV model has no strict mathematical basis. We know definitely that EBV is valid for the two extreme cases, pure cell scale traffic and pure rate-variation scale traffic. In addition, the EBV model is valid for the special case in which identical burst scale sources are aggregated with a CBR load of 0.5 because the source parameters are defined in that traffic case. All other cases require evaluation.

The following examination is based on simulation results and on a 10^{-4} cell loss probability level. For the analysing one cell scale source (C1), four burst scale sources (B1, B29, B37, B52) and three rate-variation scale sources (R3, R13, R29) have been used. The prime source parameters of each source are presented in Table 4.3 (more information about the sources can be found in Appendix A).

Table 4.3. Source parameters used in analysing, $K = 100$, $P_{loss} = 10^{-4}$

	L	$1/h$	D_{burst}	p_{burst}	p_{rv}	N_1	ρ	ε_u	ε_m
C1	1	100		1	1	100.00	1.000		
B1	10	5	2000	0.025	1	151.59	0.758	0.259	-
B29	40	5	2000	0.1	1	23.25	0.465	0.647	0.043
B37	40	30	4000	0.3	1	75.51	0.755	0.639	0.358
B52	160	5	8000	0.1	1	13.75	0.275	0.889	0.701
R3	1	100		1	0.5	166.17	0.831	1	
R13	1	20		1	0.1	98.12	0.491	1	
R29	1	50		1	0.01	3228.03	0.646	1	

Burst scale sources cover a wide range of properties. Source B29 can be classified as a cell scale source in terms of the multiplexing process although the utilisation factor is as high as 0.647. Source B52 can be classified as a rate-variation scale source and has a low allowable load (0.275) while B37 typifies a burst scale source. The special property of source B1 is that the multiplexing factor is negative. Source R3 represents a smooth

rate-variation scale source with low burstiness and high allowable load while source R29 combines a high burstiness with a moderate allowable load. Source R13 is a typical example of a source with high peak rate: the statistical multiplexing is possible only if there is a sufficiently large number of sources.

4.4.2.1 Superposition of burst scale sources with CBR load

Let us first examine the superposition of burst scale sources with CBR loads of 0.2 and 0.8 (note that because of the determination of EBV model it gives an exact result with an 0.5 CBR load). Figures 4.17, 4.18 and 4.19 reflect similar behaviour:

- with a CBR load of 0.2 EBV gives too high an allowable load for burst scale sources but the error is small;
- when a CBR load approaches 1, the real allowable load seems to be higher than that obtained by the EBV model.

The reason for the latter phenomenon is that the statistical multiplexing of deterministic sources is very efficient if the number of sources is small; the allowable load even approaches one if the number of sources is sufficiently small. Figure 4.20 provides a further insight into this phenomenon. If the number of sources B1 is at the most 20, the allowable load is one. It is possible to determine the effective bandwidth for any CBR load as:

$$k_i(\rho_{cbr}) = \frac{(1-\rho_{cbr})c}{N_i(\rho_{cbr})}, \tag{4.24}$$

where $N_i(\rho_{cbr})$ is the allowed number of sources *i* with CBR load ρ_{cbr}.

When ρ_{cbr} decreases from 1, at first $k_i(\rho_{cbr})$ is constant (= mean rate), after a certain limit which depends on buffer capacity the allowable load decreases rapidly and $k_i(\rho_{cbr})$ increases but later the decrease of the allowable load becomes smooth and finally $k_i(\rho_{cbr})$ may begin to decrease. This is the same phenomenon as Doshi (1993) has described: effective bandwidth is sometimes a non-monotonic function of a number of sources. Even EBV is unable to capture this behaviour because if we calculate $k_i(\rho_{cbr})$ from the EBV formula, we obtain either a monotonically increasing or decreasing function.

In order to avoid the problems when the number of sources is small (the point A in Figure 4.20) the use of negative values for parameter v_i^{**} in EBV model in practical implementations is not recommended.

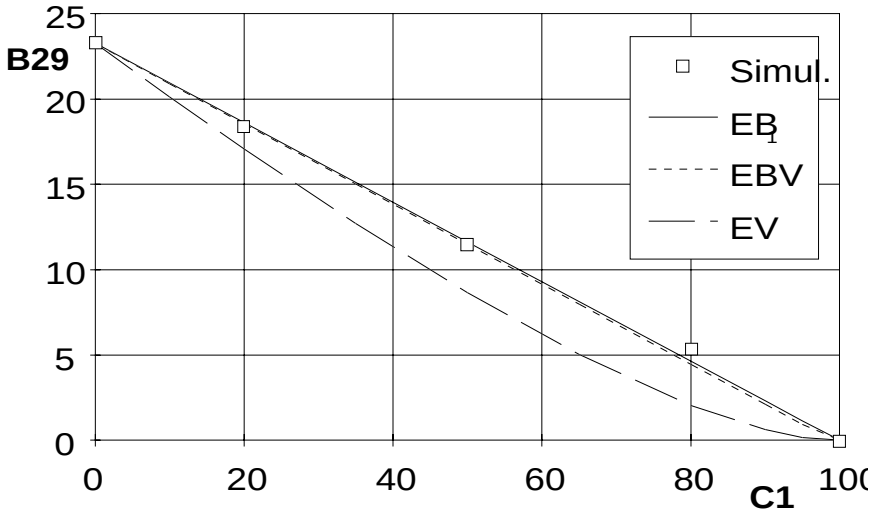

Figure 4.17. Allowable traffic mix of sources B29 and C1.

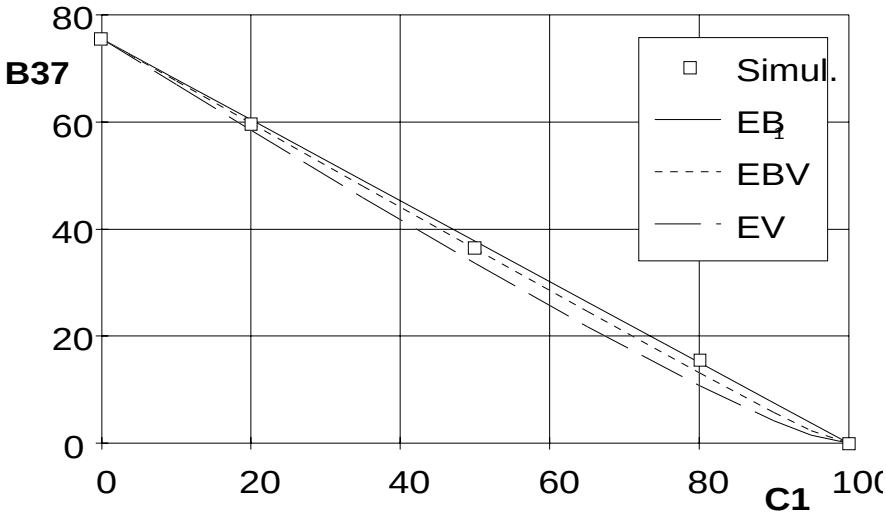

Figure 4.18. Allowable traffic mix of sources B37 and C1.

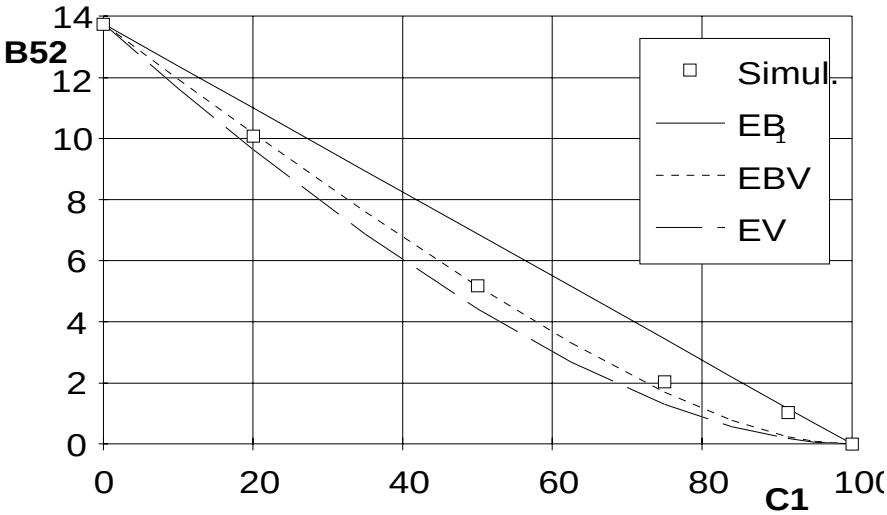

Figure 4.19. Allowable traffic mix of sources B52 and C1.

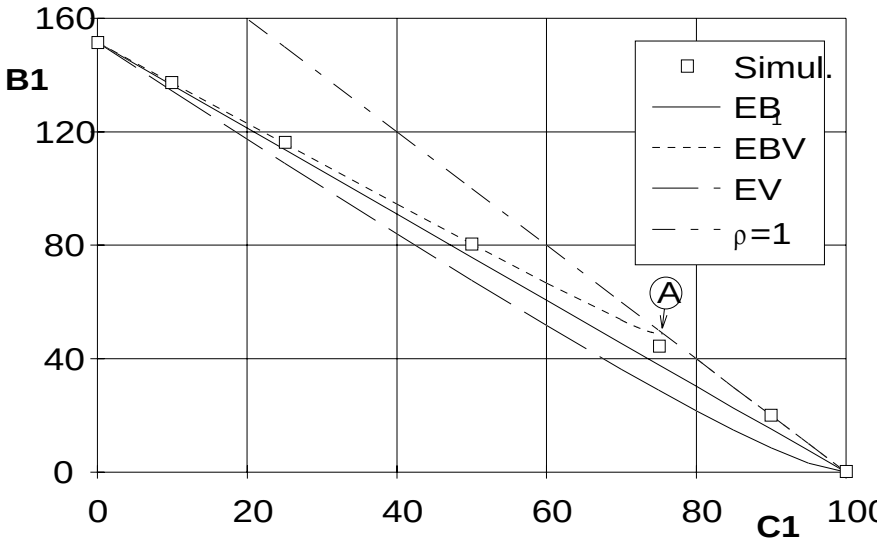

Figure 4.20. Allowable traffic mix of sources B1 and C1.

4.4.2.2 Superposition of burst scale sources with VBR sources

Since the parameters of the EBV model are determined by means of a traffic mix with a CBR load, presumably the most difficult situations occur when burst scale sources are mixed with rate-variation scale sources. In the following figures the nine combinations of burst scale sources B29, B37 and B52 with rate-variation scale sources R3, R13 and R29 are shown. The main result is that the accuracy of the EBV approximation is respectable, in particular when the load is evenly distributed between burst scale and rate-variation scale sources (the middle simulation point in each figure).

The general EBV formula (4.14) includes a free parameter γ. In this study we use value 1 partly because of the ease of solving parameters v_i^{**} and σ_i^{**}. If burst scale sources are aggregated with the CBR load, the results are almost independent of factor γ because the determination of source parameters is based on a traffic mix with the CBR load. In contrast, cases with both burst scale and rate-variation scale sources are not at all clear and we should examine whether any other choice gives a better approximation for the allowable traffic mix.

A choice γ=0.5 in (4.14) results in a simple formula:

$$\sum_i \left(m_i + \sigma_i^{***}\right) + \sqrt{\sum_i v_i^{***}} \le c\,. \tag{4.25}$$

But, as the Figure 4.24 shows, this choice leads to a considerable underestimation of the allowable load when burst scale and rate-variation scale sources are aggregated. An explanation for this phenomenon is that the choice γ=0.5 actually means a modification of the effective variance formula and therefore the characteristics of (4.25) are similar to those of the effective variance formula.

Figure 4.24 shows that the effect of increasing γ is slight, and in some cases the approximation with γ=2 is even better than that with γ=1 (i.e., with the EBV model). However, there are two strong reasons not to use a larger value than 1 for γ. Firstly, it results in increased probability that the allowed load will be overestimated, and

secondly, the implementation is more complicated if any other value than 0.5 or 1 is applied.

It is commonly held that to multiplex sources with very different characteristics is ineffective (e.g., Bonomi, Montagna & Paglino 1993). Nevertheless, we can observe especially from Figures 4.21, 4.24 and 4.27 that in some cases the superposition of burst scale sources and rate-variation scale sources results in a higher load than that given by a linear approximation gives. Thus in these cases a combining strategy for different source types yields more efficient multiplexing than a separation one. This can be explained by the fact that rate-variation scale sources exploit buffer capacity only intermittently while the allowed number of burst scale sources depends largely on buffer capacity. If the number of rate-variation scale sources is, for instance, 80% of the maximum value, the buffer capacity is free almost all the while for burst scale sources. But because the number of burst scale sources is limited, it is possible that the network is able to buffer all burst scale fluctuations. Then only rate-variation scale fluctuations are notable and the burst scale sources can be interpreted as a CBR load. This leads to an admission curve similar to cases in which burst scale sources are aggregated with CBR load (compare Figures 4.20 and 4.21).

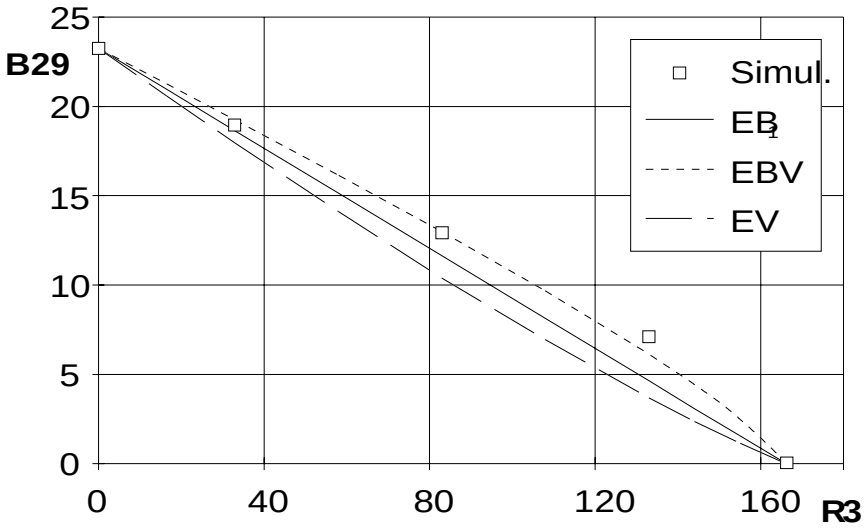

Figure 4.21. Allowable traffic mix of sources B29 and R3.

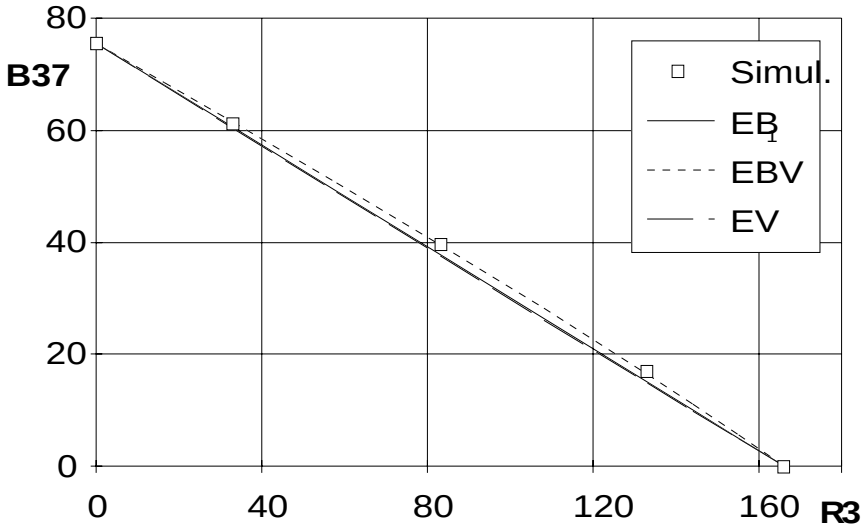

Figure 4.22. Allowable traffic mix of sources B37 and R3.

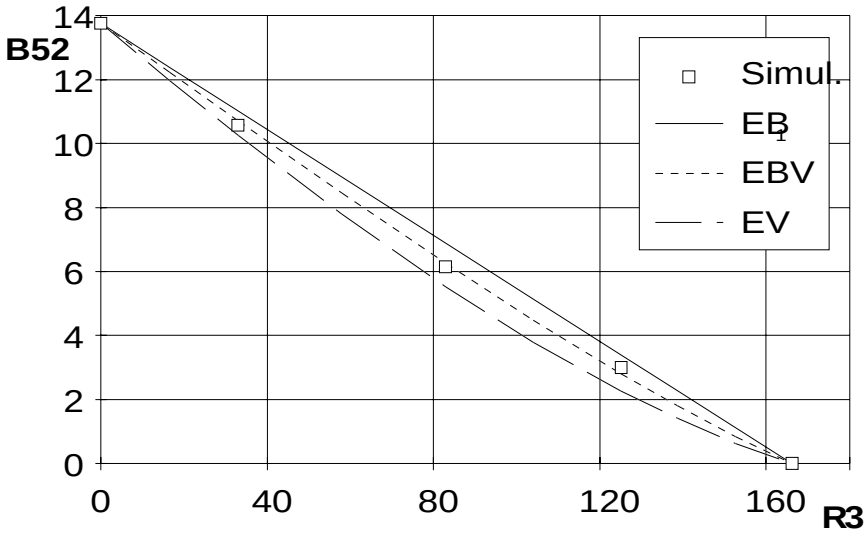

Figure 4.23. Allowable traffic mix of sources B52 and R3.

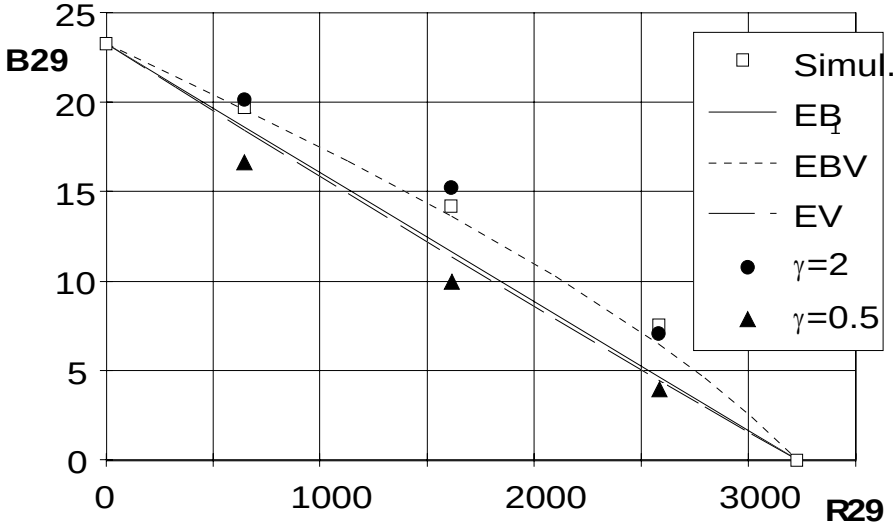

Figure 4.24. Allowable traffic mix of sources B29 and R29.

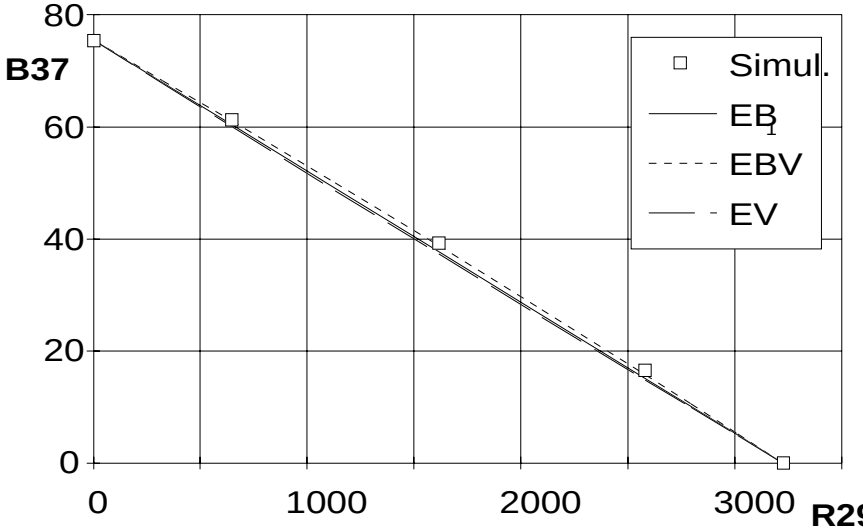

Figure 4.25. Allowable traffic mix of sources B37 and R29.

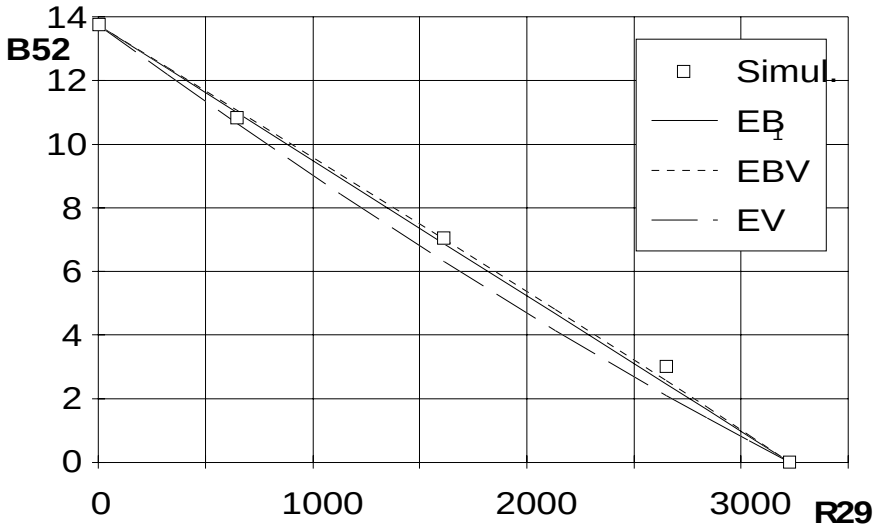

Figure 4.26. Allowable traffic mix of sources B52 and R29.

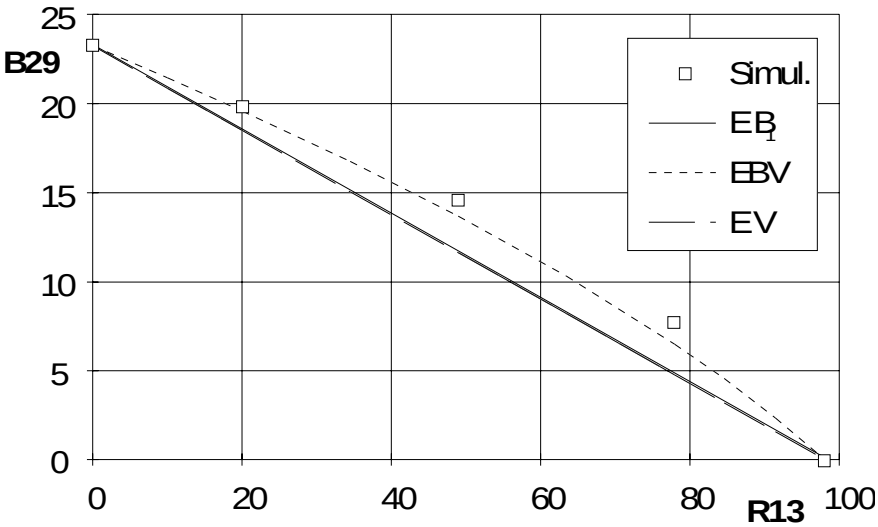

Figure 4.27. Allowable traffic mix of sources B29 and R13.

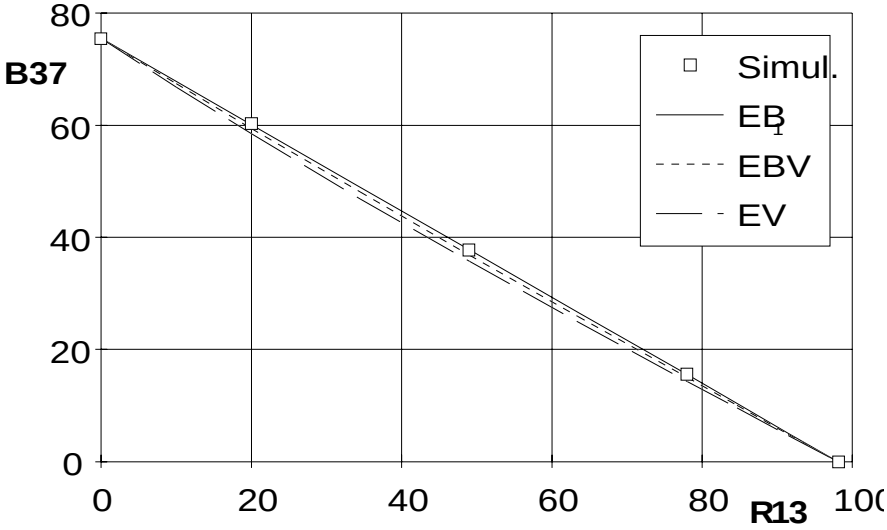

Figure 4.28. Allowable traffic mix of sources B37 and R13.

