River STONE NETWORKS™

Tutorial VPLS

GTER-16

Igor Giangrossi
igor@riverstonenet.com

River STONE NETWORKS

Agenda

- Overview
- IETF Status
- Review: Martini draft
- LDP and BGP signaling approaches
- Operation: Control Plane
- Operation: Data Plane
- Scaling VPLS: HVPLS
- Comparison: VPLS and RFC2547
- References

Agenda

- **Overview**
- IETF Status
- Review: Martini draft
- LDP and BGP signaling approaches
- Operation: Control Plane
- Operation: Data Plane
- Scaling VPLS: HVPLS
- Comparison: VPLS and RFC2547
- References

Technology Trends

- **IP is the dominant technology;**
- **Ethernet**
  - Dominant in LANs
  - New access technology for MANs
  - New backbone technology for MANs
- **MPLS as a necessary tool**
  - Traffic Engineering
  - VPNs

Traditional VPNs

- **L2 VPNs:**
  - Leased Lines
  - ATM
  - Frame Relay
  - L2TP
- **L3 VPNs:**
  - IPSec
  - GRE
  - PPTP

Ethernet VPNs

- **Native Ethernet protocols (802.1) insufficient for MANs:**
  - STP/RSTP/PVST/MSTP;
  - GARP/GVRP;
  - 802.1Q VLANs;
- **The IEEE is working on some improvements:**
  - Provider Bridges (802.1ad)
- **Ethernet alone lacks OAM, traceability, resiliency facilities**

Ethernet / IP and MPLS

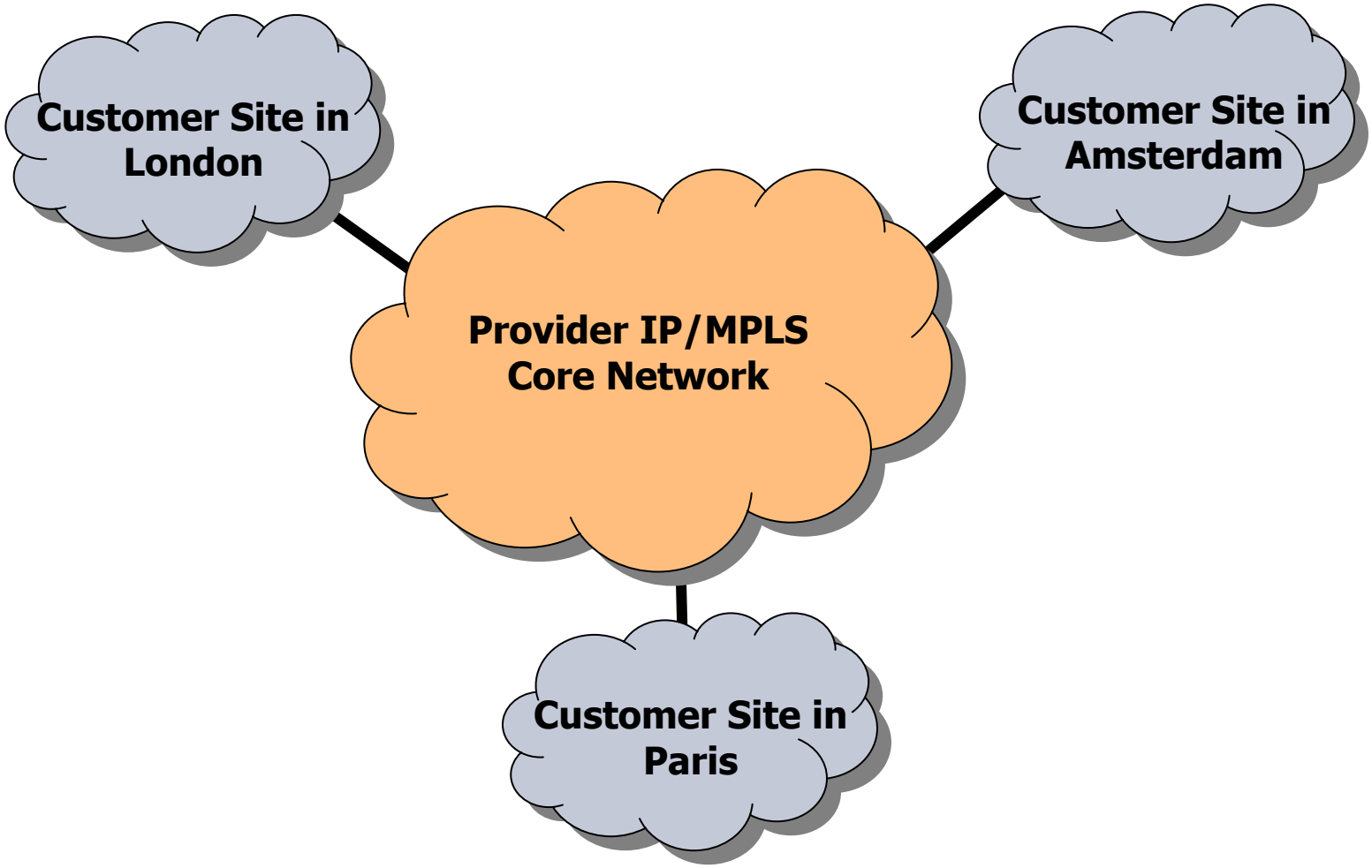
- **MPLS brings additional features to Ethernet / IP:**
  - IP Infrastructure – relies on Routing Protocols for resiliency;
  - Connection-oriented tunnels
  - Traffic Engineering tools
  - VPNs
  - Improved and unified scheme for QoS
  - Core equipments don't maintain VPN information
- **Solution: use MPLS for Ethernet VPNs!**

MPLS VPNs

- **L3 VPNs:**
  - RFC2547: BGP/MPLS VPNs
    - IP Traffic only
- **L2 VPNs:**
  - Point-to-Point: Martini tunnels
    - Generic L2 point-to-point technology
  - Multipoint: VPLS
    - Specific for Ethernet
- **More details to come...**

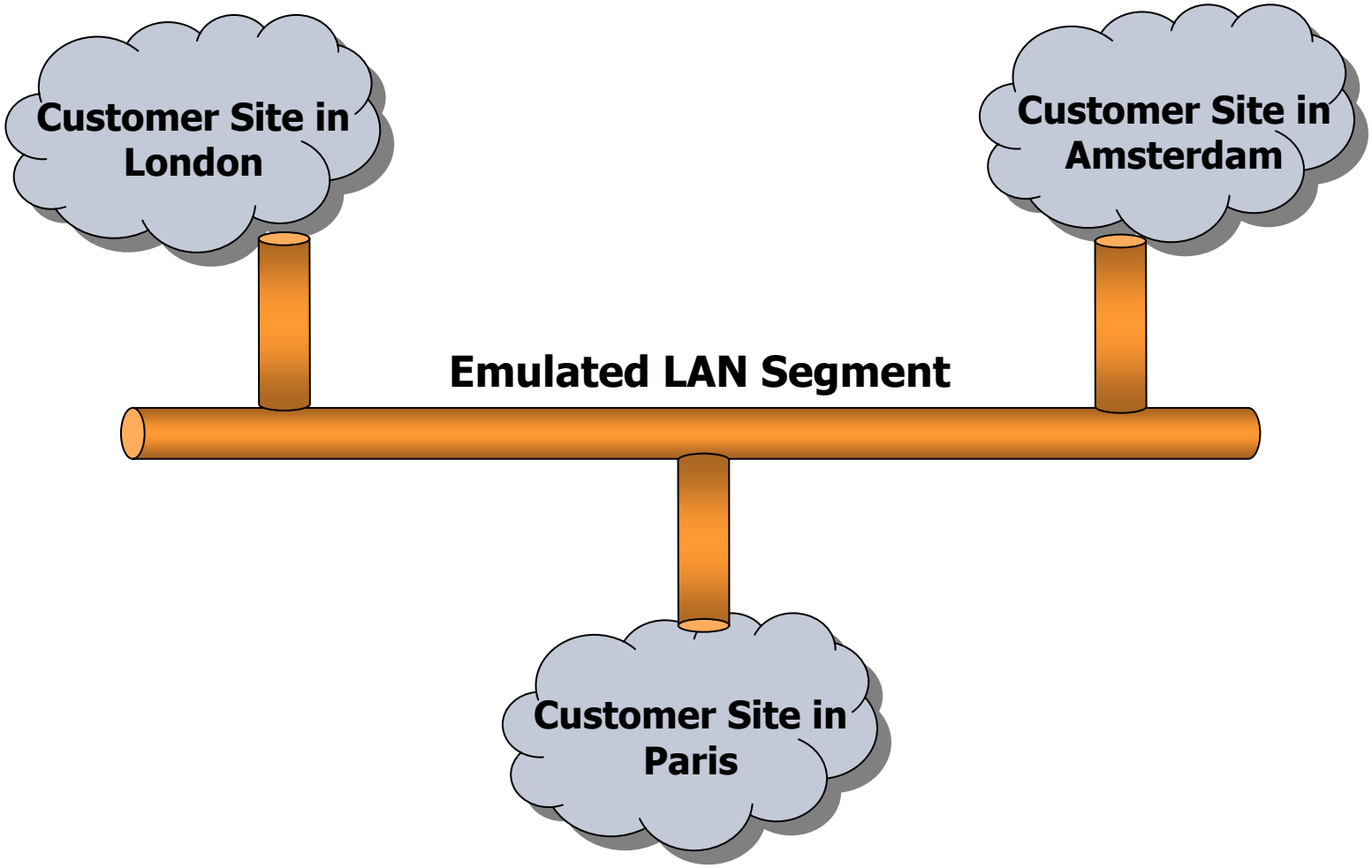
Virtual Private LAN Services

MPLS Multipoint Service



Virtual Private LAN Services

MPLS Multipoint Service



Agenda

- Overview
- **IETF Status**
- Review: Martini draft
- LDP and BGP signaling approaches
- Operation: Control Plane
- Operation: Data Plane
- Scaling VPLS: HVPLS
- Comparison: VPLS and RFC2547
- References

Point to Point Drafts

- **The Martini draft is now part of a Working Group – PWE3**
  - draft-ietf-pwe3-ethernet-encap-04.txt
  - draft-ietf-pwe3-control-protocol-04.txt
- **There are other drafts for the transport of other technologies over MPLS**
- **The Ethernet draft is very close to become an RFC**

Multipoint Drafts

- **Two solutions were chosen by the L2VPN Working Group:**
  - draft-ietf-l2vpn-vpls-ldp-01.txt
    - Former Lasserre-vKompella draft
  - draft-ietf-l2vpn-vpls-bgp-00.txt
    - Former Kompella draft

draft-ietf-l2vpn-vpls-ldp-01.txt

- **Uses LDP for signaling the VPNs**
- **It is basically an extension to the Martini draft**
- **Industry Support:**
  - Riverstone
  - Nortel
  - Alcatel/Timetra
  - Foundry
  - Extreme
  - Cisco

draft-ietf-l2vpn-vpls-bgp-00.txt

- **Uses BGP for signaling and discovery**
- **Similar to RFC2547 on signaling**
- **Similar to Martini on encapsulation**
- **Industry Support**
  - Juniper

