IEEE 1588 BMCA Re-arrangement Times

Peter Roberts
Alcatel-Lucent
Peter.roberts@alcatel-lucent.com

Geoffrey M. Garner
Consultant
gmgarner@alum.mit.edu

IEEE 802.1 TSN TG
2013.07.15

Introduction

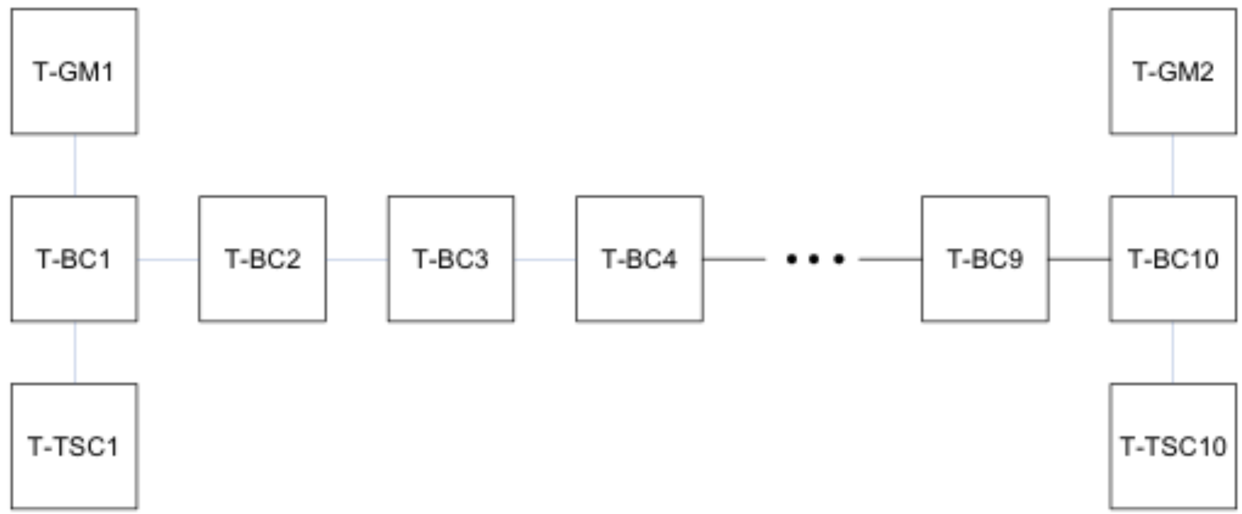
- This presentation describes work contained in Reference [1], which was presented and discussed in the July 1 – 12 ITU-T SG 15, Q13 meeting
- An Excel spreadsheet was developed that models and calculates the IEEE 1588 Best Master Clock Algorithm (BMCA) reconfiguration time for a variety of assumptions on (see slides 5 – 7 for more details):
  - Foreign master qualification
  - PRE_MASTER qualification
  - Triggering a state decision event (SDE) after receipt of Announce
  - Sending of Announce after BMCA executes
  - Time spent in UNCALIBRATED state
- Version of reference [1] is attached
  - Reference [1] had the v7 version of the spreadsheet, which is not attached
  - Instead, an updated v8 version of the spreadsheet is attached, which contains the additions done after the ITU-T Q13/15 meeting (see slide 5)

Topology re-arrangement model

- Desire to know the worst-case* re-arrangement time for a chain of Boundary Clocks
- Delays based on procedures of 1588 that trigger dataset changes and port state changes
- Interpretation of standard to get maximum wait times between changes
- Target was to have new topology configured within 60 seconds
- Parameters to adjust were AnnounceInterval and AnnounceReceiptTimeout

- * ITU-T bases performance on worst case conditions

Reference Topology



Linear chain of 10 and 20 Boundary Clocks

T-GM1 initially elected grandmaster; T-GM2 passive; Synchronization flows from T-BC1through to T-BC10/20

T-GM1 failure triggers a change to T-GM2 as grandmaster; Synchronization flows from T-BC10/20 through to T-BC1

Procedures Modeled – 1

- Detection of transport communication failure
- Update of information received from the parent clock
- Transmission of grandmaster information
- Triggering of State Decision Event
- Update of information received from a clock other than the parent (foreignMaster)
- PRE_MASTER state

- Added after ITU-T meeting
- Delay in updates due to UNCALIBRATED state (max with PRE_MASTER)
  - Updated version of the spreadsheet (v8) is attached to the pdf version of this presentation; it includes the above addition
  - **Macros must be enabled in the spreadsheet**

