IEEE 802.3 Ethernet First Mile

Optical Point to Multipoint – Objectives

Kevin Daines
Matt Goldman
Jonathan Thatcher
John Wolcott

Notes

- **All discussion in the presentation is related to point-to-multipoint optics, only**
- **“Upstream control” indicates direction of control packets or signals**
  - **N O T control of the upstream data flow**
  - **Tail end (T.E.) port sends control to Head end (H.E.) port**
- **“Downstream control” is H.E. to T.E.**

Agenda

- **Key assumptions**
- **Slot definition**
- **Link Aggregation**
- **802.3x PAUSE Flow Control**
- **Initialization / Add / Drop**
- **PMA/PMD requirements**
- **PAUSE_PORT flow control**

Key Assumptions

- **A T.E. port must be able to be added or dropped with minimal impact**
  - **Can't continuously toggle power on one T.E. port and interfere with operation of other ports**
  - **Don't drop packets; don't take other ports off line**
- **Consequence**
  - **H.E. Rx must be able to ride through darkness**
  - **H.E. Rx must be able to tolerate (T.E. to T.E.) optical power variations as seen at the H.E.**
  - **Data efficiency lowered due to extra time needed by H.E. Rx to synchronize**
    - **Trade off between cost of design to reduce time and efficiency**
- **Should this be an objective?**

Key Assumptions

- **OP2MP must auto-initialize & auto-configure**
  - **One of Ethernet's strengths is "plug & play"**
  - **Required for "Compatibility," "Broad Market Potential," and "Economic Feasibility"**
- **Consequence**
  - **No optimizations in efficiency (tuning of system) by manually changing parameters and settings**
  - **Example: no setting the optical power of a T.E. port during installation**
  - **Example: no setting of port addresses**
- **Should this be an objective?**

Key Assumptions

System needs to be synchronous

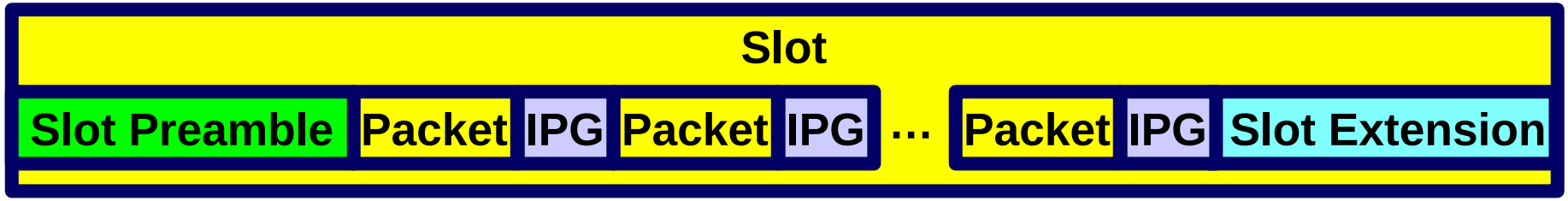
- **H.E. Tx PLL used to time all state machines**
  - **T.E. Rx PLL recovers clock to drive all control logic**
  - **It may be advantageous to have T.E. Tx run off of T.E. Rx PLL**
    - **Implementation choice?**

Slot Control – What it shouldn't be

Should H.E. "dynamically" (real time; not 'Non-realtime, or quasistatic control')" control the usage of the upstream time allocation?

- **<u>Assumption</u>: variation of time allocation is very infrequent wrt upstream packet flow**
- **<u>Assertion</u>: not insignificant overhead due to packet size, link length, and timing variations**
- **<u>Alternative</u>: "slot control" can autonomously be controlled by T.E. ports under direction of H.E.**
- **<u>Recommendation</u>: NO NO NO**

Slot Definition (1 of 2)



Slot Preamble:

- **Minimum time to guarantee H.E. Rx synchronization**
- **May or may not be same as Frame's Preamble**