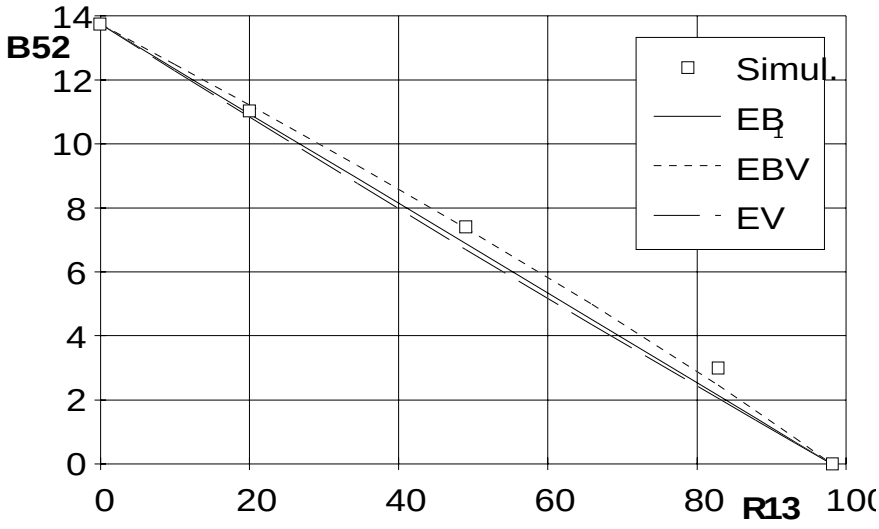

Figure 4.29. Allowable traffic mix of sources B52 and R13.

4.4.3 Rate-variation scale and effective variance

The applicability of effective variance for describing rate-variation scale processes is plausible on the ground of traffic models presented in Section 3.3. This section (together with Section 5.4) offers further evidence for this statement. In addition, the weaknesses of the effective variance model are identified.

Let us describe the difference between effective bandwidth and effective variance by a simple example using on/off sources with parameters p_{rv} = 0.1 and h = 1/20 (a similar example with a linear approximation and VBR sources of two types has been presented by Smit, 1993). If the required cell loss probability is 10^{-9}, the allowed number of sources is 50 according to (3.2). Now, if CBR load reserves 62% of the link capacity, the allowed number of sources obtained by the effective bandwidth model is:

$$N_{EB} = (1 - 0.62)/50 = 19.$$

In this case we can easily calculate the exact cell loss probability and the result is as high as $1.0\ 10^{-5}$. In contrast, when applying the effective variance model we obtain the corresponding values:

$$N_{EV} = 9 \text{ and } P_{loss} = 2.5\ 10^{-9}.$$

The superiority of effective variance is evident in the light of this example. However, there are several ways to alleviate this incompatibility problem of effective bandwidth, see in particular Sections 5.3.2 and 5.4.4.

Though the effective variance model has a mathematical basis, it is inherently an approximate model. We can see from the effective variance formula that it always gives the same allowable load for all sources that have the same allowable load in a homogeneous case. This property may cause considerable errors in certain cases. Let us take an example in which the allowed load in a homogeneous case is 0.50 for four different source types:

- h = 1/24, p_{rv} = 0.479;
- h = 1/35, p_{rv} = 0.395;
- h = 1/50, p_{rv} = 0.260;

- h = 1/70, p_{rv} = 0.058.

The admittance curve according to effective variance approximation is then identical for all the sources. The real allowed load as a function of a CBR load has been presented in Figure 4.30 together with the effective variance approximation.

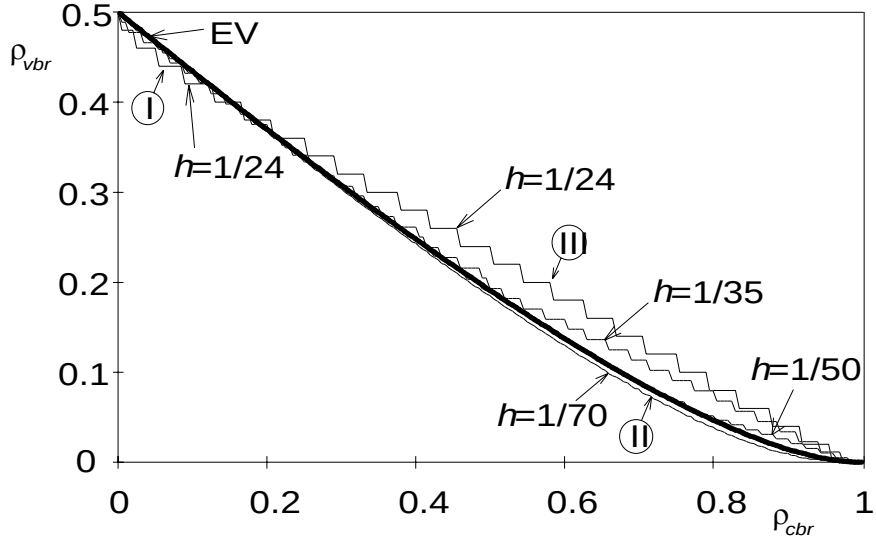

Figure 4.30. The allowable VBR load as a function of CBR load and effective variance approximation (EV).

The effective variance approximation is best when the number of sources is relatively large and *on*-probability (p_{rv}) is not very small. This result is understandable because the effective variance formula is based on Gaussian distribution and Gaussian distribution approximation is suitable if the number of independent variables is large and the distribution is symmetrical.

We can distinguish three types of approximation error:

I. overestimation of allowable load when the CBR load is small;

II. overestimation of allowable load when the CBR load is nearly one;

III. underestimation of allowable load.

Error type I occurs when the admittance curve parts from the linear curve of peak rate allocation at small values of the CBR load. In some rare events this type of approximation error may lead to a substantial exceeding of cell loss probability.

Error type II is caused by the asymmetry of real cell rate distribution, which is in contrast to the symmetry of Gaussian distribution. The reason for the overestimation is that the asymmetry is weaker when the number of sources is large and therefore the value that has been used as the basis of effective variance approximation does not capture the asymmetry, which, however, may be substantial when the number of sources is small. This phenomenon is strong in particular when traffic burstiness is very high.

The reason for error type III is similar to that of error type I: the allocation curve is linear if the peak rate is the only effective parameter. This type of error is not so serious as the previous ones because it results in an underestimation of allowed load and, therefore, in better Quality of Service for users. The under-utilisation can be partly avoided by omitting rate-variation scale fluctuations in source parameters, which means

that we replace the real mean rate by the peak rate and by that means obtain the same allocation curve as with peak rate allocation.

4.4.4 General traffic cases

The aim of this section is to assess the suitability of the EBV method for modelling complicated traffic processes. The following results are based on extensive evaluation of a wide variety of traffic cases using the simulation program presented in Section 3.6.2. In all, 133 sources from four classes have been used: CBR, burst scale, rate-variation scale, and combined burst scale and rate-variation scale. Source parameters including parameters for effective bandwidth, effective variance and EBV approximations are presented in Appendix A. In each of the 100 simulations four sources have been chosen at random and independently of each other (in a simulation all sources can be from one source class as well as from four classes). The changes of each source being chosen is as follows:

- cell scale sources (C1 - C3): 10% each;
- burst scale sources (B1 - B58): 0.52% each;
- rate-variation scale sources (R1 - R60): 0.50% each;
- combined sources (D1 - D12): 0.83% each.

This means that the proportions of C, B, R and D-sources in the traffic load are 30%, 30%, 30% and 10%, respectively. In all simulations the buffer size is 100 cells and the acceptable cell loss probability is 10^{-4}.

The number of sources is chosen with the aid of four evenly distributed random numbers x_i, i = 1,2,3,4. The number of offered sources of type j is:

$$N_j = \frac{x_j}{k_j \sum_{i=1}^{4} x_i},$$

where k_j is the effective bandwidth of source j according to formula (4.2).

The sources with the three smallest values for x_i are offered first in an ATM multiplexer and the number of sources with the largest x_i is then calculated according to four approximations: effective bandwidths (both EB_1 and EB_2), effective variance (EV) and the combined model (EBV). The calculation of effective bandwidth of the EB_2 model is based on effective variance approximation with ρ_{max} = 0.9 (see Section 5.3.2.3). The results of these approximations are compared with simulation results.

Table 4.4 shows the difference in allowed load between the approximations and simulation results. The allowed load has been calculated for real numbers of sources and consequently each simulation item consists of two separate simulations with instances both below and above the 10^{-4} cell loss level. The allowed load has been calculated by linear interpolation. The same technique has been applied when determining source parameters k_i, k_i^*, v_i^*, σ_i^{**} and v_i^{**}.

The superiority of EBV to the other models in complicated traffic cases is evident as far as approximation error is concerned. The standard deviation of relative error in the allowable load is 1.3% for EBV whereas the corresponding values for EB_1, EB_2 and EV

are 4.2%, 3.1% and 3.3%, respectively. Furthermore, the EBV model offers the best approximation in 73 out of 100 cases.

Table 4.4. Mean, standard deviation, minimum and maximum errors with models of effective bandwidth, effective variance and EBV

	simulation results		ρ(model *i*) - ρ(simulated)			
	ρ	Ψ *)	EB_1	EB_2 ρ_{max}=0.9	EBV	EV
average	0.785	0.987	0.012	-0.032	-0.005	-0.032
standard	0.104	0.040	0.033	0.024	0.010	0.026
minimum	0.500	0.893	-0.055	-0.129	-0.030	-0.149
maximum	1.000	1.082	0.105	0.006	0.042	0.005

*) Ψ (mixing efficiency) is defined in Section 5.3.2.4

Another important question, in addition to the average values, is for which cases each approximation is most suitable. In previous sections we have shown that effective variance is at its best when the multiplexing factor ε_m is small (i.e., at cell scale) and effective variance is better when ε_m is nearly 1 (at rate-variation scale). Since (4.19) is applicable only for single sources, a modified version for ε_m is applied in this connection:

$$\varepsilon_m^* = \frac{\sum_i N_i v_i^{**}}{\sum_i N_i v_i^{**} + \left(\sum_i N_i \sigma_i^{**} \right)^2}. \tag{4.26}$$

This modified formula for the multiplexing factor gives only a slightly different result from the original formula (4.19) in homogeneous cases. If ε_m^* is calculated for the maximum number of sources, the difference is less than 0.034 for all the sources presented in Appendix A with ε_m larger than -0.1.

Simulation results have been presented as a function of ε_m^* in Figure 4.31. As expected, effective variance is an adequate approximation if ε_m^* is nearly 1 but even a small decline of ε_m^* impairs the accuracy of effective variance model considerably. When the multiplexing factor ε_m^* is below 0.8, effective variance model gives too low load in nearly every case. In contrast, if ε_m^* is larger than 0.6, the effective bandwidth (EB_1) gives a substantially too high allowable load whereas with lower values it shows a moderate accuracy.

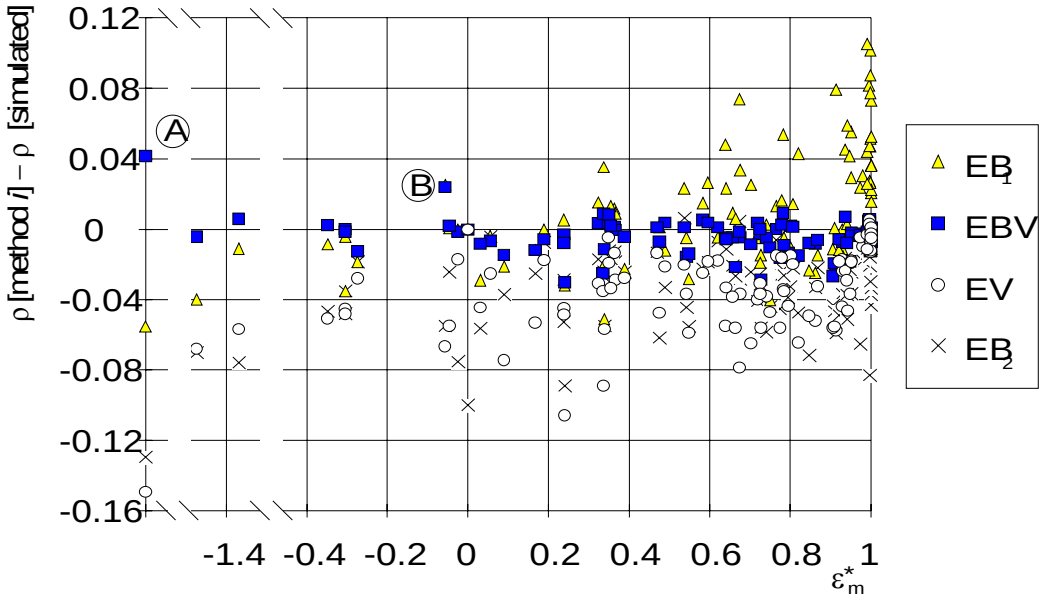

Figure 4.31. Difference in allowable load between approximate models and simulation results as a function of multiplexing factor ε_m^*.

The combined model, EBV, offers an excellent approximation for all values of ε_m^* apart from some special cases. There are two cases in the simulation material, marked with *A* and *B* in Figure 4.31, in which the EBV approximation fails to give a proper approximation of the allowed traffic mix. In both cases there is, in addition to CBR sources, only one other source type: B42 in case *A* and D2 in case *B*.

Case *A* is caused by the phenomenon depicted in Figure 4.20. Errors of this type arise when burst scale variations are so small that v_i^{**} is negative (in fact, even $\sum_i N_i v_i^{**} + \left(\sum_i N_i \sigma_i^{**} \right)^2$ is negative in this traffic case). There is an evident solution to the problem: to forbid negative values for parameter v_i^{**}. Although this limitation reduces to some degree the average load obtained by EBV, it is unavoidable in practical implementations.

Case *B* is related to the problem that arises when burst scale fluctuations and rate-variation scale fluctuations are combined in one source. Since the EBV model takes into account only two special cases (the homogeneous case and the superposition case with a 50% CBR load), it cannot entirely catch the complicated behaviour of combined sources. The information that we have about rate-variation scale fluctuations can help alleviate this problem. We can first omit burst scale fluctuations and calculate the allowed number of sources, for instance, by the large deviation approximation and then use (4.12) to calculate the effective variance. This value is used as v_i^{**} in (4.15) and (4.17) instead of (4.16) and finally σ_i^{**} is calculated by (4.17) using the value for $N_{c,i}$ that takes into account both burst scale and rate-variation scale fluctuations. After these two modifications the largest positive error (a too high allowable load) obtained by EBV from 100 simulations is less than 0.01.

Figure 4.32 shows the cell loss probability distribution that is obtained when the number of allowed sources is determined by the EBV model. Without the above-mentioned modifications the average cell loss probability in all 100 simulations is

8.25 10^{-5} and after the modifications the highest cell loss probability obtained is as low as 1.31 10^{-4}. There are certainly worse situations, especially if the 10^{-9} cell loss probability level is used, but bearing in mind that the difference in allowable load as compared with exact result is only 0.005, EBV can be regarded as an excellent technique to simplify the evaluation of complicated traffic cases.

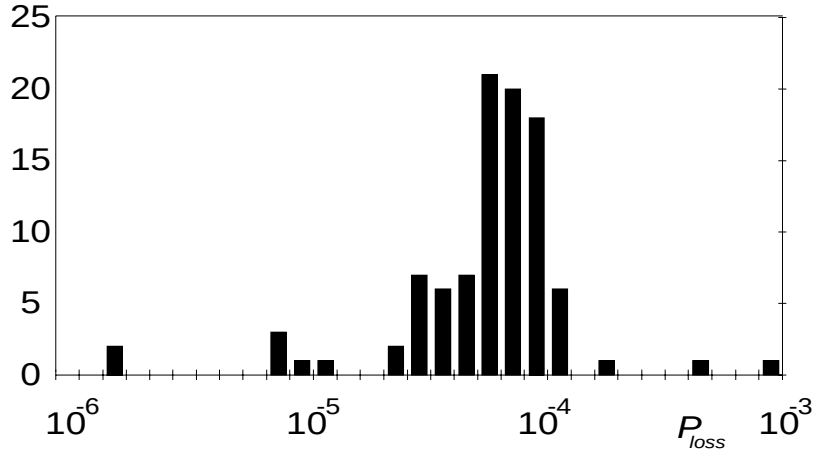

Figure 4.32. Cell loss probability distribution when allowable load is determined by EBV (100 cases).

4.4.5 Individual cell loss probabilities

The performance evaluation in the previous sections was based on the assumption that the QoS requirement is determined as an average cell loss probability from all sources (and even from all traffic cases). The reason for this is that the calculation of cell loss probability for all sources is far too complicated a process as far as practical implementations are concerned because when a connection is established or released, every individual cell loss probability changes at the same time.

One way to overcome this difficulty is to apply a tighter cell loss standard for aggregated traffic. Then we should know how much individual cell loss probabilities differ from each other and what factors have the largest influence on the differences. This issue has been studied by several authors. The general conclusion drawn from the studies is that individual loss probabilities are not a great problem, the only exception being when the streams have very different burstiness (e.g., Lindberger 1991; Virtamo & Norros 1991). However, in extreme cases the minority traffic with high burstiness experiences a loss probability which could be greater than the overall loss probability by two, three or even greater orders of magnitude (Yang & Li 1993).

Let us define $R\{i\}$ as:

$$R\{i\} = \frac{P_{loss}\{i\}}{P_{loss}\{CBR\}},$$

where $P_{loss}\{i\}$ is the cell loss probability of source i and $P_{loss}\{CBR\}$ is the cell loss probability of CBR traffic. Lindberger (1991) and Virtamo and Norros (1991) have obtained an approximation for the individual cell loss probability when load, cell loss probability and peakedness for an individual source (z_i) and for aggregated traffic (z) are known. From the formula proposed by Lindberger we can obtain the following approximation for $R\{i\}$:

$$R\{i\} = \frac{\left(\rho + (1-\rho) z_i / z \right) P_{loss}}{\rho P_{loss}}$$

$$= 1 + \frac{(1-\rho) z_i}{\rho z} \cdot \tag{4.27}$$

This approximation is only applicable as such with rate-variation scale models. However, a simple generalisation is possible on the basis of (3.3) which can be expressed in the following form:

$$\rho + \kappa \sqrt{\frac{z\rho}{c}} \leq 1. \tag{4.28}$$

If we know the allowable load, we can calculate an equivalent *z*:

$$z = \frac{c(1-\rho)^2}{\kappa^2 \rho}. \tag{4.29}$$

Finally we obtain $R\{i\}$ by applying (4.29) both to homogeneous case (load is $\rho_{hom,i}$) and to heterogeneous case (load is ρ):

$$R\{i\} = 1 + \frac{\left(1-\rho_{hom,i}\right)^2}{\rho_{hom,i}(1-\rho)} \cdot \tag{4.30}$$

It should be noted that this approximation is independent of parameters P_{loss}, κ and *c*. We can see that $R\{i\}$ is large if $\rho_{hom,i}$ is small and at the same time load ρ is near 1.

It is not evident whether (4.30) holds good for cell and burst scale sources. The simulation material used in previous sections offers the opportunity to evaluate individual loss probabilities in general traffic cases. From the 100 simulations we have picked out cases in which the CBR sources (C1 or C2) are aggregated with at least one burst scale or combined source. The value for $R\{i\}$ obtained by simulation is then compared with the value given by (4.30). Figure 4.33 shows the result as a function of the multiplexing factor in the heterogeneous case, ε_m^*.

Although there is no clear dependency between $R\{i\}$ and ε_m^*, the simulation material can be better fitted by a modification of (4.30):

$$R\{i\} = 1 + 0.77 \frac{\left(1-\rho_{hom,i}\right)^2}{\rho_{hom,i}(1-\rho)} \cdot \tag{4.31}$$

The result reveals the suitability of formula (4.30) (or the modified formula) for approximating the difference between individual cell loss probabilities with all source types. Moreover, only in some special cases the individual cell loss probability is so much higher than the average that the difference must be taken into account in QoS evaluation.

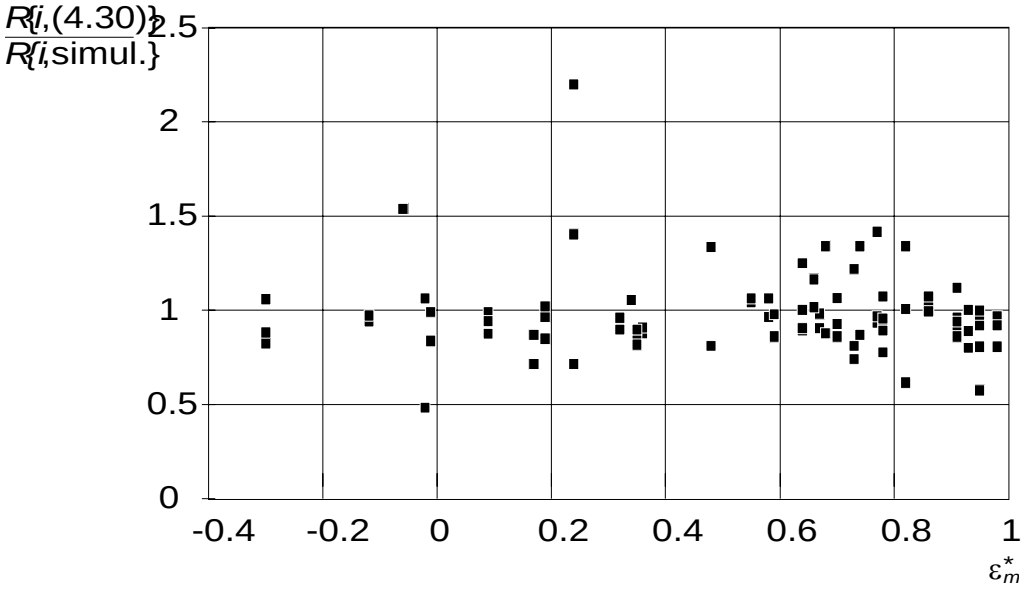

Figure 4.33. $R\{i,\text{formula (4.30)}\}/R\{i,\text{simulated}\}$ ratio as a function of multiplexing factor ε_m^*.

5 CONNECTION ADMISSION CONTROL

5.1 Framework

A general framework for the comparison of CAC-methods is depicted in Figure 5.1. This framework has been presented in the Eurescom project P105 (Eurescom 1993). The basis for CAC-methods is the Traffic Descriptor (X_{TD}). The Traffic Descriptor is the generic list of traffic parameters that can be used to capture the intrinsic characteristics of an ATM connection (ITU-T 1993a). Another important requirement is that it should be possible to control the parameters of the Traffic Descriptor. A typical Traffic Descriptor includes a mean cell rate and a peak cell rate.

The parameters that are used directly by a CAC algorithm are called here CAC-parameters (X_{CAC}). These parameters are delivered from the customer equipment to the management centre or to the network nodes. A typical CAC-parameter is an effective bandwidth. The conversion from Traffic Descriptor to CAC-parameters can be direct (functions $F_{TD\text{-}CAC}$) or it may contain several phases with intermediate parameters (X_I) and functions ($F_{TD\text{-}I}$ and $F_{I\text{-}CAC}$). It should be noted that this separation of CAC calculation into several phases (from $F_{TD\text{-}I}$ to $F_{A/R}$) follows on one hand the distinction between the mathematical and descriptive models presented in Figure 1.1 and on the other hand the distinction between the homogeneous and heterogeneous models. Complicated mathematical models are often usable for the calculation of some intermediate parameters, such as the allowed number of sources in homogeneous cases, but hardly for real time admission decision.

Further, functions $F_{TD\text{-}CAC}$ and $F_{TD\text{-}I}$ may require some knowledge of the network properties ($X_{N(U)}$) such as the link capacity, the buffer capacity, and the acceptable cell loss ratio. This separation (whether or not any information is needed about the network) is similar to the separation of derived and direct parameters (see Sections 4.1 and 4.2) and it is of great importance as regards the real implementation of CAC methods.

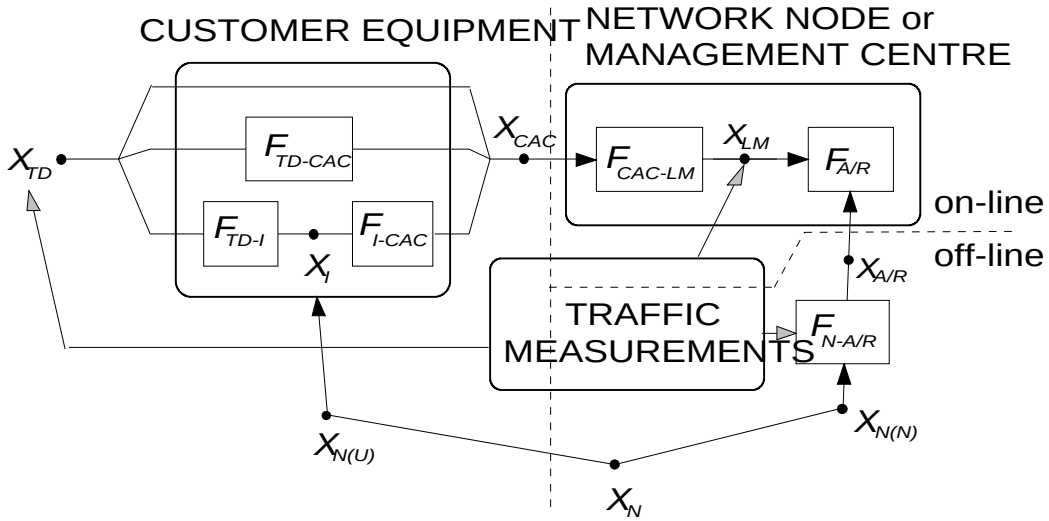

Figure 5.1. A framework for CAC methods.

The network node or the management centre maintains on each outgoing link (and if needed on the inside links of the ATM switching fabric) a link metric vector which consists of link metric parameters (X_{LM}). The link metric vector characterises the load

situation on a specific link in order to enable a simple and efficient CAC algorithm. The conversion function from a CAC to a link metric parameter is typically an addition:

$$X_{LM}[i;n+1] = X_{LM}[i;n] + X_{CAC}[i],$$

where $X_{LM}[i;n]$ is the value of the link metric parameter i before the request of a new connection and $X_{LM}[i;n+1]$ is the corresponding value after the request. More complicated functions are possible and, in addition, the link metric vector may contain some information on the actual link load (on-line measurements).

