The Bridge Queue Placement Problem

Norman Finn

Cisco Systems

IEEE Std. 802.1Q-2006

Subclause 8.6 The Forwarding Process

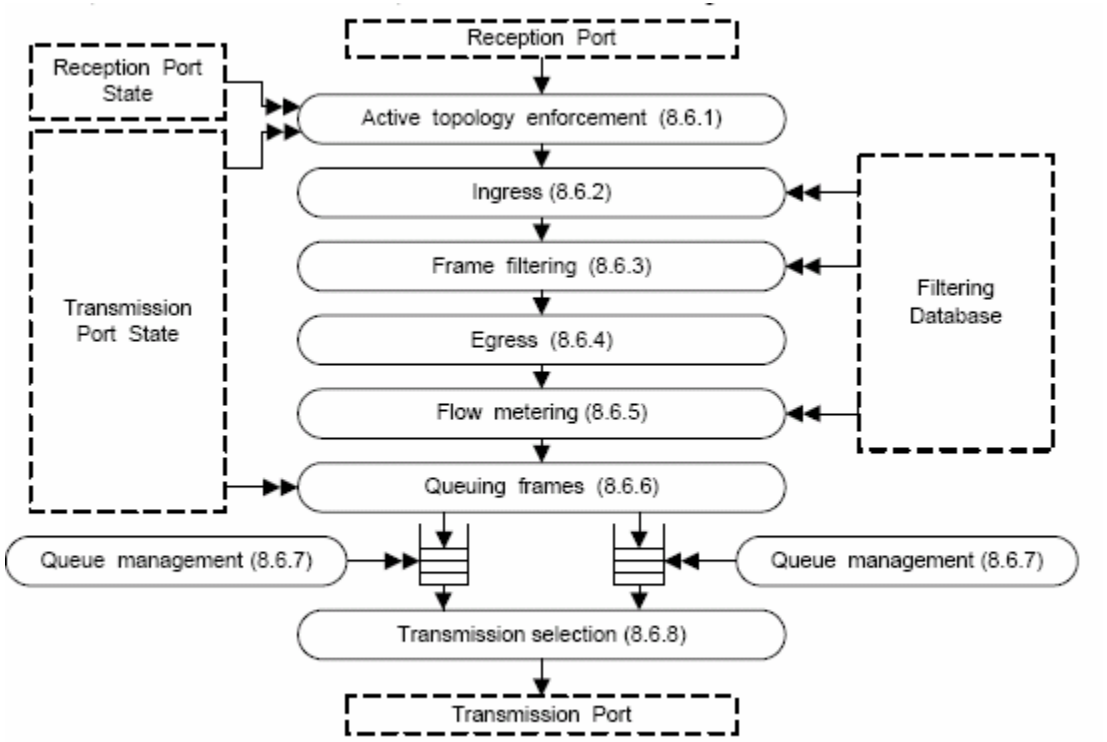


Figure 8-9—Forwarding Process functions

IEEE Std. 802.1ag-2007

Subclause 22 CFM in systems

