ELSEVIER

Computer Networks 35 (2001) 237–261

COMPUTER NETWORKS

www.elsevier.com/locate/comnet

Correctness and performance of the ATM ABR rate control scheme

David Lee [a], K.K. Ramakrishnan [b,*], W. Melody Moh [c]

[a] Bell Labs Research, Beijing, People's Republic of China
[b] AT&T Labs-Research, Room A155, 180 Park Avenue, Florham Park, NJ 07933, USA
[c] San Jose State University, San Jose, CA, USA

Received 23 December 1998; received in revised form 10 January 2000; accepted 26 May 2000

Responsible Editor: J.W. Roberts

Abstract

We study both correctness and performance of the source/destination protocols of the available bit rate (ABR) service in asynchronous transfer mode (ATM) networks. Although the basic protocol for rate-based congestion management is relatively simple, the protocol specification has to cope with several "real-world" cases such as failures and delayed/lost feedback which introduce complexity. Rigorous proof of the correct functioning of the protocol based on a formal specification is necessary. We use a formal model to show that the ABR source/destination protocol is free of livelocks, so that under all conditions both resource management (RM) and data cells will be transmitted. Furthermore, if there are data cells available, then the ABR protocol is deadlock free; otherwise, the system goes to a desirable sleep state waiting for data cells, as long as certain parameters are set appropriately at connection setup. We also show that the network options of explicit forward congestion indication (EFCI) and explicit rate (ER) interoperate correctly.

In addition to ensuring the correct functioning of the protocol, it is essential that pathological situations do not result in very poor performance, which we view as another form of "incorrect operation". We derive conditions that ensure that the source's allowed cell rate (ACR) is stable in the presence of delayed or lost feedback RM cells. We arrive at bounds on the number of consecutive RM cell losses tolerated while the ACR rate remains stable. We also provide an asymptotic estimate of ACR and the allowable RM cell loss probability to ensure that ACR is stable, statistically.

The ABR protocol contributes to the feedback delay in two ways: the source delay of sending out the probe forward RM (FRM) cells and the destination delay of turning around the backward RM (BRM) cell. We provide a worst-case analysis of the delay in turning around RM cells at the destination station and the worst-case inter-departure time of FRM cells from the source.

1. Introduction

The available bit rate (ABR) service class for asynchronous transfer mode (ATM) networks uses a rate control scheme to manage congestion. Sources adjust their rates such that the aggregate

* Corresponding author. Tel.: +1-973-360-8766; fax: +1-973-360-8871.

E-mail address: kkrama@research.att.com (K.K. Ramakrishnan).

PII: S 1 3 8 9 - 1 2 8 6 (0 0) 0 0 1 7 0 - 5

load on the network does not exceed the capacity of the network [1]. The protocol specification for the source and destination behavior to achieve overall congestion avoidance and control is specified using a relatively informal specification described in [1]. Many aspects of the protocol have been designed based on extensive performance analysis through simulation. Simulation has been used especially to determine the correct parameter settings. There is little work in the literature that rigorously analyzes the performance and examines the correctness of the ABR protocol.

Correctness relates to whether the specified protocol is free of any livelocks, deadlocks, or other undesirable properties. It is difficult, if not impossible, to prove the correctness based on an informal specification or simulation. We use a formal specification of the ABR protocol in an EFSM model [13] to show that it is livelock free in terms of transmitting data cells and resource management (RM) cells, which act as source probes of the network's congested state. Furthermore, when there are data cells available, then the ABR protocol is also deadlock free; otherwise, when there are no data cells to send, the system goes to a desirable sleep state waiting for data cells, given that certain system parameters are appropriately set at the connection setup time.

In addition, the ABR congestion management specification accommodates at least two different modes in which switches in the network may operate. The source/destination policies (the main focus of the specification) are designed to smoothly operate with any intermingling of the two types of switches in the network – these are the explicit forward congestion indication (EFCI) and explicit rate (ER) switches. EFCI switches use a single bit to communicate congestion [15], when a queue threshold is exceeded. ER switches, on the other hand, compute a max–min fair rate [4,5] and communicate this rate to the sources. The source and destination use a common ABR protocol to interface to networks with both types of switches. We show that the protocol correctly interoperates with both types of switches, using the formal specification.

Rate-based flow control mechanisms also need some form of protection against failures. For example, when the feedback from the network is not provided in a timely manner or is lost, it is desirable that the source reduce the rate so that the network is not overloaded by the source transmitting at an incorrectly high rate. ATM's ABR service can allow sources to start at a reasonably high initial rate so that higher layer protocols and applications such as remote procedure calls (RPC) may transmit a short burst without a start-up delay of a full round-trip time for feedback from the network. To avoid using this potentially high rate for too long, and thus exceeding the buffering in the network, the source rules specify a proportionate reduction in the rate in the absence of feedback from the network. This rate reduction may also be triggered in certain cases when there is a mismatch between the current transmission rate of RM cells by the source and the rate at which RM cells return from the network. This may especially be true subsequent to a source rate increase. If the parameters are set inappropriately, this may result in a net rate reduction even with feedback from the network to allow a rate increase. Some of this has been analyzed in [8] through simulation.

In this paper, we provide a quantitative analysis of worst-case conditions under which a stable source transmission rate is maintained in spite of the rule of rate reduction in the absence of timely feedback. It is particularly important to examine this in the presence of different types of RM cell loss and delay, which is one of our main contributions.

In conjunction with the problem of maintaining a stable rate, there is the need to ensure that the feedback from the network is timely. While some of the feedback delay contributed by the network is outside our framework, we quantify in this work, the contribution of the source/destination policies. Specifically, we estimate the turn around time for backward RM (BRM) cells and the inter-departure time interval for forward RM (FRM) cells. This estimation provides useful information for appropriate parameter setting as well as for understanding the relationship between the source rate and the network feedback delay.

In Section 2, we provide a brief overview of ATM's ABR service [1]. Section 3 arrives at bounds on the BRM cell turn around time at the remote destination node based on the source/destination rules for transmission of FRM, BRM, and data cells. We also examine the bounds on the time interval for transmission of FRM cells from a source, which is the carrier of the feedback information from the network back to the source. In Section 4, we prove the correctness of ABR; that it is free of deadlocks and livelocks. We also show that the EFCI and ER schemes interoperate correctly; there is no unexpected performance degradation due to their interoperation. Our main contribution is in Section 5 where we analyze the issues related to maintaining a stable rate in the context of the ABR Specification's Source Rule 6 which triggers a rate reduction in the absence of timely feedback, both when there is no loss of RM cells and when there is loss of RM cells.

The source/destination rules, the parameters and acronyms used in the paper are from [1]. For the formal specification, see [13].

2. Brief overview of available bit rate service

Rate-based congestion management has been proposed for feedback control of ATM networks [2]. The focus of this congestion management work has primarily been on ATM's ABR service for bursty data applications, where there is no clear specification of the source's characteristics [1]. These applications desire a low loss rate. There is the possibility that the demands of the sources exceed the resource capacity. Although no assurances are made of maintaining low delay or jitter, the feedback control algorithm attempts to maintain small queues and *feasible* transmission rates for the individual sources (i.e., the aggregate transmission rate of all the currently active sources utilizing a link does not exceed the link capacity). The ABR service also supports the notion of a minimum bandwidth allocation for a source.

Two components of the control algorithm are identified:

(i) the behavior of the source and destination end systems, and

(ii) the behavior of the network elements (switches).

Each source of a virtual circuit (VC) periodically transmits a special RM cell to probe the state of the network. RM cells are periodically transmitted, once every *Nrm* data cells (e.g., $Nrm = 32$), so that the overhead for carrying the probe cells is bounded, while still having a responsive control scheme. Each switch identifies and conveys its state of congestion as well as additional rate information to the source end-system in the RM cell. The source algorithm responds to the feedback information by adjusting the rate of transmission in accordance with a specified policy.

With the EFCI option specified in [1], the congestion information is a single bit that switches set in the data cells flowing in the forward direction when they determine they are congested. Destinations then feedback this information by turning around the RM cells as a BRM cell. The source responds to the feedback by adjusting ACR. It increases ACR additively when the feedback indicates that the network is uncongested, or decreases ACR multiplicatively when the network is congested.

Fig. 1 shows the operation of the ER scheme by example. With the ER option, a source specifies a "demand" or desired transmit rate in each transmitted RM cell (in addition to the currently allowed cell rate), in the *ER-field*. When an RM cell is transmitted, the ER-field is set to max(DEMAND, ACR). In the example, the ER-field is marked at 5000 cells/s, with a current cell rate (CCR) of 1000 cells/s. Switches compute the rate they may allocate to each VC, and overwrite this allocated rate in the ER-field if the computed rate is lower than what was in the received RM cell. In the example, the first switch marks the ER-field down to 4000 cells/s, representing the available capacity for *this VC* on the output link. As the RM cell progresses from the source to destination, the ER-field value reflects the smallest rate allocated by any of the switches in the path for the VC. In the example, the final value of the ER-field is 3000 cells/s, reflecting the available capacity of the bottleneck (which is the link between the second and third switches). On reaching its destination, the RM cell is returned to the source, which now sets its transmit rate based on the ER-field value in the returned BRM cell. The goal of the explicit rate-based

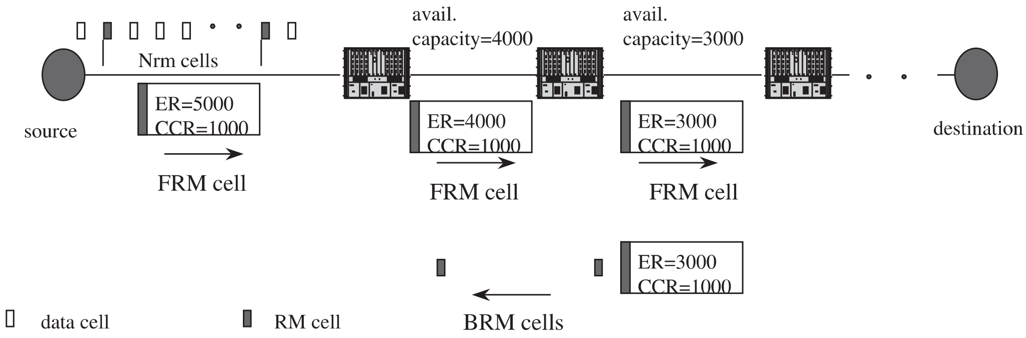

Fig. 1. Explicit rate scheme operation.

feedback control algorithm is to respond to incipient congestion, and to allocate rates to the competing sources in a max–min fair manner, while ensuring that the capacity of the network is not exceeded. There are several switch algorithms proposed for computing the rate to be allocated to a VC [4,7,9,17], with most of them attempting to achieve max–min fairness or some approximation thereof.

The source maintains a currently allowed rate, ACR, which is the rate at which queued cells are transmitted out of the source network interface. When an RM cell returns with an allocated rate ER, the source's allowed rate is changed using a set of rules specified in [1]. When the ACR is greater than or equal to the ER value returned in the RM cell, ACR is reduced to the ER value returned (subject to the minimum cell rate (MCR) constraint). However, when the allocated rate ER returned is higher than the current ACR, it increases in additive steps of RIF*PCR. RIF is an increase-factor that is a negotiated parameter, and PCR is the peak cell rate for the connection. ACR always remains above MCR. A large RIF results in convergence to the returned ER quickly, but with the potential for some transient overload on the network. To keep queues small, RIF may be chosen to be small. More details may be found in [1].

3. RM and data cell transmission delays

We first study the turn-around time of BRM cells at the receiving end-station. When an FRM cell with the current network state information is received by the destination protocol machine, it is converted into a turned-around BRM cell and the information is passed to the source protocol machine that adjusts the rate accordingly.

Informally, the source machine sends an FRM cell after $Nrm - 1$ data or BRM cells or based on the time since the last FRM cell was sent. These FRM cells arrive along with the data stream (as they are sent on the same VC) at the destination machine of the remote station. After incurring queueing delays to have the hardware (adapter) process arriving cells (which we do not consider here), the FRM cell is handed to destination machine. Our estimation of the turn-around delay starts from the point when the FRM cell is handed to the destination machine.

The destination machine follows its rules to take the contents of the received FRM cell to create information to be sent in the turned-around BRM cell. Specifically, the DIR bit in the FRM cell is changed from "forward" to "backward" and BN, CI, NI, QL and SN fields are set if necessary. This process of updating the fields of an FRM cell can usually be done in hardware and the processing time is negligible.

The relationship of the various state machines (source, destination and the scheduler machines) is shown in Fig. 2.

