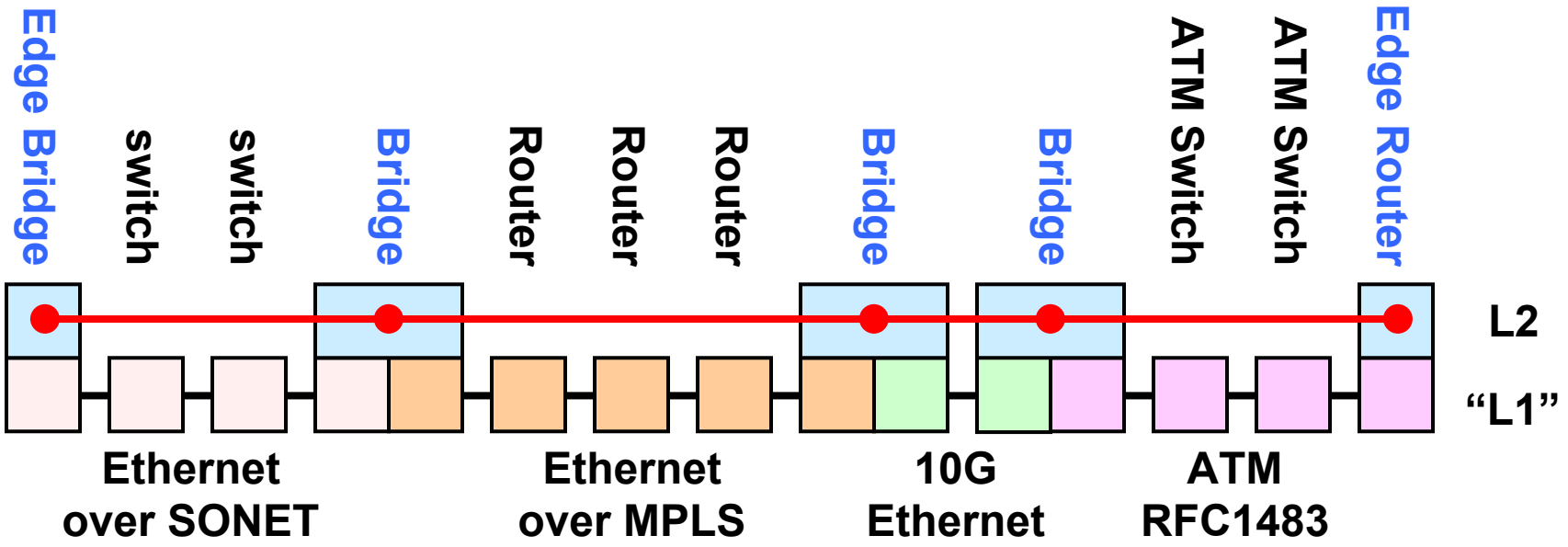
Service Provider Connection Management

Ethernet-layer “OAM” functions for Metro Ethernet Provider Services

Norman Finn

Overview

What is Metro Ethernet Connection Management, a.k.a. “End-to-End OAM”?



- **CM/OAM: Standard Ethernet frames, distinguished from ordinary data frames only by destination MAC address and/or EtherType, and seen, relayed, and/or terminated by Provider Bridges.**

Why do Service Providers want end-to-end OAM?

- Debugging Layer 2 networks is not easy in the Enterprise space.
- Debugging networks of Layer 2 networks promises to be even harder.
- Especially when the component networks belong to different organizations and/or different companies.
- Experience in providing circuit-based services provides a very useful model for managing and debugging nets of nets.

What is *not* Metro Ethernet End-to-End OAM?

- Metro E2E OAM is Ethernet frames, not MPLS, ATM, or SONET frames, cells, or sub-Ethernet control information.
- Only bridges see OAM because bridges are the only active relay functions that exist.
- MPLS, ATM, and SONET OAM are important for debugging Ethernet “wires”, but are not always end-to-end.

(Terminology issue) There is no such thing as a "Metro Ethernet Edge Router"

Cisco.com

- If interoperability is to be achieved among the various technologies used for Metro Ethernet, the IEEE 802 LAN architecture must be observed.
- There are only two kinds of active relay elements in IEEE 802: Bridges and Hubs (Repeaters).
- A "Metro Ethernet Edge Router" is a Bridge that does not need to run spanning tree, has only two ports per VLAN enabled (one trunk and one local Ethernet), and uses Pseudowires for trunks.
- (This notion saves a lot of verbiage in this document. Whether a box is a router or a bridge depends on which function you are looking at.)

Maintenance and other OAM issues not discussed

- Minimal discussion of Provider-to-Customer (single-link) IEEE 802.3ah OAM.
- No explanation of MPLS, ATM, or other OAM.
- No explanation of other techniques such as periodic confirmation of network topology and configuration, SNMP-based “traceroute”, or Layer 3 functions such as Ping.
- No explanation of other protocols such as Ethernet Line Management Interface, BPDUs, etc.

Two standards bodies are defining end-to-end OAM

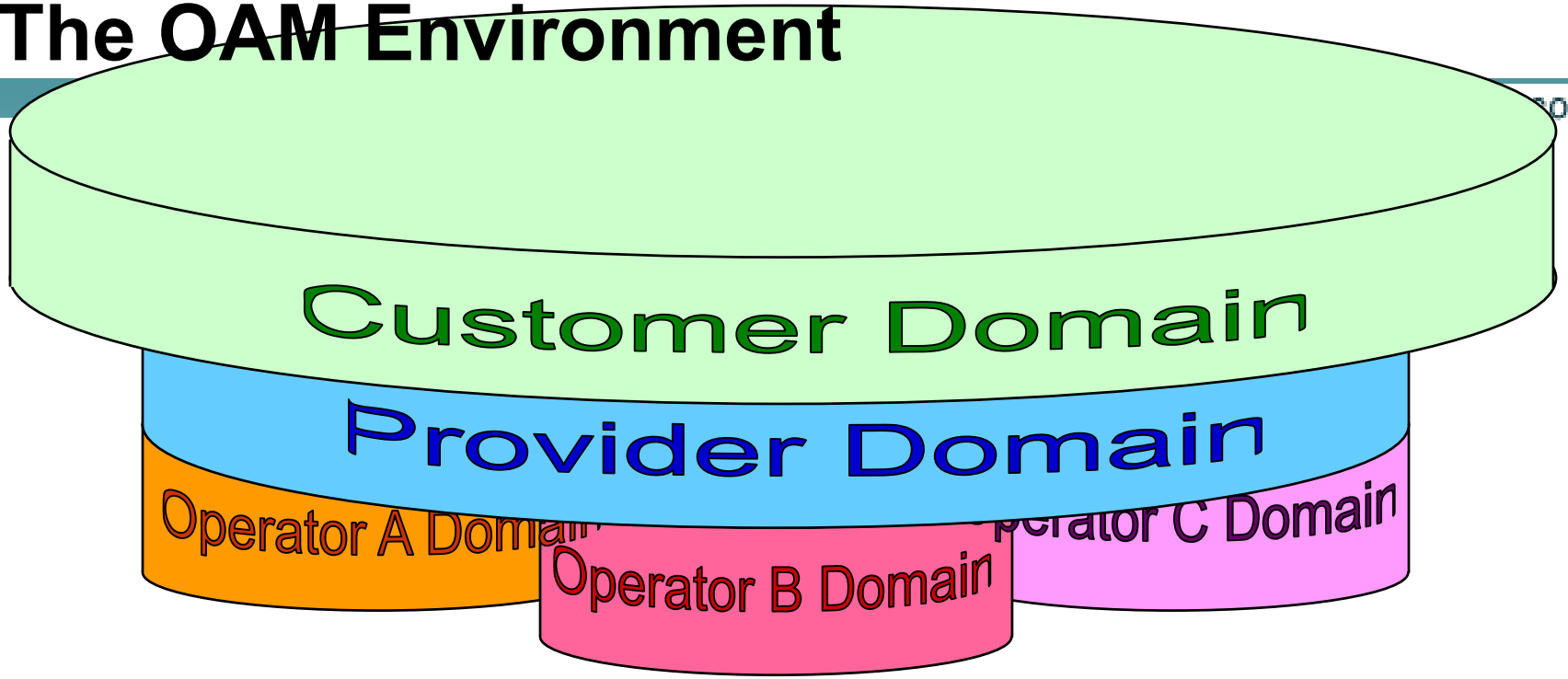
- IEEE 802.1ad defines Q-in-Q “Provider Bridges”. IEEE 802.1ag defines end-to-end Ethernet OAM for those Bridges.
- ITU-T Question 3, Study Group 13, is defining end-to-end Ethernet OAM for both circuit-switched equipment (e.g. Ether-over-SONET) and packet-switched equipment (802.1ad Bridges).
- Both 802.1 and Q.3/13 share common membership and are cooperating fully.

