7. Clock synchronization model for a bridged local area network (Informative)

7.1 General

This clause provides a model for understanding the operation of the generalized precision time protocol (gPTP), which specifies the operation of time-aware systems on a bridged LAN. Although this standard is based on the precision time protocol (PTP) described in IEEE Std 1588-2008 (and, indeed, is a proper profile of 1588 in particular configurations) there are differences which are summarized in clause 7.4.

Although this standard has been written as a stand-alone document, it is useful to understand the 1588 architecture as described in clause 6 of that document.

7.2 Architecture of a Time-aware Bridged Local Area Network

A time-aware bridged local area network consists of a number of time-aware systems interconnected by LANs that support the generalized Precision Time Protocol (gPTP) defined within this standard. A set of time-aware systems that are interconnected by gPTP-capable LANs is called a gPTP domain. There are two types of time-aware systems:

1. time-aware endpoints, one of which is the grandmaster (the source of time information), and

2. time-aware bridges, which receive time information from the grandmaster (perhaps indirectly through other time-aware bridges), apply corrections to compensate for delays in the LAN and the bridge itself, and retransmit the corrected information.

The LANs must have an IEEE 802 architecture, and there must be a way to precisely measure the message transit delay from one time-aware system to another. This standard defines mechanisms for delay measurements using standard-based procedures for:

1. IEEE Std 802.3 Ethernet using full-duplex point-to-point links (clause 11),

2. IEEE Std 802.3 Ethernet using passive optical network (EPON) links (clause 13),

3. IEEE Std 802.11 wireless (clause 12), and

4. generic coordinated shared networks (CSNs, e.g. MoCA and G.hn) (Annex E).

Figure 7-1 illustrates an example time-aware network using all those network technologies, where endpoints on several local networks are connected to a grandmaster on a backbone network via an EPON access network.

