Bridge Port Extension using PBB-TE

Paul Bottorff

Ben Mack-Crane

David Martin

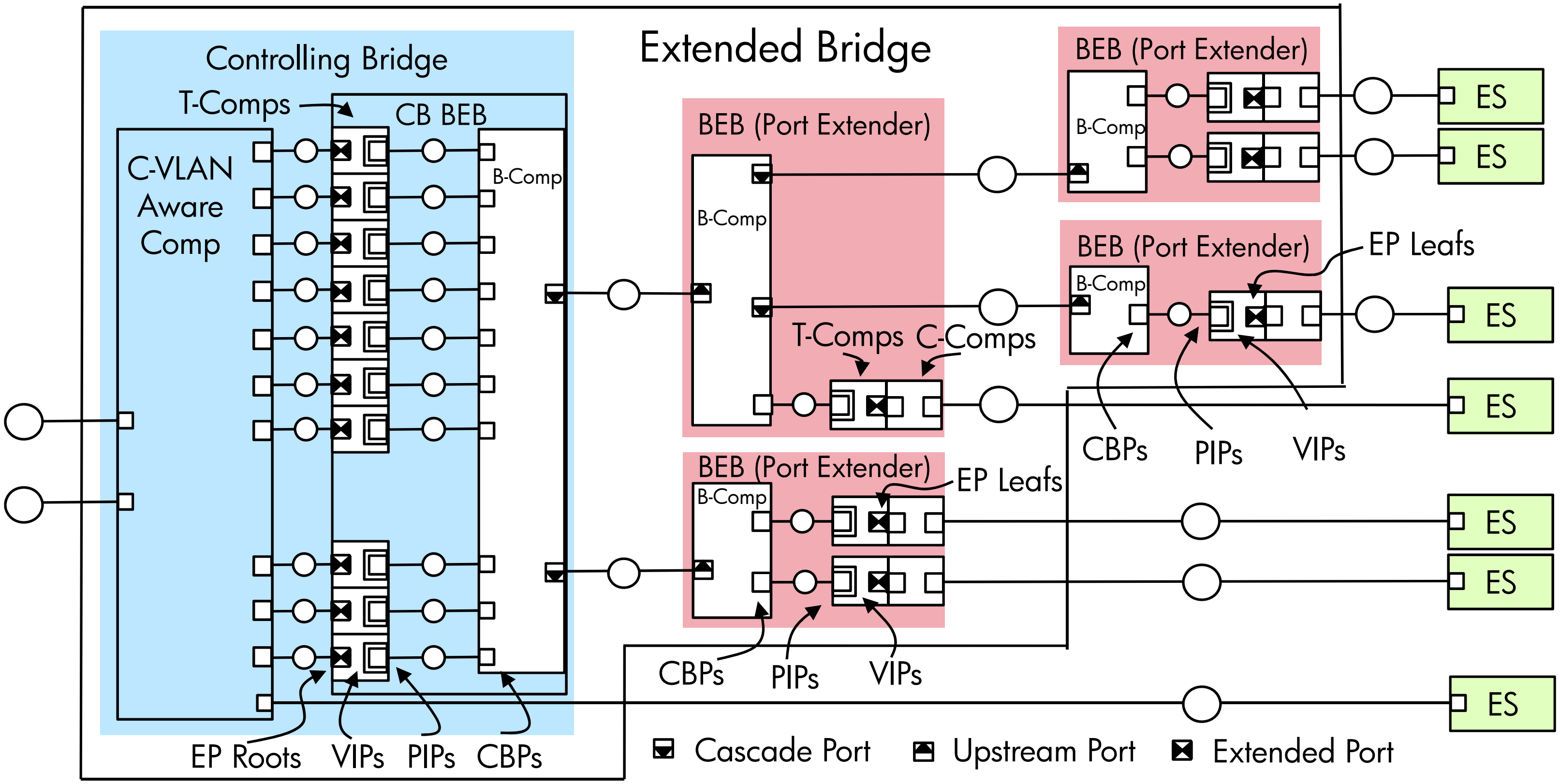
Panagiotis Saltsidis

See contribution bh-bottorff-pbbte-pe-draft-0711-v1 for further details

Comparison between E-TAG and PBB-TE Port Extenders

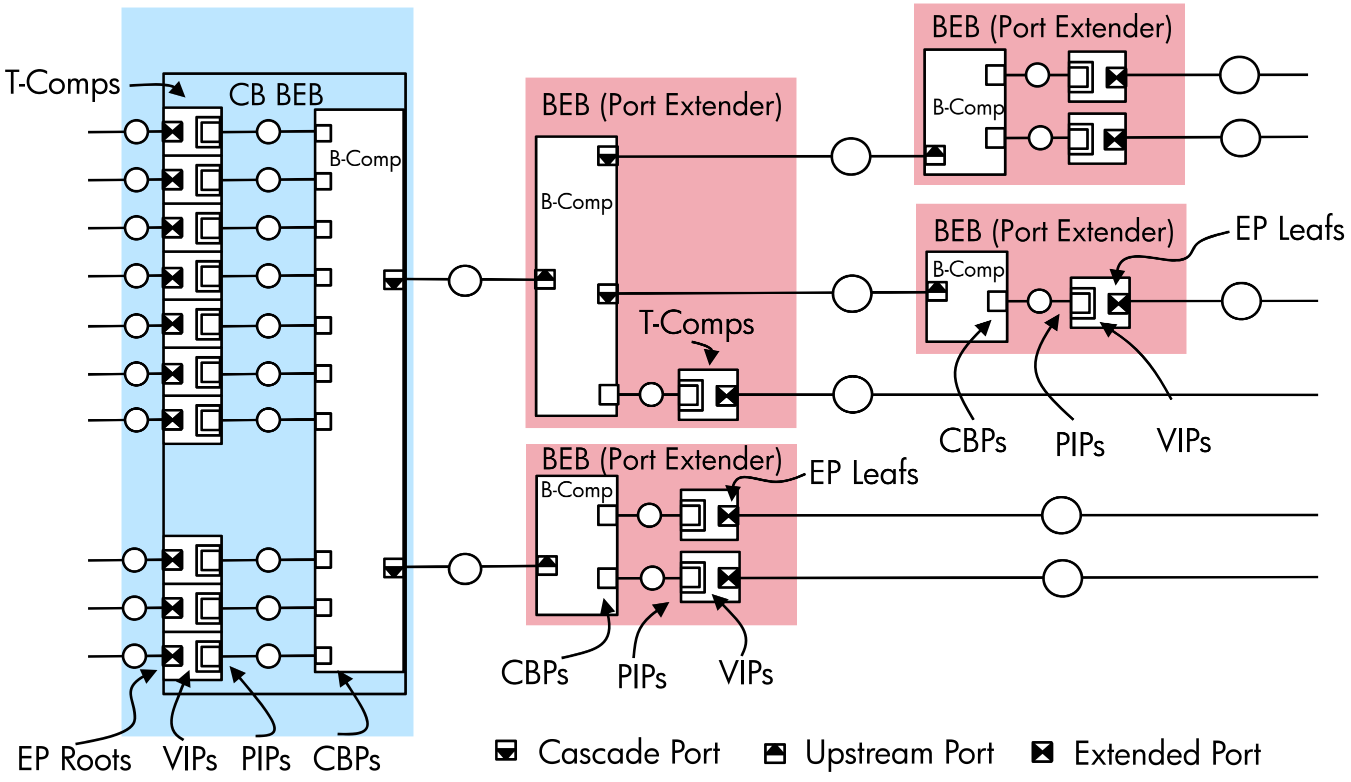
	E-TAG PEs	PBB-TE PEs
Scalability		★
Failure detection and status reporting		★
EVB synergy	☆	★
Compatibility with existing 802.1Q Bridge relay		★
No new components		★
No new tags		★
Optional support for CFM, protection and multipathing		★
Optional support for congestion notification	☆	★
Optional support for ETS and PFC	★	★
Optional support for EVB & VEPA	★	★
Lowest overhead octets	★	

Extended Bridge built from BEBs



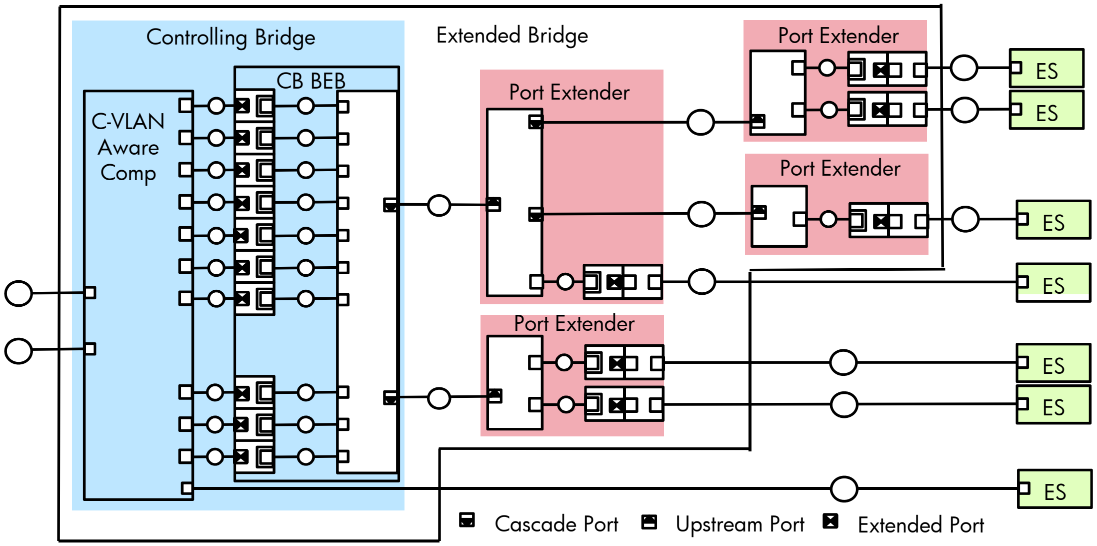
- 🟦 The EVB Controlling Bridge is composed of a BEB with a B-component and a T-component per VIP coupled to the primary C-VLAN aware component (or S-VLAN aware component)
  - Each Cascade Port is just an exterior facing PNP of the BEB
  - The VIP's of the CB-BEB are modified to form the B-DA and B-SA based on information passed from the C-Component
- 🟥 A Port Extender is a BEB composed of a primary B-component and a T-component per EP
  - The VIPs of the Port Extender BEBs use standard T-Components
  - An optional 2-Port C-Comp on each leaf Extended Port is used for C-TAG manipulations

