VSI Discovery and Configuration Protocol (VDP)
(Working Draft)

Proposed Resolutions to subset of comments against
802.1Qbg Draft-1

Chait Tumuluri (Emulex), Jeffrey Lynch (IBM), Vijoy Pandey (BNT), Rakesh Sharma (IBM), Renato Recio (IBM), Srikanth Kilaru (Juniper)

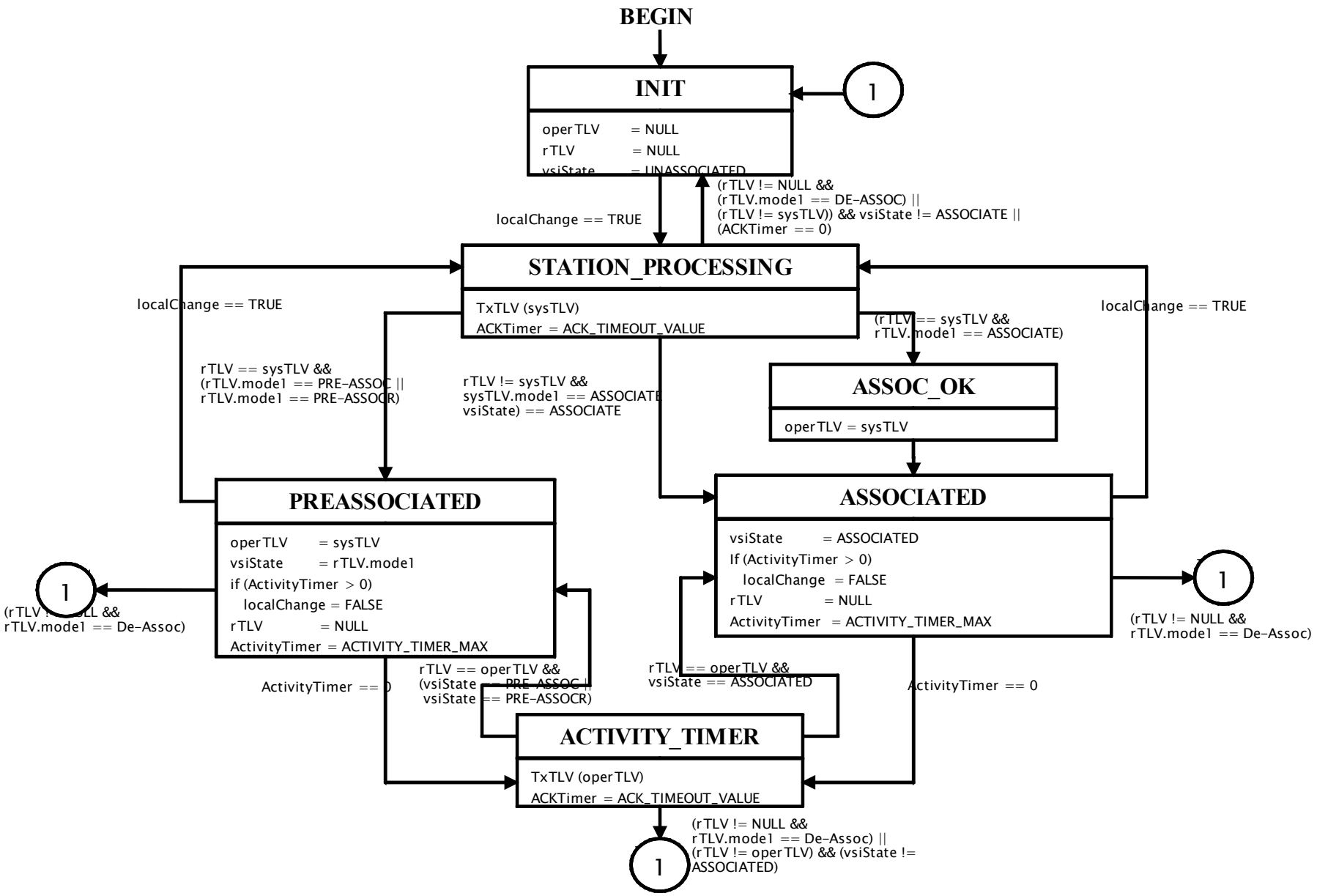
Version 05d

09/15/2010

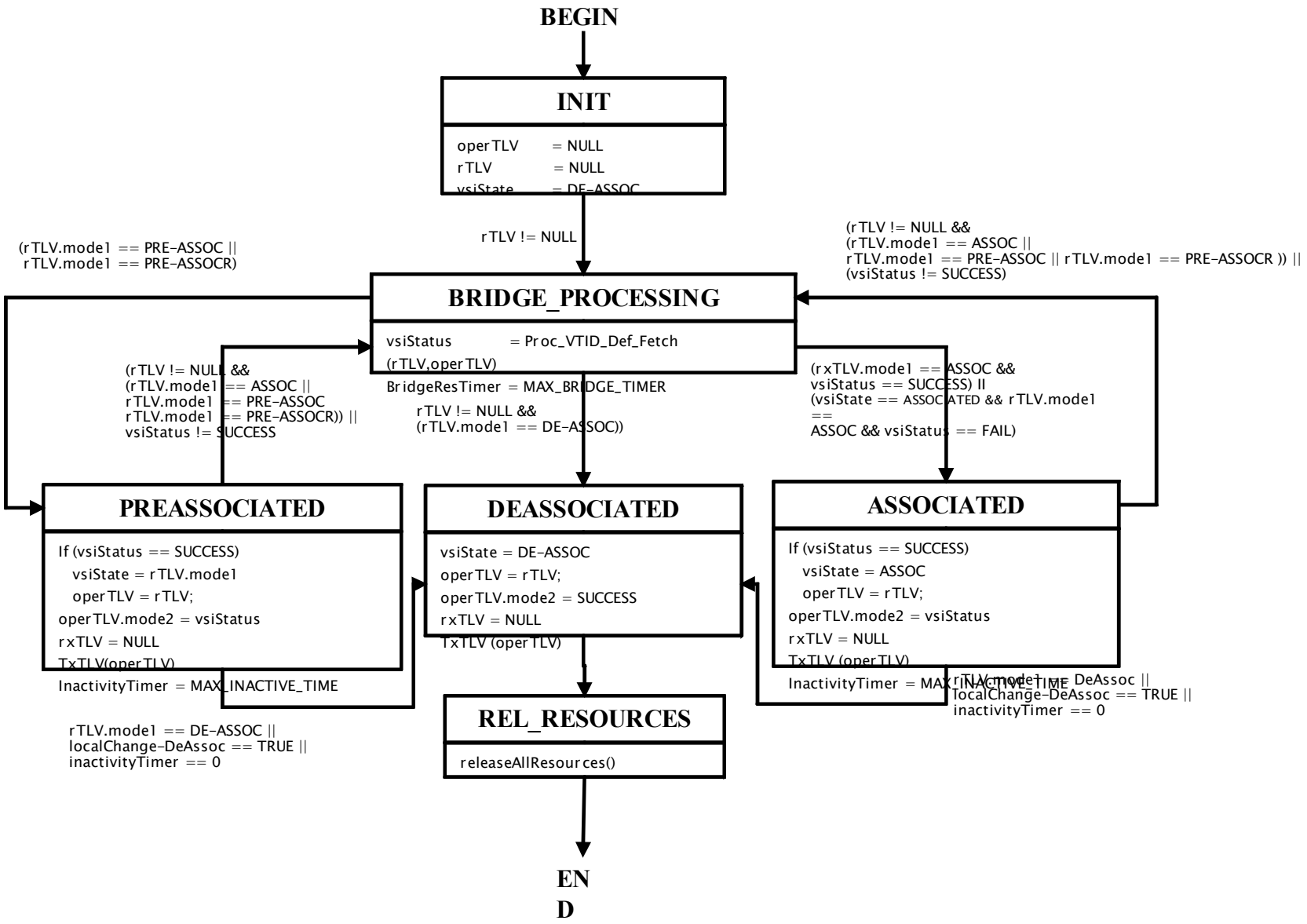
Summary of changes

1. Compliance to IEEE Protocols State Machines Conventions (Ref: IEEE 802.1AB x.xx)
2. Changed timer expiration events to comply with IEEE state machines style.
3. Changed state machine layout to match IEEE state machines.
4. Misc. other style changes in state machines to make them consistent with IEEE state machines.
5. Updated local variables and procedures names and description for additional clarification.
6. ***Ensured that changes do not impact state machines operation other than style and layout.***

Station State Machine (One Instance per VSI)



VSI State Machine – Bridge (Draft)
(One Instance per VSI)



VSI State Machine Local Variables and Procedures (1)

1. **vsiState**:
   - Local variable for current state of the station or bridge.
   - Possible states are: PRE-ASSOC: 0, PRE-ASSOCR (with resources): 1, ASSOC: 2 and DE-ASSOC: 3
2. **operTLV:**
   - Current local (active) TLV (configuration) of VSI at stattion or bridge.
3. **sysTLV:**
   - Command from station management system ("system") e.g. Hypervisor.