The rules for the destination are that the information received in the FRM cell may be used to rewrite any turned-around BRM cells already queued for transmission [1]. The queue is between the destination protocol machine and the source

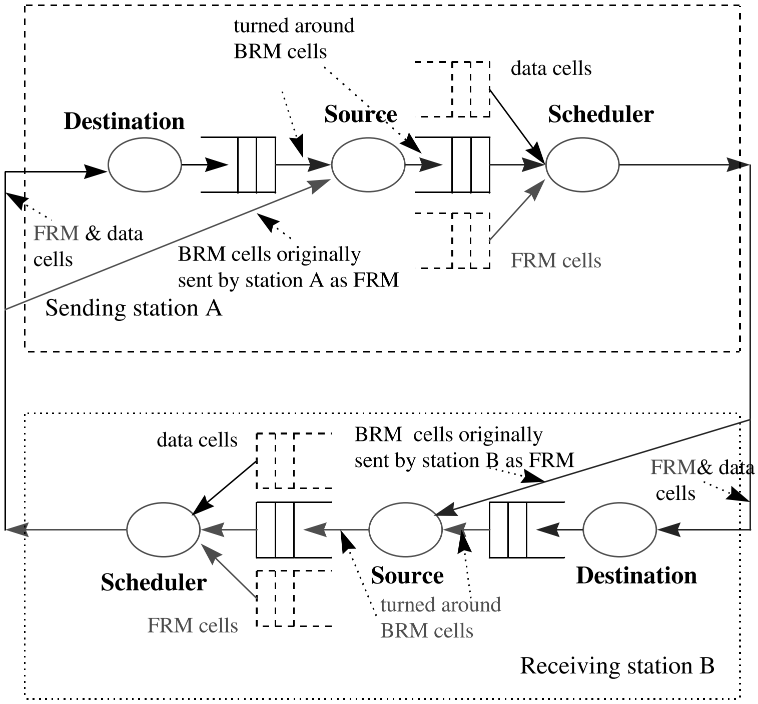

Fig. 2. Organization of state machines at sender and receiver of ABR loop.

protocol machine. For convenience, we call this *BRM re-writing*. Another alternative is to drop all queued turned-around BRM cells and queue this new BRM cell for transmission. We call this *BRM dropping*. It turns out that the turn-around time is invariant with respect to the two different rules.

The turned-around BRM cell is then handed to the source protocol machine, which adjusts the ACR based on the information in the BRM cell and then queues it to the scheduler to be sent back to the sending end-station. The scheduler follows specific rules to determine when to send this BRM cell; it is based on how many data cells have been transmitted and how much time has elapsed since the last FRM cell was sent and also on whether a BRM cell has been transmitted since the last FRM cell was sent. The turn-around delay interval ends at the point when the BRM cell is transmitted.

Our estimate of the turn-around time τ, from the moment an FRM cell is received at the destination machine to the time a corresponding BRM cell is sent to the network, is derived informally based on the original specification [1]. We chose to do so since it provides more insight. On the other hand, a formal proof based on the formal specification can also be obtained.

We estimate the delay of the BRM cell in the scheduler with the following practically reasonable assumptions. The problem becomes trivial if any of these assumptions do not hold.

Assumption 3.1.
(1) $Mrm \geqslant 1$ and $Nrm \geqslant 2$;
(2) $Mrm < Nrm$ and
(3) $Trm > 1/\mathrm{MCR}$.

The following lemma relates to the ordering of the transmission of RM cells. It is crucial for the estimation of the BRM cell turned-around time.

Lemma 3.1. *Suppose that BRM cells are waiting for transmission and that an FRM cell has just been transmitted to the network by the scheduler. Then the next cell to be transmitted by the scheduler is a BRM cell.*

Proof. From Assumption 3.1(3), the time elapsed since the FRM cell transmission is less than *Trm*. On the other hand, we have not transmitted any data cells subsequent to the FRM cell. Therefore, from Source Rule (3) [1], which determines the cell transmission ordering, only Rule (3)(b)(i) holds and hence the next cell to be transmitted is a BRM cell.

For a formal proof, note that in the scheduler machine, (see [13]), only the transition T_6 relating to the selection of a BRM cell as the next cell to be transmitted can be triggered. This causes a BRM cell to be transmitted. □

Implementations may be somewhat lax in interpreting the order in which BRM cells and queued data cells may be transmitted. However, the priority, according to the specification is given to the waiting BRM cell.

Lemma 3.2. *With ACR as the current cell rate, it takes time* τ *to turn around an FRM cell, as a BRM cell, and then to send it to the network.* τ *is specified as*

$$\begin{aligned}\frac{1}{\mathrm{PCR}} &\leqslant \frac{1}{\mathrm{ACR}} \\ &\leqslant \tau \leqslant \min\left\{\max\left\{Trm, \frac{Mrm}{\mathrm{ACR}}\right\} + \frac{1}{\mathrm{ACR}}, \frac{Nrm}{\mathrm{ACR}}\right\} \\ &\leqslant \min\left\{\max\left\{Trm, \frac{Mrm}{\mathrm{MCR}}\right\} + \frac{1}{\mathrm{MCR}}, \frac{Nrm}{\mathrm{MCR}}\right\}.\end{aligned} \tag{3.1}$$

Furthermore, all the bounds are tight.

Proof. Note that BRM cells are transmitted according to Source Rule (3)(b).

The lower bound in (3.1) is obvious; suppose that an FRM has been just sent to the network. Then by Lemma 3.1, a BRM cell is sent immediately in time $1/\mathrm{ACR}$ according to the rate constraint. Since the rate PCR is attainable by ACR, the lower bound is tight.

We now consider the upper bound. For clarity, we denote the moment the BRM cell is queued to scheduler by t_0 . Again from Lemma 3.1, the upper bound is determined by the time for a maximum number of data cells sent before sending an FRM cell and by the maximum amount of time *Trm* that may elapse before an FRM cell has to be sent. The bound can be achieved if and only if the last RM cell sent before time instant t_0 is a BRM cell. There are two cases to consider:

Case 1. If we ignore the time constraint *Trm*, then neither Source Rule (3)(a) nor (3)(b) [1] holds until $Nrm - 1$ cells have been sent since the last FRM cell was sent. To maximize the number of data cells sent after t_0, we consider that the two cells sent before t_0 are an FRM cell followed by a BRM cell. Therefore, from t_0, $Nrm - 2$ data cells are sent followed by an FRM cell and then the BRM cell. The total number of cells sent after t_0 (including this BRM cell) is *Nrm*, in time Nrm/ACR, one of the terms in Eq. (3.1).

Case 2. The timer may expire before Nrm/ACR cells are sent. By Lemma 3.1, the longest possible time duration (after t_0 and before sending the BRM cell) occurs when the two cells sent right before t_0 are an FRM cell followed by a BRM cell. The first FRM should have been sent at time $t_0 - 1/\mathrm{ACR}$, and at time $(t_0 - 1/\mathrm{ACR}) + Trm$ another FRM cell will be sent. By Lemma 3.1, it is followed by the BRM cell. Thus, it takes time $1/\mathrm{ACR}$ to send the BRM cell, and, therefore, the total time duration from t_0 to the moment the BRM cell is sent is: $Trm + 1/\mathrm{ACR}$. On the other hand, in conjunction with the expiration of the timer, at least *Mrm* "in-rate" cells have to be sent before transmitting a BRM cell, which takes time $(Mrm + 1)/\mathrm{ACR}$, including the time to send this BRM cell.

To summarize, the BRM cell is sent whenever one of the above two cases happens first. The upper bound in 3.1 is proved. Since ACR is bounded below by MCR and the bound is tight, the upper bound on the right side of 3.1 is also tight. □

RM cells act as the probes into the network to determine the capacity available for a VC. Too many probes into the network introduce overhead, and we need to ensure that this is not excessive. However, sending probes too infrequently results in exposure of the network to congestion and delay in the source's reaction to network state. Here, we attempt to understand the bounds on the time between transmitting two consecutive FRM cells. We assume that there are always data or BRM cells available for transmission. Otherwise, the inter-departure times for FRM cells is $Trm + 1/\mathrm{ACR}$, where an FRM cell is transmitted on the firing of a timer with a value of *Trm*. Examining Rule (3), we can show that:

Proposition 3.1. *If there are data or BRM cells available for in-rate transmission, then the time interval between two FRM cell transmissions is*

$$\tau = \min\left\{ \max\left\{ Trm, \frac{Mrm}{\mathrm{ACR}} \right\} + \frac{1}{\mathrm{ACR}}, \frac{Nrm}{\mathrm{ACR}} \right\}. \tag{3.2}$$

The source rate may change whenever FRM cells are sent. We now discuss the dependency of the FRM cell transmissions on the rate ACR.

Proposition 3.2. *Assume data or BRM cells are available for transmission. Then*

(1) *For* $\mathrm{ACR} \geqslant Nrm/Trm$, *each FRM cell is transmitted after sending* $(Nrm - 1)$ *data or BRM cells*;
(2) *For* $Mrm/Trm \leqslant \mathrm{ACR} < Nrm/Trm$, *FRM cells are transmitted every Trm time interval and*
(3) *For* $\mathrm{ACR} < Mrm/Trm$, *each FRM cell is transmitted after sending Mrm data or BRM cells.*

Proof. Note that by Assumption 3.1, $Mrm < Nrm$.
(1) $\mathrm{ACR} \geqslant Nrm/Trm$. In this case, $Trm \geqslant Nrm/\mathrm{ACR}$, which is the time to transmit *Nrm* cells. Source Rule (3)(a)(ii) takes effect before (3)(a)(i), therefore an FRM cell is transmitted after transmitting $(Nrm - 1)$ data or BRM cells. The timer based on *Trm* does not trigger.
(2) $Mrm/Trm \leqslant \mathrm{ACR} < Nrm/Trm$. In this case, $Mrm/\mathrm{ACR} \leqslant Trm < Nrm/\mathrm{ACR}$, thus Source Rule (3)(a)(i) takes effect when *Trm* time has elapsed, and this implies that an FRM cell is transmitted every *Trm* time interval.
(3) $\mathrm{ACR} < Mrm/Trm$. In this case, $Trm < Mrm/\mathrm{ACR} < Nrm/\mathrm{ACR}$, and Source Rule (3)(a)(i) takes effect after transmitting *Mrm* BRM or data cells, which takes longer than *Trm*. Hence each FRM cell is transmitted after *Mrm* data or BRM cells. □

If turned-around BRM cells are waiting for transmission, then by Source Rule (3)(b) a BRM cell is transmitted immediately after an FRM cell is sent. Therefore, the transmission interval of BRM cells is the same as that of FRM cells. From Proposition 3.2, we have:

Corollary 3.1. *Assume that data and turned-around BRM cells are available for transmission. Then the time intervals of transmission of FRM and BRM cells are the same*:

(1) *For* $\mathrm{ACR} \geqslant Nrm/Trm$: $(Nrm - 1)/\mathrm{ACR}$;
(2) *For* $Mrm/Trm \leqslant \mathrm{ACR} < Nrm/Trm$: *Trm and*
(3) *For* $\mathrm{ACR} < Mrm/Trm$: Mrm/ACR.

