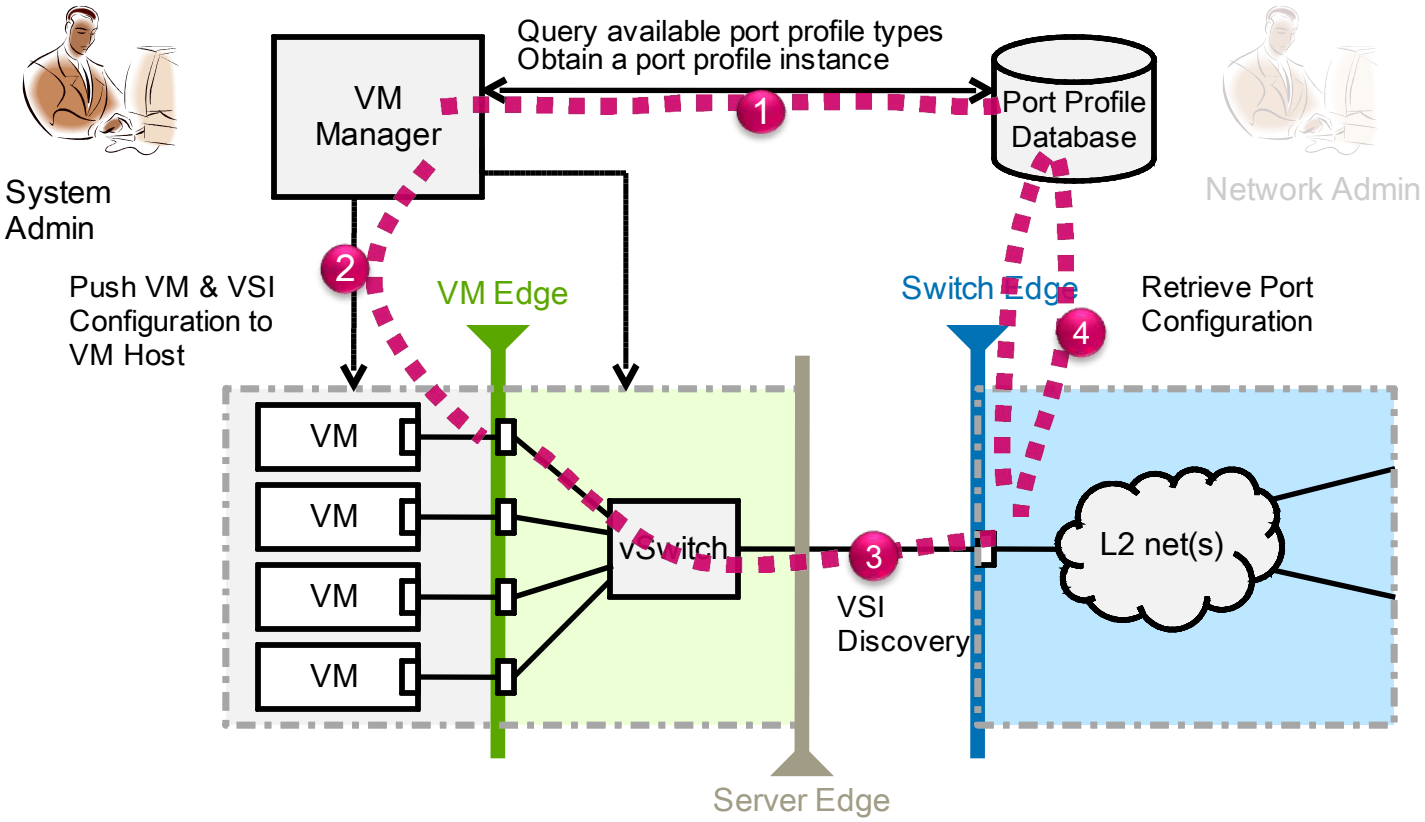
VSI Discovery Protocol using T3P/T3P-R

Draft v02

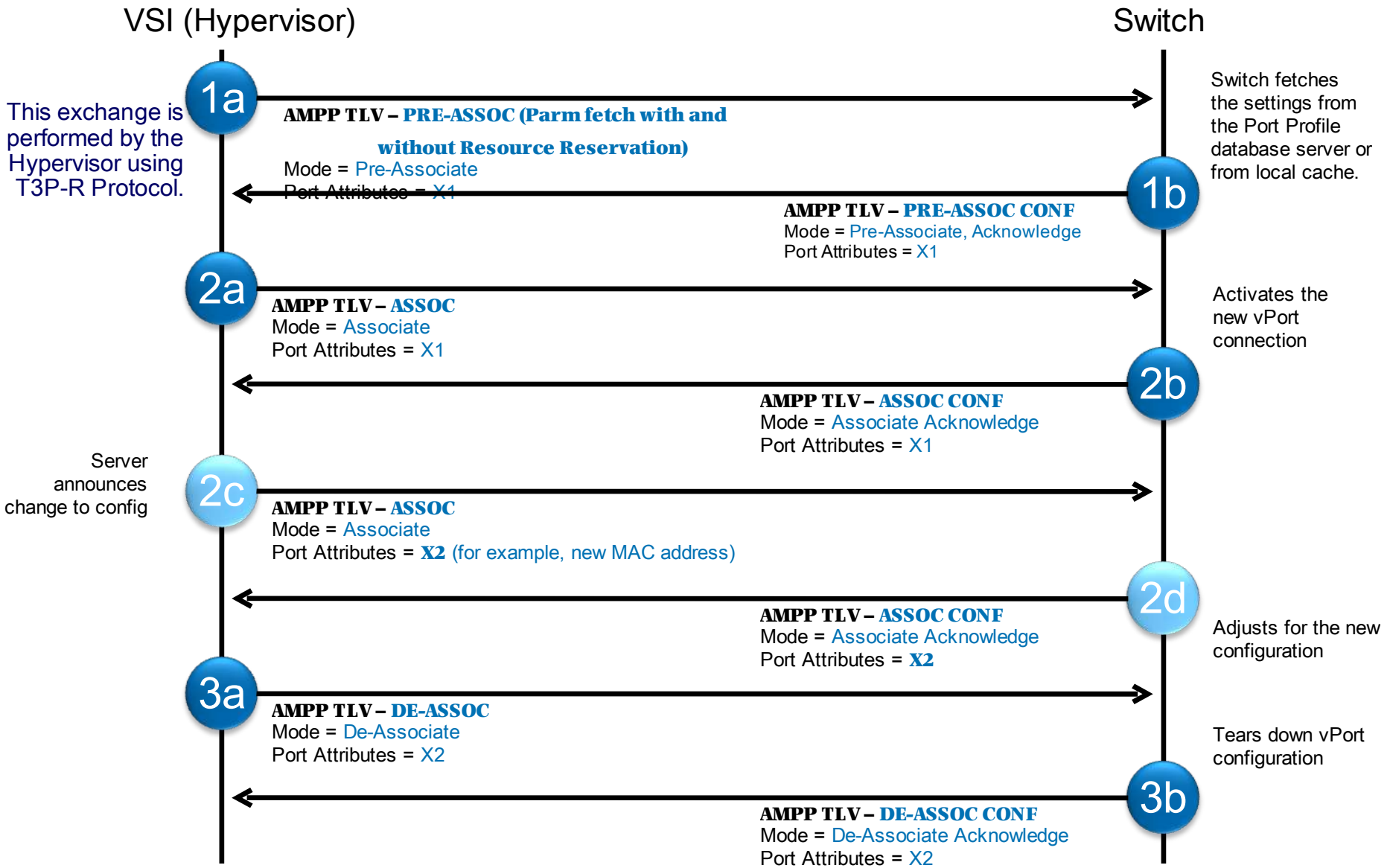
December 2nd, 2009

Authors: Rakesh Sharma, Anoop Ghanwani, Srikanth Kilaru, Manoj Wadekar, Vijoy Pandey, Chuck Hudson et. al.

