Point-to-Multipoint Bridging

Integrating Emulated LAN shared media and point-to-multipoint technologies into standard bridges

Norman Finn

Note

- This presentation is available at http://www.ieee802.org/1/files/public/docs2007/avb-nfinn-point-to-multipoint-bridging-061307.pdf.
- Much of the material in this presentation has been recycled from an earlier presentation, entitled "Generalized LAN Emulation," available at http://www.ieee802.org/1/files/public/docs2005/new-nfinn-generalized-lan-emulation-ieee-0305.pdf.

Emulated LANs

What is an “Emulated LAN”?

- An “Emulated LAN” is an attempt to make a bunch of point-to-point and/or point-to-multipoint links look, to the bridge, like a single shared medium.
- It has ports to L2 endstations and/or bridges that look like ports on a shared medium.
- It wants to look like a shared medium to the bridge, router, or host, because that is the only kind of media that Ethernet protocols know about.
- But it can have better performance than a simple shared medium, based on its knowledge of MAC addresses and VLANs.

What is an “Emulated LAN”?

- Examples:

  ATM Forum LAN Emulation

  IEEE 802.11 Wireless Access Point

  IEEE 802.3ah Ethernet Passive Optical Network

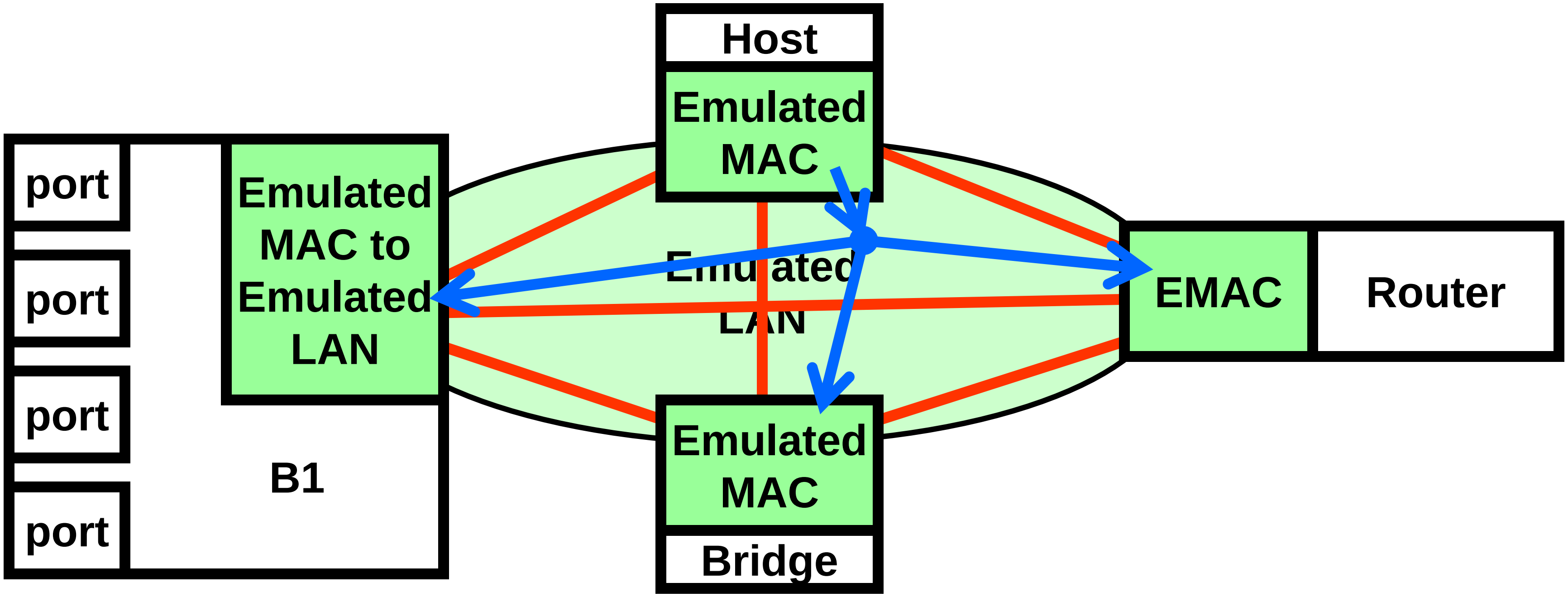
  IEEE 802.17b Spatial Reuse

  IEEE P802.1ah Provider Backbone Bridge

  IETF L2VPN Virtual Private LAN Service (VPLS)

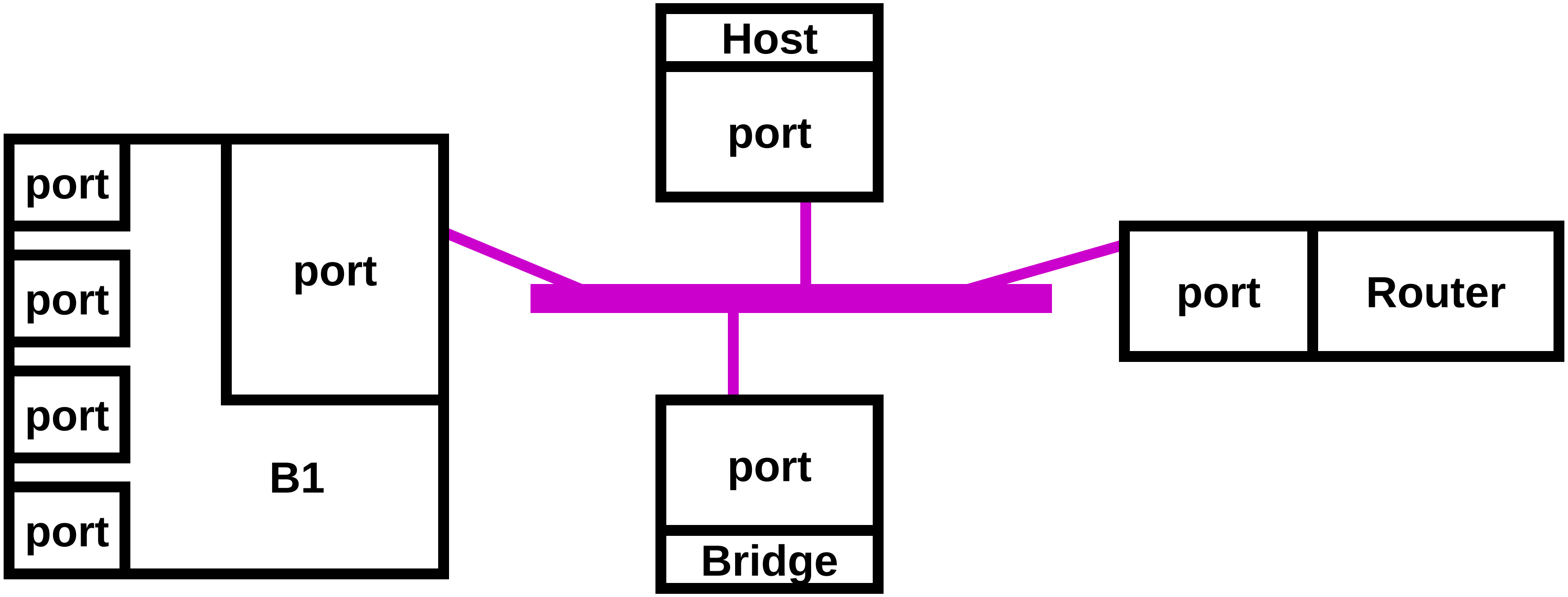
  IEEE P802.3 Secured Hub

What is an "Emulated LAN"?



- "Emulated LAN" is really a combination of point-to-point and point-to-multipoint links.

What is an “Emulated LAN”?



- But it looks to the upper layers of each device like a shared medium: a good old fat yellow coax.

WHY?

- I have a medium with special point-to-point capabilities, special multipoint capabilities, or both.
- Bridges do not understand my medium’s special capabilities, so I must either:
    - Emulate a shared medium LAN, so that bridges can understand me (hard); or
    - Accept the performance penalty of not using all of my capabilities (annoying).

WHY?

- Alternatively, one can decide:

  "I don't understand bridges and don't want to, or perhaps I just don't like/understand/want Spanning Tree, so I'm going to build something similar to a bridge, perhaps with a way-cool new feature or two, only *much* simpler than a bridge!"

- The result: a chain of discoveries as the designer painfully reinvents a bridge.

  Eventually, this "simple" solution is either
  a) just as complex as a bridge, but different, or
  b) is bypassed by the market.

Two ways to solve the problem

- Create an Emulated LAN.

  And do it again and again and again.

  This is a very attractive solution, as witnessed by the number of times it has been employed.