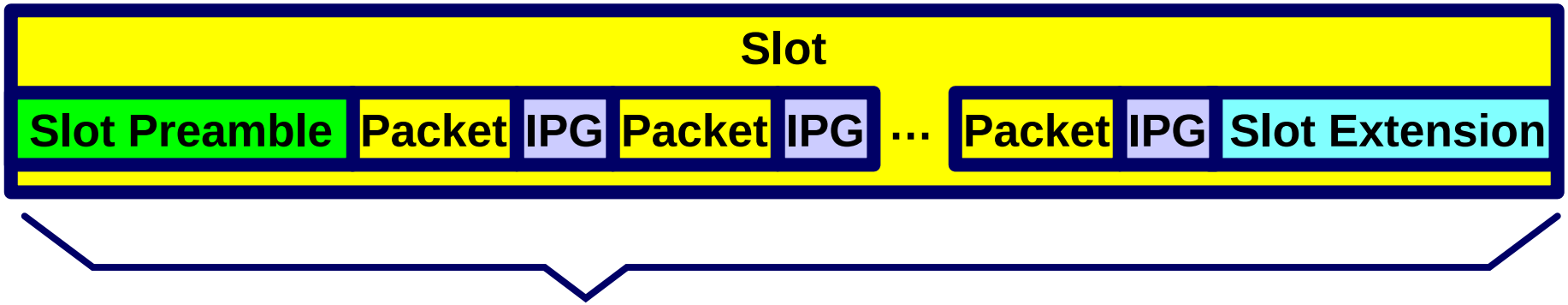
Slot Extension: idle stream to pad out slot time

- **Managed B/W: (Slot time – Slot Extension) / Slot time**
- **No growth of interslot gap due to end of packets**
- **Fills all of slot (sans the slot preamble) when no data**
- **Bit pattern might be same as Slot Preamble (/A/K/R/?)**

Interslot Gap (not shown): minimum time required to guarantee no slot overlap

Slot Definition (2 of 2)

Slot Time: when slot begins relative to beginning of slot period



Slot Duration: total number of bits of slot

- Slot Preamble + Data + Extension
- Maximum shall be entire slot period

Data Duration: time allocated within slot for packets + IPG

- Adjustment of Data_Duration controls BW (and flow ctrl?)

Slot Period: repetition rate for the start of a T.E. port's slot

Time: system assumed to be synchronous

- All timing based on the H.E. Tx bit period

Recommendation

- **Provide mechanism for H.E. to communicate all slot parameters to all T.E. ports simultaneously using a Slot_Control "signal"**
  - **Acts as synchronization pulse**
  - **Programs:**
    - **Slot period**
    - **Null slot time**
    - **Slot times for N T.E. ports**
    - **Slot durations N T.E. ports**
    - **Data durations N T.E. ports**
- **Mechanism will depend on destination**
  - **PHY**
  - **MAC Control**
  - **Other**

Some key slot issues (RFI)

- **To what degree should slot timing be implementation specific?**
  - **Max and min slot period**
    - **Shorter period: better jitter; longer period: more efficient**
    - **Remember, application jitter defined in real time, not in bit times**
  - **Min slot duration (max defined by max slot period)**
  - **H.E. must know T.E. max and min parameter values**
    - **Limits set by specification in standard**
    - **? and implementation by query during initialization ?**
- **Preamble questions**
  - **Should slot preamble be kept independent of packet preamble (are there speed upgrade considerations)?**

Link Aggregation & OP2MP

- **Assertion: 802.3ad does not support P2MP**
- **Opinion: only advantage to having a P2MP multi-link connection is for reliability (certainly not for bandwidth)**
- **Opinion: purpose of 802.3ad was aggregation, not redundancy**
- **Recommendation: clearly state in objectives and PAR that P2MP will not support 802.3ad**

Link Aggregation -- Backup

43.1.2 "Goals and Objectives."

- **n) Multipoint Aggregations**. The mechanisms specified in this clause do not support aggregations among more than two Systems.

Dot-X Flow Control

Should 802.3x MAC Control Frames for Flow Control be supported?

- **Needs to be separated into two sub-questions:**
  - **Should PAUSE flow-control be supported upstream (upstream control of downstream data flow)?**
  - **Should PAUSE flow-control be supported downstream (downstream control of upstream data flow)?**
- **<u>Remember</u>: 802.3 supports notion of asymmetric flow control**
  - **The receiving end can choose to ignore PAUSE frames**

Upstream MAC Control PAUSE?