Port Extender PBBN



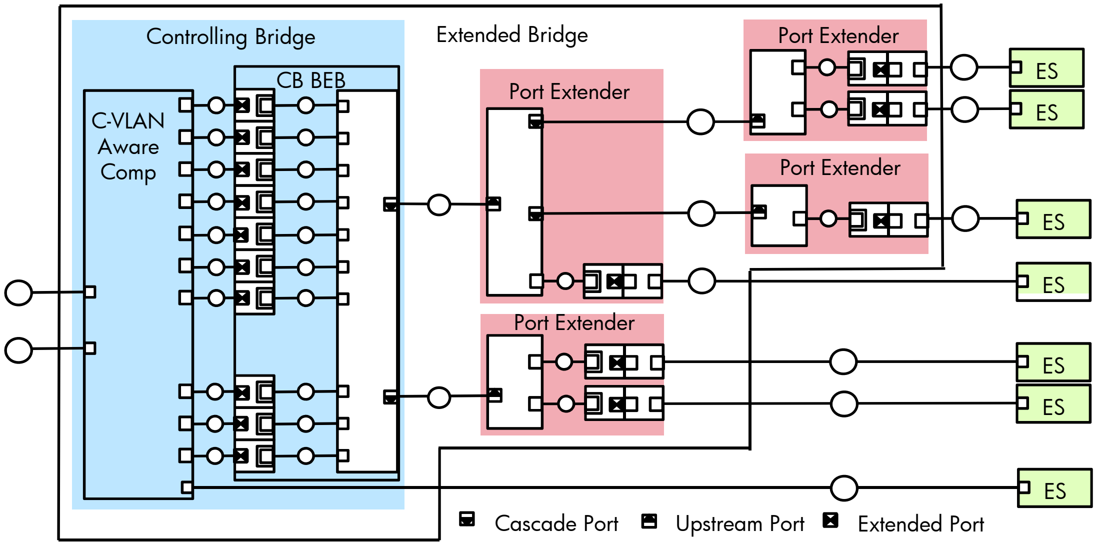
- The Port Extender components of the Controlling Bridge along with the external Port Extenders make a complete PBBN which can support PBB-TE forwarding

What is the same as PBB/PBB-TE?



- Just an application of PBB-TE with a limited topology and component organization, therefore the Controlling Bridge and Port Extender can be PBB/PBB-TE, with Extended Bridge feature additions.
  - Each leaf EP is connected to a CB-BEB VIP with a point-to-point TESI
  - Each UP is connected to a CB-BEB VIP with a point-to-point TESI
  - Each “replication group” or EP set is connected from a CB-BEB VIP with a point-to-multipoint ESP
- The Controlling Bridge’s primary component is modified as in 802.1Qbh
- The Port Extenders forward along configured TESIs
  - Each EP is attached to single VIP and PIP on a T-Comp
  - The PIP associated with an EP is identified by a unique B-MAC , which may be constructed using the E-PID

What is different from PBB/PBB-TE?

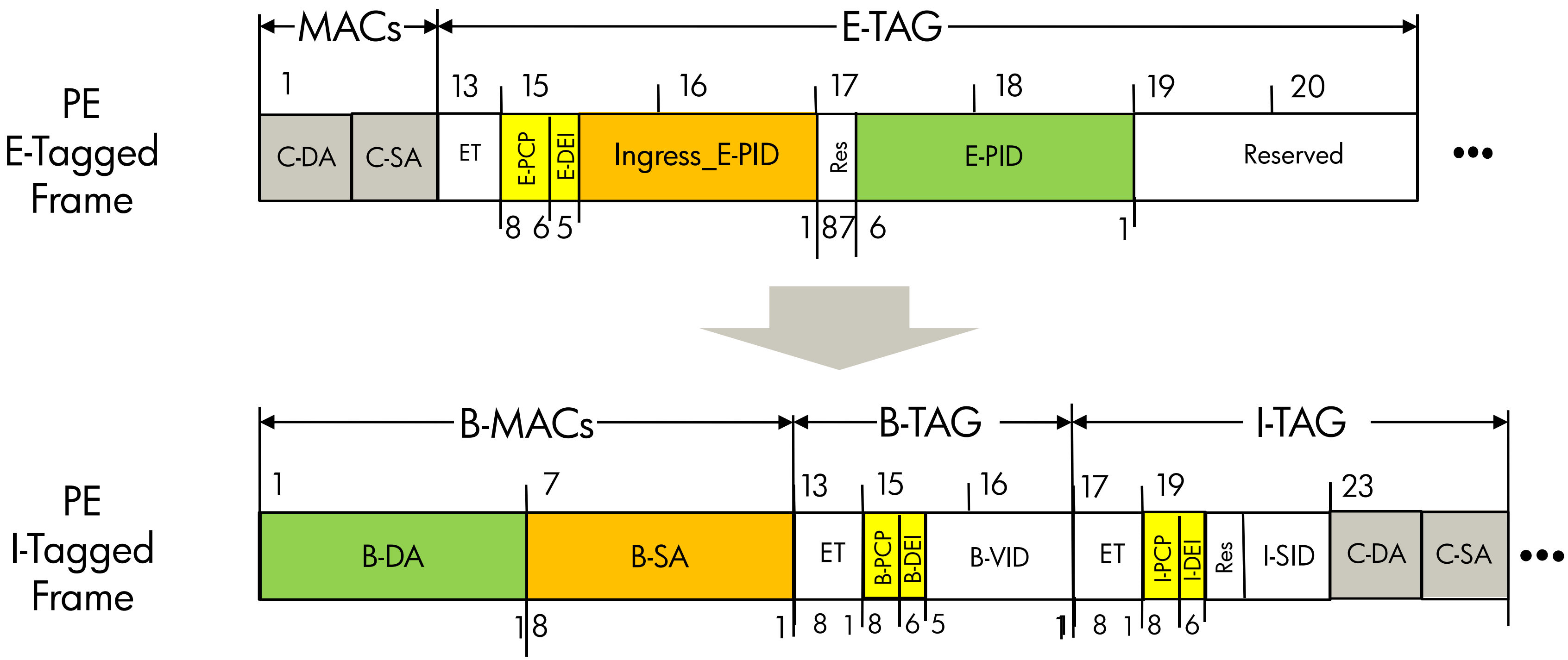


- The forwarding state for the CB-BEB and Port Extenders is configured using the Port Extender Control and Status Protocol
- The C-VLAN component relay issues one request primitive for each frame to be forwarded via the PE
  - The connection_identifier parameter carries a port map indicating the ports associated with EPs to which the frame should be forwarded
  - If the related indication primitive was received from the PE the request primitive is sent on the port from which the indication was received
  - If the related indication was not received from the PE, the request is sent on one of the ports indicated in the connection_identifier
- The CB-BEB PIPs assign B-MAC addresses selecting the ESP for each primitive according to modified rules for Port Extension
  - For "remote replication groups" the PIP selects a B-DA (E-PID) identifying a point-to-multipoint TESI (as currently in Qbh)
  - If the PIP's corresponding EP is not in the connection_identifier port map, the frame is marked for echo cancellation
- Echo cancellation is performed at the PIP associated with an EP, whenever the B-SA is equal to the corresponding root EP's CB-BEB echo cancellation B-MAC
  - Subclause paragraph 6.10.1f) is extended to provide a parameter for the B-SA which is cancelled. This parameter is set to the associated root EP