  802.1 should define low-level aspects tied closely to bridging technology.

  Q.3/13 should define high-level aspects tied to the service models.

Domains, Maintenance Levels, and Flow Points

The OAM Environment



- **Customer contracts with Provider for end-to-end service.**
- **Provider contracts with Operator(s) to provide equipment and networks.**
- **Provider and Operator(s) may or may not be the same company or same division.**

OAM Flow Points

- In ITU, domains are defined in terms of "flow points", which are "MACs" to IEEE 802, and "interfaces" or "ports" to others.
- A flow point (FP) at the edge of a domain is called a "Maintenance Point" or MP.

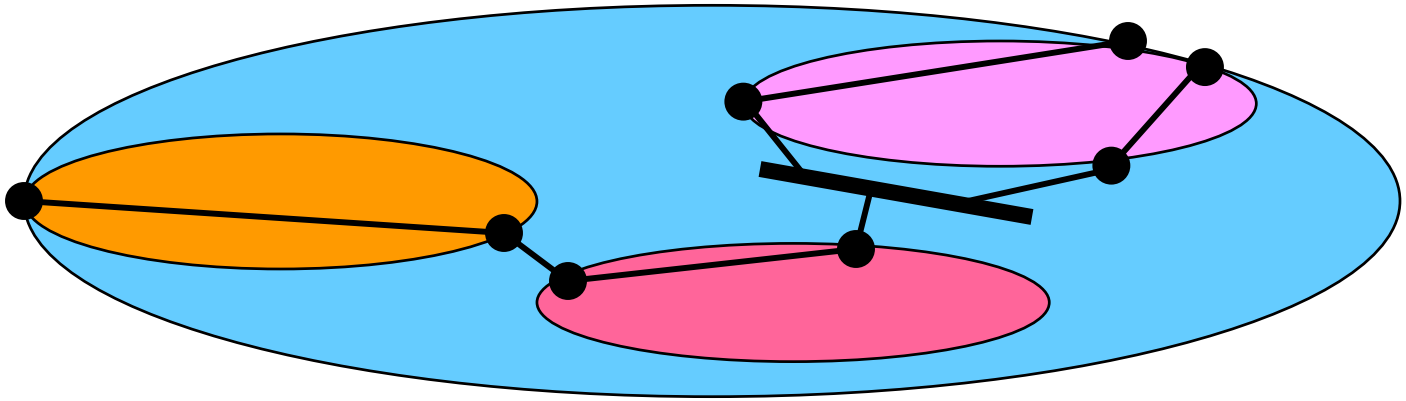
  System administrators use MPs to initiate and monitor OAM activity and report the results.

- A flow point inside a domain, and visible to an MP, is called a "Loopback Point".

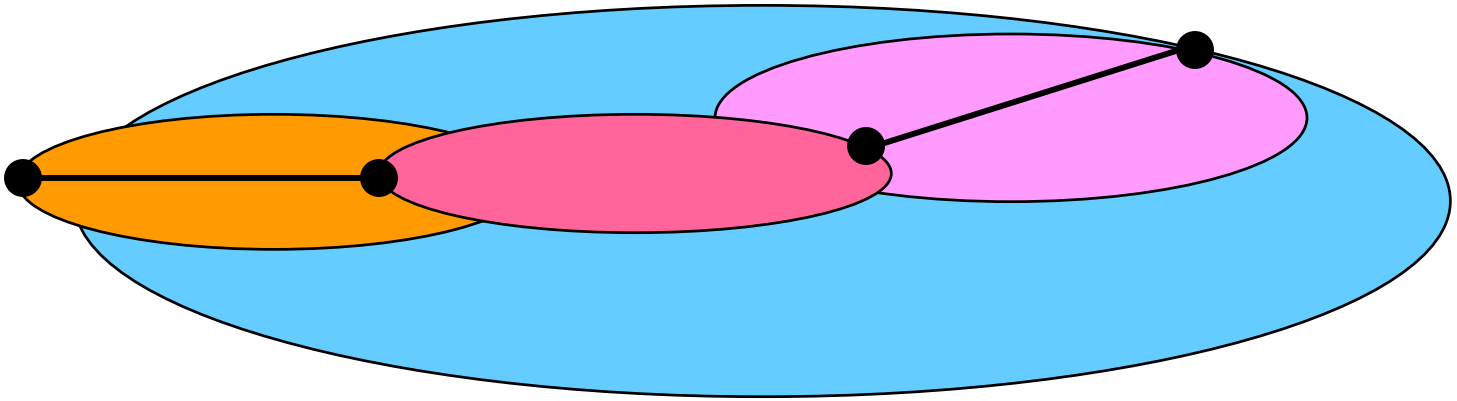
  Loopback Points (LPs) passively receive and respond to OAM packets initiated by MPs.

OAM Domains

OK

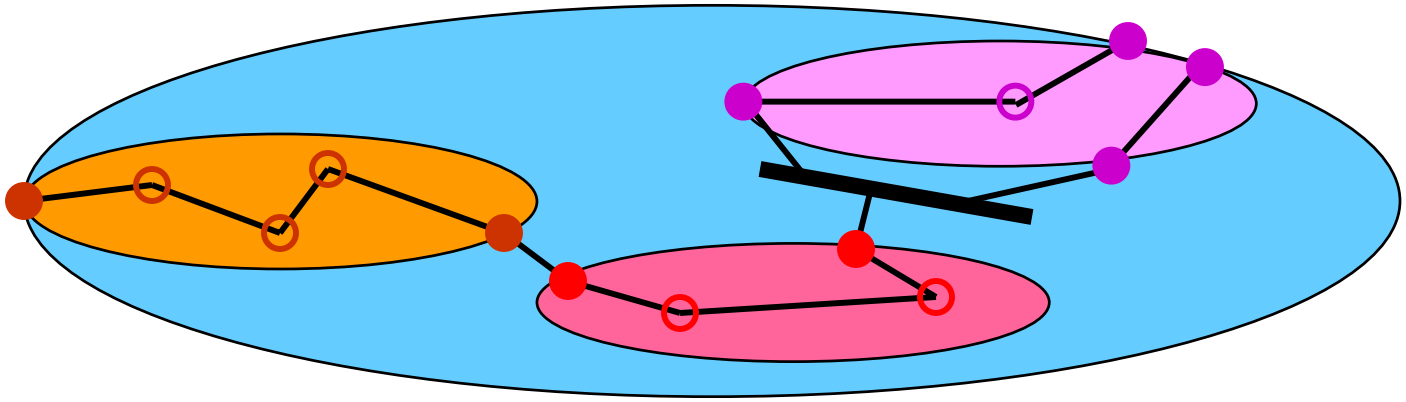


NO



- OAM Domains may nest or touch, but must never intersect.

