Ethernet Timing performance ad hoc report

Adee Ran

Intel

Purpose of this presentation

- Follow up on ran_tp_01_2017_11_15
- Summarize sources of timestamping inaccuracy
- Summarize suggested activity
- Suggest content for draft liaison response

Sources of time stamping inaccuracy (I):
FEC parity blocks

- Ethernet FEC schemes that use block codes have a *constant delay* for Encoder+Decoder and *do not cause packet jitter*.
  - Verified for FECs used in BASE-R, BASE-T, 1000BASE-H PHYs.
- Delay between xMII and MDI and vice versa *is variable*.
  - Taking that into account in min and max delay measurements (per clause 90) creates a *false inaccuracy* in time stamping.
  - Suggestion to correct that is shown below.

Sources of time stamping inaccuracy (II):
Marker insertion and deletion

- Some Ethernet sublayers periodically insert (Tx) and remove (Rx) blocks for alignment/synchronization, and may create room for these blocks by deleting/inserting idle characters, thereby changing the interpacket gaps.
  - This is specified for all multi-lane BASE-R PCS, for 50GBASE-R PCS, and for 25GBASE-R RS-FEC.
- Implementing these functions within a PHY that has xMII as the only interface to the RS creates variable delays and time stamping inaccuracy.
  - The delay variability is 10.24 ns for 25GBASE-R with RS-FEC and 12.8 ns for 100GBASE-R. (other PHY families have lower variabilities)
- There are implementation-dependent methods to avoid any time stamping inaccuracy due to this function, that are compliant and interoperable.
  - Implementations may or may not use these methods. The Ethernet standard does not prescribe implementation details.

Sources of time stamping inaccuracy (III):
Flexible buffering in extender sublayers, repeaters, and non-standard elements

- The 10 Gb/s extender sublayers (XGXS/XAUI, Clause 47) may perform rate compensation and change interpacket gaps, and therefore may cause packet jitter
  - It is suggested to avoid using PHYs that include XGXS/XAUI for applications that are sensitive to timing accuracy.
- The 200 Gb/s and 400 Gb/s extender sublayer (200GXS and 400GXS, clause 118) do not have this option specified.
- Links which include repeaters, or non-standard elements such as protocol-aware retimers, may not keep a constant synchronous data rate between the transmit and receive xMII.
  - Repeaters are defined only for half-duplex Ethernet and are not recommended for new installations (deprecated)
  - Protocol-aware retimers and other non-standard elements are outside of the scope of the 802.3 standard.

Suggested activity

- Add text to clause 90:
  - Specify the delay measurement in each direction when the PHY includes FEC functionality.
  - A note to address the inclusion of delay values resulting from separate FEC sublayers.
  - Proposed text is shown on the next slide.
- Submit a comment to the ongoing P802.3cj to include this in the next revision.

Proposed changes to Clause 90

90.7 Data delay measurement

The TimeSync capability requires measurement of data delay in the transmit and receive paths, as shown in Figure 90–3. The transmit path data delay is measured from the beginning of the SFD at the xMII input to the beginning of the SFD at the MDI output. The receive path data delay is measured from the beginning of the SFD at the MDI input to the beginning of the SFD at the xMII output.

For a PHY that includes a FEC function, the transmit and receive delays are measured for an SFD that is transmitted or received at the start of the FEC block.

The receiver of a multi-lane PHY is expected to include a buffer to compensate for skew between the lanes. This buffer selectively delays each lane such that the lanes are aligned at the buffer output. The earliest arriving lane experiences the most delay through the buffer and the latest arriving lane experiences the least delay through the buffer. The receive path data delay for a multi-lane PHY is reported as if the beginning of the SFD arrived at the MDI input on the lane with the smallest buffer delay.

(…)

NOTE 2—The data delay values represent only the data delay in the PHY sublayers. The TimeSync Client may need to adjust for delays within the gRS. For example, the TimeSync Client may need to subtract the value of the delay associated with the TS_SFD_Detect_TX function from the sum of the minimum transmit data delay values reported by individual MMD(s). Likewise, the TimeSync Client may need to add the value of the delay associated with the TS_SFD_Detect_RX function to the sum of the maximum receive data delay values reported by individual MMD(s).

NOTE 3—For a PHY that includes a FEC as a separate sublayer, the data delay for the FEC sublayer should be included in either the PCS delay registers or the PMA/PMD delay registers of the MMD in which the FEC sublayer is implemented, but not in both.

Suggested content for draft liaison response

- Respond to the 5 concerns raised in the liaison letter (discussed in slides 3-9 of ran_tp_01_2017_11_15):
  - Delay asymmetry is accounted for by using separate Tx and Rx delay measurement, as specified in clause 90.
  - FEC parity insertion creates variability in Tx and Rx delays. However, the total delay through the FEC encode+decode functions is constant, so time synchronization accuracy should not be affected. The measurement of the Tx and Rx delays should be done such that parity insertion would not add any variability. A comment will be submitted to add this specification in Clause 90.
  - Periodical insertion and deletion of markers may introduce delay variability (up to 12.8 ns in the case of 100GBASE-R). There are implementation-dependent methods to avoid any time stamping inaccuracy due to this function, that are compliant and interoperable.
  - Rate compensation is specified as a function of the Reconciliation Sublayer (RS) and does not affect the timing at the xMII. It should not be a concern for time synchronization accuracy.
- Report actions taken
  - An ad hoc was formed to address the concerns raised in the letter and recommend actions.
  - A comment was (will be) submitted to add a specification of time measurement in PHYs that include FEC functions, that will eliminate variability due to parity insertion.
- Address additional requests
  - The 802.3 standard does not enforce limits on delay variations in Ethernet PHYs. It is left for implementation choice.
  - The location of timestamping is specified by clause 90 as the xMII. This is the only reference point that is common across all PHYs. Since accurate timing at this point can be achieved by implementation, changing the location of timestamping is not considered necessary.
  - It is suggested to avoid using PHYs that include XGXS/XAUI for applications that are sensitive to timing accuracy.
  - We do not think further actions by the 802.3 working group are necessary.