



# Mission-Critical Communication over IP-based Networks

## Considerations, Technology and Components

Version: 3.0 / 2016-05-10



### Network, IT-Infrastructure and Security

**Manfred Lindner**

manfred.lindner @ frequentis.com

Senior Network & Security Architect

lindner @ ict.tuwien.ac.at (obsolete)

ml-consulting @ aon.at

Lectures: Data Communication

<https://www.ict.tuwien.ac.at/lva/384.081/index.html>

**FREQUENTIS**

# Agenda

---

- **Introduction**
- **Network Operational Model**
- **High Availability**
- **QoS**
- **VPN Technology**
- **Multicasting**
- **Summary**

# Agenda

---

- **Introduction**

- Circuit Switching (Based on Synchronous TDM)
- Packet Switching (Based on Asynchronous TDM)
- Impact Of Change To Best Effort IP For Real-Time Communication
- Relevant Areas For Design Of Mission Critical Networks

- **Network Operational Model**

- **High Availability**

- **QoS**

- **VPN Technology**

- **Multicasting**

- **Summary**

# Agenda

---

- **Introduction**
- **Network Operational Model**
  - Model M1 (Based L1-VPN)
  - Model M2 (Based L2-VPN)
  - Model M3 (Based L3-VPN)
- **High Availability**
- **QoS**
- **VPN Technology**
- **Multicasting**
- **Summary**

# Agenda

---

- **Introduction**
- **Network Operational Model**
- **High Availability**
  - Elements of HA
  - Functional Access Block Types for HA
  - Routing Aspects
- **QoS**
- **VPN Technology**
- **Multicasting**
- **Summary**

# Agenda

---

- **Introduction**
- **Network Operational Model**
- **High Availability**
- **QoS**
  - Introduction QoS
  - IP QoS Mechanism
  - QoS Handling M1, M2 or M3 Environment
- **VPN Technology**
- **Multicasting**
- **Summary**

# Agenda

---

- **Introduction**
- **Network Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
  - Introduction IT-Security
  - VPN Types
  - MPLS, MPLS-VPN
  - IPsec VPN
  - DMVPN
  - GETVPN
- **Multicasting**
- **Summary**

# Agenda

---

- **Introduction**
- **Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
- **Multicasting**
  - Introduction
  - Multicast Routing Overview
  - Multicast & HA
  - Multicast & VPN / Security
- **Summary**



# Agenda

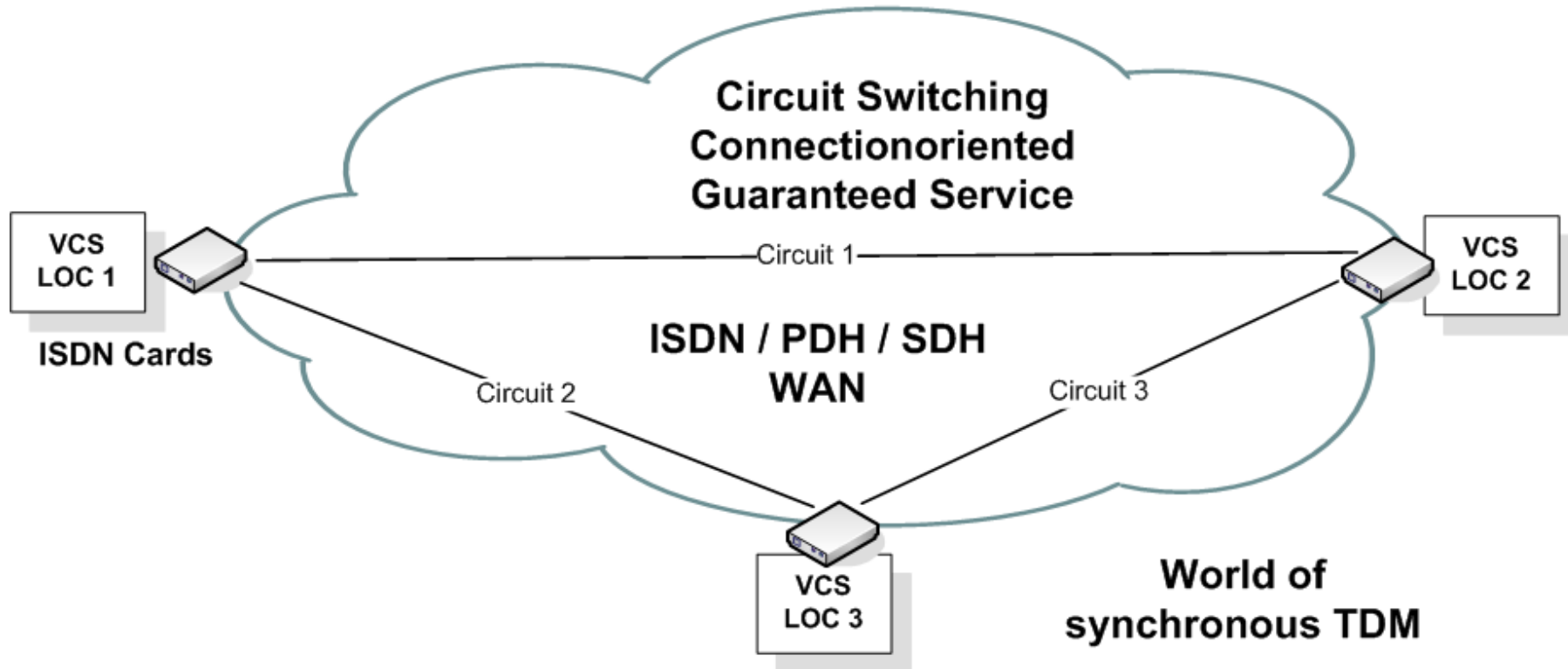
---

- **Introduction**
- **Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
- **Multicasting**
- **Summary**
  - Design Issues
  - LISP Intro
  - IP Technology Facts

# Agenda

---

- **Introduction**
- **Network Operational Model**
- **High Availability**
- **QoS**
- **VPN Technology**
- **Multicasting**
- **Summary**

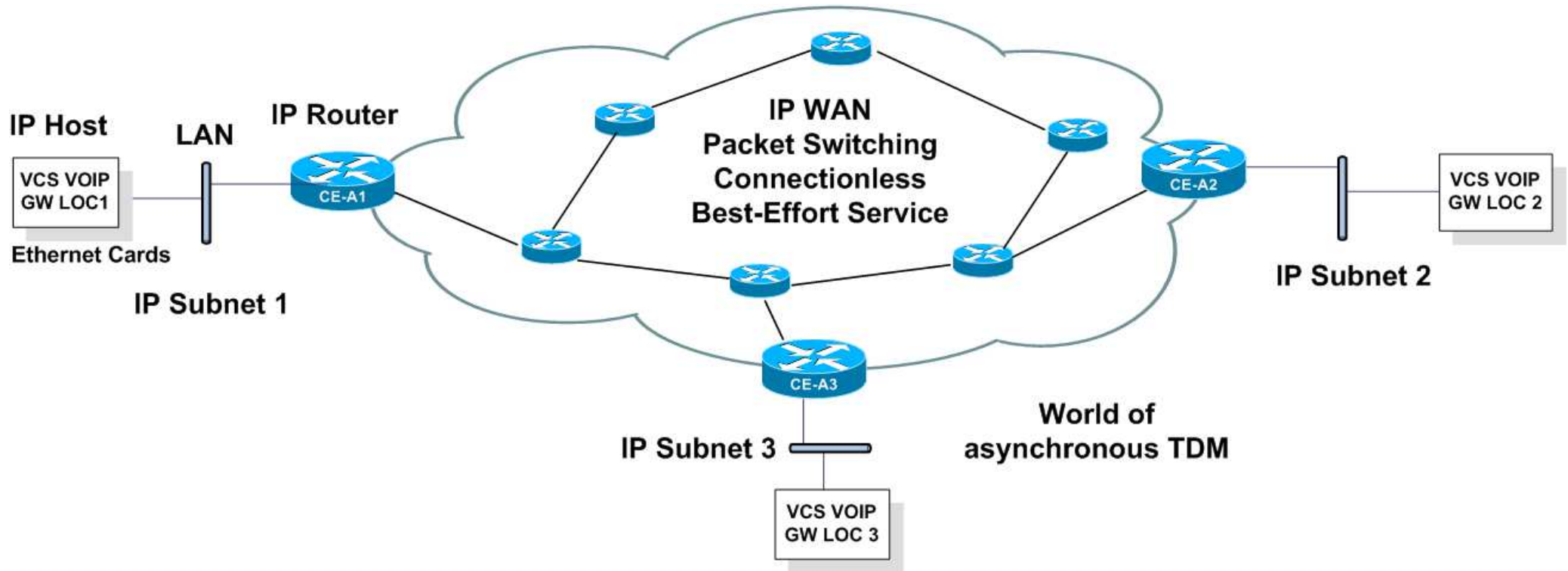


VCS ... Voice Communication System

TDM ... Time Division Multiplexing

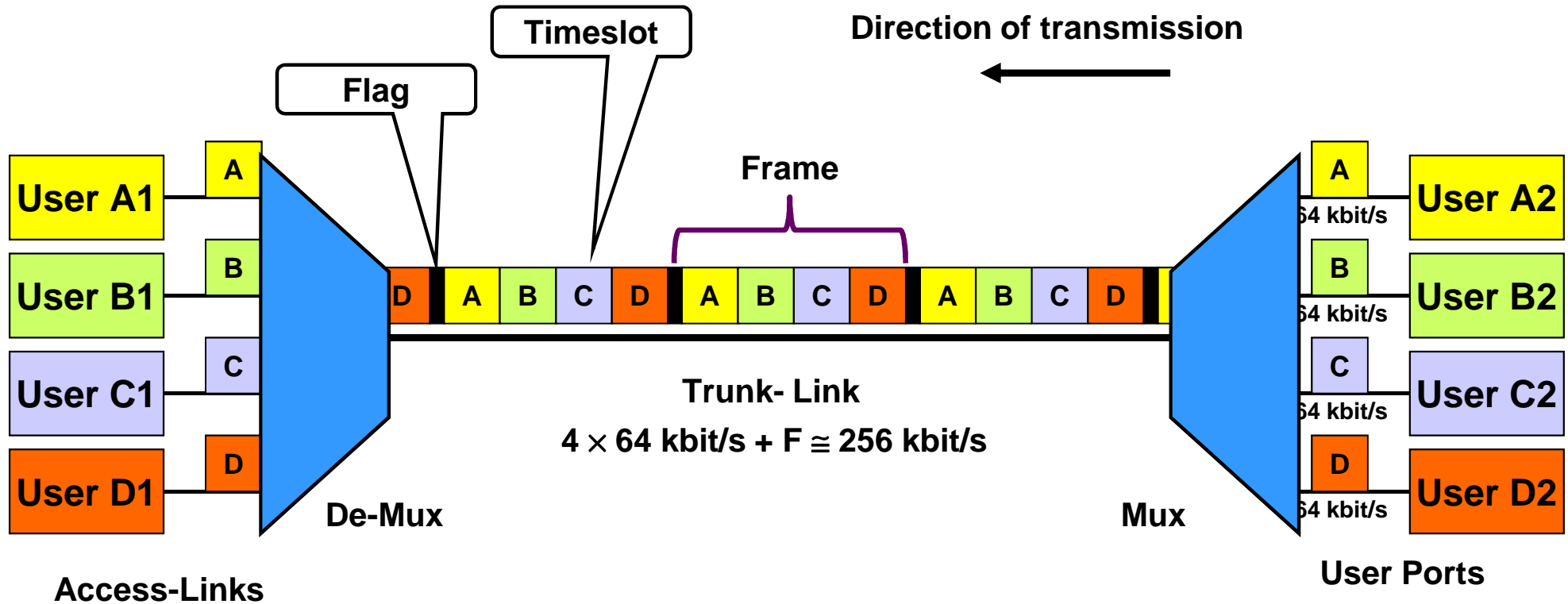
ISDN ... Integrated Services Digital Network

PDH / SDH ... Plesiochronous / Synchronous Digital Hierarchy



**VOIP ... VOice over IP**  
**GW ... Gateway**  
**LAN ... Local Area Network**  
**WAN ... Wide Area Network**

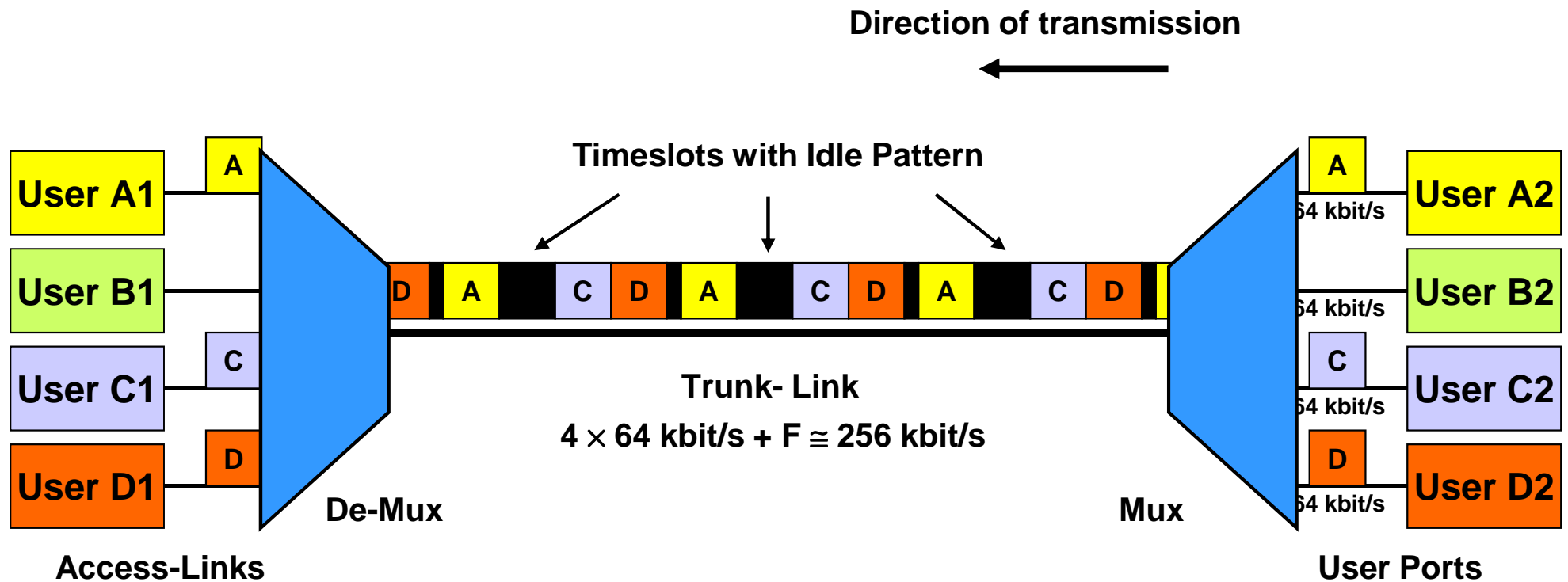
# Synchronous TDM (1)



Periodic frames consisting of a constant number of timeslots  
Every channel occupies a dedicated timeslot  
Implicit addressing given by the position of a timeslot in the frame  
Trunk rate = number of timeslots x access-link rate

Each channel experiences constant delay and no delay variation (jitter)

# Synchronous TDM (2)



Timeslot can be used for any kind of communication

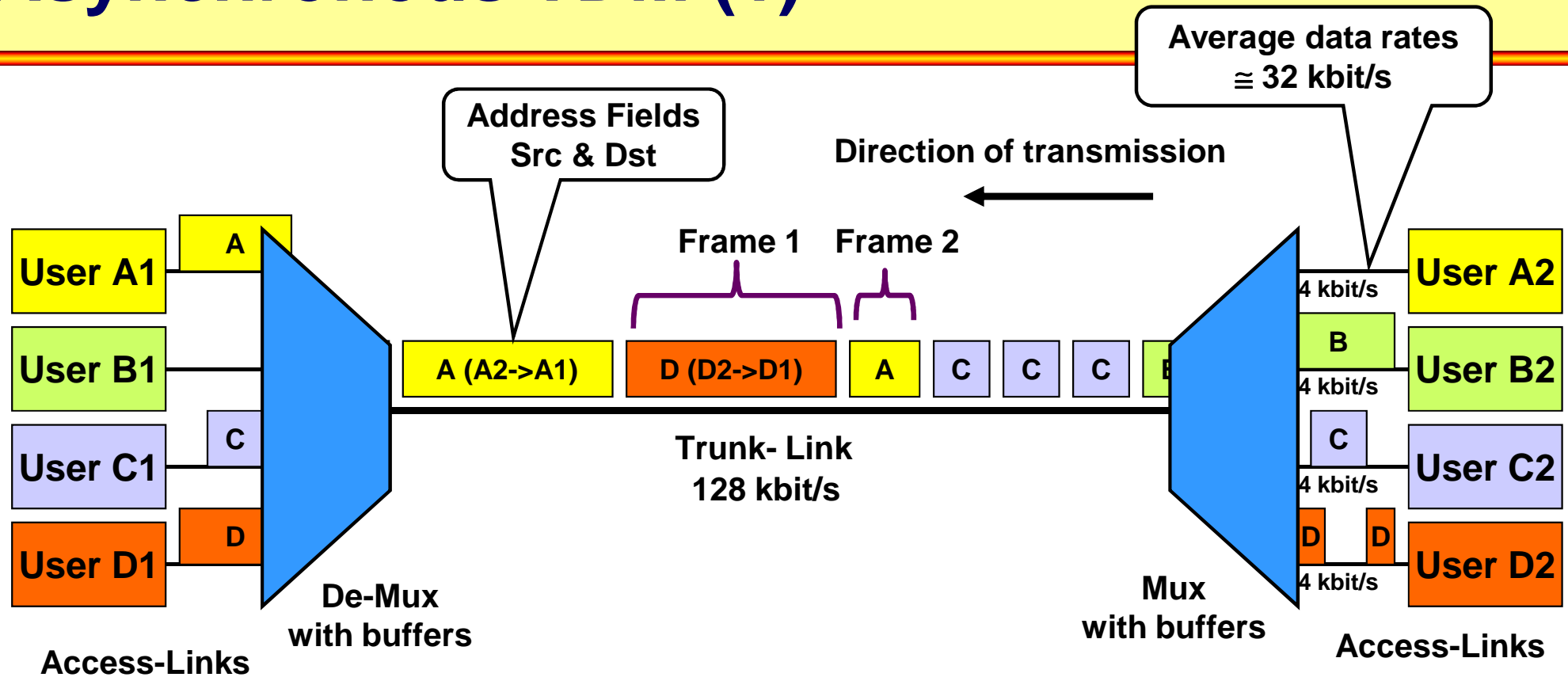
-> protocol transparency

But empty timeslots are not useable by other communication channels

-> waste of bandwidth during times of inactivity

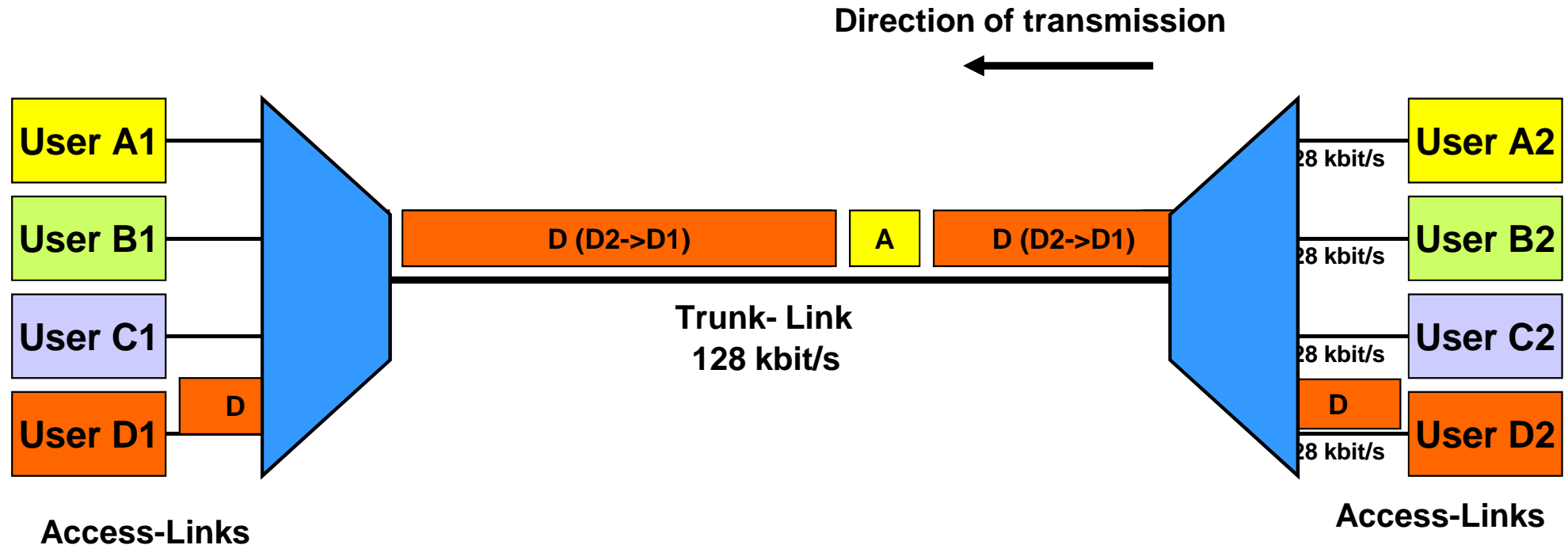
Lead to development of asynchronous/statistical multiplexing

# Asynchronous TDM (1)



- Trunk rate is dimensioned for average usage in statistical manner
- Each user channels can send packets whenever he/she wants
- Frames have different lengths
- Buffering is necessary if trunk is already occupied by another channel
- Explicit addressing by usage of address fields in the frame
- Not protocol-transparent any more

# Asynchronous TDM (2)

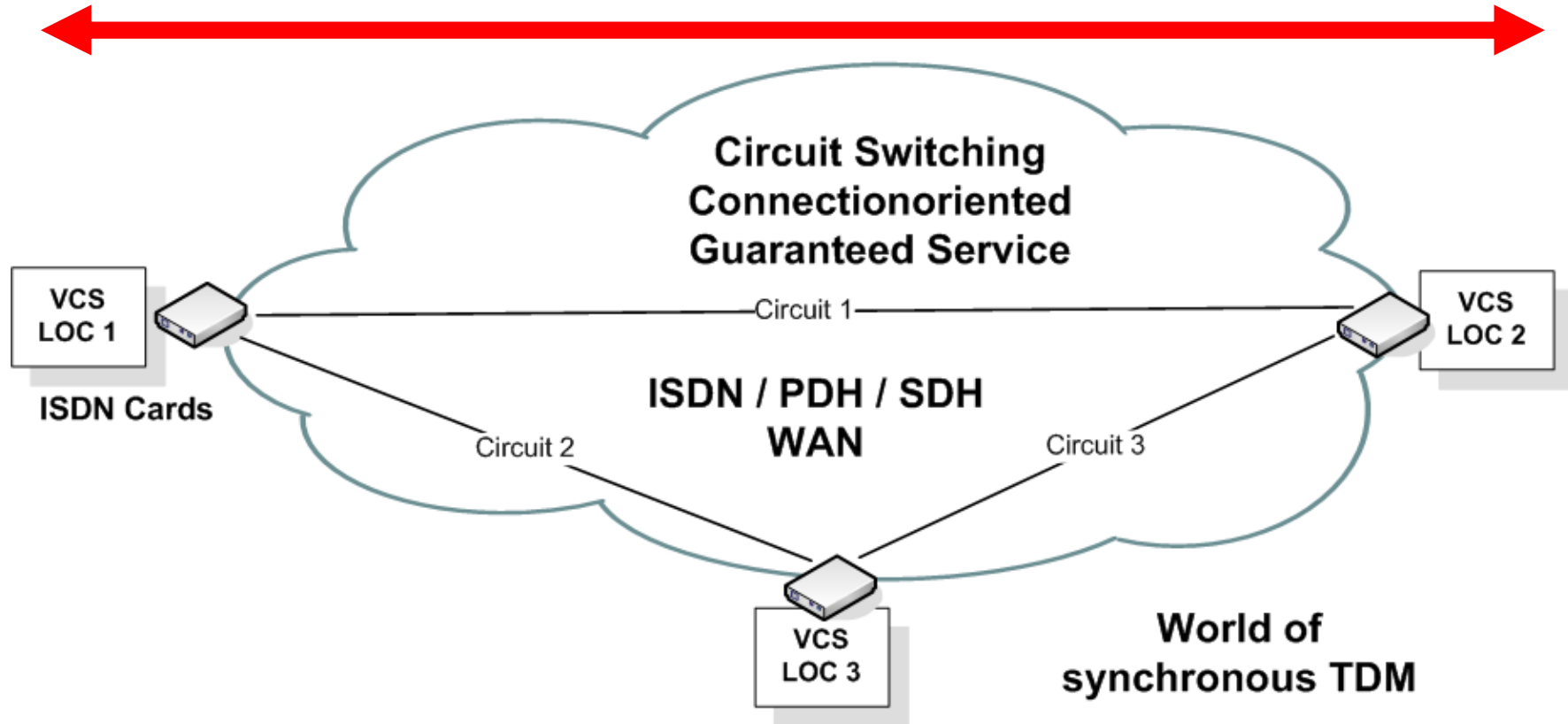


If other channels are silent, one channel can fully utilize his/her access rate  
-> better usage of network bandwidth

Variable delay and variable delay variation (jitter)  
Buffer overflow leads to loss of packets



# Impact On Applications (1a)



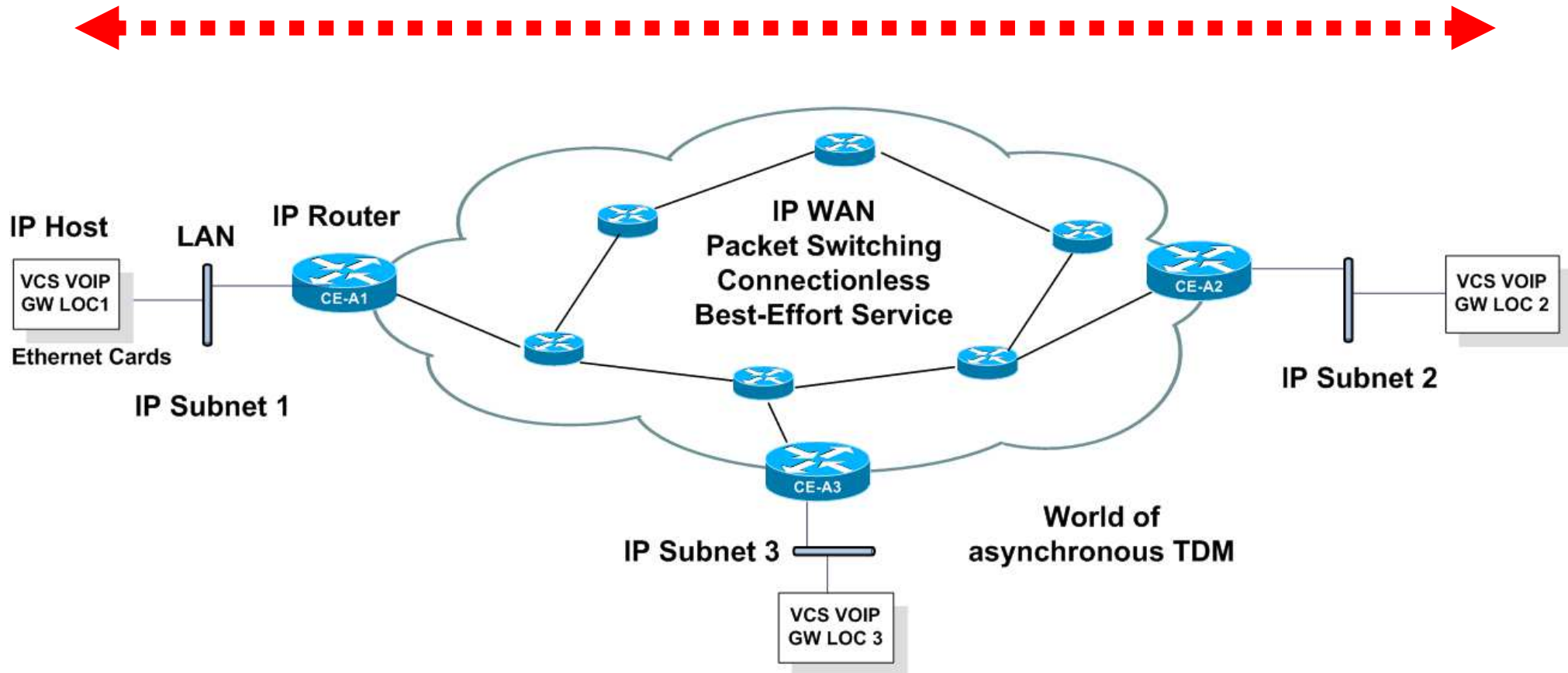
## Deterministic network behavior:

Constant bandwidth

Constant delay / no jitter per communication session

Very low bit rate / no packet (byte) drops

# Impact On Applications (1b)



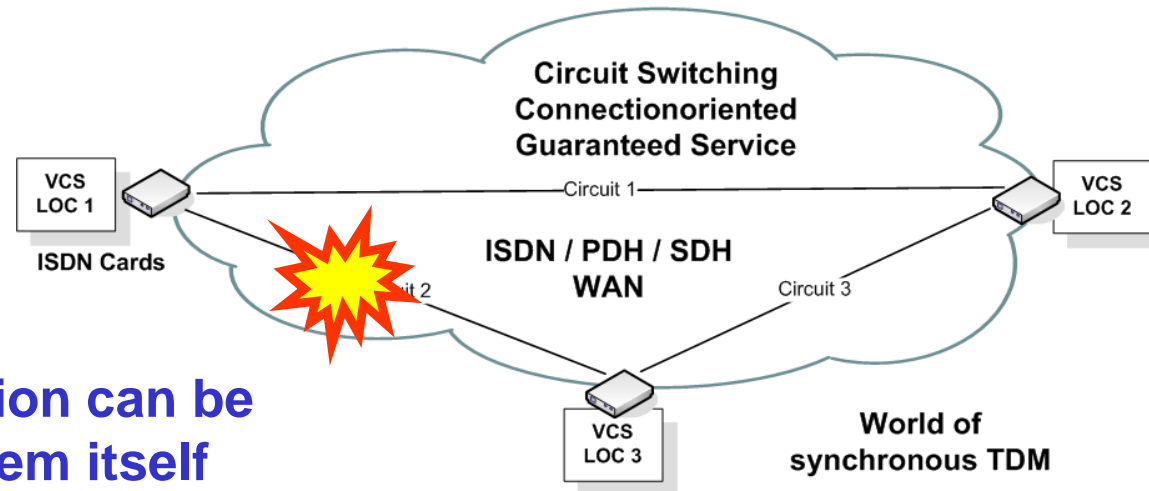
## Non-deterministic network behavior:

Variable bandwidth

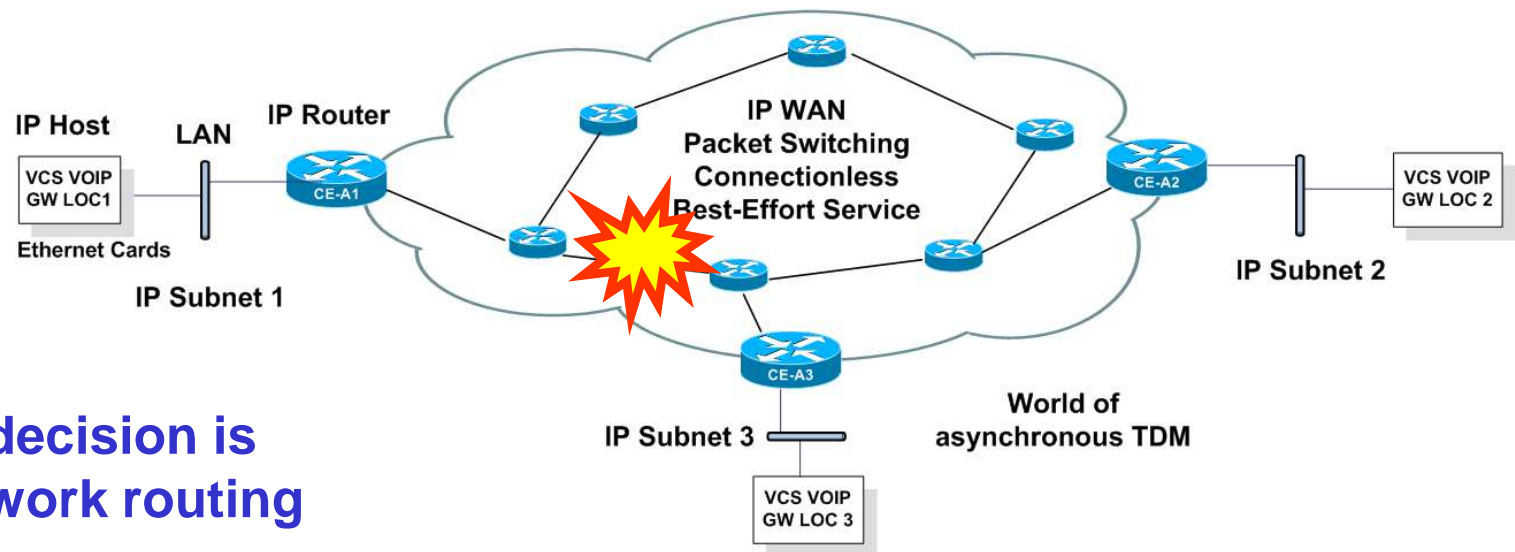
Variable delay / jitter per communication session

Because of best-effort packet loss possible

# Impact On Applications (2)

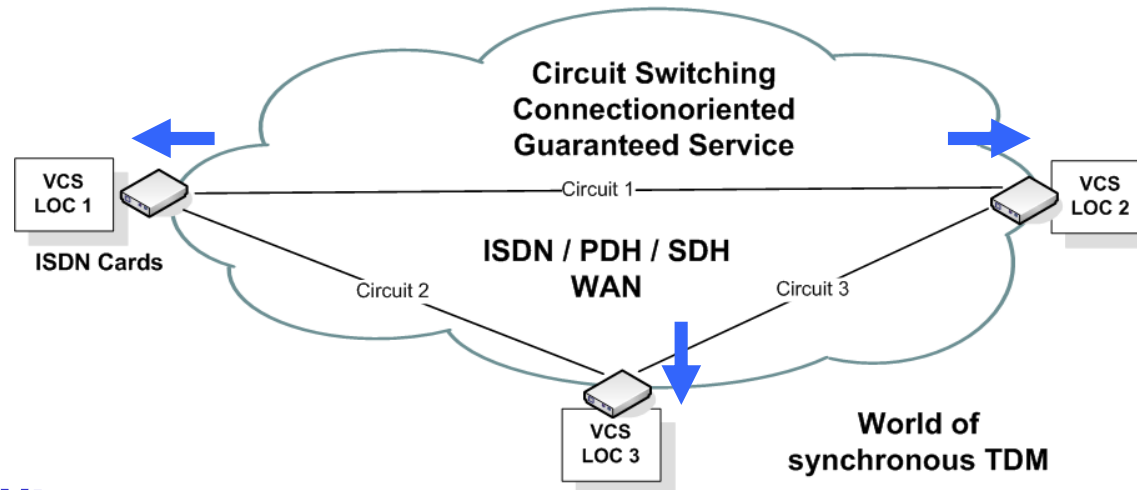


Switchover decision can be done by end-system itself

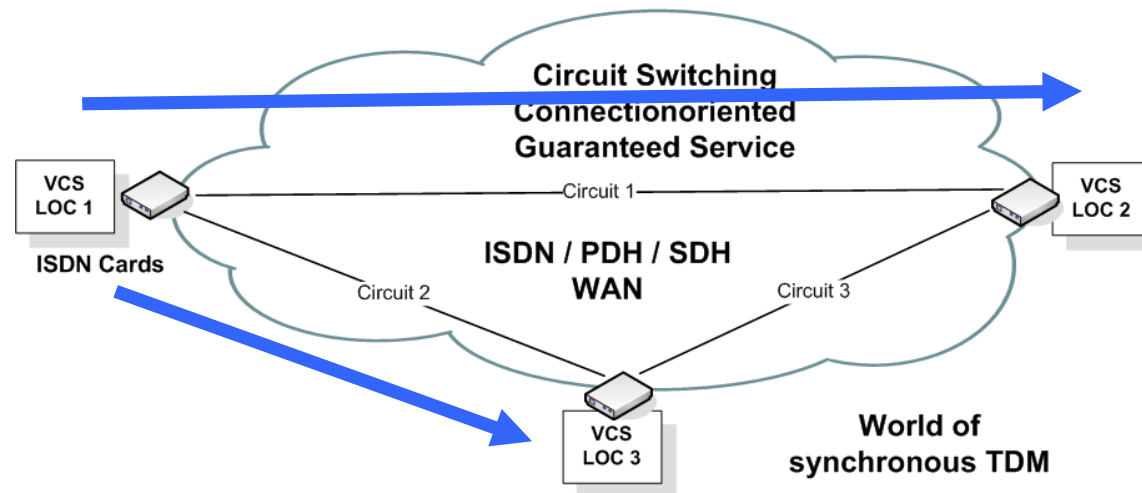


Switchover decision is done by network routing

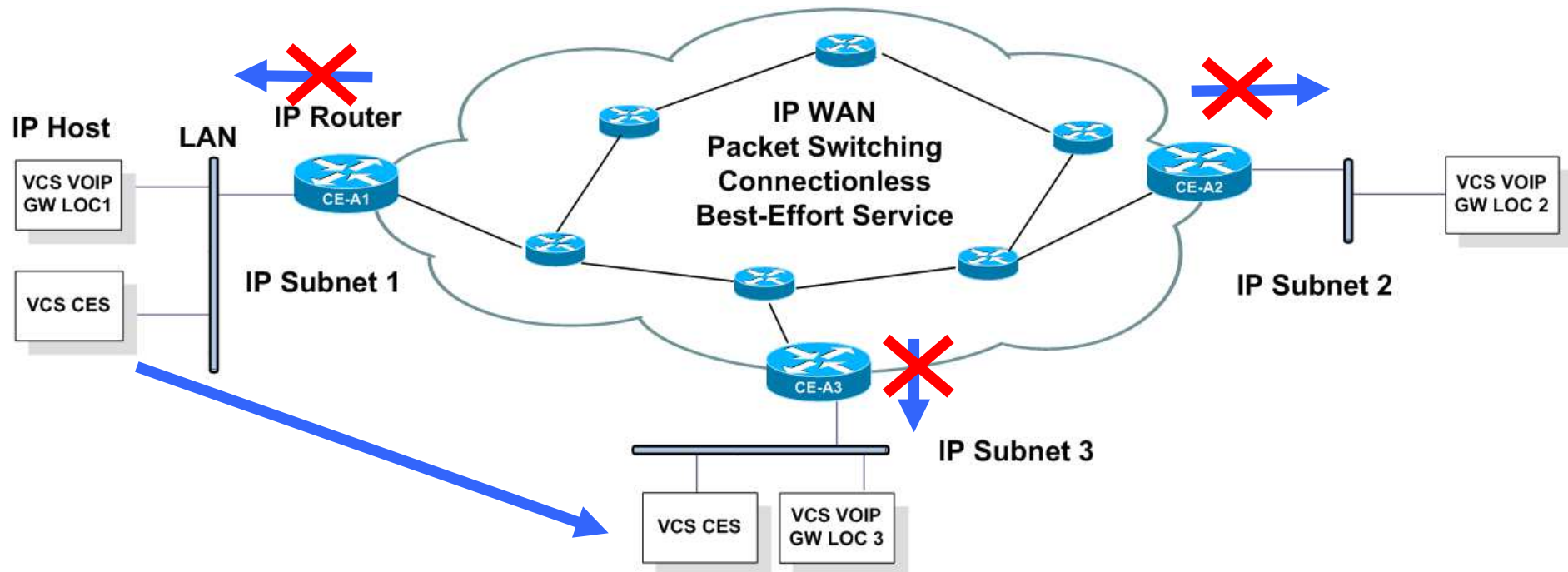
# Impact On Applications (3a)



Clock for telephony:  
Provided by the network or  
passed through the network



# Impact On Applications (3b)



## Clock for telephony?

No provision by network possible because packet switching is inherently asynchronous. Only possible solution by usage of special CES (circuit emulation services) devices (clock pass through)

## How to handle it in an asynchronous world?

Packetizing a sequence of PCM voice samples in one packet (transmitter).  
Replay buffer for jitter compensation (receiver).  
Both introduces additional delay.

- **Communication functionality and requirements of systems and applications**
  - Architectural model of overall systems
  - Transmission parameters
    - Delay, jitter, loss, guaranteed throughput. application timeouts
  
- **Communication behavior of systems and applications**
  - Who talks to whom, in which style and how much?
    - Communication matrix
    - Average bitrate / bandwidth,
    - Burstiness (duration and amount of bursts)
    - Style unicast (point-to-point or one-to-one, bidirectional or unidirectional)
    - Style multicast (point-to-multipoint or one to many, unidirectional only)
  
- **Network operational model**
  - Is network infrastructure operated by single authority?
  - Are network service provider involved?
  - If yes at what level?
    - OSI Layer 1, 2 or 3

- **High Availability (HA)**

- How to continue communication in case of failures or during time of planned service intervals by automatic switchover techniques?
  - Note: 99,99% means 52,56 minutes/year, 4,32 minutes/month, 1,01 minutes/week

- **QoS (Quality of Service)**

- How to achieve (some kind) of guarantees for mission-critical traffic over a best-effort based technology like IP?

- **Security**

- How to separate traffic of different domains (customers) ?
  - Base VPN (Virtual Private Network)
- How to protect traffic and systems against attacks?
  - Advanced VPN techniques (protection based on crypto-graphical methods)
  - Firewall techniques

- **Management**

- How to manage all that?
  - Organizational aspects
  - Technical aspects

- **Identification of distributed processes**
  - IP addresses, TCP/UDP numbers
  - Optionally usage of DNS (Domain Name System)
    - Translates symbolic names to IP addresses
  - Avoid NAT (Network Address Translation)
    - If it can not be avoided a NAT concept is needed
    - Bad design!
- **Connectivity**
  - Provided by IP routers / routing tables
  - IP routing establishes signposts for all networks to be reached
  - Avoid policy routing
    - Local decision only, does not scale
- **IP address design and IP routing concept**
  - Has to be agreed in early phase of a project



- **Network Operation Model**
  - Network infrastructure operated by single authority or involvement of service provider(s)
  - Service provider types: L1 VPN, L2 VPN, or L3 VPN
- **Management**
  - Provisioning, monitoring, alarming
  - Operation, maintenance
  - If QoS or security is involved it becomes much more complicated
- **Clarify operational model to be used and management aspects**
  - Has to be agreed in the early phase of a project

- **High Availability (HA)**

- Redundancy
- Selection of automatic switchover mechanisms
  - Rerouting to an alternate path
  - Golden-rule: The less the better
- Convergence time tuning
- HA concept

- **QoS**

- Traffic marking, traffic classifying, traffic queuing
- Traffic policing, traffic shaping
- QoS concept
  - For clarification about QoS consumer and QoS provider borders and SLAs
  - For QoS monitoring and management

- **Multicast**

- Group address plan
- Multicast routing concept
- Multicast convergence tuning

- **Security**

- Security assessment
  - Identifying of environment and threats
  - Identifying security domains / zones
- Optional risk analysis
- Security concept
  - Security domains,
  - Security responsibilities
  - Security management
- Only if security concepts is agreed
  - Identifying location of perimeter and tunnel mechanism and selection of security technology are possible

# First Summary

---

- **Holistic look to the basic and advanced topics is absolutely necessary**
  - All topics must fit together
  - Tradeoffs will be seen and compromises have to be agreed
  - Design will not emerge in straight-forward way
  - Fact-finding missions and feedback loops will be necessary

# Agenda

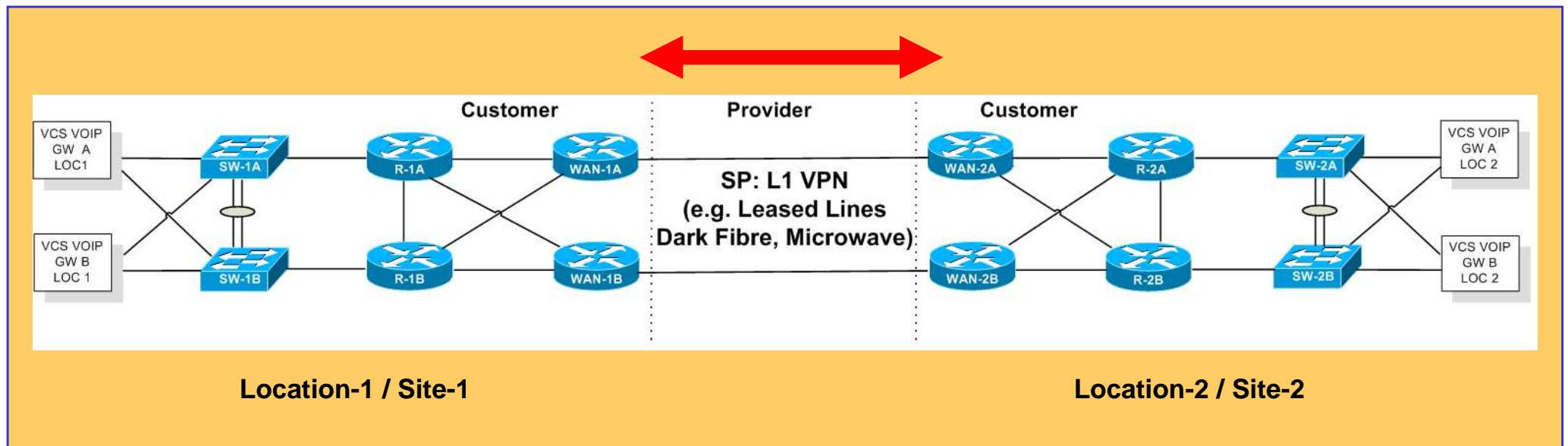
---

- **Introduction**
- **Network Operational Model**
- **High Availability**
- **QoS**
- **VPN Technology**
- **Multicasting**
- **Summary**

# Network Operational Model M1: L1-VPN



Service provider links:  
Constant bandwidth / constant delay / no jitter



IP Router / L3 Switch

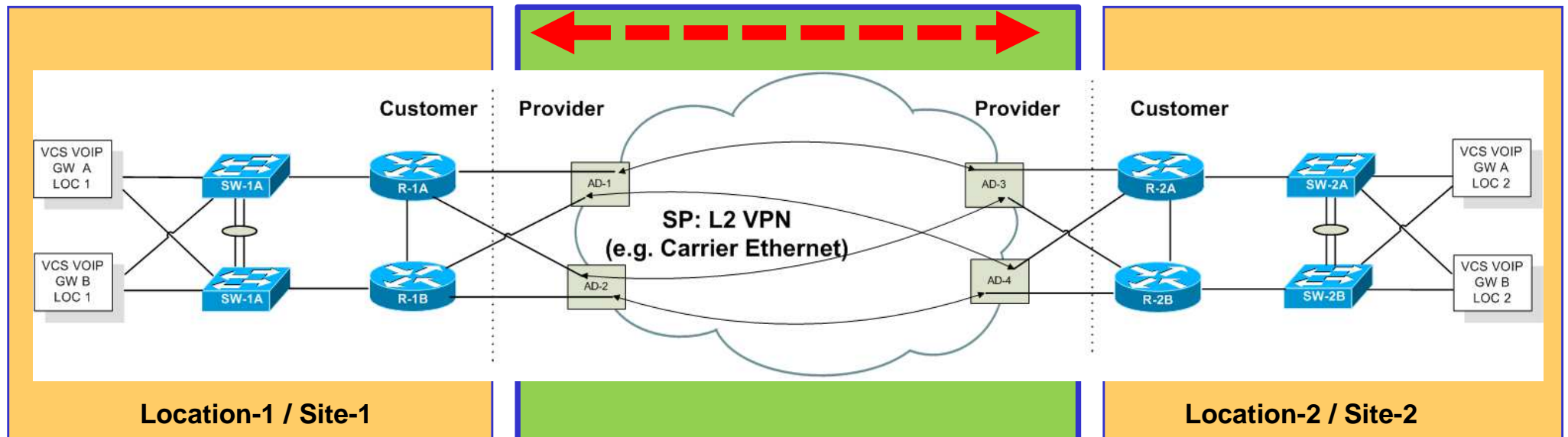


Ethernet Switch / L2 Switch

# Network Operational Model M2: L2-VPN



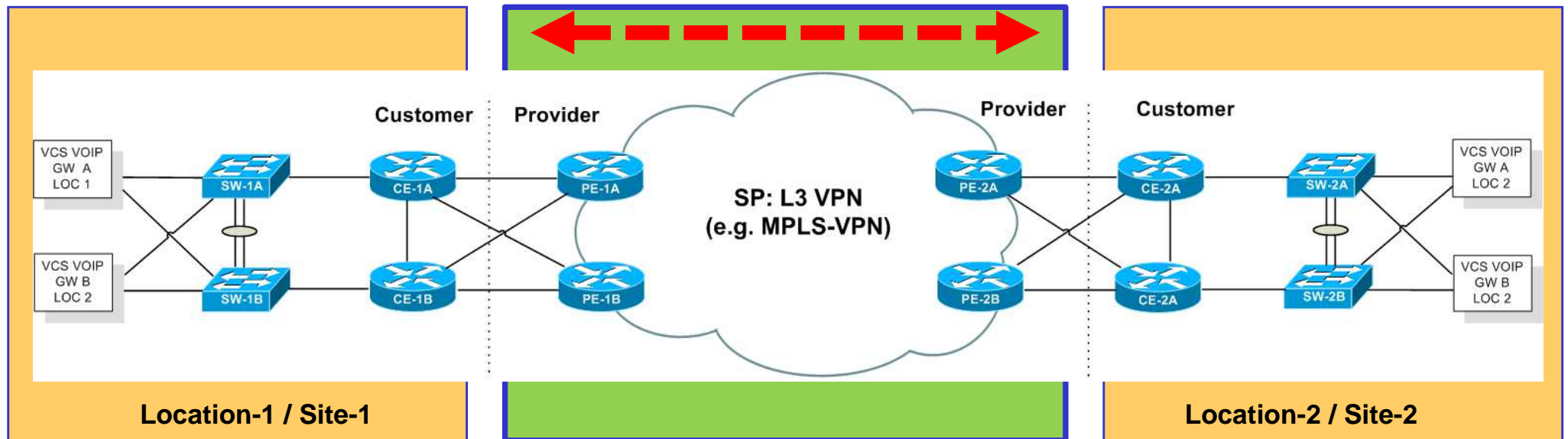
Service provider links:  
Variable bandwidth / variable delay / jitter



