Hewlett Packard
Enterprise

Edge Control Transport for LRP

September 2017

Paul Bottorff, Network System Architect, Aruba CTO's Office
Paul.Bottorff@hpe.com

Link-local Registration Protocol Choices

- Draft 1.0 is based on an IS-IS-like protocol supported by two data transport options
  - Raw 802 datalink, LRP-DT ISS
  - TCP Transport (of some type), LRP-DT TCP

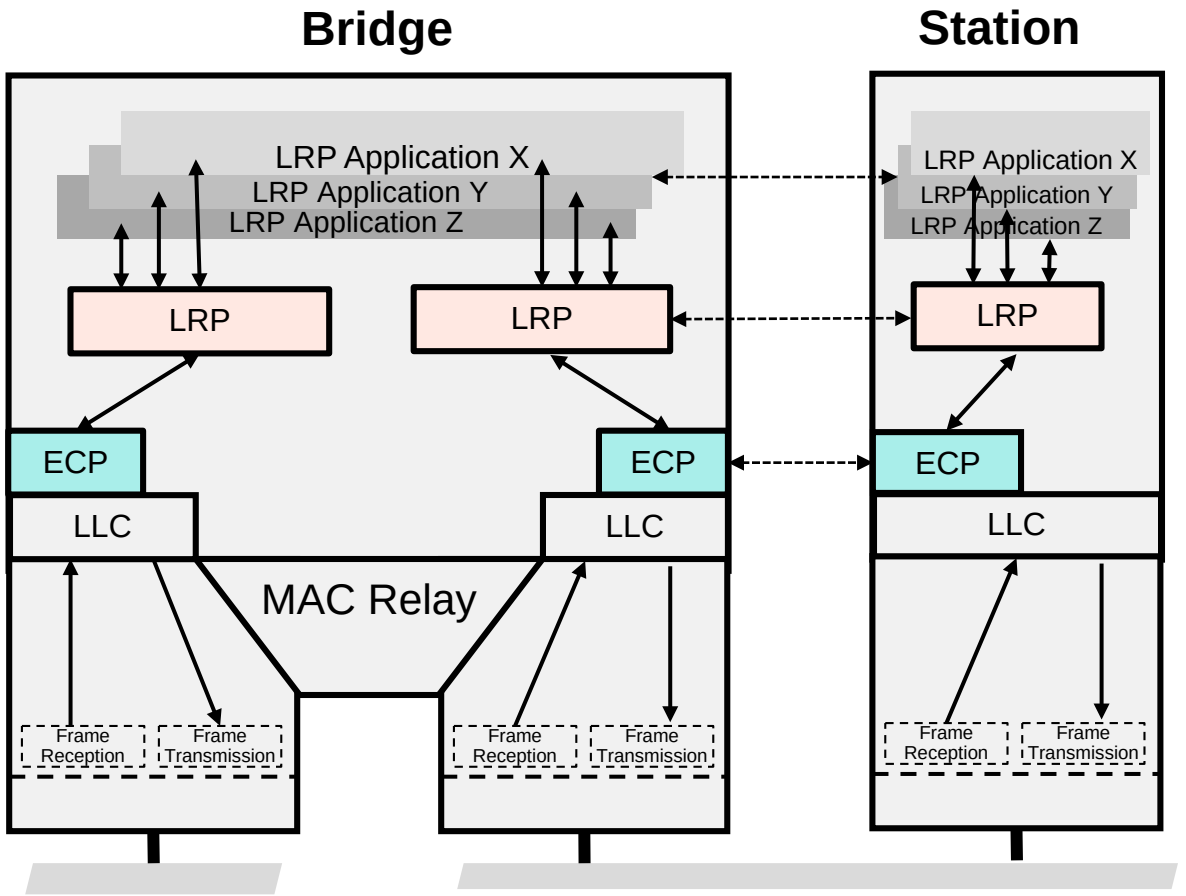
- This presentation explores the use of the Edge Control Protocol (ECP, IEEE Std 802.1Q-2014 clause 43) as a data transport for an IS-IS-like protocol
  - We propose IEEE use ECP as the transport for LRP-DT