Procedures Modeled - 2

- Foreign master qualification
  - May turn off by setting FOREIGN_MASTER_THRESHOLD to 1
- PRE_MASTER qualification may be turned on or off
  - If turned on, a port remains in the PRE_MASTER state for a time equal to 1+stepsRemoved Announce intervals before transitioning to the master state, for the case where the port is not a port on the GM
- May trigger an SDE after receipt of Announce
  - Immediately
  - After one Announce interval
  - After a random time uniformly distributed between 0 and one Announce interval

Procedures Modeled - 3

- May send Announce (after running BMCA)
  - Immediately
  - After one Announce interval
  - After a random time uniformly distributed between 0 and one Announce interval
- May set a time interval that a port remains in the UNCALIBRATED state
  - If the interval is non-zero, the port remains in the UNCALIBRATED state for the longer of
    - The specified time interval
    - The time the port would remain in the PRE_MASTER state

References

[1] Peter Roberts, Geoffrey M. Garner, and Lv Jingfei, *Re-arrangement times for G.8271.1 HRM topologies*, Alcatel-Lucent and Huawei contribution to ITU-T SG 15, Q13, Geneva, July 1 – 12, 2013, COM 15 – C 275 – E

Re-arrangement times for G.8271.1 HRM topologies

Reference Architecture

In order to analyze the worst-case re-arrangement times the topology assumes a chain of T-BCs where the T-GMs are located at each end of the chain. In the default behaviour of one active grandmaster, or in the new model of multiple active grandmasters but with topological control using port priorities, the model studies how the loss of the active grandmaster is detected and reported (i.e., how the information that the active GM is lost is propagated via Announce messages) through the chain of T-BCs to the second T-GM, and then how the T-BCs re-arrange to propagate timing from the new grandmaster.

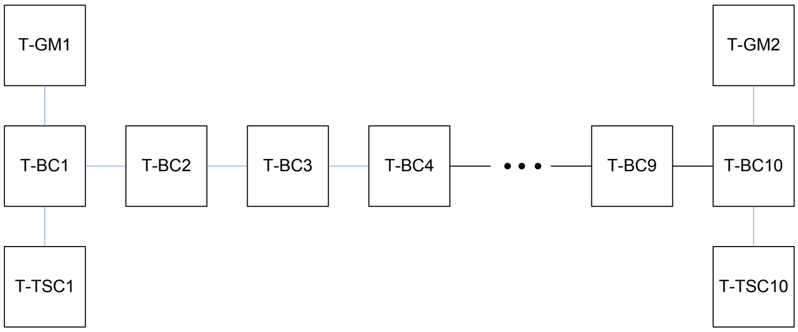


Figure 1 – 10 Node network Topology

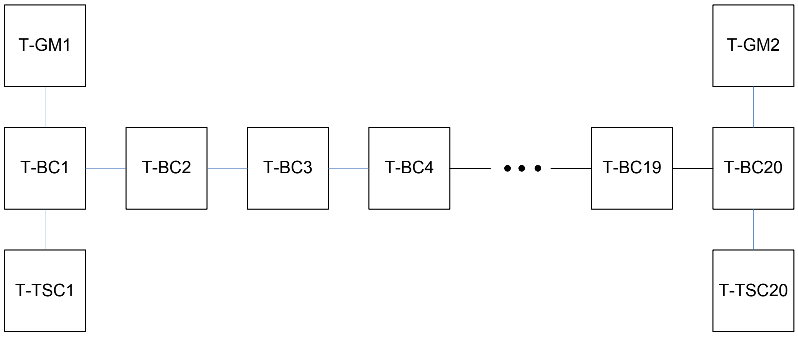


Figure 2 – 20 Node network Topology

Initial Conditions

For the studies, it is assumed that the quality information of T-GM1 and T-GM2 are identical so that the deciding factor of which T-GM is the grandmaster is the clockIdentity. For the purposes of this model, it is assumed T-GM1 has the lower clockIdentity than T-GM2. This means T-GM1 shall be a better clock than T-GM2 based on the data set comparison algorithm.

In addition, the T-TSCs are assumed to be clockclass of 255 (Slave-only) and they have only one possible parent clock, which is the directly attached T-BC.

Note: The alternate BMCA of WD114SanJose includes a localPriority parameter for a port. This could be used to force the topology to prefer the flow of Synchronization from T-BC1 through to T-BC10. In this case, the initial condition would be the same as for the default BMCA and the re-arrangement times would be the same as for the default BMCA. These numbers should be considered worst case values for the alternate BMCA as well as the default BMCA.

