

IEEE 802.3ad Link Aggregation (LAG)

what it is, and what it is not

Howard Frazier – Broadcom
Schelto Van Doorn – Intel
Robert Hays - Intel
Shimon Muller – Sun Microsystems
Bruce Tolley – Solarflare Communications
Paul Kolesar – CommScope
Geoff Thompson – Nortel
Brad Turner – Juniper Networks

17-April-2007
Ottawa

IEEE 802.3 HSSG

Outline

- Overview of 802.3ad Link Aggregation
- Current status and future plans
- Improving Link Aggregation
- Summary

802.3ad Link Aggregation

- Specified in Clause 43
- LAG is performed above the MAC
- LAG assumes all links are:
 - full duplex
 - point to point
 - same data rate
- Provides graceful recovery from link failures
- Traffic is distributed packet by packet
- All packets associated with a given “conversation” are transmitted on the same link to prevent mis-ordering

802.3ad Link Aggregation

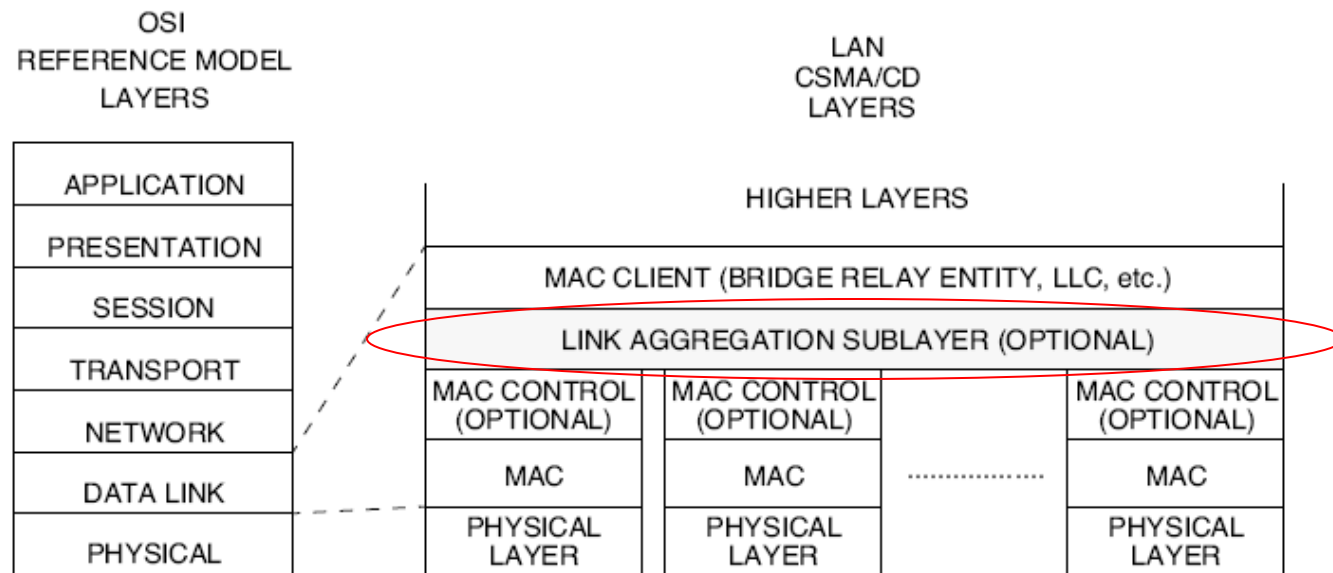
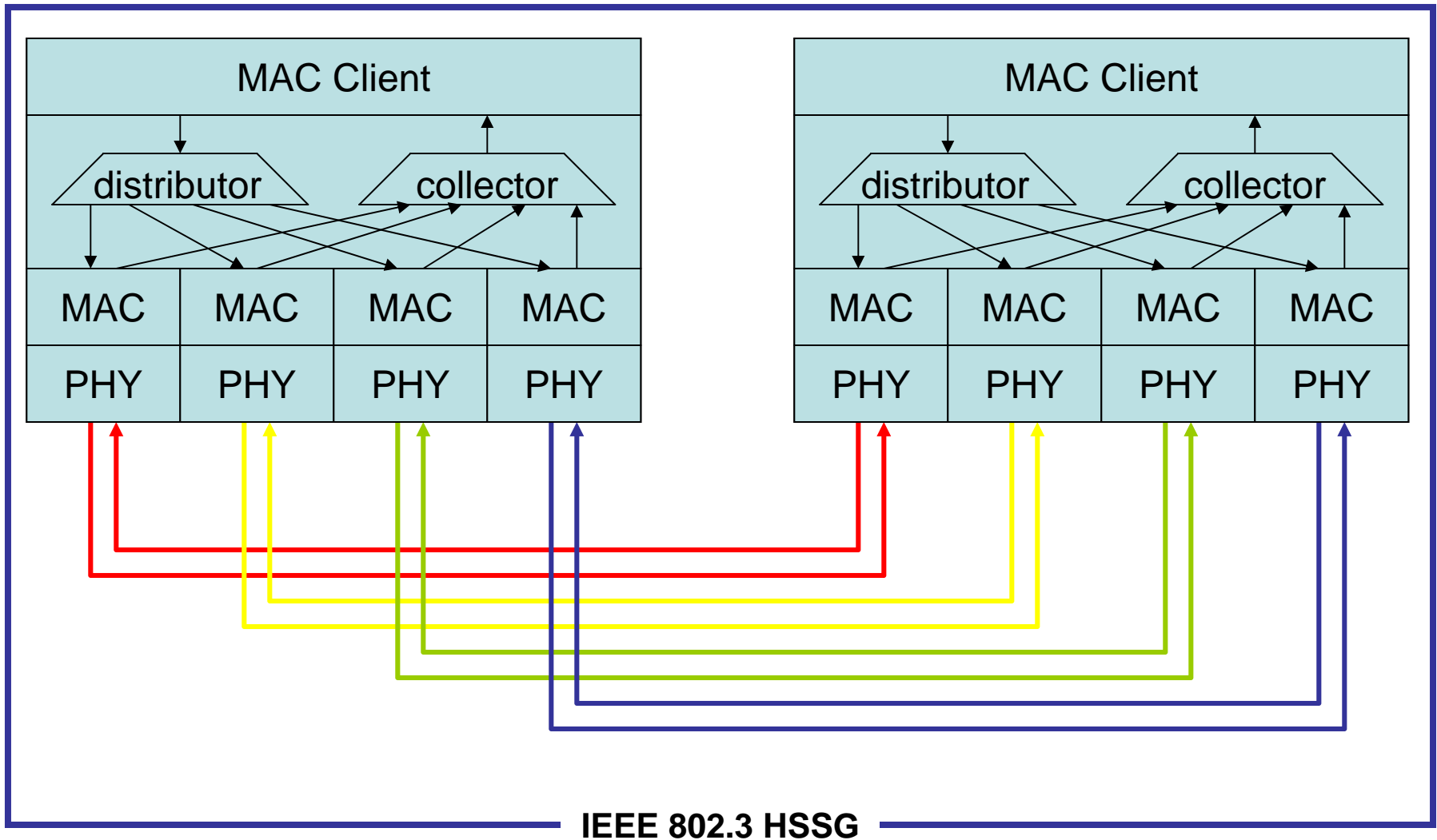