- **Assertion: the existing definition of PAUSE does not discriminate between nodes (stations / ports).**
- **Assertion: a PAUSE request from a single T.E. port would turn off all downstream traffic (not just traffic to that T.E. port)**
- **Opinion: this is not an acceptable behavior**
- **Assertion: "fixing" this requires a change to 802.3 and 802.1**
- **Recommendation: Don't change 802.3x; don't make PAUSE dependent on port type; don't support upstream PAUSE; make this explicit**

Clause 31B -- Backup

31B.1 PAUSE description

The PAUSE operation is used to inhibit transmission of data frames for a specified period of time. *A MAC Control client wishing to inhibit transmission of data frames from another station on the network generates a MA_CONTROL.request primitive specifying:*

a) *The globally assigned 48-bit multicast address 01-80-C2-00-00-01,*

b) The PAUSE opcode,

c) A request_operand indicating the length of time for which it wishes to inhibit data frame transmission.

(See 31B.2.)

The PAUSE operation cannot be used to inhibit transmission of MAC Control frames.

Downstream MAC Control PAUSE?

- **Assertion: H.E. can only use PAUSE globally, not as part of specific T.E. B/W Mgmt (“all of you, take a time out!”)**
- **Assumption: unlike B/W Mgmt, PAUSE Flow Control can be relatively frequent**
- **Assertion: Unless the T.E. PHY definition “mucks with” PAUSE from the H.E., downstream PAUSE should work as designed without modification**
- **Recommendation: Support downstream PAUSE without modification; ensure no changes required; (objective with qualification)**

Initialization – general process

- **Identify T.E. ports**
  - **H.E. directed hybrid polling-contention algorithm**
  - **H.E. should be allowed the option to use memory of prior state to speed up the process**
- **Adjust optical power of all T.E. Tx**
- **Assign slot parameters**
  - **Some, arbitrary T.E. must consume unused slot period**
- **Run**
  - **Allow “late ports” ability to enter (add)**
  - **Allow ports ability to exit (drop) cleanly**

Port Registration – Step 1

1. H.E. polls T.E. ports by attempting to turn on each "known" port in sequence with Slot_Control "signal"

- **H.E. sets slot parameters for "known" port**
  - **Could be really optimistic and do all "known" ports at once -- requires more ingenuity to be consistent with step 2.**
- **T.E. response indicates existence of T.E. to H.E.**
- **H.E. measures & remembers power at H.E. Rx**
  - **Doing this requires H.E. to set slot period so that known port gets entire slot period; all other slots would be off**

Port Registration – Step 2

2. H.E. Requests Registration from “Unknown Ports”

Use of CSMA/CD-like back-off algorithm for discovery; ref 4.1.2.2.

H.E. sets Request_Registration bit in Slot_Control to flag all non-registered (unknown) T.E. ports

- **T.E. ports know because these have no slot allocation**
- **T.E. ports respond during next null slot with short burst**
  - **If no collision, H.E. assigns slot parameters on next Slot_Control**
  - **If collision, H.E. detects collision and sets “Jam bit” in next (and only next) Slot_Control**
    - **T.E. ports “back off” and try again during random null slot**
    - **“Back off” doesn’t have to be fancy since number of T.E’s is relatively small**

Back-off

- **Back-off minimum, and jam timing determined by**
  - **Fiber length, up and back plus budget allocated for H.E. and T.E. processing**
  - **Requires extension be grown to ~10kB (20 km * 5 b/m * 0.8) = 80 microseconds**
  - **During initialization only**

Late Registration (1 of 2)

How and when does a late T.E. port "raise hand?"