   - In addition localChange variable must be set to TRUE. If localChange is TRUE or ActivityTimer has expired, sysTLV can not change (resource lock semantics).
4. **localChange:**
   - Boolean variable shared between Hypervisor and VDP state machine. localChange must be set to TRUE and sysTLV is initialized when VSI and VDP state machines are created by Hypervisor or system.
5. **rTLV:**
   - TLV received from remote (station or bridge) from ECP transport. VDP listener receives TLVs from ECP, unpacks them and places the rTLV variables of correct VSI state machines. This operation requires rTLV variable to be NULL.
   - Bridge listener will instantiate a new bridge state machine if one does not exist for VSI instance.
6. **TxTLV(vsiTLV):**
   - Transmits TLV using TLV transport (ECP) service interfaces. Includes support for aggregation of VSI TLVs.

VSI State Machine Local Variables and Procedures (2)

1. **Proc_VSI_Def_Fetch ():**
   - Bridge procedure to fetch VSI definition based on Associate and/or Pre-Associate parameters based on received Remote TLV parameters.
   - Validates parameters
   - Sets vsiStatus in case of error.
   - This function handles PreAssociate with and without resource reservation case as well.
2. **ACKtimer**:
   - ACKtimer is ACK Timer local variable. Response (ACK or NACK) is expected before timer expires.
3. **ACK_TIMEOUT_VALUE**:
   - This local variable contains ACK timeout value for the machine.