The clock port states of the initial topology shall be as shown in figure 3

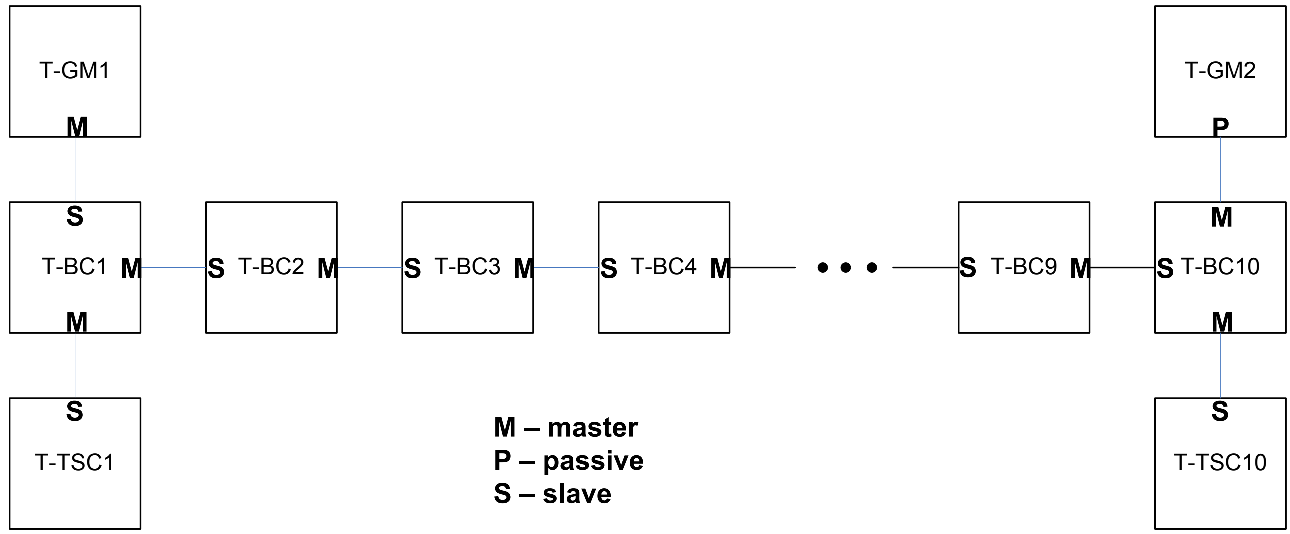


Figure 3 – Initial topology for the default BMCA with T-GM1 better than T-GM2

Failure Event

If T-GM1 enters a condition of degraded clock quality (compared to T-GM2) then it shall send announce messages to T-BC1. T-BC1 shall observe a change in grandmaster information received from its parent clock and shall take action relatively quickly (see below). If however, there is some form of communication failure between T-GM1 and T-BC1, then no action shall be taken until the ANNOUNCE_RECEIPT_TIMEOUT_EXPIRES event is triggered (§ **9.2.6.11**). Under normal conditions, this is expected to take longer than the degraded quality case and it is the one that is used in the analysis of the re-arrangement times.

NOTE: The IEEE 1588-2008 standard does not include procedures for a PTSF_lossSync event only a PTSF_lossAnnounce event. This could be added as a failure detection event but will not make a significant change to the re-arrangement times since this event would only come into play at the start of the re-arrangement.

