Comparison of Synchronization Protocol Aspects Currently Discussed for ResB with IEEE 1588

Geoffrey M. Garner (Consultant)
Gmgarner@comcast.net

SAMSUNG Electronics

IEEE 802.3 RESG 2005 San Jose
2006.01.10

Outline

- Introduction
- Grandmaster selection
- Information needed for synchronization
- Message Semantics
- Summary
- Proposals
- References

Introduction

- The current draft Residential Bridging (ResB) Timing/Synch PAR indicates that the standard will specify the use of IEEE 1588 in the context of IEEE Stds 802.1D and 802.1Q with the purpose of meeting the synch requirements for time sensitive applications such as audio and video across Bridged and Virtual Bridged LANs (both protocols and procedures will be specified)
- The draft PAR also indicates that the standard will leverage the emerging version of IEEE 1588 to develop the additional specs required to address these requirements
- In a recent ResB call, several different logistics for working with IEEE 1588 were discussed
  - The ResB group preferred the approach of individuals submitting work directly to IEEE 1588, with the work essentially being done there
  - The author was asked to communicate this to the IEEE 1588 committee, which was done in the December 1, 2005 IEEE 1588 call

Introduction (Cont.)

- The IEEE 1588 Committee indicated in the December 1 call that it needed more detail on the needs of ResB, e.g., what does “leverage the new version of 1588 to the fullest extent possible” mean?
  - The 1588 Committee said that its PAR for version 2 is already approved, and it needs to understand how the ResB work would fit into this PAR (or if a new or modified PAR would be needed)
  - Once the work is better understood, it can be determined which subcommittee the work would proceed in
- The 1588 Committee asked for a presentation from the ResB group addressing these issues
  - Two presentations were prepared, and are contained in References [1] and [2]
- The presentations [1] and [2] were presented to the 1588 Committee in their December 15, 2005 call
- In [1], the items needed for ResB timing/synchronization were divided into two categories
  - Protocol-related items
  - Profile-related items
- This was done mainly for convenience, as traditionally IEEE 1588 has specified protocol-related items, while profile-related items are specified in other documents specific to each application area

Introduction (Cont.)

- Protocol-related items include (the list is not all-inclusive)
  - Layer 2 message definitions and formats
  - Message information
  - Message semantics
  - Grandmaster (GM) selection algorithm
  - Definition of PTP domain, PTP subdomain, and PTP communication path
- Profile-related items include (the list is not all-inclusive)
  - Phase and frequency compensation algorithms
  - Filter requirements
  - Time-stamp measurement accuracy
  - Clock requirements

Introduction (Cont.)

- A possible, very rough view of a solution was provided in [1]
  - Single PTP sub-domain with one grandmaster (for one ResB)
  - Single clock quality
  - Short messages, defined at layer 2
  - Message exchange procedure is simplified version of current 1588 procedure
  - Algorithm for selecting GM is simplified version of best master clock (BMC) algorithm
  - Phase and frequency compensation algorithm will be one of several being considered
  - Filter and clock requirements TBD (except known now that clock frequency accuracy will be ± 100 ppm)
    - No or inexpensive filtering at intermediate nodes
    - Any expensive filtering limited to end nodes
  - Time stamp measurement accuracy TBD
    - Very likely that measurement will be at least (i.e., at a layer no higher than) between the PHY and MAC
- Somewhat more details on a rough solution were provided in [2]

Introduction (Cont.)

- It was concluded in [1] that the protocol-related items needed for ResB would be within the scope of the current IEEE 1588 v2 PAR
  - Possible logistics were suggested where the protocol-related work would be done in IEEE 1588 and the profile-related work would be done in IEEE 802.1 (and possibly a small amount in IEEE 802.3 if necessary)
  - With these logistics, a new or modified IEEE 1588 PAR would not be needed
- However, while the IEEE 1588 committee did not dispute that a new PAR would not be needed with this division of work (at least, no concerns on this particular item were mentioned in the December 15 call), concerns were expressed as to what impact the ResB work would have on IEEE 1588
  - Concerns were discussed briefly at the end of the December 15 call, and then provided in more detail in [3]
- Concerns included
  - Whether the ResB work would unduly delay the completion of 1588 v2
  - Whether the ResB solution would make the interoperability of 1588 v2 with v1 more difficult than it already is

Introduction (Cont.)