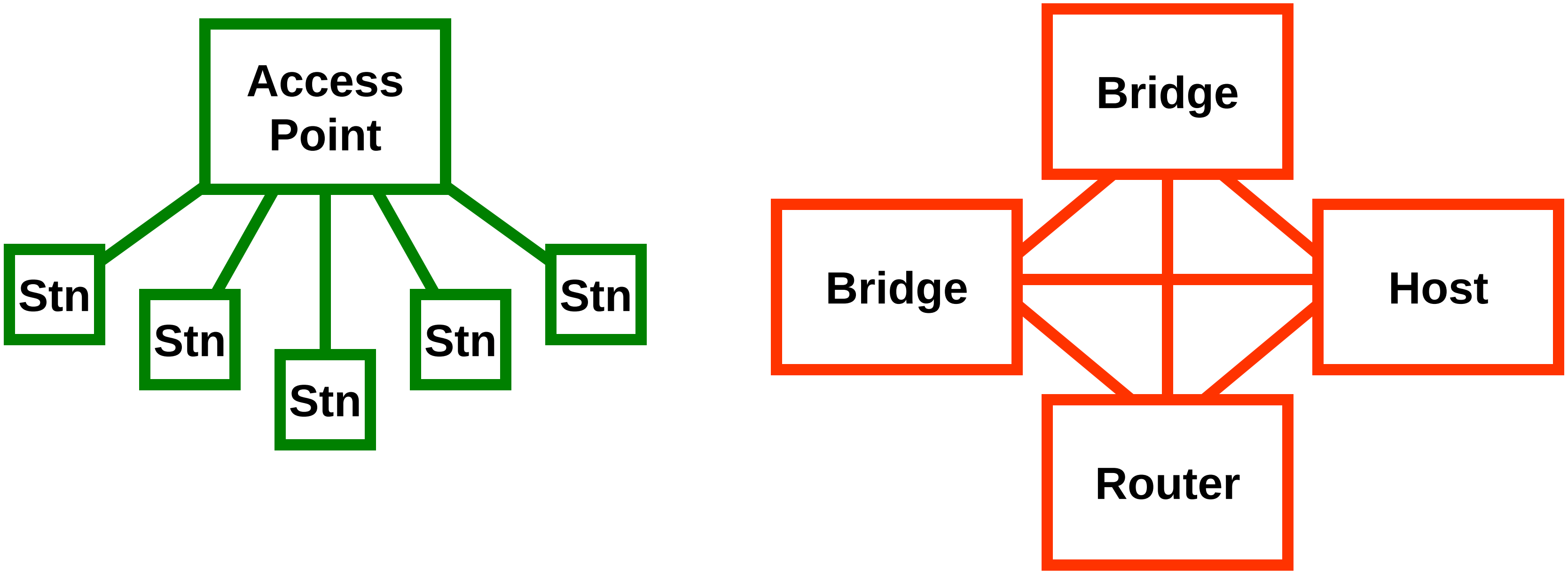
- Teach bridges to handle this common case.

  Extend the standard Bridge Relay Function so that it can directly use the point-to-point and point-to-multipoint elements from which all “Emulated LANs” are built.

  And do it once.

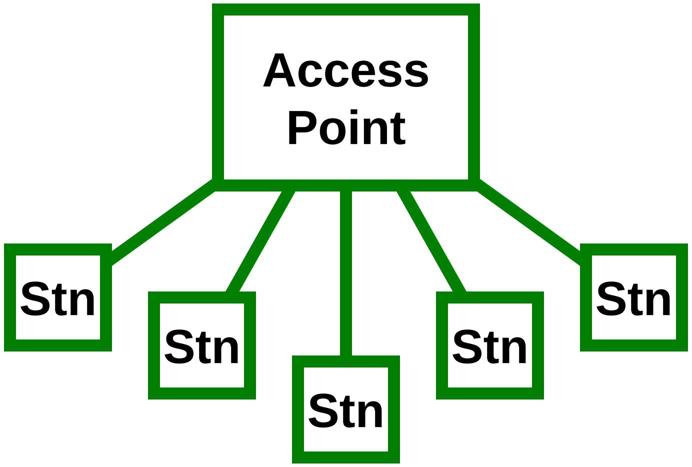
Create a Generalized LAN Emulation Module in Bridges

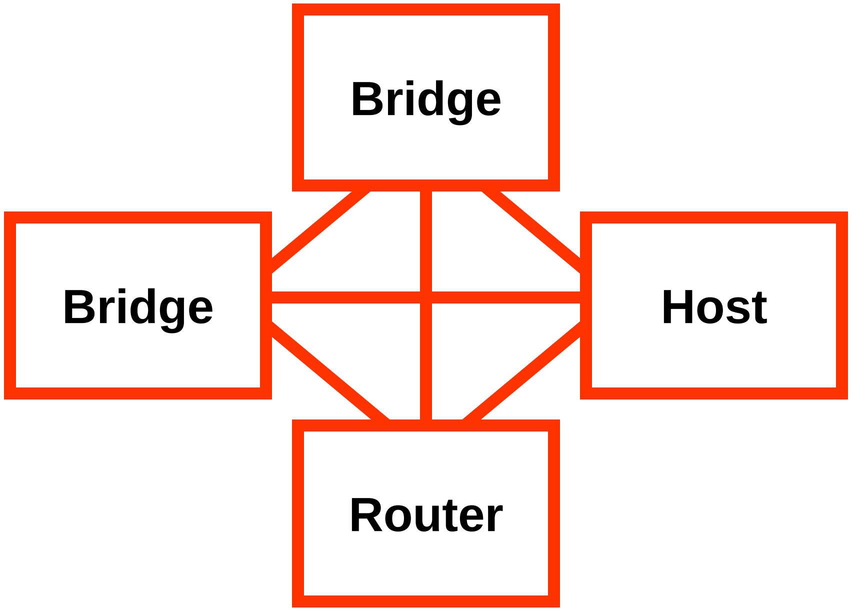
Observation



- There are two very different models for “Emulated LANs”: The Master/Slave model (EPON, 802.11) and the Full Mesh model (ATM LANE, VPLS).

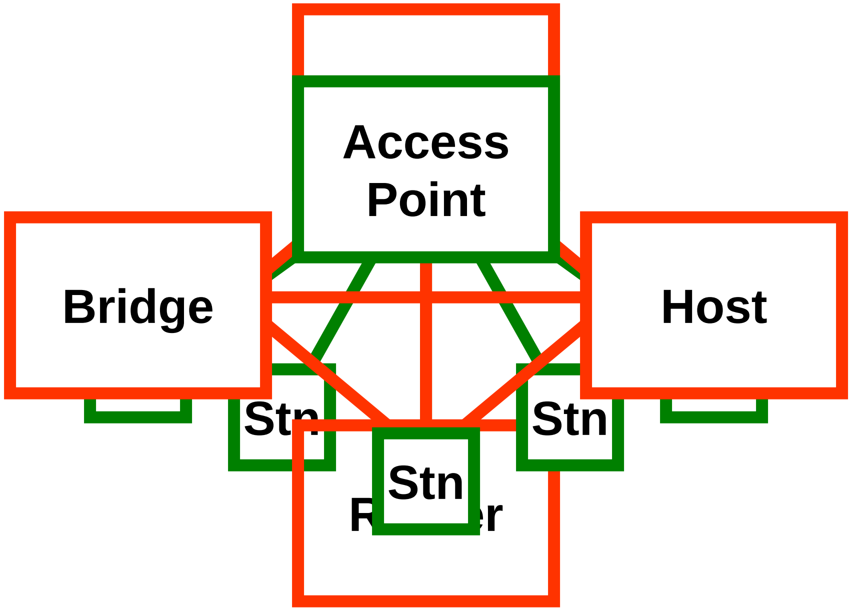
Observation





- The Master/Slave depends on the Master to relay all traffic among the Stations.
- The Full Mesh model never relays traffic inside the mesh (split horizon).

Observation



- If you can integrate those two models, you can cover all such technologies!
- The remainder of this section describes the details of this integration.

All “ELANs” have things in common.

- In every case, an “Emulation MAC” wants:

  To send frames for known unicast addresses to special point-to-point interfaces;

  To learn associations between source MAC addresses in received frames, and the special point-to-point interfaces on which they were received.

- And in many cases, it also wants to:

  Send frames for unknown unicast or multicast addresses to a point-to-multipoint interface, which is much more efficient.

A better way?

- So, suppose we teach the bridge's Relay Function to deal with:
  - A special multicast interface serving a set of point-to-point interfaces; and
  - Different kinds of point-to-point interfaces?
- Then, we can leave MAC address learning to the Relay Function, which knows it best.

Definitions

Generalized LAN Emulation: Port Associations and Portlets

- Two or more Bridge Ports in the same bridge may be associated together into a Port Association.
- Those associated Bridge Ports are then called, “Portlets”.
- A given Portlet belongs to exactly one Port Association.
- To avoid confusion, we reserve the term “Bridge Port” for non-associated ports, e.g., ordinary Ethernet Bridge Ports.

