PBBN in the Data Center

IEEE 802.1 May 2009 Interim Meeting
Pittsburgh, PA., USA
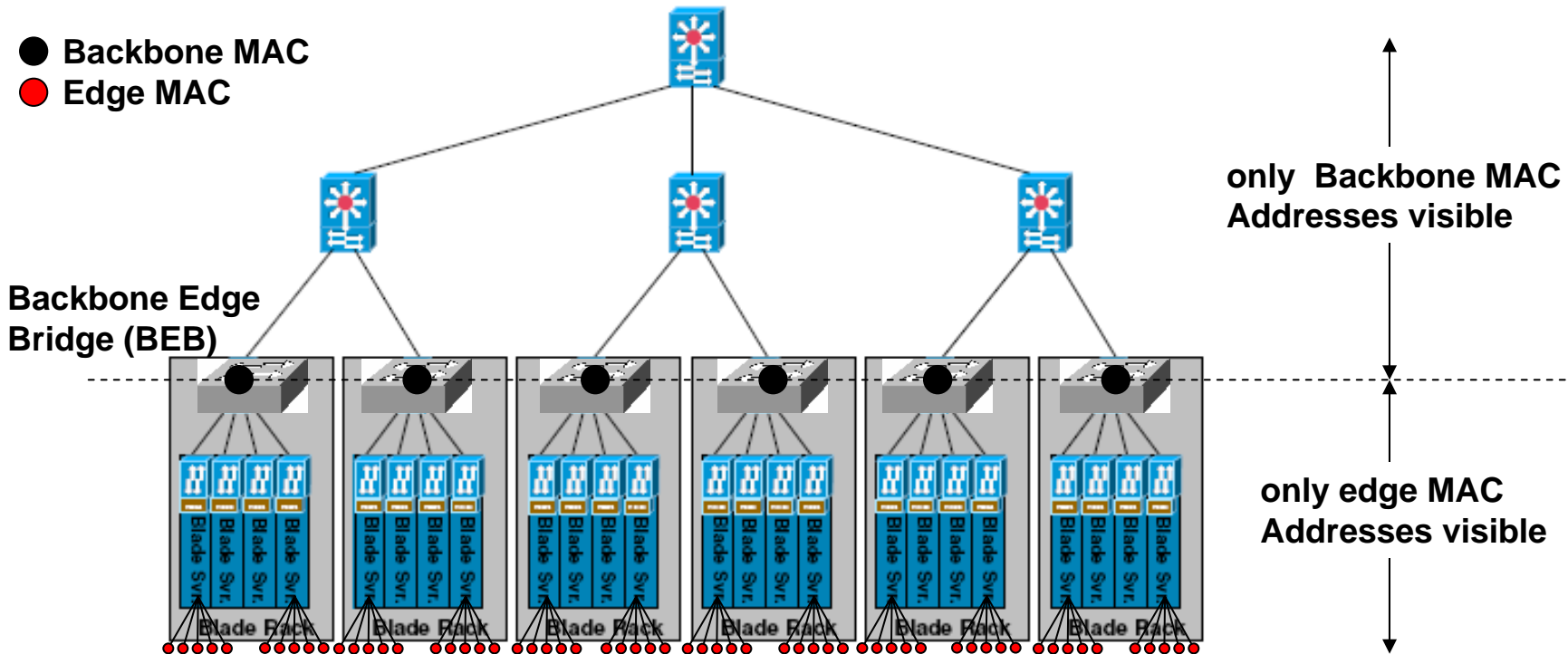
Bob Sultan