Agenda

- Overview
- IETF Status
- **Review: Martini draft**
- LDP and BGP signaling approaches
- Operation: Control Plane
- Operation: Data Plane
- Scaling VPLS: HVPLS
- Comparison: VPLS and RFC2547
- References

Martini Tunnels
MPLS Point-to-Point Service

- Point-to-Point tunnel to transport L2 frames across a MPLS backbone;
- 2 uni-directional LSPs forming a bi-directional pipe;
- There's a draft defining signaling and several drafts defining the encapsulation of frames;

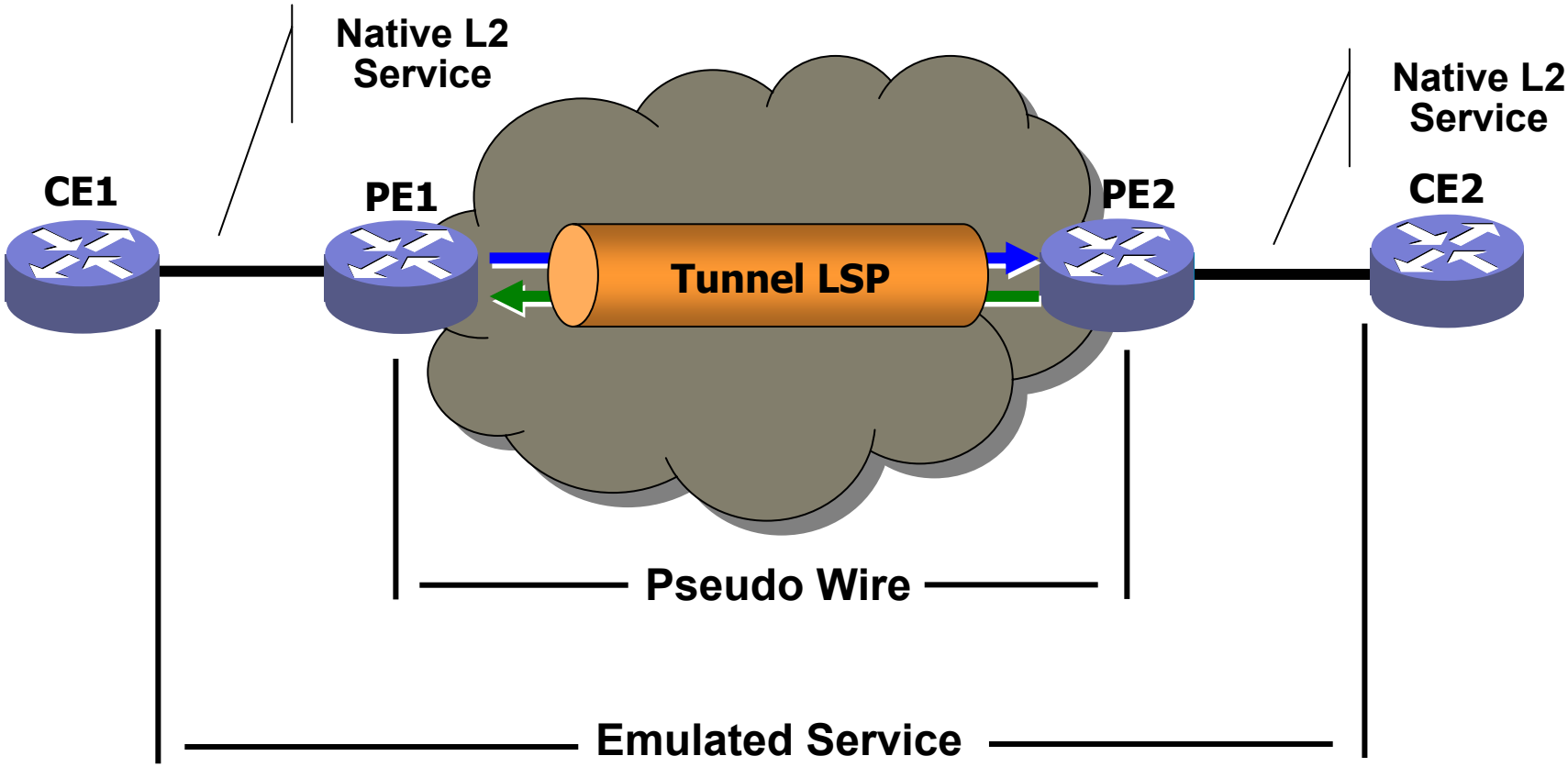
Martini Encapsulation Types

- Frame Relay
- ATM (7 modes available)
- Ethernet VLAN
- Ethernet
- PPP
- HDLC
- SONET/SDH

Martini Control Protocol

- Extends LDP to signal “demultiplexor” labels for the pseudowires;
- Uses Targeted LDP sessions for label distribution;
- Tunnel LSPs can be Traffic Engineered for specific QoS demands

Martini Reference Model



LDP Details

- **Label Mapping messages are exchanged between participating PEs to create the tunnels**
- **Message has:**
  - FEC TLV
    - PWid FEC Element or
    - Generalized ID FEC Element (not used often)
  - Label TLV
    - Generic LDP Label TLV
- **Label Withdrawal messages are used to tear down the tunnels;**

PWid FEC TLV Format

PW TLV	C	PW Type	PW Info Length
PW Group ID			
PW ID			
Interface Parameters “ “			

Interface Parameters TLV

Parameter ID	Length	Variable Length Value
Variable Length Value “		

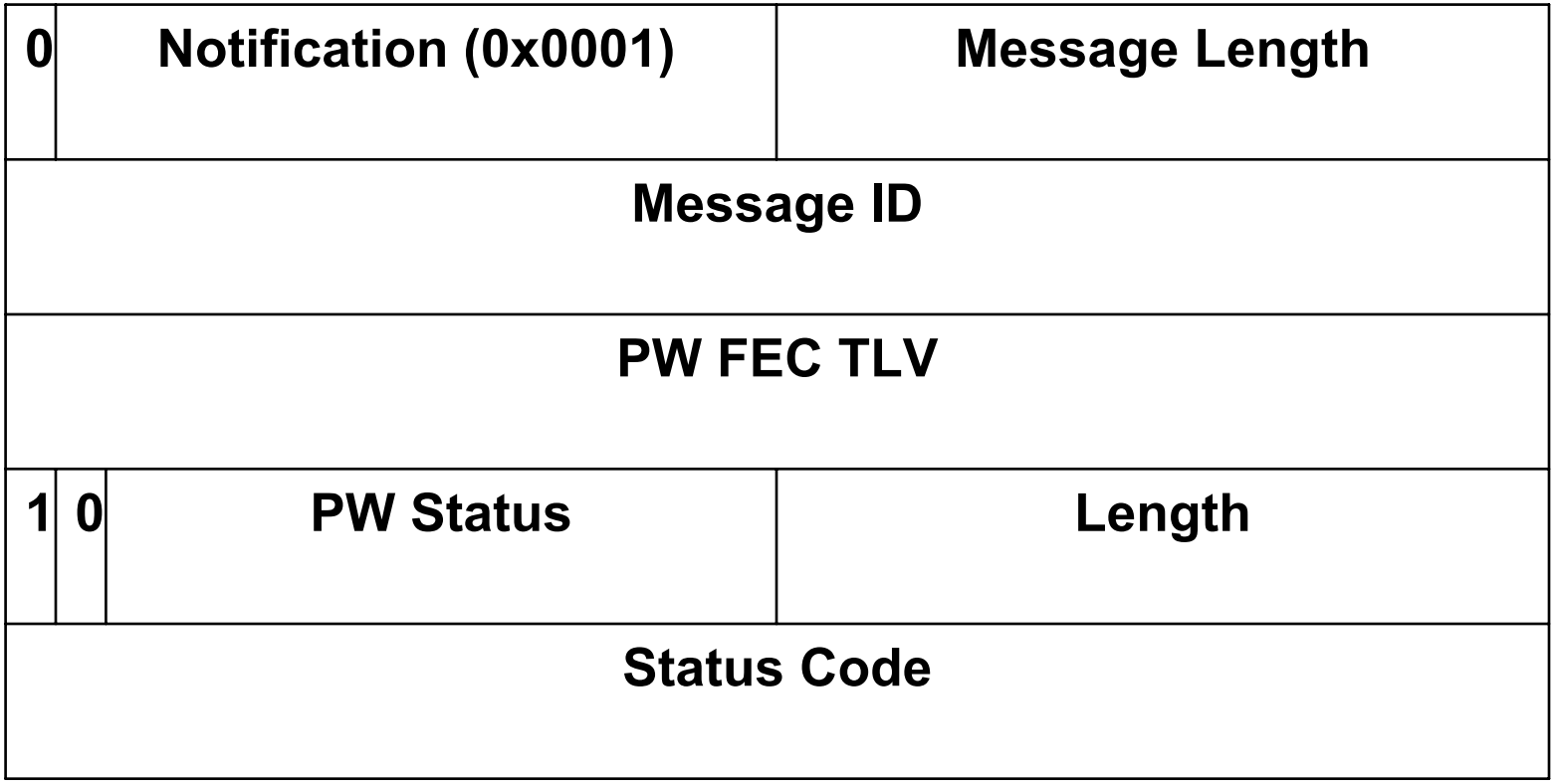
Interface Parameters TLV

- **Generic TLV format with the following possible IDs:**
  - 0x01: Interface MTU
  - 0x02: Max Number of concatenated ATM cells
  - 0x03: Interface Description
  - 0x04: CEP Payload Bytes
  - 0x05: CEP options
  - 0x06: Requested VLAN ID
  - 0x07: CEP/TDM bit Rate
  - 0x08: Frame Relay DLCI length
  - 0x09: Fragmentation Indicator

PW Status Checking

- **Uses LDP Notification Messages**
- **Optional, negotiated in the tunnel setup**
  - If TLV is present on initial PWID FEC Message, use it; else, use label mapping / withdrawal messages;
- **Includes PWID FEC TLV without the interface parameters**
- **Wildcard Status Notification uses only Group ID**

PW Status Notification Message



PW Status Codes

- **32 Bit Mapped Field:**
  - 0x00: PW Forwarding (clear all)
  - 0x01: PW not Forwarding
  - 0x02: Customer TX Fault
  - 0x04: Customer RX Fault
  - 0x08: Tunnel TX Fault
  - 0x10: Tunnel RX Fault

Agenda

- Overview
- IETF Status
- Review: Martini draft
- **LDP and BGP signaling approaches**
- Operation: Control Plane
- Operation: Data Plane
- Scaling VPLS: HVPLS
- Comparison: VPLS and RFC2547
- References

LDP and BGP drafts

- **Turned into a religious debate**
  - There will be BGP proponents
  - There will be LDP proponents
- **Two different problems at stake:**
  - Auto-discovery
  - Signaling
- **There are pros and cons with each approach**
  - Trade-off between operational comfort and efficiency

Two different problems: Discovery and Signaling

- **PE Discovery:**
  - Provisioning Application
  - BGP
  - Radius
- **Signaling:**
  - Targeted LDP
  - BGP

