```
**************************************************************************

ATM Forum Document Number: ATM Forum/94-1175R1

**************************************************************************

Title:  Current Default Proposal: Unresolved Issues

**************************************************************************
```

Abstract:
Problems with the default baseline text as circulated after the
October interim meeting are pointed out.

```
**************************************************************************
```

Source:

Raj Jain, Shiv Kalyanaraman and Ram Viswanathan

Raj Jain is now at Washington University in Saint Louis,
   jain@cse.wustl.edu http://www.cse.wustl.edu/~jain/   Jain@ACM.Org

The presentation of this contribution at the ATM Forum is sponsored by
NIST.

```
**************************************************************************

Date:   November 29-Dec 2, 1994

**************************************************************************
```

Distribution: ATM Forum Technical Working Group Members
             (Traffic Management)

```
**************************************************************************
```

```
**************************************************************************
```

The current default proposal for source/switch/destination
behavior has several problems as discussed below.

A. SOURCE BEHAVIOR DURING IDLE:

The current text (as circulated after the October Interim
meeting) for Section 5.4.3 Source Behavior paragraph 3 states:

"3. When a idle source starts transmitting, the ACR should be
decreased at least by ACR/RDF for each Nrm cell times which have
passed, down to ICR (a linear decrease)."

This is an unjustified complication. It makes both source
implementation and "policing" very difficult. Such a restriction
could have resulted from the assumption that the time to reach
the correct operating point is too high and, therefore, if the
source starts at the wrong value, it would cause a havoc on the

network. But this argument is flawed.

First, a scheme, which takes a long time to settle should be fixed so that the transient times are as small as possible. This would automatically solve the problem since the source will be out-of the step just for one round trip.

Second, the choice of *linear* decrease is arbitrary. This decrease may be too slow or too fast. Any arbitrary decision at the source without explicit feedback from the network will have this problem. Staying at the same ACR will be as wrong (or as right) as the linear decrease.

Both the load and the available bandwidth are random variables. The state (load and bandwidth) has to be continuously monitored. Memorizing previous state is not very helpful. Predicting based on the previous state may be even less helpful. In this case, starting at ACR before going idle corresponds to memorizing the previous state. While linear decrease corresponds to prediction based on it.

Due to a mismatch of the predicted state and the current state, the bandwidth, even if available, may be left unused or long queues may be formed. The correct solution is to minimize the transient interval so that the source can start at any arbitrary cell rate and come to the right operating point within one round trip.

We suggest that the text be modified as follows:

"3. When an idle source starts transmitting, it could start at the previous ACR, ICR, or any rate in between." [Motion 1]

This leaves the choice of decreasing during idle to the NIC implementors. In particular, if they feel that they should go down every Nrm cell, they can do so. The policing is also easy since it will be comparing current cell rate with the previous ACR at all times.

B. SOURCE BEHAVIOR DURING ACTIVE PERIOD:

The current text for Section 5.4.3 Source Behavior paragraph 4 states:

"4. An active source should decrease its ACR by at least ACR/RDF every Nrm cell times, down to MCR (an exponential decrease)."

The perceived logic for this requirement is probably to help during extreme congestion when the RM cells are getting lost or held up excessively. However, this particular solution violates the general principle of protocol design that the normal operation should be slowed down as little as possible by actions required to take care of abnormal events. During normal operation, loss of RM cells should be a rare event. On the other hand, RM cells sent every Nrm cells (at a particular ACR) may not be received every Nrm cells at a higher ACR. Whenever the sources are ramping up, the above rule will be triggered unnecessarily even if there is no congestion.

Also, in WAN cases, where the first RM cell will take quite a while to return, the above rule will cause sources to decrease unnecessarily.

The problem is caused by the fact that we are trying to use one parameter where two are required: the inter-RM cell sending time

and the RM-cell receiving timeout interval. The first is
determined by Nrm and ACR. While the second should be several
times that. Using the timeout value equal to (or close to) the
sending interval causes the probability of false alarms to be
100%. The normal operation for the source ramping up is to reduce
the rate and immediately go up after a cell or two. This is an
oscillation that we can live without.

