Casting of 1588 Best Master Clock Algorithm into 802.1D RSTP Formalism

Geoffrey M. Garner
Consultant

IEEE 802.1 AVB TG
2008.04.28

gmgarner@comcast.net

Introduction - 1

- Recent discussions in the AVB TG, and 802.1AS D2.0 TG ballot comments, have indicated the need for rapid reconfiguration after a change in topology or grandmaster clock (in an 802.1AS network)
  - The discussions have indicated the desire for this to be scalable, i.e., that there not be any timers whose values depend on network size (or number of hops over which synchronization is transported)
- Related to this, it has been pointed out that the spanning tree chosen by RSTP, for data transport, may not be the optimal spanning tree for transporting synchronization
  - In 802.1AS D2.0, the Best Master Clock Algorithm (BMCA) is invoked at end stations but not at bridges, and a spanning tree must be constructed outside of 802.1AS D2.0
- Reference [1] pointed out that elements of Rapid Spanning Tree Protocol (RSTP) can be used for BMC selection at the nodes, and can construct a spanning tree for synchronization
  - This would be a separate and independent spanning tree from the data spanning tree

Introduction - 2

- The full RSTP of 802.1D is not needed, because 802.1AS is only concerned about transporting synchronization using PTP and 802.11v messages, which are confined to a single hop (single LAN, using 802.1D terminology); there is no need to update a forwarding database, nor ensure that data plane frames do not loop

- In the discussion of [1], it was realized that when RSTP is simplified to do BMC selection, the result is very similar to the IEEE 1588 BMCA
  - Reference [1] describes a framework for mapping the aspects of RSTP that are needed for BMC selection to the IEEE 1588 BMCA
- The present presentation continues the work of [1] and attempts to consider how the detailed RSTP state machines might be simplified and modified to represent a simplification of the IEEE 1588 BMCA

- **This presentation is a work in progress; revision is expected after discussion**

Corresponding 1588 and RSTP Concepts - 1

IEEE 1588 Term	RSTP Term
Port State	Port Role (Note 1)
Master port	Designated port
Slave port	Root port
Passive port	Alternate port
(Note 2)	Backup port
Grandmaster	Root
clockIdentity (Note 4)	BridgeId (Note 4)
portNumber (Note 4)	PortId (Note 4)
stepsRemoved (Note 3)	RootPathCost
Announce interval	Hello Time
Announce Receipt Timeout	3*(Hello Time)
clockQuality	(Note 5)
priority1	BridgeID priority bits
priority2	(Note 5)
Announce message	Bridge PDU (BPDU)

Note 1: RSTP also defines port state, which is not needed here.

Note 2: Handled in 1588, but no term defined.

Note 3: There has been discussion in the AVB TG of generalizing this to more accurately represent phase error accumulation.

Note 4: See next two slides for more detail

Note 5: No corresponding concept in RSTP

Corresponding 1588 and RSTP Concepts - 2

- clockIdentity and BridgeID – both are 8 bytes
  - The 1588 clockIdentity is either an EUI-64, or an EUI-48 mapped to an EUI-64 using IEEE rules (802.1AS uses the latter)
  - The Bridge ID is a MAC address (EUI-48), with 2 bytes appended
    - These 2 bytes contain 4 bits that represent a settable priority, and 12 bits that represent a locally-assigned system ID extension
  - The settable priority is not needed in 802.1AS because the priority1 and priority2 fields are used for this purpose (described in subsequent slides)
  - The locally-assigned extension likely is not needed because its purpose is mainly to prevent consumption of MAC addresses in a VLAN; however, a time-aware system in 802.1AS will contain a single clock with a single clockIdentity (i.e., will not have separate clocks for different VLANs)
  - In the present presentation, the 1588 clockIdentity is used

Corresponding 1588 and RSTP Concepts - 3

- portNumber and PortID – both are 2 bytes
  - In 1588, the ports are numbered from 1 to N, where N is the number of ports
    - The portNumber is the number, in the range 1, 2, …, N, assigned to the port
  - The PortID in RSTP contains a 12 bit port number
    - A settable 4-bit priority is appended to this that allows the relative priorities of the ports to be managed
  - Use of either scheme in 802.1AS would not impact interoperability of the BMCA with other 1588 (non-802.1AS) networks, because all that is necessary in this algorithm is that the portNumber values for the different ports be unique
    - The only potential interoperability issues would arise with management messages; these would have to allow for the possibility of non-consecutive port numbers
- IEEE 1588 defines a portIdentity, which is the concatenation of clockIdentity and portNumber
  - **Not to be confused with RSTP PortID, which corresponds to 1588 portNumber**

Corresponding 1588 and RSTP Concepts - 4

- clockQuality in IEEE 1588 – consists of:
  - clockClass – characterizes the traceability of a clock, when it is grandmaster
  - clockAccuracy – characterizes the time accuracy of a clock when it is grandmaster
  - OffsetScaledLogVariance – characterizes the frequency stability of a clock, when it is grandmaster

Priority Vector - 1

- In RSTP, the priority vector consists of the following 5 items, with (a) most significant and (e) least significant:
  a) Root BridgeId
  b) RootPathCost
  c) BridgeId of the transmitting bridge
  d) PortId of the port on the transmitting bridge that transmits the message in question
  e) PortId of the port on which the message is received (where relevant)

Priority Vector - 2

- In 802.1AS, the following priority vector is consistent with the 1588 BMCA (with (a) most significant and (i) least significant, subject to discussion on a subsequent slide of the case of comparing the same two clocks)

a) grandmaster priority1

b) grandmaster clockClass

c) grandmaster clockAccuracy

d) grandmaster offsetScaledLogVariance

e) grandmaster priority2

f) grandmasterIdentity (clockIdentity of grandmaster)

g) stepsRemoved, or generalization to account for more accurate phase error accumulation (generalization might require a new Announce message field, carried in a TLV, because stepsRemoved is still needed for aging purposes)

h) portIdentity of the transmitting bridge (sourcePortIdentity in Announce message; contains clockIdentity and portNumber)

i) portNumber of the port on which the message is received (where relevant)