BGP Signalling

- **“Operational Comfort”**
  - Same signaling mechanism used in BGP VPNs
    - 1 Signaling Protocol
- **Distribution of Label Information**
  - Broadcast Mode
    - For VPLS, only a subset of BGP participants require relevant VPN information (unlike route distribution where all participants are interested for best path selection)

LDP Signaling

- **Designed specifically to set up point-to-point connections**
  - Used in Martini pseudowire services
  - The VPLS LPD draft only defines a simple extension to Martini's FEC
- **Efficient signalling of per pseudowire information that needs to be negotiated after the label exchange:**
  - Traffic parameters
  - OAM

Autodiscovery

- **There's a draft for BGP autodiscovery:**
  - draft-ietf-l3vpn-bgpvpn-auto-00.txt
  - Same mechanism as BGP VPNs
  - Can be as easily integrated with VPLS-LDP approach as with VPLS-BGP approach
- **There's another draft for RADIUS discovery:**
  - draft-heinanen-radius-pe-discovery-04.txt
  - Supports site authentication
- **Clearly the BGP approach is the preferred one**

Operating a VPLS service requires much more than autodiscovering PE members and running one signaling protocol

Operating a VPLS Service

- **OSS (Operations & Support Systems)**
  - #1 barrier to deployment
  - Need to *provision* and *manage* VPNs
    - Site specific information
    - VPN specific information
    - Fault and Performance Management

      End-to-end service management

      Fault to customer correlation

      VPN performance reports

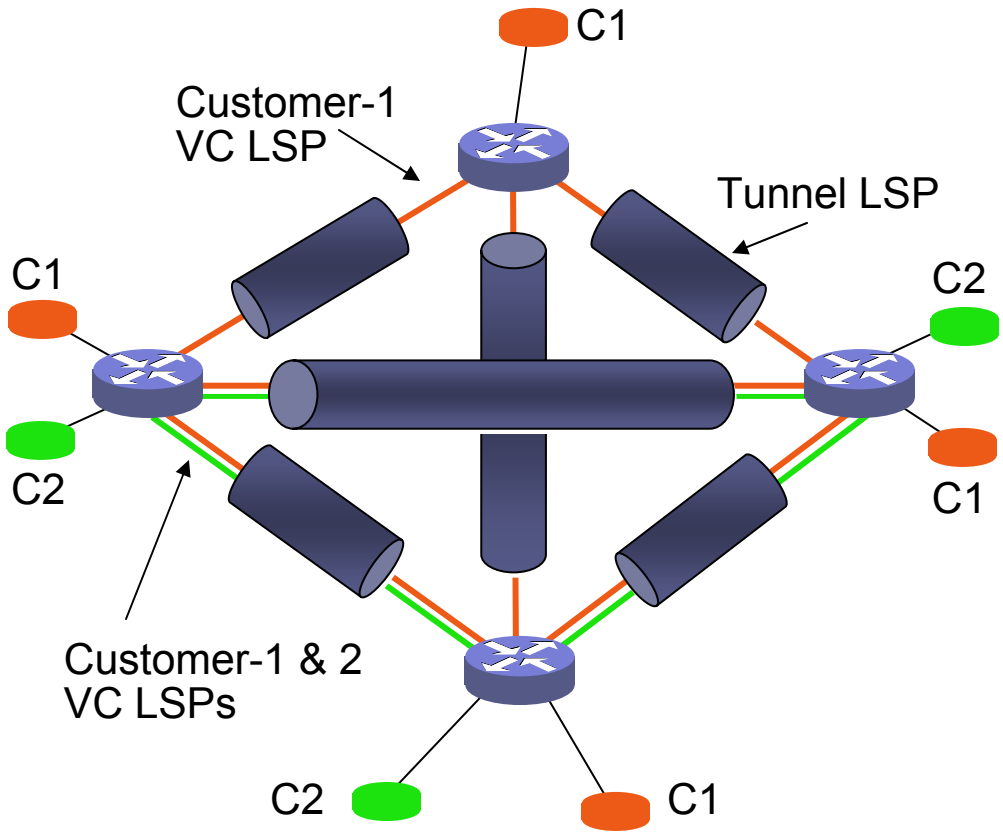
Agenda

- Overview
- IETF Status
- Review: Martini draft
- LDP and BGP signaling approaches
- **Operation: Control Plane**
- Operation: Data Plane
- Scaling VPLS: HVPLS
- Comparison: VPLS and RFC2547
- References

VPLS Control Plane

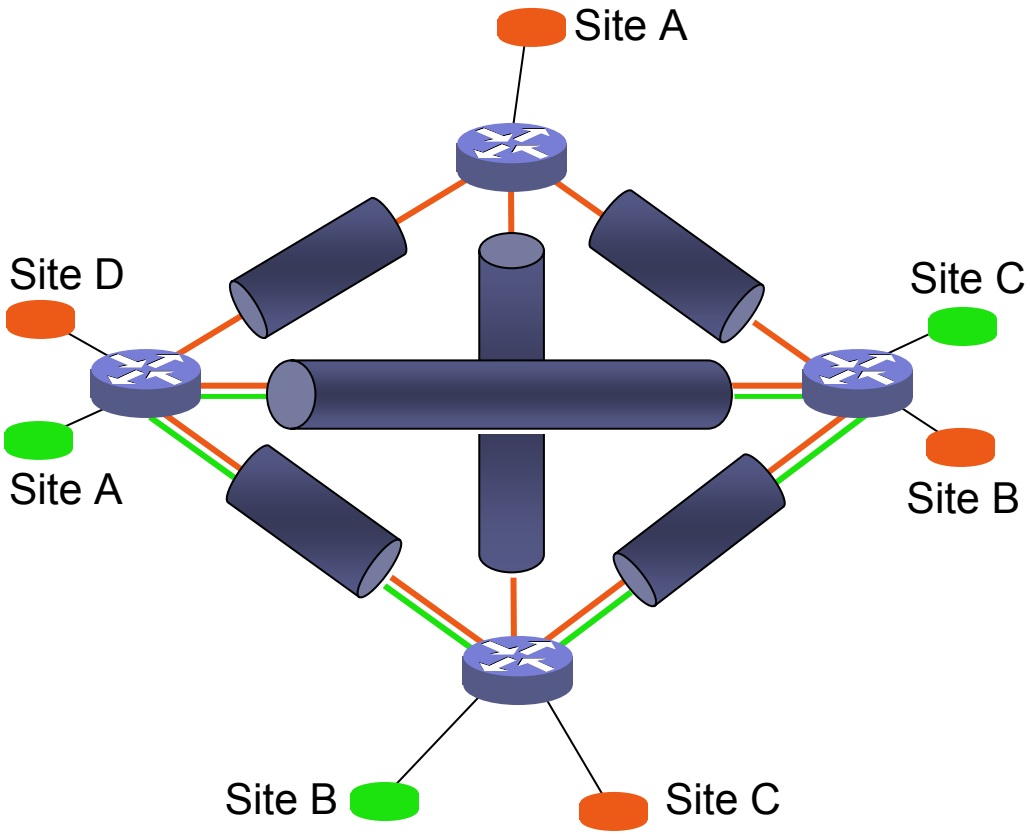
- **Differs depending on the implemented draft**
  - BGP: like BGP VPNs
  - LDP: like Martini tunnels
- **Both assume tunnel LSPs between PEs**
- **This presentation focuses on LDP signalling as it's the most implemented draft today**

Implementation Details
LSP Topology



- **Tunnel LSPs are established between PEs**
  - Full Mesh simplifies loop resolution, as Ethernet is a broadcast capable technology
- **VC LSPs are set up over Tunnel LSPs**
  - VC-ID is now VPN-ID
- **Each PE creates a rooted tree to every other PE**
- **All PEs implement a split-horizon scheme**

Loop Resolution

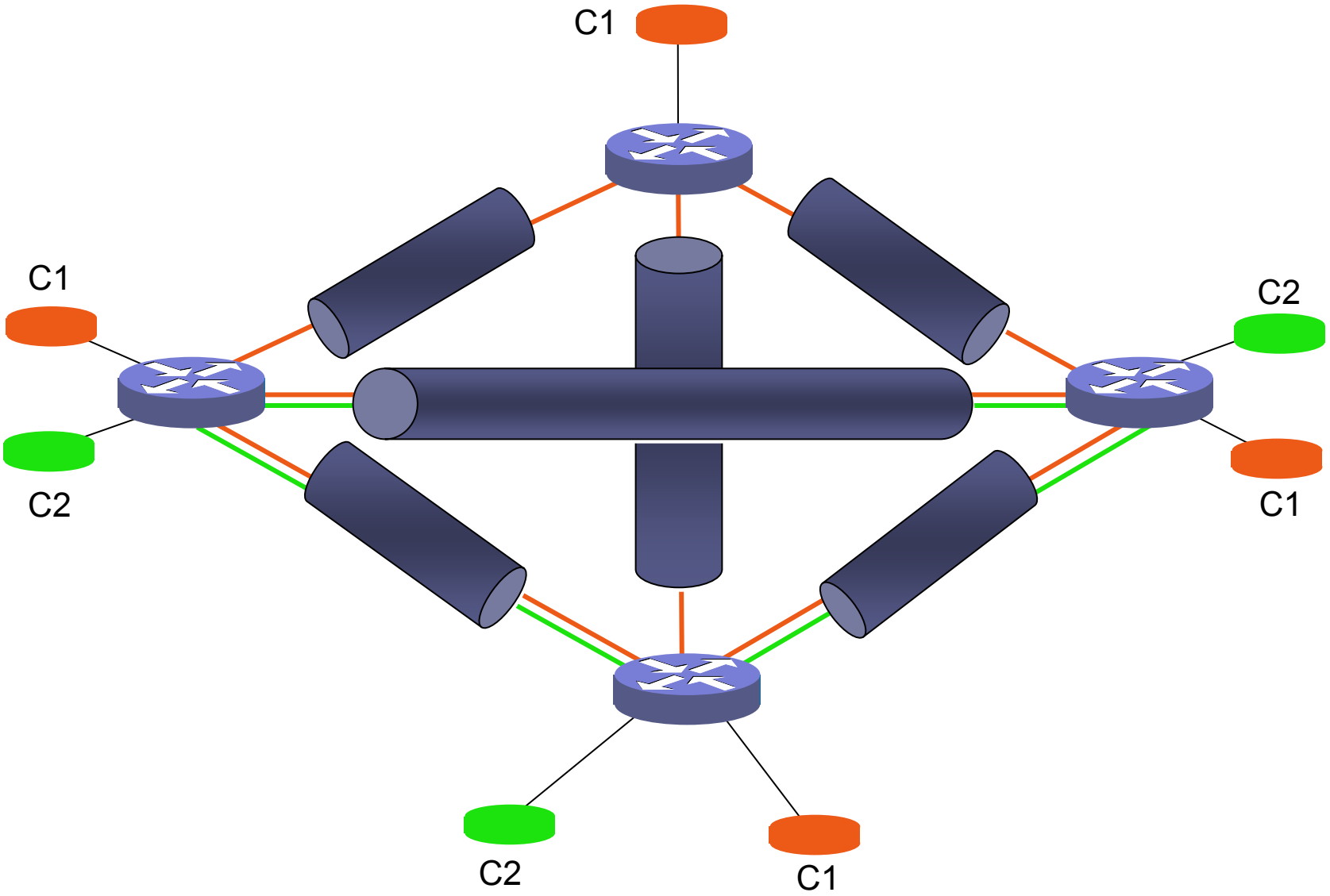


- A full mesh topology with bridges requires a loop resolution mechanism
- In VPLS, the rule of thumb is: “Don’t flood a packet received on a VC to the other VCs”
- Flooding is only done from customer facing ports to the VCs (split-horizon)
- No Spanning Tree needed!

