

---

MODULE *Controller*

---

EXTENDS *Naturals, FiniteSets, Sequences, Messages*

The set of all *ONOS* nodes

CONSTANTS *Nodes*

---

The following variables are used by the *mastership* election service. These variables represent global atomic state.

The current *mastership* term

VARIABLE *term*

The current master node *ID*

VARIABLE *master*

A sequence of standby nodes

VARIABLE *backups*

---

The following variables are per-node variables used by controller nodes in the *mastership* arbitration protocol.

A queue of events from the *mastership* service to the node

VARIABLE *events*

The current term, master, and backups known to the node

VARIABLE *mastership*

The highest term sent to the device by the node

VARIABLE *sentTerm*

Whether the node has received a *MasterArbitrationUpdate* indicating it is the master

VARIABLE *isMaster*

A counter used to generate unique stream *IDs*

VARIABLE *streamId*

---

Mastership/consensus related variables

$mastershipVars \triangleq \langle term, master, backups \rangle$

Mastership arbitration variables

$arbitrationVars \triangleq \langle streamVars, streamId \rangle$

Mastership event variables

$eventVars \triangleq \langle events, mastership \rangle$

---

Node related variables

$$nodeVars \triangleq \langle events, mastership, sentTerm, streamId, isMaster \rangle$$

This section models the *mastership* election service used by the controller to elect masters. Mastership changes through join and leave steps. Mastership is done through a consensus service, so these steps are atomic. When a node joins or leaves the *mastership* election, events are queued to notify nodes of the *mastership* change. Nodes learn of *mastership* changes independently of the state change in the consensus service.

Returns the set of values in  $f$

$$Range(f) \triangleq \{f[x] : x \in \text{DOMAIN } f\}$$

Returns a sequences with the element at the given index removed

$$Drop(q, i) \triangleq SubSeq(q, 1, i - 1) \circ SubSeq(q, i + 1, Len(q))$$

Node 'n' joins the *mastership* election

If the current 'master' is *Nil*, set the master to node 'n', increment the 'term', and send a *mastership* change event to each node. If the current 'master' is non-*Nil*, append node 'n' to the sequence of 'backups'.

$$\begin{aligned} JoinMastershipElection(n) \triangleq & \\ & \wedge \vee \wedge master = Nil \\ & \quad \wedge term' = term + 1 \\ & \quad \wedge master' = n \\ & \quad \wedge backups' = \langle \rangle \\ & \quad \wedge events' = [i \in Nodes \mapsto Append(events[i], [ \\ & \quad \quad \quad term \mapsto term', \\ & \quad \quad \quad master \mapsto master', \\ & \quad \quad \quad backups \mapsto backups'])]) \\ & \vee \wedge master \neq Nil \\ & \quad \wedge master \neq n \\ & \quad \wedge n \notin Range(backups) \\ & \quad \wedge backups' = Append(backups, n) \\ & \quad \wedge events' = [i \in Nodes \mapsto Append(events[i], [ \\ & \quad \quad \quad term \mapsto term, \\ & \quad \quad \quad master \mapsto master, \\ & \quad \quad \quad backups \mapsto backups'])]) \\ & \quad \wedge \text{UNCHANGED } \langle term, master \rangle \\ & \quad \wedge \text{UNCHANGED } \langle mastership, sentTerm, isMaster, messageVars, arbitrationVars \rangle \end{aligned}$$

Node 'n' leaves the *mastership* election

If node 'n' is the current 'master' and a backup exists, increment the 'term', promote the first backup to master, and send a *mastership* change event to each node. If node 'n' is the current 'master' and no backups exist, set the 'master' to *Nil*. If node 'n' is in the sequence of 'backups', simply remove it.