Steps for Configuring Edge Connections (vPorts)



VSI Discovery TLV Example

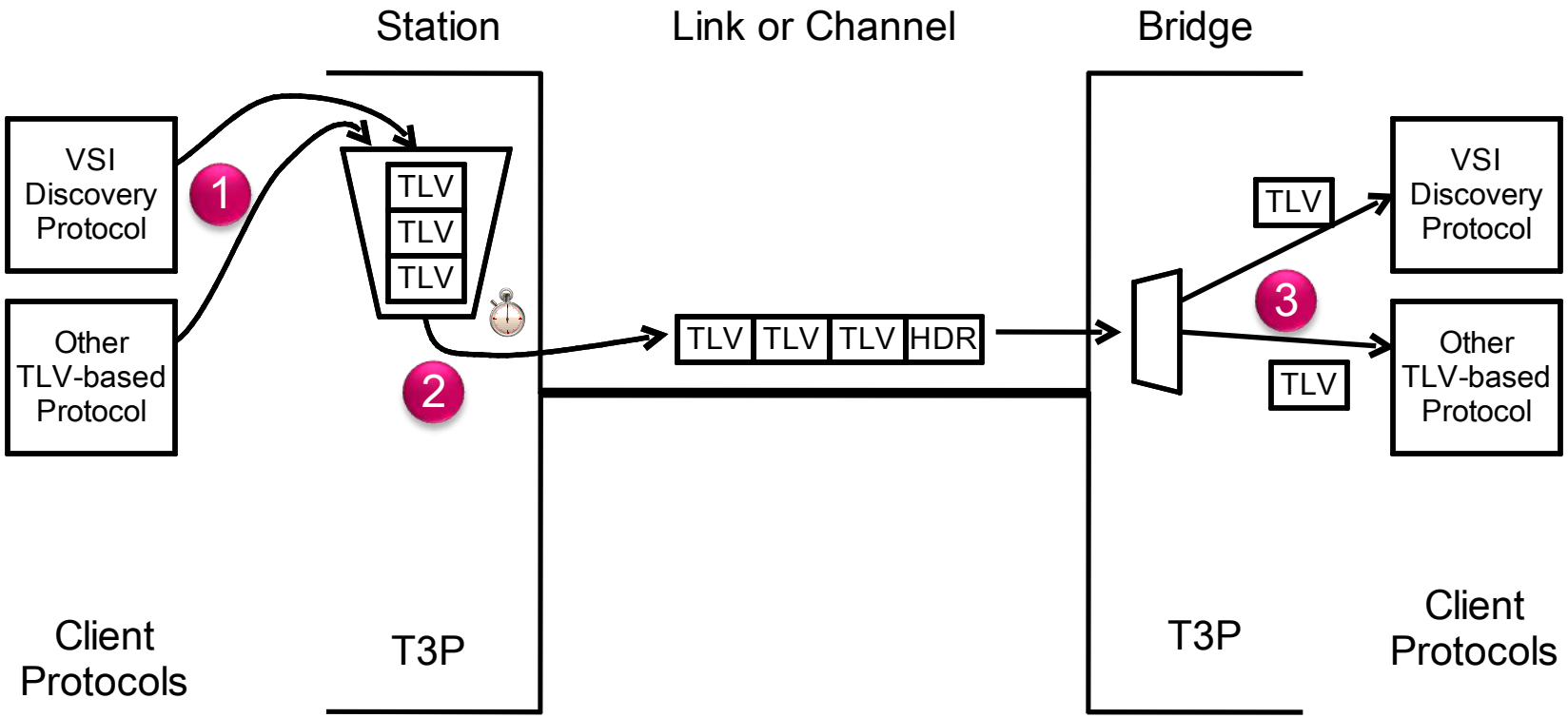


VSI Discovery Protocol needs a Trivial TLV Transport Protocol (T3P-R)

LLDP	VDP/T3P-R
Requires ALL TLVs in every LLDP exchange. Limited to 1500 bytes, no fragmentation.	Assumes subsets of TLVs can be passed in a frame.
Assumes that information is exchanged (no state maintained).	Assumes that TLV exchanges can be stateful and that state can be maintained for each TLV type.
Delivery is confirmed by regularly retransmitting (and fast retransmit) the full set of information.	Successful delivery is confirmed by other approaches: •**Optional** ACK support in T3P (T3P-R). T3P-R can enable simpler and more scalable VDP (VSI Discovery Protocol) and others •TLV-specific ACKs/NACKs may not be required •Transmission of a signature of the last TLV sent or of current state may not be required.
Requires repetition of information that is not changing.	Only transmits TLVs that are new or changed.
Assumes symmetry	Assumes edge environment with asymmetric ownership of driving state (physical stations driving identification of VSIs to bridge).

Trivial TLV Transport
(single direction shown)

Notes:
Reliability done in client protocol
Single writer, multiple readers
Sent/received in the same order as sent.



