This is a specification for the Gemini Protocol. It describes the behavior of the submodules that make up Gemini's two Top of Rack (TOR) solution for reliability. The purpose of the specification is t find all possible failure scenarios in the algorithm by exploring reachable states simulating transitions in the submodules' state machines.

EXTENDS FiniteSets

VARIABLES

torA,

 $torB, \\ mux$

Which ToR the MUX cable itself is pointing to

LinkProber knows wether the TOR it's hosted in should be active or standby by listening to the active ToR's heartbeat that is sent to both ToR's. It knows this because this ICMP "heartbeat" has the name or MAC address of the active ToR.

In the specification, this is represented by the heartbeatSender variable since the standby ToR's heartbeat will be dropped and never listened to.

```
vars \triangleq \langle torA, torB, mux \rangle
T \stackrel{\triangle}{=} \{ \text{"torA"}, \text{"torB"} \}
 Link Prober (page 9)
LPStates ≜ { "LPActive", "LPStandby", "LPWait", "LPUnknown" }
 Link Stat (page 10)
LinkStates \stackrel{\triangle}{=} \{ \text{"LinkUp"}, \text{"LinkDown"} \}
 Mux State (page 12)
\mathit{MuxStates} \triangleq \{ \text{"MuxActive"}, \text{"MuxStandby"}, \text{"MuxWait"}, \text{"MuxUnknown"} \}
\begin{array}{c} \textit{MUX\_XCVRD\_} \; (\text{page 11}) \\ \textit{XCVRDStates} \; \stackrel{\triangle}{=} \; \left\{ \text{"switch"} \, , \; \text{"check"} \, , \; \text{"-"} \, \right\} \end{array}
ToR \triangleq
      [dead : BOOLEAN]
       name: T,
       xcvrd: XCVRDStates,
       heartbeat : { "on", "off'},
       heartbeatIn : SUBSET (T \cup \{ "noResponse" \}),
       link Prober: LP States,
       linkState : LinkStates,
       muxState: MuxStates
  "Goal" state for a ToR.
Active Tor \triangleq
     [dead : \{FALSE\},
      name: T, \\ xcvrd: \{\text{``-''}\},
```

```
heartbeat : \{ "on" \},
     heartbeatIn : SUBSET (T \cup \{ \text{"noResponse"} \}),
     linkProber: \{ "LPActive" \},
     linkState : { "LinkUp" },
muxState : { "MuxActive" }]
TypeOK \triangleq
     \land torA \in ToR
     \land \quad torB \in \mathit{ToR}
         mux \in [active: T, next: T]
Init \triangleq
    LET InitialTor(name) \stackrel{\Delta}{=}
           [dead]
                                \mapsto FALSE,
            name
                                \mapsto name,
            xcvrd
                                \mapsto "check".
            heartbeat
                                \mapsto "on",
            heartbeatIn
                                \mapsto {},
            linkProber
                                \mapsto "LPUnknown",
                                \mapsto "LinkDown",
            linkState
                               \mapsto "MuxWait"]
            muxState
            \land mux \in \{f \in [active : T, next : T] : f.active = f.next\}
    IN
            \wedge torA = InitialTor("torA")
            \land torB = InitialTor("torB")
```

State machine and transition table pages 12 & 13 of the Powerpoint presentation as of 08/25/2022 XCVRD daemon described on page 11 of the Powerpoint presentation as of 08/25/2022 https://microsoft-my.sharepoint.com/:u:/p/t-gegranados/ERThXZdF5MVFusk2rP-PF0cBGguDR3Rt9yJ3WxxwAt0hpg?e

Merged LinkWait and MuxWait on Powerpoint slide 13 into Wait

MuxState State Transitions depend on LinkManager's decisions and XCVRD responses when checking or switching the MuxCable's direction

$TRIGGER_LINKMANAGER_CHECK(t) \triangleq$

Sends check request to MUX via XCVRD. Transitions muxState to MuxWait. Defined to use in subsequent MuxState actions.

```
t' = [t \text{ EXCEPT } !.muxState = "MuxWait", !.xcvrd = "check"]
```