802.1Qbh/BR Leverage

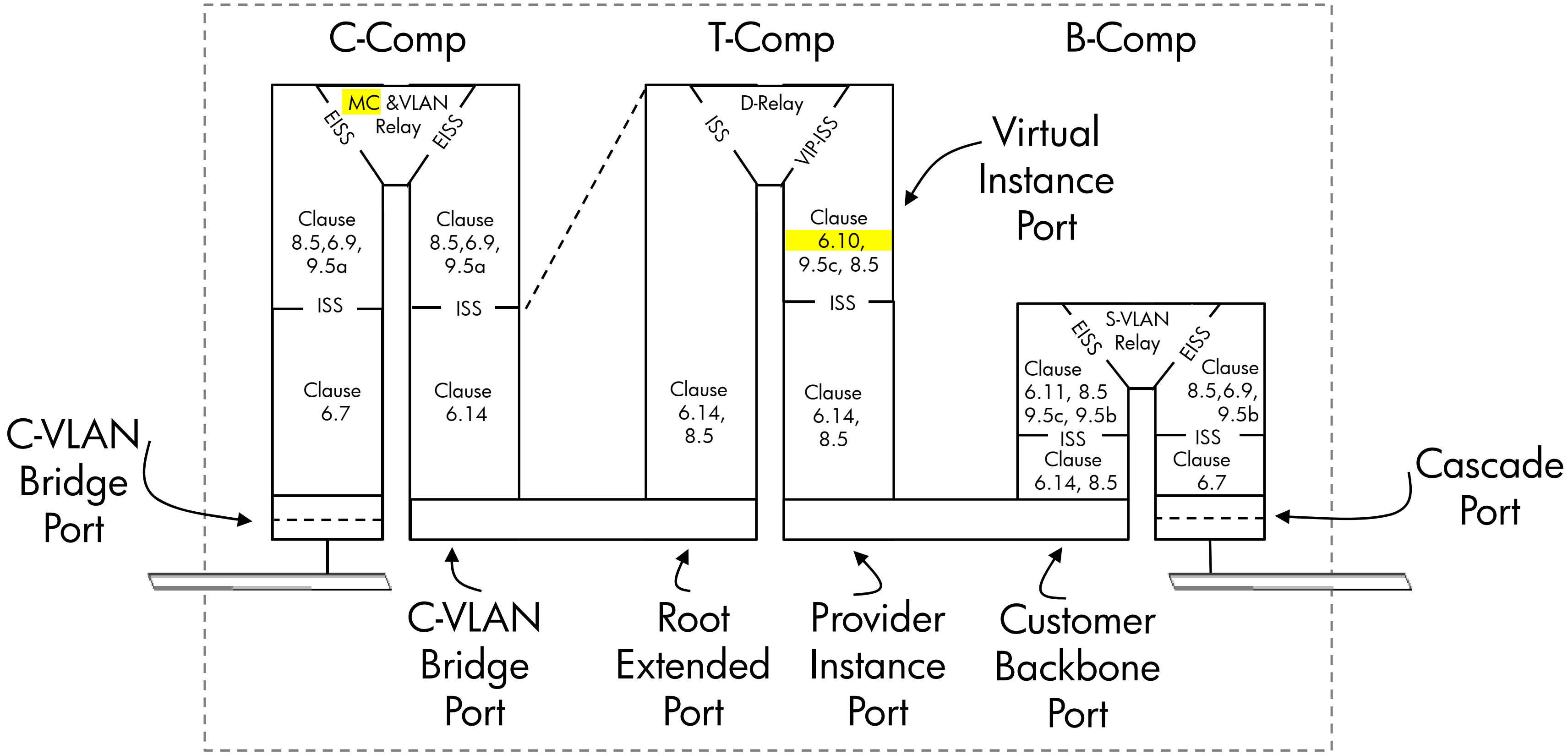
- Port Extender Control and Status Protocol from 802.1BR with perhaps some modest changes in the E-PID field definitions
- The managed object extensions for the Controlling Bridge MIB
- The PE LLDP extension TLVs for the Controlling Bridge and Port Extender

PBB Encoding Transform from E-TAG



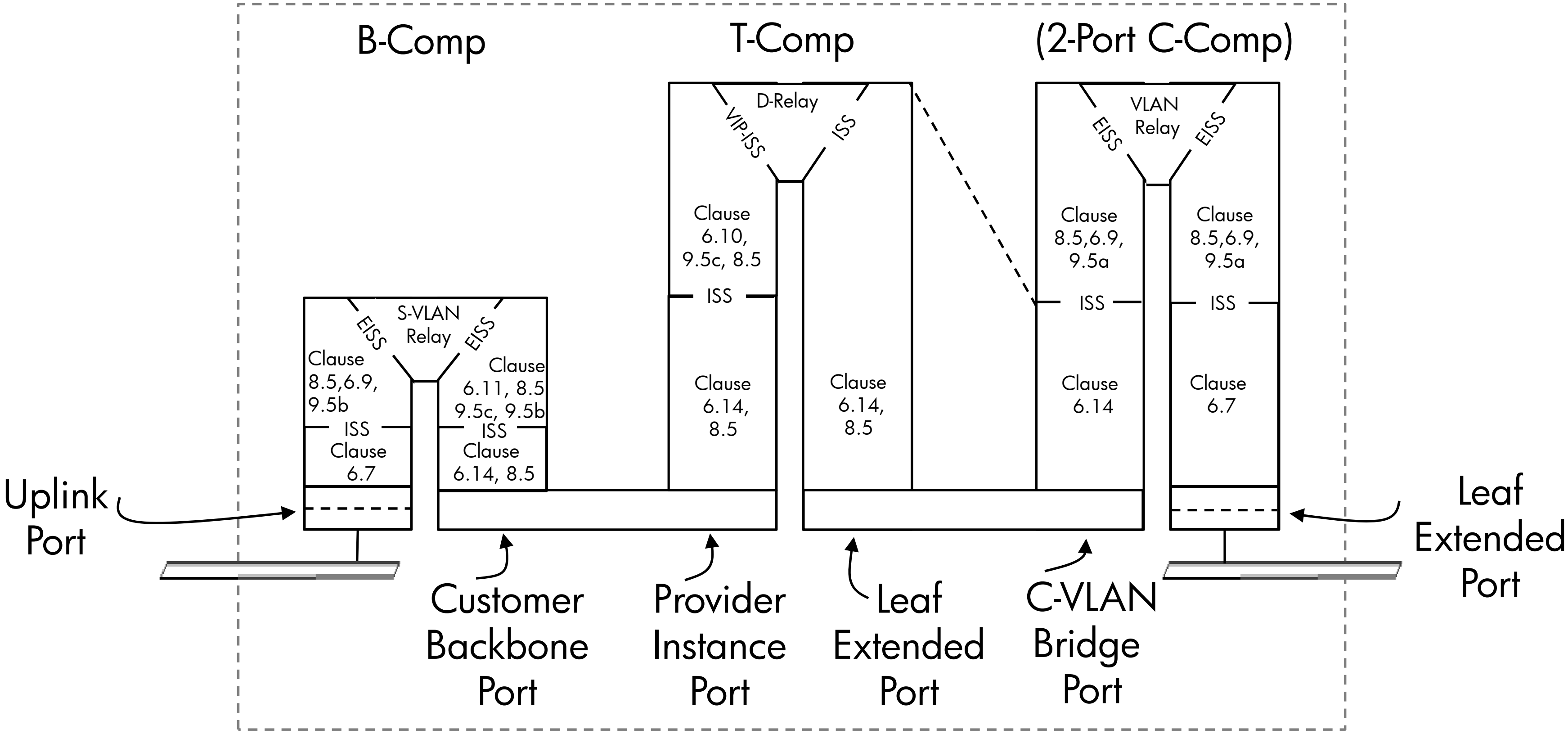
- The Ingress PE Port is identified by the B-SA rather than an Ingress_E-PID, while the PE Destination (group or unicast) is identified by the B-DA rather than an E-PID.
- The E-PCP and E-DEI are carried in I-PCP, I-DEI.
- The I-SID is not used for a PE application.

Controlling Bridge Baggy Pants Diagram



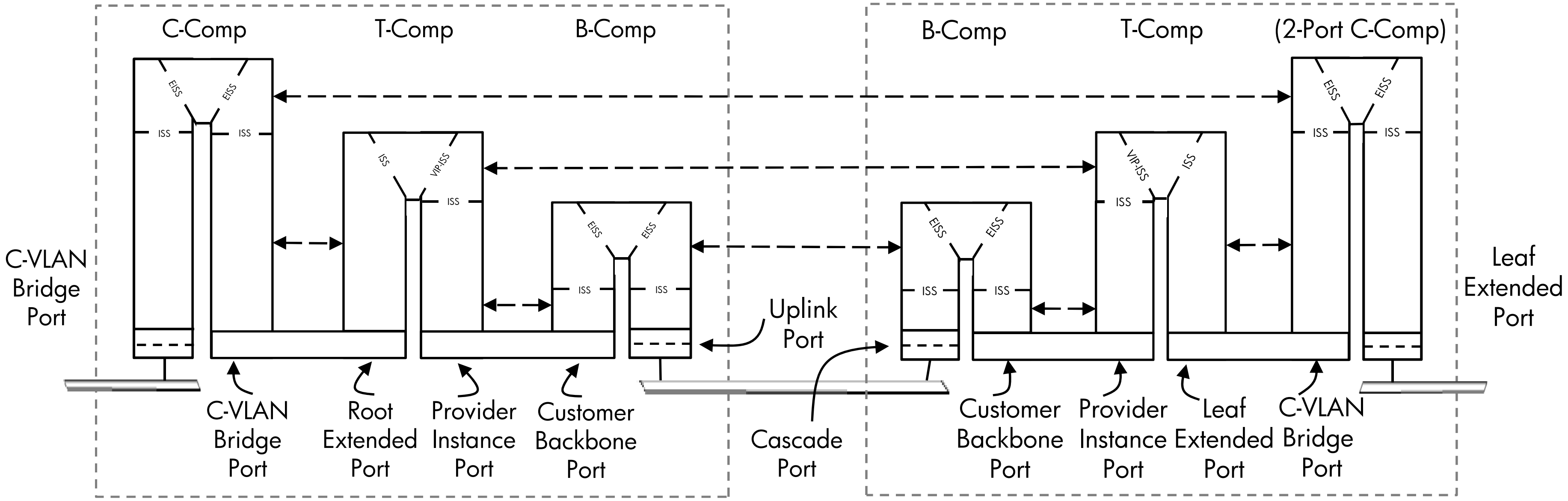
- No new relays, components, ports, or tags
- Yellow indicates subclauses requiring feature additions, other subclauses are unmodified