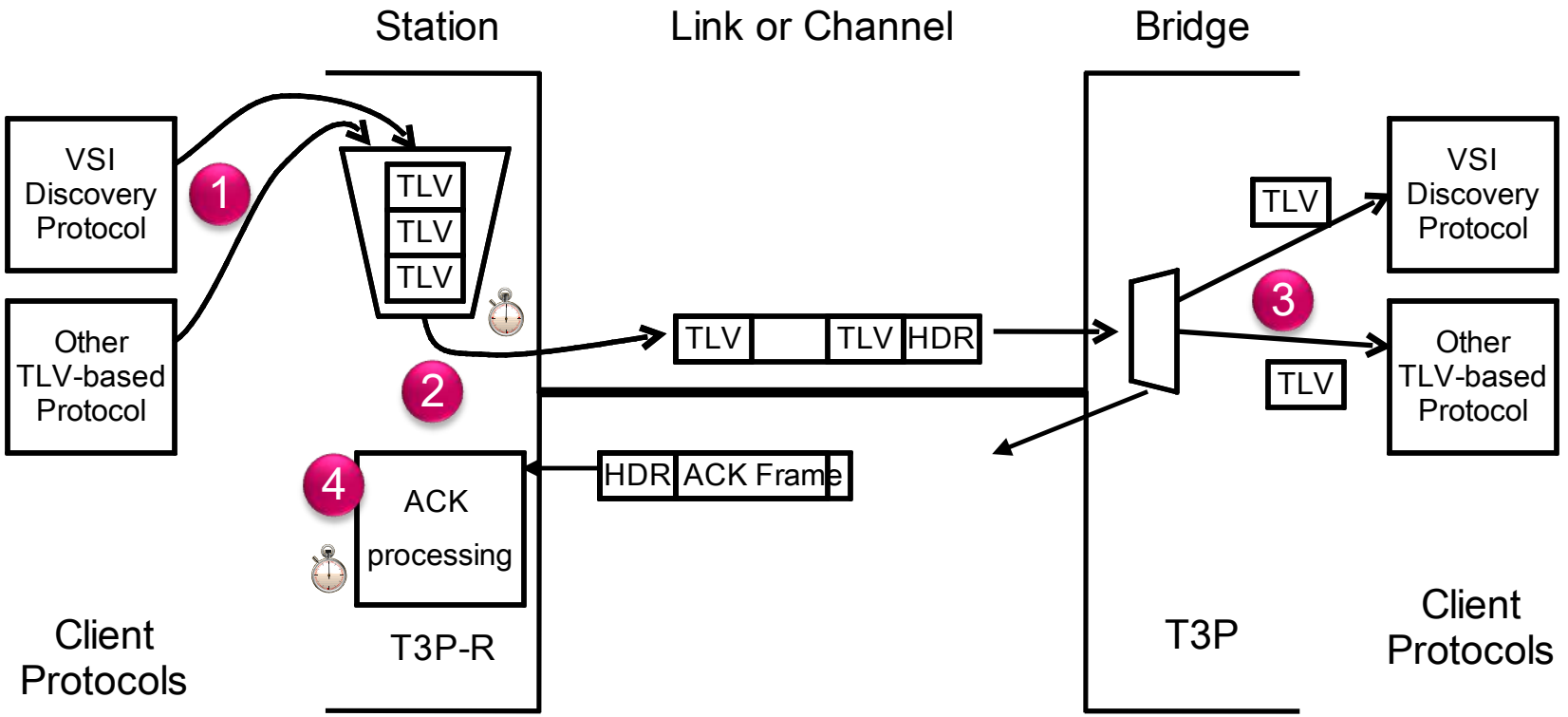
1. Client protocols pass outgoing TLVs to T3P (indicating whether the TLV is 'queued' or 'immediate').
   First TLV in queue sets queue timer.

2. If ((queue > 1500) |
   (immediate TLV) |
   (timer expired)) Then
   Send T3P frames until queue emptied

3. Client protocols subscribe* to specific TLVs, so the appropriate TLVs are forwarded to the correct client protocol. (*single sender, but multiple listeners allowed)

Trivial TLV Transport - R
(single direction shown)

Notes:
Reliability using ACK
Single writer, multiple readers
Sent/received in the same order as sent.