$TRIGGER_LINKMANAGER_SWITCH(t, target) \triangleq$

Sends write request to MUX via XCVRD. Transitions muxState to MuxWait and linkProber to LPWait target refers to ToR the Mux should point to. Defined to use in subsequent MuxState actions.

```
EXEC\_LINKMANAGER\_CHECK(t, otherTor) \stackrel{\Delta}{=}
     \land UNCHANGED \langle mux, otherTor \rangle
     \wedge t.xcvrd = \text{``check''}
     \land \lor \land mux.active = t.name
               \land t' = [t \text{ EXCEPT } !.muxState = \text{``MuxActive''}, !.heartbeat = \text{``on''}, !.xcvrd = \text{``-''}]
          \lor \land mux.active \neq t.name
              \land t' = [t \text{ EXCEPT } !.muxState = \text{``MuxStandby''}, !.heartbeat = \text{``on''}, !.xcvrd = \text{``-''}]
          \lor \land t' = [t \text{ EXCEPT } !.muxState = \text{"MuxUnknown"}, !.heartbeat = \text{"on"}, !.xcvrd = \text{"-"}]
           \lor \land t.muxState = \text{``MuxWait''}
               \wedge t' = [t \; \text{EXCEPT} \; !.muxState = \text{"MuxStandby"}, \; !.xcvrd = \text{"-"}] \; MUX\_XCVRD\_FAIL
EXEC\_LINKMANAGER\_SWITCH(t, otherTor) \triangleq
      Writing Switch direction
     \land UNCHANGED other Tor
     \wedge t.xcvrd =  "switch"
     \land t.muxState = \text{``MuxWait''}
     \land t.linkProber = "LPWait"
     \wedge mux' = [mux \ EXCEPT \ !.active = mux.next]
     \land \lor \land mux.next = t.name
               \land t' = [t \text{ EXCEPT } !.muxState = \text{"MuxActive"}, !.heartbeat = \text{"on"}, !.xcvrd = \text{"-"}]
          \lor \land mux.next \neq t.name
               \land t' = [t \text{ EXCEPT } !.muxState = \text{"MuxStandby"}, !.heartbeat = \text{"on"}, !.xcvrd = \text{"-"}]
          \lor \land t' = [t \text{ EXCEPT } !.muxState = \text{"MuxUnknown"}, !.heartbeat = \text{"on"}, !.xcvrd = \text{"-"}]
           \lor \land t.muxState = \text{``MuxWait''}
               \wedge t' = [t \; \text{EXCEPT} \; !.muxState = \text{"MuxStandby"}, \; !.xcvrd = \text{"-"}] \; MUX\_XCVRD\_FAIL
MuxStateActive(t, otherTor) \stackrel{\Delta}{=}
     \land t.muxState = "MuxActive"
     \land \lor \land t.linkState = \text{"LinkUp"}
                LinkUp MuxStateActive Row
               \land \quad \lor \quad \land \ t.linkProber \in \{ \text{ "LPStandby"}, \ \text{"LPUnknown"} \, \}
                        \land TRIGGER\_LINKMANAGER\_CHECK(t)
                   \lor \land t.linkProber = "LPWait"
                         Check and suspend heartbeat
                        \land t' = [t \text{ EXCEPT } !.muxState = \text{"MuxWait"}, !.xcvrd = \text{"check"}, !.heartbeat = \text{"off"}]
              \land UNCHANGED \langle mux, otherTor \rangle
          \lor \land t.linkState =  "LinkDown"
               Switch to Standby
               \land TRIGGER\_LINKMANAGER\_SWITCH(t, otherTor)
               \land UNCHANGED otherTor
MuxStateStandby(t, otherTor) \stackrel{\Delta}{=}
```

 $\land t.muxState = "MuxStandby"$

```
\land \lor \land t.linkState = \text{``LinkUp''}
          LinkUp\ MuxStateStandby\ Row
             \land \lor \land t.linkProber \in \{ \text{"LPActive"}, \text{"LPWait"} \}
                     \land TRIGGER\_LINKMANAGER\_CHECK(t)
                     \land UNCHANGED \langle mux, otherTor \rangle
                 \lor \land t.linkProber = \text{"LPUnknown"}
                      Switch to Active
                     \land TRIGGER\_LINKMANAGER\_SWITCH(t, t)
                     \land UNCHANGED otherTor
         \lor \land t.linkState = \text{``LinkDown''}
          LinkDown\ MuxStateStandby\ Row
             \land t.linkProber \in \{ \text{"LPUnknown, LPWait"} \}
             \wedge TRIGGER\_LINKMANAGER\_CHECK(t)
             \land UNCHANGED \langle mux, otherTor \rangle
MuxStateUnknown(t, otherTor) \stackrel{\Delta}{=}
     \land t.muxState = "MuxUnknown"
     \land \lor \land t.linkState = \text{``LinkUp''}
              LinkUp\ MuxStateStandby\ Row
             \land TRIGGER\_LINKMANAGER\_CHECK(t)
         \lor \land t.linkState = \text{``LinkDown''}
          LinkDown MuxStateStandby Row
             \land t.linkProber \in \{ \text{"LPUnknown, LPWait"} \}
             \land TRIGGER\_LINKMANAGER\_CHECK(t)
     \land UNCHANGED \langle mux, otherTor \rangle
MuxStateWait(t, otherTor) \stackrel{\Delta}{=}
      TODO Specify receiving XCVRD Response
     \land t.muxState = \text{``MuxWait''}
     \land \quad \lor \quad \land \ EXEC\_LINKMANAGER\_CHECK(t, \ otherTor)
         \vee \wedge EXEC\_LINKMANAGER\_SWITCH(t, otherTor)
MuxState(t, otherTor) \triangleq
     \wedge \neg t.dead
     \land \lor MuxStateActive(t, otherTor)
         \vee MuxStateStandby(t, otherTor)
         \vee MuxStateUnknown(t, otherTor)
         \vee MuxStateWait(t, otherTor)
 MuxXCVRD \stackrel{\triangle}{=}
    \*
```