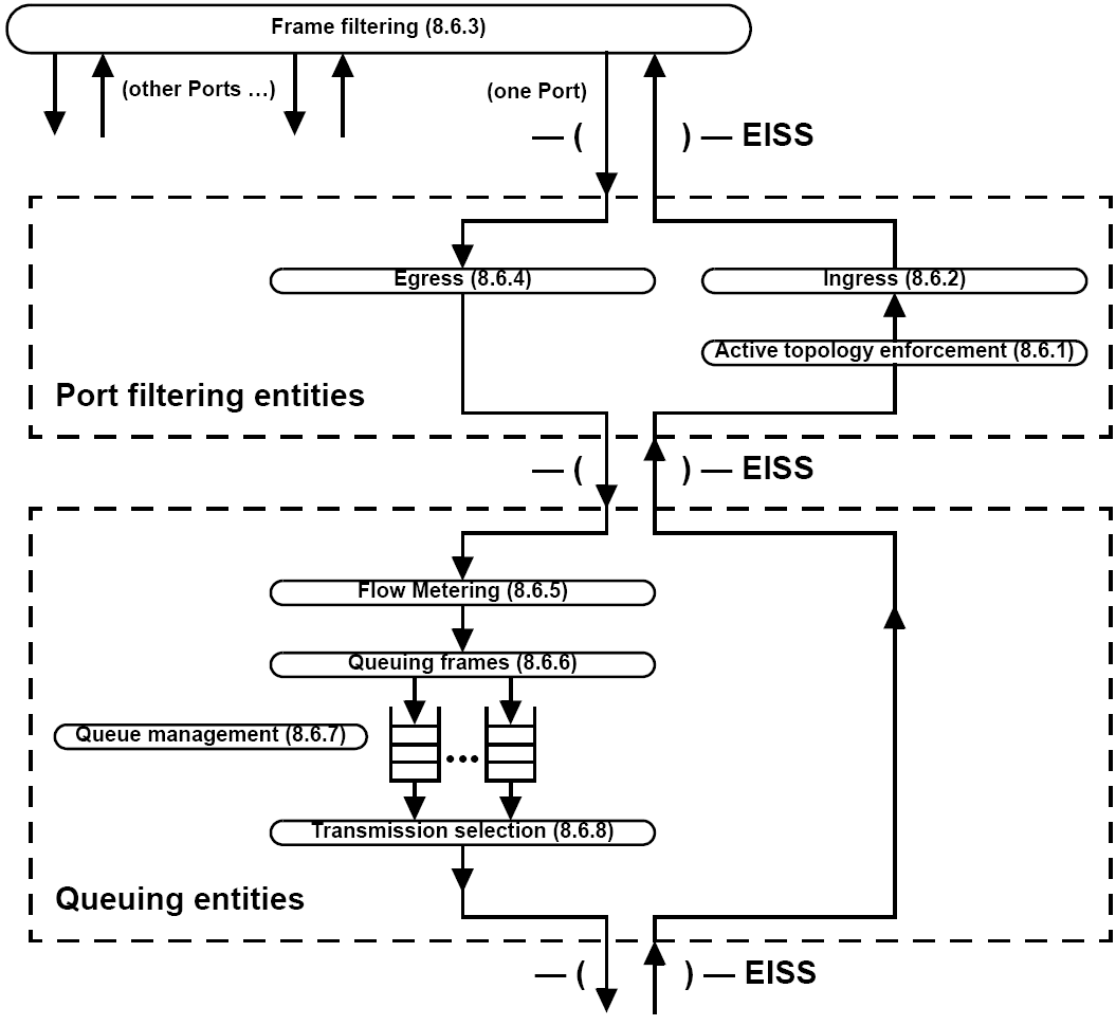


Figure 22-2—Alternate view of Forwarding process

IEEE Std. 802.1ag-2007

Subclause 22 CFM in systems

Port filtering entities are separated from queuing entities, and both are placed in the “pants leg”, instead of in the Forwarding Process