Ben Mack-Crane

Growing number of MACs in Data Center

- *Large* numbers of servers with *multiple* Virtual NICs (VNICs) per Physical NIC (PNIC);
- Bridges in core likely to see largest numbers of MAC addresses.
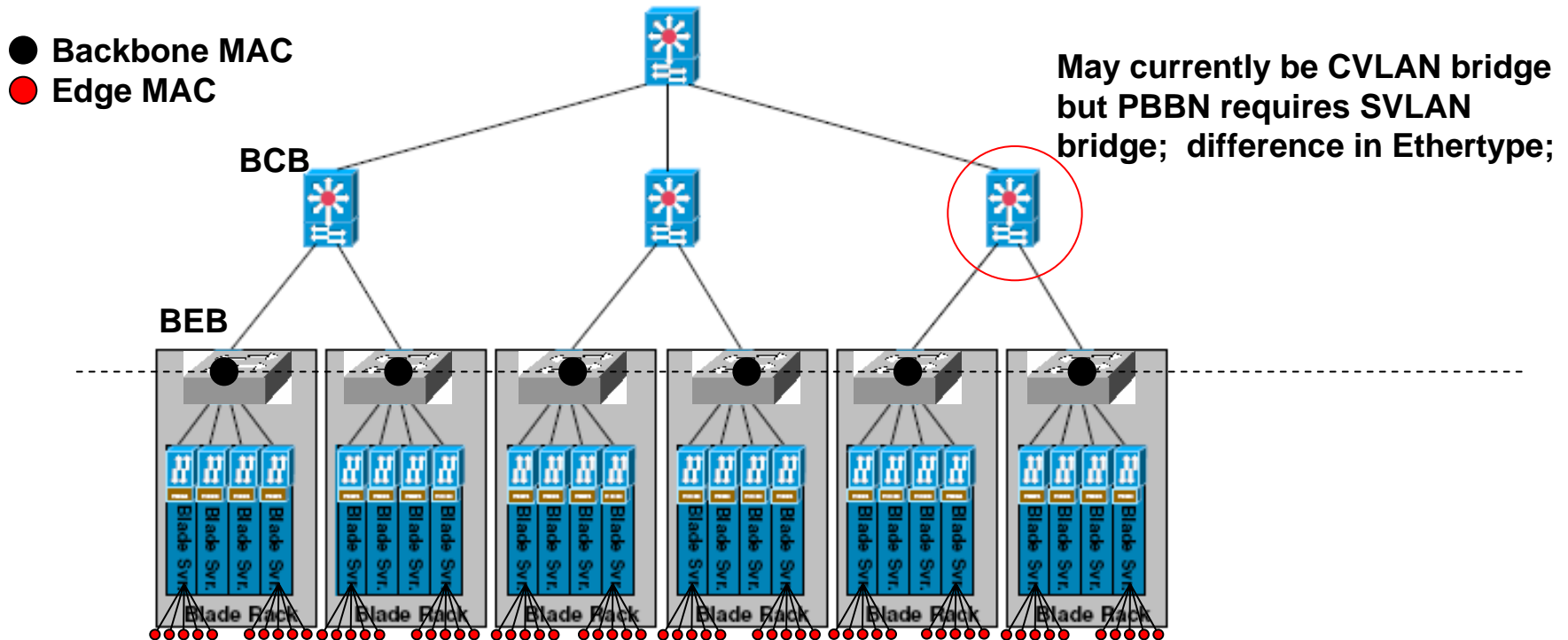
More MACs require more FDB entries

- Typical bridges support in the range of 16 to 32K MAC Addresses, with some supporting 64K;
- With external memory (TCAM), bridges can support up to 512K to 1M MAC addresses;
- Failure to support sufficient MAC addresses results in increased flooding / poor performance;
- Operators may need to upgrade to higher capacity (ie., more FDB entries) bridges *OR*
- If high-capacity bridges are already in use (e.g., in the core), performance may be compromised by excessive flooding.

PBBN can provide relief

- Backbone MAC
- Edge MAC

Backbone Edge Bridge (BEB)