Generalized LAN Emulation:
Port Associations and Portlets

- Some of the current behaviors of bridges with respect to their Bridge Ports apply directly to Portlets; the Port Associations become invisible to those behaviors.
- Other Bridge Port behaviors apply to Port Associations; the Portlets become invisible to those behaviors.

Generalized LAN Emulation:
Port Associations and PA Identifiers

- Every Port Association has a Port Association Identifier (PAID). The PAIDs must be unique over the extent of an "Emulated LAN".
- So, they must either be globally unique, or the underlying medium must guarantee that interconnections that would cause confusion are impossible.
- There are two types of Port Associations: Head PAs and Tail PAs.

Generalized LAN Emulation: Headport Associations

- Every Portlet in a Head PA falls into exactly one of four categories:

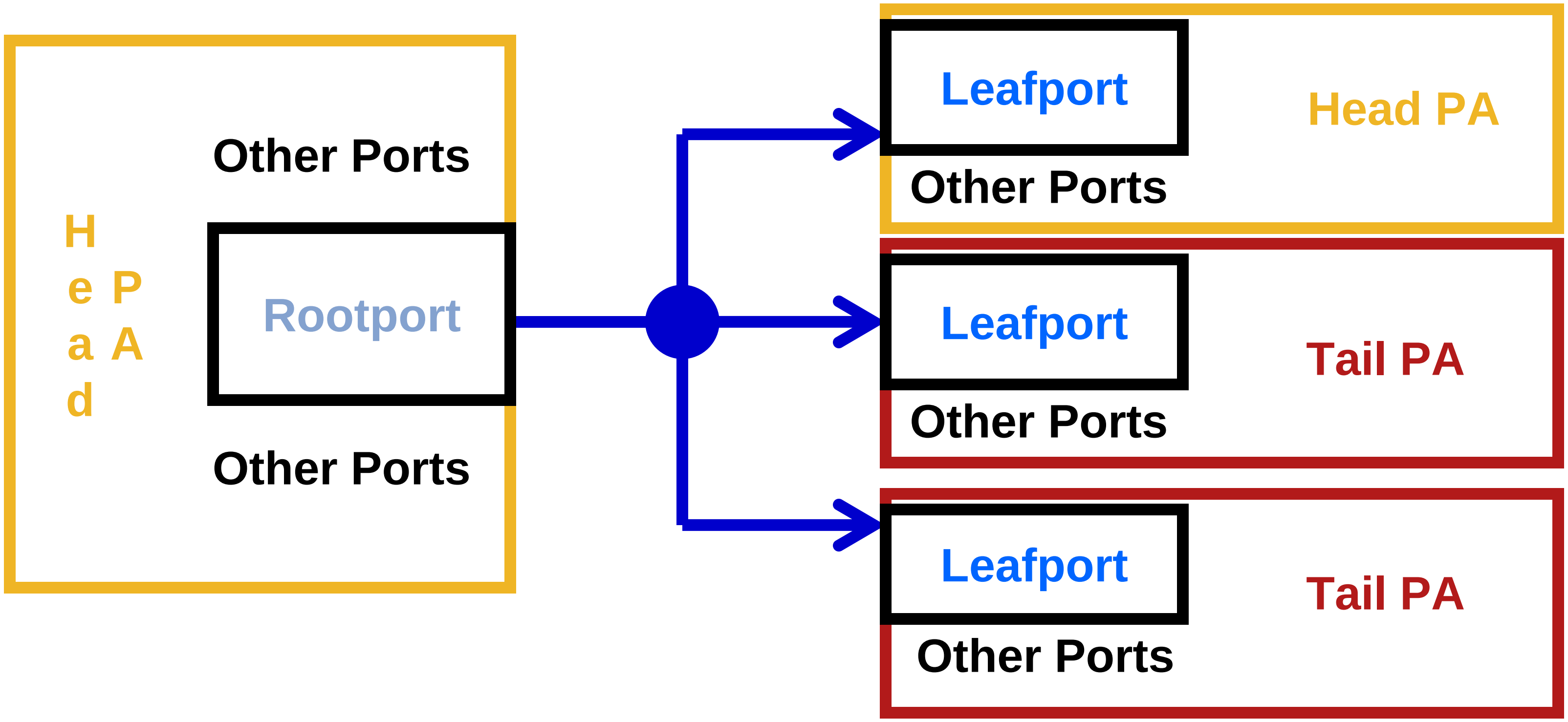
  A “Tailport” terminates a bidirectional point-to-point link (a “Taillink”) to a Tailport in a Tail PA.

  A “Headport” terminates a bidirectional point-to-point link (a “Headlink”) to a Headport in another Head PA.

  A “Rootport” takes frames outward to any number of Leafports in other Head PAs and/or Tail PAs via a unidirectional point-to-multipoint link (a “Rootlink”).

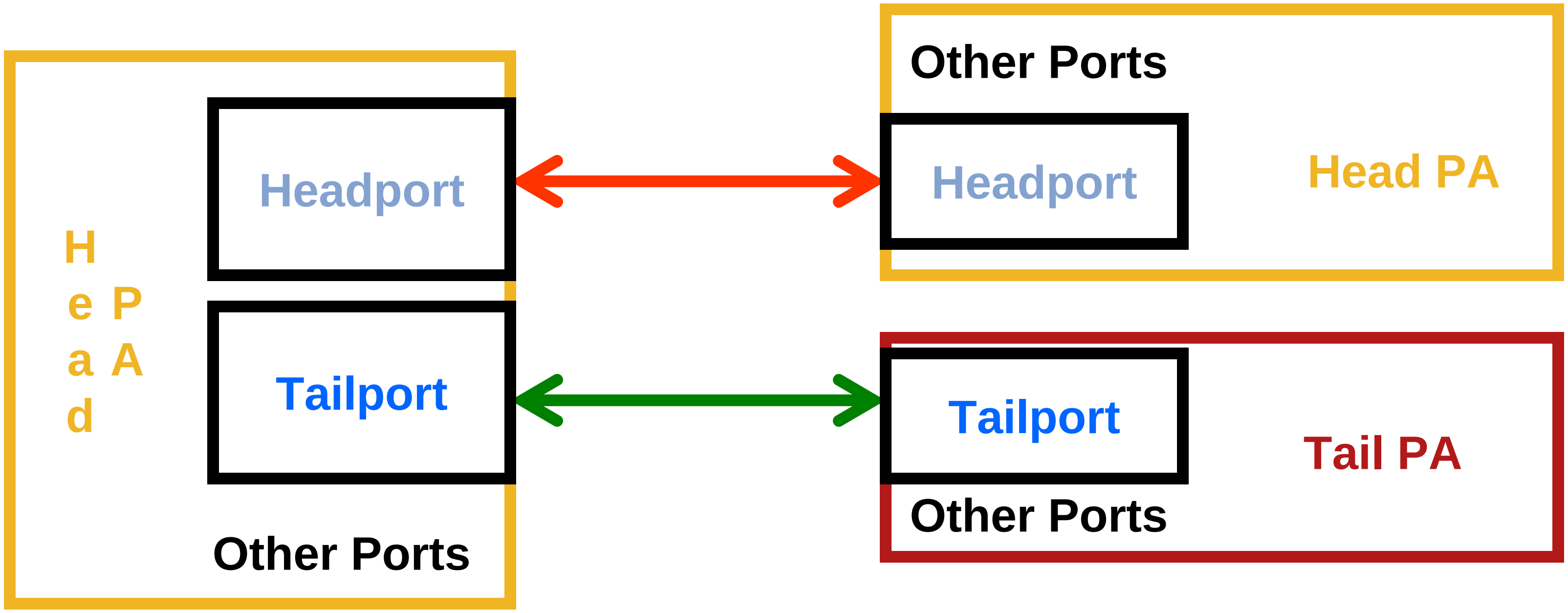
  A “Leafport” brings frames inward from a single Rootport in another Head PA via a Rootlink.

Generalized LAN Emulation



- Rootports and Leafports.
- Unidirectional Rootlink.

Generalized LAN Emulation



- Headports and Tailports.
- Bidirectional Headlink and Taillink.
- There are no Tail PA–to–Tail PA links.

Generalized LAN Emulation: Headport Associations

- A Head PA has Headports, Tailports, and Rootports:

  For each Tail PA in the Emulated LAN, one Tailport with a Taillink to the Tailport of that Tail PA.

  For each of the other Head PAs in the Emulated LAN, one Headport with a Headlink to the Headport of that Head PA.

  No more than one All Rootport with a Rootlink to zero or more Leafports, one for each PA to which the Head PA has either a Taillink or a Headlink.

  No more than one Local Rootport, with a Rootlink to zero or more Leafports, one for each Tail PA to which the Head PA has a Taillink.

  For each Headlink to another Head PA, at most one Leafport on a Rootlink to that Head PA's All Rootport.