Fig. 3 evaluates the inter-departure times for FRM cells as the source rate ACR increases, based on the analysis for Proposition 3.1 and the corresponding Eq. (3.2) for τ. We choose the parameters as follows: $Mrm = 2$, $Trm = 100$ ms, $Nrm = 32$. We observe that for very low rate sources, when $\mathrm{ACR} < Mrm/Trm$, the inter-departure time for FRM cells can be large, ranging from near 300 ms, down to the value of *Trm*. However, this is not too serious, as the rate of the source is very small at this point (less than 10 cells/s). As ACR increases further, the inter-departure time for FRM cells is then the constant time, *Trm*, of 100 ms, until $\mathrm{ACR} = Nrm/Trm$, obvious from understanding the protocol. The interesting part of Fig. 3(A) is the intermediate point when ACR is just above *Nrm*/*Trm*, from about 0.3 cells/ms to say 1 cells/ms. At 1 cell/ms, this translates to about 384 Kbits/s of payload data. For relatively low speed links, such as a T_1 (peak rate of approximately 1.5 Mbits/s), the 1 cell/ms source constitutes a significant fraction of the link rate. During this time, the inter-departure time between FRM cells is almost 33 ms. For a VC with a short round-trip time (Local Area Network or Metropolitan Area Network), the feedback to the source may *not* occur sufficiently frequently. In these cases, the opportunity to send back information occurs only once every 33 ms. The intent of showing Fig. 3(A) is to point out this "intermediate range" where the interdeparture time of the FRM cells plays a significant part in the frequency of updating network state back to the source. This is particularly true

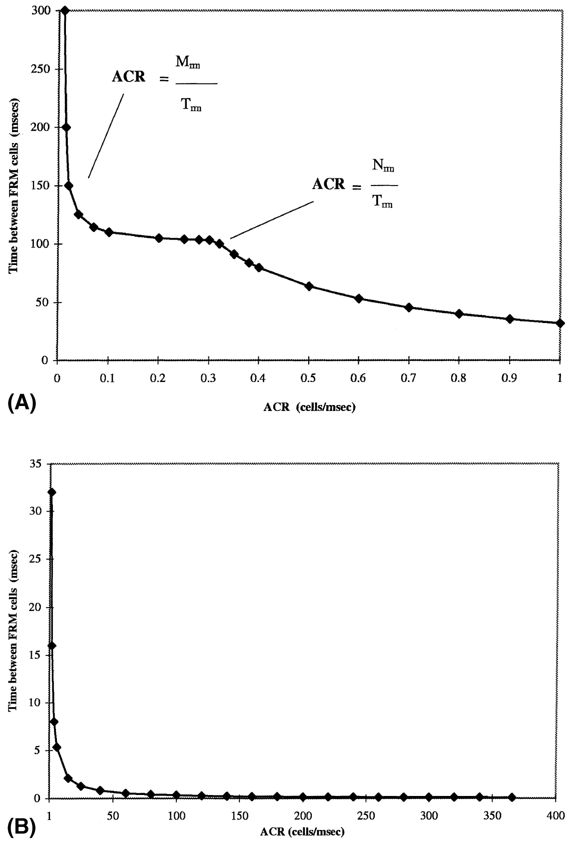

Fig. 3. (A) Variation of inter-departure time for forward RM cells with increasing ACR (low ACR). (B) Variation of inter-departure time for forward RM cells with increasing ACR (high ACR).

when there are a relatively small number of sources sharing a "slow" link such as a T_1. We need to strike a balance between overhead and timely feedback.

Fig. 3(B) shows that the inter-departure time for FRM cells then reduces with ACR as the rate goes up to the rate of 366 cells/ms (OC-3 rates). At these high rates, the primary contributor to latency in providing feedback to the source is the network round-trip time, which may include both propagation time and queueing delays for RM cells in the network. At these rates, overhead is the primary concern.

The results of this section, while primarily applicable to the ATM ABR service, are also applicable in the more general context of rate based control. The question of how frequently should a source probe the network (to obtain feedback) for its state has arisen in other contexts as well. For example, rate-based mechanisms proposed for the Internet [16] could use the insights obtained here to determine the feedback frequency.

4. Correctness

While a substantial amount of work has been done on the performance of the ABR rate control scheme, [5,7,17] there has been little attention to its correctness, especially with respect to unexpected system behaviors such as deadlock or livelock [6,11]. These system behaviors are not acceptable for the correct operation of the ABR rate control scheme. Their absence is a *necessary* condition for proper operation.

(1) *Deadlock*. An end-station enters an undesirable state, from which no further execution is possible, i.e., no source or destination rules are applicable for the system to proceed further. In such a state, data or RM cells could be waiting for transmission or processing while the system is halted.

(2) *Livelock*. In this situation, there are infinite execution sequences such that the system does not make any progress; by progress, we mean that a source is sending/processing data or RM cells or updating ACR.

We provide a formal analysis of the correctness of the ABR protocol and show that it has neither livelocks nor deadlocks if data cells are available. However, if there are no data cells waiting for transmission then anomalies may occur if we choose an improper set of system parameters, which we characterize in Section 4.2.

We focus on three specific issues for the ABR protocol: livelocks, deadlocks and the interoperation between the ER and EFCI schemes. Such an analysis is almost impossible with only the informal English specification. Our approach is based on the formal specification given in [13]. We present the main results and leave the proofs in Appendix A.

4.1. Livelocks

We first study livelocks for the ABR protocol with an assumption that data cells are available for transmission. We show that under such circumstances, data and BRM cells waiting for transmission will be eventually transmitted. We also show that a received FRM cell will be processed and turned around as a BRM cell. Then, the ABR protocol would be livelock free. Furthermore, we estimate the worst-case time delay to send data and RM cells.

Theorem 4.1. *The ABR source/destination protocol is livelock free if there are data cells available for transmission: both data cells and RM cells waiting for transmission will be sent, and received FRM cells will be processed and turned around eventually. Specifically*

(1) *FRM and turned-around BRM cells are sent in a time interval no more than*

$$\tau = \min\left\{ \max\left\{ \frac{M_{rm}}{\mathrm{MCR}}, Trm \right\} + \frac{1}{\mathrm{ACR}}, \frac{Nrm}{\mathrm{MCR}} \right\}. \tag{4.1}$$

(2) *Data cells waiting for transmission will be sent in time no more than*

$$\tau_{\mathrm{DATA}} = \frac{4}{\mathrm{MCR}}. \tag{4.2}$$

4.2. Deadlocks

With the assumption that there are data cells available for transmission, the ABR rate control scheme is deadlock free. This may be obtained as a simple corollary of Theorem 4.1. Recall that a deadlock occurs when an end-station enters an undesirable state when no execution is possible, i.e., no source or destination rules are applicable for the system to proceed further. By Theorem 4.1, waiting data cells will be transmitted. Therefore, we have the following.

Corollary 4.1. *The ABR source/destination protocol is deadlock free if there are data cells available for transmission.*

4.3. No data cells available

It must be pointed out that if there are no data cells waiting for transmission, anomalies may occur. We assume that two end-stations are transmitting data to each other over a VC. When there are no data cells to be transmitted we want both end-system's states to terminate after properly taking care of the data and RM cells that are in transmission or are being processed. The system should reach a state where no execution of further transitions is possible. This is sometimes called a "deadlock"; however, it is a *desirable* state in our case, since there is no need for cells to be transmitted further. We call this a *sleep* state, where the end-stations are waiting for the data cells from the other end-station or from the user. If, instead, given the above situation, the two end-stations keep sending RM cells to each other, then it is a waste of network resources. Such a state is neither a deadlock nor a livelock. We call it a *busy-wait* state, and is not a desirable state.

Theorem 4.2. *Suppose that there are no data cells available for transmission at either of the end-stations. Then the two end-stations on a VC will go to a sleep state eventually if*: (1) $Mrm \geqslant 2$ *and* $Nrm \geqslant 3$ *and* (2) *RM and data cells are not duplicated during processing or transmission. Otherwise, the system may enter a busy-wait state.*

It is clear from the analysis that the ABR protocol goes to a sleep state only when neither end-station on a VC has data to send. It should *not* be perceived as a protocol design fault. Instead, it is quite normal that no RM cells are sent since neither user on the two end-stations is sending data. On the other hand, if $Mrm = 1$ or $Nrm = 2$ then the two end-stations will send FRM and BRM cells in turn assuming there is no cell loss. The connection remains open, yet without any data cells being transmitted. It is a waste of network resources and such a busy-wait state is not acceptable. Therefore, when the connection is setup, one should set the system parameters so that $Mrm \geqslant 2$ and $Nrm \geqslant 3$.

4.4. Interoperability of EFCI vs ER schemes

There are two rate control schemes specified in the ABR protocol [1]: EFCI and ER. We are able to show that they interoperate correctly. The basis for showing the correct interoperation is proving that one and only one rate-change condition is satisfied. We have:

Theorem 4.3. *The EFCI and ER schemes interoperate as follows*:

(1) *If both EFCI and ER indicate congestion, then the final resulting rate is the minimum of the two rates obtained by rate reductions determined by the two schemes.*

(2) *If both EFCI and ER indicate no congestion, then the final rate increase is the minimum of the two increases obtained by applying the two schemes unless* NI = 1, *and in this case there is no rate increase.*

(3) *If there is a conflict*, (*one indicates a congestion and the other indicates no congestion*), *then the final outcome is the rate reduction obtained by the option indicating congestion.*

It is obvious that the interoperation of the two schemes always takes the lower rate from the two schemes. On the other hand, the interface between them will not degrade the performance by dictating a rate which is lower than both rates from the two schemes.

5. Rate analysis

The ABR source policy includes a mechanism that limits the use of ACR, in the absence of timely feedback from the network (Source Rule 6 [1]). When we have "steady-state" operation, where the rate does not change, and there are no RM cells lost, the source end-system should be receiving a BRM cell for every FRM sent. When a rate change occurs, we may be sending FRM cells at a different rate (based on the new ACR) than the returning rate of BRM cells. When there is no loss, the rate of returning BRM cells should match the rate of FRM cells sent approximately one round-trip time earlier. When a source starts up and is transmitting at an initial cell rate (ICR), the source runs the risk of overloading the network if this initial rate is too high. RM cells, however, return only one round-trip time (RTT) after start-up. Consequently, in the absence of feedback from the network, the source rules provide for a conservative correction of the rate that started from ICR, until the feedback arrives. The ABR policy has chosen to have sources start at an ICR that is not necessarily very small, and potentially decreases the rate if the feedback takes too long. On the other hand, we could have chosen to start at a relatively low rate (although it is sometimes difficult to define what is "low enough"), and then increase based on feedback [10]. ICR was chosen to allow for RPC-like flows to derive the benefit of a "fast-start", especially if the total amount of data transmitted in the "burst" is small enough as not to exceed the buffering in the network.

Source Rule 6 is important in a rate-controlled framework, to protect the network from sources that are continuing to transmit at an incorrect rate, due to the absence of timely feedback. Unlike a window flow-controlled environment, where there is a natural protection from persistent overload, sources in a rate-controlled environment need a self-regulator that gradually reduces a source's rate in the absence of feedback. This tries to correct incorrect rates that may overload the network.

Subsequent to start-up, even during normal operation, when a significant increase in the rate occurs, Rule 6 may be triggered. Consider the following scenario. Subsequent to a rate change at time $t = 0$, the rate at which BRM cells return will correspond to the rate at which FRM cells were sent approximately one RTT earlier. Therefore, if that rate is small, and the new rate at which the station is transmitting FRM cells is much higher, then potentially more than *Crm* (Missing RM-cell count) in-rate FRM cells may be transmitted prior to a BRM cell returning after the rate change at $t = 0$. Rule 6 will be triggered on sending *Crm* FRM cells. This reduction is multiplicative and proportionate to the current rate ACR. Refer to Fig. 4 for the relative timing of the rate reduction after a BRM cell arrives at t_0.

When a BRM cell is received, and the ER (or the CI, NI bits for EFCI operation) allows the source rate to increase, then ACR is increased additively by a fixed amount. Subsequent BRM cells received would continue to increase the ACR similarly, until the ACR reaches the ER value or reaches PCR for the connection. The ACR value from which we increase is the source rate achieved after any (zero or more) reductions caused by Rule 6.

As a result, there are two counter-acting actions taking place at the source: (1) A rate decrease as a result of Rule 6, when there are too many FRM cells transmitted before a BRM cell is received; and (2) A rate increase, potentially, when a BRM cell returns indicating that the current source rate is below its target.

We examine below whether the rate increases that take place as a result of the BRM cells returning make up for the decreases caused while "running blind" awaiting the return of BRM cells. The analysis examines the case where there are n decreases (as a result of (1)), before an increase takes place (as a result of (2)). We want to derive conditions that the rate remains stable. This implies that the rate decreases are no more than the rate increase, so that there is no "net decrease" in the rate at the source because of Rule 6.

5.1. Estimation of ACR with no RM cell loss

In this subsection we examine the behavior of ACR in a single "epoch" where an epoch is the time interval between the arrival of two BRM cells. Subsequent to the arrival of the first of these two BRM cells, we have possibly multiple decreases as a result of the repeated application of Rule 6 that causes reductions in the ACR if the next BRM cell arrives late.

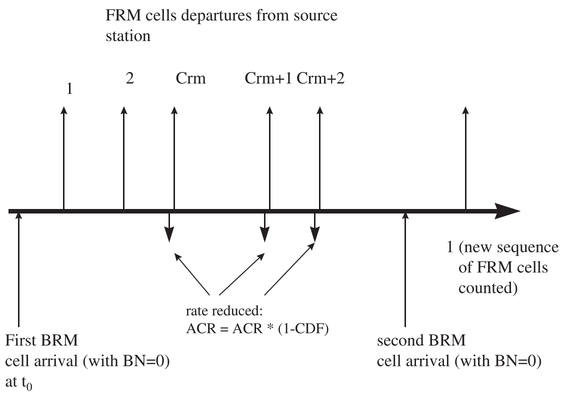

Fig. 4. Timing relationships between forward and backward RM cells at source with respect to Rule 6.