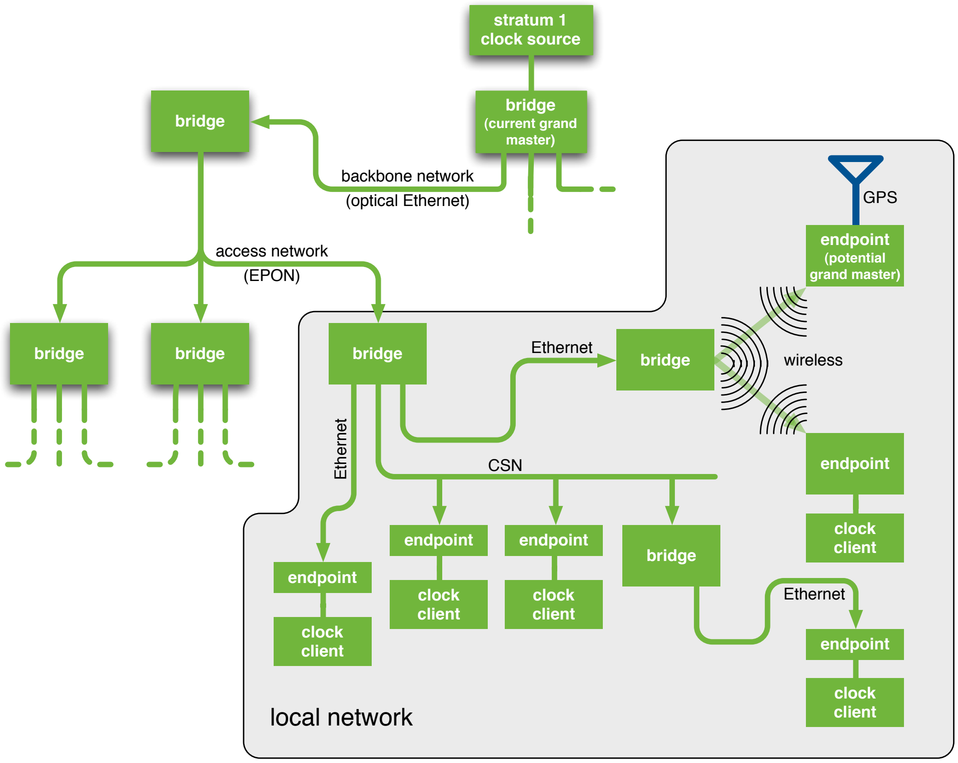

Figure 7-1. Time aware network example

Any time-aware system with clock sourcing capabilities can be a potential grandmaster, so there is a selection method (the "best master clock algorithm", or BMCA) which ensures that all of the time-aware systems in a gPTP domain use the same grandmaster.[1] The BMCA is largely identical to that used in 1588, but somewhat simplified. In Figure 7-1 the BMCA process has resulted in the grandmaster being on the network backbone. If, however the access network fails, the systems on a local network will automatically switch over to using one of the potential grandmasters on the local network that is as least as "good" as any other. For example, in Figure 7-2, the access network link has failed, so a potential grandmaster that has a GPS reference source has become the active grandmaster, and there are now two gPTP domains where there used to be one.

[1] There are, however, short periods during network reconfiguration when more than one grandmaster might be active while the BMCA process is taking place.

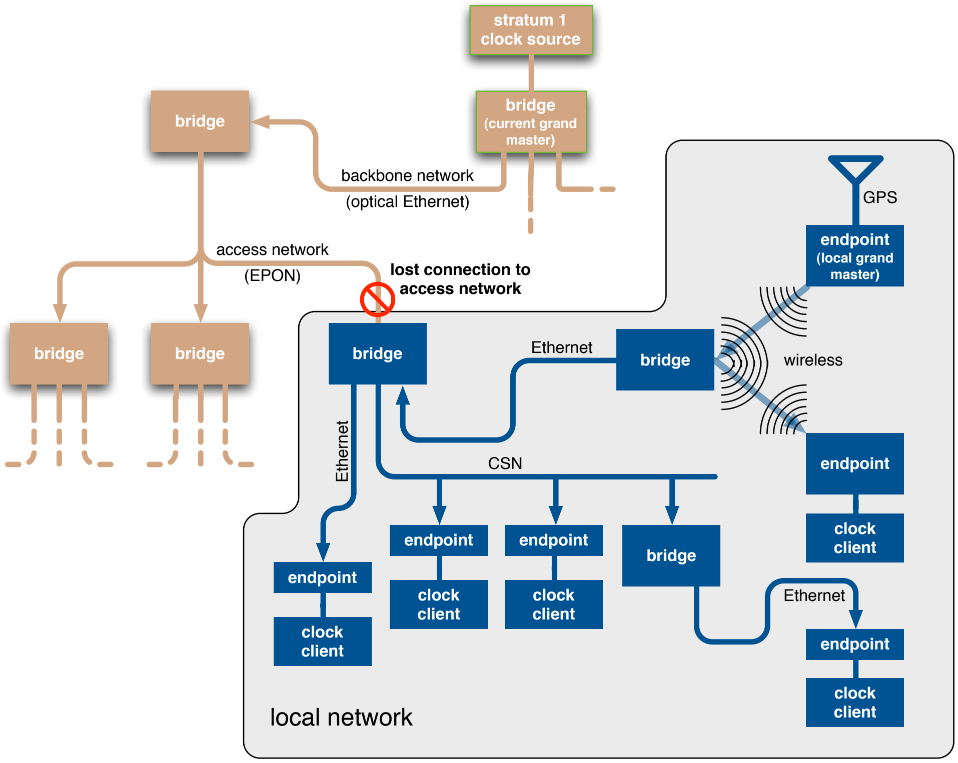

Figure 7-2. Time aware network of figure 7-1 after an access network link failure.

7.3 Time synchronization

Time synchronization in gPTP is done the same way (in the abstract) as is done in 1588: a grandmaster sends information including the current time-of-day to all directly attached time aware systems. Each of these time-aware systems must correct the received time by adding the propagation time needed for the information to transit the communication path from the grandmaster. If the time aware system is a time aware bridge, then it must forward the corrected time information (including additional corrections for delays in the forwarding process) to all the other attached time aware systems.

To make this all work, there are two time intervals that must be precisely known: the forwarding delay (called the "residence time"), and the time taken for the time-of-day information to transit the communication path between two time aware systems. The residence time measurement is local to a bridge and easy to compute, while the communication path delay is dependent on many things including media dependent properties and the length of the path.

7.3.1 Delay measurement

Each type of LAN or communication path has different methods for measuring propagation time, but they are all based on the same principal: measuring the time that a well known part of a message is transmitted from one device and the time that the same part of the same message is received by the other device, then sending another message in the opposite direction and doing the same measurement as shown in figure 7-3.