Port Extender Baggy Pants Diagram



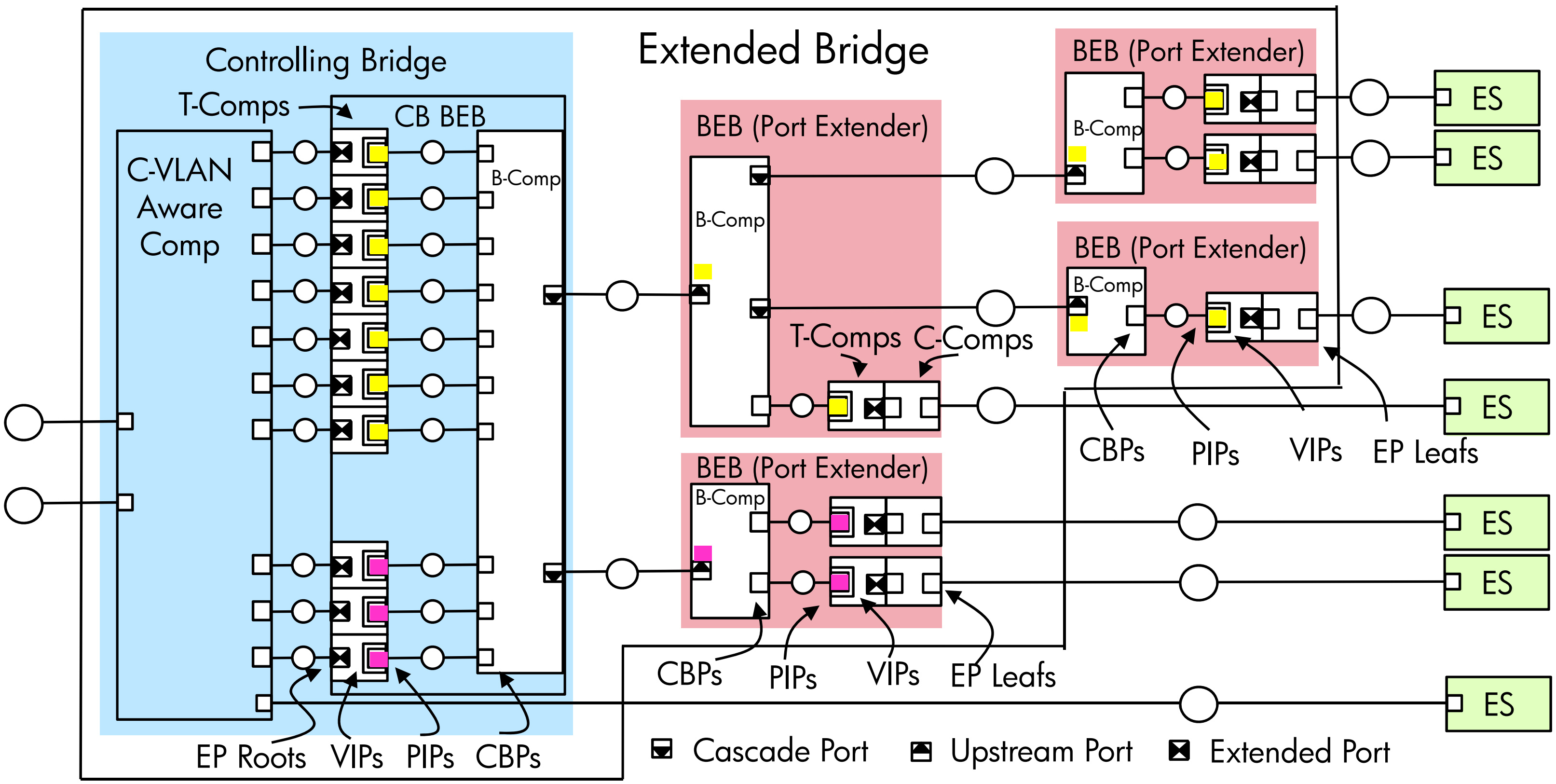
- Unmodified components form the PE relay
- One “real” filtering database at the B-Comp
- The optional 2-Port C-component allows C-tagging/untagging
- The control plane is replaced with the PE CSP

Extended Bridge component peering



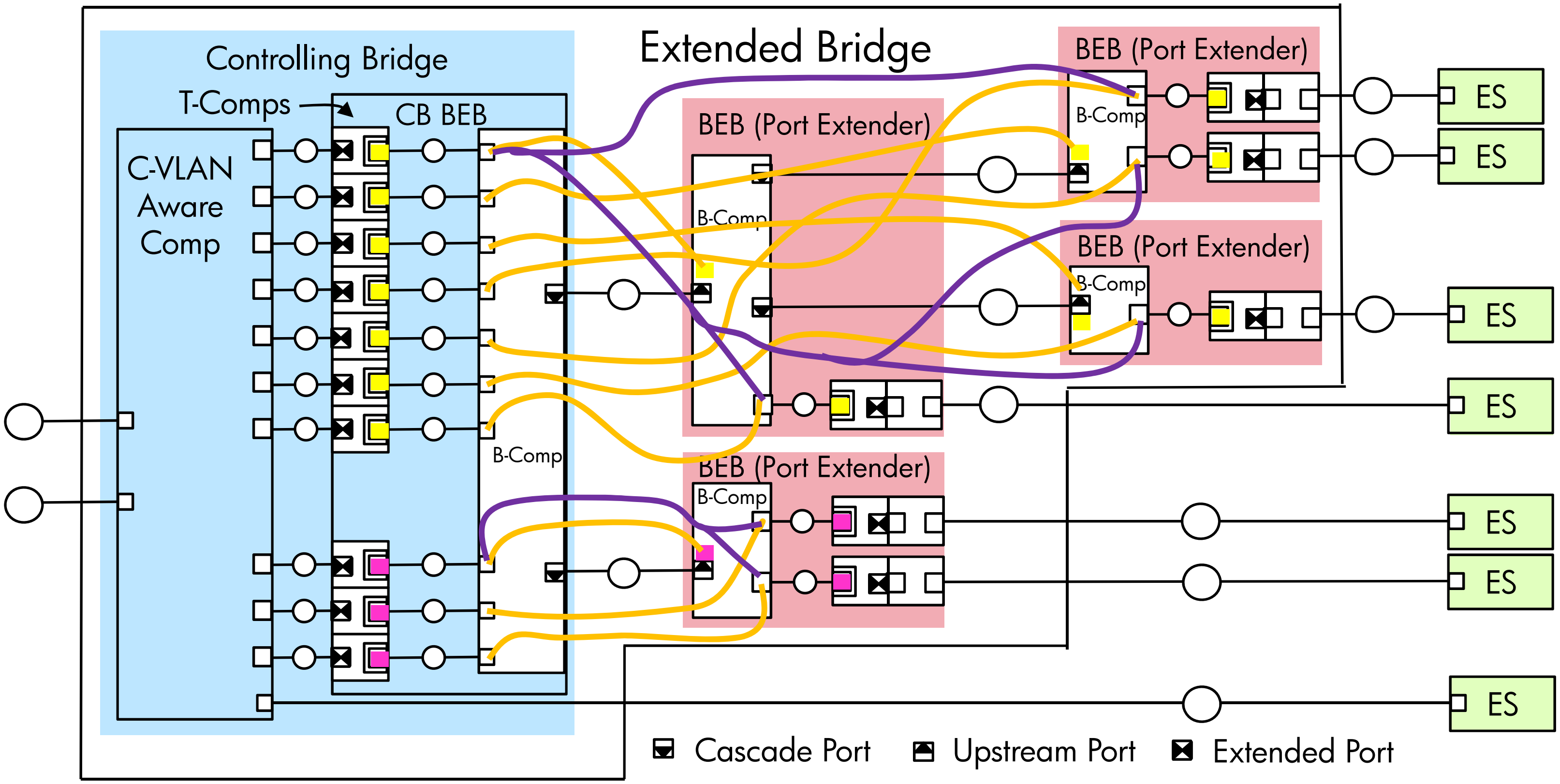
- VIPs in T-Components terminate Backbone Service Instance over Port Extender network
- C-Components in Controlling Bridge and Port Extender terminate port extensions

Extended Bridge BSIs
Backbone Service Instances



- All the VIPs of a connected PE "tree" are members of the same Backbone Service Instance (BSI) and therefore use the same I-SID value.
- In the example above we have two PE "trees" and each with a different I-SID value indicated by the yellow and pink marks
- Note that a VIP for BSI termination exists above the Uplink Port LLC layer

Extended Bridge TESIs
Traffic Engineered Service Instances



🟨 Tan lines in the diagram show the attachments of point-to-point TESIs

- One pt-pt TESI couples a Root EP's VIP to the Uplink Port's LLC on each Port Extender

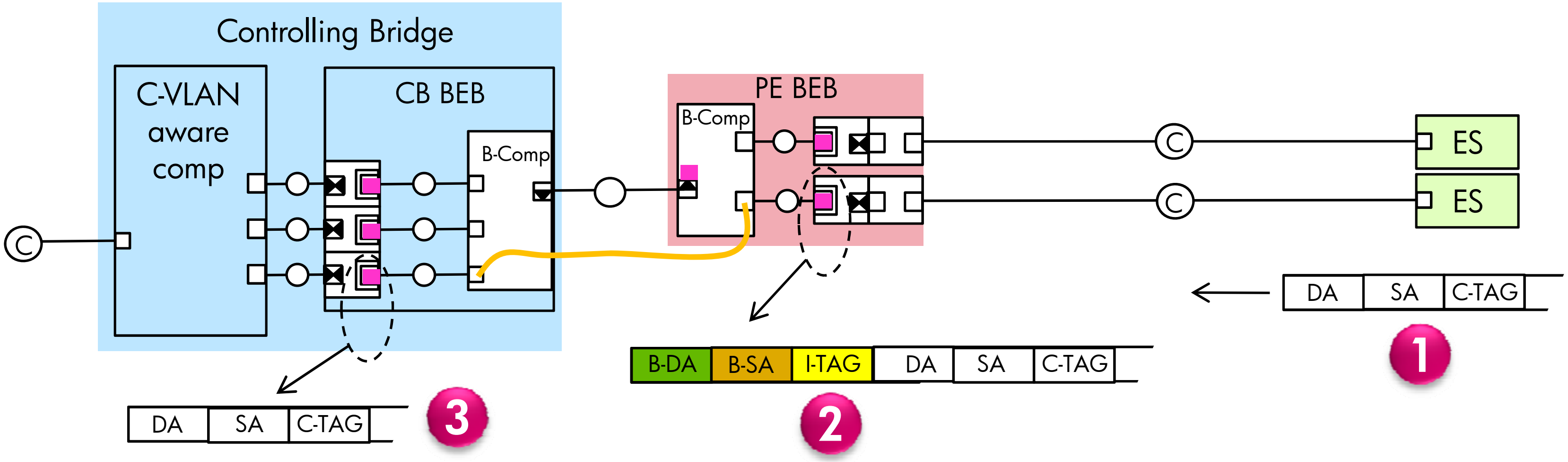
🟪 Purple lines indicate the attachments of pt-mpt TESIs within the Port Extender "trees"