- ❑Three main areas of concern are
  - ▪Message semantics
  - ▪Different frame formats and data types
  - ▪Grandmaster selection algorithm
- ❑[3] indicated that the higher sample rate poses no conflict
- ❑Message semantics concerns
  - ▪Background – ResB will use inexpensive microprocessor and would benefit if separate synch and followup messages were not needed
    - •1588 allows the use of sync only (i.e., not use followup); however, in this case the time stamp in the sync message must refer to that message
      - –This would require an accurate time stamp measurement, likely at the PHY, and might be expensive for ResB
    - •An alternative would be to allow the time stamp of the sync message to refer to the previous sync
      - –Concern that this is not compatible with 1588 v1, and would pose a problem for 1588 applications that have longer sync intervals (e.g., greater than 1 s)

Introduction (Cont.)

- Concerns on different frame formats and data types
  - ResB will use Layer 2 (L2) messages
  - While there has not yet been very extensive L2 discussion in 1588, the discussion and proposals so far have suggested using a frame format with the user portion of the frame the same as for Layer 3 (L3), and any differences confined to what is now the UDP/IP header area (see Reference [4] for latest proposal)
  - Concern also expressed with changes to data types; [3] stated that there exist numerous hardware and firmware designs with current data types
  - Note, however, that this may be impacted by the outcome of the Short Frames work (see [5] – [8])
- Concerns on grandmaster selection algorithm
  - [1] and [2] did not discuss in detail any changes needed to GM selection algorithm; general concern expressed in [3] on what changes might be needed
  - [3] indicated that ResB needs can probably be met with existing GM selection algorithm

Introduction (Cont.)

- The following slides look in more detail at the above 3 areas of concern, comparing
  - What is done in 1588 v1 and, where relevant, what is being considered in v2
  - What has been discussed in the ResB group, mainly in the context of References [9] – [12]
- The areas of concern are
  - Message semantics
  - Different frame formats and data types
  - Grandmaster selection algorithm

Grandmaster Selection Algorithm

- Grandmaster selection in 1588 is specified in [13] (mainly in Clause 7.6); a summary description is given in [14]
- GM selection is accomplished by running Best Master Clock algorithm at each clock and each port of each clock
  - Each clock computes its own state
  - Result is a synchronization distribution hierarchy
    - One master produced for each PTP communication path
    - One GM for each PTP subdomain
    - In general, may have multiple subdomains (but a ResB network will consist of one subdomain)
  - IEEE 1588 allows multi-point, half-duplex links (PTP communication paths)
    - One master, multiple slaves
    - Note that ResB links will be full-duplex, point-to-point (one master, one slave)
  - IEEE 1588 defines a hierarchy of clock qualities through stratum number (0 through 4 and 255), clock identifier (nature and expected accuracy, and epoch), and variance (Allan variance multiplied by integration time divided by 3)
    - In general, clock quality is determined first by stratum number (lower is better), then by clock identifier, and then by variance (smaller is better)
    - In ResB, there will likely be a single single clock quality with relevant requirements specified
  - IEEE 1588 allows clocks to be designated as preferred; these will be included in the set from which the GM is selected

Grandmaster Selection Algorithm (Cont.)

- IEEE 1588 maintains state information for each clock and for each port of each clock in 4 datasets for the clock as a whole and 2 data sets for each port of the clock (datasets are described in detail in Clause 7.4 of [13])
- Can designate a node as slave only; in that case it will not be chosen as master (or GM; see Clause 9.2.2 of [13])
- BMC algorithm consists of 2 parts
  - Dataset comparison algorithm
  - State decision algorithm
- Dataset comparison algorithm: compares data corresponding to 2 different clocks to determine which of those clocks is best. The comparison is made hierarchically as follows (see slide 59 of [14] for a description):
  - If one of the clocks is preferred, that clock is chosen
  - Next, if one clock has lower stratum number, that clock is chosen
  - Next, if one clock has better identifier, that clock is chosen (see [13] for clock identifiers)
  - Next, if one clock has smaller variance, that clock is chosen
  - Next, if one clock is closer to its GM or has heard from its GM more recently, that clock is chosen
  - Next, if one clock has smaller UUID, that clock is chosen
  - Next, if one dataset indicates smaller port number, that dataset is chosen

Grandmaster Selection Algorithm (Cont.)