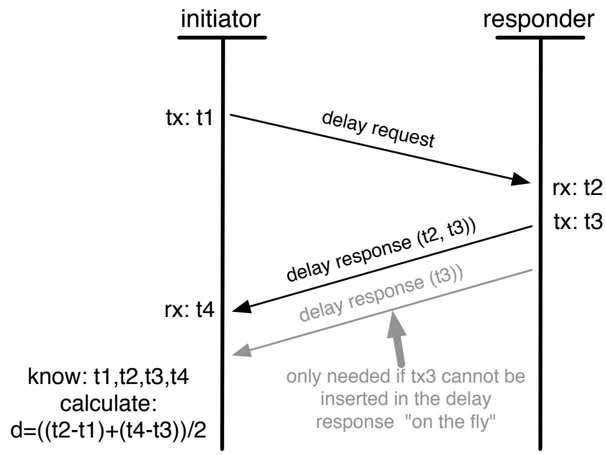

Figure 7-3. Example delay measurement

This basic mechanism is used in the various LANs in the following ways:

1. Full-duplex Ethernet LANs use the two step peer-to-peer path delay algorithm as defined in 1588, where the messages are called Pdelay, Pdelay_resp, and Pdelay_resp_follow_up.

2. 802.11 wireless LANs use the time measurement procedure defined in 802.11v, where the messages are the "timing measurement action frame" and its corresponding "ACK".

3. EPON LANs use the discovery process, where the messages are "GATE" and "REGISTER_REQ".

4. CSNs either use the same mechanism as full-duplex Ethernet, or use a method native to the particular CSN (similar to the way native methods are used by 802.11 and EPON).

7.3.2 Logical syntonization

The time synchronization correction described above is dependent on the accuracy of the delay and residence time measurements. If the clock used for this purpose is frequency locked (syntonized) to the grandmaster, then all the time interval measurements will use the same time base. Since actually adjusting the frequency of an oscillator (e.g., using a PLL) is slow and prone to gain peaking effects, time-aware bridges correct time interval measurements using the grand master frequency ratio.

Each time-aware system measures, at each port, the ratio of the frequency of its own clock to the frequency of the time-aware system at the other end of the link attached to that port. The cumulative ratio of the grandmaster frequency to the local clock frequency is accumulated in a standard organizational TLV attached to the Follow_Up message. The frequency ratio of the grandmaster relative to the local clock is used in computing synchronized time, and the frequency ratio of the neighbor relative to the local clock is used in correcting the propagation time measurement.

The grandmaster frequency ratio is measured by accumulating neighbor frequency ratios for two main reasons. First, if there is a network reconfiguration and a new grandmaster is elected, the nearest neighbor frequency ratios do not have to be newly measured as they are constantly measured using the Pdelay messages. This results in the frequency offset relative to the new grandmaster being known when the first Follow_Up message is received, which reduces the duration of any transient in synchronized time during the reconfiguration. This is beneficial to many high-end audio applications. Second, there are no gain peaking effects because an error in frequency offset at one node, and resulting residence time error, does not directly affect the frequency offset at a downstream node.

7.3.3 Grandmaster (best master) selection and network establishment

All time-aware systems participate in best master selection so that the 802.1AS protocol can determine the synchronization spanning tree. This synchronization spanning tree may be different from the forwarding spanning tree determined by IEEE 802.1 Rapid Spanning Tree Protocol (RSTP) since the spanning tree determined by RSTP may not be optimal, or even adequate for synchronization.

gPTP requires that all bridges and end-stations in the gPTP domain be time-aware-systems, i.e., the protocol will not transfer timing over “ordinary bridges” (those that meet the requirements of IEEE Std 802.1D or IEEE Std 802.1Q, but do NOT meet the requirements of this standard). A time-aware system uses the peer delay mechanism on each port to determine if an “ordinary bridge” is at the other end of the link or in between itself and the Pdelay responder. If, on sending Pdelay_Req

1. no response is received,

2. multiple responses are received, or

3. the measured propagation delay exceeds a specified threshold, the protocol concludes that an “ordinary bridge” or end-to-end TC is present.