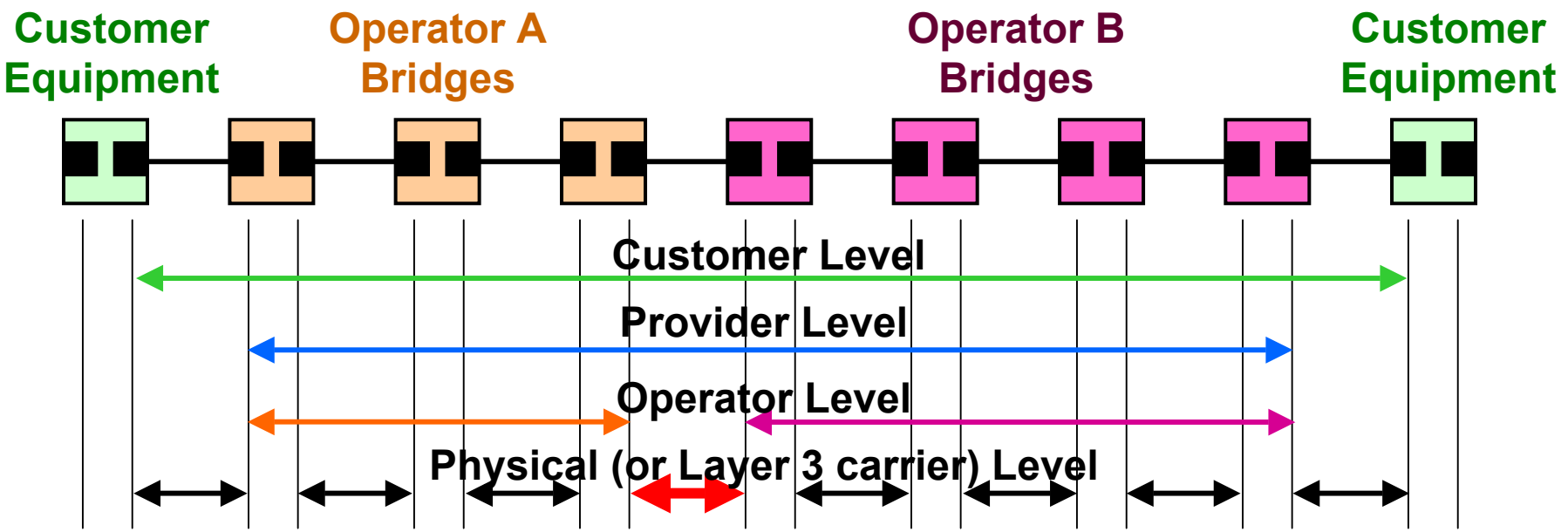
OAM Domains



● ● ● Maintenance Points ○ ○ ○ Loopback Points

- Maintenance Points are always at the edges of Domains.
- Loopback Points are always within domains.

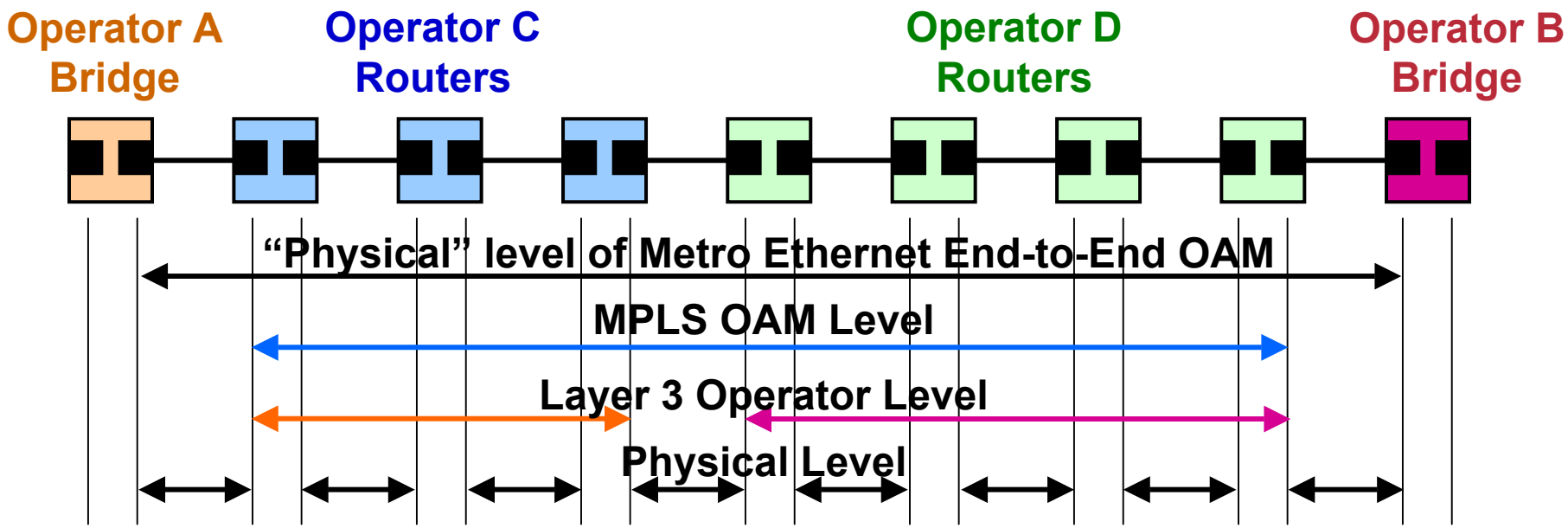
OAM Maintenance Levels



- **At (at least) Operator and Physical levels, there may be multiple Domains.**

  There *could* be multiple Domains at any level.

Lower Maintenance Levels (Expanding the red link in the previous slide)

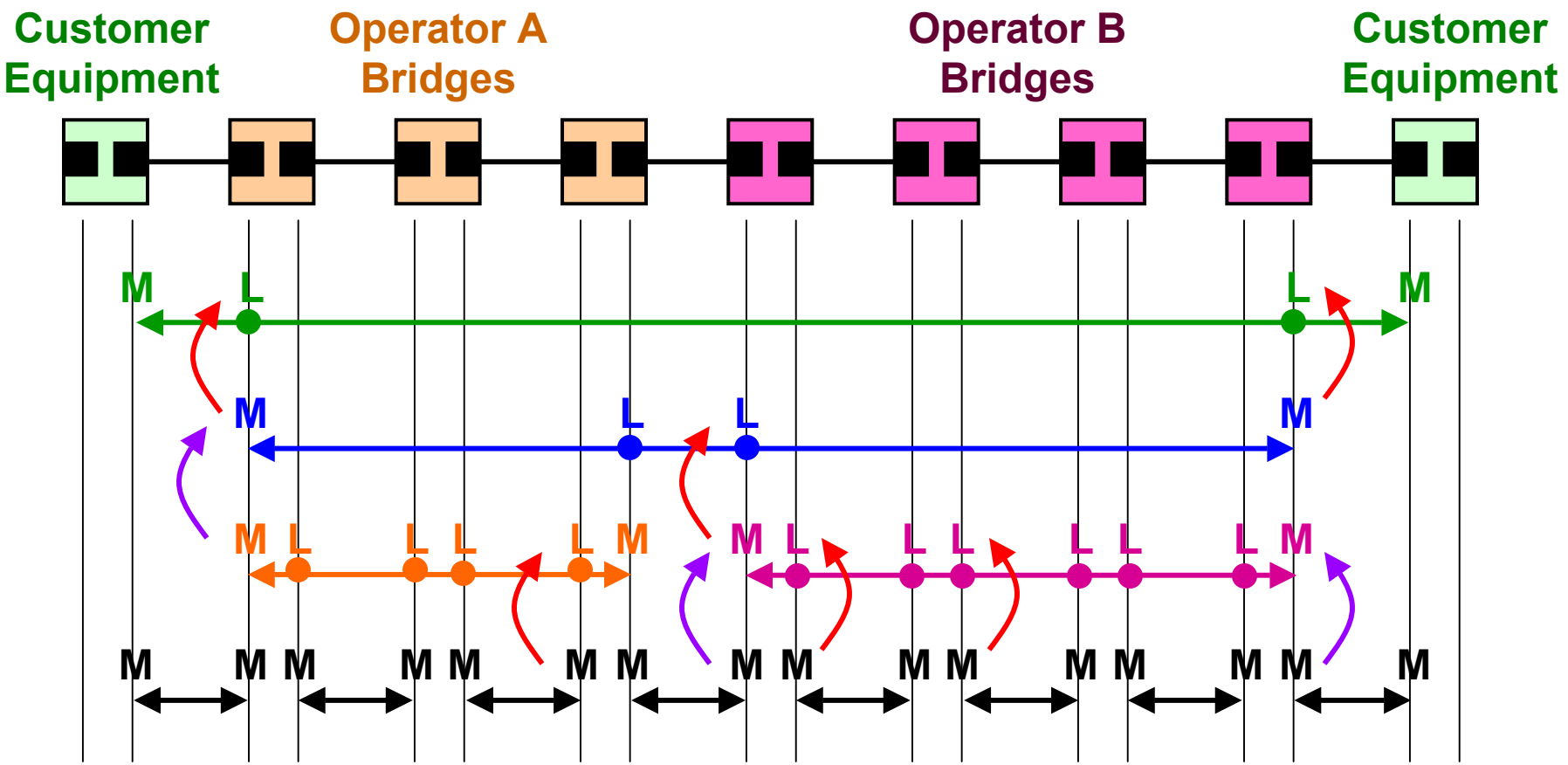


- There are levels below End-to-End OAM.

