Trivial TLV Transport Protocol (T3P-R) A Possible Transport Protocol for VSI Discovery

Paul Congdon, Chuck Hudson

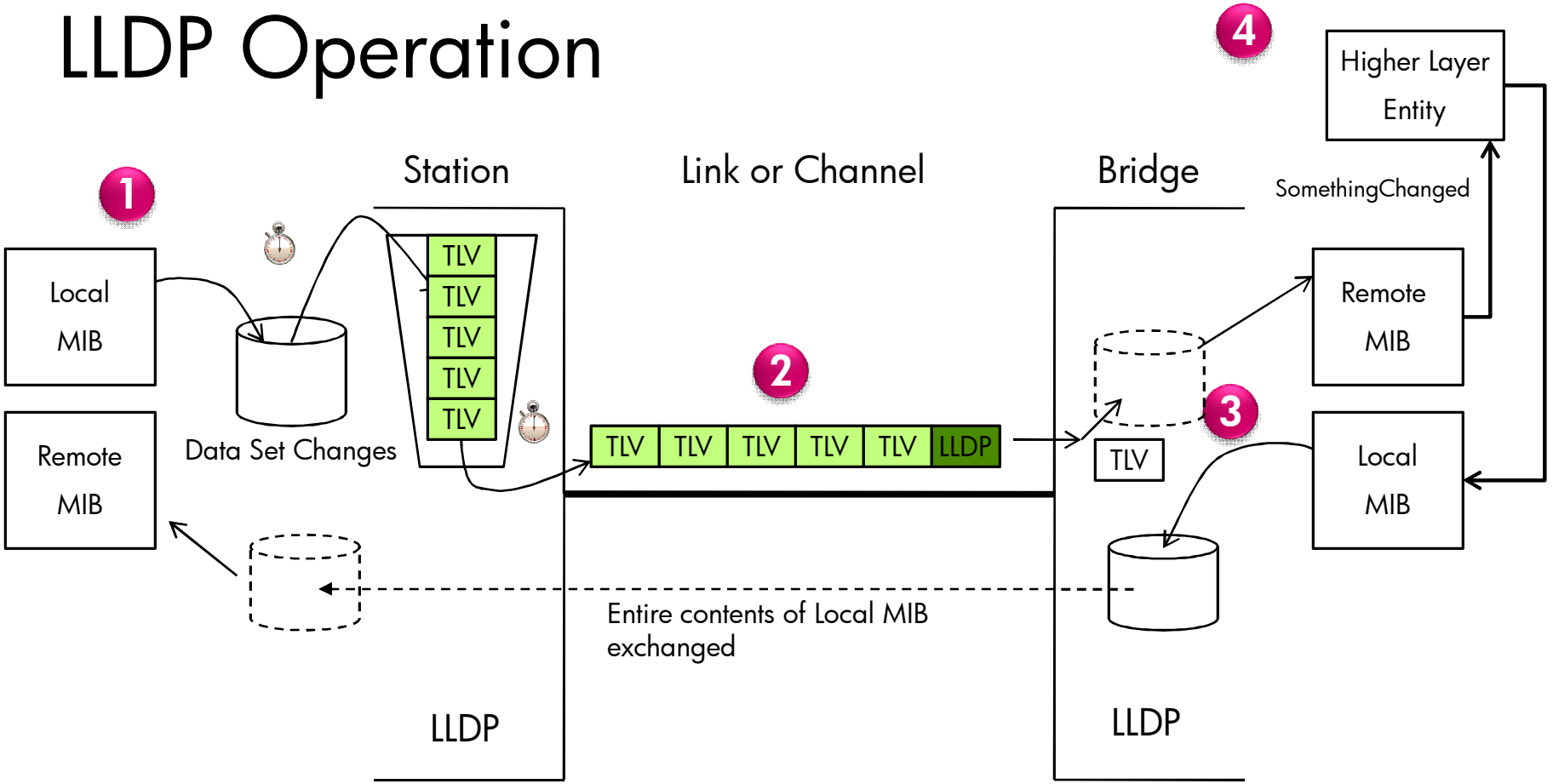
Based on discussions with Gaurish Dalvi, Norm Finn, Avi Godbole, Srikanth Kilaru, Mike Krause, Vijoy Pandey, Joe Pelissier, Renato Recio, and Rakesh Sharma

http://www.ieee802.org/1/files/public/docs2010/bg-hudson-t3pr-0110-v10.pdf

Outline

- Review LLDP Operation
- Discuss EVB/VSI Needs
- Conceptual Protocol Proposal (T3P-R)

LLDP Operation



1. Values change in Local MIB
2. All 'enabled' records in Local MIB are transmitted in a single PDU
3. Peer's entire remote MIB is replaced with contents of PDU
4. Higher layer entities are notified of changes. Higher layers may cause local changes