1. Client protocols pass outgoing TLVs to T3P (indicating whether the TLV is 'queued' or 'immediate') Reliability option can be set. First in queue sets timer.

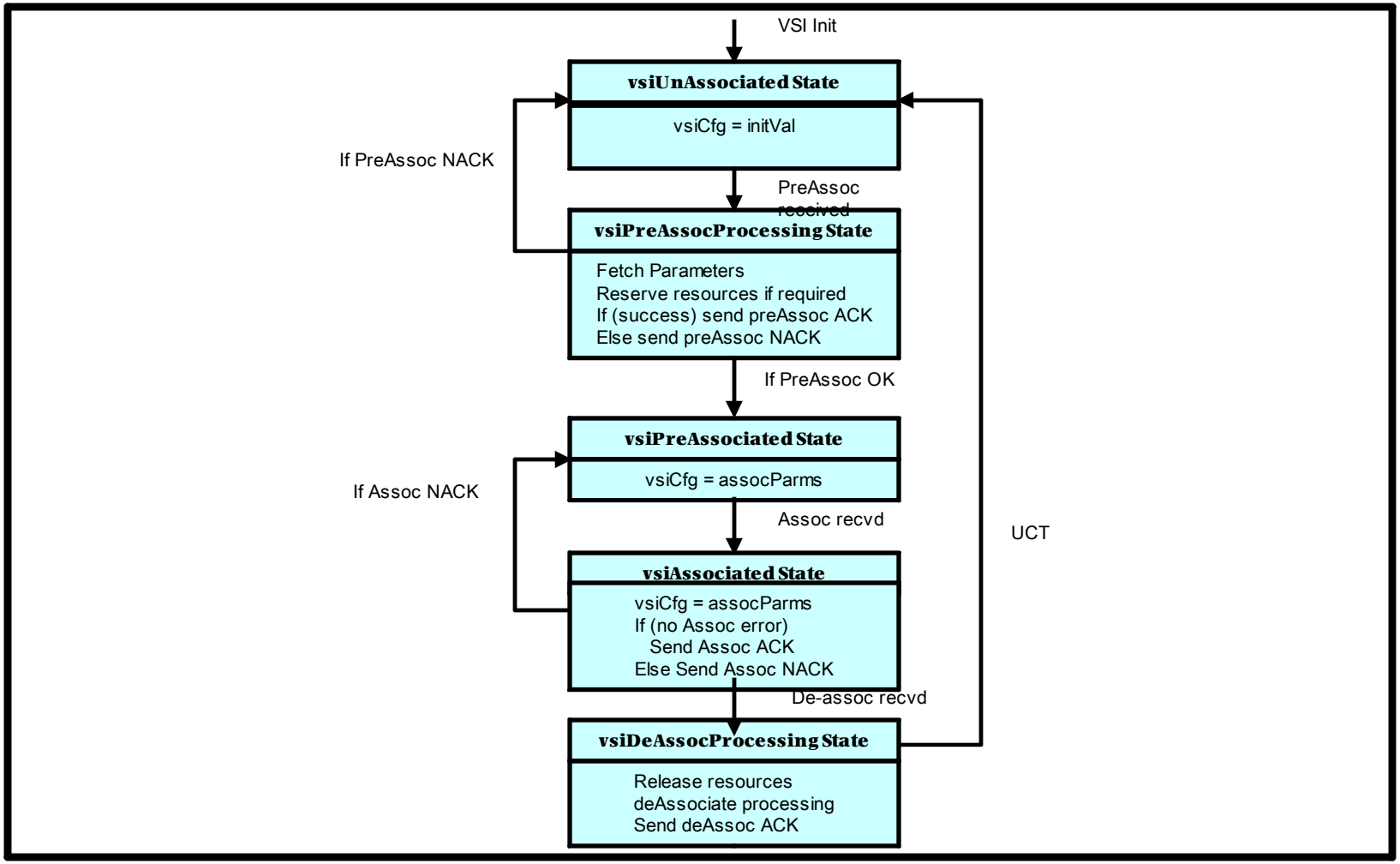
2. If ((queue > 1500) | (immediate TLV) | (timer expired)) Then Send T3P frames until queue emptied

