MODULE *paxos*

This is a specification of the paxos algorithm implemented in Ceph. The specification is based on the following source file: https://github.com/ceph/ceph/blob/master/src/mon/Paxos.cc

The main mechanism abstracted that may differ from the version implemented in Ceph are:

- The election logic. The leader is chosen randomly, and, for now, only one leader is chosen per epoch. When a new epoch begins, the messages from the previous epoch are discarded.
- Monitor quorum. The quorum is defined in the election phase, using all monitors that are up. Different epochs can have different quorums.
- The communication layer. The variable messages represents connections between monitors (e.g. messages[mon1][mon2] holds the messages sent from mon1 to mon2). Within a connection the messages are sent and received in order.
- The transactions. Transactions are simplified to represent only a change of a value in the variable monitor_store.
- Failure model. A monitor can crash if the remaining number of monitors is sufficient to form a quorum. When a monitor crashes, new elections are triggered and the monitor is marked to not be part of a quorum until he recovers.
- Timeouts. A timeout can occur at any point in the algorithm and it will trigger new elections.

For a more detailed overview of the specification: https://github.com/afonsonf/ceph-consensus-spec

EXTENDS *Integers*, *FiniteSets*, *Sequences*, *TLC*, *SequencesExt*, *FiniteSetsExt*

External libraries used on:
SequencesExt: *SetToSeq*
FiniteSetsExt: *Min*, *Max*

Constants

Set of *Monitors*.

CONSTANTS *Monitors*

$$\begin{aligned} &\mathit{MonitorsSeq} \triangleq \mathit{TLCEval}(\mathit{SetToSeq}(\mathit{Monitors})) \\ &\mathit{MonitorsLen} \triangleq \mathit{TLCEval}(\mathit{Len}(\mathit{MonitorsSeq})) \end{aligned}$$

Rank predicate, used to compute proposal numbers.

$$\mathit{rank}(\mathit{mon}) \triangleq \textsc{choose } i \in 1 \,..\, \mathit{MonitorsLen} : \mathit{MonitorsSeq}[i] = \mathit{mon}$$

Set of possible values.

CONSTANTS *Value_set*

Predicate used in the cfg file to define the symmetry set.

$$\mathit{SYMM} \triangleq \mathit{Permutations}(\mathit{Monitors}) \cup \mathit{Permutations}(\mathit{Value\_set})$$

Reserved value.

CONSTANTS *Nil*

Paxos states:

CONSTANTS $STATE\_RECOVERING$, $STATE\_ACTIVE$,
$STATE\_UPDATING$, $STATE\_UPDATING\_PREVIOUS$,
$STATE\_WRITING$, $STATE\_WRITING\_PREVIOUS$,
$STATE\_REFRESH$, $STATE\_SHUTDOWN$

$$\begin{aligned} state\_names \triangleq \{&STATE\_RECOVERING, STATE\_ACTIVE,\\ &STATE\_UPDATING, STATE\_UPDATING\_PREVIOUS,\\ &STATE\_WRITING, STATE\_WRITING\_PREVIOUS,\\ &STATE\_REFRESH, STATE\_SHUTDOWN\}\end{aligned}$$

Paxos auxiliary phase states:

They are used to force some sequence of steps.

CONSTANTS $PHASE\_ELECTION$,
$PHASE\_SEND\_COLLECT$, $PHASE\_COLLECT$,
$PHASE\_LEASE$, $PHASE\_LEASE\_DONE$,
$PHASE\_BEGIN$,
$PHASE\_COMMIT$

$$\begin{aligned} phase\_names \triangleq \{&PHASE\_ELECTION,\\ &PHASE\_SEND\_COLLECT, PHASE\_COLLECT,\\ &PHASE\_LEASE, PHASE\_LEASE\_DONE,\\ &PHASE\_BEGIN,\\ &PHASE\_COMMIT\}\end{aligned}$$

Paxos message types:

CONSTANTS $OP\_COLLECT$, OP_LAST,
OP_BEGIN, OP_ACCEPT, OP_COMMIT,
OP_LEASE, OP_LEASE_ACK

$$\begin{aligned} messages\_types \triangleq \{&OP\_COLLECT, OP\_LAST,\\ &OP\_BEGIN, OP\_ACCEPT, OP\_COMMIT,\\ &OP\_LEASE, OP\_LEASE\_ACK\}\end{aligned}$$

Global variables

Integer representing the current epoch. If is odd trigger an election.

Type: Integer

VARIABLE $epoch$

Store messages waiting to be handled.

Type: $[Monitors \mapsto [Monitors \mapsto \langle message \rangle]]$

VARIABLE $messages$

Stores history of messages. Can be useful to find specific states.

Type: $\{messages\}$

VARIABLE $message\_history$

Stores if a monitor is up or down. All available monitors, in a given epoch, are part of the quorum.
Type: $[\mathit{Monitors} \mapsto \mathit{Bool}]$

VARIABLE *quorum*

Size of the current quorum.
Type: *Int*

VARIABLE *quorum_sz*

State variables

A function that stores the current leader. *isLeader*[*mon*] is True iff *mon* is a leader, else False.
Type: $[\mathit{Monitors} \mapsto \mathit{Bool}]$

VARIABLE *isLeader*

A function that stores the state of each monitor.
Type: $[\mathit{Monitors} \mapsto \mathit{state\_names}]$

VARIABLE *state*

A function that stores the phase of each monitor.
Type: $[\mathit{Monitors} \mapsto \mathit{phase\_names}]$

VARIABLE *phase*

Restart variables

A function that stores, for each monitor, a proposal number when the commit phase starts.
This proposal number can be retrieved after a monitor crashes and restarts.
Type: $[\mathit{Monitors} \mapsto$ proposal number$]$

VARIABLE *uncommitted_pn*

A function that stores, for each monitor, a value version when the commit phase starts.
This value version can be retrieved after a monitor crashes and restarts.
Type: $[\mathit{Monitors} \mapsto$ value version$]$

VARIABLE *uncommitted_v*

A function that stores, for each monitor, a value when the commit phase starts.
This value can be retrieved after a monitor crashes and restarts.
Type: $[\mathit{Monitors} \mapsto \mathit{Value\_set}]$

VARIABLE *uncommitted_value*

Data variables

A function that stores, for each monitor, the store where the transactions are applied.
In this model, a transaction represents changing the value in the store.
Type: $[\mathit{Monitors} \mapsto \mathit{Value\_set}]$

VARIABLE *monitor_store*

A function that stores the transaction log of each monitor.
Type: $[\mathit{Monitors} \mapsto$ [value $\mathit{version} \mapsto \mathit{Value\_set}]]$

VARIABLE *values*

A function that stores the last proposal number accepted by each monitor.
Type: [$Monitors \mapsto$ proposal number]
VARIABLE *accepted_pn*

A function that stores the first value version committed by each monitor.
Type: [$Monitors \mapsto$ value version]
VARIABLE *first_committed*

A function that stores the last value version committed by each monitor.
Type: [$Monitors \mapsto$ value version]
VARIABLE *last_committed*

Collect phase variables

A function that stores the number of peers that accepted a collect request.
Type: [$Monitors \mapsto$ number of peers that accepted]
VARIABLE *num_last*

Used by leader when receiving responses in collect phase.
Type: [$Monitors \mapsto$ [$Monitors \mapsto$ value version]]
VARIABLE *peer_first_committed*

Used by leader when receiving responses in collect phase.
Type: [$Monitors \mapsto$ [$Monitors \mapsto$ value version]]
VARIABLE *peer_last_committed*

Lease phase variables

A function that stores, for each monitor, which of the peers have acked the lease request.
Type: [$Monitors \mapsto$ [$Monitors \mapsto Bool$]]
VARIABLE *acked_lease*

Commit phase variables

A function that stores, for each monitor, the value proposed by a client.
Type: [$Monitors \mapsto Value\_set \cup \{Nil\}$]
VARIABLE *pending_proposal*

A function that stores, for each monitor, the value to be committed in the begin phase.
Type: [$Monitors \mapsto Value\_set \cup \{Nil\}$]
VARIABLE *new_value*

A function that stores, for each monitor, which of the peers have acked the begin request.
Type: [$Monitord \mapsto$ [$Monitors \mapsto Bool$]]
VARIABLE *accepted*

Debug variables

Variables to help debug a behavior.
step is the diameter of a behavior/path.
step_name the current predicate being called.