- Dataset comparison algorithm is run on each port of each clock to determine which of the clocks connected to that port are best
  - In ResB, there will only be one connected clock as all links are FDX
- Dataset comparison algorithm is then run on all the best clocks selected for each port, to determine which port is connected to the best clock (i.e., the best of the best)
- State decision algorithm computes recommended state for each port of each clock and for each clock as a whole, given as input the best clock on each port, the best of the best clocks, and the local clock datasets
  - E.g., if a given clock finds that the best of the best datasets received from all its ports is better than it, then it becomes a slave to the best clock on that port
    - Other ports will either be master ports or passive ports
- If a subdomain has either no or 1 Stratum 1 or 2 clocks and no clocks with stratum > 2 that are designated as preferred, there will be exactly 1 GM
- If there are N>1 Stratum 1 or 2 clocks and no clocks with stratum >2 that are designated as preferred, there will be N subdomains, each with 1 GM
- If there are N>0 Stratum 1 or 2 clocks and additional clocks with stratum >2 that are designated as preferred, there will be at least N subdomains, each with 1 GM (there may be more than N subdomains, depending on the topology and the number of preferred clocks)

Grandmaster Selection Algorithm (Cont.)

- For ResB, the only GM selection discussed so far, besides the GM selection of IEEE 1588, has been that in [10] (see Section 7.1.5)
- Every ResB station is capable of becoming GM, but only 1 is selected
- Selection is based on preference value
- Preference value is based on the following hierarchy of components
  - System tag (this may be set to indicate a desired preference for a particular clock)
  - Unique ID, e.g., MAC address (analogous to uuid_field in IEEE 1588)
  - Hops count, i.e., distance from current GM)
  - Port tag, i.e., port id

Grandmaster Selection Algorithm (Cont.)

- Appears that IEEE 1588 GM selection algorithm can be made almost equivalent to algorithm of [10] if the following is done
  - Set all clock stratum numbers to 3 (or higher)
    - In this case clock identifier and variance fields are not relevant in algorithm
  - Allow one clock to be designated as preferred
- If the above is done, a ResB will end up as a single subdomain with one GM
- ResB devices should not be required to execute the branches of the grandmaster selection (BMC) algorithm (i.e., the dataset comparison and state decision algorithms) that pertain to stratum 1 or 2 clocks, as these branches will never be invoked
- The remaining differences between the 2 algorithms are
  - The algorithm of [10] uses a 16-bit system tag, and therefore allows 2^{16} different preferences (possibly 2^4; not clear if user can set last 12 bits)
  - The algorithm of [10] has UUID above distance from the GM in the preference hierarchy

Different Frame Formats and Data Types

- Message formats and data types in IEEE 1588 [13] and current proposal for L2 differ from message formats and data types in [10]
- The differences in representation of data are not addressed here, as arguments as to which is better depend heavily on implementation and existing hardware and firmware
  - We defer this discussion to a later date
    - As part of this, we defer discussion of how time is represented, e.g., precision, data fields, epoch, inclusion or non-inclusion of leap seconds, etc.
- However, the frequency and phase compensation algorithm of [9] and [10] (and simulated in [15]) uses additional information not transported in IEEE 1588
  - We do describe this here
- 1588 sync and delay_req messages contain an originTimestamp field, which is the measured time that the time stamp leaves, relative to the local clock after any corrections or compensation algorithms have been applied
  - E.g, if PLLs or servos are used, the time is measured relative to the clock phase time after any smoothing

Different Frame Formats and Data Types (Cont.)

- In contrast, [9], [10], and [15] use a particular phase and frequency compensation algorithm that requires
  - Time relative to the free running local oscillator (baseTime)
  - Time relative to the corrected frequency (flexTime; note that the corrected frequency can be calculated with software or firmware)
  - Cumulative time difference from the GM
  - Cumulative frequency difference from the GM
- All the above fields are contained in the sync messages

Different Frame Formats and Data Types (Cont.)