Each network node makes the decision of connection acceptance or rejection by a function ($F_{A/R}$) based on the instantaneous value of link metric parameters and on some parameters ($X_{A/R}$) which depend on network properties ($X_{N(N)}$). Typical $X_{A/R}$ parameters are ρ_{max} and κ (defined in Sections 4.2.1 and 3.3.2.2, respectively). Finally, the link metric vector should be updated when a connection is released.

Another scheme is that the customer equipment sends the Traffic Descriptor to the network without conversions and all calculations are made at the management centre or at the network nodes. This method provides an opportunity to use different CAC-methods in separate the ATM-nodes. Thus it may be practical always to send the original Traffic Descriptor to the ATM network. The main drawback of using the Traffic Descriptor as only CAC parameter lies in the complexity of conversions in many CAC-methods. The implementation may be too complicated because an ATM switch has to make a very fast acceptance/rejection decision whereas at the user interface the demand for a fast calculation is not so strict. Pre-calculated tables and off-line calculations may alleviate this problem.

5.2 Proposed methods

The CAC methods evaluated in this section are grouped either according to the approximation in heterogeneous cases (effective bandwidth, effective variance, combined models) or according to the implementation principle (convolution, measured flow, neural networks). Each section offers a brief review of references, the main characteristics of methods and example(s) of implementation.

5.2.1 Effective bandwidth

Several authors, such as Decina and Toniatti (1990), Dziong, Liao and Mason (1993), Elwalid and Mitra (1993), Gallassi, Rigolio and Fratta (1989), Griffiths (1990), Kelly (1991), Lindberger (1991) and Miyao (1993) have applied the concept of effective bandwidth in their CAC methods.

In the CAC method of Dziong et al. (1993) the calculation of parameters for each source type is based on rate-variation scale models, which results in problems because of non-linear behaviour when combining various source types (see Section 4.4.3). Dziong et al. have endeavoured to solve this problem by means of additional functions (formulae (3), (4) and (5) by Dziong et al.). The values of these functions depend on the actual traffic situation and therefore relatively complicated calculations are needed when connections are established and released. A similar approach is the class related bandwidth assignment rule proposed by Gallassi et al. (1989). With methods of this type, predefined source classes are necessary in order to achieve a simple implementation but, on the other hand, this classification is very restrictive in the context of ATM.

Lindberger (1991) has proposed an approximation for effective bandwidth (see Section 3.3.2.2). Because of the simplicity of the formula it is possible to make all calculations at the network nodes. As a result, the most likely solution using this method is that the Traffic Descriptor (m_i and h_i) is sent to the network and the effective bandwidth is calculated at every network node.

As regards the comparison with the CAC method in later sections, Lindberger's formula (3.5) can be interpreted as an EB_1 method by determining $\rho_{max} = 1/a$ and by replacing the original effective bandwidth by a new one $k_i' = k_i/a$. Lindberger's formula is, however, not a pure EB_1 method because k_i' is not only an approximation for homogeneous cases but also for heterogeneous cases. Although the formula is presented in connection with effective bandwidth, it is also a simple approximation for homogeneous cases, and therefore it can be used with other CAC-formulae.

5.2.2 Methods based on the variance of cell rate distribution

Various formulae using the variance of cell rate distribution have been applied by several authors, see for example Bermejo-Saez and Petit (1991), Guérin et al. (1991), Herzberg and Pitsillides (1993), Joos and Verbiest (1989), and Wallmeier and Hauber (1991). The method proposed by Bermejo-Saez and Petit is the same as formula (4.11) except that the load state of a connection is defined as the number of cells counted during a fixed length observation interval ΔT. Source parameters are:

- m_i = Max{$m_{\Delta T}$, all ΔT of connection i},
- v_i = Max{$v_{\Delta T}$, all ΔT of connection i}.

This definition is usable as far as the traffic control is concerned because the allowed behaviour of every source is exactly defined during a short period. However, the efficiency of this method is strongly dependent on the choice of the interval of observation.

In the method by Wallmeier and Hauber (1991) all connections are divided into two classes. The peak rate is allocated to connections of class I. The remaining bandwidth can be used for statistical multiplexing of class II connections (each connection is described by parameters m_{II}, h_{II} and v_{II}). Two different upper bounds has been presented for the variance of a source: the first one is based on an on/off model and the other one on the Gaussian model. A new connection is accepted if the sum of peak rates is less than the given level or there is enough bandwidth for statistical multiplexing and the effective variance formula allows the new connection. A new connection will be accepted (Wallmeier & Hauber):

$$\begin{aligned} &\textit{if} \\ &\quad \sum_i h_{I,i} + \sum_j h_{II,j} \le \rho_{max} c \\ &\textit{or} \\ &\quad \Big(\big(\rho^*_{max} c - \sum_i h_{I,i} \ge \rho_{II} c \big) \\ &\quad \textit{and} \\ &\quad \Big(\sum_j m_{II,j} + \kappa \sqrt{\sum_j v_{II,j}} + Max\{h_{II,j}\} \le \rho^*_{max} c - \sum_i h_{I,i} \Big) \Big). \end{aligned} \qquad (5.1)$$

The values of parameters κ, ρ_{II}, ρ_{max} and ρ^*_{max} have to be determined in advance. They depend on the overall allowable cell loss probability, the allowable cell loss probability due to cell level congestion and the definition of the two classes of connections. The basic restriction of this approach is the underlying traffic model, as the variance can be calculated only for the traffic process at rate-variation scale.

5.2.3 Combinations

The EBV model presented in Section 4.2.3 is an example of the combination of effective bandwidth and effective variance. A different approach has been applied by Guérin et al. (1991). Their method differs from the previous methods in the characterisation of traffic sources: the method itself assumes a burst scale traffic model. If both burst and idle periods are exponentially distributed, formula (3.1) can be used for the calculation of the needed bandwidth of a single separate source (k_i). A new connection is then accepted:

if

$$\sum_i k_i \le c$$

or

$$\sum_i m_i + \kappa\sqrt{\sum_i v_i} \le c\,, \tag{5.2}$$

where κ is obtained from (3.4). The first part of the acceptance procedure takes into account the burst scale behaviour of the traffic process. In homogeneous cases the sum of effective bandwidths (Σk_i) can be used as an upper limit for the needed bandwidth (Guérin et al. 1991). It should be stressed that this upper limit rule is not generally valid because with a deterministic source the effective bandwidth may be a non-monotonic function of the number of sources (see Section 4.4.2.1). For example, if the buffer size is larger than the burst size, the needed bandwidth of a single deterministic source is equal to mean rate. Consequently, (5.2) is not a suitable method for admission control if the traffic process of a single source is deterministic at burst scale.

The second part of the CAC procedure has been added to avoid overestimation of the needed bandwidth when the number of sources is great. In this part of the procedure Guérin et al. suppose that the traffic fluctuations are in rate-variation scale and that they can be modelled by Gaussian distribution. As we have earlier noticed, these assumptions lead to the formula (3.3).

5.2.4 Convolutions

Esaki (1992) and Saito (1992b) have presented CAC methods with a convolution algorithm. The determination of source descriptor in Esaki's method is based on a limited period T which is defined as the inverse of largest peak rate among all connections. In this case all sources can be presented by one parameter, the mean cell rate, which determines the probability that the connection produces a cell during the period T. This calculation may result in an inefficient use of network resources and therefore additional techniques are needed in real implementations, see appendices in (Esaki).

In the method proposed by Saito (1992b) the length of the observation interval is equal to the time at which $K/2$ cells are transmitted (K is buffer size in cells). Then the probability that exactly n cells arrive during this period is calculated by the aid of

convolutions and cell loss probability is obtained by a formula similar to (3.2) (in addition, traffic measurements can be applied, see Section 5.2.5). Each source is determined by two parameters: mean and maximum numbers during the observation interval.

However, approximation errors may occur because the maximum number of cells must be an integer. For example, if peak rate $h = c/40$, mean rate $m = h/10$, buffer size $K = 100$ and acceptable cell loss ratio $P_{loss} = 10^{-9}$, the exact allowed number of source is 159. If we apply Saito's method, the maximum number of cells during an observation period is 2, which means that the peak rate used in Saito's method is $c/25$ instead of $c/40$. After this modification the allowed number of sources is 112. Consequently, despite the theoretical accuracy of the convolution method, practical implementations may cause a considerable underestimation of allowable load.

5.2.5 Measured flow

Traffic measurements can be used for many purposes as Figure 5.1 illustrates: evaluation of traffic parameters, performance of UPC and CAC methods, etc. Measurements of these types have only a minor effect on the function and structure of CAC methods whereas on-line measurements may markedly affect the structure of the CAC procedure.

Some parameters describing the traffic process, such as mean rate and the intensity of variations, can be estimated with the aid of measuring results instead of the theoretical values which have been calculated from declared source parameters. The most important limitation of this approach is that it is very difficult to discover a suitable time scale for measuring. If there are long range fluctuations, for example because of scene changes of video source, the measuring period should be very long in order to capture all fluctuations. On the other hand, during a long measuring period connections will be established and released, and we cannot suppose that the behaviour of sources with rapid fluctuations remains unchanged.

An approach to combine traffic measurements and CAC has been proposed by Saito (1992a). Saito's method is based on an estimated distribution of the number of cells arriving during a renewal period, $\hat{\mathbf{p}}(t) = (\hat{p}(0;t), \hat{p}(1,t),...)$ and on the measured distribution of arrived cells during N periods, $\mathbf{q}(t) = (q(0;t), q(1;t),...)$. The estimated distribution for the period (t+1) is then:

$$\hat{\mathbf{p}}(t+1) = \alpha\mathbf{q}(t) + (1-\alpha)\hat{\mathbf{p}}(t). \tag{5.3}$$

When a new connection request is connected in the t^{th} renewal period and the maximum number of cells during a renewal period is R, the renewing procedure is:

$$\hat{p}(k;t+1) = \begin{cases} \hat{p}(k-R;t) & \text{for } k \geq R, \\ 0 & \text{for } k < R. \end{cases} \tag{5.4}$$

This equation means that the number of cells arriving from the new connection is a priori assumed to be R. The value R is given by the peak cell rate declared by the user. The CAC procedure applied has been presented in Section 5.2.4.

In addition, traffic measurement may be useful for detecting and predicting exceptional traffic events because according to traffic measurement the most congested periods are preceded by signs of impending danger (Fowler & Leland 1991).

5.2.6 Neural networks

Neural networks can be used for both source parameter determination and CAC procedure. In extreme cases every individual source has its own input to the neural network and this huge network makes the decision of connection admission. In this case the neural network should be connected directly to every user interface (points X_{TD} in Figure 5.1). This is not a practical solution because the number of sources may be very large and it is very difficult to train the neural network if there is a very large amount of source combinations.

The next approach is to implement the $F_{TD\text{-}CAC}$ box in Figure 5.1 by a neural network as in Takahashi and Hiramatsu (1990 Section 5.2). However, the same method that is required to train the neural network to recognise acceptable patterns can be used for the determination of effective bandwidth or effective variance of a conventional CAC. If predefined source types are used, there is presumably no reason to use a neural network for *on-line* calculation between X_{TD} and X_{CAC}.

Furthermore, the CAC procedure (functions $F_{CAC\text{-}LM}$ and $F_{A/R}$ in Figure 5.1) can be realised by a neural network as presented in Takahashi and Hiramatsu (1990 Section 5.3). In this case the system state seen by the network is $X = \{n_1, n_2, \ldots n_M\}$ where n_i denotes the number of active connections of type i in the system. According to Fritsch, Mittler and Tran-Gia (1992) the CAC problem can be formulated as a pattern recognition problem: upon recognition of the load pattern X, a yes/no decision has to be made to accept/reject the connection request. This pattern recognition replaces effective bandwidth or effective variance approximations in conventional CAC methods. In order to justify this application of neural network it should be either simpler or more efficient than conventional methods.

The last identified approach is a combination of on-line traffic measurements, neural networks and a suitable CAC procedure. In this scheme neural networks are applied to refine the measuring results regarding fluctuations of incoming traffic process, queue length, etc. The output of this process together with the parameter of the request source is then used as input for CAC procedure. This seems to be the most promising application of neural networks in traffic control of ATM networks.

5.3 Efficiency comparison

5.3.1 Selection of methods for analysis

Although the CAC methods presented in Section 5.2 contained many useful ideas for solving the CAC problem of ATM networks, they all have limitations. The main difficulty is related to the limitation of traffic models since most of the methods are valid only for rate-variation scale traffic models and, moreover, they are often tied to a certain technique of determining traffic parameters and to certain methods to approximate homogeneous cases. These underlying assumptions make it difficult to compare different types of CAC method and, as a result, the published comparisons are typically restricted in some aspects. For example, Pettersen (1993) has evaluated various approaches based on the large deviation approximation.

In this study we attempt to make as a general comparison as possible although there are certain limitations. The performance evaluation covers only rate-variation scale models because of the lack of simple and accurate methods for traffic models at burst scale.

Another reason for this restriction is that the burst scale models offer a significant gain in utilisation in comparison with rate-variation scale models only if the average burst size is smaller than buffer size (see Section 4.3). On the other hand, if the burst size is small, it seems to be more efficient to stretch the burst at user interface than to exploit the statistical gain at burst scale. If burst scale processes are utilised in CAC procedures, the results presented in Section 4.4.4 provide an insight into the efficiency of different CAC approaches.

Methods based on neural networks and measured traffic have been omitted in the following examination because the source description with these methods will apparently be dissimilar to that of the other methods. Convolutions have been applied only for the calculation of exact cell loss probabilities.

The traffic models presented in Section 4.2 offer an opportunity to achieve a general comparison because they make it possible to combine various homogeneous and heterogeneous models in numerous ways. The prime property of a CAC method is the principle used for the heterogeneous approximation; the main approaches are (note that EBV gives the same results as effective variance model at rate-variation scale):

- EB_1: effective bandwidth, the first version, formulae (4.2) and (4.3);
- EB_2: effective bandwidth, the second version, (4.6), (4.7);
- EV: effective variance (4.11), (4.12).

For homogeneous cases the following methods have been applied:

- GD: Gaussian distribution approximation (3.3), (3.4);
- LF: Lindberger's formula (3.5) with the modification presented in Section 5.2.1;
- LD: the large deviation approximation (3.8);
- KF: Kelly's formula (3.12).

In the case of EB_2 three different modifications are examined (see a detailed account in Section 5.3.2.3):

- KF: Kelly's formula;
- LD-EV: the large deviation is used for the homogeneous solution and an approximation based on effective variance is applied for the calculation of $\psi_{max,i}$;
- LD-LD: the large deviation is used for the homogeneous solution as well as for the calculation of $\psi_{max,i}$.

Using peak rate allocation (PR) and an exact formula as extreme cases, the following selection of CAC methods has been chosen for the comparison:

- PR;
- EB_1-LF, EB_1-LD;
- EB_2-KF, EB_2-LD-EV, EB_2-LD-LD;
- EV-GD, EV-LD;
- exact.

where the notation xx-yy-zz means:

- xx: the method used for approximation of heterogeneous cases;
- yy: the method used for approximation of homogeneous case;
- zz: the method used in the determination of $\psi_{max,i}$ (only with EB_2).

5.3.2 Application of regulation factors

As regards the CAC methods there are various sources of error. For example in EB_2-LD-EV method the homogeneous solution is based on some intrinsic traffic parameters (X_{TD} in Figure 5.1) of which we have in reality only an approximate knowledge. Though this phenomenon may be very important in real implementations, it is common in all CAC methods and presumably its effect is nearly the same in all methods. Therefore, this phenomenon is omitted in the following comparison of CAC methods.

The second cause of error is the method used in the determination of CAC parameters (X_{CAC} in Figure 5.1). The EB_2-LD-EV method consists of the large deviation approximation in homogeneous cases and the effective variance approximation used with the CBR load. The next cause of error is the heterogeneous approximation, the effective bandwidth in EB_2-LD-EV. In addition there are some other errors, such as the difference between individual cell loss probabilities (see Section 4.4.5), which are for the most part common to all CAC methods.

Since these causes of error may have a considerable effect on the obtained cell loss probabilities, every CAC method should have a proper technique for managing the errors so that the required QoS can be reached. The simplest technique is to regulate the maximum attainable load. This is a feasible solution with the EB_1, EV and EBV methods whereas the methods of EB_2 type methods are more difficult because the formulae (4.6) and (4.7) already include the factor for maximum load (ρ_{max}). Therefore the attainable load of the EB_2 methods should be regulated by adjusting the effective bandwidth of each source.

Moreover, it should be stressed that with all models applied in this section the allowed number of sources in homogeneous cases takes into account the effect of limited buffer capacity, and consequently there is no need to take the buffer capacity into account during the determination of ρ_{max}. This also means that the link capacity used in the performance calculation is the real maximum capacity on offer to ATM connections.

5.3.2.1 Factor ρ_{max} in EB_1, EV and EBV methods

A common problem of the traffic models presented in Section 4.2 is that if they are used as the basis for the CAC method, the obtained average QoS does not necessary fulfil the required QoS standard. This property is especially clear with EB_1 models because they may result in a much higher cell loss probability than what is required by the rate-variation scale traffic process. The other methods also fail to model the behaviour of some complicated traffic cases. Therefore we need systematic tools to regulate the allowed number of sources for each model in order to obtain the desired value for cell loss probability.

The simplest way to regulate the allowed number of sources is to add an extra factor (ρ_{max}) that determines the maximum load in all possible traffic cases (see Section 4.2.1).

Using this factor the first version of the effective bandwidth method (EB_1) can be presented in the following form:

$$\sum_i k_i \le \rho_{max} c, \tag{5.5}$$

where k_i is determined by (4.2). The same principle can be applied to the effective variance and EBV models:

$$\sum_i m_i + \sqrt{\sum_i v_i^*} \le \rho_{max} c, \tag{5.6}$$

$$\sum_i m_i + \sqrt{\left(\left(\sum_i \sigma_i^{**}\right)^2 + \sum_i v_i^{**}\right)^+} \le \rho_{max} c. \tag{5.7}$$

Source parameters v_i^*, v_i^{**} and σ_i^{**} are obtained from (4.12), (4.16) and (4.17), respectively (in practice, v_i^{**} should be 0 if (4.16) gives a negative value for v_i^{**}). By changing ρ_{max} it is always possible to obtain the desired level for the average cell loss probability.

5.3.2.2 Factors ψ_{adj} in EB_2 methods

The second effective bandwidth model (EB_2) is more problematic than the other models because it already contains the factor ρ_{max} and, in addition, ρ_{max} has a considerable influence on the effective bandwidth of each source. Therefore it is difficult to use ρ_{max} as a regulating parameter, but it is still possible to keep the definition of ρ_{max} unchanged by regulating the values of the effective bandwidth with an additional factor ψ_{adj}. We obtain the following formula for the effective bandwidth (see Figure 5.2):

$$k_i^* = \max\left\{\frac{\rho_{max} c}{\psi_{adj}\psi_{max,i} N_{c,i}}, m_i\right\}. \tag{5.8}$$

The determination of factor $\psi_{max,i}$ is similar to the original formula (4.6), only the factor ψ_{adj} is added to the denominator of (4.6). The formula for acceptance decision is the same as that of EB_1 method:

$$\sum_i k_i^* \le \rho_{max} c. \tag{5.9}$$

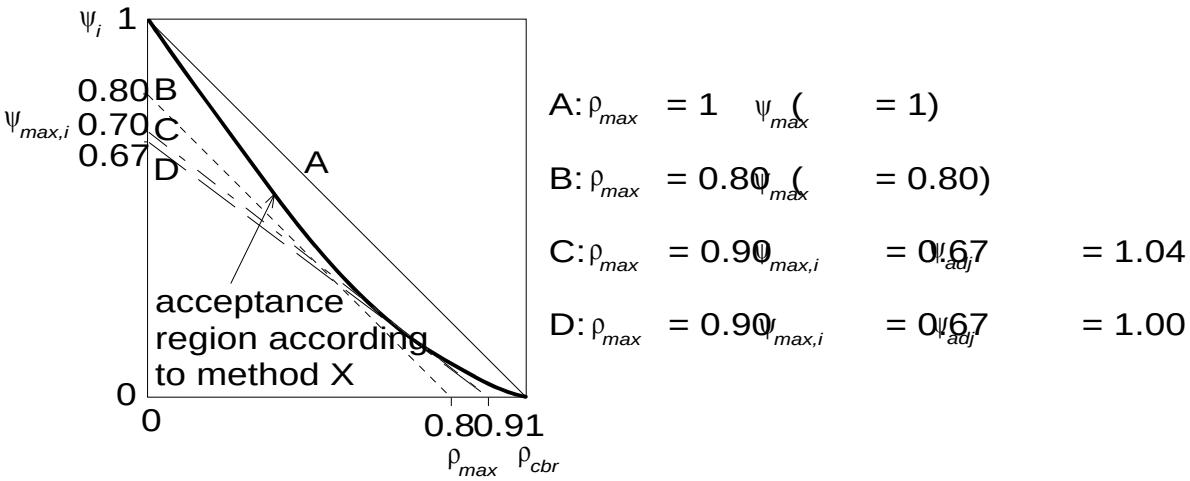

Figure 5.2. The regulating parameters of EB_1 methods (cases *A* and *B*) and EB_2 methods (cases *C* and *D*).

5.3.2.3 The determination of $\psi_{max,i}$ in EB_2 methods by means of effective variance model

With regards to EB_2 methods the homogeneous case can be solved by any appropriate method, for instance by the large deviation approximation. In the case of mixing with a CBR load we have two approaches: an independent calculation of each case with a different CBR load (EB_2-LD-LD method), and the use of a homogeneous case as a starting point. In the latter approach it is possible to apply the effective variance model for the approximation of heterogeneous cases because it offers a good approximation for the real acceptance region in the case of rate-variation scale traffic models (EB_2-LD-EV method). The effective variance model (4.11) can be written in the following form when sources of type *i* are multiplexed with a CBR load (ρ_{cbr}):

$$\rho_{hom,i}\psi_i + \left(1-\rho_{hom,i}\right)\sqrt{\psi_i} + \rho_{cbr} \le 1, \tag{5.10}$$

where $\rho_{hom,i} = \dfrac{mN_{c,i}}{c}$ and $\psi_i = \dfrac{N_i}{N_{c,i}}$.

Formula (5.10) has a considerable advantage, namely, the value of factor ψ_i depends only on $\rho_{hom,i}$ and ρ_{cbr}, and, consequently, factor $\psi_{max,i}$ depends merely on ρ_{max} and $\rho_{hom,i}$. Therefore it is possible to use relatively small pre-calculated tables in practical implementations.

Figure 5.3 depicts the dependency between $\psi_{max,i}$ and $\rho_{hom,i}$ with four different values of ρ_{max}. If the allowed load in the homogeneous case is high, a high value of ρ_{max} is advantageous and, correspondingly, with a small $\rho_{hom,i}$ a small ρ_{max} is recommendable. The choice ρ_{max} = 0.8 seems to be appropriate to a wide range of source parameters. This inference is confirmed by the examination presented in Section 5.4.

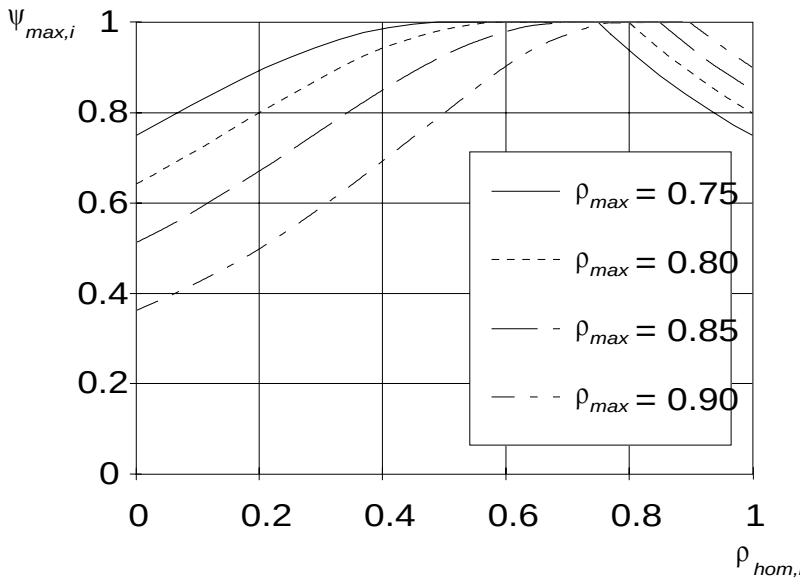

Figure 5.3. Factor $\psi_{max,i}$ as function of the allowed homogeneous load for four values of ρ_{max}.

It should be noted that Figure 5.3 illustrates a fundamental property of the rate-variation scale traffic process and not even detailed information on the traffic process, such as a complicated distribution for the needed cell rate, has any significant influence on the result.

At rate-variation the behaviour of ψ_i is in most cases defined quite accurately by $\rho_{hom,i}$ (compare figure 4.30). In contrast, with cell and burst scale models the traffic process is different and it is not possible to apply a formula similar to (5.10) because the primary parameter as regards the multiplexing process is not $\rho_{hom,i}$ but the $N_{c/2,i}/N_{c,i}$ ratio (see Figure 4.2). However, for cell and burst scale sources we can define an effective $\rho_{hom,i}$ based on the $N_{c/2,i}/N_{c,i}$ ratio and the effective variance approximation (i.e., formulae (5.10)):

$$\rho^*_{hom,i} = \frac{1-\frac{1}{2}\sqrt{N_{c,i}/N_{c/2,i}}}{1-\sqrt{N_{c/2,i}/N_{c,i}}}. \tag{5.11}$$

In practical use of (5.11) it is better to avoid values below 0 and above 1 by determining:

- $\rho^*_{hom,i} = 0$ when $N_{c/2,i}/N_{c,i} < 0.25$;
- $\rho^*_{hom,i} = 1$ when $N_{c/2,i}/N_{c,i} > 0.5$.

