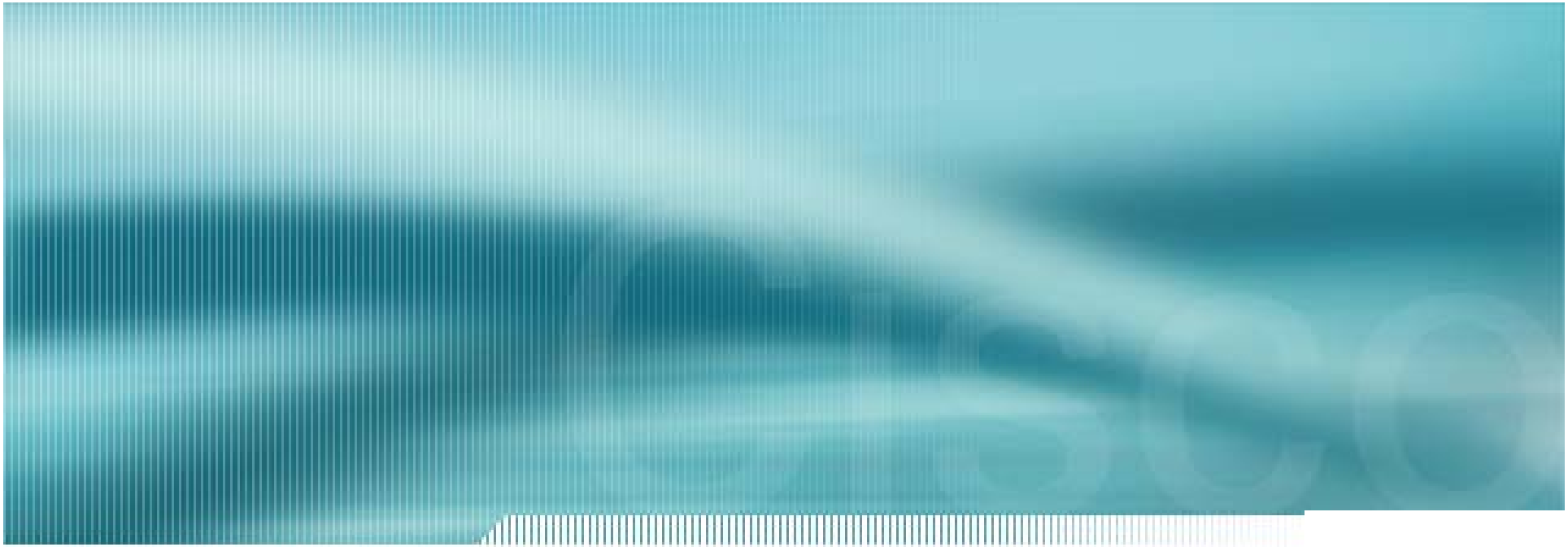


Rate control for Ethernet congestion management II

Hugh Barrass (Cisco Systems)

Agenda

- **Recap : changes proposed**
- **Possible optimization for frame rate definition**
- **Interaction with CRS and deference**
- **Remote rate control request**
- **Conclusions and proposals**

3 types of rate control – refresher…

Rate control will fix a link at a reduced rate

- **Bits transmitted at the same rate; packet rates reduced**
- **There are 3 types of rate control to define**

a) Constant (per packet) overhead

- **Effectively increases min IPG**

b) Limited (payload) bit rate

- **Dependant on packet length (as in 802.3ae)**

c) Limited packet rate

- **Counts the number of SFD's per second**

Changes to Clause 4A

- Changes are only considered for full duplex
  - … edit Clause 4A – echo in Clause 4 at WG ballot
- 4A.2.3 Frame transmission model
  - Includes 4A.2.3.2.2 interframe spacing
- Propose to use similar method to 802.3ae
  - Look in Clause 4 for details
  - Redefine interFrameSpacing into 2 parts
  - Add specification for 2nd part of IFS
  - Needs additional state & counter for frame rate method
- Change details are in barrass_1_0503