Comparison of two priority vectors - 1

- When 2 priority vectors are compared in RSTP, they are compared as though they were each single unsigned integers
  - The vector with the smaller numerical value represents the better bridge
- When 2 priority vectors are compared in the IEEE 1588 BMCA, a modification is needed to handle the case where the two vectors represent the same grandmaster clock (i.e., grandmasterIdentity is the same in both vectors)
  - If the two vectors represent different grandmaster clocks, i.e., different values of grandmasterIdentity, they are compared as though they were each single unsigned integers, and the vector with the smaller numerical value represents the better clock (just like RSTP)
  - If the two vectors represent the same grandmaster clock, i.e., grandmasterIdentity is the same in both, then components (a) – (e) are ignored (priority1, clockClass, clockAccuracy, scaledLogVariance, priority2), and the comparison is done on the subvectors consisting of components (g) – (i) (stepsRemoved, portIdentity of transmitting bridge, portNumber of receiving bridge (where relevant))

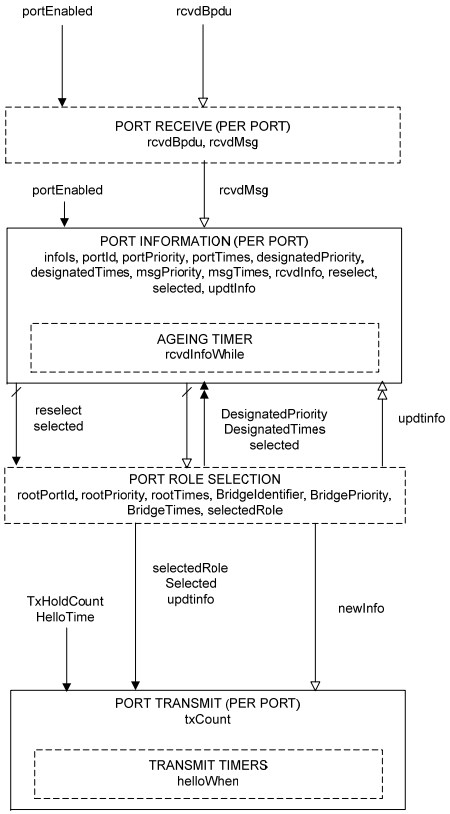
Comparison of two priority vectors - 2

- The reason the 1588 BMCA ignores priorty1, clockClass, clockAccuracy, scaledLogVariance, and priority2 for the case where the grandmasterIdentities are the same is that this represents a case where a clock is upgraded or downgraded
  - In this case, the values of one or more of these ignored parameters in one of the two vectors are no longer relevant
  - If these parameters were used in the comparison, the algorithm might favor a worse path based on parameters that had changed
    - Note that if the 1588 BMCA did use the RSTP comparison, eventually an Announce message with the updated parameters would be received on the better path, and the spanning tree would be changed to use that path
    - The 1588 approach would avoid having a spanning tree with the worse path in effect for a short time

RSTP State Machines Needed for 1588 BMCA

- The following slides contain simplifications of the subset of RSTP state machines needed to represent the 1588 BMCA
- The state machines include
  - Overview and interrelationships (802.1D/Figure 17-12)
  - Port Timers state machine (802.1D/Figure 17-13)
  - Port Receive state machine (802.1D/Figure 17-14)
  - Port Transmit state machine (802.1D/Figure 17-17)
  - Port Information state machine (802.1D/Figure 17-18)
  - Port Role Selection state machine (802.1D/Figure 17-19)
- Other RSTP state machines, having to do with detecting the edge of the network (this will be done using the Pdelay mechanism in 802.1AS), updating the forwarding data base in a way that does not cause loops, notifying of topology changes, and compatibility between RSTP and STP, are not needed for the BMCA
- For now, RSTP nomenclature (e.g., variable and function names) is retained; this will be replaced by more meaningful nomenclature for BMCA

Overview and Interrelationships of state machines

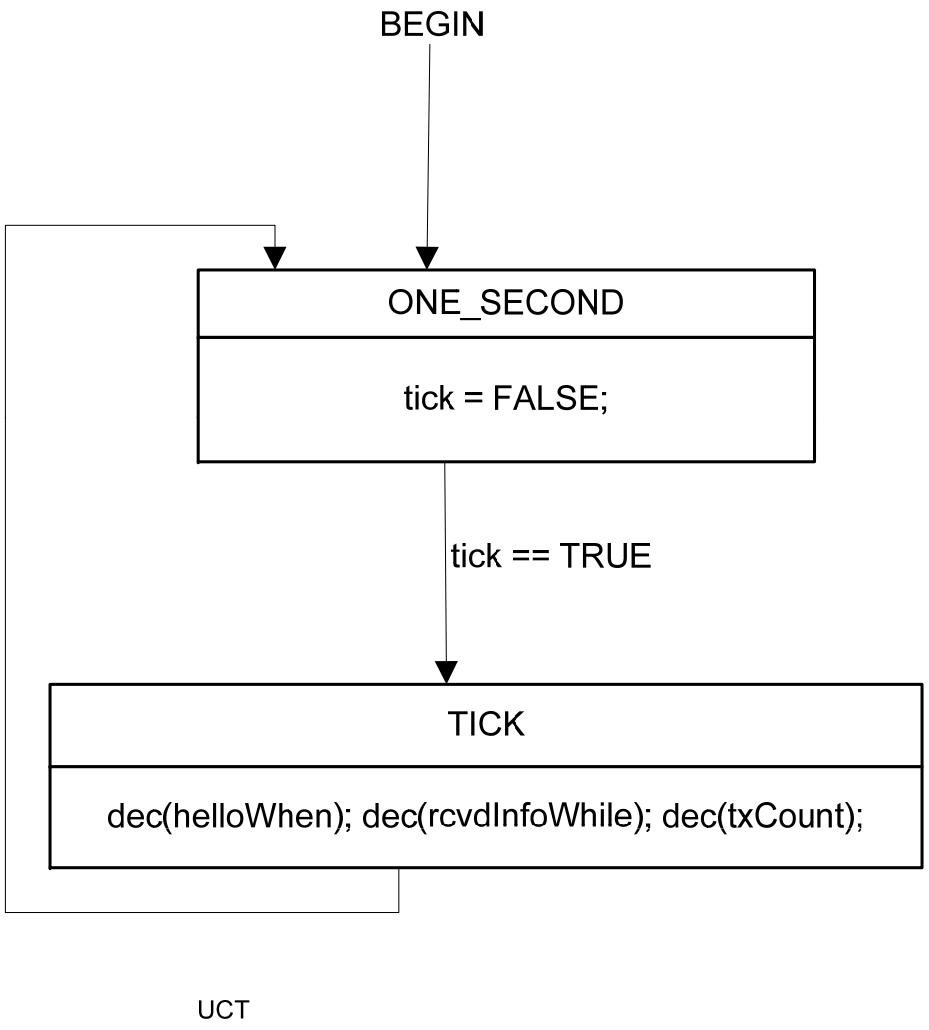


Note: It may be more convenient to combine the Port Receive and Port Information state machines

