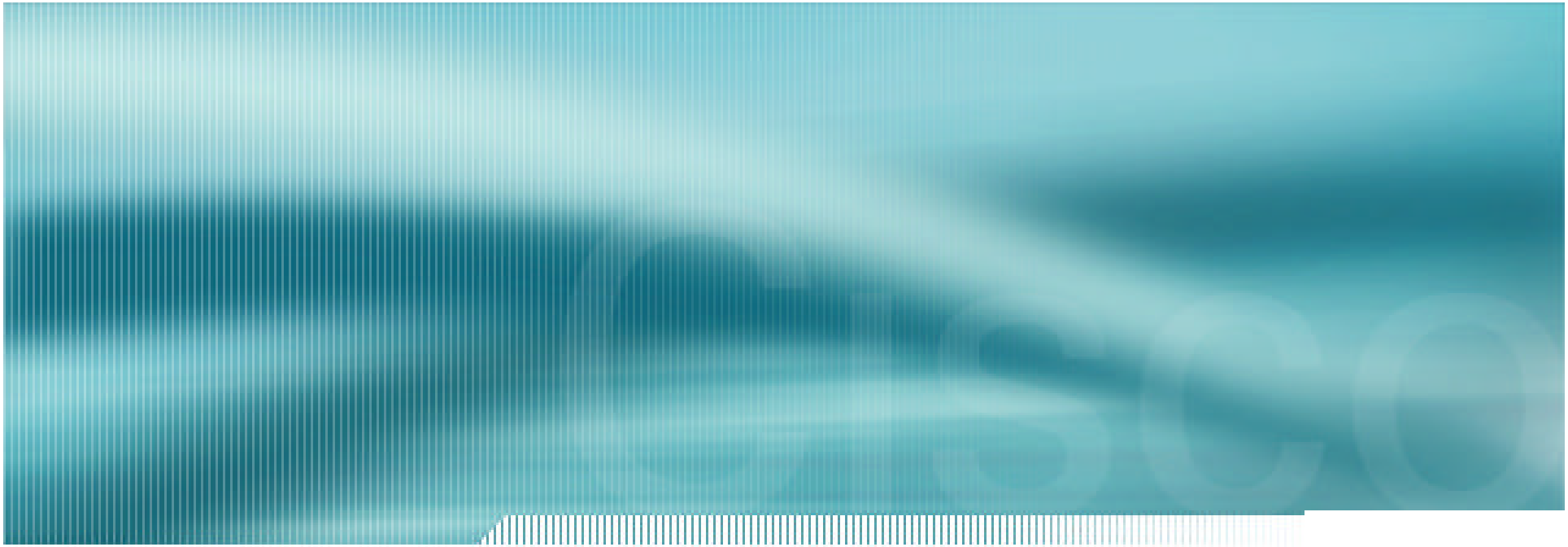


Rate control for Ethernet congestion management

Hugh Barrass (Cisco Systems)

Agenda

- 3 (& ½) types of rate control
- Interface & Management
- Remote rate control request
- Conclusions and proposals

3 (& ½) types of rate control

Rate control will fix a link at a reduced rate

- **There are 3 distinct applications that require this**
- **Hence the need for 3 types of rate control…**
- **… plus a hybrid between two of these types**