Generalized LAN Emulation: Tailport Associations

- A Tail PA is associated with exactly one Head PA.
- A Tail PA has Tailports and Leafports:

  Exactly one Tailport linked to a Tailport in a Head PA with a (bi-directional) Taillink.

  At most one Leafport linked to each Rootport in the Head PA. The Head PA’s All Rootport and Local Rootport may connect to the Tail PA either separately, through two Leafports, or together, through the same Leafport.

Generalized LAN Emulation: Head PAs vs. Tail PAs

- Note that “Head” and “Tail” are not synonymous with “Bridge” or “Station”.
- A Station can have a Head PA, but the Head PA is, in effect, a Bridge.
- A Bridge can connect to an emulated LAN through a Tail PA or through a Head PA.

  Connecting via a Tail PA is actually easier, because the Tail PA looks like a normal Bridge Port.

Generalized LAN Emulation: Head PAs vs. Tail PAs

- The Head PAs in an Emulated LAN must be connected via a full mesh of Headlinks.
- The Head PAs may have a full mesh of point-to-multpoint Rootlinks.
- Note that these full meshes are not trivial to ensure when the underlying medium is point-to-point, and when there are more than one or two interconnected Head PAs.
- A Tail PA is linked to exactly one Head PA.

Generalized LAN Emulation: PAID Marking and Learning

- The frame format used on the point-to-multpoint links must include a provision for carrying the PAID in addition to the standard IEEE 802.1D ISS MAC addresses (the “Ethernet MAC addresses”).
- Source MAC addresses received by a Bridge on a Portlet are learned according to the standard rules, except that Source MAC addresses received on a Leafport are learned as if received from the Headport or Tailport connected to that same PA.

Free Rootlinks

Free Rootlinks

- All Rootports and Local Rootports are tied to PAs attached to Emulated LANs. This clarifies their use.
- Generalizing this idea, a Bridge could have "Free Rootports" that connect via Free Rootlinks to Free Leafports in other devices.

Free Rootlinks

- Like other Leafports, each Free Leafport must be associated with exactly one Headport, Tailport, and/or “other” Bridge Port for MAC address learning.
- A Bridge has the option of sending a frame to a Free Rootport, instead of sending it to each individual member of its associated set of Ports and Portlets, as long as this would not violate the normal rules for Bridges forwarding frames.

Free Rootlinks

- Like other Rootports, every Free Rootport is associated with a set of Headports, Tailports and/or “other” Bridge Ports to which its Free Rootlink optimizes distribution.

  Sending a frame to a Free Rootport is exactly equivalent to sending the frame to every one of its associated ports.

  If a Free Rootlink can be used to reflect a frame back towards its source, then its Rootport and each of its Leafports must have a PAID, and that PAID must be sent from the Rootport so that the source Leafport can discard the frame.

Free Rootlinks

- If a Free Rootlink is associated with more than one Emulated LAN and/or “other” Bridge Port, then either:

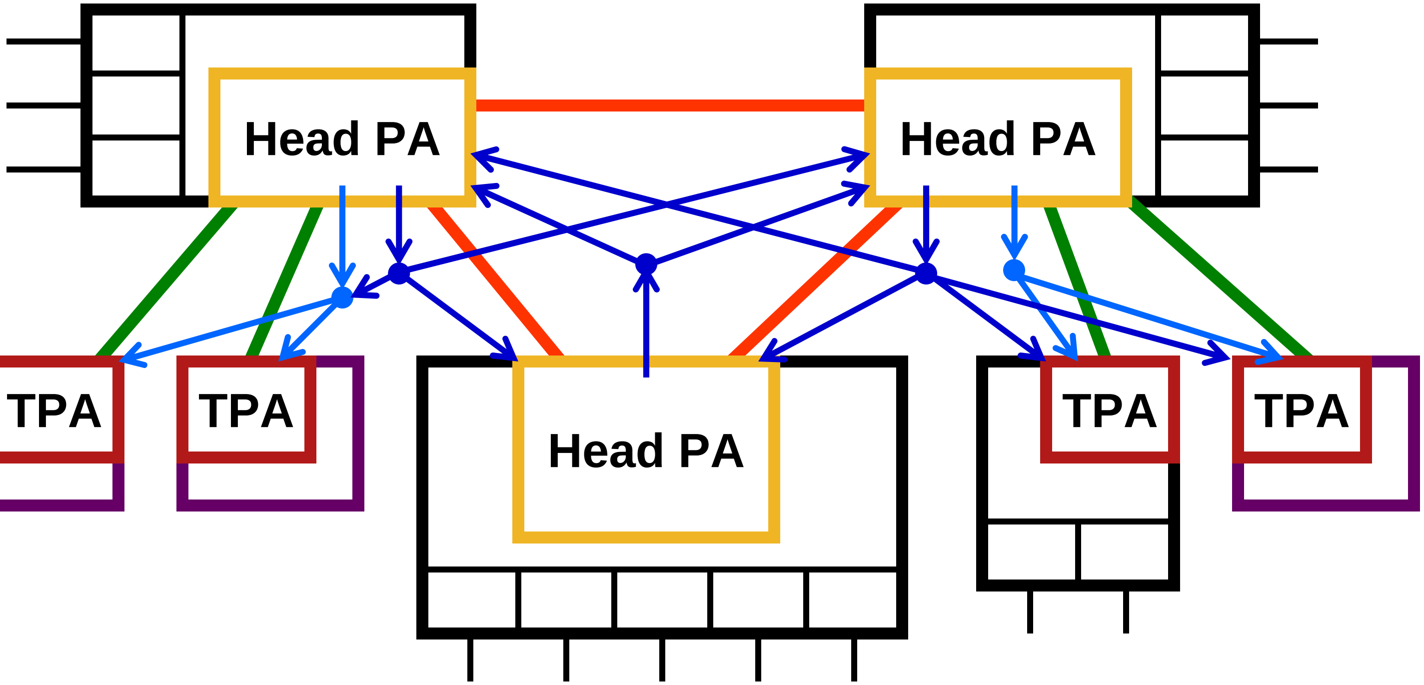
    It must change its topology as the Spanning Tree topology changes; or

    There must be a set of overlapping Free Rootlinks, so that the one appropriate to the Spanning Tree state can be used; or

    Some combination of both techniques can be employed.

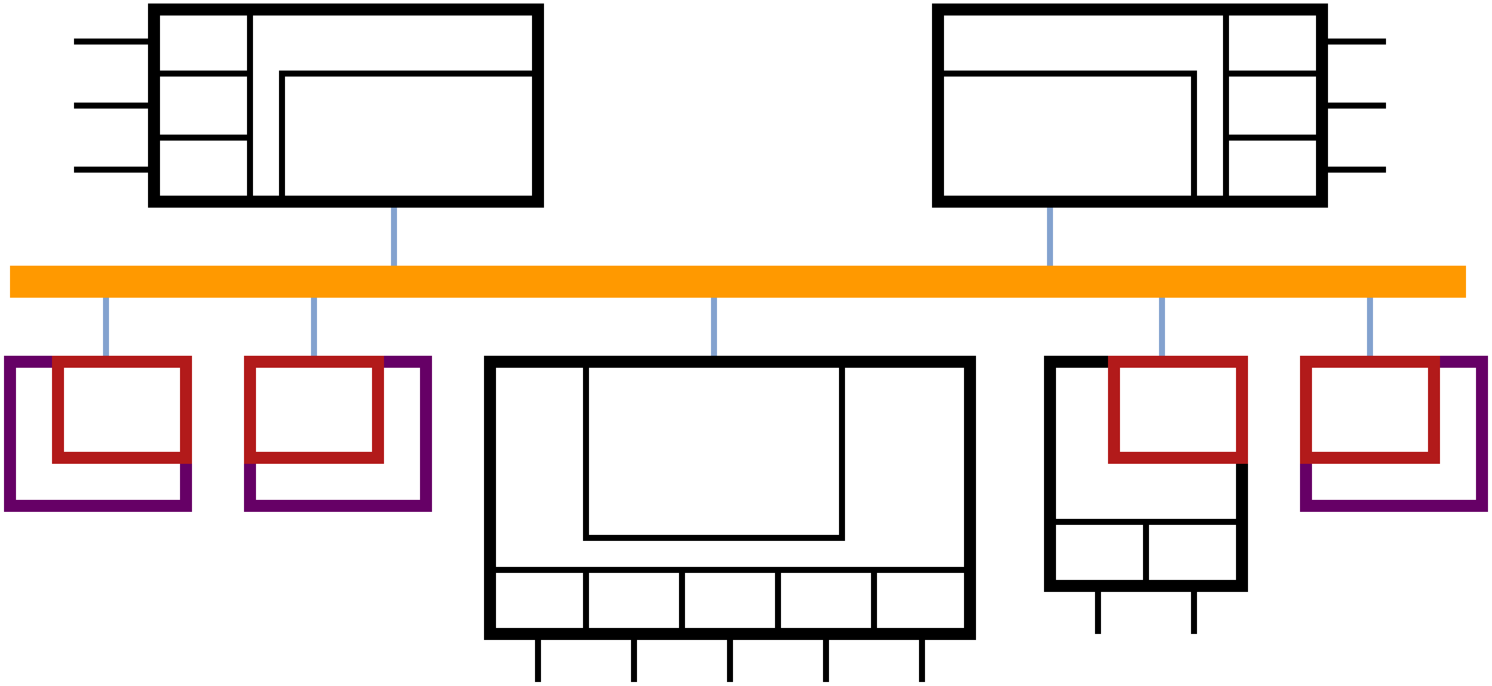
1000 Words

A small “Emulated LAN”



