95-1660R1

A Fix for Source End System Rule 5

Raj Jain, Shiv Kalyanaraman, Rohit Goyal, Sonia Fahmy, Fang Lu

The Ohio State University

Saragur M. Srinidhi

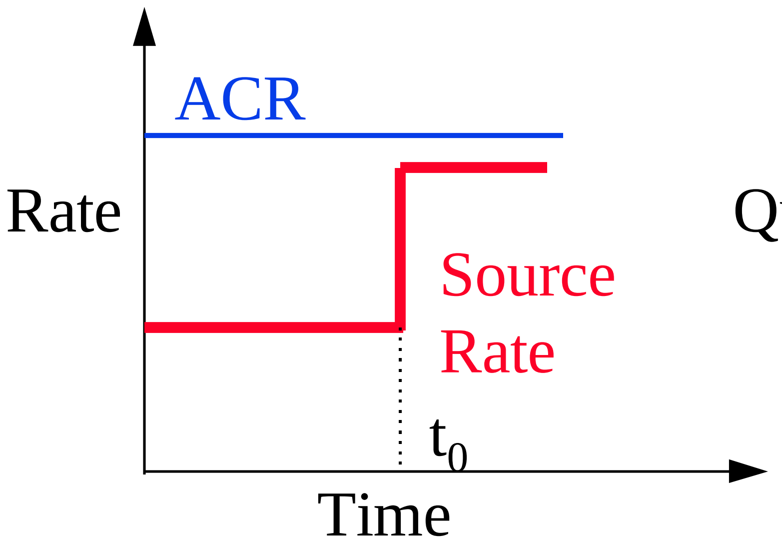
Sterling Software and NASA Lewis Research Center

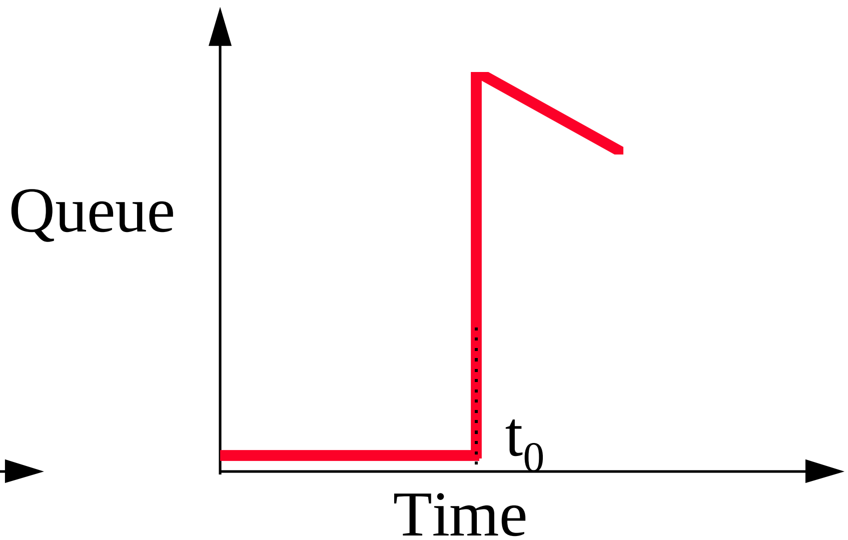
Raj Jain is now at Washington University in Saint Louis,
jain@cse.wustl.edu **http://www.cse.wustl.edu/~jain/**

- Do we really need rule 5?
- Key Design Issues
  - Multiplicative vs additive
  - Rule 5b
- Our proposal
- Comparison with other four proposals
- Parameter selection guidelines for our proposal