# Network Operational Model M3: L3-VPN



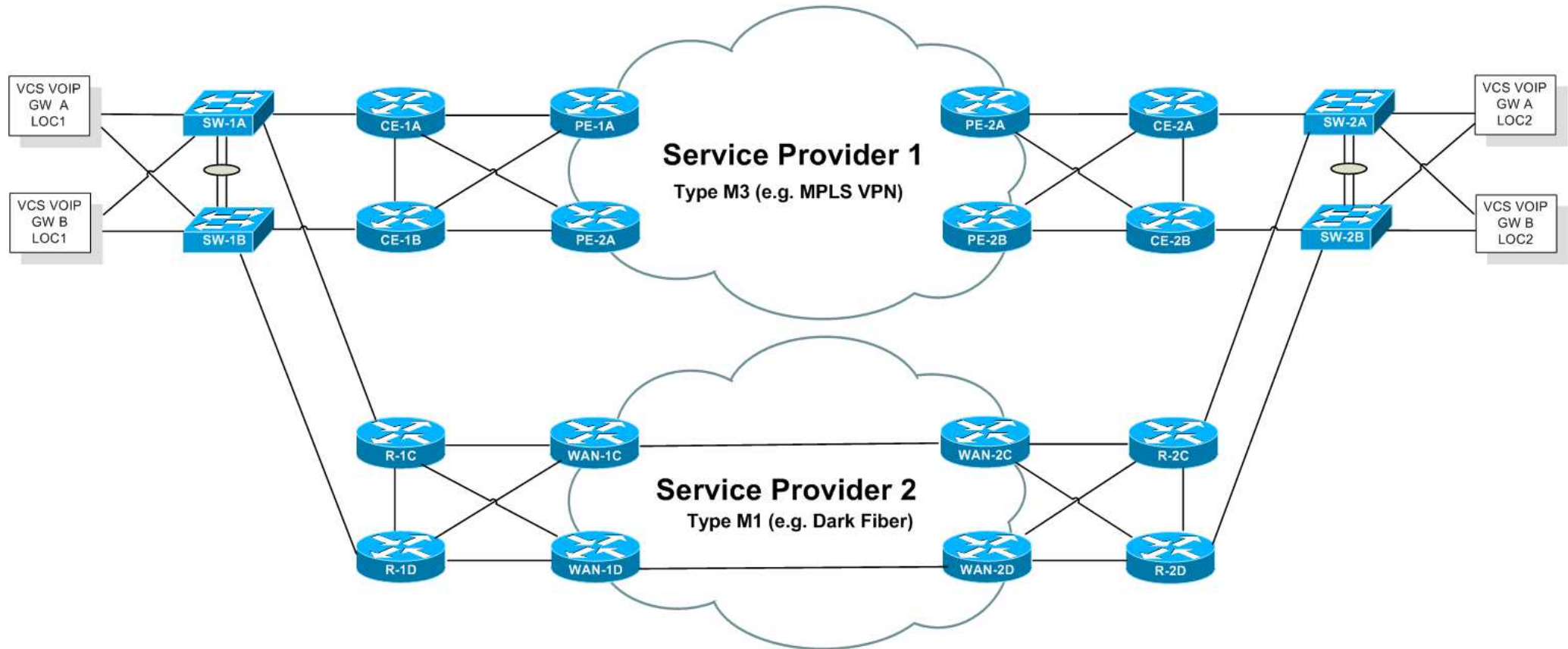
Service provider links:  
Variable bandwidth / variable delay / jitter



CE ... Customer Edge  
PE ... Provider Edge  
SP ... Service Provider



# Example: Dual Network Service Providers



# Agenda

---

- **Introduction**
- **Network Operational Model**
- **High Availability**
  - Elements of HA
  - Functional Access Block Types for HA
  - Routing Aspects
- **QoS**
- **VPN Technology**
- **Multicasting**
- **Summary**

# Elements For High Availability 1

- **Restore from backup**
  - Reconstruction of repaired or changed components
- **Redundancy**
  - Alternate paths / components in order to switchover in case of failure or to be used for load balancing
- **Automatic rerouting**
  - Usage of dynamic routing techniques found on different OSI layers (1, 2, 3 and 7)
- **Convergence Time**
  - Time to detect and to react locally
  - Time to propagate the event to other components
  - Time until all other components have recognized and reacted and a consistent state is reached again

# Elements For High Availability 2

- **Examples of rerouting techniques**

- L2 LACP, Linux-Bonding, Intel-Teaming
- L2 Rapid Spanning Tree
- L2 BFD Bidirectional Forwarding Detection
- L3 dynamic IP routing protocols (OSPF, IS-IS, BGP, MPLS-LDP),
- L3 First-hop routing (HSRP/VRRP)
- L3 Equal Cost Multiple Path (ECMP)

- **Dynamic rerouting techniques**

- Tuning necessary to achieve (sub)second convergence time
- The less different techniques used the better
- Needs to be harmonized
- Failure repair may also lead to interruption until convergence

- **Traditional IP mechanism**

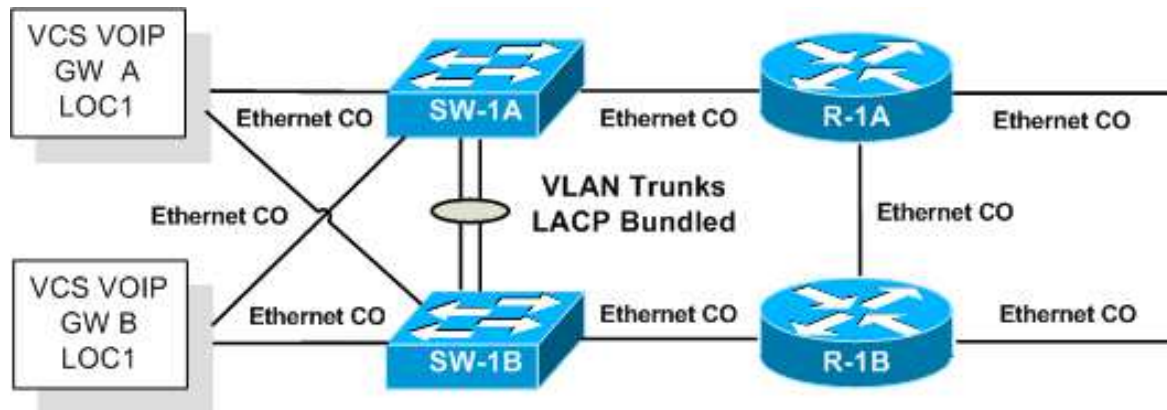
- Good at “Black-Outs” (e.g. link down, router down)
- Bad at “Brown-Outs” (e.g. packet loss increases)

# Agenda

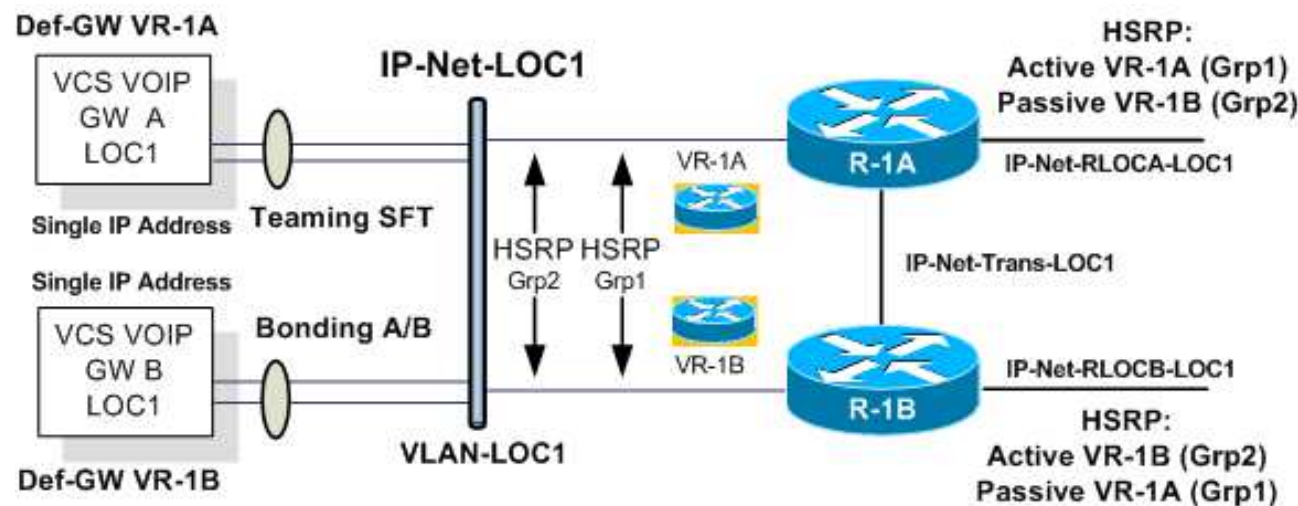
---

- **Introduction**
- **Network Operational Model**
- **High Availability**
  - Elements of HA
  - Functional Access Block Types for HA
  - Routing Aspects
- **QoS**
- **VPN Technology**
- **Multicasting**
- **Summary**

# HA Functional Access Block Type 1



Access Network Type 1: Physical Topology



Access Network Type 1: VLAN / IP Topology

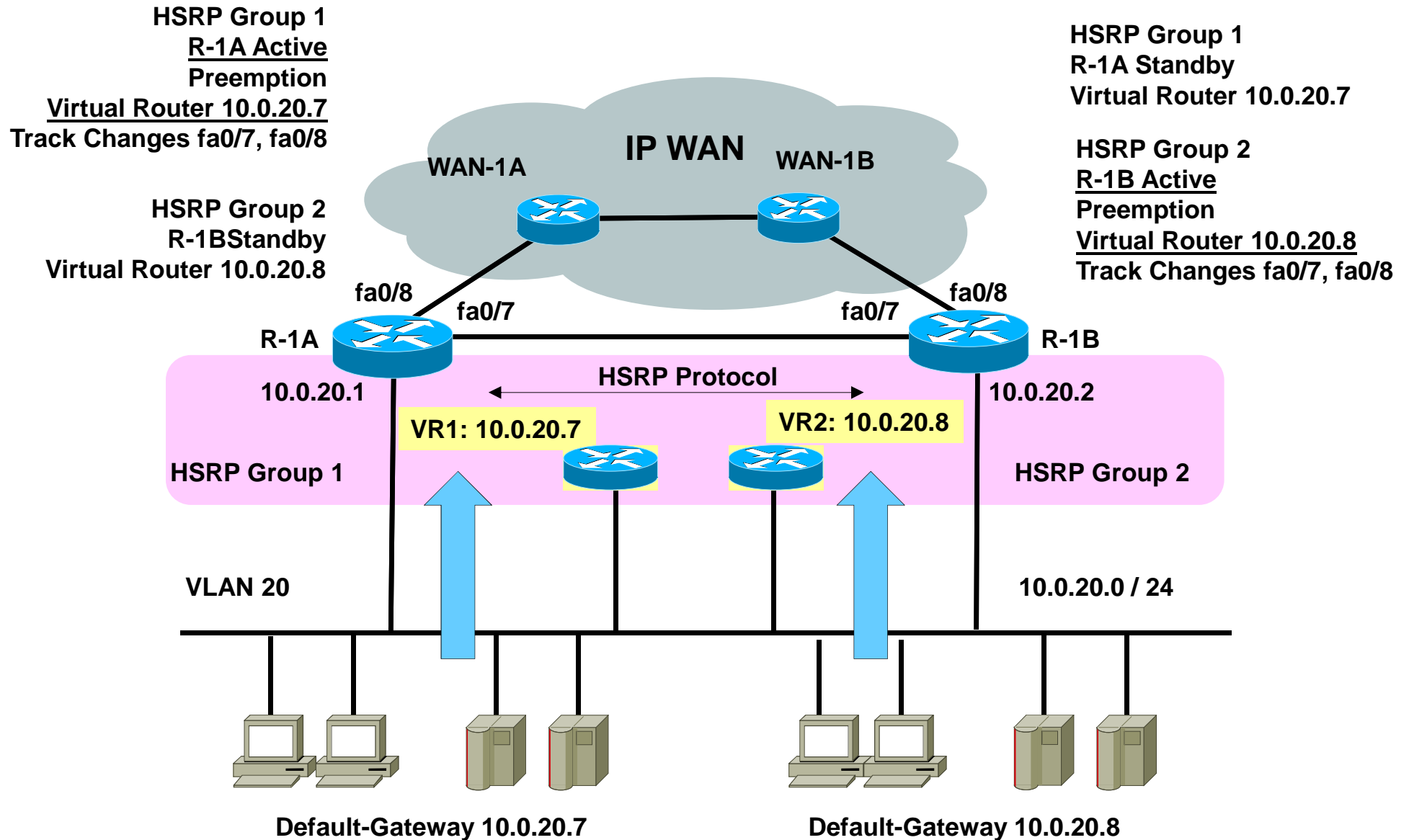
## Physical Topology

- Redundant VOIP interfaces bundled (multi-homed) by techniques like
  - Intel Teaming (Switch Fault Tolerance -> SFT)
  - Linux Bonding (Active-Backup)
- Redundant Ethernet switches
  - Trunks grouped by LACP
- Redundant routers
- Redundant PSUs
- CO (Copper) Ethernet links only
  - All components housed in one cabinet /rack or room (100m limit for cables)

## IP Topology

- HSRP
  - Create one virtual router for the IP hosts used as default gateway
  - Optional two HSRP groups for directing A and B to different routers

# HSRP Example



## HSRP ... Hot Standby Router Protocol

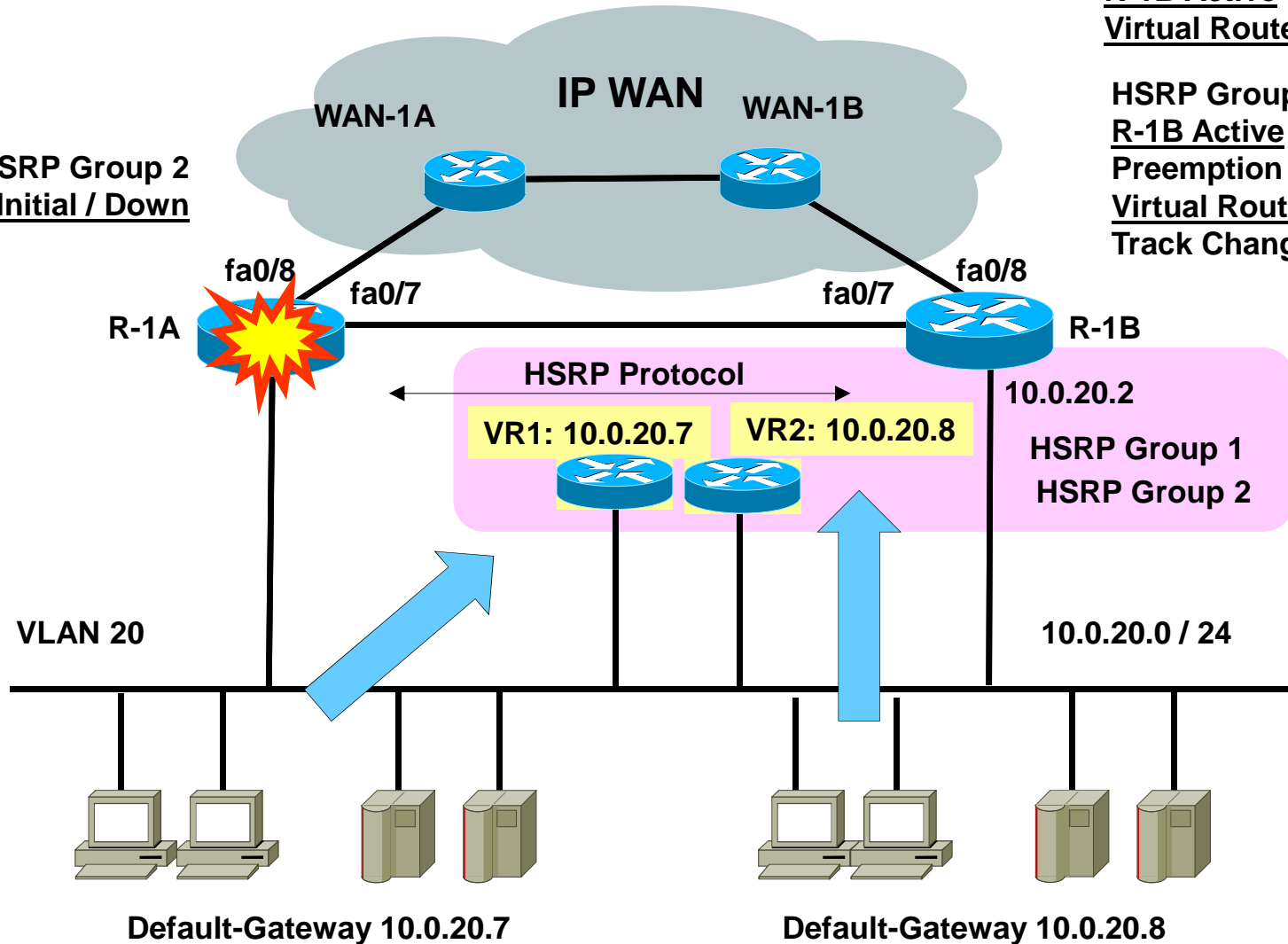
# HSRP Failover 1 (Router R-1A Down)

HSRP Group 1  
R-1A Initial / Down

HSRP Group 2  
R-1B Initial / Down

HSRP Group 1  
R-1B Active  
Virtual Router 10.0.20.7

HSRP Group 2  
R-1B Active  
Preemption  
Virtual Router 10.0.20.8  
Track Changes fa0/7, fa0/8



**HSRP ... Hot Standby Router Protocol**



# HSRP Failover 2

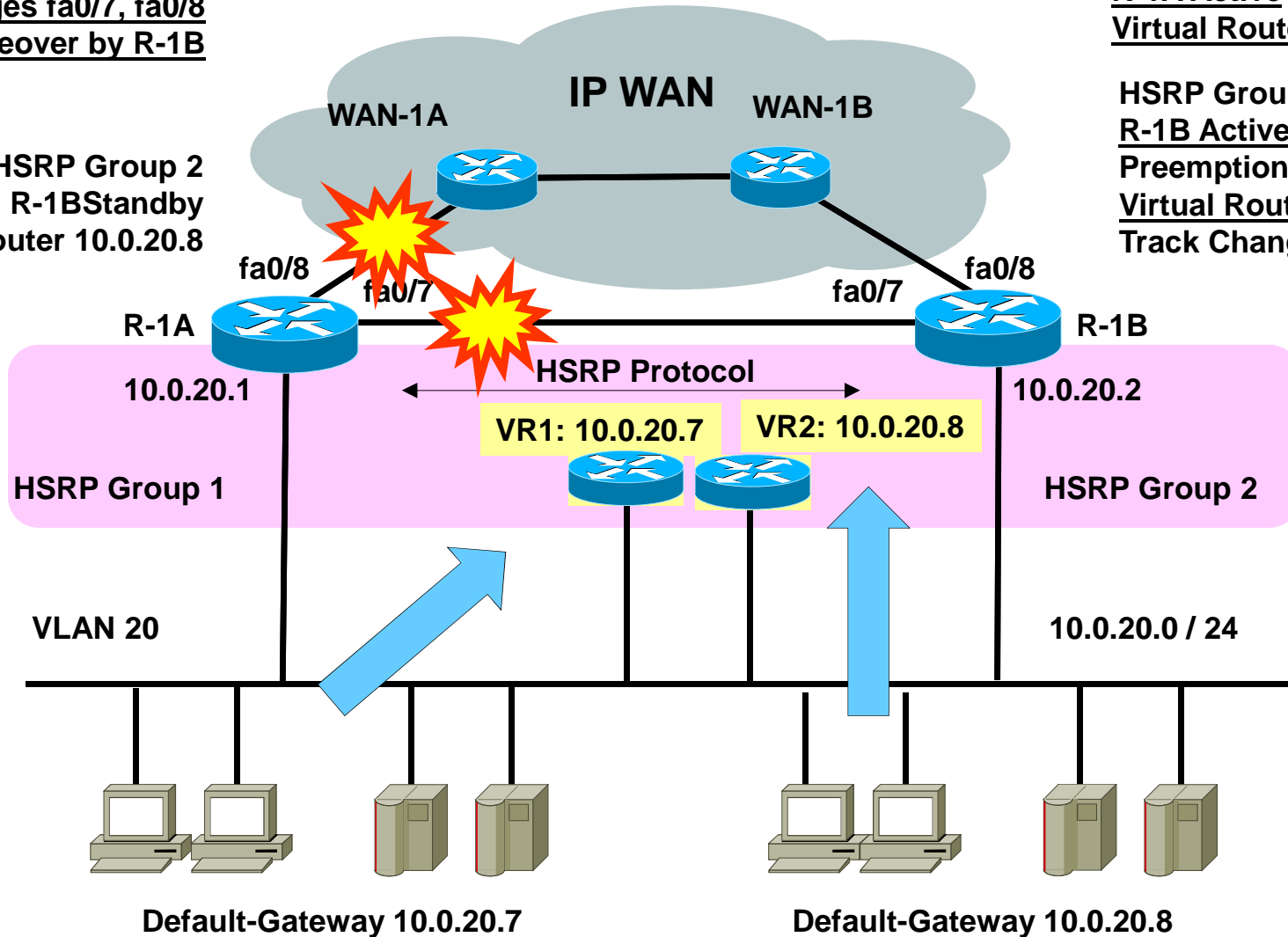
## (Router R-1A All WAN Links Down)

HSRP Group 1  
R-1A Standby  
Track Changes fa0/7, fa0/8  
Causes takeover by R-1B

HSRP Group 1  
R-1A Active  
Virtual Router 10.0.20.7

HSRP Group 2  
R-1B Standby  
Virtual Router 10.0.20.8

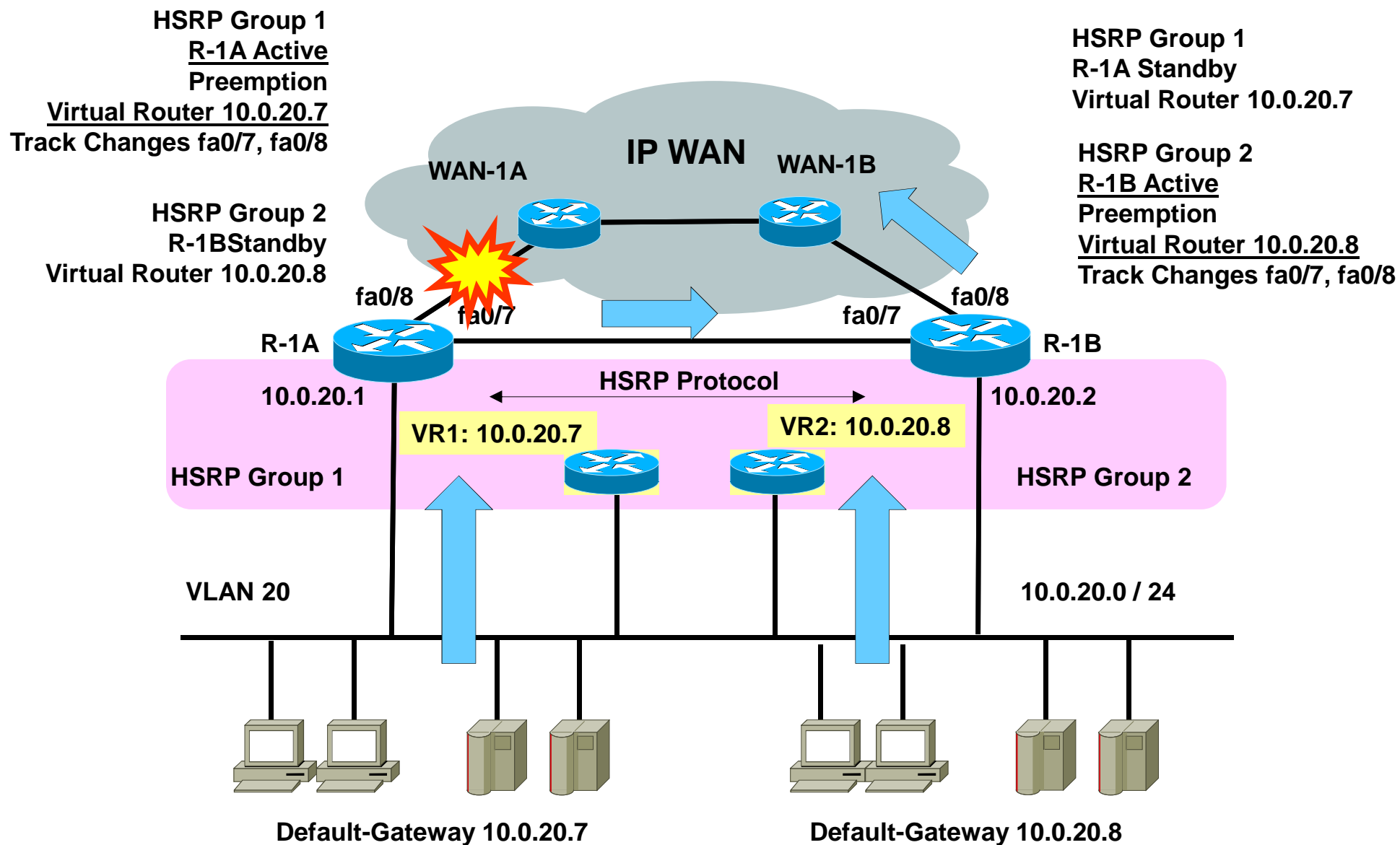
HSRP Group 2  
R-1B Active  
Preemption  
Virtual Router 10.0.20.8  
Track Changes fa0/7, fa0/8



### HSRP ... Hot Standby Router Protocol

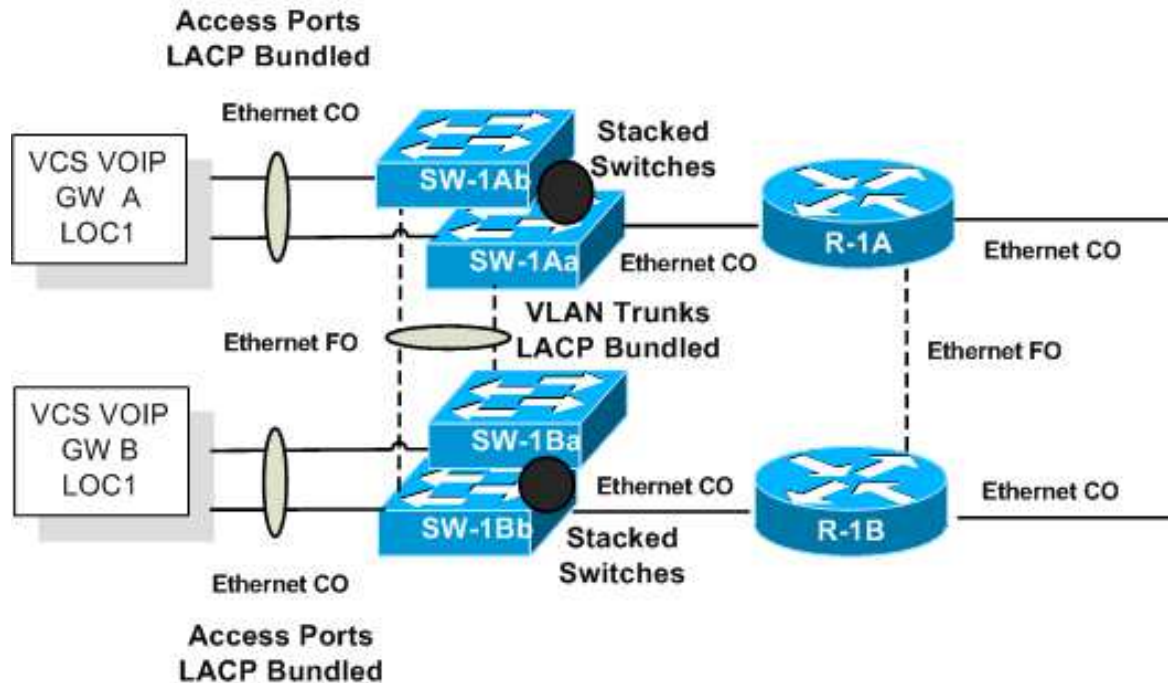
# HSRP Failover 3

## (Router R-1A Single WAN Link Down)



### HSRP ... Hot Standby Router Protocol

# HA Functional Access Block Type 2

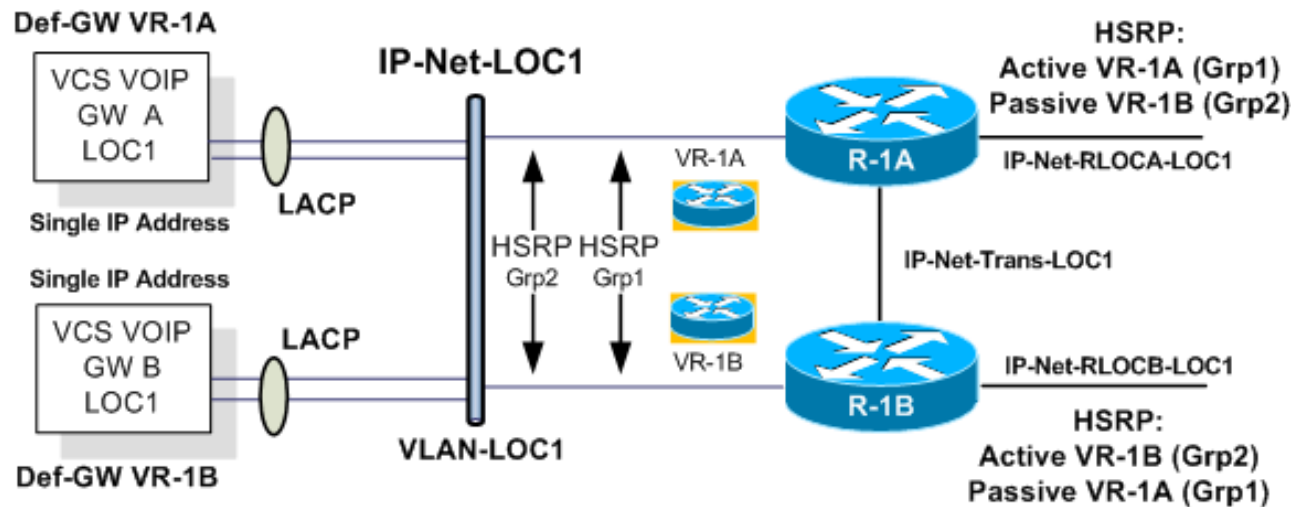


Access Network Type 2: Physical Topology

## Physical Topology

- Redundant VOIP interfaces bundled (multi-homed) by LACP to different physical members of a stacked Ethernet switch
- Redundant Ethernet switches
  - Trunks grouped by LACP
- Redundant routers to IP WAN
- Redundant PSUs
- Copper Ethernet links within a cabinet / rack
- Fiber optic Ethernet links between rooms or buildings (if cable distance is larger as 100m)

# HA Functional Access Block Type 2 (cont.)



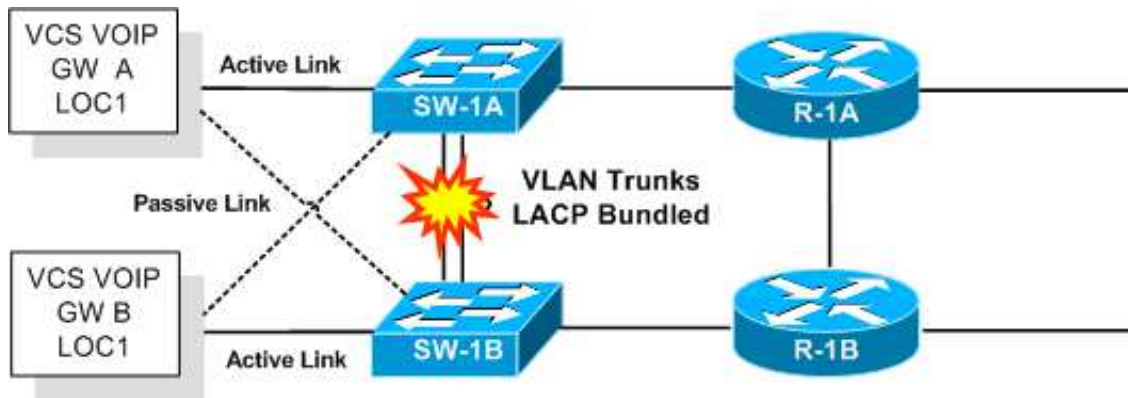
Access Network Type 2: VLAN / IP Topology

## IP Topology

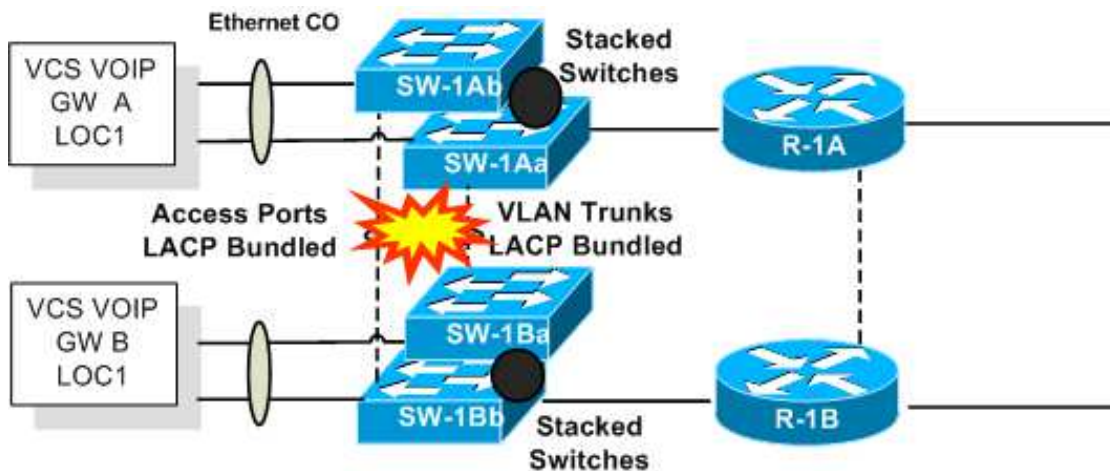
- LACP instead Teaming / Bonding
- Other elements are same as HA type 1

# HA Type 1 / Type 2 Problem

1



## Worst Case Scenarios: Split L2 Connectivity (Type 1)



## Worst Case Scenarios: Split L2 Connectivity (Type 2)

### Dual point of failures:

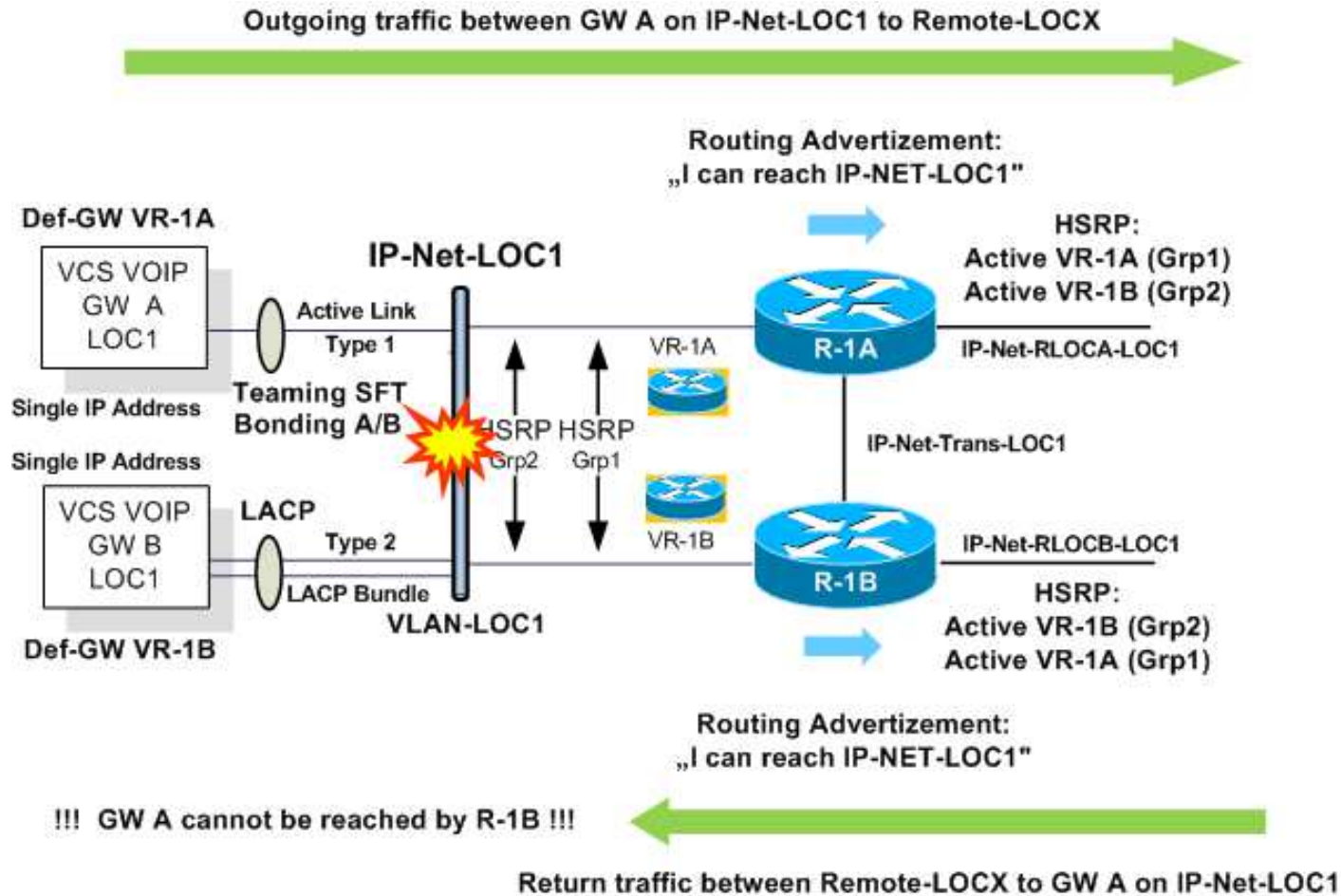
- Both VLAN trunks are broken
- Switch runs amok concerning VLAN trunk or LACP

### Result:

- Split Ethernet connectivity

# HA Type 1 / Type 2 Problem

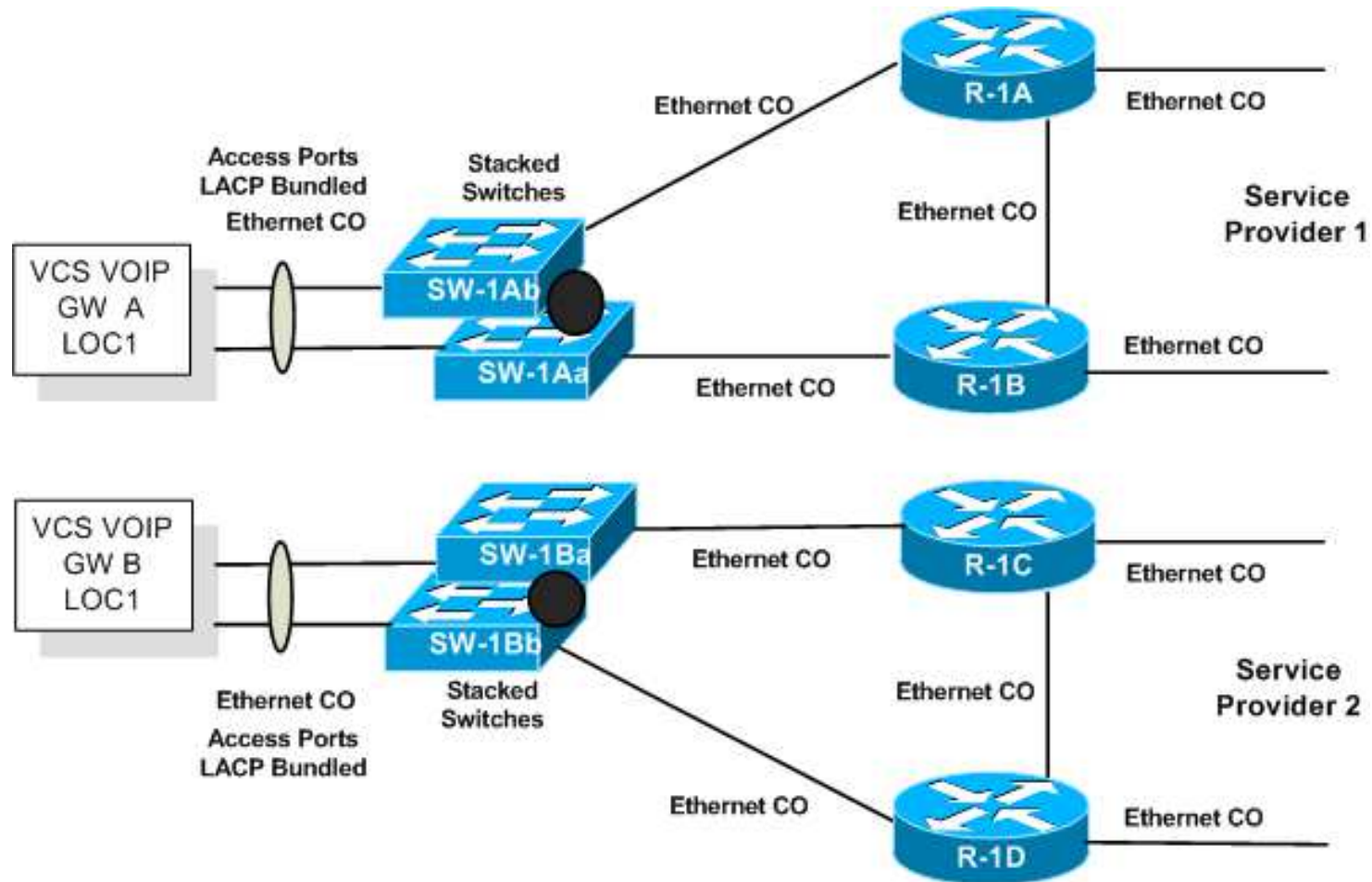
2



## Worst Case Scenario: Splitted IP-Net-LOC1 (Type 1)

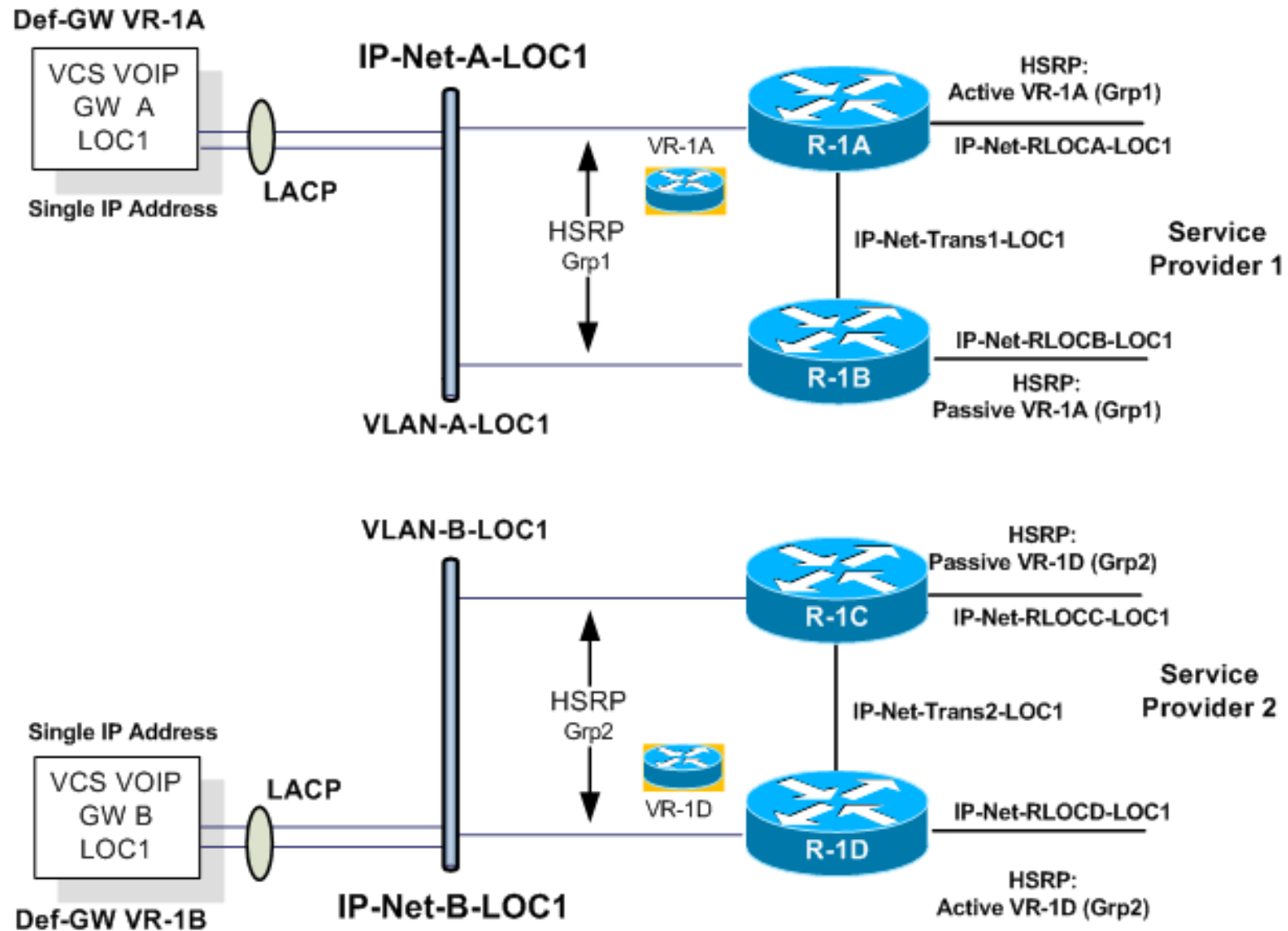


# HA Functional Access Block Type 3



**Access Network Type 3: Physical Topology Dual Provider**

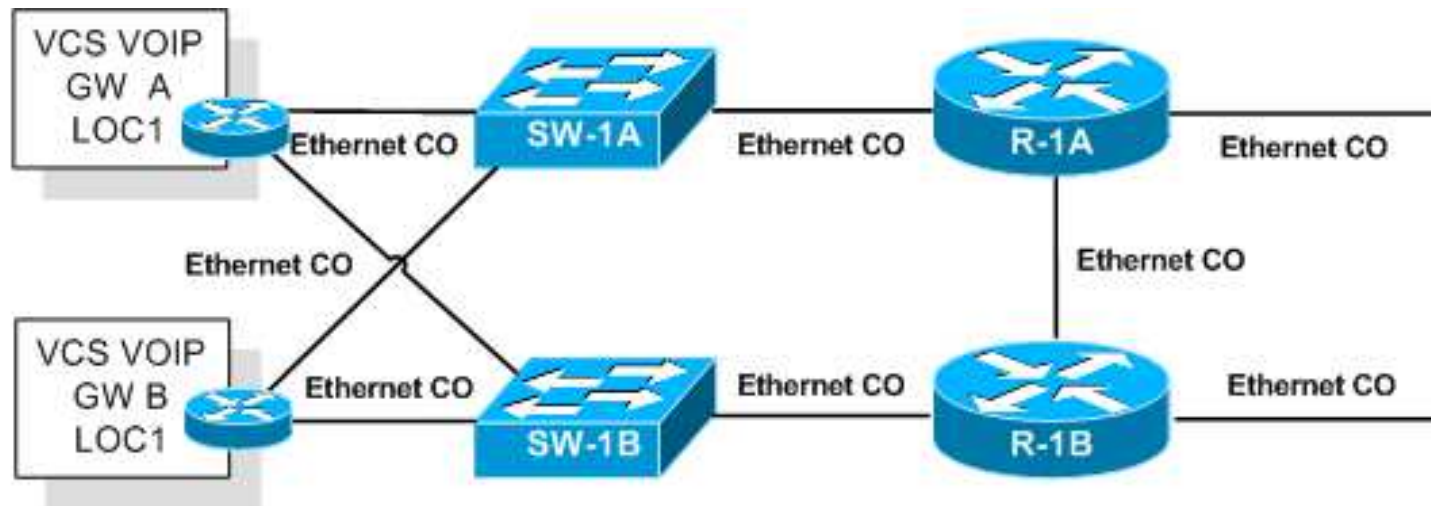
# HA Functional Access Block Type 3 (cont.)