Denote the current $\mathrm{ACR} \equiv \mathrm{ACR}_0$ and the ACR after k consecutive rate reductions, before the arrival of the second BRM cell, from Rule 6 as:

$$\mathrm{ACR}_k = \mathrm{ACR}_0(1-\mathrm{CDF})^k, \quad k = 1,2,\ldots \tag{5.1}$$

Lemma 5.1. *The time interval t_k for the kth consecutive rate reduction from Rule* 6 *is*

$$t_1 = Crm \cdot \min\left\{ \max\left\{ \frac{Mrm}{\mathrm{ACR}_0}, Trm \right\} + \frac{1}{\mathrm{ACR}_0}, \frac{Nrm}{\mathrm{ACR}_0} \right\} + t_0, \tag{5.2a}$$

$$t_k = \min\left\{ \max\left\{ \frac{Mrm}{\mathrm{ACR}_{k-1}}, Trm \right\} + \frac{1}{\mathrm{ACR}_{k-1}}, \frac{Nrm}{\mathrm{ACR}_{k-1}} \right\}, \quad k = 2,\ldots, \tag{5.2b}$$

where

$$0 \leqslant t_0 \leqslant \min\left\{ \max\left\{ \frac{Mrm}{\mathrm{ACR}_0}, Trm \right\}, \frac{Nrm-1}{\mathrm{ACR}_0} \right\}.$$

Proof. There are two cases when FRM cells are sent.

Case 1. At a rate ACR_j, the source sends an FRM cell out after every *Nrm*-1 cell, as long as there are data cells to be sent. This corresponds to the second factor in Eqs. (5.2a) and (5.2b).

Case 2. If instead, we have *Mrm* data cells sent out ($Mrm < Nrm - 1$) and the timer also expires after time *Trm* then an FRM cell is sent. This is the first factor in Eqs. (5.2a) and (5.2b).

The first rate reduction occurs after *Crm* FRM cells have been sent after receiving a BRM cell with $\mathrm{BN} = 0$. The time interval of sending the *Crm* FRM cells is given in the first term of (5.2a). On the other hand, a BRM cell can be received either right before the first FRM cell is to be transmitted or after the transmission of the previous FRM cell, during normal operation. This is time interval t_0. The sum of the two terms is the time interval t_1 in (5.2a). Afterwards, each FRM cell sent results in a rate reduction and that gives the time intervals t_k in (5.2b) for the consecutive rate reductions. □

Before we proceed with the rate analysis, observe that rate ACR can never be reduced to less than the MCR and consequently, the accumulated rate reduction before a rate increase from receiving a BRM cell is no more than ACR–MCR. On the other hand, a received BRM cell allows a rate increase (by Source Rule 8) of:

$$r_{\text{inc}} = \text{RIF} \cdot \text{PCR}. \tag{5.3}$$

Also note that $\text{ACR} \leqslant \text{ER}$.

Proposition 5.1. *Suppose that* $\text{ACR}_0 - \text{MCR} \leqslant r_{\text{inc}}$. *Then, the accumulated rate reduction is less than the increase permitted by the received BRM cell. Thus, there is no net rate reduction after a BRM cell with* $BN = 0$ *and rate ER is received in an epoch. Furthermore, if* $\text{ER} - \text{MCR} \leqslant r_{\text{inc}}$, *then there are no rate reductions after a BRM cell with* $BN = 0$ *is received.*

One may also look at this single epoch in the context of a steady-state environment, where the network returns a fixed rate, ER, in each RM cell. Then, the maximum accumulated reduction is $\text{ER} - \text{MCR}$. We can extend the analysis to the case of multiple epochs, still with the assumption of no loss of RM cells. For multiple epochs, we have to ensure that at all times, the total reduction that occurred previously is made up by the total increase. For this, we now discuss the rate decrease in between two consecutive BRM cells with $\text{BN} = 0$. Suppose that we just had a possible rate increase at time t_0 as a result of receiving a BRM cell. We examine what would be the maximum number of rate decreases occurring as a result of Rule 6, prior to the arrival of another BRM cell at time $(T_0 + \tau)$. Here τ is typically the inter-departure time of FRM cells from the source about one RTT time earlier, at $t_0 - \text{RTT}$. For instance, $\tau = (Nrm - 1)/(\text{ACR}_{T_0 - \text{RTT}})$. For simplicity, for the rest of this subsection we assume that the total amount of rate decrease is no more than $\text{ACR} - \text{MCR}$. Otherwise, the exact amount of rate decrease is $\text{ACR} - \text{MCR}$ and we can easily determine whether each received BRM cell will increase ACR back by a comparison of the reduced rate $\text{ACR} - \text{MCR}$ with the increase in the rate, by $\text{RIF} \cdot \text{PCR}$, in (5.3).

Lemma 5.2. *Suppose that two consecutive BRM cells with* $BN = 0$ *are received in time period* τ. *Then after n consecutive reductions, a rate increase occurs, where*

$$n = \max\left\{k\colon \sum_{i=1}^{k} t_i \leqslant \tau\right\}, \tag{5.4}$$

where t_k *is given in* (5.2a) *and* (5.2b). *Before the rate increase, the total rate reduction is*

$$\text{ACR}_0 - \text{ACR}_n = \text{ACR}_0[1 - (1 - \text{CDF})^n]. \tag{5.5}$$

After n consecutive rate reductions, comparing the rate increase in (5.3) with the rate decrease in (5.5) and from Proposition 5.1, we have the following:

Theorem 5.1. *If* $\text{ER} - \text{MCR} \leqslant \text{RIF} \cdot \text{PCR}$, *then there is no net rate reduction after a BRM cell with* $\text{BN} = 0$ *is received. Otherwise, a necessary and sufficient condition that there is no net rate reduction after a BRM cell with* $BN = 0$ *is received is*

$$\frac{\text{RIF} \cdot \text{PCR}}{\text{ACR}_0} + (1 - \text{CDF})^n \geqslant 1, \tag{5.6}$$

where n, given in (5.4), *is the number of rate reductions since the previous BRM cell with* $BN = 0$ *was received.*

Theorem 5.1 provides a relationship between the amount of decreases in the rate as a result of successive application of Rule 6, prior to a rate increase that compensates for the decreases. The number of decreases, n, between two consecutive BRM cells received is given in Eq. (5.4).

5.2. Asymptotic estimation of ACR with no RM cell loss

The previous subsection examined the behavior of the rate over a single BRM inter-arrival time. If there is variability in the inter-arrival time of the BRM cells at the source, the actions at the source may be different from one epoch to the next. As before, an epoch is the time between two consecutive BRM cell arrivals. In some cases, a BRM cell may arrive before Rule 6 triggers a rate reduction. We now consider the asymptotic behavior of the rate when the network is consistently returning a fixed value of ER in the BRM cells, and examine if the reductions due to Rule 6 result in an overall decrease in the source rate, when observed over a long period of a number of BRM cell arrivals.

Suppose that there are k rate reductions from Rule 6 in between two BRM cells with $\mathrm{BN}=0$. Then the amount of rate reduction is

$$\mathrm{ACR}_0[1-(1-\mathrm{CDF})^k] \leqslant r_{\mathrm{dec}} \leqslant \min\{\mathrm{ER}-\mathrm{MCR}, \mathrm{ER}[1-(1-\mathrm{CDF})^k]\}. \quad (5.7)$$

The additional complexity comes primarily from the fact that in some epochs we may be limited by MCR. Also, because of the variability in the inter-arrival time, some BRM cells prevent the triggering of a reduction. In other epochs, a reduction may be triggered when the next FRM cell is sent. On the other hand, each rate increase from receiving a BRM with $\mathrm{BN}=0$ is given in (5.3).

Suppose that during a (long) period of time, a large number, N, of FRM cells have been sent out and n BRM cells have been received. We now analyze the extent of the rate change. Denote $(x)_+ = \max\{0, x\}$. We have the following:

Lemma 5.3. *The number of rate reductions d due to the N FRM cell transmissions is*

$$[N-(n+1)\cdot Crm]_+ \leqslant d \leqslant N - Crm. \quad (5.8)$$

The accumulated amount of rate reduction r_{dec} *is*

$$\mathrm{ER}[1-(1-\mathrm{CDF})^d] \leqslant r_{\mathrm{dec}} \leqslant \min\{\mathrm{ER}-\mathrm{MCR}, d\cdot\mathrm{ER}\cdot\mathrm{CDF}\}. \quad (5.9)$$

Proof. After first sending *Crm* FRM cells, each transmission of an FRM cell causes a rate reduction and that contributes to the upper bound in (5.8).

After having sent *Crm* FRM cells, a BRM cell with $\mathrm{BN}=0$ arrives and that prevents a rate reduction. Therefore, the rate reductions are minimized when this happens n times where n is the number of BRM cells with $\mathrm{BN}=0$ received during the period of observation, and this accounts for $n\cdot Crm$ FRM cells sent without contributing to rate reductions. The remaining FRM cells, $N-n\cdot Crm$, if any, contribute to rate reductions after an additional *Crm* FRM cell is sent. This is a worst-case scenario for rate reductions and provides the lower bound in (5.8).

The maximal amount of the reductions in the rate is based on the starting value for the rate for each of the reductions being $\mathrm{ACR}=\mathrm{ER}$ when $k=1$ in (5.7). Thus, the total accumulated reduction from d rate reductions is $d\cdot\mathrm{ER}\cdot\mathrm{CDF}$; however it should not exceed $\mathrm{ER}-\mathrm{MCR}$. This gives the upper bound in (5.9).

The minimal amount of reductions is from d consecutive reductions, estimated in (5.4). This minimal amount of reduction in the rate is from one epoch of reductions starting at ER and suffering d consecutive reductions, which is the lower bound in (5.9). □

We have discussed rate reductions during a long period with a large number, N, of FRM cells transmitted. During this period, n BRM cells are received, resulting in rate increases. From (5.3), the total rate increase is $n\cdot\mathrm{RIF}\cdot\mathrm{PCR}$. On the other hand, from (5.8) and (5.9), the total rate reduction during this period is no more than $[N-Crm]\cdot\mathrm{ER}\cdot\mathrm{CDF}$, starting from ER. Since the rate is bounded above by ER, we have the following:

Lemma 5.4. *The total rate increase* r_{inc} *from receiving the n BRM cells is*

$$0 \leqslant r_{\mathrm{inc}} \leqslant \min\{n\cdot\mathrm{RIF}\cdot\mathrm{PCR}, \mathrm{ER}-\mathrm{MCR}, (N-Crm)\cdot\mathrm{ER}\cdot\mathrm{CDF}\}. \quad (5.10)$$

The above discussion is deterministic, looking at the case of increases and decreases across multiple epochs. To ensure rate stability, the total amount of reduction r_{dec} should be no more than the total amount of increase r_{inc}.

5.3. Rate stability in the presence of RM cell loss

Suppose that an end system is operating at a steady state with a transmission rate ACR = ER. While FRM cells are being transmitted and BRM cells returned, we want to derive a set of "broad" conditions to ensure that the cumulative reductions due to Rule 6 are overcome by the aggregate increase achieved by the returning BRM cells. In steady state, the total number of FRM cells transmitted should correspond to the number of BRM cells received. In a sense, we are "integrating" the effect across multiple BRM cells returning to the source without necessarily maintaining the constraint that reductions have to occur between each and every arriving BRM cell. In some cases, a BRM cell may arrive prior to Rule 6 being triggered.

In Section 5.1 we considered the case when there were no RM cells lost in the network. However, with RM cell loss there is an increasing likelihood of rate reductions due to Rule 6. RM cell loss both in the forward (FRM cells) and backward (BRM cells) directions can result in the source experiencing a rate reduction. For ease of exposition, in the rest of this subsection we will treat all RM cell losses as in effect being BRM cell losses, since from the source's perspective, it is reflected as a BRM cell not arriving in time. For clarity, we assume that with all possible reductions,

$$\mathrm{MCR} \leqslant \mathrm{ACR} \leqslant \mathrm{ER} \leqslant \mathrm{PCR}. \tag{5.11}$$

In the following rate analysis, we model the worst-case scenarios for rate reductions, and take $t_0 = 0$.

Suppose that after the initial setup the source rate ACR has converged to ER as a result of a constant allocation of ER for the VC being returned by the network. We derive the conditions such that ACR remains stable at ER in the event of BRM cell losses.