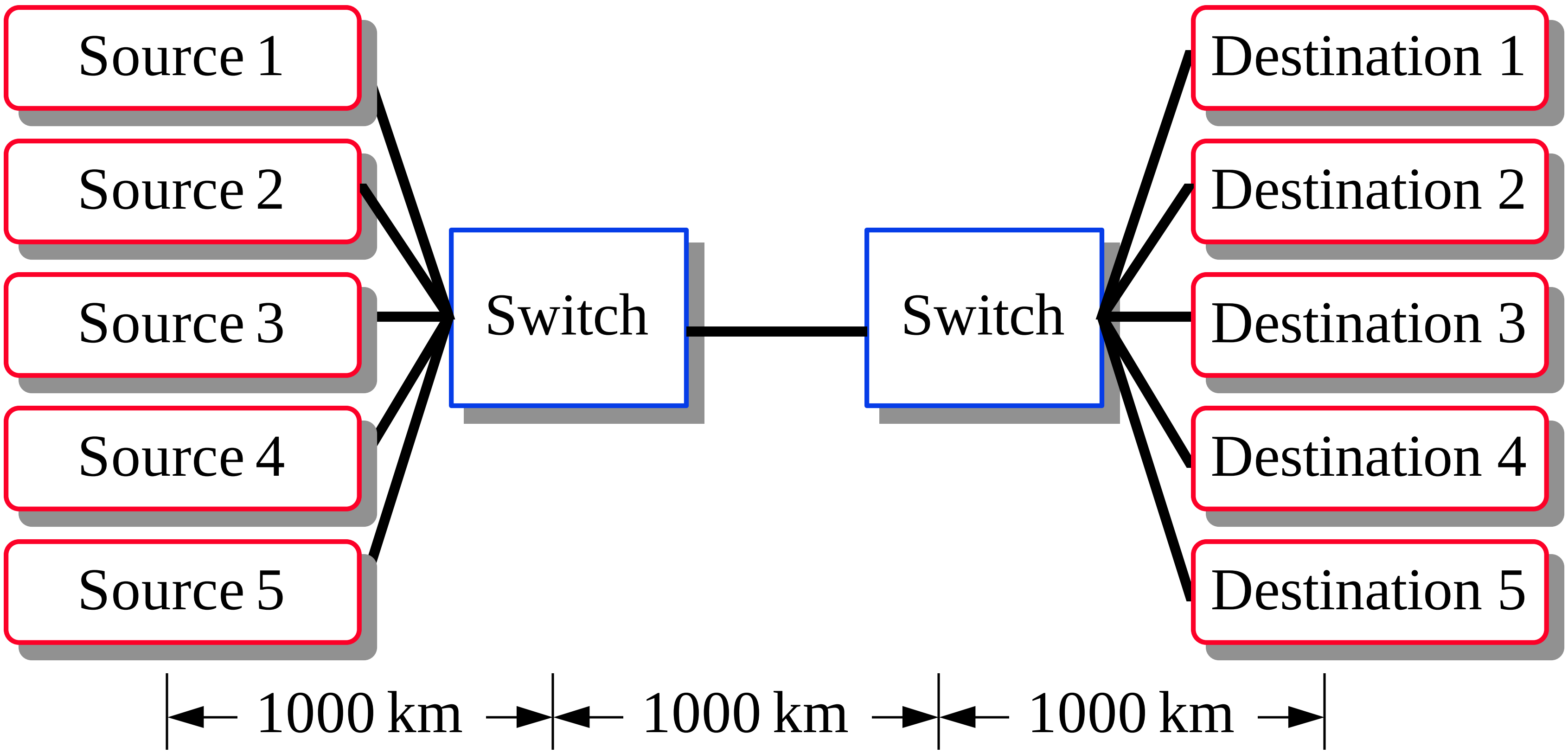
Do We Really Need Rule 5?