only Backbone MAC Addresses visible
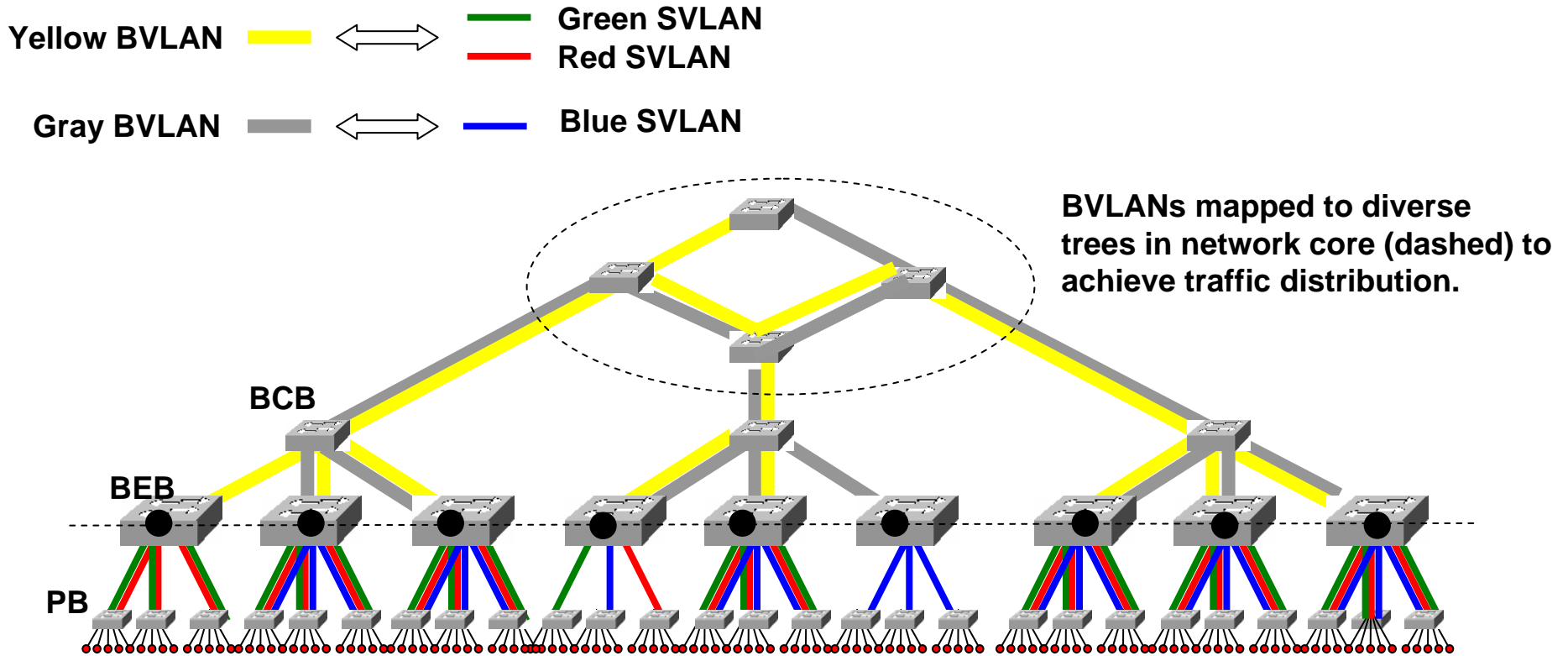
only edge MAC Addresses visible

- Eg., top-of-rack deploys Backbone Edge Bridge;
- Others bridges within rack see fraction of large edge address space;
- External switches see smaller BB address space.