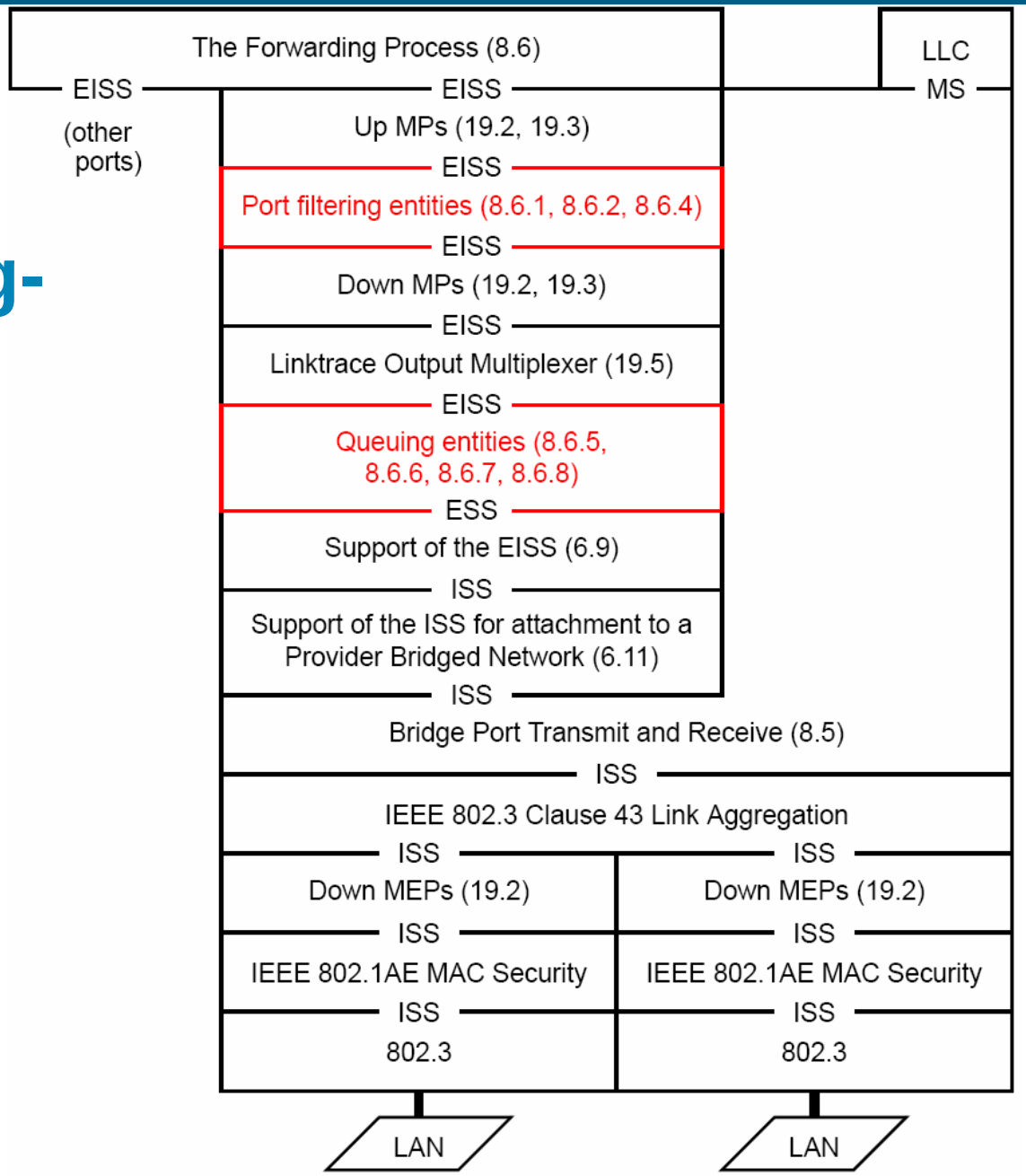


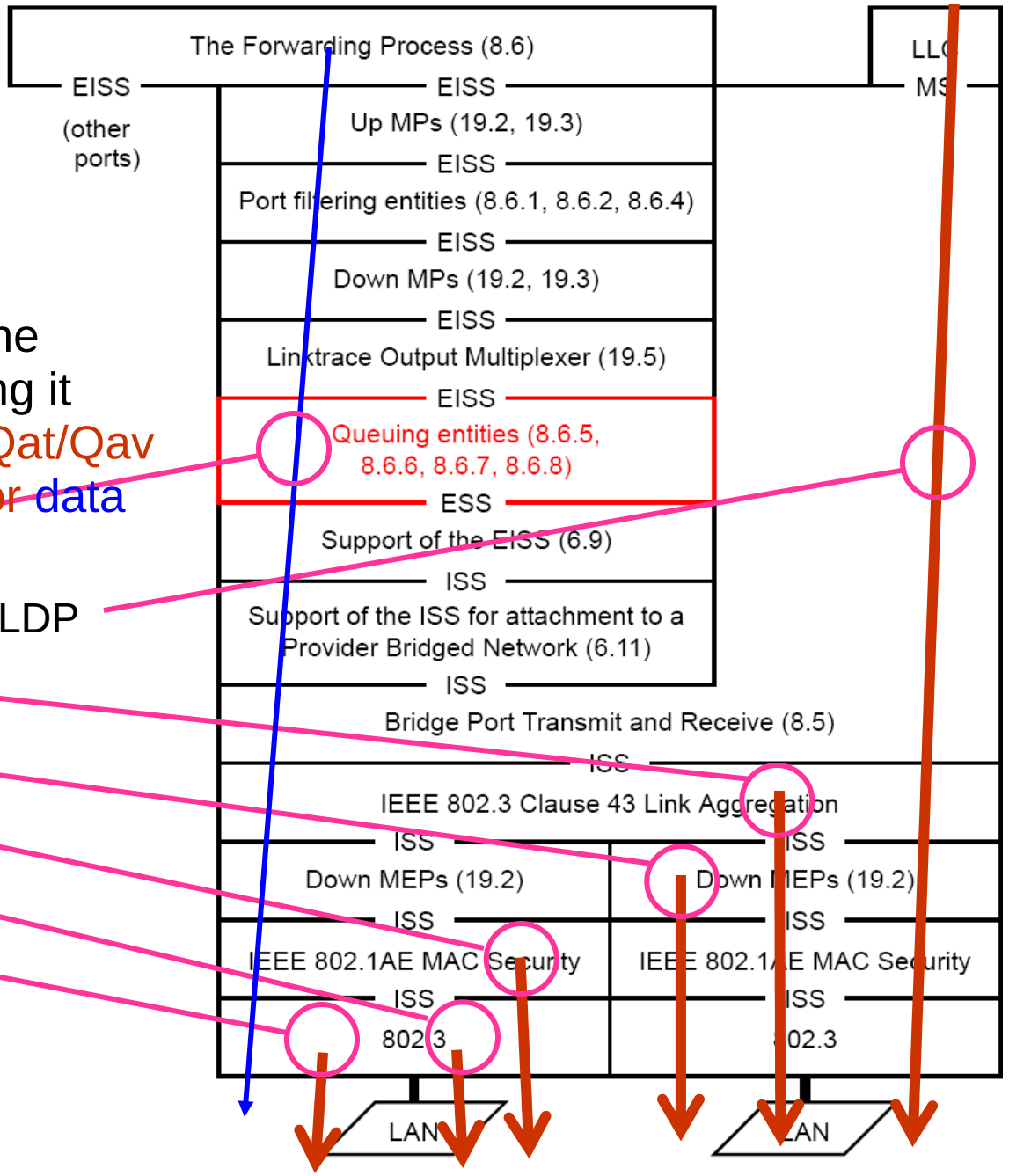
Figure 22-8—Maintenance Point placement relative to other standards

Two Projects

- P802.1Qat specifies a protocol to reserve bandwidth.
- P802.1Qav specifies transmission selection algorithms for the queues to support bandwidth-reserved data streams.
- P802.1au specifies a protocol to generate congestion notifications.
- P802.1Qaz specifies transmission selection algorithms for the queues to support congestion-controlled data streams.

Problem:

- Many frames bypass the queuing entities, making it impossible for P802.1Qat/Qav to guarantee latency for data streams:
  - 802.1 xSTP, MxRP, LLDP
  - 802.3 LACP
  - 802.1 CFM
  - 802.1 Security
  - 802.3 OAM
  - 802.1 Pause