VARIABLE $step\_name$

Variables to limit the number of monitors crashes that can occur over a behavior.
This variable is used to limit the search space.

VARIABLE $number\_crashes$

Variables initialization

$$\begin{aligned}
&global\_vars &&\triangleq \langle epoch, messages, message\_history, quorum, quorum\_sz\rangle\\
&state\_vars &&\triangleq \langle isLeader, state, phase\rangle\\
&restart\_vars &&\triangleq \langle uncommitted\_pn, uncommitted\_v, uncommitted\_value\rangle\\
&data\_vars &&\triangleq \langle monitor\_store, values, accepted\_pn, first\_committed, last\_committed\rangle\\
&collect\_vars &&\triangleq \langle num\_last, peer\_first\_committed, peer\_last\_committed\rangle\\
&lease\_vars &&\triangleq \langle acked\_lease\rangle\\
&commit\_vars &&\triangleq \langle pending\_proposal, new\_value, accepted\rangle
\end{aligned}$$

$$\begin{aligned}
vars \triangleq \langle &global\_vars, state\_vars, restart\_vars, data\_vars, collect\_vars,\\
&lease\_vars, commit\_vars\rangle
\end{aligned}$$

$$\begin{aligned}
&Init\_global\_vars \triangleq\\
&\quad \land epoch = 1\\
&\quad \land messages = [mon1 \in Monitors \mapsto [mon2 \in Monitors \mapsto \langle\rangle]]\\
&\quad \land message\_history = \{\}\\
&\quad \land quorum = [mon \in Monitors \mapsto \text{TRUE}]\\
&\quad \land quorum\_sz = MonitorsLen
\end{aligned}$$

$$\begin{aligned}
&Init\_state\_vars \triangleq\\
&\quad \land isLeader = [mon \in Monitors \mapsto \text{FALSE}]\\
&\quad \land state \ = [mon \in Monitors \mapsto Nil]\\
&\quad \land phase = [mon \in Monitors \mapsto Nil]
\end{aligned}$$

$$\begin{aligned}
&Init\_restart\_vars \triangleq\\
&\quad \land uncommitted\_pn = [mon \in Monitors \mapsto 0]\\
&\quad \land uncommitted\_v = [mon \in Monitors \mapsto 0]\\
&\quad \land uncommitted\_value = [mon \in Monitors \mapsto Nil]
\end{aligned}$$

$$\begin{aligned}
&Init\_data\_vars \triangleq\\
&\quad \land monitor\_store \ = [mon \in Monitors \mapsto Nil]\\
&\quad \land values = [mon \in Monitors \mapsto [version \in \{\} \mapsto Nil]]\\
&\quad \land accepted\_pn = [mon \in Monitors \mapsto 0]\\
&\quad \land first\_committed = [mon \in Monitors \mapsto 0]\\
&\quad \land last\_committed = [mon \in Monitors \mapsto 0]
\end{aligned}$$

$$\begin{aligned}
&Init\_collect\_vars \triangleq\\
&\quad \land num\_last = [mon \in Monitors \mapsto 0]
\end{aligned}$$

```
    $\land\ peer\_first\_committed = [mon1 \in Monitors \mapsto [mon2 \in Monitors \mapsto\ -1]]$
    $\land\ peer\_last\_committed = [mon1 \in Monitors \mapsto [mon2 \in Monitors \mapsto\ -1]]$

$Init\_lease\_vars\ \triangleq$
    $\land\ acked\_lease = [mon1 \in Monitors \mapsto [mon2 \in Monitors \mapsto \text{FALSE}]]$

$Init\_commit\_vars\ \triangleq$
    $\land\ pending\_proposal\ \ = [mon \in Monitors \mapsto Nil]$
    $\land\ new\_value = [mon \in Monitors \mapsto Nil]$
    $\land\ accepted = [mon1\ \ \in Monitors \mapsto [mon2 \in Monitors \mapsto \text{FALSE}]]$

$Init\ \triangleq$
    $\land\ Init\_global\_vars$
    $\land\ Init\_state\_vars$
    $\land\ Init\_restart\_vars$
    $\land\ Init\_data\_vars$
    $\land\ Init\_collect\_vars$
    $\land\ Init\_lease\_vars$
    $\land\ Init\_commit\_vars$
    $\land\ step\_name = \text{“init”} \land number\_crashes = 0$
```

Message manipulation

Note: Variable *message_history* has impact in performace, update only when debugging.

Converts a set with at most one element to a sequence.

```
$SingleMessageSetToSeq(S)\ \triangleq$
   $\text{IF}\ \exists\, elem \in S : \text{TRUE THEN LET}\ elem\ \triangleq\ \text{CHOOSE}\ x \in S : \text{TRUE}$
                                  $\text{IN}\quad \langle elem \rangle$
                            $\text{ELSE}\ \ \langle\rangle$
```

Add message *m* to the network *msgs*.

```
$WithMessage(m,\ msgs)\ \triangleq$
   $[msgs\ \text{EXCEPT}\ ![m.from] =$
      $[msgs[m.from]\ \text{EXCEPT}\ ![m.dest] = Append(msgs[m.from][m.dest],\ m)]]$
```

Remove message *m* from the network *msgs*.

```
$WithoutMessage(m,\ msgs)\ \triangleq$
   $[msgs\ \text{EXCEPT}\ ![m.from] =$
      $[msgs[m.from]\ \text{EXCEPT}\ ![m.dest] = Tail(msgs[m.from][m.dest])]]$
```

Adds the message *m* to the network.