- However, reference [11] describes two phase and frequency compensation algorithms
  - Offset and Frequency Compensated Clock (OFCC)
    - This is similar to the algorithm of [9], [10], and [15], but needs only times relative to the corrected frequency (it assumes the IEEE 1588 message exchange format)
    - In principle, it should be equivalent to algorithm of [9], [10], and [15]; if each slave measures its time relative to its master's compensated time, then the resulting frequency is relative to the GM (rather than the immediate master) and the cumulative frequency need not be transported)
    - Enhancement described in [11], which does not require the transport of additional information
  - Frequency Compensated Clock (FCC)
    - This algorithm was first described in [16]
    - Several enhancements described in [11]; one of the enhancements requires the transport of cumulative frequency offset
  - Reference [11] contained experimental results for OFCC and simulation results for FCC; performance with the enhancements described there was promising
- Therefore, it may be possible to omit some or all of the additional fields
- If OFCC is used, the existing IEEE 1588 frame format can be used
  - If FCC is used, then only the cumulative frequency offset must be added to the information transported

Message Semantics

- IEEE 1588 uses a one-way with less frequent two-way messaging scheme
  - Master sends sync to slave with its current time
  - If configured for followup, master will send followup at later time with better estimate of time sync was sent
  - Less frequently, slave sends delay_req to master with its current time
  - Master sends delay_resp back to slave, with time it received the delay_req
  - Sync (and followup messages if configured for followup) allow computation of master to slave delay
  - Delay_req and delay_resp allow computation of slave to master delay
  - Sync (and followup messages if configured for followup) and both delays allow computation of time offset for slave (under the assumption that the delays are the same in both directions)
    - Note that slave to master and master to slave delays are saved and used in successive offset computations until new delays are computed; this is valid because the delays change very slowly

Message Semantics (Cont.)

- The algorithm of [9], [10], and [15] uses a two-way messaging scheme
  - Slave send messages to master at nominal rate; time in each message refers to the message two messages ago
  - Master notes when it receives each successive message
  - Master responds to slave with time it sent previous response, and times of sending and receipt of message from slave two messages ago
  - Slave notes when it receives each message from master
  - Slave computes its time offset from the four times in the response message
    - Equivalent to computing both offset and delays each time a message is received
  - Having message times refer to previous messages is analogous to use of followup in IEEE 1588
- The compensation scheme of [9], [10], and [15] can work with both the two-way message scheme and the one-way with less frequent two-way message scheme of IEEE 1588
  - This is because the propagation delays are largely static

Message Semantics (Cont.)

- However, a bigger issue is the concern raised in [3]
- ResB will very likely need to have message time stamps refer to previous messages
  - Having a time stamp refer to the current message will likely require that the measurement be made at the PHY, which will likely be expensive
  - If the 1588 scheme is used, this means that the followup capability must be used
  - However, this means that two messages will be sent at each sync interval (both sync and followup)
  - ResB will require a short sync interval to obtain acceptable jitter and wander performance for end-to-end applications
    - Sync interval will be relatively small (i.e., 10 ms or less)
  - It therefore would be desirable to send only one message from master to slave at each sync interval
    - A suggestion was made that only sync be sent, but that its time stamp refer to the previous sync message
  - However, it was indicated in [3] that this would pose a problem for applications (i.e., non-ResB applications of IEEE 1588) that use a long sync interval, e.g., 1 s or longer
    - For these applications, the time stamp contained in a sync message would be 1 s old; this could result in unacceptable performance (i.e., excessive wander)

Message Semantics (Cont.)

- Note that this issue does not directly affect interoperability with v1 IEEE 1588 applications, because v1 covers only L3 operation
  - ResB will operate at L2, and L2 is not standardized in v1
- However, any non-ResB v2 L2 applications that have a long sync interval would be affected if the sync time stamp referred to the previous sync message
- One possible solution might be to add an option for L2 sync messages in v2 to refer to either previous or current sync message
  - An additional field could indicate which option is being used
    - One possibility would be to use one of the reserved flag bits (in conjunction with the PTP_ASSIST bit that indicates whether or not a clock is followup capable)
- While this would mean a change for v2 messages at L2 relative to current L2 proposal in [4], it should be noted that short frames work in [5] – [8] may also result in a change to message format
  - In any case, this would need only apply to L2 (at least, to satisfy the needs of ResB) and not L3