VPLS Signaling

- **Full Mesh of tunnel LSPs between VPLS PEs**
  - Best Effort via LDP
  - Traffic Engineered via RSVP-TE
- **Per-Service VC labels are negotiated using the same mechanism used in Martini tunnels**
  - Targeted LDP

VPLS Control Plane Setup



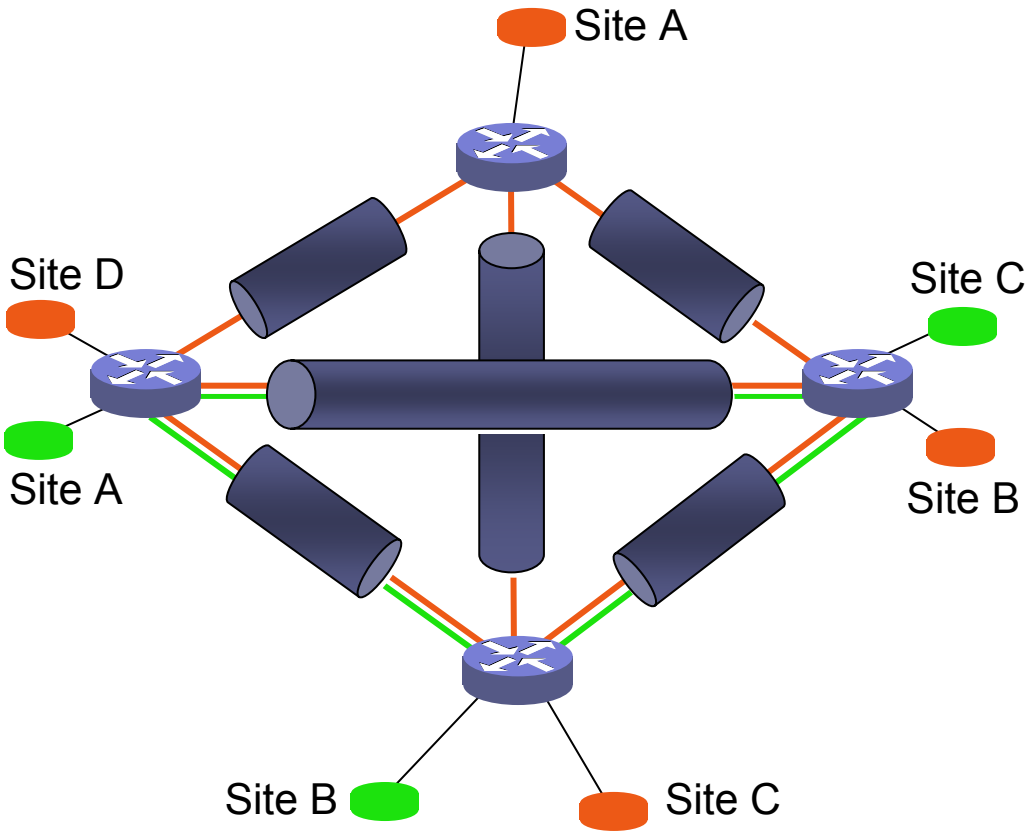
VLL/VPLS Provisioning

- **Tunnel LSPs**
  - Typically traffic engineered via RSVP-TE
  - Typically protected
    - Backup paths
    - Fast Reroute
  - Established between POPs
- **VC LSPs**
  - Signaled via LDP
  - Established between customer sites in the same VPN
  - Nested within tunnel LSPs
    - RSVP routers configured to tunnel LDP messages for end-to-end LDP sessions

Learning and Forwarding

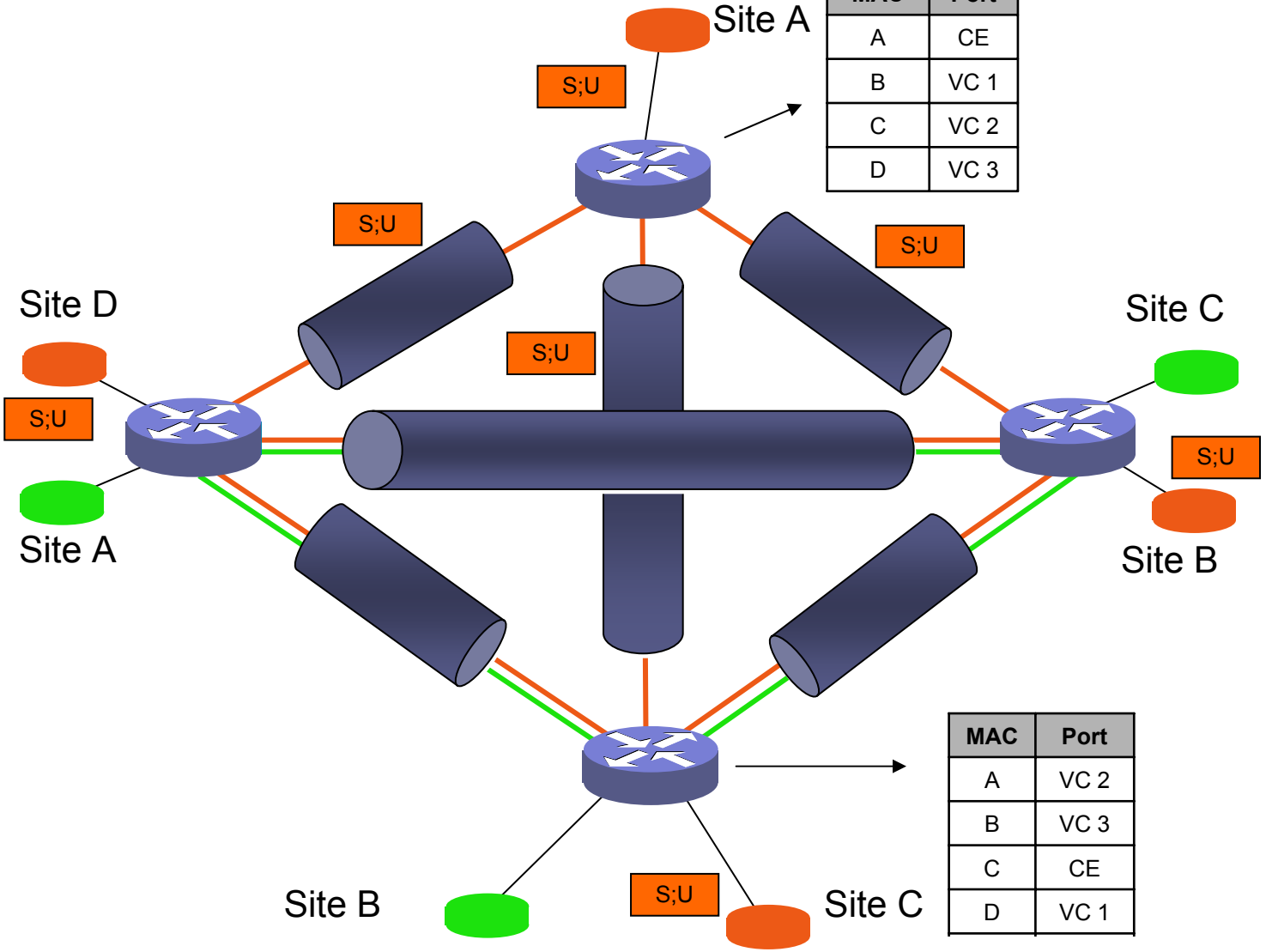
- **VPLS network looks like a L2 switch to the customer**
- **As a L2 switch, the VPLS cloud must:**
  - Learn MAC addresses
  - Flood packets with unknown addresses
  - Flood Multicast packets
  - Flood Broadcast packets
  - Age out MAC addresses
- **The PEs create a VSI per VPLS instance**

Address Learning



- **Dynamic MAC address learning on PEs**
- **Each PE must learn**
  - On customer facing ports
  - On VC LSPs
- **Each PE must age out MAC addresses**
- **Packets are forwarded based on the MAC table**

Example


Site A
S;U
MAC
Port
A
CE
B
VC 1
C
VC 2
D
VC 3
S;U
S;U
Site D
S;U
Site C
S;U
S;U
Site A
Site B
Site B
S;U
Site C
MAC
Port
A
VC 2
B
VC 3
C
CE
D
VC 1

Fast Convergence

- An optional MAC Withdrawal Message to communicate MAC withdrawals between PEs is defined
- Uses LDP Address Withdrawal Messages with a FEC TLV and a new MAC TLV
- This scheme can be used to improve the convergence time in the case of a failure
- Useful mainly for multi-homed MTU in hierarchical topologies or multi-homed CE