As indicated in Proposition 5.1, if $\mathrm{ER} - \mathrm{MCR} \leqslant r_{\mathrm{inc}}$, then the rate remains stable, where r_{inc} is the amount of rate increase from receiving a BRM cell with BN = 0, given in (5.3). In the following analysis, we assume that $\mathrm{ER} - \mathrm{MCR} > r_{\mathrm{inc}}$ and derive conditions such that the accumulated rate reductions in the time interval τ is no more than $r$inc. Under these conditions, the rate ACR remains larger than MCR. We have the following conditions for rate stability.

Proposition 5.2. *Suppose that the time interval between two successive BRM cells with* BN = 0 *is* τ. *Then the rate ACR is stable if and only if*

$$n \leqslant \frac{\log\left[1 - \frac{\mathrm{RIF}\cdot\mathrm{PCR}}{\mathrm{ER}}\right]}{\log(1-\mathrm{CDF})}, \tag{5.12}$$

where

$$n = \max\left\{k : \sum_{i=1}^{k} t_i \leqslant \tau\right\}, \tag{5.13}$$

where t_i *is the time interval for the ith consecutive rate reduction, given in* (5.3).

Proof. The rate is stable if an increment of ACR from the arrival of a BRM cell with BN = 0 offsets the decrements that occur before its arrival. This can be derived directly from Lemma 5.2 and Theorem 5.1 with $\mathrm{ACR}_0 = \mathrm{ER}$. □

In the above proposition, the rate stability conditions depend on two intermediate parameters: t_i and n. The condition and intuition derived from these are related to the number of reductions due to Rule 6 and the time for these, which are similar to the conditions on the system parameters that are specified in the ATM Forum Specification [1]. The larger the number of reductions, n, the harder it is to maintain stability in the rate. We now go a significant further step in presenting one of the main results of our paper: a quantification of how many BRM cells could be lost consecutively without affecting the rate stability, based on the system parameters only. This is based on the following results that estimate the number of rate decreases in a time period τ between two consecutive BRM cells with BN = 0 received.

Theorem 5.2. *The number of rate decreases in a time period* τ *between two consecutive BRM cells with* BN = 0 *received is* [1]

[1] More precisely, n is the largest integer which is no more than the expression in (5.14). For clarity, we leave out the details in this and some of the subsequent expressions.

$$n = \frac{1}{Mrm+1} \cdot \{\mathrm{ER} \cdot \tau - [(\mathrm{ER} \cdot Trm + 1) \cdot (r-1) + Nrm \cdot (Crm + l - 1)]\} + r, \tag{5.14}$$

where

$$l = \max\left\{0, \frac{\log \frac{Nrm}{Trm \cdot \mathrm{ER}}}{\log(1-\mathrm{CDF})}\right\} \tag{5.15}$$

and

$$r = \max\left\{0, \frac{\log \frac{Mrm}{Trm \cdot \mathrm{ER}}}{\log(1-\mathrm{CDF})}\right\}. \tag{5.16}$$

Proof. We first estimate the time interval t_k between consecutive rate decreases. We assume $Nrm - 1 \geqslant Mrm$, based on generally accepted ranges for these parameters.

We order the decreases in the rate in the following manner: for time instants $t_i = 0, \ldots, l$, when the rate is still high, FRM cells are sent after every $Nrm - 1$ data cells and BRM cell transmissions. Subsequently, the source rate ACR is sufficiently low so that the FRM cells are sent by the timer, *Trm*, expiring. This occurs at time instants $t_i = l+1, \ldots, r$. Finally, when the rate has reduced to small enough values, the FRM cells have to wait for *Mrm* cells to be sent out, which occur after the timer *Trm* has expired.

$$l = \max\left\{i\colon \frac{Nrm}{\mathrm{ACR}_{i-1}} \leqslant Trm\right\},$$
$$r = \max\left\{i\colon \frac{Mrm}{\mathrm{ACR}_{i-1}} \leqslant Trm\right\}. \tag{5.17}$$

The condition $l = 0$ implies that it is always true that $Nrm/\mathrm{ACR}_i > Trm$ and similarly, the condition $r = 0$ implies that $Mrm/\mathrm{ACR}_i > Trm$ is always true.

For large enough rates of ACR, Rule 6 is triggered based on sending *Crm* FRM cells, each of which takes Nrm/ACR_{i-1} cell times. For lower (intermediate) rates of ACR, when the source has sent more than *Mrm* cells but less than *Nrm* cells in time *Trm*, the decrease is triggered by *Crm* FRM cells transmitted on expiration of the timer *Trm*. For "very low rates", when *Trm* has already expired, FRM cell transmission occurs only after the requisite minimum number of cells *Mrm* has been sent. From (5.2a) and 5.2b, we can derive

$$t_i = \begin{cases} \frac{Crm \cdot Nrm}{\mathrm{ACR}_0}, & i = 1, \\ \frac{Nrm}{\mathrm{ACR}_{i-1}}, & i = 2, \ldots, l, \\ Trm + \frac{1}{\mathrm{ACR}_{i-1}}, & i = l+1, \ldots, r, \\ \frac{Mrm+1}{\mathrm{ACR}_{i-1}}, & i = r+1, \ldots, k. \end{cases} \tag{5.18}$$

Therefore

$$\sum_{i=1}^{k} t_i = Crm \cdot \frac{Nrm}{\mathrm{ACR}_0} + \left[\sum_{i=2}^{l} \frac{Nrm}{\mathrm{ACR}_{i-1}} + (r-l) \cdot Trm + \sum_{i=l+1}^{r} \frac{1}{\mathrm{ACR}_{i-1}} + \sum_{i=r+1}^{k} \frac{Mrm+1}{\mathrm{ACR}_{i-1}}\right].$$

Using (5.1), $\mathrm{ACR}_{i-1} = \mathrm{ER}(1-\mathrm{CDF})^{i-1}$, we have

$$\sum_{i=2}^{l} \frac{Nrm}{\mathrm{ACR}_{i-1}} = \frac{Nrm}{\mathrm{ER} \cdot \mathrm{CDF}}\left[\left[\frac{1}{1-\mathrm{CDF}}\right]^{l-1} - 1\right],$$

$$\sum_{i=l+1}^{r} \frac{1}{\mathrm{ACR}_{i-1}} = \frac{1}{\mathrm{ER} \cdot \mathrm{CDF} \cdot (1-\mathrm{CDF})^{l}} \times \left[\left[\frac{1}{1-\mathrm{CDF}}\right]^{r-l} - 1\right]$$

and

$$\sum_{i=r+1}^{k} \frac{Mrm+1}{\mathrm{ACR}_{i-1}} = \frac{Mrm+1}{\mathrm{ER} \cdot \mathrm{CDF} \cdot (1-\mathrm{CDF})^{r}}\left[\left[\frac{1}{1-\mathrm{CDF}}\right]^{k-r} - 1\right].$$

We have

$$\sum_{i=1}^{k} t_i = Crm \cdot \frac{Nrm}{\mathrm{ER}} + \frac{Nrm}{\mathrm{ER} \cdot \mathrm{CDF}} \times \left[\left[\frac{1}{1-\mathrm{CDF}}\right]^{l-1} - 1\right] + (r-l) \cdot Trm + \frac{1}{ER \cdot \mathrm{CDF}}\left[\left[\frac{1}{1-\mathrm{CDF}}\right]^{r} - \left[\frac{1}{1-\mathrm{CDF}}\right]^{l}\right] + \frac{Mrm+1}{\mathrm{ER} \cdot \mathrm{CDF}}\left[\left[\frac{1}{1-\mathrm{CDF}}\right]^{k} - \left[\frac{1}{1-\mathrm{CDF}}\right]^{r}\right]. \tag{5.19}$$

If we consider the boundary cases in (5.17), so that $Nrm/\mathrm{ACR}_l = Trm$ and $Mrm/\mathrm{ACR}_r = Trm$, then we obtain (5.15) and (5.16).

Since $\mathrm{CDF} \ll 1$, we have

$$\left(\frac{1}{1-\mathrm{CDF}}\right)^k \approx 1 + \frac{k\cdot \mathrm{CDF}}{1-\mathrm{CDF}}. \tag{5.20}$$

Based on (5.15) and (5.16), we have

$$\begin{aligned}\sum_{i=1}^{k} t_i &\approx Crm\cdot\frac{Nrm}{\mathrm{ER}} + \frac{(l-1)\cdot Nrm\cdot \mathrm{CDF}}{\mathrm{ER}\cdot\mathrm{CDF}\cdot(1-\mathrm{CDF})} \\ &\quad + (r-l)\cdot Trm + \frac{r\cdot\mathrm{CDF} - l\cdot\mathrm{CDF}}{\mathrm{ER}\cdot\mathrm{CDF}\cdot(1-\mathrm{CDF})} \\ &\quad + \frac{(Mrm+1)\cdot(k\cdot\mathrm{CDF}) - r\cdot\mathrm{CDF})}{\mathrm{ER}\cdot\mathrm{CDF}\cdot(1-\mathrm{CDF})} \\ &\approx Crm\cdot\frac{Nrm}{\mathrm{ER}} + \frac{(l-1)\cdot Nrm}{\mathrm{ER}} + \frac{r-l}{\mathrm{ER}} \\ &\quad + \frac{(k-r)\cdot(Mrm+1)}{\mathrm{ER}} + (r-l)\cdot Trm \leqslant \tau,\end{aligned}$$

that is

$$\begin{aligned}\sum_{i=1}^{k} t_i &\approx Crm\cdot\frac{Nrm}{\mathrm{ER}} + \frac{(l-1)\cdot Nrm}{\mathrm{ER}} + \frac{r-l}{\mathrm{ER}} \\ &\quad + \frac{(k-r)\cdot(Mrm+1)}{\mathrm{ER}} + (r-l)\cdot Trm \\ &\leqslant \tau.\end{aligned} \tag{5.21}$$

A routine computation from (5.21) yields (5.14). □

Theorem 5.2 provides a foundation for rate stability analysis. As a case study, we examine here the effect of losing $v-1$ *consecutive* BRM cells; the epoch between BRM cell arrivals is now increased by an amount equal to $(v-1)$ BRM cell inter-arrival times. We assume that ACR is still high enough such that $Nrm/\mathrm{ACR} < Trm$. Consequently, the time interval between two consecutive BRM cells received with $\mathrm{BN}=0$ is $\tau = v\cdot Nrm/\mathrm{ER}$. From Theorem 5.2 we have

Corollary 5.1. *If* $\mathrm{ER}-\mathrm{MCR} \leqslant \mathrm{RIF}\cdot\mathrm{PCR}$, *then the rate remains stable. Otherwise, the rate ACR remains stable with no more than* $v-1$ *consecutive BRM cells lost, where*

$$\begin{aligned}v \leqslant \frac{1}{Nrm}\Bigg\{&(Mrm+1)\cdot\left[\frac{\log\left[\frac{1-\mathrm{RIF}\cdot\mathrm{PCR}}{\mathrm{ER}}\right]}{\log(1-\mathrm{CDF})} - r\right] \\ &+ Crm\cdot Nrm + (l-1)\cdot Nrm \\ &+ (r-1)\cdot(\mathrm{ER}\cdot Trm+1)\Bigg\}.\end{aligned} \tag{5.22}$$

Proof. The first part of the corollary is from Proposition 5.1 and the second part can be obtained by a straightforward computation from Theorem 5.2 and Proposition 5.2 with $\tau = v\cdot Nrm/\mathrm{ER}$. □

If the rate remains high, so that $Nrm/\mathrm{ACR} \leqslant Trm$, then the rate decreases are triggered after *Crm* FRM cells are sent. In this case, from (5.18), $t_1 = Crm\cdot Nrm/\mathrm{ACR}_0$ and $t_i = Nrm/\mathrm{ACR}_{i-1}, i = 2,\cdots,l$. In this case, (5.21) only contains the first two terms and we have the following:

$N = v - Crm + 1.$

Given Proposition 5.2, we have

Corollary 5.2. *If* $\mathrm{ER}-\mathrm{MCR} \leqslant \mathrm{RIF}\cdot\mathrm{PCR}$, *then the rate remains stable. Otherwise, suppose that the rate ACR satisfies* $Nrm/\mathrm{ACR} < Trm$. *Then the number of allowed consecutive BRM cell losses is*

$$v \leqslant \frac{\log\left[1-\frac{\mathrm{RIF}\cdot\mathrm{PCR}}{\mathrm{ER}}\right]}{\log(1-\mathrm{CDF})} + Crm - 1. \tag{5.23}$$

When the rate is low, i.e., $\mathrm{ER}-\mathrm{MCR} \leqslant \mathrm{RIF}\cdot\mathrm{PCR}$, then the rate remains stable because a single rate increase from a BRM cell overcomes any accumulated effects of rate decreases from Rule 6. The interesting region for the rate stability issue is when $\mathrm{ER}-\mathrm{MCR} > \mathrm{RIF}\cdot\mathrm{PCR}$. In this case, the number v, of consecutive losses tolerated depends on several parameters and the network feedback rate ER. We examine the values of v as the rate ER varies, below.