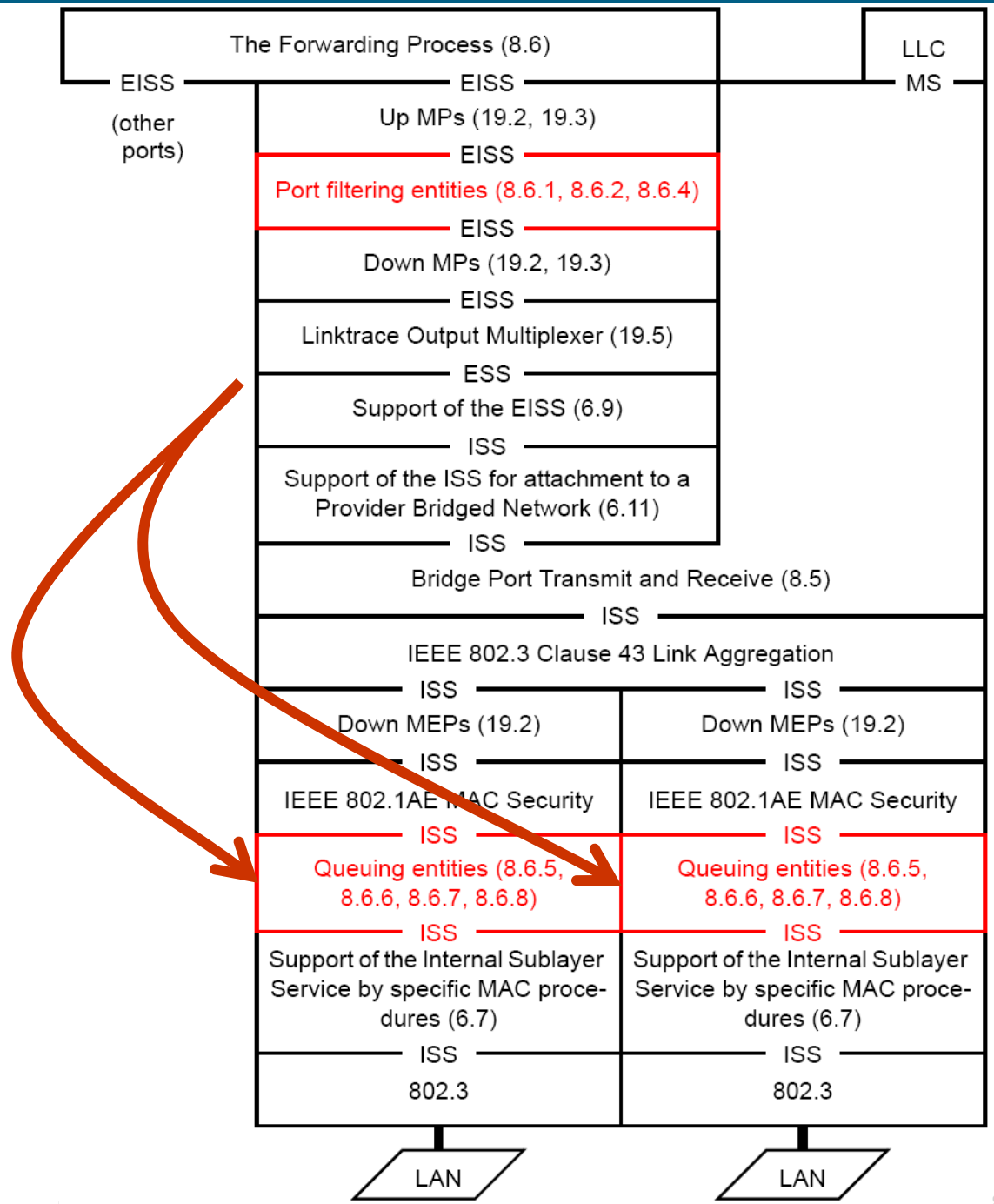
Current queue position kills P802.1Qat/Qav

- Frames can be introduced into the stack a numerous places around and/or below the priority queues.
- Given this situation, we cannot leave the difficult task of ensuring the AVB TG's latency guarantees as an exercise to the reader, and expect the guarantees to be met.

Wrong answer:

If we move all of the queuing down to a lower position in the diagram, P802.1Qat/Qav can guarantee latency.

But, it becomes impossible for P802.1au/Qaz to do Backward Congestion Notification.