The idea of (5.11) is that the shape of the acceptance curve is exactly determined by $\rho_{hom,i}$ (or by $N_{c/2,i}/N_{c,i}$ ratio) if the effective variance model is used. If we know the $N_{c/2,i}/N_{c,i}$ ratio for another type of source, we can calculate an effective $\rho_{hom,i}$ and then use the corresponding rate-variation scale model as an approximation during the determination of $\psi_{max,i}$.

5.3.2.4 Optimisation of ρ_{max} in EB_2 methods

Although the factor ρ_{max} in EB_2 methods is not used for regulating the average cell loss probability level, it is used as another type of regulation factor. As Figure 5.3 depicts, if the average load (in a homogeneous case) is low and at the same time ρ_{max} is high, the resultant efficiency might be much lower than the maximum efficiency. In order to avoid the inefficient use of resources, we need a simple and efficient way to optimise ρ_{max} in a wide variety of traffic cases. This section attempts to capture the essence of the problem by using a simple approximation for the traffic process and thus to develop algorithms to ascertain the optimum value for ρ_{max}.

As can be seen from Figure 5.3 the factor $\psi_{max,i}$ has the lowest value when $\rho_{hom,i}$ is small. In this case we can obtain a simple relation between ρ_{cbr} and ψ_i, and by that means we can find an approximation for the optimum ρ_{max}. When CBR traffic needs a proportion ρ_{cbr} of the link capacity, we obtain ψ_i from (5.10):

$$\psi^*_{i,EV}(\rho_{cbr}) \le (1-\rho_{cbr})^2, \tag{5.12}$$

where the asterisk (*) refers to the system with VBR sources with very small $\rho_{hom,i}$. As (5.12) is a second order function, we can easily obtain the solution for $\psi^*_{max,i}$ (see Section 4.2.1 and Figure 4.1):

$$\psi^*_{max,i} = \begin{cases} 1 & \text{for} \quad \rho_{max} \le 0.5, \\ 4\rho_{max}(1-\rho_{max}) & \text{for} \quad 0.5 < \rho_{max} < 1, \\ 0 & \text{for} \quad \rho_{max} \ge 1. \end{cases} \tag{5.13}$$

We can the apply (5.13) and the EB_2 model in order to determine the allowed number of VBR sources as a function of ρ_{cbr}:

$$\psi^*_{i,EB2}(\rho_{cbr}) \le \psi^*_{max,i}\left(1-\frac{\rho_{cbr}}{\rho_{max}}\right). \tag{5.14}$$

We can attempt to maximise the achievable *load* using (5.14). However, the result is not at all satisfactory because the load induced by VBR connections were presumed to be very small even in a homogeneous case; in fact, the maximisation of an average load in this case means only that the CBR load is maximised. Consequently, we need a different approach. Let us define a new concept, mixing efficiency:

$$\begin{aligned} \Psi\{l\} &= \frac{1}{M}\sum_{j=1}^{M}\sum_{i}\frac{N_{i,j}\{l\}}{N_{c,i}} \\ &= \frac{1}{M}\sum_{j=1}^{M}\sum_{i}\psi_{i,j}\{l\}\ , \end{aligned} \tag{5.15}$$

where $N_{c,i}$ is the allowed number of sources of type i in a homogeneous case (using the best available method to solve the homogeneous case) and $N_{i,j}\{l\}$ is the allowed number of sources of type i in traffic case j according to method l. This factor depicts the real efficiency of a CAC method provided that the charging of an ATM connection is based on the effective bandwidth of each connection rather than the number of transmitted cells. In other words, by $\Psi\{l\}$ we can compare the revenues achieved by different CAC-

methods if the charging is based on the effective bandwidths determined in a homogeneous case.

Now we can maximise the mixing efficiency when $\rho_{hom,i}$ of VBR sources is small. We obtain the mixing efficiency as a function of a CBR load by combining (5.13), (5.14) and (5.15):

$$\begin{aligned}\Psi^*_{EB2}(\rho_{cbr}) &= \psi^*_{max,i}\left(1-\frac{\rho_{cbr}}{\rho_{max}}\right)+\rho_{cbr} \\ &= 4(1-\rho_{max})(\rho_{max}-\rho_{cbr})+\rho_{cbr}\,. \end{aligned} \tag{5.16}$$

The maximum of mixing efficiency is obtained when:

$$\rho^*_{max,opt} = \frac{1+\rho_{cbr}}{2}. \tag{5.17}$$

In fact, we can obtain this result by using the tangent of the allocation formula (5.12) at point (ρ_{cbr}, $(1\text{-}\rho_{cbr})^2$). However, formulae (5.16) and (5.17) can be applied more generally than merely for this special traffic mix. Since (5.16) is a linear function of ρ_{cbr}, ρ_{cbr} can be interpreted as an average value of a CBR load and still keep the analysis valid. It should be stressed that (5.17) results in a maximum mixing efficiency but it does not maximise the load.

Although the optimum choice of ρ_{max} is clear in this simple case, the prime target of this section is to find a simple way to determine the optimum ρ_{max} for a general source combination. For this purpose we should find a parameter which determines the equivalence of various traffic combinations in terms of the optimisation of ρ_{max}. There are many alternatives. The average *allowable load in different heterogeneous cases*:

$$\rho_{het,ave} = \tfrac{1}{M}\sum_{j=1}^{M}\sum_{i} m_i N_{i,j}$$

leads to a simple formula

$$\rho^*_{max}\{het\} = \frac{1+\rho_{het,ave}}{2}\,, \tag{5.18}$$

where $N_{i,j}$ is the allowed number of sources of type i in traffic case j. Note that the product m_iN_i of VBR traffic has been presumed to be very small.

There are several problems as regards (5.18). Firstly, the influence of sources with a low $\rho_{hom,i}$ might be too small in relation to the proportion of these sources in the revenues. Moreover, the definition of $\rho_{het,ave}$ depends on the method used in the determination of the allowed number of sources and, finally, we can apply $\rho_{het,ave}$ only with rate-variation scale sources. In order to avoid these problems, let us define the average value of *allowable load in a homogeneous case* as:

$$\rho_{hom,ave} = \frac{\sum_i \frac{N_i \rho_{hom,i}}{N_{c,i}}}{\sum_i \frac{N_i}{N_{c,i}}} = \frac{\sum_i \psi_i \rho_{hom,i}}{\sum_i \psi_i}, \quad (5.19)$$

where N_i is the number of sources i over all traffic cases under consideration. Because in (5.19) the number of sources is weighted by ψ_i, we can assume that $\rho_{hom,ave}$ is more suitable for optimising ρ_{max} than $\rho_{het,ave}$ if the charging is based on effective bandwidths.

If we again suppose that the $\rho_{hom,i}$ of VBR sources is small, we obtain the following $\rho_{hom,ave}$ by using the effective variance approximation, (5.12), and (5.19):

$$\rho^*_{hom,ave} = \frac{\rho_{cbr}}{1 - \rho_{cbr} + \rho^2_{cbr}}, \quad (5.20)$$

since we supposed that with VBR sources $\rho_{hom,i} \approx 0$ and with CBR load $\rho_{hom,i} = 1$.

Finally we can solve ρ_{max} from (5.17) and (5.20):

$$\rho^*_{max}\{hom\} = \frac{3\rho^*_{hom,ave} + 1 - \sqrt{1 + 2\rho^*_{hom,ave} - 3\left(\rho^*_{hom,ave}\right)^2}}{4\rho^*_{hom,ave}}. \quad (5.21)$$

We can apply this result as a general approximation for the optimum ρ_{max}. Firstly we calculate $\rho_{hom,ave}$ for a given traffic combination. Then we replace the original traffic combination with a simple approximation (CBR load and VBR sources with small $\rho_{hom,i}$) that has the same $\rho_{hom,ave}$. Finally, (5.21) provides an approximation for the optimum value of ρ_{max} in the original system. Figure 5.4 shows both approximations for optimum ρ_{max}, (5.18) and (5.21). The suitability of these approximations for optimum is ρ_{max} is investigated in Section 5.4.4.

It should be noted that with other types of model the primary parameter is not $\rho_{hom,i}$ but the $N_{c/2,i}/N_{c,i}$ ratio and therefore (5.11) should be used instead of the direct application of $\rho_{hom,i}$.

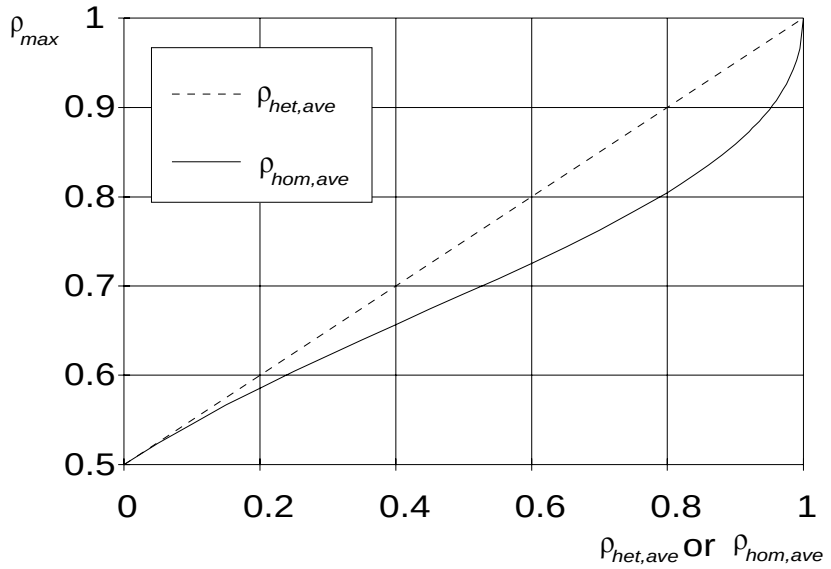

Figure 5.4. Optimum ρ_{max} as a function of $\rho_{het,ave}$ and $\rho_{hom,ave}$.

5.3.2.5 Factor ρ_{max} in Kelly's formula

As mentioned in Section 3.3.2.3, Kelly's formula (3.12) has one free parameter, β^*. Since (3.12) determines an effective bandwidth in the same way as (5.5), it is obvious that the free parameters of these two methods, ρ_{max} and β^*, have a fixed relation. Using the requirement that the effective bandwidth of a CBR source should be equal to peak rate, we obtain:

$$\beta^* = \frac{-\ln(P_{loss})}{(1-\rho_{max})c},$$

$$k_i^* = \frac{1}{\beta^*}\ln\left(\mathrm{E}\left\{e^{\beta^* \lambda}\right\}\right). \tag{5.22}$$

Since Kelly's formula takes into account the properties of each source of a traffic mix, it can be grouped as an EB_2 formula and consequently it contains, in a sense, a similar factor to $\psi_{max,i}$. Although this parameter can be calculated backwards using homogeneous and heterogeneous cases, it is by no means necessary for the application of Kelly's formula.

5.3.3 Criteria for comparison

There are two important requirements for the criteria used to compare CAC methods. Firstly, for the purpose of a fair comparison the same criterion should be applicable to all CAC methods, and secondly, the comparison should make it possible to find the weaknesses of each method. In consequence, a very wide range of source types is needed. It should be stressed that the criteria presented in this section are not intended to be used in the performance evaluation of real ATM traffic but only as a basis of comparison of different CAC methods.

The main criterion for the following comparison is the efficient use of network resources. In addition, the result of the comparison depends on the determination of the QoS offered to customers. In this study we use two different criteria for QoS:

Mean: the average cell loss probability of *M* cases must fulfil the following condition:

$$Mean\{P_{loss}\} = \frac{\sum_{j=1}^{M} \rho_j P_{loss,j}}{\sum_{j=1}^{M} \rho_j} \le P_{req},$$

where ρ_j is the allowed offered load in a traffic case *j* and P_{req} is the required cell loss probability.

Max: the average cell loss probability should be less than the required cell loss probability P_{req}:

$P_{loss,i} \le P_{req}$ among all examined cases (but it is not intended to find the most difficult cases of all possible traffic mixes).

The mean-criterion allows some occasional degradation in QoS while guaranteeing the average QoS during a relatively long period (e.g., over some minutes) whereas the max-criterion is more sensitive to excessive cell losses (although during a short period the actual cell loss probability may exceed the given level). It is possible to combine these two criteria by determining a maximum value for the highest allowed cell loss probability, for example one or two orders of magnitude higher than that of allowed mean value.

An issue of great importance is what we attempt to optimise. The average value of allowable load may seem to be a natural choice. However, this choice is feasible only if the charging in ATM networks is based on the total amount of cells delivered. This is not a probable charging policy in ATM networks; a more probable scheme is that the effective bandwidths of connections form the basis of charging. In this case the starting point is the number of sources that can be aggregated in homogeneous cases (or with a typical background traffic). As regards the comparison of CAC methods, this criterion means that we compare the abilities of CAC methods to mix various source types. This comparison can be made with the aid of the mixing efficiency defined in Section 5.3.2.4. Furthermore, the concept of mixing efficiency is suitable for assessing whether it is more useful to separate different source types on different links than to mix them. If the mixing efficiency is less than one, the separation principle is advantageous, whereas if it is higher than one, the mixing of different source types results in higher efficiency.

5.4 Comparison with rate-variation scale traffic

In this section we attempt to clarify the properties of CAC methods in different traffic situations: homogeneous traffic, cases with VBR sources of one type aggregated with a CBR load, and the superposition of up to four different source types. In addition, the optimisation formulae for ρ_{max} in EB_2 methods are assessed with simulation results.

The investigation of CAC methods is based on rate-variation scale sources, both on/off sources (from R1 to R30 in Appendix A) and sources with three cell rate levels (from R31 to R60). The cell loss probability standard is presumed to be 10^{-9} (note that in the appendix derived parameters are calculated for $P_{loss} = 10^{-4}$ which is the P_{loss} level used in all simulations). The results of investigation have been presented in tables from 5.1 to 5.8. In all tables the following items are presented: the definition of the CAC method

using the notation presented in Section 5.3.1, parameters ρ_{max} and ψ_{adj}, the average of allowable load, average value of cell loss probability, the largest value for cell loss probability, and the number of cases in which P_{loss} is in a certain order of magnitude.

5.4.1 Homogeneous cases

The first stage is to evaluate the accuracy of the different methods in homogeneous cases. The results with on/off sources and 3-level sources are presented in Tables 5.1 and 5.2, respectively. With Gaussian distribution and Lindberger's formula parameter ρ_{max} has been chosen so that either mean or max-criterion is satisfied; the exception is Kelly's formula which has no regulation parameter. The regulation parameter, either ρ_{max} or ψ_{adj}, is determined with an accuracy of 0.01 (e.g., the EV-GD method with ρ_{max} = 0.96 gives an average cell loss probability of 7.98 10^{-10} whereas if ρ_{max} is 0.97, the average cell loss probability is higher than 10^{-9}).

The results show the remarkable accuracy of the large deviation approximation; there is evidently no reason to use the exact formula (3.2) for calculating cell loss probability in homogeneous cases. This inference is even clearer with complicated heterogeneous cases because the implementation of the large deviation approximation is essentially simpler than that of the exact formula.

Gaussian distribution approximation leads roughly to a 8% lower load than the exact method when mean-criterion is used but it is inefficient with max-criterion. Lindberger's approximation is slightly better than Gaussian distribution approximation in respect of allowed load, Max{P_{loss}} and the width of P_{loss}-distribution, especially with on/off sources. In contrast, with 3-level sources parameter ρ_{max} in Lindberger's formula has to be rather low (0.78) because of some difficult sources (those cases are beyond the range in which Lindberger's approximation had originally been planned).

Kelly's formula yields a substantially lower load than other approximations with a homogeneous load if mean-criterion is applied. The strength of Kelly's formula emerges predominantly in heterogeneous cases; this property is typical of all EB_2 approximations and, to some degree, of Lindberger's approximation.

Table 5.1. The accuracy of approximations for cell loss probability in homogeneous cases, 30 on/off sources (R1-R30), $P_{loss} = 10^{-9}$

Method			ρ_{max}	ρ	P_{loss}	P_{loss}			<10^{-8}	<10^{-9}	<10^{-10}
het.	hom.	crit.		mean	mean	max	>10^{-8}	>10^{-9}	>10^{-10}	>10^{-11}	
PR	PR	max	1.00	0.172	0	0	0	0	0	0	
EV	GD	mean	0.96	0.535	7.98 10^{-10}	2.01 10^{-8}	3	1	5	5	
		max	0.85	0.459	2.77 10^{-11}	8.56 10^{-10}	0	0	1	2	
EB	LF	mean	0.86	0.569	6.96 10^{-10}	5.54 10^{-9}	0	10	10	4	
		max	0.77	0.509	2.98 10^{-11}	8.91 10^{-10}	0	0	3	4	
EB_2	KF	max	0.80	0.486	1.88 10^{-13}	8.25 10^{-13}	0	0	0	0	
-	LD	max	1.00	0.579	7.79 10^{-10}	8.92 10^{-10}	0	0	29	0	
-	Ex.	max	1.00	0.581	9.21 10^{-10}	9.99 10^{-10}	0	0	29	0	

Table 5.2. The accuracy of approximations for cell loss probability in homogeneous cases, 30 sources with three cell rate levels (R31-R60), $P_{loss} = 10^{-9}$

Method			ρ_{max}	ρ	P_{loss}	P_{loss}		<10^{-8}	<10^{-9}	<10^{-10}
het.	hom.	crit.		mean	mean	max	>10^{-8}	>10^{-9}	>10^{-10}	>10^{-11}
PR	PR	max	1.00	0.174	0	0	0	0	0	0
EV	GD	mean	0.94	0.479	$9.18\ 10^{-10}$	$1.14\ 10^{-8}$	2	5	1	8
		max	0.85	0.420	$3.67\ 10^{-11}$	$6.80\ 10^{-10}$	0	0	4	3
EB	LF	mean	0.78	0.481	$7.02\ 10^{-10}$	$1.24\ 10^{-8}$	1	3	6	5
		max	0.73	0.451	$6.52\ 10^{-11}$	$7.64\ 10^{-10}$	0	0	5	4
EB_2	KF	max	0.80	0.422	$2.49\ 10^{-13}$	$8.03\ 10^{-13}$	0	0	0	0
-	LD	max	1.00	0.521	$7.43\ 10^{-10}$	$9.87\ 10^{-10}$	0	0	29	0
-	Ex.	max	1.00	0.522	$8.32\ 10^{-10}$	$9.97\ 10^{-10}$	0	0	29	0

5.4.2 The combination of VBR and CBR sources

The next stage of evaluation is to combine VBR and CBR sources. Each VBR source has been mixed with 19 CBR loads of 0.05, 0.10, ... , 0.95. The results can be summarised by dividing the methods into three groups based on the decrease in the allowable load as compared with the exact method:

- $\Delta\rho \approx 0.02$: EV-LD;
- $\Delta\rho \approx 0.04$: EB_2-LD-EV, EB_2-LD-LD, EV-GD;
- $\Delta\rho \approx 0.08$: EB_1-LD, EB_1-LF, EB_2-KF.

The changes from EV to EB_2, from EB_2 to EB_1, and from the large deviation to other approximations have a similar effect on the allowed load: each step doubles the difference from the exact method. This effect is also valid for a combination of EB_2-GD-EV which a gives roughly equal load to EB_1-LD (EB_2-GD-EV is not presented in the tables).

The approximate nature of the EB_2-LD-EV model reflects in the required value for parameter ψ_{adj}. If ρ_{max} is large, ψ_{adj} should be relatively small because the most difficult cases for EV approximation arise when the CBR load is large (the error type II in Figure 4.30). As can be seen from Figure 4.30 and from the analytical results, EB_2 methods have problems with these errors only if ρ_{max} is larger than 0.8. If ρ_{max} is smaller, say 0.75, the adjusting factor ψ_{adj} can be even bigger than 1 because the underestimation of the allowed load with a large CBR load makes an over-utilisation possible with a small CBR load still keeping the average cell loss probability at the required level.

An interesting point is that the accuracy of EB_2-LD-EV is nearly the same as that of EB_2-LD-LD as far as the mean-criterion is concerned although the latter is based on a more accurate model. A possible interpretation of this phenomenon is that the effective variance approximation (EB_2-xx-EV) captures the essence of the rate-variation scale behaviour but not all exceptional cases with high cell loss probabilities. Consequently, because the mean-criterion is not sensitive to the rare cases with high cell loss probability, an approximate model can yield a better average accuracy than a method which is better in exceptional cases.

The results are quite different when a max-criterion is applied. EB_2-LD-LD guarantees the required cell loss level because of the basic principle applied, whereas EB_2-LD-EV is almost unsuitable with max-criterion. Lindberger's approximation and EB_2-LD-EV do not offer any considerable advantage when compared with peak rate allocation. The best versions of the EB_2 methods are even better than the methods based on effective variance. Kelly's method is comparable to the other methods when max-criterion is used.

The results with on/off and 3-level sources are similar, only some minor distinctions can be observed. The adjusting parameters ρ_{max} (for EB_1 and EV) and ψ_{adj} (for EB_2) are slightly smaller with 3-level models than with on/off models. The most important exception is with the EB_2-LD-EV model when ρ_{max} is 0.85 or 0.90, when the accuracy of EB_2-LD-EV is insufficient for some sources with three cell rate levels and therefore ψ_{adj} should be quite small.

The optimum value for ρ_{max} with EB_2 models is fairly high, either 0.85 or 0.9, evidently because the average proportion of CBR traffic is high (see Figure 5.3). The optimisation of ρ_{max} is dealt with further in Section 5.4.4.

Table 5.3. The accuracy of CAC formulae when on/off sources are aggregated with CBR sources, mean-criterion, 30 on/off sources * 19 CBR load levels, $P_{loss} = 10^{-9}$

Method					ρ	P_{loss}	P_{loss}		<10^{-8}	<10^{-9}	<10^{-10}
het.	hom.	with CBR	ρ_{max}	ψ_{adj}	mean	mean	max	>10^{-8}	>10^{-9}	>10^{-10}	>10^{-11}
PR	PR	-	1.00	-	0.586	0	0	0	0	0	0
EB_1	LF	-	0.77	-	0.671	$6.59\ 10^{-10}$	$3.58\ 10^{-8}$	15	25	29	35
	LD	-	0.82	-	0.679	$6.45\ 10^{-10}$	$4.33\ 10^{-8}$	12	17	38	35
EB_2	LD	EV	0.75	1.05	0.684	$8.92\ 10^{-10}$	$1.78\ 10^{-8}$	6	123	59	19
		EV	0.80	1.03	0.709	$8.85\ 10^{-10}$	$2.10\ 10^{-8}$	8	138	70	44
		EV	0.85	0.99	0.722	$7.46\ 10^{-10}$	$4.91\ 10^{-8}$	9	43	149	77
		EV	0.90	0.90	0.706	$9.60\ 10^{-10}$	$1.18\ 10^{-7}$	8	19	28	42
	LD	LD	0.75	1.05	0.682	$6.70\ 10^{-10}$	$1.78\ 10^{-8}$	16	91	69	35
		LD	0.80	1.04	0.707	$7.05\ 10^{-10}$	$1.26\ 10^{-8}$	5	106	91	41
		LD	0.85	1.03	0.724	$8.52\ 10^{-10}$	$1.97\ 10^{-8}$	6	115	85	58
		LD	0.90	1.02	0.727	$8.37\ 10^{-10}$	$1.85\ 10^{-10}$	6	100	102	48
EV	GD	-	0.94	-	0.710	$8.74\ 10^{-10}$	$3.53\ 10^{-8}$	29	23	31	41
	LD	-	0.97	-	0.736	$5.54\ 10^{-10}$	$5.55\ 10^{-8}$	7	27	138	111
Ex.	Ex.	-	1.00	-	0.755	$7.54\ 10^{-10}$	$1.00\ 10^{-9}$	0	0	500	0

Table 5.4. The accuracy of CAC formulae when rate-variation scale sources with three cell rate levels are aggregated with CBR sources, mean-criterion, 30 sources * 19 CBR load levels, $P_{loss} = 10^{-9}$

Method					ρ	P_{loss}	P_{loss}		$<10^{-8}$	$<10^{-9}$	$<10^{-10}$
het.	hom.	with CBR	ρ_{max}	ψ_{adj}	mean	mean	max	$>10^{-8}$	$>10^{-9}$	$>10^{-10}$	$>10^{-11}$
	PR	-	1.00	-	0.587	0	0	0	0	0	0
EB 1	LF	-	0.76	-	0.650	$7.43\ 10^{-10}$	$3.13\ 10^{-8}$	14	50	34	38
	LD	-	0.82	-	0.659	$8.95\ 10^{-10}$	$5.21\ 10^{-8}$	14	27	29	40
EB 2	LD	EV	0.75	1.04	0.668	$8.27\ 10^{-10}$	$1.24\ 10^{-8}$	5	139	84	32
		EV	0.80	1.01	0.686	$8.40\ 10^{-10}$	$2.95\ 10^{-8}$	10	123	117	56
		EV	0.85	0.95	0.687	$9.36\ 10^{-10}$	$8.19\ 10^{-8}$	10	35	86	149
		EV	0.90	0.81	0.658	$9.93\ 10^{-10}$	$1.22\ 10^{-7}$	7	12	22	33
	LD	LD	0.75	1.05	0.668	$8.21\ 10^{-10}$	$1.34\ 10^{-8}$	6	128	92	39
		LD	0.80	1.04	0.687	$8.70\ 10^{-10}$	$1.39\ 10^{-8}$	1	149	104	46
		LD	0.85	1.03	0.694	$8.37\ 10^{-10}$	$1.81\ 10^{-8}$	3	125	109	67
		LD	0.90	1.03	0.688	$8.49\ 10^{-10}$	$2.84\ 10^{-8}$	13	68	104	54
EV	GD	-	0.92	-	0.677	$5.79\ 10^{-10}$	$2.03\ 10^{-8}$	9	62	38	23
	LD	-	0.96	-	0.700	$4.81\ 10^{-10}$	$6.01\ 10^{-8}$	4	24	101	184
Ex.	Ex.	-	1.00	-	0.721	$5.82\ 10^{-10}$	$9.98\ 10^{-10}$	0	0	446	3

