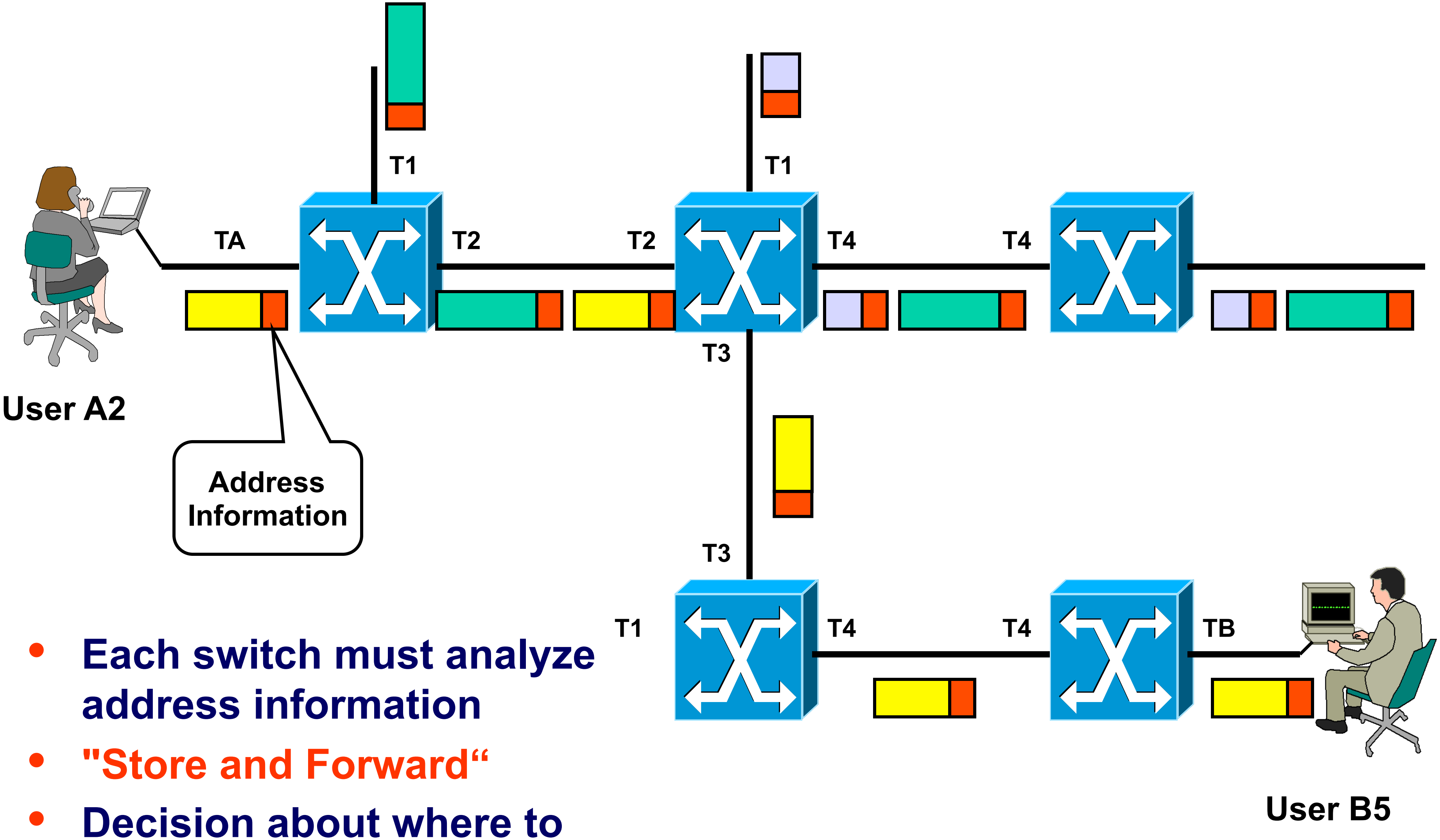
Primer IP Technology

L2 Ethernet Switching versus L3 routing
IP Protocol, IP Addressing, IP Forwarding
ARP and ICMP
IP Routing, OSPF Basics
First Hop Redundancy (HSRP)

Agenda

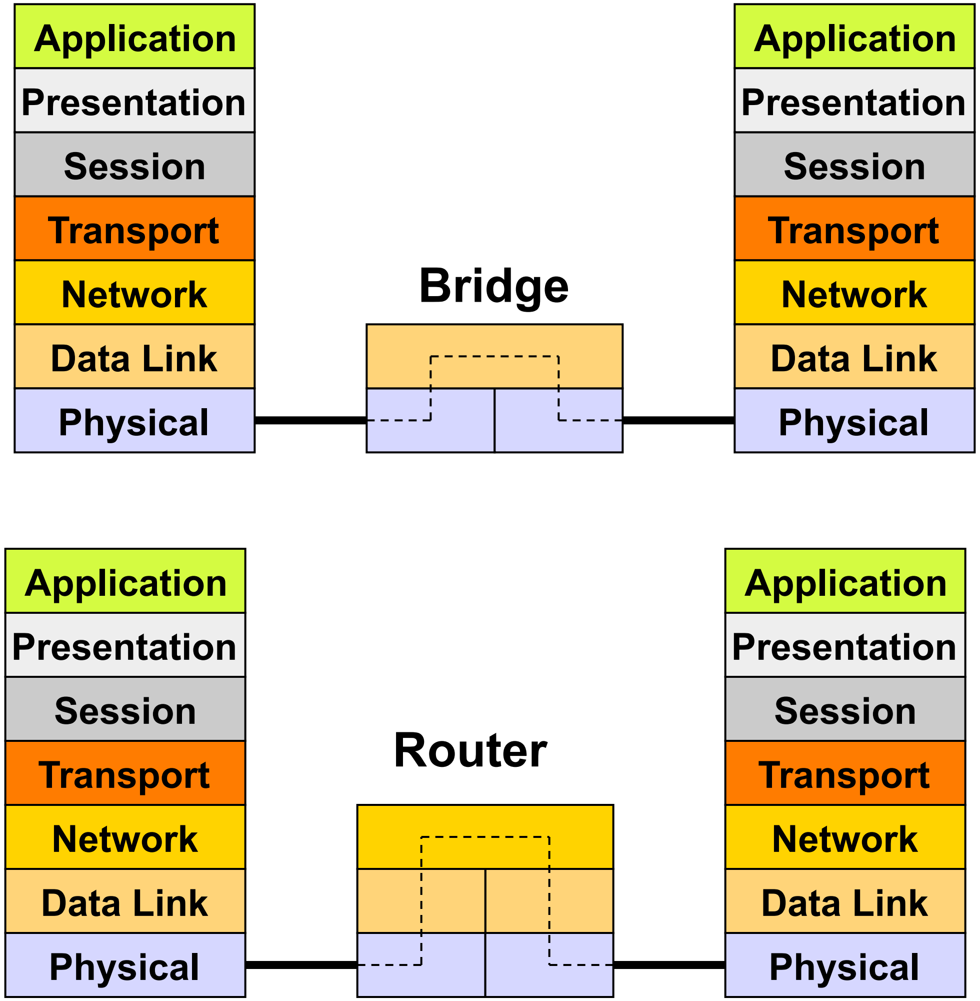
- L2 versus L3 Switching
- IP Protocol, IP Addressing
- IP Forwarding
- ARP and ICMP
- IP Routing
- First Hop Redundancy