3. Client protocols subscribe to specific TLVs, so the appropriate TLVs are forwarded to the correct client protocol

4. ACK Processing to Provide reliability. If ACK Timer expires, re-xmit. Only 1 TLV block tx at time.

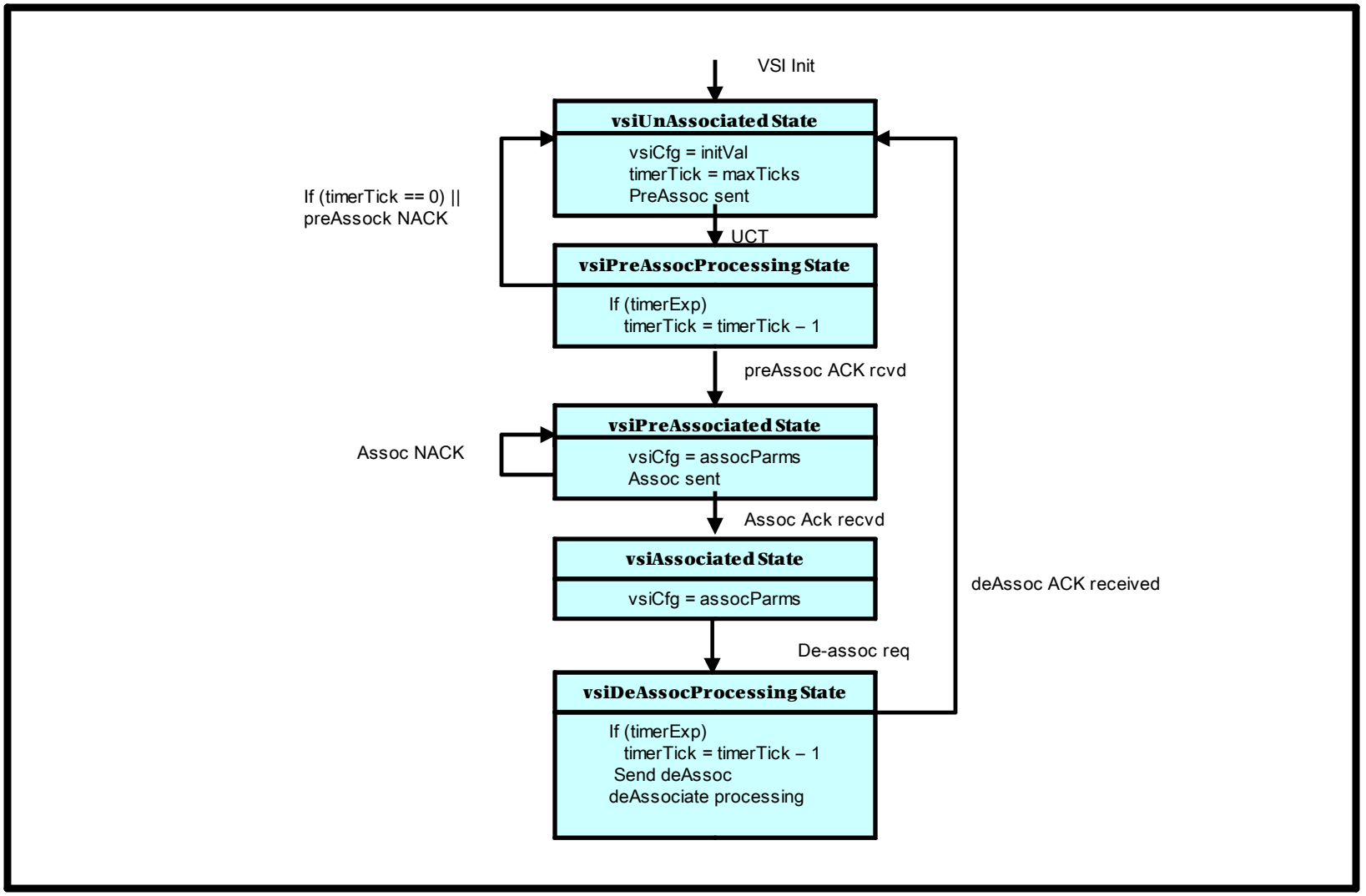
VSI State Machine – Adjacent Bridge

(One Instance per VSI)



VSI State Machine – Hypervisor

(One Instance per VSI)