Notation:
Variables are shown both within the machine where they are principally used and between machines where they are used to communicate information. In the latter case a variety of arrow styles, running from one machine to another, show how each is typically used:

- Not changed by the target machine. Where the mcahines are both per Port, this variable communicates between instances for the same port
- Set (or cleared) by the originating machine, cleared (or set) by the target machine. Where the machines are both per Port, this communicates between instances for the same port.
- As above, except that the originating per port machine instance communicates with multiple port machine instances (by setting or clearing variables owned by those ports).
- As above, except that multiple per Port instances communicate with (an)other instance(s) (by setting or clearing variables owned by the originating ports).

Port Timers State Machine - 1



The following abbreviation is used in this state diagram:

Dec(x)
{ if (x != 0) x = x-1;
}

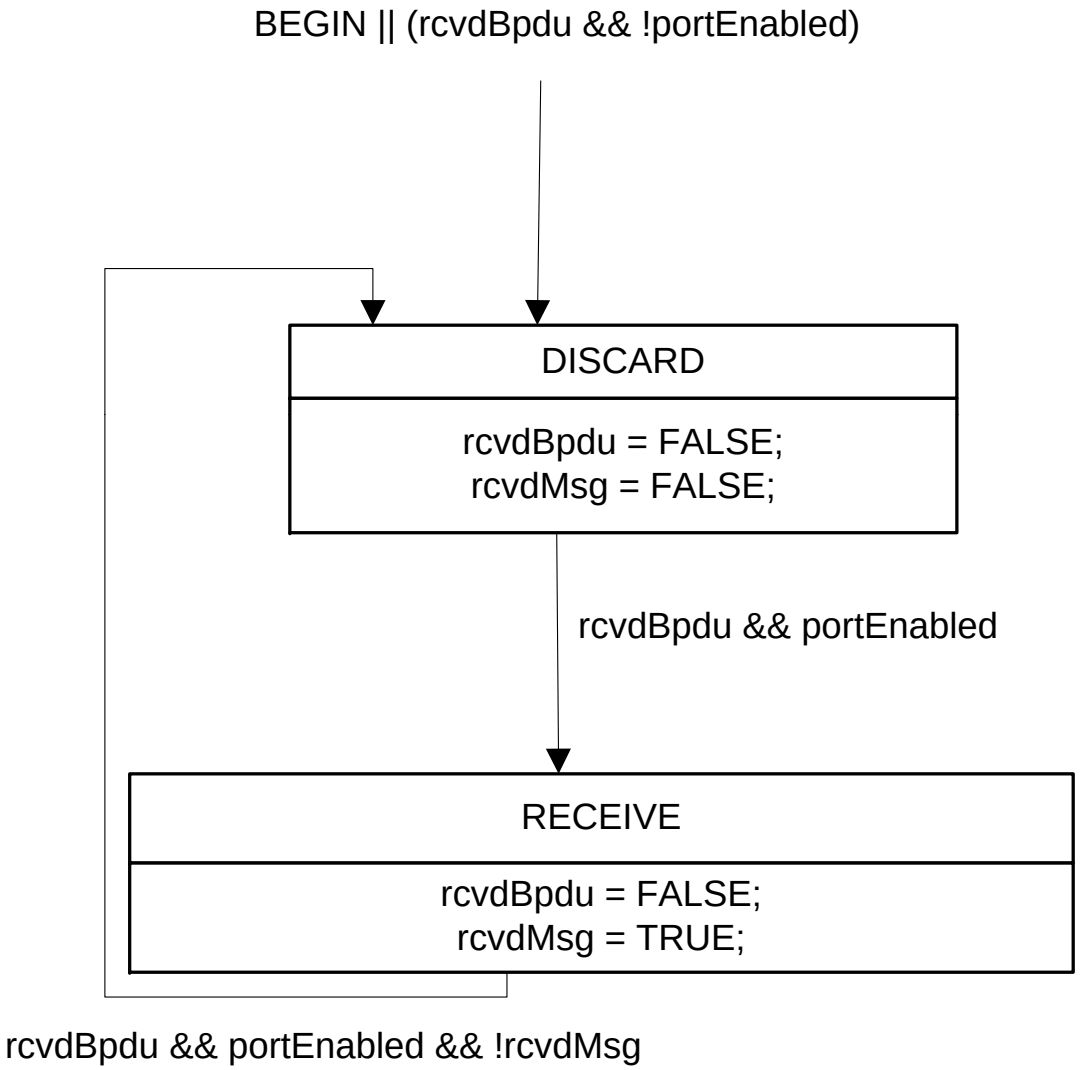
If there is more than one instance of x, e.g., per port, all instances are separately decremented.