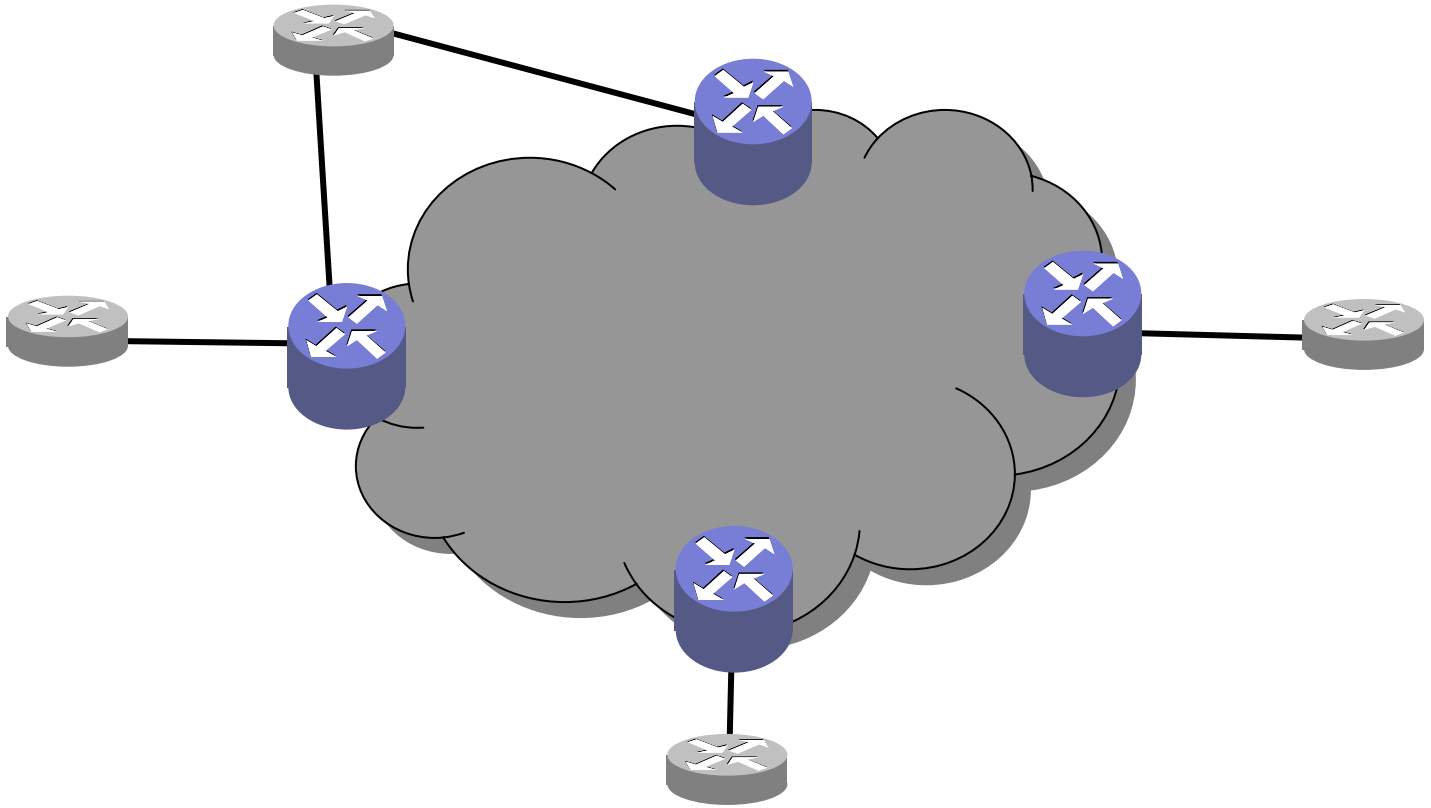
MAC TLV

U	F	Type (0x0404)	Length
MAC Address #1			
MAC Address #2			
MAC Address #n			

- If the message has a list of MAC addresses, they must be relearned on the received pseudo-wire
- If the message has an empty list, all MAC addresses must be flushed from the VPLS table except the ones already learned through the pseudo-wire

Multi-Homed CE Topology

- Spanning Tree transparently tunneled across the VPLS domain
- PE could look for Topology Change messages to flush the MAC table using the MAC Withdrawal TLV



Agenda

- Overview
- IETF Status
- Review: Martini draft
- LDP and BGP signaling approaches
- Operation: Control Plane
- **Operation: Data Plane**
- Scaling VPLS: HVPLS
- Comparison: VPLS and RFC2547
- References

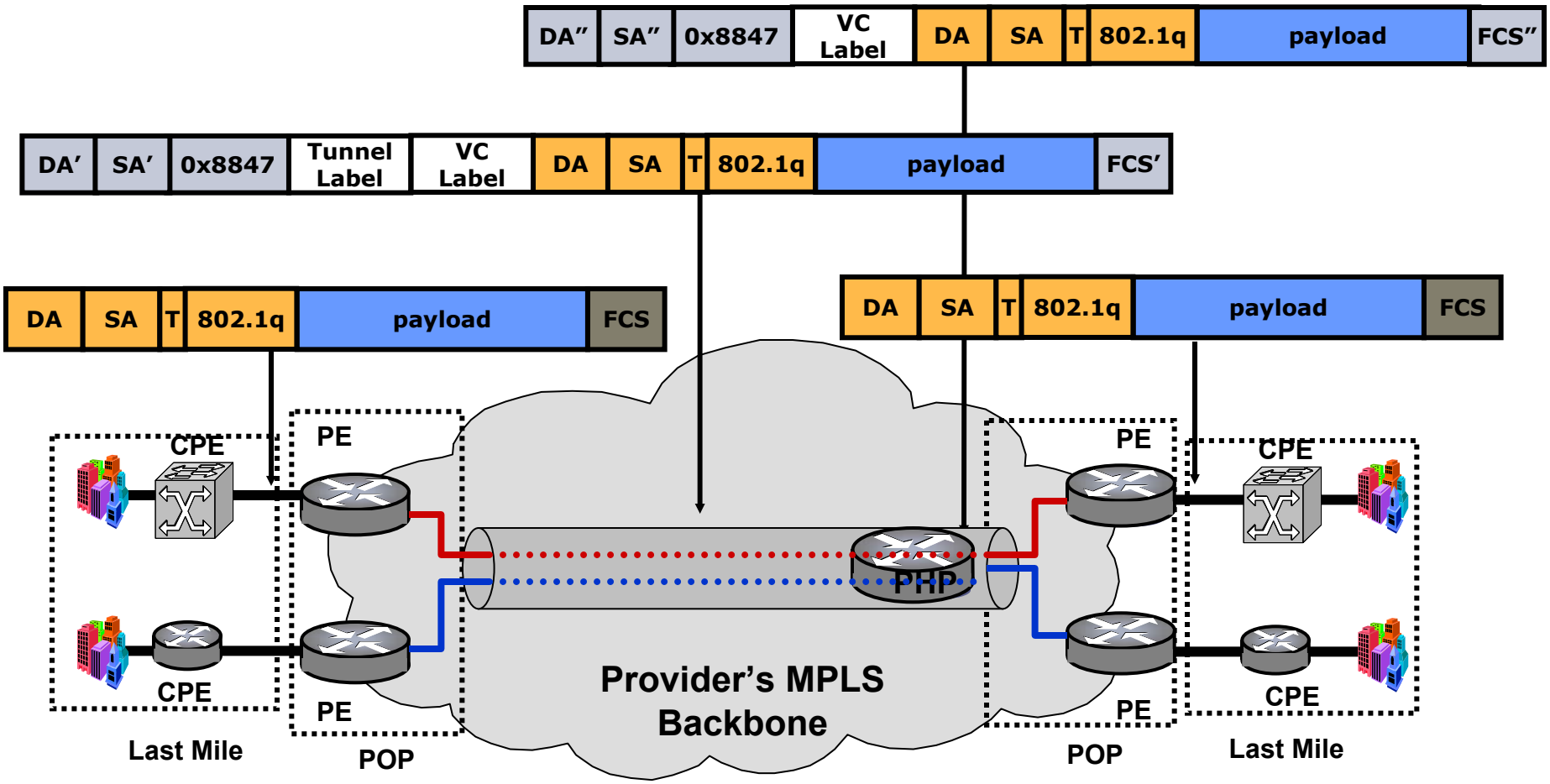
VPLS Data Plane

- Uses the same encapsulation method defined by Martini (draft-ietf-pwe3-ethernet-encap-02.txt)
- Preamble and FCS are stripped from original Ethernet frame, which is then encapsulated into a MPLS frame
- Transparently transports the Ethernet frame through the MPLS Network

Service Delimiting VLANs

- **An important concept is the "Service Delimiting VLAN"**
- **If the VLAN was defined by the provider to identify the customer or the service, it is a Service Delimiting VLAN;**
  - The VLAN tag should be stripped from the frame
- **If the VLAN is used to define multiple L2 domains inside the customer network, it is not a Service Delimiting VLAN;**
  - The VLAN tag should be kept in the frame

Life of a Frame



VPLS PE Tasks

- **At ingress:**
  - Map port or port/VLAN-id to Service-id/FIB
  - Look up dest. MAC in FIB -> dest. PE
  - Apply VC-label to customer packet
  - Apply tunnel label & send packet
- **At egress:**
  - Tunnel label popped to reveal VC-label
  - Look up VC-label -> Service-id/FIB
  - Map dest. MAC in FIB -> Egress port
  - Send original Ethernet frame

Agenda

- Overview
- IETF Status
- Review: Martini draft
- LDP and BGP signaling approaches
- Operation: Control Plane
- Operation: Data Plane
- **Scaling VPLS: HVPLS**
- Comparison: VPLS and RFC2547
- References

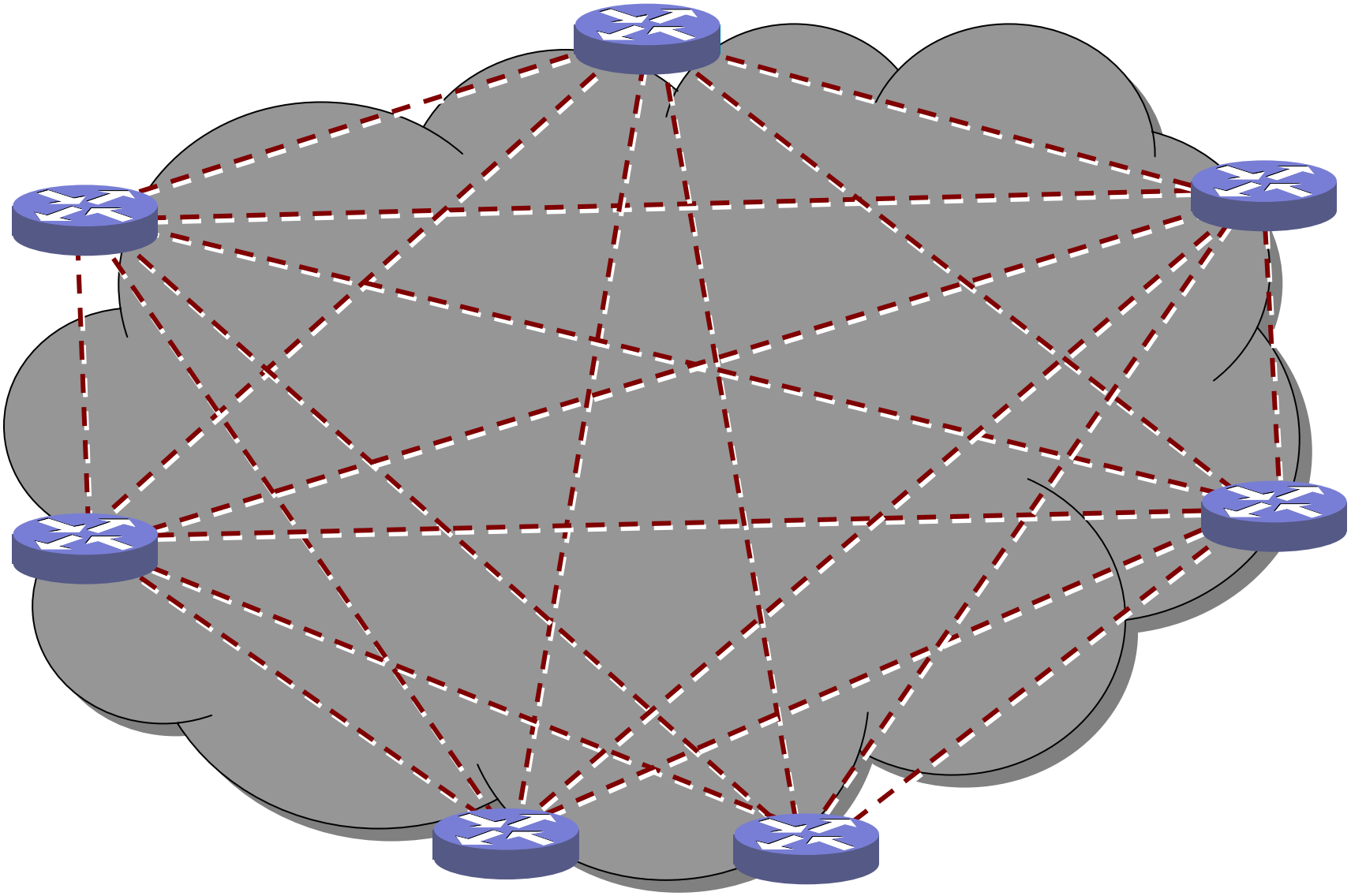
VPLS Scaling Aspects

- **Signalling**
  - Number of peers
  - Number of LSPs
- **Number of packet replications**
- **MAC Address Learning**
- **Provisioning**