In this case, the link attached to the port is deemed not capable of running gPTP and BMCA ignores it. However, the port continues to attempt the measurement of propagation delay using the peer delay mechanism (for full-duplex, 802.3 links), xxx MPCP messages (for EPON), or 802.11v messages (for 802.11 links), and periodically checks whether the link is or is not capable of running 802.1AS.

<<editor’s note: it’s not clear how EPON will participate here … I don’t think we can ignore this. Also, should I put in a few words about CSN discovery>>

7.3.4 Energy efficiency

Sending PTP messages at relatively high rates when there is otherwise little or no traffic is counter to the goal of reducing energy consumption. This standard specifies a way to request that a neighbor port reduce the rate of sending Sync (and Follow_Up), peer delay, and Announce messages, and also to inform the neighbor not to compute neighbor rate ratio and/or propagation delay on this link. A time-aware system could do this when it enters low-power mode, but this standard does not specify the conditions under which this is done; it specifies only the actions a time aware system takes.

7.4 Time aware system architecture

The model of a time-aware system is shown in figure 7-4.

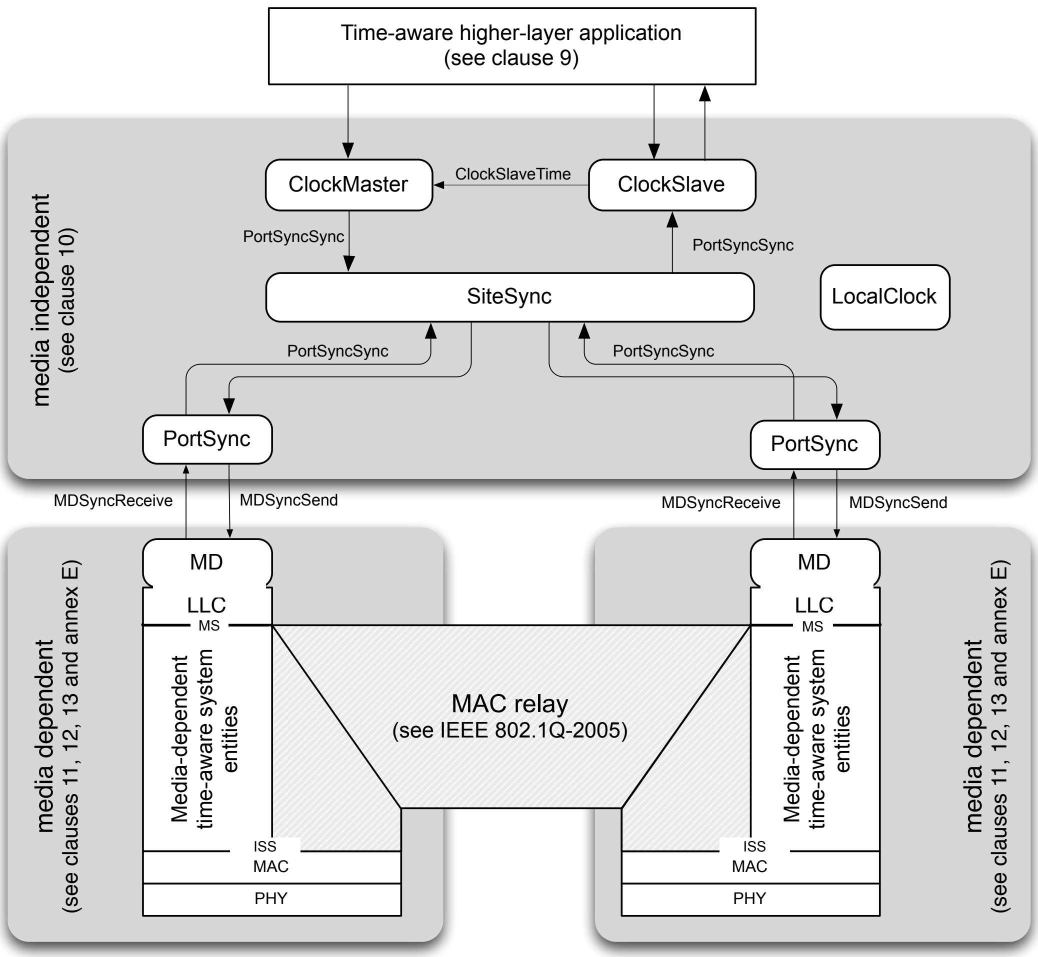

Figure 7-4. Time-aware system model

A time-aware system consists of the following major parts:

1. If the time-aware system includes application(s) that either use or source time information, then they interface with the gPTP information using the service interfaces specified in clause 9.