Variables changed: messages, *message_history*.

```
$Send(m)\ \triangleq$
     $\land\ messages' = WithMessage(m,\ messages)$
      $\land\ message\_history' = message\_history \cup \{m\}$
     $\land\ \text{UNCHANGED}\ \ message\_history$
```

Adds a set of messages to the network.
Variables changed: messages, *message_history*.

```
$Send\_set(from, m\_set) \triangleq$
    $\land messages' = [messages \text{ EXCEPT } ![from] =$
        $[mon \in Monitors \mapsto$
            $messages[from][mon] \circ SingleMessageSetToSeq(\{m \in m\_set : m.dest = mon\})]]$
    $\land message\_history' = message\_history \cup m\_set$
    $\land \text{UNCHANGED } message\_history$
```

Removes the request from network and adds the response.
Variables changed: messages, *message_history*.

```
$Reply(response, request) \triangleq$
    $\land messages' = WithoutMessage(request, WithMessage(response, messages))$
    $\land message\_history' = message\_history \cup \{response\}$
    $\land \text{UNCHANGED } message\_history$
```

Removes the request from network and adds a set of messages.
Variables changed: messages, *message_history*.

```
$Reply\_set(from, response\_set, request) \triangleq$
    $\land \text{LET } msgs \triangleq WithoutMessage(request, messages)$
      $\text{IN } messages' = [msgs \text{ EXCEPT } ![from] =$
            $[mon \in Monitors \mapsto$
                $msgs[from][mon] \circ SingleMessageSetToSeq(\{m \in response\_set : m.dest = mon\})]]$
    $\land message\_history' = message\_history \cup response\_set$
    $\land \text{UNCHANGED } message\_history$
```

Removes message m from the network.
Variables changed: messages, *message_history*.

```
$Discard(m) \triangleq$
    $\land messages' = WithoutMessage(m, messages)$
    $\land \text{UNCHANGED } message\_history$
```

Helper predicates

Computes a new unique proposal number for a given monitor.

Version A - Equal to the one in the source.
This version breaks the symmetry of the monitor set.
Example: $oldpn = 305$, $rank(mon) = 5$, $newpn = 405$.
$get\_new\_proposal\_number(mon, oldpn) \triangleq ((oldpn \div 100) + 1) * 100 + rank(mon)$

Version $B -$ Adapted to not break symmetry.
Example: $oldpn = 300$, $rank(mon) = 5$, $newpn = 400$.

$$get\_new\_proposal\_number(mon, oldpn) \triangleq ((oldpn \div 100) + 1) * 100$$

Clear the variable *peer_first_committed*.

Variables changed: *peer_first_committed*.

```
$clear\_peer\_first\_committed(mon) \triangleq$
    $peer\_first\_committed' = [peer\_first\_committed$ EXCEPT $![mon] =$
                                     $[m \in Monitors \mapsto -1]]$
```

Clear the variable *peer_last_committed*.

Variables changed: *peer_last_committed*.

```
$clear\_peer\_last\_committed(mon) \triangleq$
    $peer\_last\_committed' = [peer\_last\_committed$ EXCEPT $![mon] =$
                                     $[m \in Monitors \mapsto -1]]$
```

Store peer values and update *first_committed*, *last_committed* and *monitor_store* accordingly.

Variables changed: values, *first_committed*, *last_committed*, *monitor_store*.

```
$store\_state(mon, msg) \triangleq$
      Choose peer values from $mon$ last committed $+1$ to peer last committed.
    $\land$ LET $logs \triangleq$ (DOMAIN $msg.values) \cap (last\_committed[mon] + 1 \,..\, msg.last\_committed)$
       IN   $\land$ $values' = [values$ EXCEPT $![mon] =$
                 $[i \in$ DOMAIN $values[mon] \cup logs \mapsto$
                    IF $i \in logs$
                     THEN $msg.values[i]$
                     ELSE $values[mon][i]]]$
              Update last committed and first committed.
            $\land$ $last\_committed' = [last\_committed$ EXCEPT $![mon] = Max(logs \cup \{last\_committed[mon]\})]$
            $\land$ IF $logs \neq \{\} \land first\_committed[mon] = 0$
                THEN $first\_committed' =$
                          $[first\_committed$ EXCEPT $![mon] = Min(logs)]$
                ELSE $first\_committed' =$
                          $[first\_committed$ EXCEPT $![mon] = Min(logs \cup \{first\_committed[mon]\})]$
      Update monitor store.
    $\land$ IF $last\_committed'[mon] = 0$
       THEN UNCHANGED $monitor\_store$
       ELSE $monitor\_store' = [monitor\_store$ EXCEPT $![mon] = values'[mon][last\_committed'[mon]]]$
```

Check if uncommitted value version is still valid, else reset it.

Variables changed: *uncommitted_pn*, *uncommitted_v*, *uncommitted_value*.

```
$check\_and\_correct\_uncommitted(mon) \triangleq$
    IF $uncommitted\_v[mon] \leq last\_committed'[mon]$
     THEN $\land$ $uncommitted\_v' = [uncommitted\_v$ EXCEPT $![mon] = 0]$
          $\land$ $uncommitted\_pn' = [uncommitted\_pn$ EXCEPT $![mon] = 0]$
          $\land$ $uncommitted\_value' = [uncommitted\_value$ EXCEPT $![mon] = Nil]$
     ELSE UNCHANGED $\langle uncommitted\_pn, uncommitted\_v, uncommitted\_value \rangle$
```

Trigger new election by incrementing epoch.

Variables changed: epoch.

```
$bootstrap \triangleq$
    $\land$ $epoch' = epoch + 1$
```

Lease phase predicates

Changes *mon* state to *STATE_ACTIVE*.

Variables changed: state.

```
$finish\_round(mon) \triangleq$
    $\land\ isLeader[mon] = \text{TRUE}$
    $\land\ state' = [state\ \text{EXCEPT}\ ![mon] = STATE\_ACTIVE]$
```

Resets the variable acked lease and send lease messages to peers.

Variables changed: *acked_lease*, messages, *message_history*, phase.

```
$extend\_lease(mon) \triangleq$
    $\land\ isLeader[mon] = \text{TRUE}$
    $\land\ acked\_lease' = [acked\_lease\ \text{EXCEPT}\ ![mon] =$
        $[m \in Monitors \mapsto \text{IF}\ m = mon\ \text{THEN}\ \text{TRUE}\ \text{ELSE}\ \text{FALSE}]]$
    $\land\ Send\_set(mon,$
        $\{[type \mapsto OP\_LEASE,$
          $from \mapsto mon,$
          $dest \mapsto dest,$
          $last\_committed \mapsto last\_committed[mon]] : dest \in \{m \in Monitors \setminus \{mon\} : quorum[m]\}$
         $\})$
    $\land\ phase' = [phase\ \text{EXCEPT}\ ![mon] = PHASE\_LEASE]$
```

Handle a lease message. The peon changes his state and replies with a lease ack message.

The reply is commented because the lease ack is only used to check if all peers are up.

In the model this is done by “randomly” triggering the predicate *Timeout*. In this way, the search space is reduced.

Variables changed: messages, *message_history*, state.

```
$handle\_lease(mon, msg) \triangleq$
    $\land$  discard if not peon or peon is behind
      $\text{IF}\ \lor\ isLeader[mon] = \text{TRUE}$
          $\lor\ last\_committed[mon] \neq msg.last\_committed$
       $\text{THEN}\ \land\ Discard(msg)$
              $\land\ \text{UNCHANGED}\ state$
       $\text{ELSE}\ \ \land\ state' = [state\ \text{EXCEPT}\ ![mon] = STATE\_ACTIVE]$
          $\land\ Reply([type \mapsto OP\_LEASE\_ACK,$
              $from \mapsto mon,$
              $dest \mapsto msg.from,$
              $first\_committed \mapsto first\_committed[mon],$
              $last\_committed \mapsto last\_committed[mon]], msg)$
            $\land\ Discard(msg)$
    $\land\ \text{UNCHANGED}\ \langle epoch, quorum, quorum\_sz, isLeader, phase\rangle$
    $\land\ \text{UNCHANGED}\ \langle restart\_vars, data\_vars, collect\_vars, lease\_vars, commit\_vars\rangle$
```

Handle a lease ack message. The leader updates the *acked_lease* variable.

