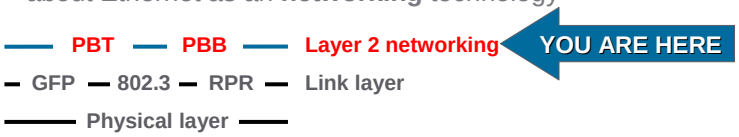
GMPLS and Ethernet “Provider Backbone Transport”

Dave Allan. Nortel
dallan@nortel.com
MPLSCON 2006

What you should get out of this

An appreciation of new innovations in the Ethernet networking space

> “Ethernet” means many different things; this overview is about Ethernet as an **networking** technology

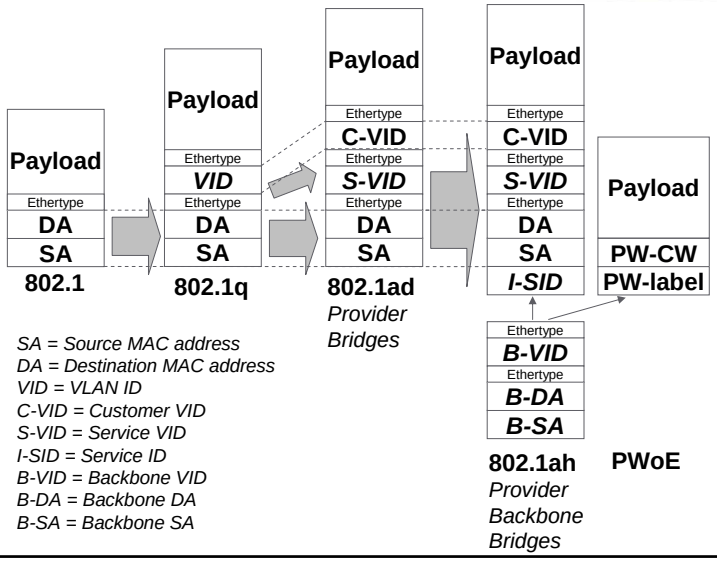


Specifically:

1. 802.1ah Provider Backbone Bridging (PBB)
2. Provider Backbone Transport (PBT)
3. Applying GMPLS to Ethernet and specifically PBT

A brief history of Ethernet stacking

Separating infrastructure end points from service end points has profound implications as to what Ethernet is capable of



SA = Source MAC address
DA = Destination MAC address
VID = VLAN ID
C-VID = Customer VID
S-VID = Service VID
I-SID = Service ID
B-VID = Backbone VID
B-DA = Backbone DA
B-SA = Backbone SA

What is PBT?

> Normal Ethernet
- establishes a constrained, loop free topology
- Broadcasts traffic within the loop free connectivity
- Learns to prune forwarding by observing where traffic comes from
- Problems with STP operation leads to perception that Ethernet “needs help”

> PBT
- Replaces learning to populate “relay filter database” with explicit configuration
- Provides complete route freedom for MAC forwarding as loop free constraint is removed
  - *Loop free is a control plane problem not an STP problem*
- Permits creation of “Ethernet switched paths” or ESPs

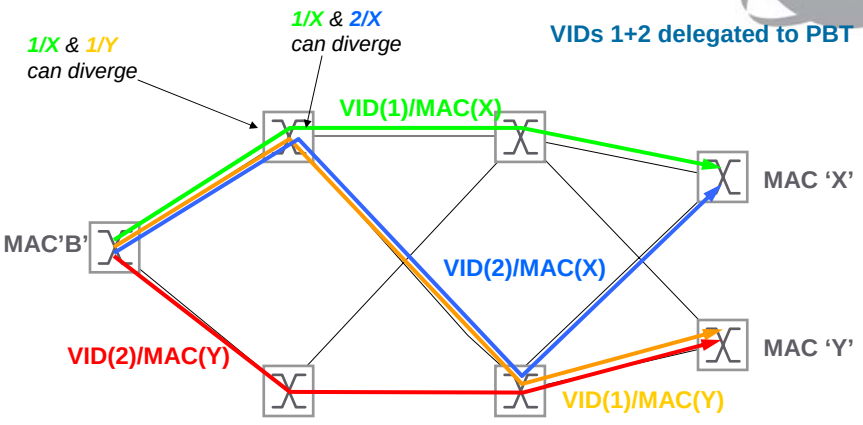
PBT addresses Ethernet’s “Gaping Head Wound”

How can we do this?