Port Timers State Machine - 2

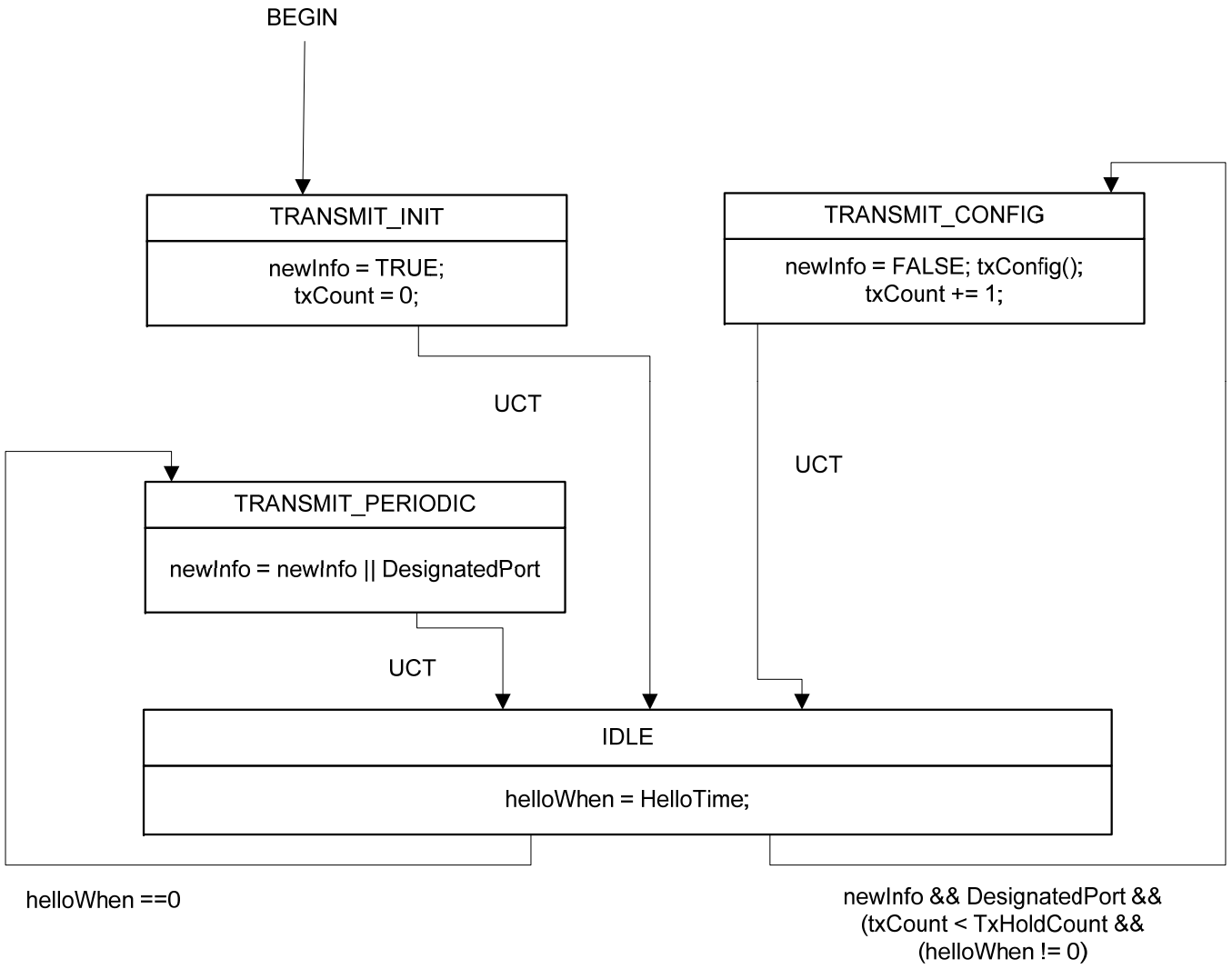
- This state machine may be replaced by timer model in 802.1AS that makes use of currentTime variable to represent current value of time
  - At any given time, timer expiration time is

- Port Timers state machine may be replaced by timer model in 802.1AS that makes use of currentTime variable to represent current value of tiem
- When it is desired to set a timer, set the expirationTime equal to currentTime plus timerValue
- In each relevant state where it must be checked whether timer has expired, check whether currentTime >= expirationTime