Final Conditions after re-arrangement

The clock port states of the final topology shall be as shown in figure 4

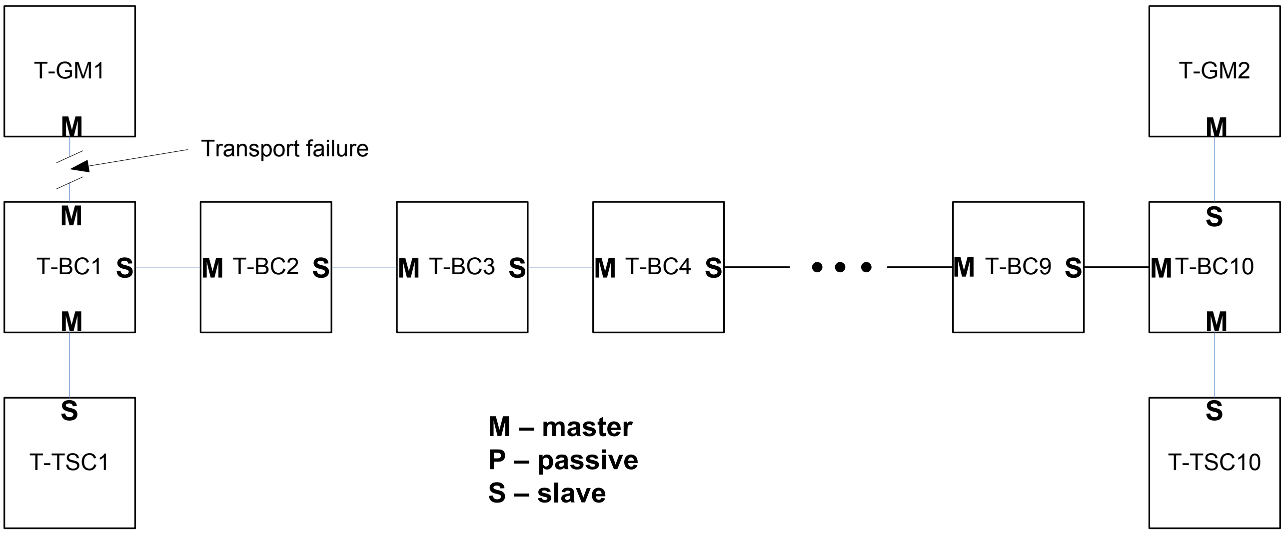


Figure 4 – Final topology for the default BMCA with loss of T-GM1

In the model it is assumed the T-TSCs continue to maintain their adjacent T-BC as their parentClock. This should be correct if they use clockclass = 255 compared to the worst case of 248 for a T-BC.

Steps involved in re-arrangement

Steps that will occur:

0) Transport failure occurs between T-GM1 and T-BC1

1) T-BC1 shall eventually trigger an ANNOUNCE_RECEIPT_TIMEOUT event on the port facing T-GM1

2) T-BC1 shall trigger a STATE_DECISION_EVENT and execute the BMCA

3) The best clock data shall be the defaultDS since there is no other Announce data

4) T-BC1 shall update currentDS and parentDS, change port facing T-GM1 to be in MASTER state.

5) At some time after the update of the currentDS and parentDS, an Announce shall be sent out the port of T-BC1 facing T-BC2.

6) T-BC2 receives an Announce message from its parent clock and immediately updates the "data sets for the ingress port". This means that the information on the new GM (T-BC1) shall be advertised at the next Announce transmission.

7) Similar to step (5) at some time after the update of the data sets, an Announce shall be sent out the port of T-BC2 facing T-BC3.

8) Steps 6 and 7 are repeated through to T-BC10 (or T-BC20).

9) T-GM2 receives an Announce message from T-BC10 and immediately updates the "data sets for the ingress port". When the next STATE_DECISION_EVENT is triggered, the BMCA is run. The outcome of the BMCA shall be to change the state of the port facing T-BC10 from PASSIVE to

MASTER. And on the next Announce interval, an Announce shall be sent toward T-BC10 containing the clockId and attributes of T-GM2 as the grandmaster.

NOTE: In order to study the worst case delays in propagation along the T-BC chain, the assumption is the chain of T-BCs maintains the original parent-child associations of the initial configuration until the information from T-GM2 is available.

10) T-BC10 receives an Announce message from T-GM2. This is not from its current parent clock, and it therefore treats T-GM2 as a FOREIGN MASTER. The Announce information must be qualified first through receipt of additional identical Announce messages (Foreign master qualification). After the information is qualified, then when the next STATE_DECISION_EVENT is triggered the BMCA is run. The outcome of the BMCA shall be to change the state of the port facing T-GM2 to UNCALIBRATED. The port facing T-BC9 shall enter the PRE_MASTER state. The port facing T-TSC10 remains in MASTER state throughout. On the next Announce interval, an Announce shall be sent toward T-TSC10 containing the clockId and attributes of T-GM2 as the grandmaster.

11) In T-BC10, at the port facing T-BC9, the port shall remain in PRE_MASTER state for a period of time and then transition to MASTER state. On the next Announce interval, after the port transitions to the MASTER state, an Announce shall be sent toward T-BC9 containing the clockId and attributes of T-GM2 as the grandmaster.

12) T-BC9 through T-BC1 shall repeat steps 10 and 11 . At each T-BC, the stepsRemoved shall be one greater that at the previous T-BC, and each T-BC shall remain in the PRE_MASTER state for a longer time interval.

13) Once T-BC1 has sent out its new GM information to T-TSC1, the re-arrangement is complete.