But, it's *not* a perfect fit

● Backbone MAC
● Edge MAC

May currently be CVLAN bridge but PBBN requires SVLAN bridge; difference in Ethertype;
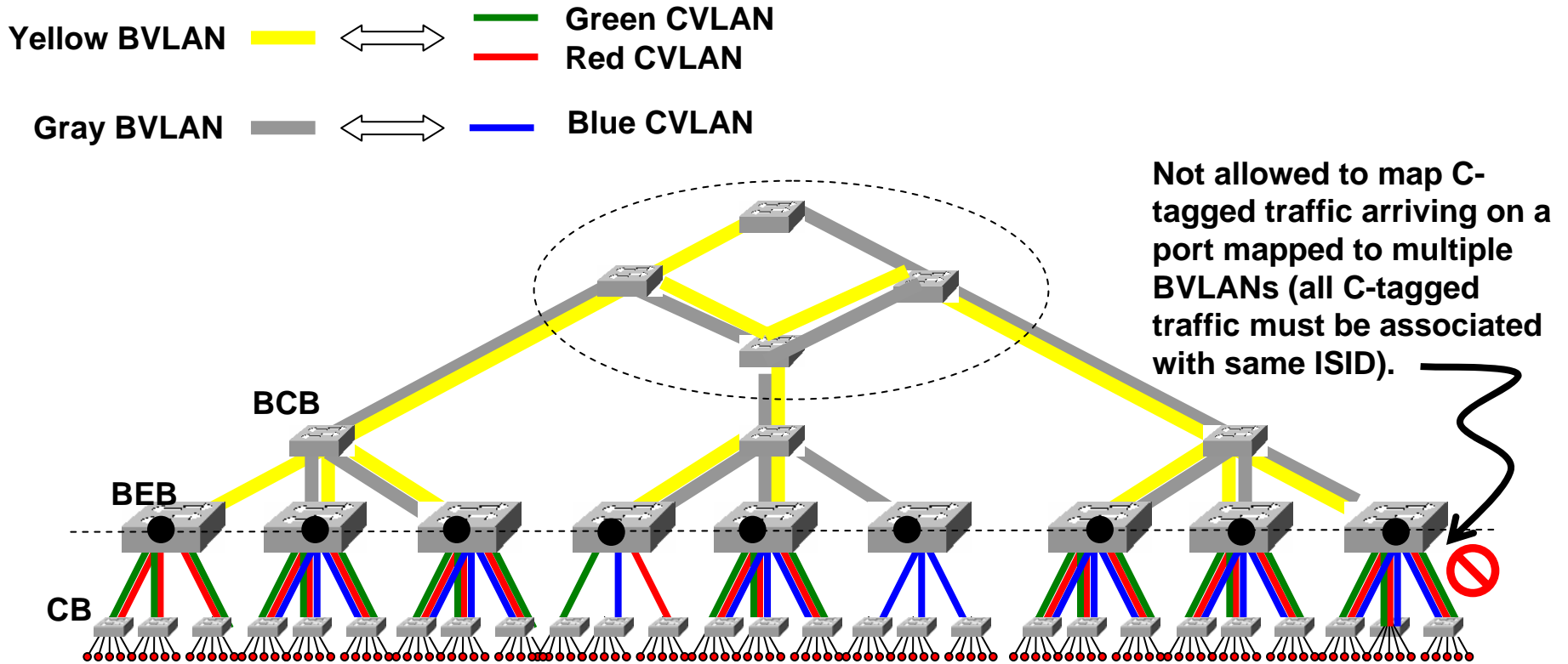
BCB

BEB

- PBBN BCB is identical to a Provider Bridge;
- But DCB core may be *CVLAN* bridges;
- Not a problem if core bridges .1ad capable.

PBBN intended to interconnect *PBN*s

Yellow BVLAN — ⟺ — Green SVLAN
                    — Red SVLAN

Gray BVLAN — ⟺ — Blue SVLAN

BVLANs mapped to diverse trees in network core (dashed) to achieve traffic distribution.

BCB

BEB

PB

- No problem if access to backbone is PB;
- BVLANs can be used to distribute load.

Problem if BB access is by cust Bridge

Yellow BVLAN ⟷   Green CVLAN
                          Red CVLAN

Gray BVLAN ⟷   Blue CVLAN

Not allowed to map C-tagged traffic arriving on a port mapped to multiple BVLANs (all C-tagged traffic must be associated with same ISID).

BCB

BEB

CB

- C-tagged (or untagged) traffic must access BB via Port-based interface;
- Cannot distribute port traffic to multiple BVLANs.