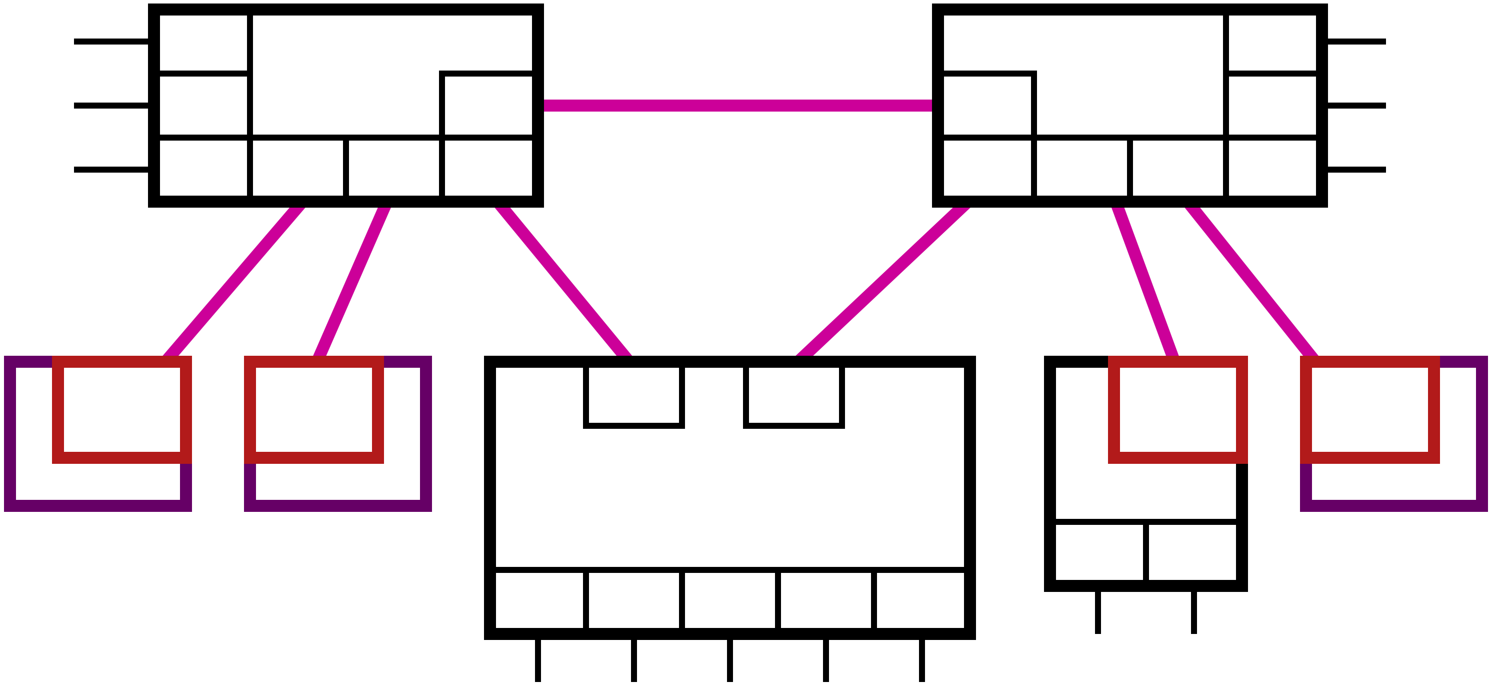
- Bridges, Endstations, Head Port Associations, Tail Port Associations, Headlinks, Taillinks, All Rootlinks, and Local Rootlinks.
- (Note the two ways of doing the All Rootlink.)

A small “Emulated LAN”



- To BPDUs or MRP in PA Mode, this looks like a Fat Yellow Coax.

A small "Emulated LAN"



- To MRP in Separated Mode (see below), this looks like a lot of point-to-point links.

Data Forwarding

Data Forwarding rules

- “Access” Tailports may be exempted from the transmission of unknown unicast frames.
- Heuristics may be employed to transmit frames through the All Rootport, if the nature of the underlying medium makes this more efficient than transmitting frames through multiple point-to-point links.
- Rootports and their point-to-multipoint Rootlinks are an optional optimization for each individual Head PA.

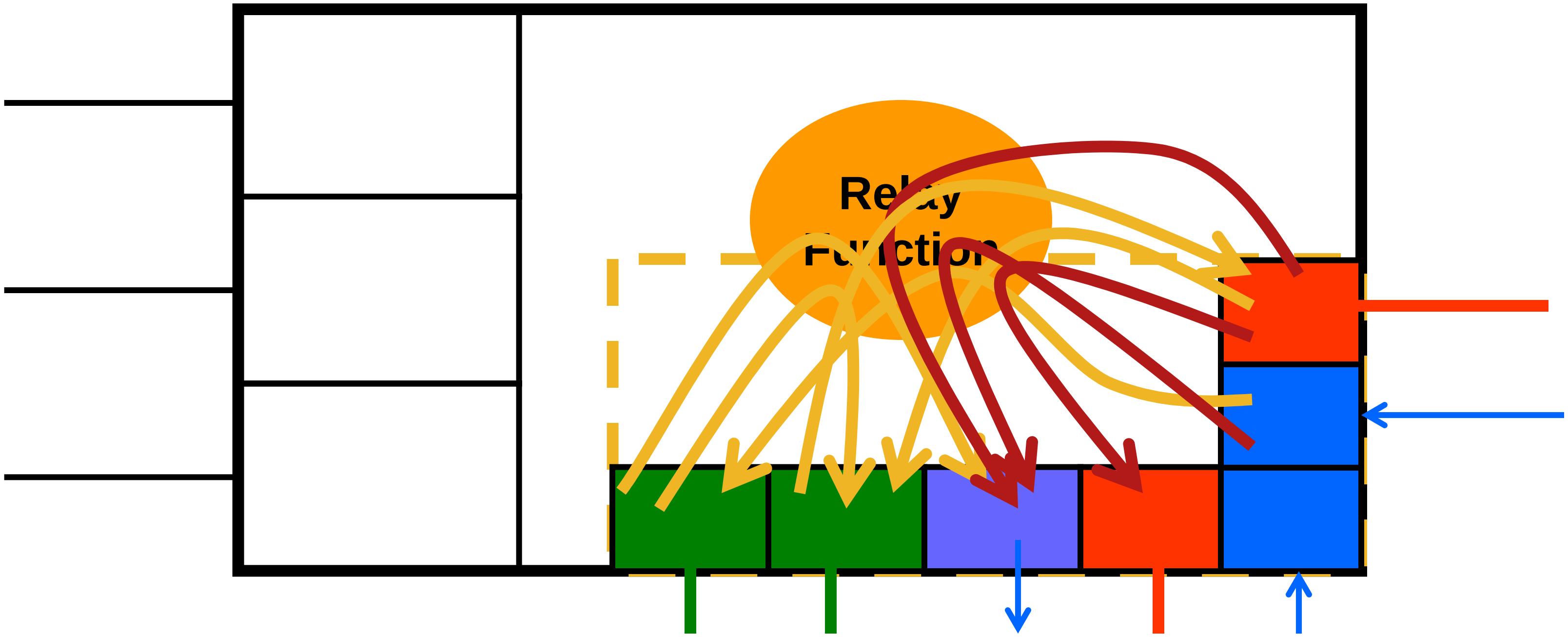
Data Forwarding rules

- No frame received on a Headport or Leafport can ever be relayed on any Headport belonging to the same Port Association, or to the All Rootport.

  It is the job of the transmitting Head PA to ensure that the frame gets to the right Head PAs in the Emulated LAN.

- Any frame received from a Tailport, and relayed to a Rootport of the same Head PA, is marked with the PAID of the Tail PA from which it was received.
- All other frames transmitted to a Rootport are marked with the PAID of the transmitting Head PA.
- A frame received on a Leafport, marked with the PAID of the receiving Tail PA, must be discarded.

Head PA Relaying Rules



- Some relaying is allowed, some is not allowed.