Because the *lease_ack* messages are not sent, this predicate is never called.

The reasoning for this is given in *handle_lease* comment.

Variables changed: *acked_lease*, messages, *message_history*.

```
$handle\_lease\_ack(mon, msg) \triangleq$
    $\land\ phase[mon] = PHASE\_LEASE$
    $\land\ acked\_lease' = [acked\_lease \text{ EXCEPT } ![mon] =$
        $[acked\_lease[mon] \text{ EXCEPT } ![msg.from] = \text{TRUE}]]$
    $\land\ Discard(msg)$
    $\land\ \text{UNCHANGED } \langle epoch, quorum, quorum\_sz \rangle$
    $\land\ \text{UNCHANGED } \langle state\_vars, restart\_vars, data\_vars, collect\_vars, commit\_vars \rangle$
```

Predicate that is called when all peers ack the lease. The phase is changed to prevent loops.
Because the *lease_ack* messages are not sent, this predicate is never called.
The reasoning for this is given in *handle_lease* comment.
Variables changed: phase.

```
$post\_lease\_ack(mon) \triangleq$
    $\land\ phase[mon] = PHASE\_LEASE$
    $\land\ phase' = [phase \text{ EXCEPT } ![mon] = PHASE\_LEASE\_DONE]$
    $\land\ \forall m \in Monitors : quorum[m] \Rightarrow acked\_lease[mon][m] = \text{TRUE}$
    $\land\ \text{UNCHANGED } \langle isLeader, state \rangle$
    $\land\ \text{UNCHANGED } \langle global\_vars, restart\_vars, data\_vars, collect\_vars,$
                          $lease\_vars, commit\_vars \rangle$
```

Commit phase predicates

Start a commit phase by the leader. The variable *new_value* is assigned. Send begin messages to the peers.
The value of *uncommitted_v* and *uncommitted_value* are assigned in order for the leader to be
able to recover from a crash.
Variables changed: accepted, *new_value*, phase, messages, *message_history*, values, *uncommitted_pn*, *uncommitted_v*, *uncomm*

```
$begin(mon, v) \triangleq$
    $\land\ isLeader[mon] = \text{TRUE}$
    $\land\ \lor\ state'[mon] = STATE\_UPDATING$
      $\lor\ state'[mon] = STATE\_UPDATING\_PREVIOUS$
    $\land\ quorum\_sz = 1 \lor num\_last[mon] > MonitorsLen \div 2$
    $\land\ new\_value[mon] = Nil$
    $\land\ accepted' = [accepted \text{ EXCEPT } ![mon] =$
        $[m \in Monitors \mapsto \text{IF } m = mon \text{ THEN TRUE ELSE FALSE}]]$
    $\land\ new\_value' = [new\_value \text{ EXCEPT } ![mon] = v]$
    $\land\ phase' = [phase \text{ EXCEPT } ![mon] = PHASE\_BEGIN]$
    $\land\ values' = [values \text{ EXCEPT } ![mon] =$
        $((last\_committed[mon] + 1) :> new\_value'[mon]) \ @@ \ values[mon]]$
    $\land\ Send\_set(mon,$
        $\{[type            \mapsto OP\_BEGIN,$
          $from            \mapsto mon,$
          $dest            \mapsto dest,$
          $last\_committed  \mapsto last\_committed[mon],$
          $values          \mapsto values'[mon],$
    $pn \qquad \mapsto accepted\_pn[mon]] : dest \in \{m \in Monitors \setminus \{mon\} : quorum[m]\}$
    $\})$
  $\land uncommitted\_pn' = [uncommitted\_pn \text{ EXCEPT } ![mon] = accepted\_pn[mon]]$
  $\land uncommitted\_v' = [uncommitted\_v \text{ EXCEPT } ![mon] = last\_committed[mon] + 1]$
  $\land uncommitted\_value' = [uncommitted\_value \text{ EXCEPT } ![mon] = v]$
```

Handle a begin message. The monitor will accept if the proposal number in the message is greater or equal than the one he accepted.

Similar to what happens in begin, *uncommitted_v* and *uncommitted_value* are assigned in order for the monitor to recover in case of a crash.

Variables changed: messages, *message_history*, state, values, *uncommitted_pn*, *uncommitted_v*, *uncommitted_value*.

```
$handle\_begin(mon, msg) \triangleq$
  $\land isLeader[mon] = \text{FALSE}$
  $\land \text{IF } msg.pn < accepted\_pn[mon]$
     THEN
      $\land Discard(msg)$
      $\land \text{UNCHANGED } \langle state, values, restart\_vars \rangle$
     ELSE
      $\land msg.pn = accepted\_pn[mon]$
      $\land msg.last\_committed = last\_committed[mon]$

      assign $values[mon][last\_committed[mon] + 1]$
      $\land values' = [values \text{ EXCEPT } ![mon] =$
          $((last\_committed[mon] + 1) :> msg.values[last\_committed[mon] + 1])$ @@ $values[mon]]$

      $\land state' = [state \text{ EXCEPT } ![mon] = STATE\_UPDATING]$
      $\land uncommitted\_pn' = [uncommitted\_pn \text{ EXCEPT } ![mon] = accepted\_pn[mon]]$
      $\land uncommitted\_v' = [uncommitted\_v \text{ EXCEPT } ![mon] = last\_committed[mon] + 1]$
      $\land uncommitted\_value' = [uncommitted\_value \text{ EXCEPT } ![mon] =$
          $values'[mon][last\_committed[mon] + 1]]$
      $\land Reply([type \qquad \mapsto OP\_ACCEPT,$
                $from \qquad \mapsto mon,$
                $dest \qquad \mapsto msg.from,$
                $last\_committed \mapsto last\_committed[mon],$
                $pn \qquad \mapsto accepted\_pn[mon]], msg)$
  $\land \text{UNCHANGED } \langle epoch, quorum, quorum\_sz, isLeader, phase, monitor\_store,$
                 $accepted\_pn, first\_committed, last\_committed \rangle$
  $\land \text{UNCHANGED } \langle collect\_vars, lease\_vars, commit\_vars \rangle$
```

Handle an accept message. If the leader receives a positive response from the peer, it will add it to the variable accepted.

Variables changed: messages, *message_history*, accepted

```
$handle\_accept(mon, msg) \triangleq$
  $\land isLeader[mon] = \text{TRUE}$
  $\land \lor state[mon] \; = STATE\_UPDATING\_PREVIOUS$
     $\lor state[mon] \; = STATE\_UPDATING$
```

```
    $\land\ phase[mon] = PHASE\_BEGIN$
    $\land\ new\_value[mon] \neq Nil$
    $\land$ IF $\lor\ msg.pn \neq accepted\_pn[mon]$
           $\lor\ \land\ last\_committed[mon] > 0$
              $\land\ msg.last\_committed < last\_committed[mon] - 1$
       THEN UNCHANGED $accepted$
       ELSE $accepted' = [accepted$ EXCEPT $![mon] =$
                 $[accepted[mon]$ EXCEPT $![msg.from] =$ TRUE$]]$
    $\land\ Discard(msg)$
    $\land$ UNCHANGED $\langle epoch, quorum, quorum\_sz, pending\_proposal, new\_value\rangle$
    $\land$ UNCHANGED $\langle restart\_vars, state\_vars, data\_vars, collect\_vars, lease\_vars\rangle$
```