5.4. Rate reductions and allowable consecutive RM cell losses

We examine, numerically, the number of consecutive RM cell losses (called BRM cell loss for simplicity), for typical parameter values, as the source rate ACR varies. We restrict ourselves to the condition that ER − MCR > RIF · PCR, and in particular when Corollary 5.2 applies. In Fig. 5(A), we examine first the number of consecutive RM cell losses that may be tolerated, when the steady-state rate, ER, returned by the network to the source is a small fraction (less than 0.1) of the peak rate, PCR. The more significant case, for higher values of ER up to PCR, is examined in Fig. 5(B).

The values of the parameters chosen for the figures are as follows: The values chosen are $Crm = 32$, $Nrm = 32$, $Mrm = 2$ and $Trm =$ 100 ms. We have chosen a somewhat conservative value of RIF $= 1/64$, based on our understanding of the need to minimize the burst load on the network when the source rate increases. We have chosen to vary the ratio of ER/PCR, so that it is applicable for a range of link speeds, and examine how many consecutive RM cell losses may be tolerated while keeping the source rate stable.

Fig. 5(A) shows the numerical evaluation of Eq. (5.23) for different values of CDF, varying in powers of 2 from 1/2 to 1/128. The value of ER/PCR ranges from 1/64 to 0.1. When CDF $=$ $1/2$, a few consecutive RM cell losses are tolerated (up to approximately 30), even at these low rates of ER. Increasing ER only reduces the tolerance for these losses. But, when we consider more realistic values of CDF (we believe 1/64 or higher), a slightly larger number of RM cell losses are tol-

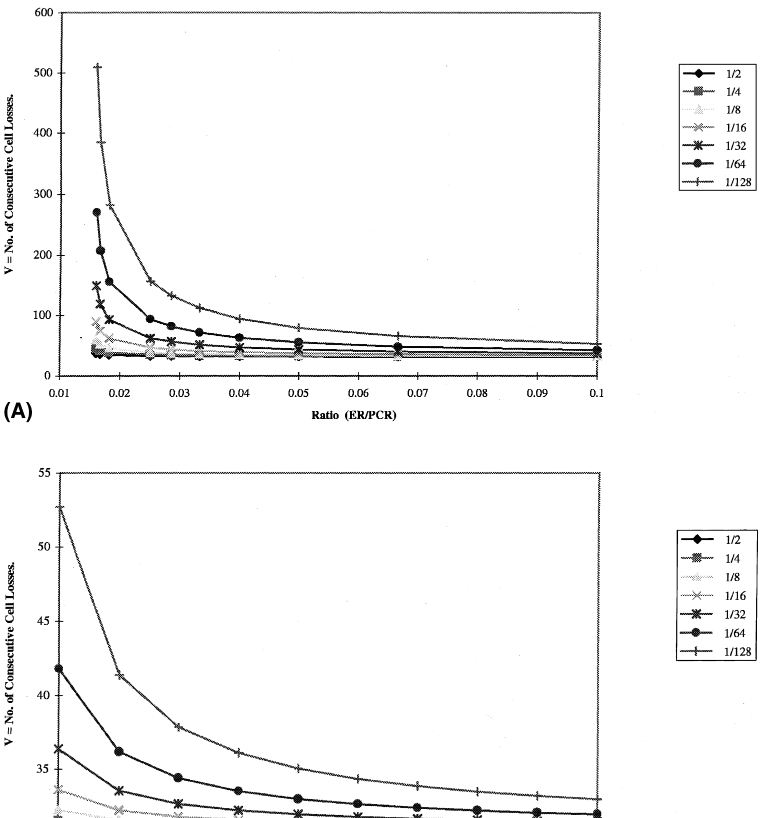

Fig. 5. (A) Bounds on allowable consecutive RM cell losses as a function of ER (low values of ER/PCR). (B) Bounds on allowable consecutive RM cell losses as a function of ER (high values of ER/PCR).

erated for small values of *ER*. For an intermediate $\mathrm{CDF} = 1/32$, at a ratio of $\mathrm{ER}/\mathrm{PCR} = 0.1$, the number of consecutive losses tolerated is higher at the small values of ER. Hence, for these low rates, for reasonable values of the parameters, RM cell losses are not such a substantial concern.

We now examine the tolerance to RM cell losses in the more realistic situation of having a higher ER rate. The ratio ER/PCR ranges from 0.1 to 1.0. The *shape* of the curves, as seen in Fig. 5(B) for the number of consecutive RM cell losses tolerated is similar to the previous figure. However, the absolute number of RM cell losses tolerated is fewer for the larger values of CDF. For a typical value of $\mathrm{CDF} = 1/32$, when $\mathrm{ER}/\mathrm{PCR} = 1.0$, the number of consecutive RM cell losses tolerated is approximately 30, with little reduction in the tolerance as the ratio of ER/PCR goes from 0.1 to 1.0. This gives us some guidance on how large CDF can be: for the set of "default parameter" values we chose, and for $\mathrm{RIF} = 1/64$, it would be advisable to keep CDF to be approximately 1/32 or lower.

There are trade-offs unfortunately. The motivation of keeping the rate stable suggests a small value of CDF, such as 1/32 or 1/64. There is the conflicting desire not to overload the network with a source sending at an inappropriately high rate, either at start-up or in the presence of delayed or lost feedback: thus wanting the source to rapidly reduce the source rate every time Rule 6 is triggered. This suggests a larger value of CDF than 1/32. A further practical constraint to keep in mind is that RM cell loss is likely to be highest when the rate of transmission is high. A processor within the switch performing the RM cell processing may be unable to keep up with this processing when all the switch links are consistently heavily utilized. Our analysis, and Fig. 5(B), suggest that we need to ensure that there is no burst loss of 30 or more RM cells. This ensures that the source rate remains stable, and Rule 6 does not unnecessarily cause the steady-state rate to be reduced unnecessarily.

We now examine how the ACR varies with time, due to repeated triggering of Rule (6). By using Eq. (5.18), we can numerically solve for ACR, given an initial value of ACR. Figs. 6(A)–(C) show the progress of ACR for varying ER. Fig. 6(A) is for ER starting at a very small value of 3.5 cells/ms, equivalent to a T_1 link speed of 1.5 Mbits/s. Fig. 6(B) is for an intermediate value of ER of 100 cells/ms, equivalent to a T3 link speed of 45 Mbits/s. Finally, Fig. 6(C) is for an OC-3 link speed of 155 Mbits/s or 360 cells/ms. We examine the effect of varying the value of CDF from 1/2 to 1/256. The figures are obtained by having the initial value of ACR_0 equal to ER. We then repeatedly apply Eq. (5.18) to derive the steps of reduction of ACR. *Nrm* was chosen to be 32, *Trm* was 100 ms, $Mrm = 2$ and *Crm* was 16. The numerical iteration stopped when ACR reduced down to MCR.

In Fig. 6(A), with $\mathrm{ER} = 3.5$ cells/ms, ACR goes down relatively slowly, down to $\mathrm{MCR} = 0$, for small values of $\mathrm{CDF} = 1/256$. When CDF is large ($= 1/2$), ACR goes down very rapidly, and it only takes a small number of steps to go down to MCR ($\mathrm{MCR} = 0$, for this case).

As the starting value of ER gets larger, Figs. 6(B) and (C) reveal that ACR goes down to MCR much more rapidly, taking only 40 ms to reach an $\mathrm{MCR} = 50$ cells/ms, when $\mathrm{CDF} = 1/256$. When ER is larger, each occurrence of the reduction due to Rule (6) results in a larger absolute magnitude of the reduction of ACR. This, in conjunction with the more frequent triggering of Rule (6) results in a quick correction of ACR, reducing it down to MCR quickly. This is precisely the desired effect of Rule (6), which is to protect the network in the absence of feedback from the network.

Figs. 7(A)–(C) are another way of examining the effects of Rule 6. These show the number of reductions (due to triggering of Rule (6)) before ACR goes down to MCR. This potentially is a measure of the tolerance to prolonged loss of RM cells in the network. With a small value of ER (initial value of ACR), it takes a long time before ACR reduces to MCR, and a longer interval passes between each reduction, as shown in Fig. 7(A). This is reasonably harmless, because ACR is small. As ACR gets larger (Fig. 7(B)), even though the number of reductions is smaller to reduce the rate from 100 to 50 cells/ms, it takes a lot less time-going down from seconds to tens of milliseconds. As CDF increases, it takes a lot less time and fewer reductions to bring the rate down to MCR. Finally, in Fig. 7(C), we

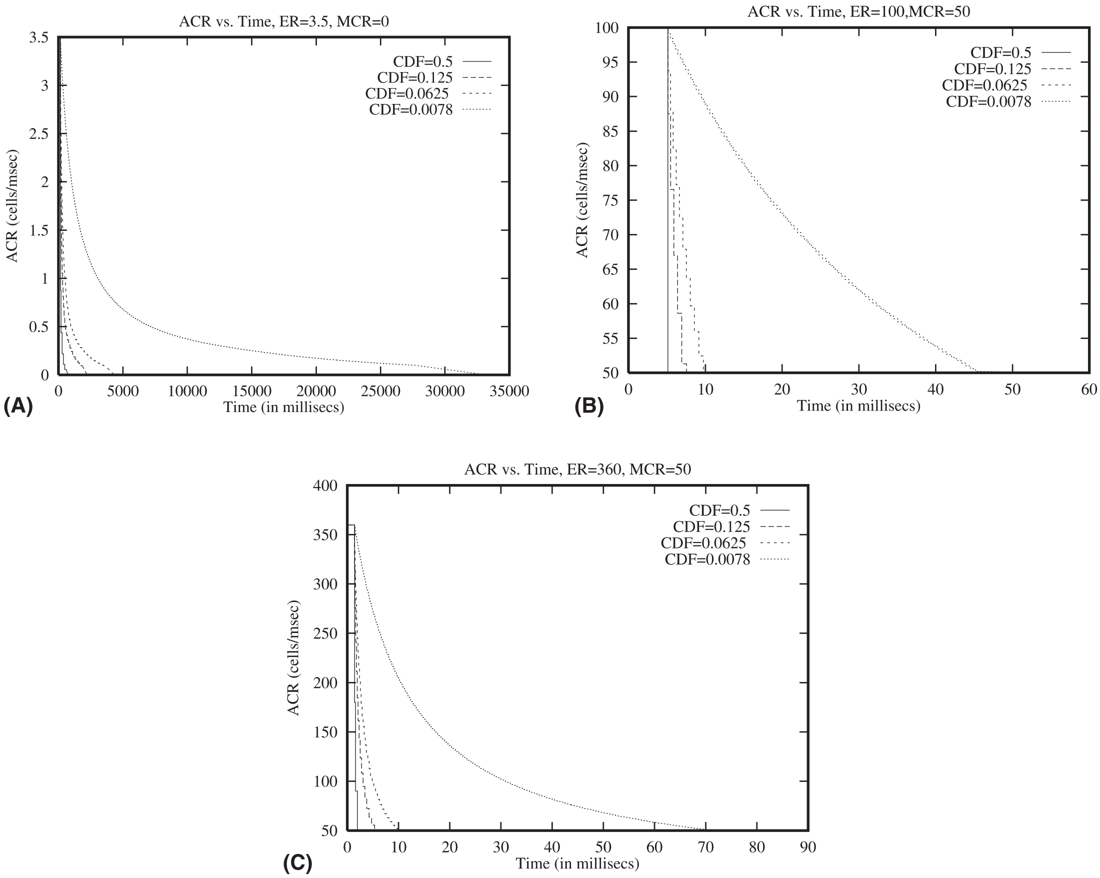

Fig. 6. (A) Variation of ACR with time, from repeated application of Rule 6: ER = 1.5 Mbps, MCR = 0. (B) Variation of ACR with time, from repeated application of Rule 6: ER = 45 Mbps, MCR = 22.5 Mbps. (C) Variation of ACR with time, from repeated application of Rule 6: ER = 155 Mbps, MCR = 22.5 Mbps.

observe the increased time it takes to reduce ACR from 360 cells/ms down to the MCR of 50 cells/ms. Note however, the increased initial slope, indicating the more rapid reduction in ACR based on the higher frequency at which Rule (6) is triggered at higher initial value of ACR. This is also an important, desired feature of a rate control protocol, where the reductions in the rate occur more quickly (in addition to the larger magnitude as shown in Fig. 6(C)) when the rate is high.