Head PA Relaying Rules

(Normal, unblocked,
Head Port Associations)

TO

F
R
O
M

	Head	All Root	Local Root	Tail	Other
Head	NO	NO	YES	YES	YES
Leaf	NO	NO	YES	YES	YES
Tail	YES	YES	YES	YES	YES
Other	YES	YES	YES	YES	normal bridging

Control Protocols

Control Protocols

- The control protocols include:

  Spanning tree Bridge Protocol Data Units (802.1Q BPDUs).

  Multiple Routing Protocol PDUs (802.1ak MRP PDUs).

  Link Aggregation Control Protocol (802.3ad LACP PDUs).

  Link Layer Discovery Protocol (802.1AB LLCP PDUs).

  Others

- There are two ways to handle these protocols:

  Separated Mode: Treat each Portlet separate Bridge Port, and send different information on each.

  Combined Mode: Treat the whole Port Association's as a single Bridge Port, and distribute identical PDUs to all attached systems.

Control Protocols:
Separated Mode and Combined Mode

- Interestingly, the same decision need not be made for all control protocols.
- For example, one could run [M,R]STP in Combined Mode on the Port Association, and run M{V,M}RP in Separated Mode on the Portlets.
- Note, however, that if spanning tree runs in Separated Mode, all other Bridge protocols must run in Separated Mode; they cannot run at a coarser grain than spanning tree.

Combined Mode

- In Combined Mode, as far as the protocol sending and receiving the PDUs is concerned, the entire Port Association is one Bridge Port.
- A single database of distribution state information is maintained for all of the Portlets in the PA.
- All received control information is handled as if it arrived on the All Rootport.

Combined Mode

- Control PDUs received by a Head PA on a Tailport are accepted by the receiving Bridge, and relayed within the Emulated LAN just like a normal multicast, either:
  - Relayed individually to all Headports and (other) Tailports; or
  - Relayed to all connected PAs through the All Rootport.
- Control PDUs received on Headports or Leafports are accepted by the receiving Bridge, and applied to the whole Port Association, but are relayed only to Tailports and/or the Local Rootport.
- The forwarding of control PDUs among the Portlets of a PA (as well as data frames) apply even if the Port Association is blocked by the Spanning Tree Protocol. (See Blocked Bridge Ports, below.)

Separated Mode

- In Separated Mode, each individual Portlet is treated as a Bridge Port.
- Separate databases and state machines are maintained for each of the Headports and Tailports in the Port Association.
- All control protocol information received from a Leafport is handled as if it arrived from the corresponding Headport or Tailport.

Separated Mode

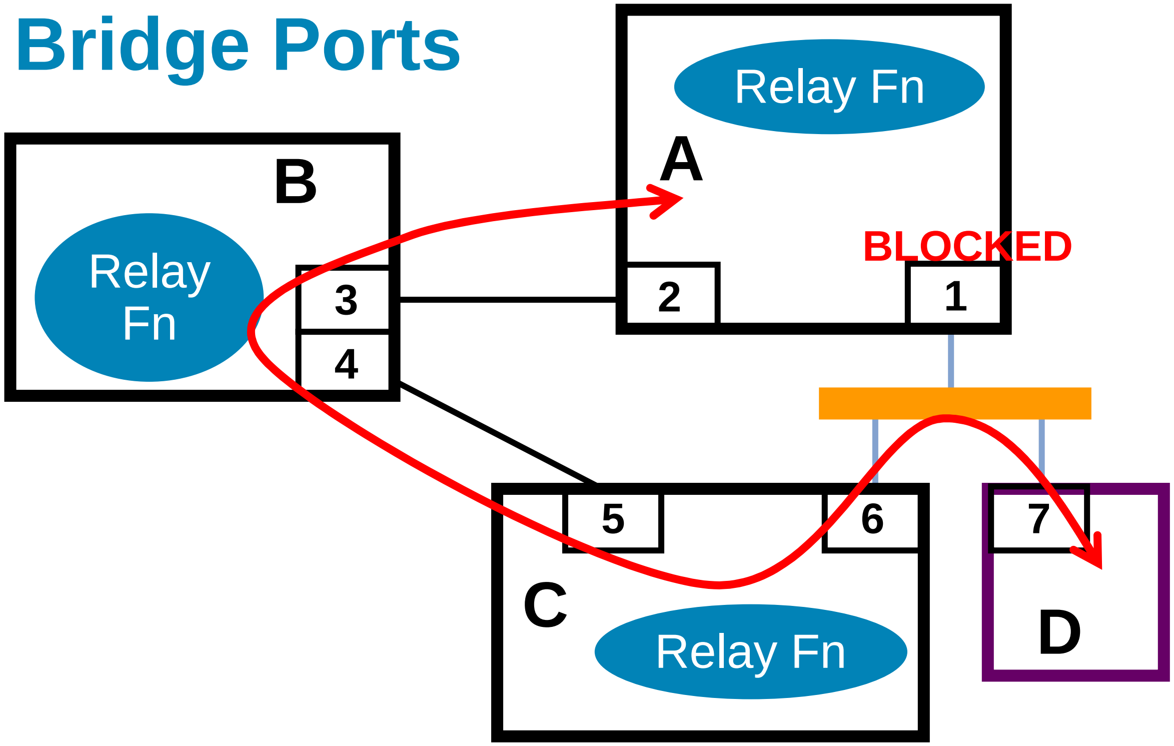
- Control PDUs are transmitted on a Tailport or Headport, although, if a control protocol did transmit identical PDUs to multiple Portlets, a Rootport could be used.
- Propagation of information follows rules for data relaying, e.g., an MVRP registration received on a Headport is not relayed to other Headports, but is relayed to Tailports.
- No relaying of control PDUs is done in Separated Mode; data relaying is controlled by the usual rules for Bridges.

Blocked Bridge Ports

Blocked Bridge Ports

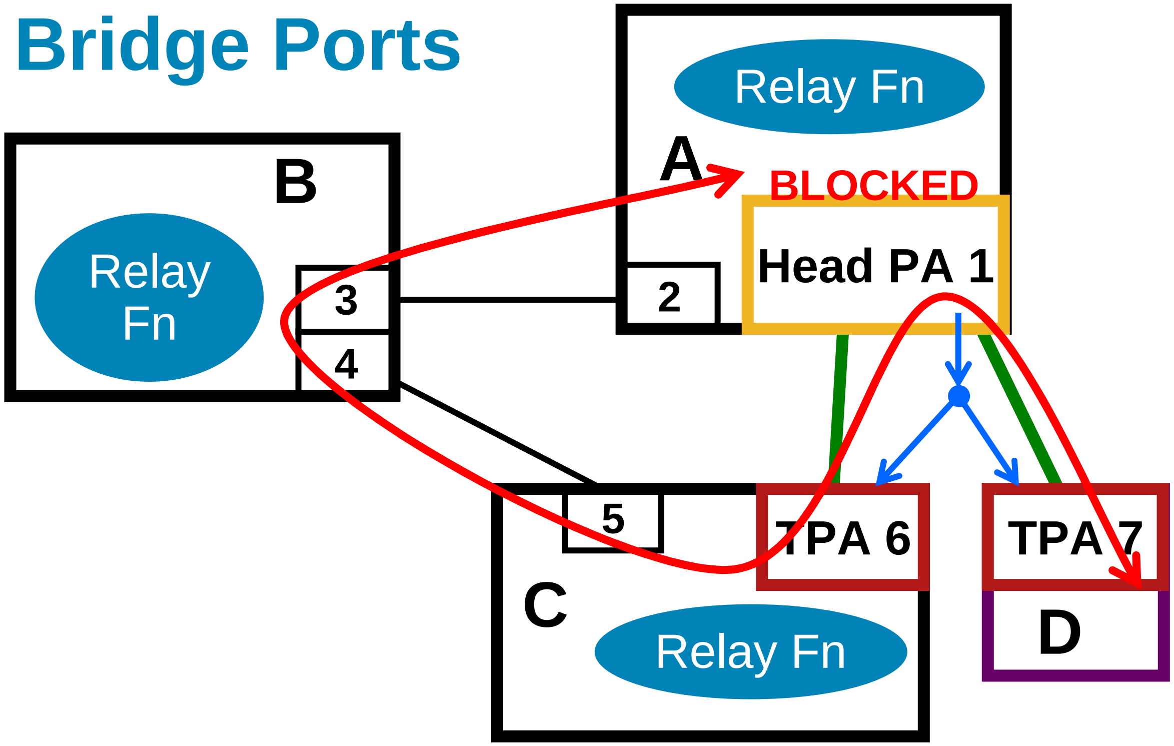
- If a device has both point-to-point and point-to-multipoint capability, such as that available to an 802.11 Access Point, there are issues with blocking ports.
- In Combined Mode, where PA is considered a single emulated shared medium, then there are cases where the PA must act as a sub-Bridge, relaying data from Tailport to Tailport , even though the Port Association is blocked. However, connectivity among the Stations behaves normally.
- In Separated Mode, where each link Headlink or Taillink is an individual Bridge Port, then no sub-Bridge is needed. But, the Bridge wants to send multicast data to a subset of the stations reached by its Rootport(s).