a) Constant (per packet) overhead

- **Covers encapsulation cases**

b) Limited (payload) bit rate

- **For non 10x bit rates or per bit overhead**

c) Limited packet rate

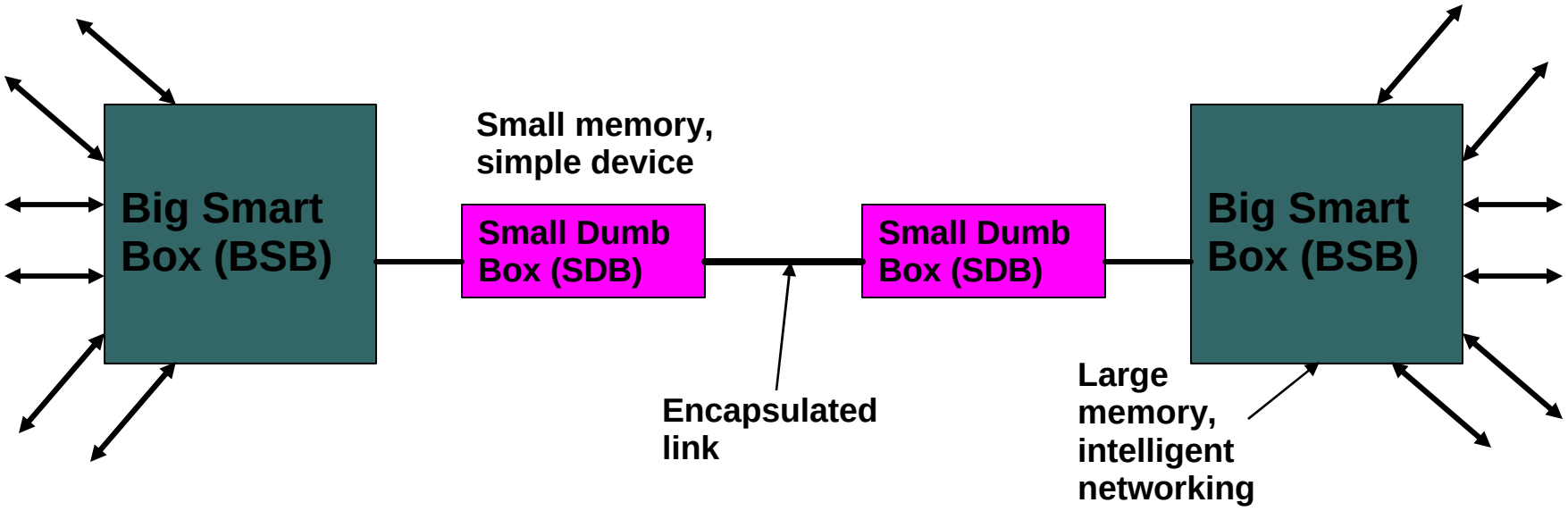
- **For packet processing limitations**

NB – FEC requires a hybrid of a) & b)

- **Includes fixed plus per bit overhead**

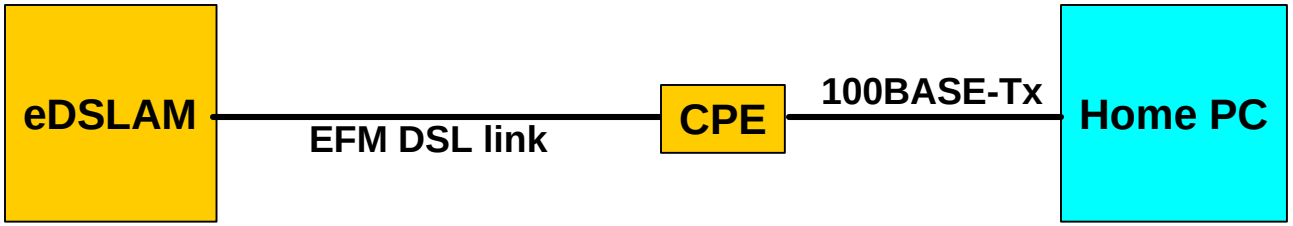
Constant packet overhead

- Explored in detail in daines_cmsg_1_0409.pdf (thanks Kevin)
  - Becomes a significant problem for inline MACsec implementations
  - … or other “dongle” encapsulator applications
- Inline encapsulators (dongles) must be economic devices
  - Small buffers, limited smarts (maybe line powered)
  - Network performance across constricted link sucks!



Limited payload bit rate

- Example in barrass_1_0704.pdf for high speed NIC
  - Ethernet link rate exceeds NIC bus rate, creating constriction
  - Limited intelligence & buffer in NIC – arbitrary packet drop
- Also applies for .3ah (EFM-DSL) CPE devices
  - Very simple CPE, with limited buffering,
  - Bridges between (e.g.) 100Mb LAN & 30Mb WAN links
- Can be used as a friendlier way of enforcing SLA
  - Link is limited to customer bit rate instead of policing & packet drop
  - Better overall network performance (if customer makes use of it)



Limited packet rate

- **No specific demand for this as yet, but…**
  - **… applications are easy to imagine**
- **Device with limited lookup engine rate**
  - **(e.g. cheap 10G)**
- **Interrupt driven or microcoded NIC**
  - **Must service each packet before proceeding with next**
  - **DMA allows high bit rate for large packets**

Agenda

- 3 (& ½) types of rate control
- Interface & Management
- Remote rate control request
- Conclusions and proposals

Interface to MAC client (now)

- **Currently specified in Clauses 4 & 2 (31)**
  - **… differently in each ☹**
- **4.3.2 function TransmitFrame**
  - **Includes TransmitStatus – to indicate success**
- **Clause 2 defines MA_DATA.request**
  - **Used in 31 (MAC control sublayer)**
  - **North & South interfaces to optional sublayer are different …**
  - **… causing historical anomaly**
- **Timing / pipelining not defined**
  - **Literal interpretation of standard would make QOS impossible!**