Review Packet Switching



- Each switch must analyze address information
- "Store and Forward“
- Decision about where to forward is based on tables

Bridging (Ethernet Switching) versus IP Routing

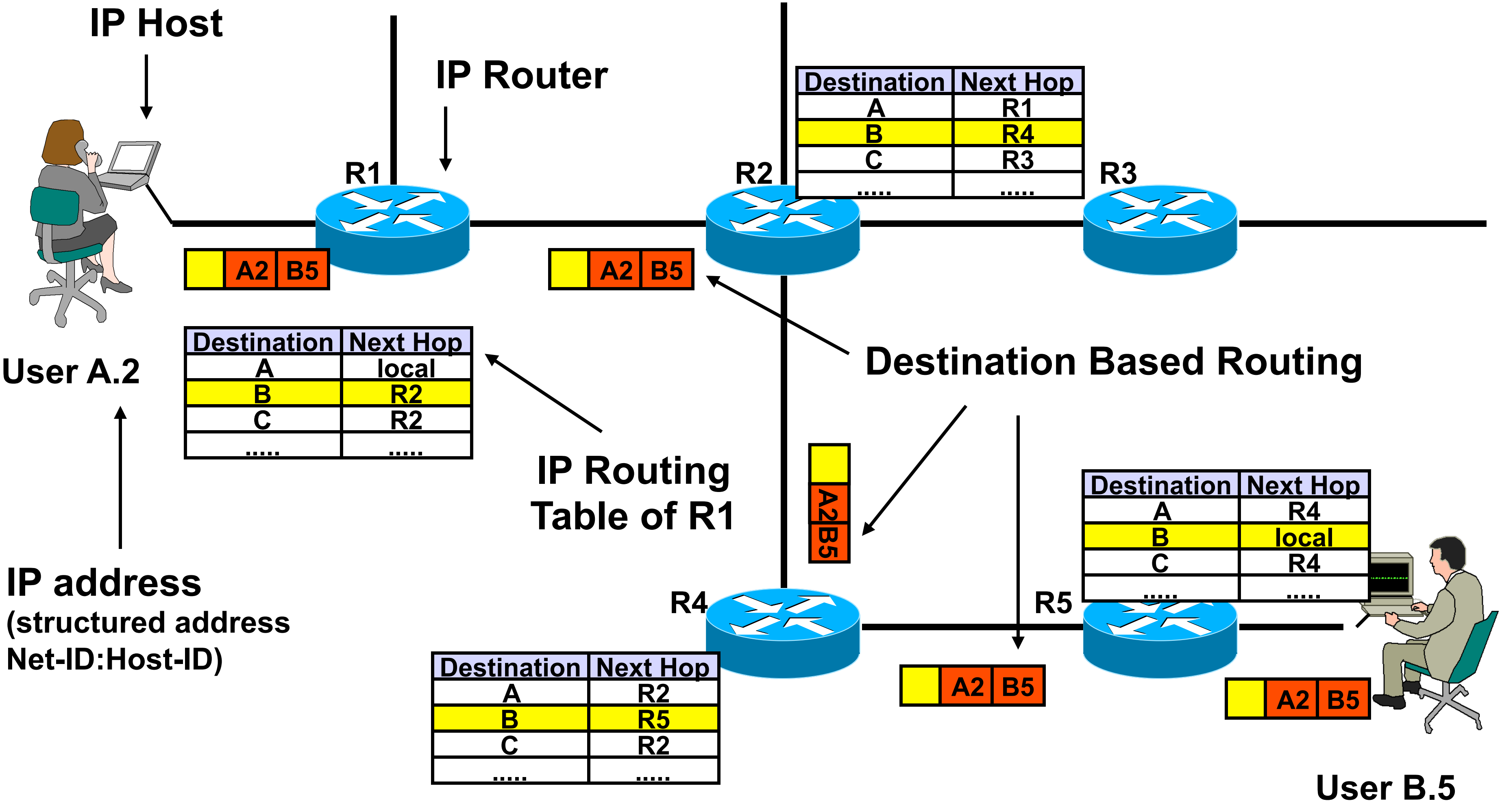


- **Bridging is**
  - Connectionless packet switching on OSI layer 2 using unique but unstructured MAC addresses without any topology information
  - Signpost in the MAC address table
- **Routing is**
  - Connectionless packet switching on OSI layer 3 using unique and structured addresses which contain topology information
  - Signpost in the routing table

IP Technology

- **IP (Internet Protocol)**
  - Packet switching technology
    - Packet switch is called router or gateway (IETF terminology)
    - End system is called IP host
    - Structured layer 3 address (IP address)
- **Datagram service**
  - Connectionless
    - Datagrams are sent without establishing a connection in advance
  - Best effort delivery
    - Datagrams may be discarded due to transmission errors or network congestion

IP Datagram Service



IP Host

IP Router

R1

R2

R3

R4

R5

Destination	Next Hop
A	R1
B	R4
C	R3
.....	.....

User A.2

Destination	Next Hop
A	local
B	R2
C	R2
.....	.....

IP Routing Table of R1

Destination Based Routing

Destination	Next Hop
A	R4
B	local
C	R4
.....	.....

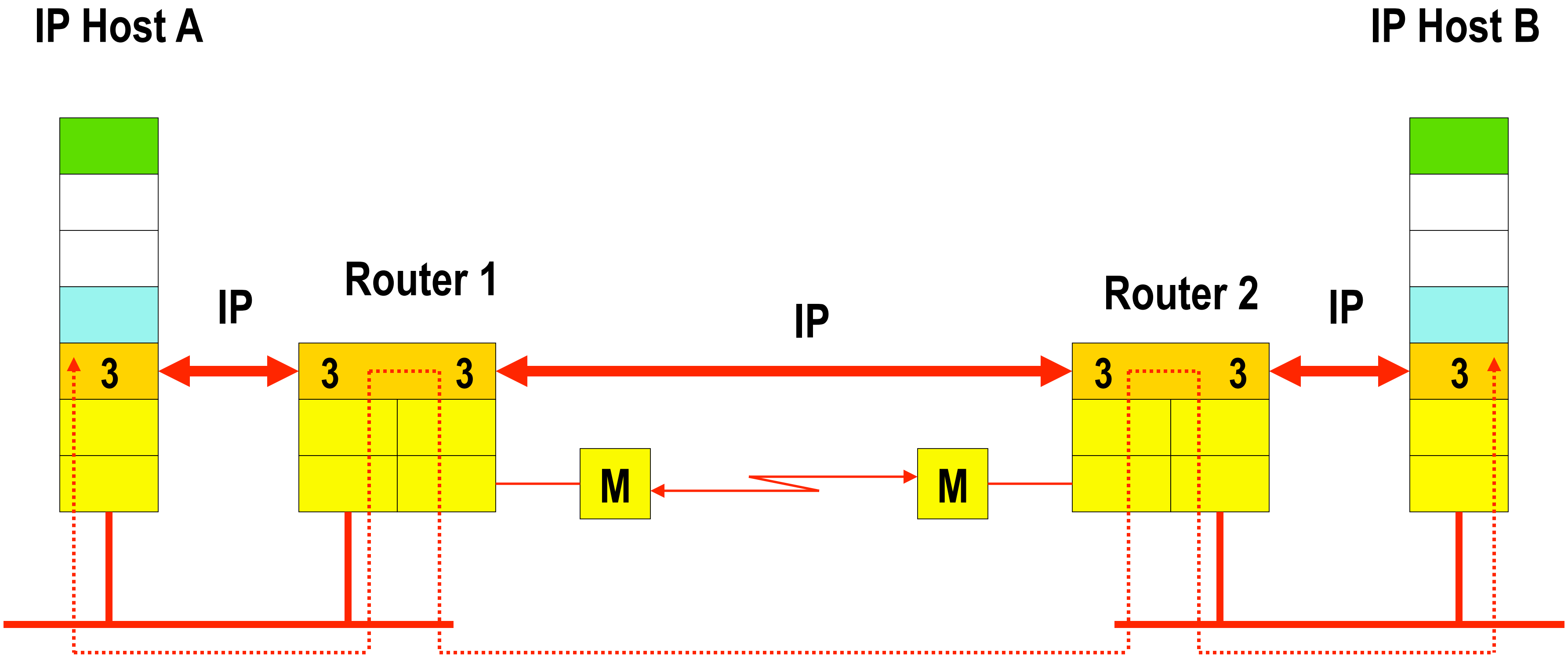
Destination	Next Hop
A	R2
B	R5
C	R2
.....	.....