Attributes of LLDP that we like…

- It is very simple – we made it so on purpose
- Enables a synchronized view of a local database. Entire local database transmitted in single PDU
- Single periodic timer for all data in local database
- Unacknowledged delivery. Achieves reliability through periodic retransmission. Handles the 'silent reset' problem well.
- Extensible record format (TLVs)
- Multiple higher layer entities can subscribed to advertised information

NOTE: Higher layer protocols are built above LLDP by converging on a common view of the local database and acting upon the advertised state

Attributes of LLDP that we don't like so much for EVB…

- One way protocol (??)
- Single PDU to transmit all local data
- All 'enabled' local data must be transmitted in each PDU
- Receiver 'forgets' all previously received information
- Unacknowledged delivery. Achieves reliability through periodic retransmission
- All higher layer protocols are subject to LLDP's transmit timer
- Difficult to implement Query/Response type protocols

Overall EVB/PE Needs

1. Need to communicate the bindings of VSIs to VSI Profiles
2. Need to communicate the bindings of S-Tags to a channel and a pair of (v)Ports
3. Need to communicate the bindings of M-Tags to (v)Port Sets (could be communicated on LLDP)
4. Query/Response of individual records of information (e.g. statistics)

NOTE: Almost all of these could be considered a synchronization of 'data set' state between the Edge Device and the Adjacent Bridge

Protocol Philosophy

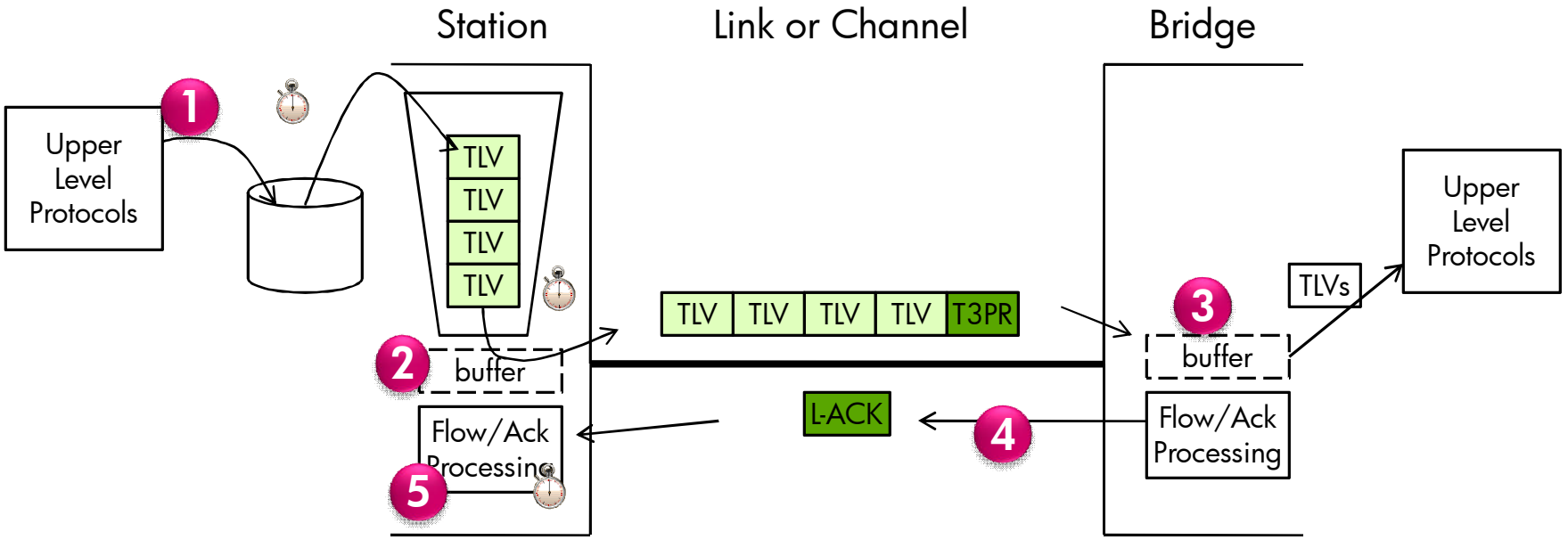
- Separate into two layers,
  - Lower layer bus for reliable delivery,
  - Higher layer state exchange
- Exchange state, not commands
- State is represented as a set of attributes (e.g. data pairs, bindings, individual values)
- Indicate when state exchange is complete or in progress (higher layer issue)
- When possible, exchange only the partial changes to the state, not always the entire state
- Allow the transport of multiple, independent sets of state (e.g. multiple higher layer protocols)

Protocol Concepts

- Lower layer delivery bus
  - Efficiently packs higher layer messages into PDUs
  - Provides reliable delivery of individual PDUs
  - Simple ACK flow control (window size of 1)
  - Minimizes complexity of higher layer protocols (i.e. avoid higher layer timeouts, retransmissions, etc)
- Higher layer data set sync
  - Transmits data set records to remote peer. Entire data set may require several PDUs
  - Transmissions may include a 'digest' of all previous transmitted records, per data set, or since last digest
  - Supports ability to detect the need for, and invoke, a re-transmit when digest doesn't match at receiver
  - Data set digests are periodically transmitted, depending upon higher layer's needs
  - Multiple higher layer entities may share lower layer bus

T3P-R
Lower-level Transport Overview

T3P-R has symmetric behavior, but only a single direction shown here.