2. A single media-independent part that consist of ClockMaster, ClockSlave, SiteSync logical entities and one or more PortSync entities. The BMCA and forwarding of time information between abstract ports and the ClockSlave and ClockMaster is done by the SiteSync entity, while the computation of port-specific delays needed for time synchronization correction is done by the PortSync entities.

3. Media dependent ports, which translate the abstract “MDSynchSend” and “MDSyncReceive” structures sent to or received from the media independent layer and corresponding methods used for the particular LAN attached to the port.

A. In the case of full-duplex Ethernet ports, 1588 Sync and Follow_Up messages are used, except for the use of different addresses and ethertypes and with an additional TLV in the Follow_up used for rate ratio communication. The path delay is measured using the two-step 1588 Pdelay mechanism. This is defined in clause 11.

B. For 802.11 ports, timing information is communicated using the MAC Layer Management Entity to request a “timing measurement” (as defined in IEEE 802.11v) that also sends everything that would be included in the Follow_up message for full-duplex Ethernet. The timing measurement result includes all the information to determine the path delay. This is defined in clause 12.

C. EPON **<< editor’s note: TBD, how about some help, Frank/Yuanqiu? >>** This is defined in clause 13.

D. CSN **<< editor’s note: TBD, how about some help, Philippe? >>** This is defined in Annex E.

7.5 Differences between gPTP and 1588 PTP

1. gPTP assumes all communication between time-aware systems is done only using 802 MAC PDUs and addressing, while 1588 supports various layer 2 and layer 3-4 communication methods.

2. gPTP specifies a media-independent sublayer that simplifies the integration within a single timing domain of multiple different networking technologies with radically different media access protocols. The information exchanged between time-aware systems has been generalized to support different packet formats and management schemes appropriate to the particular networking technology. 1588, on the other hand, is fully specified only for Ethernet-type LANs and similar technology.

3. In gPTP there are only two types of time-aware systems: end-points and bridges, while 1588 has ordinary clocks, boundary clocks, end-to-end transparent bridges and peer-to-peer transparent bridges. A time-aware endpoint corresponds to a 1588 ordinary clock, and a time-aware bridge is a type of 1588 boundary clock where its operation is very tightly defined … so much so that a time-aware bridge with Ethernet ports can be shown to be mathematically equivalent to a peer-to-peer transparent bridge, as shown in clause 11.1.3.

4. Time-aware systems only communicate gPTP information directly with other time-aware systems. I.e, a gPTP domain consists ONLY of time-aware systems. Non-time-aware bridges cannot be used to relay gPTP information. In 1588 it is possible to use non-1588-aware bridges in a 1588 domain, although this will slow timing convergence and introduce extra jitter that must be filtered by any 1588 clock.

5. For Ethernet full-duplex links, gPTP requires the use of the peer delay mechanism, while 1588 also allows the use of end-to-end delay measurement.

6. For Ethernet full-duplex links, gPTP requires the use of two-step processing (use of Follow_up and Pdelay_resp_follow_up messages to communicate timestamps), while 1588 allows single step processing (embedding timestamps in messages “on the fly” as they are being transmitted).

7. In steady state, there is only a single active grandmaster in a time-aware network. I.e., there is only a single gPTP domain, whereas 1588 allows multiple overlapping timing domains.

8. All time-aware systems in a gPTP domain are logically syntonized, meaning that they all measure time intervals using the same frequency. This is done by the process described in x, and is mandatory. Syntonization in 1588 is optional, and the method used is not as direct and takes longer to converge.

9. The BMCA used in gPTP is the same as that used in 1588 with the following exceptions: (i) Announce messages received on a slave port that were not sent by the receiving time-aware system are used immediately, i.e., there is no foreign-master qualification, (ii) a port that the BMCA determines should be a master port enters the master state immediately, i.e., there is no pre-master state, (iii) the uncalibrated state is not needed and, therefore, not used, and (iv) all time-aware systems are required to participate in best master selection (even if it is not grandmaster capable).

10. Finally, this standard includes a formal services definition for the time-aware applications. (See clause 9.) 1588 does not define how an application provides or obtains time information.