- Though a single pt-mpt TESI attaching a Root EP's VIP to all Leaf EPs VIPs of the PE "tree" is shown, additional pt-mpt TESIs attaching to limited groups of Leaf EPs are possible

Port Extension B-VIDs
Path Selection

- Without redundant links the Port Extender can use a single default B-VID
- By using multiple B-VIDs to engineer alternate ESPs it would be possible to support extended features
  - The B-VID can be used to enhance the Port Extenders with protection support
  - The B-VID can be used to enhance the Port Extenders with multi-pathing support

Frame forwarding from Leaf EP



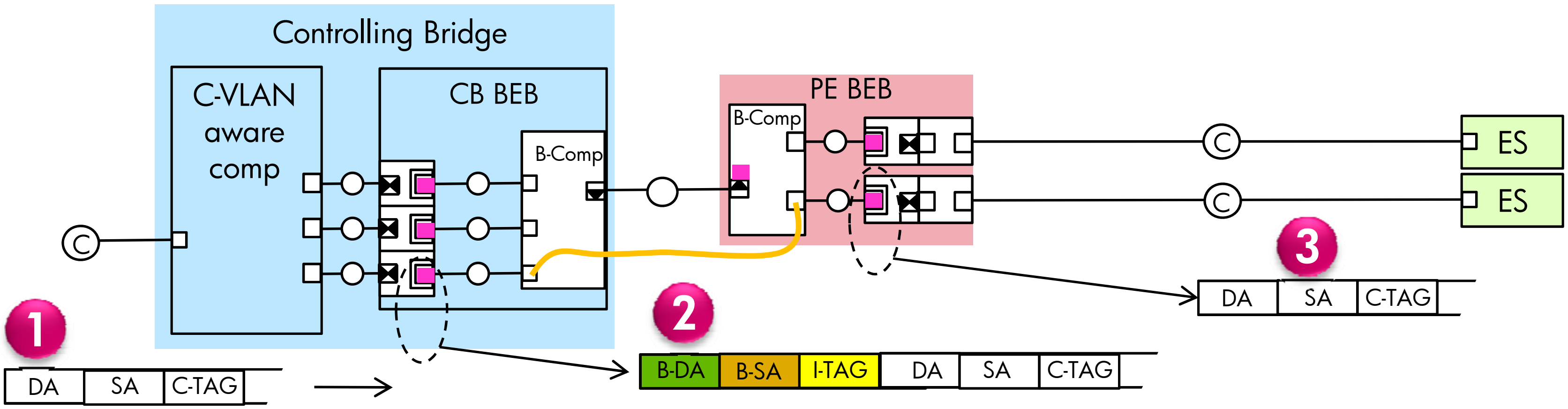
- Before frame transmission the PIP of the T-component is programmed using PE CSP with:
  - It's SA as a leaf EP address constructed from the E-PID
  - It's Default Backbone Destination parameter set to a root EP address constructed from the E-PID
  - The enableConnectionIdentifier parameter is set to FALSE
  - The I-SID parameter is set to default value

1 A frame is transmitted from the ES attached to an Extended Port with DA/SA/C-TAG

2 The frame is received at a leaf PE of a T-component within the Port Extender who delivers it over the VIP-ISS to the PIP. The PIP builds a frame with B-DA = root EP and B-SA = leaf EP sending it to the CBP of the B-Comp who forwards it along the TESI

3 The frame is de-encapsulated at the PIP of the T-component within the CB-BEB and delivered over the internal LAN to an internal port of the C-VLAN aware component

Frame forwarding from the root EP Individual B-DAs



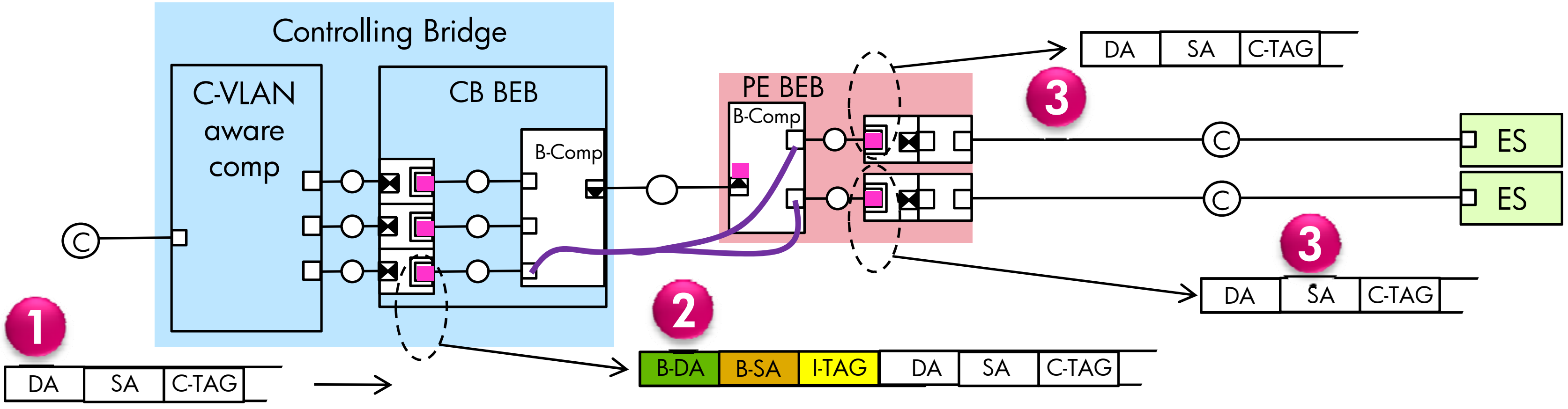
- Before frame transmission the PIP of the CB-BEB T-component is programmed with:
  - It's SA as a root EP address constructed from the E-PID
  - It's Default Backbone Destination parameter set to a leaf EP address constructed from the E-PID
  - The I-SID parameter is set to identify the PE "Tree"

1 A frame is sent from a C-Comp Port to a root EP of the CB-BEB with DA/SA/C-TAG

2 The frame is received at a root EP of the T-component within the CB-BEB and delivered over the VIP-ISS to the PIP. The PIP builds a frame with B-DA = leaf EP and B-SA = root EP sending it to the CBP of the B-Comp who forwards it along the TESI

3 The frame is de-encapsulated at the PIP of a T-component of an EP and delivered to a LAN

Frame forwarding from the root EP Group B-DAs no Echo Cancellation



- Before frame transmission the PIP of the CB-BEB T-component is programmed:
  - Is programmed as in the Individual address case
  - The T-component supports passing a connection_identifier containing a destination port map
  - The PIP is modified to use the connection_identifier to select a B-DA using the destination port map

1 A frame is sent from a C-Comp Port to a root EP of the CB-BEB with DA/SA/C-TAG

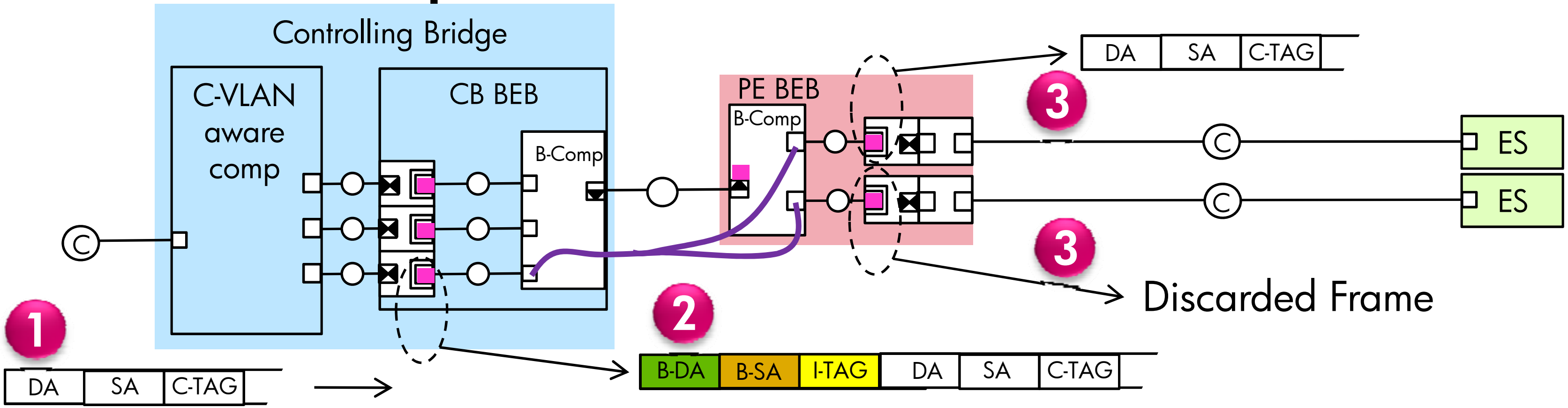
- The frame was sent from outside the "replication group" and so the connection_identifier contains a destination port map which includes the CB-BEB PIP used to forward the frame (only a single request is sent to the "replication group").