User B.5

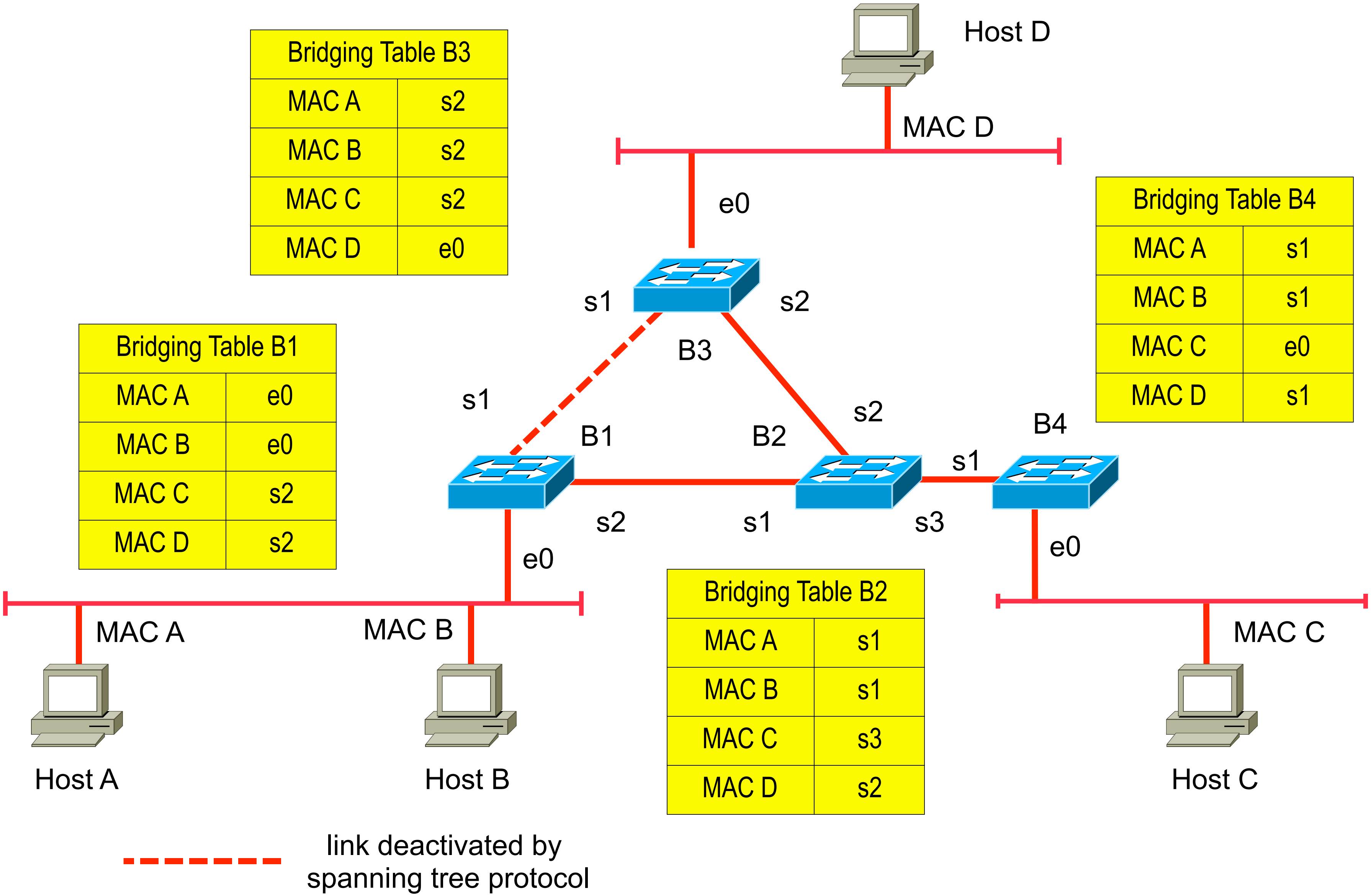
IP address
(structured address
Net-ID:Host-ID)

