

VSI Discovery Protocol Options

v04

December 8, 2009

Caitlin Bestler, Aprius

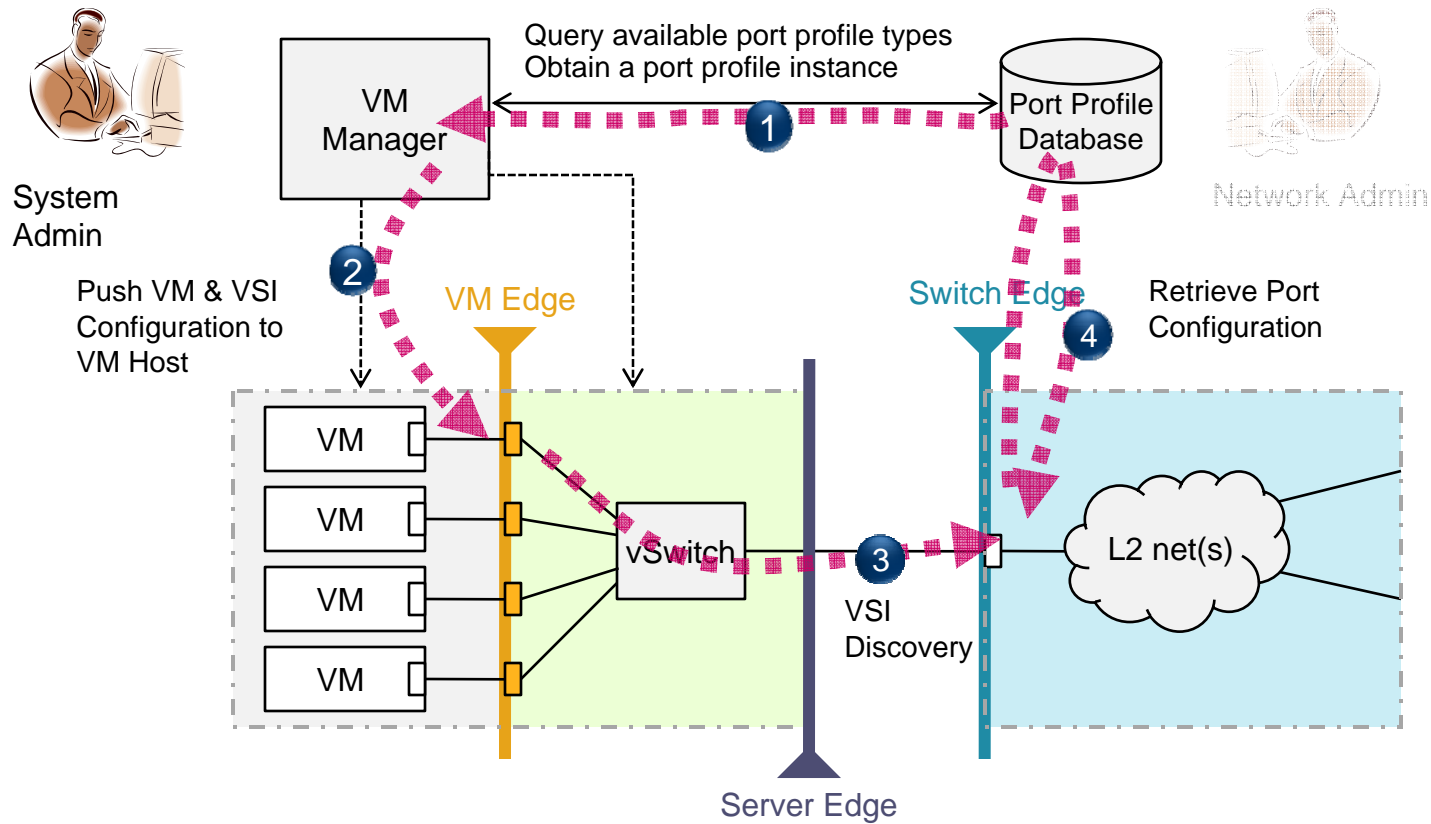
Uri Elzur, Broadcom

Ilango Ganga, Intel

Changes from v03

- Updated diagram on slide #3 by adding the Stable State TLV as part of the LLDPDU. Note that this slide deck merely contains observations and more information, but is NOT a departure from content of [bg-elzur-vPort-Discovery-LLDP-1109-v03.pdf](#)