Access Network Type 3: VLAN / IP Topology Dual Provider

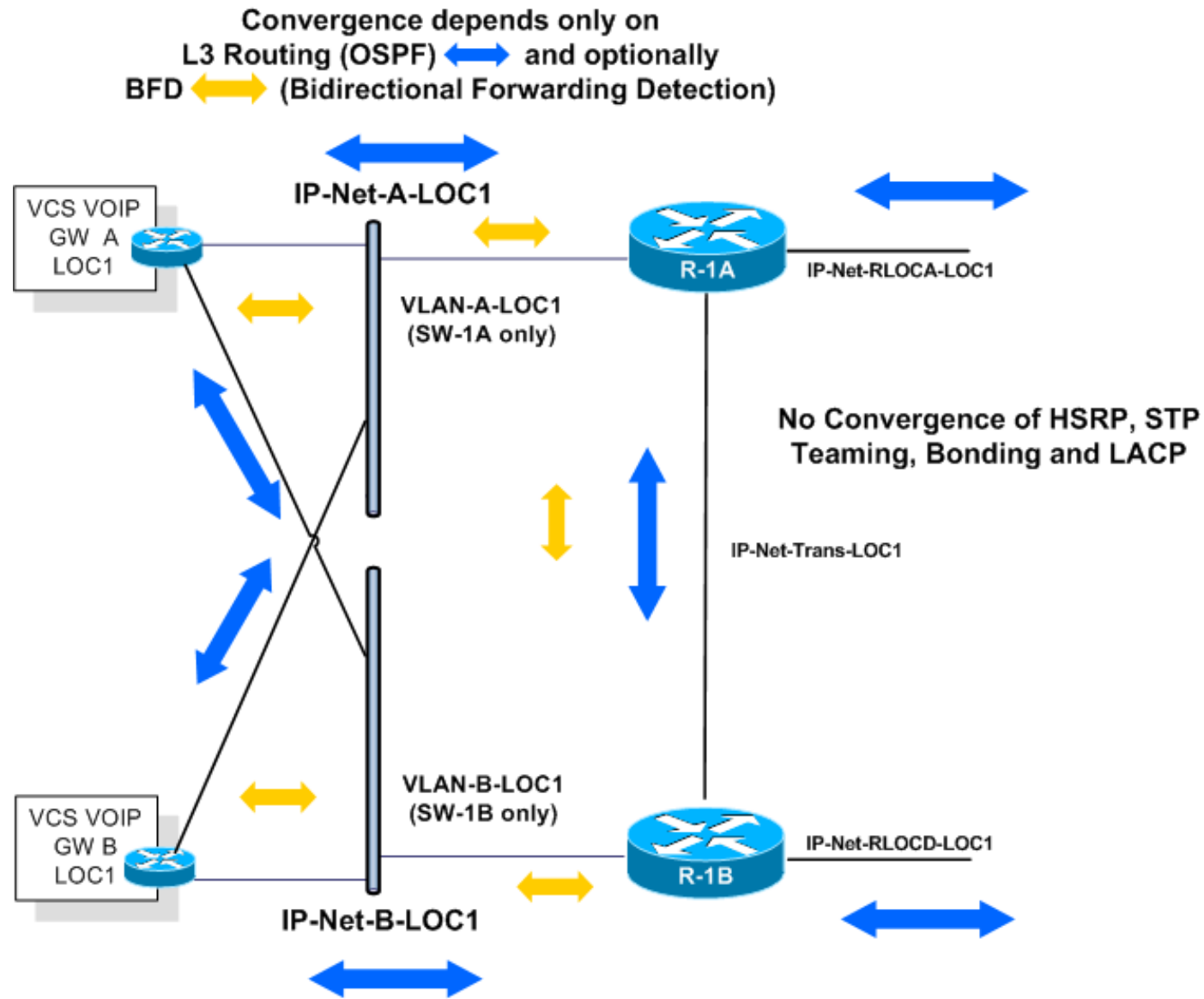


# HA Functional Access Block Type 4



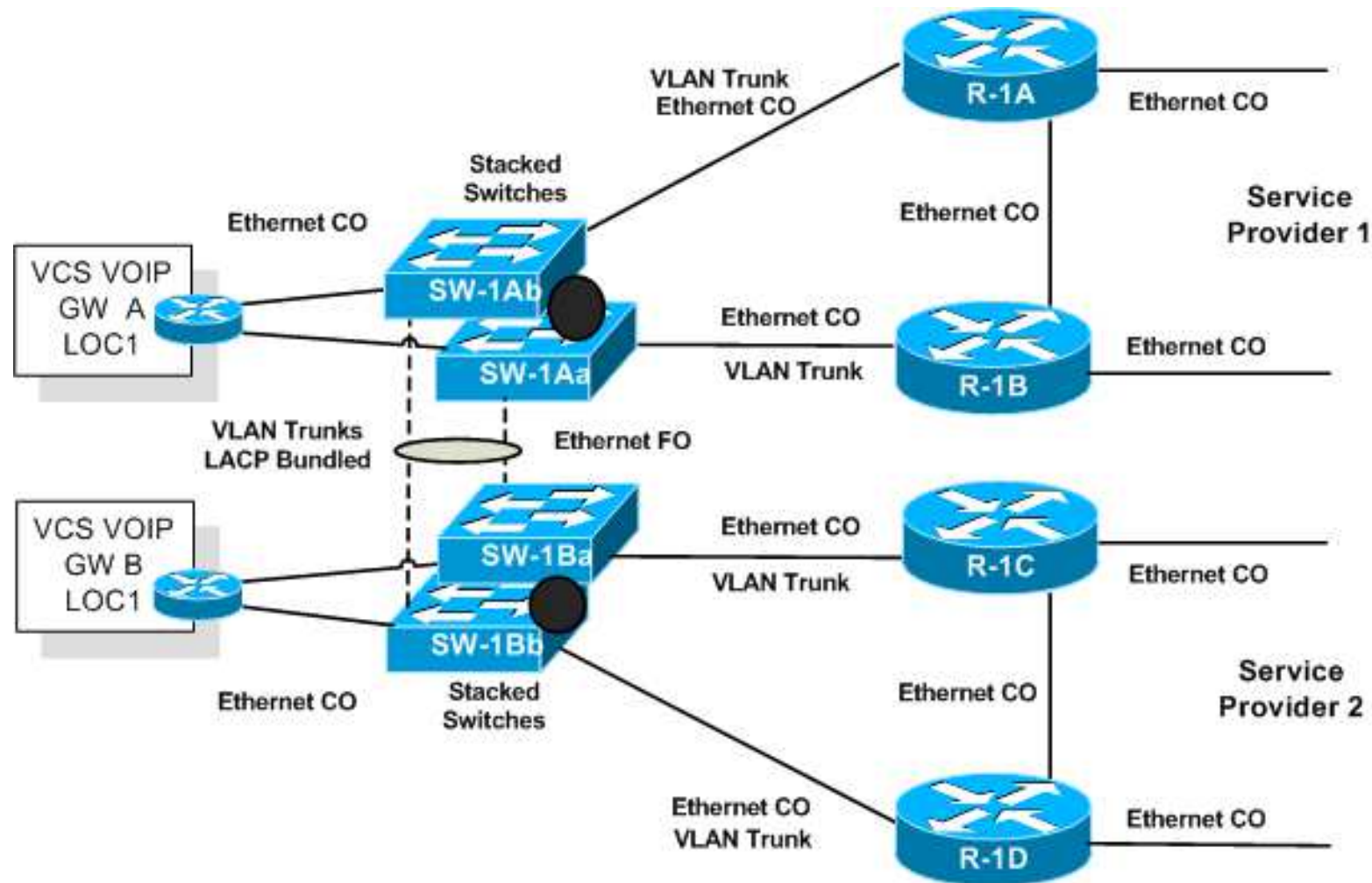
**Access Network Type 4 (Internal Router): Physical Topology**

# HA Functional Access Block Type 4 (cont.)



Access Network Type 4: VLAN / IP Topology

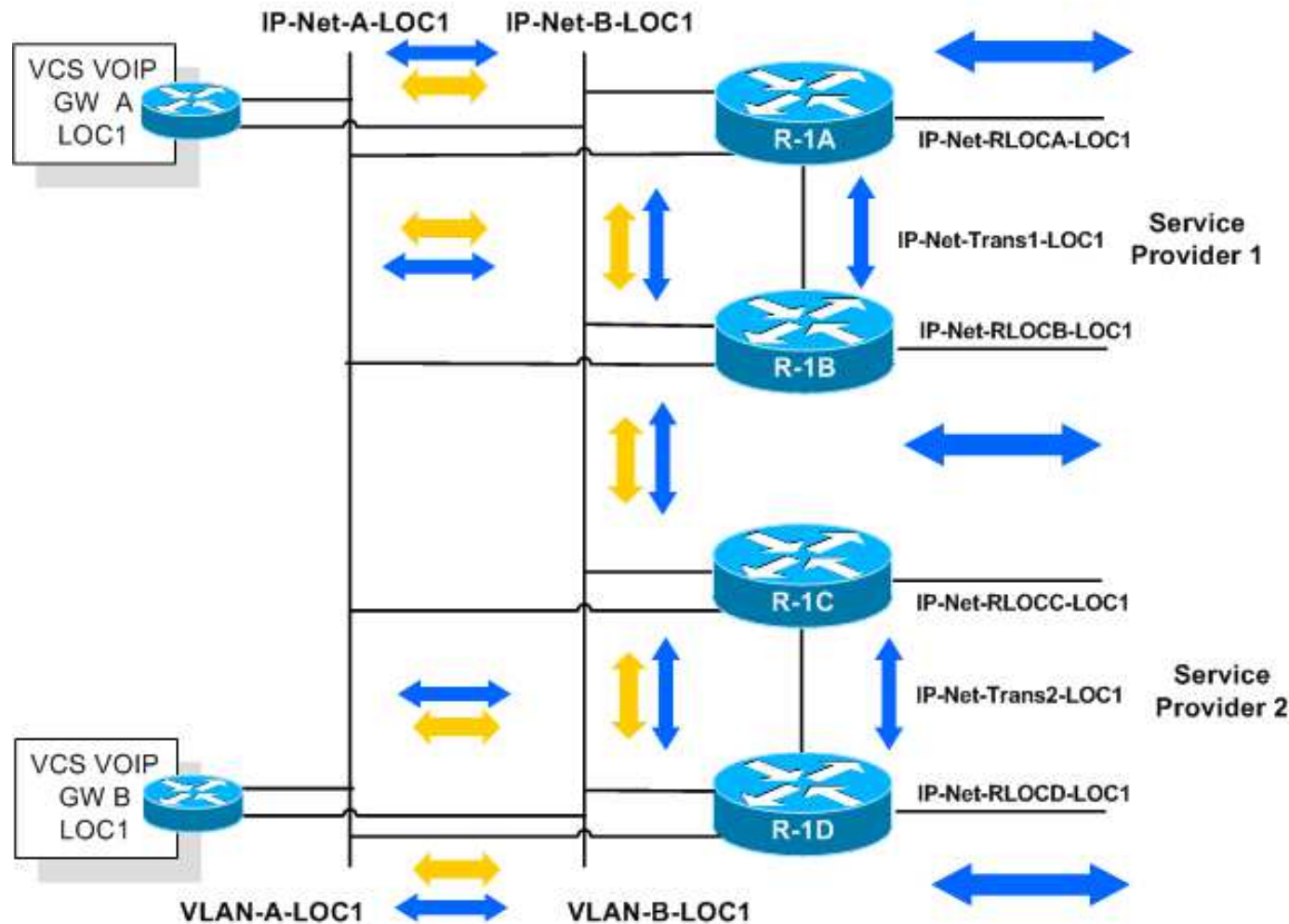
# HA Functional Access Block Type 5



**Access Network Type 5: Physical Topology Dual Provider**

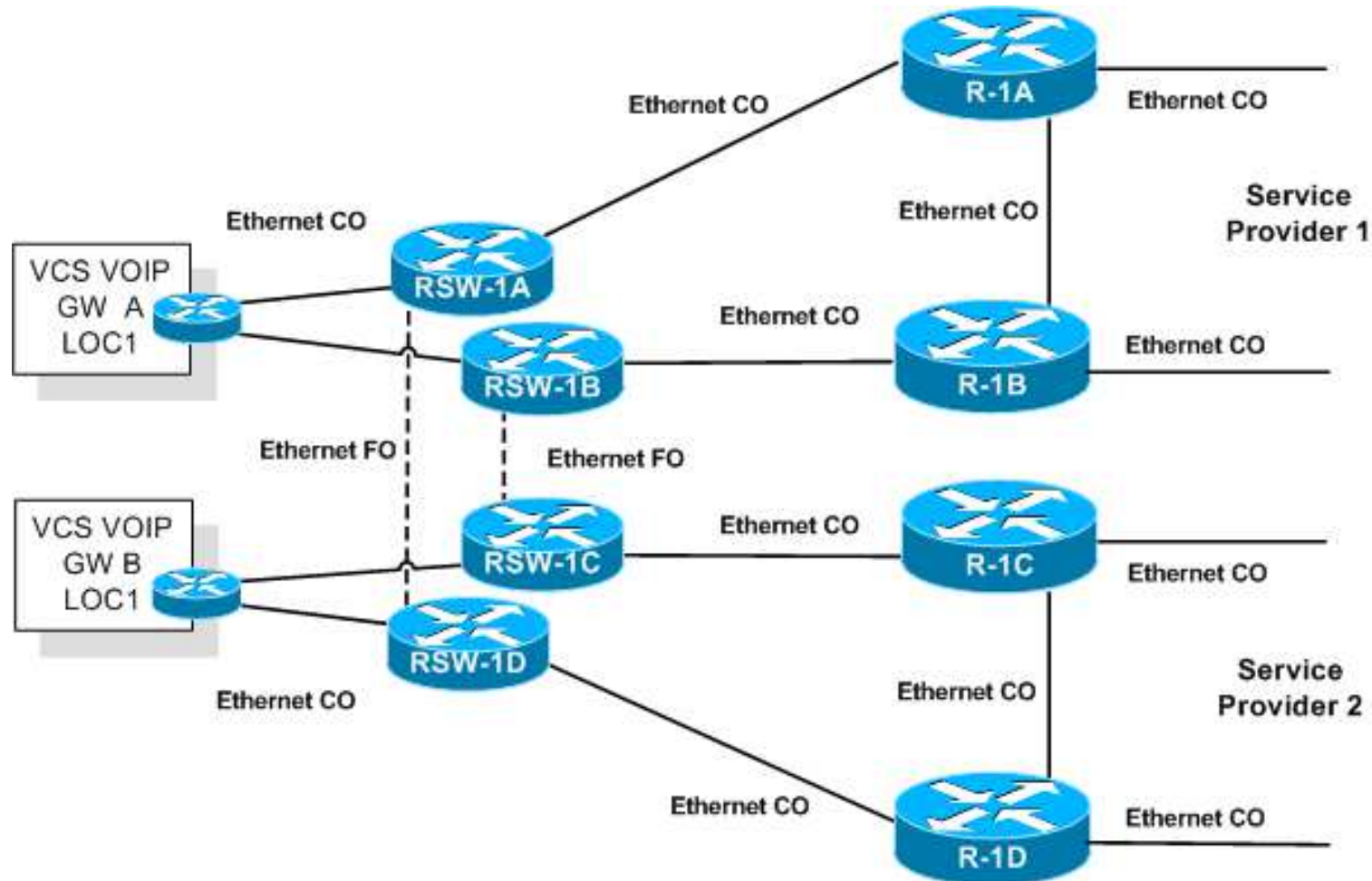
# HA Functional Access Block Type 5 (cont.)

Convergence depends only on  
L3 Routing (OSPF) ↔ and optionally  
BFD ↔ (Bidirectional Forwarding Detection)



Access Network Type 5: VLAN / IP Topology Dual Provider

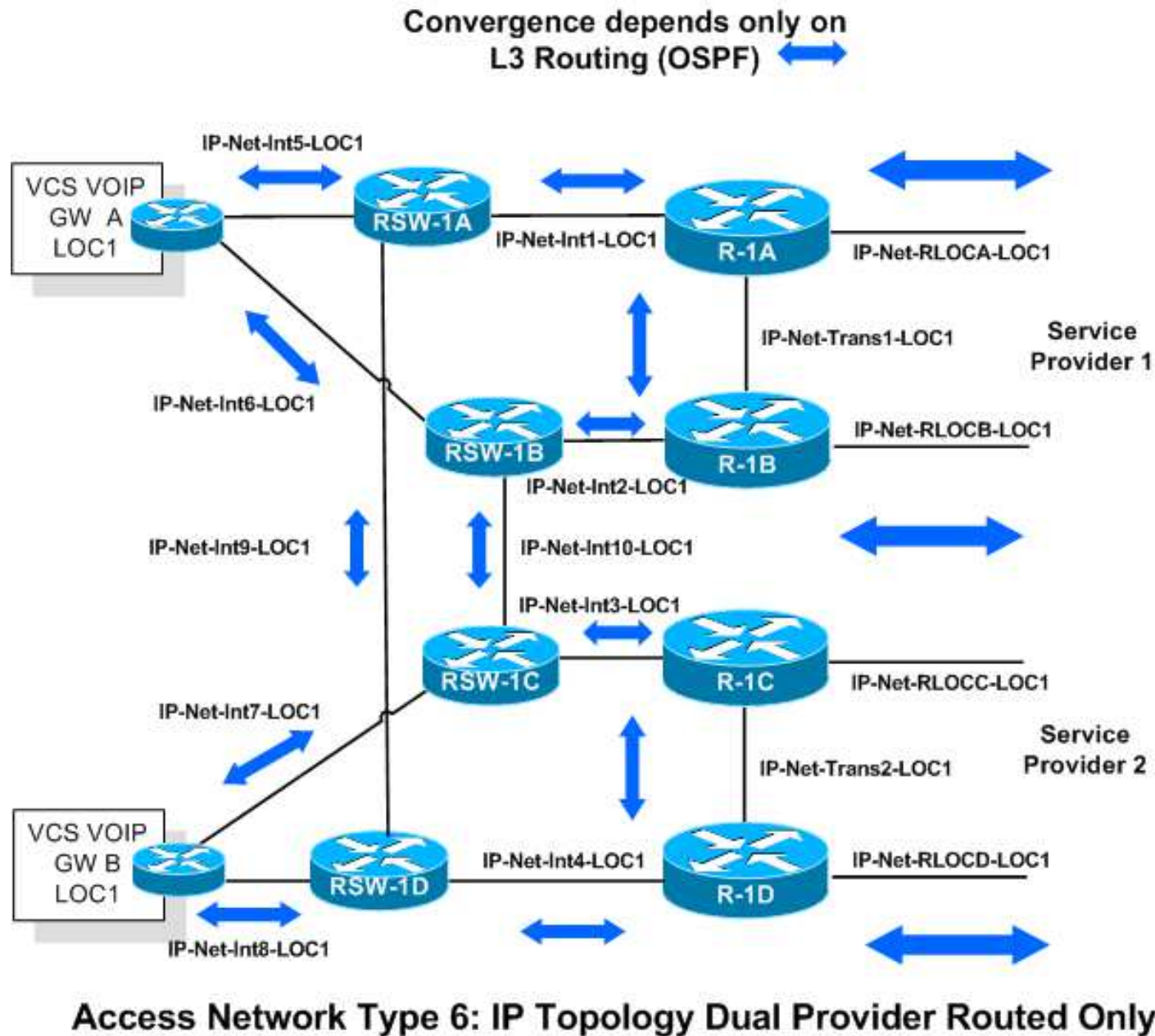
# HA Functional Access Block Type 6



**Access Network Type 6: Physical Topology Routed Only**



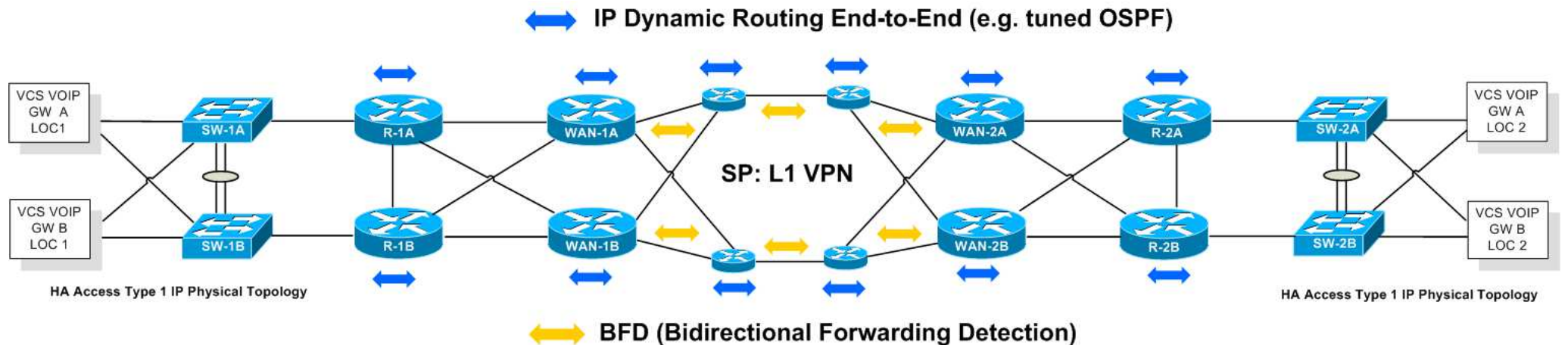
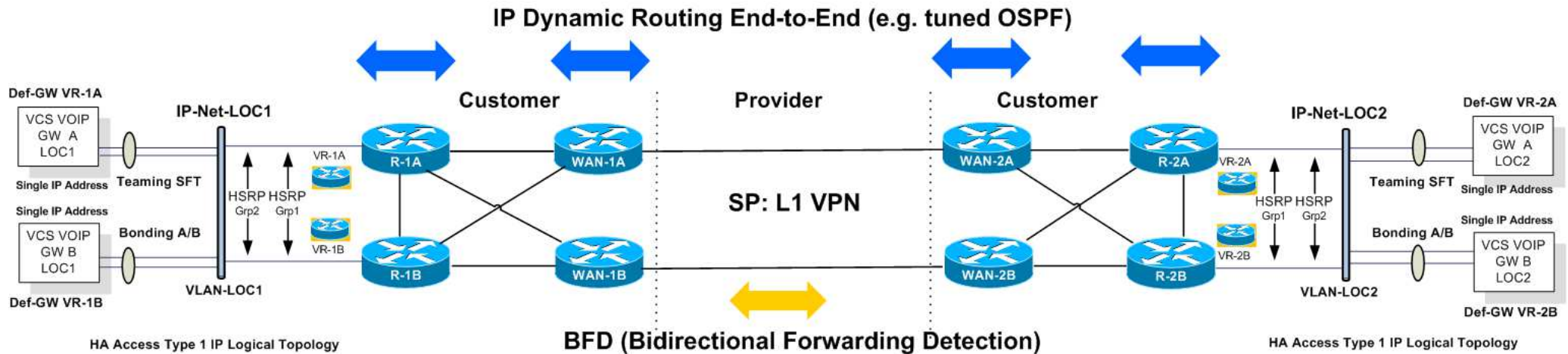
# HA Functional Access Block Type 6 (cont.)



# Agenda

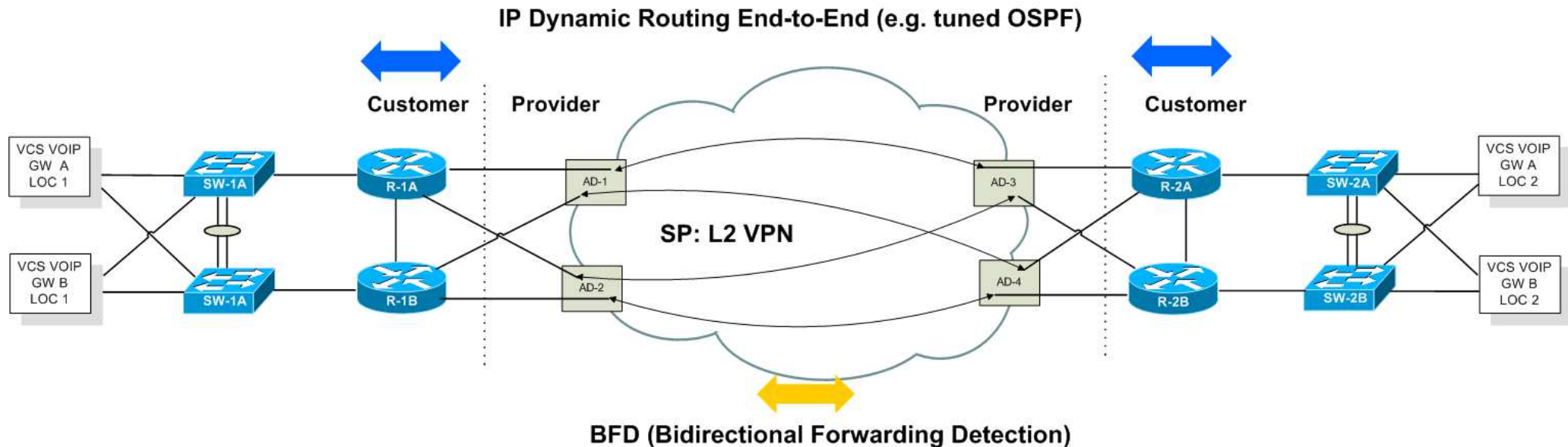
---

- **Introduction**
- **Network Operational Model**
- **High Availability**
  - Elements of HA
  - Functional Access Block Types for HA
  - Routing Aspects
- **QoS**
- **VPN Technology**
- **Multicasting**
- **Summary**

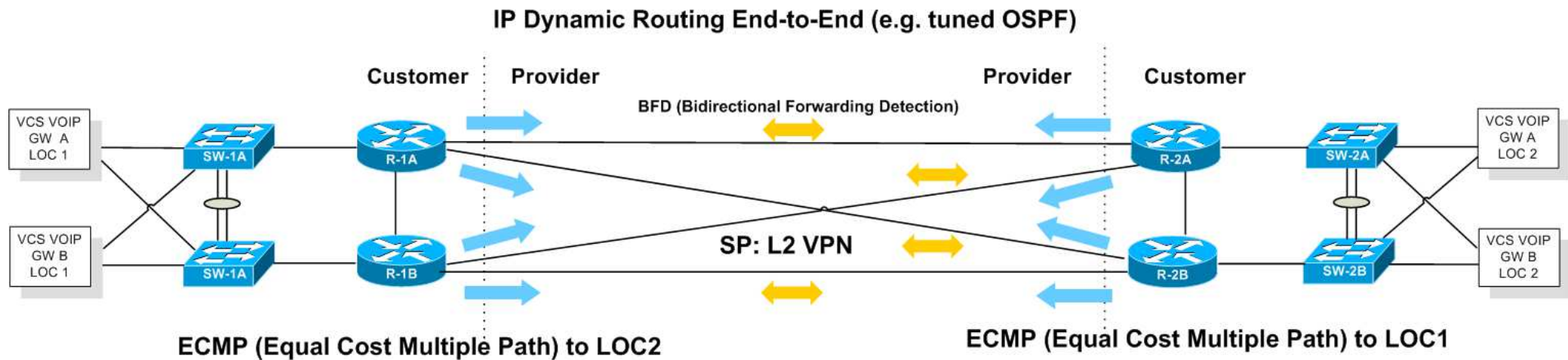


- Full control over IP connectivity and IP routing convergence

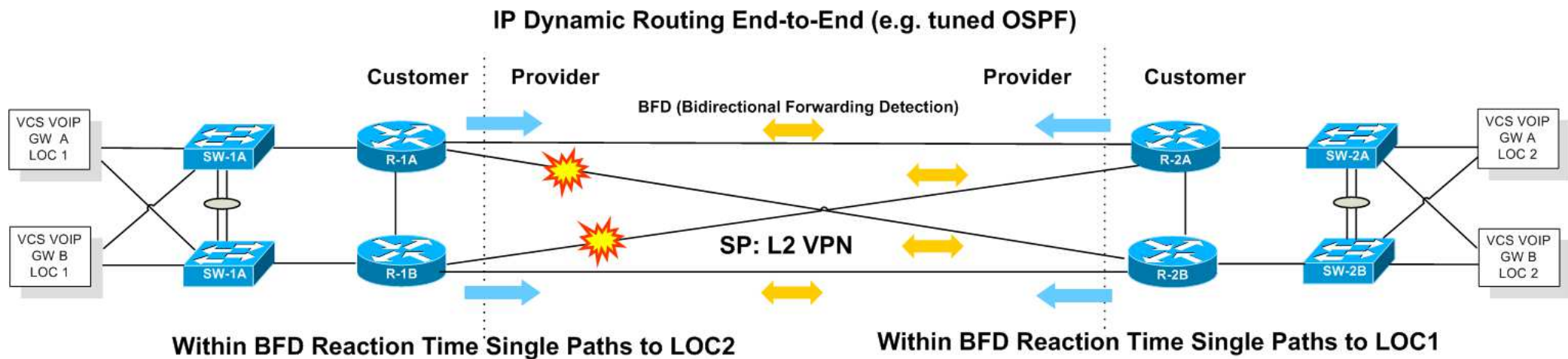




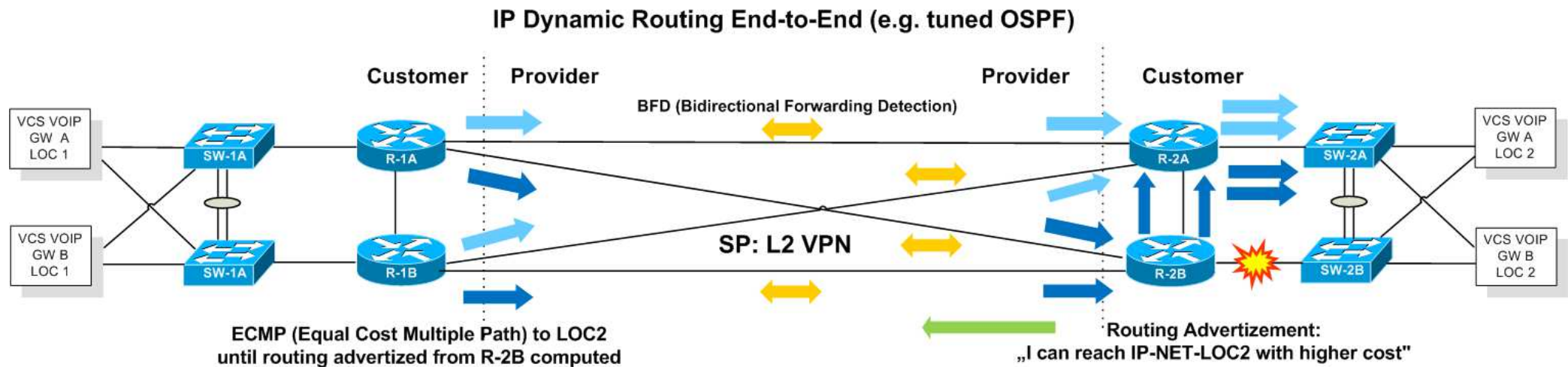
- **You have full control over**
  - IP connectivity and IP routing convergence in case of failures (seconds range)
- **Techniques to be used**
  - Timer tuning of routing protocols to speed up convergence
  - Bidirectional Forward Detection (BFD) to detect indirect failures
  - Equal Cost Multiple Path (ECMP) to load balance and fast switchover



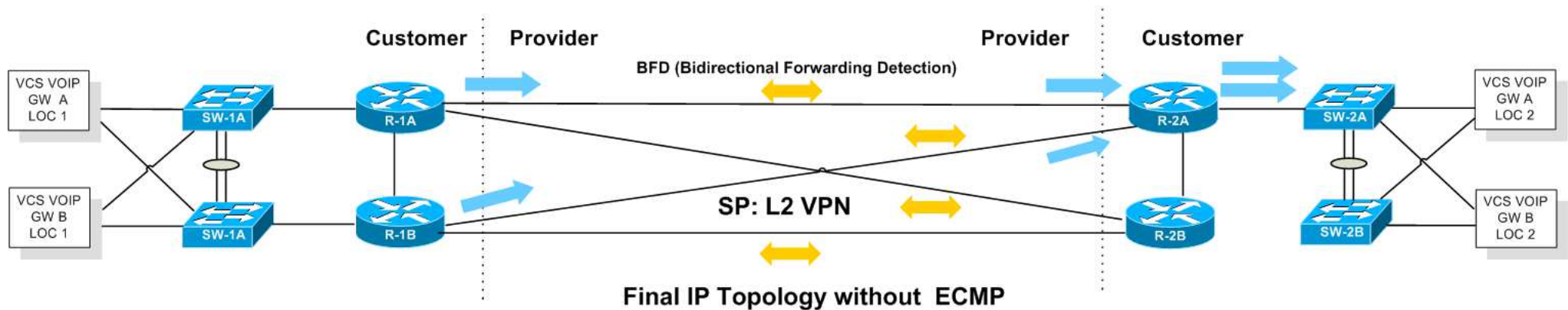
- **ECMP balances traffic session-based over IP paths with equal routing metric**
- **Attention: Links seen need also physical separation in the service provider domain to overcome any single point-of-failures**



- Convergence time depends only on BFD timeout
- Fastest way to direct traffic to remaining links
- Routing updates will inform routers about new topology but not necessary for rerouting



- Protection against single failure at the inside
- Interconnection link between local routers at location 2 allows router R-2B to redirect arriving packets (dark blue) without waiting for convergence of IP routing at routers R-1A and R-1B

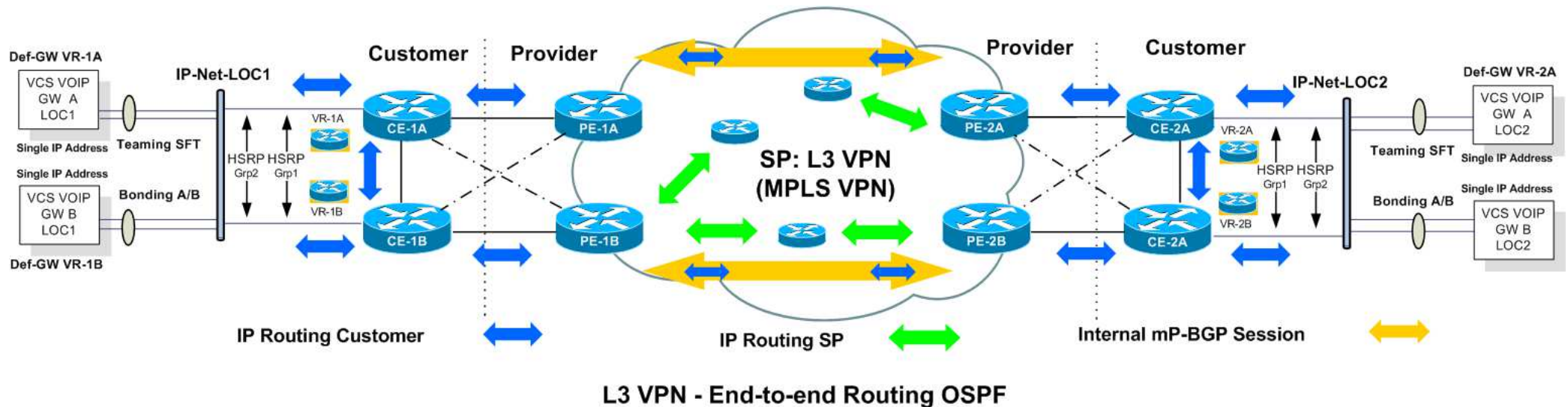


- **Final topology after full routing convergence**
  - No ECMP in such a situation for traffic from CE-1A and CE-1B to IP-Net-LOC2
  - Only a single path remain for routers at location 1





# M3: End-to-End Routing - OSPF



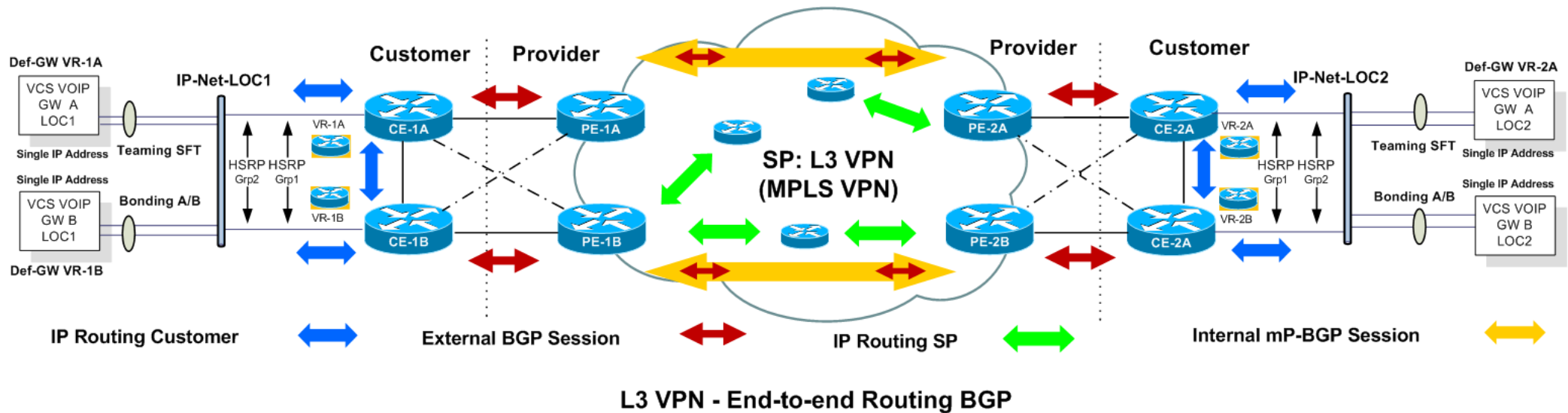
- **End-to-end routing**

- OSPF between CE and PE
- Customers sees OSPF end-to-end with WAN backbone as OSPF area 0
- Customer OSPF is translated into internal mP-BGP to be transported over MPLS-VPN infrastructure
- Internal mP-BGP needs full mesh among all PE routers (scalability issues)

- **Redundancy causes additional complexity**

- Dashed links often not supported

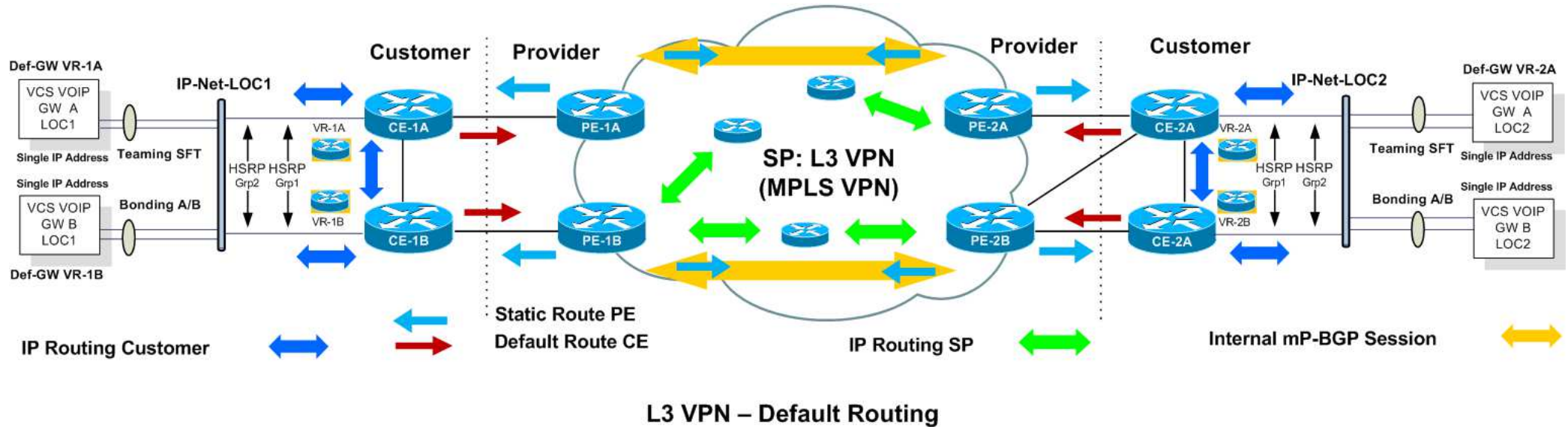
# M3: End-to-End Routing - Ext. BGP



- **End-to-end routing**
  - External BGP between CE and PE
  - Customers sees other locations as different AS (autonomous systems)
  - External BGP is translated into internal mP-BGP to be transported over MPLS-VPN infrastructure
  - Internal mP-BGP needs full mesh among all PE routers (scalability issue)
- **Incoming load balancing adds additional complexity**
  - Dashed links often not supported

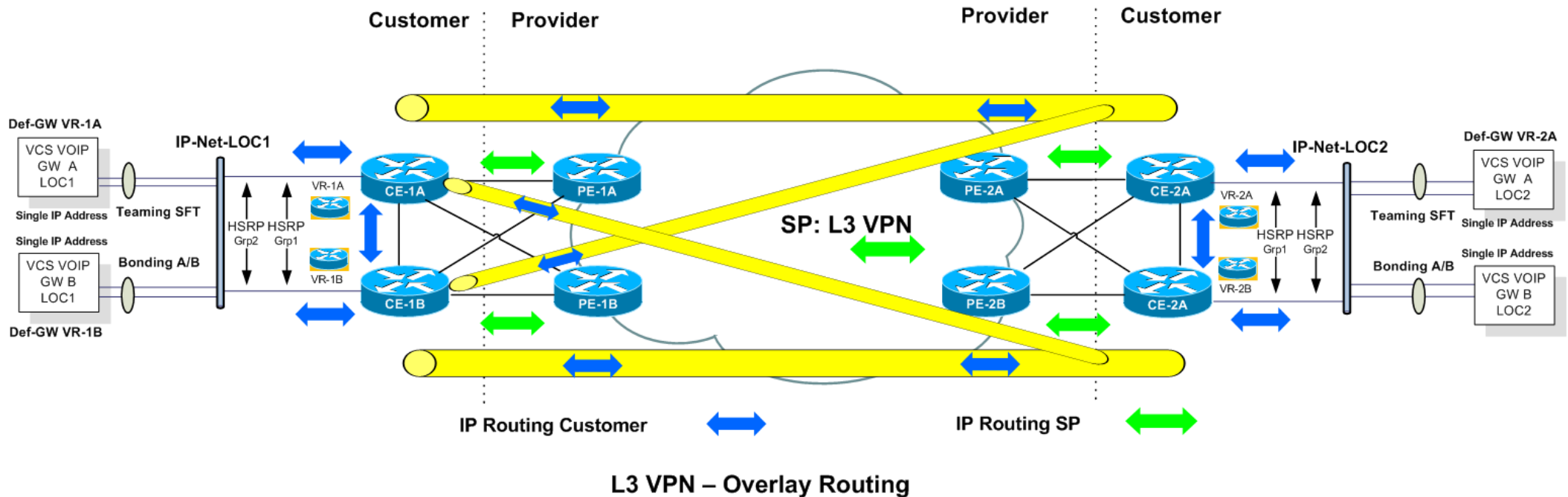


# M3: End-to-End Routing - Default Routes



- **No end-to-end routing**
  - No topology view from customer side
  - Default route at CE points to corresponding PE
  - Static routes at PE points to IP subnets of locations
  - Static routes have to be redistributed to internal mP-BGP in order to be transported over MPLS-VPN infrastructure
  - Internal mP-BGP needs full mesh among all PE routers
- **Incoming load balancing is not supported**

# M3: Overlay Routing



- **Overlay routing**

- Topology view of overlay tunnels from customer side
- GRE tunnels, standalone site-to-Site IPsec tunnels or GRE into site-to-Site IPsec tunnels
- Dynamic routing and routing tuning possible in the overlay
- Scalability issues (full mesh of tunnels, duplication of routing updates on single physical interface)

- **LISP as an alternative technology**

- Locator / Identifier Separation Protocol

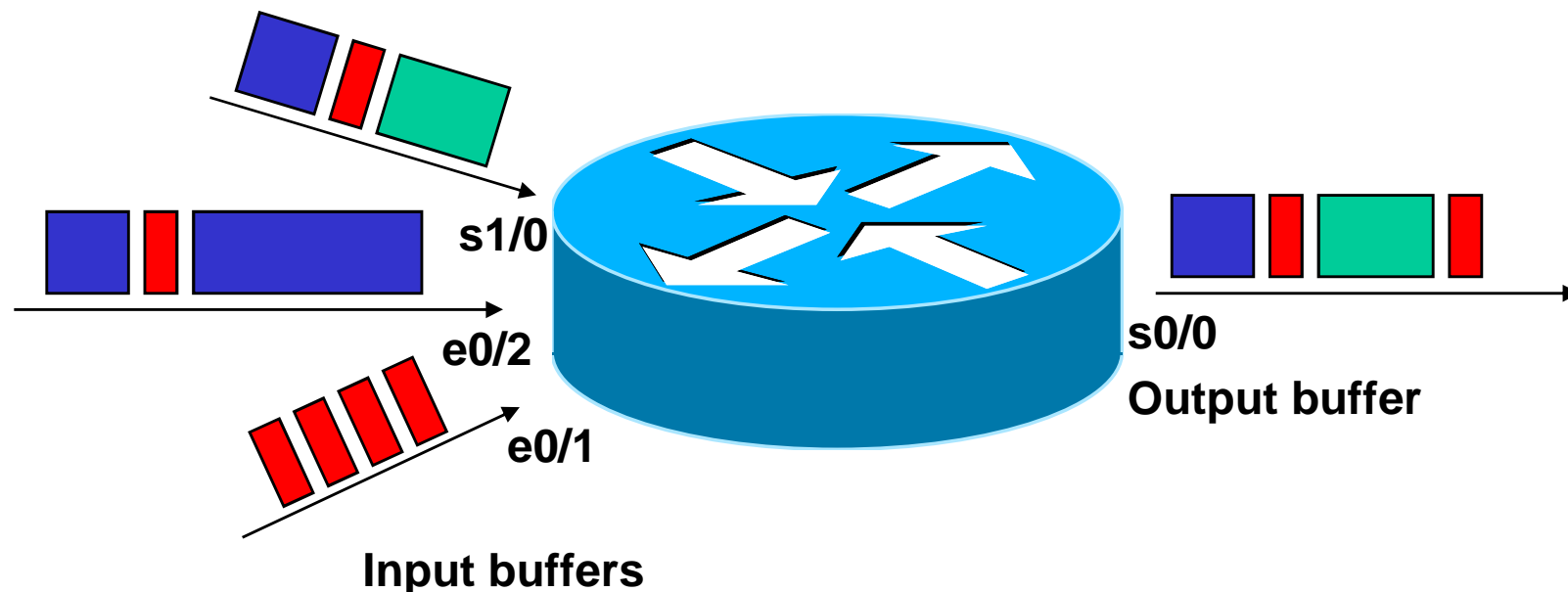
# Agenda

---

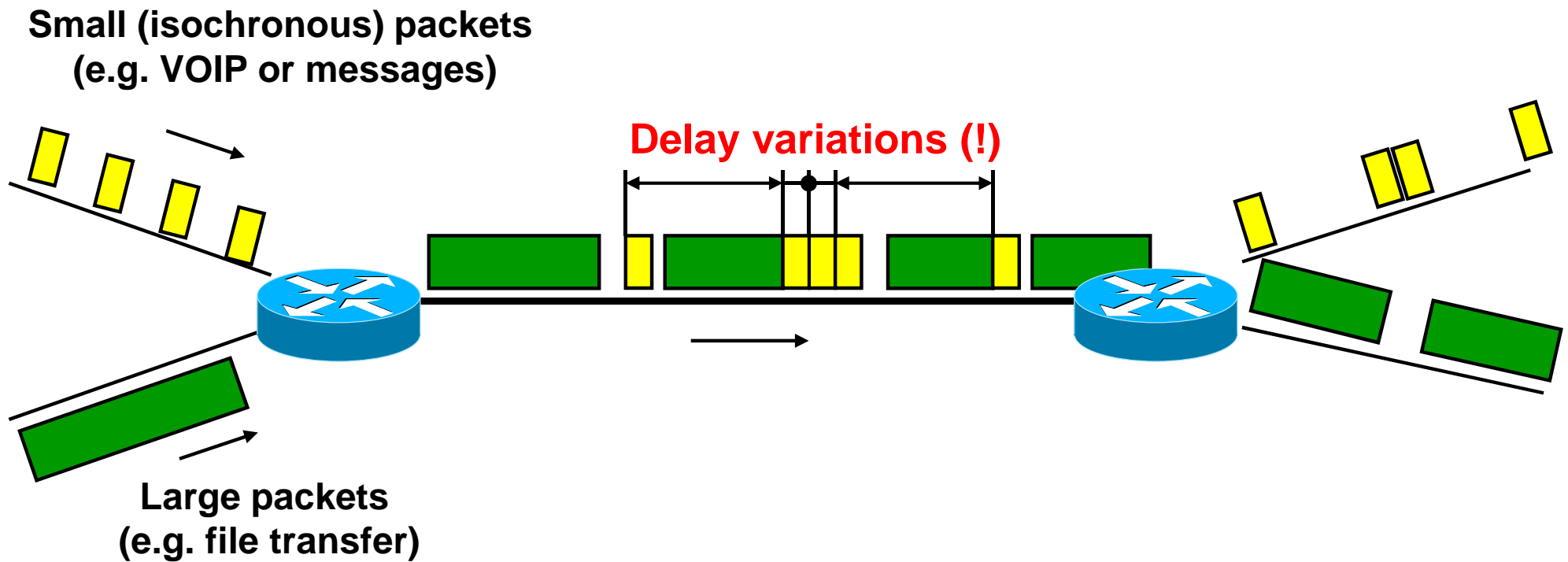
- **Introduction**
- **Network Operational Model**
- **High Availability**
- **QoS**
- **VPN Technology**
- **Multicasting**
- **Summary**

# Packet Switching Needs Buffering

- Packet delivery and switching processes work at different (and varying) rates
- Buffers are needed to interface between those asynchronous processes
  - Too large buffers: Introduce more delay
  - Too small buffers: Packets might get lost during bursts

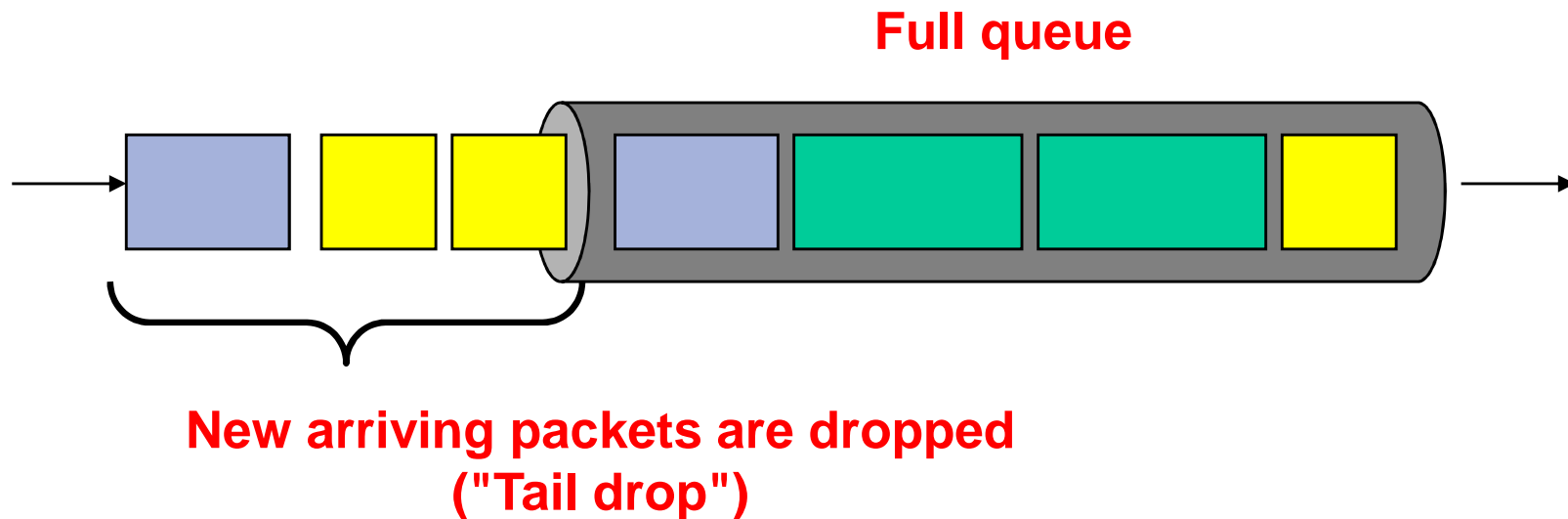


# Jitter = Delay Variation Caused By Serialization Delays



# FIFO Queuing / No - QoS

- ***Tail-drop queuing*** is the standard dropping behavior in FIFO queues
  - If queue is full all subsequent packets are dropped
- **Of course that is not sufficient to implement any kind of QoS**

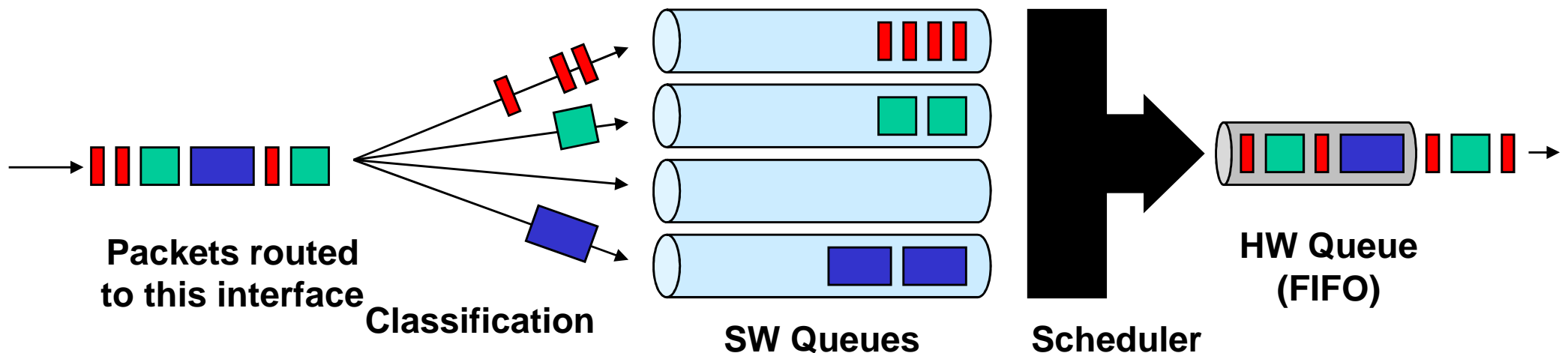


# IP Quality of Service

- **No QoS is necessary in case of over-provisioning**
  - But can you economically justify it?
- **Manages available bandwidth in case of congestion**
  - But cannot create additional bandwidth on the fly
- **Ensures certain upper limits for transmission parameters**
  - Bounded maximum delay, jitter and loss
  - Assured minimum throughput
- **Needs more performance at the network components**
  - Hardware (ASIC), CPU, memory at Ethernet switches, IP routers, firewalls, etc.
- **Needs monitoring**
  - To understand what is going on in your network
  - To recognize trends for deploying additional bandwidth in time

# IP Routers With QoS Support

- **Queuing actually encompasses two parts: SW and HW queues!**
- **SW queuing is typically more sophisticated**
  - WRR (Weighted Round Robin)
  - CBWFQ (Class Based Weighted Fair Queuing)
  - Priority Queuing, LLQ (Low Latency Queuing)
  - These kind of techniques are an important part of any QoS implementation
- **HW queuing is typically only FIFO**
- **SW queue only needed if HW-queue full**
  - Otherwise packet bypasses SW-queue





# Building Blocks for QoS

## QoS Policy

Signaling  
(RSVP)