$$\begin{aligned} LeaveMastershipElection(n) \triangleq & \\ & \wedge \vee \wedge master = n \\ & \quad \wedge \vee \wedge Len(backups) > 0 \end{aligned}$$

$$\begin{aligned}
& \wedge term' = term + 1 \\
& \wedge master' = backups[1] \\
& \wedge backups' = Pop(backups) \\
& \wedge events' = [i \in Nodes \mapsto Append(events[i], [ \\
& \quad \quad \quad term \mapsto term', \\
& \quad \quad \quad master \mapsto master', \\
& \quad \quad \quad backups \mapsto backups'])] \\
& \vee \wedge Len(backups) = 0 \\
& \quad \wedge master' = Nil \\
& \quad \wedge UNCHANGED \langle term, backups, events \rangle \\
& \vee \wedge n \in Range(backups) \\
& \quad \wedge backups' = Drop(backups, CHOOSE j \in DOMAIN backups : backups[j] = n) \\
& \quad \wedge UNCHANGED \langle term, master, events \rangle \\
& \wedge UNCHANGED \langle mastership, sentTerm, isMaster, messageVars, arbitrationVars \rangle
\end{aligned}$$


---

This section models controller-side stream management.

Opens a new stream on the controller side

$$\begin{aligned}
OpenStream(n) & \triangleq \\
& \wedge requestStream[n].state = Closed \\
& \wedge streamId' = streamId + 1 \\
& \wedge requestStream' = [requestStream \text{ EXCEPT } ![n] = [id \mapsto streamId', state \mapsto Open]] \\
& \wedge requests' = [requests \text{ EXCEPT } ![n] = \langle \rangle] \\
& \wedge responses' = [responses \text{ EXCEPT } ![n] = \langle \rangle] \\
& \wedge UNCHANGED \langle mastershipVars, eventVars, sentTerm, isMaster, responseStream \rangle
\end{aligned}$$

Closes an open stream on the controller side

$$\begin{aligned}
CloseStream(n) & \triangleq \\
& \wedge requestStream[n].state = Open \\
& \wedge requestStream' = [requestStream \text{ EXCEPT } ![n].state = Closed] \\
& \wedge sentTerm' = [sentTerm \text{ EXCEPT } ![n] = 0] \\
& \wedge isMaster' = [isMaster \text{ EXCEPT } ![n] = FALSE] \\
& \wedge UNCHANGED \langle mastershipVars, eventVars, responseStream, messageVars, streamId \rangle
\end{aligned}$$


---

This section models controller-side *mastership* arbitration. The controller nodes receive *mastership* change events from the *mastership* service and send master arbitration requests to the device. Additionally, master nodes can send write requests to the device.

Returns master node 'n' *election\_id* for *mastership* term 'm'

$$MasterElectionId(m) \triangleq m.term + Cardinality(Nodes)$$

Returns the index of node 'n' in the sequence of 'm' backups

$$BackupIndex(n, m) \triangleq CHOOSE i \in DOMAIN m.backups : m.backups[i] = n$$

Returns backup node 'n' *election\_id* for *mastership* term 'm'

$$\text{BackupElectionId}(n, m) \triangleq \text{MasterElectionId}(m) - \text{BackupIndex}(n, m)$$

Returns the *mastership* term for *MasterArbitrationUpdate* 'm'  
 $\text{MasterTerm}(m) \triangleq m.\text{election\_id} - \text{Cardinality}(\text{Nodes})$

Node 'n' receives a *mastership* change event from the *mastership* service

When a *mastership* change event is received, the node's local *mastership* state is updated. If the *mastership* term has changed, the node will set a flag to push the *mastership* change to the device in the master arbitration step.