Steps for Configuring Edge Connections (vPorts)



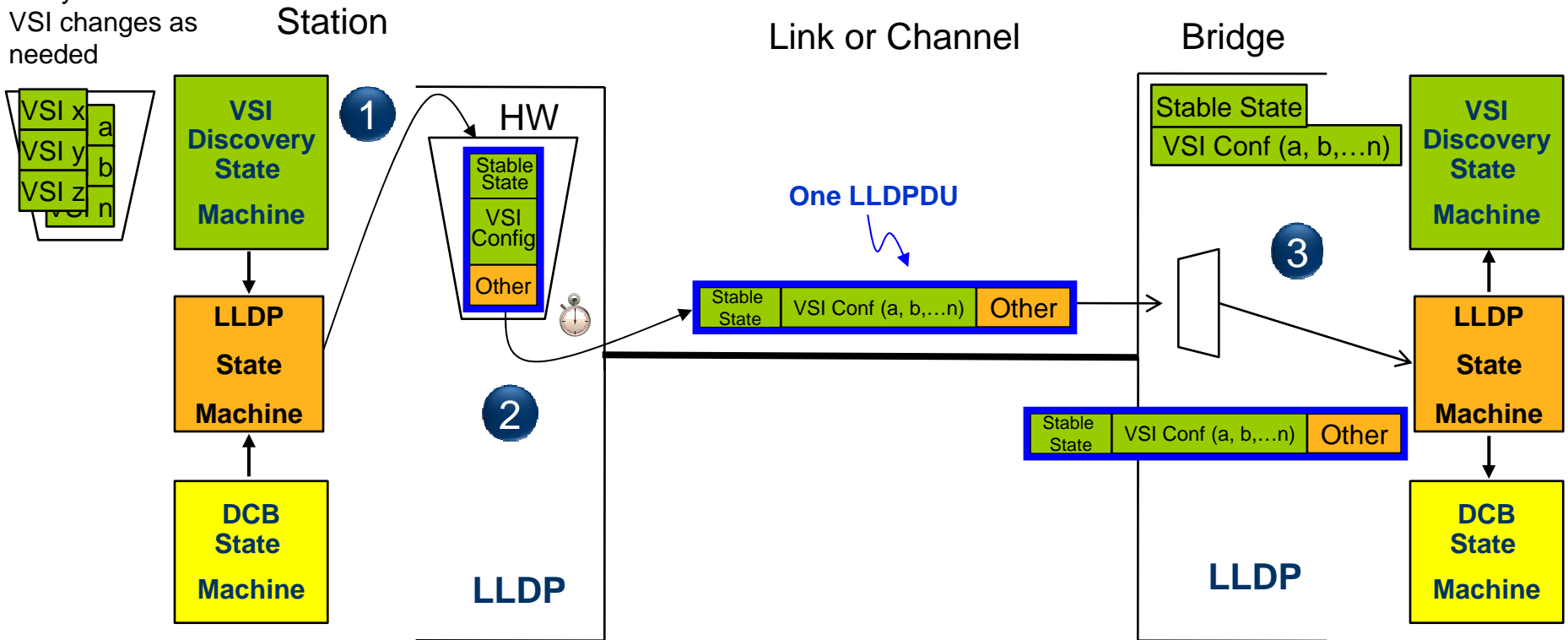
Legend: vPort (or VSI) 

LLDP

(single direction shown)

Notes:
Reliability relies on LLDP re-transmitted

Queue up as many desired VSI changes as needed



- 1 LLDP state machine creates a PDU with Required TLVs and a single VSI Config TLV containing all change requests from the VSI Discovery State Machine, that fit into the LLDPDU. LLDP State machines queue up LLDPDU for transmission

- 2 LLDP supports an "immediate PUSH" in case of MIB change. LLDP Timer is used. Only ONE outstanding LLDPDU, Device storage limited to 1500B.