Table 5.5. The accuracy of CAC formulae when on/off sources are aggregated with CBR sources, max-criterion, 30 on/off sources * 19 CBR load levels, $P_{loss} = 10^{-9}$

Method					ρ	P_{loss}	P_{loss}	$<10^{-8}$	$<10^{-9}$	$<10^{-10}$
het.	hom.	with CBR	ρ_{max}	ψ_{adj}	mean	mean	max	$>10^{-9}$	$>10^{-10}$	$>10^{-11}$
	PR	-	1.00	-	0.586	0	0	0	0	0
EB_1	LF	-	0.68	-	0.603	$1.09\ 10^{-11}$	$7.01\ 10^{-10}$	26	9	9
	LD	-	0.74	-	0.626	$8.48\ 10^{-12}$	$6.13\ 10^{-10}$	18	14	16
EB_2	KF	KF	0.75	-	0.652	$8.81\ 10^{-14}$	$1.14\ 10^{-12}$	0	0	14
		KF	0.80	-	0.675	$8.05\ 10^{-14}$	$8.37\ 10^{-13}$	0	0	0
		KF	0.85	-	0.690	$6.10\ 10^{-14}$	$6.64\ 10^{-13}$	0	0	0
		KF	0.90	-	0.690	$4.00\ 10^{-14}$	$3.18\ 10^{-13}$	0	0	0
	LD	EV	0.85	0.54	0.617	$4.80\ 10^{-12}$	$7.54\ 10^{-10}$	6	4	5
	LD	LD	0.75	1.00	0.676	$8.61\ 10^{-10}$	$9.41\ 10^{-10}$	130	49	29
		LD	0.80	1.00	0.701	$1.16\ 10^{-10}$	$9.45\ 10^{-10}$	154	64	35
		LD	0.85	1.00	0.719	$1.41\ 10^{-10}$	$9.23\ 10^{-10}$	167	72	44
		LD	0.90	1.00	0.722	$1.51\ 10^{-10}$	$8.93\ 10^{-10}$	166	64	44
EV	GD	-	0.85	-	0.657	$1.06\ 10^{-11}$	$7.58\ 10^{-10}$	30	13	5
	LD	-	0.92	-	0.706	$1.06\ 10^{-11}$	$7.54\ 10^{-10}$	20	61	63
Ex.	Ex.	-	1.00	-	0.755	$7.54\ 10^{-10}$	$1.00\ 10^{-9}$	500	0	0

Table 5.6. The accuracy CAC formulae when rate-variation scale sources with three cell rate levels are aggregated with CBR sources, max-criterion, 30 sources * 19 CBR load levels, $P_{loss} = 10^{-9}$

Method					ρ	P_{loss}	P_{loss}	$<10^{-8}$	$<10^{-9}$	$<10^{-10}$
het.	hom.	with CBR	ρ_{max}	ψ_{adj}	mean	mean	max	$>10^{-9}$	$>10^{-10}$	$>10^{-11}$
	PR	-	1.00	-	0.587	0	0	0	0	0
EB_1	LF	-	0.68	-	0.593	$1.64\ 10^{-11}$	$8.11\ 10^{-10}$	30	26	22
	LD	-	0.74	-	0.610	$9.78\ 10^{-12}$	$6.43\ 10^{-10}$	15	25	29
EB_2	KF	KF	0.75	-	0.633	$1.17\ 10^{-13}$	$1.32\ 10^{-12}$	0	0	9
		KF	0.80	-	0.650	$1.17\ 10^{-13}$	$1.02\ 10^{-12}$	0	0	1
		KF	0.85	-	0.656	$6.19\ 10^{-14}$	$5.50\ 10^{-13}$	0	0	0
		KF	0.90	-	0.646	$2.47\ 10^{-14}$	$3.49\ 10^{-13}$	0	0	0
	LD	EV	0.85	0.50	0.594	$5.29\ 10^{-12}$	$7.40\ 10^{-10}$	6	3	4
	LD	LD	0.75	0.99	0.658	$8.29\ 10^{-11}$	$9.08\ 10^{-10}$	155	71	34
		LD	0.80	1.00	0.679	$1.52\ 10^{-10}$	$9.76\ 10^{-10}$	203	71	35
		LD	0.85	1.00	0.688	$1.72\ 10^{-10}$	$8.91\ 10^{-10}$	191	84	45
		LD	0.90	1.00	0.682	$1.22\ 10^{-10}$	$9.33\ 10^{-10}$	134	66	56
EV	GD	-	0.86	-	0.644	$2.43\ 10^{-11}$	$9.24\ 10^{-10}$	39	35	19
	LD	-	0.91	-	0.673	$1.14\ 10^{-11}$	$7.40\ 10^{-10}$	18	53	80
Ex.	Ex.	-	1.00	-	0.721	$5.82\ 10^{-10}$	$9.98\ 10^{-10}$	446	3	0

5.4.3 Combination of different VBR sources

The next step is to evaluate the combination of various VBR and CBR sources. The results are presented in Tables 5.7 and 5.8. The required value for ρ_{max} in Lindberger's formula is noticeably lower than the value (=1/1.2) proposed by Lindberger (1991). The main reason for this difference is that the source parameter area in this study is wider than that originally intended with Lindberger's approximation (as regards the application region of Lindberger's formula see Roberts 1992a p. 43). Despite the approximate character of Lindberger's formula, it gives a higher load than EB_1-LD and, moreover, the highest observed P_{loss} is lower with Lindberger's approximation.

The allowable load obtained by EB_2-LD-EV is higher than that obtained by EB_2-LD-LD if a mean-criterion is applied but it is not suitable with a max-criterion. Although EB_2-LD-LD is feasible with a max-criterion, it does not guarantee P_{loss} in complicated cases without the application of ψ_{adj}.

With Kelly's method the highest observed P_{loss} is as small as $3.64\ 10^{-12}$. Such low value is due to the difference between the formulae for saturation probability (3.6) and cell loss probability (3.8). This difference is typically of the order 100 (Roberts 1992a p. 154), which explains fairly well the smallness of P_{loss} together with the intrinsic property of Kelly's formula to guarantee cell loss probability even in the worst cases.

Table 5.7. The accuracy of CAC formulae when on/off sources, rate-variation scale sources with three cell rate levels and CBR sources are aggregated, mean-criterion, 3000 cases, $P_{loss} = 10^{-9}$

Method					ρ	P_{loss}	P_{loss}				
het.	hom.	with CBR	ρ_{max}	ψ_{adj}	mean	mean	max	$>10^{-8}$	$<10^{-8}$ $>10^{-9}$	$<10^{-9}$ $>10^{-10}$	$<10^{-10}$ $>10^{-11}$
	PR	-	1.00	-	0.239	0	0	0	0	0	0
EB_1	LF	-	0.76	-	0.561	$7.58\ 10^{-10}$	$3.47\ 10^{-8}$	62	341	449	391
	LD	-	0.83	-	0.551	$6.16\ 10^{-10}$	$6.22\ 10^{-8}$	42	148	266	483
EB_2	LD	EV	0.75	0.99	0.592	$9.90\ 10^{-10}$	$3.09\ 10^{-8}$	30	769	1142	268
		EV	0.80	0.98	0.596	$9.59\ 10^{-10}$	$9.00\ 10^{-8}$	39	505	1437	443
		EV	0.85	0.97	0.585	$7.39\ 10^{-10}$	$9.59\ 10^{-8}$	27	246	938	898
		EV	0.90	0.95	0.541	$9.48\ 10^{-10}$	$2.79\ 10^{-7}$	18	46	221	451
	LD	LD	0.75	1.01	0.591	$7.89\ 10^{-10}$	$1.20\ 10^{-8}$	1	874	1038	352
		LD	0.80	1.01	0.591	$8.39\ 10^{-10}$	$2.24\ 10^{-8}$	8	764	1061	423
		LD	0.85	1.01	0.568	$7.12\ 10^{-10}$	$1.98\ 10^{-8}$	8	468	914	442
		LD	0.90	1.02	0.526	$9.37\ 10^{-10}$	$3.55\ 10^{-8}$	44	302	427	323
EV	GD	-	0.93	-	0.564	$7.12\ 10^{-10}$	$3.27\ 10^{-8}$	46	412	406	350
	LD	-	0.98	-	0.604	$9.76\ 10^{-10}$	$1.52\ 10^{-7}$	25	416	1808	431
Ex.	Ex.	-	1.00	-	0.615	$7.57\ 10^{-10}$	$1.00\ 10^{-9}$	0	0	2878	35

Table 5.8. The accuracy of CAC formulae when on/off sources, rate-variation scale sources with three cell rate levels and CBR sources are aggregated, max-criterion, 3000 cases, $P_{loss} = 10^{-9}$

Method					ρ	P_{loss}	P_{loss}	$<10^{-8}$	$<10^{-9}$	$<10^{-10}$
het.	hom.	with CBR	ρ_{max}	ψ_{adj}	mean	mean	max	$>10^{-9}$	$>10^{-10}$	$>10^{-11}$
	PR	-	1.00	-	0.239	0	0	0	0	0
EB_1	LF	-	0.67	-	0.495	$1.26\ 10^{-11}$	$6.95\ 10^{-10}$	130	304	259
	LD	-	0.74	-	0.491	$5.08\ 10^{-12}$	$5.44\ 10^{-10}$	50	155	276
EB_2	KF	KF	0.75	-	0.518	$3.03\ 10^{-13}$	$3.64\ 10^{-12}$	0	0	231
		KF	0.80	-	0.511	$2.03\ 10^{-13}$	$3.64\ 10^{-12}$	0	0	34
		KF	0.85	-	0.488	$9.94\ 10^{-14}$	$3.64\ 10^{-12}$	0	0	10
		KF	0.90	-	0.441	$2.90\ 10^{-14}$	$1.28\ 10^{-12}$	0	0	1
	LD	EV	0.85	0.51	0.364	$1.93\ 10^{-12}$	$8.76\ 10^{-10}$	7	2	6
	LD	LD	0.75	0.90	0.552	$1.74\ 10^{-11}$	$9.46\ 10^{-10}$	118	978	811
		LD	0.80	0.92	0.556	$1.92\ 10^{-11}$	$6.31\ 10^{-10}$	88	1006	868
		LD	0.85	0.95	0.545	$2.71\ 10^{-11}$	$8.76\ 10^{-10}$	157	956	578
		LD	0.90	0.98	0.512	$3.62\ 10^{-11}$	$7.44\ 10^{-10}$	265	460	341
EV	GD	-	0.85	-	0.505	$1.67\ 10^{-11}$	$9.45\ 10^{-10}$	188	321	257
	LD	-	0.91	-	0.551	$1.24\ 10^{-11}$	$8.04\ 10^{-10}$	85	690	700
Ex.	Ex.	-	1.00	-	0.615	$7.57\ 10^{-10}$	$1.00\ 10^{-9}$	2878	35	4

The approximations obtained by EV-LD and EB_2-LD-EV (ρ_{max}=0.8) give nearly symmetric distributions with a clear peak in the range between 10^{-10} and 10^{-9}. EV-GD and EB_1-LF also give symmetric P_{loss} distributions but the peaks are lower. EB_2-LD-LD results in a different type of distribution. The tail of distribution towards zero P_{loss} is relatively strong while the other half of the distribution drops rapidly after 10^{-8}. This phenomenon is caused by the fact that in a great majority of cases EB_2-LD-LD fulfils the cell loss requirement when $\psi_{adj} = 1$. The allowed value for ψ_{adj} (1.01) shifts part of distribution above the required P_{loss} level but still the probability that P_{loss} considerably exceeds the allowed level is very small. For the same reason, the Max{P_{loss}} to Mean{P_{loss}} ratio is smaller for EB_2-LD-LD than for EV-LD although the latter usually gives a better average accuracy.

5.4.4 Optimisation of ρ_{max} in EB_2 methods

A proper choice of parameter ρ_{max} is important with all EB_2 type of methods. In practical implementations it is not reasonable to adjust ρ_{max} continuously according to the current traffic situation because the effective bandwidth of each source depends on ρ_{max}. Therefore the selection of ρ_{max} should be based on a typical traffic mix of each ATM link. However, there is not much knowledge of the proportion of different traffic types in real ATM networks and thus the approach in this section is to assess the accuracy of optimising formulae (5.18) and (5.21) when the traffic consists of a wide variety of sources. The results concerning ρ_{max} are gathered into Figures 5.5 and 5.6. The following weighting coefficients have been used:

- CBR sources: 0.1;
- on/off sources (Table 5.1): 0.1;
- 3-level sources (Table 5.2): 0.1;
- on/off and CBR (Table 5.3 or 5.5): 0.1;
- 3-level and CBR (Table 5.4 or 5.6): 0.1;
- combination (Table 5.7 or 5.8): 0.5.

With these weighting coefficients we obtain the following values (see Section 5.3.2.4):

$\rho_{het,ave}$= 0.628,

$\rho_{hom,ave}$ = 0.707.

These values depend to some extent on the CAC method applied, here EB_2-LD-EV with ρ_{max} = 0.8 and ψ_{adj} = 1 has been used. By applying (5.18) and (5.21) we obtain two approximations for the optimum ρ_{max}:

ρ_{max}{het} = 0.81,

ρ_{max}{hom} = 0.77.

This result is in accordance with Figures 5.5 and 5.6. The figures show the allowable load and mixing efficiency as a function ρ_{max} for three methods. Mean-criterion has been used with EB_2-LD-EV, EB_2-LD-LD whereas with Kelly's method only max-criterion is applicable. With EB_2-LD-EV the maximum load and maximum mixing efficiency are achieved with values 0.82 and 0.79 for ρ_{max}. The corresponding values for EB_2-LD-LD are 0.80 and 0.77. This example supports the assumption that ρ_{max}{het} is valid as far as the average load is concerned and ρ_{max}{hom} works better if mixing efficiency is used as the maximising criterion.

The value of 0.8 for ρ_{max} in EB_2 methods may be recommended as a safety choice for a wide range of traffic combinations and CAC models, particularly as far as rate-variation scale models are concerned. Furthermore, (5.18) and (5.21) offer simple and efficient ways to optimise ρ_{max} provided that there is enough information on the proportion of different traffic types.

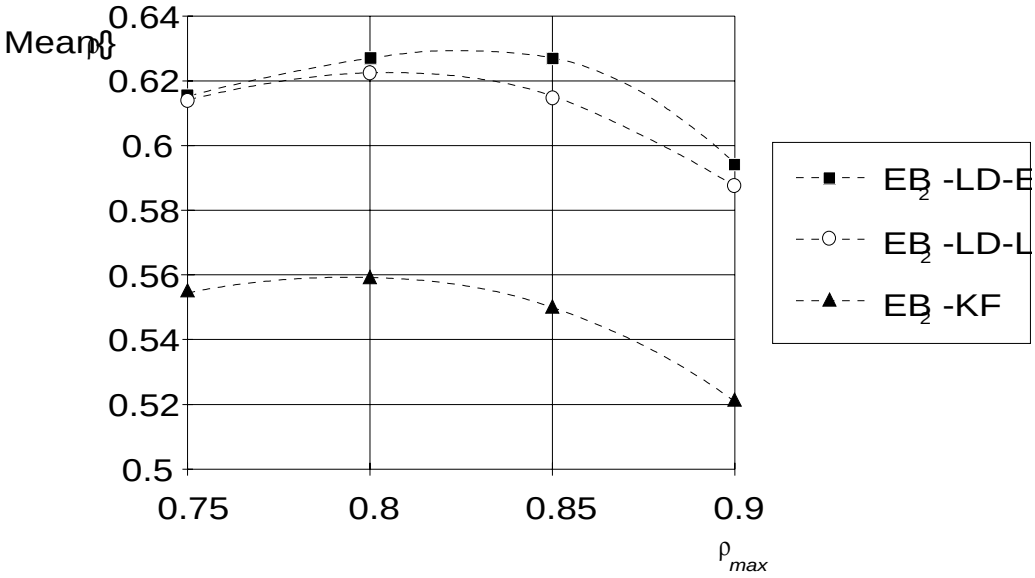

Figure 5.5. The average allowable load as a function of ρ_{max} for EB_2 methods.

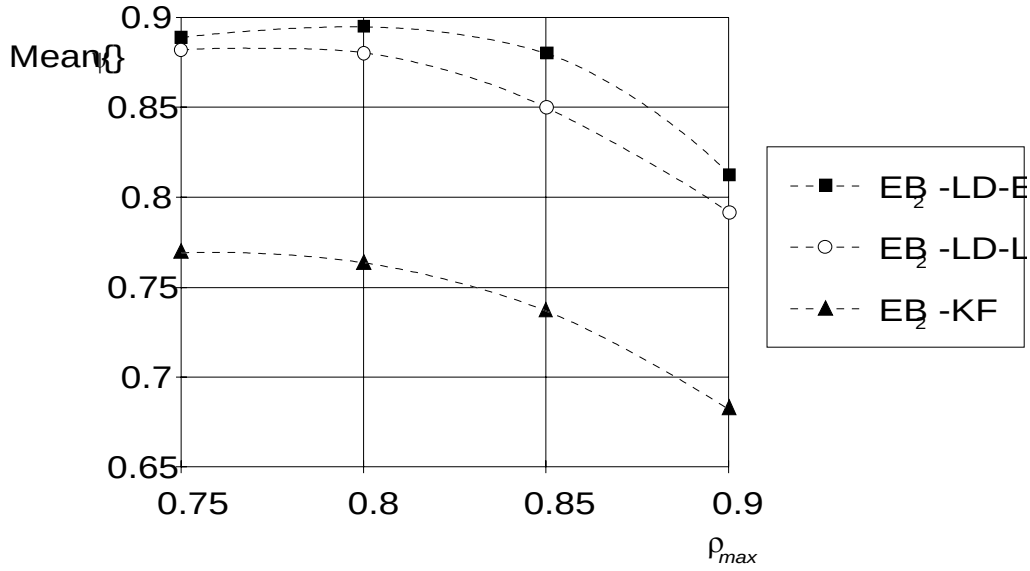

Figure 5.6. Mixing efficiency as a function of ρ_{max} for EB_2 methods.

5.4.5 Summary of the efficiency with rate-variation scale models

In the previous sections we have evaluated the properties of different CAC methods in various traffic cases in order to develop as efficient a CAC method as possible. Figures 5.7 and 5.8 summarise the results using the same weighting coefficients as in the previous section. Factor ρ_{max} is 0.8 in all EB_2 methods.

The accuracy of effective variance with large deviation approximation is excellent in great majority of traffic combinations. Because of the small difference between EV-LD and exact method (0.604 vs. 0.615) there seems to be no practical reason to use more complicated methods with rate-variation scale traffic than:

- the large deviation approximation for homogeneous cases;
- the effective variance for heterogeneous cases.

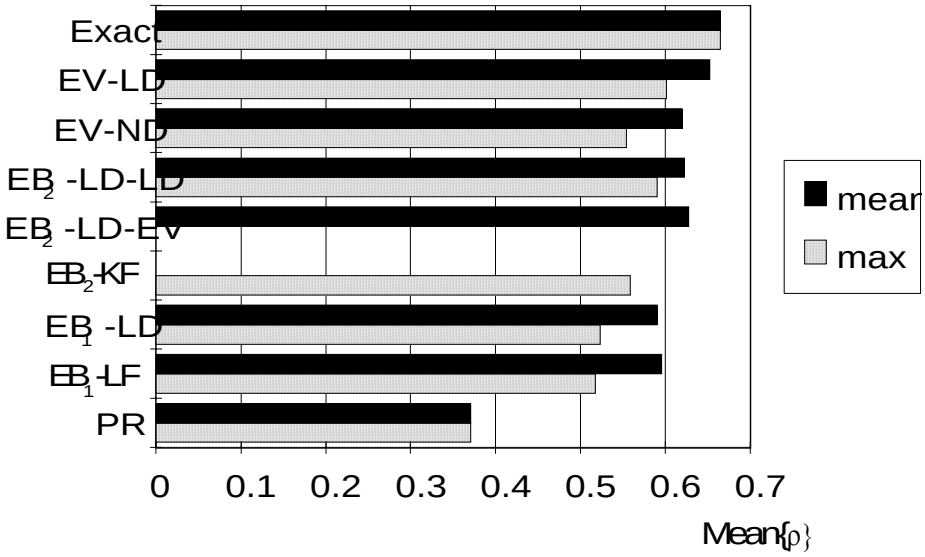

Figure 5.7. The average allowable load of different CAC-methods; mean and max-criteria.

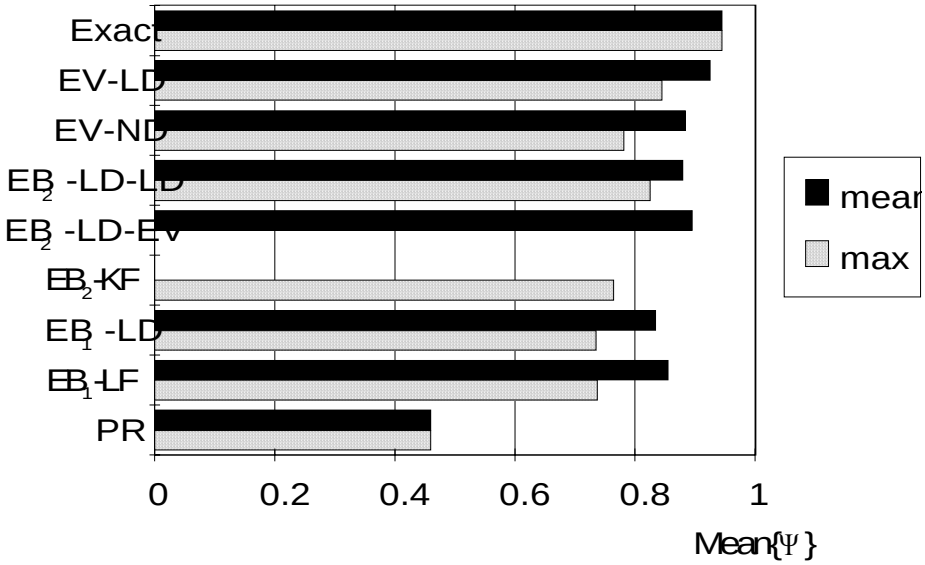

Figure 5.8. The mixing efficiency of different CAC methods; mean and max-criteria.

The next question to be answered is whether the efficiency of any simpler approximation is sufficient for practical purposes. Lindberger's approximation works even better than the large deviation approximation when mixing efficiency is concerned and when EB_1 is used for heterogeneous cases. A possible explanation for this somewhat surprising phenomenon is that the formula (3.5), which determines effective bandwidths, contains similar properties to the EB_2 methods. Similarly, effective variance method gives better results with Lindberger's approximation than with Gaussian distribution approximation (Kilkki 1992). Consequently, Lindberger's approximation is a noteworthy alternative for approximating homogeneous cases at rate-variation scale.

If we take into account the previous remarks, three major candidates for a practical CAC method can be proposed: EB_1-LF, EB_2-LD-EV and EV-LD. In order to clarify the significant differences of these methods let us use as two reference points peak rate allocation (relative efficiency = 0) and the exact method (relative efficiency = 1). Other methods can be placed on the linear scale determined by these two reference points.

Table 5.9 summarises the comparison of CAC methods. The gains measured in the scale from peak rate allocation to exact method are roughly 80%, 90% and 95% for EB_1-LF, EB_2-LD-EV and EV-LD, respectively. This result and the result presented in Section 5.4.2 have an obvious similarity although the latter is based on a limited evaluation including only the superposition of identical VBR sources and a CBR load. This similarity implies that the main result concerning the relative efficiency of different CAC methods is due to the intrinsic behaviour of the traffic process at rate-variation scale and does not depend much on the weighting of different traffic combinations.

Table 5.9. Summary on the efficiency CAC methods with rate-variation scale traffic

Method	ρ_{max}	ψ_{adj}	ρ mean	Ψ mean	relative ρ mean	relative Ψ mean
PR			0.371	0.460	0.000	0.000
EB_1-LF	0.775		0.596	0.855	**0.765**	**0.815**
EB_1-LD	0.839		0.591	0.834	0.746	0.771
EB_2-KF	0.800		0.559	0.764	0.639	0.627
EB_2-LD-EV	0.800	1.00	0.627	0.895	**0.870**	**0.897**
EB_2-LD-LD	0.800	1.02	0.622	0.880	0.854	0.866
EV-GD	0.941		0.620	0.884	0.846	0.874
EV-LD	0.981		0.653	0.925	**0.957**	**0.959**
Exact			0.665	0.945	1.000	1.000

5.5 Other aspects for comparison

5.5.1 Efficiency with burst scale traffic

The evaluation in the previous section was based only on rate-variation scale models, mainly because there is no established method to determine the allowed number of sources with burst scale traffic. A way of solving this problem is to classify the sources into a limited number of predefined groups. In this case it is possible to apply complicated models to determine the required parameters for CAC methods.

The average accuracy of effective bandwidth and effective variance methods are similar although the errors occur in different cases. According to the results presented in Section 4.3 effective bandwidth model forms a feasible alternative provided that the maximum burst size is small (say, less than five cells) whereas when burst size is larger than buffer size the traffic behaviour is similar to that of rate-variation scale and, consequently, effective variance is the most accurate approximation.