$$\begin{aligned} \text{LearnMastership}(n) &\triangleq \\ &\wedge \text{Len}(\text{events}[n]) > 0 \\ &\wedge \text{LET } e \triangleq \text{events}[n][1] \\ &\quad m \triangleq \text{mastership}[n] \\ &\text{IN} \\ &\quad \vee \wedge e.\text{term} > m.\text{term} \\ &\quad \quad \wedge \text{mastership}' = [\text{mastership} \text{ EXCEPT } ![n] = [ \\ &\quad \quad \quad \text{term} \mapsto e.\text{term}, \\ &\quad \quad \quad \text{master} \mapsto e.\text{master}, \\ &\quad \quad \quad \text{backups} \mapsto e.\text{backups}] \\ &\quad \vee \wedge e.\text{term} = m.\text{term} \\ &\quad \quad \wedge \text{mastership}' = [\text{mastership} \text{ EXCEPT } ![n] = [ \\ &\quad \quad \quad \text{term} \mapsto e.\text{term}, \\ &\quad \quad \quad \text{master} \mapsto e.\text{master}, \\ &\quad \quad \quad \text{backups} \mapsto e.\text{backups}] \\ &\quad \wedge \text{events}' = [\text{events} \text{ EXCEPT } ![n] = \text{Pop}(\text{events}[n])] \\ &\quad \wedge \text{UNCHANGED } \langle \text{mastershipVars}, \text{sentTerm}, \text{isMaster}, \text{messageVars}, \text{arbitrationVars} \rangle \end{aligned}$$

Node 'n' sends a *MasterArbitrationUpdate* to the device

If the node has an open stream to the device and a valid *mastership* state, a *MasterArbitrationUpdate* is sent to the device. If the node is a backup, the request's 'election\_id' is set to (*mastership* term) + (number of nodes) - (backup index). If the node is the master, the 'election\_id' is set to (*mastership* term) + (number of nodes). This is done to avoid *election\_ids*  $\leq 0$ . Note that the actual protocol requires a (*device\_id*, *role\_id*, *election\_id*) tuple, but (*device\_id*, *role\_id*) have been excluded from this model as we're modelling interaction only within a single (*device\_id*, *role\_id*) and thus they're irrelevant to correctness. The *mastership* term is sent in *MasterArbitrationUpdate* requests for model checking.

$$\begin{aligned} \text{SendMasterArbitrationUpdate}(n) &\triangleq \\ &\wedge \text{requestStream}[n].\text{state} = \text{Open} \\ &\wedge \text{LET } m \triangleq \text{mastership}[n] \\ &\text{IN} \\ &\quad \wedge m.\text{term} > 0 \\ &\quad \wedge \text{sentTerm}[n] < m.\text{term} \\ &\quad \wedge \vee \wedge m.\text{master} = n \\ &\quad \quad \wedge \text{SendRequest}(n, [ \\ &\quad \quad \quad \text{type} \mapsto \text{MasterArbitrationUpdate}, \\ &\quad \quad \quad \text{election\_id} \mapsto \text{MasterElectionId}(m)]) \\ &\quad \vee \wedge m.\text{master} \neq n \end{aligned}$$

$$\begin{aligned}
& \wedge n \in \text{Range}(m.\text{backups}) \\
& \wedge \text{SendRequest}(n, [ \\
& \quad \text{type} \mapsto \text{MasterArbitrationUpdate}, \\
& \quad \text{election\_id} \mapsto \text{BackupElectionId}(n, m)]) \\
& \wedge \text{sentTerm}' = [\text{sentTerm} \text{ EXCEPT } ![n] = m.\text{term}] \\
& \wedge \text{UNCHANGED } \langle \text{mastershipVars}, \text{eventVars}, \text{isMaster}, \text{arbitrationVars}, \text{responses} \rangle
\end{aligned}$$

Node 'n' receives a *MasterArbitrationUpdate* from the device

If the node has an open stream with a *MasterArbitrationUpdate*, determine whether the local node is the master. If the *MasterArbitrationUpdate* 'status' is *Ok*, the 'election\_id' matches the last requested *mastership* term, and 'n' is the master for that term, update the node's state to master. Otherwise, the *mastership* request is considered out of date.