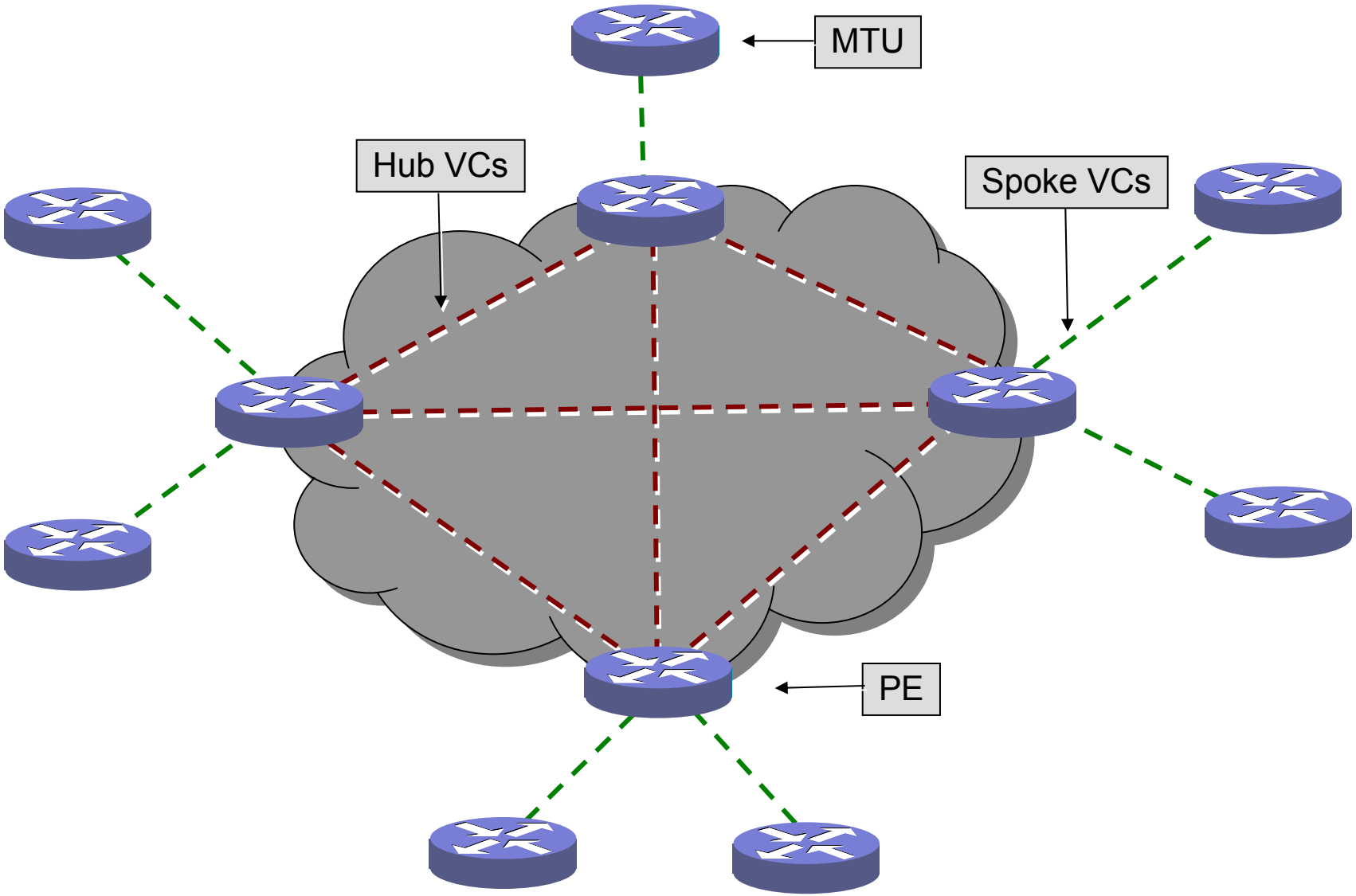
Hierarchical VPLS

- **In order to better scale a VPLS network, hierarchy is introduced: HVPLS**
- **Hierarchy achieved through a hub and spoke topology between MTUs and PEs, reducing the number of full mesh tunnels**
- **Enhanced scaling in the following areas:**
  - Signaling
  - Packet Replication
  - Provisioning

Plain VPLS Topology

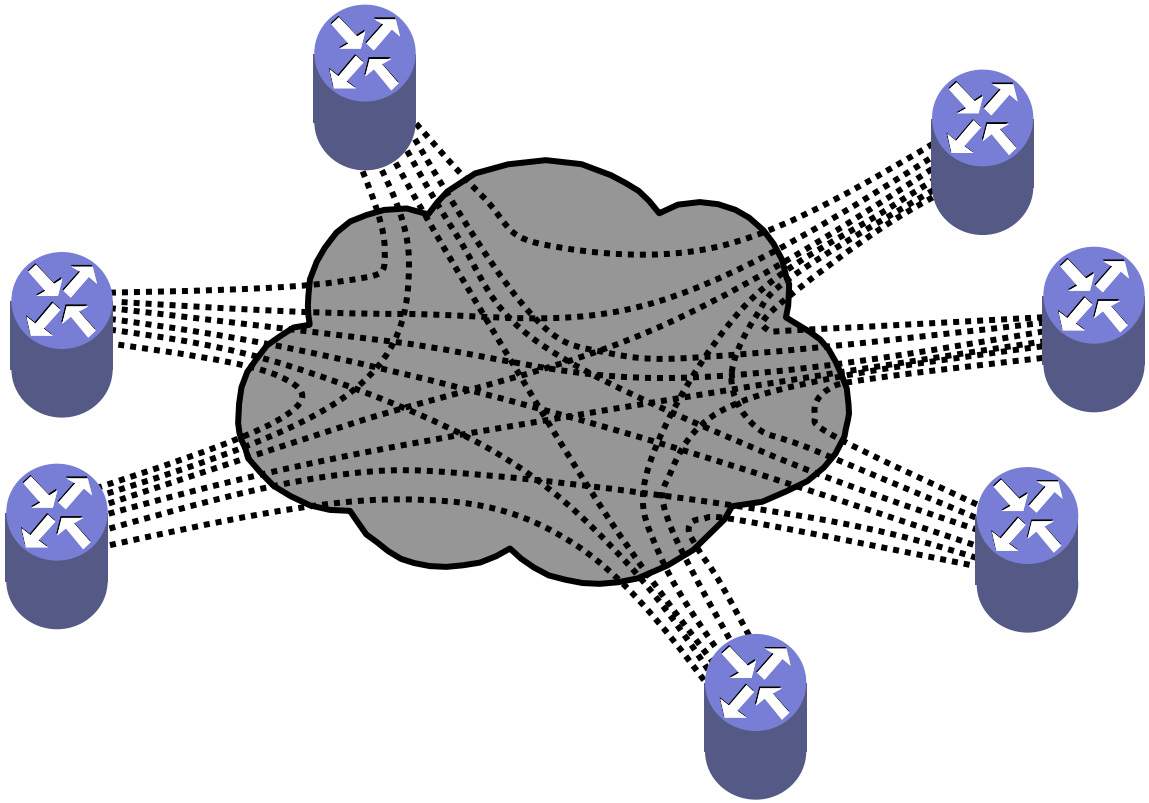


H-VPLS Topology



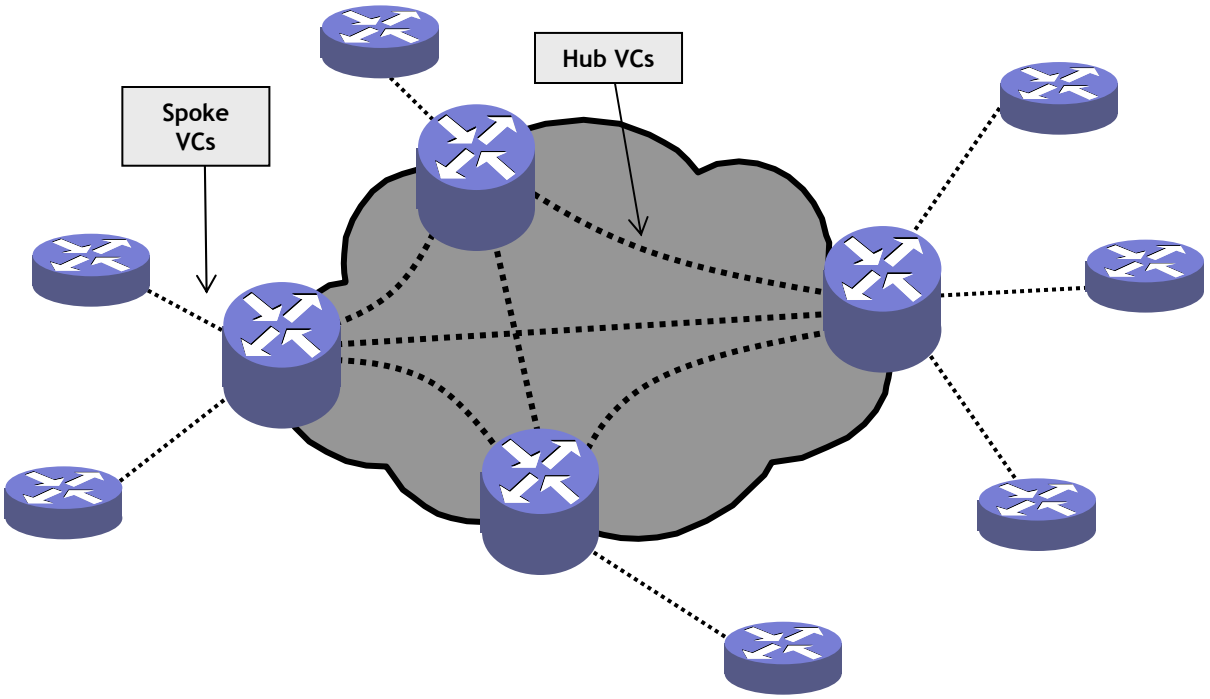
Scaling VPLS: Signaling

- **Flat Topology (Basic VPLS architecture)**
  - N^2 T-LDP sessions
  - N^2 Tunnels (RSVP-TE or LDP)
  - N^2 VC LSPs