Interface to MAC client (needed)

- **Cleaner definition should include .acknowledge**
  - **Indicates that frame transmission is inevitable**
  - **Client may assert, remove or change request until acknowledge**
- **Addition of acknowledge controls timing**
  - **MAC/PHY layer can specify pipelining**
  - **MAC client can define queue draining**
- **MAC client has no concept of time**
  - **MAC & PHY defines real time frame timing**
  - **MAC is ideal place to define rate control**

Management

Rate control needs a management interface

- A means of telling the MAC what rate is required
- Management can currently set rate by choosing PHY

New MIB object(s) – Clause 30

- 3 parameters to define:
- Per packet overhead (IPG increase)
- Maximum payload rate (IPG stretch)
- Maximum packet rate

Outstanding question – real time or relative

- Could be % of max PHY rate but real time more straightforward

Agenda

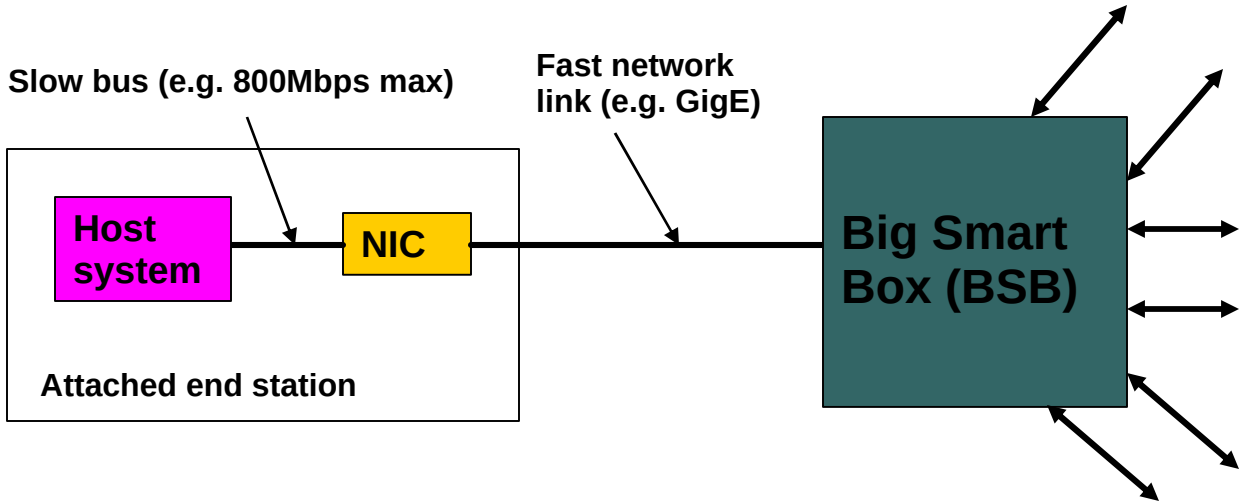
- 3 (& ½) types of rate control
- Interface & Management
- Remote rate control request
- Conclusions and proposals

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**
  - **Two suggestions (so far)**
- **Slow protocol frame**
  - **Similar to .3ah OAM**
  - **Defined entirely within 802.3 (OAM layer?)**
- **Piggy-back on LLDP**
  - **Discovered device parameter includes rate limit**
  - **Would need modification to 802.1**
- **Both cases need remote MIB attributes**
  - **Identical to Tx rate limit – but specifies max Rx rate**

Agenda

- 3 (& ½) types of rate control
- Interface & Management
- Remote rate control request
- Conclusions and proposals

Summary

- 3 (& ½) types of rate control
- Changes needed to MAC client interface
- MIB attributes for rate control
- Define remote rate control request

Proposals

- **Agree that description of 3 rate control mechanisms be added to Clause 4 (& 4A)**
- **Agree that Clause 4 & Clause 2 (31) be changed to clean up MAC client interface**
- **Agree that MIB attributes be added to Clause 30**
- **Agree that remote rate control request be defined**

Outstanding issues

- **Rate control definition in real time or relative to PHY speed?**
- **Remote request based on OAM or LLDP?**
- **Definition of client interface to include pipelining restrictions?**

Finally…

- **This slide set is 2/3 towards a baseline**
  - **With consensus, baseline could be prepared**
  - **Ideally ready for March Plenary**
- **Close open issues**
  - **& address any new ones…**
- **Baseline must be sufficient to start draft**
  - **Complete description of technical solution**
  - **Leaves editorial control to editor**