Summary

- Appears that IEEE 1588 GM selection algorithm can be made almost equivalent to algorithm of [10] if all clock stratum numbers are set to 3 (or higher) and allow one clock to be designated as preferred
  - Differ in number of possible GM preference values and precedence of UUID versus distance from GM in determining BMC
- Current compensation scheme of white paper needs both free-running and compensated phase values, as well as cumulative phase and frequency values, in messages; IEEE 1588 needs only compensated phase values
  - OFCC scheme in [11] is likely equivalent to [9] and [10], and needs only the information transferred in IEEE 1588
- Compensation scheme of [9] and [10] can work with IEEE 1588 one-way scheme (and OFCC and FCC schemes of [11] work with 1588)
- Main difference in message semantics is due to need for ResB to have short sync interval; low cost requirement implies having sync refer to previous sync
  - Suggested possible solution of allowing this as an option for Layer 2

Proposals

- Use IEEE 1588 GM selection algorithm, with all clock stratum numbers set to 3 and allow one clock to be designated as preferred
- Further investigate the performance and stability of the OFCC scheme of [11], because if this scheme is acceptable it can be used with the existing IEEE 1588 messages
- For Layer 2 operation in IEEE 1588 v2, add an option to have the sync message time stamp refer to the previous sync message

References

1. *Logistics for ResB Timing/Synchronization Work*, prepared by Geoffrey M. Garner and edited in ResB conference calls, Rev. 2, December 14, 2005.
2. *Residential Ethernet: Time-of-Day Timer Synchronization*, maintained by David V. James, latest version is December 17, 2005 (December 12, 2005 presented in IEEE 1588 call on December 15, 2005).
3. John Eidson, Comments on Residential Ethernet Discussion of December 15, 2005, email attachment sent to IEEE 1588 reflector December 19, 2005.
4. Matthias Wenk, *PTP on Layer-2*, Contribution to Precise Networked Clock Synchronization Working Group – IEEE 1588 Revision, December 22, 2005.
5. Ron Cohen, Silvana Rodrigues, *Unifying Short and Long Messages*, Contribution to Precise Networked Clock Synchronization Working Group – IEEE 1588 Revision, November 30, 2005.
6. John Eidson, Comments on Short Frames Discussion of December 15, 2005, email attachment sent to IEEE 1588 reflector December 19, 2005.
7. David Tonks, Comments on Short Frames Discussion of December 15, email sent to IEEE 1588 reflector December 20, 2005.
8. Ron Cohen, *Comments on Unifying Short and Long Messages*, Contribution to Precise Networked Clock Synchronization Working Group – IEEE 1588 Revision, December 28, 2005.
9. David V. James and other contributors, *Residential Ethernet (RE) (a working paper)*, November 16, 2005 (latest version).

References

10. David V. James, Extract of material on timing/synchronization from [9] and possible update and addition of detail, December 4, 2005. Available via http://dvjames.com/esync/dvjTimeSync2005Dec04.pdf.
11. Sihai Wang, *Improvements to Boundary Clock Based Time Synchronization Through Cascaded Switches*, Samsung presentation at November, 2005 IEEE 802.3 ResE SG meeting, Vancouver, BC, November 14, 2005. Available via http://www.ieee802.org/3/re_study/public/index.html.
12. Geoffrey M. Garner, *Summary of Potential Choices for Synchronization Transport in Residential Ethernet*, Samsung presentation at July, 2005 IEEE 802.3 ResE SG meeting, San Francisco, CA, July 18, 2005. Available via http://www.ieee802.org/3/re_study/public/index.html.
13. IEEE Std™ 1588-2002, *IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems*, IEEE, November 8, 2002.
14. John Eidson, *IEEE-1588 Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems,* Tutorial, September 27, 2004.
15. Geoffrey M. Garner and Kees den Hollander, *Analysis of Clock Synchronization Approaches for Residential Ethernet*, 2005 IEEE 1588 Conference, Winterthur, Switzerland, October 10 – 12, 2005. Paper and VGs available via http://www.ieee802.org/3/re_study/public/index.html.
16. S. Balasubramanian, K.R. Harris, and A. Moldovansky, *A frequency compensated clock for precision synchronization using IEEE 1588 protocol and its application to Ethernet*, Workshop on IEEE 1588, 2003