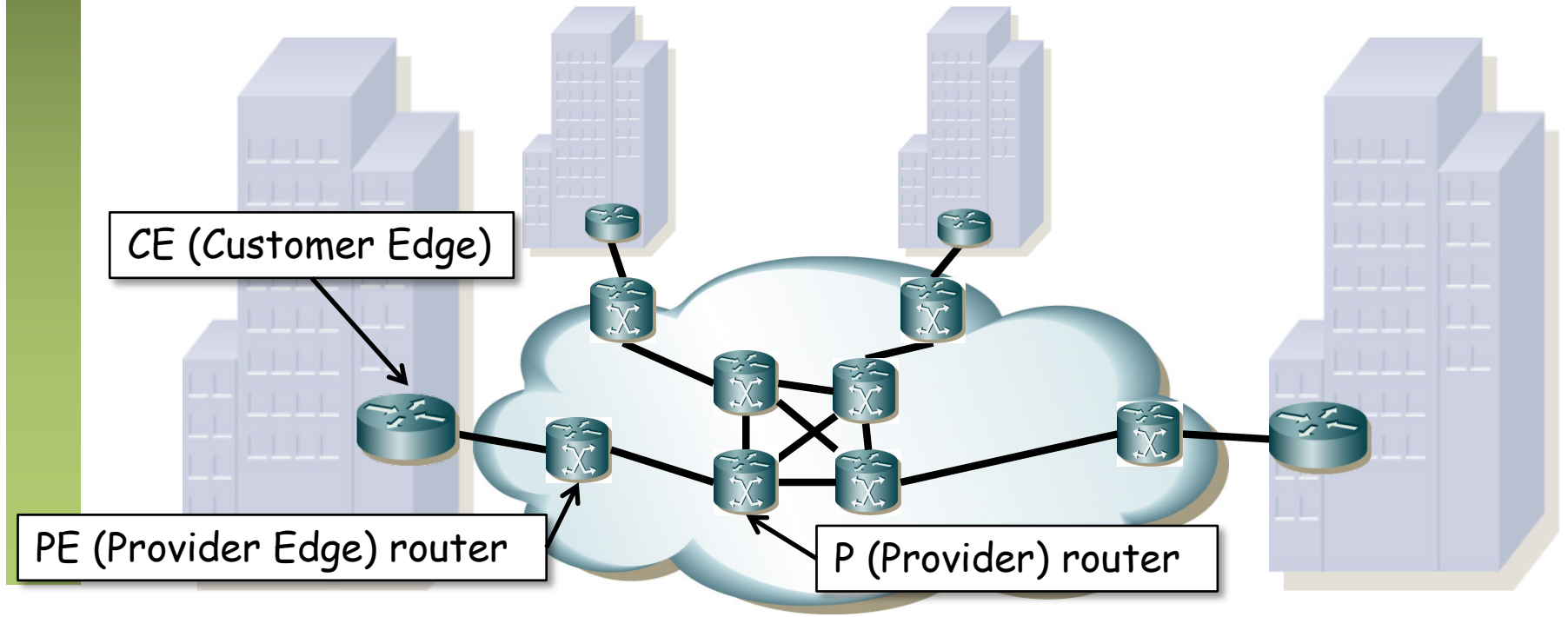
# Layer 3 MPLS VPNs

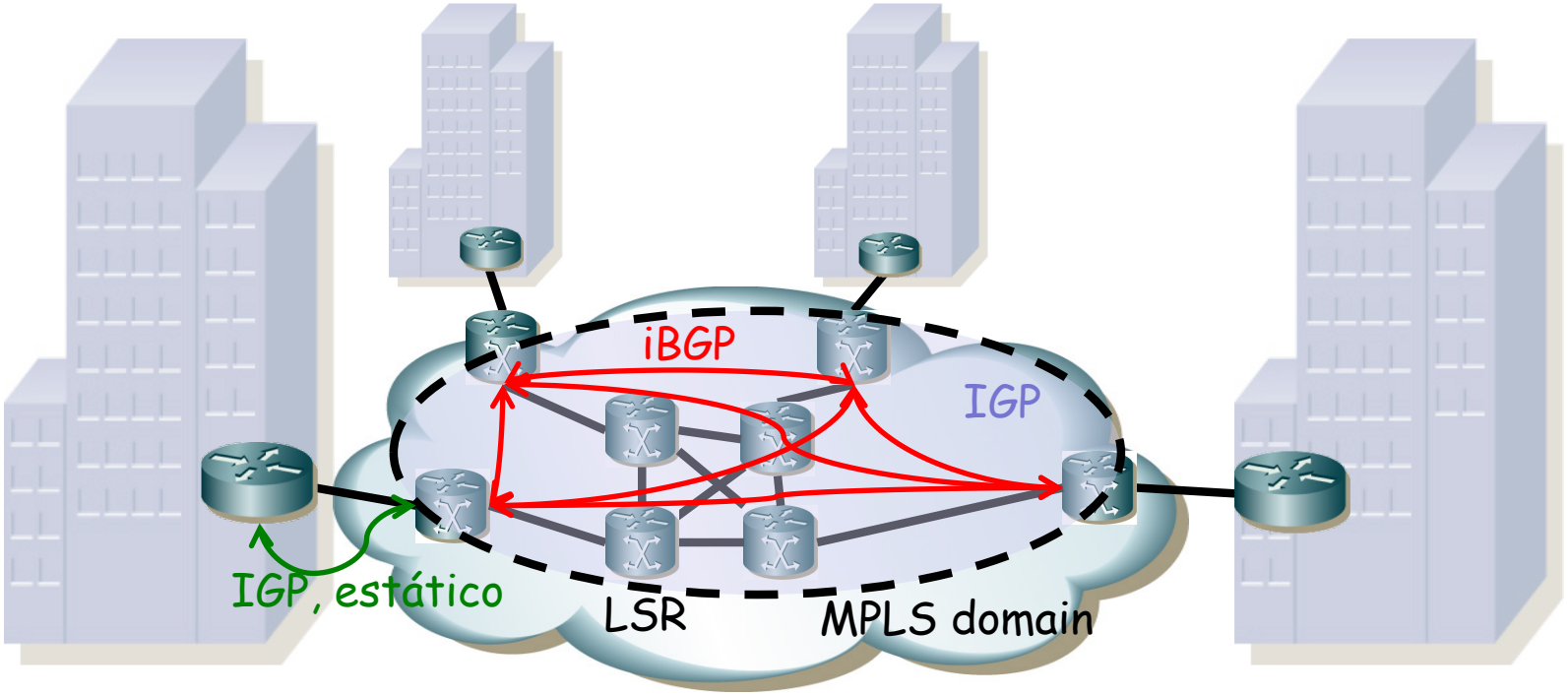
# Layer 3 VPNs

- RFC 4364 “BGP/MPLS IP Virtual Private Networks (VPNs)” (Cisco Systems y Juniper Networks, 2006)
  - *“This document describes a method by which a Service Provider may use an IP backbone to provide IP Virtual Private Networks (VPNs) for its customers.”*
- VPN para el transporte de paquetes IP entre sedes (*sites*)
- El backbone del proveedor de servicio es una red IP MPLS
- RFC 4760 “Multiprotocol Extensions for BGP-4” (Cisco, Sanoa, Juniper, 2007)
- Extensiones a BGP-4 para poder transportar información de otros protocolos de nivel de red: IPv6, IPX, L3VPN, etc
- En este caso, en lugar de transportar rutas IPv4 transportará rutas “VPN-IPv4”



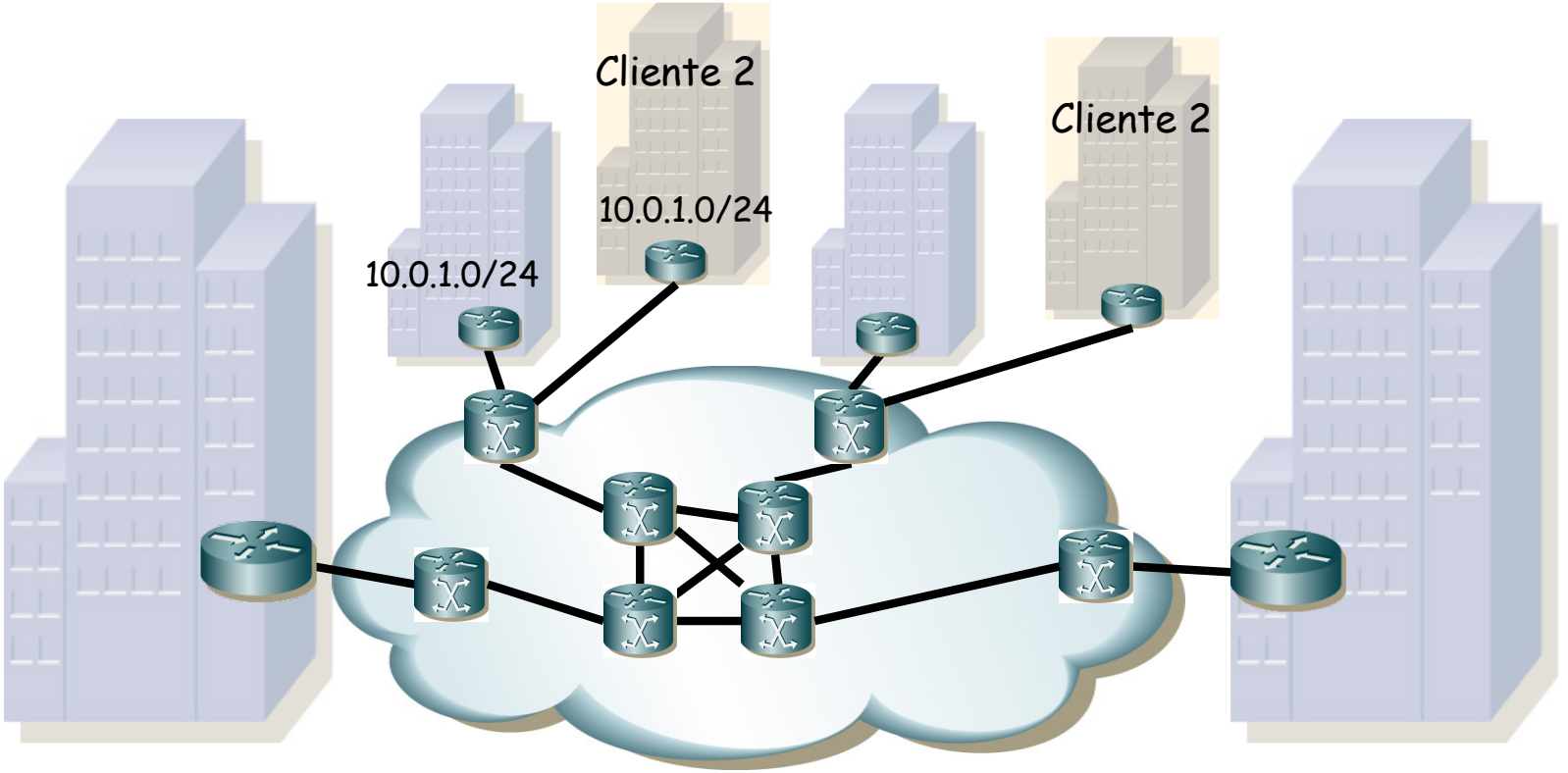
# L3VPN: Routing

- Los CEs anuncian sus rutas a los PEs (con un IGP o rutas estáticas)
- Los PEs emplean MP-BGP para intercambiarse esas rutas (iBGP)
- El PE la distribuye al CE del mismo cliente (de la misma VPN)
- Los P y PE corren un IGP para tener alcanzabilidad interna
- Los CE son routers convencionales, no necesitan ninguna configuración de VPN ni emplean MPLS
- Los CEs no intercambian información de routing entre ellos, no son adyacentes
- La VPN no actúa como un overlay sino una red IP con otro gestor



