Thoughts on EVB/PE protocols

Paul Congdon

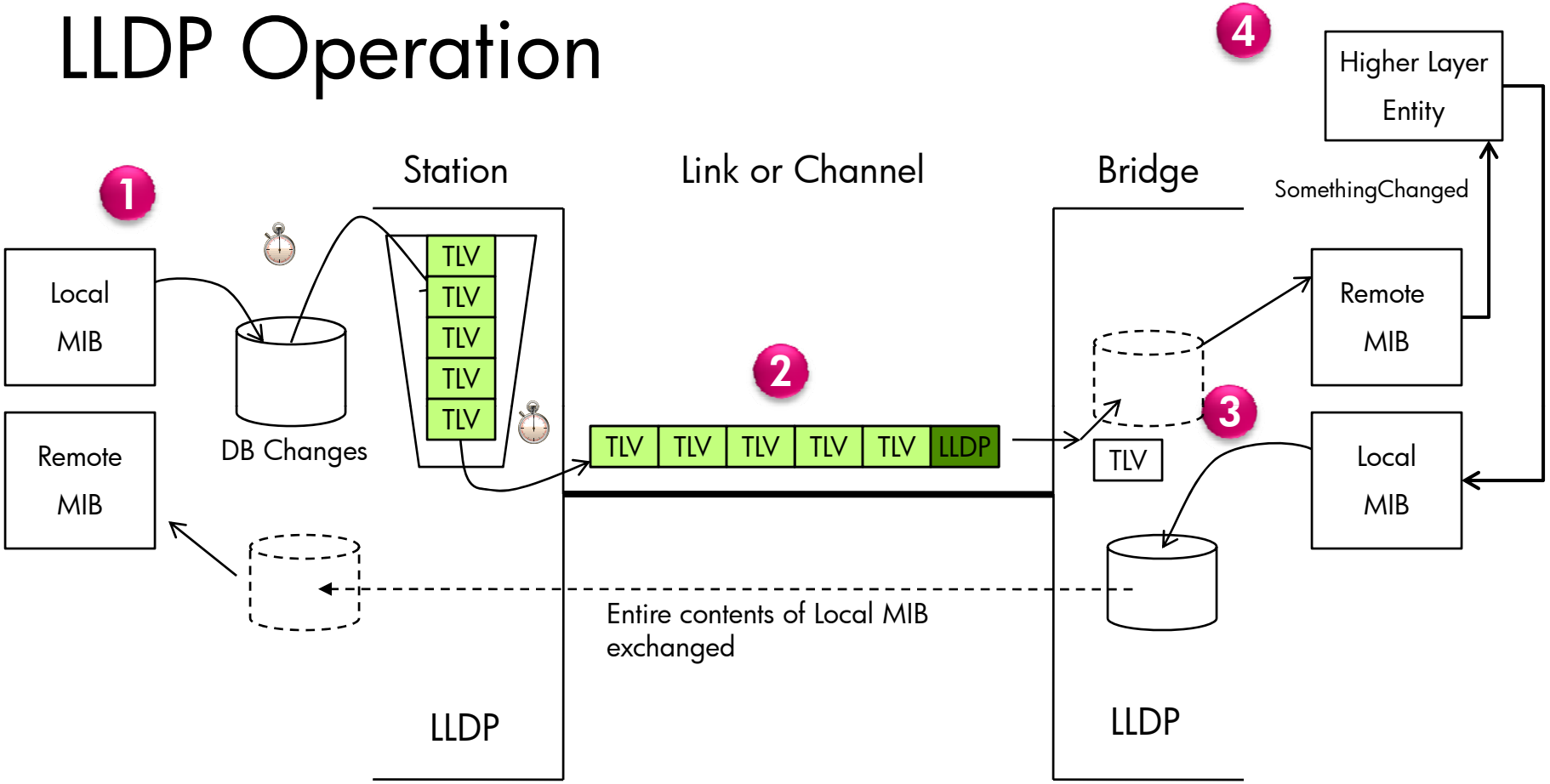
Based on discussions with Chuck Hudson, Norm Finn and Joe Pelissier

Agenda

- Objective: To discuss, at a conceptual level, the characteristics of a possible protocol for EVB/PE

- Review LLDP Operation
- Discuss EVB/PE Needs
- Conceptual protocol proposal

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…

- One way protocol (??)
- Single PDU to transmit all local data
- All 'enabled' local data must be transmitted 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

Current Use of LLDP

- Basic Discovery
  - Chassis-ID, Port-ID, system names, interface names, versions, etc..
- 802.1 Extensions
  - VLAN consistency
- 802.3 Extensions
  - PoE+ negotiation
  - Duplex, Jumbo, LAG setting consistency
- LLDP-MED
  - MED capabilities, Network QoS, MED Inventory
- DCBX
  - PFC, ETS and ECN settings

Current LLDP TLVs, more on the way…

TLV Type	Where Defined	ANSI / TIA standard	IEEE 802.1AB standard
Chassis ID	Base	M	M
End of LLDPDU	Base	M	M
Extended Power via MDI	TIA	Y-if PoE	NA
Inventory management	TIA	O	NA
Link Aggregation	802.3	O	O
LLDP-MED Capabilities	TIA	M	NA
Location Identification	TIA	O	NA
MAC/PHY configuration/status	802.3	M	O
Management address	Base	O	O
Maximum Frame Size	802.3	O	O
Network Policy	TIA	M	NA
Port and Protocol VLAN ID	802.1	O	O
Port description	Base	O	O
Port ID	Base	M	M
Port VLAN ID	802.1	O	O
Power via MDI	802.3	Not recommended	O
Protocol Identity	802.1	O	O
System capabilities	Base	M	O
System description	Base	O	O
System name	Base	O	O
Time to Live	Base	M	M
VLAN Name	802.1	O	O