Low queue position kills P802.1au/Qaz

- If placed below MACsec:

  The Reaction Point (source of the frame) will not recognize the congestion-causing frame when the first part of the frame is returned as part of the data in a Backward Congestion Notification; and

  The BCN would have to use the trusted port of the SecY.

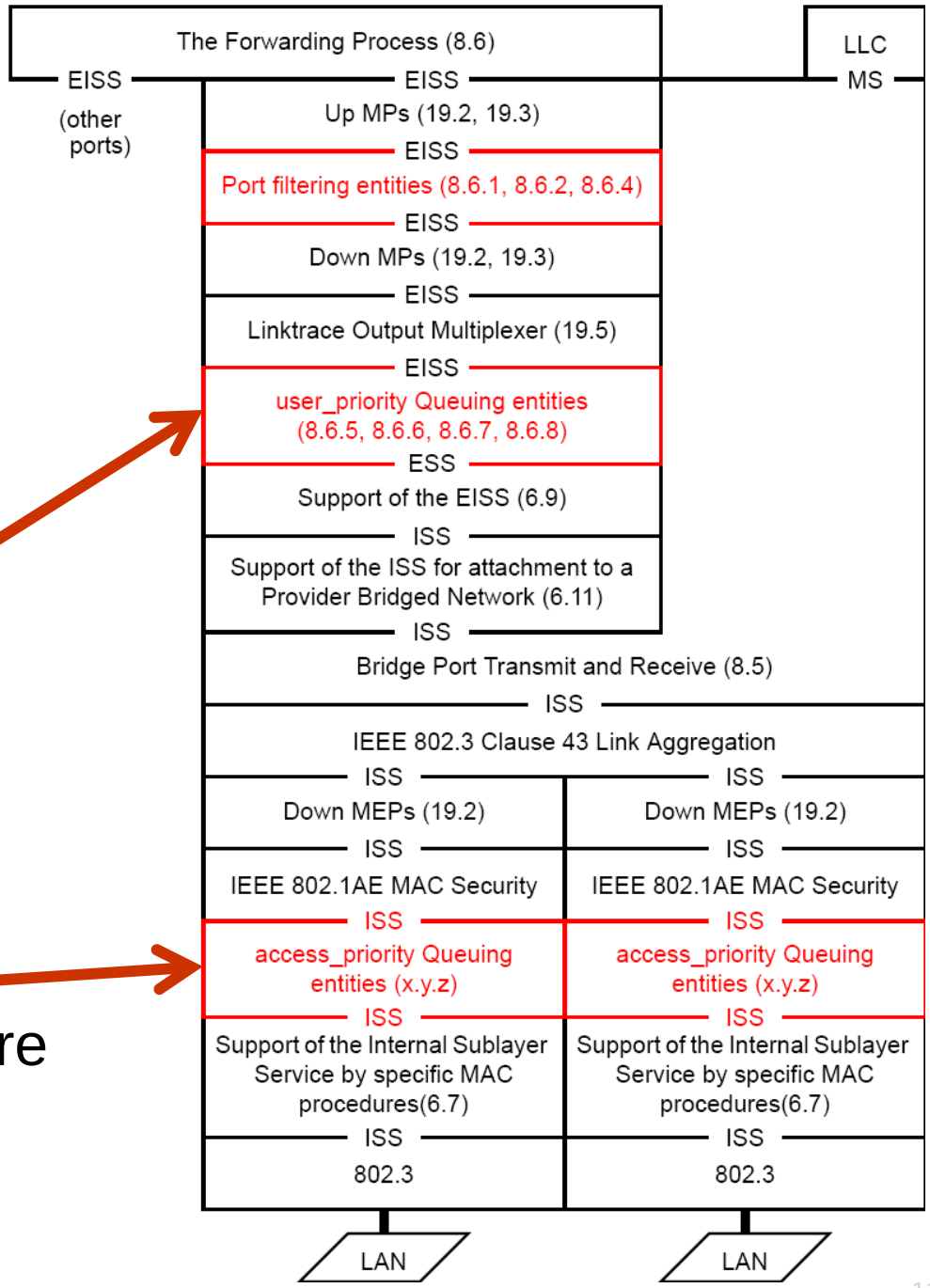
- If placed above MACsec, there are still a number of flows of data that can disrupt the queuing and guarantees provided by both P802.1Qat/Qav and by P802.1au/Qaz.