Admission  
Control

Classification  
and Marking

Integrated  
Service

Differentiated  
Service

Connection-oriented QoS

Per-hop Behavior

Traffic  
Shaping  
& Policing

Congestion  
Control

Queuing

Fragmentation  
and Interleaving

Token Buckets

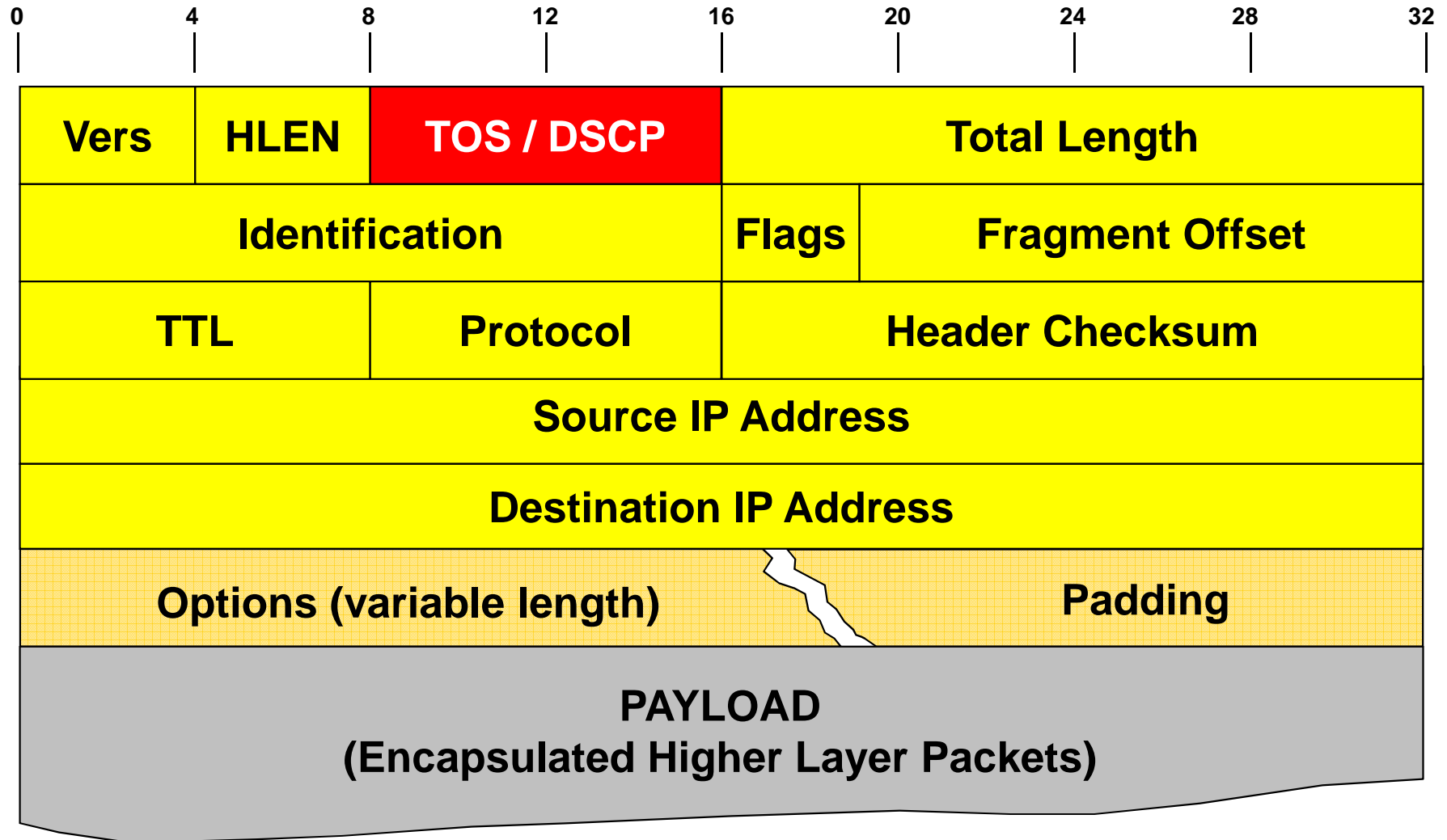
RED, WRED  
ECN

Priority  
Weighted Fair,  
Class-based

MLP, LFI

# IP Header Field TOS / DSCP

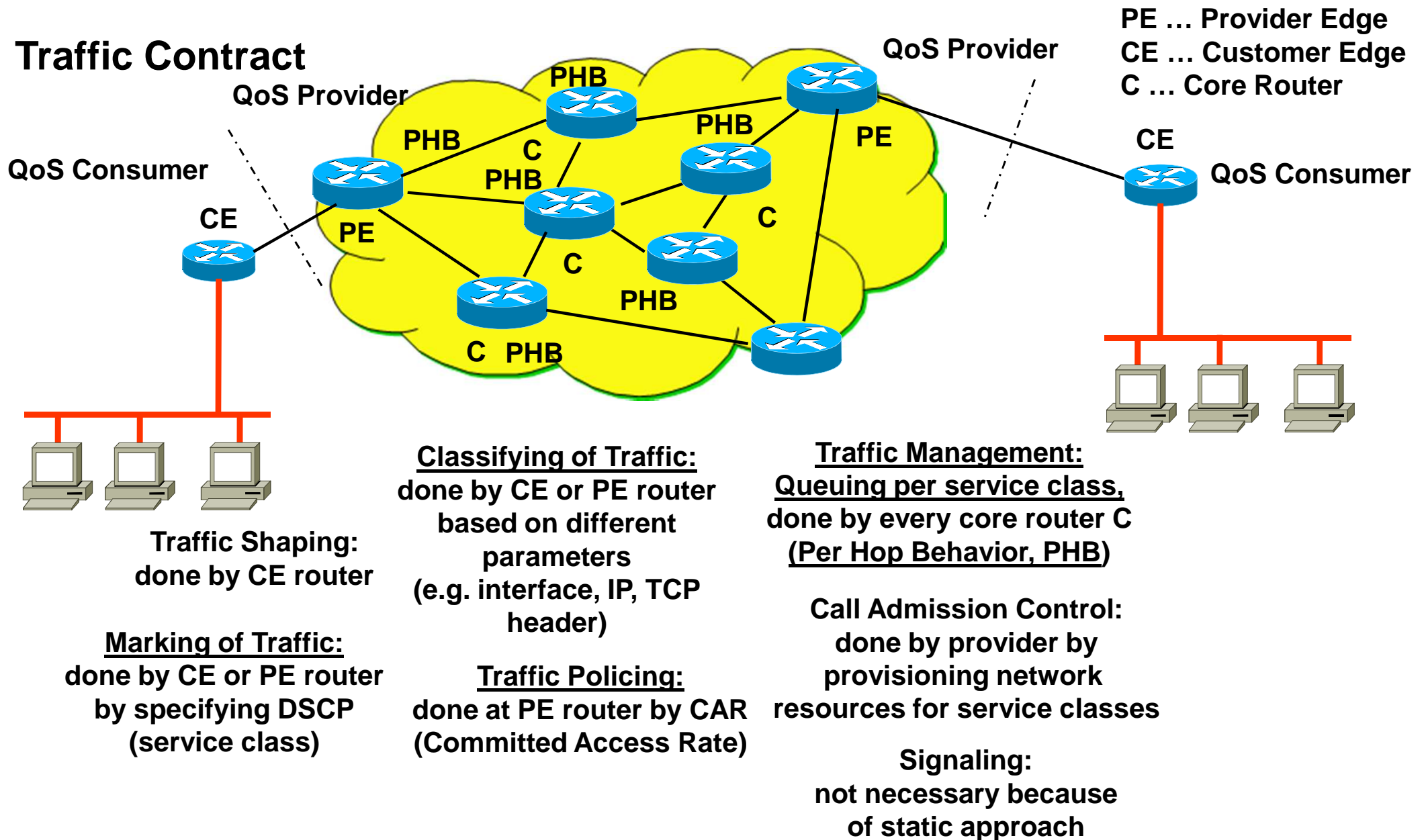
(Used as Indication of QoS Service Class)



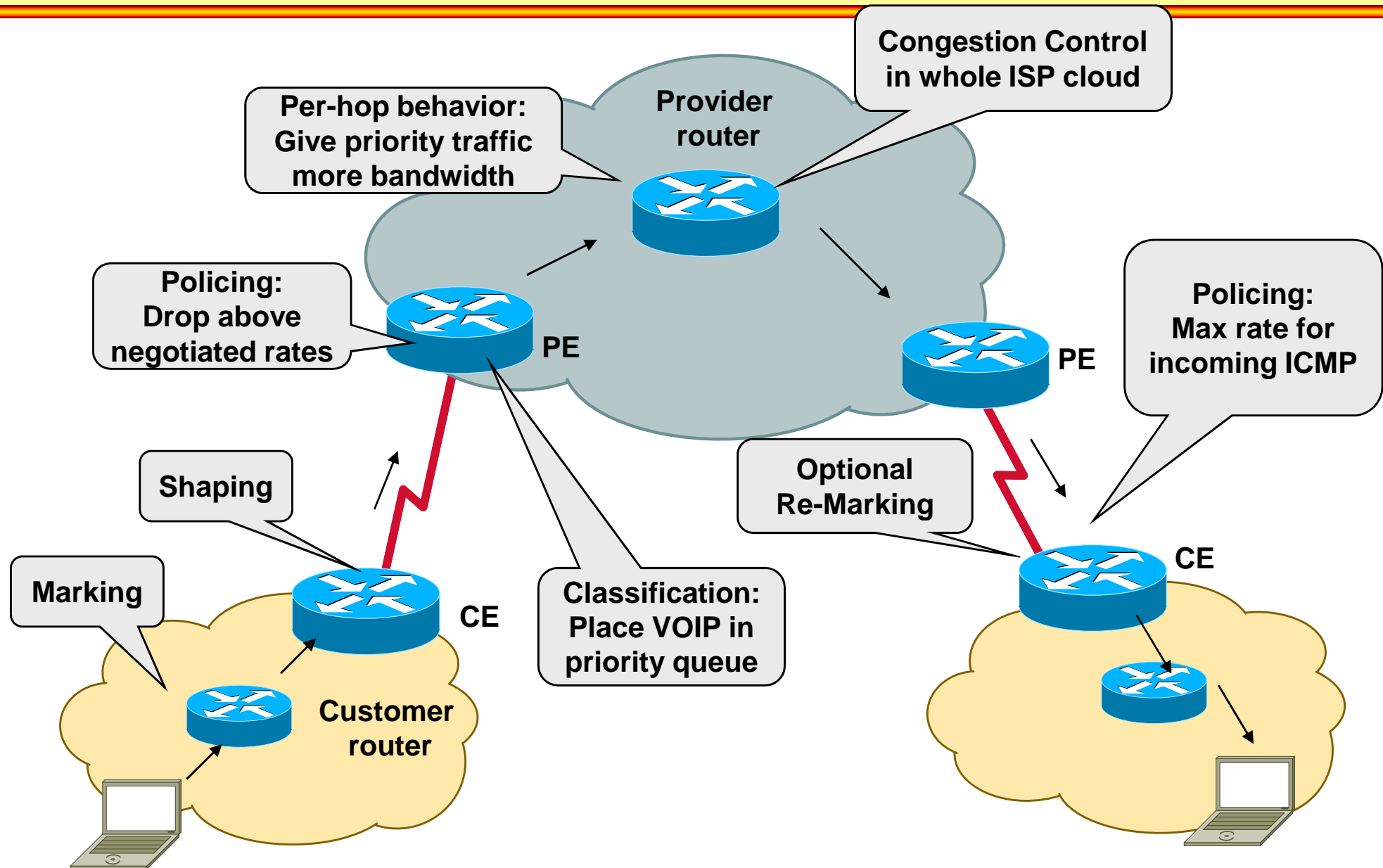
# DSCP Values Overview

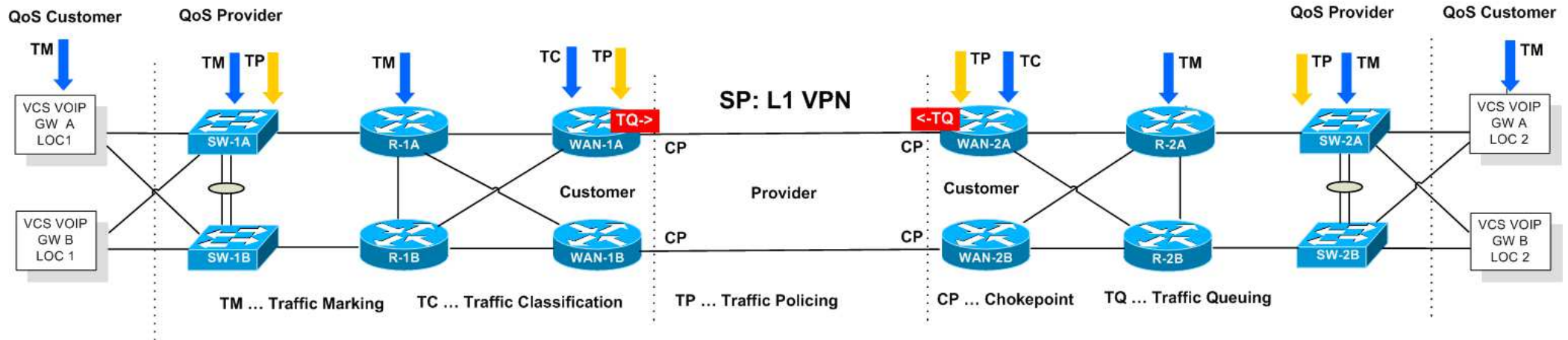
Code Point Name	DSCP		Whole IP TOS byte		
	hex	dec	binary	hex	dec
EF	0x2e	46	10111000	0xb8	184
AF41	0x22	34	10001000	0x88	136
AF42	0x24	36	10010000	0x90	144
AF43	0x26	38	10011000	0x98	152
AF31	0x1a	26	01101000	0x68	104
AF32	0x1c	28	01110000	0x70	112
AF33	0x1e	30	01111000	0x78	120
AF21	0x12	18	01001000	0x48	72
AF22	0x14	20	01010000	0x50	80
AF23	0x16	22	01011000	0x58	88
AF11	0x0a	10	00101000	0x28	40
AF12	0x0c	12	00110000	0x30	48
AF13	0x0e	14	00111000	0x38	56
CS7	0x38	56	11100000	0xe0	224
CS6	0x30	48	11000000	0xc0	192
CS5	0x28	40	10100000	0xa0	160
CS4	0x20	32	10000000	0x80	128
CS3	0x18	24	01100000	0x60	96
CS2	0x10	16	01000000	0x40	64
CS1	0x08	8	00100000	0x20	32
CS0 = BE	0x00	0	00000000	0x00	0

# Differentiated Services Model: Elements

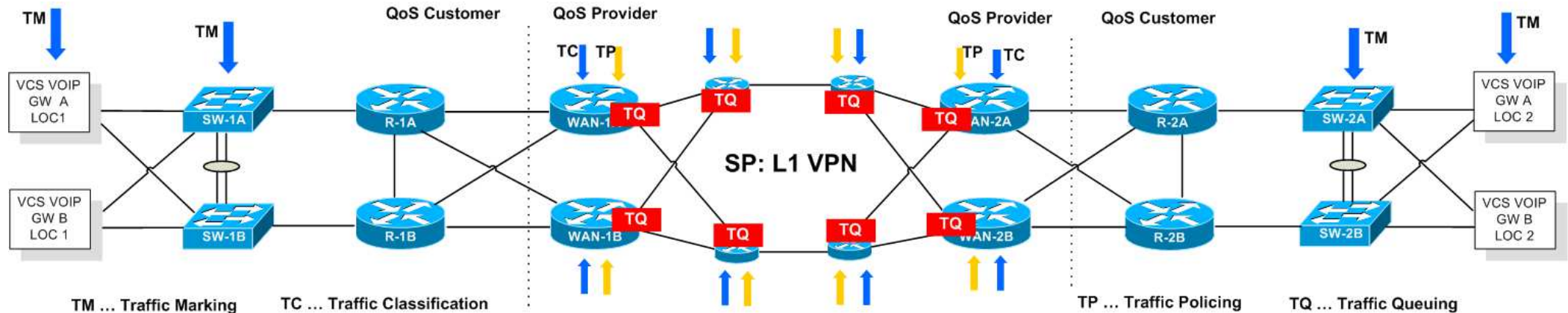


# Differentiated Services Model: In Action





- **You have full control over**
  - QoS tuning based on necessary communication matrix
  - QoS consumer -> your applications using the network infrastructure
  - QoS provider -> network team establishes the necessary QoS behaviour in the network
- **Techniques to be used**
  - Traffic marking at the QoS edge (end-system if trusted, first Ethernet switch if un-trusted)
  - Traffic classification, traffic policing and traffic queuing on choke points of WAN backbone
  - Traffic policing optionally at the QoS edge (first Ethernet switch) to implement a kind of admission control
- **You need QoS Monitoring / Management**
  - To find out or verify communication behaviour (matrix) of the QoS consumer (e.g. with the help of NetFlow)
  - To recognizes trends for additional bandwidth needed in the network

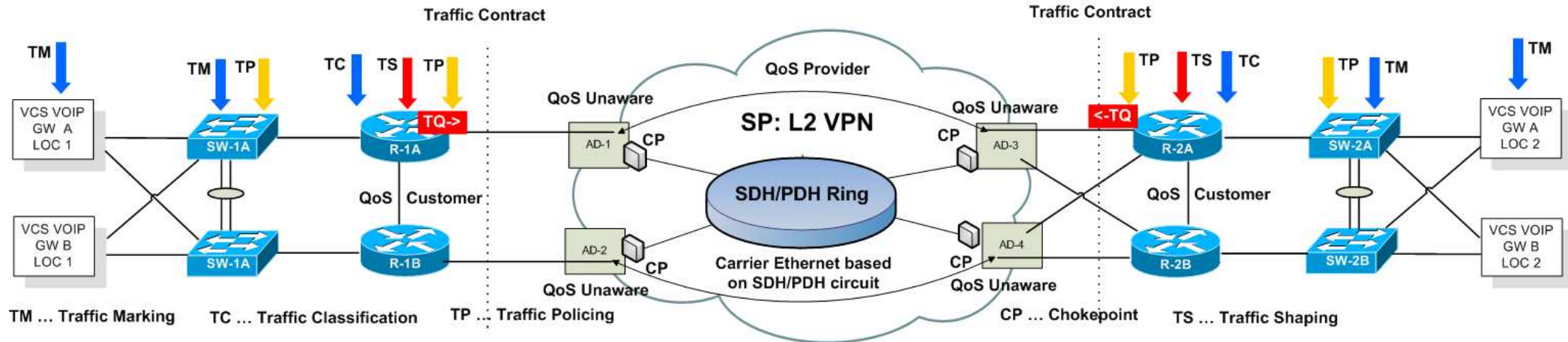


## ● Points to be kept in mind:

- Load balancing like ECMP will split traffic session-based
- In the case of single point of failure after routing convergence the load balancing will stop
- Hence QoS tuning of a single link must calculate the summary bandwidth for the most critical traffic in such a situation
  - Otherwise service degradation might happen for the most critical traffic (real-time voice, real-time video)
- Critical traffic typically used LLC (priority-queue based)
  - Priority queue should always be policed to avoid starvation of the network for other traffic in case a erroneous system produces huge amount of critical traffic
- Regarding amount of traffic classes: less is better than more
- Do not use MLP to bundle physical links to an aggregate link (e.g. 4 x 2 Mbps E1 -> 8 Mbps)
  - Problems with QoS parameters, routing metrics, BFD, fast routing convergence





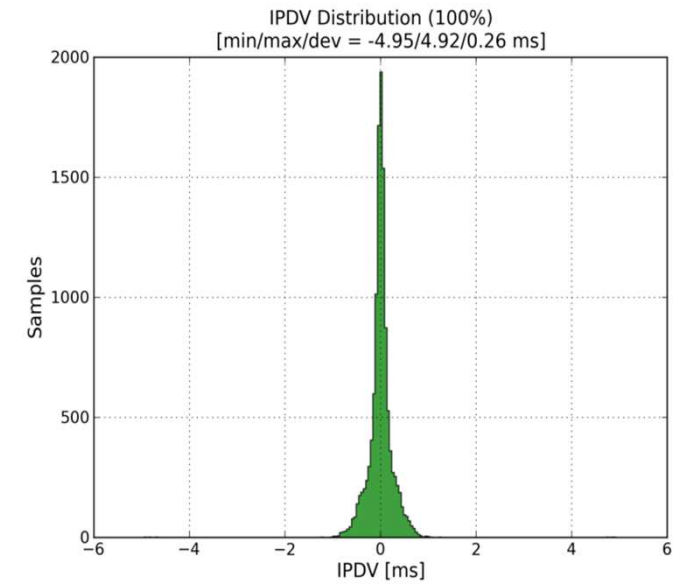
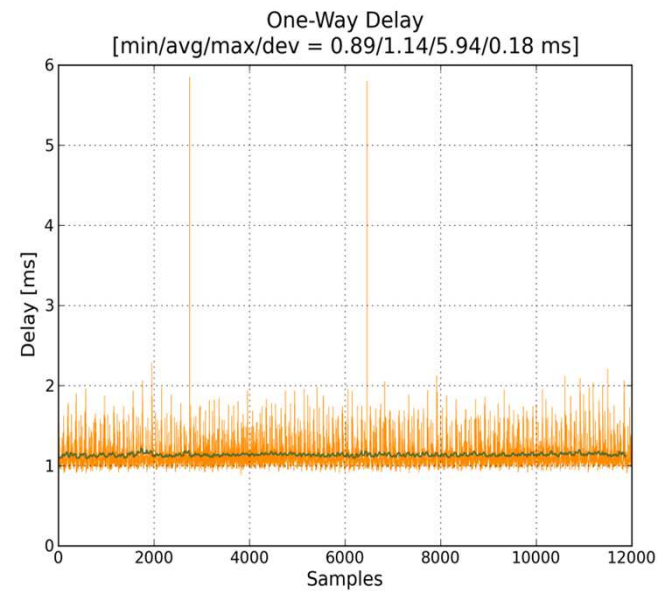


## ● Bandwidth mismatch on internal carrier edge

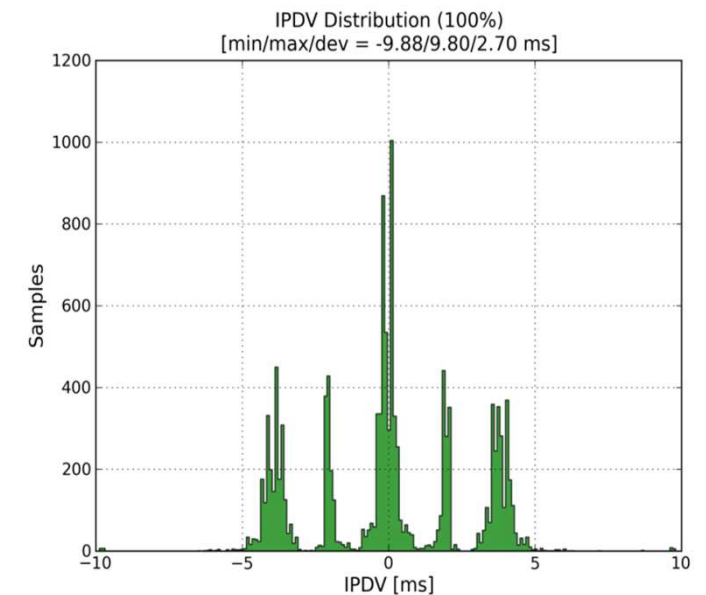
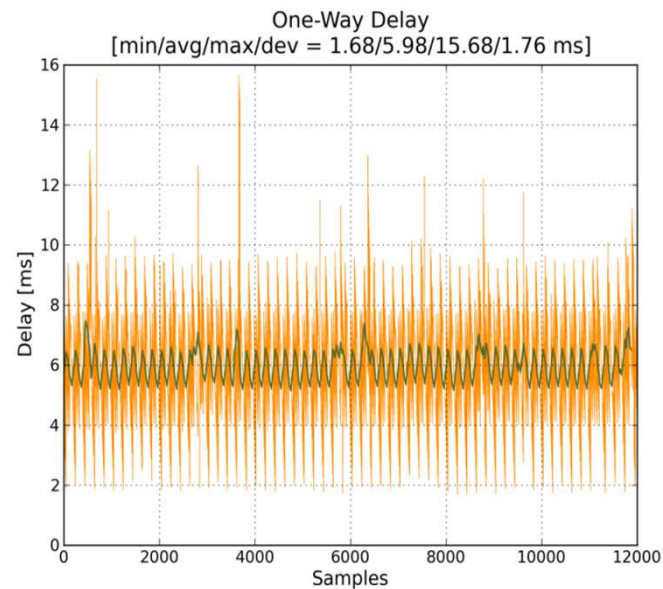
- E.g. putting 100Mbps Ethernet onto a 34Mbps PDH circuit
- E.g. putting 10Mbps Ethernet onto a 6Mbps microwave circuit
- Carrier edge equipment typically implements Ethernet switch functionality (= transparent bridging) with less sophisticated QoS tuning instrumentation or even no QoS support
- But for R routers it looks just a normal Ethernet LAN giving nominal Ethernet speed.
- It is must to implement traffic shaping on the corresponding R routers to avoid any uncontrolled packet drops at the carrier edge
- But traffic shaping introduces larger variance of delay variation (jitter) and sums up if there are several shapers in a queue

# Voice Jitter Tests (1 Shaper)

No shaper active

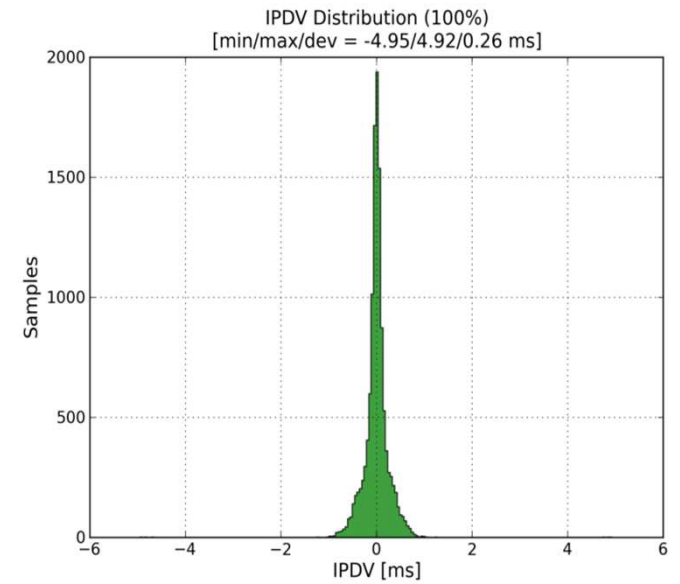
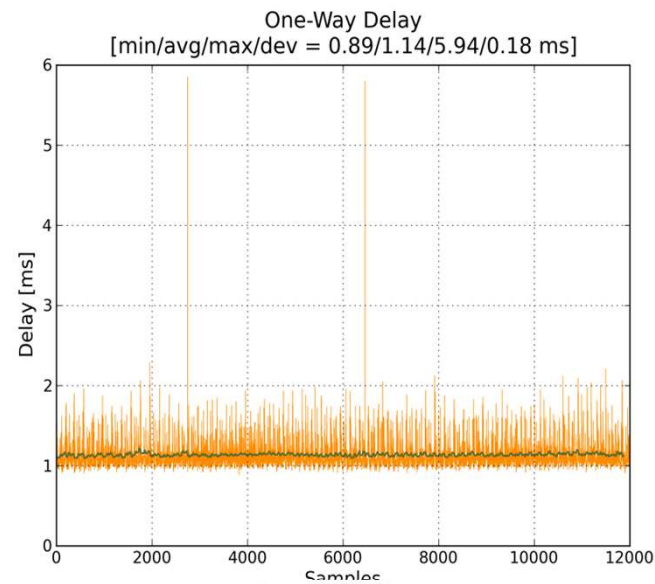


1 shaper active

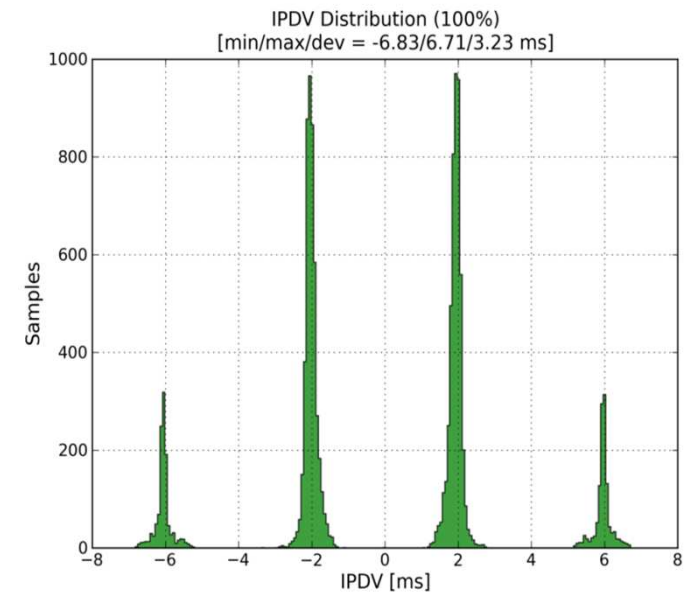
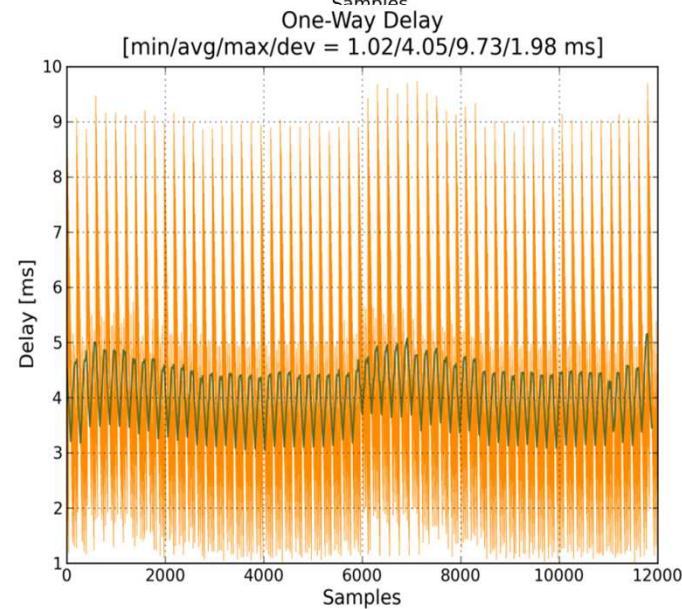


# Voice Jitter Tests (2 Shaper)

No shaper active

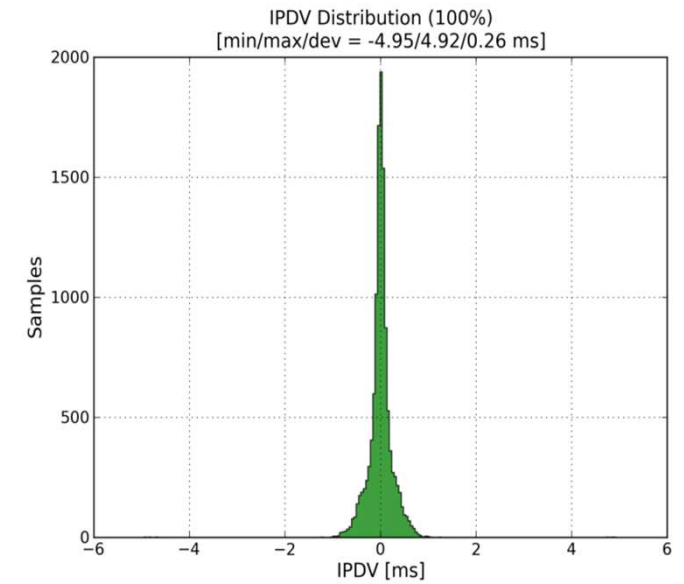
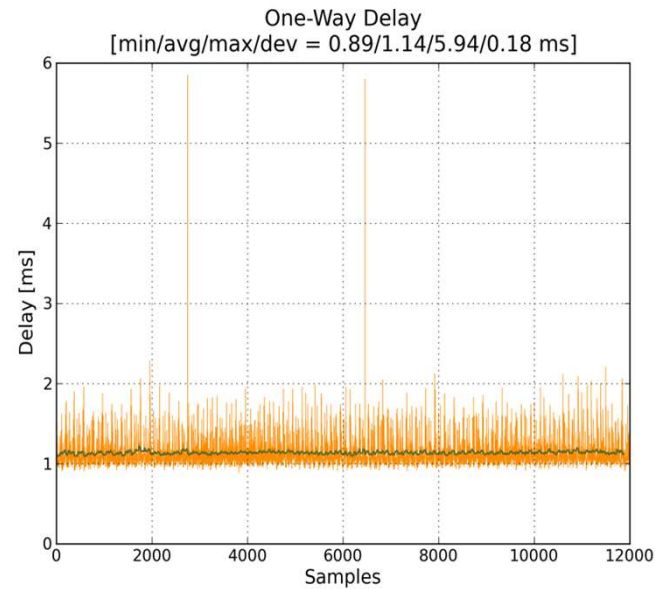


2 shaper active

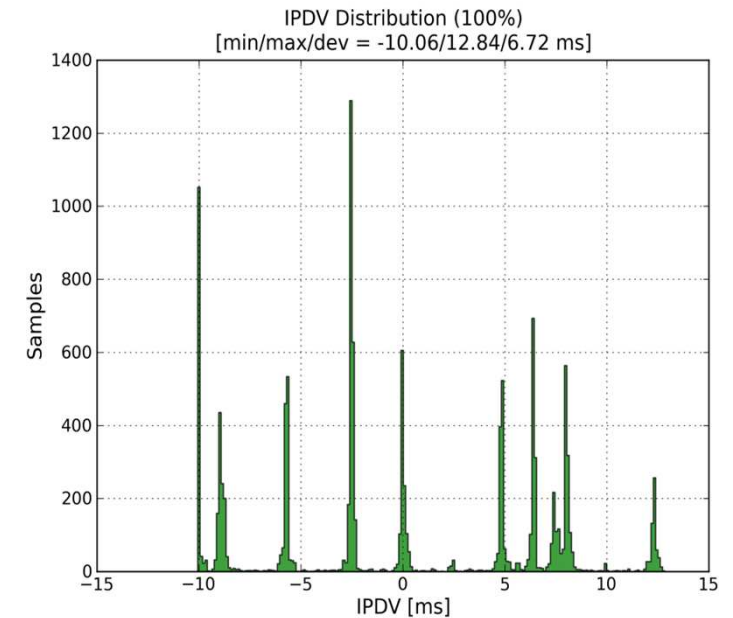
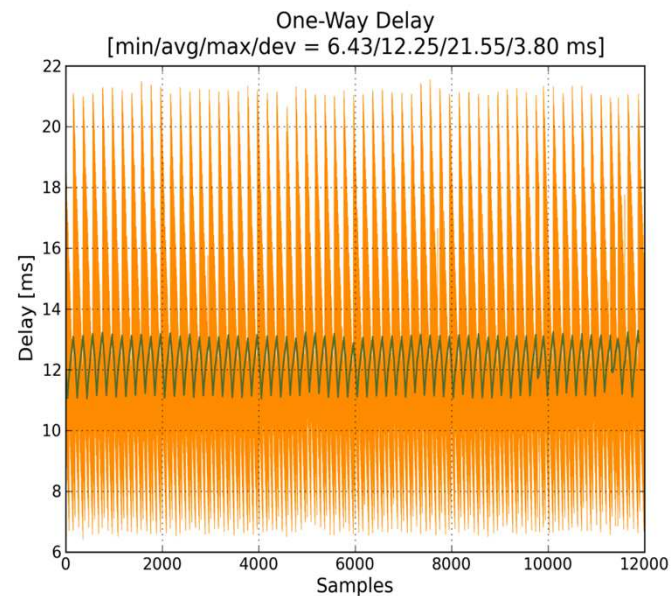


# Voice Jitter Tests (3 Shaper)

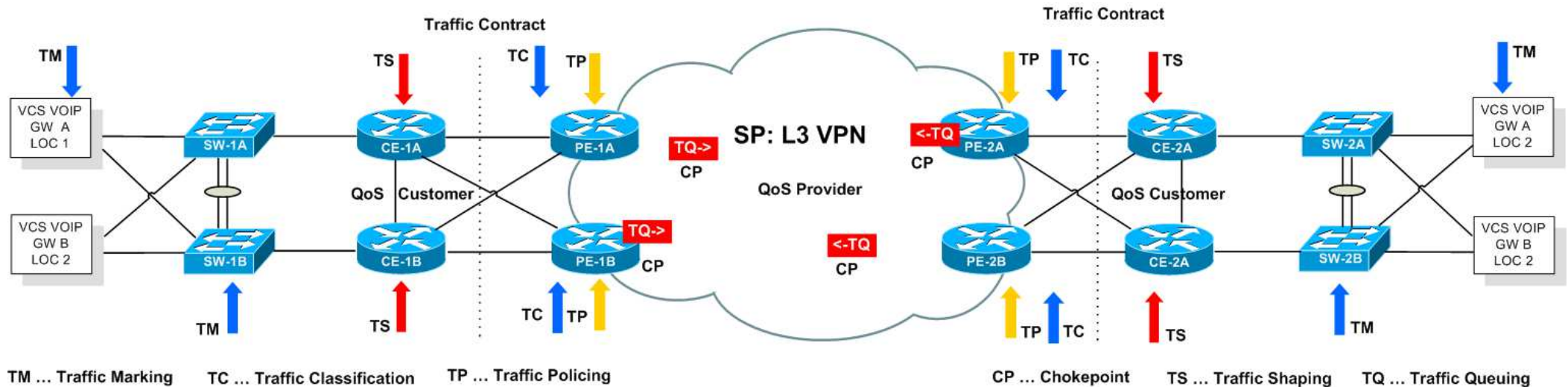
No shaper active



3 shaper active







- **Traffic contract (static)**

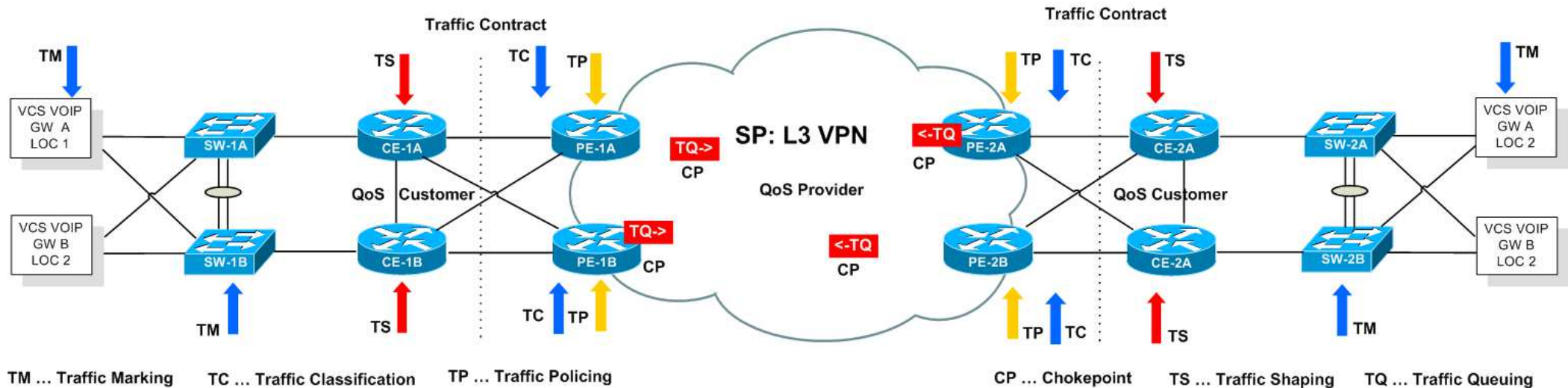
- Between QoS consumer and QoS provider
- QoS consumer relies on the correct QoS implementation at the provider

- **As customer you have only limited control over**

- QoS tuning (-> just marking, maybe shaping if you want to obey the traffic contract in the case your communication matrix is not fully known)
- TC, TP and TQ is done at provider routers which cannot be controlled by the customers

- **QoS monitoring and management**

- Have to be done by both parties
  - Provider justifies SLAs are obeyed
  - Consumer proofs if SLAs are fulfilled
- Otherwise as customer you completely have to trust your provider

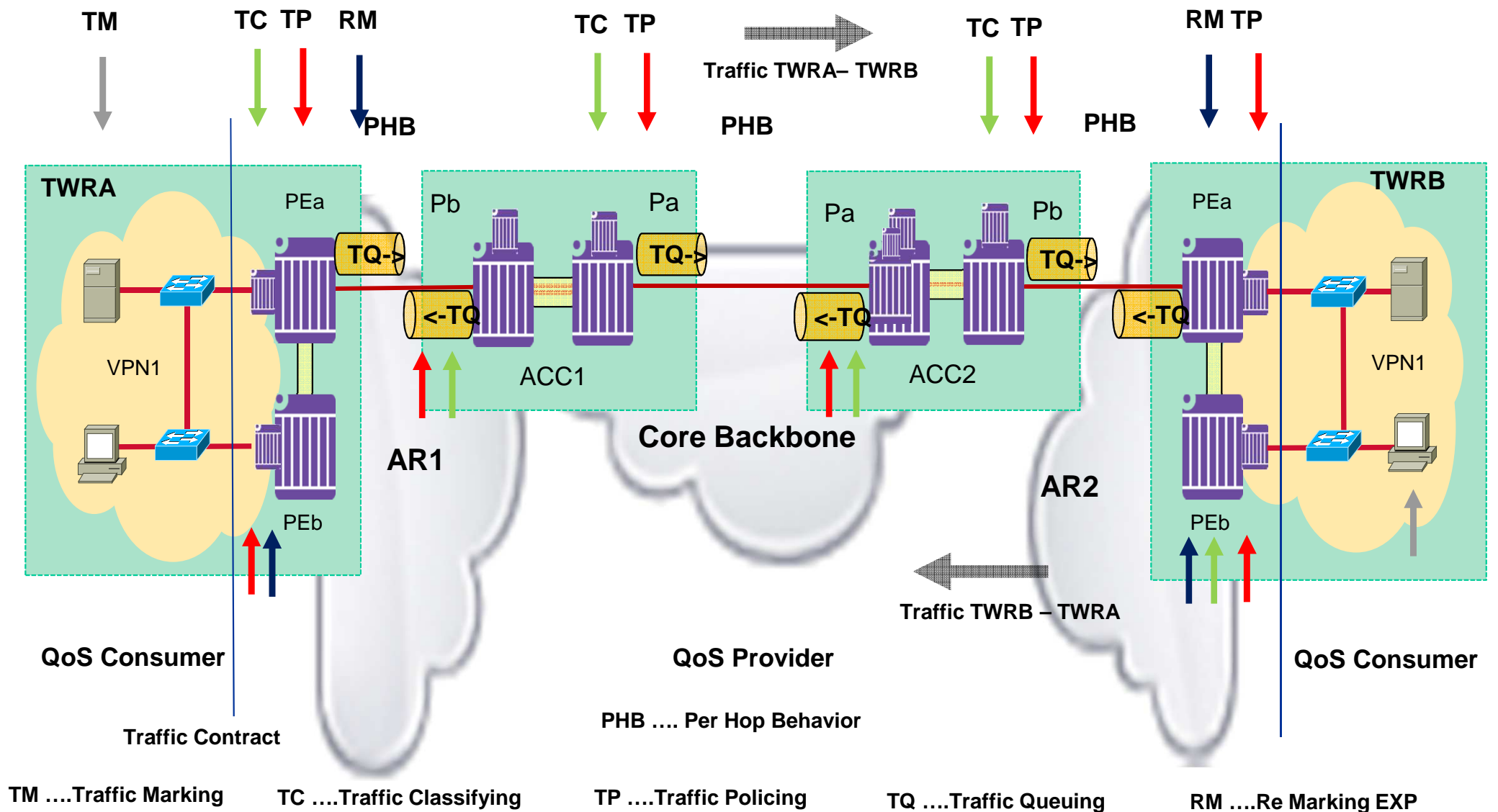


## ● SP tasks and challenges:

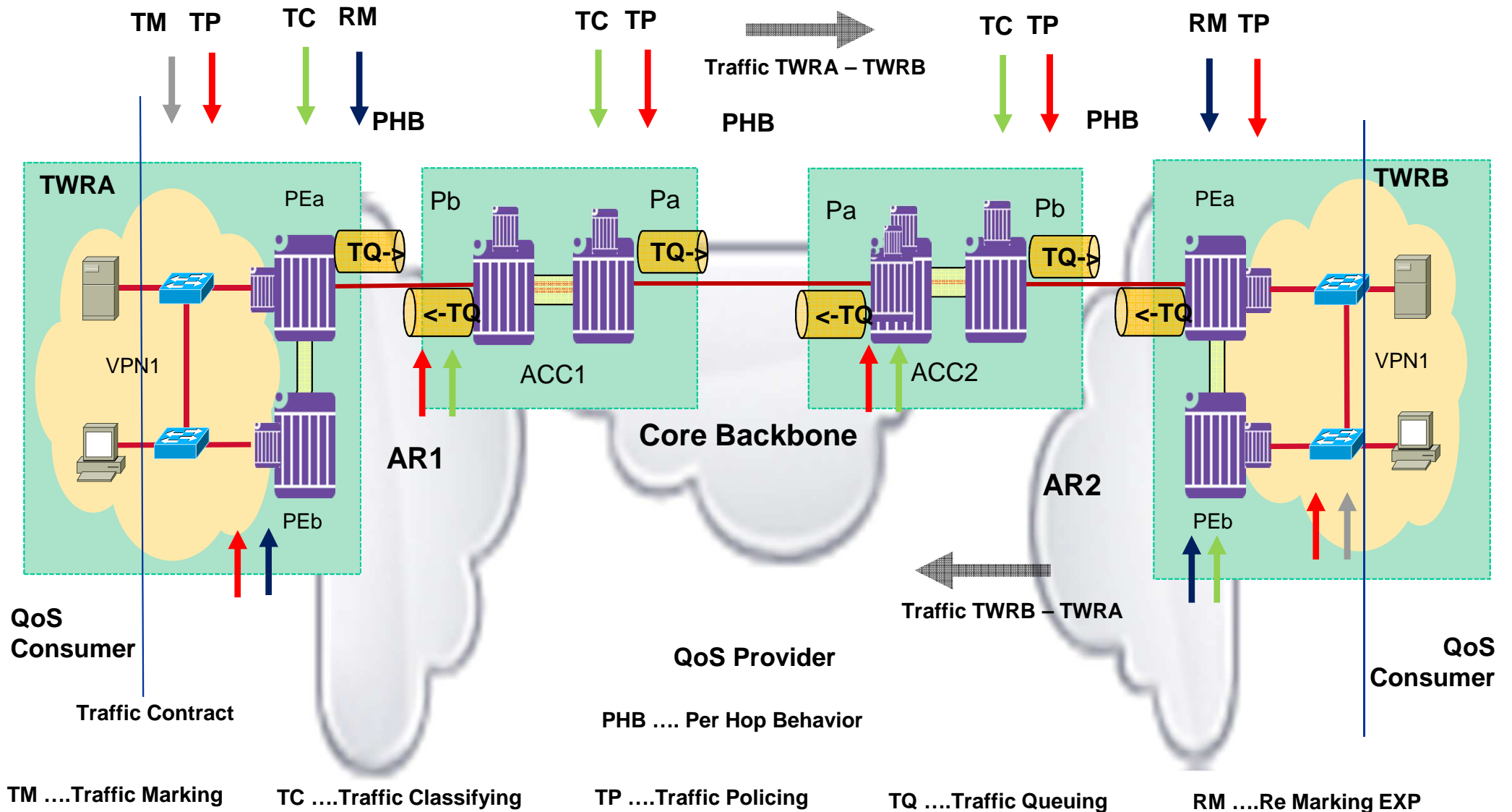
- Implementation of DiffServ Model for multiple customer usages
- Mapping of customer services classes onto internal service classes (= remarking)
  - E.g. for MPLS-VPN you have only 8 possible values for QoS tagging
  - Critical traffic of different customers will use the same internal service class
- Policing every customer down to the agreed values
  - Otherwise one customer can influence another customer by not obeying the rules
  - Important for priority queue carrying the most critical traffic
- Finding an appropriate network topology and bandwidth provisioning
  - To guarantee high availability and QoS
  - At least there must be enough bandwidth for the sum of all critical traffic streams of all customers
- But L3 VPN SP needs a kind of under provisioning and some statistical traffic behaviour of his customers to economically survive

# Example1: QoS Functions Overview

## MPLS-VPN Based

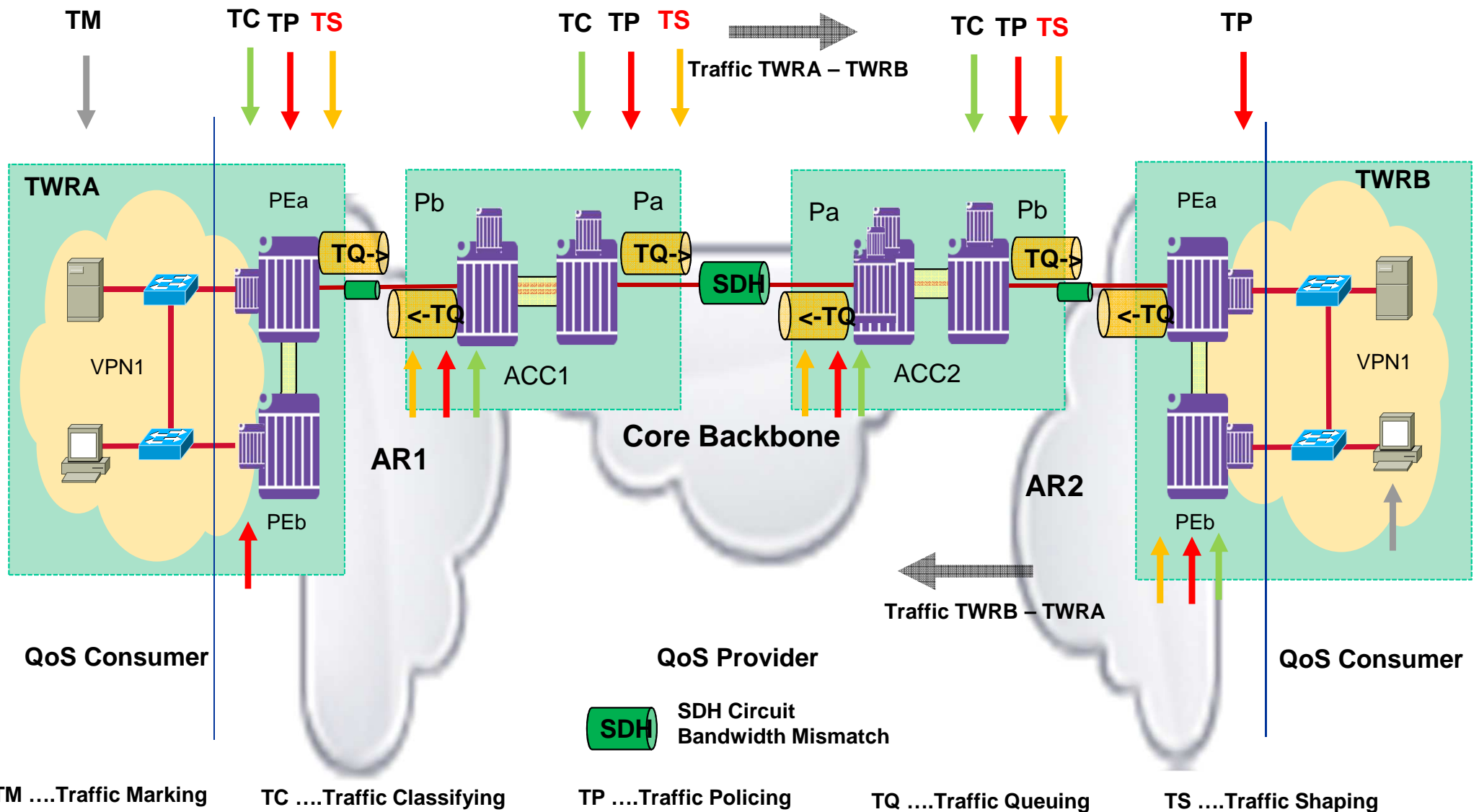


# Example2: QoS Alternate Control Closer To Endsystem





# Example 3: QoS Functions With Bandwidth Mismatch



# Agenda

---

- **Introduction**
- **Network Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
  - Introduction IT-Security
  - VPN Types
  - MPLS, MPLS-VPN
  - IPsec VPN
  - DMVPN
  - GETVPN
- **Multicasting**
- **Summary**