- Is ACR Retention a real problem?
- **Answer: It depends!**
- ACR retention ⇒ ACR >> Source Rate
  ⇒ Source rate can go high <u>any moment</u>
  ⇒ Long Queue length if Σ ACR > Available Bandwidth





Five Source Configuration



- All links 155 Mbps
- All source bottlenecked at 10 Mbps upto 200 ms
  At t = 200 ms, the sources are able to use their full ACR

Simulation Parameters

- Source: Parameters selected to maximize ACR
  ICR = {155.2 Mbps, 1 Mbps}
  TBE = 4096 $\Rightarrow$ Rule 6 disabled
  CRM (Xrm) = Min{TBE/Nrm, PCR $\times$ FRTT/Nrm}
  TDF = {0, 1/8}
  PNI = {0, 1}
  TOF = 2
  PCR = 155.52 Mbps, MCR= 0, RIF (AIR) = 1, Nrm = 32,
  Mrm = 2, RDF = 1/512, Trm = 100ms, CDF (XDF) = 0.5,
  TCR = 10 c/s
- Traffic: Bi-directional, infinite. Source bottlenecked initially.
- Switch: ERICA modified
  Target Utilization = 90%
  Averaging interval = min{30 cells, 200 μs}

Conclusion 1

- ACR Retention can cause sudden queue growth of (ACR-Source Rate)$\times$Feedback delay $\times$(Number of Sources-1)
- Some form of Rule 5 is required
- ACR Retention even for a small interval
  $\Rightarrow$ Switches are exposed to "sudden arrivals"
- Rule 5 proposals that allow ACR retention for some time are vulnerable to such "sudden arrivals" during those times
- VCs that disable Rule 5 (e.g., by setting ICR = PCR) can cause such "sudden arrivals"
- On LANs: Feedback delay is lower than Inter-RM time
  Network feedback arrives faster than source sending FRMs
  $\Rightarrow$ Rule 5 is not required on LANs

Rule 5 Proposals

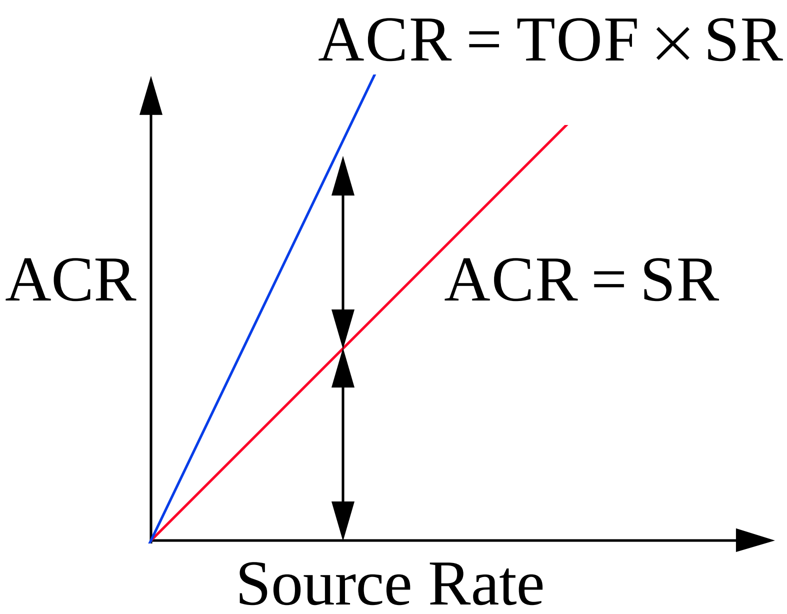
- No Rule 5
- Baseline: As in ATM Forum 95-0013/R9
  - Multiplicative headroom
  - ACR = Max{ICR, ACR $\times$ T $\times$ (1-TDF)}
- October: As proposed by Barnhart and Jain et al
  - Additive headroom
  - ACR = Max{ICR, ACR $\times$ (1-TDF)}
- AF-TM 1614
  - Additive headroom
  - ACR = Max{SR+ICR, ACR $\times$ (1-T/Tc)}
- This Proposal: Modification of our "October" proposal