May limit route diversity in DC core

Gray BVLAN

Green CVLAN
Red CVLAN
Blue CVLAN

BCB

BEB

CB
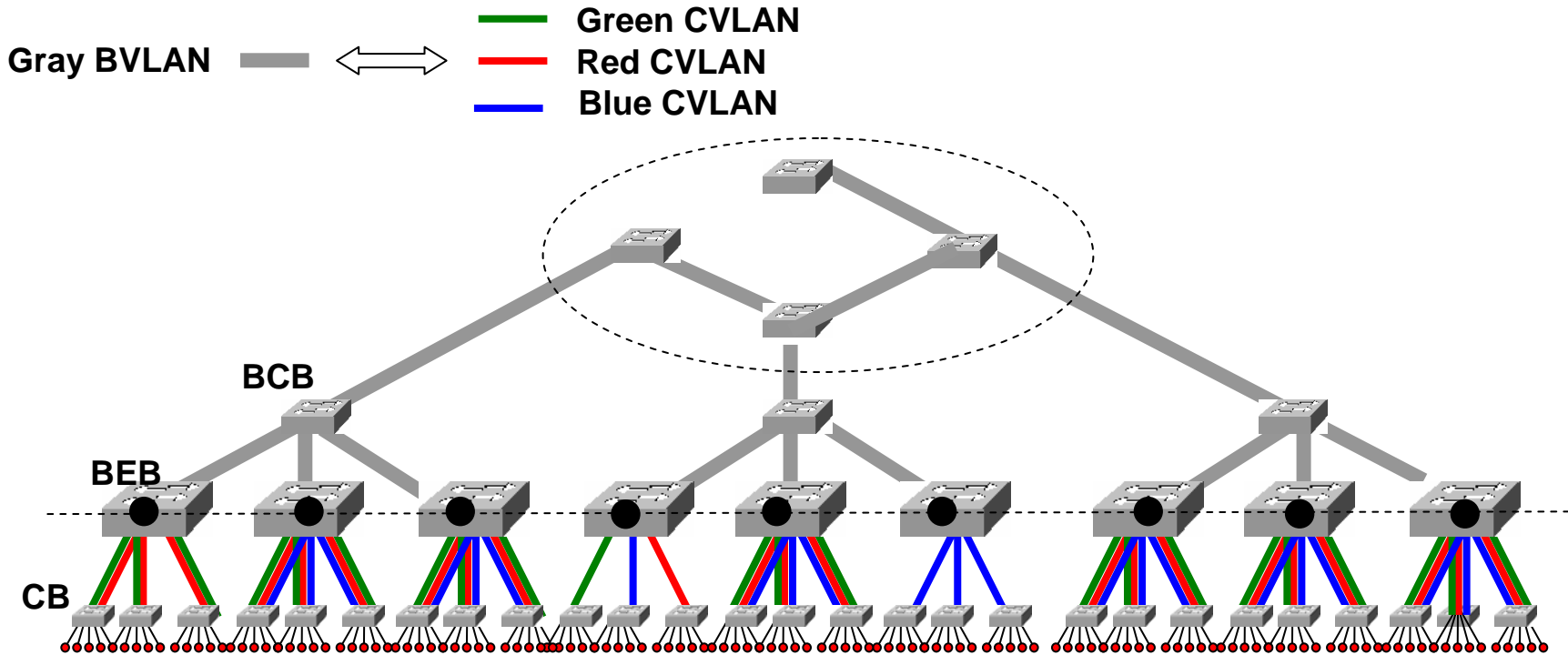
- If we want to allow any port to participate in any CVLAN then we are limited to one BVLAN;
- But then we lose route diversity provided by multiple BVLANs.

Possible solution

- Amend PBB to support a C-tagged interface;
- Would allow C-tagged traffic to be processed by BEB in much the same way as S-tagged traffic.

Data center vs. Provider Network

Data center admin is likely to assign same 'VID' (say '6') to Red VLAN in *all* locations.