# Information Security (Definition ISO 27001:2005)

- **Preservation of confidentiality, integrity and availability of information**
  - In addition other properties such as authenticity, accountability, non-repudiation and reliability can also be involved
- **Confidentiality (Privacy)**
  - The property that information is not made available or disclosed to unauthorized individuals, entities or processes
  - Intuitive: the information can be read only by intended persons, field of “encryption”
- **Integrity**
  - The property of safeguarding the accuracy and completeness of assets
  - Intuitive: we can trust in the information, it is not changed unintentionally, field of “fingerprint and cryptographic checksum/hashes”
- **Availability**
  - The property of being accessible and usable on demand by an authorized entity
  - Intuitive: the information is accessible when it is really needed

# Information Security

- **Confidentiality, Integrity, Availability (CIA)**
  - Different views on security for information in transit (IIT) or information at rest (IAR)
  - Different areas: network security, computer security,
- **Security is a process with a life-cycle**
  - And not just the implementation of security functions by technology
  - 20% technology related, 80% organization related
- **Topics included**
  - Security assessment, risk analysis
  - Security concept identifying domains, borders between domains, organization of responsibilities
  - Security implementation (technological and organizational)
  - Security management
    - Policies, controls, audits

# Computer Security

- **Information At Rest (IAR)**

- Availability

- Downsizing to required functionality
- Hardening and access control
- Redundancy
- Backup

- Confidentiality and integrity

- Access control (in most cases generic functionality of the OS)
- Authentication (e.g. username / password)
- Authorization (e.g. ACLs – access control list)
- (Encryption)

# Network Security

- **Information In Transit (IIT)**

- Availability

- Redundancy of network components (links, switches, routers)
- Path redundancy (backup paths)
- Simultaneous transmission over separated paths

- Confidentiality

- Encryption (secret key technology e.g. 3DES, AES)

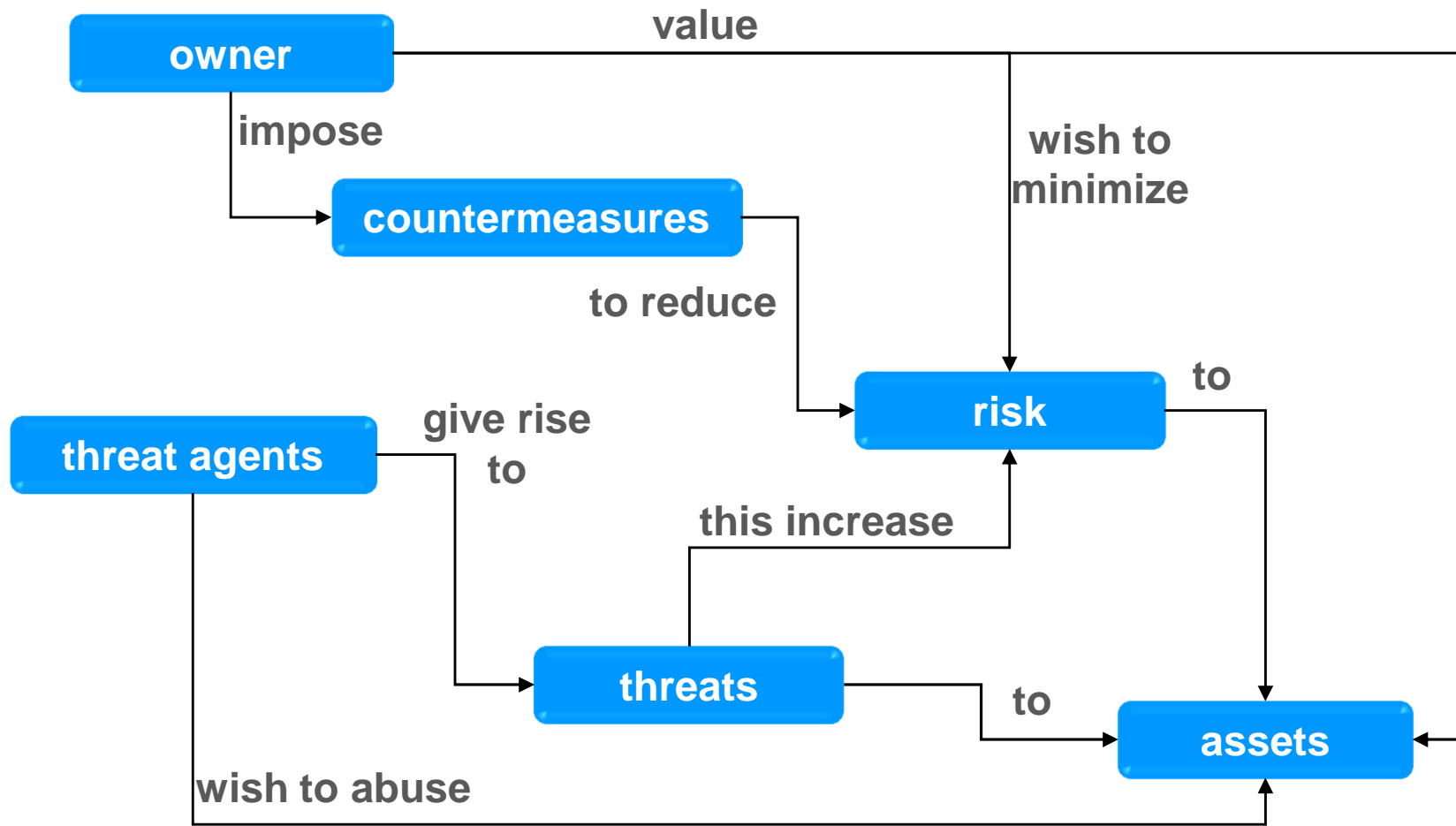
- Integrity and identity

- Cryptographic checksums (e.g. keyed MD5, keyed-SHA1)
- Digital Signature (public/private key technology e.g. RSA, certificates)

- Key management

- Keys are necessary for authentication procedures/protocols
- Keys are necessary for cryptographic operations
- Preshared key versus PKI (Public Key Infrastructure)

# The Principle Of Security Evaluation



Source CC v3.1 / 2006 

***Without any assets to be protected, there is no need for security ever!  
100% security is impossible => you need to decide, what to secure how well!***

# Security Assessment / Analysis

***Security Assessment ...assess security weaknesses in the product or system by identifying and addressing security risks in the system and in the system environment.***

consolidation

- critical devices or sensitive network connections are identified and the system is **structured in security zones**

requirements

- **confidentiality, availability, integrity**, (access control, auditing, network separation, remote access...)

assumptions

- Indicate requirements applicable only at the **customer** premises under his **responsibility**

assets

- **Information or services** be protected by the countermeasures of a system.

threats

- A potential cause of an **incident** , which may result in harm to a system or organization

assess risk

- Systematic use of information (assets, threats, assumptions, requirements) to identify and **estimate the risk**.

objectives

- Abstract statement of the **intended solution** to the security problem.

measures

- Technical, operative and procedural measures which support the objectives and **lead to protection mechanism**

remaining risk

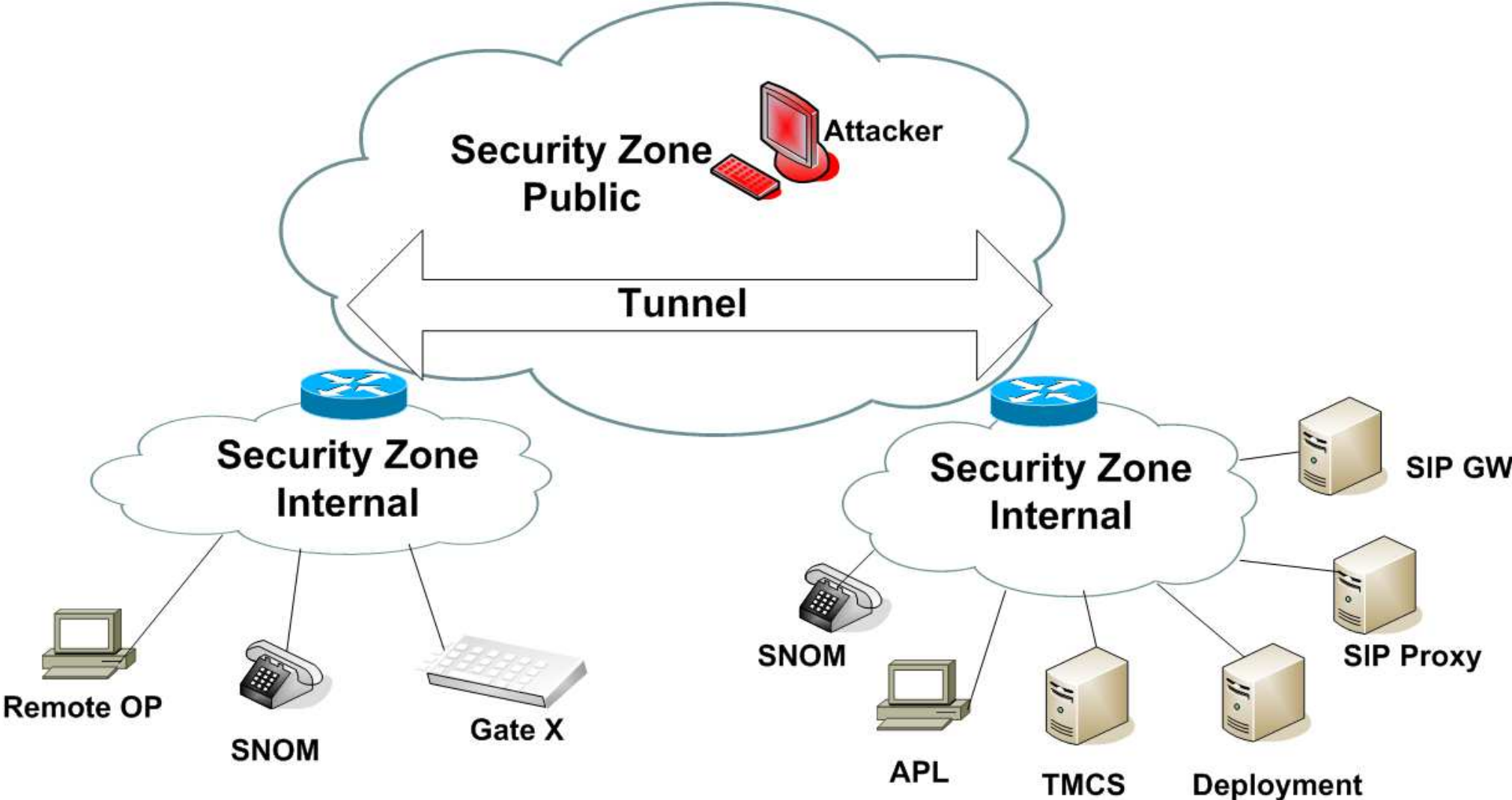
- Management of the residual risk so that the **residual security risk is tolerable** and as low as reasonably practicable.



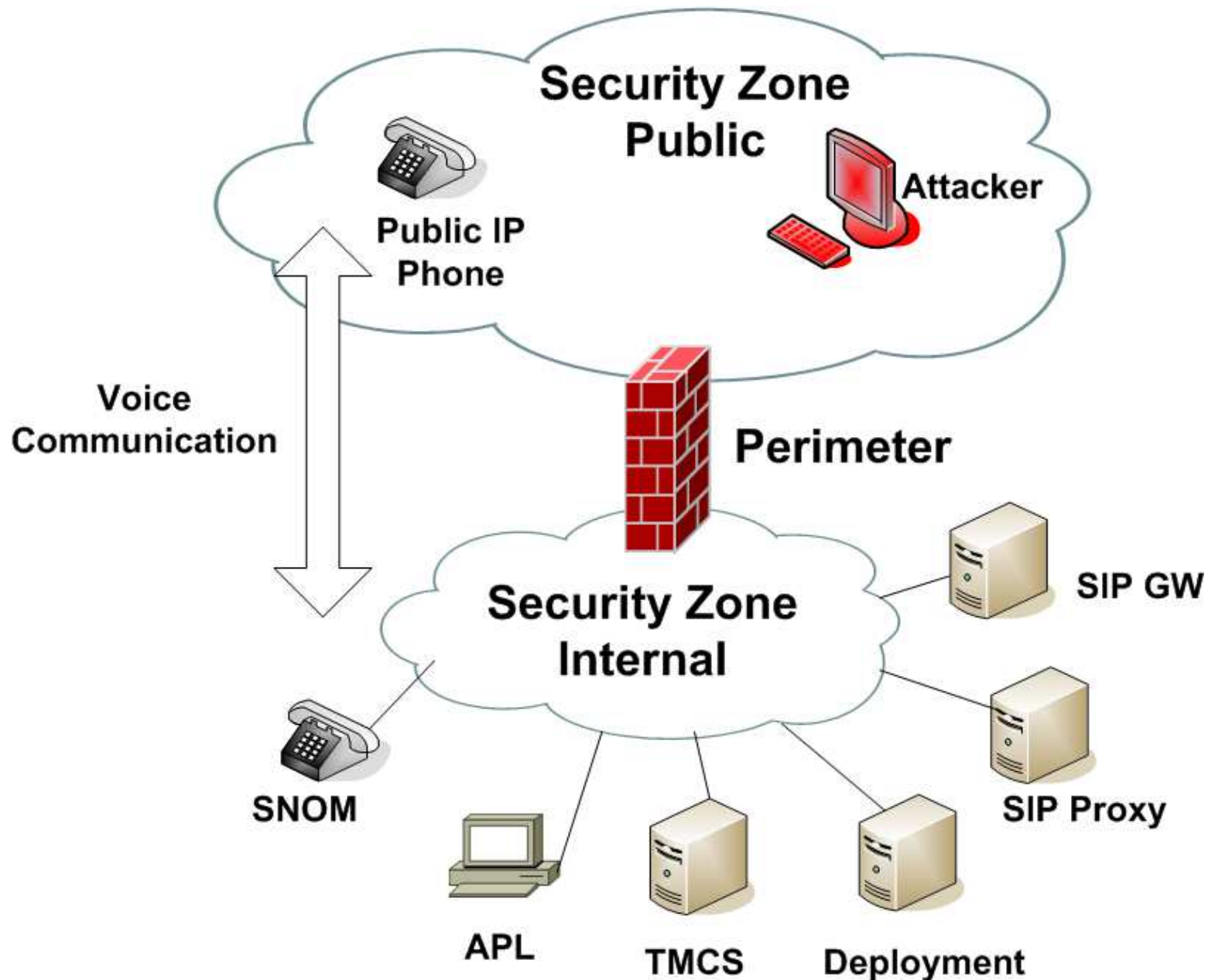
# IT-Security – Network Security Elements

- **“Security zones / domains”**
  - Definition of the environment systems or system parts are operating in
- **Summarization of assumptions about**
  - Access to system physically protected
  - Personal access to system protected by physical access control and strong authentication techniques
  - ....
- **“Multiple Barriers”**
  - In the network infrastructure and at the end system
- **Generic security function “Tunnel”**
  - System parts are in the same security zone
  - Ensures protected communication between dispersed system parts over non protected network infrastructure
  - E.g. site-to-site IPsec VPN, client-to-site IPsec VPN, SSL-VPN
- **Generic security function “Perimeter”**
  - System parts are in different security zones
  - Ensures controlled communication between systems with different functionality and authorization rights
  - E.g. firewall with stateful inspection

# Security Function Tunnel



# Security Function Perimeter



# Agenda

---

- **Introduction**
- **Network Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
  - Introduction IT-Security
  - VPN Types
  - MPLS, MPLS-VPN
  - IPsec VPN
  - DMVPN
  - GETVPN
- **Multicasting**
- **Summary**

# VPN (Virtual Private Network) Types

- **VPN != Encryption (Confidentiality and Integrity)**
- **Three basic VPN types**
  - Classical VPNs
    - Separation of traffic of different customers over a shared network infrastructure
    - Crypto-graphical support is not available
    - Non-encrypted VPNs
  - Overlay VPNs
    - Tunnelling of traffic over a given network infrastructure
    - Inherent crypto-graphical support for encryption and integrity checking is possible
    - Encrypted VPNs
  - Proxy VPNs
    - No separation of traffic of different customers
    - Optional crypto-graphical support for encryption and integrity
    - Encrypted VPNs

# Classical VPNs

- **Legacy techniques:**

- **X.25 or Frame Relay PVCs (L2-VPN):**

- Multiplexing of virtual circuits across a shared X.25 or FR packet switching infrastructure

- **X.25 or Frame Relay SVCs with closed user group feature (L2-VPN):**

- Multiplexing of virtual circuits across a public X.25 or FR packet switching infrastructure

- **ISDN with closed user group feature (L2-VPN):**

- Multiplexing of virtual circuits across a public ISDN circuit switching infrastructure (TDM)

- **Current techniques:**

- **VLAN (L2-VPN):**

- Multiplexing of LANs across a shared L2 Ethernet switching infrastructure

- **MPLS-VPN (L3-VPN):**

- Multiplexing of IP nets across a shared L3 IP/MPLS infrastructure

- **Pseudowire: (L2-VPN):**

- Transporting a wire (Frame-Relay, ATM, Ethernet) using L2TPv3 or ATOM (MPLS)
- Carrier Ethernet

- **VPLS (Virtual Private LAN Service; L2-VPN):**

- Multiport Ethernet bridging across a MPLS backbone

# Overlay VPNs

- **GRE (Generic Route Encapsulation)**
  - Old technique often used in the Internet for transporting multiprotocol traffic (e.g. IPv4 multicast, IPv6 unicast or IPX-Novell) over an IPv4 unicast-only backbone
  - No encryption support but multicast is possible
- **IPsec VPN**
  - Site-to-site VPN between VPN gateways or client-to-site VPN between an end-system with VPN-client-SW and a VPN concentrator
  - Point-to-point security associations
  - Currently for unicast only, scalability problem for full mesh
- **SSL VPN**
  - Alternative to IPsec client-to-site VPNs
  - Originally based on HTTP over SSL
- **DMVPN (Dynamic Multipoint VPN)**
  - Cisco implementation for large scale IPsec VPN
  - Combines mGRE, dynamic NHRP/NHS and IPsec protection
  - Multicast support is possible but could be suboptimal (Hub and Spoke)

# Proxy VPNs / Alternate VPNs

- **GETVPN (Group Encrypted Transport VPN)**
  - Cisco implementation
  - Point-to-multipoint security associations using group keys, tunnel less technology
  - Multicast possible if backbone supports it
- **LISP (Locator / Identifier Separation Protocol)**
  - Cisco novel approach for separation of identity (“Who I am”, EID address space) from location (“Where I am”, RLOC address space)
    - Identity and location is normally represented by a just single IP address
  - Network based solution
    - Available already in Cisco IOS and NX-OS
  - Open specifications and implementations
    - Experimental RFCs 6830 - 6836
    - OpenLISP (open-source for FreeBSD)
    - LISP mobile node (open-source for Linux and Android)
  - Base VPN behavior
    - By separating EIDs from RLOCs of IP WAN service provider
  - Encrypted VPN possible
    - By combining LISP with GETVPN



# Agenda

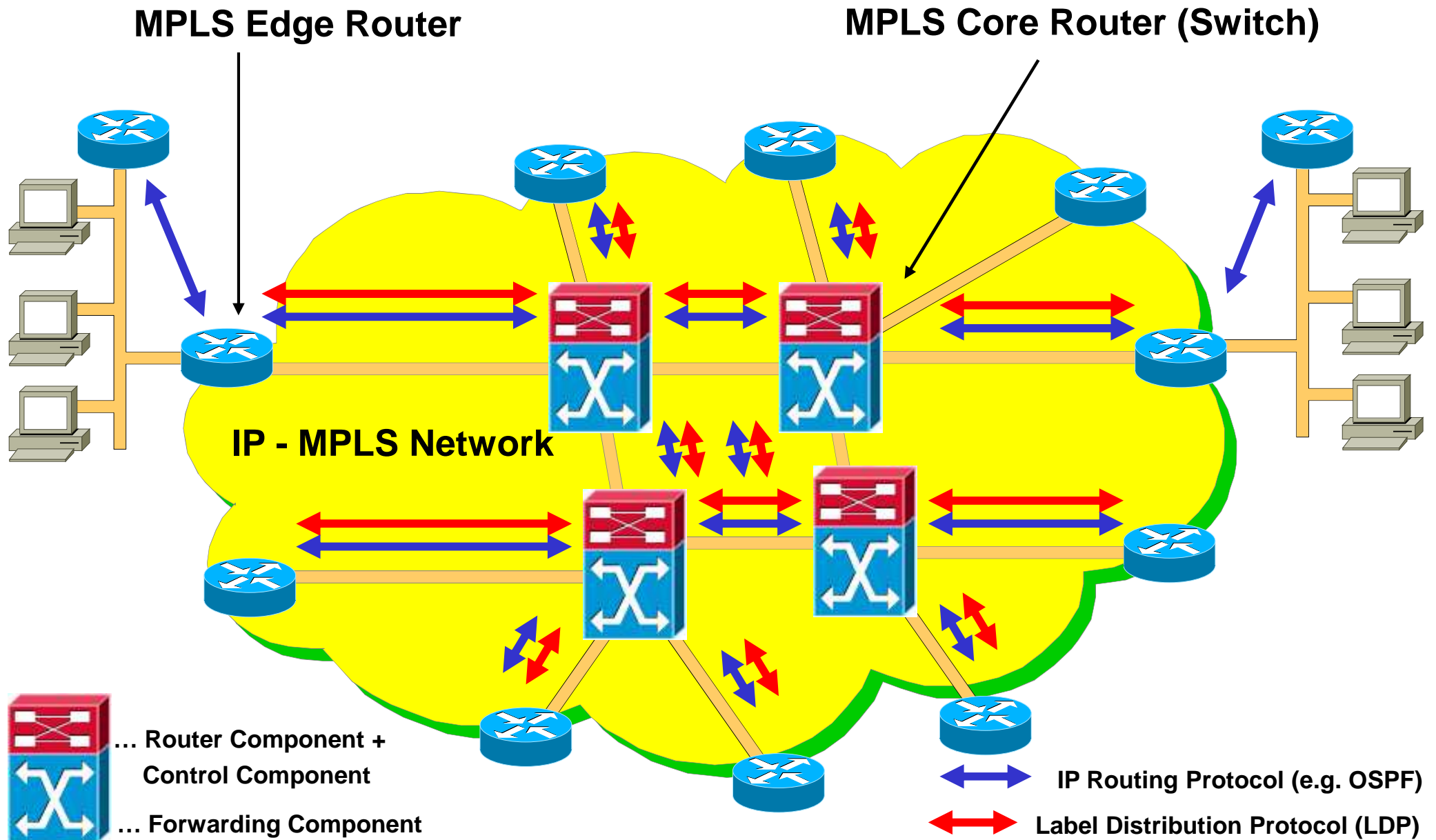
---

- **Introduction**
- **Network Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
  - Introduction IT-Security
  - VPN Types
  - MPLS, MPLS-VPN
  - IPsec VPN
  - DMVPN
  - GETVPN
- **Multicasting**
- **Summary**

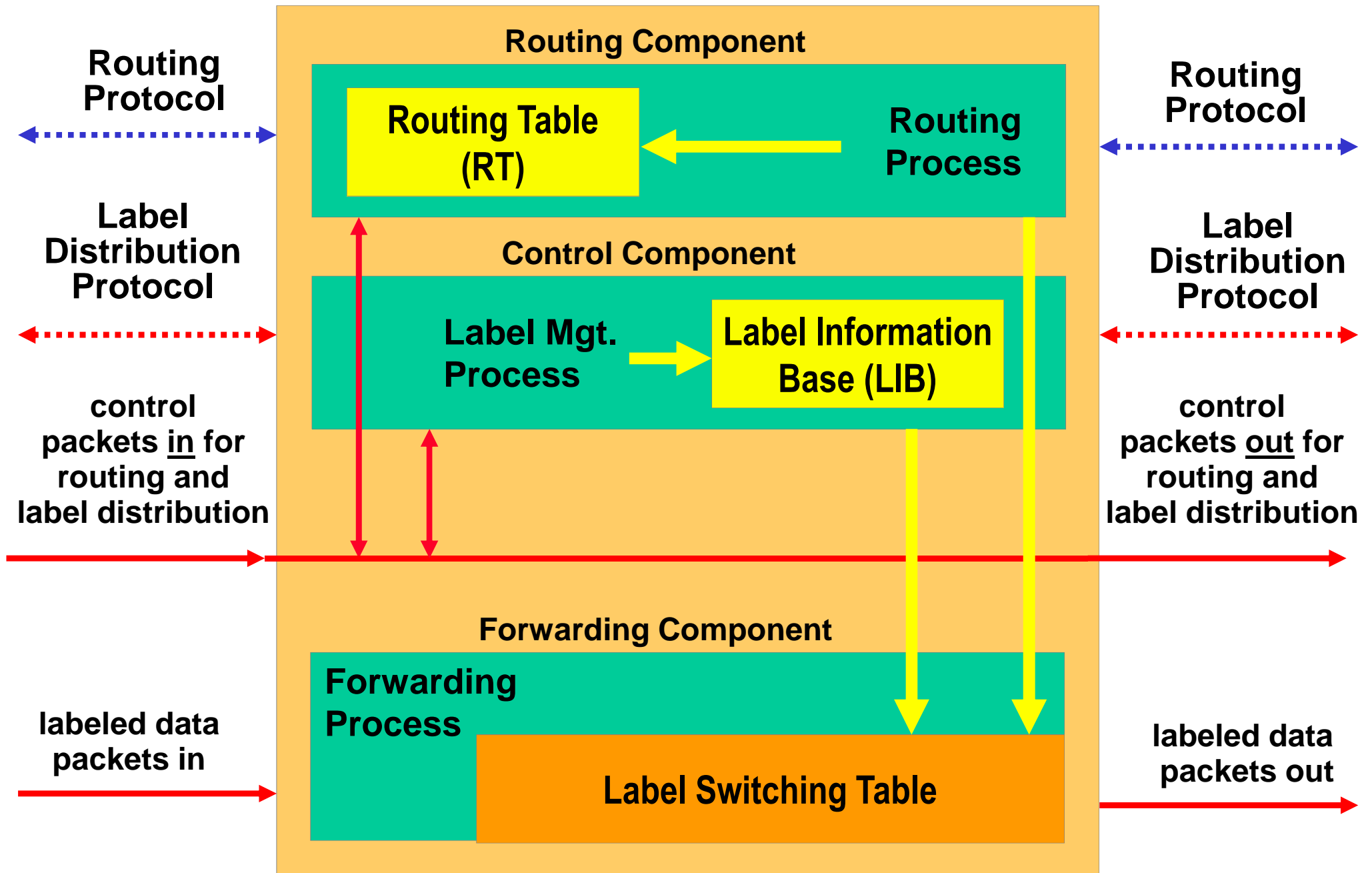
# MPLS Principle

- **Traditional IP uses the same information for**
  - Path determination (routing)
  - Packet forwarding (switching)
- **MPLS separates the tasks**
  - L3 addresses used for path determination
  - Labels used for switching
- **MPLS network consists of**
  - MPLS edge routers and MPLS core routers
- **Edge routers and core routers**
  - Exchange routing information about L3 IP networks using classical IP routing protocols (OSPF, IS-IS)
  - Exchange forwarding information about the actual usage of labels using label distribution protocol (LDP)

# MPLS Network

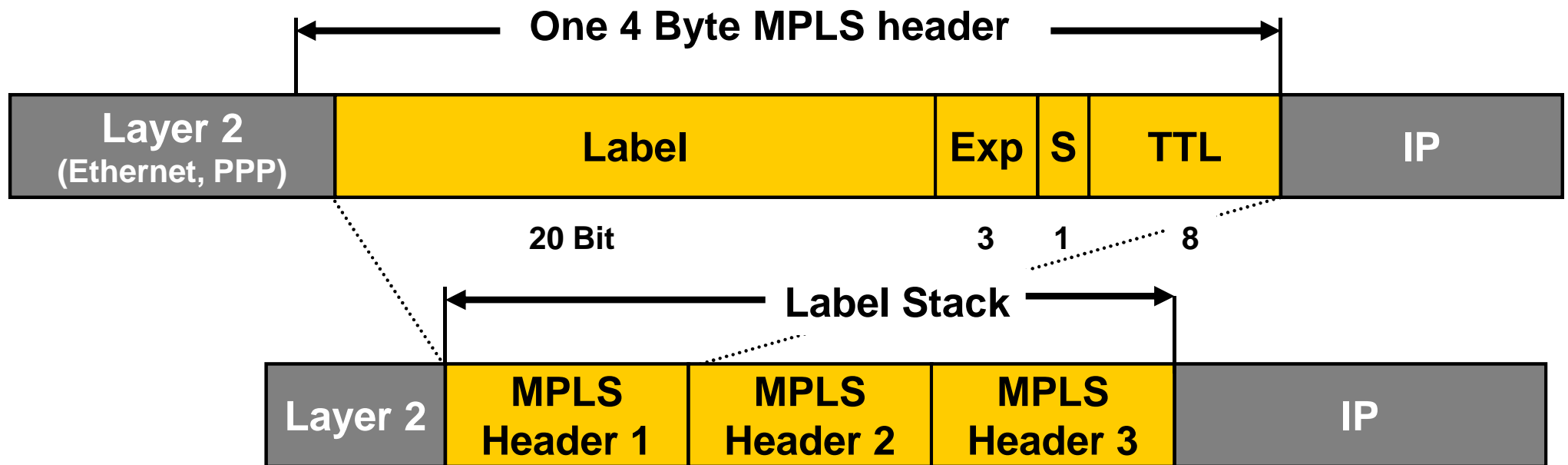


# MPLS Router Internals



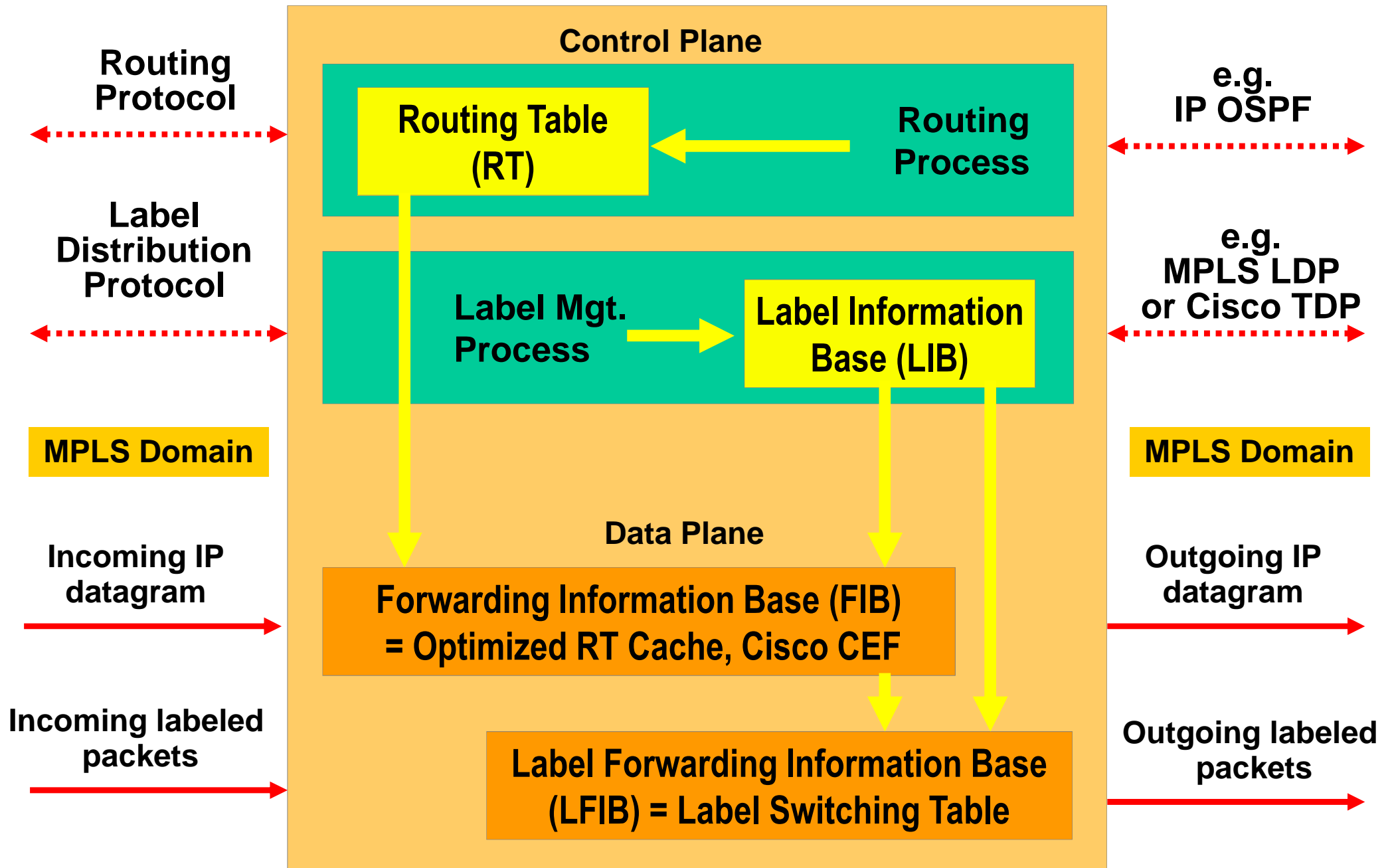


# MPLS Header

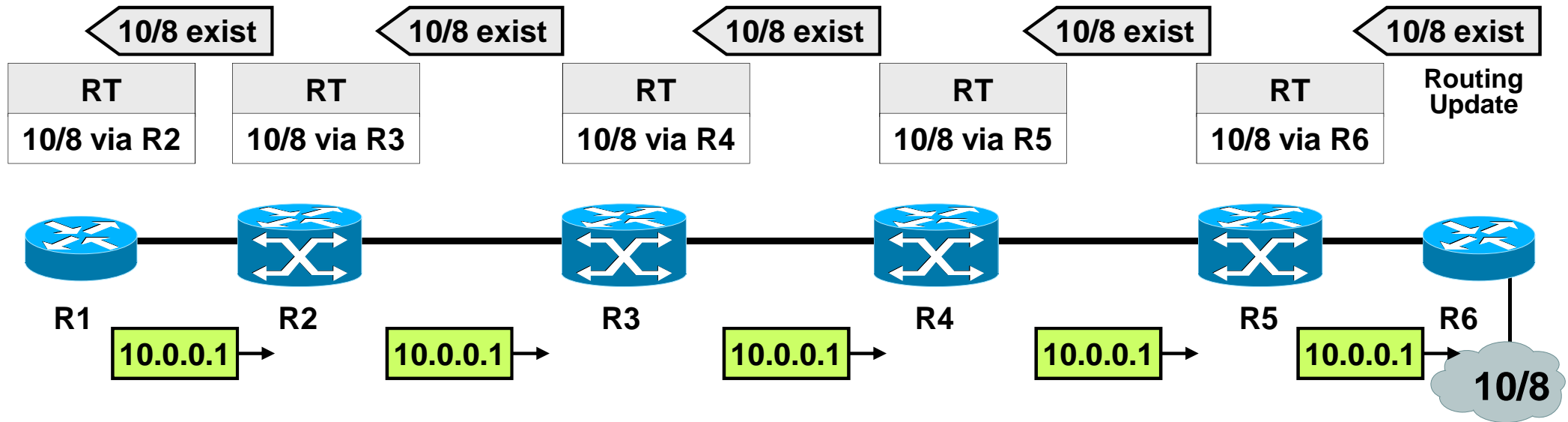


- **20-bit MPLS label (Label-Bits)**
- **3-bit experimental field (Exp-Bits)**
  - Could be copy of IP Precedence -> MPLS QoS like IP QoS with DiffServ Model based on DSCP
- **1-bit bottom-of-stack indicator (S)**
  - Labels could be stacked (Push & Pop)
  - MPLS switching performed always on the first label of the stack
- **8-bit time-to-live field (TTL)**

# MPLS Router Internals (Cisco)

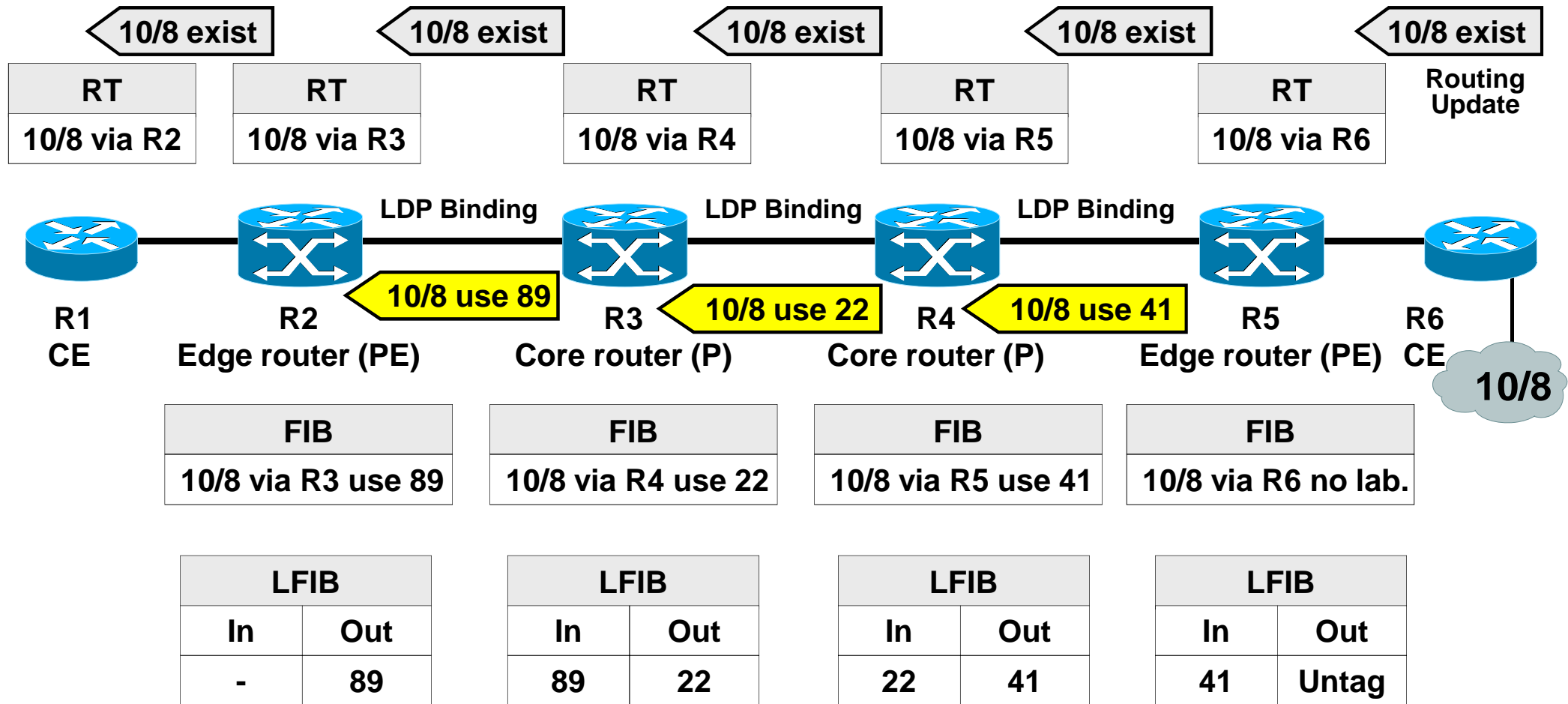


# Classical IP Forwarding: Hop by Hop Forwarding



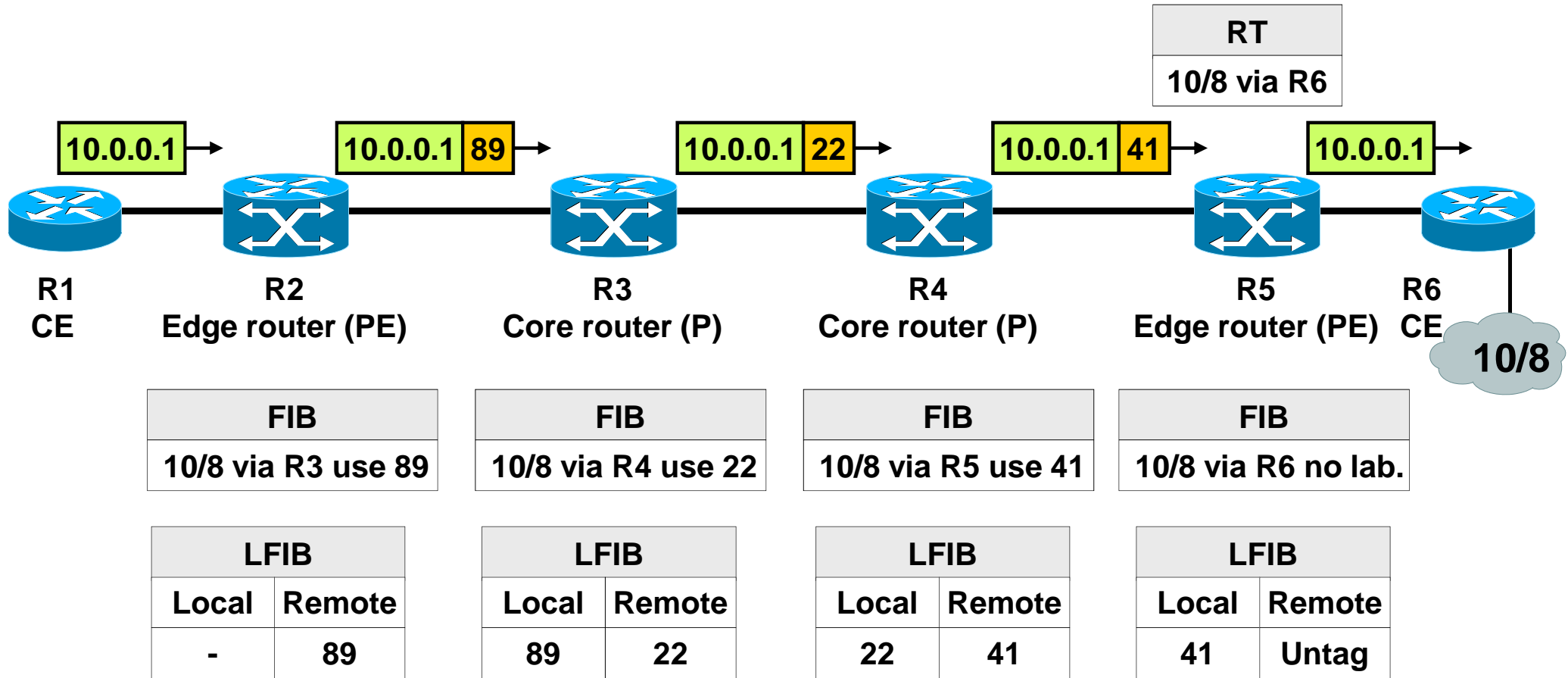


# MPLS Switching In Action: Label Distribution

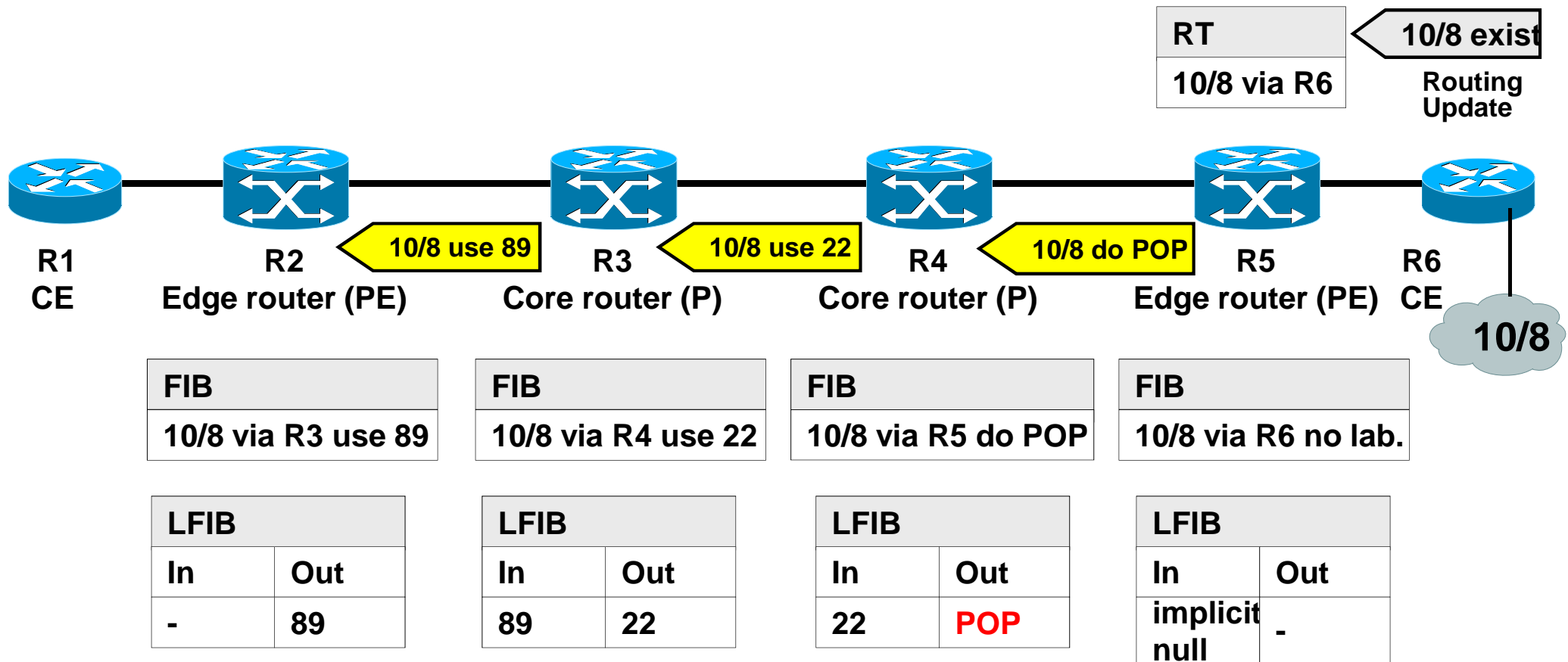


- Both routing updates and LDP/TDP distribute reachability information
- “in” = local label created by the router itself and advertized
- “out” = remote label received from other routers

# MPLS Switching In Action: Label Swapping

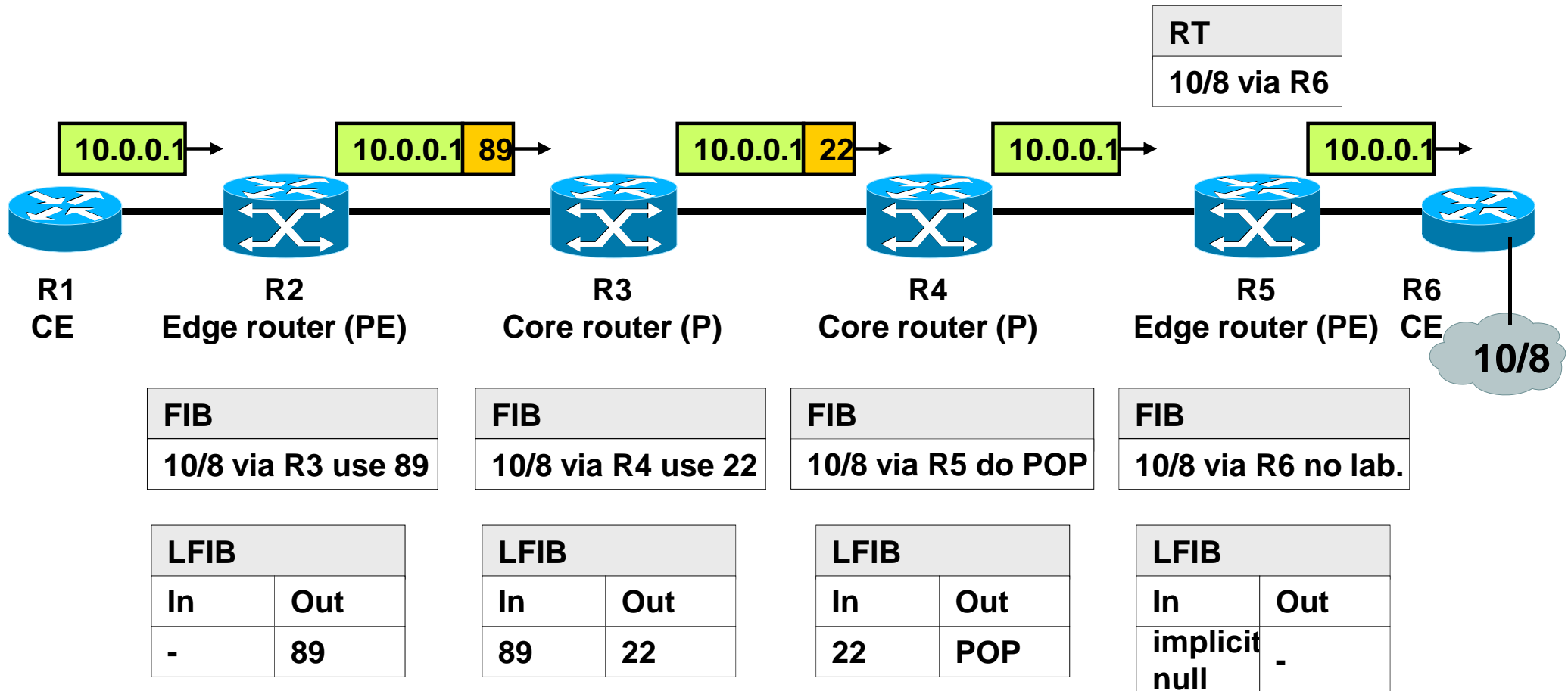


# MPLS Switching In Action: Penultimate Hop Popping



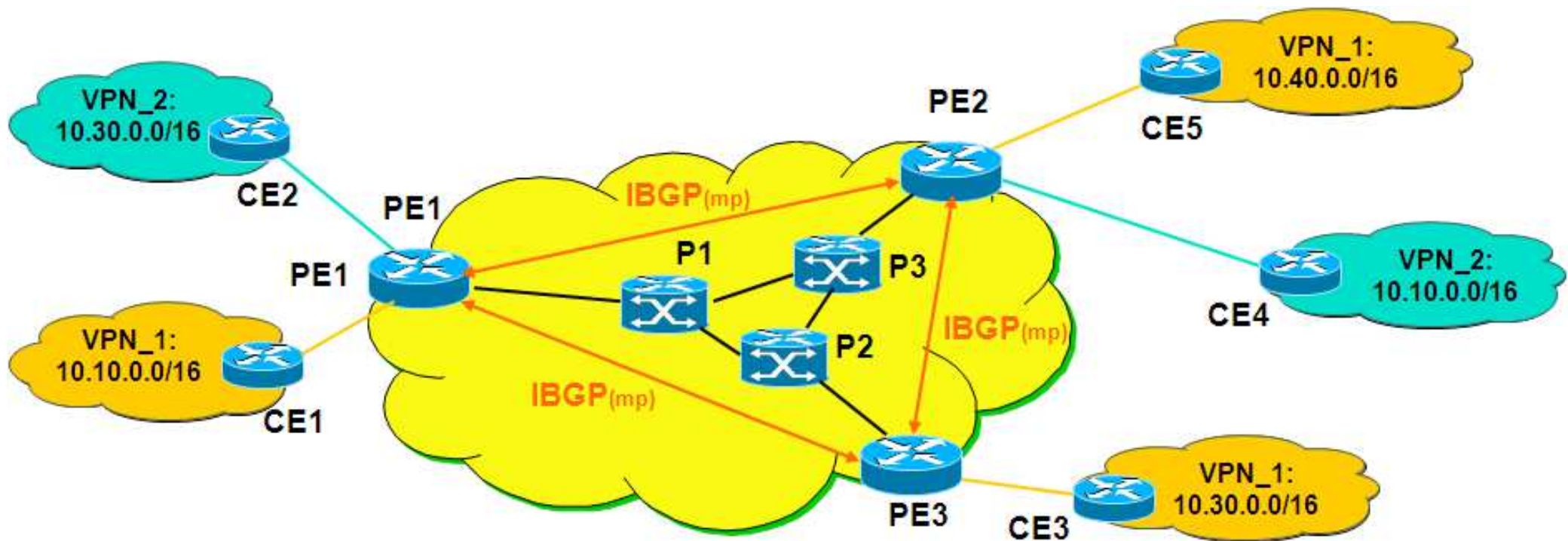
- Last hop router (R5) tells penultimate router (R4) to remove label
  - "Penultimate Hop Popping" (PHP)
  - Also called "Implicit Null Label"