 \land UNCHANGED $\langle torA, torB \rangle$

 $\land mux' = [mux \ EXCEPT \ !.active = mux.next]$

```
State machine page 10 of the Powerpoint presentation as of 08/25/2022
```

```
LinkState(t, otherTor) \stackrel{\Delta}{=}
     \wedge \neg t.dead
     \wedge t.linkState = "LinkDown" unnecessary, because going from LinkUp to LinkUp is just (finite) stuttering. However
     \land UNCHANGED \langle otherTor, mux \rangle
     \wedge t' = [t \text{ EXCEPT } !.linkState = \text{``LinkUp''}]
 State machine page 09 of the Powerpoint presentation as of 08/25/2022
 SendHeartbeat(t) \triangleq
     \wedge \neg t.dead
     \wedge t.linkState = "LinkUp"
     \land \ t.heartbeat = "on"
    Active ToR sends heartbeat to server. MUX duplicates packet and sends it to both ToR's
     \land mux.active = t.name The MUX will drop traffic from ToR if it is not pointing to it
     \land torA' = [torA \ EXCEPT \ !.heartbeatIn = @ \cup \{t.name\}]
     \land torB' = [torB \ EXCEPT \ !.heartbeatIn = @ \cup \{t.name\}]
     \land UNCHANGED \langle mux \rangle
LinkProberWait(t, otherTor) \stackrel{\Delta}{=}
     \land UNCHANGED \langle otherTor, mux \rangle
     \wedge t.linkState = "LinkUp"
     \land t.linkProber = "LPWait"
     \land \exists heartbeat \in t.heartbeatIn :
         \lor \land t.name = heartbeat
            \land t' = [t \text{ EXCEPT } !.linkProber = \text{"LPActive"}, !.heartbeatIn = @ \setminus \{heartbeat\}]
         \lor \land otherTor.name = heartbeat
            \land t' = [t \text{ EXCEPT } !.linkProber = \text{``LPStandby''}, !.heartbeatIn = @ \setminus \{heartbeat\}]
         \lor \land "noResponse" = heartbeat
            \land t' = [t \text{ EXCEPT } !.heartbeatIn = @ \setminus \{heartbeat\}]
LinkProberUnknown(t, otherTor) \stackrel{\Delta}{=}
     \land UNCHANGED \langle otherTor, mux \rangle
     \land t.linkState = "LinkUp"
     \land t.linkProber = "LPUnknown"
     \land \exists heartbeat \in t.heartbeatIn :
         \lor \land t.name = heartbeat
            \land t' = [t \text{ EXCEPT } !.linkProber = \text{"LPActive"}, !.heartbeatIn = @ \setminus \{heartbeat\}]
         \lor \land otherTor.name = heartbeat
            \land t' = [t \text{ EXCEPT } !.linkProber = \text{``LPStandby''}, !.heartbeatIn = @ \setminus \{heartbeat\}]
         \lor \land "noResponse" = heartbeat
            \wedge t' = [t \text{ EXCEPT } !.heartbeatIn = @ \setminus \{heartbeat\}]
```

```
LinkProberStandby(t, otherTor) \stackrel{\Delta}{=}
     \land UNCHANGED \langle otherTor, mux \rangle
     \land t.linkState = "LinkUp"
     \land \ t. \mathit{linkProber} = \text{``LPStandby''}
     \land \exists heartbeat \in t.heartbeatIn :
        \lor \land t.name = heartbeat
           \land t' = [t \text{ EXCEPT } !.linkProber = \text{"LPActive"}, !.heartbeatIn = @ \setminus \{heartbeat\}]
        \lor \land otherTor.name = heartbeat
           \wedge t' = [t \text{ EXCEPT } !.heartbeatIn = @ \setminus \{heartbeat\}]
        \lor \land "noResponse" = heartbeat
           \land t' = [t \text{ EXCEPT } !.linkProber = \text{``LPUnknown''}, !.heartbeatIn = @ \setminus \{heartbeat\}]
LinkProberActive(t, otherTor) \stackrel{\Delta}{=}
     \land UNCHANGED \langle otherTor, mux \rangle
     \wedge t.linkState = "LinkUp"
     \wedge t.linkProber =  "LPActive"
     \land \exists heartbeat \in t.heartbeatIn :
        \lor \land t.name = heartbeat
           \wedge t' = [t \text{ EXCEPT } !.heartbeatIn = @ \setminus \{heartbeat\}]
        \lor \ \land \ other Tor.name = heart beat