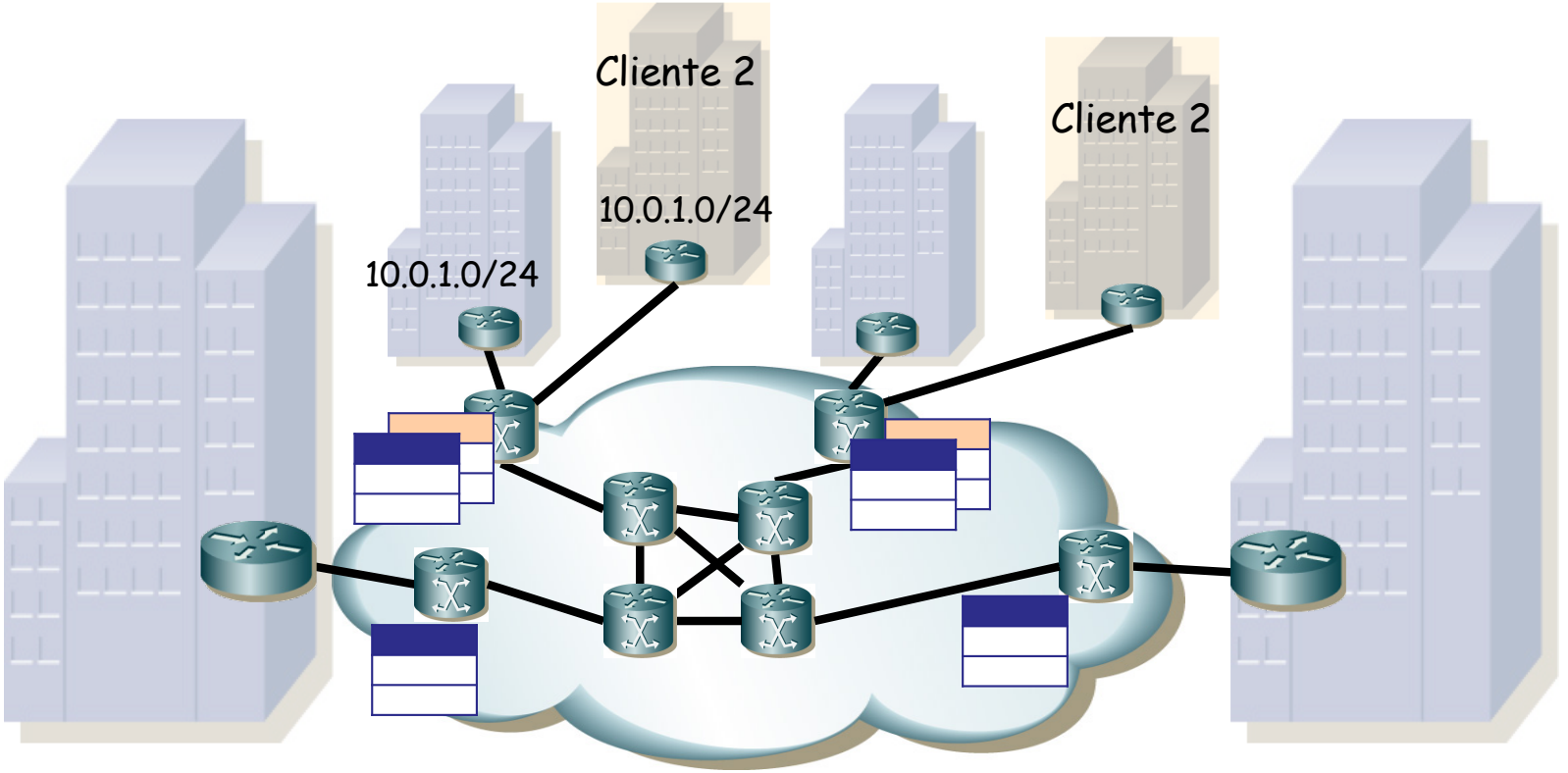
# L3VPN: Routing

- Dos VPNs pueden emplear espacios de direcciones IP que se solapen
- Los anuncios VPN-IPv4 de esas subredes mediante BGP incluyen un identificador (*Route Distinguisher = RD, 8 bytes)* que las diferencia
- Cada service provider tiene su espacio de valores RD
- Los P no ven las rutas de las VPNs (evita problemas de escalabilidad)
- ¿Cómo se enruta si hay direcciones duplicadas y los routers centrales no ven esas rutas?



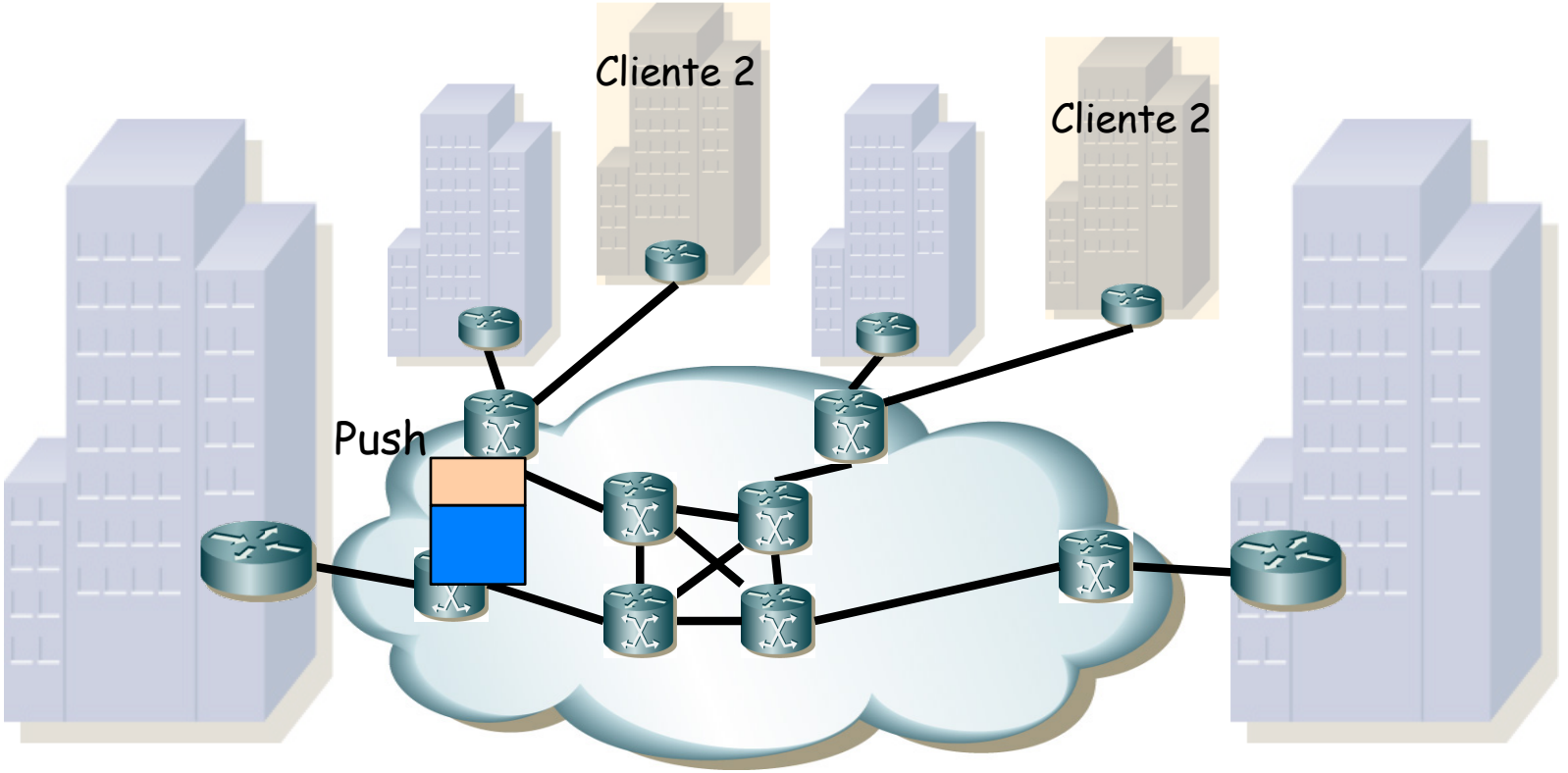
# L3VPN: Forwarding

- Cada PE mantiene una tabla de rutas para cada VPN o *VPN/Virtual Routing and Forwarding tables (VRFs) y a*demás una tabla por defecto
- Cara VRF está asociada a un valor o más de *"Route Target" (RT)*
- Al recibir un paquete IP de un cliente consulta la VRF correspondiente
- Las rutas VPN-IPv4 se anuncian con un (o más) valor de RT
- Incluye una etiqueta MPLS (para el plano de datos)
- Una VRF importa las rutas con unos RT que desee (plano de control)



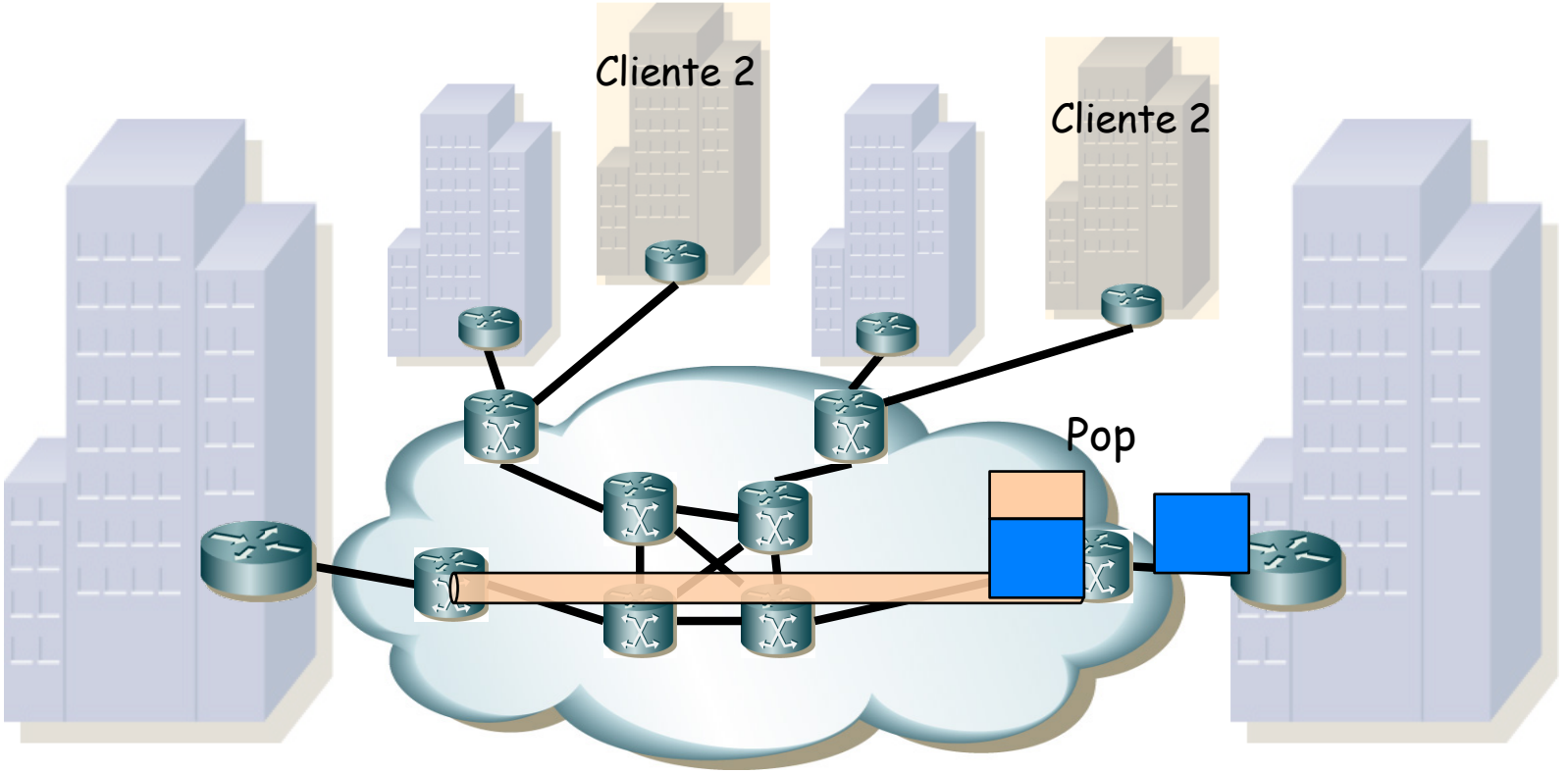
# L3VPN: MPLS

- ¿Para qué esa etiqueta?
- (…)