VSI State Machine Local Variables and Procedures (3)

1. **INACTIVITY_TIMER**
   - Bridge timer local variable for inactivity timer.
2. **MAX_INACTIVE_TIME**
   - Max inactivity time on the bridge. It is set by bridge based on configured value and complies to range specified in 802.1Qbg.
3. **BridgeResTimer**
   - Bridge resource fetch timer local timer.
4. **localChange-DeAssoc**
   - Local variable for de-associating VSI. It is set to TRUE by bridge management when VSI to be disconnected.
5. **MAX_BRIDGE_TIMER**
   - Local variable that stores max bridge timer value. It is set by bridge based on configured value and complies to range specified in 802.1Qbg.
6. **vsiStatus = ProcRxAndSetCfg(rTLV,operTLV,sysTLV)**
   - Validate received TLV, set configuration and return vsiStatus

Pre-Associate Response Processing (Station)

- If Pre-Associate or Pre-Associate with Resource Response is success
  - operTLV is updated to same values as sysTLV
  - vsiState is set to PREASSOCIATED/PREASSOCIATED-R
- If Pre-Associate or Pre-Associate with Resource response has an error,
  - system command is failed and VDP goes to init state.
- If Pre-Associate Response is unexpected i.e. does not match sent TLV
  - system command failed and VDP goes to init state.
- When Station is in Pre-Associated state and a TLV is received
  - This TLV is unexpected
  - TLV is dropped and station remains Pre-Associated without any change to configuration.