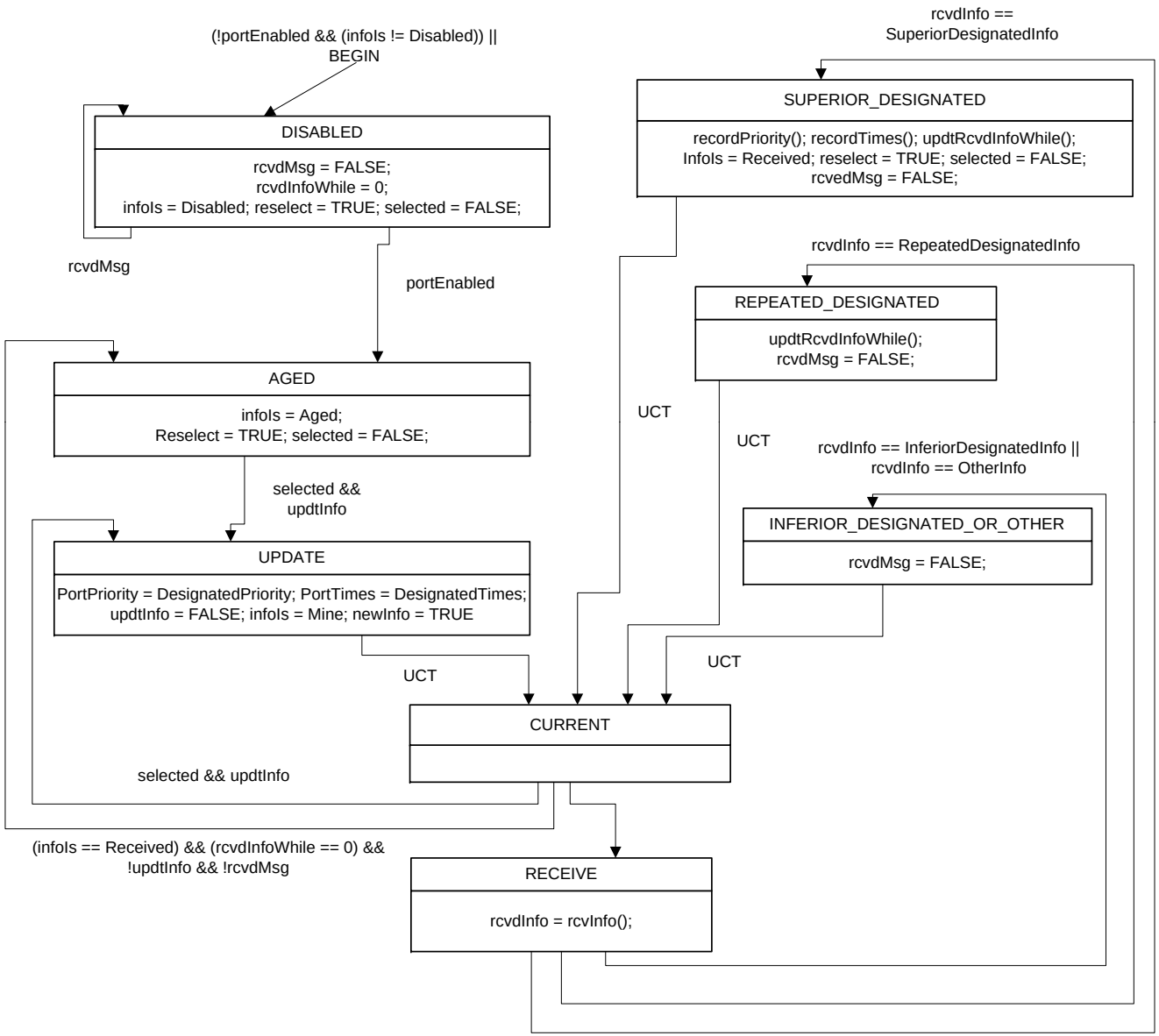
Port Receive State Machine



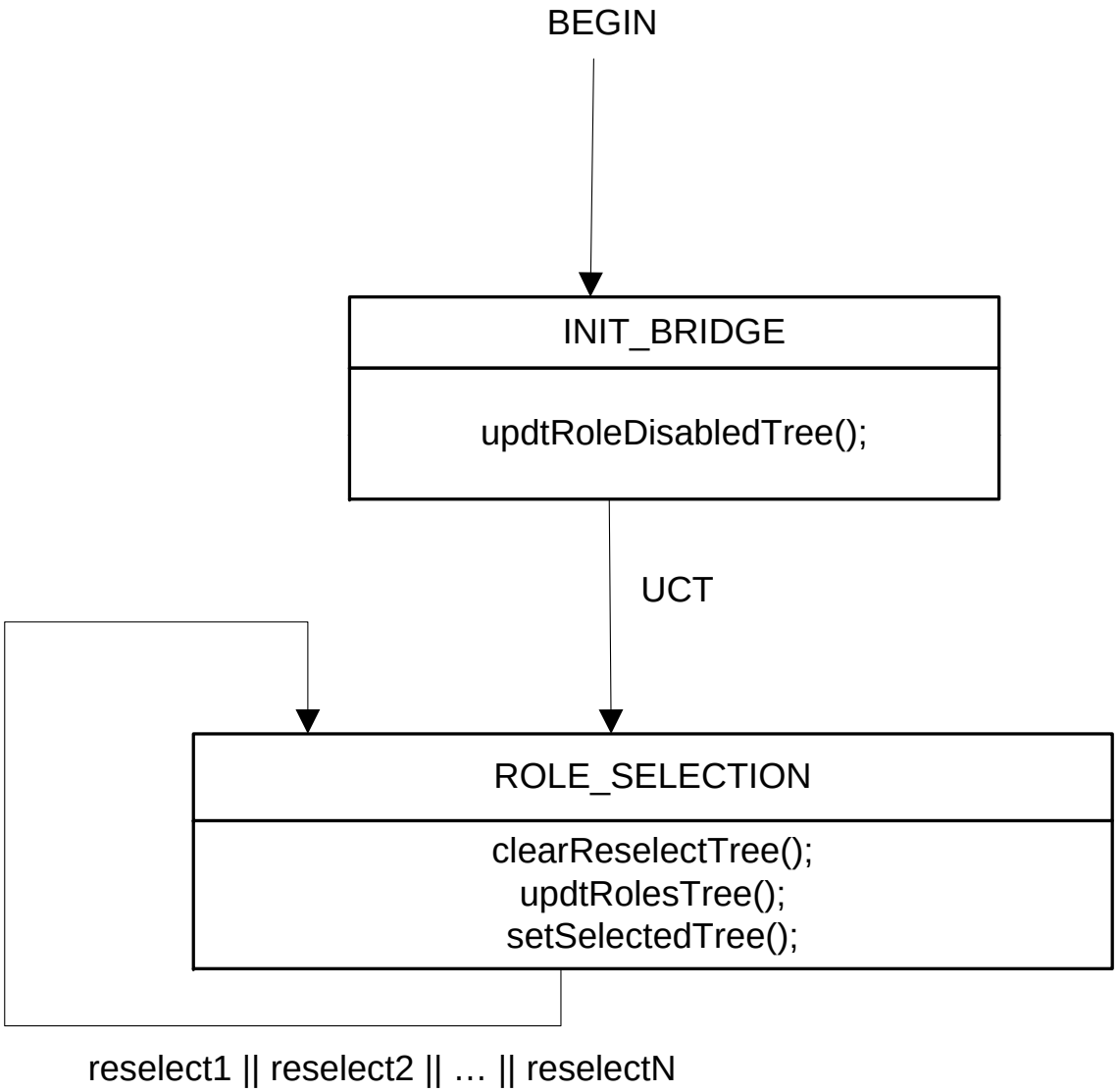
Port Transmit State Machine



Port Information State Machine



Port Role Selection State Machine



rcvInfo() function

- Equivalent to 1588 computation of Erbest using data set comparison algorithm
- Decodes the message priority vector and timer values from the received BPDU, storing them in the msgPriority and msgTimes variables
- Returns SuperiorDesignatedInfo if the received message conveys a Designated Port Role, and
  - The message priority vector is superior to the Port's priority vector, or
  - The message priority vector is the same as the port's priority vector and any of the received timer values differ from those already held for he port
- Returns RepeatedDesignatedInfo if the received message conveys a Designated Port Role and message priority vector and timer values that are the same as the port's priority vector and timer values
- Returns InferiorDesignatedInfo if the received message conveys a Designated Port Role and message priority vector that is worse than the port's priority vector and timer values
- Otherwise, returns OtherInfo

updtRolesTree() function - 1

- Equivalent to 1588 computation of Ebest using dataset comparison algorithm, followed by 1588 state decision algorithm, followed by updating of data sets and setting of states
- Computes
  - Root path priority vector for each port
  - Root priority vector for bridge (1588 Ebest)
  - rootTimes for bridge
  - First 4 components of designated priority vector for each port
  - Designated times for each port
- The above is highly abbreviated; see 802.1D-2004, subclauses 17.6 and 17.21.25 for details