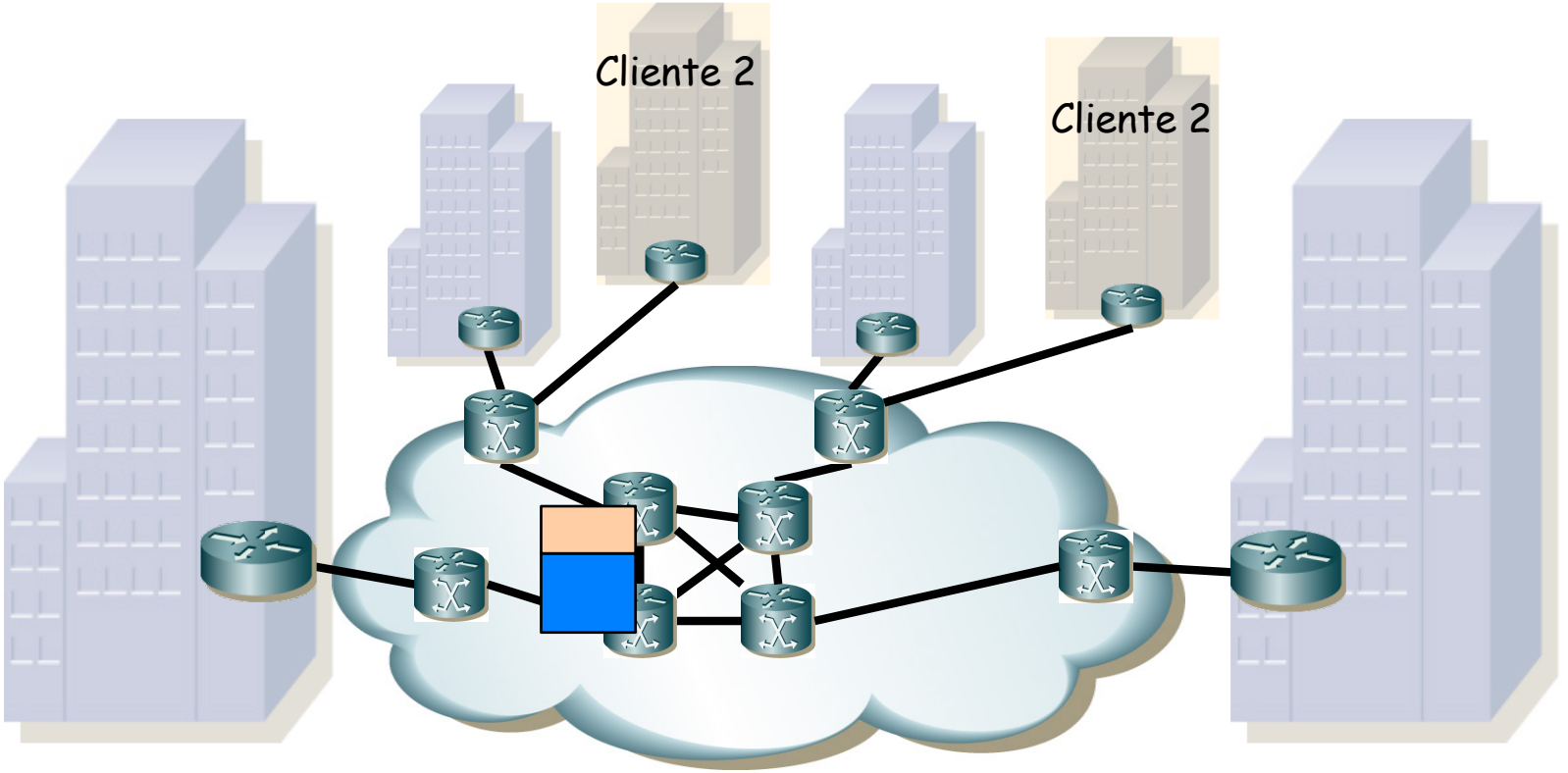
# L3VPN: MPLS

- ¿Para qué esa etiqueta?
- Para que el PE de salida sepa a qué VRF pertenece el paquete
- No puede basarse en la dirección IP destino pues pueden estar duplicadas



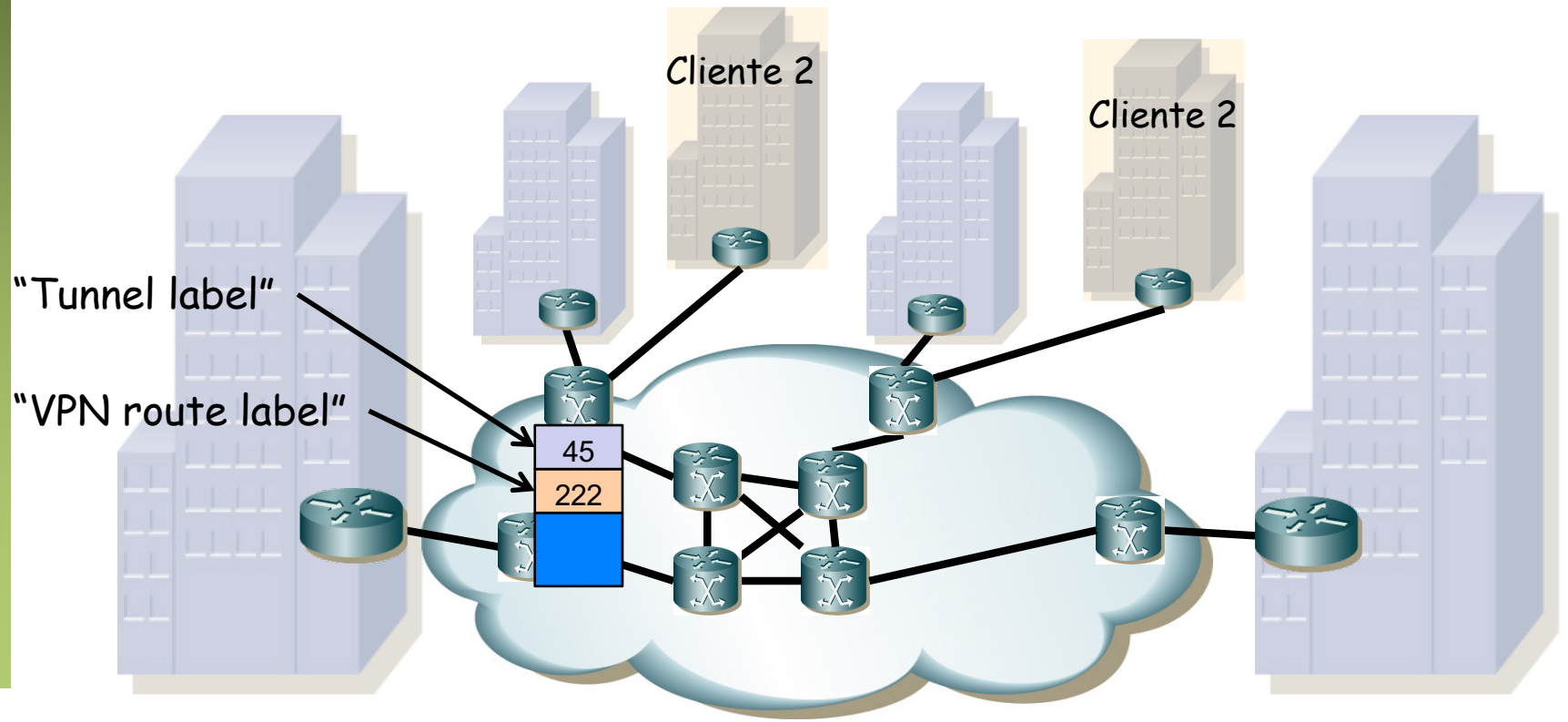
# L3VPN: MPLS

- ¿Y en los P routers? ¿Reenvían en función de esa etiqueta?
- Tendríamos en ellos una gran cantidad de LSPs, para todas las VPNs
- Mala escalabilidad
- (…)



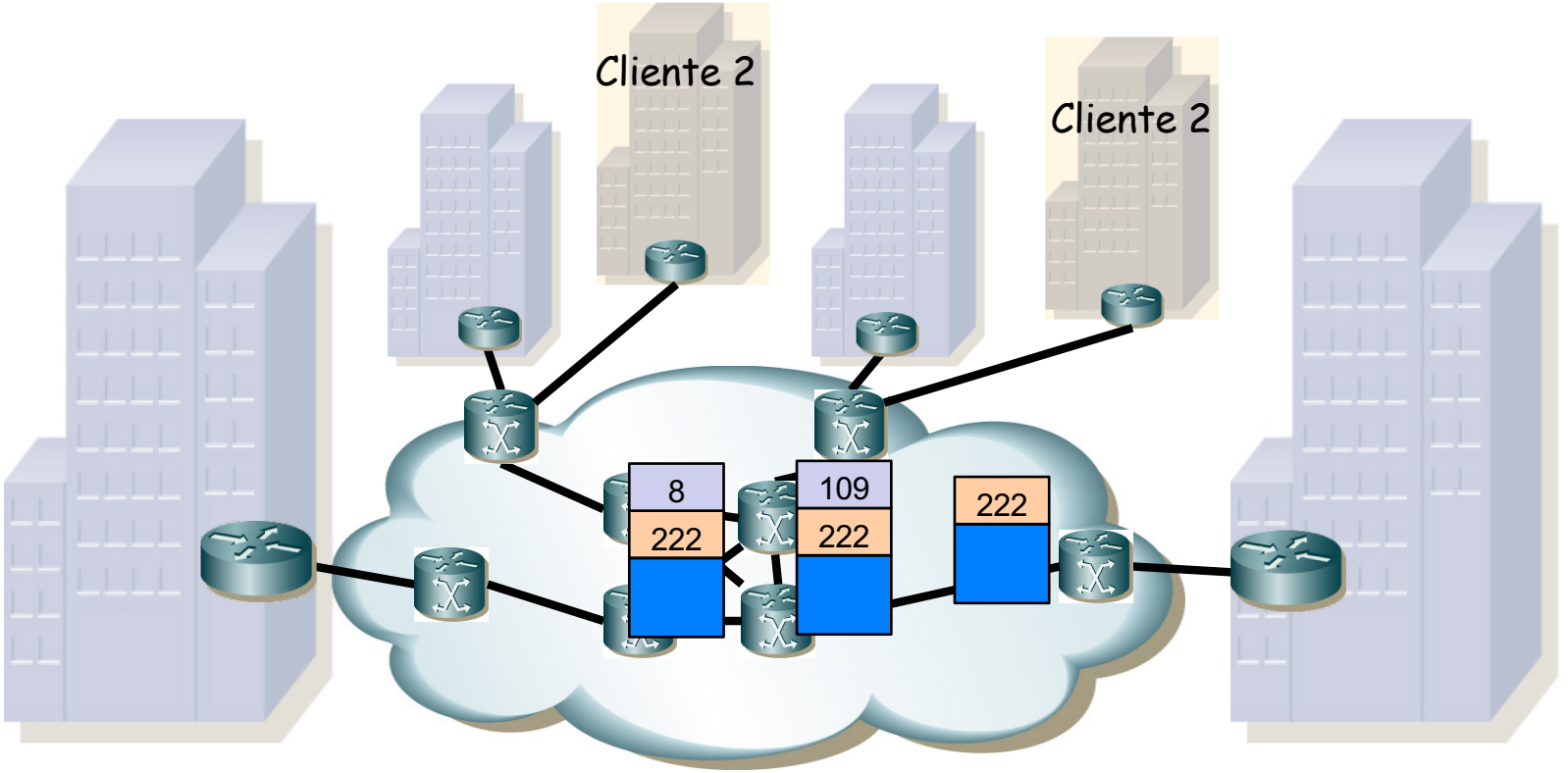
# L3VPN: MPLS

- ¿Y en los P routers? ¿Reenvían en función de esa etiqueta?
- Se crean LSPs entre los PEs ("Tunnel LSPs")
- Los P routers reenvían en función de esa etiqueta externa (...)