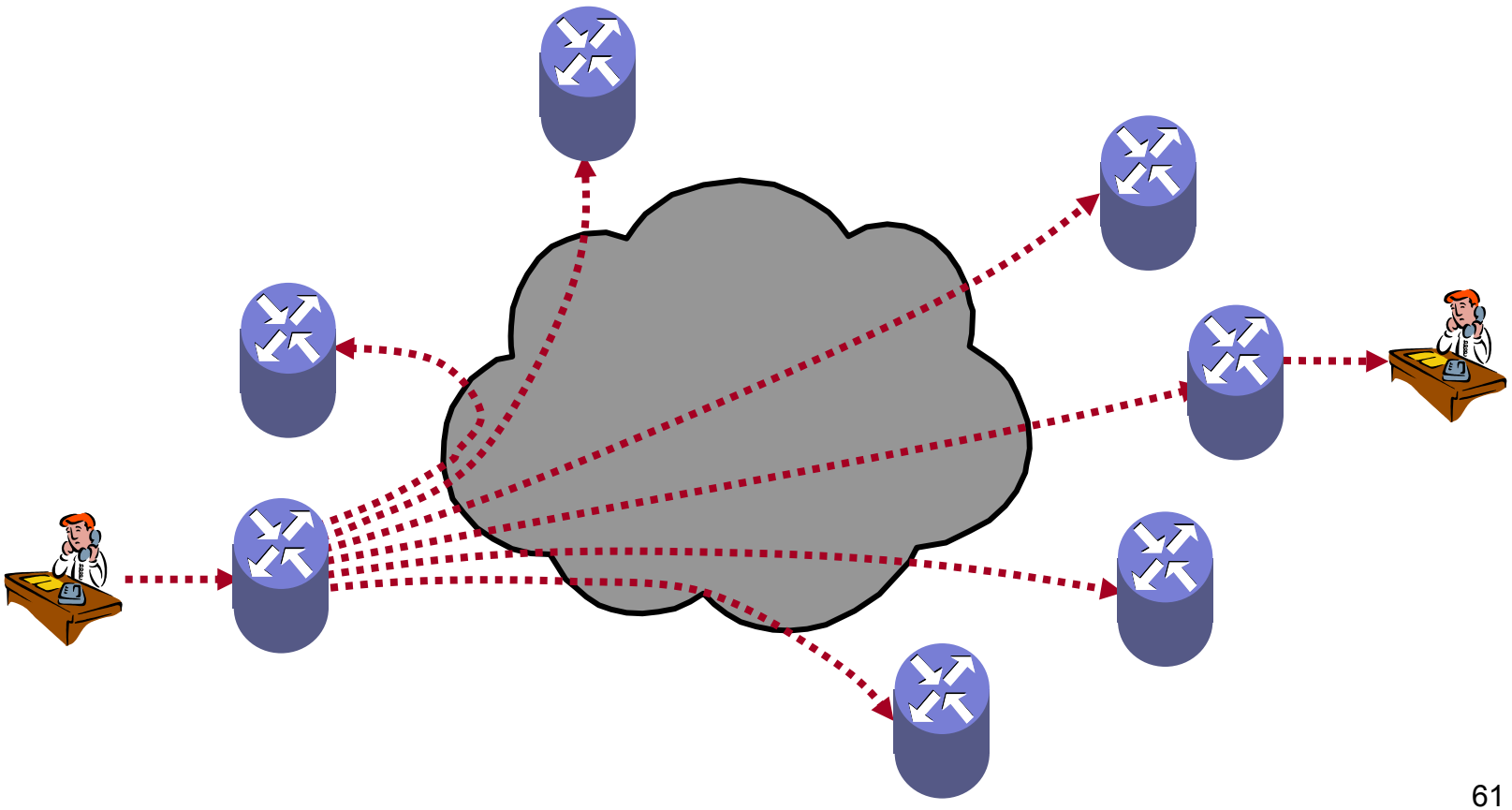
Scaling VPLS: Signaling

- **Tree Topology: Hierarchical VPLS (HVPLS)**
  - O(N) T-LDP sessions
  - O(N) Tunnels (RSVP-TE or LDP)
  - O(N) VC LSPs



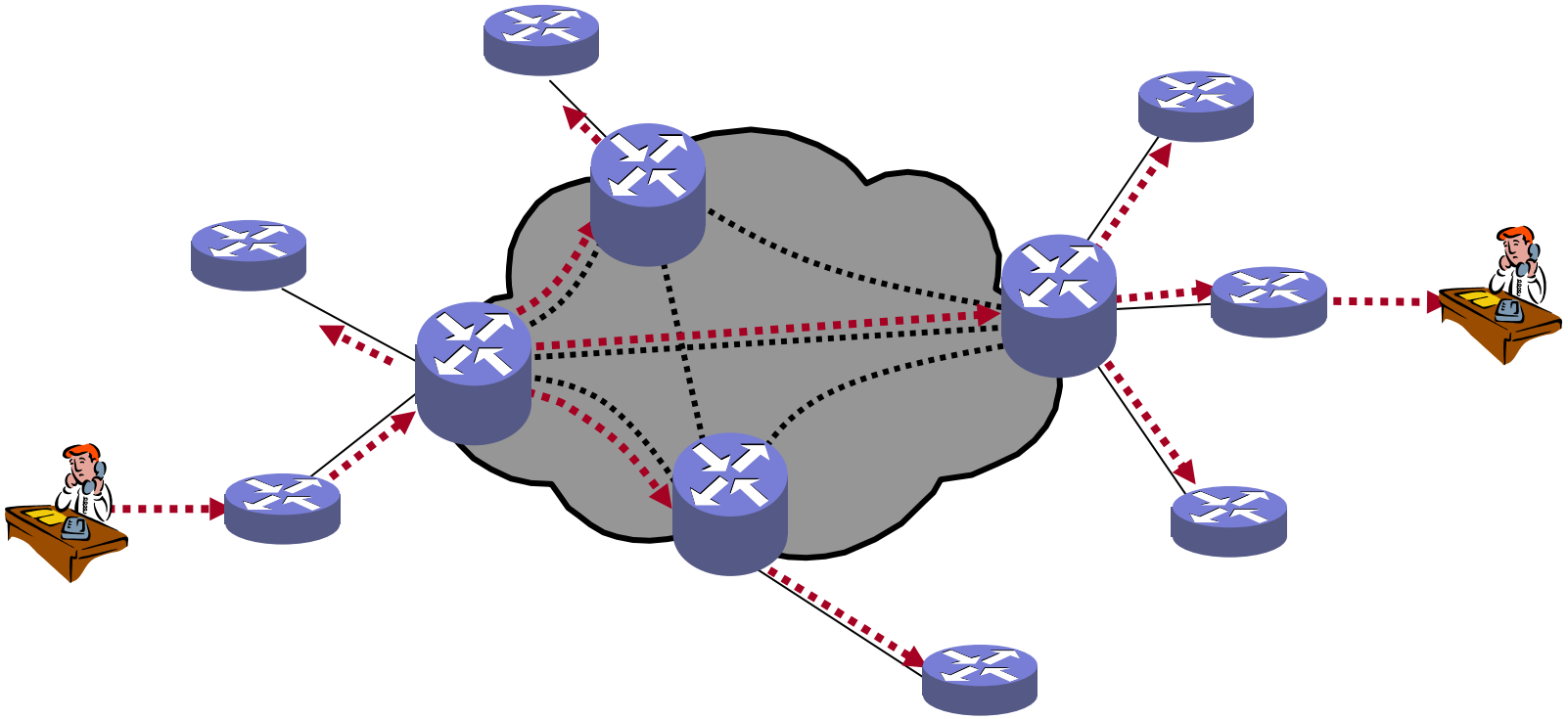
Scaling VPLS: Packet Replication

- **Flat Topology (Basic VPLS architecture)**
  - Replication at the very edge of the network
    - Close to the source



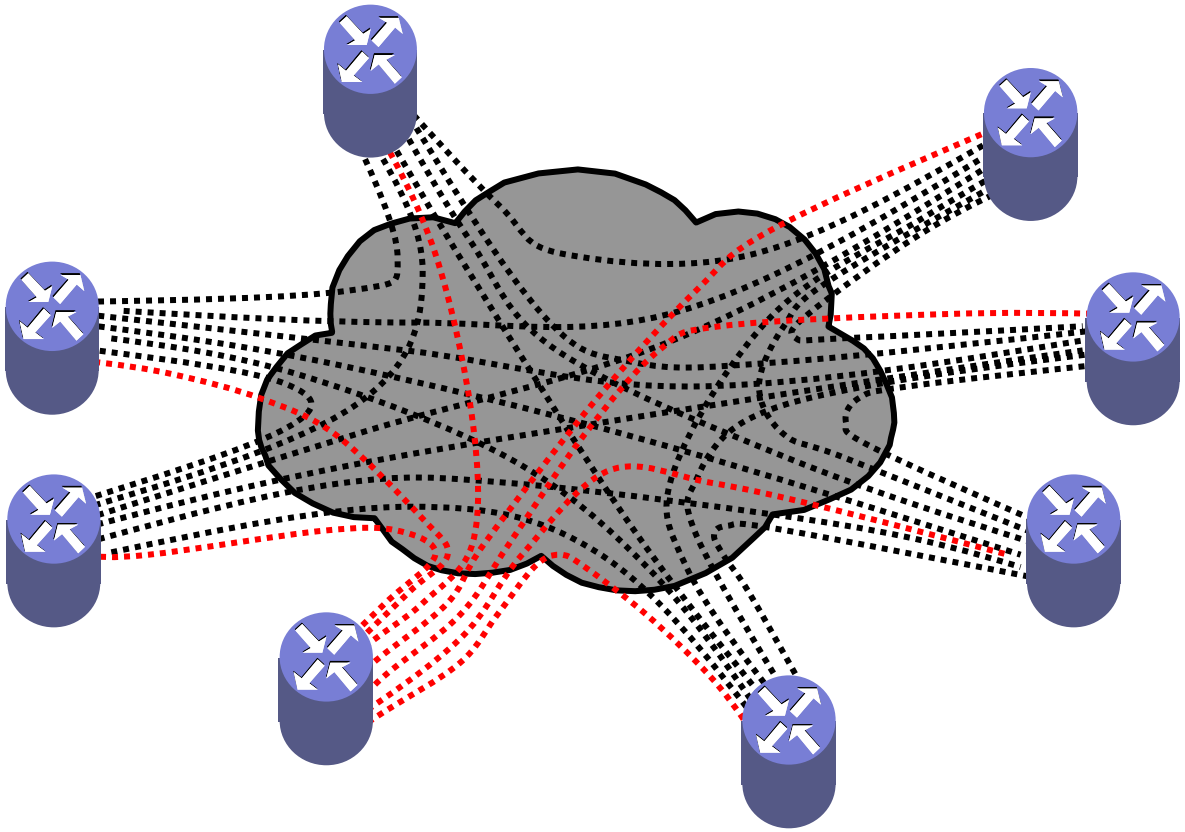
Scaling VPLS: Packet Replication

- **Tree Topology: Hierarchical VPLS (HVPLS)**
  - Distributed replication across spoke and hub PEs
    - Limited to directly adjacent connections
  - Replication as close to destination as possible