Although these insights were gained based on the detailed analysis of Rule 6 of the ABR specification, we believe these considerations should be generally applicable even to a transport protocol that uses rate-based flow and congestion control scheme.

5.5. Rate analysis under probabilistic assumptions of RM cell losses

In Section 5.3, we were primarily concerned with maintaining the source rate stability with RM cell loss, but in a deterministic sense. We now examine the effect of RM cells losses, when we lose them in a probabilistic manner. We assume that our observation of the effect on the source rate ACR is over a sufficiently long period, covering multiple epochs. As before, over this period, the total number of BRM cells (*lost* as well as those

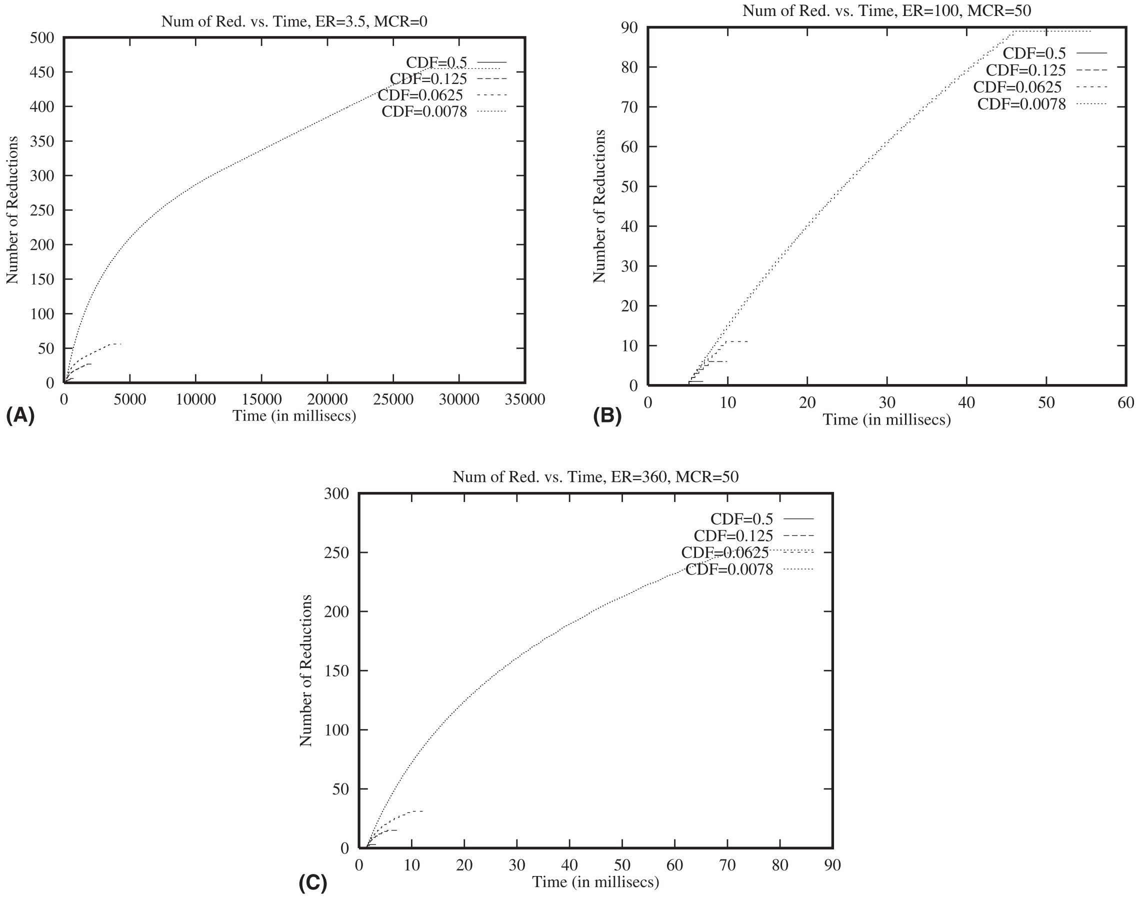

Fig. 7. (A) Number of successive rate reductions for ACR, from repeated application of Rule 6: ER = 1.5 Mbps, MCR = 0. (B) Number of successive rate reductions for ACR, from repeated application of Rule 6: ER = 45 Mbps, MCR = 22.5 Mbps. (C) Number of successive rate reductions for ACR, from repeated application of Rule 6: ER = 155 Mbps, MCR = 22.5 Mbps.

returned to source) is equal to the FRM cells transmitted. We determine for a given probability ρ of RM cells' loss, the relationship between the parameters associated with Rule 6, CDF and *Crm*, and the rate increase factor RIF so that the expected rate is not below the starting value of ACR.

As before, we assume that the target rate ER returned to the source is constant over the interval. Knowing a starting rate ACR at time $t = 0$ which is equal to the steady-state rate for the VC (i.e., ACR = ER at $t = 0$), and assuming that the network continues to provide feedback of the same ER value with all BRM cells, we examine the effect on ACR when RM cells are lost.

Suppose that the RM cell loss probability is $0 \leqslant \rho \leqslant 1$. If we send N FRM cells over a time period, then we receive $N \cdot (1-\rho)$ BRM cells. Therefore, the frequency of sending FRM cells is $1/(2-\rho)$ and that of receiving BRM cells is $(1-\rho)/(2-\rho)$. Consequently, the expected amount of rate increases in the steady state over a fixed time interval is $\mathrm{RIF} \cdot \mathrm{PCR}[1 - 1/(2-\rho)]$. Corresponding to the same time interval, a decrease is caused if *Crm* or more consecutive FRM cells are sent. The probability of sending $(Crm + i)$ FRM cells consecutively (i.e., without an intervening event of a BRM cell returning) is $1/(2-\rho)^{Crm+i}$, $i = 0, 1, \ldots,$ and each of these causes a rate decrease of $\mathrm{ACR}[1-(1-\mathrm{CDF})^{i+1}]$. The expected rate reduction is then

$$\sum_{i=0}^{\infty} \frac{\mathrm{ACR}\left[1-(1-\mathrm{CDF})^{i+1}\right]}{(2-\rho)^{Crm+i}} = \frac{\mathrm{ACR}}{(2-\rho)^{Crm-1}}\left[\frac{1}{1-\rho}-\frac{1-\mathrm{CDF}}{1-\rho+\mathrm{CDF}}\right].$$

In summary, the expected rate change in the steady state with an initial rate of ACR is

$$\mathrm{RIF}\cdot\mathrm{PCR}\left[1-\frac{1}{2-\rho}\right] - \frac{\mathrm{ACR}}{(2-\rho)^{Crm-1}}\left[\frac{1}{1-\rho}-\frac{1-\mathrm{CDF}}{1-\rho+\mathrm{CDF}}\right].$$

For the rate to be stable, statistically, the expected rate change should be non-negative. Thus, we have the following:

Proposition 5.3. *With an RM cell loss probability $0 \leqslant \rho \leqslant 1$ and current transmission rate ACR, there is no expected rate decrease if*

$$\mathrm{RIF}\cdot\mathrm{PCR}\left(1-\frac{1}{2-\rho}\right) \geqslant \frac{\mathrm{ACR}}{(2-\rho)^{Crm-1}}\left(\frac{1}{1-\rho}-\frac{1-\mathrm{CDF}}{1-\rho+\mathrm{CDF}}\right). \tag{5.24}$$

We examine, numerically, the tolerable RM cell loss probability for a set of system parameter values. We choose an OC-3 link, with PCR = 360,000 cells/s, for varying values of *Crm*, RIF, CDF and cell rate ACR. Note here that we are not modeling the effect of feedback delays. Table 1 shows the statistically allowable RM cell loss probability, ρ.

The tolerance to RM cell loss is very low (i.e., ρ is small) when the rate increase factor RIF is small, but the rate decrease due to the triggering of Rule 6 (CDF) is large. When the rate is high, e.g., ACR = PCR = 360,000 cells/s, RIF is small (say 1/64), and CDF is large (1/4), and the window, *Crm*, of the missing RM cells is small (16), ρ is only 0.1623. As we make the increases smaller and the decreases (due to Rule 6) larger, the tolerance to RM cell loss gets smaller and smaller, with ρ tending to 0. For the representative set of values of $Crm = 16, \mathrm{RIF} = 1/64$ and $\mathrm{CDF} = 1/128$, we find that the tolerance to RM cell losses can be quite high: up to 72% of the RM cells may be lost, without Rule 6 exhibiting an undesirable effect.

We are not modeling effects of feedback delay. Therefore, when CDF is approximately RIF, we should need only about 1 out of every *Crm* RM cells to be returned not to trigger Rule 6. Thus, our tolerance to RM cell loss may be quite high, as seen in the fifth and last rows of Table 1. In fact, for typical values of RIF and CDF, and for a reasonably large value of *Crm*, the system can tolerate a fairly substantial probability of RM cell loss.

However, since we do not model feedback delays, we should in fact ensure that the *Crm* value

Table 1
Allowable RM cell loss probabilities

ρ	ACR	PCR	*Crm*	RIF	CDF
0.1623	360 000	360 000	8	1/64	1/4
0.6580	360 000	360 000	16	1/64	1/32
0.6924	360 000	360 000	16	1/64	1/64
0.7253	360 000	360 000	16	1/64	1/128
0.7977	360 000	360 000	32	1/64	1/32
0.8135	360 000	360 000	32	1/64	1/64
0.8294	360 000	360 000	32	1/64	1/128
0.7708	36 000	360 000	16	1/64	1/32
0.7968	36 000	360 000	16	1/64	1/64
0.8217	36 000	360 000	16	1/64	1/128
0.8536	36 000	360 000	32	1/64	1/32
0.8666	36 000	360 000	32	1/64	1/64
0.8797	36 000	360 000	32	1/64	1/128

that is stated above in Table 1 is corrected for the practical case when a large number of FRM cells is sent in the first round-trip time. *Crm* only affects the first rate reduction of Rule 6. Thus, the actual value of *Crm* used should be the value of *Crm* given in Table 1, plus the expected number of FRM cells sent during one round-trip time. This will allow us to tolerate the loss probability, ρ, shown in Table 1. The critical parameters for tolerating RM cell loss are:

(1) A reasonably large value of *Crm* (implying the need to have a reasonable amount of buffering in the network, to allow for a large initial (or transiently incorrect) rate to persist.

(2) A small value of CDF, so that the reduction caused by Rule 6 triggering is not substantial.

6. Conclusions

Protocol design has traditionally been an evolutionary process. First is the genesis of the algorithms and an informal specification of the algorithm using an English description. The protocol is subsequently refined as a result of additional scrutiny and performance analysis using simulations and other techniques. In the process, the protocol begins to accommodate complexity to address real-life situations where the initial design was inadequate. This has also been the evolution of the ATM Forum's ABR service description and congestion management protocol. Our study based on formal methods is useful to understand the correctness and performance issues of the resulting protocol.

Based on a careful analysis of the source rules in the ABR specification, we derived the conditions to ensure that the source's ACR is stable in the presence of delayed or lost feedback RM cells. We arrived at bounds on the number of consecutive RM cell losses tolerated while the ACR rate remains stable. In addition, we gave an asymptotic estimate of the value of ACR and the allowable RM cell loss probability to ensure that ACR is statistically stable.

While the details of our work are focused on the ABR service for ATM, we believe the learning here should be applicable in general for rate-based feedback control protocols. Recent work exploring the applicability of rate-based approaches to the Internet [16] may use the methods proposed for ABR to tolerate delayed or lost feedback from the destination. The analysis we provided in this paper indicates that we can make a rate-based flow control protocol stable under delayed or lost feedback information. However, the parameters for the reduction in the source rate during the time that there is no feedback have to be chosen carefully.

Uncited references

[3,12,14]

Acknowledgements

Melody Moh is supported in part by the NSF grant NCR-9714700, the SJSU Faculty Development Award, and a grant funded by Silicon Valley Research, Inc.

Appendix A

The correctness proofs in Section 4 use the formal specification in [13].

Proof of Theorem 4.1. Note that the source machine queues data and RM cells instantly to the scheduler machine whenever they are available. Therefore, delay and "lockup" of cells can only occur in the scheduler. We first examine the scheduler behavior. A routine proof shows that transitions from the active state S_1 are mutually exclusive and inclusive. □

Lemma A.1. *In the scheduler machine, at any moment one and exactly one of the transitions from the active state* S_1 *is executable. A similar statement also holds for the rate change state* S_2.