Not to mention ...

- The always bothersome fact that, even though 802.1D says that queues are up in the forwarding process, many bridge builders place the queues very close to the physical LAN, and it works just fine.

Solution: there are two places where queues can go.

- Queues based on user_priority, (which belongs to bridges) go here.
- Queues based on access_priority, (which belongs to media) go here

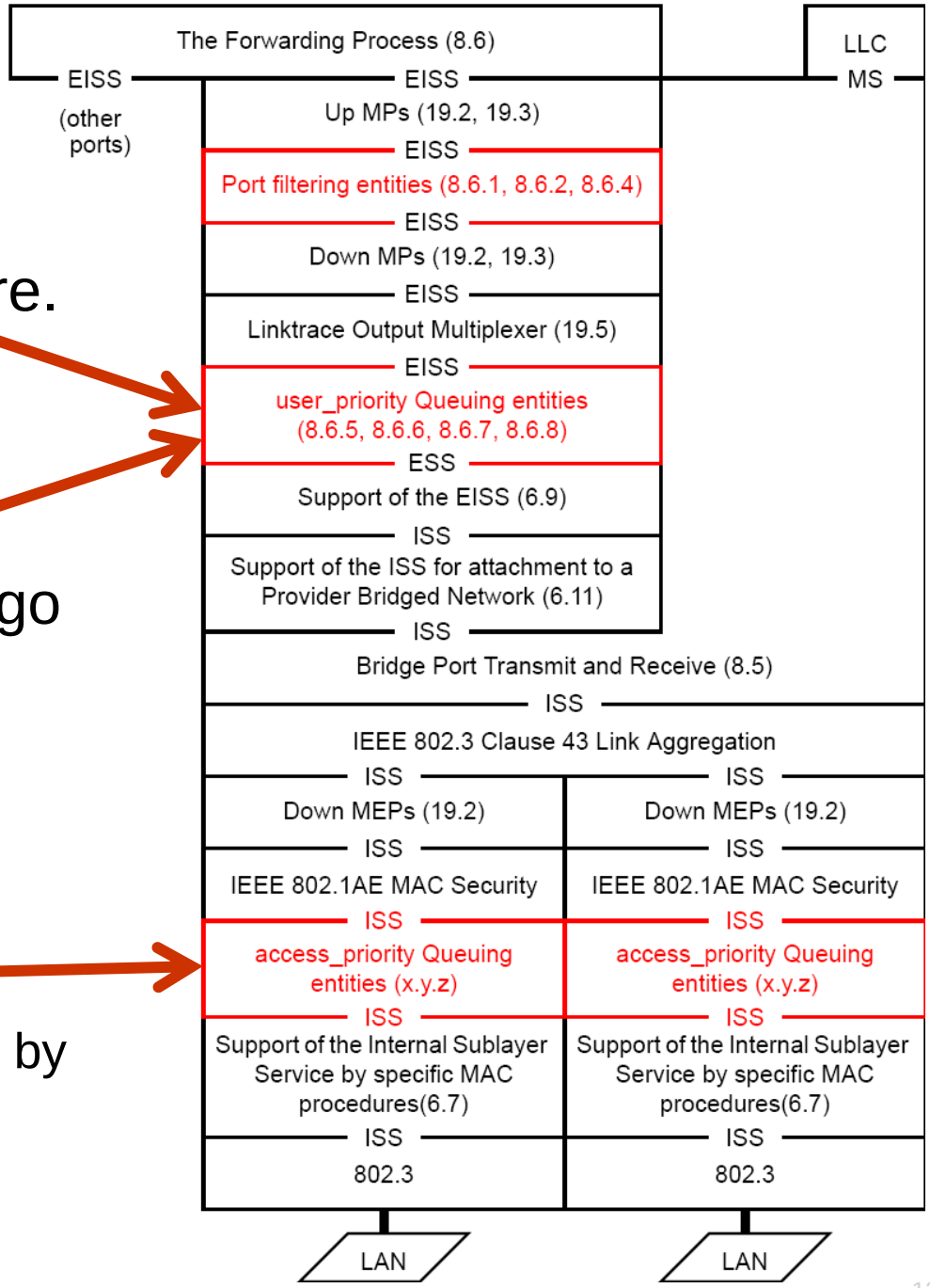


access_priority

- There are two priority parameters in the ISS, user_priority and access_priority.
- user_priority is follows the frame through the forwarding process, and can be mapped on both input and output.
- user_priority from the ISS is replaced by priority from the 802.1Q/ad tag (if any) on the way up, and inserted into the 802.1Q/ad tag (if any) on the way down.
- access_priority is discarded on the way up, and derived from user_priority (by a fixed table) on the way down

How is this a "solution"?

- P802.1au queues go here.
  - BCN generation works properly in these queues.
  - Non-flow controlled traffic bypasses these queues.
- P802.1Qat queues also go here.
  - Rate shaper queues only.
  - Non-guaranteed traffic bypasses these queues.
- All frames then hit these queues.
  - These queues are drained by strict access_priority.



How does this help?

- The special queues (.1Qat/Qav or .1au/Qaz) serve only the special data.
- All traffic leaving the .1Qat/Qav queues is assigned the highest access_priority.

  This is because, unlike the current transmission selection algorithm, the .1Qav transmission selection algorithm understands time.

  .1Qav queues are drained at intervals, so that lower-priority traffic is guaranteed bandwidth.

  Even BPDUs are lower priority than latency-guaranteed flows! The BPDU will be transmitted, but the audio latency is also important, and the configuration will guarantee that the audio rate is never high enough that the BPDU delays will disrupt the spanning tree.

How does this help?

purpose	user_ priority	access_ priority
control	7	5
	6	4
latency guaranteed	5	7
	4	7
best-effort	3	3
	0	0
	2	2
	1	1