Management

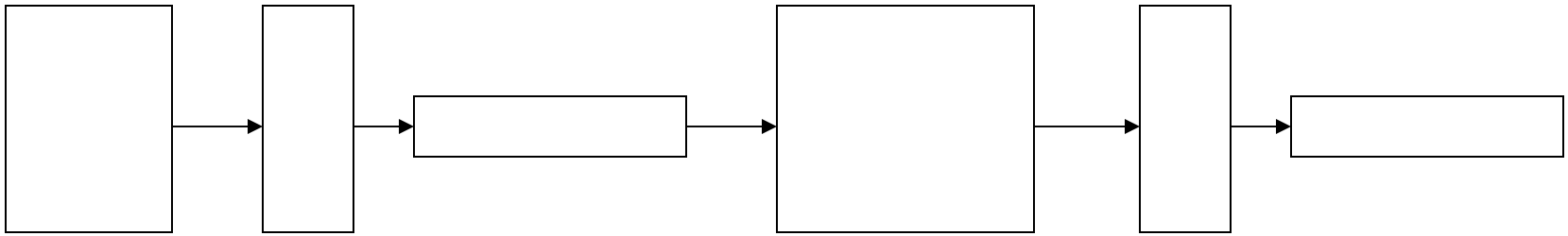
- Clause 30 definitions are needed for the following variables:
  - rateLimitPacketOverheadEnable, rateLimitPayloadRateEnable, rateLimitFrameRateEnable {enables for the rate limiter features}
  - additionalPacketOverhead {for fixed packet overhead}
  - ifsStretchRatio {for fixed payload rate limit}
  - frameRateControlStart {for frame rate limit}
- Additionally, objects are needed for local receive capabilities
- Analogous to the 6 objects for transmit rate limits
  - 3 boolean objects to indicate a limitation
  - 3 parameters for the 3 types of rate limit

Agenda

- Recap : changes proposed
- Possible optimization for frame rate definition
- Interaction with CRS and deference
- Remote rate control request
- Conclusions and proposals

Per frame rate limiting

- The definition proposed limits the packet rate by adding to each ifs (if necessary)
  - This could be optimized for certain cases by allowing bursting



Optimization of frame rate limit

- Allow bursting, with a definition of burst window size & net rate
  - Burst size must not exceed ingress buffer size
- If ingress buffer >= max frame size, gain for mixed frames
  - Up to 33% in limit case
- Applicable for certain architectures
  - Separate header vs payload pipeline
  - Faster than line rate (data) draining into next element

Reasons not to optimize

- **Requires multi-frame state & definition**
  - **Not compatible with current MAC architecture**
- **Only benefits for corner case of corner case**
  - **Any deviation from optimal architecture➔ no gain**
- **Leaves opening for misconfiguration**
  - **Size of Rx ingress buffer doesn't = burst capability**
- **Need to consider proposal vs benefits**
  - **Weigh both sides…**

Agenda

- Recap : changes proposed
- Possible optimization for frame rate definition
- Interaction with CRS and deference
- Remote rate control request
- Conclusions and proposals

Interaction with CRS and deference

- The definition in the previous section follows the same format used in Clause 4 for carrier extension and IFS stretch (802.3ae)
  - The IFS is stretched regardless of state of CRS mode …
  - … but only if deference mode is enabled
  - Except that the frame rate timer continues in any case
- Need to consider the interaction of rate control with other mechanisms
  - External frame spacing (EPON)
  - Ethernet over DSL (CRS mode)
  - Etc…

Interaction with (non) deference

- The EPON (or other) external scheduler determines when a frame is available to send
  - This would normally be in bursts (but not necessarily)
  - Using rate limiter would enable reduction in buffering requirements
  - Steady state case should be handled through scheduler
  - Scheduler may not (easily) be able to handle frame rate limit
- Propose to leave as defined
  - Frame rate limiter will operate
  - Other rate limiters will not