Associate Response Processing (Station)

- If Associate Response is success
  - operTLV is updated to same values as sysTLV
  - vsiState is set to ASSOCIATED
- If Associate response has an error,
  - System command is failed
  - If Station is already in ASSOCIATED state
    - Station remains in ASSOCIATED state and configuration does not change.
- If Associate Response is unexpected i.e. does not match sent TLV
  - Handled same as ASSOCIATE error case (see above)

Pre-Associate/Pre-AssociateR Processing (Bridge)

- If Pre-Associate or Pre-Associate with Resource is received.
  - Pre-Assoc parameters are validated.
  - If Pre-AssocR is received, VTID definition is fetched and resources allocated
  - If no errors
    - vsiState is set to PRE-ASSOC or PRE-ASSOCR based on type of pre-associate received.
    - Pre-associated response is sent with SUCCESS in mode2 field.
- If Pre-Associate or Pre-AssociateR validation or resource fetch fails
  - system command is failed and VDP goes to init state.
- If Pre-Associate Response is unexpected i.e. does not match sent TLV
  - system command failed and VDP goes to init state.
- When Station is in Pre-Associated state and a TLV is received
  - This TLV is unexpected
  - TLV is dropped and station remains Pre-Associated without any change to configuration.

Associate Response Processing (Bridge)

- If Associate is success (parameters are valid and resources are available)
  - operTLV is updated to same values as sysTLV
  - vsiState is set to ASSOCIATED
- If Associate has an error,
  - Associate response with error code is sent.
  - If Station is already in ASSOCIATED state
    - Station remains in ASSOCIATED state
    - Configuration does not change.