- Ethernet as a whole does not fully exploit the standards
  - Independent VLAN Learning (IVL) switches perform a full 60 bit lookup (VID+MAC)
- IVL switches do not need both VLAN AND MAC each to be unique, just the concatenation of both
- We delegate some VLAN IDs (VIDs) to a control plane
  - We only need a few
  - Much of Ethernet today uses MAC/VID paradigm, don't mess with it
- Moving *to* configuration *from* flooding and learning permits complete route freedom for labelled Ethernet Switched Paths (ESPs)
  - Loop free constraints for VLAN are removed
  - Instead of VID referring to a loop free spanning tree, it refers to a mesh instance, and we configure loop free MAC forwarding within each mesh instance…
  - Semantics effectively work out to provide
    - ***60 bit globally unique, destination administered "label"***

MPLScon

Dataplane Example



Note that MACs and VIDs can overlap, it is the combination of both that is unique and allows diverse routing

MPLScon

Is PBT different from Ethernet Today?

- Ethernet standards currently allow
  - MAC learning to be disabled (by VID range)
  - STP to be disabled (by VID range)
  - Forwarding table configuration
- What is needed?
  - Discard unconfigured MACs instead of flooding them
    - detect misconfiguration errors
    - Many proprietary solutions today
      - Unknown Unicast Flood Blocking
      - Flood rate = 0
- Then the dataplane is complete and we can add configuration or a control plane
  - Run bridging and ESPs side by side

MPLScon

Useful Properties

- Destination based forwarding
  - Scales O(N) in the core
- Class based Queuing (.1P)
  - Packet level pre-emption,
  - Connectivity and QoS decoupled, analogous to E-ESPs
- Knowledge of source preserved (SA-MAC)
  - Can instrument connectivity as if P2P
- Global 60 bit label (VID+DA-MAC)
  - dataplane forwards directly on “connection ID”

All of this is largely for free! A combination of what exists with how you use it!

MPLScon

Scaling

> Destination based forwarding means network can be ***fully meshed with one VID*** delegated to PBT

> Constraint routing & Engineering means probability of needing several paths to a given destintion
- Likely in proportion to network diameter

> So we will typically delegate a small number of VIDs to PBT

> Example
- Delegate 16 VIDs to PBT (2^4)
  - Leave **4076 VIDs for bridging,** content delivery etc.
- Provides 2^{50} theoretical MP2P trees in network (16 unique MP2P trees to each of 2^{46} end points)
  - Note that these are PEs , not CEs
- Actual scaling constrained by current switch implementation, not by dataplane design

HUGE connection space with trivial VID consumption

MPLScon

Explicit Design Decisions

> ESPs are always bi-directional
- Both directions have common routing
  - Can have different traffic descriptors in each direction
- Gives us a closed system for OAM purposes
  - Dataplane is self contained and resilient without a control plane
    - can be ***managed OR used with control plane***
- Is consistent with how Ethernet is specified today
  - Ethernet dependent on both directions of a link fate sharing
  - Carry that paradigm over into virtual links

> VID assignment may be different in each direction
- Otherwise
  - Would impose administrative constraints (require pairwise agreement on VID value)
  - Would complicate and constrain MP2P (require pairwise agreement on routing)

MPLScon

Multicast

> Configuring VID/MAC technique extends into multicast

> Using MPLS p2mp procedures and M'cast MAC addresses we can configure (S,G) SSM trees
- Can configure a MP2P return path at the same time…

> Same procedures can be used for VIDs if requirement is to engineer a bundle of groups from a single source

> Addition of OAM allows relatively simply 1:1 and 1+1 protection and maintenance schemes
- On failure, switch to backup tree and then restore original tree
- Also avoids issues with live "fiddling" with broadcast domains…

MPLScon

Applying OAM to PBT

IEEE 802.1-SG13/Q5-SG15/Q9

> FM (802.1ag/Y.1731 *a.k.a. Y.17ethoam*)
- "ME level" concepts
  - PBT uses a single ME level for trunking,
  - PBB uses a single ME level for backbone bridging
  - Client ME levels unaltered…
- Basic Transactions
  - CCM – continuity check (heartbeat)
  - Eth-LB – loop back
  - Eth-LT – link trace
  - Eth-AIS – alarm inhibit signal

PBT is a different way of populating the FIB, it does not change forwarding of packets, so OAM "just works"

> PM (Y.1731)
- ETH-LM – Loss measurement
- Procedures for delay and jitter measurement using ETH-LB