           \land t' = [t \text{ EXCEPT } !.linkProber = \text{``LPStandby''}, !.heartbeatIn = @ \setminus \{heartbeat\}]
        \lor \land "noResponse" = heartbeat
           \land t' = [t \text{ EXCEPT } !.linkProber = \text{"LPUnknown"}, !.heartbeatIn = @ \setminus \{heartbeat\}]
LinkProber(t, otherTor) \triangleq
     \wedge \neg t.dead
     \land \lor LinkProberActive(t, otherTor)
         \vee LinkProberStandby(t, otherTor)
         \vee LinkProberWait(t, otherTor)
         \vee LinkProberUnknown(t, otherTor)
System \stackrel{\triangle}{=}
    Mux handling a switch or check command.
     \vee EXEC\_LINKMANAGER\_SWITCH(torA, torB)
     \vee EXEC\_LINKMANAGER\_SWITCH(torB, torA)
     \vee EXEC\_LINKMANAGER\_CHECK(torA, torB)
     \vee EXEC_LINKMANAGER_CHECK(torB, torA)
    XCVRD and LinkMqrd
     \vee MuxState(torA, torB)
     \vee MuxState(torB, torA)
     ToR periodically send heartbeats via the mux to the server.
     \vee SendHeartbeat(torA)
     \vee SendHeartbeat(torB)
```

```
\vee LinkProber(torB, torA)
    Notification from the kernel that a physical link (L1) came up.
     \vee LinkState(torA, torB)
     \vee LinkState(torB, torA)
FailHeartbeat \triangleq
    Sender fails to send heartbeat to ToR's making them go into unknown state
     \land \lor \land \exists \ heartbeat \in SUBSET \ tor A. heartbeat In:
                \wedge torA' = [torA \ EXCEPT \ !.heartbeatIn = heartbeat]
                 \land UNCHANGED torB
        \lor \land \exists \ heartbeat \in SUBSET \ tor B. heartbeat In:
                \wedge torB' = [torB \ EXCEPT \ !.heartbeatIn = heartbeat]
                 \land UNCHANGED torA
     \wedge UNCHANGED mux
FailMux \triangleq
    Failure Action for inconsistent MUX States with MuxCable State
     \land mux' \in [active : T, next : T]
     \land UNCHANGED \langle torA, torB \rangle
FailTor(t, otherTor) \triangleq
     \land t' = [t \text{ EXCEPT } !.dead = \text{TRUE}]
     \land UNCHANGED \langle otherTor, mux \rangle
FailXCVRD(t, otherTor) \triangleq
     \land UNCHANGED \langle otherTor, mux \rangle
      According to Vaibhav Dahiya, the mux returns "Unknown" in case of failure,
      which, subsequently, causes the ToR to go to "Standby".
     \wedge t' = [t \text{ EXCEPT } !.xcvrd = "-"]
FailLinkState(t, otherTor) \triangleq
     \wedge \neg t.dead
     \land UNCHANGED \langle otherTor, mux \rangle
     \land t' = [t \text{ EXCEPT } !.linkState = \text{``LinkDown''}]
Environment \triangleq
     \vee FailMux
     \vee FailHeartbeat
     \vee FailTor(torA, torB)
     \vee FailTor(torB, torA)
     \vee FailXCVRD(torA, torB)
     \vee FailXCVRD(torB, torA)
```

ToR receives heartbeat and triggers appropriate transition in LinkProber

 $\vee LinkProber(torA, torB)$

```
\vee FailLinkState(torA, torB)
      \vee FailLinkState(torB, torA)
Next \triangleq
       \lor Environment
      \vee \, System
Spec \triangleq
     Init \wedge \Box [Next]_{vars} \wedge WF_{vars}(System)
 AtMostOneActive \stackrel{\Delta}{=}
     \Box\neg(torA \in ActiveTor \land torB \in ActiveTor)
RepeatedlyOneActive \; \stackrel{\triangle}{=} \;
     \Box \Diamond (\exists t \in \{torA, torB\} : \neg t.dead \Rightarrow (torA \in ActiveTor \lor torB \in ActiveTor))
Theorem Spec \Rightarrow
       \land \ AtMostOneActive
      \land \ Repeatedly One Active
Alias \triangleq
          torA \mapsto torA, \ torB \mapsto torB, \ mux \mapsto mux,
           active \mapsto \{t.name : t \in \{t \in \{torA, torB\} : t \in ActiveTor\}\}
```