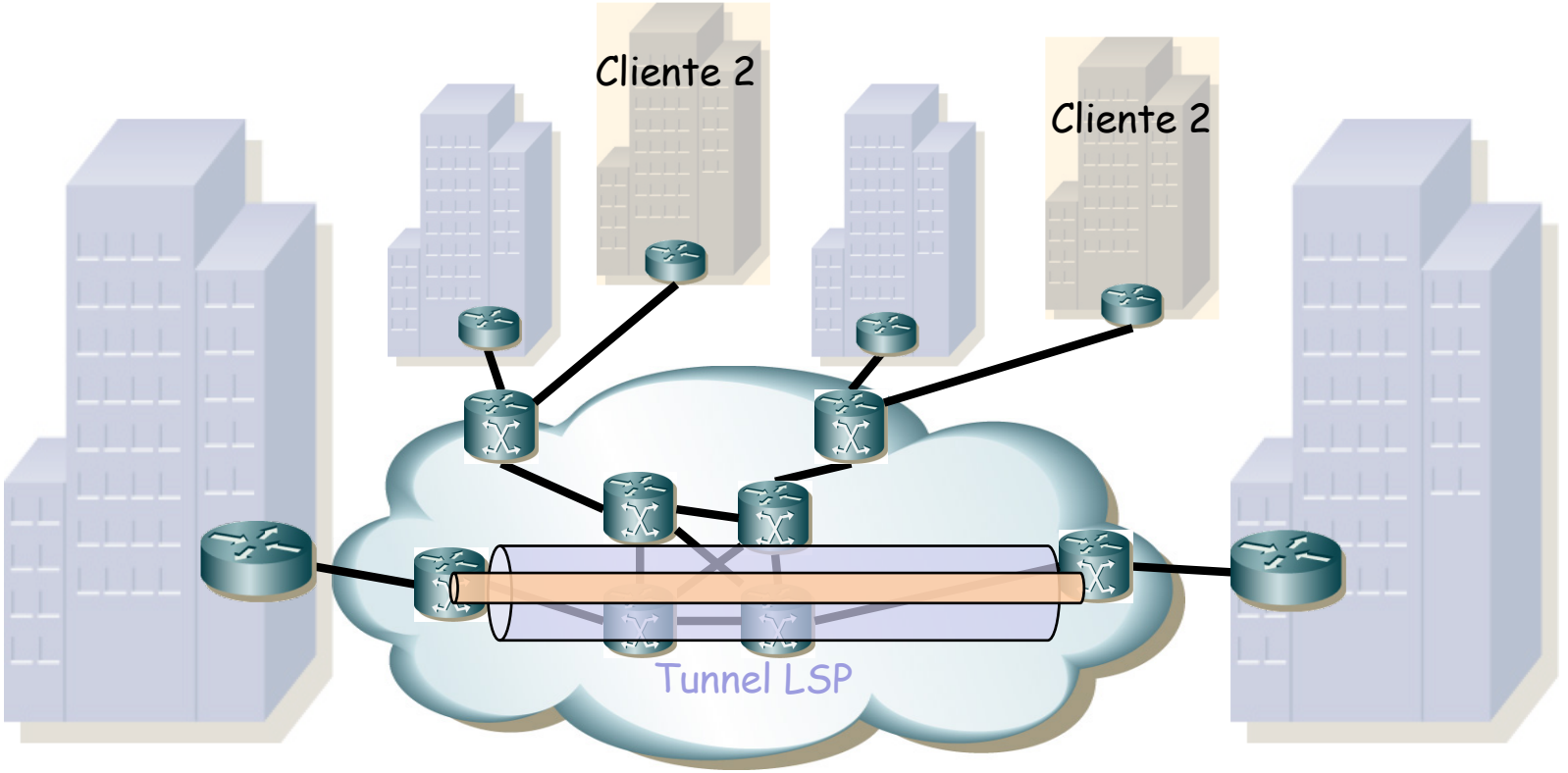
# L3VPN: MPLS

- ¿Y en los P routers? ¿Reenvían en función de esa etiqueta?
- Se crean LSPs entre los PEs (“Tunnel LSPs”)
- Los P routers reenvían en función de esa etiqueta externa
- (…)



# L3VPN: MPLS

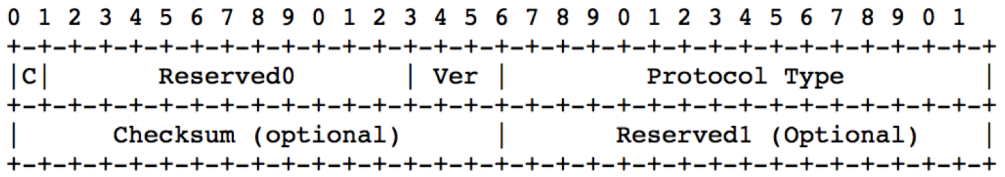
- ¿Y en los P routers? ¿Reenvían en función de esa etiqueta?
- Se crean LSPs entre los PEs (“Tunnel LSPs”)
- Los P routers reenvían en función de esa etiqueta externa
- Un full-mesh entre los PEs que compartan VRF
- Podrían ser otro tipo de túneles (GRE o IP en IP, RFC 4797), lo cual elimina el requerimiento de una red de transporte MPLS

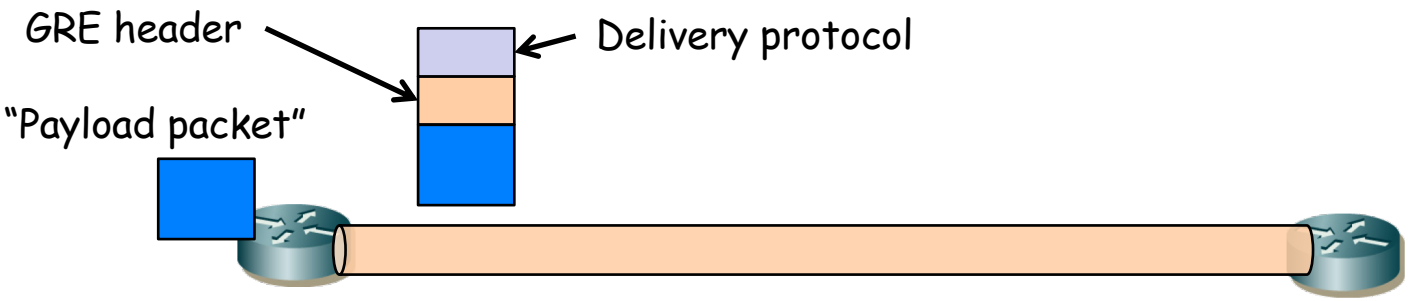


# GRE

# GRE

- RFC 2784 “Generic Routing Encapsulation (GRE)” (Procket Networks, Enron Communications, Cisco Systems, Juniper Networks, 2000)
- Encapsular un nivel de red en otro nivel de red
- PPTP (Point-to-Point tunneling Protocol) usa algo similar a GRE
- La cabecera básica GRE ocupa 8 bytes
- Uno de los campos es un Ethertype (*Protocol Type*)
- La versión anterior (RFC 1701) tenía más campos que desaparecen en esta
- Aunque algunos se recuperan en la RFC 2890 “Key and Sequence Number Extensions to GRE” (Cisco, 2000) (…)

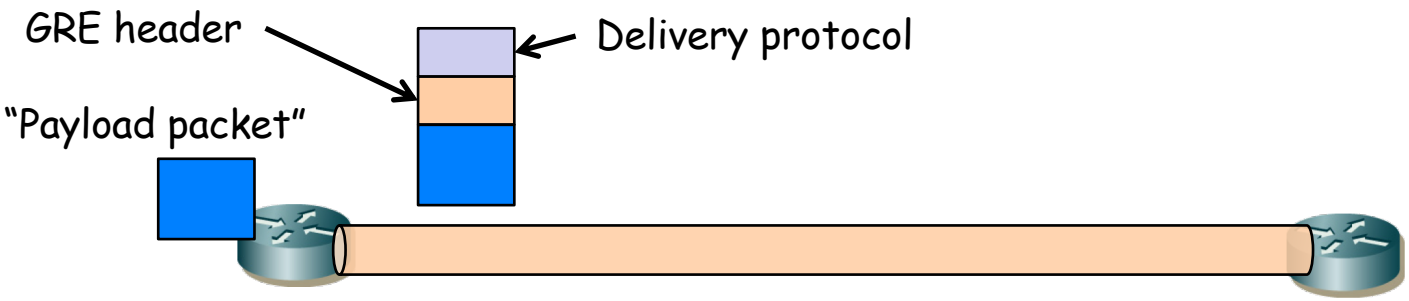


# GRE

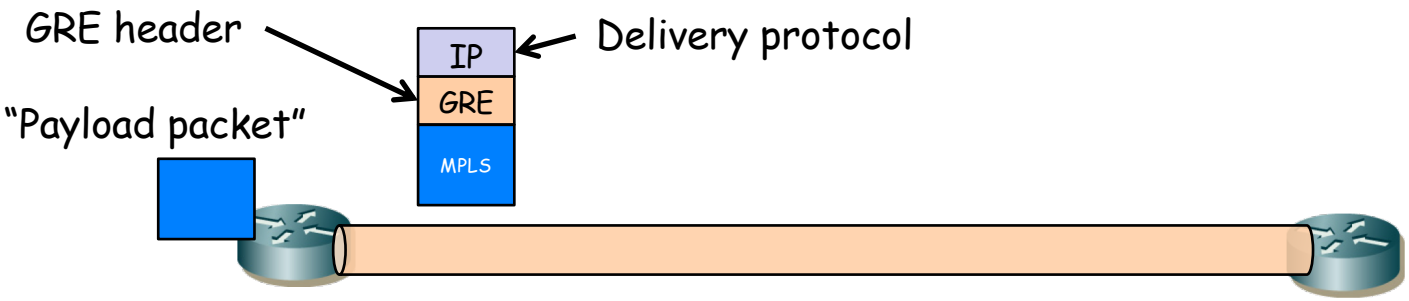
- RFC 2890 “Key and Sequence Number Extensions to GRE”
- “Key” sirve para distinguir flujos dentro del túnel
- “Sequence Number”
  - Si hay “key” entonces el número de secuencia es por “key”
  - Permite dar entrega en orden (aunque no fiable)
  - Si llega uno “anterior” lo descarta
  - Si llega uno que deja un hueco puede guardarlo intentando reconstruir la secuencia
  - Pasado cierto tiempo sin lograr reconstruir los reenvía

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|C| |K|S| Reserved0       | Ver |         Protocol Type         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Checksum (optional)      |       Reserved1 (Optional)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Key (optional)                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Sequence Number (Optional)                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```