> PS Coordination (G.8031)
- Synchronizes PS state at both ends of a path

MPLScon

Applying GMPLS to PBT

draft-fedyk-gmpls-ethernet-ivl-01

Relatively few modifications are required to GMPLS in order to apply it to PBT

> 60 bit VLAN/DA MAC "label" is invariant
- Different from GMPLS today

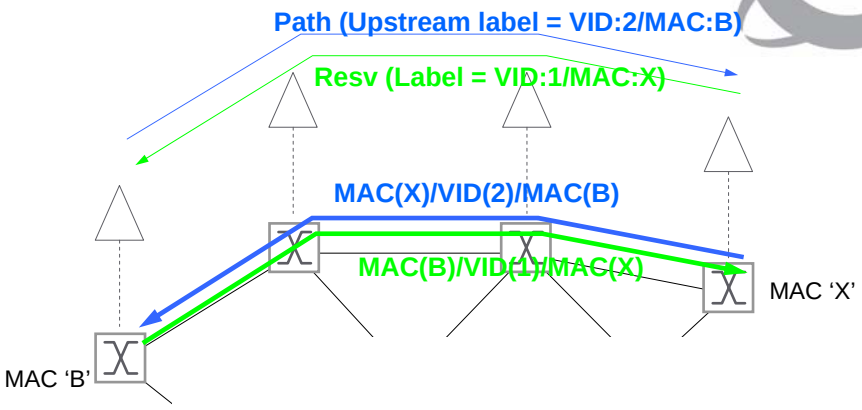
> VLAN (local to DA), DA (global to network) means destination can administer labels
- Destination label administration as per GMPLS today

> Bi-directional ESPs w. common routing preferred
- No impacts on Ethernet OAM (802.1ag/Y.17ethoam)
- No impacts on Ethernet clients (e.g. 802.1ah)
- GMPLS supports today (Upstream label)
- Full 108 bit label SA/VID/DA inferred from upstream label (which will be mandatory)

01 version to be posted shortly

MPLScon

Signalling Bi-Directional Paths



'B' offers preferred label for 'X' in upstream label object

'X' replies with offer of preferred label

60 bit entries populated in the FIB

Full 108 bit connection IDs constructed from both components

MPLScon

Applying PBT to PWs

draft-allan-pw-o-pbt-00

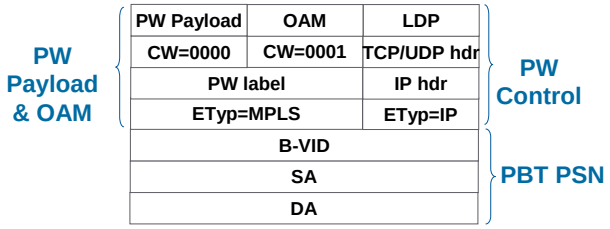
> PBT assumes function of PSN in PW architecture

> Topology defined by PBT trunks/protection groups

- Appear as single PSN/control hop to PW segments
- Additional PW layer discovery/configuration procedures not required

> Channel associated signaling LLC multiplexed with PW

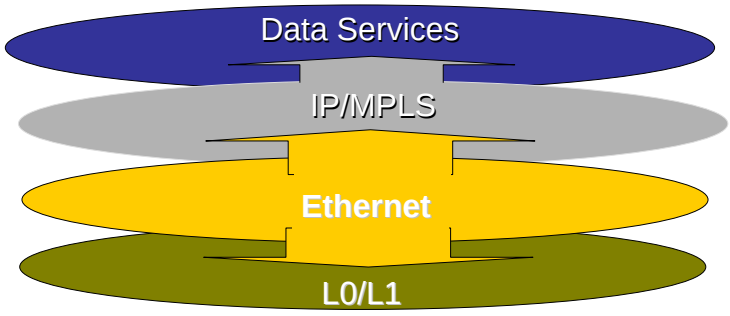
- Control adjacency per PBT trunk/protection group
- PW labels explicitly bound to PBT trunk/protection group



PWoPBT Stack

And the result is…still… *Ethernet*

...but with capabilities akin to MPLS-TE and operational attributes akin to SONET/SDH ***added***



Providing the opportunity for radical delaying of the network

Summary

- Ethernet switches are the telecommunications infrastructure with the lowest cost point
- We can re-purpose these switches with new control software to meet provider needs
- This can be done with minimal changes to Ethernet standards
  - *And they are well underway*
- This provides an opportunity to delayer the network using low cost infrastructure, offering both CAPEX and OPEX benefits

Questions?

For Further Reading

> "GMPLS Control of Ethernet". May 2006
- http://www.ietf.org/internet-drafts/draft-fedyk-gmpls-ethernet-ivl-01.txt

> Pseudo Wires over PBT, March 2006
- http://www.ietf.org/internet-drafts/draft-allan-pw-o-pbt-00.txt

> "Ethernet as Carrier Transport Infrastructure"
- IEEE Communications magazine, February 2006

> www.ieee802.org/1/files/public/docs2005/ah-bottorff-pbt-for-iee-v41-0905.pdf