Edge Control Protocol (ECP) Service Characteristics

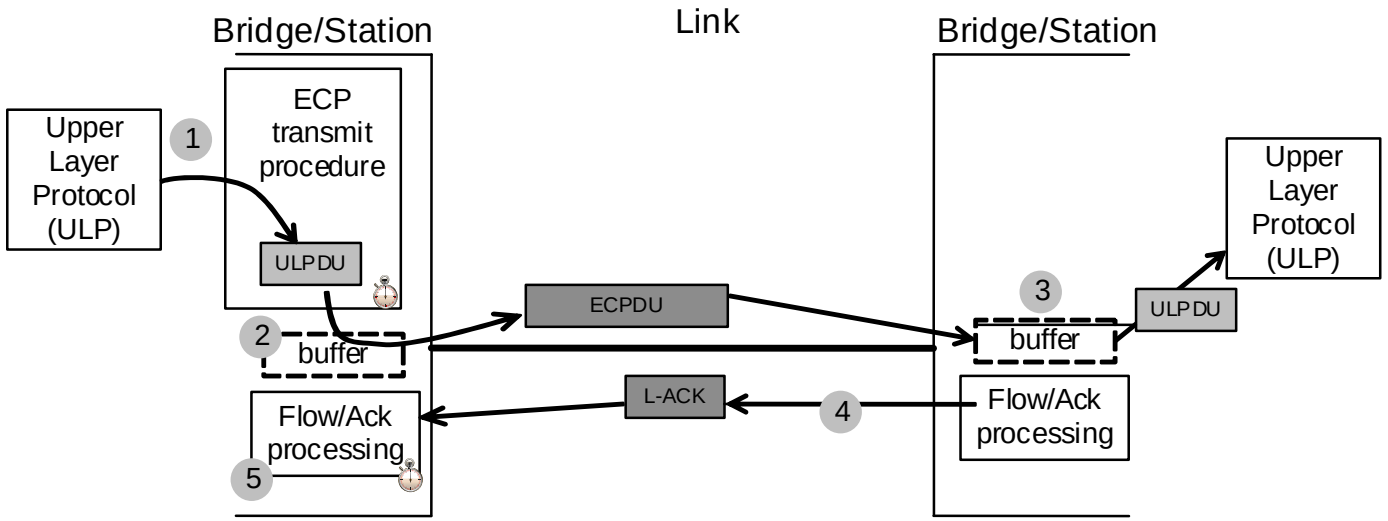
- ECP as defined in 802.1Q-2014 clause 43 supports transfer of control information between bridges and stations over 802 links
- ECP is currently used by two 802.1 protocols:
  - Virtual Network Interface Discovery and Configuration Protocol (VDP, 802.1Q Clause 41)
  - Port Extender Control and Status Protocol (PE CSP, 802.1BR)
- ECP can provide service to multiple Upper Layer Protocols (ULPs)
- Basic services provided by ECP are:
  - Reliable delivery of ULPDUs, resilient against frame loss.
  - In order delivery of ULPDUs to the recipient ULP
  - Delivery of a single copy of each ULPDU to the recipient
  - Flow control that provides protection against buffer overrun and congestion

Edge Control Protocol (ECP) in Bridges and End Stations

- ECP is an L2 transport service sitting on top of LLC
- Each ECP dialogs with a peer determined by the destination MAC address used in the ECP frame
- ECP layer provides a service interface for client layers which can be used by any Upper Layer Protocol (i.e. LRP in this example)
- By using ECP rather than LLC for LDP-DT, LRP can be based on a reliable delivery service



Edge Control Protocol (ECP) Operation

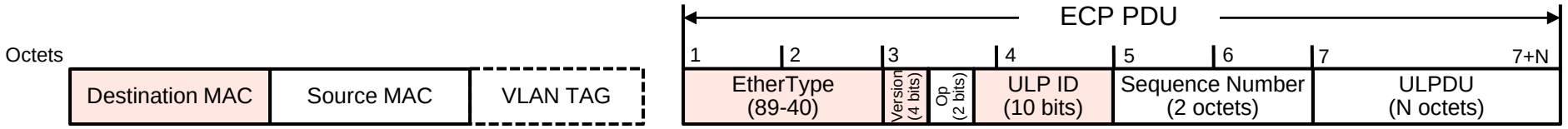


– ECP is a simple stop and wait automatic repeat request protocol supporting frame by frame transport of data from multiple Upper Layer Protocols (i.e. LRP database management protocol).