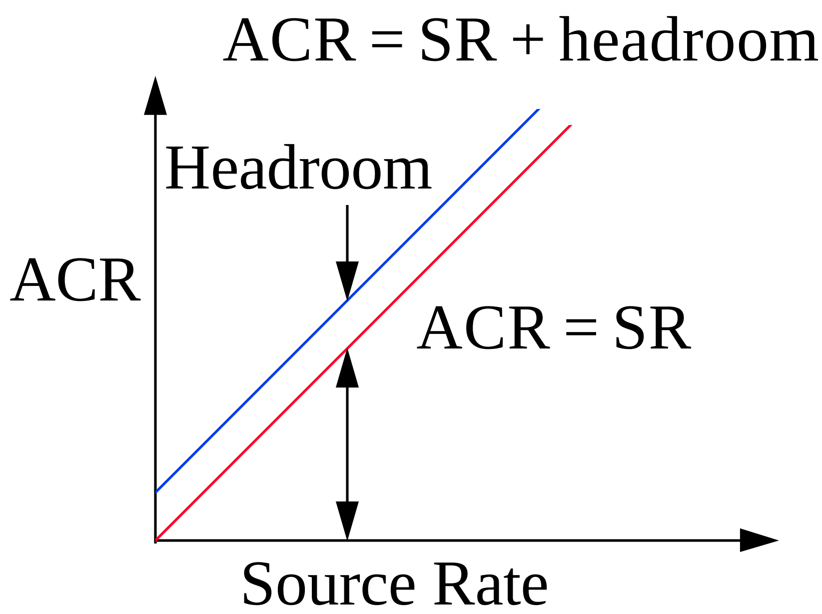
Rule 5 Design Issues

- Additive headroom vs multiplicative headroom
- Action on BRM:
  How long should the feedback be ignored after an adjustment?
  - Do not ignore (as obtained by PNI = 1)
  - Ignore once (as in baseline and other proposals)
  - Ignore as long as there is any ACR retention
- Floor of Reduction:
  - ICR
  - Source Rate + Headroom
- Decrease proportional to T vs fixed decrease
- Do we need two parameters? ICR and headroom