2 The frame is received at a root EP of the T-component within the CB-BEB and delivered to the PIP. The PIP builds a frame and sends it to the CBP of the B-Comp who forwards it along a TESI

- B-DA is selected based on the connection_identifier destination port map
- B-SA = root EP B-MAC without Echo Cancellation (since the source is outside the replication group)

3 The frame is replicated over the TESI and de-encapsulated at the PIPs of the T-components, delivered to the leaf EPs and then the attached LANs

Frame forwarding from the root EP Group B-DAs with Echo Cancellation



- The PIP of the CB-BEB T-component is programmed:
  - Is programmed as in the Individual address case
  - The T-component supports passing a connection_identifier containing a destination port map
  - The PIP is modified to use the connection_identifier to select a B-DA using the destination port map
  - The PIP is modified to use the connection_identifier to select the B-SA using both the destination port map and source port
- The PIP of all Leaf EPs are modified to filter out frames matching a B-SA filter parameter (6.10f)
  - Each root EP has two B-MACs one echo cancelled and one not. The B-SA filter parameter of the each leaf EP PIP is set to the echo cancelled B-MAC of it's root EP

1 A frame is sent from a C-Comp Port to a root EP of the CB-BEB with DA/SA/C-TAG

- The frame was sent from within the "replication group" and so the connection_identifier contains a destination port map which excludes the CB-BEB PIP used to forward the frame (only a single request is sent to the "replication group").

2 The frame is received at a root EP of the T-component within the CB-BEB and delivered to the PIP. The PIP builds a frame and sends it to the CBP of the B-Comp who forwards it along a TESI

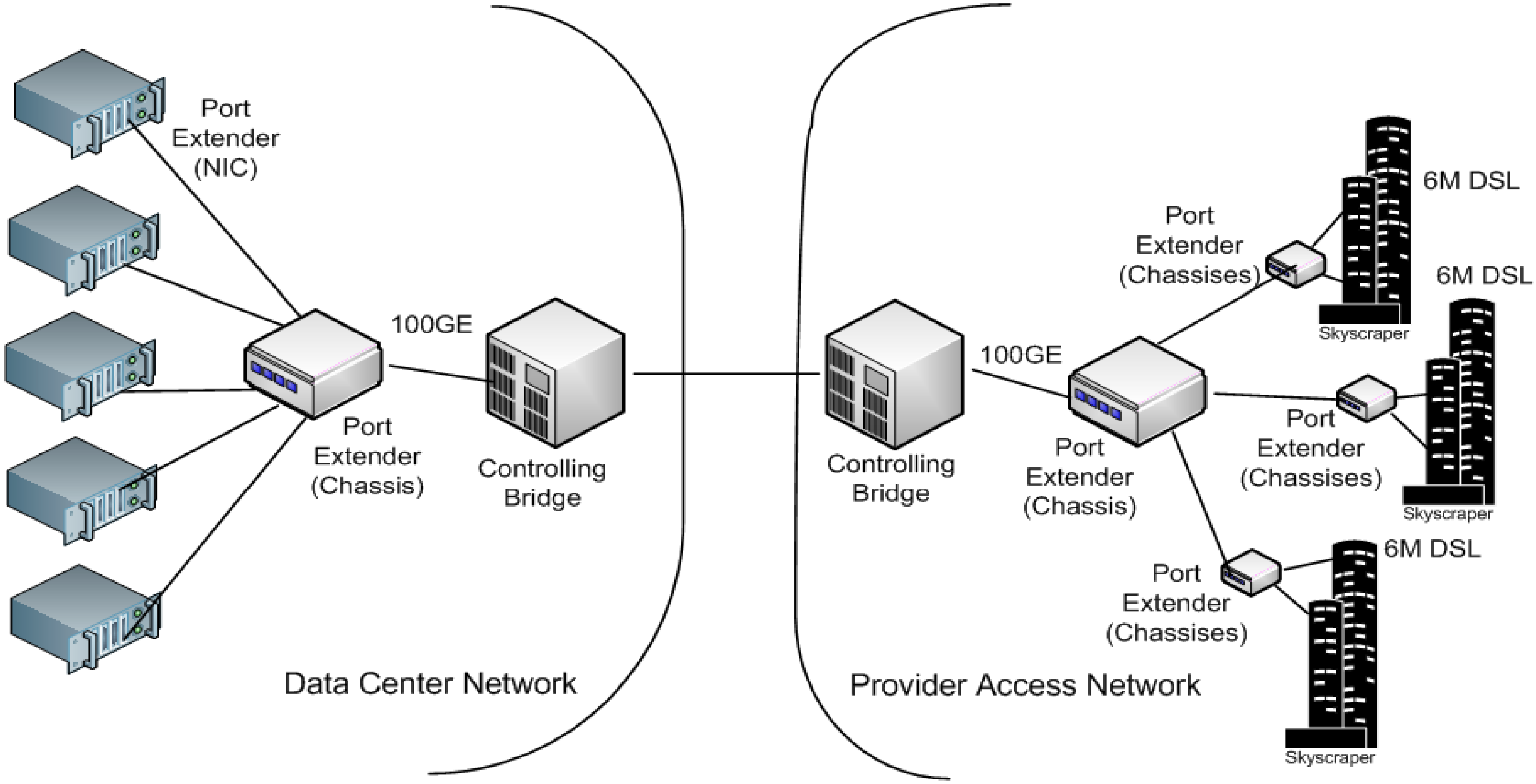
- B-DA is selected based on the connection_identifier destination port map
- B-SA = root EP B-MAC with Echo Cancellation of the source port from the connection_identifier (should be this root EP port)

3 The frame is de-encapsulated at the PIPs of the T-components of the PEs and delivered to the LANs which are not echo cancelled.

What needs to be specified

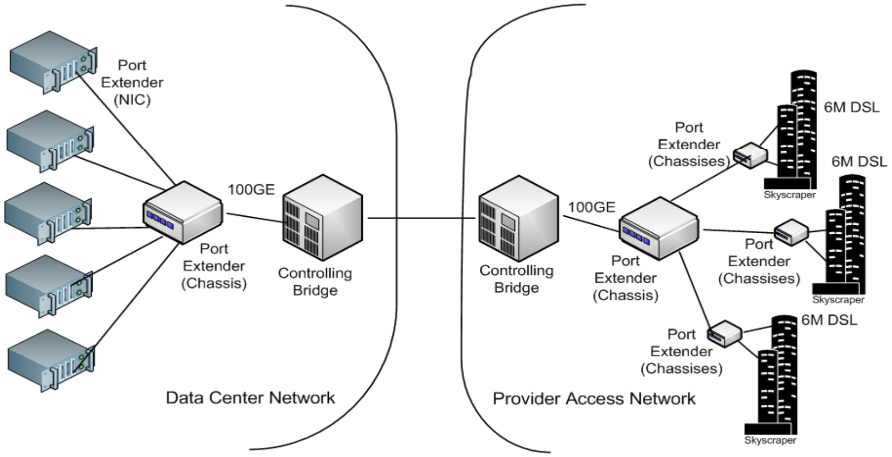
- Contribution bh-bottorff-pbbte-pe-draft-0711-v1.pdf provides a complete proposed draft (or course needs review)
- Move clause 8, 7.12-7.14 (PE CSP) of 802.1BR into a new 802.1Qbh clause 45 using 7.12-7.14 as part of the protocol introduction.
- Port Extender can be defined by a new conformance subclause specifying a Port Extender as a specific type of BEB and including the PE CSP
- The Controlling Bridge can be defined using the current conformance statement from 802.1Qbh replacing the PE requirements with the requirements defining a CB-BEB

Scaling: EtherSlam application with up to 16,666 Extended Ports per PE tree



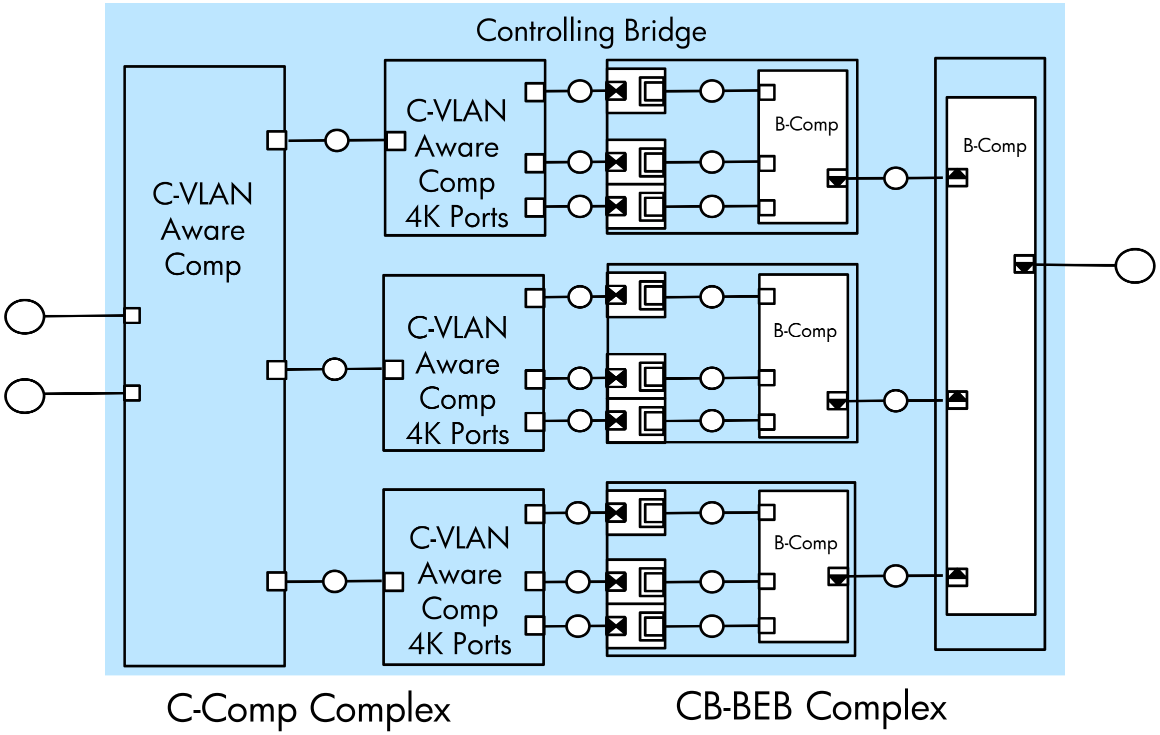
- Each 100 GE Cascade Port may support up to 16,666 Extended Ports at 6 Mbit each (100G/6M)