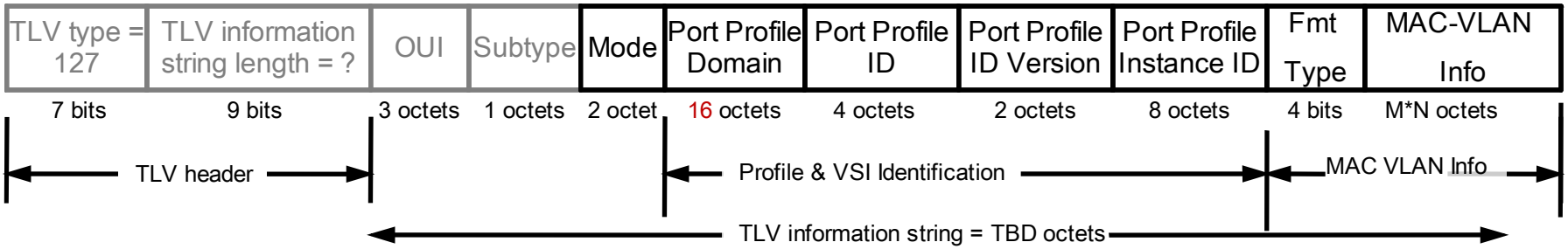


VSI Discovery and Configuration Protocol (VDP) Module

VSI	VSI TLV Fields	VSI state	VSI Timer Ticks	VSI statistics
VSI0	VSI0 TLV fields	VSI0 state	timerTicks	TBD
VSI1	VSI1 TLV fields	VSI1 state	timerTicks	TBD
VSIn	VSIn TLV fields	VSIn state	timerTicks	TBD

- VDP module implementation can be single module that supports all VSIs that
  - Contains table of VSI variables
  - timerTicks updated by single timer with long duration (for example, every 10 or more seconds
  - VDP module get requests (create VSI, pre-Assoc, Assoc and de-Assoc) from VDP user (Hypervisor/Bridge OS) and sends events to the user.
  - Transmits TLVs by queueing to T3P module and receives TLVs from T3P for processing.

Proposed EVB Port Discovery TLV (AMPP)



- Mode – First octet Indicates whether the TLV is a pre-associate (with or without resource reservation), associate, de-associate, or the corresponding confirmation or rejection for each.

  The second octet is used during a rejection to indicate the reason for the pre-assoc or assoc rejection.

- Port Profile Domain – Identifies the Port Profile Database that holds the detailed port profile definition. (Could be implemented as a simple integer or as the IP address of the database server).

- Port Profile ID – The integer identifier of the port profile type. The station and switch environments and their common understanding of the meaning of a port profile ID is outside the scope of this effort. A single port profile ID may be used to describe the port profile configuration of multiple ports.

- Port Profile ID Version (optional?) – The version of the port profile currently expected/desired. Selected pattern could mean “latest version”

- Port Profile Instance ID – A globally unique ID for the connection instance.

- Fmt – MAC-VLAN Information Format ID. Several formats are possible. Additional can be defined as needed.

- MAC-VLAN Information – The specific MAC addresses and VLANs associated with the vPort/VSI being discovered.

- Assumes that the VLAN ID of the port being discovered is in a VLAN TLV in the same LLDP packet or else that it can be obtained from the Port Profile Database. MAC-VLAN info the TLV, overrides that.

Backup

T3P-R is different than LLDP – but following should stay the same

- TLV format
- Addressing
- Strong push to minimize traffic

VSI (vPort) Discovery Protocols Overview

- VSI (vPort) Discovery (AMPP) Protocol (VDP)
  - Announce the arrival status of a VSI connection and communicate information to allow the edge switch to retrieve the appropriate configuration for the connection.
  - VSI protocol uses LLDP TLVs and T3P-R
- VSI transport protocol uses T3P-R attributes as follows
  - Registers VSI TLV type to Receive with T3P-R
  - TLV delivery reliability using ACK
  - Uses TLV exchange for capabilities advertisement and configure peers.
- VSI Protocol Stacks are used by Hypervisor and Adjacent Bridge as peer users.
  - Possible realizations examples are given below (others are possible)
    - EVB Discovery, VSI Discovery and LLDP transport in hypervisor+virtual switch (VMware, KVM, PowerVM etc.) over physical NIC and in Adjacent Bridge OS.
    - EVB Discovery, VSI Discovery and LLDP transport in hypervisor (VMware, KVM, PowerVM etc.) + VEB in SR-IOV NIC and in Adjacent Bridge OS.