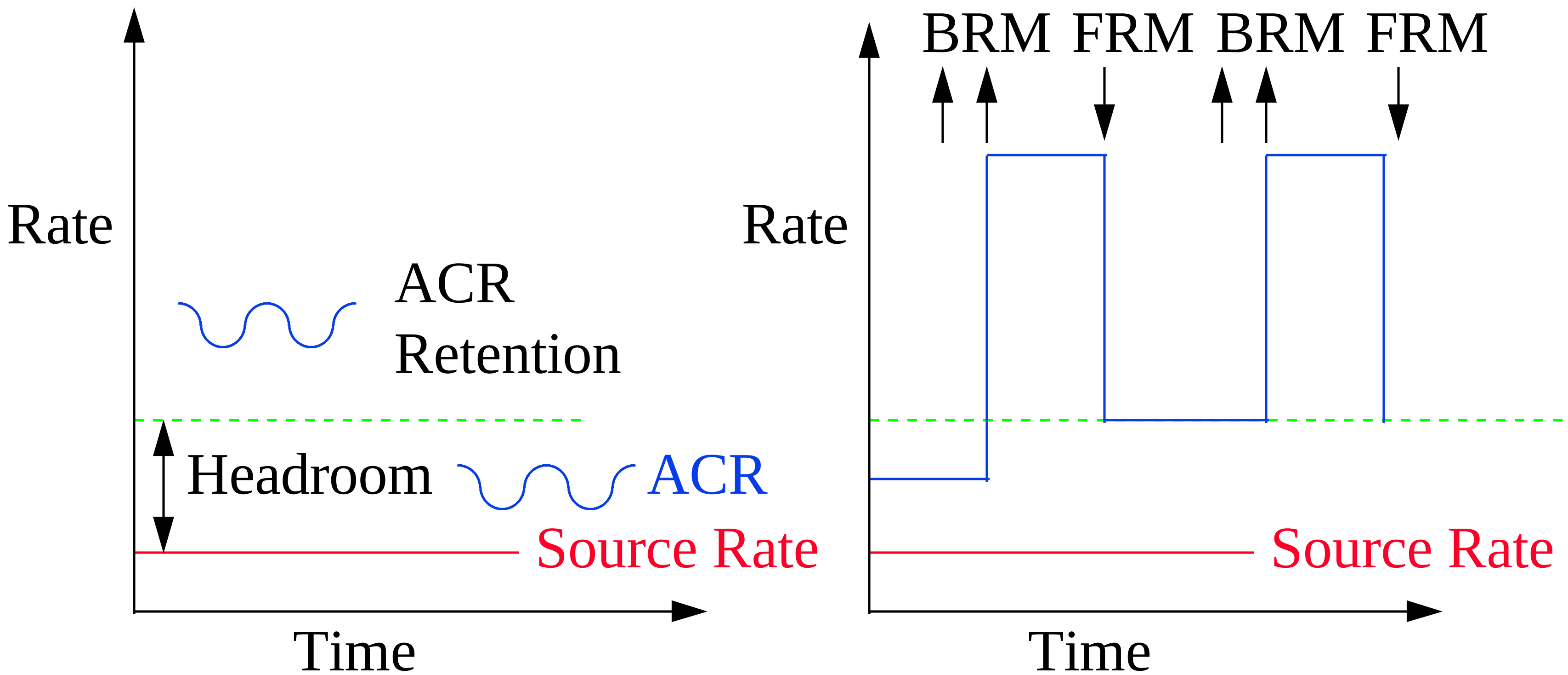
Multiplicative vs Additive Headroom

- Multiplicative headroom $\Rightarrow$ Large (ACR-SR) for some sources $\Rightarrow$ Large queue growth possible
- Additive headroom $\Rightarrow$ ACR-SR same regardless of SR
- Queue growth = (ACR-SR) $\times$ Feedback delay $\times$ # of sources
  $\Rightarrow$ Additive headroom provides a better protection