1. Upper layer (ULP) passes an outgoing ULP Data unit to ECP by invoking a transmit request procedure
2. The ULPDU (perhaps containing a set of TLVs) is transmitted and an ECP low-level acknowledgement (ACK) timer is set. The ULPDU is retained in a local ECP buffer.
3. The ECP frame is received into a receive buffer, here it is held until it is removed by an ECP procedure that passes the ULP Data Unit to the associated ULP.
4. When the receive buffer is emptied, a L-ACK is sent to the sender
5. If the L-ACK is received before the L-ACK timer expires, then the transmit buffer is cleared and ECP can process another ULPDU
5. If the L-ACK timer expires before the L-ACK is received, then the frame in the transmit buffer is re-sent and the L-ACK timer is re-initialized

– Simple enough to allow low level hardware/firmware implementations with reserved frame buffer

Edge Control Protocol (ECP) Frame Format

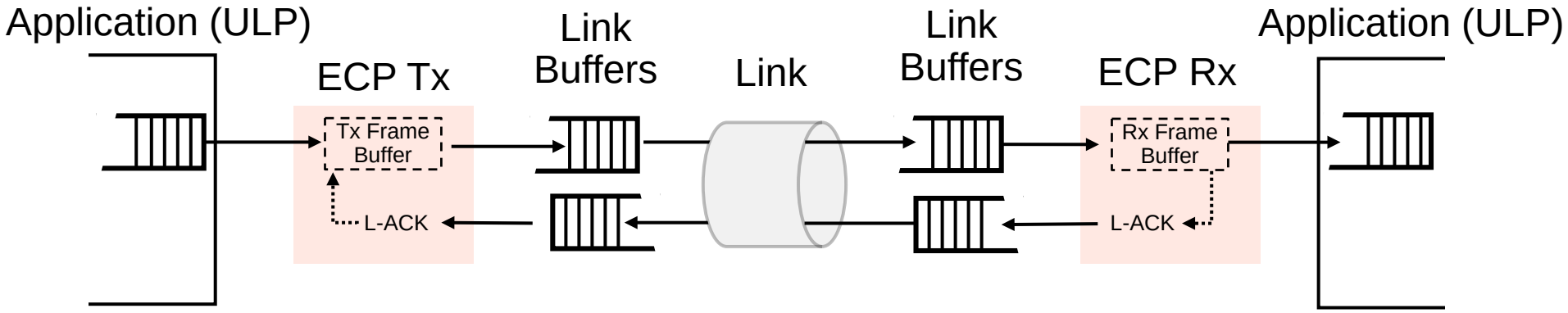


- ECP frame destinations are addressed by destination MAC + EtherType + Version + Upper Layer Protocol ID
  - Destination MAC addresses may be Group Addresses or Unicast Addresses
    - Currently the VDP protocol uses the Nearest Customer Bridge group address
    - Currently the 802.1BR PE-CSP protocol uses either the Nearest Non-TPMR Bridge group address or a unicast address determined by and LLDP exchange
    - Current work on VDP for NVO3 is expected to use a unicast ECP address
  - LLC decodes ECP by EtherType 89-40 and Version (0x1)
  - Upper Layer Protocol (ULP) identifiers are encoded in the subtype field
    - Currently ulpID 0x001 identifies the VDP protocol and ulpID 0x002 identifies the 802.1BR PE-CSP protocol
    - IEEE has reserved all values above 3 for future assignment to new protocol (like LRP)
- ECP Operation types are Request and Acknowledgement
- The sequence number is used for acknowledgement and error checking
- The Upper Layer PDU is opaque data passed to the ULP

ECP Configuration

- ECP has two configuration parameters which must be set:
  1. maxRetries
  2. ackTimerInit (in units of 10 usec)
- For the Edge Virtual Bridge application (IEEE Std 802.1Q clause 40-42) these parameters are negotiated between the link ends using the LLDP EVB Discovery and Configuration TLV (IEEE Std 802.1Q clause D.2.12)
- ECP can be used without the LLDP by setting the parameters either by some other discovery protocol or by provisioning
- Adding a Configuration and Discovery LLDP TLV for the LRP protocol with ECP configuration parameters is desirable since it allows the negotiation of the ECP parameters between the link ends