BCB

BEB
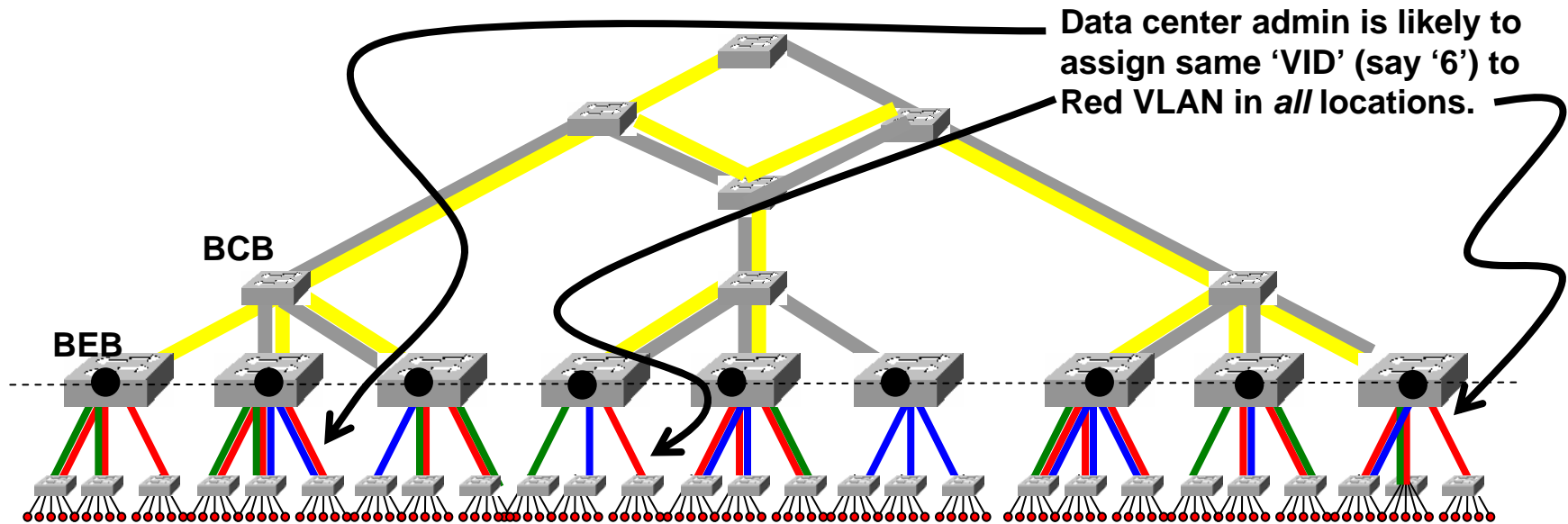
- PBNs interconnected by PBBN, in general, use different VID to identify the same 'customer service instance';
- Data Center admin, in general, does not assign different VID to same VLAN in different racks.

So BVLAN can be *same* as cust VLAN
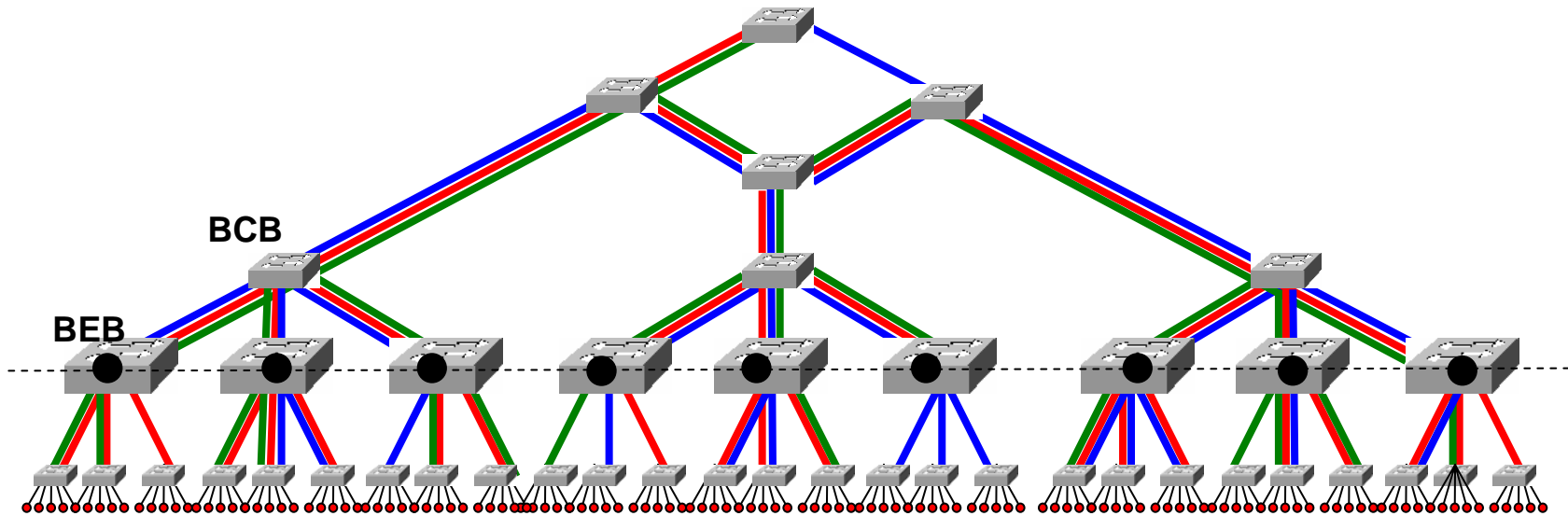
- Thus, in the Data Center it may be desirable to allow customer VLANs to extend *through* the backbone;
- Advantage: The Data Center admin does not have to alter the VLAN structure at all when using PBB to reduce MAC visibility in the core;