  This is where MPLS OAM, ATM OAM, etc., live.

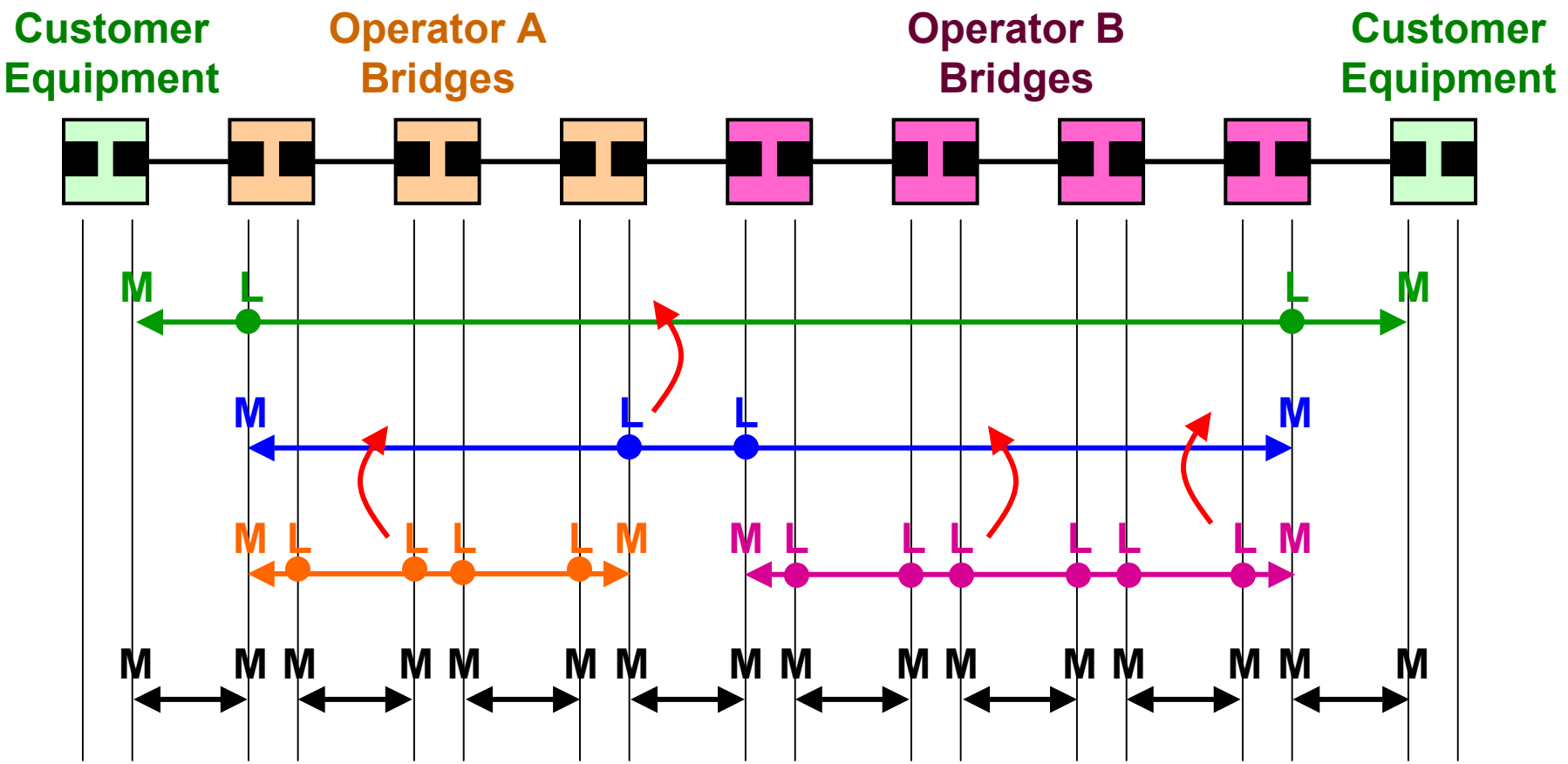
  These OAM functions *ought to*, but do not *necessarily*, follow the MP/LP/FP model.

Loopback Points and Maintenance Points



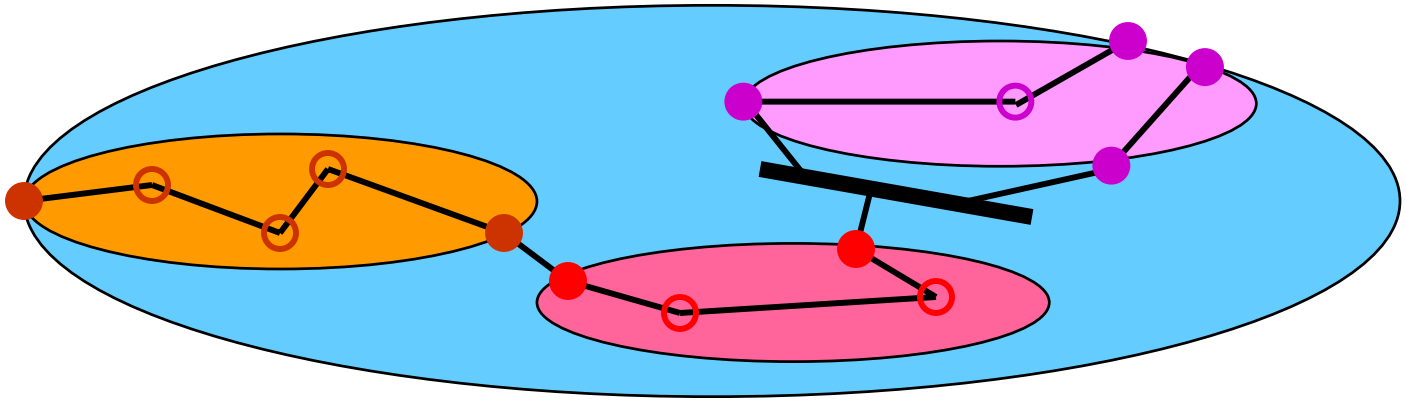
- Each Level's Maintenance Points are the next-higher-level's Maintenance or Loopback Points.

Loopback Points and Maintenance Points



- Each Level's Loopback Points are invisible to all higher levels.