The solution is to take the recovery action only when it is
somewhat likely that the problem has happened. Thus, if a source
has not received an RM cell for two, three, or k times the
sending interval (k*Nrm cells), it should assume that the RM cell
has been lost and decrease. Here, k is a parameter. Fixing k for
ever at one is not justified.

Thus, the proposed fix is to replace the text as follows:

"4. If an RM cell is not received in k*Nrm cell times, the source
should decrease its ACR by at least ACR/RDF down to MCR (an
exponential decrease). Here k is a parameter negotiated at
connection setup. No decrease is required if an RM cell is
received." [Motion 2]

C. SOURCE BEHAVIOR DURING NO CONGESTION:

This section relates to the paragraph 5 of Section 5.4.3 Source
Behavior:

"5. Only when a backward RM cell is received with CI=0, may the
source increase the ACR by an amount, AIR, negotiated at call
setup and restores any previous decrease since receipt of the
previous RM cell."

AIR is useful only if there are EFCI switches. If there are no
EFCI switches, AIR unnecessarily limits the increase even if the
network has specified a higher explicit rate in the ER field.
This hurts the transient performance. One solution to this
problem is to allow ER-based switches the ability to leave AIR at
PCR at the time of connection setup. Thereby, nullifying the
effect of this parameter. EFCI-based switches, could set it at
some fraction of PCR. Of course, the minimum of that allowed by
all switches is what will be eventually passed on to the source.

Although the current text does not disallow this, the possibility
that AIR can be set at PCR should be explicitly stated since that
would be the normal mode of operation eventually.

Thus, we suggest that the following sentence be added to the
paragraph 5:

"Setting AIR at PCR at the connection setup eliminates the effect
of this parameter and is allowed." [Motion 3]

D. SWITCH BEHAVIOR:

The text in Section 5.4.5 Switch Behavior is:

"b. The Switch queuing point can optionally set CI=1 in the
backward RM cells to ensure the source does not increase its
rate."

and

"b. The Switch queuing point may optionally set CI=1 in the
backward RM cells to ensure the source does not increase its

rate, in addition to modifying the ER field in the backward RM cell to a lower value."

It is not clear why the switches could not do this in the forward direction. The choice of forward or backward direction should be left to the switches.

At least in the case of BECNs or switch generated RM cells, it has been pointed out [1, 2] that the switches would modify CI and ER in the forward direction.

Secondly (and more importantly), the queue state at the time of the RM cell arrival in the forward direction is related to the CCR field in the RM cell. By the time the cell returns in the backward direction, the VC's cell rate may have changed and the queue state has no relation to the CCR field.

Those schemes that use CCR field of the RM cell in determining the feedback would perform better if the state and feedback relationship is maintained. It must be pointed out that the EPRCA as described in 94-735R2 is one such scheme. The MACR is calculated based on CCR and is used to determine ER along with the queue state.

The conclusion is that the switches should be able to modify the ER field in the forward, backward, or both directions. A general advantage of the explicit rate approach is that the switches have considerable flexibility in their operation. If this flexibility has to be maintained then every single restriction on the switches should be carefully justified.

We suggest that the phrase "backward RM cell" be replaced by "RM Cell" resulting in the following corrected text:

"b. The Switch queuing point can optionally set CI=1 in the RM cells to ensure the source does not increase its rate."

"b. The Switch queuing point may optionally set CI=1 in the RM cells to ensure the source does not increase its rate, in addition to modifying the ER field in the RM cell to a lower value." [Motion 4]

REFERENCES:

[1] Raj Jain, Shiv Kalyanaraman and Ram Viswanathan, "Ordered BECN: Why we need a timestamp or sequence number in the RM Cell" ATM Forum Traffic Management Group Interim Meeting Contribution 94-0987, October 25-26, 1994.

[2] Andy Barnhart, "Proposal for Switch-Generated RM Cells for Explicit Rate," ATM Forum Traffic Management Group Meeting Contribution 94-1112, November 28 - December 2, 1994.