Interaction with CRS mode

- **The PHY is controlling payload data rate using CRS**
  - **Additional IFS stretch due to rate limiter appears redundant**
  - **Furthermore EFM copper PHY will compress and eliminate IFS in any case**
  - **However, there may be a case for MAC rate limiting in addition to PHY rate limiting (inline security tagger with E-DSL)**
- **Propose to leave as defined**
  - **If rate limiter produces net b/w more than PHY, CRS will determine rate**
  - **If rate limiter produces net b/w less than PHY, CRS will not be asserted (& rate limiter functions as normal)**
  - **Combination space operates at sub-optimal efficiency (sometimes packets are delayed unnecessarily)**

Agenda

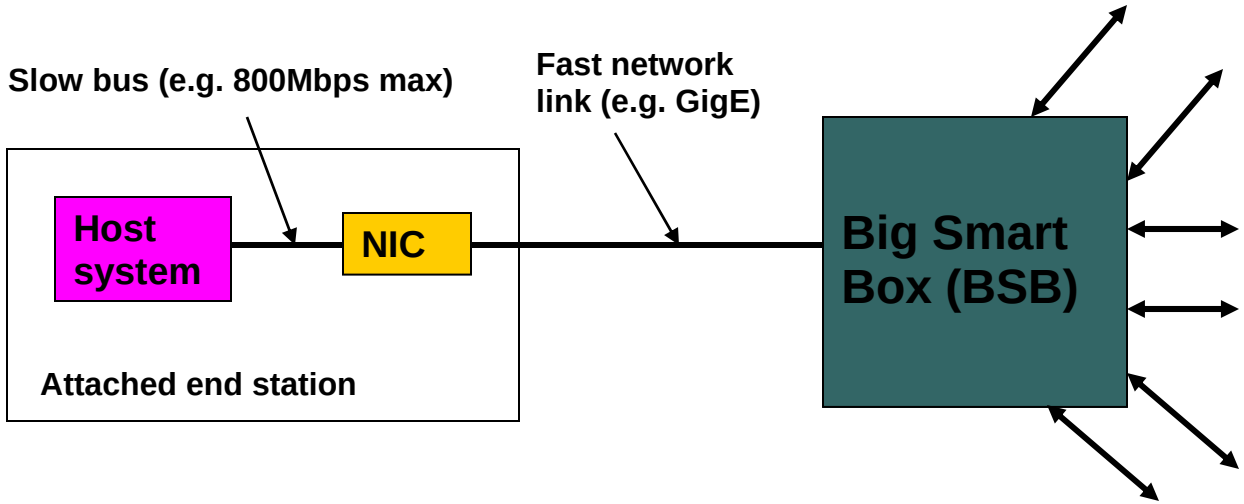
- Recap : changes proposed
- Possible optimization for frame rate definition
- Interaction with CRS and deference
- Remote rate control request
- Conclusions and proposals

Remote rate control request

A case can be made for defining a remote mechanism

- A device can tell its link partner to limit the Tx rate
- In addition to the MIB method

Example configuration (justification for rate control)



Network management could set egress rate control on BSB

- But end station may be moved arbitrarily
- Much more convenient for end station to signal its requirement

Request definition

- **Rate control is pseudo static**
  - **No real-time requirement**
- **Use slow protocol frame**
  - **Similar to LACP & .3ah OAM – see 43B**
  - **Ethertype x8809, multicast MAC addr**
  - **Subtype = 4 (next available)**
- **Propose: request and acknowledge**
  - **Nb – legacy partners will not acknowledge**

Request definition

- **Request opcode says “these are my capabilities”**
  - **Local Clause 30 attributes**
- **Acknowledge says “these are my setting”**
  - **May match request or not**
- **Management interface includes:**
  - **Send requests (every ??? Seconds)**
  - **Remote request received & remote capabilities**
  - **Acknowledge received & settings**
  - **Allow / disallow auto local setting at remote capability**

Agenda

- Recap : changes proposed
- Possible optimization for frame rate definition
- Interaction with CRS and deference
- Remote rate control request
- Conclusions and proposals

Summary

- Changes to Clause 4A & 30
- 3 types of mechanisms for f-d MACs
- MIB attributes for rate control
- Remote request

Proposals

- **Adopt changes to Annex 4A & Clause 30 described in this presentation**
- **Consider optimization as modification**
  - **Needs definition of mechanism**
- **Accept principle of remote request or…**
- **Adopt definition of remote request**