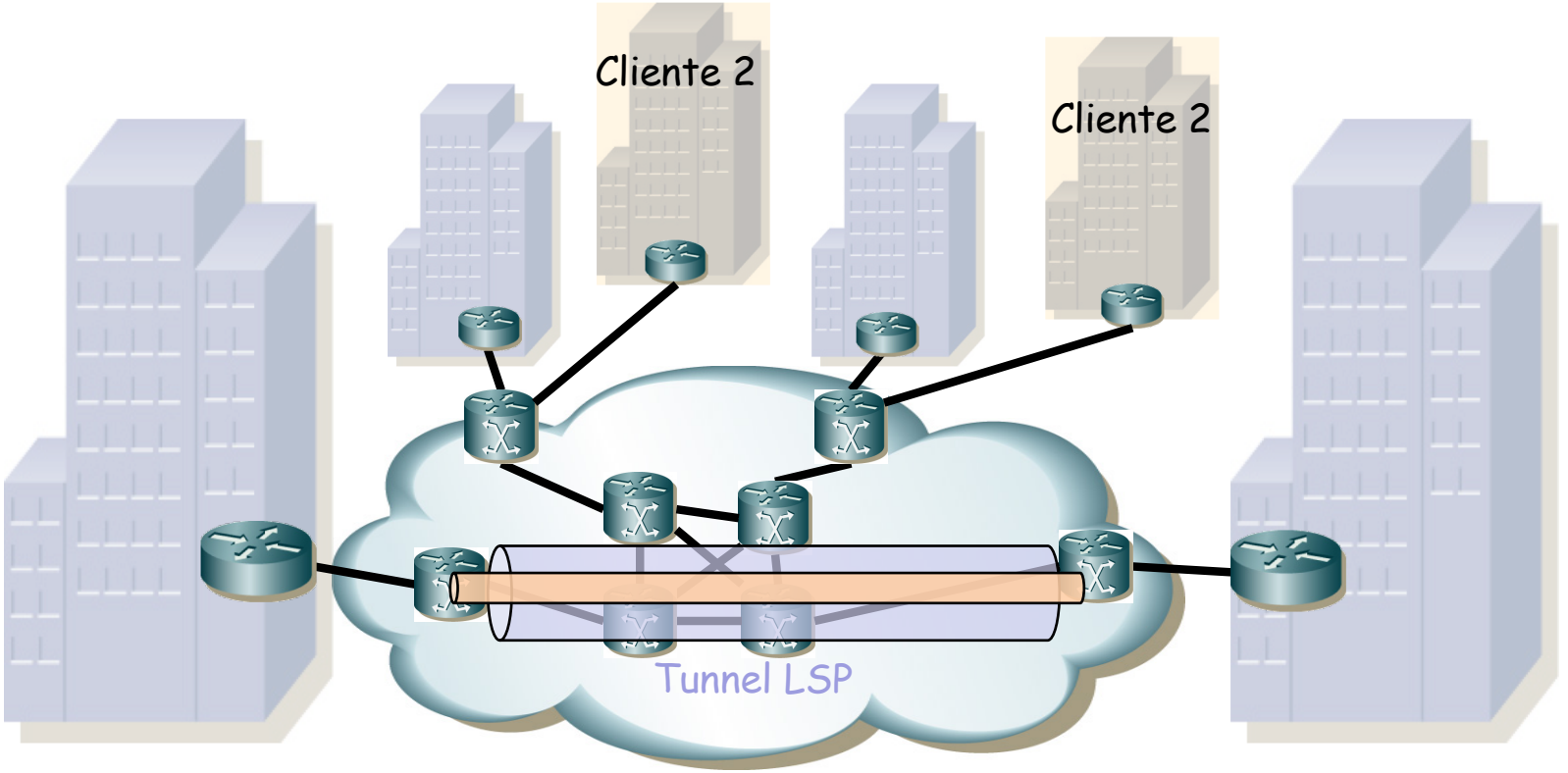
# *MPLS in GRE in IP*

- RFC 4023 “Encapsulating MPLS in IP or Generic Ruting Encapsulation (GRE)” (Motorola, Juniper, Cisco, 2005)
- El “*delivery protocol*” podría ser IP (protocol = 47 = GRE)
- El “*payload packet*” podría ser MPLS (Ethertype 0x8847 para unicast y ese mismo ó 0x8848 para multicast, RFC 5332)
- EoMPLSoGRE = Ethernet over MPLS over GRE
- Al transportarse sobre IP puede emplear IPSec
- RFC 4023 contempla también que MPLS se transporte directamente sobre IP, lo cual es más eficiente (sin GRE, protocolo 137 sobre IP)
- Puede haber motivos para tener GRE (exista el túnel con anterioridad, la implementación del equipo lo requiera en su fastpath, etc)



# L3VPN con GRE

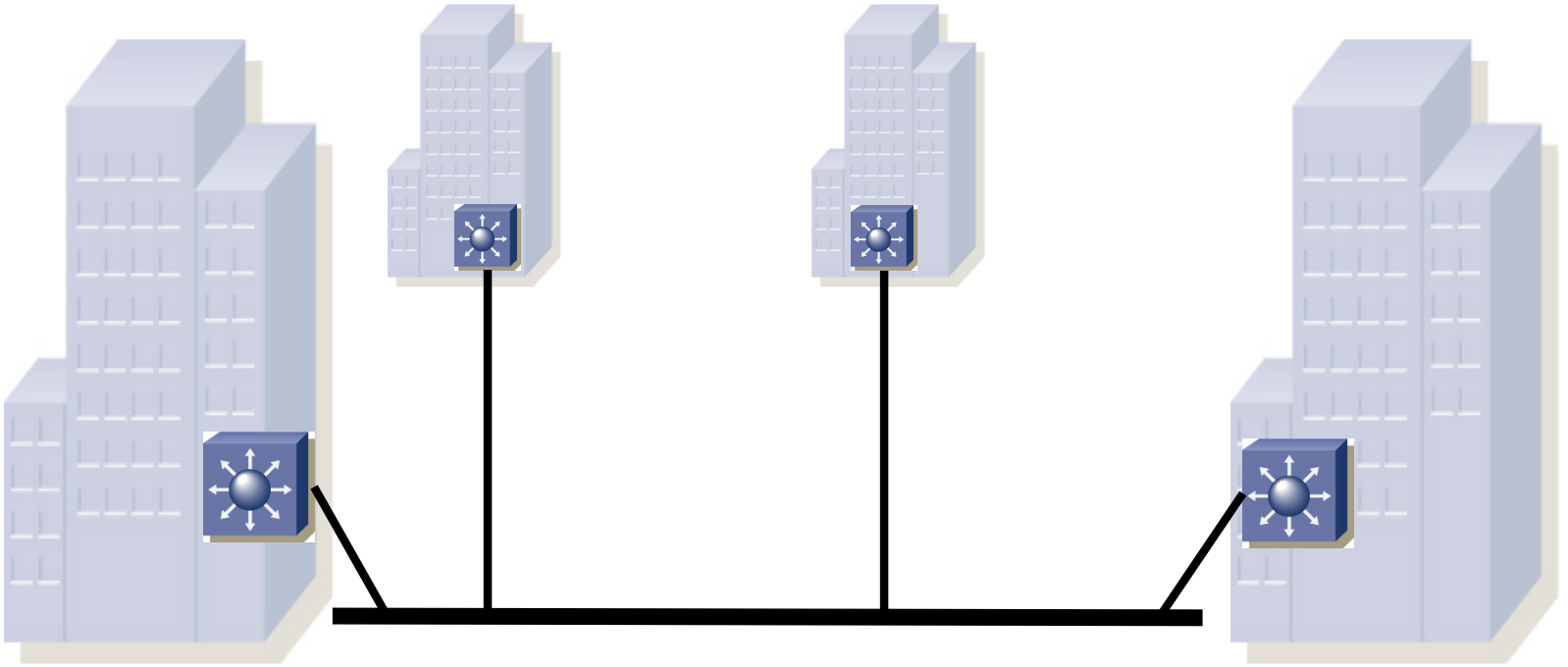
- Como decíamos, los túneles entre los PE pueden ser túneles GRE o simple IPoIP
- Entonces la red de transporte ya no necesita ser MPLS, es simple IP



# VPLS

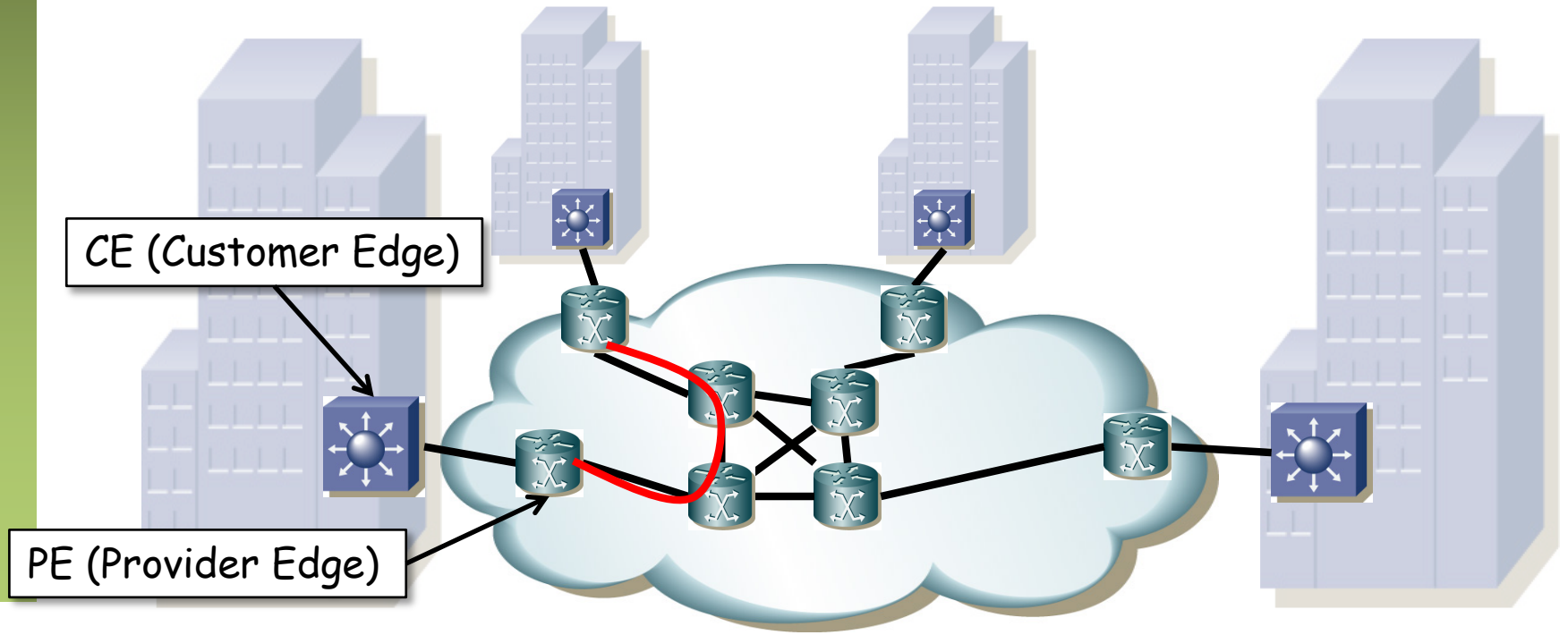
# MPLS y VPLS

- *"Virtual Private LAN Service",* una VPN layer 2 (RFC 4664, Acreo y Cisco, 2006)
- Interconecta múltiples *sites* en un solo dominio puenteado
- Todos los extremos se comportan como si estuvieran en una LAN
- *E-LAN Service*
- Transporta Ethernet así que sobre ella el cliente puede usar IP o cualquier otro protocolo
- Los equipos de usuario (Customer Edge) pueden ser switches o routers
- Transporte MPLS u otra solución de túneles (GRE, L2TP, IPsec)



