Accredited Standards Committee
X3, Information Processing Systems

Doc: X3T10.1/95a210R3
Date: 10 May, 1996
Project: 1147D
Ref Doc.: X3T10.1/1147D
Reply to: Mark DeWilde

To: X3T10.1 Membership
From: Mark DeWilde

Subject: Interlocked Election of Master Process, Rev 3, SIMPLIFIED VERSION

BACKGROUND

At the October SSA Week meetings, Revision 0 of this proposal entitled "loss of initiator table space" was discussed. With further discussion it became apparent that the loss of table space and several other problems were being caused by the common problem of multiple Masters existing on a web for a transient period. In order to eliminate this transient period, I proposed an interlocked protocol verbally. I took the action item to produce a written proposal and forward to the committee for comment and consideration. That proposal (which also cures the problem described in r0) follows. It became clear while working on rev 1 and rev 2 that the dynamic nature of the web and the requirement that the master perform Third Party Quiesces on behalf of missing Configutors was causing considerable complexity in the process. In order to eliminate the need to keep track of all web changes prior to stabilization, a method was devised to perform configutor table maintenance and proposed in a separate document, which permitted simplification of this proposal.

PROPOSAL

Add master priority to the QUERY NODE SMS so that a current master receives immediate notification that it must resign. Add a "Current Master" bit to the QUERY NODE REPLY so that the pending master "knows" to delay the assumption of mastership. Use the QUERY NODE with the "MA" bit set to indicate to the pending master that the resigning master has resigned. This set of changes provides an interlock mechanism that prevents multiple masters for any interval. Add a timer to detect loss of resigning master in case the resigning master fails to activate the pending master with it's QUERY NODE with MA set.

The following sections are changed as indicated:

12.8 Master Negotiation Process

During the configuration process, the initiator with the highest MASTER PRIORITY is elected to be the Master (see 11.2.6). If more than one initiator is set at the same highest MASTER PRIORITY, then the initiator with the highest Unique ID among them is elected to be the Master.

Each initiator compares its own Master priority with the Master priority it receives from each other initiator in the QUERY NODE REPLY SMS. Each initiator in the group with the highest MASTER PRIORITY then compares its own Unique ID with those of the other initiators in the group. The initiator within the group with the highest Unique ID becomes the Master. If two Webs that both contain a Master are joined by a new link, then each Master invokes the Configuration process for the new link. Since they may invoke the Configuration Process at different times, an interlock mechanism is necessary to allow them to complete outstanding Async Alert actions on their portions of the web prior to one of them assuming mastership of the entire web. This is accomplished by the use of the "CM" bit in the QUERY NODE REPLY SMS, and the "My Master Priority" field in the QUERY NODE SMS. The following sections describe the usage of the fields and messages.

Definitions of terms used in the description:

Sub-Web:

A string of nodes with target functions, initiator functions, or a combination of both that are added to another web as a unit.

Resignation Request:

Received by a Master or Resigning master of a sub-web, this is a QUERY NODE SMS that contains either a higher master priority than the receiving node's, or the same Master priority and a higher Unique ID. A resignation request is answered with a QUERY NODE REPLY with the CM bit set.

Pending Resignation:

A Resignation is Pending whenever a QUERY NODE REPLY SMS is sent with the CM bit set. The Pending Resignation is satisfied when the Formal Resignation is performed.

Resigning Master:

The (temporary) Master node of a Sub-Web that has received one or more Resignation Requests. It no longer accepts ASYNC ALERT SMS's, but has not yet finished processing those received prior to the QUERY NODE SMS.

Formal Resignation:

A Resigning Master Formally Resigns to a node with which it has completed a Resignation Request transaction at an earlier time when it sends a QUERY NODE SMS with the MA bit set to that node.

Pending Master:

The current Master of a sub-web that, after walking a newly attached sub web containing one or more active Masters, has discovered that it has won mastership of that new sub web. It waits to assume mastership of the new sub-web until all of the resigning masters formally resign.

Assuming Master:

A Pending Master that has received all of the Formal Resignations from the Resigning Masters is the Assuming Master until it has completed the web configuration and is then the web Master. The Assuming Master may receive a Resignation Request and not become the web Master.

Rules for Interlocked Master Election

1. If two sub webs containing Masters are joined, then the master of each must walk the other sub web to determine Mastership of the new joined web. The configuration of the new web section by the winning master does not proceed until the old Master of the sub web has completed it's handling of outstanding ASYNC ALERTS and formally resigned to the new Master.

2. When the Master of a sub-web , or a pending master receives a Resignation Request, it becomes a Resigning Master. It immediately stops accepting new ASYNC ALERT SMSs, Sends a QUERY NODE REPLY SMS with the CM bit set, logs a Pending Resignation, and continues finishing the actions required by the remaining pending ASYNC ALERTs. Any QUERY NODE SMSs that would not require resignation are always answered with QUERY NODE REPLY SMSs with the CM bit clear.