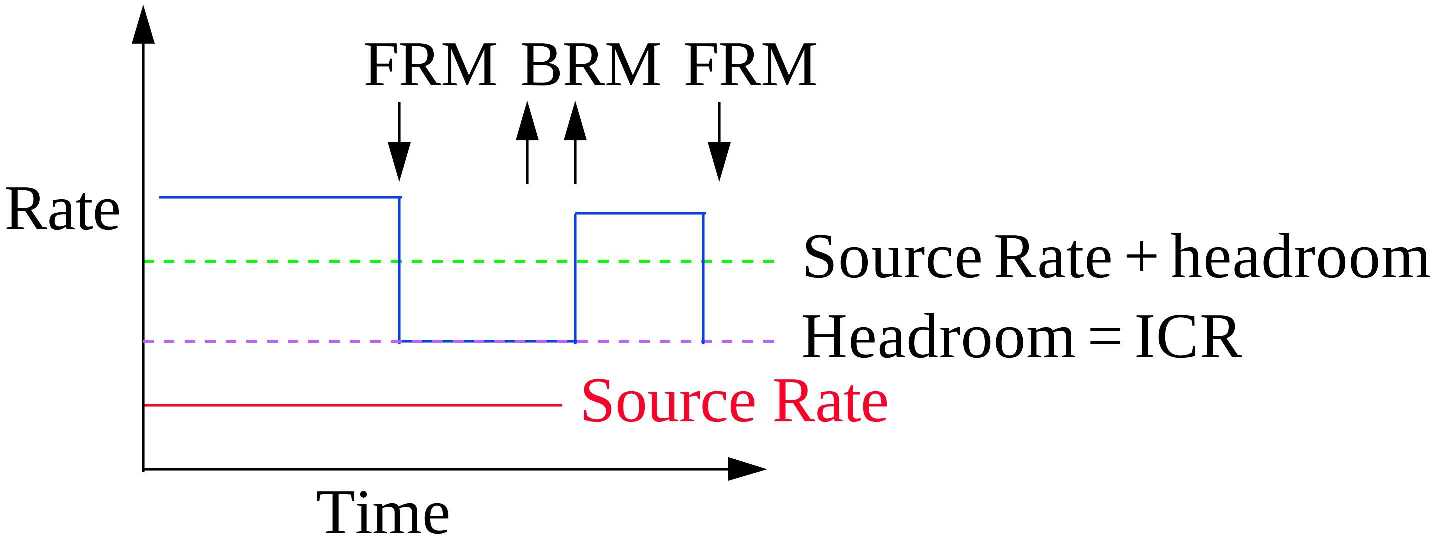


Action on BRM



- Should the explicit feedback increases be ignored?
  - Not ignoring will cause oscillations
  - Ignoring once still causes oscillations
  - Ignore as long as source has not used its previous ACR

Floor of ACR



- ACR = Max{ICR, ACR(1-TDF)}
- For large TDF, ACR may be reduced below Source rate+headroom
- **Conclusion**: Do not go below source rate + headroom
  ACR = Max{Source Rate + headroom, ACR(1-TDF)}

Our Proposal: Pseudocode

- **At FRM Send event:**

```
SR = Nrm/T;
ACR_ok = ((ACR <= SR)||(TDF == 0.0));
IF (PR5 == False)
        IF (ACR > SR + headroom)
                ACR = Max(SR + headroom, ACR × (1.0 - TDF));
        ENDIF
ELSE
        PR5 = False;
```

- **At BRM Receive Event:**

```
IF (NI = 0 AND ACR_ok )
        IF (ACR < ER) PR5 = True ELSE PR5=False;
        ACR = Min(ACR + AIR × PCR, PCR);
ENDIF
ACR = Min(ACR, ER);
ACR = Max(ACR, MCR);
```

Initialization:

```
        ACR_ok = True;
        PR5 = False;
```

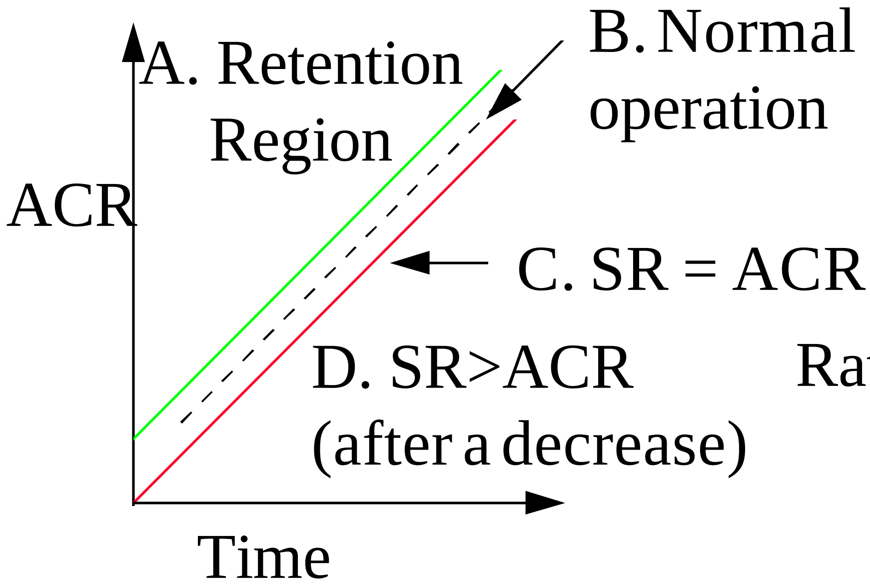
ACR_ok $\Rightarrow$ VC has used its ACR