Predicate that is enabled and called when all peers in the quorum accept begin request from leader.
The leader commits the transaction in *new_value* and sends commit messages to his peers.
Variables changed: *first_committed*, *last_committed*, *monitor_store*, *new_value*, messages, *message_history*, state, phase

```
$post\_accept(mon) \triangleq$
    $\land\ phase[mon] = PHASE\_BEGIN$
    $\land\ \forall m \in Monitors : quorum[m] \Rightarrow accepted[mon][m] =$ TRUE
    $\land\ new\_value[mon] \neq Nil$
    $\land\ \lor\ state[mon] = STATE\_UPDATING\_PREVIOUS$
       $\lor\ state[mon] = STATE\_UPDATING$
    $\land\ last\_committed' = [last\_committed$ EXCEPT $![mon] = last\_committed[mon] + 1]$

    $\land$ IF $first\_committed[mon] = 0$
       THEN $first\_committed' = [first\_committed$ EXCEPT $![mon] = first\_committed[mon] + 1]$
       ELSE UNCHANGED $first\_committed$

    $\land\ monitor\_store' = [monitor\_store$ EXCEPT $![mon] = values[mon][last\_committed[mon] + 1]]$
    $\land\ new\_value' = [new\_value$ EXCEPT $![mon] = Nil]$
    $\land\ Send\_set(mon,$
       $\{[type \mapsto OP\_COMMIT,$
         $from \mapsto mon,$
         $dest \mapsto dest,$
         $last\_committed \mapsto last\_committed'[mon],$
         $pn \mapsto accepted\_pn[mon],$
         $values \mapsto values[mon]] : dest \in \{m \in Monitors \setminus \{mon\} : quorum[m]\}$
       $\})$
    $\land\ state' = [state$ EXCEPT $![mon] = STATE\_REFRESH]$
    $\land\ phase' = [phase$ EXCEPT $![mon] = PHASE\_COMMIT]$
    $\land$ UNCHANGED $\langle isLeader, values, accepted\_pn, pending\_proposal, accepted\rangle$
    $\land$ UNCHANGED $\langle epoch, quorum, quorum\_sz, restart\_vars, collect\_vars, lease\_vars\rangle$
```

Predicate that is called after *post_accept*. The leader finishes the commit phase by updating his state to *STATE_ACTIVE* and by extending the lease to his peers.
Variables changed: state, phase, *acked_lease*, messages, *message_history*.

```
$finish\_commit(mon) \triangleq$
```

$$
\begin{array}{l}
\quad \land state[mon]\ = STATE\_REFRESH \\
\quad \land phase[mon] = PHASE\_COMMIT \\
\quad \land finish\_round(mon) \\
\quad \land extend\_lease(mon) \\
\quad \land \text{UNCHANGED } \langle epoch, quorum, quorum\_sz, isLeader \rangle \\
\quad \land \text{UNCHANGED } \langle restart\_vars, data\_vars, collect\_vars, commit\_vars \rangle
\end{array}
$$

Handle a commit message. The monitor stores the values sent by the leader commit message.
Variables changed: messages, *message_history*, values, *first_committed*, *last_committed*, *monitor_store*, *uncommitted_v*, *uncommitted_pn*, *uncommitted_value*.

$$
\begin{array}{l}
handle\_commit(mon, msg) \triangleq \\
\quad \land isLeader[mon] = \text{FALSE} \\
\quad \land store\_state(mon, msg) \\
\quad \land check\_and\_correct\_uncommitted(mon) \\
\quad \land Discard(msg) \\
\quad \land \text{UNCHANGED } \langle epoch, quorum, quorum\_sz, accepted\_pn \rangle \\
\quad \land \text{UNCHANGED } \langle state\_vars, collect\_vars, lease\_vars, commit\_vars \rangle
\end{array}
$$

Client Request

Request a transaction v to the monitor. The transaction is saved on pending proposal to be committed in the next available commit phase.
Variables changed: *pending_proposal*.

$$
\begin{array}{l}
client\_request(mon, v) \triangleq \\
\quad \land isLeader[mon] = \text{TRUE} \\
\quad \land state[mon] = STATE\_ACTIVE \\
\quad \land pending\_proposal[mon] = Nil \\
\quad \land pending\_proposal' = [pending\_proposal \text{ EXCEPT } ![mon] = v] \\
\quad \land \text{UNCHANGED } \langle new\_value, accepted \rangle \\
\quad \land \text{UNCHANGED } \langle global\_vars, state\_vars, restart\_vars, data\_vars, collect\_vars, lease\_vars \rangle
\end{array}
$$

Start a commit phase with the value on pending proposal.
Variables changed: state, *pending_proposal*, accepted, *new_value*, phase, messages, *message_history*, values, *uncommitted_pn*, *uncommitted_v*, *uncommitted_value*.

$$
\begin{array}{l}
propose\_pending(mon) \triangleq \\
\quad \land phase[mon] = PHASE\_LEASE \lor phase[mon] = PHASE\_ELECTION \\
\quad \land state[mon]\ = STATE\_ACTIVE \\
\quad \land pending\_proposal[mon] \neq Nil \\
\quad \land pending\_proposal' = [pending\_proposal \text{ EXCEPT } ![mon] = Nil] \\
\quad \land state' = [state \text{ EXCEPT } ![mon] = STATE\_UPDATING] \\
\quad \land begin(mon, pending\_proposal[mon]) \\
\quad \land \text{UNCHANGED } \langle isLeader, monitor\_store, accepted\_pn, first\_committed, last\_committed \rangle \\
\quad \land \text{UNCHANGED } \langle epoch, quorum, quorum\_sz, collect\_vars, lease\_vars \rangle
\end{array}
$$

Collect phase predicates

Start collect phase. This first part of the collect phase is divided in two parts (collect and *send_collect*) in order to simplify variable changes (when collect is triggered from *handle_last*).

Variables changed: *accepted_pn*, phase.

```
$collect(mon, oldpn) \triangleq$
    $\land state[mon] = STATE\_RECOVERING$
    $\land isLeader[mon] = \text{TRUE}$
    $\land \text{LET } new\_pn \triangleq get\_new\_proposal\_number(mon, Max(\{oldpn, accepted\_pn[mon]\}))$
       $\text{IN} \quad \land accepted\_pn' = [accepted\_pn \text{ EXCEPT } ![mon] = new\_pn]$
    $\land phase' = [phase \text{ EXCEPT } ![mon] = PHASE\_SEND\_COLLECT]$
```

Continue the start of the collect phase. Initialize the number of peers that accepted the proposal (*num_last*) and the variables with peers version numbers. Check if there is an uncommitted value.

Send collect messages to the peers.