- 3 Bridge consumes LLDP Required TLV and the VSI Config TLV and responds when it chooses. Bridge storage is limited to 1500B per port!

Consensus and New requirements

- **Consensus?**

- To have ALL EXPLICIT VSI state in one message, MAY require more than 1500B
- No more than one 1500B message outstanding between link partners at any time
- No ACK or Sequencing required
- Stay as close to LLDP as possible
- Must run LLDP anyhow for DCBx
- Limits to how many VSI state changes/Second can be realistic to supported by HV/Server/switch/Network/Storage/Management subsystems
- LLDP can pass config information to a link partner e.g. DCBx
- LLDP can pass information to another state machine e.g. DCBx

- **New Requirements**

- Desire to send State Changes only + Digest of all VSI state
- Faster reaction time

- **Solution space: Require new protocol or minor adjustments to / interpretation of LLDP?**

Why LLDP may be of interest

- **No backwards compatibility issue with new proposed LLDP interpretation**
 - Any “change” affects only products targeting Virtualization in Data Center
 - No impact to devices not supporting virtualization TLVs
- **Changes in last 802.1AB-REV**
 - Allow immediate PUSH (transmit) in case of MIB changes
 - Allow Fast Re-Transmit
 - Added tags for channels (S-Components) to accommodate needs of Provider Bridging
 - Above makes one optimistic that LLDP may allow some additional flexibility
- **With VDP/T3P – Two (2) protocols need to be run simultaneously !!!**
 - Can’t eliminate DCBx that requires LLDP
 - 2 protocols mean higher product complexity and wire activity
- **Time line – may take longer to get IEEE to approve a new protocol**
 - Should be faster to get a read on IEEE802.1 willingness to update LLDP again
 - NO CHANGE to LLDP is needed, maybe change the Organizationally Defined TLV restrictions
 - Will take longer time to finalize new protocol
 - Will require IEEE802.1 complete review/approval of the new protocol
 - Not clear how DCBx/LLDP and new protocol interact/interfere with existing products

Other LLDP Misnomers

- How many events per Second can the whole system handle?
- Scalability
 - Similar to T3P, assuming 1500B outstanding between the 2 link partners
 - Can handle more than 64 Full State changes at one time but will not communicate more than 64 Full State changes at one time, till switch responds (note: a Simple state change to De-Associate is not limited)
 - With assumptions presented, LDDP with 64 VSI consume – vPort Config TLV has head room for more information per VSI
 - LLDP based vPort TLV for 1024 vPorts with 64 changing : $11 + SC\text{-}Max * 17B + 128B = 11 + 64 * 17 + 128 = 11 + 1068 + 128 = 1207B$
- Does not require multiple instances of same TLV in an LLPDU – One TLV contains a list of entries (sub-TLV) each for a different VSI
- LLDP doesn't REQUIRE multi channel, it simply allows higher scalability if Multi Channel is already used
- LLDP does not preclude any separation of DCB and VSI to separate agents, nor does it require it

Implications of multiple outstanding PDUs

- **Protocol**

- Requires a more complete Transport Protocol
- Negotiation of how many outstanding LLDPDs (or 1500B frames) are supported
- Complexity of ACK and sequencing
- Complexity of dealing with partial processing State Change with each PDU

- **Switch Memory requirements (64 port switch)**

- LLDP, one outstanding LLDPDU: 64 port * 1500B = **96KB**
- T3P, VSI change for 1000VMs: 64 port * 1000 VM * 20B = **1.28MB**
- T3P, VSI change for 1000VMs, AMPP TLV: 64 port switch * 1000 VM * 50B = **3.2 MB !**