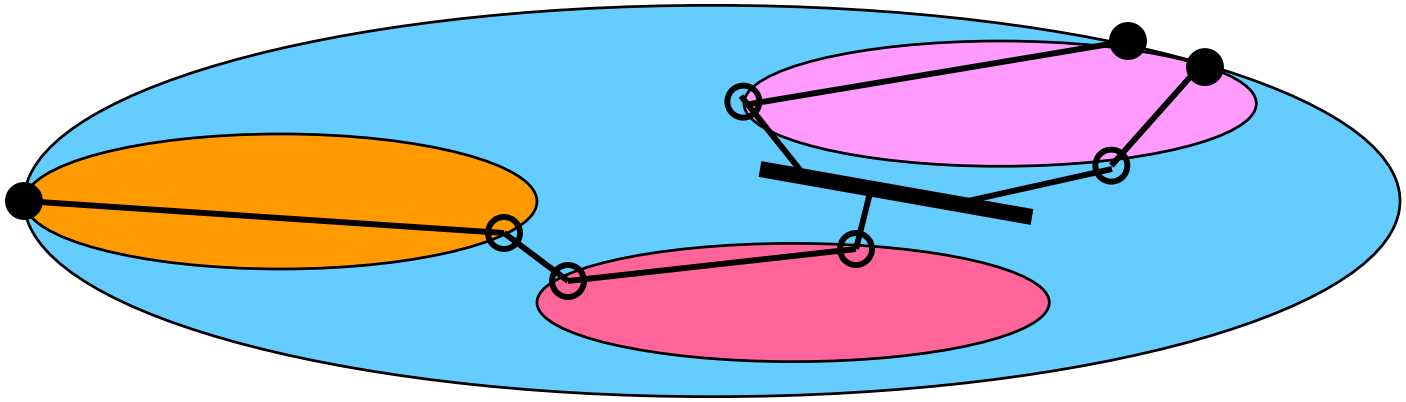
Loopback Points and Maintenance Points: Operator Levels (x3)



● ● ● Maintenance Points ○ ○ ○ Loopback Points

- Each Level's Maintenance Points are the next-higher-level's Maintenance or Loopback Points.
- Each Level's Loopback Points are invisible to all higher levels.

Loopback Points and Maintenance Points: Provider Level

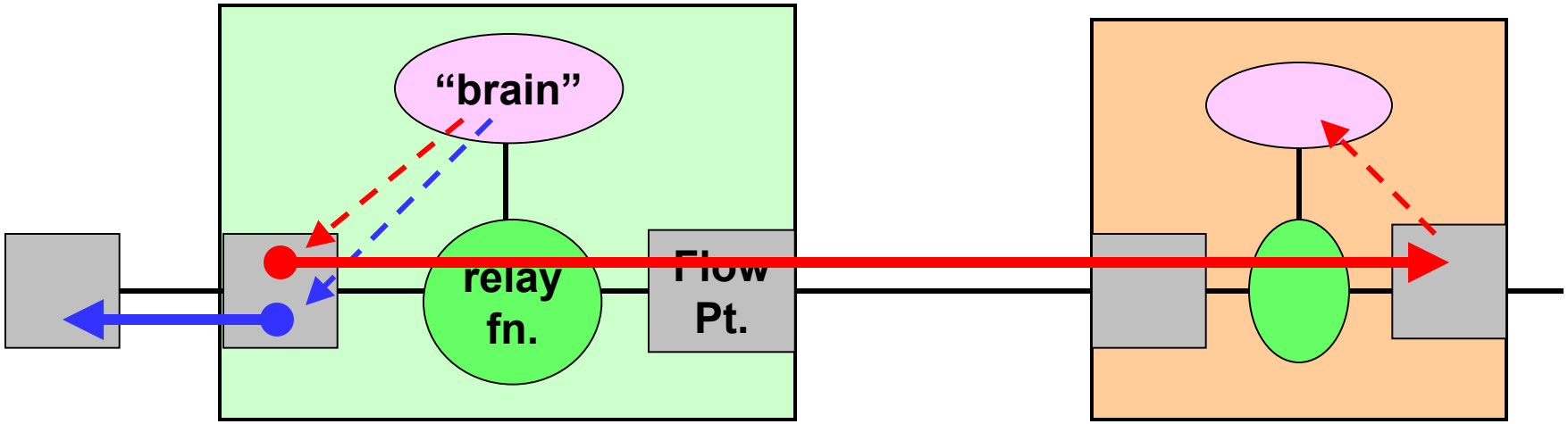


● Maintenance Points ○ Loopback Points

- Each Level's Maintenance Points are the next-higher-level's Maintenance or Loopback Points.
- Each Level's Loopback Points are invisible to all higher levels.

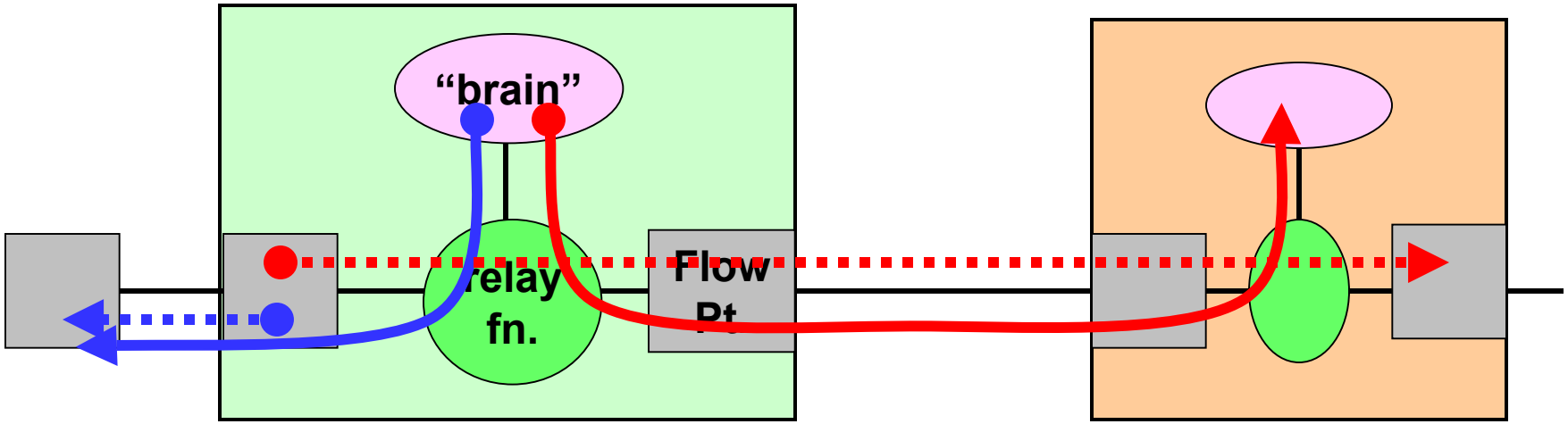
Using the “Back Side” of the MAC

Using the “back side” of the MAC



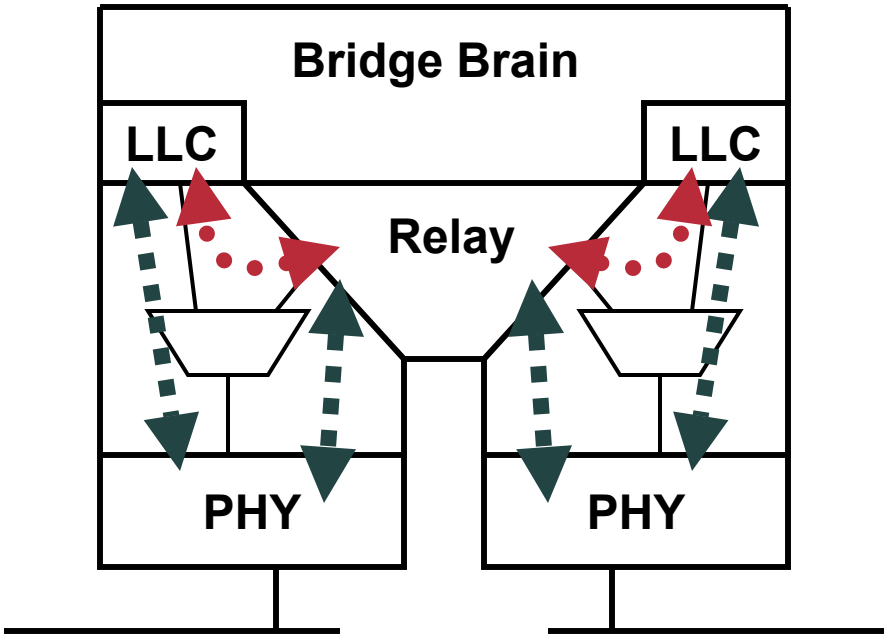
- *At least in theory,* an Ethernet frame may be sent or received from either direction in a MAC: to/from the PHY (the normal case), or to/from the relay function!
- E2E OAM uses this “back door” capability.