EVB/PE Needs

1. Need to communicate the bindings of VSIs to Connection 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
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

EVB/PE Data Set Sizes

1. VSI/Profile mapping
   - Scales with the number of VMs (Potentially 100s per channel)
   - May include MAC address, VLAN, identifier
   - Consistent view of the complete set useful for ACL implemention
2. S-Tag/Channel mapping
   - Scales with the number of vPorts (0-4094), but likely less than 48
   - The set of mappings determines the set of active vPorts
3. M-Tag/vPort Set mapping
   - Combinations of vPorts for bcast, mcast and floods
   - A function of the number of active vPorts
   - A consistent table needed for proper frame delivery
4. Query/Response
   - Statistical counters

Why LLDP isn't the best choice for EVB/PE

- Already many users vying for space in the LLDPDU
- The need to exchange more than 1 PDUs worth of information
- Knowledge of reliable reception at the far end
- Entire local database sent on each transmission (inefficient)
- State changes in any higher layer cause transmission of entire local database
- Working around LLDP restrictions may create an explosion in the number of LLDP agents within an entity
- Difficult to implement Query/Response support

Thoughts on improvements

- Separate the lower layer protocol from the higher layer users.
- Support acknowledge transfer at the lower layer.
- Support the ability to exchange a related set of information (e.g. database, data set), larger than a single PDU
- Allow an efficient (compressed) way of assuring synchronization without having to periodically transmit the entire data set.

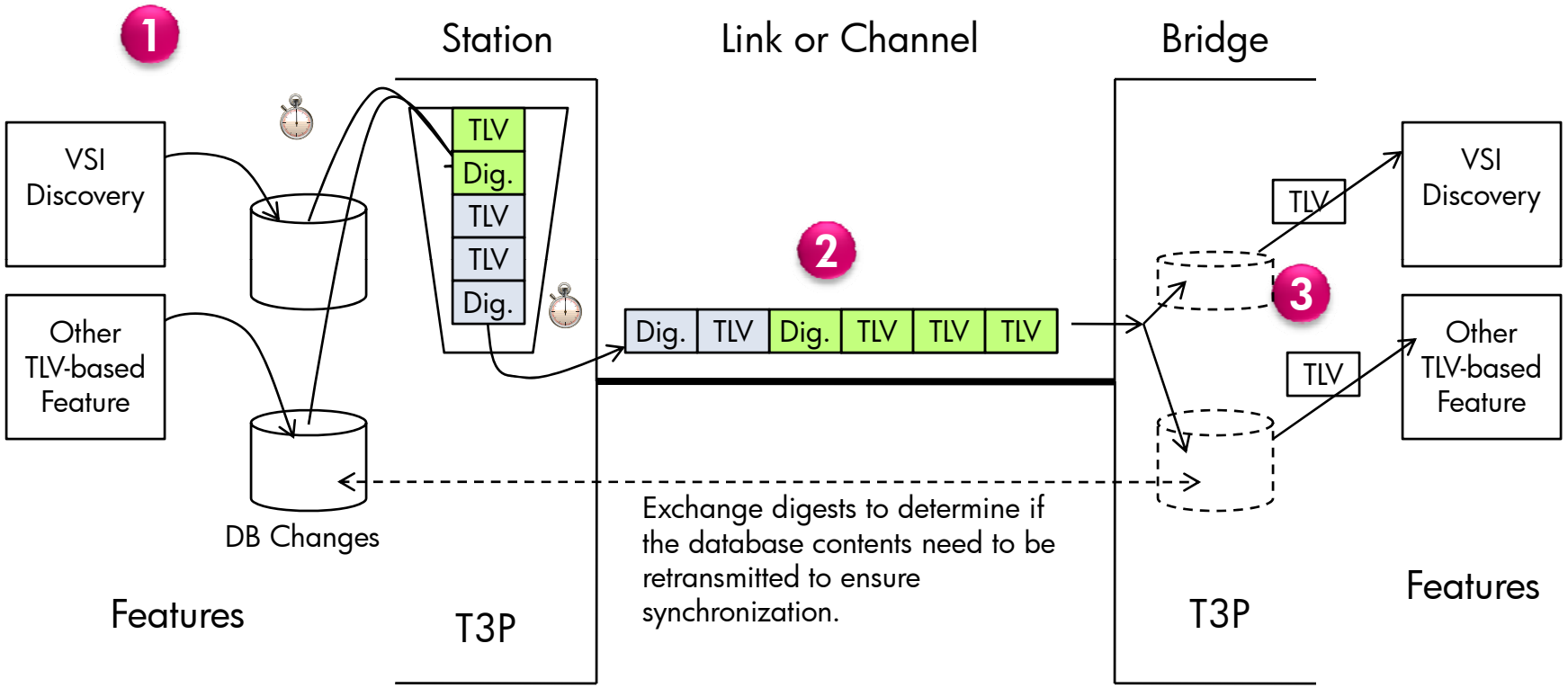
Protocol Philosophy

- 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 include a 'digest' of all previous transmitted records, per data set, since last re-sync
  - Supports ability to detect the need for, and invoke, a re-sync 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

Data Set Exchange with Sync



Basic bus protocol 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 EVB/PE needs is challenging
- A new protocol is proposed that:
  - Maintains many of the 'good' things about LLDP
  - 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