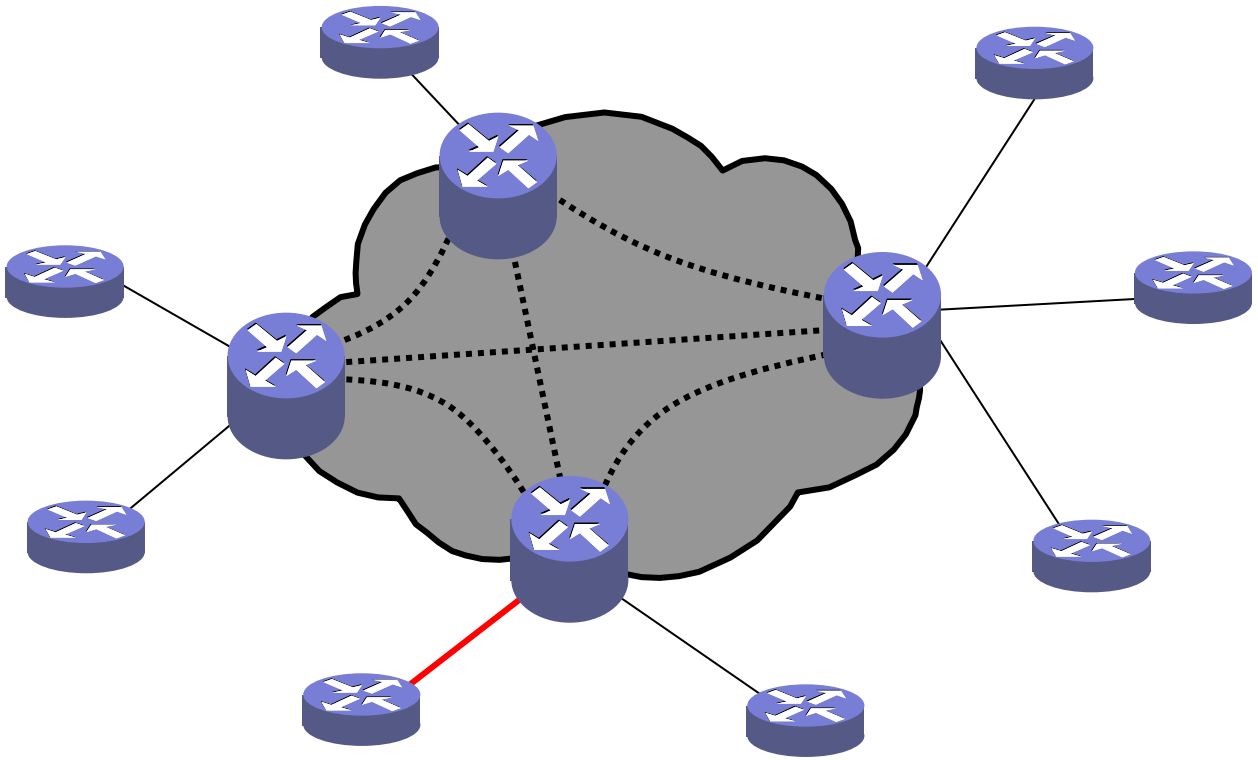
Scaling VPLS: Provisioning

- **O(N) effort to add a new site**
  - Configuration of all PEs participating in VPLS Instance



Scaling VPLS: Provisioning

- **O(1) effort to add a new site**
  - Configure new spoke on corresponding PE



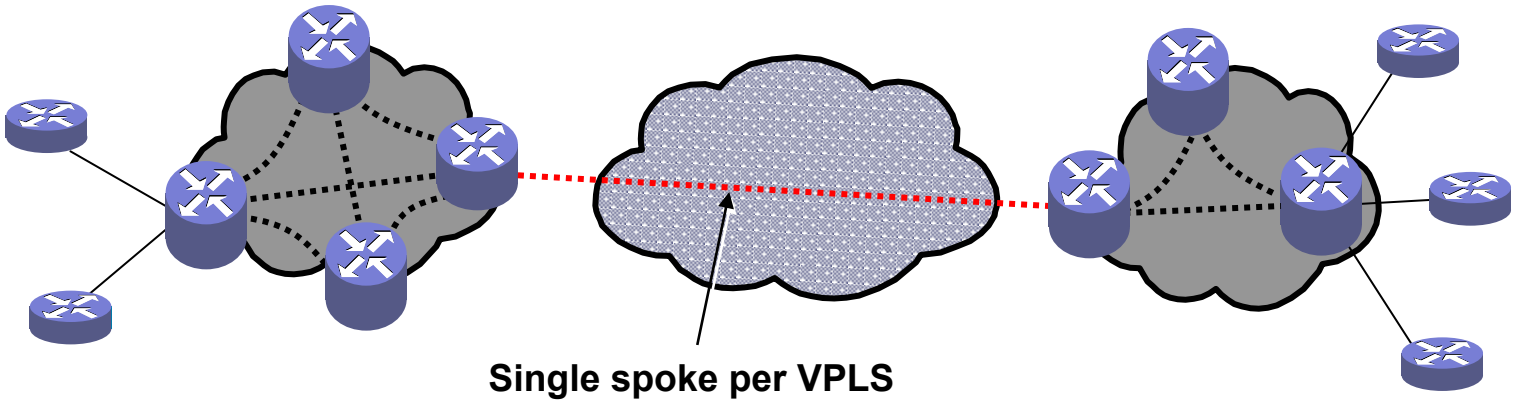
Scaling VPLS: MAC Addresses

- **VPLS FIB Size depends on the type of Service Offering:**
  - Switch interconnect
    - Multiple MAC addresses per site

      MAC limiting per access circuit
  - Router Interconnect
    - One MAC address per site
- **Same Network Design principles apply for**
  - MAC FIB Size of VPLS Service
  - Route Table Size of RFC2547 Service

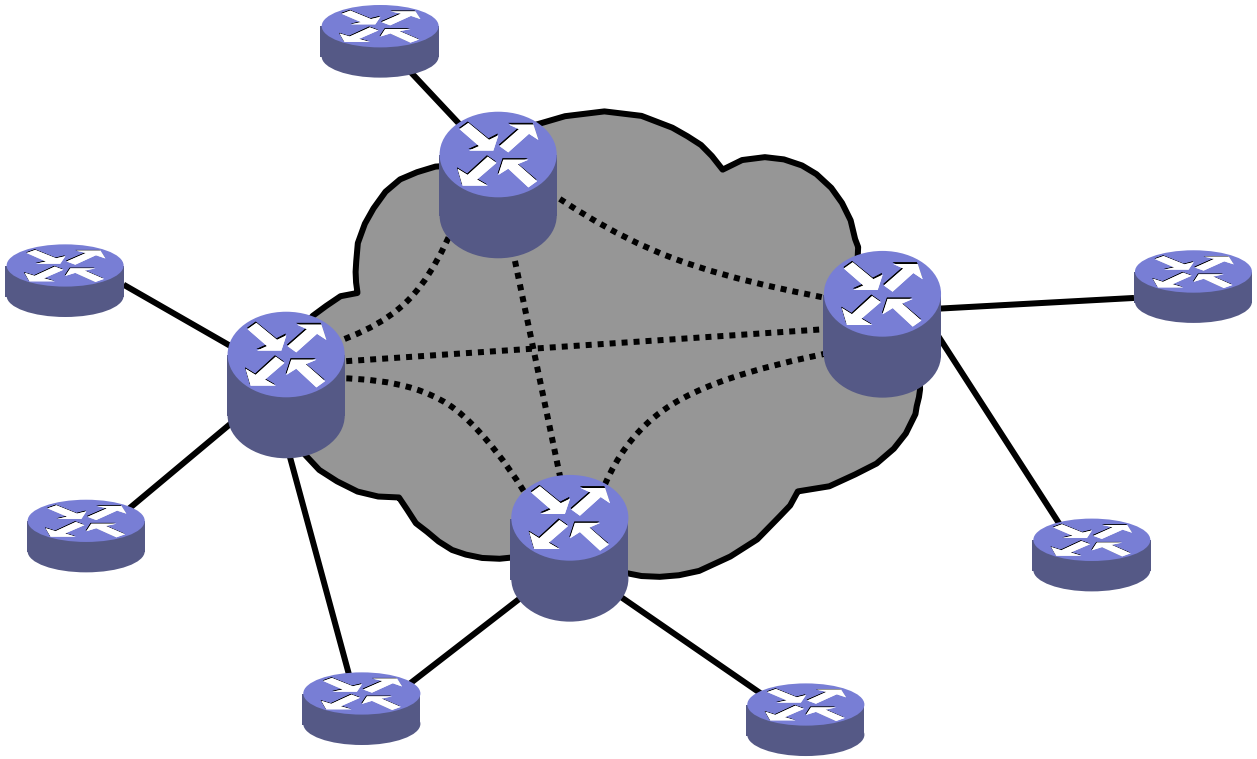
Inter Domain HVPLS

- Single spoke LSP between 2 domains
- Specific VPLS Gateway functions to interconnect multiple domains to be defined in the future



Multi-Homed MTU with Martini

- Two Martini tunnels used for redundancy
- No Spanning Tree needed: one active, one stand-by
- MAC Withdrawal Messages speed up convergence



VPLS OAM

- **Work in progress**
  - draft-stokes-vkompella-ppvpn-hvpls-oam-02.txt
- **Uses data plane initially, and then the control plane to verify errors**
- **Another draft to be created on VPLS MIBs**

VPLS OAM Facilities

- **VPLS Ping**
  - Extension to draft-ietf-mpls-lsp-ping-04.txt
  - Similar to IP Ping
- **VPLS Traceroute**
  - Used to trace the data path
  - Similar to IP Traceroute

Agenda

- Overview
- IETF Status
- Review: Martini draft
- LDP and BGP signaling approaches
- Operation: Control Plane
- Operation: Data Plane
- Scaling VPLS: HVPLS
- **Comparison: VPLS and RFC2547**
- References

VPLS and BGP VPNs

- **VPLS:**
  - L2 VPNs
  - Transports Ethernet
  - Needs LDP, may use RSVP-TE for tunnels
  - Creates a VSI per VPN
  - Forwarding based on MAC tables
  - CE can be a router or a switch
- **RFC2547:**
  - L3 VPNs
  - Transports IP
  - Needs BGP, plus LDP or RSVP-TE
  - Creates a VRF per VPN
  - Forwarding based on IP route tables
  - CE must be a router

Agenda

- Overview
- IETF Status
- Review: Martini draft
- LDP and BGP signaling approaches
- Operation: Control Plane
- Operation: Data Plane
- Scaling VPLS: HVPLS
- Comparison: VPLS and RFC2547
- **References**

References

- www.rfc-editor.org for IETF drafts
- http://www.riverstonenet.com/technology/tls.shtml for a whitepaper on VPLS/TLS
- http://www.riverstonenet.com/technology/mpls_ethernet.shtml for a whitepaper on Metro Ethernet using MPLS

River
STONE
NETWORKS™

Obrigado!