However, the traffic offered to an ATM network consists of different traffic types and it is not feasible to use different CAC methods on different links. Therefore, the CAC scheme should be applicable to all traffic cases. The results presented in Section 4.4.4 makes it possibility to make some preliminary assessments with complicated traffic mixes. The most promising candidates for CAC methods are EB_2 (simple CAC procedure) and EBV (efficient with all types of traffic). According to Table 4.4 the allowed load of EBV is some percentage higher than that of the EB_2 method. This profitability of the EBV method requires that source parameters for burst scale traffic can be determined with a reasonable accuracy.

5.5.2 Implementation aspects

In addition to efficiency, the simplicity of implementation is the main requirement for a CAC method. Table 5.10 offers an assessment regarding four parts of the implementation: calculation of source parameters, CAC procedure, adjusting of ρ_{max}, and the additional requirements for routing and dimensioning. The figures are, of course, only indicative.

Table 5.10. Complexity of CAC methods

Method	Complexity of parameter calculation	CAC calculation	adjusting of ρ_{max}	routing and dimensioning	Σ
PR	0	1	0	1	2
EB_1-LF	1	1	1	1	4
EB_1-LD	3	1	1	1	6
EB_2-KF	3	1	2	1	7
EB_2-LD-EV	4	1	2	1	8
EB_2-LD-LD	6	1	2	1	10
EV-GD	1	2	0	3	6
EV-LD	3	2	0	3	8
EBV	4	3	0	4	11
Exact	0	15	0	5	20

As regards the parameter calculation, the acceptability of complicated procedures in real implementations depends essentially on the structure of the whole traffic control system in ATM networks (see Figure 5.1). In any event, the simplicity of parameter determination is a major benefit for a CAC method. Gaussian distribution approximation and Lindberger's formula are based directly on the variance of cell rate distribution. Large deviation approximation is simple but still harder to calculate than variance. With EB_2 methods an additional difficulty is to estimate the allowable number of sources with superposition of a CBR load. With EB_2-LD-LD at least 20 different values of CBR load should be calculated in order to obtain an appropriate value for $\psi_{max,i}$. In addition, since all traffic parameters affect the result in the case of large deviation approximation, it is difficult to use any pre-calculated tables in the same way as with EB_2-LD-EV method.

The implementation of a CAC procedure (function $F_{CAC\text{-}LM}$ in Figure 5.1) should be very simple because the calculation is always needed when a new connection is established or released and even during a connection if an FRM protocol is used. Methods applying effective bandwidth are simple as only an addition is needed. A square root calculation is the additional procedure required with effective variance .

The adjusting of factor ρ_{max} to the actual traffic mixture may cause additional procedures in particular if the efficiency of the method is dependent on the proper value of this parameter. With EB_2 methods any change of ρ_{max} affects the effective bandwidth of every source whereas with EB_1 the effect is limited to the CAC procedure. With methods based on effective variance there is presumably no need for adjusting ρ_{max}.

As regards the routing and dimensioning, the simplest possible case is effective bandwidth because it is possible to use circuit-switched methods to analyse and design

ATM networks when effective bandwidth concept is applied (e.g., Girard & Lessard 1992; Griffiths 1990). Effective variance may result in more complicated dimensioning methods although it might be possible to convert effective variance to effective bandwidth for dimensioning and routing purposes and by that means to exploit the methods developed for circuit switched networks.

The sum-column in Table 5.10 should be viewed circumspectly because the appraisal of different aspects is very difficult and, moreover, it is not at all clear whether an addition is the best way to combine the results of different aspects. In fact, if the difference in sum-points between two CAC methods is small (one or two) and the order of the methods is different in respect of different aspects, it is not possible to positively infer the order of these methods in terms of common feasibility. For instance, it is difficult to conclude whether EB_2-LD-EV is simpler than EV-LD. The answer depends on the emphasis of different aspects: in some cases the simplicity of parameter calculation is important while in other cases a simple CAC principle as regards network dimensioning is needed.

We can draw three obvious inferences when the results of Tables 5.9 and 5.10 are combined. If the simpler effective bandwidth, EB_1, is applied, there is no need to use more complicated methods than Lindberger's approximation for calculating effective bandwidths. Secondly, with the other effective bandwidth principle, EB_2, the approximation based on effective variance (EB_2-LD-EV) is better regarding both the attainable load and complexity than the other combination EB_2-LD-LD. Kelly's method is, without modification, appropriate only when max-criterion is applied.

5.5.3 Selection of CAC method

As a final conclusion to be drawn from the evaluation, which includes both performance and implementation aspects, the most promising CAC methods in ATM networks are:

- EB_1-LF: when a simple implementation is the most important aspect;
- EB_2-LD-EV: when a simple CAC procedure is needed but source parameter determination can be rather complicated;
- EV-LD: when a high utilisation is preferred.

This conclusion is valid mainly for traffic models at rate-variation scale whereas, if burst scale fluctuations are also concerned, the situation is somewhat different. In that case the most promising candidates are:

- EB_2-X-EV: the main advantage is a simple CAC procedure;
- EBV-X: efficient with all types of traffic process.

However, without any suitable method (X) for determining the allowed number of sources in homogeneous case for burst scale traffic, it is not possible to make an extensive performance evaluation at the cell loss probability level of 10^{-9}. Moreover, the difficulties of controlling source parameters may reduce the gain in efficiency obtained by applying of burst scale parameters.

5.6 Real traffic aspects

The basis of the previous evaluation has been mathematical models, whereas the requirements and properties of real traffic have been mostly ignored. In this section we return to the themes of Sections 2.3 and 2.4. The first subsection considers the uncertainty of real traffic in ATM networks. Then we attempt to clarify the complicated relationship between CAC and other traffic control functions by presenting a simplified scheme for traffic control in ATM networks. In the last subsections two prime service types, video and LAN traffic, are assessed using the control scheme and the results of performance evaluation of CAC methods.

5.6.1 Uncertainty

The most important source of uncertainty concerning the previous evaluation of CAC methods is that the whole analysis was based on theoretical models with exact source parameters, not on real traffic in ATM networks. In other words, we have assumed that all the uncertainty of traffic variations of one connection is inside the model whereas the traffic model itself and its parameters are exactly known. By contrast, with a real traffic process we do not have any simple model that totally describes the behaviour of the traffic process. The most precise description of a real traffic process in ATM networks is the $\Sigma D_i/D/1/K$ model, but only if the traffic process consists of independent CBR connections, whereas in all other cases theoretical models have fundamental limitations (see e.g., Minoli 1993 Section 4.6).

A way to capture the uncertainty of real traffic is to construct more complicated models, metamodels, which take into account the uncertainty of source models. If the sources are separate and independent of each other, this kind of model is even practicable. Actually, this phenomenon is much the same as the rate-variation scale variations particularly if the uncertainty concerns mean rate (see e.g., Burgin 1990). We can continue the modelling: if the uncertainties are themselves uncertain, results for different levels of uncertainty can be generated to see sensitivities and tradeoffs (Holtzman 1990).

A further problem is that the traffic process may consist of groups of sources such as telephone and video calls. The information we have on the properties of traffic is usually common to the whole group of sources and consequently they have about the same predicting errors. The effect of this phenomenon is especially strong when there are a large number of sources with small mean rate. If the predicting errors of source parameters are independent and non-biased, the effect is usually negligible but if the error is common to all sources, the effect may be much larger than expected. This situation may arise on account of some occasional external reason, for instance, when a great number of video sources are showing the same event. The underlying problem is common to all CAC approximations: sources are supposed to be independent of each other (both the determination of source parameters and instantaneous source behaviour). Without this assumption any estimation of traffic behaviour becomes very difficult.

Some dependencies between connections are intentional. Intentional dependencies may cause real problems only if there is a substantial advantage to be achieved by the user. A typical example is that by splitting a call with high peak rate and burstiness into several smaller connections, the customer can deceive the operator into believing that connections can be effectively multiplexed (Norros 1992). If the charging is based purely on the connection's effective bandwidth, the user may gain a notable advantage.

However, this kind of user behaviour can presumably be avoided by an appropriate charging principle (Lindberger 1992a).

If statistical parameters are applied, there is an obvious risk because the unpredictability of individual events is characteristic for all statistical quantities. It may be possible only after a large number of cases to assess whether the source behaviour has been acceptable, but by then the possible damages have already occurred. The charging policy is once again of great importance. Charging can be planned so that the most profitable strategy for the user is to estimate source parameters as exactly as possible (see Kelly 1993; Roberts 1992a Section 3.4). A proper tariff scheme is needed especially when the amount of transferred cells is crucial for the user (e.g., when the connection is used for file transfers).

Another approach is to use such tight control methods that the source behaviour cannot exceed predefined limits. The idea is that the traffic patterns that are allowed to go through the controlling device are definitely determined, typically by controlling the mean and peak rate and the maximum burst size of each source. The problem is that it is not easy to infer what the worst traffic pattern is in terms of statistical multiplexing. A deterministic on/off source with maximum peak and mean rates is frequently supposed to be the worst case source. However, this assumption is valid only in some special cases, usually the worst case pattern is more complicated, resulting in a higher cell loss probability than an on/off pattern (see Section 3.2.1). On the other hand, we can ask whether the user can benefit by producing these complicated patterns since both mean rate and peak rate are tightly restricted—it is very unlikely that a large number of users would intentionally produce at the same time the worst possible traffic patterns.

5.6.2 The relationship between CAC and other control functions

Connection Admission Control is not a separate function but it must work seamlessly with other control functions, such as Fast Resource Management (see Section 2.3.2). We can say that FRM is the technique that shifts the uncertainty of the ATM traffic process from one level to other. Burst scale uncertainty relates to arrival times of packets (or a group of packets) on the ATM network interface whereas after the arrival the cell scale process is usually predictable. If the burst size is large enough, it is possible to apply a fast CAC type of procedure for every individual burst and by that means alleviate the problem of controlling statistical parameters. This means that the uncertainty problem is shifted from the core of the ATM network to the interface. An outline for a control structure of the ATM network with a FRM procedure is presented in the following paragraphs.

Traffic with a strict cell loss requirement (e.g., 10^{-9}) may use a CAC method based on statistical source parameters. If a high priority source wishes to modify a traffic parameter during the connection, the modification is made by aid of the same CAC procedure that is used for connection acceptance. This procedure may be named the FRP/DT procedure but it should be closely integrated with the CAC procedure. In consequence, the CAC method should be fast enough to be suitable to in-call traffic parameter modifications. All high priority traffic is managed in this way. In this scheme modified traffic parameters have no priority over new connections and therefore the probability that a modification request will be rejected is equal to the probability of instantaneous call blocking (it should be noted that these probabilities may depend on the required cell rate).

High and low priority cells are separated into different connections. The remaining capacity left by high priority connections is offered to low priority connections with a two level acceptance procedure. Firstly, there is a connection admission procedure similar to that of high priority traffic, but this time concerning burst congestion. The probability of burst congestion might be of the order of 10^{-4} (Roberts 1993a) and if this cannot be guaranteed the connection is rejected.

Secondly, FRP/IT procedure is always used when a low priority source has something to send. The acceptance procedure should be as simple as possible, presumably based only on peak rate, in order to enable a very fast decision. In addition, FRP/IT procedure should be consistent with the CAC procedure of high priority traffic because it needs information on the actual capacity needed by the high priority traffic. The remaining capacity left to low priority connections varies depending on the changes in high priority flows. Thus if the network guarantees a certain level of burst congestion in low priority traffic, the CAC high priority traffic procedure has somehow to take into account the low priority connections in progress.

All these aspects of admission procedure, burst congestion evaluation and FRP/IT for low priority connections, and CAC for high priority connections, may apply the same principles such as effective bandwidth and effective variance. Mathematical models for evaluating burst congestion are similar to the rate-variation scale models for cell loss probability. If the burst congestion probability is relatively small, the effect of re-attempts of congested bursts can presumably be ignored and therefore the burst congestion probability depends only on the sufficiency of link capacity to carry the requested peak rate.

If a further exploitation of network resources is wanted, an essentially different approach is needed. A real best effort traffic, one without any guarantee for QoS and without any restriction on traffic variations, can be integrated into the ATM network under certain conditions. First, the best effort cells should be separated from other cells at the first switching stage (dedicated for this purpose) and routed to large dedicated buffers. Then the best effort cells are allowed to use the capacity of outgoing links only if there is no other cell to be delivered to the link. Finally, in order to avoid huge buffers a BECN procedure is needed for informing sources when there is no more capacity for best effort traffic (see Section 2.3.2). With a physical separation of this type it is possible to guarantee that other traffic streams are not disturbed even though there is an excessive amount of best effort traffic.

Another alternative is to place large buffers at network interfaces since there is the best knowledge of the requirement and properties of each application. Then a high utilisation of network capacity is achieved by means of smoothing out the traffic process in the ATM network. However, this smoothing-out process is not appropriate to all applications.

5.6.3 Requirements of VBR video sources

The traffic process of a VBR video source may consist of four phenomena (see Section 2.4.3): (a) permanent basic level during scene; (b) small variations during scene; (c) considerably variations in the needed cell rate from scene to other; (d) high peaks at scene changes.

The primary issue in respect of Connection Admission Control is the predictability of variations. Since there is no method to predict the needed bandwidth in advance, inter-

scene variations must be managed by statistical means. If there are inter-scene variations, the variations during a scene usually have no appreciable effect on the allowed load. In this case the traffic models of rate-variation scale are doubtless appropriate because the time scales of the variations are very long. The main difficulty is to find traffic parameters with a sufficient accuracy.

If the required cell rate can be predicted at scene changes, in-call modification of traffic parameters may be applicable although this makes high demands on the coding method (note that prediction is quite possible with a recorded video). A possible scheme is that the basic cell rate level remains constant during the connection and all variations in the needed cell rate are controlled either by a high priority connection and an in-call modification (the FRP/DT procedure), or by a separate low priority connection with FRP/IT. In both cases there may be a relatively high probability that an in-call modification request will be rejected and therefore the coding method must be able to manage these situations without substantial impairment of picture quality.

In addition, we must take into account high peaks at scene changes. If a layered coding scheme is used, these peaks are supposedly manageable by the FRM procedure since the required cell rate is predictable. It is possible to apply an FRM procedure, perhaps even without a permanent reservation, because the duration of the peak is short and it might occasionally be acceptable to decrease QoS level at some scene changes. If we take these peaks into account in a statistical CAC procedure, the increase in required (effective) bandwidth may be unreasonable high when compared with the improvement of average QoS achieved.

Because most VBR applications will know the whole frame content before sending (Aagesen 1993), it is highly recommended that sources should send the total frame content equally stretched within the frame. At least, the effect of intra-frame variations should be much lower than the inter-scene variations, which cannot usually be modified. This issue can be analysed by the methods presented in Section 4.3.4.

5.6.4 Traffic between Local Area Networks

The traffic between non-ATM Local Area Networks is perhaps the hardest situation for the traffic control of ATM networks. The first task is to assess what kind of traffic model is suitable for LAN traffic. The original traffic from a LAN consists of variable length packets which should be converted to ATM cells at LAN/ATM interface. The burst size depends largely on the application but an average packet size of 500 bytes or 10 cells may be used as a starting point. If ATM cells are sent to the ATM network at the speed of LAN, the peak rate may be 1/20 of the ATM link capacity. The mean rate to peak rate ratio of this type of connection is small (e.g., 0.1). From these values we conclude that the inter arrival time of bursts is about 2000 time slots.

Using the result presented in Section 4.3 we obtain allowable loads of 0.73 and 0.48 for Poisson bursts with cell loss requirement of 10^{-4} and 10^{-9}, respectively. When a Markov model is applied, the scale factors are $\varepsilon_u = 0.5$ and $\varepsilon_m = 0.3$ if $P_{loss} = 10^{-4}$, and $\varepsilon_u = 0.65$ and $\varepsilon_m = 0.62$ if $P_{loss} = 10^{-9}$ (see Figure 4.13). Thus LAN interconnection traffic may be classified as a burst scale source. However, this is not an adequate evaluation because of the long range dependency peculiar to LAN traffic. A further approach is to modulate the above-mentioned process by an upper on/off process. According to the result presented in Section 4.3.4 the primary issue is the attainable loads of two limit cases:

the arrival process of Poisson bursts and the modulating rate-variation scale process without burst scale fluctuations.

In our example the peak rate of *rate-variation scale* is 1/200 of link capacity (this is determined by the burst scale parameters). If the *on* probability in rate-variation scale is 0.1, the allowable loads are 0.83 and 0.69 for cell loss probabilities 10^{-4} and 10^{-9}, respectively. We can see that in this example the allowable load for the burst scale process is lower than that of the rate-variation scale. On the other hand, if we decrease the peak rate in burst scale, we can easily get a situation in which the rate-variation scale process has a lower allowable load. For example, if the peak rate (at burst scale) is 1/50 instead of 1/20, the allowable loads of Poisson bursts are 0.92 and 0.82 for cell loss probabilities 10^{-4} and 10^{-9}, and consequently the rate-variation scale fluctuations are dominant. The main consequence is that a proper traffic evaluation should include the fluctuations both in burst scale and rate-variation scale.

Rate-variation scale fluctuations cannot usually be smoothed out in the same way as those at burst scale. If the rate-variation scale burstiness is high, the allowable load may be very low and there is an obvious demand for methods of increasing the utilisation of network resources. FRM may be a good solution but again predictability and relatively long periods are the main conditions for the application of FRM. If the required cell loss probability is moderately high, a low priority connection and a FRP/IT procedure may be practicable. In contrast, a high priority connection with in-call modifications may lead to an insignificant gain compared with peak rate allocation if the predictability of traffic process is poor or the peak rate is high.

In addition, the connections between end-users may cause strong variations in the traffic between Local Area Networks and these variations are difficult to predict and may even be invisible to traffic control of the ATM network. Peak rate may be the only known parameter and another fact is that traffic burstiness is very high. In this situation traffic models should include the uncertainty aspect of several levels and the outcome can be very complicated, and however sophisticated a traffic model we have, the achievable load may remain very low. The conclusion can be expressed as Lindberger (1992b): The internal operator of the LAN should identify subusers and subcalls, analyse burst and call scale problems separately and so on, if he is really interested in having a better control of the LAN traffic than just regarding it as one strange user with very complicated variations.

6 SUMMARY

The greatly variable requirements of different applications, particularly those of video and data sources, make high demands on the development of traffic control in ATM networks. In this study simple and efficient traffic models have been developed in order to obtain a clear view of the traffic process in ATM networks. These traffic models form a solid basis for the development of Connection Admission Control (CAC) methods.

The traffic process in ATM networks may be extremely complicated. A principal tool used to analyse this process is a division into three time scales: cell, burst and rate-variation scales. The main property in cell scale is traffic variations due to the asynchronous arrival of cells from distinct connections. Traffic processes in the rate-variation scale consist of long-range variations that cannot be buffered in ATM network nodes. All traffic processes that cannot be properly described by these two extreme processes belong to the burst scale. This time scale division has been applied throughout the study.

In order to apply the time scale division efficiently we should find a way to classify a traffic source into the proper time scale. In this study two new factors, utilisation factor and multiplexing factor, have been introduced. The utilisation factor depicts the multiplexing efficiency in homogeneous cases as compared with pure cell scale traffic and pure rate-variation scale traffic. The multiplexing factor utilises the same extreme cases and determines the characteristic of a source according to the type of multiplexing process. The most important source parameter for the classification is burst size. Even bursts with two or three cells influence the utilisation factor, which means that the methods of analysis for cell scale traffic are not valid in these cases. If the ratio of the burst size to buffer size is more than four, there is no need to use complicated burst scale models but relative simple rate-variation scale models are adequate for analysing QoS.

There is an essential difference in characteristic behaviour between the traffic processes of cell scale and rate-variation scale. A pure cell scale traffic flow consists of an arrival process of independent cells. The bandwidth required by a source of this type is almost independent of the other traffic components, and consequently, a linear model, called effective bandwidth, is a suitable traffic model. If rate-variation scale fluctuations are prominent, the main issue is whether there is enough link capacity at any given instant. In this case the multiplexing process is different and another approach, called effective variance, is much more accurate than the effective bandwidth model.

The ambiguous area between the two extreme cases is the most challenging. In this study a combination of effective bandwidth and effective variance, the EBV model, has been developed to describe burst scale sources. Moreover, EBV is an adequate model when diverse source types are mixed. The validity of the EBV model has been evaluated by extensive simulations. The standard deviation of error in allowable load obtained by EBV is only 1.3% while the corresponding values for two effective bandwidth formulae (EB_1 and EB_2), and effective variance are 4.2%, 3.1% and 3.3%, respectively.

The main requirements for the CAC method are an efficient use of network resources, and simplicity concerning both parameter determination and CAC calculation at network nodes. The last requirement is fulfilled in this study by separating the CAC

method into two parts: approximation of the source parameters based mainly on homogeneous traffic and the approximation for the combination of various source types. Using this separation any method suitable for homogeneous cases can be applied to any of the traffic models appropriate to heterogeneous traffic cases. This separation offers a very flexible framework to develop CAC methods.

Since the effect of burst size on the allowed load is very difficult to evaluate precisely, most CAC methods have been based on rate-variation scale models. At rate-variation scale the most promising principles for the heterogeneous part of CAC procedure are effective bandwidth and effective variance although essentially different approaches have been used, such as on-line traffic measurements and neural networks. The effective bandwidth method has two basic modifications. In the first one, the effective bandwidth of each source is calculated purely from a homogeneous case and cell loss probability is adjusted by a parameter common to all sources. In the second one a higher utilisation is achieved by determining the effective bandwidth separately for each source type using a background traffic.

Homogeneous and heterogeneous approximations can be combined in numerous ways. In this study seven different combinations have been thoroughly evaluated with various rate-variation scale models. The results of performance evaluation can be summarised in the gain achieved from peak rate allocation to ideal allocation. Effective variance with a large deviation approximation offers at best 95% of the possible gain. With effective bandwidth methods, values from 75 to 90% can be attained.

The simplicity of implementation including parameter calculation and routing aspects is the other main requirement for a practical CAC method. By combining the results of evaluation concerning both efficiency and implementation, three promising candidates for CAC have been identified. Lindberger's approximation is appropriate when the simplicity of implementation is the most important aspect. Effective bandwidth with a large deviation approximation and an optimisation technique using effective variance approximation is suitable when a simple CAC procedure is necessary but source parameter determination can be rather complicated. Effective variance combined with large deviation approximation results in the highest utilisation with rate-variation scale traffic.

The definitive selection between different CAC methods depends on the assessment of different aspects and it cannot be made without knowledge of the development of ATM technology and the behaviour of real ATM traffic. There are many sources of uncertainty and ways of dependencies which are very difficult to take into account in CAC methods. Some of these problems may be alleviated by additional control functions such as Fast Resource Management. However, the actual capability of traffic control functions can be tested only in a real environment with various traffic sources.

REFERENCES

Aagesen, F. A. 1993. A Flow Management Architecture for B-ISDN, Proc. IBCN&S, Copenhagen, pp. 14.3.1-14.

Aarstad, E. 1993. A Comment on Worst Case Traffic, COST 242, doc. TD(93)39.

Addie, R. D. & M. Zukerman 1993. A Gaussian Characterization of Correlated ATM Multiplexed Traffic and Related Queueing Studies, Proc. IEEE ICC '93, Geneva, pp. 1404-1408.

Akar, N. & E. Arikan 1993. Markov Modulated Periodic Arrival Process Offered to an ATM Multiplexer, Proc. IEEE ICC '93, Geneva, pp. 783-787.

Alparone, L., F. Argenti, L. Capriotti & G. Benelli 1992. Models for ATM Video Packet Transmission, European Transactions on Telecommunications and Related Technologies, Vol. 3, No. 5, pp. 491-497.

Andrade, J. & M. Villen 1993. Including the Second Moment of the Cell Rate in the Source Traffic Descriptor, COST 242, doc. TD(93)38.

Appleton, J. 1991. Performance Related Issues Concerning the Contract Between Network and Customer in ATM Networks, BT Technol. J., Vol. 9, No. 4, pp. 57-60.

Bae, J. J. & T. Suda 1991. Survey of Traffic Control Schemes and Protocols in ATM Networks, Proc. IEEE, Vol. 79, No. 2, pp. 170-189.

Baiocchi, A., N. B. Melazzi, M. Listanti, A. Roveri & R. Winkler 1991. Loss Performance Analysis of an ATM Multiplexer Loaded with High-Speed on-off Sources, IEEE J. Selected Areas Commun., Vol. SAC-9, No. 3, pp. 388-393.

Bean, N. G. 1993. Effective Bandwidths with Different Quality of Service Requirements, Proc. IBCN&S, Copenhagen, pp. 13.3.1-12.

Beneš, V. E. 1963. *General Stochastic Processes in the Theory of Queues*, Addison Wesley.

Bensaou, B., J. Guibert & J. W. Roberts 1990. Fluid Queueing Models for a Superposition of on/off Sources, 7th ITC Specialist Seminar, Morristown, NJ.

Bermejo-Saez, L. & G. H. Petit 1991. Bandwidth Resource Dimensioning in ATM Networks: A Theoretical Approach and Some Study Cases, 13th ITC, Copenhagen, Vol. 14, pp. 929-934.

Blaabjerg, S. 1991. On the Approximation of Heterogeneous Fluid Queues, COST 242 Seminar, Paris.

Bonomi, F., S. Montagna & R. Paglino 1993. A Further Look at Statistical Multiplexing in ATM Networks, Computer Networks and ISDN systems, Vol. 26, pp. 119-138.

Brady, P. T. 1969. A Model for Generating on-off Speech in Two-Way Conversations, Bell Syst. Tech. J., Vol. 48, Sept.