# MPLS Switching In Action: Penultimate Hop Popping



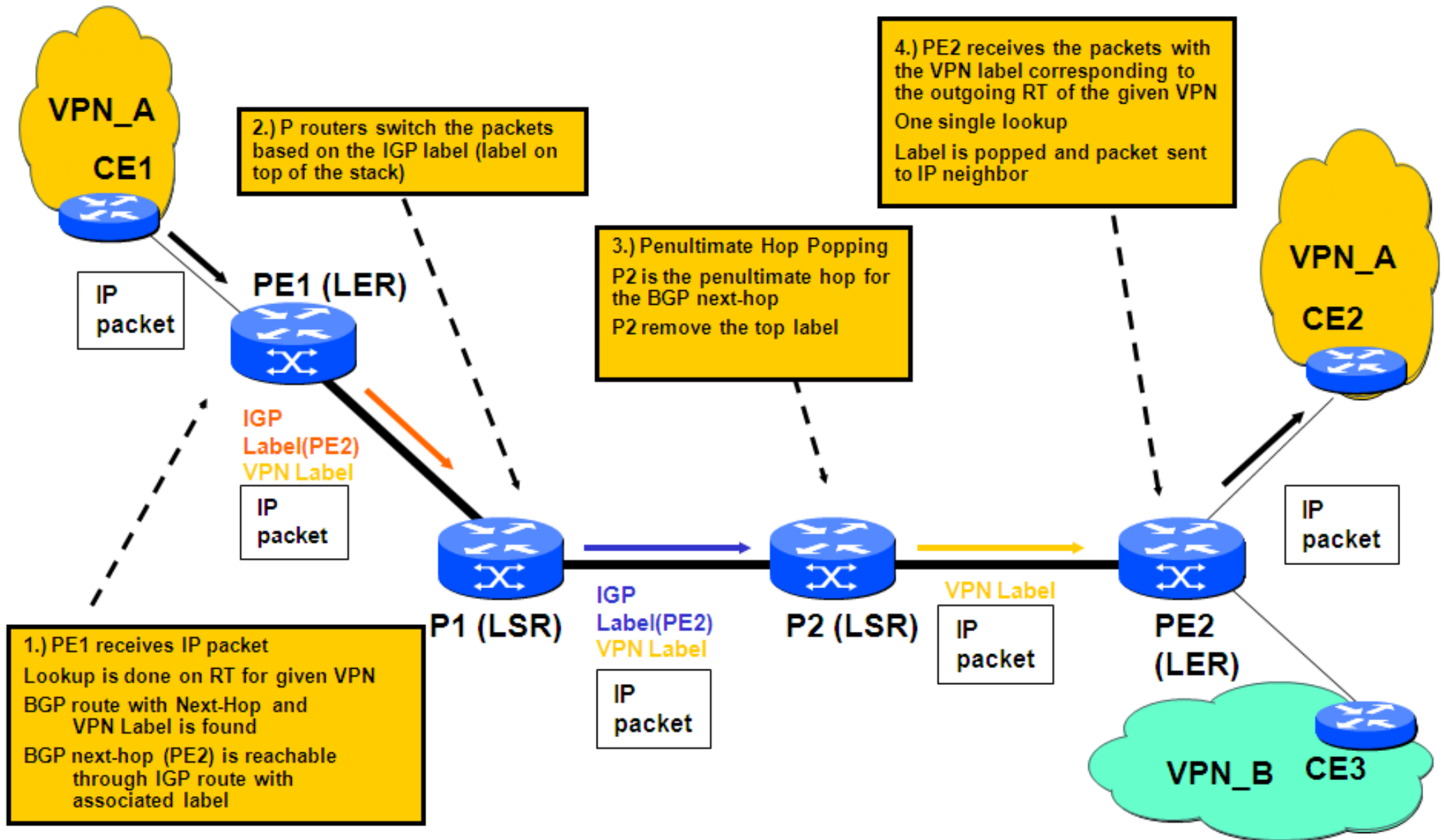
- R5 only performs single lookup in FIB

# MPLS VPN Architecture



- **Service provider offers MPLS-VPN based on internal MPLS switching infrastructure**
  - PE ... provider edge MPLS edge router, P ... provider internal MPLS core router
  - CE ... customer edge, conventional, IP router
- **MPLS-VPN requires full mesh of internal multiprotocol (mp) BGP sessions**
  - Could lead to a scalability problem in large environments
- **Customers receiving an IP VPN service**
  - Customer "Orange" and "Green"
  - Each customer has its own IP address space (VPN-1 or VPN\_2) which is separated by MPLS-VPN
  - Address may overlap

# MPLS VPN In Action Using MPLS Labelstack

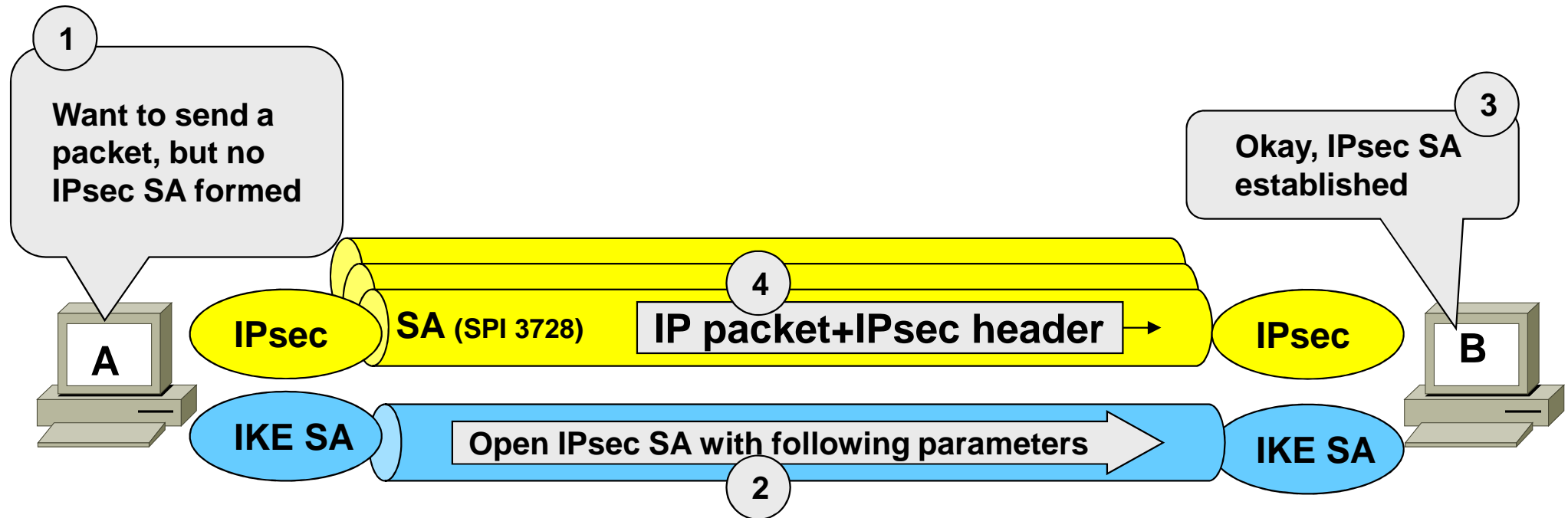


# Agenda

---

- **Introduction**
- **Network Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
  - Introduction IT-Security
  - VPN Types
  - MPLS, MPLS-VPN
  - IPsec VPN
  - DMVPN
  - GETVPN
- **Multicasting**
- **Summary**

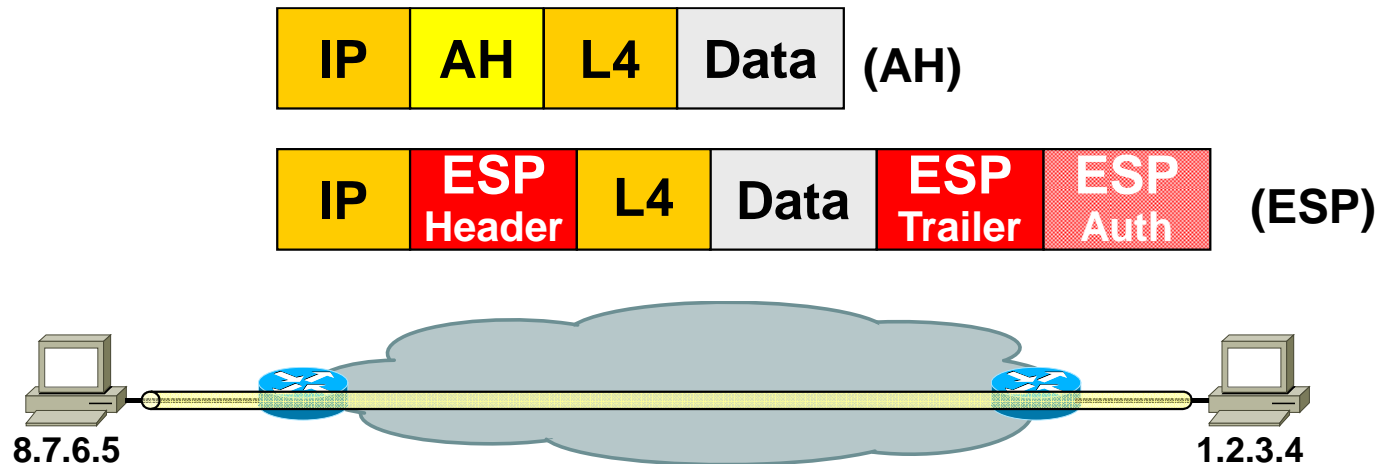
# Security Association (SA) Internet Key Exchange (IKE)



- **IKE SA (bidirectional control channel)**
  - Establishes an authenticated and encrypted and integrity protected tunnel (blue pipe)
  - Authentication based on security credentials like pre-shared secret, public signature key, public key encryption techniques
  - Used for securely establishing IPsec SAs and the initial key material valid for a certain lifetime
- **IPsec SAs (unidirectional data channel)**
  - Are created on demand (yellow pipes)
  - Rekeying is done again by usage of IKE before lifetime exceeds



# IPsec Transport Mode

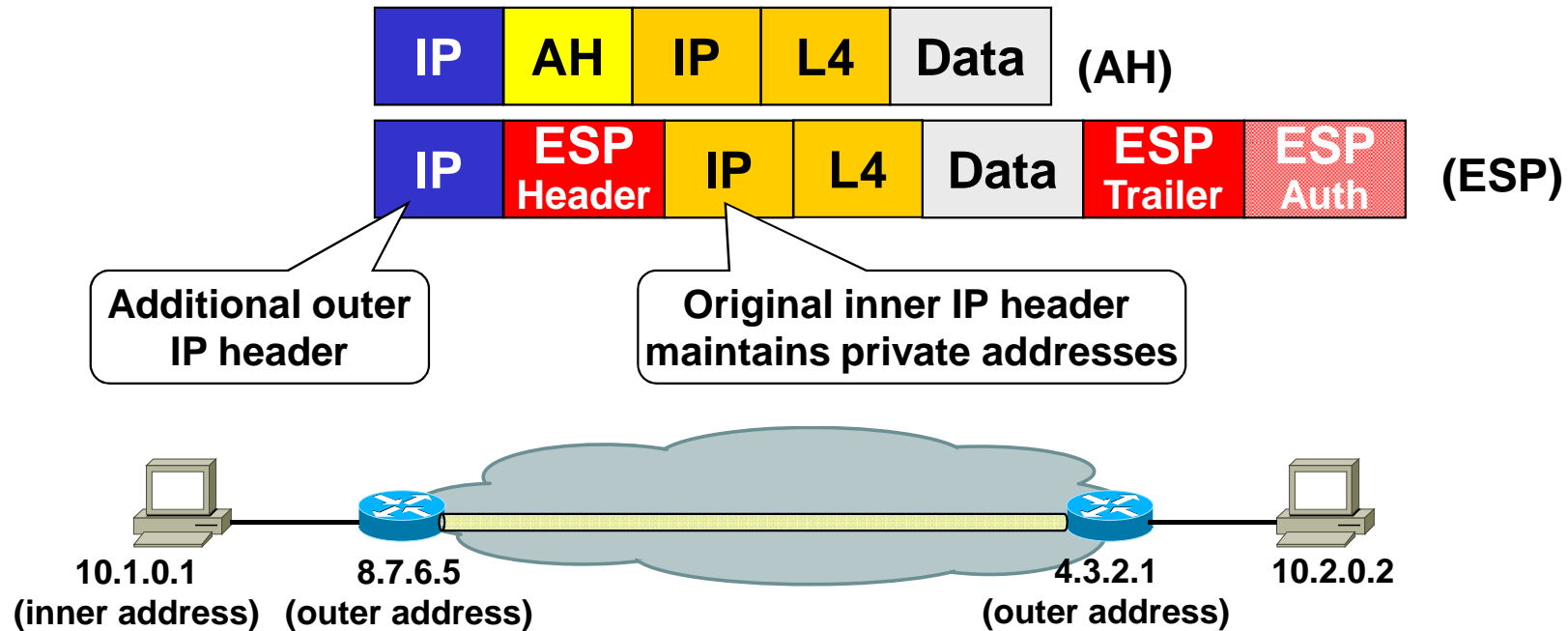


- **IPsec headers**

- AH and ESP Auth Header for integrity protection (crypto fingerprints)
- ESP for privacy protection (encryption)

- **Used for end-to-end sessions**

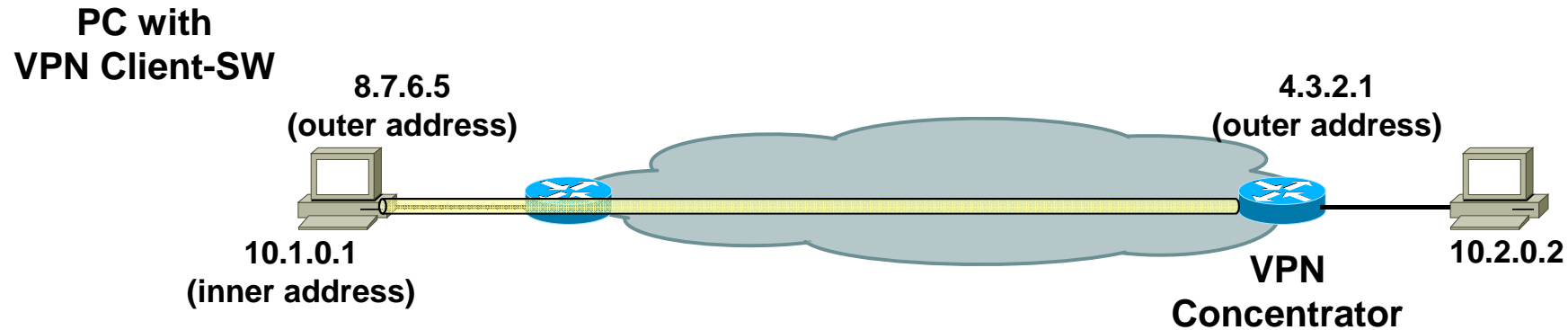
- Does not hide communication statistics because of network header containing IP addresses of the end systems is sent in clear



- **Used for site-to-site VPN**

- Between security gateways like firewalls, routers with IPsec support, VPN concentrators
- Does hide communication statistics because original IP packet is IPsec encapsulated

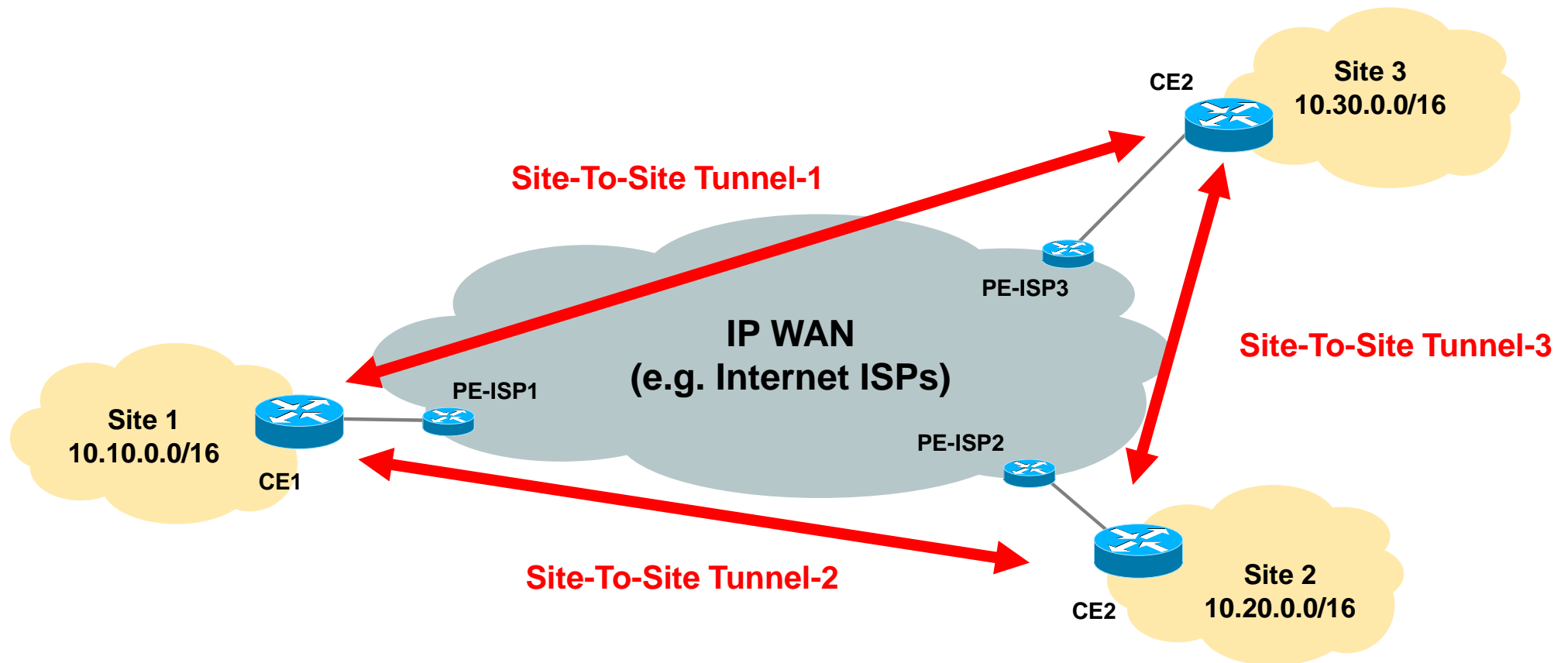
## Tunnel Mode for Client-to-Site VPN



- **Used for client-to-site VPN**

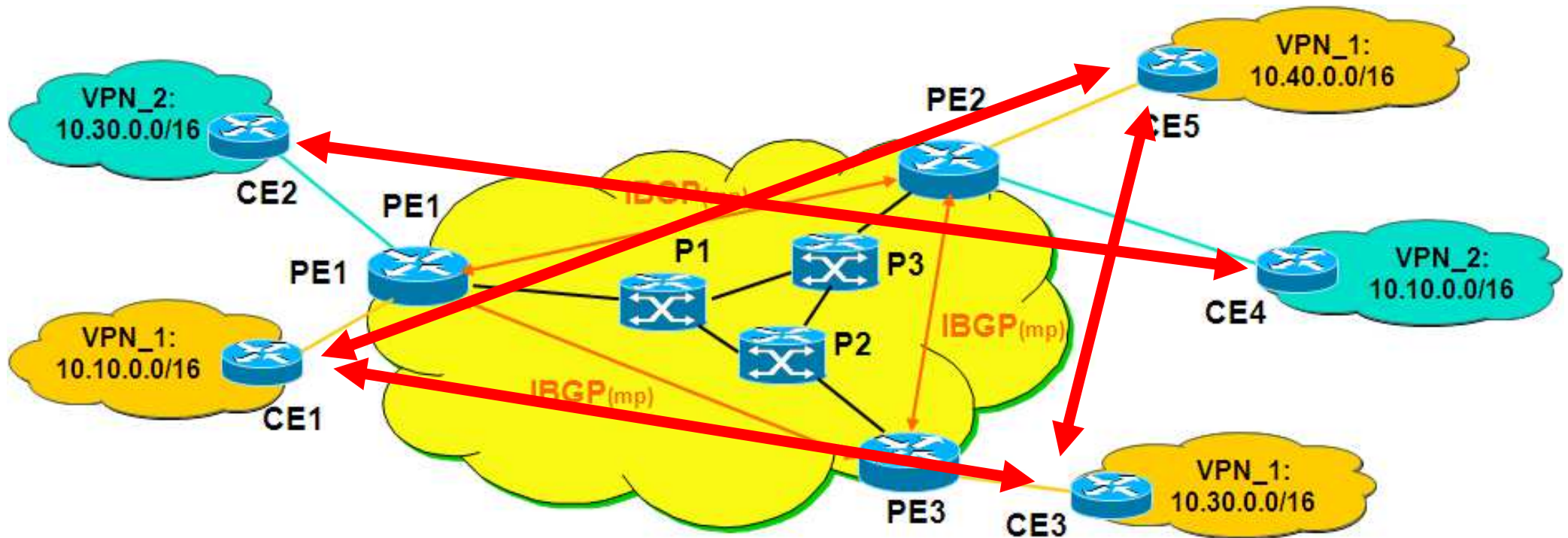
- Between PC client with VPN Dial-In software and VPN concentrators
- Does hide communication statistics because original IP packet is IPsec encapsulated

# IPsec Site-To-Site VPN Scalability

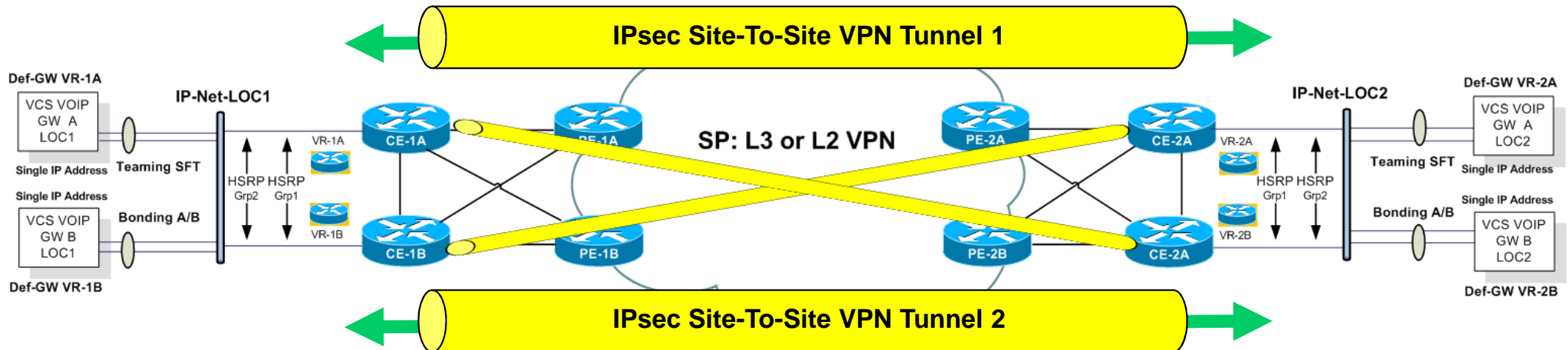


- **Because of point-to-point behavior of IPsec SAs**
  - IPsec site-to-site VPN requires full mesh of IPsec tunnels
  - That causes a scalability problem in large environments

# Combining MPLS-VPN And IPsec-VPN




- **MPLS-VPN requires a full mesh of internal multiprotocol (mp) BGP sessions**
- **IPsec site-to-site VPN requires a full mesh of IPsec tunnels**



- **IPsec Management**

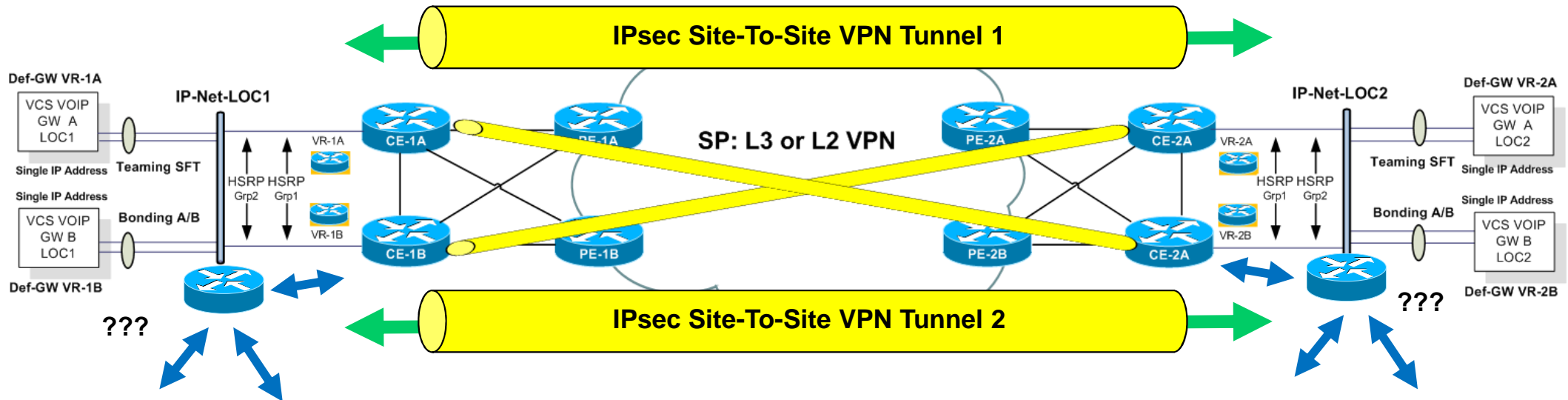
- Task of the customer on the CE routers

- **Traffic between IP-Net-LOC1 and IP-Net-LOC2** ↔

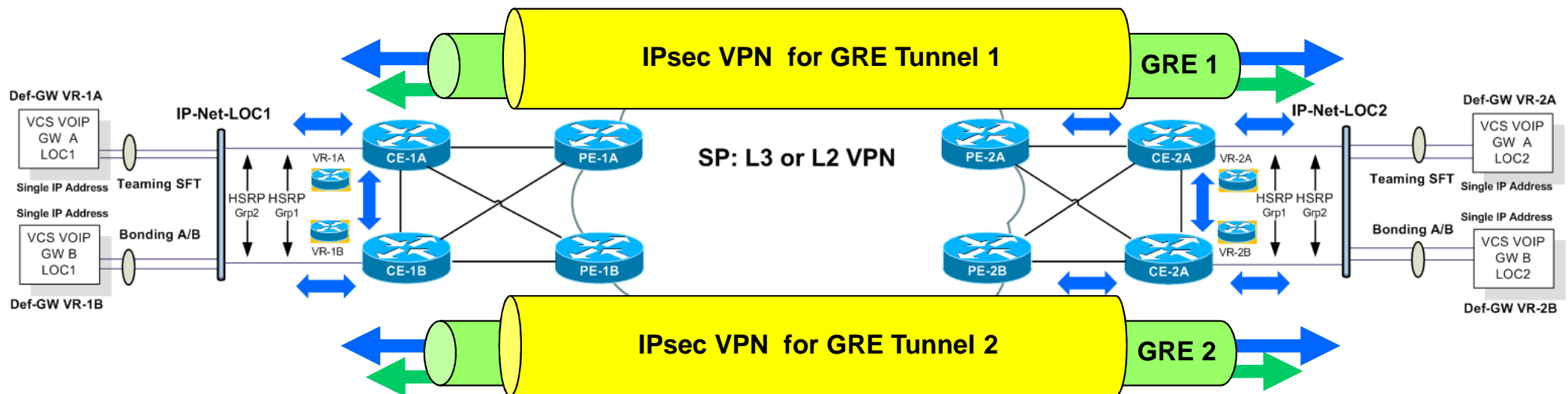
- Will be protected by IPsec site-site VPN (tunnel-mode) from CE-1A to CE-2A 
- Other tunnels on picture above for redundancy
- Interesting traffic (= to be encrypted traffic) has to be specified (worst case: every net-id combination)

- **Attention: IPsec is a kind of “Dial-Up” technique**

- Set-up of an IPsec tunnel is triggered by “interesting traffic”
- Hence the problem of set-up delay of so far not used tunnels arises in case of a failure



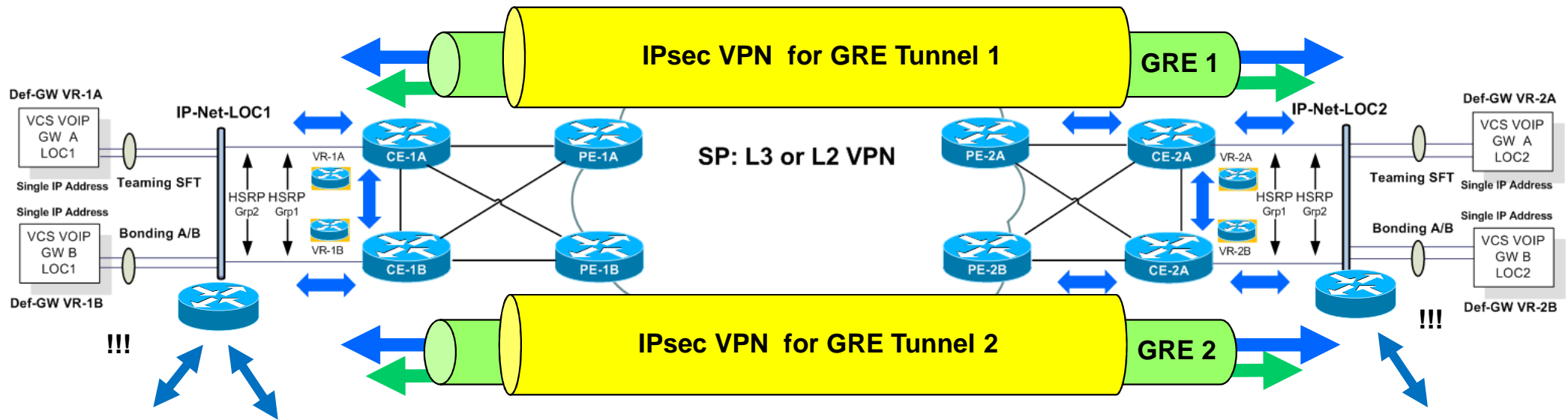
- **Static IP Routing between customer sites only**
  - IPsec can not transport multicast (broadcast) routing messages
  - Routes advertised by an internal router need special treatment at the CE routers and loses the IP dynamic routing style ↔
- **Scalability problem if many sites have to communicate without a “Hub and Spoke” style**
  - Full mesh of tunnels is necessary  $[n * (n-1) / 2]$
  - Administration and router performance is the challenge
- **Bandwidth requirements especially for small packets (e.g. VOIP) are higher than without security**
  - Double IP headers plus IPsec headers



- **GRE in combination with IPsec (transport mode)**

- Solves the problem of routing (now we have end-to-end routing)
  - Note: GRE can transport multicast (limited broadcast) routing messages
- Solves the problem of set-up delay
  - Routing messages act as keepalive for IPsec tunnels hence IPsec tunnels will not timeout during periods with no user traffic
- Eases management of IPsec tunnels
  - IPsec tunnel endpoints are the GRE tunnel addresses but not all the possible networks behind a site (interesting traffic identified by GRE tunnel addresses only)
- Additionally solves the problem of transport of multicast traffic
- But does not scale in large environments (many location, fully meshed tunnels)





## ● Dynamic Routing in the Overlay Network

- Eases management of routing
  - No static routes necessary
  - Full view of all sites and their networks in the overlay network
  - Service provider independent routing
- Can improve routing convergence
  - Even if the routing convergence of the SP provider is too slow for the applications of the customer
  - By tuning routing parameters of the end-to.-end overlay routing protocol

# Agenda

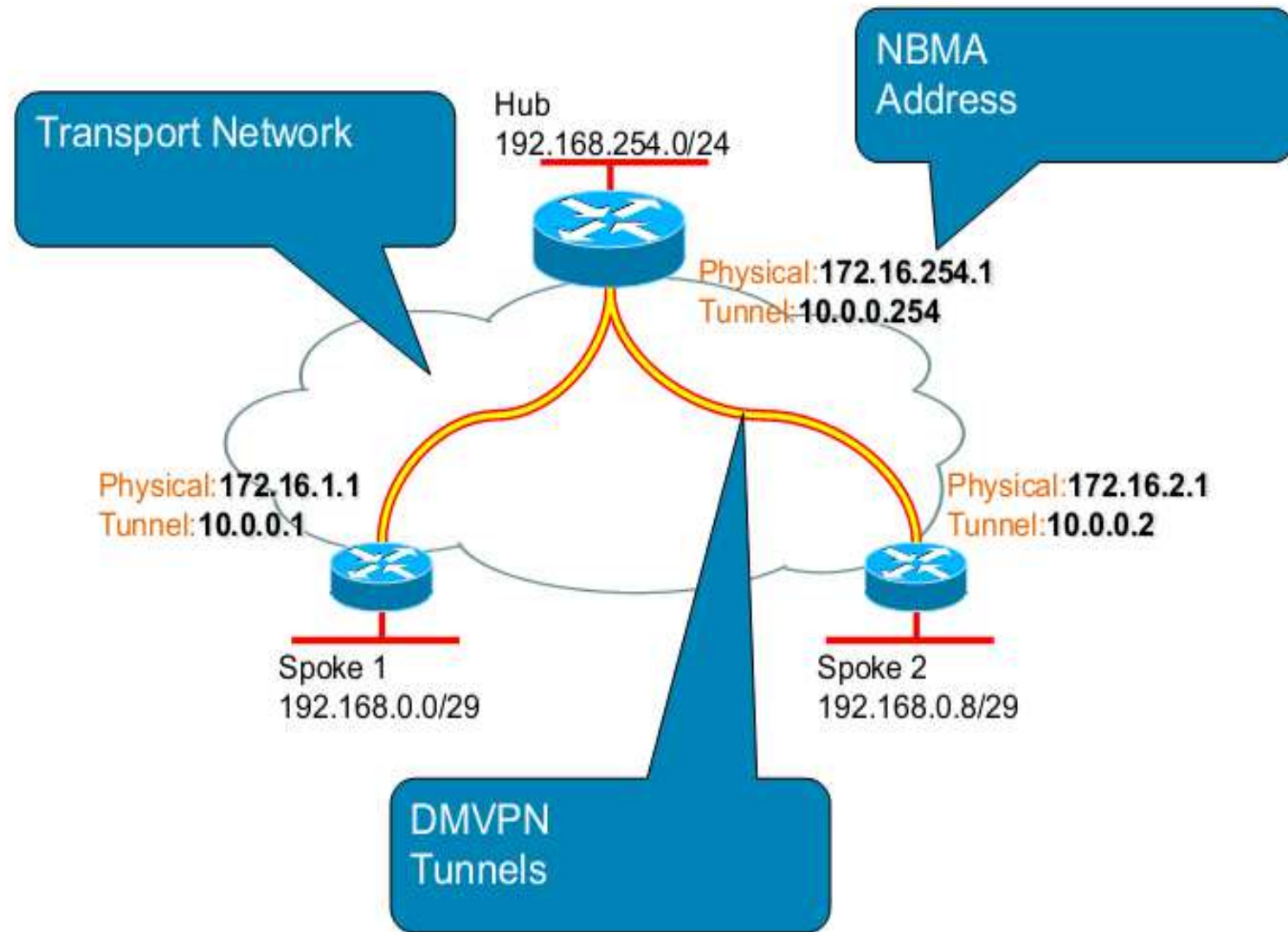
---

- **Introduction**
- **Network Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
  - Introduction IT-Security
  - VPN Types
  - MPLS, MPLS-VPN
  - IPsec VPN
  - DMVPN
  - GETVPN
- **Multicasting**
- **Summary**

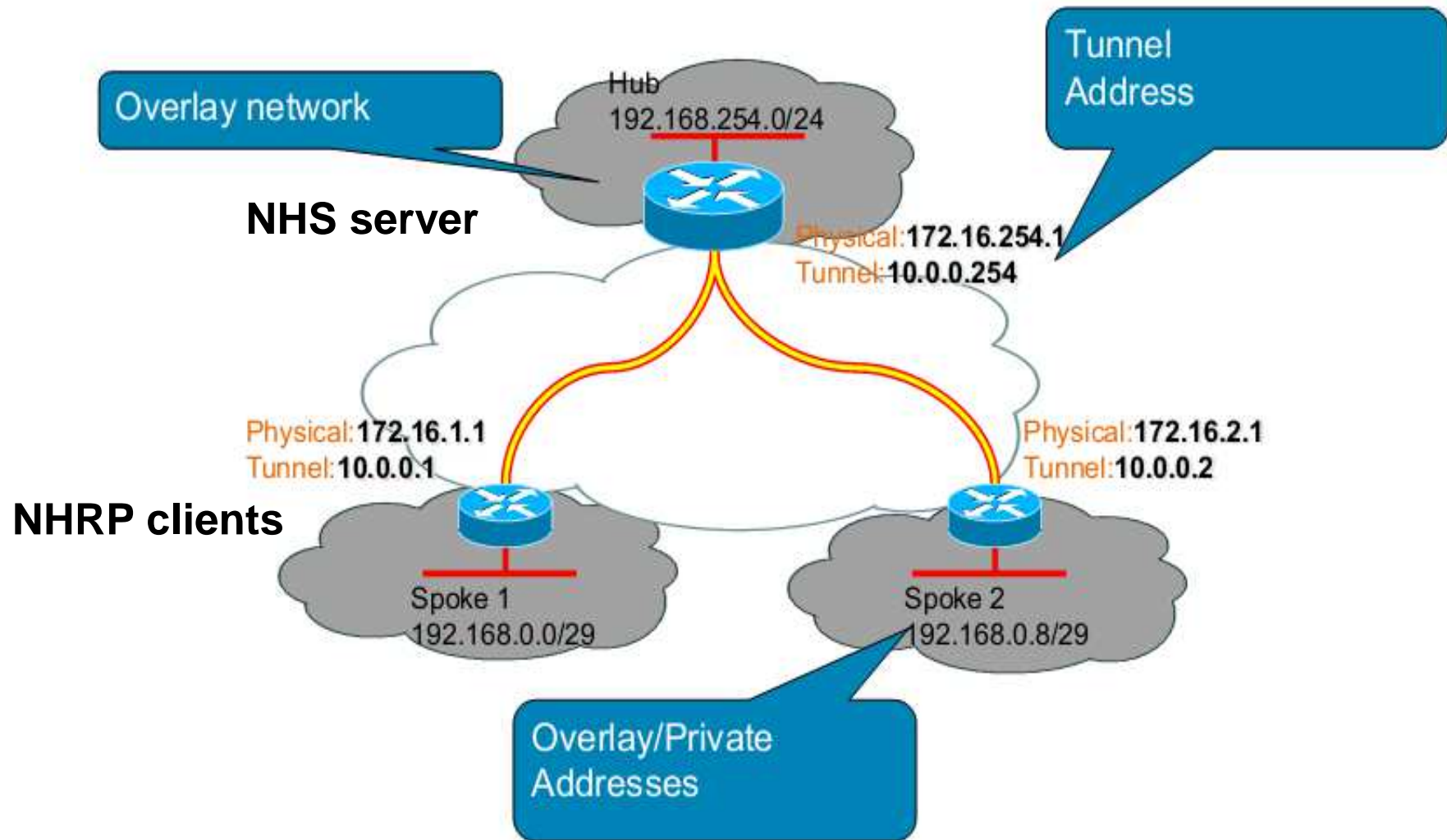
# Overlay VPN - DMVPN

- **Basic IPsec or IPsec+GRE**
  - Sufficient if you have to cover a small number of sites with these techniques
  - Maybe acceptable for larger number of sites if applications on sites requires a “Hub and Spoke” communication style
  - Not scalable for a large number of sites or “Any To Any” communication style
- **DMVPN (Dynamic Multipoint VPN)**
  - Serves large scale IPsec VPNs with overlay IP routing between sites
  - Combines IPsec protection with GRE and NHRP/NHS (Next Hop Resolution Protocol / Next Hop Server); NHS located at the hub site
  - IPsec tunnels between hub and spokes are activated automatically
  - IPsec tunnels between spokes are activated on demand and ceases after interesting unicast traffic has gone (DMVPN Phase 2)
  - Multicast replication is possible on hub site only
  - Configuration for multi-homed sites, redundant hubs and redundant service providers could be tricky and complex
  - Two independent convergence processes
    - Overlay routing and NHRP

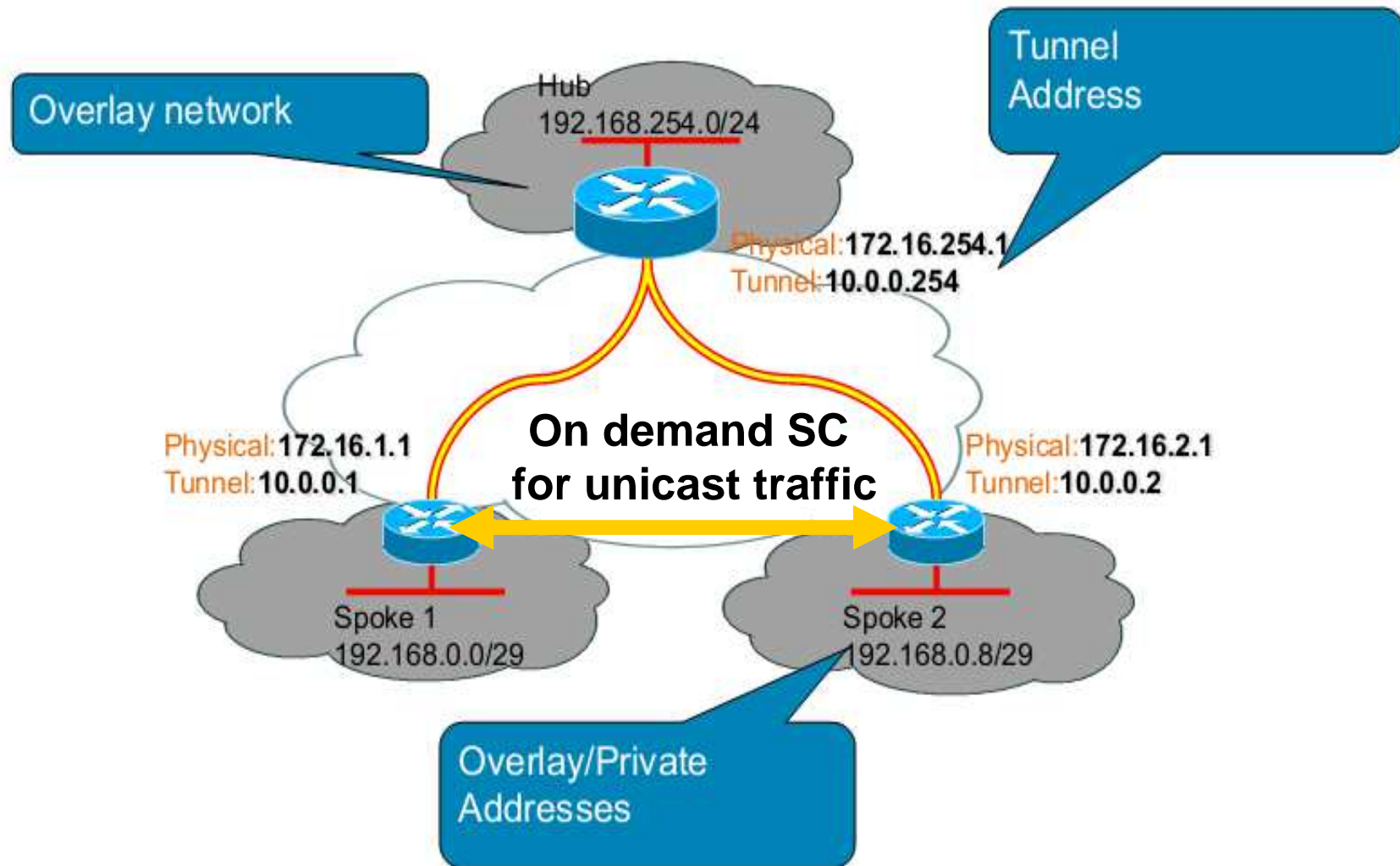
# DMVPN - Transport Network Aspects



# DMVPN - Overlay Network Aspects



# DMVPN Shortcut (SC) for Spoke to Spoke



# Agenda

---

- **Introduction**
- **Network Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
  - Introduction IT-Security
  - VPN Types
  - MPLS, MPLS-VPN
  - IPsec VPN
  - DMVPN
  - GETVPN
- **Multicasting**
- **Summary**



# Proxy VPN - GETVPN

- **IPsec technology**

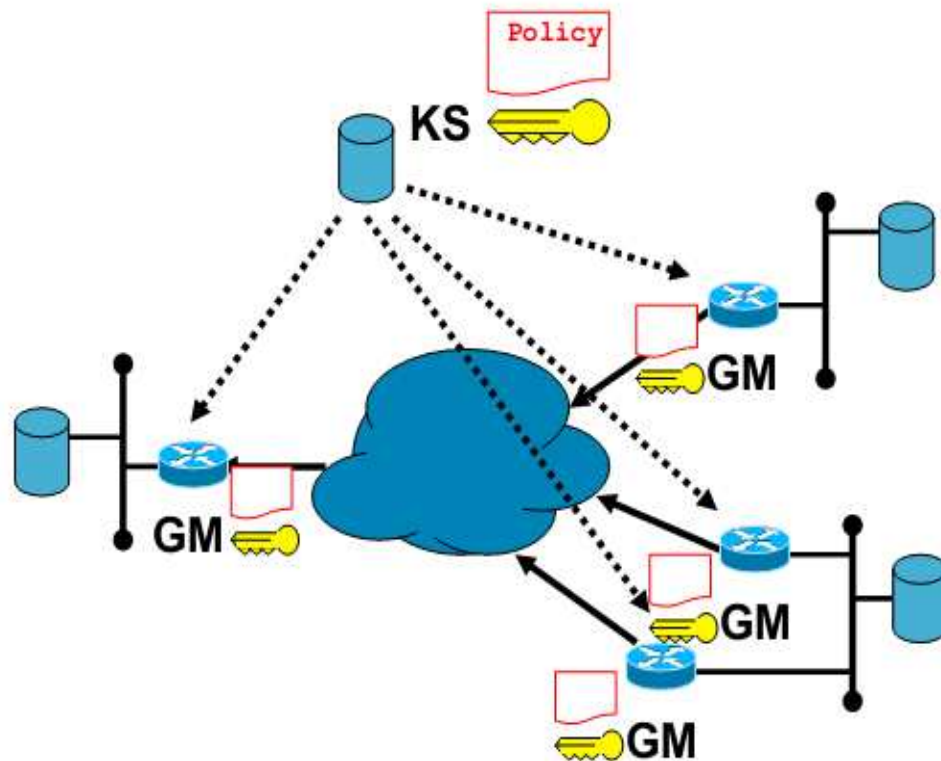
- Established security associations between two partners only
  - IKE tunnel for authentication and rekeying
  - IPsec tunnel for user data protection

- **GETVPN (Group Encrypted Transport VPN)**

- Breaks the basic IPsec concept of point-to-point security associations
  - There are no security associations anymore
  - A GETVPN endpoint just take the group key to encrypt the messages in tunnel mode and passes it on
- All partners are getting their key material from a group key server which is used for rekeying too
- There is no IP address and routing separation between sites and the backbone
  - All IP addresses of sites will be seen in packets from the backbone
  - All encrypted messages are proceeded by the original source and destination addresses hence communication statistics will be seen in the backbone
- Multicasting is possible
  - If backbone supports multicast routing and multicast forwarding
- Group key servers located in the backbone need to be well protected

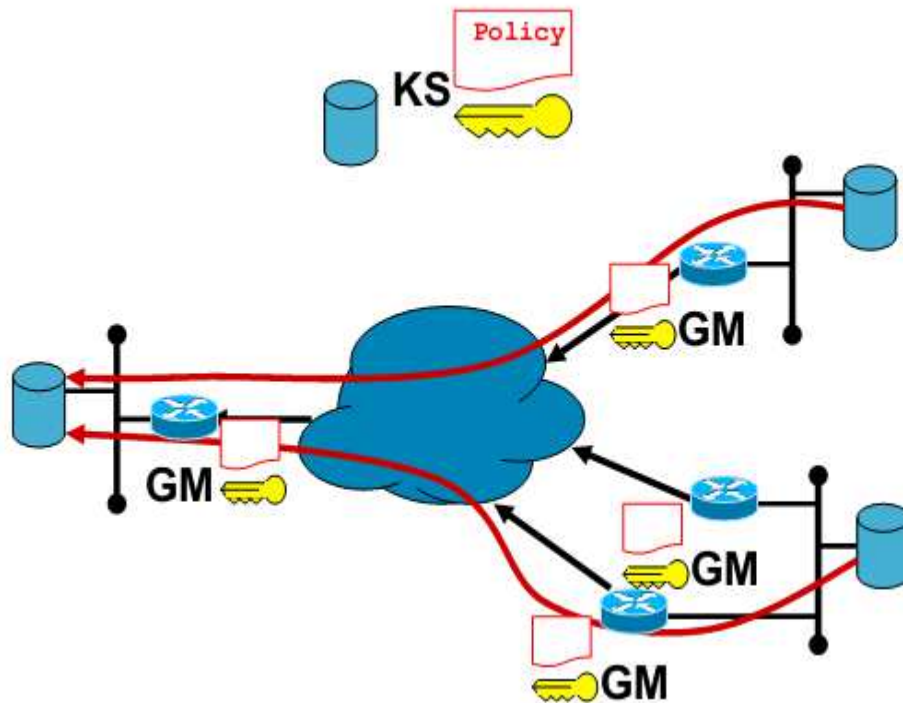


# GETVPN Key Server / Group Members



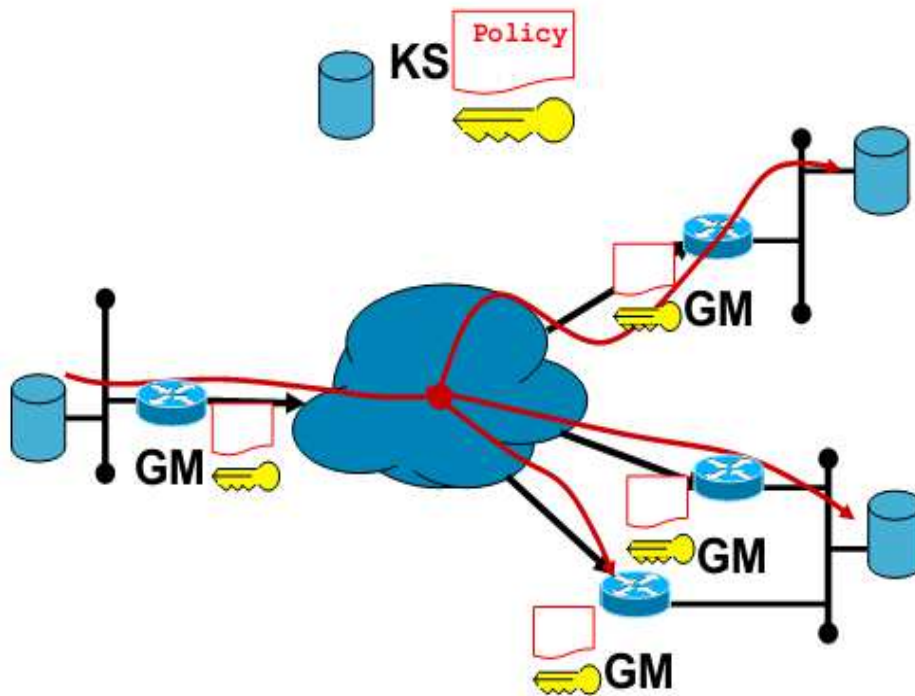
- **Key Server (KS):**
  - Device which distributes keys & policies to group members
- **Group Member (GM):**
  - Device which registers with a group controlled by the KS to communicate securely with other GMs