Model Results

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Announce Rate	0.5 pps	1 pps	2 pps	4 pps	1pps	1 pps
Pre_Master M3 processing	Enabled	Enabled	Enabled	Enabled	Disabled	Enabled
Foreign Master Processing	Enabled	Enabled	Enabled	Enabled	Enabled	Disabled
Re-arrangement Time	220	110	55	27.5	45	100

Table 1. Re-arrangement times for the 10 T-BC network

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Announce Rate	0.5 pps	1 pps	2 pps	4 pps	2pps	2 pps
Pre_Master M3 processing	Enabled	Enabled	Enabled	Enabled	Disabled	Enabled
Foreign Master Processing	Enabled	Enabled	Enabled	Enabled	Enabled	Disabled
Re-arrangement Time	630	315	157.5	78.75	42.5	157.5

Table 2. Re-arrangement times for the 20 T-BC network

Conclusions

The model shows that in networks of up to 10 T-BCs and without modifying existing 1588 procedures, in order to ensure network re-arrangements occur in less than 60 seconds, the Announce message rate must be at least 2 pps. If 1588 procedural changes are considered, then some modification of the PRE_MASTER decision point M3 processing could be considered.

This contribution is for information.

Appendix 1: Worksheet

Note: Worksheet uses Visual Basic Macros. Ensure Visual Basic is enabled with Excel before opening the file.

Note: When either of the random distribution options is selected, the F9 key can be used to force re-computations.

Rearrangements_v7.xls

Appendix 2: IEEE 1588-2008 Procedures Relevant to topological re-arrangements

Detection of transport communication failure

This is governed by the rules of **§ 9.2.6.11**. After the triggering of the ANNOUNCE_RECEIPT_TIMEOUT_EXPIRES event on a port in SLAVE state, and assuming there is no other port in SLAVE state, a clock shall immediately update its data sets based on it being the best clock (decision M1) and change the state of the port at which the event was triggered to MASTER state.

Transmission of grandmaster information

The distribution of clock information is done using the Announce messages. The information included in the Announce messages is extracted from the timePropertiesDS, parentDS, currentDS, and timePropertiesDS. These data sets are updated as part of the processing of the STATE_DECISION_EVENT at the conclusion of the execution of the BMCA or on specific events as indicated in the standard. But the transmission of the Announce messages is not required to be synchronized to the triggering of the STATE_DECISION_EVENT. So there can be a time lag of up to one transmission interval between the update of the data sets and the transmission of an Announce with the new information.

The transmission interval for Announce messages using multicast is required to meet a mean value with some permitted deviation on individual messages (**§ 9.5.8**). In practice, there would not appear to be any benefit in the deliberate jittering of these messages as they are PTP general messages and are not time sensitive. The requirement in the IEEE 1588-2008 standard would appear to be more of a test requirement than a suggestion for implementation (i.e., some variation in the Announce intervals is allowed because it is realized that the Announce interval cannot be exactly constant in a real implementation). The model used for this study has used a consistent interval in the transmission of Announce message. Randomization could be added.

Update of information received from the parent clock

On reception of an Announce message from the parent clock, the relevant data sets are immediately updated (**§ 9.5.3**). This means that new information is immediately made available for any new Announce messages transmitted by the node.

Update of information received from a clock other than the parent

When a clock receives Announce messages on a port and the sender of the Announce is not the parent clock, then the rules governing the foreign masters (**§ 9.5.3**) and Announce message qualification (**§ 9.3.2.5**) comes into play. A port shall maintain an 'implementation-specific foreignMasterDS data set' and shall track the identities of the ports which generated the announce messages and the number of Announce messages received from those ports.

The Announce message shall only be qualified for use into the data set comparison of the BMCA if it there have been at least FOREIGN_MASTER_THRESHOLD Announce messages received from the sending port within the FOREIGN_MASTER_TIME_WINDOW interval. It appears that a window must be started on reception of an Announce message from a 'previously silent master clock' but there are not more details on what this means.

The model used in this study considers an implementation where the FOREIGN_MASTER_THRESHOLD number of announce messages are met within the FOREIGN_MASTER_TIME_WINDOW interval. That is to say, there is no condition of crossing an implementation specific FOREIGN_MASTER_TIME_WINDOW interval boundary that might flush the contents of the foreignMasterDS and add additional delay.

PRE_MASTER state

When the BMCA generates a recommended state of BMC_MASTER and the port is not already in MASTER state, then the port shall move to PRE_MASTER state and shall start a timer for a duration of qualificationTimeoutInterval. When the decision point is M3, this qualificationTimeoutInterval has a duration (1 + currentDS.stepsRemoved) * announce interval. It is only after expiry of this timer that the port shall move to the MASTER state and begin sending Announce messages and Sync messages. Decision point M3 occurs on clocks that are not the Grandmaster clock, T-BCs that have identified a parentClock.

___