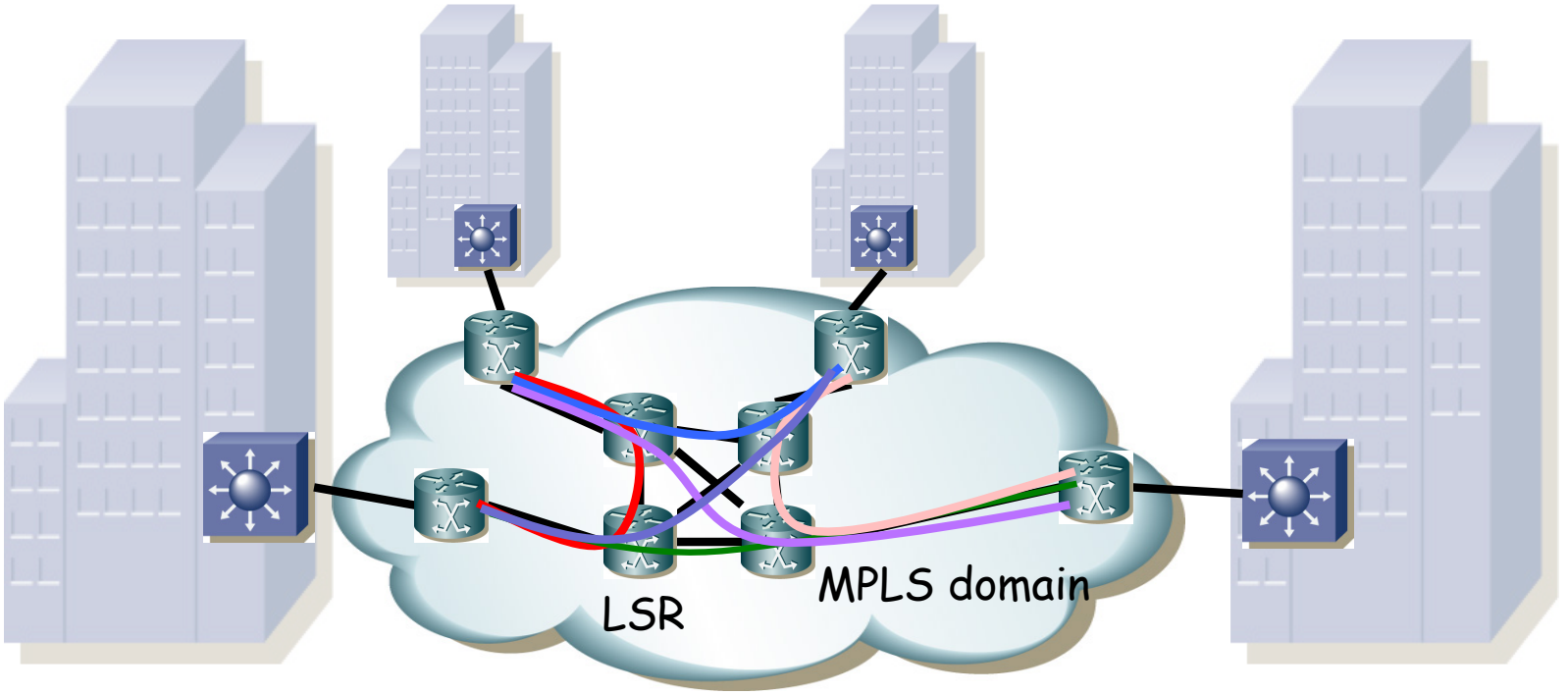
# MPLS y VPLS

- El dominio MPLS puede transportar las tramas MPLS sobre IP o sobre otra tecnología
- La red puede dar servicio VPLS a más de un cliente
- El PE hace aprendizaje de direcciones MAC y replicación de tramas de forma independiente para cada cliente
- No interfiere el servicio de un usuario al otro (pueden por ejemplo emplear el mismo direccionamiento IP)
- Los equipos frontera establecen entre ellos los LSPs necesarios para el servicio multiacceso



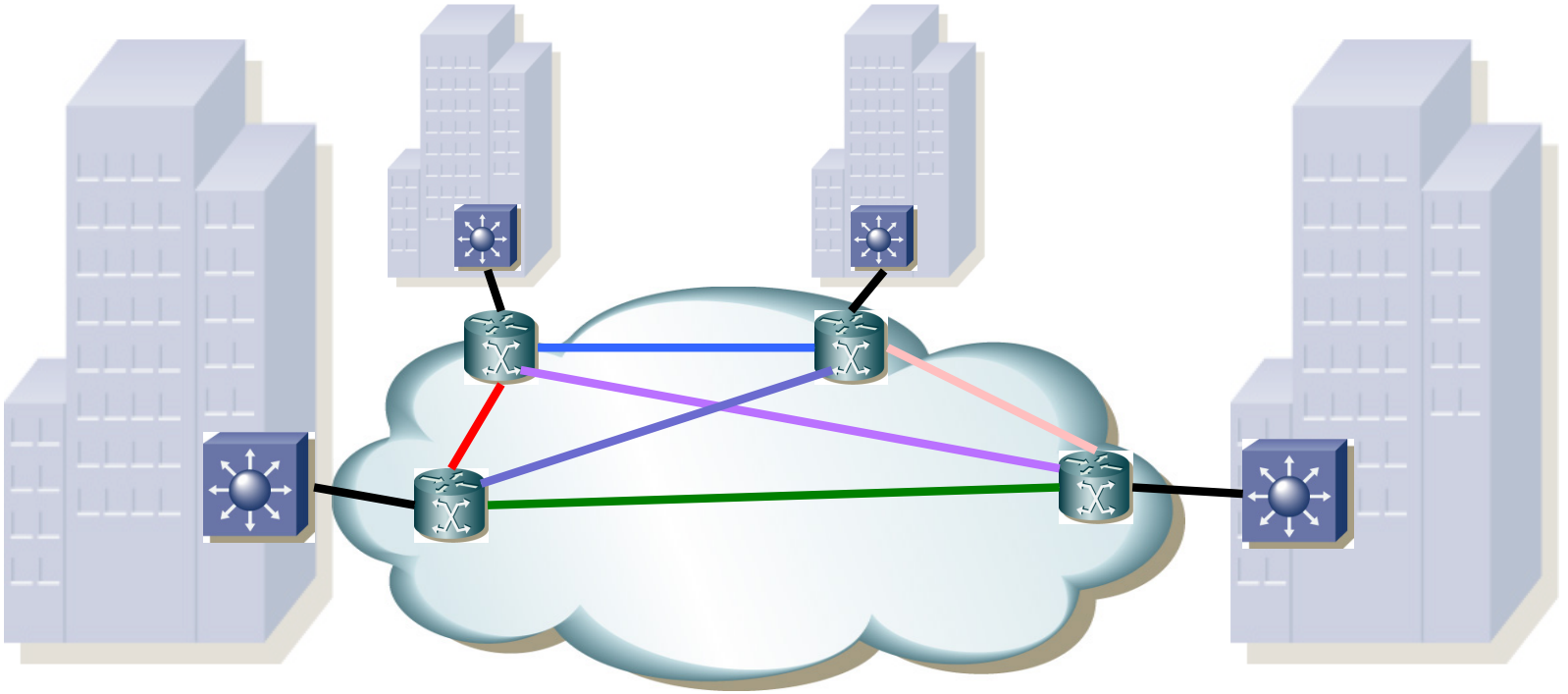
# MPLS y VPLS

- Establecen un *full-mesh* entre los nodos frontera
- Para ello disponen de RSVP-TE (o LDP)
- (…)



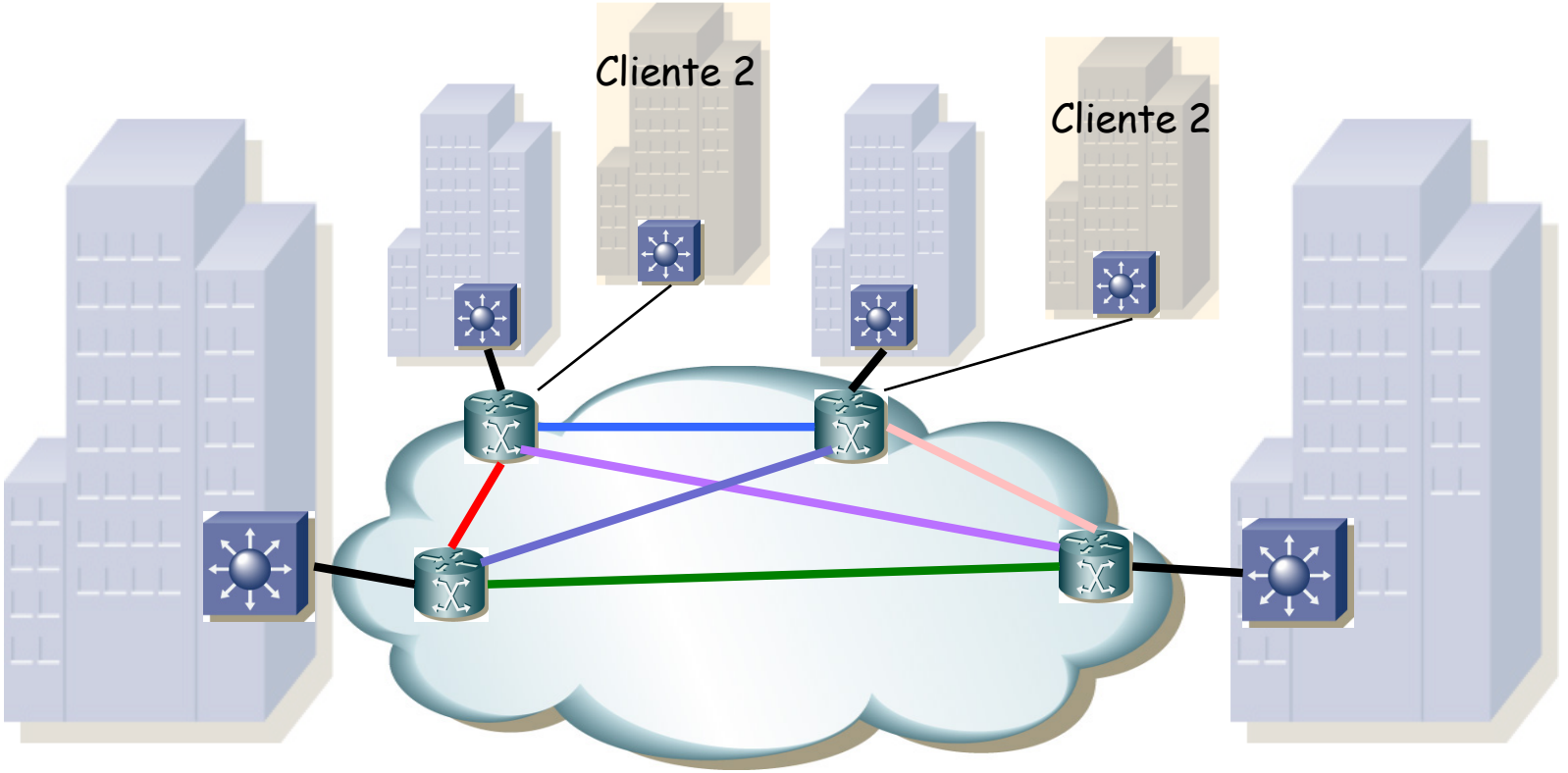
# MPLS y VPLS

- Establecen un *full-mesh* entre los nodos frontera
- Para ello disponen de RSVP-TE (o LDP)
- Esos LSPs son globales al servicio VPLS, no particulares para cada cliente
- Es decir, puede haber otras LANs creadas con VPLS, para las sedes de otra empresa, y emplearán los mismos LSPs (…)