Backup

VDP Objectives

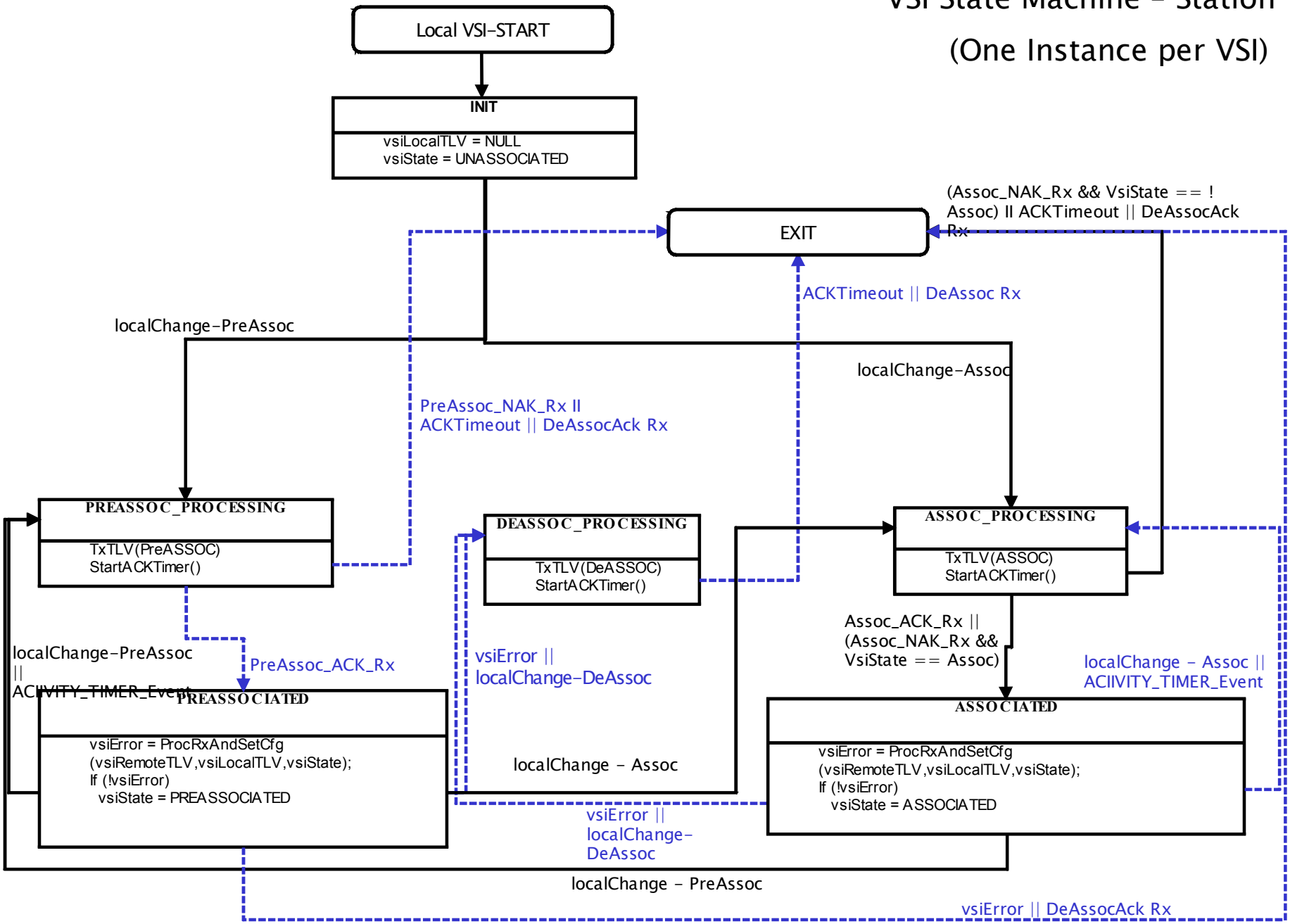
1. Support VSI preAssociate (with and without resource reservations), Associate and deAssociate.
2. ASSOCIATE, PreAssociate and DeAssociate are Idempotent i.e. can be repeated.
3. Capability to Associate skipping PreAssociate.
4. VDP will work both for VEPA and VEB environments.
5. Timeout mechanism to ensure:
    a. Bridge resources are not reserved too long for inactive VSIs (lease semantics)
    b. Allow removing resources from inactive VSIs with the goal of
        a. Conserve bridges resources (Number VSIs being handled by bridge can be large).
        b. Prevent inactive or VMs in error state to continue to hold resources.
6. Timeout out values to be negotiated on per channel between station and bridge. One timeout used for all ULPs on the channel negotiated using EVB TLV.

VDP Objectives (continued)