IP and OSI Network Layer 3

Layer 3 Protocol = IP
Layer 3 Routing Protocols = RIP, OSPF, EIGRP, BGP



Example Topology for Review Bridging



Bridging Table B3	
MAC A	s2
MAC B	s2
MAC C	s2
MAC D	e0

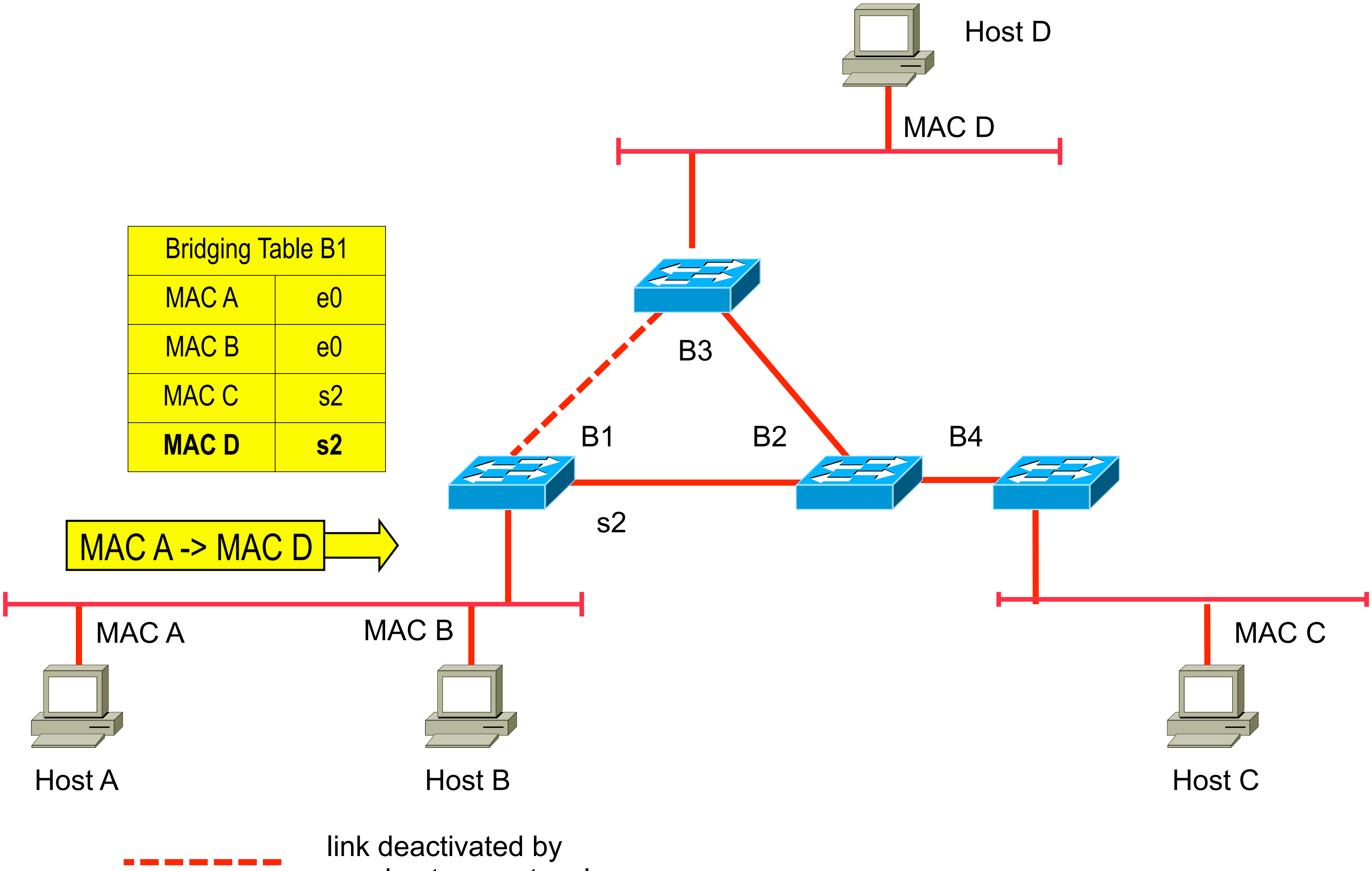
Bridging Table B1	
MAC A	e0
MAC B	e0
MAC C	s2
MAC D	s2

Bridging Table B2	
MAC A	s1
MAC B	s1
MAC C	s3
MAC D	s2

Bridging Table B4	
MAC A	s1
MAC B	s1
MAC C	e0
MAC D	s1

link deactivated by
spanning tree protocol

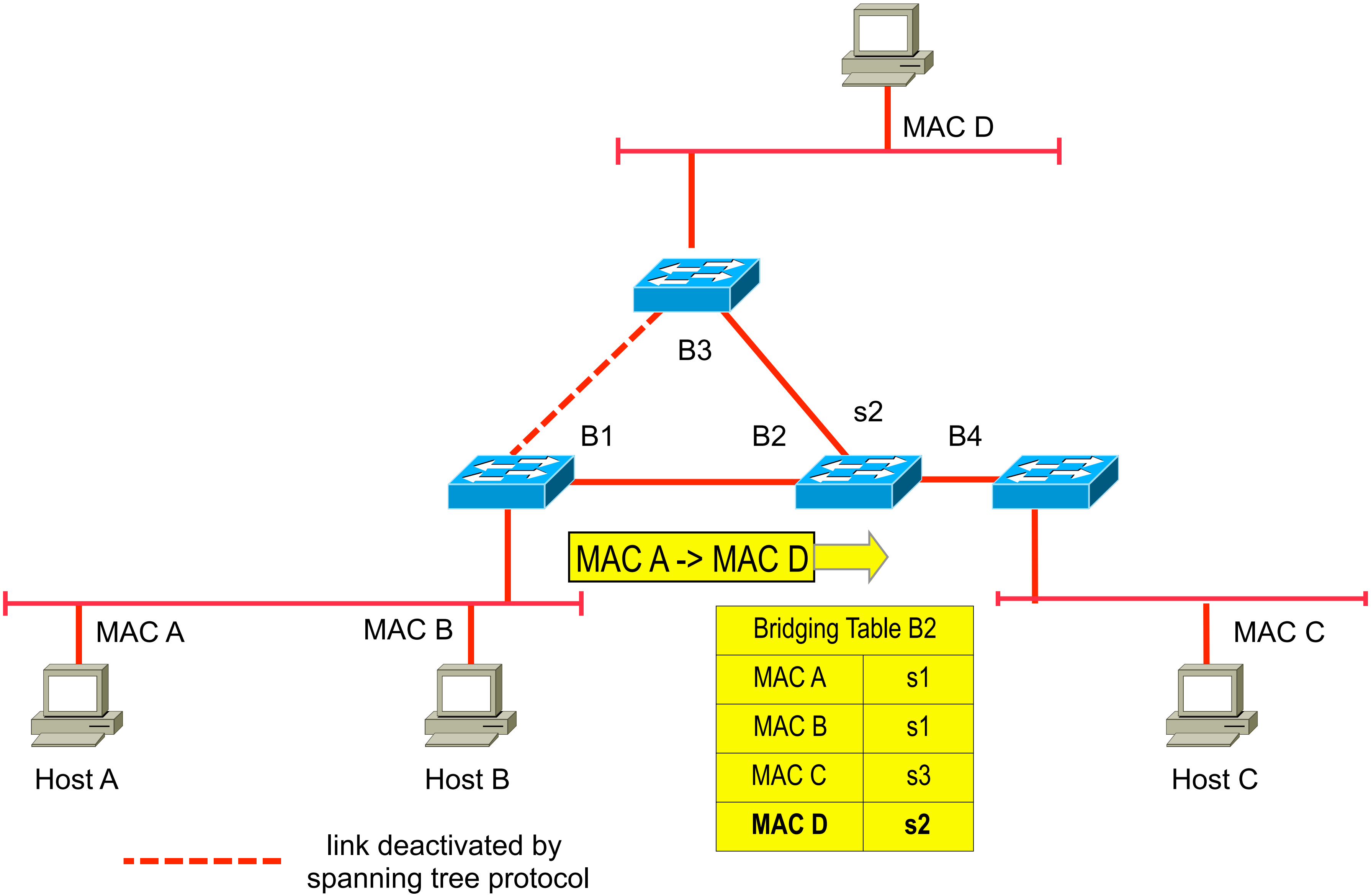
Frame MAC A to MAC D (1)