16K EP load refinements



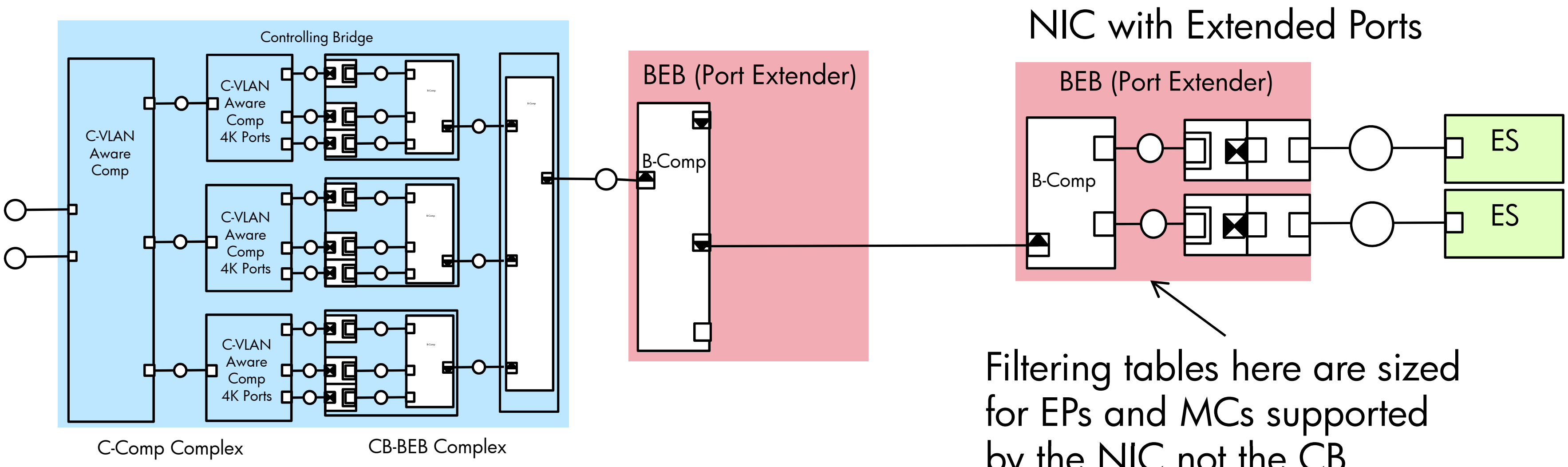
- In the data center each external frame typically would generate 10 DCN frames (10x expansion)
- However the aggregate throughput from the 6 Mbit DSL lines will be 1/100 or less the line rate giving a total aggregate bandwidth of 1 G rather than 100 G
- 16K VM interfaced through the single 100 GE link would then run at about 10% utilization giving headroom for bursting

Scaling a PBB-TE Controlling Bridge



- To Scale a PBB-TE Controlling Bridge we simply add stages
- E-channels can be identified by the pair <B-DA,B-SA>
  - Total number of filtering table entries per CB Cascade port is 2 x Number of Extended Ports + Number of Group Destinations
  - For example if we have 16K Extended Ports and 16K Group Addresses then we have 2 x 16K + 16K = 36K filtering table entries

PBB-TE filtering database allows the NIC state table to be independent from the CB



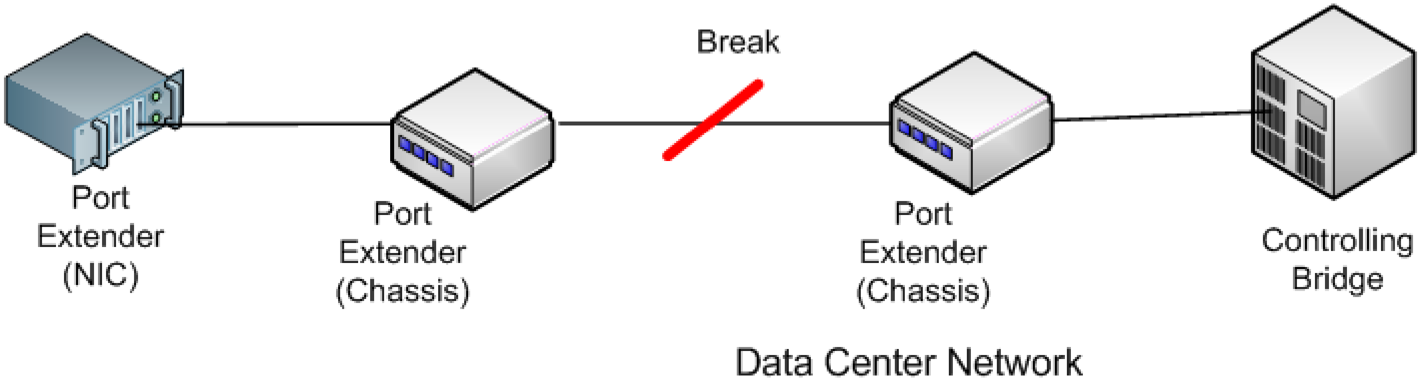
Filtering tables here are sized for EPs and MCs supported by the NIC not the CB

- Filtering tables at the edges of the PE network need to be large enough to hold the E-channels actually passing through
- For instance a NIC with 128 ports and 128 group requires 2*128 + 128 = 384 filtering table entries (note if additional multicast sources pass through NIC then these also need filtering entries)
- The NIC filter table size requirement is independent from the CB filter table size requirement

PBB-TE PEs easily scale to 16K Extended Ports while minimizing state and table size

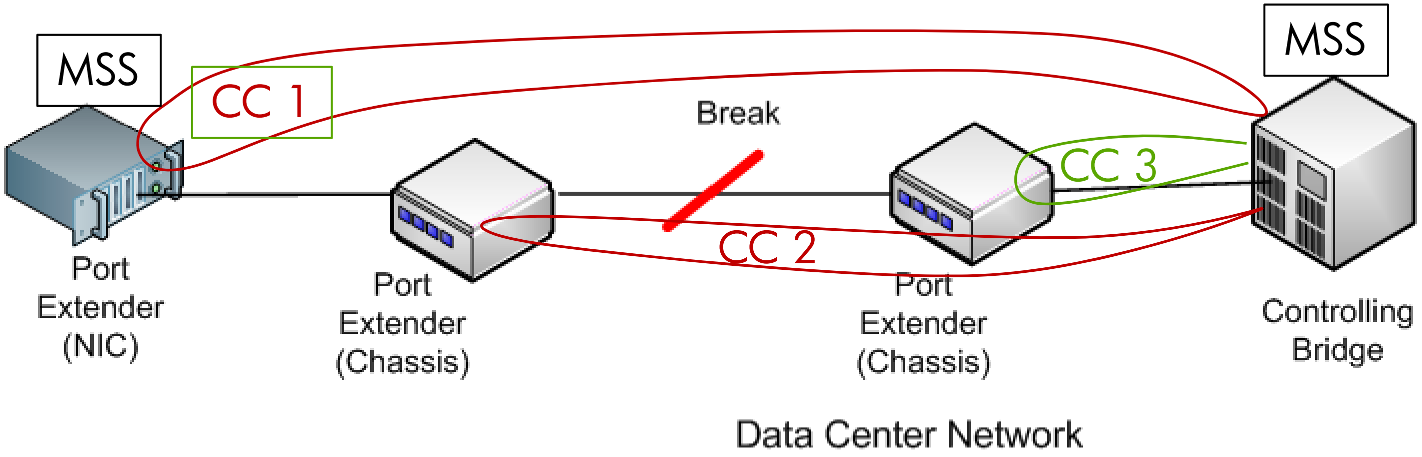
- PBB-TE PE uses existing filtering DB tables without any size increases
  - For Controlling Bridge10K-100K filtering entries are common and sufficient
    - For a CB-BEB supporting 16K Extended Ports we would need a total of 36K filtering entries providing:
      - Source and Destination for each Extended Port
      - 16K group addresses
    - Allows component cascading for Controlling Bridge port expansion
  - For Port Extenders we don't need the as many filtering entries
    - In an adapter we need two entries for each Extended Port plus the number required for multicast
    - NIC filter table size is independent from CB filter table size
- Number of Extended Ports is limited by E-TAG Ingress_E-CID and by the E-CID table size
  - Requires new tables for switches and chips
  - Changes in proposed E-TAG size to support 16K Extended Ports
  - Both Port Extenders and Controlling Bridge must support full sized tables