# GETVPN - Security Protection



- Receiver does not know the potential encryption sources
- Receiver assumes that legitimate group members obtain Traffic Encryption Key from key server for the group
- Receiver can authenticate the group membership

# GETVPN - Multicasting



- IP address preservation for end-to-end IP unicast and multicast routing
- Encrypt multicast traffic with IP address preservation
- Replication In the backbone is based on original (S,G) states built by multicast routing protocols

# Comparison DMVPN versus GETVPN

- **DMVPN is an overlay VPN**
  - Creates tunnels over the transport network
    - Isolates protected networks from transport network
    - Allows private protected addresses over a public transport network
  - Hubs concentrate connections - all spokes must connect
    - Hubs concentrate part of the spoke-spoke traffic
    - Hubs need to know about all the private networks
  - Multicast requires replication before encryption - usually on hubs
- **GETVPN is a “proxy VPN”**
  - Encrypted packets have the same addresses as the protected packets
    - Does not isolate address spaces hence requires end-to-end routing
  - Key servers concentrate connections - all group members must connect
    - Key servers do not concentrate any traffic
  - Transport network takes care of routing packets
  - Multicast can happen in the core if core supports it

# Agenda

---

- **Introduction**
- **Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
- **Multicasting**
  - Introduction
  - Multicast Routing Overview
  - Multicast & HA
  - Multicast & VPN / Security
- **Summary**

# Communication Behavior

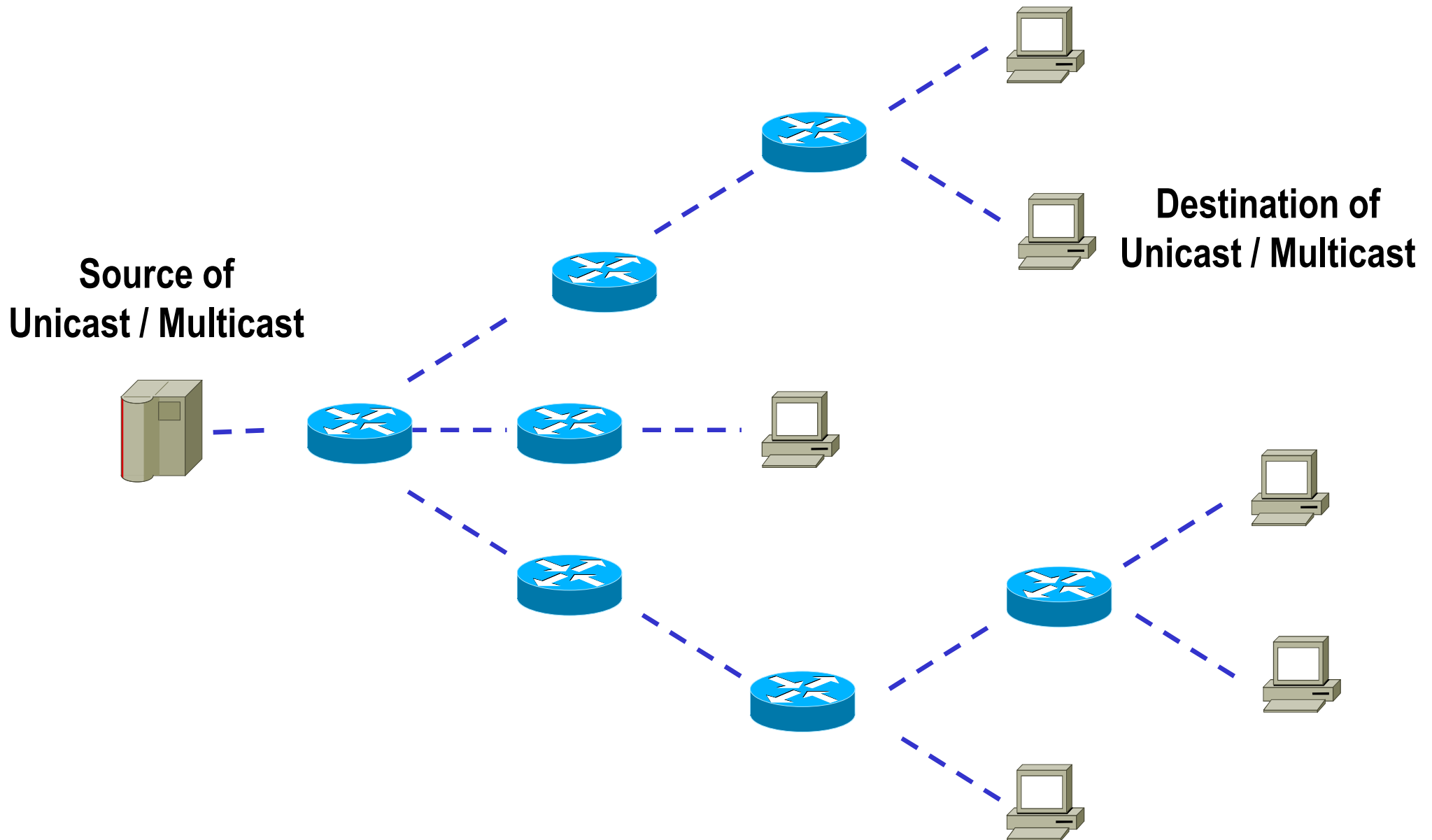
- **IP Unicast**

- Natively implemented in an IP network
- For individual communication style (like video on demand)
- State of the art in the “Internet”

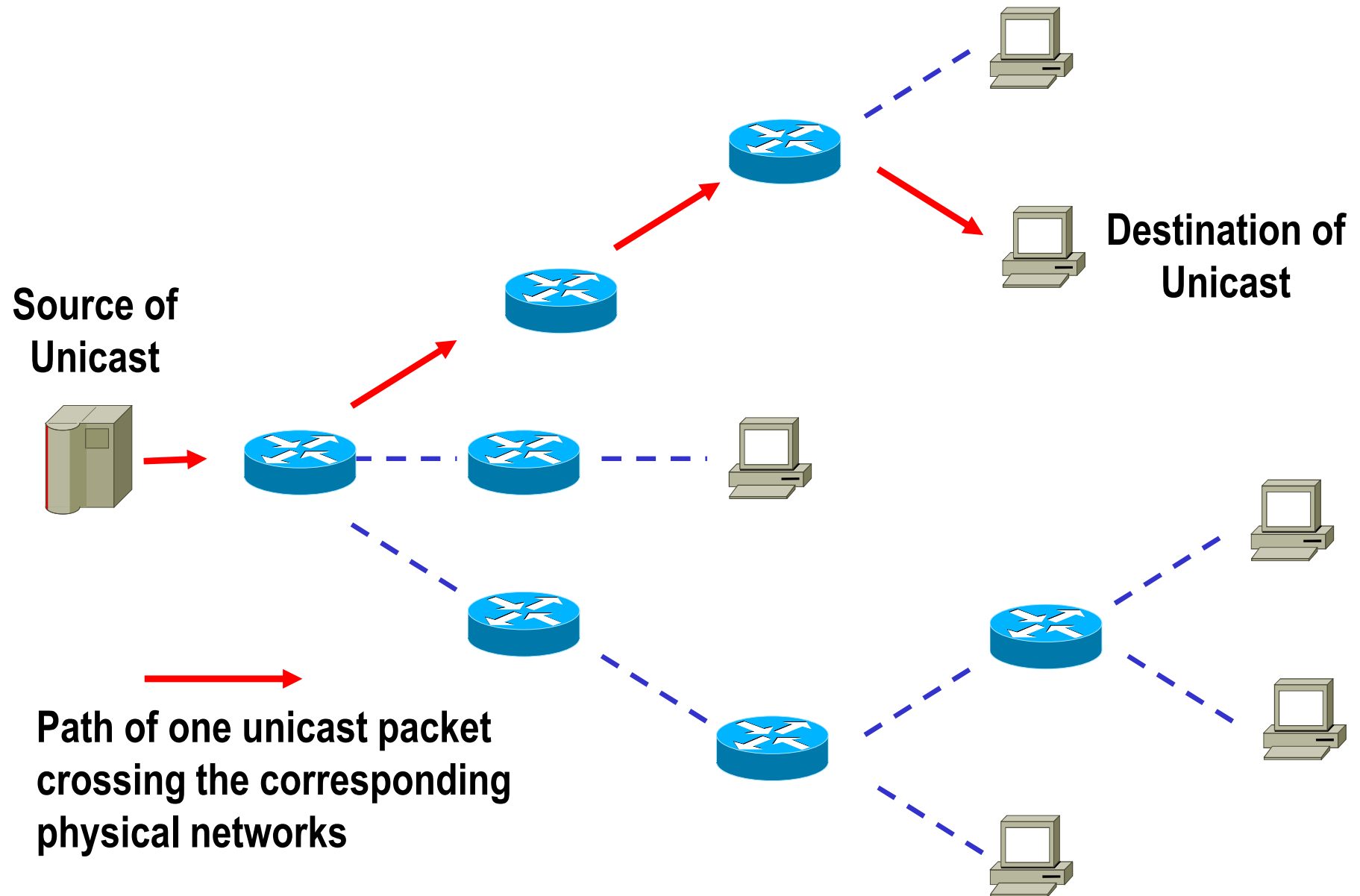
- **IP Multicast**

- For broadcast communication style (like television)
- No global multicast in the “Internet” today
- Promises to save bandwidth in the network
  - Only true if you have complete control over the infrastructure
  - Could be suboptimal or even not possible if you base your network on service provider technology or in case of security
  - Note: IPsec do not support multicast so far, you need additional functionality and/or tricks to do it. MPLS-VPN supports multicast in a sophisticated list of methods only recently.

# Physical Network Topology (Example 1)

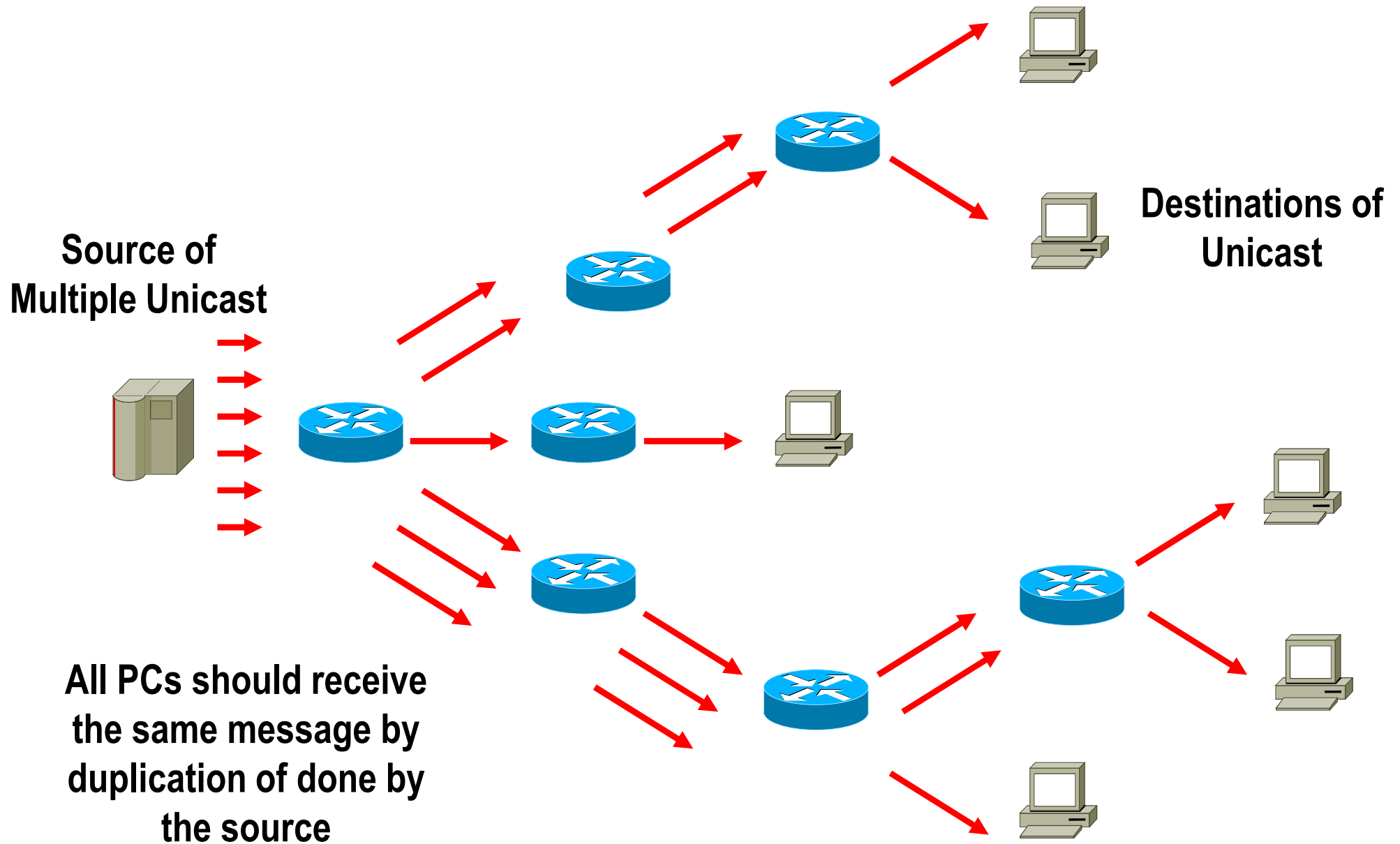


# Unicast Transmission (Example1)

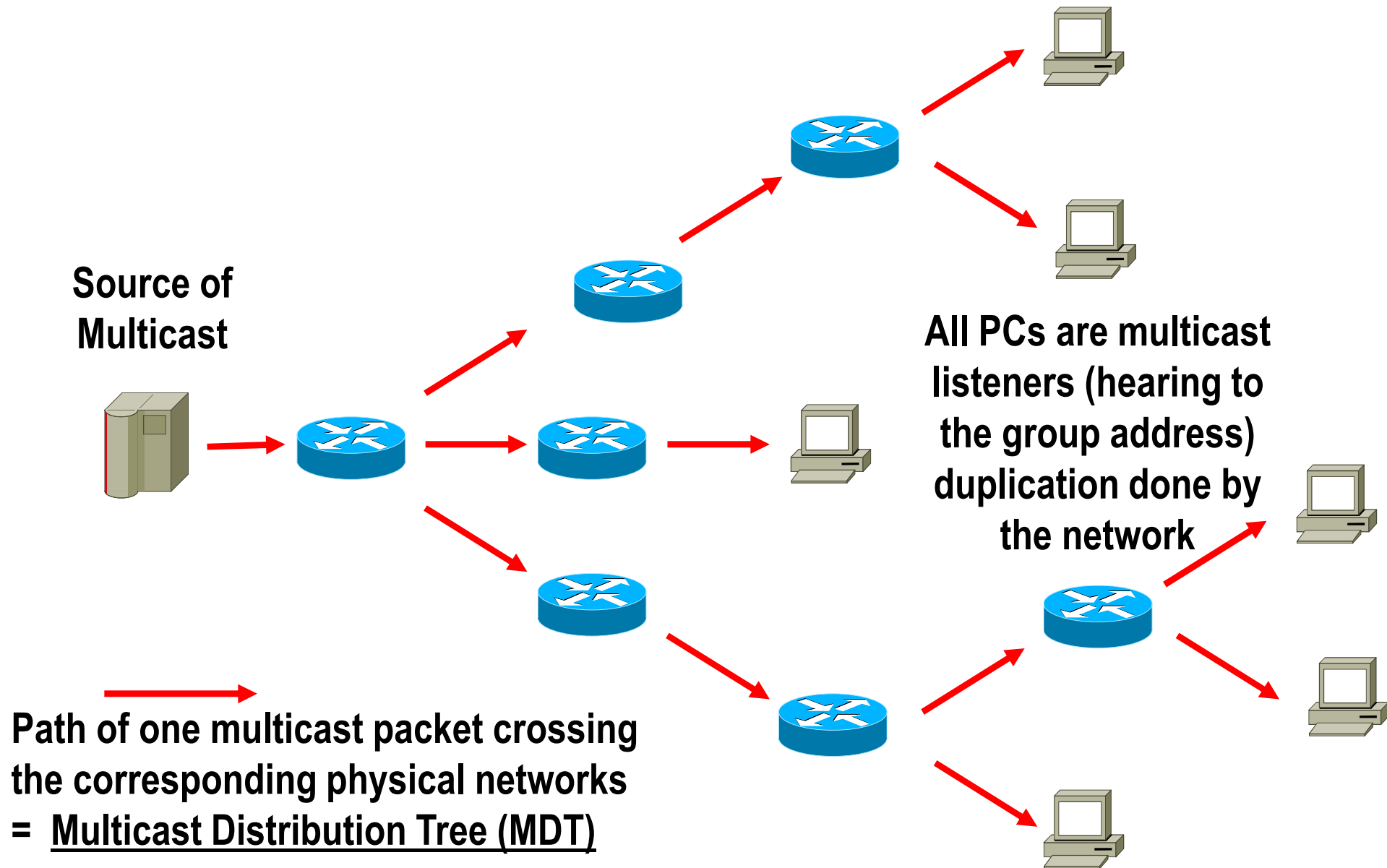




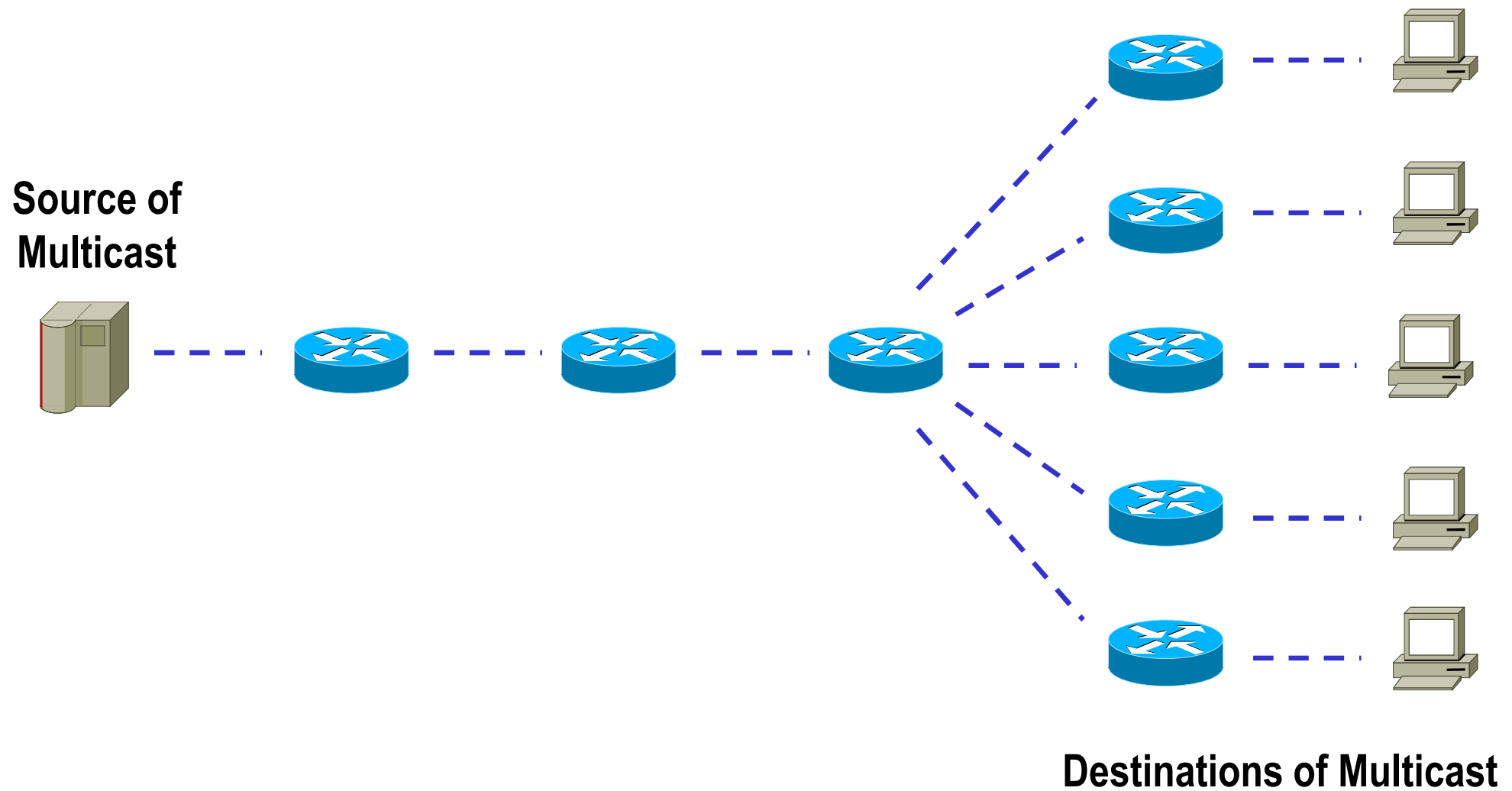
# Multiple Unicast Packets (Example1)



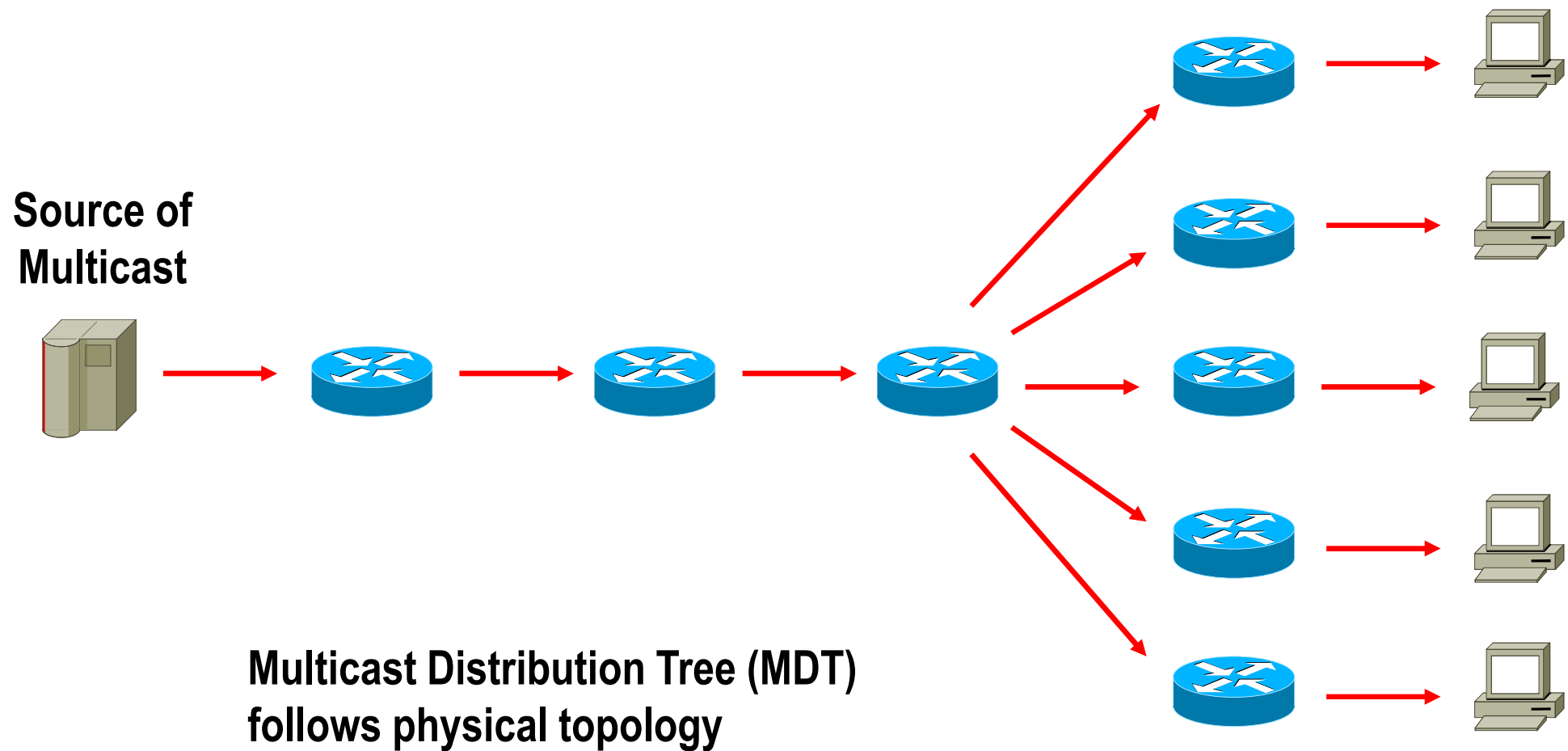
# Multicast Transmission (Example1)



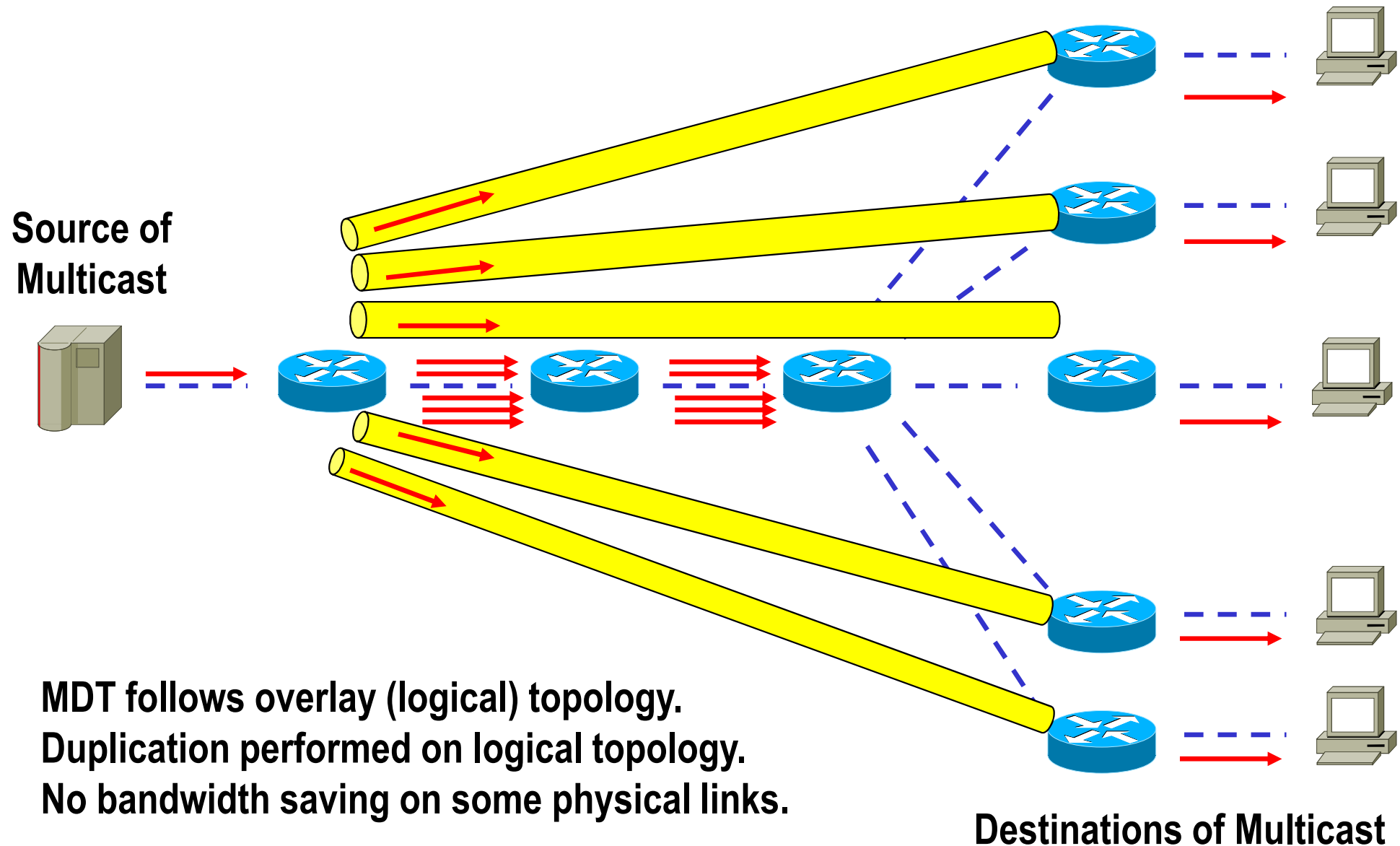
# Physical Network Topology (Example2)



# Multicast Without Overlay VPN (Example2)



# Multicast With Overlay VPN (Example2)



# Agenda

---

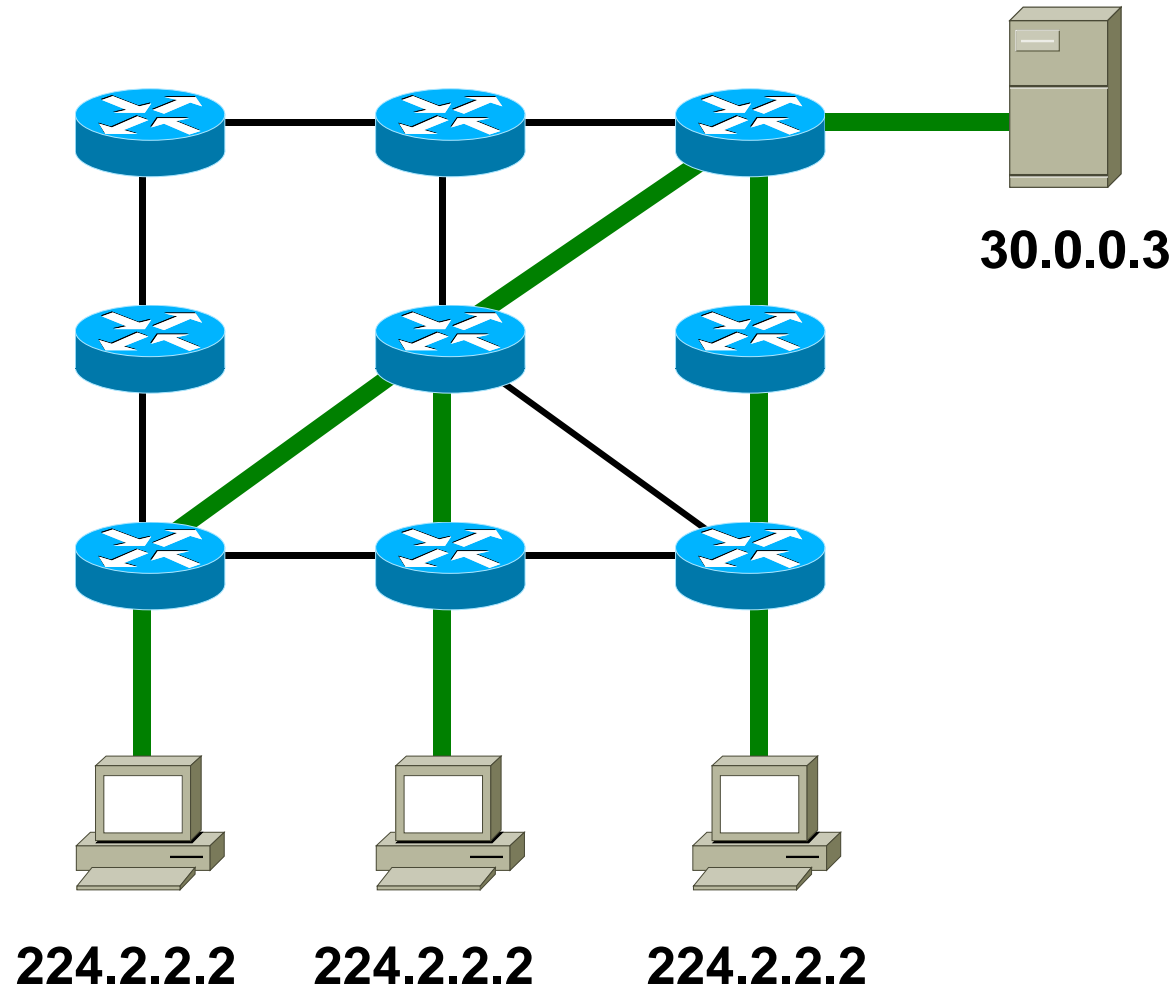
- **Introduction**
- **Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
- **Multicasting**
  - Introduction
  - Multicast Routing Overview
  - Multicast & HA
  - Multicast & VPN / Security
- **Summary**



# MDT Types - Shortest Path Tree (2)

Also called "Source Distribution Tree" or "Source (-based) Tree"

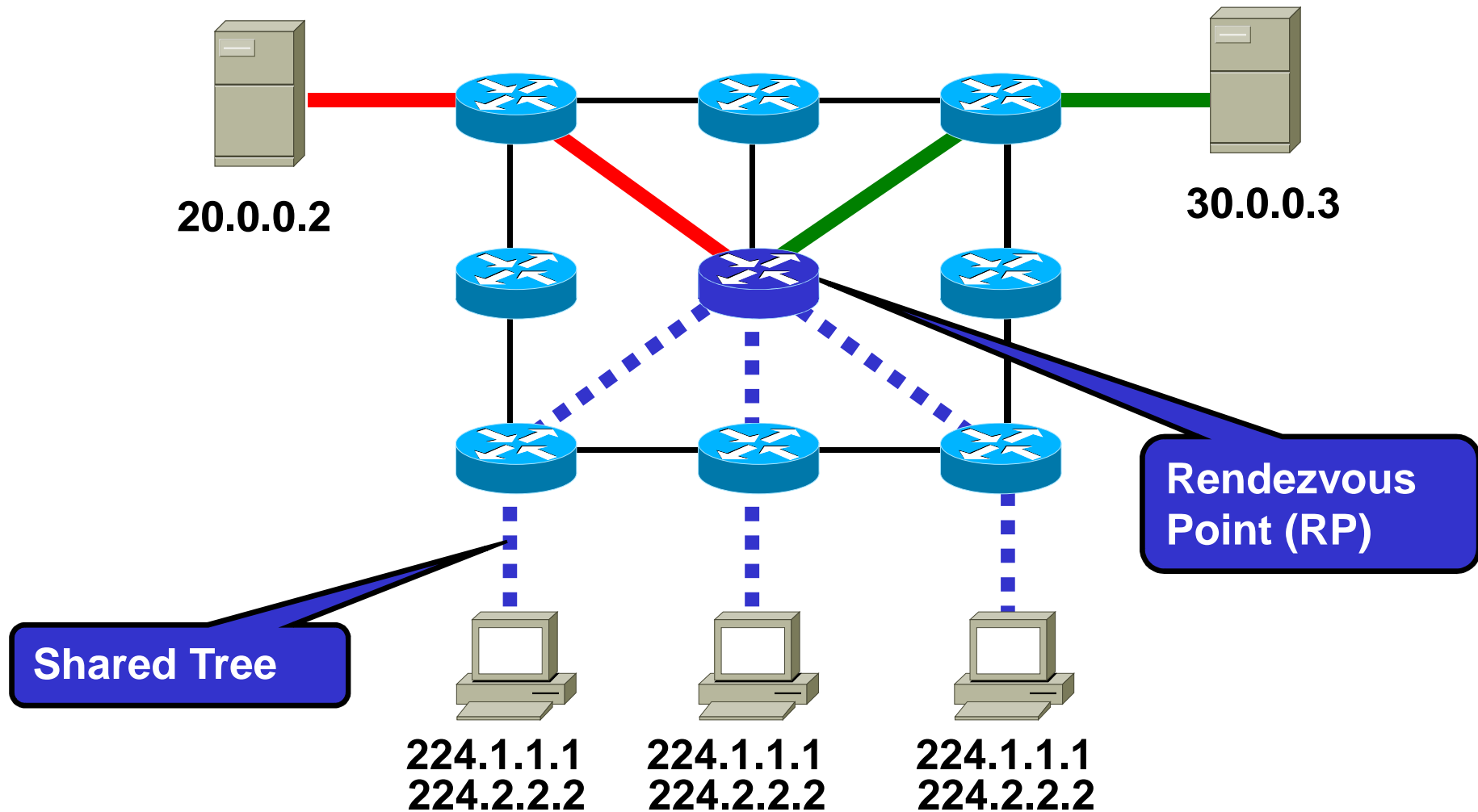
$(S, G) = (30.0.0.3, 224.2.2.2)$





# MDT Types - Shared Tree

(\*, G) = (\*, 224.1.1.1) and (\*, 224.2.2.2)



# Multicast Routing Protocol Types

- **Dense Mode: Push method**

- Initial traffic is flooded through whole network
- Branches without receivers are pruned (for a limited time period only)
  - DVMRP            Distance Vector Multicast Routing Protocol
  - MOSPF            Multicast OSPF (deprecated RFC)
  - PIM-DM           Protocol Independent Multicast – Dense Mode

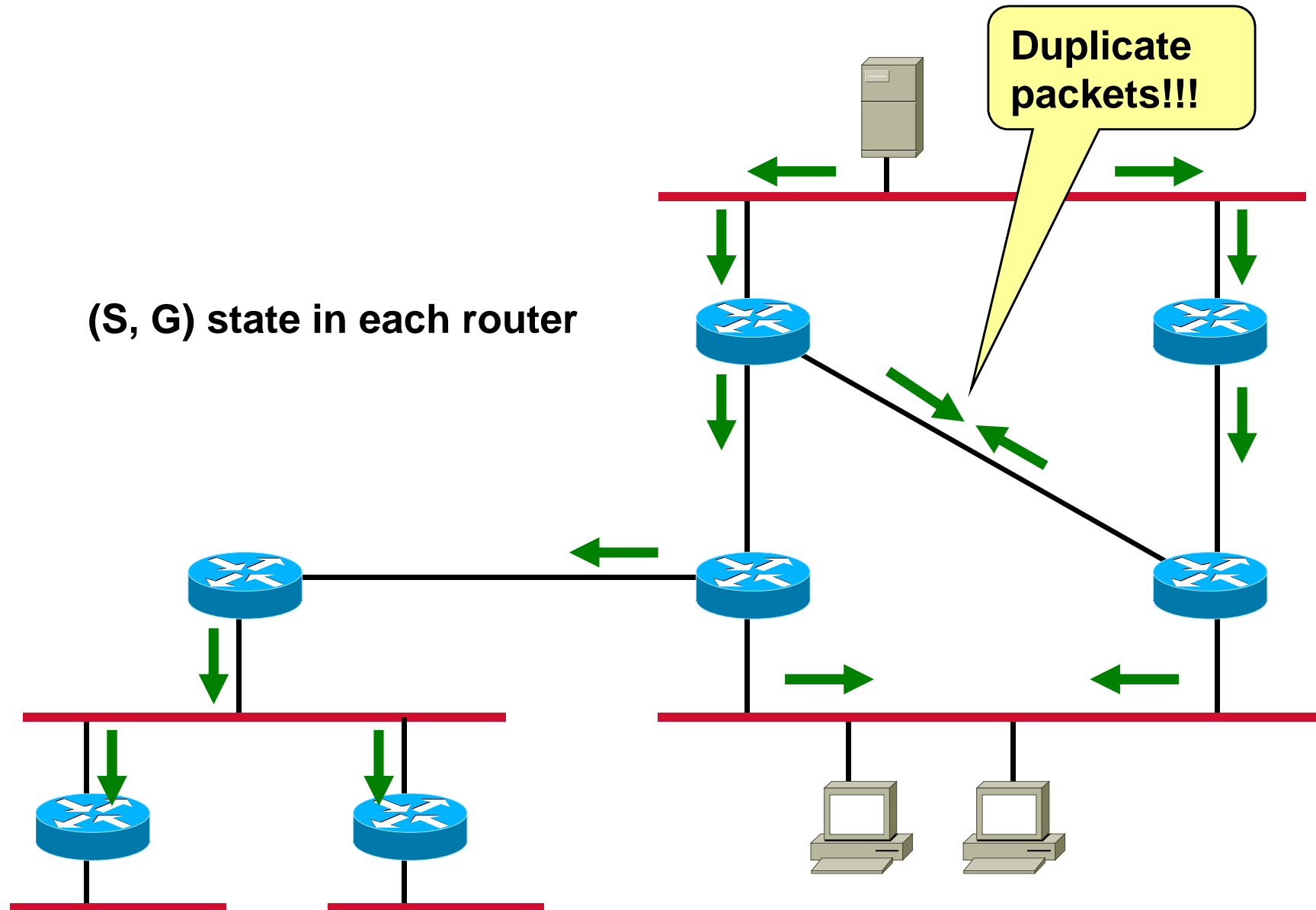
- **Sparse Mode: Pull method**

- Explicit join messages
- Last-hop routers pull the traffic from the rendezvous point (RP) or directly from the source
  - PIM-SM            Protocol Independent Multicast – Sparse Mode
  - CBT                Core Based Trees

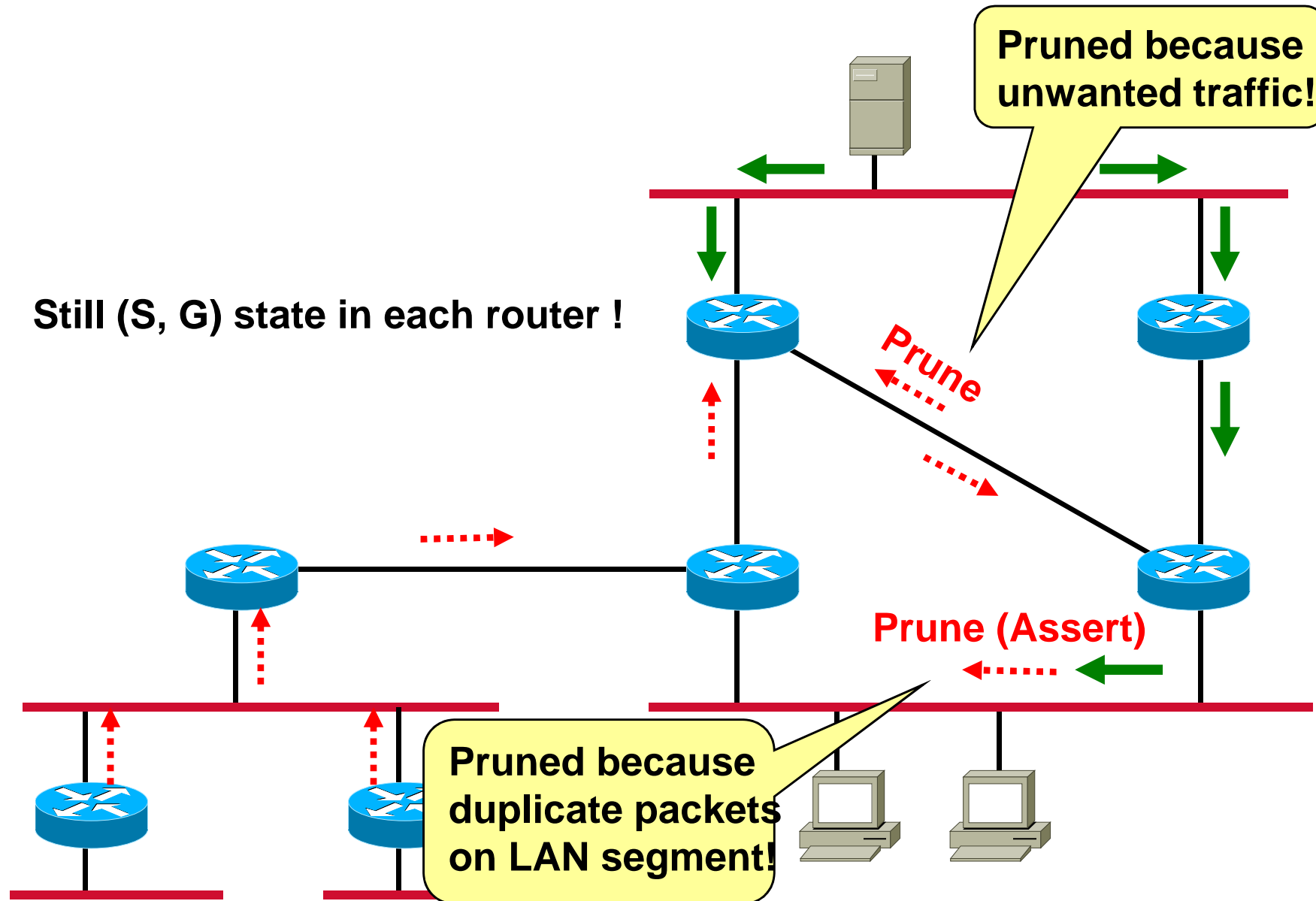
# PIM - DM

- **Protocol Independent**
  - Utilizes any underlying unicast routing protocol
- **Method**
  - No dedicated multicast routing protocol in use
  - RPF, flood and prune is performed
- **For small networks only**
  - Every router maintains (S, G) states
  - Initial flooding causes duplicate packets on some links
- **Easy to configure**
  - Two command lines
  - Useful for small trial networks

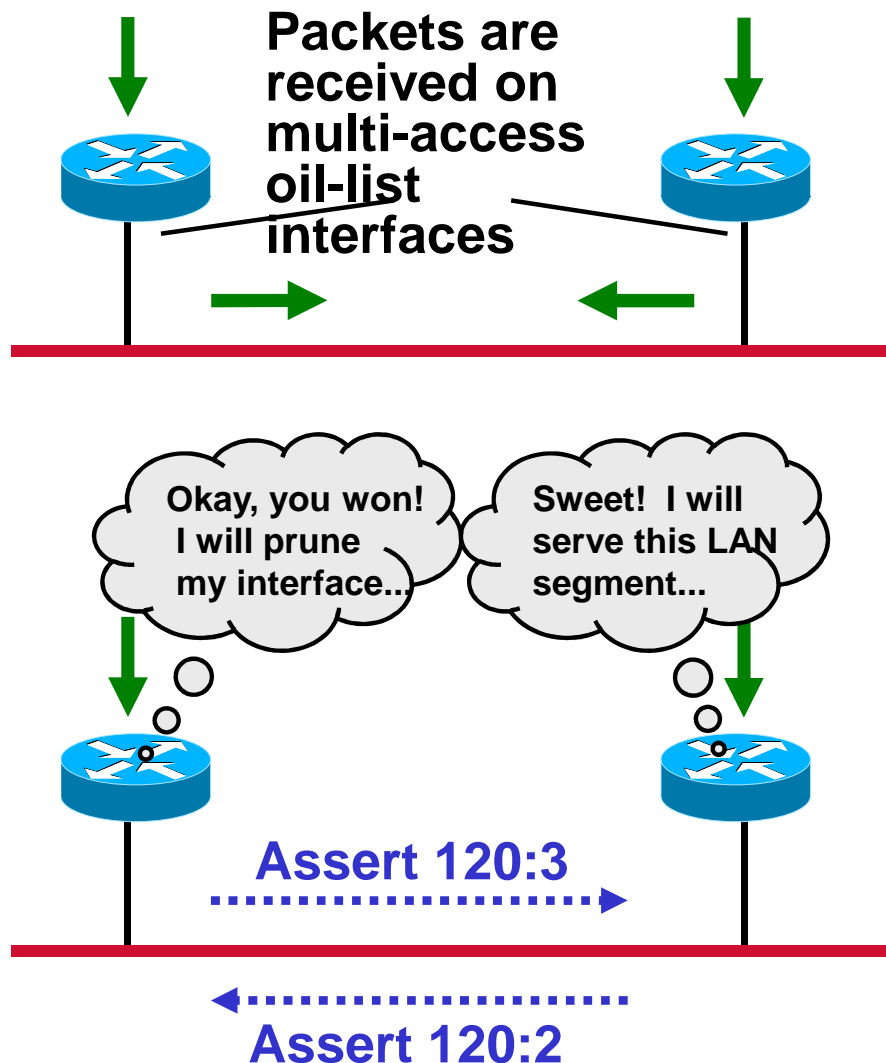
# PIM-DM: Initial Flooding



# PIM-DM: Pruning



# PIM-DM: Assert Mechanism

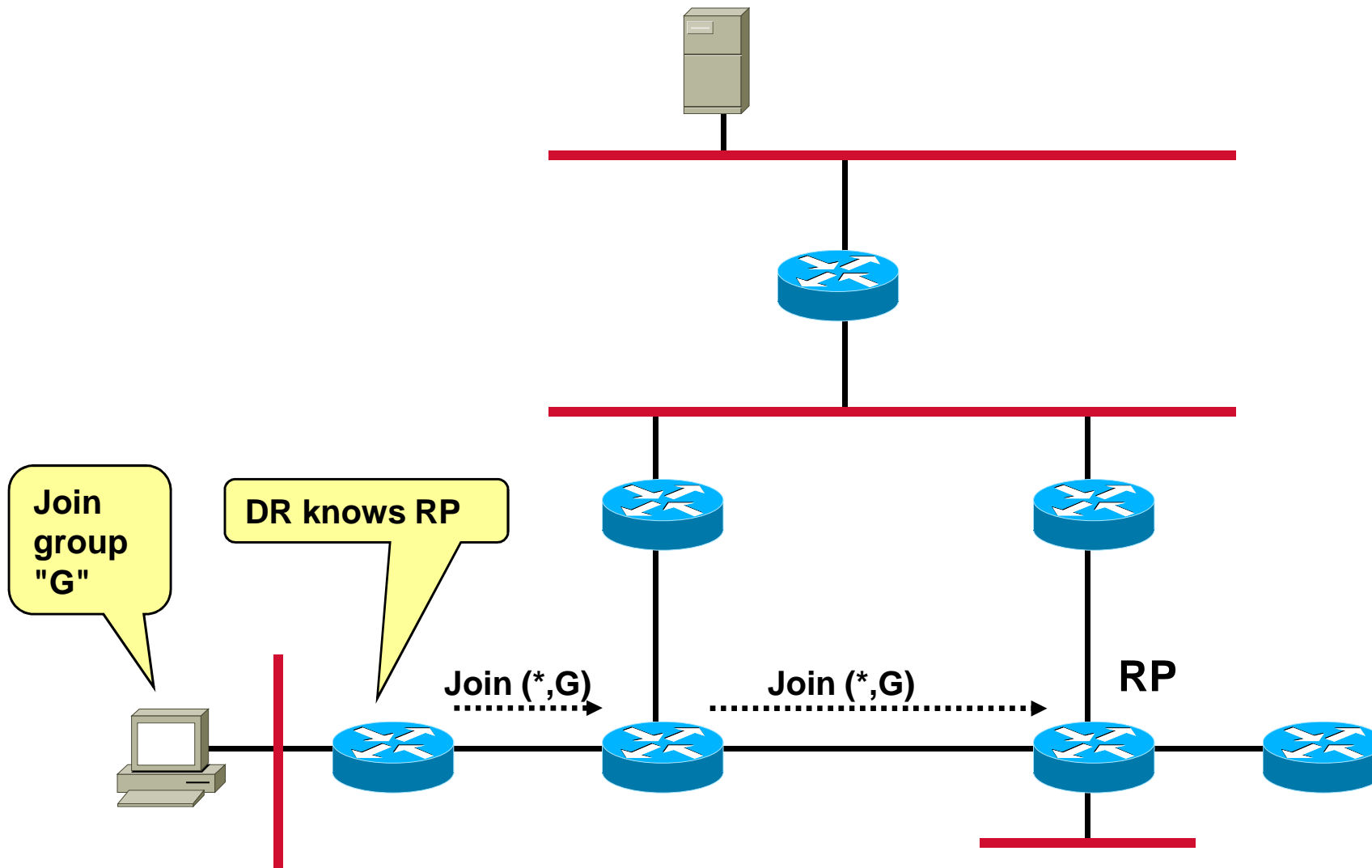


- **Each router receives the same (S, G) packet through an interface listed in the oil-list**
  - Only one router should continue sending
- **Both routers send "PIM assert" messages**
  - To compare administrative distance and metric to source
- **If assert values are equal, the highest IP address wins**