3. When a resigning master has completed all actions required by pending ASYNC ALERT SMSs, it performs Formal Resignations for all logged Pending Resignations. After resignation, the node will respond to any subsequent QUERY NODE SMSs by issuing QUERY NODE REPLY SMSs with the CM bit clear.

4. A Pending Master will wait until all Resigning Masters have Formally Resigned prior to configuring the ports on the new Sub-Web. When it has received all Formal Resignations from all Resigning Masters, it becomes the Assuming Master. The Assuming Master then Configures the ports in the web.

5. If the Assuming Master receives an ASYNC ALERT SMS indicating “PORT NOW OPERATIONAL”, from a port during configuration, it must check it’s web topology map to see if there has been a new sub-web added to the web. If not, then the ASYNC ALERT was an old queued one and the Assuming Master replies to it, discards it, and continues it’s configuration of the ports in the web. If there has been a new sub-web added, then the Assuming Master walks the new sub-web.

6. Whenever a sub-web containing no configutors is added to an existing web, the master of the existing web shall reset the nodes in the sub-web. The sub-web reset is required since the sub-web had no master when attached, and there could be outstanding commands and stale initiator table entries in the nodes.

7. If the Assuming Master receives a Resignation Request from a node in a newly attached sub-web, it sends a QUERY NODE REPLY with the CM bit set, and becomes a Resigning Master. Since the ASYNC ALERT SMS Processing caused by the addition of the sub web includes walking the new web, this will be completed prior to resignation. If configuration of ports was in progress, this operation ceases since the new master will re-configure all of the web ports.

8. Each Pending Master shall start a timer when it receives a reply confirming the receipt of a Resignation Request. If the Resigning Master has not issued a Formal Resignation in 5 seconds, then the Pending Master shall send a QUERY NODE SMS to the Resigning Master. If the Resigning Master returns a QUERY NODE REPLY with the CM bit set, the Pending Master restarts the Timer. If the Resigning Master returns a QUERY NODE REPLY SMS with the CM bit clear, then the Pending Master accepts this reply as the Formal Resignation, but still replies to any subsequent Formal Resignations from that node. If no reply is received within another 5 seconds, the Pending Master issues a TOTAL RESET to the node, followed by re-configuration of the node.

[The remaining portions of the current text regarding change of master priority remain unchanged.]

11.2.5 QUERY NODE SMS

This SMS is sent from a Configutor node to every other Operational node during the Configuration process. QUERY NODE is also used as a remote wrap test to verify the integrity of the Path.

The QUERY NODE SMS is defined in Table 31.

Table 31 - QUERY NODE SMS

Byte	Bit 7	6	5	4	3	2	1	Bit 0
0	SMS CODE (00h)							
1	SSA-TL2 VERSION (10h)							
2	TAG							
3	TAG							
4	RETURN PATH							
5	RETURN PATH							
6	RETURN PATH							
7	RETURN PATH							
8	UNIQUE ID							
9	UNIQUE ID							
10	UNIQUE ID							
11	UNIQUE ID							
12	UNIQUE ID							
13	UNIQUE ID							
14	UNIQUE ID							
15	UNIQUE ID							
16	DR	MA	My Master Priority			reserved		

The SSA-TL2 VERSION field defined in Table 32 identifies the version of SSA-TL2 being used by the sender.

Table 32 - SSA-TL2 VERSION field values

Version	Description
00h-0Fh	reserved for SSA-TL1
10h	This standard
11h-FFh	reserved

The TAG field is returned in the QUERY NODE REPLY SMS. The TAG is assigned by the Configutor node and it shall be unique among the TAG values that are currently active from that Configutor node.

The RETURN PATH field specifies the path component that shall be placed in the Configutor table entry created in response to this QUERY NODE SMS, if an entry is created. This value is used for the address field of the resulting QUERY NODE REPLY SMS, and is used for the address field of any future application SMS that utilizes this Configutor table entry.

The UNIQUE ID field contains the Unique ID of the Configutor node that issued QUERY NODE SMS.

The DR bit (Disable Registration) controls the updating of the Configutor table in the node. If the DR bit is cleared then the node shall enter the specified RETURN PATH and UNIQUE ID into its Configutor table. If a Configutor node intends to use several alternative paths to the same node then it shall issue QUERY NODE SMS with the DR bit is cleared once over each path. If the DR bit is set the Configutor table shall not be updated.

The MA bit (Master Alive) is set when the Master invokes the Master Alive process (see 12.2), and is cleared all other times. When a node receives a QUERY NODE SMS with the MA bit set, it restarts the 10 second Master Alive timer (see 12.2).