Burgin, J. 1990. Broadband ISDN Resource Management, Computer Networks and ISDN Systems, Vol. 20, pp. 323-331.

Castelli, P., E. Cavallero & A. Tonietti 1991. Policing and Call Admission Problems in ATM Networks, Proc. 13th ITC, Copenhagen, Vol. 14, pp. 847-852.

Coudreuse, J.-P. 1983. Les Réseaux Temporels Asynchrones: du Transfert de Données à L'Image Animée, L'Echo des Recherches, No. 112, pp. 33-48.

Coudreuse, J.-P. 1991. General Principles of ATM, L'Echo des Recherches, English Issue, pp. 5-16.

Decina, M. & T. Toniatti 1990. On Bandwidth Allocation to Bursty Virtual Connections in ATM Networks, Proc. IEEE ICC '90, Atlanta, GA, pp. 844-851.

Doshi, B. 1993. Deterministic Rule Based Traffic Descriptors for Broadband ISDN: Worst Case Behavior and Connection Acceptance Control, Proc. IEEE ICC '93, Geneva, pp. 1759-1764.

Doshi, B., S. Dravida, P. Johri & G. Ramamurthy 1991. Memory, Bandwidth, Processing and Fairness Considerations in Real Time Congestion Controls for Broadband Networks, Proc. 13th ITC, Copenhagen, Vol. 14, pp. 153-159.

Drakopoulos, E. 1993. Performance and Traffic Analysis of Network-Based Distributed System, Computer Communications, Vol. 16, No. 3, pp. 155-167.

Dziong, Z., K.-Q. Liao & L. Mason 1993. Effective Bandwidth Allocation and Buffer Dimensioning in ATM Based Networks with Priorities, Computer Networks and ISDN Systems, Vol. 25, pp. 1065-1078.

Eckberg, A. E., D. M. Lucantoni & P. K. Prasanna 1991. Congestion Control Issues and Strategies Associated with B-ISDN/ATM Access and Network Transport, Proc. ISSLS '91, pp. 196-202.

Elwalid, A. I. & D. Mitra 1993. Effective Bandwidth of Bursty, Variable Rate Sources for Admission Control to B-ISDN, Proc. IEEE ICC '93, Geneva, pp. 1325-1330.

Esaki, H. 1992. Call Admission Control Method in ATM Networks, Proc. IEEE ICC '92, Geneva, pp. 1628-1633.

Eurescom 1993. Project P105, European ATM Network Studies, Deliverable No. 3, Vol. 3 of 4: VP Connection Admission Control (Task 7.1).

Falaki, S. O. & S.-A. Sørensen 1992. Traffic Measurements on a Local Area Computer Network, Computer Communications, Vol. 15, No. 3, pp. 192-197.

Fowler, H. J. & W. E. Leland 1991. Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management, IEEE J. Selected Areas Commun., Vol. SAC-9, No. 7, pp. 1139-1149.

Fritsch, T., M. Mittler & P. Tran-Gia 1992. Artificial Neural Net Applications in Telecommunication Systems, Univ. of Würzburg, Institute of Computer Science, Report 52.

Fuhrmann, S. & J.-Y. Le Boudec 1991. Burst and Cell Level Models for ATM Buffers, Proc. 13th ITC, Copenhagen, Vol. 14, pp. 975-980.

Gallassi, G., G. Rigolio & L. Fratta 1989. ATM: Bandwidth Assignment and Bandwidth Enforcement Policies, Proc. Globecom '89, Dallas, TX, pp. 1788-1793.

Gilbert, H., O. Aboul-Magd & V. Phung 1991. Developing a Cohesive Traffic Management Strategy for ATM Networks, IEEE Commun. Mag., Oct., pp. 36-45.

Girard A. & N. Lessard 1992. Revenue Optimization of Virtual Circuit ATM Networks, Proc. Networks '92, Kobe, pp. 183-188.

Griffiths, T. R. 1990. Analysis of a Connection Acceptance Strategy for Asynchronous Transfer Mode Networks, Proc. Globecom '90, San Diego, pp. 862-868.

Gropp, O. 1993. Modeling Concepts for VBR Video Sources, COST 242, doc. TD(93)24.

Grünenfelder, R., J. P. Cosmas, S. Manthorpe & A. Odinma-Okafor 1991. Characterization of Video Codecs as Autoregressive Moving Average Processes and Related Queueing System Performance, IEEE J. Selected Areas Commun., Vol. SAC-9, No. 3, pp. 284-293.

Guérin, R., H. Ahmadi & M. Naghshineh 1991. Equivalent Capacity and Its Application to Bandwidth Allocation in High-Speed Networks, IEEE J. Selected Areas Commun., Vol. SAC-9, No. 7, pp. 968-981.

Gusella, R. 1991. Characterizing the Variability of Arrival Processes with Indexes of Dispersion, IEEE J. Selected Areas Commun., Vol. SAC-9, No. 2, pp. 203-211.

Heegaard, P. E. & B. E. Helvik 1993. Establishing ATM Source Models for Traffic Measurement, 11th Nordic Teletraffic Seminar, Stockholm.

Heffes, H. & D. M. Lucantoni 1986. A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance, IEEE J. Selected Areas Commun., Vol. SAC-4, No. 6, pp. 856-868.

Helvik, B. E., P. Hokstad & N. Stol 1991. Correlation in ATM Traffic Streams - Some Results, Proc. 13th ITC, Copenhagen, Vol. 15, pp. 25-32.

Herzberg, M. & A. Pitsillides 1993. A Hierarchical Approach for the Bandwidth Allocation, Management and Control in B-ISDN, Proc. IEEE ICC '93, Geneva, pp. 1320-1324.

Holtzman, J. M. 1990. Coping with Broadband Traffic Uncertainties: Statistical Uncertainty, Fuzziness, Neural Networks, Proc. Globecom '90, San Diego, pp. 7-11.

Hui, J. Y. 1990. *Switching and Traffic Theory for Integrated Broadband Networks*, Kluwer Academic Publishers, Boston.

Hui, J. Y., M. B. Gursoy, N. Moayeri & R. D. Yates 1991. A Layered Broadband Switching Architecture with Physical or Virtual Path Configurations, IEEE J. Selected Areas Commun., Vol. SAC-9, No. 9, pp. 1416-1426.

Hübner, F. & P. Tran-Gia 1991. Quasi-Stationary Analysis of a Finite Capacity Asynchronous Multiplexer with Modulated Deterministic Input, Proc. 13th ITC, Copenhagen, Vol. 14, pp. 723 -729.

Hyman, J. M., A. A. Lazar & G. Pacifici 1993. A Separation Principle Between Scheduling and Admission Control for Broadband Switching, IEEE J. Selected Areas Commun., Vol. SAC-11. No. 4, pp. 605-616.

International Telecommunication Union, Telecommunication Standardization Sector (ITU-T) 1993a. Recommendation I.371, Traffic Control and Congestion Control in B-ISDN, Geneva.

International Telecommunication Union, Telecommunication Standardization Sector (ITU-T) 1993b. ITU-T COM 13-R 4-E, Annex 10, I.371 Living list, Geneva.

Iversen, V. B. & A. Bohn Nielsen 1992. Traffic Descriptors for B-ISDN, 10th Nordic Teletraffic Seminar, Århus.

Jain, R. 1990. Congestion Control in Computer Networks: Issues and Trends, IEEE Network Magazine, May, pp. 24-30.

Joos, P. & W. Verbiest 1989. Statistical Bandwidth Allocation and Usage Monitoring Algorithm for ATM Networks, Proc. IEEE ICC '89, Boston, MA, Vol. 1, pp. 415-422.

Kawashima, K. & H. Saito 1990. Teletraffic Issues in ATM Networks, Computer Networks and ISDN Systems, Vol. 20, pp. 369-375.

Kelly, F. P. 1991. Effective Bandwidths at Multi-Class Queues, Queueing Systems, Sept., pp. 5-16.

Kelly, F. P. 1993. On Tariffs, Policing and Admission Control for Multiservice Networks, IFIP 7.3 Workshop, Vitznau, Switzerland.

Kilkki, K. 1992. Connection Admission Control Methods for ATM Networks, COST 242, doc. TD(92)61.

Kleinrock, L. 1975. *Queueing Systems. Vol. I: Theory*, John Wiley & Sons, New York, NY.

Kleinrock, L. 1992. The Latency/Bandwidth Tradeoff in Gigabit Networks, IEEE Commun. Mag., April pp. 36-40.

Kröner, H. 1991. Statistical Multiplexing of Sporadic Sources - Exact and Approximative Performance Analysis, Proc. 13th ITC, Copenhagen, Vol. 14, pp. 787-792.

Lindberger, K. 1991. Analytical Methods for the Traffical Problems with Statistical Multiplexing in ATM-Networks, Proc. 13th ITC, Copenhagen, Vol. 14, pp. 807-813.

Lindberger, K. 1992a. Charging Principles Against Tricky Users in ATM Networks, COST 242, doc. TD(92)43.

Lindberger, K. 1992b. LAN-Traffic, Public Networks and GOS, COST 242, doc. TD(92)56.

Lyons, M. H., K. O. Jensen & I. Hawker 1993. Traffic Scenarios for the 21st Century, BT Technol. J., Vol. 11, No. 4, pp. 73-84.

Minoli, D. 1993. *Broadband Network Analysis and Design*, Artech House, Boston.

Miyao, Y. 1993. Bandwidth Allocation in ATM Networks That Guarantee Multiple QoS Requirements, Proc. IEEE ICC '93, Geneva, pp. 1398-1402.

Newman, P. 1993. Backward Explicit Congestion Notification for ATM Local Area Networks, Proc. IEEE ICC '93, Geneva, pp. 719-723.

Norros, I. 1992. The Tricky User and Other Causes of Dependence Between ATM Streams, 10th Nordic Teletraffic Seminar, Århus.

Norros, I. 1993. A Simple Model for Connectionless Traffic with Long-Range Correlations, 11th Nordic Teletraffic Seminar, Stockholm.

Norros, I., J. W. Roberts, A. Simonian & J. T. Virtamo 1991. The Superposition of Variable Bit Rate Sources in an ATM Multiplexer, IEEE J. Selected Areas Commun., Vol. SAC-9, No. 3, pp. 378-387.

Okuda, T., H. Akimaru & K. Nagai 1992. Performance Evaluation for Multiclass Traffic in ATM Systems, Proc. IEEE ICC '92, Chicago, pp. 207-210.

Pettersen, H. 1993. Connection Admission Control and Resource Allocation - Some Case Studies, 11th Nordic Teletraffic Seminar, Stockholm.

Rahko, K. 1967. The Dimensioning of Local Telephone Traffic Routes Based on the Distribution of the Traffic Carried, Acta Polytechnica Scandinavica, Electrical Engineering Series.

Rahko, K. 1976. Kauko Rahko's Publications on Traffic Theory from 1964 to 1976, assembled by S. Hertzberg, Helsinki University of Technology, Telecommunications Switching Laboratory, Report No. 6/76.

Rahko, K. 1983. Kauko Rahko's Publications on Traffic Theory from 1977 to 1983, assembled by S. Hertzberg, Helsinki University of Technology, Telecommunications Switching Laboratory, Report No. 4/83.

Rahko, K. & S. Hertzberg 1988. *Traffic Measurements, A TETRAPRO Specialized Course*, Otakustantamo, Espoo.

Ramamurthy, G. & R. S. Dighe 1991. A Multidimensional Framework for Congestion Control in B-ISDN, IEEE J. Selected Areas Commun., Vol. SAC-9, No. 9, pp. 1440-1451.

Ramamurthy, G. & B. Sengupta 1990. Modeling and Analysis of a Variable Bit Rate Video Multiplexer, 7th ITC Specialist Seminar, Morristown, NJ.

Rasmussen, C., J. H. Sørensen, K. S. Kvols & S. B. Jacobsen 1991. Source-Independent Call Acceptance Procedures in ATM Networks, IEEE J. Selected Areas Commun., Vol. SAC-9, No. 3, pp. 351-358.

Riordan, J. 1951. Telephone Traffic Time Averages, B.S.T.J, Vol. 30, No. 4, 1951, part II, pp. 1129-1144.

Roberts, J. W. (Ed.) 1992a. *Performance Evaluation and Design of Multiservice Networks*, COST 224 Final Report, CEC, Luxembourg.

Roberts, J. W. 1992b. LAN-LAN Interconnection and B-ISDN, COST 242, doc. TD(92)10.

Roberts, J. W. 1992c. Queueing Models for Variable Bit Rate Traffic Streams, COST 242, doc. TD(92)67.

Roberts, J. W. 1993a. Traffic Control in the B-ISDN, Computer Networks and ISDN Systems, Vol. 25, pp. 1055-1064.

Roberts, J. W. 1993b. Virtual Spacing for Integrated High Speed Data and Real Time Services, COST 242, doc. TD(93)34.

Roberts, J. W. 1993c. Resource Allocation Using Sustainable Cell Rate and Burst Tolerance Traffic Parameters, COST 242 TD(93)35.

Roberts, J. W., B. Bensaou & Y. Canetti 1992. A Traffic Control Framework for High Speed Data Transmission, COST 242, doc. TD(92)51, Issue 2.

Roberts, J. W., J. Guibert & A. Simonian 1991. Network Performance Considerations in the Design of a VBR Codec, Proc. 13th ITC, Copenhagen, Vol. 15, pp. 77-82.

Saito, H. 1992a. Hybrid Connection Admission Control in ATM Networks, Proc. IEEE ICC '92, Chicago, pp. 699-703.

Saito, H. 1992b. Call Admission Control in an ATM Networks Using Upper Bound of Cell Loss Probability, IEEE Trans. on Comm., Vol. 40, No. 9, pp. 1512-1521.

Smit, T. A. 1993. The Poor Gain from Statistical Multiplexing in the Homogeneous and the Heterogeneous Case, Proc. IBCN&S, Copenhagen, pp. 13.1-11.

Sriram, K. 1993. Methodologies for Bandwidth Allocation, Transmission Scheduling, and Congestion Avoidance in Broadband ATM Networks, Computer Networks and ISDN Systems, Vol. 26, pp. 43-69.

Sriram, K. & W. Whitt 1986. Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data, IEEE J. Selected Areas Commun., Vol. SAC-4, No. 6, pp. 833-846.

Sykas, E., K. M. Vlakos & N. G. Anerousis 1991. Performance Evaluation of Statistical Multiplexing Schemes in ATM Networks, Computer Communications, Vol. 14, No. 5, pp. 273-286.

Takahashi, T. & A. Hiramatsu 1990. Integrated ATM Traffic Control by Distributed Neural Networks, Proc. ISS '90, Stockholm, Vol. 3, pp. 59-65.

Tidblom, S.-E. 1992. Complete Equivalent Bandwidth Formulae for Various Cell Loss Ratios, COST 242 TD(92)36.

Tranchier, D. P., P. E. Boyer, Y. M. Rouaud & J.-Y. Mazeas 1992. Fast Bandwidth Allocation in ATM Networks, Proc. ISS '92, Yokohama, Vol. 2, pp. 7-11.

Uose, H., S. Shioda & K. Mase 1990. Fast Cell Loss Rate Evaluation Methods and Their Application to ATM Network Control, 7th ITC Specialist Seminar, Morristown, NJ.

Virtamo, J. T. & I. Norros 1991. Who Loses Cells in the Case of Burst Scale Congestion? Proc. 13th ITC, Copenhagen, Vol. 14, pp. 829-833.

Virtamo, J. T. & J. W. Roberts 1989. Evaluating Buffer Requirements in an ATM Multiplexer, Globecom '89, Dallas, TX, pp. 1473-1477.

Wallmeier, E. & C. M. Hauber 1991. Blocking Probabilities in ATM Pipes Controlled by a Connection Acceptance Algorithm Based on Mean and Peak Bit Rates, Proc. 13th ITC, Copenhagen, Vol. 15, pp. 137-142.

Yang, T. & H. Li 1993. Individual Cell Loss Probabilities and Background Effects in ATM networks, Proc. IEEE ICC '93, Geneva, pp. 1373-1379.

Yegenoglu, Y. & B. Jabbari 1993. Modeling of Aggregated Bursty Traffic Sources in ATM Multiplexers, Proc. IEEE ICC '93, Geneva, pp. 1703-1707.

APPENDIX A. SOURCES USED IN SIMULATIONS

K = 100, P_{loss} = 10^{-4}, c = 1, ρ_{max} = 0.9 in EB_2 models.

Cell scale sources

	D	m	N_c	ρ_{hom}	ε_u	ε_m	k	k^*	v^*	σ^{**}	v^{**}
C	100	0.01	100.00	1.000			1.00E-	1.00E-	0	0	0
C	1000	0.001	1000.0	1.000			1.00E-	1.00E-	0	0	0
C	4000	0.0002	3931.2	0.983			2.54E-	2.50E-	7.52E-	6.24E-	-7.79E-

Burst scale sources

	$1/h$	L	D	p_{burst}	m	N_c	ρ_{hom}	ε_u	ε_m	k	k^*	v^*	σ^{**}	v^{**}
B1	5	10	2000	0.025	0.005	151.59	0.758	0.26	-0.57	6.60E-3	5.94E-3	3.86E-4	1.94E-3	-1.85E-
B2	5	10	8000	0.00625	0.00125		0.722	0.30	-0.15	1.73E-3	1.56E-3	1.34E-4	5.13E-4	-1.81E-
B3	10	10	2000	0.05	0.005	158.10	0.790	0.28	-0.62	6.33E-3	5.69E-3	2.78E-4	1.63E-3	-1.40E-
B4	10	10	8000	0.0125	0.00125		0.751	0.33	-0.03	1.66E-3	1.50E-3	1.03E-4	4.20E-4	-3.07E-
B5	15	10	2000	0.075	0.005	162.47	0.812	0.29	-0.59	6.16E-3	5.54E-3	2.17E-4	1.40E-3	-1.02E-
B6	15	10	4000	0.0375	0.0025	312.42	0.781	0.34	-0.15	3.20E-3	2.88E-3	1.53E-4	7.45E-4	-2.00E-
B7	15	10	8000	0.01875	0.00125		0.765	0.36	-0.04	1.63E-3	1.47E-3	8.99E-5	3.89E-4	-2.80E-
B8	30	10	2000	0.15	0.005	171.81	0.859	0.30	-0.81	5.82E-3	5.24E-3	1.16E-4	1.04E-3	-7.12E-
B9	30	10	4000	0.075	0.0025	327.58	0.819	0.38	-0.11	3.05E-3	2.75E-3	1.00E-4	5.77E-4	-9.11E-
B10	30	10	8000	0.0375	0.00125		0.801	0.41	0.07	1.56E-3	1.40E-3	6.17E-5	3.01E-4	3.56E-6
B11	60	10	4000	0.15	0.0025	345.45	0.864	0.40	-0.11	2.89E-3	2.61E-3	5.38E-5	4.13E-4	-5.04E-
B12	60	10	8000	0.075	0.00125		0.841	0.46	0.10	1.49E-3	1.34E-3	3.77E-5	2.27E-4	3.16E-6
B13	100	10	2000	0.5	0.005	193.74	0.969	0.03	-2.47	5.16E-3	5.00E-3	5.05E-6	2.32E-4	-5.39E-
B14	100	10	8000	0.125	0.00125		0.871	0.49	0.14	1.43E-3	1.29E-3	2.37E-5	1.73E-4	2.87E-6
B15	5	20	1000	0.1	0.02	33.18	0.664	0.39	-0.68	3.01E-2	2.71E-2	3.41E-3	1.28E-2	-2.04E-
B16	5	20	4000	0.025	0.005	112.31	0.562	0.50	-0.11	8.90E-3	8.01E-3	1.71E-3	4.10E-3	-1.76E-
B17	10	20	1000	0.2	0.02	38.12	0.762	0.36	-0.68	2.62E-2	2.36E-2	1.48E-3	7.77E-3	-8.20E-
B18	10	20	4000	0.05	0.005	127.21	0.636	0.51	0.08	7.86E-3	7.07E-3	1.04E-3	2.75E-3	7.78E-5
B19	15	20	1000	0.3	0.02	41.49	0.830	0.31	-1.02	2.41E-2	2.17E-2	6.99E-4	5.44E-3	-5.31E-
B20	15	20	2000	0.15	0.01	72.85	0.729	0.46	-0.12	1.37E-2	1.24E-2	1.01E-3	3.91E-3	-1.04E-
B21	15	20	4000	0.075	0.005	135.81	0.679	0.53	0.11	7.36E-3	6.63E-3	7.58E-4	2.24E-3	7.76E-5
B22	15	20	8000	0.0375	0.0025	263.14	0.658	0.55	0.21	3.80E-3	3.42E-3	4.45E-4	1.17E-3	8.66E-5
B23	30	20	2000	0.3	0.01	82.67	0.827	0.43	-0.28	1.21E-2	1.09E-2	3.63E-4	2.33E-3	-8.48E-
B24	30	20	4000	0.15	0.005	152.34	0.762	0.55	0.17	6.56E-3	5.91E-3	3.73E-4	1.44E-3	5.58E-5
B25	30	20	8000	0.075	0.0025	296.48	0.741	0.57	0.41	3.37E-3	3.04E-3	2.26E-4	6.88E-4	8.57E-5
B26	60	20	4000	0.3	0.005	169.65	0.848	0.52	0.15	5.89E-3	5.31E-3	1.36E-4	8.31E-4	1.86E-5
B27	60	20	8000	0.15	0.0025	322.43	0.806	0.62	0.35	3.10E-3	2.79E-3	1.17E-4	4.97E-4	3.71E-5
B28	100	20	4000	0.5	0.005	183.87	0.919	0.38	-0.23	5.44E-3	5.00E-3	3.54E-5	4.79E-4	-6.72E-
B29	5	40	2000	0.1	0.02	23.25	0.465	0.65	0.04	4.30E-2	3.87E-2	1.23E-2	2.25E-2	5.08E-4
B30	5	40	8000	0.025	0.005	78.35	0.392	0.70	0.26	1.28E-2	1.15E-2	4.72E-3	6.72E-3	1.18E-3

	$1/h$	L	D	p_{burst}	m	N_c	ρ_{hom}	ε_u	ε_m	k	k^*	v^*	σ^{**}	v^{**}
B31	10	40	2000	0.2	0.02	30.39	0.608	0.62	0.10	3.29E-	2.96E-	5.06E-	1.23E-	4.65E-4
B32	10	40	8000	0.05	0.005	103.10	0.516	0.69	0.42	9.70E-	8.74E-	2.28E-	3.64E-	9.14E-4
B33	15	40	2000	0.3	0.02	35.21	0.704	0.58	0.08	2.84E-	2.56E-	2.49E-	8.10E-	1.78E-4
B34	15	40	4000	0.15	0.01	62.08	0.621	0.67	0.36	1.61E-	1.45E-	2.32E-	4.98E-	7.73E-4
B35	15	40	8000	0.075	0.005	116.18	0.581	0.71	0.47	8.61E-	7.75E-	1.51E-	2.68E-	6.75E-4
B36	30	40	2000	0.6	0.02	44.89	0.898	0.32	-0.65	2.23E-	2.01E-	2.33E-	2.78E-	-1.15E-
B37	30	40	4000	0.3	0.01	75.51	0.755	0.64	0.36	1.32E-	1.19E-	7.94E-	2.65E-	2.63E-4
B38	30	40	8000	0.15	0.005	140.15	0.701	0.71	0.55	7.14E-	6.42E-	6.39E-	1.48E-	3.31E-4
B39	60	40	4000	0.6	0.01	90.19	0.902	0.44	-0.13	1.11E-	1.00E-	1.07E-	1.15E-	-1.15E-
B40	60	40	8000	0.3	0.005	161.31	0.807	0.70	0.45	6.20E-	5.58E-	2.32E-	9.12E-	9.78E-5
B41	100	40	8000	0.5	0.005	177.45	0.887	0.60	0.32	5.64E-	5.07E-	7.16E-	5.37E-	2.05E-5
B42	5	80	1000	0.4	0.08	7.79	0.623	0.61	-0.48	1.28E-	1.16E-	1.82E-	5.79E-	-7.91E-
B43	5	80	4000	0.1	0.02	17.23	0.345	0.80	0.43	5.80E-	5.39E-	2.49E-	2.91E-	1.04E-2
B44	10	80	4000	0.2	0.02	25.88	0.518	0.77	0.47	3.86E-	3.49E-	8.99E-	1.38E-	4.04E-3
B45	15	80	2000	0.6	0.04	20.53	0.821	0.49	-0.25	4.87E-	4.38E-	1.56E-	9.59E-	-3.29E-
B46	15	80	4000	0.3	0.02	31.34	0.627	0.75	0.45	3.19E-	2.87E-	4.45E-	9.02E-	1.90E-3
B47	15	80	8000	0.15	0.01	54.90	0.549	0.80	0.56	1.82E-	1.66E-	3.70E-	5.60E-	1.99E-3
B48	30	80	4000	0.6	0.02	41.97	0.839	0.57	0.09	2.38E-	2.14E-	6.15E-	3.67E-	5.05E-5
B49	30	80	8000	0.3	0.01	70.93	0.709	0.77	0.60	1.41E-	1.27E-	1.19E-	2.68E-	6.80E-4
B50	60	80	8000	0.6	0.01	86.94	0.869	0.63	0.33	1.15E-	1.04E-	1.96E-	1.26E-	5.79E-5
B51	5	160	2000	0.4	0.08	6.19	0.495	0.83	-0.21	1.61E-	1.45E-	4.11E-	8.91E-	-8.00E-
B52	5	160	8000	0.1	0.02	13.75	0.275	0.89	0.70	7.27E-	8.47E-	3.82E-	3.00E-	2.59E-2
B53	10	160	8000	0.2	0.02	23.32	0.466	0.86	0.67	4.29E-	4.16E-	1.22E-	1.36E-	7.87E-3
B54	15	160	4000	0.6	0.04	19.00	0.760	0.69	0.14	5.26E-	4.74E-	3.04E-	1.18E-	3.96E-4
B55	15	160	8000	0.3	0.02	29.05	0.581	0.84	0.59	3.44E-	3.13E-	6.04E-	9.48E-	3.44E-3
B56	30	160	8000	0.6	0.02	40.18	0.804	0.73	0.39	2.49E-	2.24E-	9.60E-	3.90E-	3.49E-4
B57	5	320	4000	0.4	0.08	5.69	0.456	0.90	0.39	1.76E-	1.58E-	5.21E-	7.55E-	1.96E-2
B58	15	320	8000	0.6	0.04	18.05	0.722	0.81	0.21	5.54E-	4.99E-	4.28E-	1.38E-	8.21E-4
B59	5	10	500	0.1	0.02	46.17	0.923	0.06	< -1	2.17E-	2.00E-	1.27E-	-	-
B60	10	10	500	0.2	0.02	49.07	0.981	-0.01	< -1	2.04E-	2.00E-	7.04E-	-	-
B61	15	10	500	0.3	0.02	50.00	1.000	-0.06	< -1	2.00E-	2.00E-	0	-	-
B62	15	10	1000	0.15	0.01	88.00	0.880	0.18	< -1	1.14E-	1.02E-	1.64E-	-	-
B63	5	20	16000	0.00625	0.00125		0.542	0.51	0.06	2.31E-	2.08E-	4.85E-	1.03E-	2.90E-5
B64	15	20	16000	0.01875	0.00125		0.648	0.57	0.28	1.93E-	1.74E-	2.40E-	5.85E-	6.25E-5
B65	30	20	1000	0.6	0.02	49.58	0.992	-0.08	< -1	2.02E-	2.00E-	1.42E-	-	-
B66	60	20	16000	0.075	0.00125		0.789	0.65	0.48	1.58E-	1.43E-	7.05E-	2.48E-	3.18E-5
B67	15	40	16000	0.0375	0.0025	225.17	0.563	0.72	0.53	4.44E-	4.02E-	8.48E-	1.36E-	4.34E-4
B68	5	80	16000	0.025	0.005	57.07	0.285	0.83	0.58	1.75E-	1.83E-	8.95E-	8.29E-	5.03E-3
B69	15	80	16000	0.075	0.005	103.39	0.517	0.82	0.65	9.67E-	9.10E-	2.26E-	2.83E-	1.43E-3
B70	15	160	16000	0.15	0.01	51.19	0.512	0.88	0.73	1.95E-	1.89E-	4.65E-	5.16E-	3.29E-3