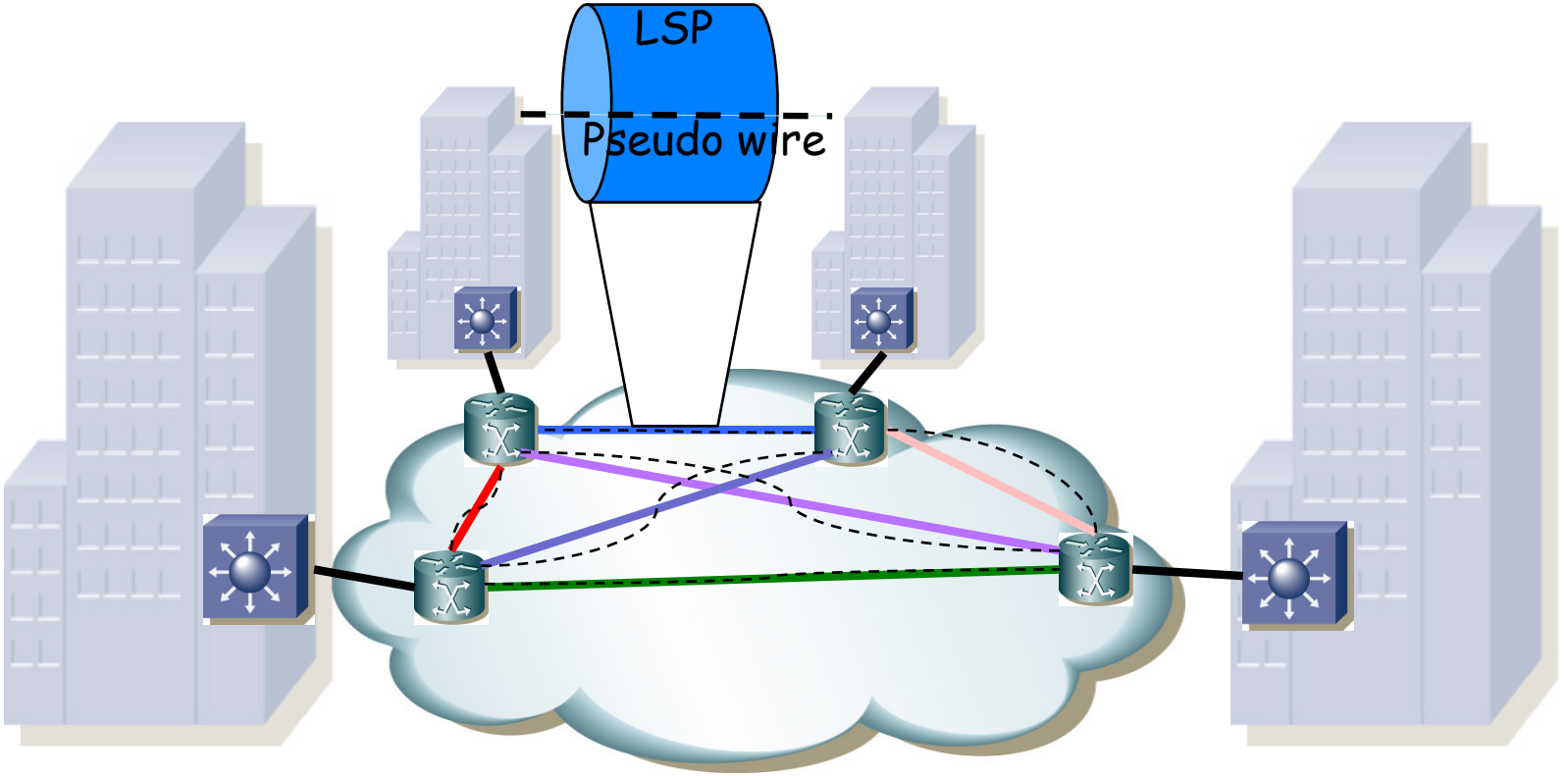
# MPLS y VPLS

- Establecen un *full-mesh* entre los nodos frontera
- Para ello disponen de RSVP-TE (o LDP)
- Esos LSPs son globales al servicio VPLS, no particulares para cada cliente
- Es decir, puede haber otras LANs creadas con VPLS, para las sedes de otra empresa, y emplearán los mismos LSPs
- ¿Y para diferenciar a los clientes?



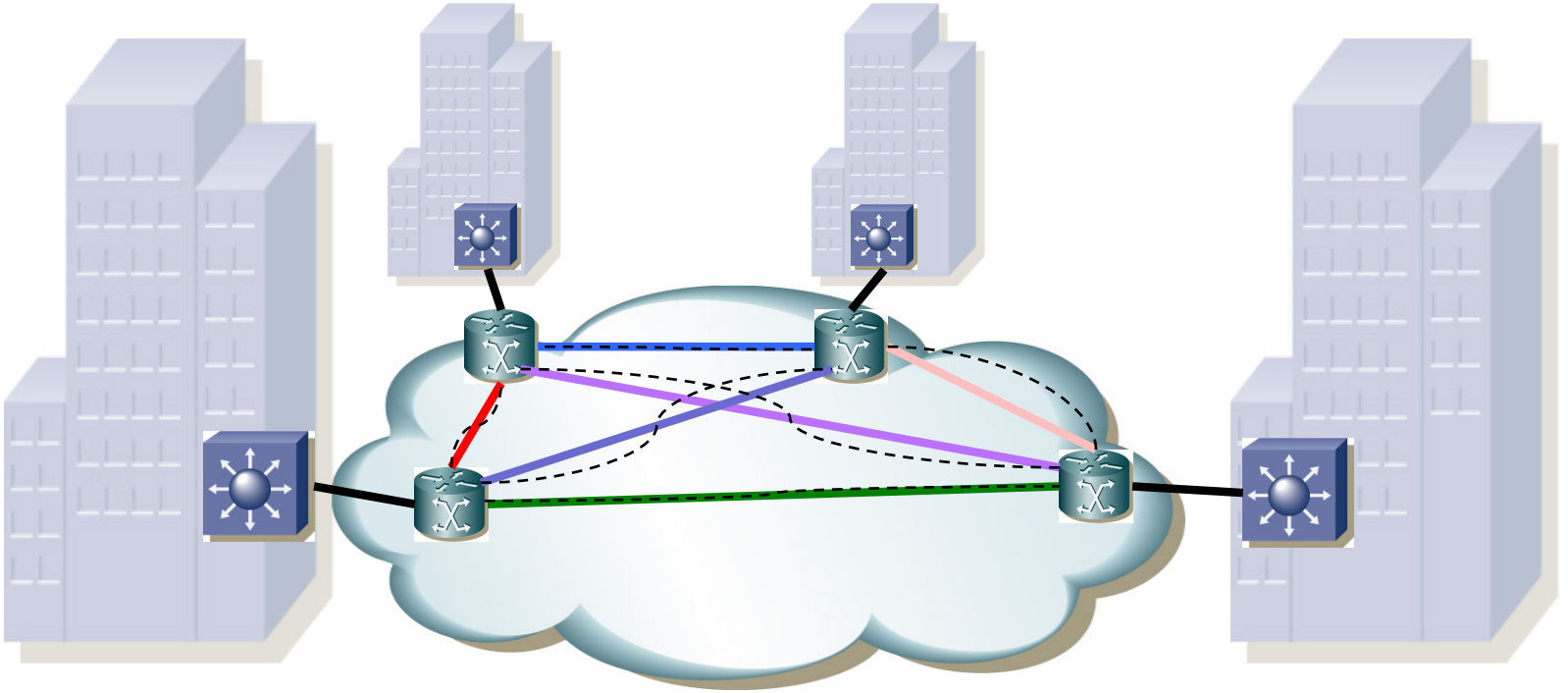
# VPLS y PWE

- Por cada instancia VPLS (cada cliente) se establece un full mesh de *pseudo-wires* (PWs) entre los PEs
- RFC 3985 "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture" (Cisco Systems, Overture Networks, 2005)
- Un PW emula un circuito, por ejemplo para transportar un E1 o un PVC ATM
- También puede transportar Ethernet, AAL5, SDH, etc



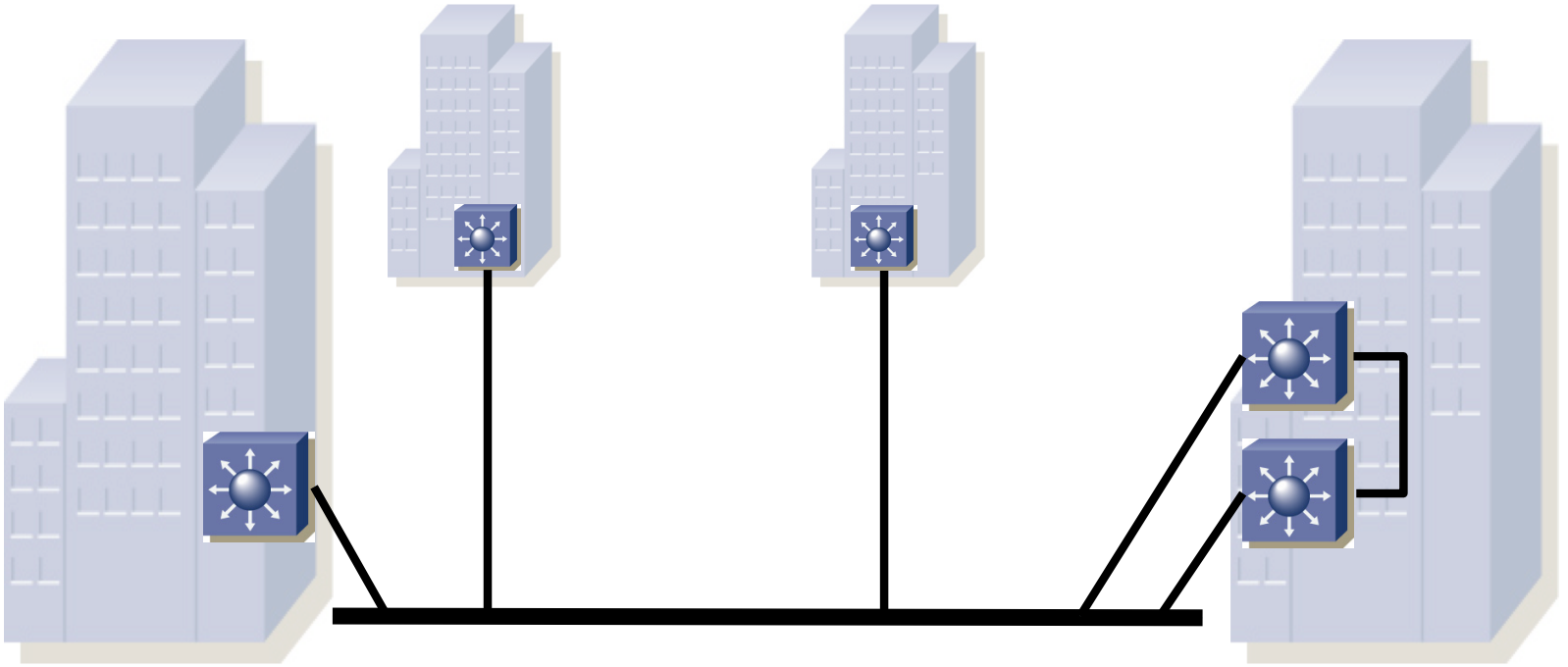
# VPLS y PWE

- El full-mesh de PWs hace que los PE puedan enviarse directamente los unos a los otros
- No necesitan hacer reenvío y no hace falta resolver posibles bucles
- Simplemente se implementa una solución que se llama de "*split horizon*":
  - Un PE no debe reenviar tráfico de un PW a otro en el mismo mesh VPLS
- El aprendizaje de direcciones MAC se hace en el plano de datos (con la llegada de tramas Ethernet)



# VPLS y PWE

- Sí puede haber ciclos, pero creados por el usuario para obtener redundancia
- En ese caso podrá emplear STP
- Las BPDUs se transportarían normalmente por el mesh VPLS



# VPLS Control Plane

- Dos alternativas para el establecimiento de los pseudo-wires:
  - RFC 4761 “Virtual Private LAN Service (VPLS) Using BGP or Auto-Discovery and Signaling”
  - RFC 4762 “Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling”
- Repito: el aprendizaje de direcciones MAC se hace en el plano de datos, es decir, con la dirección MAC origen de la trama recibida