Bridging Table B1	
MAC A	e0
MAC B	e0
MAC C	s2
MAC D	**s2**

- - - - - - link deactivated by spanning tree protocol

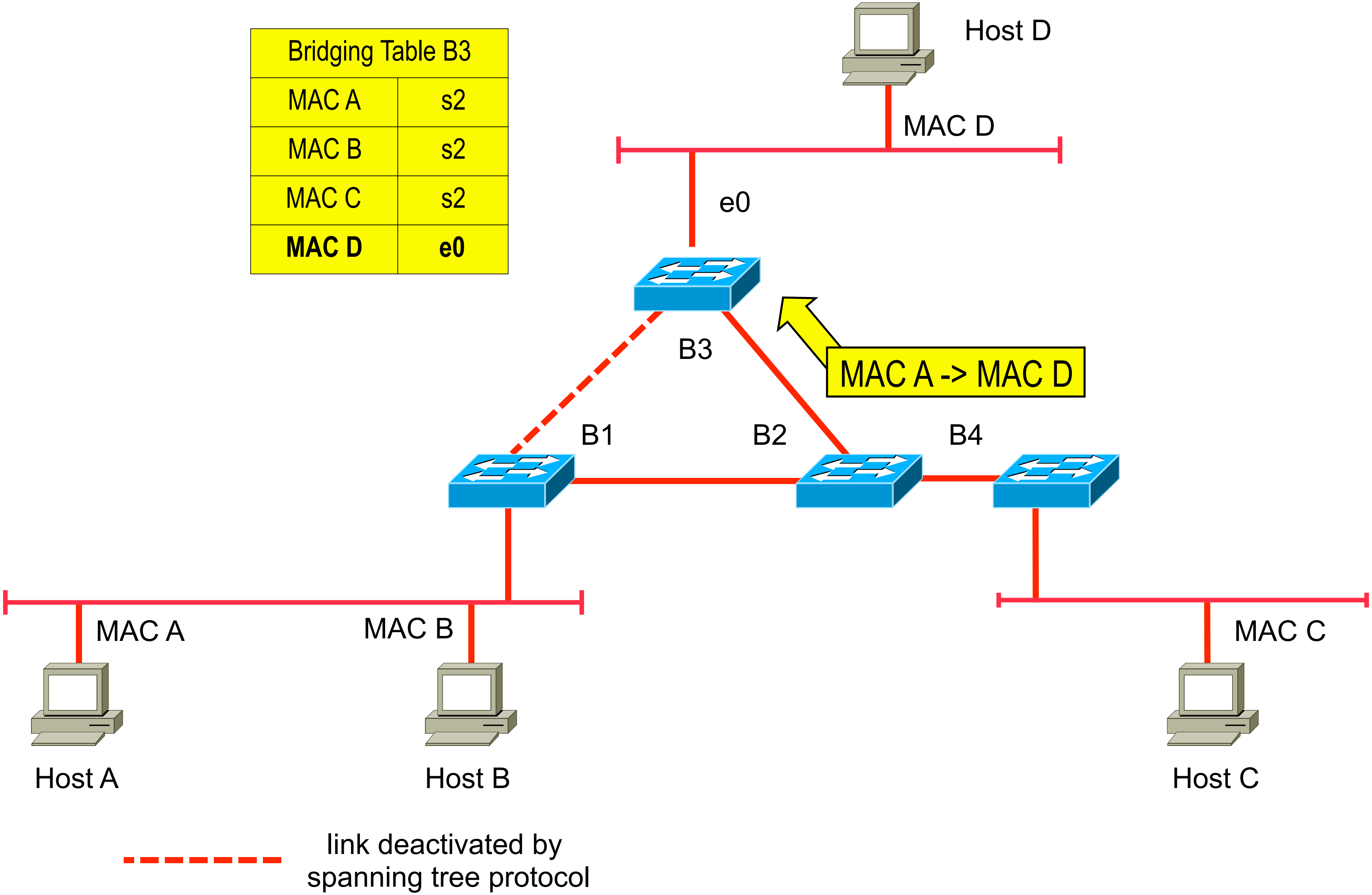
Frame MAC A to MAC D (2)



Bridging Table B2	
MAC A	s1
MAC B	s1
MAC C	s3
MAC D	**s2**

link deactivated by spanning tree protocol

Frame MAC A to MAC D (3)