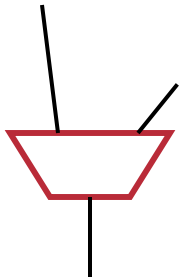
Using the “brain” of the bridge



- *At least for now*, most bridges’ “brains” actually emit and receive the OAM packets via the Relay Function.
- Additional hardware paths may needed to implement the E2E OAM model.

The 802.1Q “baggy pants” diagram:
As it is often implemented

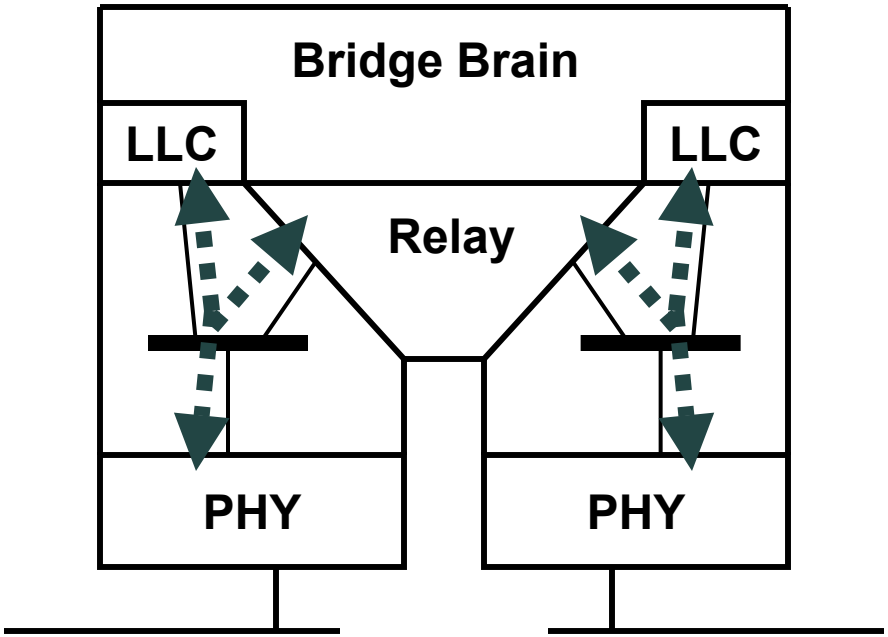


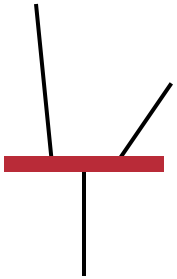


This is a **two-to-one mux**.

- **Disallowed Paths:**
  - LLC to Relay
  - Relay to LLC
- **Allowed Paths:**
  - LLC to PHY
  - PHY to LLC
  - Relay to PHY
  - PHY to Relay
  - PHY to LLC & Relay

The 802.1Q “baggy pants” diagram:
As IEEE 802.1D and Q specify it



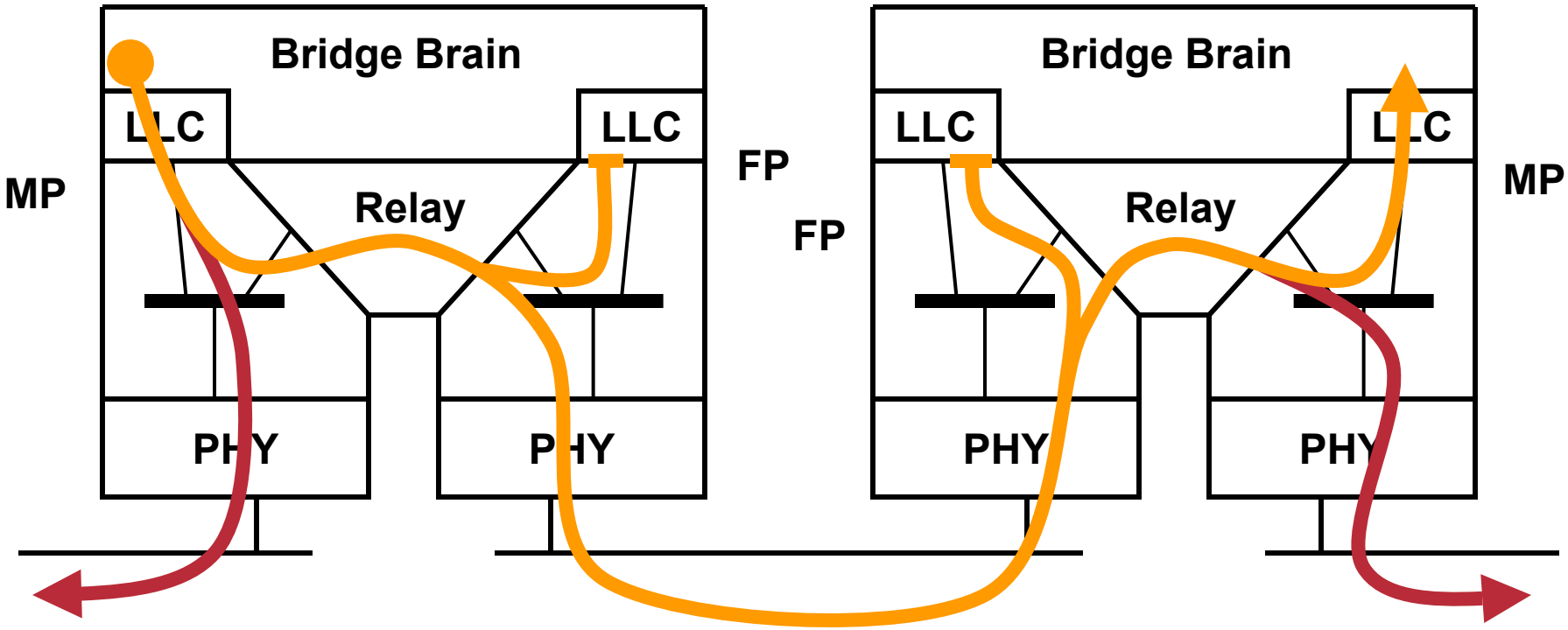


This is a shared medium.

- **Disallowed Paths:**
  - **Point-to-point!**
- **Allowed Paths:**
  - **LLC to PHY & Relay**
  - **Relay to PHY & LLC**
  - **PHY to LLC & Relay**

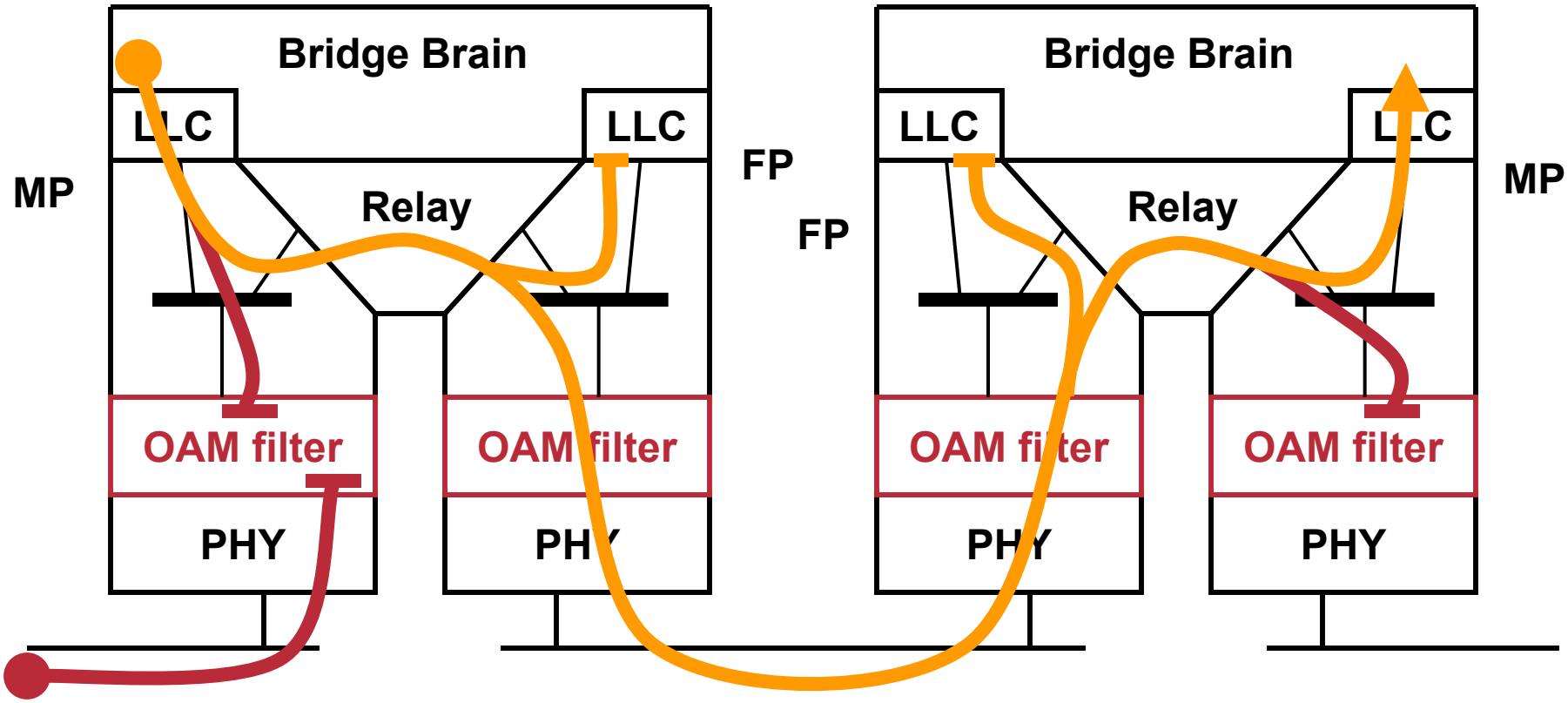
The 802.1Q “baggy pants” diagram: Frame flow for an ordinary multicast