- **Option 1: Any time, simply force a collision**
  - **Can't occur without disturbing the operating data flows**
- **Option 2: H.E. opens (reserves) a window on periodic basis to request acknowledgement from all unregistered T.E. ports (null slot)**
  - **Fixed period (within slot period)**
  - **Dynamic, based on BW utilization / allocation**
  - **Fixed or dynamic can be implementation choice**
  - **Under previous assumptions, H.E. Rx does not require additional time for synchronization during or following a null slot**

Late Registration (2 of 2)

- **Recommendation: Option 2, reserved registration window**
  - **It is less efficient than Option 1, but doesn't disrupt other T.E. upstream packet flows (and reliability)**
    - **Efficiency loss is approximately same as back-off: ~10kB or 80 microseconds every slot period**
    - **If slot is set at 5 ms, then this represents ~1.6 % loss in bandwidth**
  - **Optical operation consistent with dropped port requirements**
  - **Method is same as with Step 2 of initialization.**
    - **H.E. sends out Request_Registration "signal" on periodic basis**
    - **T.E. responds with a short burst**
    - **H.E. registers port by setting slot control parameters**

Dropped Port

Given assumptions made before

- **No big deal**
- **H.E. can choose to recover missing slot, or not**
  - **Implementation might want to be intelligent and watch slot usage and recover unused B/W**
  - **Implementation might want to simply ignore the fact that the slot is unused until the same port comes back and requests its B/W back**
- **T.E. should never return and assume it can just resume its previous slot**
  - **Unless the Slot_Control parameters allow it**
  - **Else, it must wait for a Request_Registration and follow the process for a late registration**

Other issues

- **EPON support of peer-to-peer NOT possible if we assume that peer-to-peer means that any port on the shared media must be able to communicate directly with any other port on the media.**
- **<u>Recommendation</u>: identify this in PAR and ask 802.3 and 802 for exception**

PMD/PMA Requirements

- **H.E. Rx optical dynamic range must handle T.E. power range**
- **T.E. Tx turn-on / turn-off timing specifications required**
  - **Maximum time from Tx_disable=1 till lights out**
  - **Maximum time from Tx_disable=0 till bit stream meets specification**
- **H.E. Rx PLL specifications are required for maximum time to recover clock from “lights out” condition and phase recover**
  - **Roll over from the local H.E. Tx PLL**
  - **Includes the H.E. detector and amplification stabilization time max (e.g. jitter conditions on first bits of live slot following “lights out”)**

PMD/PMA Requirements

- **T.E. Tx must "come up" into an off state; there can be no optical glitches during power on or reset**
- **Will rapid Signal_Detect be required?**
- **Jitter budget analysis and feasibility needed ASAP**
- **Initial link budget analysis needed by July**
  - **Must include assumptions on reflections**
  - **Must indicate budget allocation required to support a downstream analog video signal at a different wavelength (cross talk; reflections; noise; etc)**

Want T.E.-Specific Flow Control?

- **PAUSE_PORT is a reasonable addition if desired**
- **Easily supported using features already described for downstream control**
- **Upstream control requires interaction with MAC Client**
- **Will need a new MAC Control Frame**
- **Will need an 802.1 project**

Why this decision is important

- **Assertion: There would need to be a communication path between T.E. MAC Client and the PHY**
  - **T.E. PMD needs to know when to power off and on**
  - **Upstream over subscription will be greater in access space than in enterprise space**
    - **=> dynamic flow control is more important**
    - **For efficient use of slot – especially if the slot duration is dynamically changing – the T.E. MAC Client needs to know slot parameters**
- **This affects many architectural decisions**

Summary

- It should be clear that the initialization / add / drop scheme will work with either a PCS based or a MAC Control based mechanism
- It should also be clear that a path between the MAC Client and the PHY for controlling the slot control engine is required for efficient use of the slot
  - Does the MAC Client tell the PHY when to turn on or off or
  - Does the PHY tell the MAC Client what the slot parameters are?
  - This is especially true if we want PAUSE_PORT function
- Note: it is not a big change to the text of Clause 31 to allow a MAC Control Frame to be both captured by the MAC Control layer AND forwarded to the MAC Client
- Think about it. Till next time….

BACKUP

Initialization – Late Port Power Adj

How to do power adjustment after initial adjustment?

- **Minimal Interference**
  - **All power settings set to minimum during original initialization**
    - **New ports can't be less than the previous least**
    - **No requirement to do them all again**
    - **Generally operates at degraded BER**
- **Maximal simplicity**
  - **Reset all power settings based on new information**
  - **Have scheme that allows reset of all ports simultaneously**
  - **Impacts need study ASAP**
    - **Interference of new port on existing operation (assume it runs at maximum power at minimum distance….)**
    - **Ability to modify power settings during "normal" operation**