Note that the separate 'isMaster' state is maintained to indicate whether the \*device\* considers this node to be the current master, and this is necessary for the safety of the algorithm. Both the node and the device must agree on the role of the node.

$$\begin{aligned}
& \text{ReceiveMasterArbitrationUpdate}(n) \triangleq \\
& \quad \wedge \text{requestStream}[n].\text{state} = \text{Open} \\
& \quad \wedge \text{HasResponse}(n, \text{MasterArbitrationUpdate}) \\
& \quad \wedge \text{LET } r \triangleq \text{NextResponse}(n) \\
& \quad \quad m \triangleq \text{mastership}[n] \\
& \quad \text{IN} \\
& \quad \quad \vee \wedge r.\text{status} = \text{Ok} \\
& \quad \quad \quad \wedge m.\text{master} = n \\
& \quad \quad \quad \wedge m.\text{term} = \text{MasterTerm}(r) \\
& \quad \quad \quad \wedge \text{sentTerm}[n] = m.\text{term} \\
& \quad \quad \quad \wedge \text{isMaster}' = [\text{isMaster} \text{ EXCEPT } ![n] = \text{TRUE}] \\
& \quad \quad \vee \wedge \vee r.\text{status} \neq \text{Ok} \\
& \quad \quad \quad \vee m.\text{master} \neq n \\
& \quad \quad \quad \vee \text{sentTerm}[n] \neq m.\text{term} \\
& \quad \quad \quad \vee m.\text{term} \neq \text{MasterTerm}(r) \\
& \quad \quad \quad \wedge \text{isMaster}' = [\text{isMaster} \text{ EXCEPT } ![n] = \text{FALSE}] \\
& \quad \wedge \text{DiscardResponse}(n) \\
& \quad \wedge \text{UNCHANGED } \langle \text{mastershipVars}, \text{eventVars}, \text{sentTerm}, \text{arbitrationVars}, \text{requests} \rangle
\end{aligned}$$

Master node 'n' sends a *WriteRequest* to the device

To write to the device, the node must have an open stream, must have received a *mastership* change event from the *mastership* service (stored in 'mastership') indicating it is the master, and must have received a *MasterArbitrationUpdate* from the switch indicating it is the master (stored in 'isMaster') for the same term as was indicated by the *mastership* service. Additionally, the node's current term is sent as the *WriteRequest* 'term' to avoid writes from a master that has since been superseded by a newer master.

$$\begin{aligned}
& \text{SendWriteRequest}(n) \triangleq \\
& \quad \wedge \text{requestStream}[n].\text{state} = \text{Open} \\
& \quad \wedge \text{LET } m \triangleq \text{mastership}[n] \\
& \quad \text{IN} \\
& \quad \quad \wedge m.\text{term} > 0
\end{aligned}$$

$$\begin{aligned}
& \wedge m.master = n \\
& \wedge isMaster[n] \\
& \wedge SendRequest(n, [ \\
& \quad type \mapsto WriteRequest, \\
& \quad election\_id \mapsto MasterElectionId(m), \\
& \quad term \mapsto m.term]) \\
& \wedge UNCHANGED \langle mastershipVars, eventVars, arbitrationVars, isMaster, sentTerm, responses \rangle
\end{aligned}$$

Node 'n' receives a write response from the device

$$\begin{aligned}
ReceiveWriteResponse(n) & \triangleq \\
& \wedge requestStream[n].state = Open \\
& \wedge HasResponse(n, WriteResponse) \\
& \wedge LET \ m \triangleq NextResponse(n) \\
& \quad IN \\
& \quad \vee m.status = Ok \\
& \quad \vee m.status = PermissionDenied \\
& \wedge DiscardResponse(n) \\
& \wedge UNCHANGED \langle mastershipVars, nodeVars, arbitrationVars, requests \rangle
\end{aligned}$$


---

\ \* Modification History  
\ \* Last modified Tue Feb 26 13:24:44 PST 2019 by jordanhalterman  
\ \* Created Wed Feb 20 23:49:08 PST 2019 by jordanhalterman