Cisco.com



- But the OAM packets would escape from both ends of the path and another Domain (e.g. the Customer) would receive them!

The 802.1Q “baggy pants” diagram: As E2E OAM would like to see it



- An OAM filter layer blocks OAM frames based on Maintenance Level.

The “Baggy Pants” OAM Filter

- **OAM Filter stops OAM packets that are at or below a certain Maintenance Level.**
- **Required for Maintenance Points only.**
- **Multicast OAM packets can be filtered by destination MAC address to achieve *almost* the same result.**
- **But the OAM Filter is needed to prevent the inadvertent distribution of OAM packets destined for unknown unicast addresses.**

End-to-End OAM Packet Types

Three Kinds of OAM Packets

- **Continuity Check**

  Multicast from MP. Received by MPs and LPs. Catalogued by receiving MPs.

- **Traceroute**

  Next-hop Multicast from MP to next MP or LP along route. Receiver both replies with unicast to original MP, and sends Traceroute to next MP/LP.

- **Loopback (Ping)**

  Unicast from MP to LP or MP, which replies with unicast to originating MP.

All OAM Packets must include

Cisco.com

- Destination MAC addr (uni- or multicast).
- Source MAC addr of sending Flow Point.
- Customer Service Instance (CSI) indicator, typically an S-VLAN ID in an 802.1ad tag.
- EtherType, Version, and Op-code.
- Maintenance Level indicator (an integer).
- Transaction Identifier (an integer).
- Universally unique CSI identifier string (UUCSIID).

All OAM Packets may include

- **Arbitrary amount of arbitrary data.**

  Data MUST NOT be interpreted by replying FP.

  Supports testing frame size limitations and testing of Service Level Agreements.

- **Internal Checksum computed on all of frame except .1Q/.1ad tags, which are subject to alteration.**

  This may detect bad FCS hardware.

Continuity Check: Transmission

- Transmitted periodically from each Maintenance Point on each CSI (S-VLAN).
- Timing depends on administrators' needs; perhaps 1 second intervals, perhaps 600.
- Contains the basic info, plus an MP Identifier (MPID) unique over CSI, a "validity time" of 2.5 or 3.5 times the transmission period, and port state info.
- There is one standard reserved Continuity Check multicast destination MAC address defined for ***each*** Maintenance Level.

Continuity Check: Transmission

- Separate CCs are used at every level, from each MP (one bridge may have multiple MPs for same CSI), for each CSI.
- Transaction ID is initialized to random value, incremented once per transmission from a given MP, CSI, and Level.
- Failure detection speed must be balanced against CC loss rate and overhead.

  E.g. Validity time of 2.5 transmission interval allows loss of 1 CC.

Continuity Check: Reception

- Received in each Maintenance Point at the Maintenance Level and on the CSI of the sender.
- No reply is made to a Continuity Check.
- Received information is saved in a table indexed by MPID for use by Traceroute or Ethernet Line Management Interface (ELMI).
- If validity time of a table entry expires, saved information is discarded.

Continuity Check: Errors Detected

Cisco.com

- Receiver may compare list of saved MPIDs to a configured list, and complain of mismatches.
- A CC received with a UUCSIID / S-VLAN mismatch indicates cross-connected CSIs.
- Duplicate or out-of-order Transaction IDs may indicate a forwarding loop in Provider space.
- Port state is communicated to all MPs for relay to the customer via ELMI.

Continuity Check: Loopback Points

Cisco.com

- Loopback Points also receive Continuity Checks.
- Unlike MPs, LPs *need not* log CCs by source MP.
- LPs remember only the source MAC address, Maintenance Level, S-VLAN, and ingress bridge port in a (software?) CC database separate from the bridge's main Filtering Database (L2 forwarding table).
- This information is used by the Traceroute function.

Traceroute Message Initiation

- Originator MP initiates Traceroute Message (TM). Included in packet are:

  Target MAC address, always a Maintenance Point

  Originator's MAC address

- TM also includes a Hop Count.
- There is one standard reserved Traceroute multicast destination MAC address defined for each Maintenance Level.

  (For tracing at higher Maintenance Levels.)

Traceroute Message forwarding

- All MPs and LPs at a given level intercept all TMs at that level, using destination MAC address.
- Each MP or LP intercepting a TM:
  1. If TTL is not 0, decrements TTL and:
  2. Sends unicast Traceroute Reply (TR) to Originator MP.
  3. If TTL still not 0, forwards the TM to next hop along path to target MAC address.

Traceroute Message
Next Hop determination

- The bridge looks up the target MAC address in the Filtering Database to find the egress FP.
- If not found, the Continuity Check database is examined.
- If the egress FP matches the ingress FP, then no Traceroute Reply is returned.

  (This prevents multiple answers after hopping over a shared medium.)
- Else, the next hop cannot be found.

Traceroute Reply

- **Traceroute Reply always contains the TTL value, after decrementing, of the intercepted Traceroute Message, the Transaction ID, the device ID of the replying bridge that owns the FP, and the replying bridge's Layer 3 management address, all in 802.1ab LLDP format.**
- **In theory, TR comes from a Flow Point, not a Bridge's brain. Therefore, a Bridge may issue either two TRs for the ingress and egress FPs, or one TR for both.**

Traceroute Reply Contents:
Flow Point at ingress to Bridge

- Ingress Flow Point ID and Type and Egress Flow Point ID.
- Reason for not forwarding TM, e.g.:

  Ingress FP's STP/VLAN state not Forwarding;
  Ingress FP's ifOperState is not UP;
  Traceroute Message's TTL expired; or
  No egress FP could be identified.

- If forwarded, method for finding Egress FP, either:

  Egress FP from Filtering Database; or
  Egress FP from Continuity Check database.