Figure 43-1 – Architectural positioning of Link Aggregation sublayer

802.3ad Link Aggregation



802.3ad Link Aggregation

43.2.4 Frame Distributor

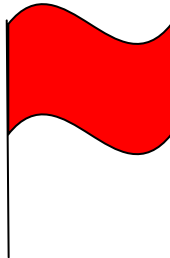
...

This standard does not mandate any particular distribution algorithm(s); however, any distribution algorithm shall ensure that, when frames are received by a Frame Collector as specified in 43.2.3, the algorithm shall not cause

- a) Mis-ordering of frames that are part of any given conversation, or
- b) Duplication of frames.

The above requirement to maintain frame ordering is met by ensuring that all frames that compose a given conversation are transmitted on a single link in the order that they are generated by the MAC Client; hence, this requirement does not involve the addition (or modification) of any information to the MAC frame, nor any buffering or processing on the part of the corresponding Frame Collector in order to re-order frames.

...



802.3ad Link Aggregation

- Does not change packet format
 - No added headers or sequence numbers
 - Type/Length interpretation unchanged
- Does not require added buffers
 - No fragmentation or reassembly
- Does not re-order or mis-order packets
- Does not add significant latency
- Does not increase the bandwidth for a single conversation
- Achieves high utilization only when carrying multiple simultaneous conversations
- Is not transparent to some 802.1 sub-layers