Variables changed: *peer_first_committed*, *peer_last_committed*, *uncommitted_pn*, *uncommitted_v*, *uncommitted_value*, *num_las*

messages, *message_history*, phase.

```
$send\_collect(mon) \triangleq$
    $\land state[mon] = STATE\_RECOVERING$
    $\land isLeader[mon] = \text{TRUE}$
    $\land phase[mon] = PHASE\_SEND\_COLLECT$
    $\land clear\_peer\_first\_committed(mon)$
    $\land clear\_peer\_last\_committed(mon)$

    $\land \text{IF } last\_committed[mon] + 1 \in \text{DOMAIN } values[mon]$
       $\text{THEN} \land uncommitted\_v' =$
                 $[uncommitted\_v \text{ EXCEPT } ![mon] = last\_committed[mon] + 1]$
             $\land uncommitted\_value' =$
                 $[uncommitted\_value \text{ EXCEPT } ![mon] = values[mon][last\_committed[mon] + 1]]$
             $\land uncommitted\_pn' = uncommitted\_pn$
       $\text{ELSE UNCHANGED } \langle restart\_vars \rangle$

    $\land num\_last' = [num\_last \text{ EXCEPT } ![mon] = 1]$
    $\land Send\_set(mon,$
       $\{[type \mapsto OP\_COLLECT,$
         $from \mapsto mon,$
         $dest \mapsto dest,$
         $first\_committed \mapsto first\_committed[mon],$
         $last\_committed \mapsto last\_committed[mon],$
         $pn \mapsto accepted\_pn[mon]] : dest \in \{m \in Monitors \setminus \{mon\} : quorum[m]\}$
       $\})$
    $\land phase' = [phase \text{ EXCEPT } ![mon] = PHASE\_COLLECT]$
    $\land \text{UNCHANGED } \langle isLeader, state \rangle$
    $\land \text{UNCHANGED } \langle epoch, quorum, quorum\_sz, data\_vars, lease\_vars, commit\_vars \rangle$
```

Handle a collect message. The peer will accept the proposal number from the leader if it is bigger than the last proposal number he accepted.

Variables changed: messages, *message_history*, epoch, state, *accepted_pn*

```
$handle\_collect(mon, msg) \triangleq$
    $\land isLeader[mon] = \text{FALSE}$
    $\land state' = [state \text{ EXCEPT } ![mon] = STATE\_RECOVERING]$
    $\land \lor \land msg.first\_committed > last\_committed[mon] + 1$
            $\land bootstrap$
            $\land Discard(msg)$
            $\land \text{UNCHANGED } \langle accepted\_pn \rangle$
       $\lor \land msg.first\_committed \leq last\_committed[mon] + 1$
            $\land \text{IF } msg.pn > accepted\_pn[mon]$
                $\text{THEN } accepted\_pn' = [accepted\_pn \text{ EXCEPT } ![mon] = msg.pn]$
                $\text{ELSE } \text{UNCHANGED } accepted\_pn$
            $\land Reply([type                \mapsto OP\_LAST,$
                       $from                \mapsto mon,$
                       $dest                \mapsto msg.from,$
                       $first\_committed     \mapsto first\_committed[mon],$
                       $last\_committed      \mapsto last\_committed[mon],$
                       $values              \mapsto values[mon],$
                       $uncommitted\_pn      \mapsto uncommitted\_pn[mon],$
                       $pn                  \mapsto accepted\_pn'[mon]], msg)$
            $\land \text{UNCHANGED } epoch$
    $\land \text{UNCHANGED } \langle isLeader, phase, values, first\_committed, last\_committed, monitor\_store \rangle$
    $\land \text{UNCHANGED } \langle quorum, quorum\_sz, restart\_vars, collect\_vars, lease\_vars, commit\_vars \rangle$
```

Handle a last message (response from a peer to the leader collect message).
The peers first and last committed version are stored. If the leader is behind, bootstraps. Stores any value that the peer may have committed (*store_state*). If peer is behind send commit message with leader values.
If peer accepted proposal number increase num last, if he sent a bigger proposal number start a new collect phase.
Variables changed: messages, *message_history*, epoch, phase, *uncommitted_pn*, *uncommitted_v*, *uncommitted_value*, *monitor_s*
accepted_pn, *first_committed*, *last_committed*, *num_last*, *peer_first_committed*, *peer_last_committed*.

```
$handle\_last(mon, msg) \triangleq$
    $\land isLeader[mon] = \text{TRUE}$

    $\land peer\_first\_committed' = [peer\_first\_committed \text{ EXCEPT } ![mon] =$
        $[peer\_first\_committed[mon] \text{ EXCEPT } ![msg.from] = msg.first\_committed]]$
    $\land peer\_last\_committed' = [peer\_last\_committed \text{ EXCEPT } ![mon] =$
        $[peer\_last\_committed[mon] \text{ EXCEPT } ![msg.from] = msg.last\_committed]]$

    $\land \text{IF } msg.first\_committed > last\_committed[mon] + 1$
        $\text{THEN}$
         $\land bootstrap$
         $\land Discard(msg)$
         $\land \text{UNCHANGED } \langle num\_last, accepted\_pn, values, phase, monitor\_store \rangle$
         $\land \text{UNCHANGED } \langle first\_committed, last\_committed, restart\_vars \rangle$
        $\text{ELSE}$
         $\land store\_state(mon, msg)$
         $\land \text{IF } \exists peer \in Monitors :$
```

```
      $\land peer \neq mon$
      $\land peer\_last\_committed'[mon][peer] \neq -1$
      $\land peer\_last\_committed'[mon][peer] + 1 < first\_committed[mon]$
      $\land first\_committed[mon] > 1$
THEN
 $\land bootstrap$
 $\land check\_and\_correct\_uncommitted(mon)$
 $\land Discard(msg)$
 $\land$ UNCHANGED $\langle phase,\ accepted\_pn,\ num\_last \rangle$
ELSE
 $\land$ LET $monitors\_behind \triangleq \{peer \in Monitors :$
             $\land peer \neq mon$
             $\land peer\_last\_committed'[mon][peer] \neq -1$
             $\land peer\_last\_committed'[mon][peer] < last\_committed[mon]$
             $\land quorum[peer]\}$
    IN   $Reply\_set(mon,$
            $\{[type \mapsto OP\_COMMIT,$
              $from \mapsto mon,$
              $dest \mapsto dest,$
              $last\_committed \mapsto last\_committed'[mon],$
              $pn \mapsto accepted\_pn[mon],$
              $values \mapsto values[mon]] : dest \in monitors\_behind$
            $\},\ msg)$
 $\land \lor \land msg.pn > accepted\_pn[mon]$
      $\land collect(mon,\ msg.pn)$
      $\land check\_and\_correct\_uncommitted(mon)$
      $\land$ UNCHANGED $num\_last$

    $\lor \land msg.pn = accepted\_pn[mon]$
      $\land num\_last' = [num\_last$ EXCEPT $![mon] = num\_last[mon] + 1]$
      $\land$ IF $\land msg.last\_committed + 1 \in$ DOMAIN $msg.values$
            $\land msg.last\_committed \geq last\_committed'[mon]$
            $\land msg.last\_committed + 1 \geq uncommitted\_v[mon]$
            $\land msg.uncommitted\_pn \geq uncommitted\_pn[mon]$
         THEN $\land uncommitted\_v' =$
                   $[uncommitted\_v$ EXCEPT $![mon] = msg.last\_committed + 1]$
              $\land uncommitted\_pn' =$
                   $[uncommitted\_pn$ EXCEPT $![mon] = msg.uncommitted\_pn]$
              $\land uncommitted\_value' =$
                   $[uncommitted\_value$ EXCEPT $![mon] = msg.values[msg.last\_committed + 1]]$
         ELSE $check\_and\_correct\_uncommitted(mon)$
      $\land$ UNCHANGED $\langle phase,\ accepted\_pn \rangle$

    $\lor \land msg.pn < accepted\_pn[mon]$
      $\land check\_and\_correct\_uncommitted(mon)$
```

```
            $\land$ UNCHANGED $\langle phase, accepted\_pn, num\_last \rangle$
        $\land$ UNCHANGED $epoch$
    $\land$ UNCHANGED $\langle epoch \rangle$

  $\land$ UNCHANGED $\langle quorum, quorum\_sz, isLeader, state \rangle$
  $\land$ UNCHANGED $\langle lease\_vars, commit\_vars \rangle$
```

Predicate that is enabled and called when all peers in quorum accept collect request from leader. If there is an uncommitted value, a commit phase is started with that value, else the leader changes to *ACTIVE_STATE* and extends the lease to his peers.