	$1/h$	L	D	p_{burst}	m	N_c	ρ_{hom}	ε_u	ε_m	k	k^*	v^*	σ^{**}	v^{**}
B71	60	160	1600	0.6	0.01	84.74	0.847	0.77	0.54	1.18E-2	1.06E-2	2.75E-4	1.26E-3	1.40E-4
B72	5	320	1600	0.1	0.02	11.62	0.232	0.94	0.78	8.61E-2	1.14E-1	5.07E-2	3.27E-2	3.83E-2
B73	10	320	1600	0.2	0.02	21.56	0.431	0.92	0.73	4.64E-2	4.78E-2	1.50E-2	1.42E-2	1.07E-2
B74	15	320	1600	0.3	0.02	27.75	0.555	0.90	0.71	3.60E-2	3.39E-2	7.14E-3	8.89E-3	4.95E-3
B75	30	320	1600	0.6	0.02	38.97	0.779	0.83	0.53	2.57E-2	2.31E-2	1.25E-3	4.01E-3	6.23E-4

Rate-variation scale sources

	$1/h_1$	$1/h_2$	$1/h_3$	p_1	p_2	m	N_c	ρ_{hom}	k	k^*	v^*	σ^{**}	v^{**}
R1	20			0.5		0.025	25.82	0.64	3.87E-	3.72E-	4.87E-	0	4.87E-
R2	50			0.5		0.01	76.23	0.76	1.31E-	1.19E-	7.41E-	0	7.41E-
R3	100			0.5		0.005	166.17	0.83	6.02E-	5.42E-	1.72E-	0	1.72E-
R4	200			0.5		0.0025	352.77	0.88	2.83E-	2.55E-	3.95E-	0	3.95E-
R5	500			0.5		0.001	928.10	0.92	1.08E-	1.00E-	5.57E-	0	5.57E-
R6	1000			0.5		0.0005	1902.35	0.95	5.26E-	5.00E-	1.25E-	0	1.25E-
R7	20			0.2		0.01	52.49	0.52	1.91E-	2.08E-	4.30E-	0	4.30E-
R8	50			0.2		0.004	171.60	0.68	5.83E-	5.43E-	5.73E-	0	5.73E-
R9	100			0.2		0.002	388.88	0.77	2.57E-	2.32E-	1.27E-	0	1.27E-
R10	200			0.2		0.001	845.12	0.84	1.18E-	1.06E-	2.84E-	0	2.84E-
R11	500			0.2		0.0004	2264.13	0.90	4.42E-	4.00E-	3.93E-	0	3.93E-
R12	1000			0.2		0.0002	4679.15	0.93	2.14E-	2.00E-	8.80E-	0	8.80E-
R13	20			0.1		0.005	98.12	0.49	1.02E-	1.16E-	2.64E-	0	2.64E-
R14	50			0.1		0.002	332.19	0.66	3.01E-	2.85E-	3.39E-	0	3.39E-
R15	100			0.1		0.001	762.28	0.76	1.31E-	1.19E-	7.41E-	0	7.41E-
R16	200			0.1		0.0005	1668.72	0.83	5.99E-	5.39E-	1.64E-	0	1.64E-
R17	500			0.1		0.0002	4495.30	0.89	2.22E-	2.00E-	2.27E-	0	2.27E-
R18	1000			0.1		0.0001	9313.15	0.93	1.07E-	1.00E-	5.07E-	0	5.07E-
R19	20			0.05		0.0025	189.70	0.47	5.27E-	6.14E-	1.46E-	0	1.46E-
R20	50			0.05		0.001	653.83	0.65	1.53E-	1.46E-	1.83E-	0	1.83E-
R21	100			0.05		0.0005	1509.71	0.75	6.62E-	5.99E-	3.98E-	0	3.98E-
R22	200			0.05		0.00025	3316.73	0.82	3.02E-	2.71E-	8.80E-	0	8.80E-
R23	500			0.05		0.0001	8958.89	0.89	1.12E-	1.00E-	1.21E-	0	1.21E-
R24	20			0.02		0.001	464.69	0.46	2.15E-	2.54E-	6.17E-	0	6.17E-
R25	50			0.02		0.0004	1619.12	0.64	6.18E-	5.92E-	7.67E-	0	7.67E-
R26	100			0.02		0.0002	3752.45	0.75	2.66E-	2.41E-	1.66E-	0	1.66E-
R27	200			0.02		0.0001	8261.39	0.82	1.21E-	1.09E-	3.66E-	0	3.66E-
R28	20			0.01		0.0005	1046.20	0.52	9.56E-	1.04E-	2.17E-	0	2.17E-
R29	50			0.01		0.0002	3228.03	0.64	3.10E-	2.97E-	3.89E-	0	3.89E-
R30	100			0.01		0.0001	7490.49	0.74	1.34E-	1.21E-	8.41E-	0	8.41E-

	$1/h_1$	$1/h_2$	$1/h_3$	p_1	p_2	m	N_c	ρ_{hom}	k	k^*	v^*	σ^{**}	v^{**}
R31	20	200	0	0.5	0.5	0.0275	24.92	0.68	4.01E-2	3.74E-	3.97E-	0	3.97E-
R32	50	1000	0	0.5	0.1	0.0101	75.69	0.76	1.32E-2	1.19E-	7.33E-	0	7.33E-
R33	20	100	0	0.01	0.5	0.0055	137.72	0.75	7.26E-3	6.57E-	4.27E-	0	4.27E-
R34	50	200	500	0.2	0.5	0.0071	112.97	0.80	8.85E-3	7.97E-	3.47E-	0	3.47E-
R35	200	500	0	0.1	0.5	0.0015	605.81	0.90	1.65E-3	1.50E-	1.38E-	0	1.38E-
R36	100	1000	0	0.05	0.5	0.001	829.67	0.83	1.21E-3	1.08E-	3.50E-	0	3.50E-
R37	20	500	0	0.2	0.05	0.0101	52.07	0.52	1.92E-2	2.09E-	4.32E-	0	4.32E-
R38	50	500	0	0.2	0.2	0.0044	160.17	0.70	6.24E-3	5.76E-	5.44E-	0	5.44E-
R39	50	100	200	0.5	0.2	0.0135	64.02	0.86	1.56E-2	1.41E-	2.88E-	0	2.88E-
R40	50	200	0	0.02	0.2	0.0014	550.76	0.77	1.82E-3	1.64E-	9.52E-	0	9.52E-
R41	50	500	0	0.1	0.2	0.0024	289.08	0.69	3.46E-3	3.21E-	3.24E-	0	3.24E-
R42	100	1000	0	0.02	0.2	0.0004	2045.10	0.81	4.89E-4	4.40E-	1.62E-	0	1.62E-
R43	20	500	0	0.1	0.5	0.006	87.94	0.52	1.14E-2	1.23E-	2.54E-	0	2.54E-
R44	50	100	1000	0.1	0.01	0.00299	244.30	0.73	4.09E-3	3.73E-	2.97E-	0	2.97E-
R45	50	100	0	0.05	0.1	0.002	355.96	0.71	2.81E-3	2.58E-	2.33E-	0	2.33E-
R46	100	200	0	0.2	0.1	0.0025	320.21	0.80	3.12E-3	2.81E-	1.24E-	0	1.24E-
R47	100	500	0	0.01	0.1	0.0003	2775.99	0.83	3.60E-4	3.24E-	1.01E-	0	1.01E-
R48	20	1000	0	0.1	0.1	0.0051	96.52	0.49	1.04E-2	1.18E-	2.67E-	0	2.67E-
R49	20	200	500	0.05	0.02	0.00446	131.62	0.58	7.60E-3	7.70E-	1.30E-	0	1.30E-
R50	20	50	0	0.05	0.5	0.0125	56.66	0.70	1.76E-2	1.63E-	1.50E-	0	1.50E-
R51	100	200	0	0.05	0.2	0.0015	547.75	0.82	1.83E-3	1.64E-	5.81E-	0	5.81E-
R52	20	200	0	0.02	0.05	0.00125	400.53	0.50	2.50E-3	2.81E-	6.23E-	0	6.23E-
R53	100	500	0	0.1	0.05	0.0011	702.73	0.77	1.42E-3	1.28E-	7.33E-	0	7.33E-
R54	20	100	200	0.02	0.5	0.0084	94.28	0.79	1.06E-2	9.55E-	4.59E-	0	4.59E-
R55	50	1000	0	0.02	0.5	0.0009	839.85	0.75	1.19E-3	1.08E-	7.10E-	0	7.10E-
R56	20	100	0	0.2	0.02	0.0102	51.70	0.52	1.93E-2	2.10E-	4.32E-	0	4.32E-
R57	200	1000	0	0.02	0.2	0.0003	2967.32	0.89	3.37E-4	3.03E-	4.06E-	0	4.06E-
R58	20	200	0	0.01	0.05	0.00075	701.21	0.52	1.43E-3	1.55E-	3.21E-	0	3.21E-
R59	50	500	1000	0.01	0.1	0.00129	662.25	0.85	1.51E-3	1.36E-	3.21E-	0	3.21E-
R60	20	50	0	0.05	0.01	0.0027	178.65	0.48	5.60E-3	6.45E-	1.50E-	0	1.50E-

Combined sources

	$1/h$	L	D	p_{burst}	p_{rv}	m	N_c	ρ_{hom}	ε_u	ε_m	k	k^*	v^*	σ^{**}	v^{**}
D1	5	20	500	0.2	0.0312	0.0012	404.6	0.506	0.56	0.85	2.47E-	2.54E-	6.04E-	5.00E-	5.02E-
D2	5	20	100	0.1	0.25	0.005	113.3	0.567	0.49	0.00	8.83E-	7.94E-	1.66E-	3.83E-	-6.72E-
D3	5	20	100	0.1	0.0625	0.0012	433.1	0.541	0.51	0.16	2.31E-	2.08E-	4.86E-	9.78E-	7.14E-
D4	5	20	400	0.025	0.25	0.0012	431.9	0.540	0.52	0.07	2.32E-	2.08E-	4.90E-	1.03E-	3.07E-
D5	10	20	500	0.4	0.0312	0.0012	413.7	0.517	0.67	0.97	2.42E-	2.61E-	5.63E-	2.31E-	5.41E-
D6	10	20	100	0.2	0.25	0.005	126.2	0.631	0.52	0.30	7.92E-	7.13E-	1.08E-	2.48E-	2.97E-
D7	10	20	100	0.2	0.0625	0.0012	481.2	0.602	0.55	0.45	2.08E-	1.87E-	3.30E-	6.25E-	1.42E-
D8	10	20	400	0.05	0.25	0.0012	486.6	0.608	0.54	0.15	2.05E-	1.85E-	3.15E-	7.46E-	4.45E-
D9	15	20	500	0.6	0.0312	0.0012	413.7	0.517	0.79	0.97	2.42E-	2.61E-	5.63E-	2.27E-	5.42E-
D10	15	20	100	0.3	0.25	0.005	134.0	0.670	0.55	0.49	7.46E-	6.71E-	8.11E-	1.80E-	3.77E-
D11	15	20	100	0.3	0.0625	0.0012	508.6	0.636	0.59	0.70	1.97E-	1.79E-	2.61E-	4.04E-	1.78E-
D12	15	20	400	0.075	0.25	0.0012	517.0	0.646	0.57	0.26	1.93E-	1.74E-	2.42E-	5.97E-	5.74E-
D13	5	20	160	0.625	0.01	0.0012	192.2	0.240	0.87	0.95	5.20E-	8.24E-	3.00E-	9.77E-	2.82E-
D14	5	20	320	0.3125	0.02	0.0012	330.3	0.413	0.67	0.93	3.03E-	3.65E-	1.04E-	4.98E-	9.61E-
D15	5	20	400	0.25	0.025	0.0012	369.8	0.462	0.61	0.89	2.70E-	2.98E-	7.81E-	5.06E-	6.87E-
D16	5	20	640	0.1562	0.04	0.0012	423.8	0.530	0.53	0.58	2.36E-	2.16E-	5.21E-	7.41E-	2.89E-
D17	5	20	800	0.125	0.05	0.0012	431.9	0.540	0.52	0.31	2.31E-	2.08E-	4.90E-	8.93E-	1.45E-
D18	5	20	160	0.0625	0.1	0.0012	432.5	0.541	0.51	0.06	2.31E-	2.08E-	4.88E-	1.03E-	2.57E-
D19	10	20	320	0.625	0.02	0.0012	330.3	0.413	0.83	0.93	3.03E-	3.65E-	1.04E-	4.98E-	9.61E-
D20	10	20	400	0.5	0.025	0.0012	373.0	0.466	0.75	0.92	2.68E-	3.00E-	7.64E-	4.18E-	6.98E-
D21	10	20	640	0.3125	0.04	0.0012	449.7	0.562	0.61	0.84	2.22E-	2.16E-	4.26E-	4.12E-	3.50E-
D22	10	20	800	0.25	0.05	0.0012	475.0	0.594	0.56	0.70	2.11E-	1.94E-	3.47E-	4.81E-	2.37E-
D23	10	20	160	0.125	0.1	0.0012	485.3	0.607	0.54	0.22	2.06E-	1.85E-	3.19E-	7.21E-	6.67E-
D24	15	20	320	0.9375	0.02	0.0012	330.3	0.413	0.97	0.93	3.03E-	3.65E-	1.04E-	4.98E-	9.61E-
D25	15	20	400	0.75	0.025	0.0012	373.0	0.466	0.88	0.92	2.68E-	3.00E-	7.64E-	4.18E-	6.98E-
D26	15	20	640	0.4687	0.04	0.0012	456.1	0.570	0.70	0.91	2.19E-	2.18E-	4.05E-	3.04E-	3.63E-
D27	15	20	800	0.375	0.05	0.0012	486.0	0.608	0.63	0.82	2.06E-	1.94E-	3.17E-	3.55E-	2.56E-
D28	15	20	160	0.1875	0.1	0.0012	518.6	0.648	0.56	0.34	1.93E-	1.74E-	2.39E-	5.60E-	7.57E-

APPENDIX B. SIMULATION PROGRAM

All simulation material presented in this study has been attained by a dedicated simulation program. The simulation program consists of three parts:

- main program including user interface (2879 rows Pascal code);
- simulation unit (2042 rows);
- auxiliary units, including CAC formulae (2244 rows).

The network structure used in the simulation program is presented in Figure B.1. Only one ATM node with one switching stage with output buffers has been implemented in the program. However, because the traffic generation part of the program is completely separated from the part that contains switching and queuing procedures, more complicated systems are possible to implement. The size of the node is restricted to the following values:

- the number of incoming links: $2 \le M_{in} \le 64$;
- the number of outgoing links: $2 \le M_{out} \le 64$;
- the buffer size in cells: $2 \le K \le 400$.

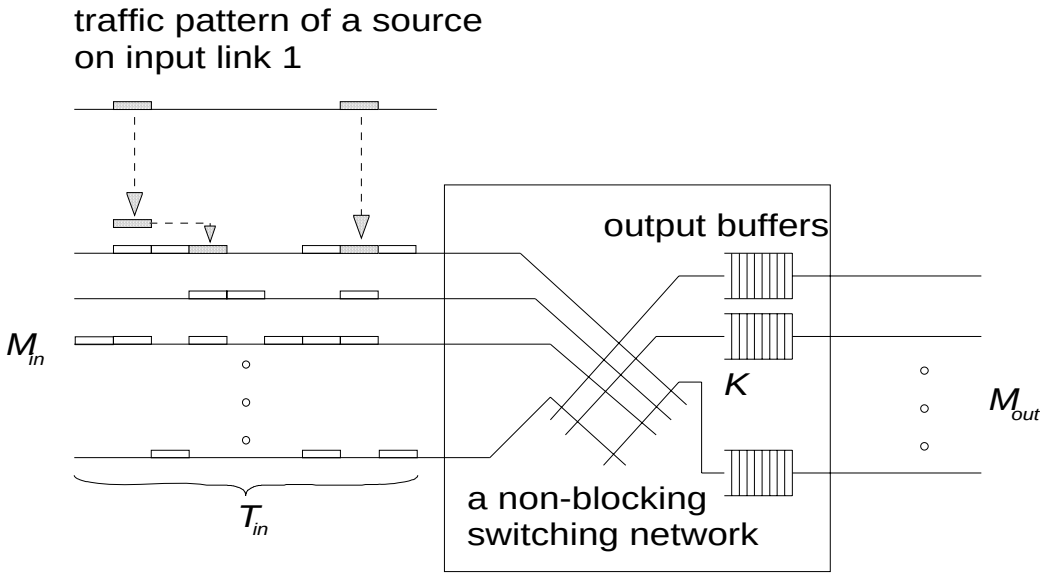

Figure B.1. Network structure used in simulation program.

Each traffic source either randomly chooses an input and an output link, or the connections are distributed to different links is as evenly as possible. All simulation results presented in this study are based on even distribution. The original traffic pattern of a source is first generated on the basis of traffic parameters. All traffic pattern cells are placed into the traffic pattern of the corresponding input link. If the desired time-slot is already reserved, the next free time-slot is selected. If all time-slots are reserved, the cell will be rejected.

The suitable number of time-slots (T_{in}) of a generation period varies typically from 1000 to 16000 depending on the type of source. Because of restricted memory available for

the simulation program the allowed length of traffic generation period depends on the number of links and buffer size (see Table B.1).

Table B.1. The maximum allowed traffic generation period

M_{in}	M_{out}	K	Allowed T_{in}
64	64	400	2000
64	32	100	3000
32	**16**	**100**	**8000**
16	8	100	16000
8	4	100	32000

In all cases presented in this study the M_{in}/M_{out} ratio is 2 and buffer size is 100 cells. The number of incoming links is 32 except for those cases whose the required T_{in} is larger than 8000, then M_{in} is 16.

The following traffic processes have been implemented:

- Cell scale:
  - deterministic process;
  - Poisson process.
- Arrival processes on burst scale:
  - constant interarrival time;
  - geometrical interarrival time distribution.
- Burst size:
  - constant;
  - geometric distribution;
  - even distribution between minimum and maximum values;
  - a truncated geometrical distribution with minimum and maximum values.
- Rate-variation scale:
  - three different average bit rate levels;
    the same burst size distribution at each level.

The implementation of simulation consists of two modes:

- periodic mode in which all traffic sources are supposed to be periodic;
- continuous mode in which at least one of the sources is not periodic.

There is a fundamental difference between these two modes. In the periodic mode we have a periodic traffic process and therefore there is no need to simulate more than two periods for each traffic combination. The first period starts with an empty buffer and by means of that period we can determine the state of buffers at the beginning of each period. All performance calculations are made during the second period. The required traffic generation period T_{in} is the largest period of sources under study.

In continuous mode the simulation process is continuous and the state of buffers remains unchanged between traffic generation periods. This principle is suitable for Markov sources. Usually, the largest possible value for T_{in} is recommendable because the boundary between traffic generation periods may disturb the traffic process. A mixing of periodic and Markov sources is possible but it leads to many difficulties because of the different nature of the traffic processes.

The results of this study are based chiefly on periodic source types, while sources of other types are used only as material for comparison and validation of mathematical formulae.

APPENDIX C. THE ACCURACY OF DETERMINING SOURCE PARAMETERS

Figure C.1 illustrates the determining process of scale factors. The starting point is that we have simulation results for two numbers of sources, $N_{i,1}$ and $N_{i,2}$, and we know the cell loss probability with a certain accuracy (see Section 3.6.2). Because usually the difference between $N_{i,1}$ and $N_{i,2}$ is small, we can presumably use a linear approximation for the dependency between N_i and $\ln(P_{loss})$ and simple discrete approximations for P_{loss} distribution (three values in Figure C.1). By connecting every possible pair we obtain a discrete distribution for the allowed number of sources i.

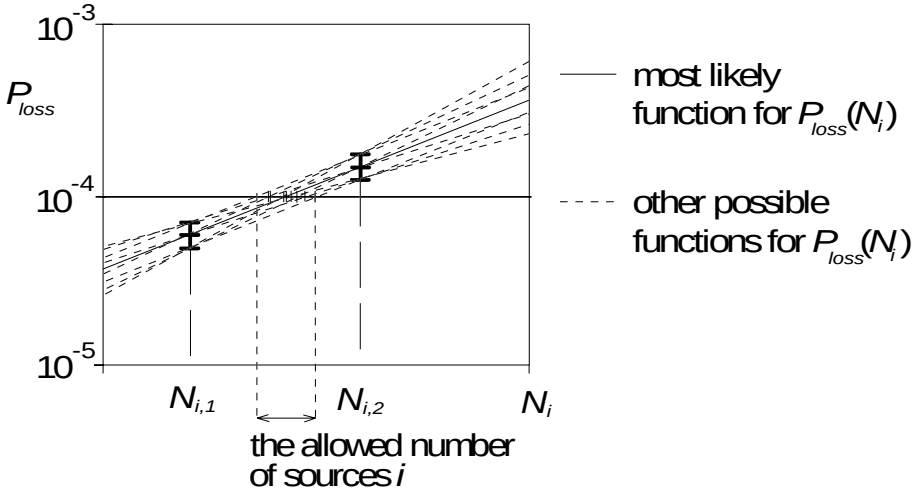

Figure C.1. The accuracy of determination of the allowed number of sources.

This is usually a feasible approach. However, the situation is more difficult if we have chosen the number of sources in a way that the two distributions for cell loss probability ($N_{i,3}$ and $N_{i,4}$ in Figure C.2) come close to each other. A direct approach may lead to such a curious consequence as point *A* in Figure C.2. The reason to this phenomenon is the assumption that all possible functions, even those with a negative slope, have the same a priori probability. Evidently, we do have a prior knowledge of the dependency between P_{loss} and N_i. Furthermore, we can suppose that the real accuracy in Figure C.2 is usually better than that in Figure C.1 provided that the traffic processes are similar.

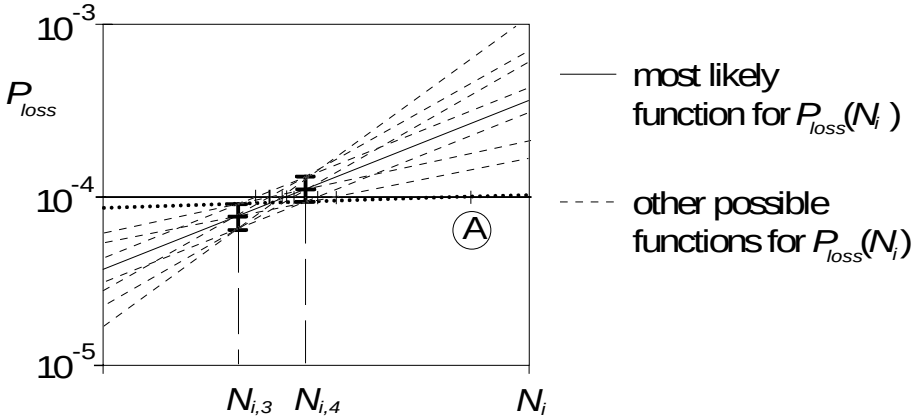

Figure C.2. A difficult case for determining the accuracy of allowed number of sources.