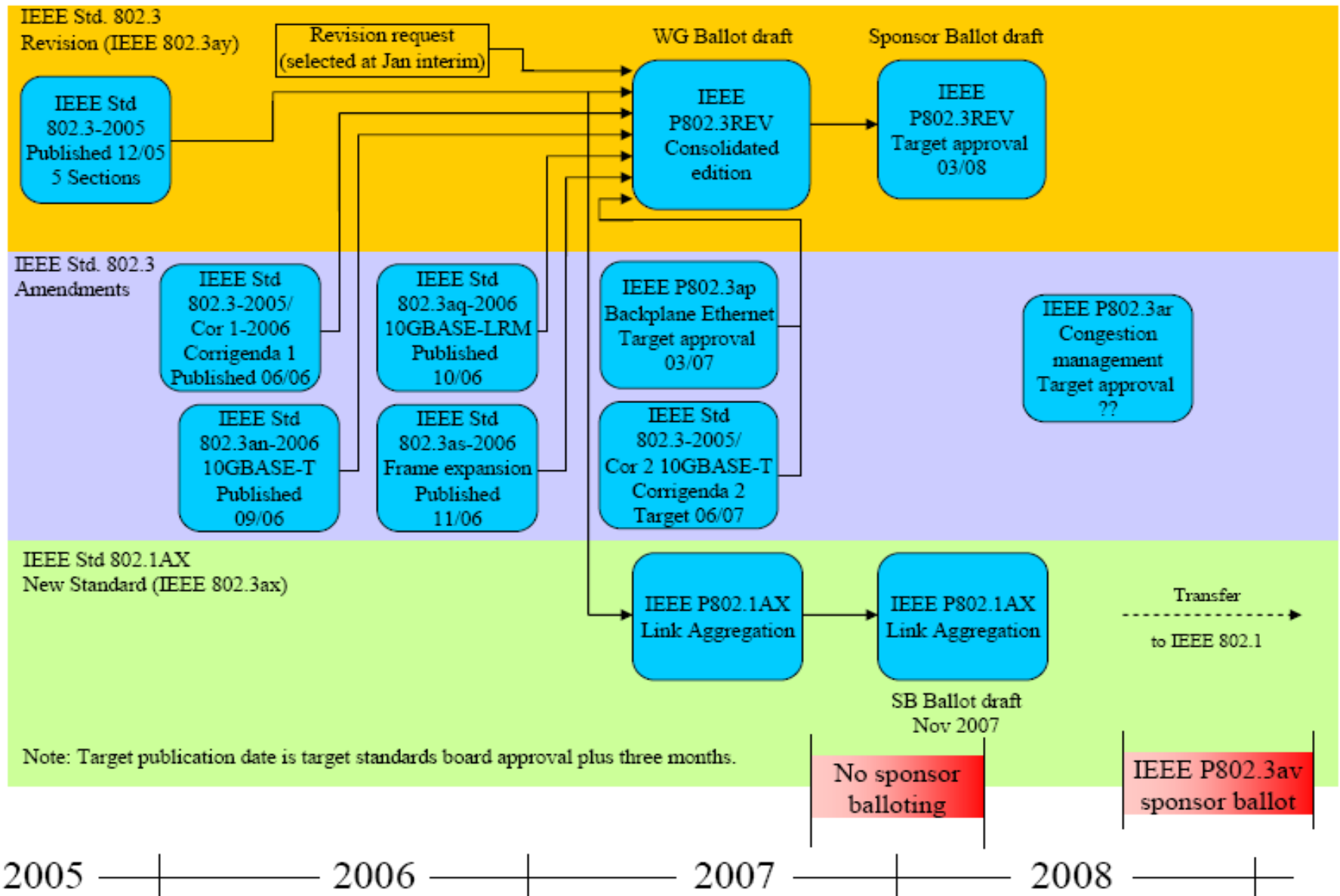
802.3ad Link Aggregation

- Is a very good thing
 - It does what it was intended to do
 - It is relatively easy to implement and use
- Does not always provide a linear multiple of the data rate of a single link
 - N aggregated links usually do not provide N times the bandwidth
- Incurs a linear multiple of the cost of a single link
 - N aggregated links cost N times as much as a single link, because everything must be replicated
- Appears to the user as N individual links, which must be individually managed

Current status

- In the process of being “spun out” of 802.3
- 802.3ax project creating IEEE P802.1AX
- No technical changes
- After the transfer, 802.1 may undertake enhancements or revisions

IEEE Std 802.3 revision plan



Improving Link Aggregation

- Can link aggregation be “fixed”?
 - Inspect headers deep into packet
 - or
 - Add sequence number to packet
 - Change the packet format
 - and
 - Add LARGE buffers to receiver

Improving Link Aggregation

- Why not inspect headers deep into a packet?
 - Futile if packets are encrypted
 - Layer 2 LAG would need to parse upper layer protocol header formats (layering violation)
- Why not change the packet format?
 - Requires a new 802.3 MAC definition
 - Requires a new 802.3 MAC client interface
- Why not add LARGE buffers to the receiver?
 - Adds LONG delay
- Fixing LAG is neither easy, or rewarding

Summary

LAG is good, but it's not as good as a fatter pipe