Desirable to allow STP to cross BB edge
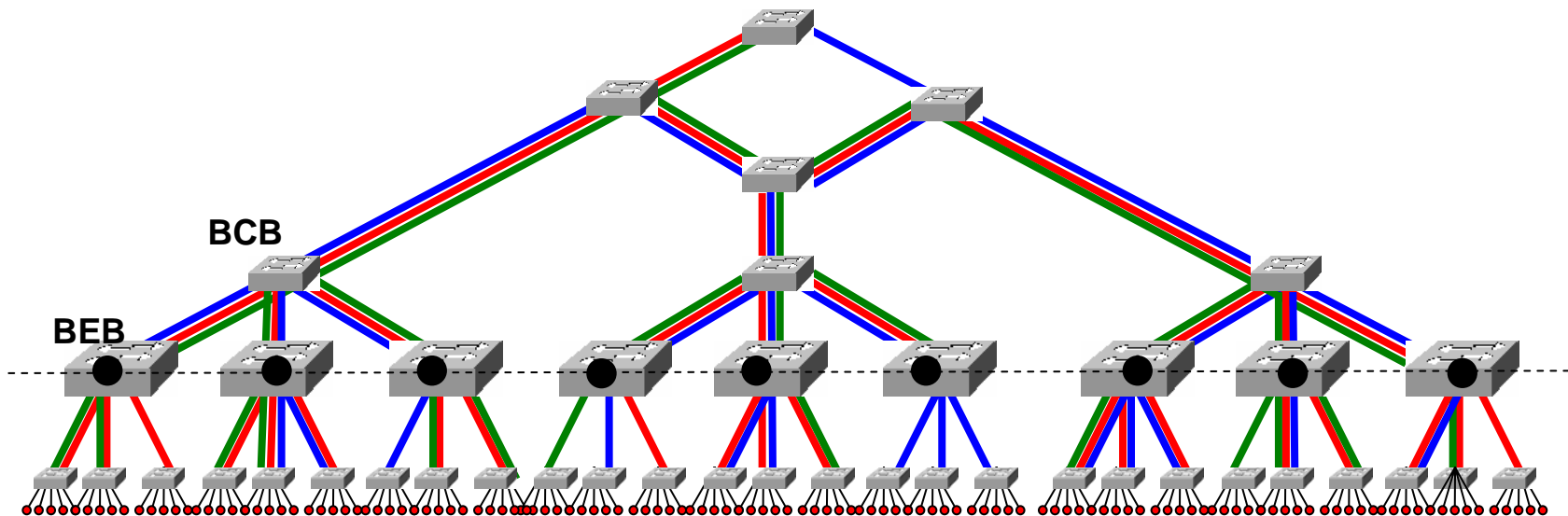
- PBB currently restricts BPDUs from being exchanged across the backbone;
- It would be useful to lift this restriction when the Data Center deploys STP and customer VLANs are allowed to extend through the backbone.
- It is similarly desirable to allow the exchange of Multiple Registration Protocol (MRP) PDUs.