# PIM-SM

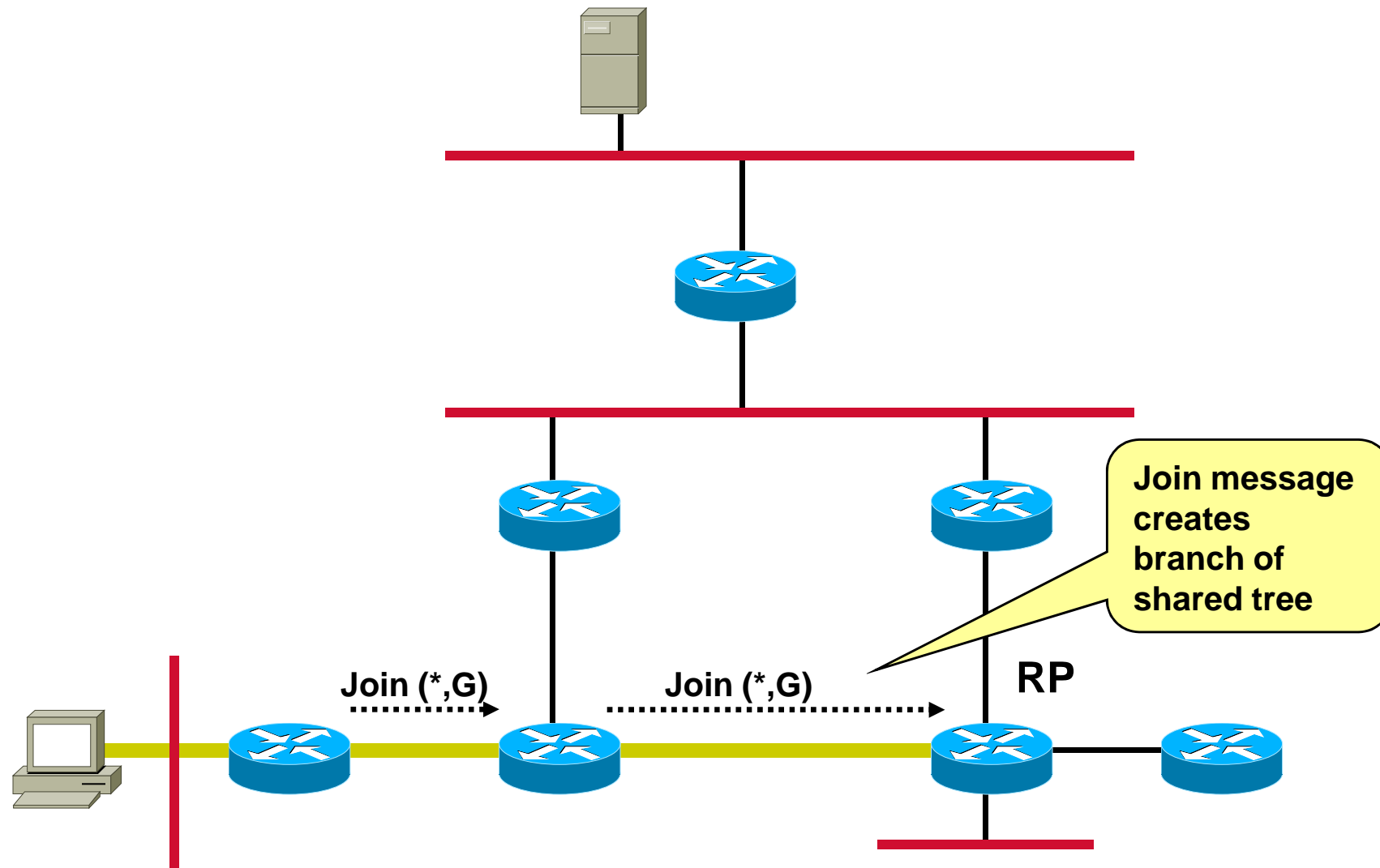
- **Protocol Independent**
  - Utilizes any underlying unicast routing protocol
- **Supports both source and shared trees**
- **Uses a Rendezvous Point (RP)**
  - Sources are registered at RP by their first-hop router
  - Groups are joined by their local designated router (DR) to the shared tree, which is rooted at the RP
- **Best solution today**
  - Optimal solution regardless of size and membership density
- **Variants**
  - Bidirectional mode (PIM-bidir)
  - Source Specific Multicast (SSM)

# PIM-SM / User Becomes Active

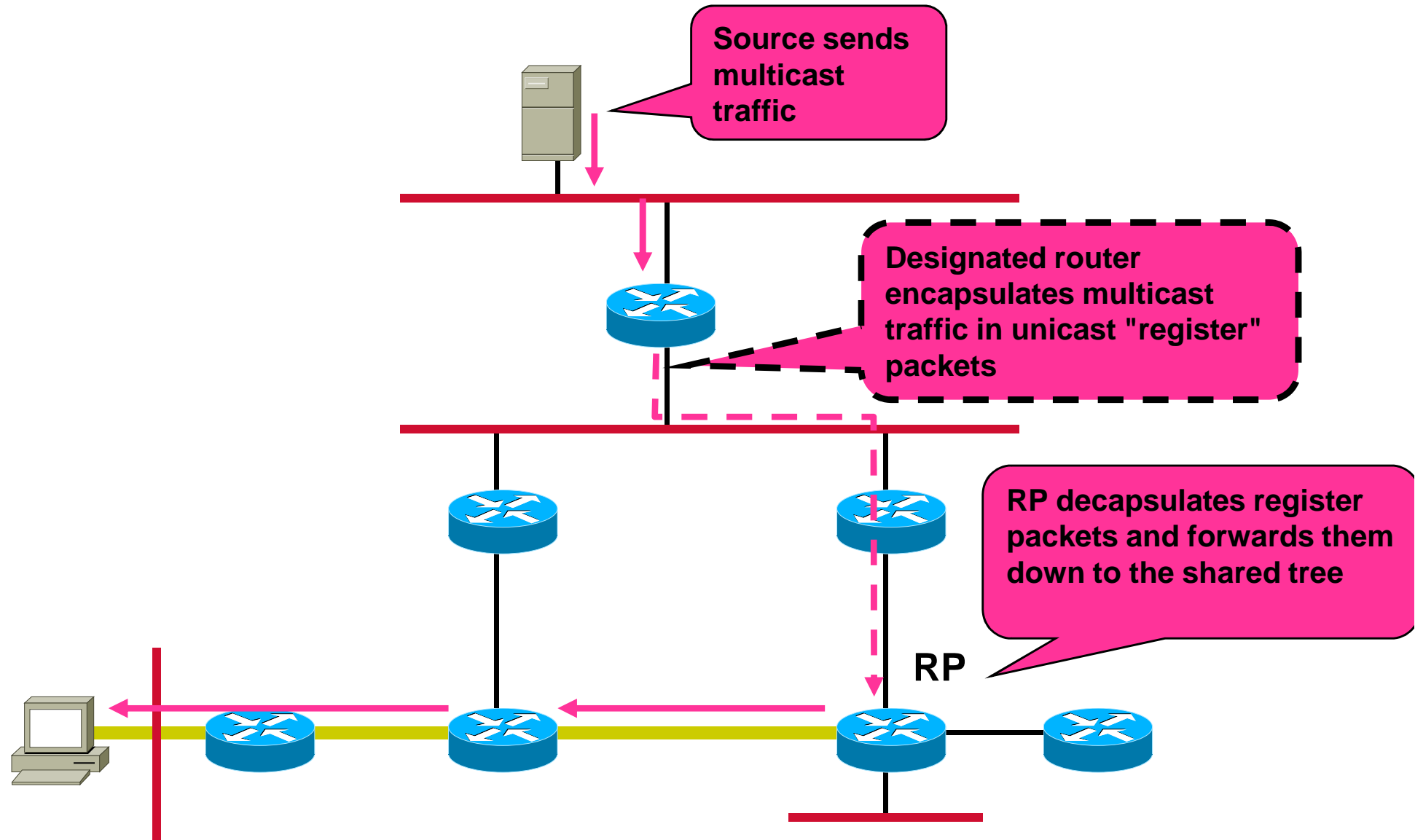




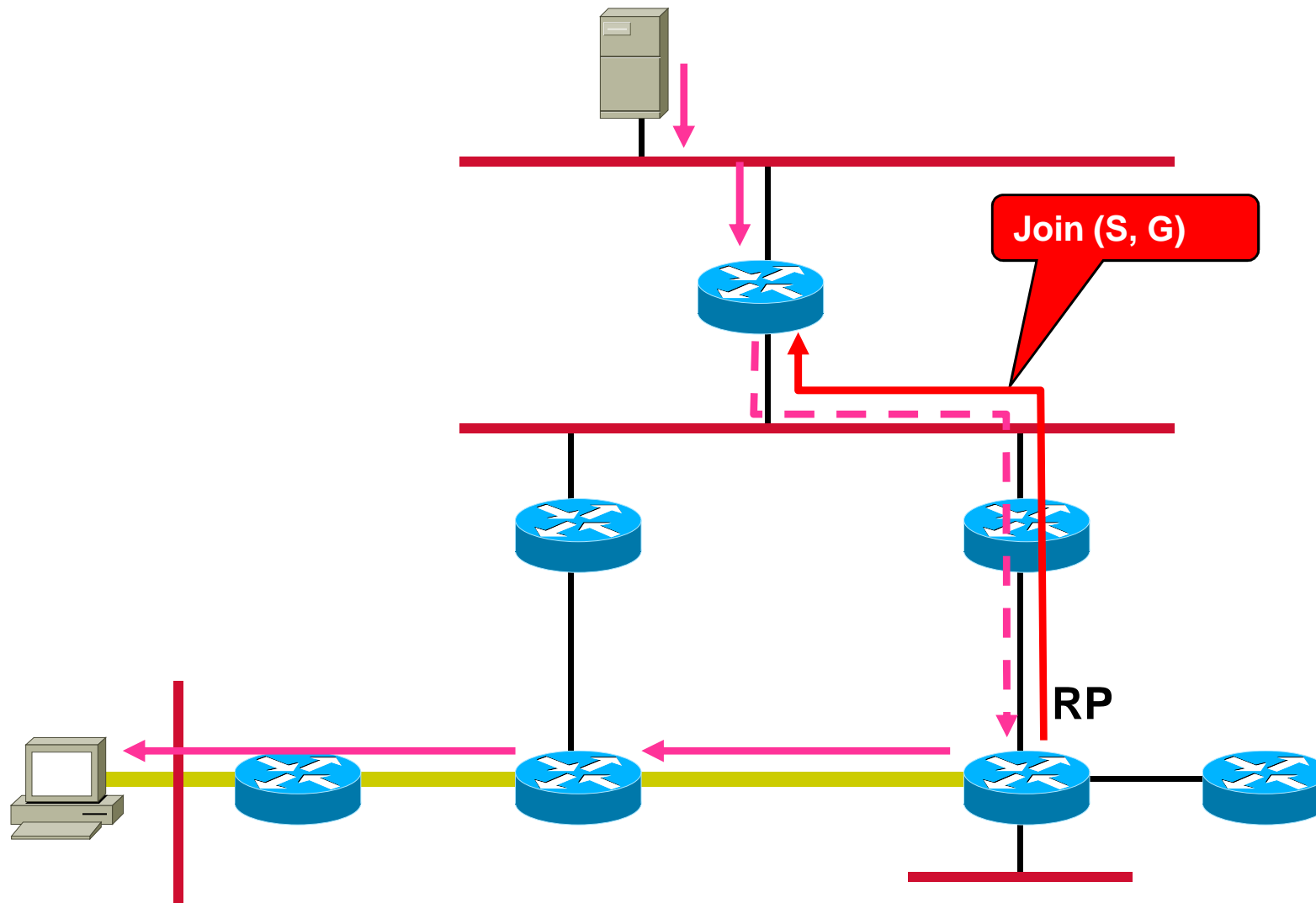
# PIM-SM / Create Shared Tree



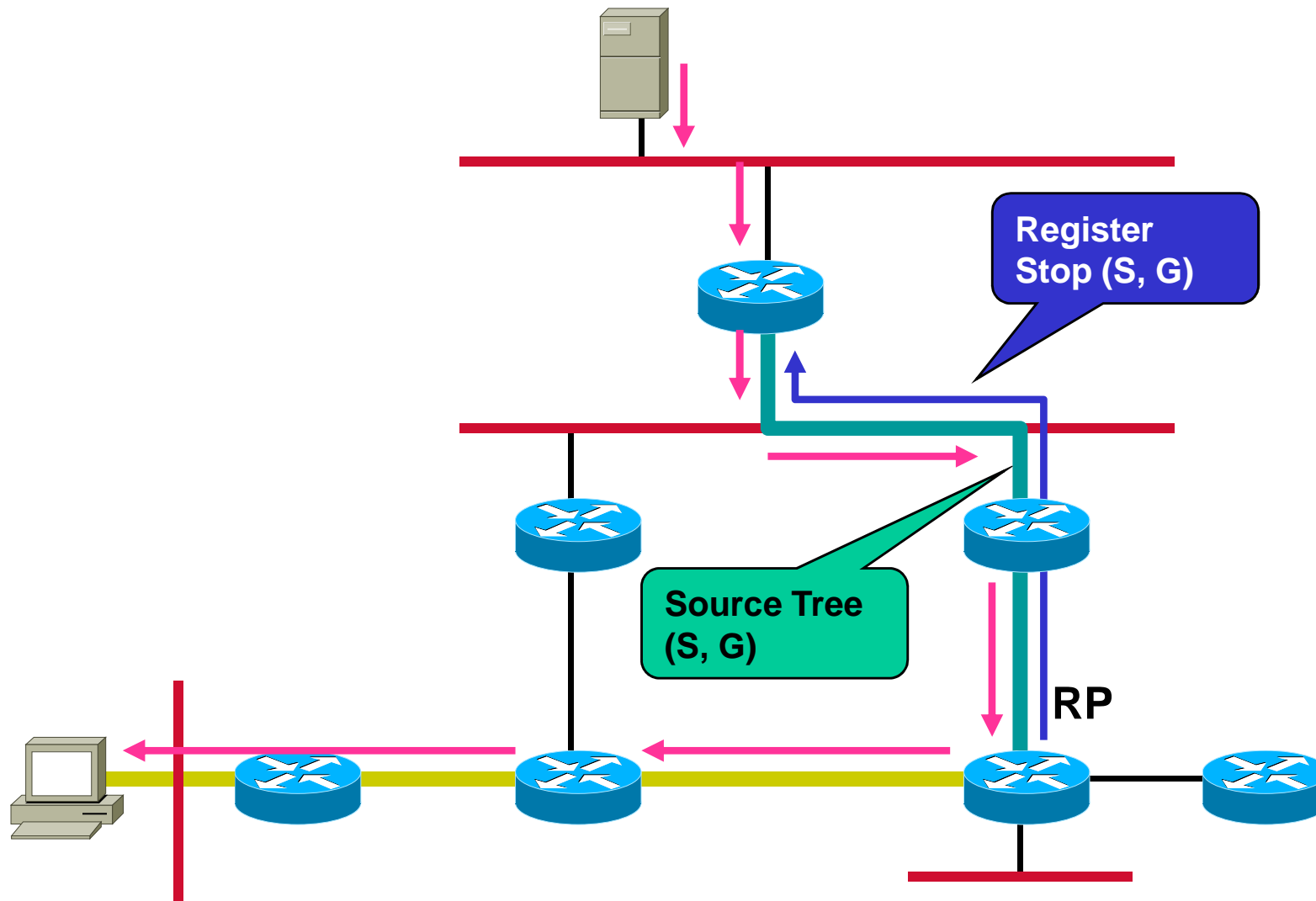
# PIM-SM / Register Source



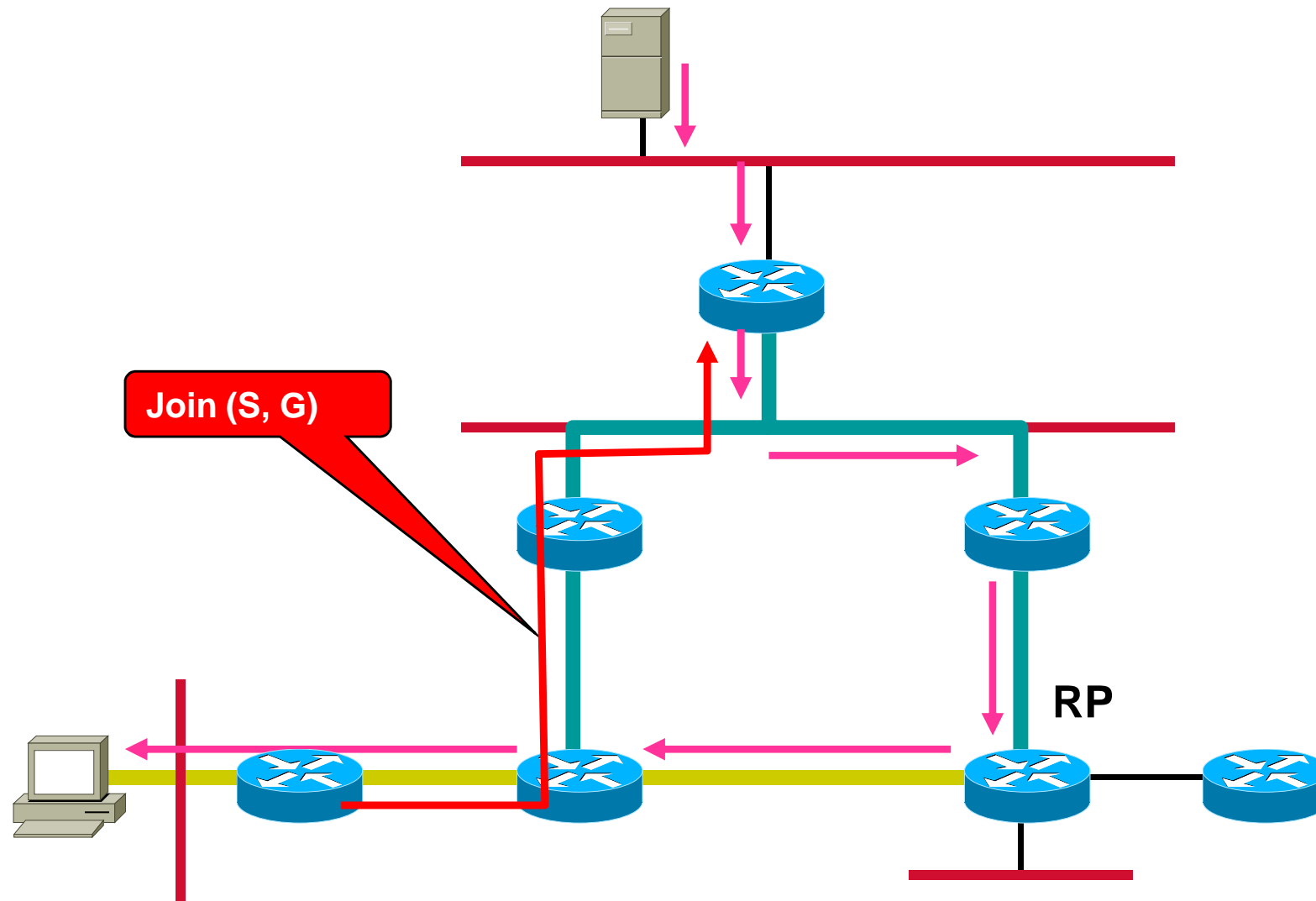
# PIM-SM / Create Source Tree



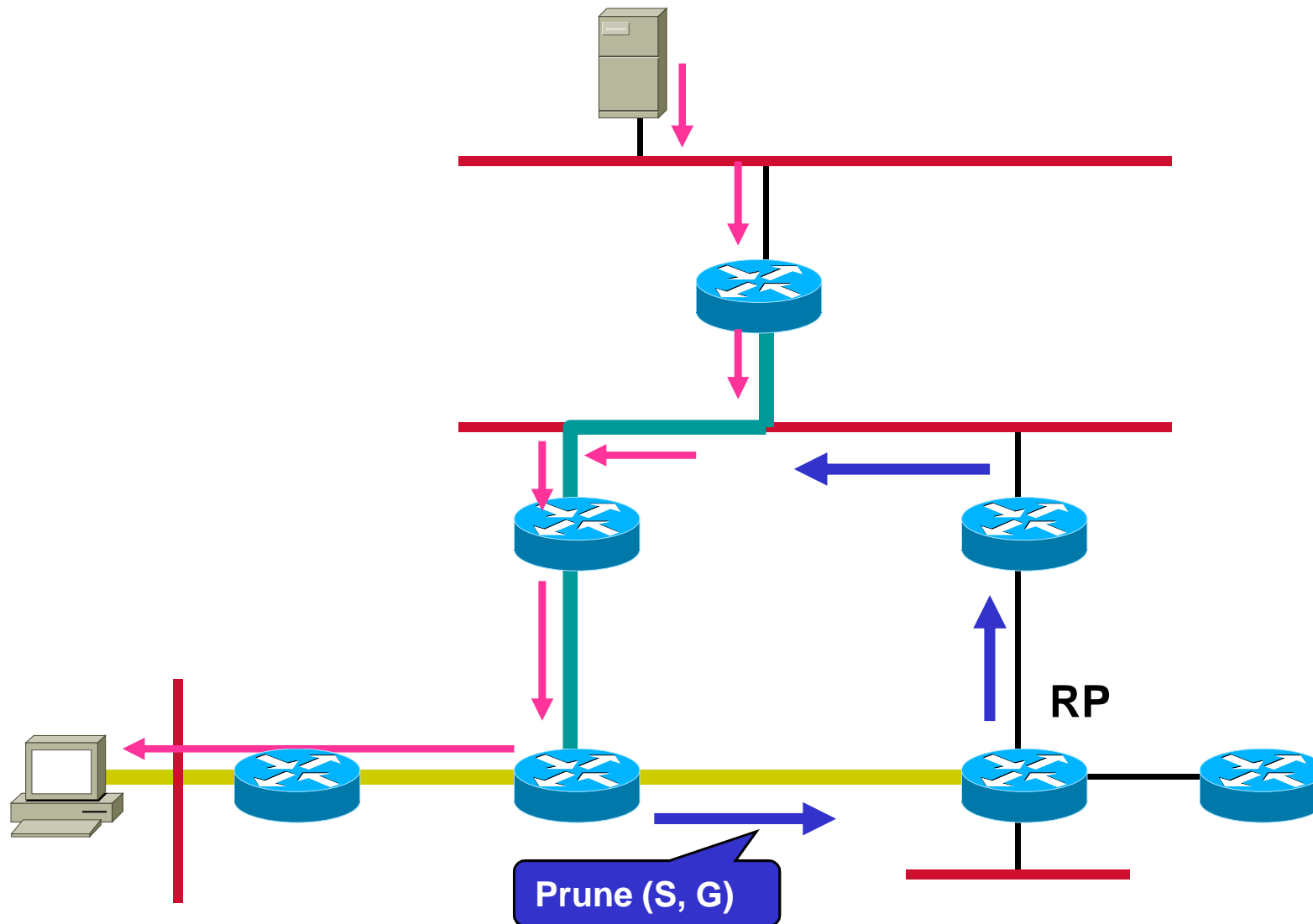
# PIM-SM / Create Source Tree



# PIM-SM / Switchover



# PIM-SM / Pruning



# Agenda

---

- **Introduction**
- **Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
- **Multicasting**
  - Introduction
  - Multicast Routing Overview
  - Multicast & HA
  - Multicast & VPN / Security
- **Summary**

# Multicast & HA / Convergence

## ● Basic problems

- MDT establishment may be caused only by presence of multicast traffic from given source(s)
  - A kind of setup delay which adds on in case of switchover to redundant paths not so far used for multicast transmission
- MDT (states) will be removed after a timeout when multicast traffic from given source(s) stops
  - To reduce amount of necessary states to be kept or to heal a tree in case of network topology changes
- Multicast routing is “data-driven” versus “topology-driven” style of IP unicast routing
- Most multicast routing protocols depend on underlying unicast routing protocol
  - Hence multicast convergence can only be what unicast convergence will give



# PIM - DM Operation Summary

- **Implementation of RFP, flood and prune**
- **Shortest path trees (SPT or S,G trees) are built on demand**
  - When multicast source start sending such traffic
  - “Data-driven”
- **States for pruning**
  - Are established in the multicast routers
- **States are removed and need to be refreshed**
  - To adapt to network topology changes
  - To adapt to new multicast listeners on so far pruned locations
- **RPF check**
  - Not done for every multicast packet but be periodically proofed based on RFP timeout value or change of the unicast routing table concerning active sources

# PIM - DM Convergence

- **Depends on**
  - IGMP timing and timeouts in case new multicast listener appears in the network
  - On timing for grafting in case the location was pruned so far
  - Periodically flooding if grafting is not supported
  - Active multicast sources otherwise MDT states time out
  - Unicast IP routing convergence together with RPF check timeout in case of network topology change
- **High complexity**
  - For troubleshooting and understanding
  - For building test cases for verification
    - All the above parameters influence the actual behavior
- **Do not use PIM-DM for mission critical communication**

# PIM - SM Operation Summary

- **Presence of multicast listeners**
  - Creates shared trees or \*,G trees towards a rendezvous point (RP)
- **States for joining**
  - Established in the multicast routers
  - Time out if not periodically refreshed
- **Multicast source traffic**
  - First hop router uses register encapsulation to reach the RP via unicast transport system
- **Optional:**
  - Creation of S,G tree from RP to source with join messages if multicast transport system is available toward source
  - This stops register procedure
- **First hop routers of multicast listeners**
  - Create individual S,G tree towards the source
  - Prune from the \*,G tree towards RP
  - If all multicast listeners have built their individual S,G tree the RP is not necessary anymore for that particular source/group combination
- **Hence**
  - RP for meeting to establish individual S,G trees on the fly

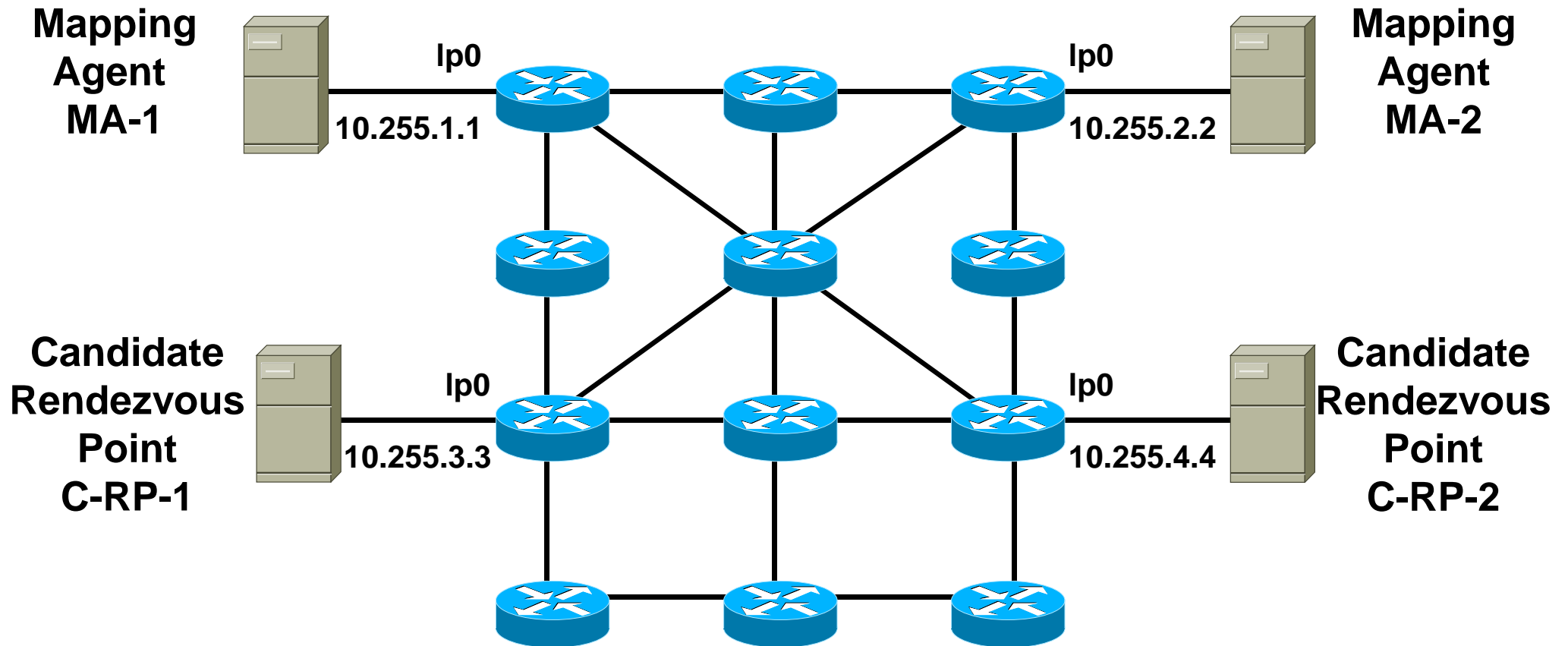
# PIM - SM Convergence

- **Depends on**
  - IGMP timing and timeouts in case new multicast listener appears in the network
  - On timing for joining in case the location had no multicast listener so far
  - Timing for selection new rendezvous point (RP) in case a RP is not available any longer
  - Unicast IP routing convergence together with \*,G and/or S,G building in case of network topology change
- **Less complexity**
  - For troubleshooting and understanding
  - For building test cases for verification
- **Recommended method for mission critical communication**
  - Decoupling done by the individual S,G trees from availability of RP ensures that ongoing traffic will continue if there is a RP switchover

# RP Redundancy - RP Auto Discovery

**MA: Listening to 224.0.1.39 (Cisco-RP-Announce)**

**MA: Sending on 224.0.1.40 (Cisco-RP-Discovery)**

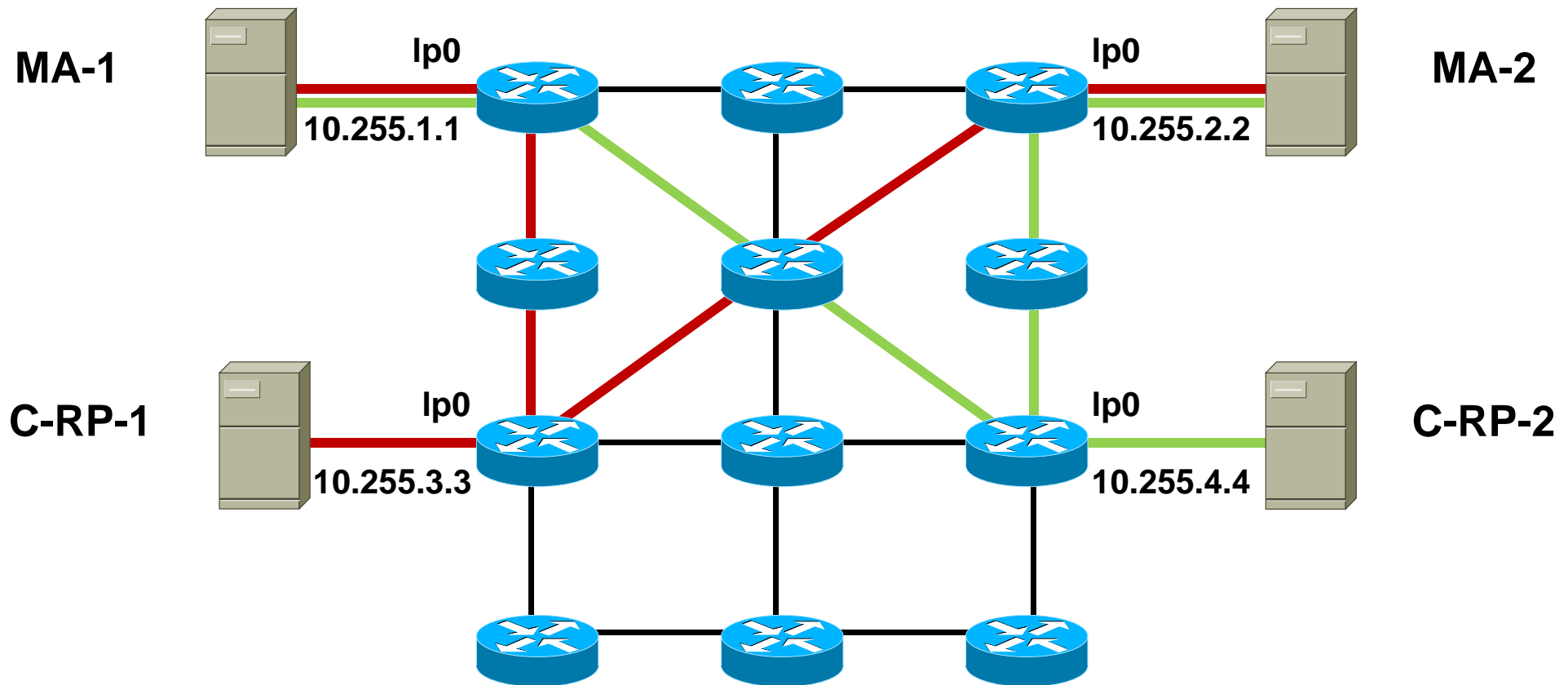


**C-RP: Sending on 224.0.1.39 (Cisco-RP-Announce)**  
**All MC: Listening to 224.0.1.40 (Cisco-RP-Discovery)**

# Cisco-RP-Announce Trees (DM, pruned)

C-RP-1 creates dense mode (S, G) = (10.255.3.3, 224.0.1.39)

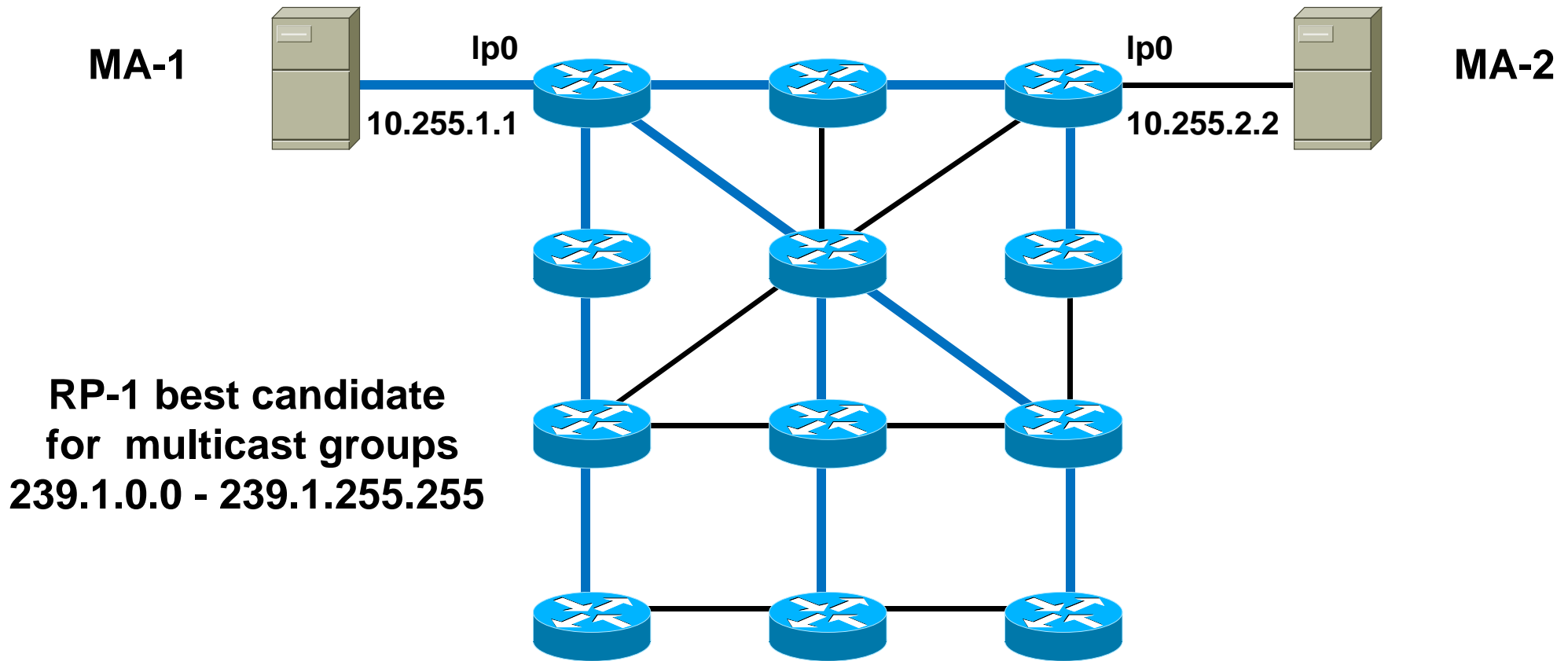
C-RP-2 creates dense mode (S, G) = (10.255.4.4, 224.0.1.39)



MA-1 and MA-2 are multicast listeners for 224.0.1.39

# Cisco-RP-Discovery Tree 1 (DM)

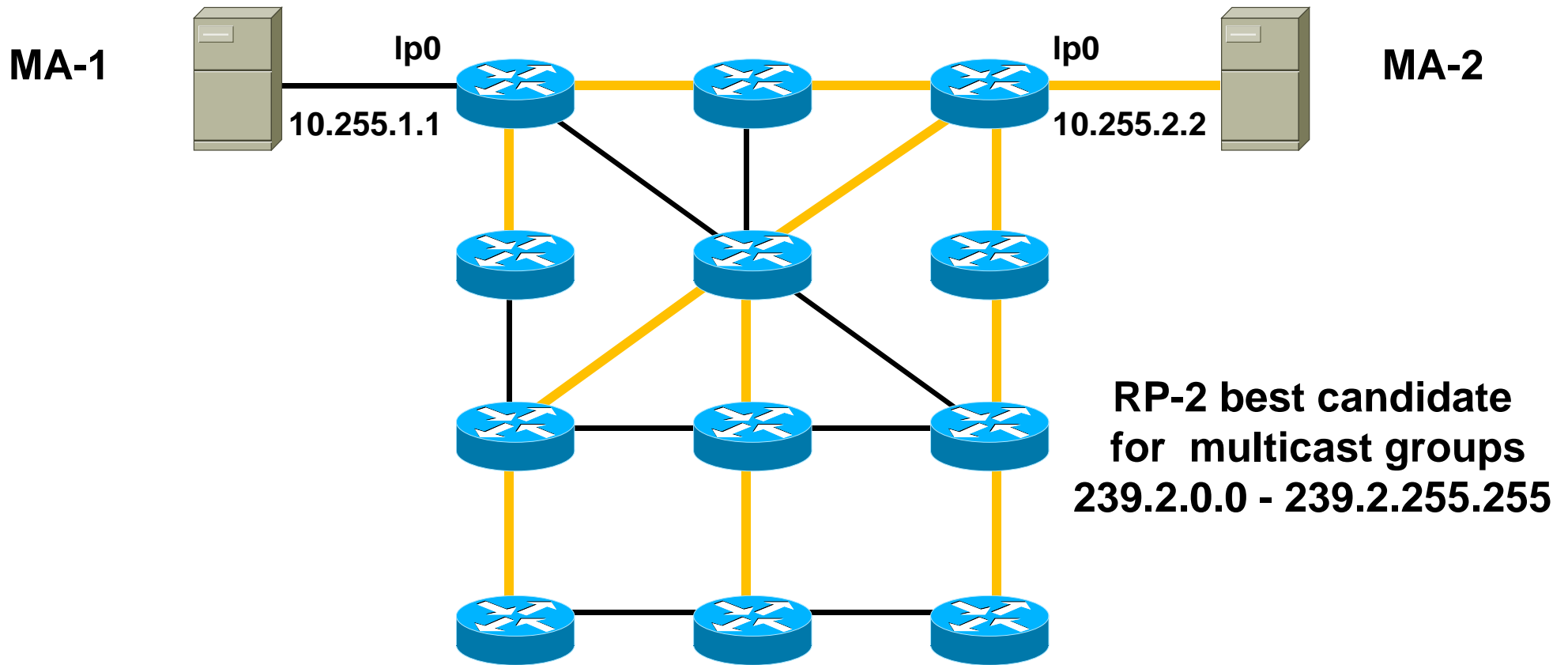
MA-1 creates dense mode (S, G) = (10.255.1.1, 224.0.1.40)



All multicast routers are multicast listeners for 224.0.1.40

# Cisco-RP-Discovery Tree 2 (DM)

MA-2 creates dense mode (S, G) = (10.255.2.2, 224.0.1.40)



All multicast routers are multicast listeners for 224.0.1.40



# Agenda

---

- **Introduction**
- **Operational Model**
- **High Availability**
- **QoS**
- **VPN Technologies**
- **Multicasting**
  - Introduction
  - Multicast Routing Overview
  - Multicast & HA / Convergence
  - Multicast & VPN / Security
- **Summary**

# Multicasting and MPLS

- **MPLS (Multi Protocol Label Switching)**

- Unicast IP/MPLS as backbone technology widely deployed by SPs nowadays
  - LDP for label distribution in conjunction with traditional IP unicast routing
  - RSVP-TE for Traffic Engineering and Fast Reroute
- Often the base for MPLS-VPN
- 15 years of development and experience

- **MPLS and Multicasting**

- Relatively new compared to unicast MPLS
  - MLDP (Multicast LDP) for P2MP/M2MP-> draft-ietf-mpls-ldp-p2mp-08
  - RSVP-TE P2MP -> RFC 4875

- **MPLS-VPN and Multicasting**

- Huge complexity caused by decoupling customer and provider multicast routing techniques and by separation of customers multicast domains (default-MDT)
- Therefore very seldom deployed by SPs

# Multicasting and Security

- **Classical IPsec VPNs**
  - Do not support multicast transport
- **DMVPN (Dynamic Multipoint VPN)**
  - Multicast replication on hub site only
  - Suboptimal concerning bandwidth savings enjoyed by using multicast techniques in a full self-controlled IP environment
- **GETVPN (Group Encrypted Transport VPN)**
  - Supports multicast transport if backbone is enabled for multicast routing/forwarding

# Agenda

---

- **Introduction**
- **Basic Building Block**
- **Routing and HA**
- **QoS**
- **VPN Technologies**
- **Multicasting**
- **Summary**

- **Unicast Connectivity**

- IP Address Plan
- Routing Concept
- NAT Concept (optional if necessary)

- **Network Operation Model**

- Complete infrastructure owned and self-operated
- Service Provider (L1 VPN, L2 VPN, L3 VPN)

- **High Availability (HA)**

- Selection of Automatic Switchover Mechanisms (the less the better)
- Routing Convergence Tuning

- **QoS**

- QoS Concept -> Consumer/Provider Clarification, QoS Monitoring, and QoS Management

- **Security**

- Security Concept -> Security Domains, Security Responsibilities
- Identifying Location of Perimeter and Tunnel Mechanism
- Agree on Security Management

- **Multicast (optional if appropriate)**

- Group Address Plan
- Multicast Routing Concept
- Routing Convergence Tuning

- **Management**

- Monitoring
- Security
- QoS

- **Holistic looking to all these topics is necessary**
  - All these topics must fit together
  - Tradeoffs will be seen and compromises have to be agreed
  - Design will not emerge in straight-forward way
  - Fact-finding missions and feedback loops will be necessary

# Hope for the Future – The Big Unifier ? !!!

- **LISP (Location / Identifier Separation Protocol)**
- **Open Standard**
  - Currently experimental RFCs and IETF drafts only
    - RFCs 6830 - 6836
  - Driven mainly by Cisco Network based solution
- **Original driven**
  - By routing scalability issues caused by PI (provider independent) addressing and PA (provider assigned) addressing in case of multi-homing to two or more ISPs



# LISP Base Ideas

- Separation of identity and location of an IP device / IP service
  - Remark: IP address covers both. Change of location means change of IP address and hence change of identity.
- LISP mapping system
  - Consists of mapping server(s) and resolver(s)
- LISP border routers
  - Separate EID (endsystem identifier) address domain from RLOC (routing locator) address domain
- Dynamic unidirectional encapsulation
  - Performed by LISP border routers
- Dynamic based caching
  - Triggered by data traffic between LISP sites

# LISP Results

- **What comes out:**
  - Multi-homing and routing scalability
  - Ingress traffic engineering (TE) in case of multi-homing without complex BGP configuration
- **But also a lot of other use cases:**
  - Especially interesting for enterprises
  - Disaster recovery, deployable systems
  - Mobility and GEO-redundancy
  - Connection of IPv6 islands over IPv4 infrastructure, transition to IPv6
  - Virtualization, VPN
    - Cloud computing as combination of mobility, multi-tenancy and segmentation (VPN)
    - VM mobility (VM move across IP subnets instead of subnet extension)
  - LISP mobile node
  - And many others to be discovered
- **Easy start**
  - No changes at the end systems
  - No changes in the IP WAN (service provider) infrastructure
  - LISP capable routers at the border only
  - Incremental deployment possible with benefiting from LISP day-one by usage of proxies

# Information about LISP

---

- [www.lisp4.net](http://www.lisp4.net)
- [lisp.cisco.com](http://lisp.cisco.com)
- IETF RFC 6830 - 6836
- [OpenLISP.org](http://OpenLISP.org)
- [LISPmob.org](http://LISPmob.org)

# IP Paradigms and their Consequences 1

- **Connectionless (CL) Packet Switching**

- “Store and Forward” of IP datagrams

- Queues in case more traffic arrives at a router than can be passed on (forwarded)
- Forwarding decision based on “signposts”
  - Routing table contains next hop in order to reach a given IP prefix
- Distributed control -> Forwarding decision of every router is based on own routing table
- Efficient and scalable routing
  - Needs unique and structured (and aggregate-able) addressing
- IP datagram contains
  - Global destination address for the forwarding decision per router

- Best effort service for IP datagrams

- No error recovery performed by routers
- No sequence guarantee
- Protection of the network against endless looping of IP datagrams by using TTL (Time-To-Live) field in the IP header

# IP Paradigms and their Consequences 2

- **Destination Based Routing**

- Destination IP prefix has to be in the routing table
  - Otherwise IP datagrams for that destination are deleted
- Exception: Default Route
  - Have to point to regions where IP prefix is known
  - Otherwise routing loops can occur
- To achieve line speed forwarding
  - Routing table lookup nowadays is hardware optimized
  - FIB (Forwarding Information Base)

- **Best Path Routing**

- Decision about best path based on metrics
- Metrics have static character only
  - e.g. link costs, physical bitrate, router hops, AS hops, ...

# IP Paradigms and their Consequences 3

- **More than one best path**
  - ECMP (Equal cost multiple path) can be used for loadbalancing
  - Loadbalancing has to ensure that IP datagrams of a given flow take the same path
  - ECMP support in hardware to achieve line speed forwarding
    - Lookup of certain fields within the IP datagram to create a hash number
    - Hash numbers are mapped to one of the multiple paths (next hop)
  - Implicit flow awareness of ECMP
- **Loadbalancing for unequal paths**
  - Supported by some routing protocols

# IP Paradigms and their Consequences 4

- **Dynamic routing**

- Discovering of network topology and changes by exchange of routing protocol messages among routers
  - Routing messages must be handled with highest priority
- Decision about best path(s) in case of redundancy
- Best path(s) stored in routing table
- Changes discovered
  - New IP prefix (new network)
  - Previously known IP prefix not reachable anymore
  - Failure of a link
  - Failure of a router node
  - “Blackouts”
- Changes not discovered
  - Dynamic parameters like congestion, bit error rate, link utilization
  - “Brownouts”

# IP Paradigms and their Consequences 5

- **Routing convergence**

- Time to achieve consistent routing tables in all routers of a domain
- Convergence time sums up time for
  - Detection of failures
    - Direct failures by detecting loss of physics
    - Indirect failures by timeout of certain control messages (e.g. routing hellos, BFD, ...)
  - Local decision for path switchover
  - Propagation of failures to other routers
  - Decision at other routers for path switchover
- Routing loops may occur during convergence time
  - Can lead to temporary congestion on remaining links



# IP Paradigms and their Consequences 6

- **On failure repair**
  - Automatic rerouting to former best path again
  - May lead to a temporary disruption again
- **Validation tests**
  - should include failure repair scenarios

# MPLS and IP Unicast 1

- **Brings kind of connection oriented (CO) approach into the CL IP world**
  - LSP (Label Switched Path)
- **MPLS forwarding decision**
  - Based on local labels versus global addresses
  - CO inheritance of legacy packet switching techniques
    - Local connection identifier
      - e.g. X.25 LCN, FR DLCI, ATM VPI/VCI
  - Mapping / Swapping of incoming to outgoing labels
    - Based on label switching table

# MPLS and IP Unicast 2

- **MPLS switches can forward packets**
  - Without any IP routing table lookup
- **This MPLS behavior enables useful applications**
  - Transport of Internet transit traffic within an AS without explicit knowledge about IP prefixes on internal routers
    - Internet SP
  - Transport of IPv6 traffic across an IPv4 domain
    - Internet SP, enterprise backbone network
  - Multiplexing of different IP networks over a common IP/MPLS infrastructure
    - VPNv4 service, VPNv6 service
    - Label stack technique used for transport label and service label
    - Usage of mP-BGP for label distribution of service labels

# MPLS and IP Unicast 3

---

- **MPLS is an architectural framework**
  - That decouples transport from service
- **MPLS instructing stacking**
  - Allows services that go beyond simple connectivity

# MPLS and IP Unicast 4

- **Label switching table**

- Created by LDP together with IP routing
  - Unsolicited, downstream label distribution
  - Liberal label retention mode
  - Topology driven
  - Results in MP2P paths
- Created by RSVP-TE
  - Reuses RSVP signaling system of IntServ for label mapping
    - PATH and RESV messages
  - PATH triggered by headend of LSP
  - Downstream-on-demand label distribution by RESV messages
  - Configuration driven
  - Constraint-based routing
  - Results in P2P paths

# MPLS and IP Unicast 5

- **LDP method**

- MP2P LSPs are built according IP routing
- LSPs will follow the IP traditional best path
- LDP sessions protected by TCP, maintained by LDP hellos (UDP)

- **RSVP-TE method**

- Overcomes IP traditional best path for all traffic
- Traffic splitting across alternate paths is possible
  - P2P LSPs are built according to constraints
  - Headend router builds ERO (Explicit Route Object) list based on TED (Traffic Engineering Database) constraints
    - TED is built by OSPF or IS-IS TE Extensions
    - Constraints are TE metric (different from IGP metric), link coloring, shared risk link group (SRLG), bandwidth (maximum reservable bandwidth, unreserved bandwidth per priority, setup and hold priorities, preemption)
  - RSVP establishes label mapping
  - LSPs maintained by periodical PATH/RESV messages (ip protocol 46)
  - Traffic Policing / Admission control is not performed by basic RSVP-TE in case of bandwidth constraints
  - Auto bandwidth (traffic rate measurements and periodic adjustments) as enhancement possible but optimization not capable for real-time

# MPLS and IP Unicast 6

- **RSVP-TE method (cont.)**
  - Primary LSP protected by Standby LSP for link / node protection
  - Forwarding information already established in the label switching table for alternate (backup) path
  - Fast switchover (max. 50ms) in case of failover
  - Overcome longer convergence time of IP routing protocols
  - Fast-Reroute (FRR)
- **Basic LDP discovery**
  - establishes adjacencies between directly connected neighbors
- **Targeted LDP**
  - establishes adjacencies between not directly connected neighbors, used for FRR in RLFA (remote-LFA)
- **BGP Labeled Unicast**
  - Interprovider VPN, MPLS in data center, Seamless MPLS