The “My Master Priority” field is valid only for Master Capable Nodes, and contains the Master Priority of the node that issued the QUERY NODE SMS. It is used by the node being queried to decide if it needs to resign mastership.

11.2.6 QUERY NODE REPLY SMS

The QUERY NODE REPLY SMS as defined in Table 33 is returned when a QUERY NODE SMS is received. The QUERY NODE REPLY SMS is returned on the same port that received the corresponding QUERY NODE SMS.

Table 33 - QUERY NODE REPLY SMS

Byte	Bit 7	6	5	4	3	2	1	Bit 0
0	SMS CODE (01h)							
1	PORT							
2	TAG							
3	TAG							
4	UPPER LEVEL PROTOCOL							
5	ITF	MASTER PRIORITY			reserved			
6	TOTAL OTHER PORTS							
7	SSA-TL2 VERSION (10h)							
8	UNIQUE ID							
9	UNIQUE ID							
10	UNIQUE ID							
11	UNIQUE ID							
12	UNIQUE ID							
13	UNIQUE ID							
14	UNIQUE ID							
15	UNIQUE ID							
16	RETURN PATH ID							
17	RETURN PATH ID							
18	RETURN PATH ID							
19	RETURN PATH ID							
20	P1O	P2O	CM	reserved				

The PORT field indicates the number of the port currently being used.

The TAG field is copied from the QUERY NODE SMS. It identifies the QUERY NODE SMS that this reply is being generated for.

The UPPER LEVEL PROTOCOL field identifies the upper-level protocol that the node shall respond to. The UPPER LEVEL PROTOCOL field shall contain a value from Table 34.

Table 34 - Upper-level protocol code values

ULP code	Protocol	Notes
00h	USE THE QUERY PROTOCOL SMS TO LIST UPPER LEVEL PROTOCOLS SUPPORTED	1
01h	SHALL RESPOND TO NO UPPER-LEVEL PROTOCOL	1
02h	Vendor specific	1,2
03h-7Fh	reserved	
80h	SSA-IA / 95SP	1,2
81h	reserved	
82h	SSA-S2P	1,2
83h	SSA-S3P	1,2
84h-FAh	reserved	
FBh-FFh	Vendor specific	2
Notes 1) Valid in QUERY NODE REPLY. 2) Valid in QUERY PROTOCOL REPLY		

The Configutor table full (ITF) bit is set when there is no space left in the Configutor table to make an entry for the Configutor node that sent QUERY NODE SMS.

The MASTER PRIORITY field defines the priority of the node for becoming the Web Master. A value of zero indicates that the node is not capable of functioning as a Master. A value of one is used by a node that is not Master Capable, but wishes to participate in the Healthy Web process. Any value greater than 1 indicates the node's priority for becoming the Master. A value of two is the lowest priority and seven the highest

> Implementer's note 11: A Configutor node optionally fixes its MASTER PRIORITY at 4 (the default priority) or provides some mechanism outside of SSA to change its MASTER PRIORITY dynamically.

The TOTAL OTHER PORTS field contains a value that is one less than the number of ports implemented. If this value exceeds two, the QUERY SWITCH SMS is used to retrieve a port mask for the switch.

> Implementer's note 12: The TOTAL OTHER PORTS field value shall be 0 (single port), 1 (dual port) or an odd number greater than 1 (switch) as per clause 8.2.

The SSA-TL2 field defined inTable 32 identifies the version of SSA-TL2 being used by the sender.

The UNIQUE ID field contains the node's Unique ID.

The RETURN PATH ID contains a value created by the node and returned to the Configutor node. This field shall be used by the Configutor node in any future application SMSs that are utilizing the same path as used by the QUERY NODE SMS that caused this QUERY NODE REPLY SMS is be generated.

The P1O bit is only valid for single port and dual port nodes, Switch nodes shall clear the P1O bit. For single port or dual port nodes, the P1O bit is set if port 1 is operational, and is cleared if port 1 is not operational.

The P2O bit is only valid for single port and dual port nodes, Switch nodes shall clear the P2O bit. For single port or dual port nodes, the P2O bit is set if port 2 is operational, and is cleared if port 2 is not operational.

The Path component of the ADDRESS field in a QUERY NODE REPLY SMS frame is copied from the RETURN PATH field in the corresponding QUERY NODE SMS. All padding bytes shall be discarded.

The CM bit is used by a Master-Capable node to indicate to the querying master that it is currently master of a Sub-web. If the master priority of the replying node is lower than that of the querying node, then the replying node becomes a resigning master if this bit is set.

Sincerely,
Mark A. DeWilde
Principal System Architect
Pathlight Technologies
Voice: (607)266-4000 X-403
FAX: (607)266-0352
Email: mark@pathlight.com