The Structure for Congestion Management

Bob Grow
Principal Architect, Intel Corporation
May 24, 2004

intel

Topics

- Structuring the congestion management problem
- The constraints and opportunities for congestion management in the 802.3 architecture

Observations on CM

- The best solution for low latency and high priority is bandwidth (throughput)
- “Congestion management” is ultimately allocation of throughput
- Bandwidth can be given to one traffic classification at the expense of other traffic classifications
- Low latency can be provided to some traffic at the expense of greater latency for other traffic through preferential access to bandwidth

Stating the obvious

- There is only as much bandwidth as there is bandwidth
  - Congestion management won't create bandwidth
  - Congestion management will waste some bandwidth
- Bandwidth demand will exceed availability
  - Long term excess demand can only be satisfied with more bandwidth
  - Short term excess demand can be buffered
  - We will probably never all agree on a time value for short term

Definitions

- **Congestion is of two general types:**
  - **Transitory**

    Traffic which can be smoothed over time, without frame drop, because average bandwidth demand is less than capacity and peak demand can be buffered
  - **Oversubscription**

    Traffic which cannot be smoothed over time and results is either not being admitted to the network (e.g., admission control) or either results in frame drop (e.g., buffer overflow, RED), or back-up into source buffers

Transitory congestion

- Does not result in frame loss
- It does add latency, and consequently latency jitter
- Multiple causes (e.g., simultaneous requests, clock rate disparity, addition of flows)
- Improved latency can be given to “high” priority through differentiated service
- High priority latency improvement will generally provide greater benefit than efforts to improve the minimum latency of the network

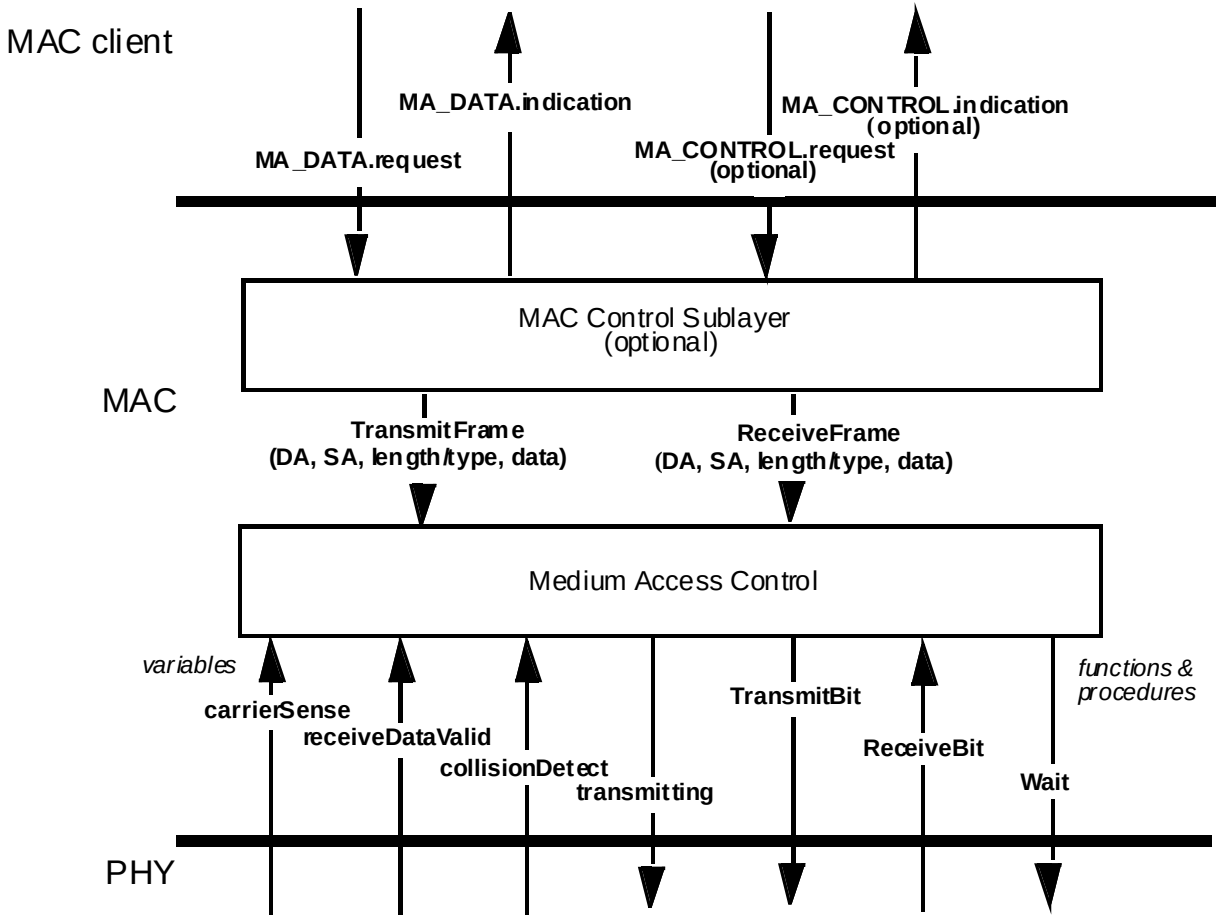
Oversubscription

- Where (in the path) frames are dropped can effect overall system performance
- Rate limiting is used to decrease frame drops
- End systems can benefit from network rate limiting
- Pushing congestion to the source provides more options for rate limiting
- If not pushed to the source, will result in unintended consequences
- Network rate limiting protocols that only push congestion toward the source may or may not provide acceptable cost/benefit

Two solution spaces

- **Differentiated service (priorities)**
  - **Provides preferential latency performance to high priority**
  - **Naturally pushes rate limiting onto lower priorities**
- **Rate limiting has potential to reduce frame drops within the network**
  - **How rate is limited will affect jitter characteristics**
  - **How rate is limited can limit or add to latency jitter**
  - **All practical rate limit proposals will have undesirable side effects**

Service Interfaces



MAC Service Interface

- **Four primitives**
  - **MA_DATA.request**
  - **MA_DATA.indication**
  - **MA_CONTROL.request**
  - **MA_CONTROL.indication**
- **Not mapped well to MAC definition**
  - **Not specifically bound to actual frame transmission/reception**
  - **No acknowledgement**
- **MAC Control sublayer arbitrates between requests**
  - **Assumes MA_DATA.request is HOL in a queue**
  - **Assumes MA_CONTROL is HOL in a queue**

Where is full duplex rate limited?

- MAC data rate
- Rate and size of MAC client requests
  - MPMC for 802.3ah
- MAC Control
  - Pause
  - Pre-emption by MAC Control frames
- PHY
  - 10GBASE-W
  - 10PASS-TS, 10BASE-TL

Study Group questions

- **Is differentiated service (priorities) too obvious to specify?**
  - **Already specified by other groups**
  - **Not specified for 802.3 “end-points”**
- **Will MAC Control or MAC rate control improve system performance?**
  - **Is there a general layer 2 solution?**
  - **Is there something that layer 2 can do to aid a general solution?**
  - **Is there sufficient market potential for a limited topology solution?**

Summary

- A clear separation of congestion will aid evaluation of proposals
- The 802.3 architecture already includes arbitration between transmit requests
- The 802.3 architecture has multiple locations where rate is limited
- We still have a significant study task to evaluate the cost/benefit of congestion management alternatives