Traceroute Reply Contents:
Flow Point at egress from Bridge

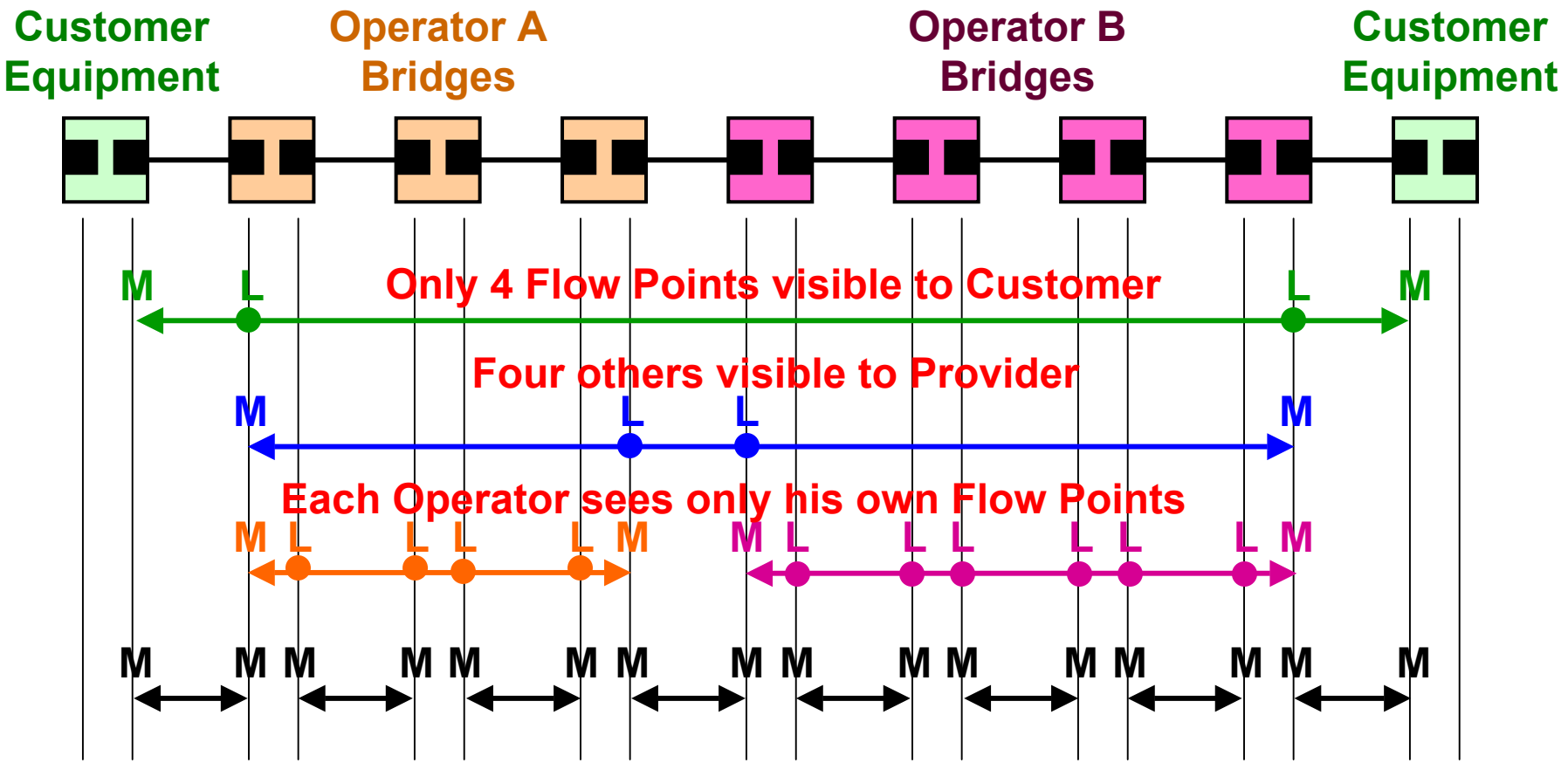
Cisco.com

- Egress Flow Point ID and Type.
- Device and Flow Point IDs of next hop bridge (or other device), if unambiguous.
- Reason for not forwarding TM, e.g.:

  TM was forwarded; OR

  Egress FP's ifOperState is not UP;
  Egress FP/VLAN STP state not Forwarding; or
  Traceroute Message's TTL expired.

Traceroutes at different Maintenance Levels see different pictures



- Each Level's Maintenance Points see only that Level's MPs and LPs.

Traceroute: It's never there when you need it!

- Just when you need it most, a Traceroute based *solely* on the Filtering Database would likely yield no answer at all!

  Five minutes after an MP disappears, its MAC address is forgotten.

  Immediately after a Topology Change, a lost MP's MAC address may be deleted.

- Hence, Loopback Points remember Continuity Checks' source MAC addresses, and Traceroute uses them.

Loopback Message and Loopback Reply

- **Always sent by an MP *only* to a unicast MAC address of another MP or LP.**

  Address is obtained from a previous Continuity Check or Traceroute operation.

- **Receiving MP or LP responds with a Loopback Reply, which has a different Op-code.**

  All optional fields are returned in the Loopback Reply.

- **Transaction ID allows originating MP to match Loopback Messages to Replies.**

Controlling OAM

A Layer 3 overlay is required

- At least one “Customer Service Instance” will be used by the Provider(s) and/or Operator(s) to provider Layer 3 services for managing the bridges.
- This may be a private network, or may be attached to the Big-I Internet.
- Routing functions may be embedded in the Provider Bridges, or may be separate devices.
- This is of no concern to IEEE 802.

Different organizations may control the same Bridge

- One Operator might provide two UNI ports for two Customers of two different Providers on the same Provider Bridge.
- Although the Operator has the primary responsibility of configuring the Bridge, each Provider requires control of the OAM functions at the Provider Level on those Ports assigned to that Provider as Maintenance Points.
- SNMPv3 is probably the best way to allow a Provider to access OAM controls.

Continuity Check Management

Cisco.com

- **Controls:**
  - Transmission period and validity time
  - Optional information (e.g. interior checksum)
  - Configured list of peer MPs
  - Layer 3 address and options for error reporting
- **Reports:**
  - Maintenance Level (read-only for CC operator, read/write for Bridge owner/operator)
  - CC database of discovered MP addresses
  - Error conditions

Traceroute Management

- **Controls:**
  - Initiate Traceroute
  - Select MP (MAC address) to trace
  - TTL values
  - Layer 3 address and options for error reporting
- **Reports:**
  - Traceroute Replies
  - Error conditions

Loopback Management

- **Controls:**

  Initiate Loopback sequence

  Select FP or MP (MAC address) to ping

  Optional information (e.g. arbitrary data)

  Layer 3 address and options for error reporting

- **Reports:**

  Successful Loopback Replies

  Error conditions

Usage Scenarios

Continuity Check is the backbone of OAM

- **Continuity Check** enables the Ethernet Line Management Interface (ELMI) to report the status of the other UNIs of the service to the Customer.
- **Continuity Check** also allows MPs to catalog each other, and to associate a MAC address with each other MP.
- MAC addresses of other devices must be discovered via **CC**, not configured.

  (Line cards may be swapped.)

Traceroute discovers the Loopback Points

- If CC indicates loss of another MP, Traceroute Message is sent to trace the path from the MP detecting the failure to the saved MAC address of the failed MP.

  This may yield the reason for the loss of connectivity.

  This will also yield a list of Loopback Points' MAC addresses to use in Loopback Messages.

- Traceroute may also be used to obtain the list of Loopback Points' MAC addresses when CC is ***not*** indicating any error.

Loopback allows detailed checking of paths

- **Loopback Messages** to successive Loopback Points along the path to another Maintenance Point may detect failures not found by **CC** or **Traceroute**.

  TM is hop-by-hop, and **CC** is a multicast. **LM/LR** are unicast.

  LM/LR can be used to detect links with too-small MTUs.

- **Loopback** may also be used for testing Service Level Agreement compliance: throughput, packet loss rate, etc.

Drilling Down

- **If the failure can be isolated to a particular link using the above methods, then that link can be debugged.**
- **If the failing link is not a simple physical link, e.g. it is an Emulated LAN using Pseudowires over MPLS, then the OAM capabilities of that link must be used to debug that link.**

Summary

Metro Ethernet End-to-End OAM on one slide

- Different organizations' Domains are nested in Levels.
- Maintenance Points are active, Loopback Points are passive, other Flow Points are transparent.
- At least these five OAM packets are needed: Continuity Check, Traceroute Message, Traceroute Reply, Loopback Message, and Loopback Reply.
- An "OAM Filter" function is needed in the Baggy Pants diagram.


CISCO SYSTEMS