Variables changed: *peer_first_committed*, *peer_last_committed*, state, accepted, *new_value*, phase, messages, *message_history*, values, *uncommitted_pn*, *uncommitted_v*, *uncommitted_value*, *acked_lease*.

```
$post\_last(mon) \triangleq$
  $\land$ $isLeader[mon] = $ TRUE
  $\land$ $num\_last[mon] = quorum\_sz$
  $\land$ $phase[mon] = PHASE\_COLLECT$

  $\land$ $clear\_peer\_first\_committed(mon)$
  $\land$ $clear\_peer\_last\_committed(mon)$

  $\land$ IF $\land$ $uncommitted\_v[mon] = last\_committed[mon] + 1$
         $\land$ $uncommitted\_value[mon] \neq Nil$
      THEN $\land$ $state' = [state$ EXCEPT $![mon] = STATE\_UPDATING\_PREVIOUS]$
           $\land$ $begin(mon, uncommitted\_value[mon])$
           $\land$ UNCHANGED $\langle acked\_lease \rangle$
      ELSE $\land$ $finish\_round(mon)$
           $\land$ $extend\_lease(mon)$
           $\land$ UNCHANGED $\langle accepted, new\_value, values, restart\_vars \rangle$

  $\land$ UNCHANGED $\langle isLeader, monitor\_store, accepted\_pn, first\_committed, last\_committed \rangle$
  $\land$ UNCHANGED $\langle epoch, quorum, quorum\_sz, num\_last, pending\_proposal \rangle$
```

Leader election

Elect one monitor as a leader and initialize the remaining ones as peons.

Variables changed: *isLeader*, state, phase, *new_value*, *pending_proposal*, epoch.

```
$leader\_election \triangleq$
  $\land$ $\exists\, mon \in Monitors :$
      $\land$ $quorum[mon]$
      $\land$ $isLeader' = [m \in Monitors \mapsto$ IF $m = mon$ THEN TRUE ELSE FALSE$]$
      $\land$ $state' = [m \in Monitors \mapsto$
          IF $quorum\_sz = 1$ THEN $STATE\_ACTIVE$ ELSE $STATE\_RECOVERING]$
  $\land$ $phase' = [m \in Monitors \mapsto PHASE\_ELECTION]$
  $\land$ $new\_value' = [m \in Monitors \mapsto Nil]$
  $\land$ $pending\_proposal' = [m \in Monitors \mapsto Nil]$
  $\land$ $epoch' = epoch + 1$
```

```
    $\land\ messages' = [mon1 \in Monitors \mapsto [mon2 \in Monitors \mapsto \langle\rangle]]$
    $\land$ UNCHANGED $\langle quorum, quorum\_sz, accepted, message\_history\rangle$
    $\land$ UNCHANGED $\langle data\_vars, restart\_vars, collect\_vars, lease\_vars\rangle$
```

Start recovery phase if number of monitors in quorum is greater than 1.

Variables changed: *accepted_pn*, phase.

```
$election\_recover(mon) \triangleq$
    $\land\ quorum\_sz > 1$
    $\land\ phase[mon] = PHASE\_ELECTION$
    $\land\ collect(mon, 0)$
    $\land$ UNCHANGED $\langle isLeader, state, values, first\_committed, last\_committed, monitor\_store\rangle$
    $\land$ UNCHANGED $\langle global\_vars, restart\_vars, collect\_vars, lease\_vars, commit\_vars\rangle$
```

Timeouts and restart

```
$crash\_mon(mon) \triangleq$
    $\land\ quorum\_sz > (MonitorsLen \div 2) + 1$
    $\land\ quorum[mon] = $ TRUE
    $\land\ quorum' = [quorum$ EXCEPT $![mon] = $ FALSE$]$
    $\land\ quorum\_sz' = quorum\_sz - 1$
    $\land\ bootstrap$
     $\land\ number\_crashes' = number\_crashes + 1$
    $\land$ UNCHANGED $\langle messages, message\_history\rangle$
    $\land$ UNCHANGED $\langle state\_vars, restart\_vars, data\_vars, collect\_vars, lease\_vars, commit\_vars\rangle$

$restore\_mon(mon) \triangleq$
    $\land\ quorum[mon] = $ FALSE
    $\land\ quorum' = [quorum$ EXCEPT $![mon] = $ TRUE$]$
    $\land\ quorum\_sz' = quorum\_sz + 1$
    $\land\ bootstrap$
    $\land$ UNCHANGED $\langle messages, message\_history\rangle$
    $\land$ UNCHANGED $\langle state\_vars, restart\_vars, data\_vars, collect\_vars, lease\_vars, commit\_vars\rangle$
```

Monitor timeout (simulate the various timeouts that can occur). Triggers new elections.

Variables changed: epoch.

```
$Timeout(mon) \triangleq$
    $\land$   $bootstrap$
    $\land$   UNCHANGED $\langle messages, quorum, quorum\_sz, message\_history, state\_vars, restart\_vars,$
                         $data\_vars, collect\_vars, lease\_vars, commit\_vars\rangle$
```