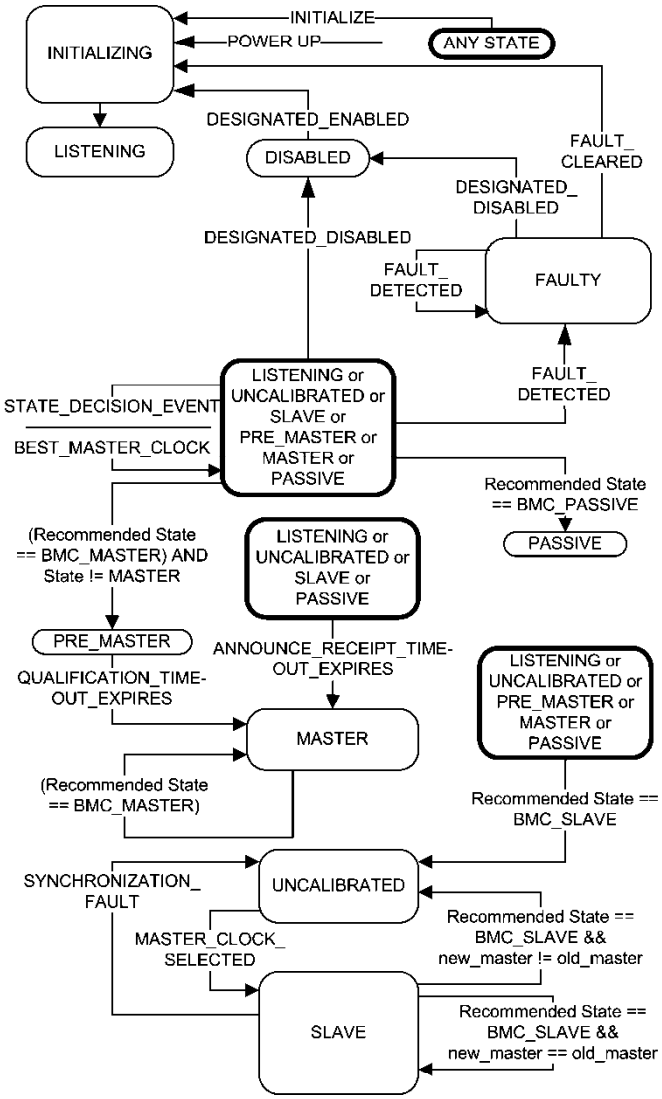
updtRolesTree() function - 2

- Assigns port Role for each port, and updates port priority vectors and timer information (1588 state decision algorithm, updating of states and data sets)
  - If port is disabled, selectedRole set to DisabledPort; otherwise
  - If port priority vector was aged, updtInfo set and selected Role set to DesignatedPort (1588 Announce Receipt Timeout)
  - If port priority vector was derived from another port on the bridge or from the bridge itself as the root bridge, selectedRole set to DesignatedPort (1588 master port)
  - If port priority vector was received in a configuration message (BPDU) and not aged, and root priority vector is now derived from it, selectedRole set to RootPort (1588 slave port)
  - If port priority vector was received in a configuration message (BPDU) and not aged, and root priority vector is not now derived from it, the designated priority vector is not higher than the port priority vector, and designated bridge and port components of port prioroity vector *do not* reflect another port on the bridge, selectedRole set to AlternatePort (1588 passive port)

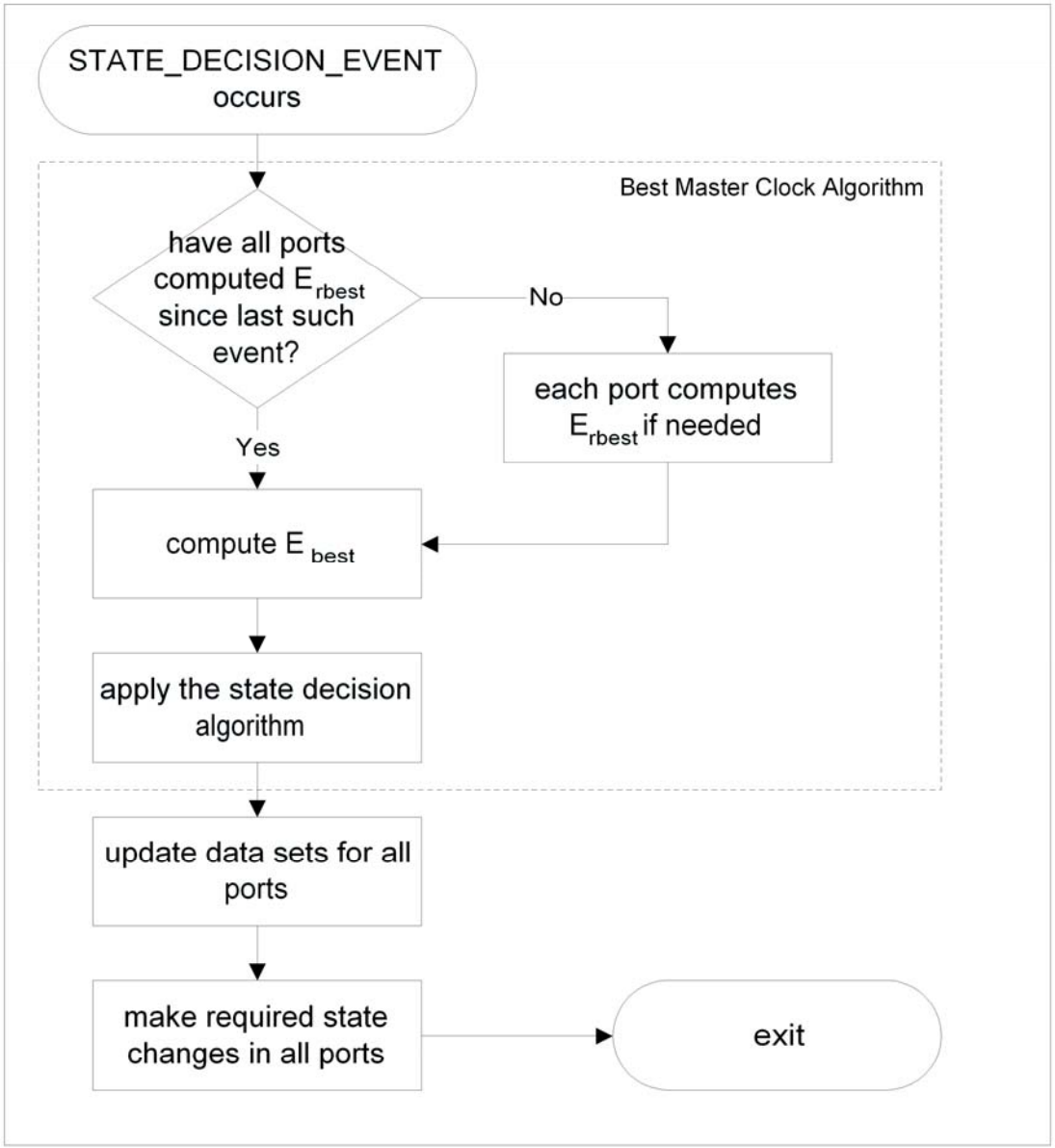
updtRolesTree() function - 3

- Port Role assignment and port priority vector and timer information update (continued)
  - If port priority vector was received in a configuration message (BPDU) and not aged, and root priority vector is not now derived from it, the designated priority vector is not higher than the port priority vector, and designated bridge and port components of port priority vector *do* reflect another port on the bridge, selectedRole set to BackupPort (1588 passive port)
  - If the port priority vector was received in a configuration message (BPDU) and not aged, and root priority vector is not now derived from it, the designated priority vector is higher than the port priority vector, selectedRole set to DesignatedPort (1588 master port)
- The above is highly abbreviated; see 802.1D-2004, subclauses 17.6 and 17.21.25 for details