Combined Mode
Blocked Bridge Ports



- A shared medium connects Bridge Ports 1, 6, and 7.
- There is a loop that must be broken by blocking a Bridge Port.
- Let’s say 1 is blocked. The arrow shows the A—D path.

Combined Mode
Blocked Bridge Ports



- Bridge **A** in Combined Mode. Perhaps because the 3—2 link is slow, the Head PA is blocked.
- The Head PA in A must relay data from C to D, even though its port is blocked.

Combined Mode Blocked Bridge Ports

- The blocking of Bridge Ports takes place above the Head PA or Tail PA functions in a Bridge; a whole PA is blocked or unblocked by spanning tree, not individual Portlets.
- Blocking a Tail PA is trivial, since the Tail PA does not make choices based on MAC addresses.
- Blocking a Head PA is somewhat more complex, since the Head PA still forwards data among its Tailports.

  Blocking a Bridge Port's access to a shared medium does not prevent the medium's stations from communicating.

  The Head PA must relay traffic among Taillinks and the Local Rootlink so the emulated medium's stations can communicate.

  No Headport forwarding is needed, however.

Combined Mode
Blocked Bridge Ports: Learning and Forwarding

- On the one hand, it is desirable to continue learning MAC addresses, since multiple Stations can be Bridges, and we want to forward frames efficiently.
- However, the Bridge cannot learn MAC addresses from a blocked Bridge Port, or it will try to forward data from an “other” Ethernet port to the blocked Head PA, and blackhole it, when another path through the network is available.

Combined Mode
Blocked Bridge Ports: Learning and Forwarding

- Unless is simply floods all data, the blocked Head PA must maintain an independent Filtering Database (FDB), separate from the main Bridge’s FDB, and separate from any other blocked Head PA’s FDB, and perform address learning and forwarding on its own FDB while the PA is blocked.
- If there is only one Head PA, as for an 802.11 Access Point, this extra FDB can be a single bit added to the normal FDB, so is feasible. For more complex devices, it becomes a burden.

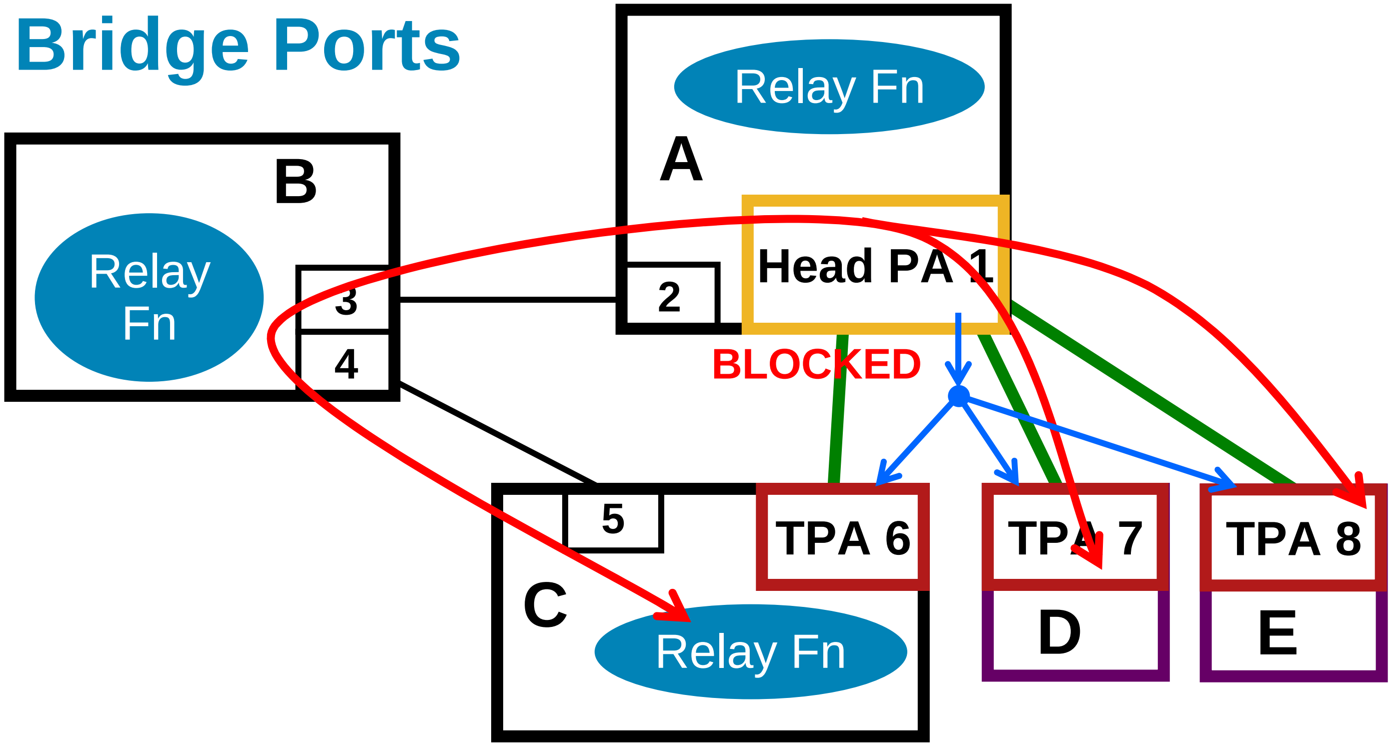
Combined Mode Head PA Relaying Rules

(For Blocked Head Port Associations)

TO

		Head	All Root	Local Root	Tail	Other
FROM	Head	NO	NO	NO	NO	NO
	Leaf	NO	NO	NO	NO	NO
	Tail	NO	NO	YES	YES	NO
	Other	NO	NO	NO	NO	NO

Separated Mode
Blocked Bridge Ports



- Bridge **A** in Separated Mode. Perhaps because the 3—2 link is slow, the Head PA blocks the Tailport to 6.
- The Head PA's Rootport cannot be used, now; it would send data to C, as well as D and E.

Separated Mode
Blocked Bridge Ports: Forwarding

- Use of a Rootport is highly desirable, especially if there are many Tailports and/or Headports, because it saves bandwidth.
- But, the Head PA’s Bridge cannot send data through a blocked Bridge Port.
- So, there must be a means for only distributing data to those other PAs that should get it.

Separated Mode
Blocked Bridge Ports: Forwarding

- The Head PA can:
  - Alter the scope of the Rootport as the spanning tree state changes.
  - Maintain multiple Rootports, one for each possible combination of blocked ports.
  - Use special “Group PAIDs” to identify the subset of Leafports to which a given frame is to be received.

Separated Mode
Blocked Bridge Ports: Forwarding

- A less desirable option:

  Invent a protocol to indicate the Head PA’s port blocking state to the Leafports, so they can filter it.

  This protocol would have to be coordinated with the Spanning Tree Protocol to prevent temporary loops.

  It would be likely to extend, somewhat, the time required for the spanning tree state to converge after a topology change.

Name your poison

- Combined Mode requires forwarding among the Portlets when the Port Association is blocked.
- Separated Mode requires the Rootlink be able to adjust quickly to the state of the Spanning Tree Protocol on the various Portlets.
- TAANSTAAFL*, when it comes to teaching Bridges to take advantage of point-to-multipoint links.

* There Ain't No Such Thing As A Free Lunch: from *The Moon Is A Harsh Mistress* by Robert Heinlein, G. P. Putnam's Sons, 1966.

Mapping pseudo-Bridges to Generalized LAN Emulation

ATM LAN Emulation

- Every node in the ELAN is a Head PA.
- A (potential) full mesh is created and maintained dynamically using the ATM LAN control protocols.
- There are no Tail PAs nor Tailports.
- The Rootport is the point-to-point link to the Broadcast and Unknown Server (BUS).
- The mesh of point-to-multpoint links is emulated via the BUS, and its single point-to-multpoint link.

EPON

- The IDs in the preamble serve as PAIDs.
- There is one Head PA (the OLT) and some number of Tail PAs (the ONUs).
- There is only one Rootport, and no Headports.
- The Rootport is not used by a bridge at present, but should be.

802.17b Spatial Reuse

- Every node is a Head PA.
- The full mesh is provided by ring connectivity.
- Non-SR nodes:
  - All frames sent Rootport.
- Non-bridge SR nodes:
  - Filtering Database determined by ring exploration.
  - PA Mode used for multicast and unknown unicast.
- Bridge SR nodes:
  - Filtering Database determined by learning.
  - Either PA Mode or Separated Mode is acceptable.

802.1ah Provider Backbone Bridge

- Each Virtual Instance Port (VIP) is a Head PA; there are no Tail PAs.
- A Headlink, consisting of a source / destination MAC address pair, connects each pair of VIPs across the backbone.
- The All Rootlink is the default outer MAC address.
- The full mesh of Headlinks needs no explicit creation; the mesh is the byproduct of normal backbone connectivity.
- Free Rootlinks can be used to optimize control packet distribution across VIPs, e.g. “MIRP”.