- All special 802.1Qat/Qav data is assigned to a single access_priority queue, with the highest access_priority.
- All other data is assigned to other access_priority values, and thus other queues, according to how many queues are available, similar to the current assignments.
- Since the user_priority 802.1Qat/Qav transmission selection knows about time, and spools the data out accordingly, the access_priority queue(7) is almost always empty.
- Both latency queues get the same access_priority, because their frames are sequenced by the user_priority queues.

What about congestion management?

- Clearly, the P802.1au/Qaz congestion management queues must be at the user_priority position, so that meaningful BCNs can be generated.
- The relationship between the user_priority and access_priority queues for P802.1au/Qaz congestion management is TBD, but this author asserts that even if the concept of two queue layers proves not to be useful to that project, the concept will not injure P802.1Qaz.

What about current implementations?

- access_priority for an 802.3 medium must be 0, according to 802.1D-2004 subclause 7.7.5, but 802.3 implementations don't often have or use access_priority, anyway, so that is meaningless.
- If access_priority = user_priority, then you could place your queues in either location, and they would work just the same, except that they would serve all frames. (Which, in fact, is what many implementations do!)
- For media that actually use access_priority, (Token Ring, FDDI) this author asserts that the

Summary

- P802.1Qav and P802.1Qaz should be written to incorporate this concept:

  The upper queues are in the position specified in 802.1ag, and are controlled by user_priority.

  The lower queues are immediately above the MAC specific functions, controlled by access_priority.

  The user_priority to access_priority table is changed as described here, when P802.1Qat is used, in some manner (TBD) when P802.1au is (or both are) used, and is a 1:1 no-op when neither is used.

- 802.1D/Q should be modified to make the lower queue position the standard queues when neither P802.1au or P802.1Qat are used.


CISCO