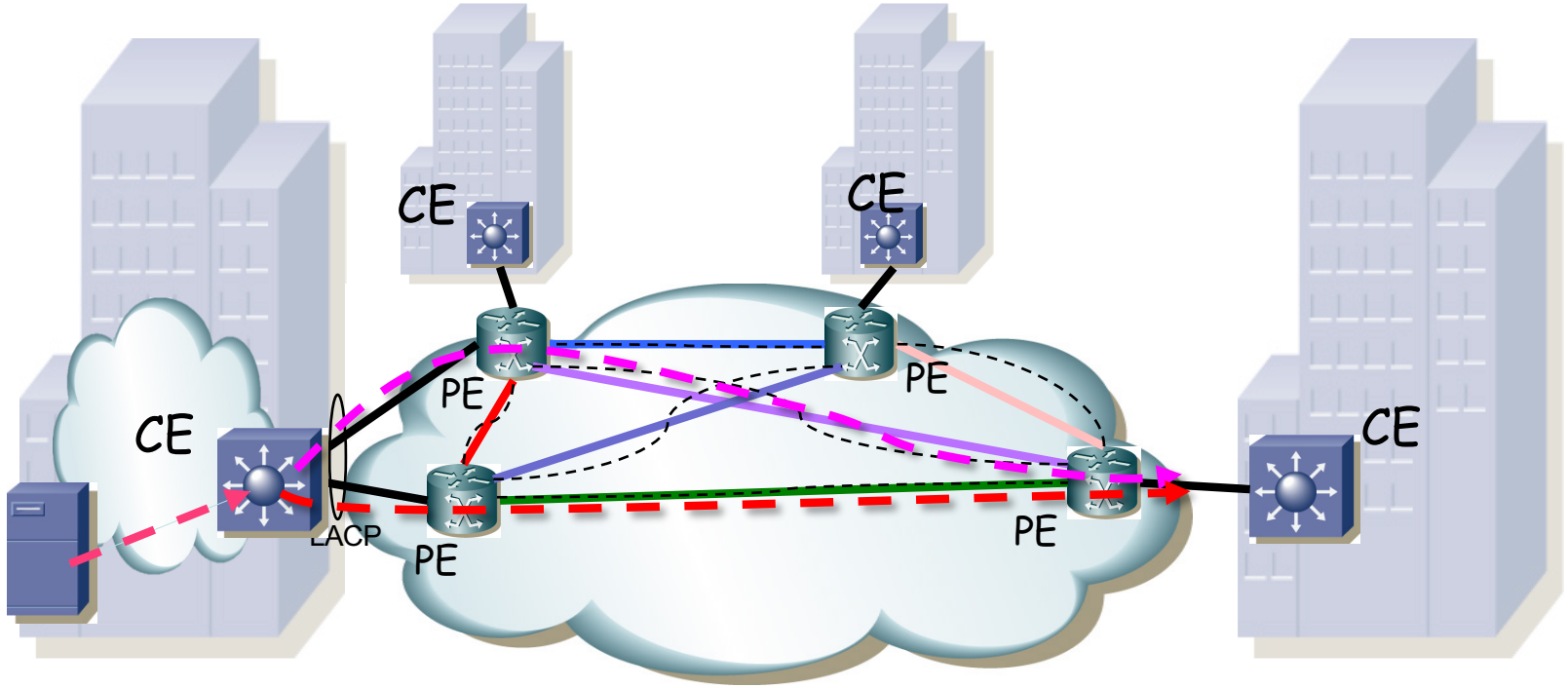
# Problemas en VPLS

- Se deben establecer Nx(N-1)/2 pseudo-wires
- Problema de escalabilidad (cantidad de tráfico de control)
- Replicación de paquetes que sufren inundación:
  - Se lleva a cabo en el PE de entrada
  - Se dirigen punto-a-punto a cada otro PE del servicio
  - Mayor trabajo en el PE
  - Más uso de capacidad
  - Mayor retardo (si hay que enviar N veces la trama por N PWs que se implementan sobre el mismo LSP irán en serie)
- Si se añade un acceso del cliente, a un PE diferente, se deben crear los PWs, lo cual implica reconfigurar los demás PEs
- Para despliegues pequeños
- Mejoras:
  - H-VPLS (Hierarchical VPLS)
  - Hierarchical BGP VPLS

# EVPN

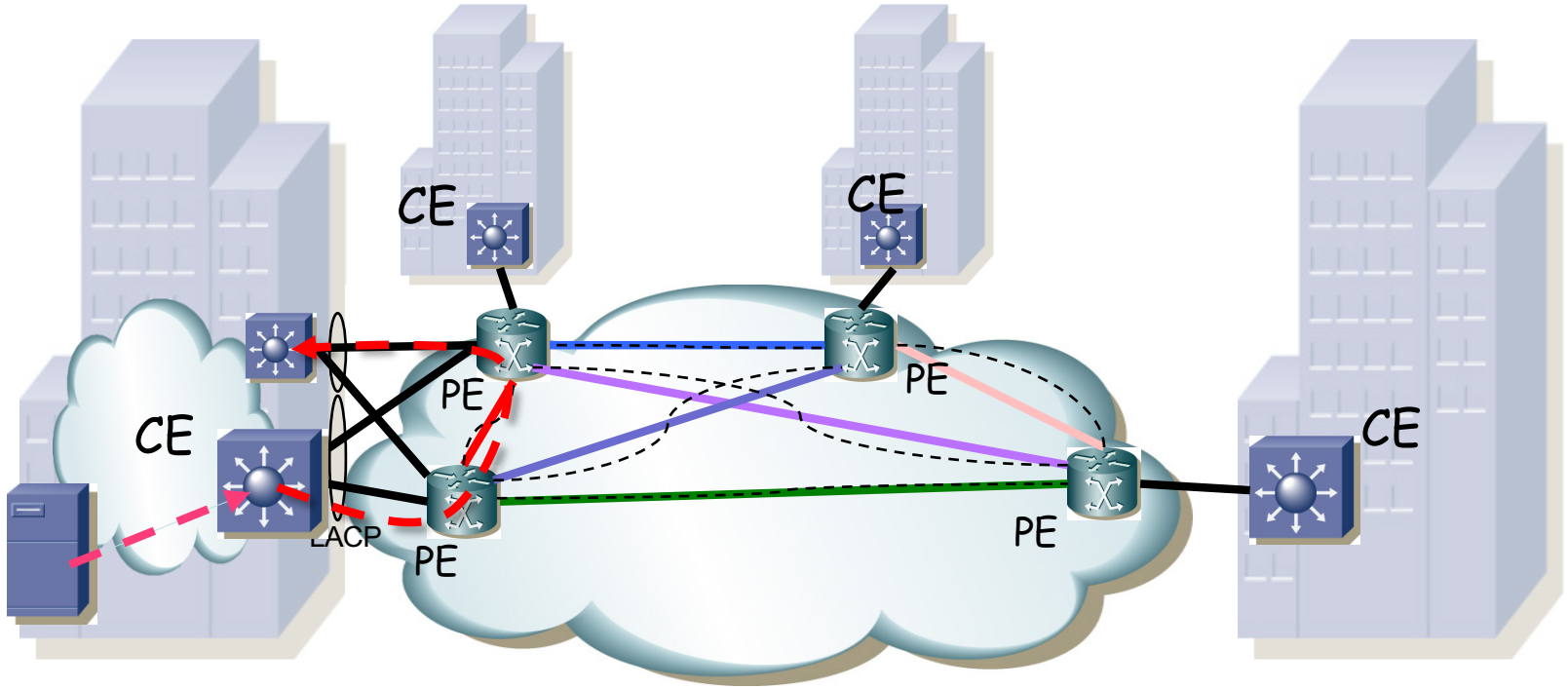
# Limitaciones de VPLS

- Solo soporta *Single-Active Redundancy Mode* (no *all-active*)
- Ejemplo:
  - CE un LAG hacia dos PEs
  - Reparte tráfico entre ellos
  - Al llegar a otro PE llega la misma dirección MAC origen por dos PWs, saltando la MAC aprendida de uno a otro (...)
- Active-Active solo mediante soluciones propietarias (vPC, VSS, etc)



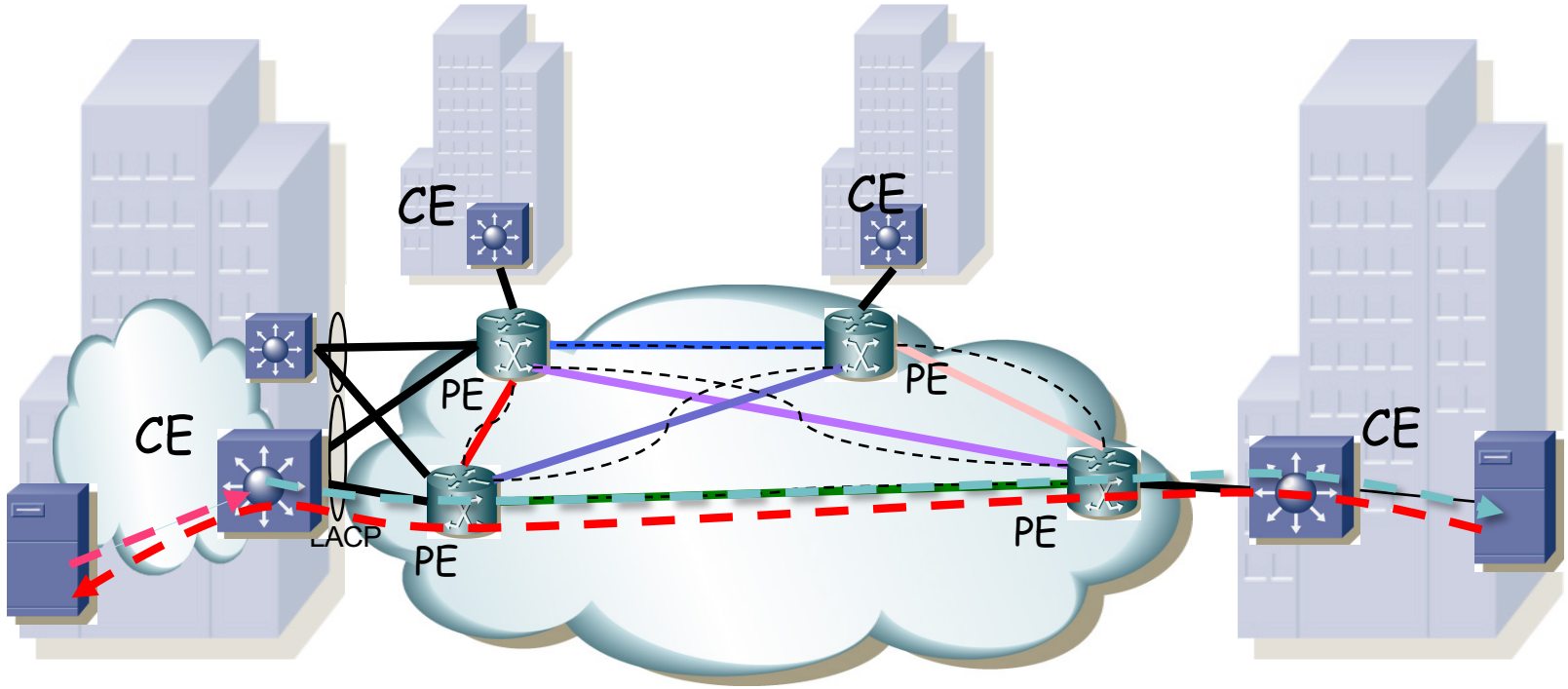
# Limitaciones de VPLS

- Solo soporta *Single-Active Redundancy Mode* (no *all-active*)
- Ejemplo:
  - Dos CEs con LAGs (redundancia en el CE)
  - BUM es enviado por PE a todos los demás PEs y puede volver por el otro CE (...)



# Limitaciones de VPLS

- Solo soporta *Single-Active Redundancy Mode* (no *all-active*)
- Ejemplo:
  - No hay load balancing en el sentido de respuesta si el hash lleva a emplear uno solo de los PEs (...)
  - El otro extremo solo aprende la dirección MAC por un PW
  - En cualquier caso no la puede aprender por dos PWs (...)



# EVPN

- RFC 7209 “Requirements for Ethernet VPN (EVPN)”, Cisco, Arktan, AT&T, Verizon, Alcatel-Lucent, Bloomberg (2014)
- RFC 7432 “BGP MPLS-Based Ethernet VPN”, Cisco, Arktan, Verizon, Bloomberg, AT&T, Juniper, Alcatel-Lucent (2015)
- Ofrece una VPN capa 2, como VPLS
- Los PEs puede estar conectados mediante LSPs o túneles (IP/GRE)
- Emplea un plano de control **BGP** como una L3VPN
- Aprendizaje de direcciones MAC en el plano de control en lugar de en el plano de datos
- Es decir, BGP distribuye Ethernet MACs (opcional el par MAC-IP)
- PEs anuncian direcciones MAC aprendidas del CE, junto con una etiqueta MPLS, al resto de PEs mediante MP-BGP
- Queda abierto cómo aprende el PE del CE (puede ser plano de datos)
- Igual que las L3VPNs emplea *route distinguishers* y *route targets*
- Su uso principal es en DCI