ECP For LAN/MAN Control Applications



- ECP transmits a single frame per round trip time therefore the round trip time limits throughput
- Round trip time depends on transmission time, link latency, chip pipeline delay, link scheduling delay, queueing delay, ECP scheduling delay, and ECP execution time
- Assuming uncongested links with zero link scheduling delay, queueing delay, and ECP scheduling delay we can approximate the best case bandwidth

ECP is sufficient for LAN/MAN LRP-DT

Approximate Best Case Throughput*

Link	Dist	Max f/s	utilization	1 Mbyte
1 GigE	100 m	60,000	75%	10 msec
1 GigE	1 Km	40,000	50%	20 msec
1 GigE	10 Km	8,500	10%	80 msec
1 Gig	100 Km	1000	1%	700 msec

*Considering only link latency, link transmission time, and using 1 usec for chip pipeline delay. Calculations assume 1 Mbyte of data is delivered in 1518 octet frames.

TCP For WAN Links

- Some advantages to TCP are:
  - Bulk transfer rates are mostly independent from link distance
  - Bulk transfers can utilize full link bandwidth
  - The ends of a TCP connection could be separated by an L3 rather than L2 network (assuming we are using IP addressing)
  - TCP provides fragmentation from 64K datagrams into packets
  - TCP is commonly available in current Bridge control stacks
- Some problems with TCP
  - Using IP addresses will require a way to distribute them over the links
  - There are many TCPs to choose from
  - For single frame transfers TCP will not perform better than ECP
  - TCP does not enable low level hardware/firmware implementation
  - If LRP provides fragmentation service, then TCP fragmentation is of no added benefit
  - Usually it is desirable to limit the bandwidth consumed by control protocols
- TCP is indicated if WAN links and L3 transit networks are a requirement

Do we need TCP or will ECP do?

ECP

- Simple, light weight, fully specified by 802.1
- L2 MAC/Ethertype/ULP Addressed
- Flow Controlled, Sequenced, Reliable Delivery
- Congestion Mgmt effectively always in Slow Start
- No fragmentation service
- Throughput is less than link bandwidth and is distance dependent

TCP

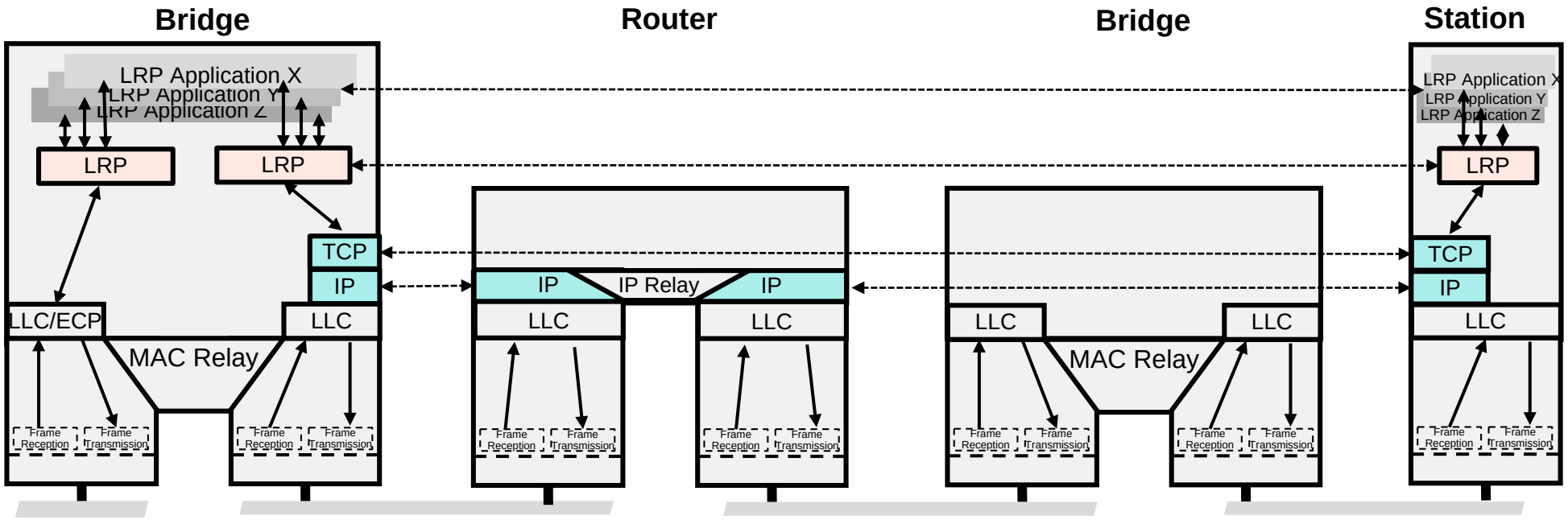
- Heavy weight, full service, many versions
- L3/L4 IP/Port Addressed
- Flow Controlled, Sequenced, Reliable Delivery
- Congestion Mgmt adjusts to fastest possible rate
- Fragmentation service from 64K datagrams to packets
- Bulk transfers at full link bandwidth mostly independent from distance