1. Client protocols pass outgoing TLVs to lower-level transport. TLVs are queued until frame is ready to be sent.**

2. Frame with TLVs is transmitted but the frame is not yet deleted from the transmit buffer. A L-ACK timer is set.

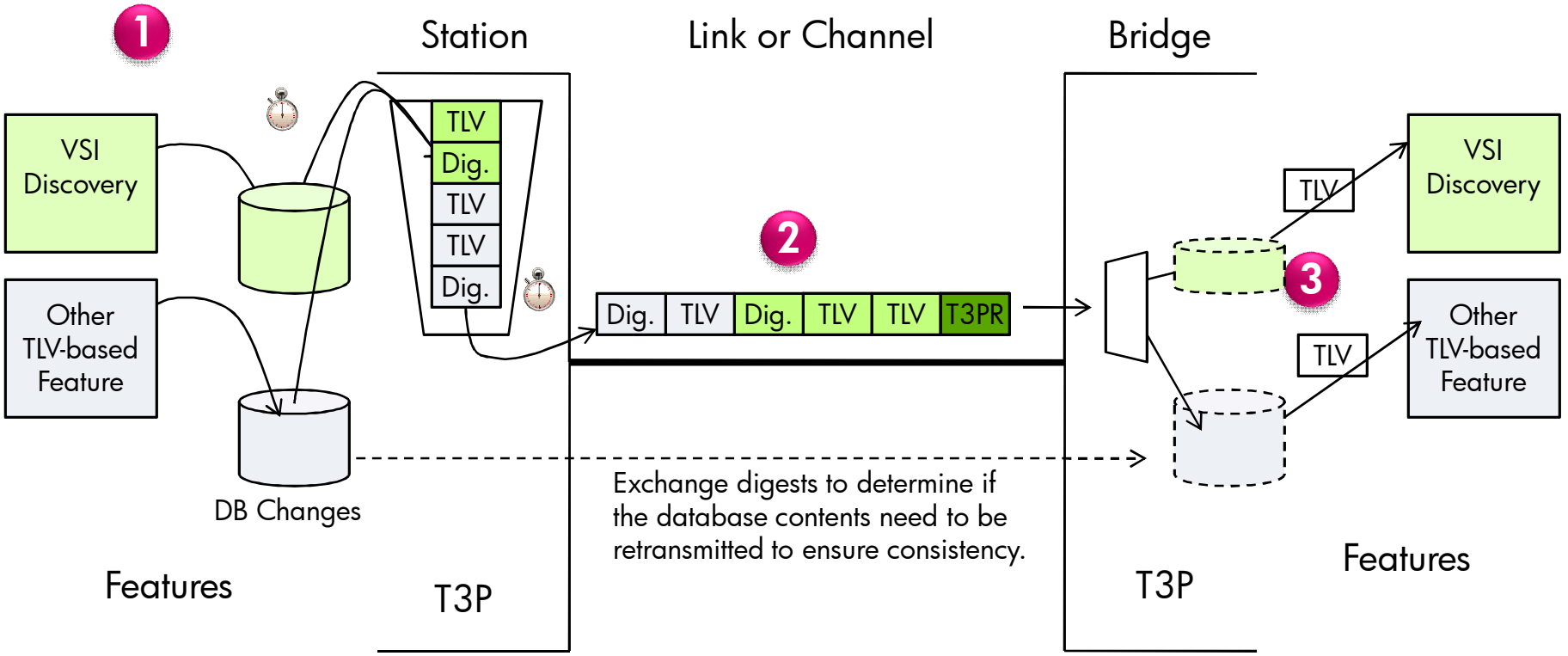
3. Arriving frame is received into a receive 'buffer', where it is held until it is removed by frame/TLV processing to pass TLVs to the upper level protocols.

4. When the receive buffer is emptied, a low-level acknowledge (L-ACK) is sent to the sender.

5. If the L-ACK is received before the timer expires, then the transmit buffer is cleared and the next TLV can be transmitted from the queue.

   If the L-ACK timer expires before the L-ACK is received, then the frame in the transmit buffer is resent.

Upper-level View: Verifying Consistent Station/Bridge State



Lower-level protocol is used for grouping TLVs into frames. Includes bus-level frame acknowledge. Also provides flow control. Notes gain/loss of overall connectivity.

Summary

- There are many attractive things about LLDP
- Using LLDP for all EVB needs is challenging
- A new protocol is proposed that:
  - Maintains many of the 'good' things about LLDP
  - Addresses short comings of LLDP for EVB use
  - Separates the lower layer transport from the higher layer users
  - Provides an efficient mechanism for multiple higher layers to exchange and synchronize views of data sets