Distinct BB group address not needed

- The number of group addresses in the Data Center is unlikely to represent a significant fraction of FDB entries in core bridges;
- In cases where customer VLANs are allowed to extend through the backbone, mapping customer Group C-DAs to a Backbone Service Instance Group Address (BSIGA) may adversely impact performance;
- Thus, it would be useful to amend PBB to allow the option to use a Group C-DA for the Group B-DA in the frame;

Customer VID = ISID = BVID

BCB

BEB
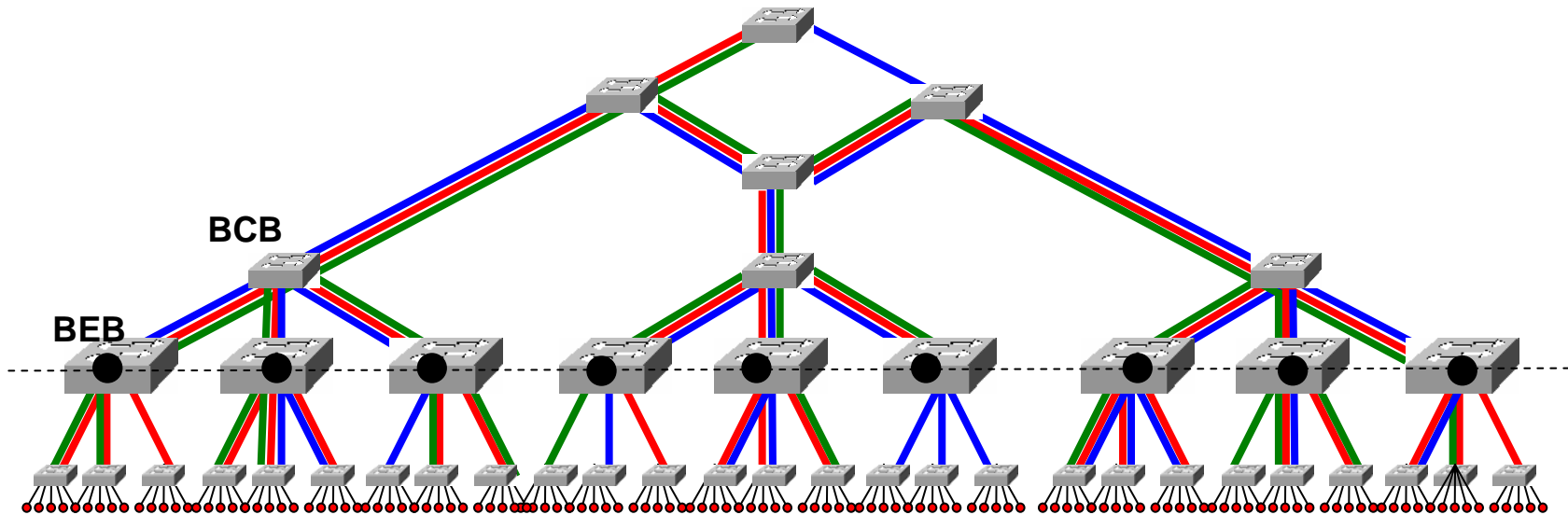
- If the customer VLAN is allowed to extend through the backbone, then the customer service instance (SVID or, if allowed, the CVID) is the same as the backbone service instance (ISID), is the same as the backbone VLAN (BVID);

This suggests several optimizations

- Since the ISID identifies the same entity as the BVID, the ISID could be omitted (saving four bytes) by replacing the I-tag with the Encapsulated Address Type; this would require lifting the PBB restriction preventing use of the EAT on the LAN;
- The B-component ISID → BVID mapping table need not be provisioned; the ISID and the BVID are identical;
- The I-component customer VID → ISID mapping table need not be provisioned; the customer VID and the ISID are identical.

Possible useful extensions to PBB

- Extend the PNP associated with a BEB to optionally support the sending of C-tagged traffic when the BCB is not a Provider Bridge;
- Add C-tagged interface to BEB (to support Data Center CVLAN bridges);
- Support option to remove restriction on BPDUs crossing backbone edge (to allow use of STP in Data Center);
- Support option to allow C-DMAC Group address to be used as B-DMAC;
- Support option to remove restriction on sending Encapsulated Address Type on the LAN (to allow omission of ISID when ISID is identical to BVID.

What's next?

- Would like to hear from DCB or Interworking participants interested in further discussion of requirements for PBBN in the Data Center;
- Could have phone discussion between meetings.