Dispatchers and next statement

Handle a message.

```
$Receive(msg) \triangleq$
    $\land$   $\lor\ \land\ msg.type = OP\_COLLECT$
```

```
        $\land handle\_collect(msg.dest, msg)$
        $\land step\_name' =$ “receive collect”

      $\lor \land msg.type = OP\_LAST$
        $\land handle\_last(msg.dest, msg)$
        $\land step\_name' =$ “receive last”

      $\lor \land msg.type = OP\_LEASE$
        $\land handle\_lease(msg.dest, msg)$
        $\land step\_name' =$ “receive lease”

      $\lor \land msg.type = OP\_LEASE\_ACK$
        $\land handle\_lease\_ack(msg.dest, msg)$
        $\land step\_name' =$ “receive lease_ack”

      $\lor \land msg.type = OP\_BEGIN$
        $\land handle\_begin(msg.dest, msg)$
        $\land step\_name' =$ “receive begin”

      $\lor \land msg.type = OP\_ACCEPT$
        $\land handle\_accept(msg.dest, msg)$
        $\land step\_name' =$ “receive accept”

      $\lor \land msg.type = OP\_COMMIT$
        $\land handle\_commit(msg.dest, msg)$
        $\land step\_name' =$ “receive commit”

 Limit some variables to reduce search space.
$reduce\_search\_space \triangleq$
   $\land epoch \neq 8$
   $\land \lor \forall mon \in Monitors : last\_committed[mon] < 2$
       $\lor \forall mon2 \in Monitors: new\_value[mon2] = Nil$
   $\land \forall mon \in Monitors : accepted\_pn[mon] < 300$
    $\land number\_crashes \neq 4$

 State transitions.
$Next \triangleq$
   $\land reduce\_search\_space$
   $\land$ IF $epoch\%2 = 1$ THEN
      $\land leader\_election$
      $\land step\_name' =$ “election”
      $\land$ UNCHANGED $number\_crashes$
     ELSE
      $\lor \land \exists mon \in Monitors : election\_recover(mon)$
        $\land step\_name' =$ “election_recover”
        $\land$ UNCHANGED $number\_crashes$

      $\lor \land \exists mon \in Monitors : send\_collect(mon)$
    $\land\ step\_name' = \text{“send\_collect”}$
    $\land\ \text{UNCHANGED}\ number\_crashes$

$\lor$  $\land\ \exists\, mon \in Monitors : post\_last(mon)$
    $\land\ step\_name' = \text{“post\_last”}$
    $\land\ \text{UNCHANGED}\ number\_crashes$

$\lor$  $\land\ \exists\, mon \in Monitors : post\_lease\_ack(mon)$
    $\land\ step\_name' = \text{“post\_lease\_ack”}$
    $\land\ \text{UNCHANGED}\ number\_crashes$

$\lor$  $\land\ \exists\, mon \in Monitors : post\_accept(mon)$
    $\land\ step\_name' = \text{“post\_accept”}$
    $\land\ \text{UNCHANGED}\ number\_crashes$

$\lor$  $\land\ \exists\, mon \in Monitors : finish\_commit(mon)$
    $\land\ step\_name' = \text{“finish\_commit”}$
    $\land\ \text{UNCHANGED}\ number\_crashes$

$\lor$  $\land\ \exists\, mon \in Monitors : \exists\, v \in Value\_set : client\_request(mon, v)$
    $\land\ step\_name' = \text{“client\_request”}$
    $\land\ \text{UNCHANGED}\ number\_crashes$

$\lor$  $\land\ \exists\, mon \in Monitors : propose\_pending(mon)$
    $\land\ step\_name' = \text{“propose\_pending”}$
    $\land\ \text{UNCHANGED}\ number\_crashes$

$\lor$  $\land\ \exists\, mon1, mon2 \in Monitors :$
            $\land\ mon1 \neq mon2$
            $\land\ Len(messages[mon1][mon2]) > 0$
            $\land\ Receive(messages[mon1][mon2][1])$
    $\land\ \text{UNCHANGED}\ number\_crashes$

$\lor$  $\land\ \exists\, mon \in Monitors : crash\_mon(mon)$
    $\land\ step\_name' = \text{“crash\_mon”}$
    $\land\ \text{UNCHANGED}\ number\_crashes$

$\lor$  $\land\ \exists\, mon \in Monitors : restore\_mon(mon)$
    $\land\ step\_name' = \text{“restore\_mon”}$
    $\land\ \text{UNCHANGED}\ number\_crashes$

$\lor$  $\land\ \exists\, mon \in Monitors : Timeout(mon)$
    $\land\ step\_name' = \text{“timeout\_and\_restart”}$
    $\land\ \text{UNCHANGED}\ number\_crashes$
```

Safety invariants

If two monitors are in state active then their *monitor_store* must have the same value.

$$\begin{array}{l}
\mathit{same\_monitor\_store} \triangleq \forall\, \mathit{mon}1,\ \mathit{mon}2 \in \mathit{Monitors} : \\
\quad \mathit{state}[\mathit{mon}1] = \mathit{STATE\_ACTIVE} \land \mathit{state}[\mathit{mon}2] = \mathit{STATE\_ACTIVE} \\
\quad\ \Rightarrow \mathit{monitor\_store}[\mathit{mon}1] = \mathit{monitor\_store}[\mathit{mon}2]
\end{array}$$

$$\mathit{Inv} \triangleq \ \land\ \mathit{same\_monitor\_store}$$

Test/Debug invariants

Invariant used to search for a state where 'x' happens.

$$\mathit{Inv\_find\_state}(x) \triangleq \neg x$$

Invariant used to search for a behavior of diameter equal to 'size'.

$$\mathit{Inv\_diam}(\mathit{size}) \triangleq \mathit{TLCGet}(\text{“level”}) \neq \mathit{size} - 1$$

Invariants to test in model check

$$\begin{array}{ll}
\mathit{DEBUG\_Inv} \triangleq & \land\ \textsc{true} \\
& \land\ \mathit{Inv\_diam}(20)
\end{array}$$

Examples:

Find a behavior with a diameter of size 60.
$\mathit{Inv\_diam}(60)$

Find a behavior where two different monitors assume the role of a leader.

$$\begin{array}{l}
\mathit{Inv\_find\_state}(\\
\quad \exists\, \mathit{msg}1,\ \mathit{msg}2 \in \mathit{message\_history} : \\
\qquad \land\ \mathit{msg}1.\mathit{type} = \mathit{OP\_COLLECT} \land \mathit{msg}2.\mathit{type} = \mathit{OP\_COLLECT} \\
\qquad \land\ \mathit{msg}1.\mathit{from} \neq \mathit{msg}2.\mathit{from} \\
)
\end{array}$$

Find a state where a monitor crashed during the collect phase and fails to send a $\mathit{OP\_LAST}$ message.

$$\begin{array}{l}
\mathit{Inv\_find\_state}(\\
\quad \land\ \mathit{step\_name} = \text{“}\mathit{crash\ mon}\text{”} \\
\\
\quad \backslash * \ \textit{The system is in collect phase and no } \mathit{OP\_LAST} \textit{ message has been received.} \\
\quad \backslash * \ \mathit{isLeader}[\mathit{mon}] = \textsc{true} \textit{ assures that the leader was not the one that crashed.} \\
\quad \land\ \exists\, \mathit{mon} \in \mathit{Monitors} : \\
\qquad \land\ \mathit{isLeader}[\mathit{mon}] = \textsc{true} \\
\qquad \land\ \mathit{phase}[\mathit{mon}] = \mathit{PHASE\_COLLECT} \\
\qquad \land\ \mathit{num\_last}[\mathit{mon}] = 1 \\
\\
\quad \backslash * \ \textit{All the collect requests have been handled by the peers.} \\
\quad \land\ \forall\, \mathit{mon}1,\ \mathit{mon}2 \in \mathit{Monitors} : \\
\qquad \forall\, i \in 1 \mathrel{..} \mathit{Len}(\mathit{messages}[\mathit{mon}1][\mathit{mon}2]) : \mathit{messages}[\mathit{mon}1][\mathit{mon}2][i].\mathit{type} \neq \mathit{OP\_COLLECT} \\
\\
\quad \land\ \mathit{epoch} = 2 \\
)
\end{array}$$

Find a state where the leader crashes during the commit phase, failing to complete the commit.

$$\begin{array}{l}
\mathit{Inv\_find\_state}(\\
\quad \land\ \mathit{step\_name} = \text{“}\mathit{crash\ mon}\text{”} \\
\quad \land\ \exists\, \mathit{mon}1,\ \mathit{mon}2 \in \mathit{Monitors} :
\end{array}$$

```
        $\exists\, i \in 1 \,..\, Len(messages[mon1][mon2]) : messages[mon1][mon2][i].type = OP\_ACCEPT$
    $\land\, \forall\, mon \in Monitors :$
        $isLeader[mon] = \text{FALSE}$
    $\land\, epoch = 2$
)
```

Note: After finding a state, that complete state can be used as an initial state to analyze behaviors from there.

\ ∗ Modification History
\ ∗ Last modified *Wed Apr* 14 14:21:13 WEST 2021 by *afonsonf*
\ ∗ Created *Mon Jan* 11 16:15:26 WET 2021 by *afonsonf*