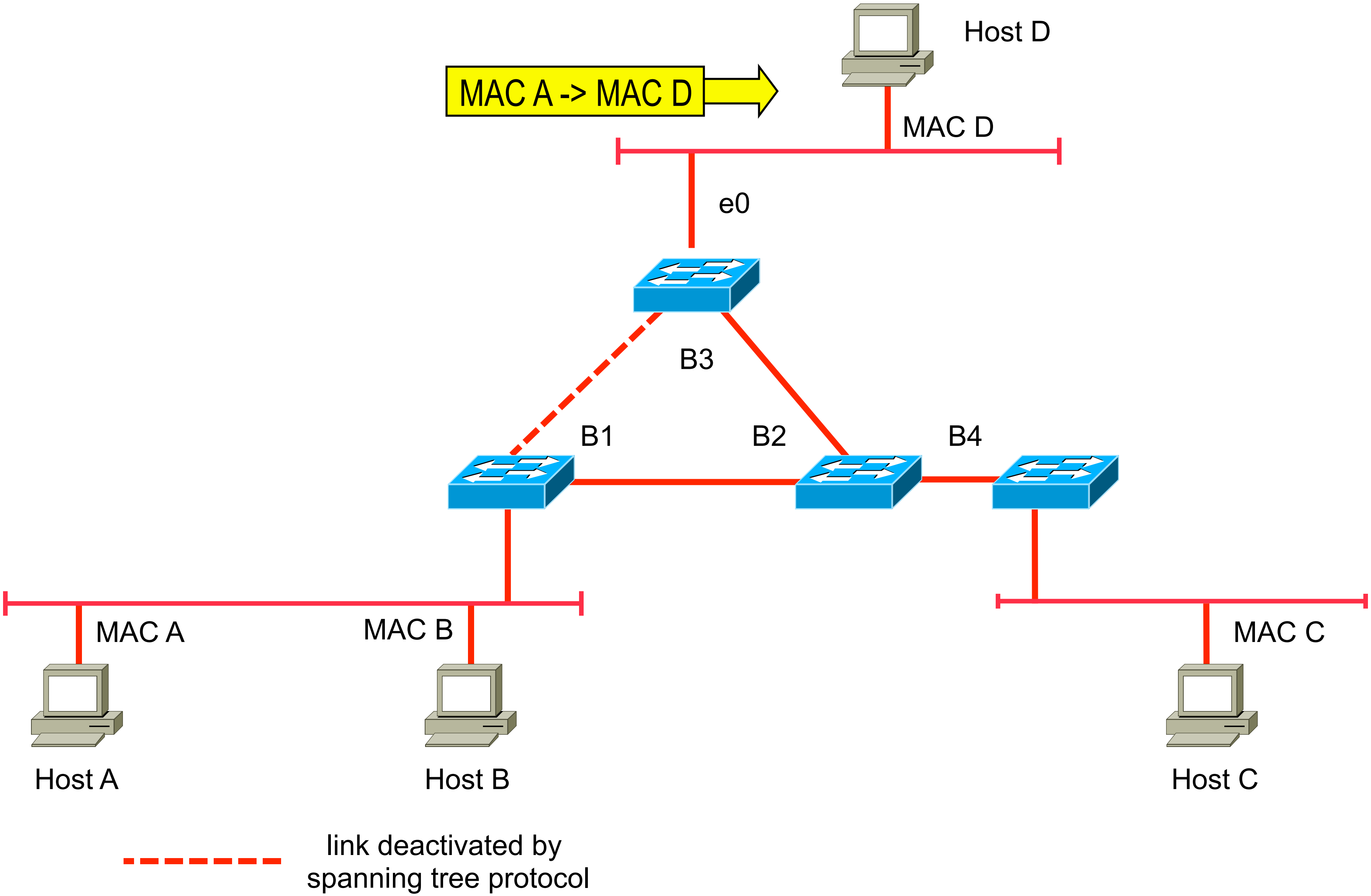


Bridging Table B3	
MAC A	s2
MAC B	s2
MAC C	s2
MAC D	**e0**

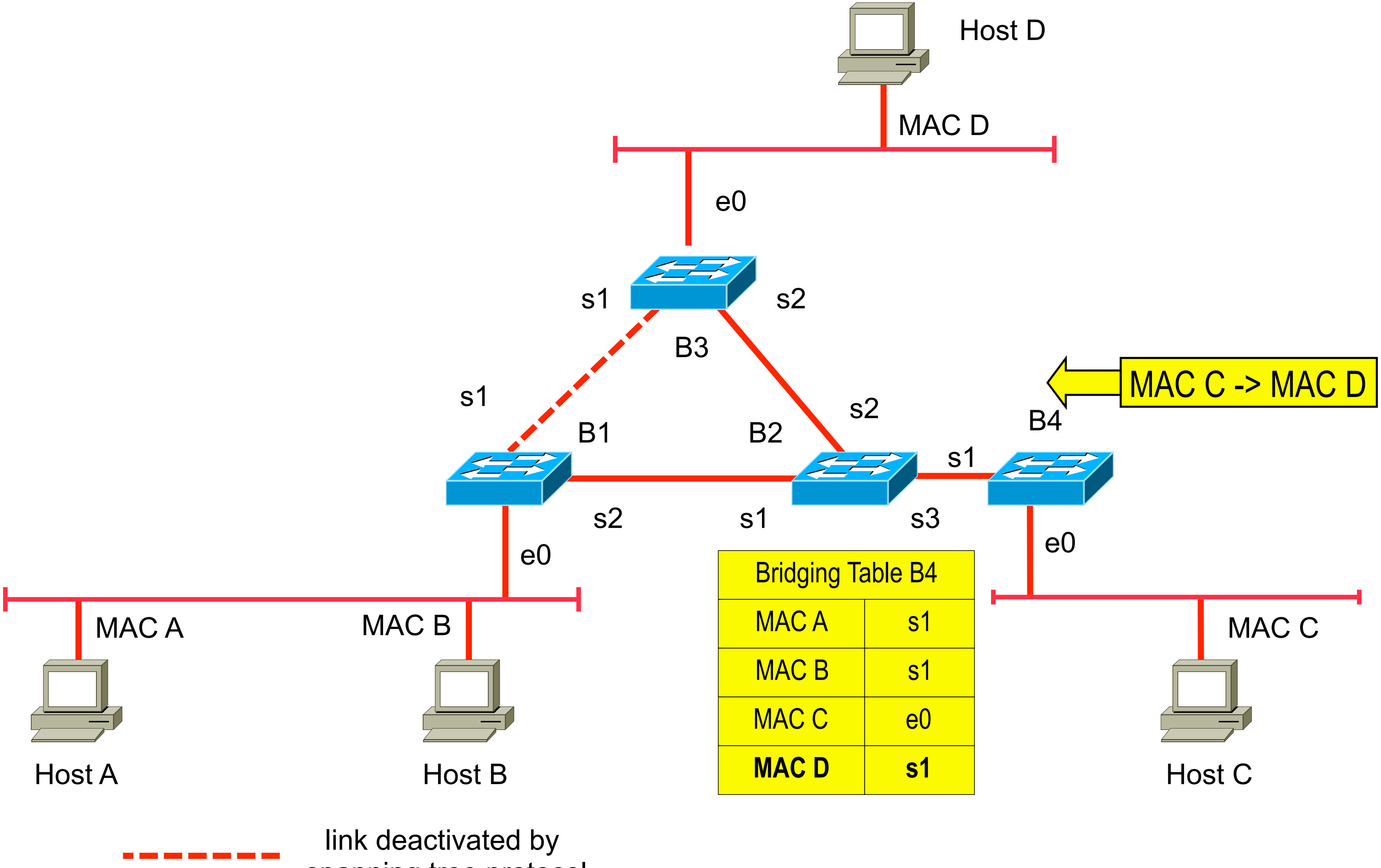
MAC A -> MAC D

link deactivated by spanning tree protocol

Frame MAC A to MAC D (4)



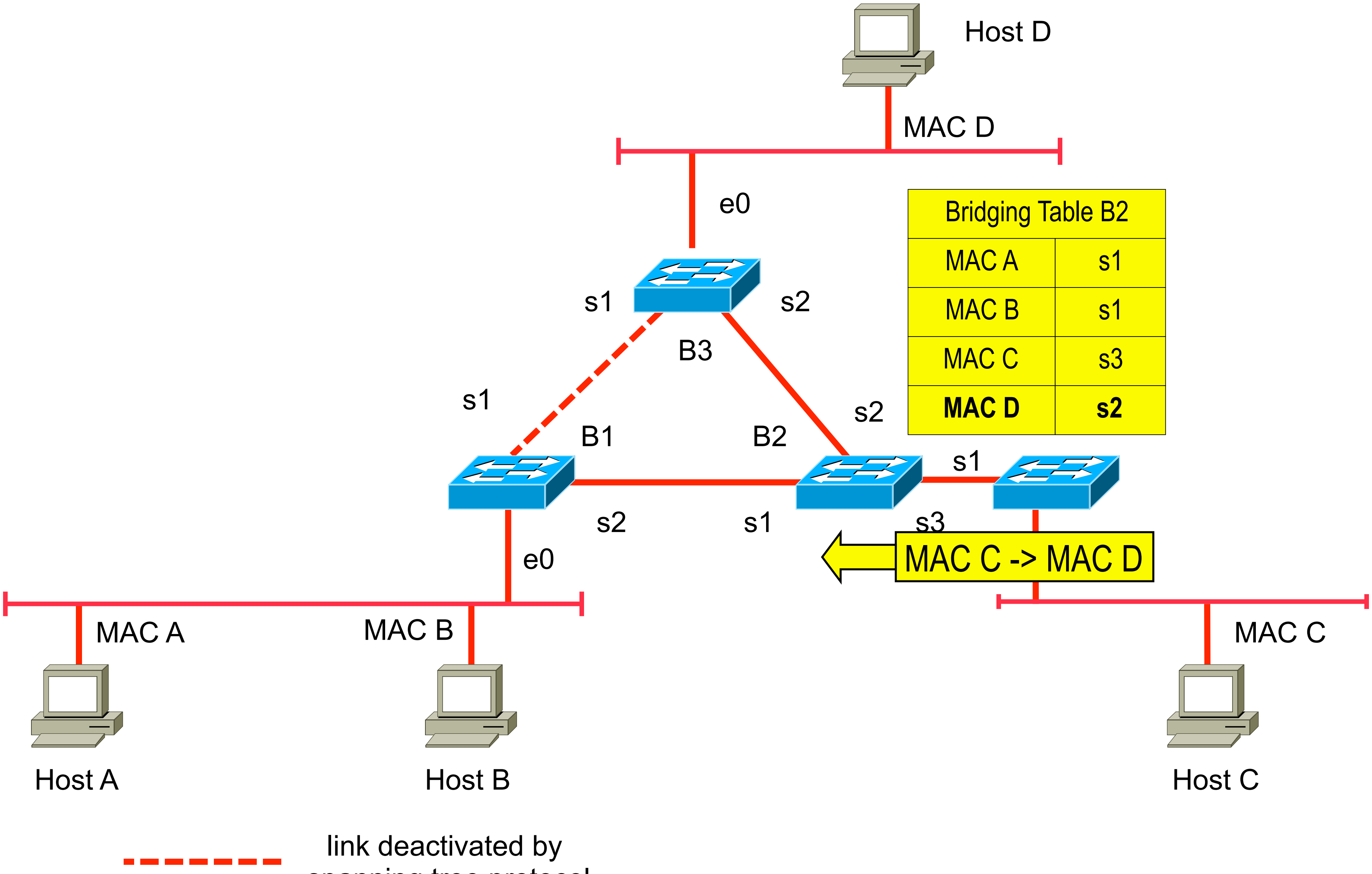
Frame MAC C to MAC D (1)