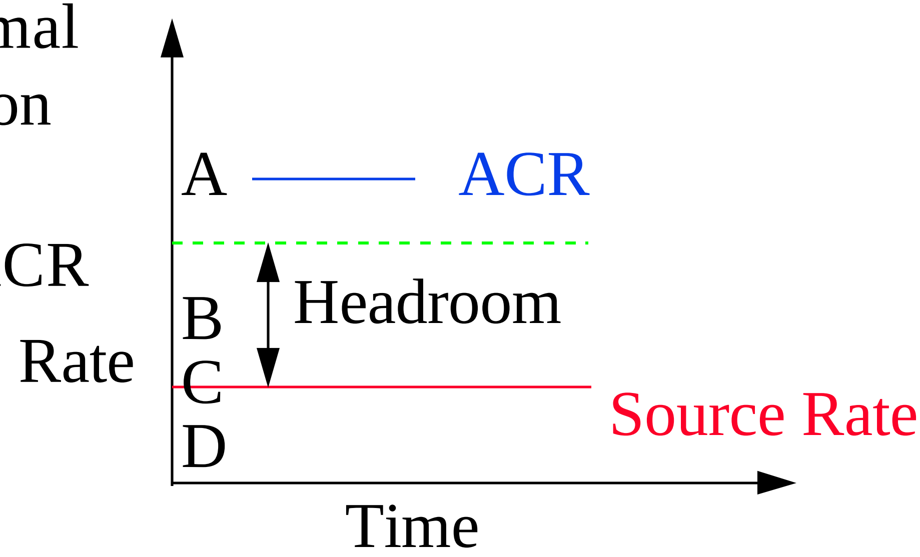
PR5 $\Rightarrow$ Network directed increase

Key Features

- SR = Source Rate is a temporary variable. It is not stored between successive execution of the code.
- ACR formula does not allow decrease below ACR + headroom
- ACR_ok ⇒ Source Rate is at or above ACR
  ⇒ No ACR Retention
- PR5 is set to "true" on network directred increase and set to "false" on all other BRMs
- PR5 = Prohibit next Rule 5 decrease
  ⇒ Ignore rule 5 once on next FRM
  Named similar to "Prohibit Next Increase (PNI)"
- Source has to use its current allocation before rising (ACR_ok ⇒ Source has used its previous allocation)

Regions of Operation





	Trigger Rule 5	**Increase On BRM**	**Decrease On BRM**
A	Yes unless PR5	No	Yes
B	No	No	Yes
C	No	Yes	Yes
D	No	Yes	Yes

Simulation Results: Source Bottleneck Case

- Queue lengths are large if there is "no rule 5" or if rule 5 is disabled due to high ICR
- Ignoring feedback just once causes oscillations. Network is susceptible to large queues during oscillations. Our proposal eliminates such oscillations.
- Oscillations caused by low ICR in Baseline are eliminated in our proposal.
- Low headroom and high ICR is feasible with our proposal.
- Time to converge depends upon:
  - TDF (Large TDF $\Rightarrow$ Faster convergence)
  - Inter-FRM cell time
- Continuous flow of BRMs $\Rightarrow$ No protection due to Rule 6

Other Cases

- Normal Rise
- Short and long distance VCs sharing the same switch
- Bursty sources
  - Metric = Number of bursts transmitted per unit time

Bursty Sources

- Queue is not a problem for small/medium bursts
- Any attempt to reduce/eliminate ACR retention will reduce burst throughput
  $\Rightarrow$ Bursty sources want to retain ACR until they need it
- With our scheme, bursty sources are eventually allocated ACR = average source rate + headroom
  $\Rightarrow$ headroom should be high for bursty sources
  headroom should be as low as possible for infinite sources
- Possible solutions for small bursts:
  - Use GCRA type of burst tolerance mechanism to allow small bursts at link rate
  - Use a small TDF
  - Use larger headroom

Decrease α T vs Fixed

- For LANs: Tc is small $\Rightarrow$ 1/Tc is large $\Rightarrow$ Decrease is too large
- For WANs: Tc is large $\Rightarrow$ 1/Tc is small $\Rightarrow$ Decrease is small.
- On WANs, deasonable decreases obtained only if T (time between FRMs and idle time) is large. Does not protect against source bottleneck case (no idle time).
  More opportunity for ACR retention.
- The issue is that of decreasing every ΔT vs every N cells
- Tc depends upon round trip time. The feedback delay (round-trip between source and the bottleneck) affects the performance more than round-trip time.