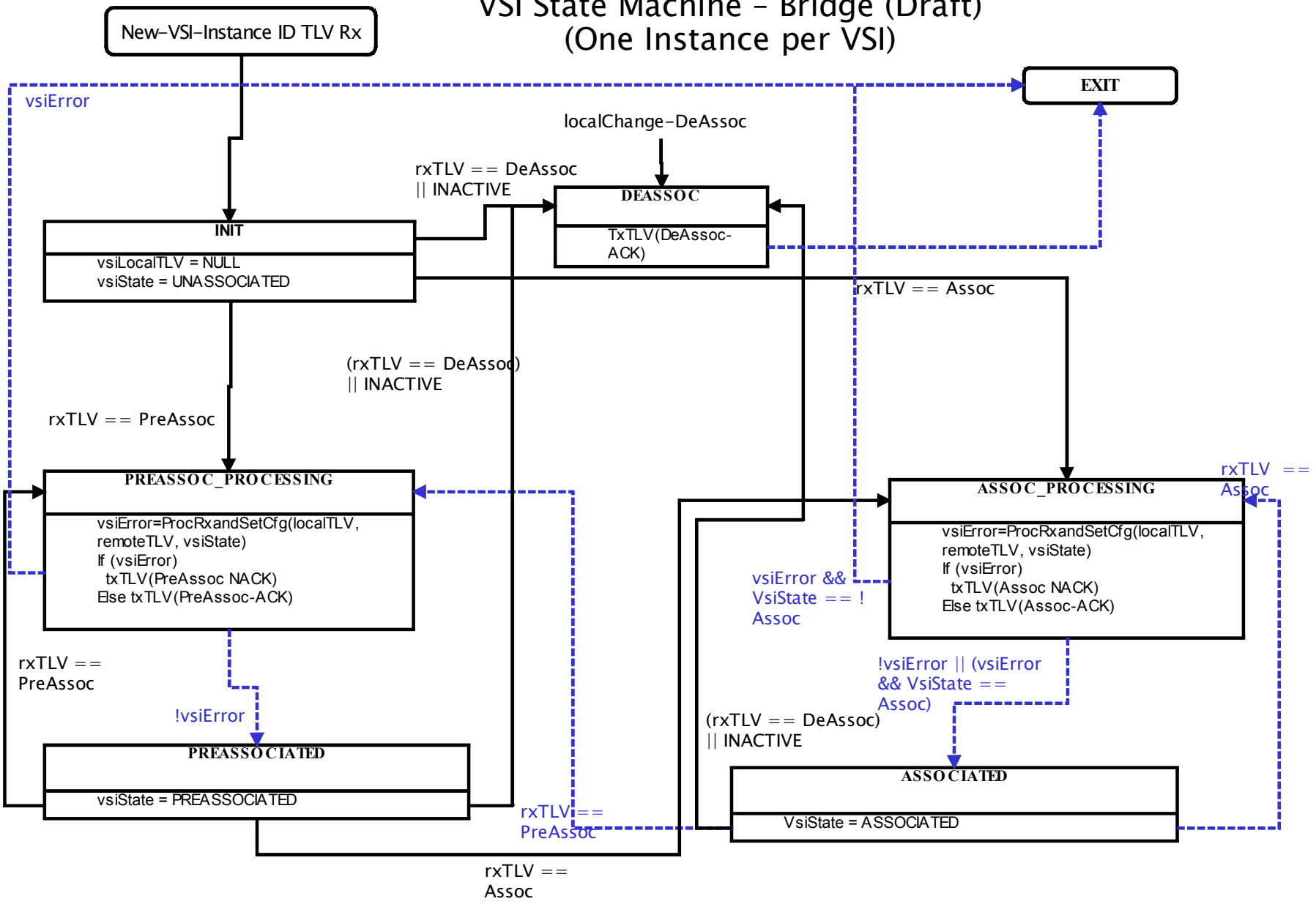
7. Manageability and Robustness
   a. Ensure VSI state and configuration between the Station and the Bridge remains consistent.
   b. Hard errors at the Bridge or the Hypervisor that can impact individual VSI or Hypervisor/Bridge as a whole. All VSI configuration goes away in this situation.
   c. Bridge and Station Errors are detected through one or more of the following mechanisms.
      - VSI KEEP-ALIVE (periodic transmission of VSI TLV from station and response from Bridge)
      - ACK Timer
      - Transport (ECP and LLDP) status indications.
   d. Bridge and Hypervisor administrative capability to force VSI deAssociate.
   e. ASSOCIATED VSI remains associated with existing configuration if a new associate is not successfully completed.
   f. Statistics and logging support (need specific proposal).

VSI State Machine – Station

(One Instance per VSI)



VSI State Machine – Bridge (Draft)
(One Instance per VSI)



VSI State Machine – Bridge (Draft)
(One Instance per VSI)