Bridging Table B4	
MAC A	s1
MAC B	s1
MAC C	e0
MAC D	**s1**

link deactivated by
spanning tree protocol

Frame MAC C to MAC D (2)

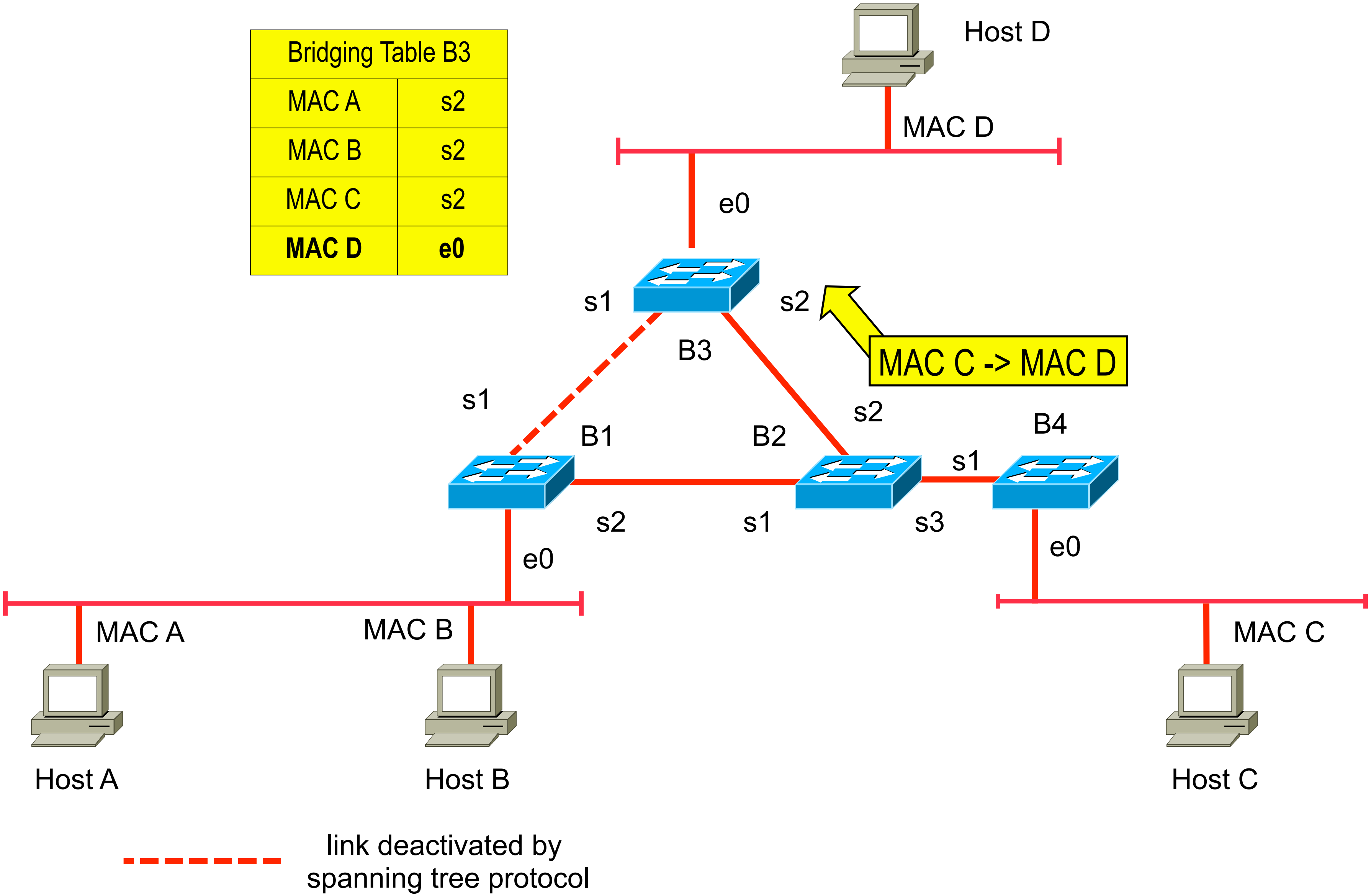


Bridging Table B2	
MAC A	s1
MAC B	s1
MAC C	s3
MAC D	**s2**

link deactivated by
spanning tree protocol

Frame MAC C to MAC D (3)