ICR Selection Guidelines

- Used just once on connection setup
- Equivalent to ER in the first RM cell
  but available right after the connection is set up
  As if a BRM was tagged to "Connect-Confirm" message.
- For switches, ICR is a short term decision like any other ACR
- ICR can be high or low depending upon current congestion

Headroom Selection Guidelines

- Long term decision
  Applies throughout the life of the VC
  The duration could be several years (for PVCs)
- Controls how much the sources can lie at any time
- Determines how many cells you may receive at once
- Must be as low as possible
- Too low headroom is not good for bursty sources
- Recommended value = 10 Mbps
  Allows LANE traffic at full Ethernet speed
  Use smaller values for WANs
- Sources, which have been idle for long, will send an RM cell $\Rightarrow$ Rule 5 will be triggered $\Rightarrow$ Will start at headroom

TDF Selection Guidelines

- Determines the speed of convergence
  Determines the duration for which network is susceptible to burst arrival due to ACR Retention
- Larger value $\Rightarrow$ Faster convergence
- Should be as high as possible
- Low value preferred for bursty sources
- TDF = 0.0 disables rule 5
- Recommended value = 1/8

Key Features of Our Proposal

- No oscillations even when ICR is low
- No oscillations during normal operation
- Less parameters: No TOF or PNI or Tc
- Separates out the role of ICR and headroom
- One parameter TDF can enable/disable the scheme
- Parameters are easy to select and negotiate

Motion

- Change pseudocode on page 74 section I.1 of R9 as follows and update text and flow chart accordingly
- **At FRM Send event:**

```
SR = Nrm/T;
ACR_ok = ((ACR <= SR)||(TDF == 0.0));
IF (PR5 == False)
    IF (ACR > SR + headroom)
        ACR = Max(SR + headroom, ACR × (1.0 - TDF));
    ENDIF
ELSE PR5 = False;
```

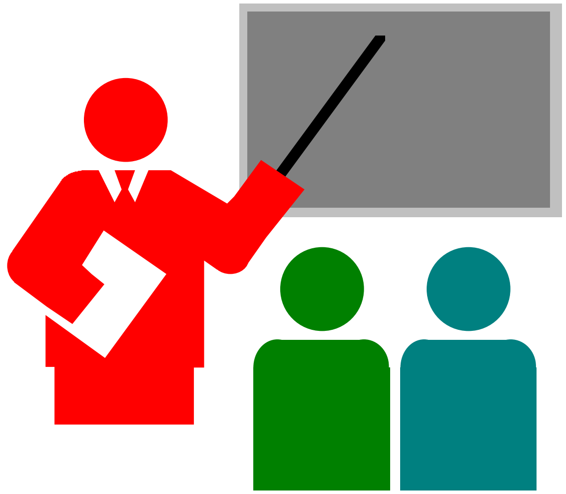
- **At BRM Receive Event:**

```
IF (NI = 0 AND ACR_ok )
    IF (ACR < ER) PR5 = True ELSE PR5=False;
    ACR = Min(ACR + AIR × PCR, PCR);
ENDIF
ACR = Min(ACR, ER);
ACR = Max(ACR, MCR);
```

Initialization:

```
    ACR_ok = True;
    PR5 = False;
```

Summary



- Rule 5 is required for large bursts or ACR retaining infinite sources on WANs
- ACR retention, even for a short duration, can be dangerous.
- ICR and headroom have different roles.
  Headroom is a long term commitment and should be allocated conservatively.