  Free Rootlinks do not carry data.

VPLS

- Every node is a Head PA, so a full mesh is both required, and difficult to guarantee.
- If there is no Rootport, every multicast, broadcast, or unknown unicast must be sent via replication on all Headports.
- Since all links are Headports, intra-PA relaying is disallowed. (This is also called, “split horizon”.)

Secure Hub (e.g., 802.1aj TPMR Phone+PC)

- This will make a Bridge Port attached to a hub operate much like an 802.11 Access Point.
- One (or more bridges) could have Head PAs, the rest (if any) Tail PAs. One is easier to set up, two quicker to failover.
- No Rootports are allowed; there is no place for the PAID, and they would reduce the level of security, assuming the Bridge trusts no station.

Mapping 802.11 Access Points to Generalized LAN Emulation

802.11 Access Point as a Bridge

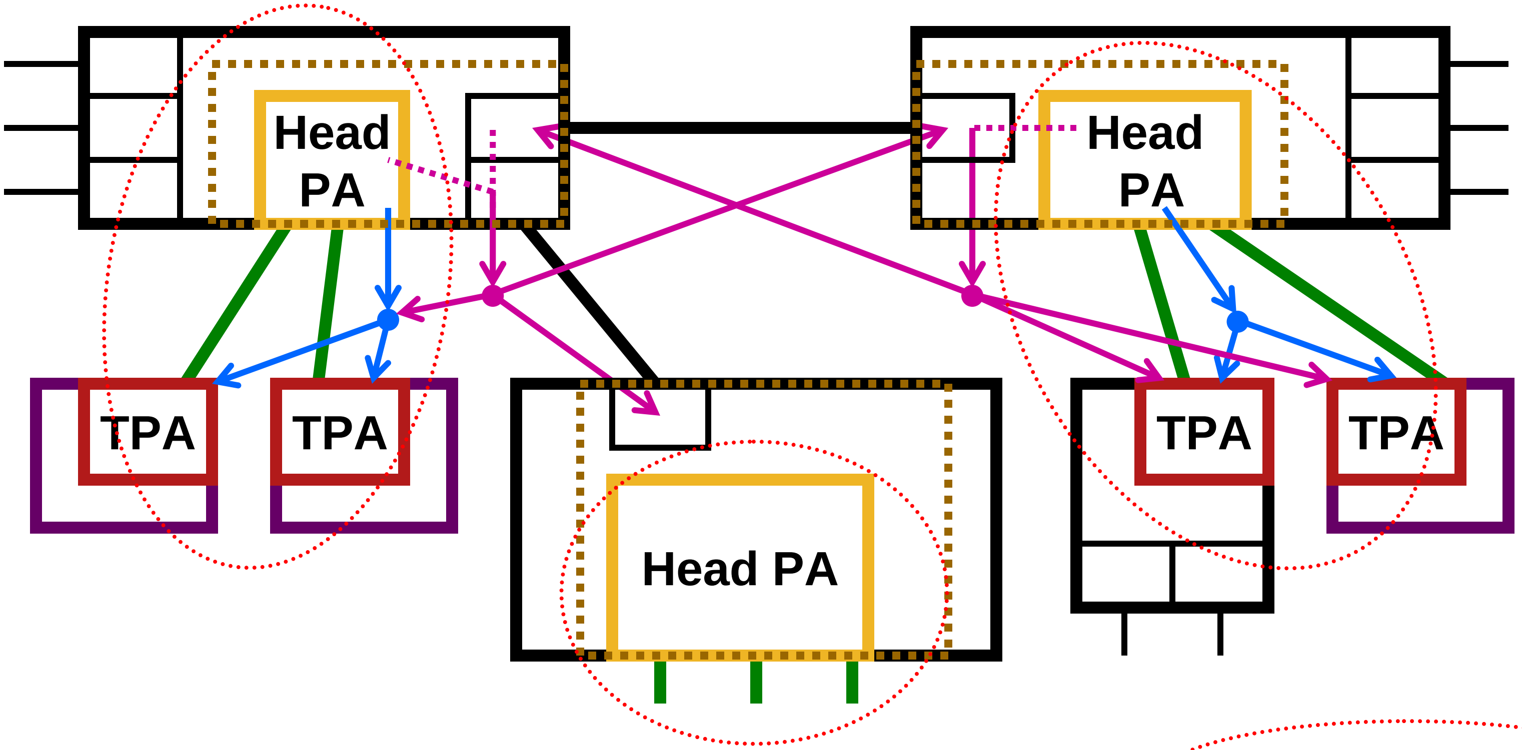
- Wireless inter-AP channels are ordinary point-to-point “other” Bridge Ports, not PAs.
- The AP’s collection of attached Stations are all attached to the AP / Bridge through a single Head Port Association.
- Every Station’s wireless link is a Tail PA.
- The AP and its associated Stations form a single Emulated LAN. The AP’s Head PA has a Local Rootport to all its Stations’ Tail PAs.

802.11 Access Point as a Bridge

- MAC addresses are used as PAIDs in the three- and four-address formats.
- As a Bridge, the AP must forward unknown unicast addresses to at least those Stations that are Bridges.
- As a Bridge, the AP can omit forwarding unknown unicast addresses to Stations that are not Bridges.
- Because of the nature of the medium, the Inter-AP channels, outside the Emulated LAN, can be easily included into a Free Rootlink to optimize bandwidth.

  There must be a means (perhaps one of the four MAC addresses) to determine whether a given frame is being sent to the Local Rootlink or to some particular Free Rootlink.

802.11 Access Point as a Bridge



- These three Access Points each have an Emulated LAN, and use Free Rootlinks to improve bandwidth utilization.
- There must be a means to distinguish between the Local Rootlinks and the Free Rootlinks within the wireless world.

802.11 Four address format

- A Station can be a Bridge!
- Wireless stations normally use a three-address format:

  Either the unicast address of the AP or station by which the frame is to be accepted, or a broadcast address, indicating all stations.

  The Ethernet destination MAC address of the frame.

  The Ethernet source MAC address of the frame.

- A Station sends a multicast by using the AP unicast address as one address, and the usual Ethernet addresses as the other two.
- If the Ethernet destination address is a broadcast or multicast, the AP reflects the frame back to the Stations using the broadcast to indicate “all stations should look at this frame.”
- The originating non-Bridge Station uses the Ethernet source MAC address to discard the reflected frame.

802.11 Four address format

- But, a Station / Bridge cannot discard based on source MAC address; it uses the source MAC address to learn the direction in which to forward frames! (Not doing so is the reason that roaming from AP to AP can cause a 5-minute delay, while the MAC address times out.)
- Therefore the four-address format for 802.11 frames must be used.

  The Station / Bridge puts its own MAC address as the fourth, “origin” address in any frame with a broadcast or multicast Ethernet destination address.

  The AP, when reflecting the broadcast or multicast, preserves the origin address; when transmitting a broadcast or multicast that is not being reflected, the AP either puts its own MAC address as the fourth address, or uses the three-address format.

  The originating Station / Bridge discards any frame whose fourth MAC address matches its own address, and accepts all others and learns from them.

  There are details, here, to be determined.

802.11 Blocked Port Problem

- As described above, the Blocked Bridge Port problem applies to Access Points.
- Either the Combined Mode or the Separated Mode model must be adopted by any given Bridge / Access Point.
- The choice is local to the Bridge / Access Point, as long as the mechanisms used (e.g. the means to disable Leafports) is known to the Tail PAs.

What is left for Emulated LAN and pseudo-Bridge designers?

Still plenty of details to solve

- Obvious things:
  - Encapsulation formats.
  - Rootlink topology, e.g. broadcast server vs. mesh of point-to-multipoints.
  - Creating backup Head PAs for important Tail PAs, and switching over when appropriate.

Still plenty of details to solve

- Not-so-obvious things:

  Frame ordering, e.g. replicate unknown unicasts or use “flush” mechanism on Rootlink.

  Either guarantee full mesh among Head PAs, or create Headlinks dynamically and use static Rootlinks.

  Guarantee that unidirectional Headlinks and Taillinks cannot persist.

  Either ensure that a Tail PA can identify and discard its own multicasts received on the Leafport, even if the Tail PA is part of a bridge, or use Tailports to distribute multicasts from other Tailports.

What Emulated LANs do **not** need to solve

- The relationships between MAC addresses and VLANs.
- The relationships between Emulated LANs and the Spanning Tree Protocols.
- When to forget learned MAC addresses.
- How to signal VLAN distribution.

Summary

Summary

- Adding “Generalized LAN Emulation” rules to IEEE 802.1Q, integrating ELANs into the Relay Function, would take most of the fun out of reinventing bridging, but would do the world a service.

CISCO