Summary

- Edge Control Protocol (ECP) is perfect for LAN/MAN LRP-DT applications
  - Simple protocol using MAC/Ethertype/ULP addressing
  - Supports group or unicast MAC addressing
  - Fully specified by IEEE 802.1Q cl43
  - Good performance at LAN/MAN distances
  - Practical for low level hardware/firmware implementations
  - Configuration using LLDP TLV or other discovery methods to setup parameters
- If we need to support WAN links then TCP could be supported as an optional feature
  - A TCP connection may be provisioned at each link end using TCP socket addressing
  - IEEE would need a TCP Port Number for 802 protocols over TCP
- ECP and TCP both provide:
  - Flow control
  - Reliable frame delivery
  - Sequenced frame delivery
- The Link-Local Registration Protocol should be designed assuming reliable sequenced frame delivery

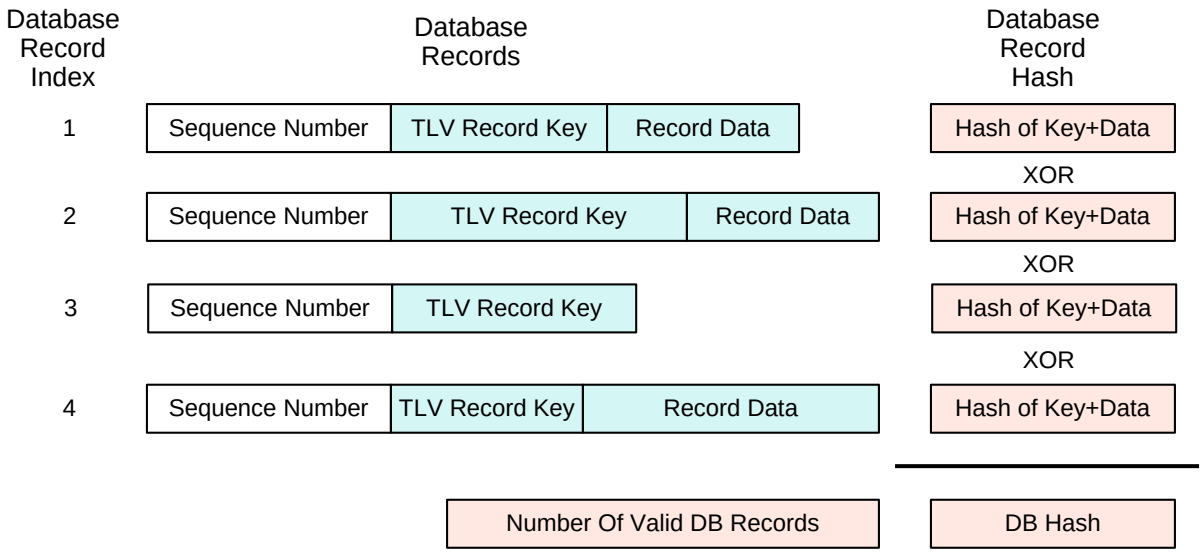
Backup Slides

TCP in Routers, Bridges and End Stations



- What TCP is best at is traveling through a series of relays
- Standard TCP addressing (IP/Port) allows passage through L2 and L3 relays
- Adjacent relay does not have enough information to configure a destination IP address

Digests For Verifying Synchronization



- Here each record is accompanied by a hash (such as SHA-256 or MD5) rather than a CRC
- This allows simple calculation of a hash for the entire (valid) database by XOR or record hashes
- Insertions and deletions from the database update the DB hash by XOR
- By combining the DB hash with the number of valid records in the DB we have a compact and high reliability database digest
- The digest can be exchanged between the applicant and registrar to verify database synchronization