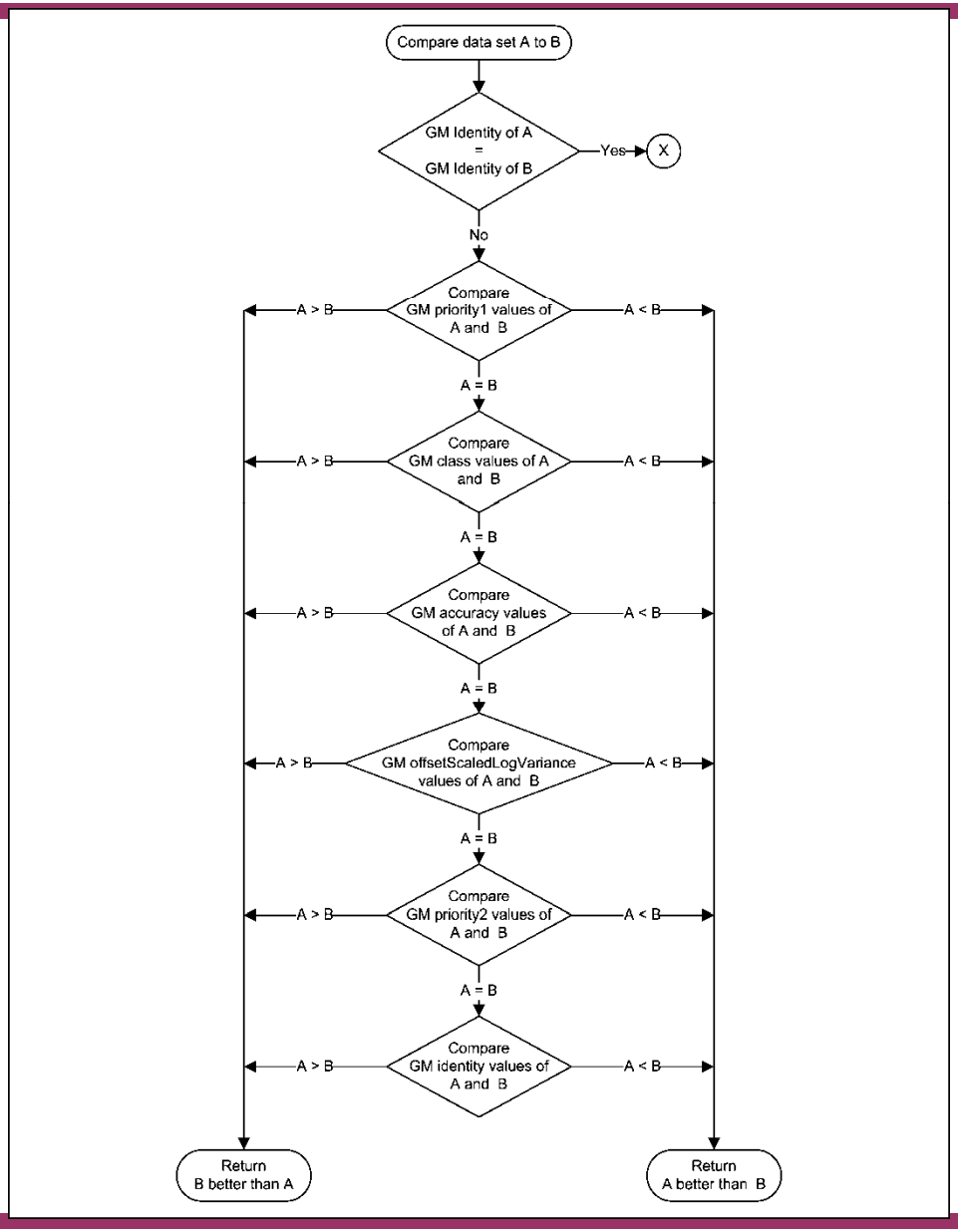
1588 Port State Machine



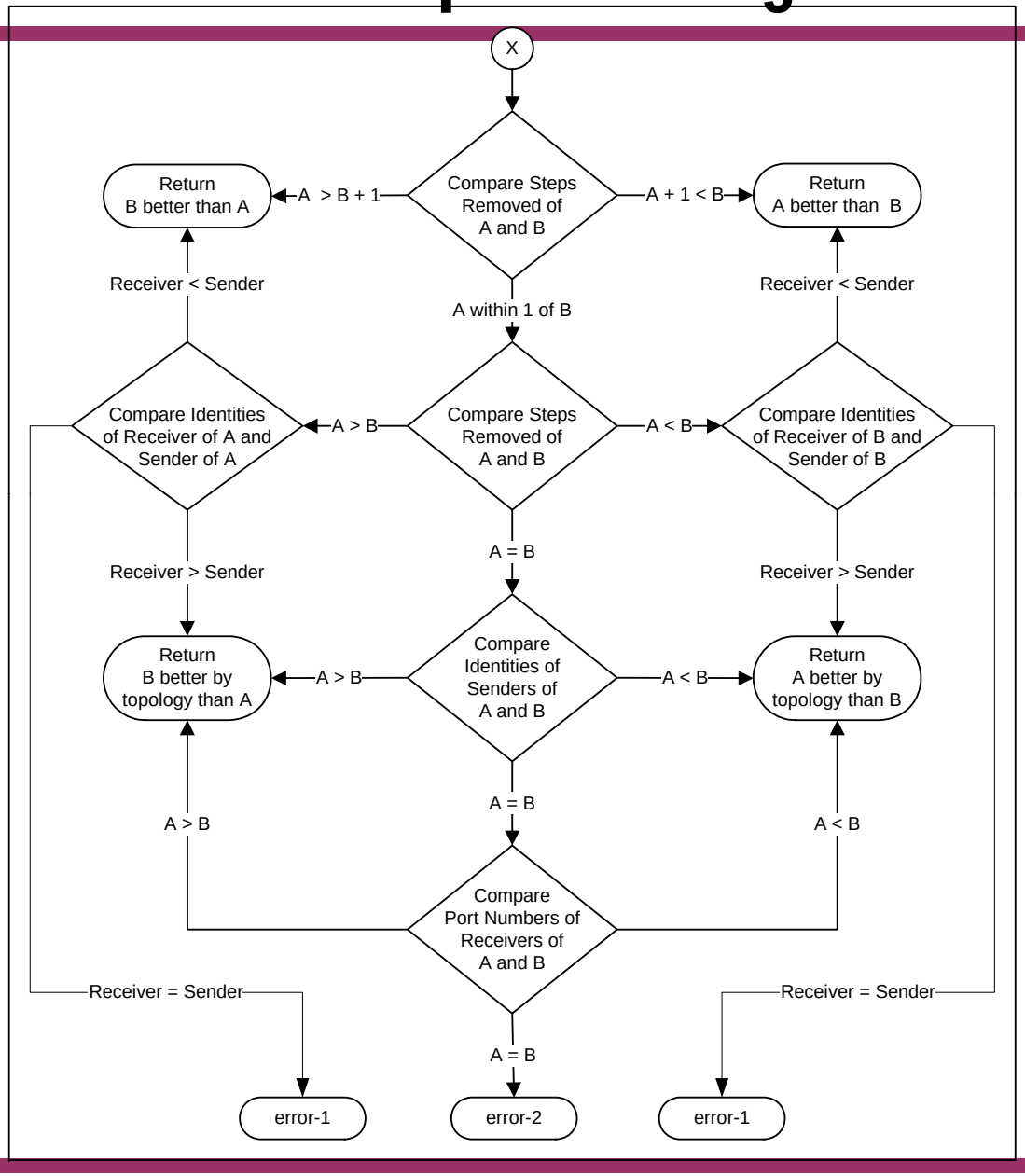
1588 State Decision Event Logic



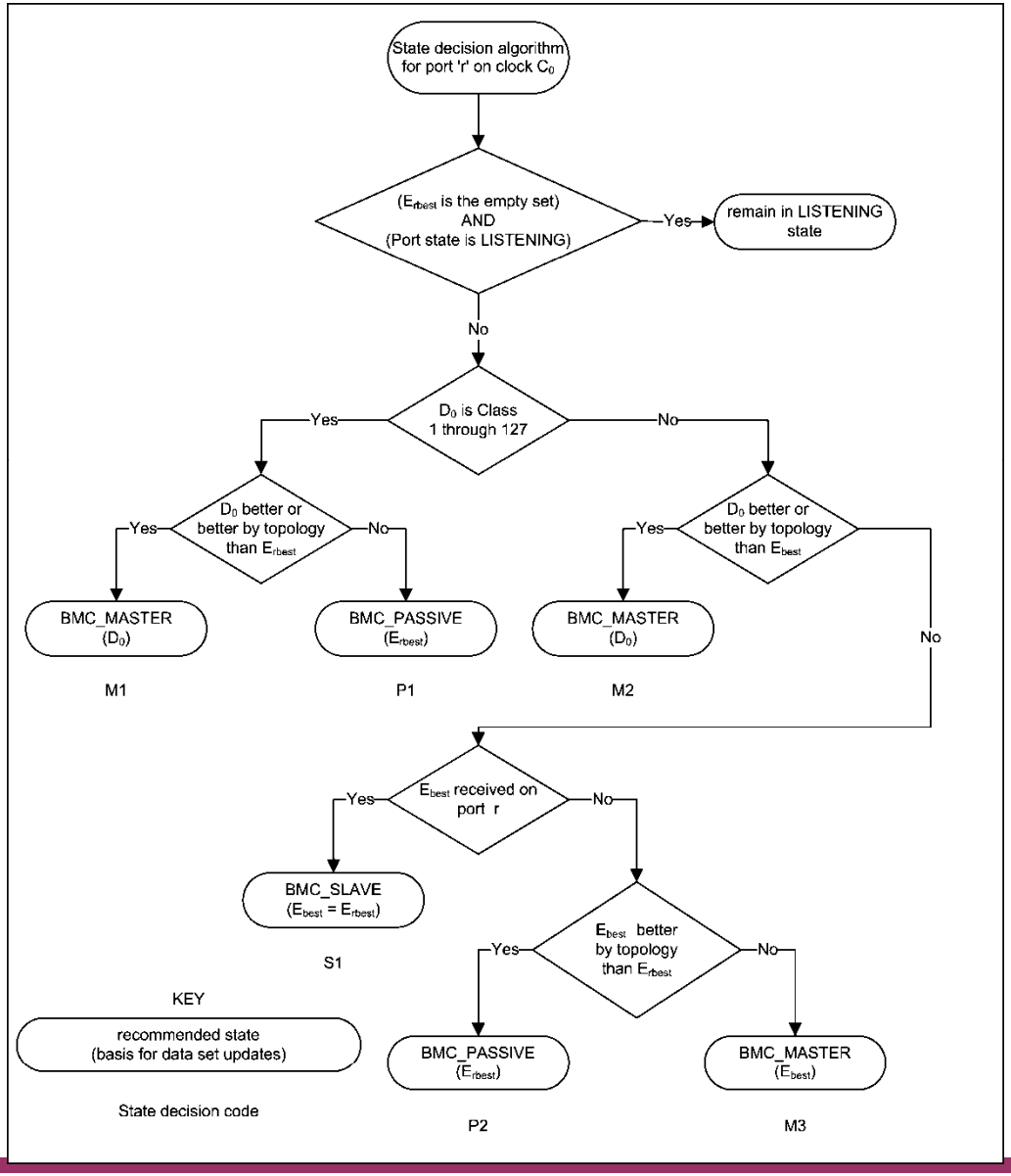
1588 Data Set Comparison Algorithm – 1



1588 Data Set Comparison Algorithm – 2



1588 State Decision Algorithm



References

1. Norman Finn, *802.1AS Fast Master Clock Selection, Moving 802.1AS closer to RSTP*, Version 2, April, 2008.