The above lemma shows that in the active state one and only one of the outgoing transitions is executable. It does not imply that the ABR protocol is livelock free; for that, we have to show that the transitions which transmit data cells and RM cells will always be executed within a finite amount of time. The following analysis proves this fact and also provides a bound on the time for the data and RM cells to be transmitted.

Since MCR $\leqslant$ ACR, the time interval in (4.1) for FRM cells can be easily derived from Proposition 3.1. On the other hand, the bound for BRM cells is provided in Lemma 3.2. Note that the bounds in (3.1). is obtained directly from the specification without any assumptions that the protocol is deadlock or livelock free.

We now analyze the waiting time for a data cell ready for transmission and that completes the proof Data cells are queued to scheduler immediately, and, therefore, the only delay occurs in the scheduler. There are three cases. All transitions refer to those in the scheduler machine in [13].

Case 1. *The last cell sent was an FRM cell.* The corresponding transitions executed were either T_1; or T_3 followed by T_4; or T_3 followed by T_5. In all the cases, we are back in state S_1 with $X_1 + X_2 = 0$. There are two subcases.

(A) $X = 0$. In this case, among all the transitions from S_1, only T_7 is executable, since event E is TRUE: there are data cells waiting for transmission, $X = 0$, and $X_1 + X_2 = 0 < Nrm - 1$. Therefore, T_7 is executed and the data cell waiting for transmission is sent. In this case, the delay is 1/ACR.

(B) $X > 0$. In this case, only T_6 is executable, since $X > 0$ and $X_1 = 0$. Therefore, a BRM cell is sent and we are still in state S_1 with $X_1 = 1$. Now by a similar argument among all the transitions from S_1, only T_7 is executable. Note that T_6 is no longer executable since $(X_1 = 0) \parallel\ !\ E = \text{FALSE}$. Therefore, T_7 is executed and the data cell is sent. The delay is 2/ACR.

Case 2. *The last cell sent was a BRM cell.* The corresponding transition for the transmission was T_6. After the execution of T_6, $X_1 > 0$ and T_6 is no longer executable since $(X_1 = 0) \parallel\ !\ E = \text{FALSE}$. From Lemma A.1, one and only one of the transitions from S_1 and also S_2 is executable and consequently, either an FRM cell or a data cell is sent as a result of the execution. If an FRM cell is sent, then from Case 1 the data cell will be sent in time no more than 3/ACR. Otherwise, the data cell is transmitted in time 1/ACR.

Case 3. *The last cell sent was a data cell.* The corresponding transition executed was T_7. Similar to Case 1, among all the transitions from S_1 and S_2, one and only one of them is executable. Therefore, either T_7 is executed next and in this case the data cell is sent, or an FRM cell or a BRM cell is sent. By the analysis in Cases 1 and 2, the data cell is sent in time no more than $1/\text{ACR} + 3/\text{ACR} = 4/\text{ACR} \leqslant 4/\text{MCR}$.

Proof of Theorem 4.2. Since RM cells and data cells are not duplicated, eventually they will all be processed and we will arrive at a situation where there are no data or RM cells in transmission. The only possible actions subsequently are: sending new FRM cells and corresponding BRM cells being turned around and transmitted. □

If there are no BRM cells waiting for transmission at both the end-stations then only FRM cells can be sent. After transmitting an FRM cell by Source Rule (3)(a) no more FRM cells can be sent until min $\{Mrm, Nrm - 1\} \geqslant 2$ data or BRM cells are transmitted before the next FRM cell can be sent. Therefore, the system goes to a sleep state.

If there are BRM cells waiting for transmission then each FRM cell is sent after at least min $\{Mrm, Nrm - 1\} \geqslant 2$ BRM cells are transmitted according to Source Rules (3)(a) and (b). Since RM cells in transmission are not duplicated, the total number of RM cells flowing on the VC is strictly monotonically decreasing as FRM cells are sent and received at the remote station. Eventually, both end-stations are in the following situation: each of them sends an FRM cell followed by n BRM cells where $n < \min\{Mrm, Nrm - 1\}$. Eventually, there are no RM cells in transmission on the VC. Neither end-station can proceed; there are no FRM cells in transmission that are to be turned around and neither end-station can send FRM cells by Source Rule (3)(a). We reach a sleep state.

If the conditions $Mrm \geqslant 2$ and $Nrm \geqslant 3$ are not satisfied, it is obvious from Source Rule (3)(a) that each end-system is in a state that sends an FRM cell after turning around a BRM cell from the other station and *Trm* time has elapsed, since $Mrm = 1$ or $Nrm - 1 = 1$. It is a busy-wait state.

Proof of Theorem 4.3. Upon receiving BRM cells, the source adjusts ACR in state S_2 of the source machine. The four transitions involving rate changes are T_4, T_5, T_6, and T_7; they all move from state S_2 to S_1. These four rate-change transitions are mutually exclusive and inclusive.

Lemma A.2. *In the source machine and on arrival of a BRM cell, one and only one rate-change transition is executable from the rate-change state* S_2.

We use the above results to discuss the interoperation of EFCI and ER in determining the rate ACR, and that completes the proof. □

The EFCI scheme indicates congestion by setting the CI bit; ER scheme indicates congestion by setting ER such that $\mathrm{ER} \leqslant \mathrm{ACR}$. Obviously, there are three cases.

Case 1. *Both EFCI and ER indicate a congestion.* In this case, $(\mathrm{CI} = 1)\&\&(\mathrm{ER_{BRM}} \leqslant \mathrm{ACR})$. Source tranistion T_4 [13] is enabled, and by Lemma A.2 it is the one and only transition enabled; the result is the following rate change:

$$\mathrm{ACR} := \max\{\mathrm{MCR}, \min\{\mathrm{ACR} \cdot (1 - \mathrm{RDF}), \mathrm{ER_{BRM}}\}\}.$$

Thus the final rate is the minimum of the two rates, determined by the two schemes, respectively.

Case 2. *Both EFCI and ER indicate no congestion.* In this case, $(\mathrm{CI} = 0)\&\&(\mathrm{ACR} < \mathrm{ER_{BRM}})$, we consider two sub-cases:

(A) $\mathrm{NI} = 0$. In this case, $(\mathrm{CI} = 0)\&\&(\mathrm{NI} = 0)\&\&(\mathrm{ACR} < \mathrm{ER_{BRM}})$, source T_5 is the one and only transition enabled; this results in the following rate change:

$$\mathrm{ACR} := \min\{\mathrm{ER_{BRM}}, \mathrm{PCR}, \mathrm{ACR} + \mathrm{RIF} \cdot \mathrm{PCR}\}.$$

Therefore the final rate increase is bounded by the minimum of the increases allowed by the two options.

(B) $\mathrm{NI} = 1$. In this case, $(\mathrm{CI} = 0)\&\&(\mathrm{NI} = 1)\&\&(\mathrm{ACR} < \mathrm{ER_{BRM}})$, source T_6 is the one and only transition enabled; the result is that there is no rate increase.

Case 3. *One indicates congestion and the other indicates no congestion.* Consider two sub-cases:

(A) *EFCI indicates congestion but ER indicates no congestion.* In this case, $(\mathrm{CI} = 1)\&\&(\mathrm{ER_{BRM}} > \mathrm{ACR})$, source T_4 is the one and only transition enabled, the result is the following rate change:

$$\mathrm{ACR} := \max\{\mathrm{MCR}, \min\{\mathrm{ACR} \cdot (1 - \mathrm{RDF}), \mathrm{ER_{BRM}}\}\}.$$

Hence the outcome is the rate reduction due to EFCI scheme since $\mathrm{ER_{BRM}} > \mathrm{ACR}$.

(B) *EFCI indicates no congestion but ER indicates congestion.* In this case, $\mathrm{CI} = 0\&\&\mathrm{ACR} \geqslant \mathrm{ER_{BRM}}$, source T_7 is the one and only transition enabled, this results in the following new rate:

$$\mathrm{ACR} := \max\{\mathrm{MCR}, \mathrm{ER_{BRM}}\}.$$

Thus the outcome is that there is a rate reduction due to the ER scheme.

References

[1] ATM Forum Traffic Management Specification Version 4.0, af-tm-0056.00, ATM Forum, April 1996, available as: ftp://ftp.atmforum.com/pub/approved-specs/af-tm-0056.000.pdf.

[2] F. Bonomi, K.W. Fendick, The rate-based flow control framework for the available bit rate ATM service, IEEE Network 9 (2) (1995) 25–39.

[3] Y. Chang, N. Golmie, D. Su, Study of interoperability between EFCI and ER switch mechanisms for ABR traffic in an ATM network, in: Proceedings of the Fourth International Conference on Computer Communications and Networks, 1995, pp. 310–315.

[4] A. Charny, D.D. Clark, R. Jain, Congestion control with explicit rate indication, in: Proceedings of ICC, June 1995.

[5] A. Charny, K.K. Ramakrishnan, Time scale analysis of explicit rate allocation in ATM networks, in: Proceedings of Infocom '96, March 1996.

[6] G.J. Holzmann, Design and Validation of Computer Protocols, Prentice-Hall, Englewood Cliffs, NJ, 1991.

[7] R. Jain, Congestion control and traffic management in ATM networks: recent advances and a survey, Computer Networks and ISDN Systems 28 (13) (1996) 1723–1738.

[8] R. Jain, S. Kalyanaraman, S. Fahmy, F. Lu, Parameter Values for Satellite Links, ATM Forum Contribution ATM Forum/95–972, August 1995.

[9] L. Kalampoukas, A. Varma, K.K. Ramakrishnan, An efficient rate allocation algorithm for ATM networks

providing max–min fairness, in: Proceedings of Sixth IFIP International Conference on High Performance Networking, HPN'95, Spain, 11–15 September 1995.

[10] S. Keshav, A control-theoretic approach to flow control, in: Proceedings of ACM Sigcomm'91, September 1991.

[11] R.P. Kurshan, Computer-Aided Verification of Coordinating Processes, Princeton University Press, Princeton, NJ, 1995.

[12] D. Lee, M. Yannakakis, Principle and methods of testing finite state machines – a survey, Proceedings of the IEEE 84 (8) (1996).

[13] D. Lee, K.K. Ramakrishnan, W.M. Moh, A formal specification of the ATM ABR rate based control scheme, Comput. Networks ISDN Syst. 30 (1998) 1735–1748.

[14] D. Lee, K.K. Ramakrishnan, W.M. Moh, A.U. Shankar, Performance and correctness of the ATM ABR rate control scheme, in: Proceedings of IEEE INFOCOM'97, April 1997.

[15] K.K. Ramakrishnan, R. Jain, A binary feedback scheme for congestion avoidance in computer networks, ACM Trans. Comput. Syst. (1990).

[16] R. Rejaie, M. Handley, D. Estrin, An end-to-end rate-based congestion control mechanism for realtime streams in the Internet, in: Proceedings of IEEE INFOCOM'99, New York, March 1999.

[17] L. Roberts, Enhanced PRCA (Proportional Rate-Control Algorithm), ATM Forum Contribution, AF-TM 94-0735R1, August 1994.

David Lee (IEEE M'89-SM'91-F'96) was born in Hong Kong and grew up in Soozhow, China. He received a Master of Arts degree in Mathematics from Hunter College of City University of New York in 1982 and a Ph.D. degree in Computer Science from Columbia University in 1985. Since 1985 he has been with Bell Laboratories Research. He is the Director of Bell Laboratories Research, China, and was Head of Networking Research Department. He has been an adjunct professor at Columbia University, a Visiting Faculty of National Institute of Standards and Technology, and a Visiting Professor of Institut National des Telecommunications, France. His current research interests are data networking, communication protocols and software, and complexity theory. He is a Fellow of The IEEE and has the Bell Laboratories Research Distinguished Member of Technical Staff Award in 1996. He is an editor of IEEE/ACM Transactions on Networking and an associate editor of Journal of Complexity.

K.K. Ramakrishnan is a Founder and Vice President at Tera-Optic Networks. Until recently, he was a Technology Leader at AT&T Labs Research. He holds a BS in Electrical Engineering from Bangalore University in India in 1976, an MS in Automation from the Indian Institute of Science in 1978 and a Ph.D. in Computer Science from the University of Maryland in 1983. From 1983 to 1994, he was with Digital Equipment Corporation. He is an editor for the IEEE/ACM Transactions on Networking and IEEE Network. His research interests are in design and performance of computer and communication network protocols and algorithms.