MSP: MAC operational propagation both up and down from a break



- A break between two port extenders should be reflected in MAC operational status at both the Controlling Bridge and Network Interface Port Extenders
- PE CSP has no connectivity to the station Port Extenders during a break and so can’t control the MAC operational status from the Controlling Bridge

One CC MEP pair per Port Extender

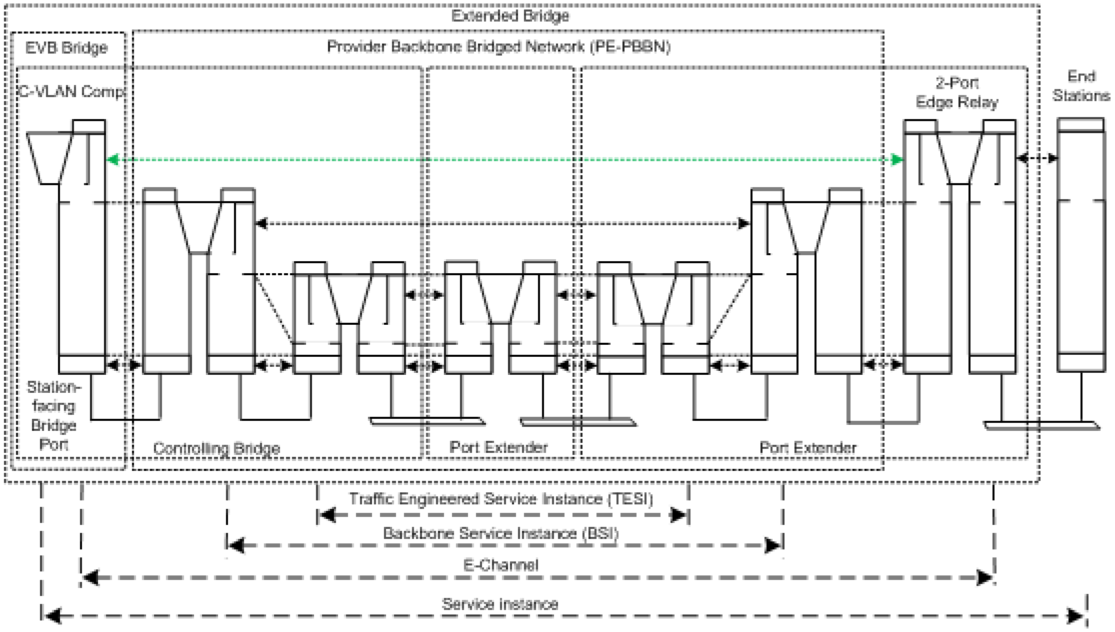


- Here the break is detected by CFM using CC flows between the Controlling Bridge and Port Extenders
- Both the Controlling Bridge and all affected Port Extenders receive indications form CFM
- Once a break is detected by CFM the MSS on both ends can be used to set MAC_operational status on individual affected ports

Mid-span failure detection and reporting solved using CFM and MSP

- CFM used to detect mid-span breaks to PEs
  - Required since we don't have RSTP or SPB
  - Run CCs over the control E-channel to each Port Extender
  - Any failure will be reported both to the CB and to the Port Extender affected
  - The CB will see all mid-span failures
- Each Port Extender can set MAC enable based on the connectivity state to the CB
- The MSP protocol co-ordinates MAC_operational state between each external Extended Port and each internal Extended Port

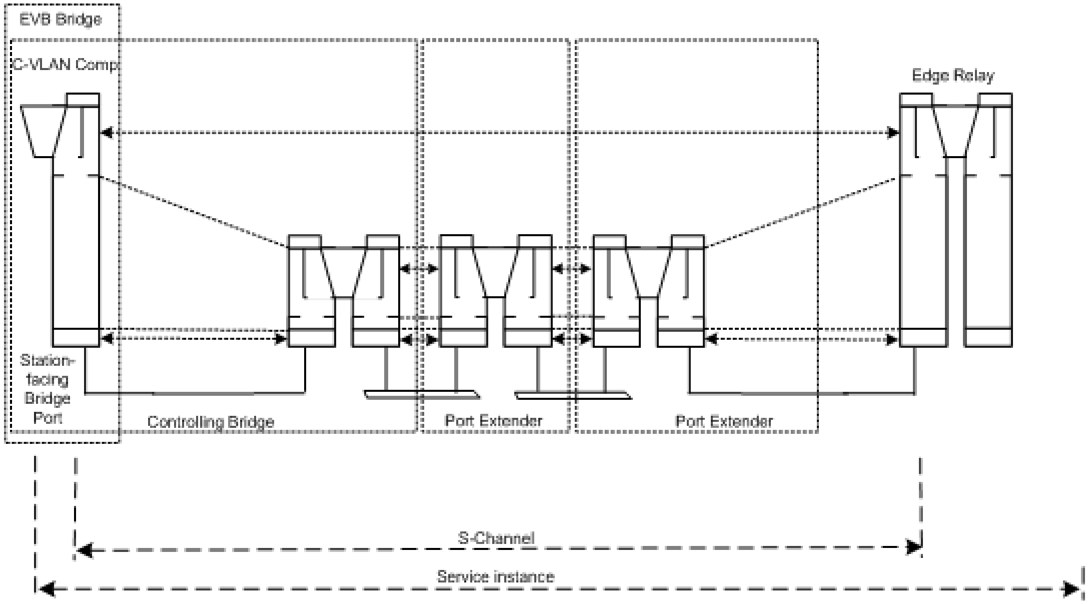
PBB-TE PE and EVB synergy integrated Edge Relay



Extended Bridge
EVB Bridge
Provider Backbone Bridged Network (PE-PBBN)
C-VLAN Comp
2-Port Edge Relay
End Stations
Station-facing Bridge Port
Controlling Bridge
Port Extender
Port Extender
Traffic Engineered Service Instance (TESI)
Backbone Service Instance (BSI)
E-Channel
Service instance

- PE-PBBN provides transparent extension using T-components between an EVB Bridge and a 2-Port Edge Relay

PBB-TE PE and EVB synergy S-channel compatibility



- Each PBB-TE PE B-component is an S-VLAN component
- All S-VLANs are available except the one used for Port Extension
- S-channel service couples the B-comps direct to C-VLAN comp and ER
- Configuration of S-channels is easily automated using the existing LLDP exchanges

PE CSP for PBB-TE (a fringe benefit)

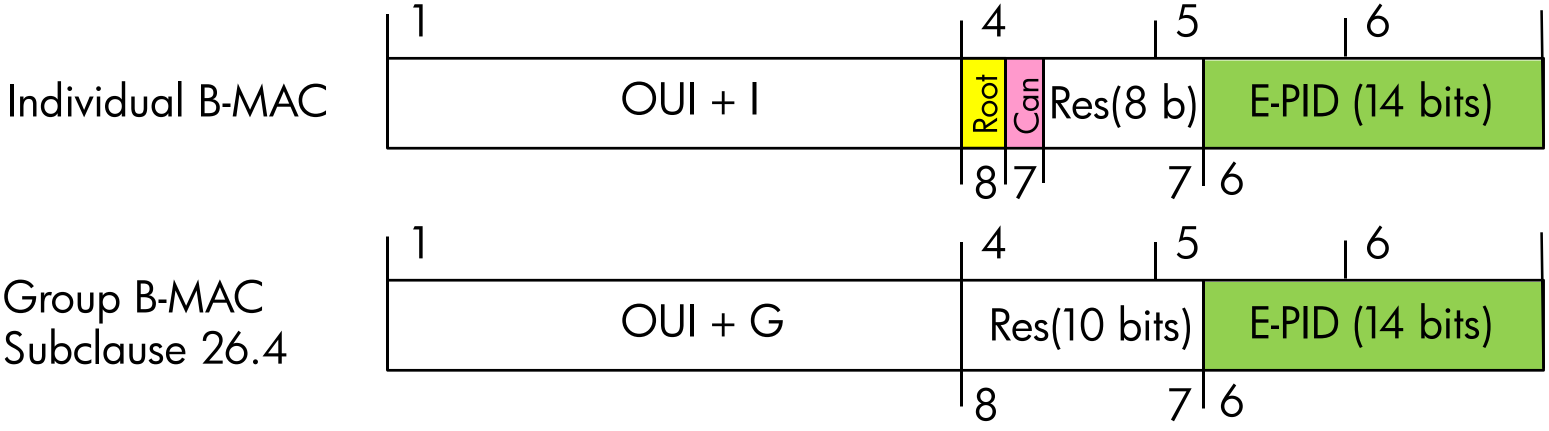
- The PE CSP protocol could be expanded as a control protocol for provisioning PBB-TE networks
- To do this it would be desirable to expand PE CSP to support generalized TESID programming
- This would provide broader utility for the protocol

Comparison between E-TAG and PBB-TE Port Extenders

	E-TAG PEs	PBB-TE PEs
Scalability		★
Failure detection and status reporting		★
EVB synergy	☆	★
Compatibility with existing 802.1Q Bridge relay		★
No new components		★
No new tags		★
Optional support for CFM, protection and multipathing		★
Optional support for congestion notification	☆	★
Optional support for ETS and PFC	★	★
Optional support for EVB & VEPA	★	★
Lowest overhead octets	★	

BACKUP SLIDES

Constructed B-MACs



- Globally assigned B-MACs also could be used by simply increasing the E-PID size to a full TESI address.
- Constructed individual B-MACs use the Root indicator to differentiate between the CB-PIPs and the PE-PIPs
- Constructed addresses use the Can indicator to differentiate frames which can be echo cancelled and those which can not
- Constructed group B-MACs could use the Backbone Service Instance Group Address OUI
- Since the Controlling Bridge is co-ordinating the selection of E-PIDs the assignments would be locally unique
- Since the B-MACs don't extend beyond a single PE mesh they would never interact with a general purpose system