Takes Same Path as Frame from MAC A to MAC D -> No Load Distribution

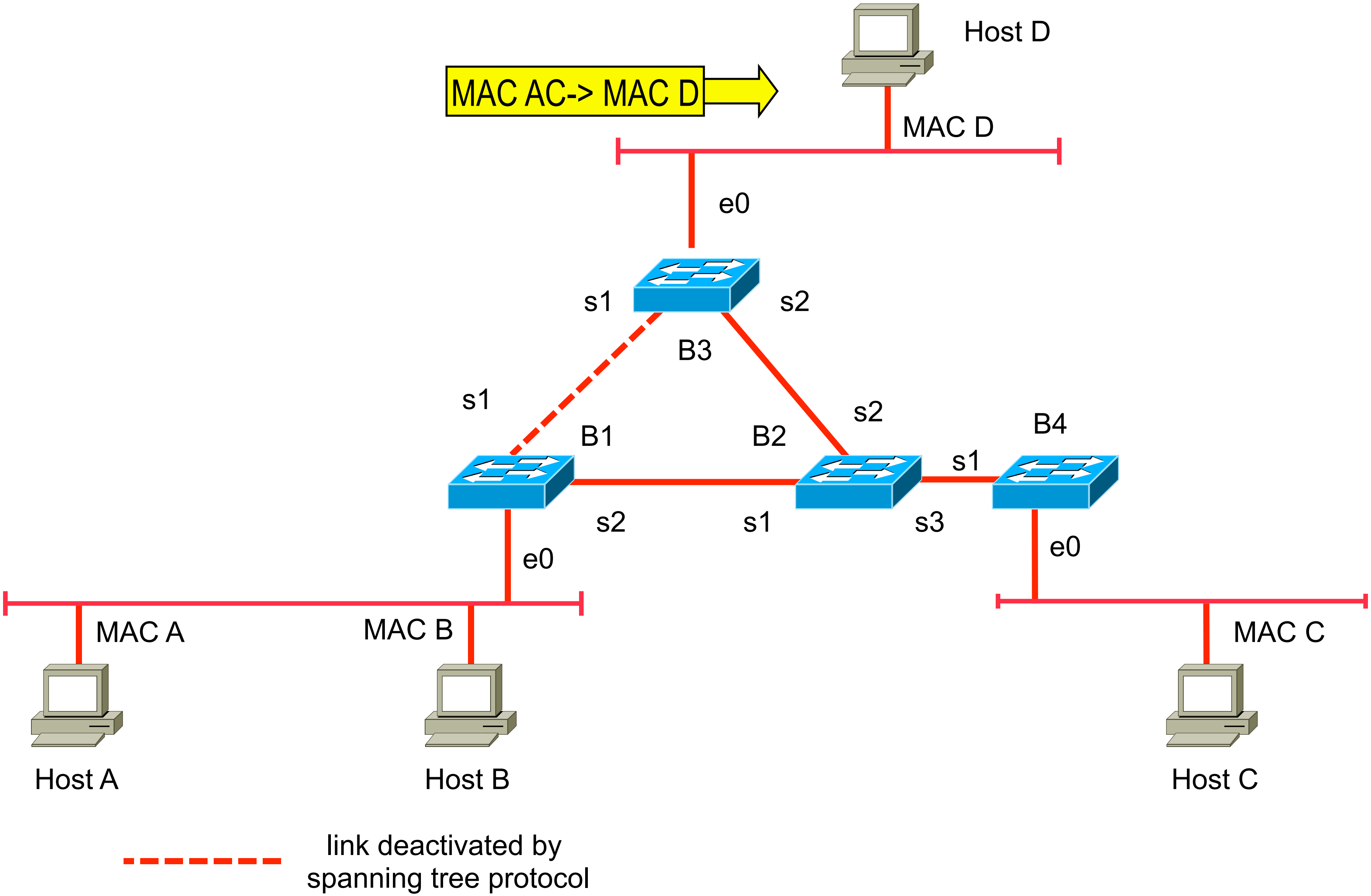


Bridging Table B3	
MAC A	s2
MAC B	s2
MAC C	s2
MAC D	**e0**

MAC C -> MAC D

link deactivated by spanning tree protocol

Frame MAC C to MAC D (4)



Requirements for Routing

- **Consistent layer-3 functionality**
  - For entire transport system
  - From one end-system over all routers in between to the other end-system
  - Hence routing is not protocol-transparent
    - all elements must speak the same „language“
- **End-system**
  - Must know about default router
  - On location change, end-system must adjust its layer 3 address
- **To keep the routing tables consistent**
  - Routers must exchange information about the network topology by using routing-protocols or network administrator has to configure static routes in all routers

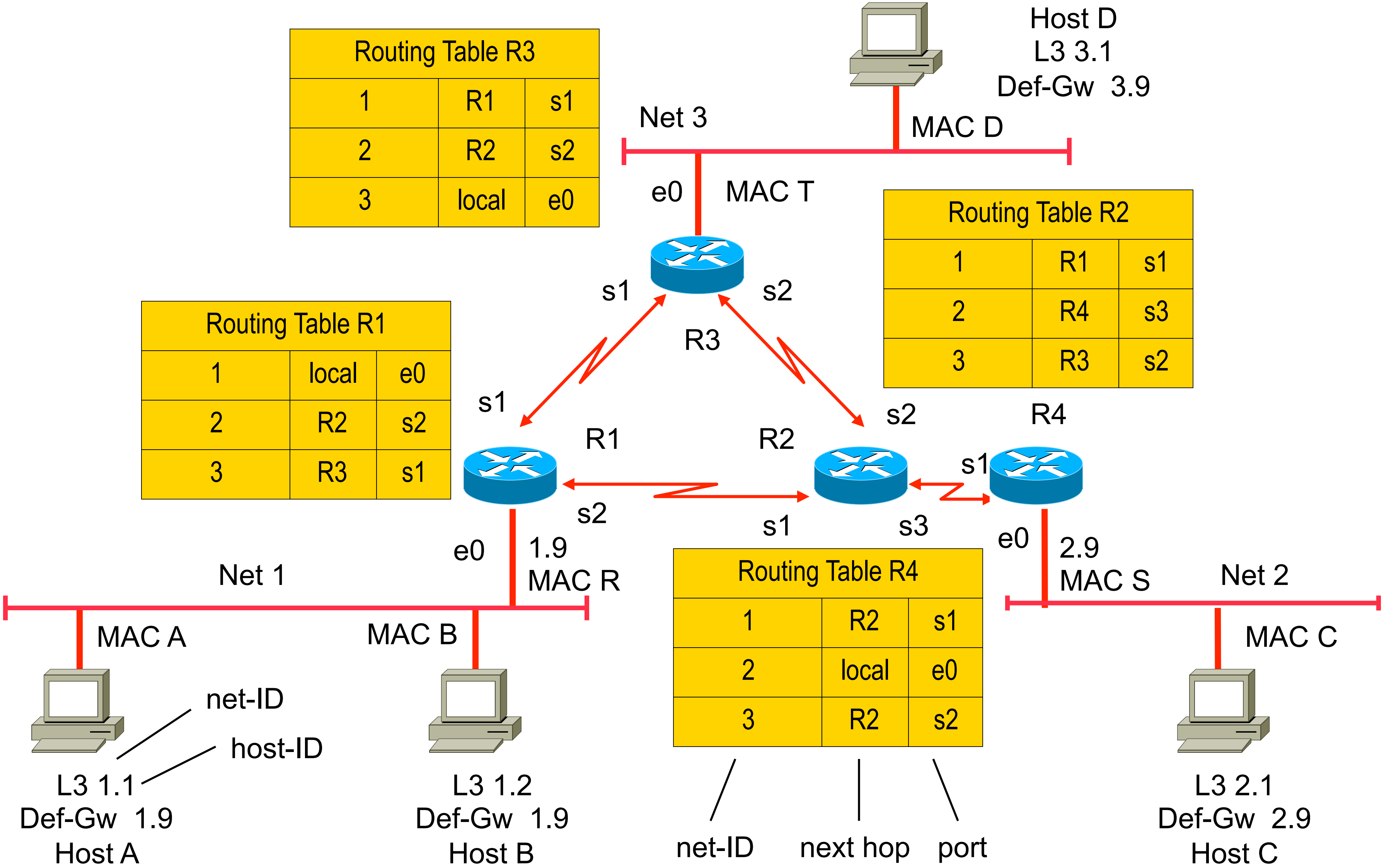
Routing Facts 1

- **In contrast to bridges**
  - Router maintains only the Net-ID of the layer 3 addresses in its routing table
  - The routing table size is direct proportional to the number of Net-IDs and not to the number of end-systems
- **Transport on a given subnet**
  - Still relies on layer 2 addresses
- **End systems forward data packets for remote destinations**
  - To a selected router (default gateway, default router) using the router's MAC-address as destination
  - Only these (unicast MAC addressed) packets must be processed by the router

Routing Facts 2

- **L2 Broadcast/multicast-packets in the particular subnet**
  - Are blocked by the router so L2 broad/multicast traffic on the subnets doesn't stress WAN connections
- **Independent of layer 1, 2**
  - so coupling of heterogeneous networks is possible
- **Routers can use redundant paths**
  - meshed topologies are usual
- **Routers can use parallel paths for load balancing**

Example Topology for Intro Routing



Routing Table R3		
1	R1	s1
2	R2	s2
3	local	e0

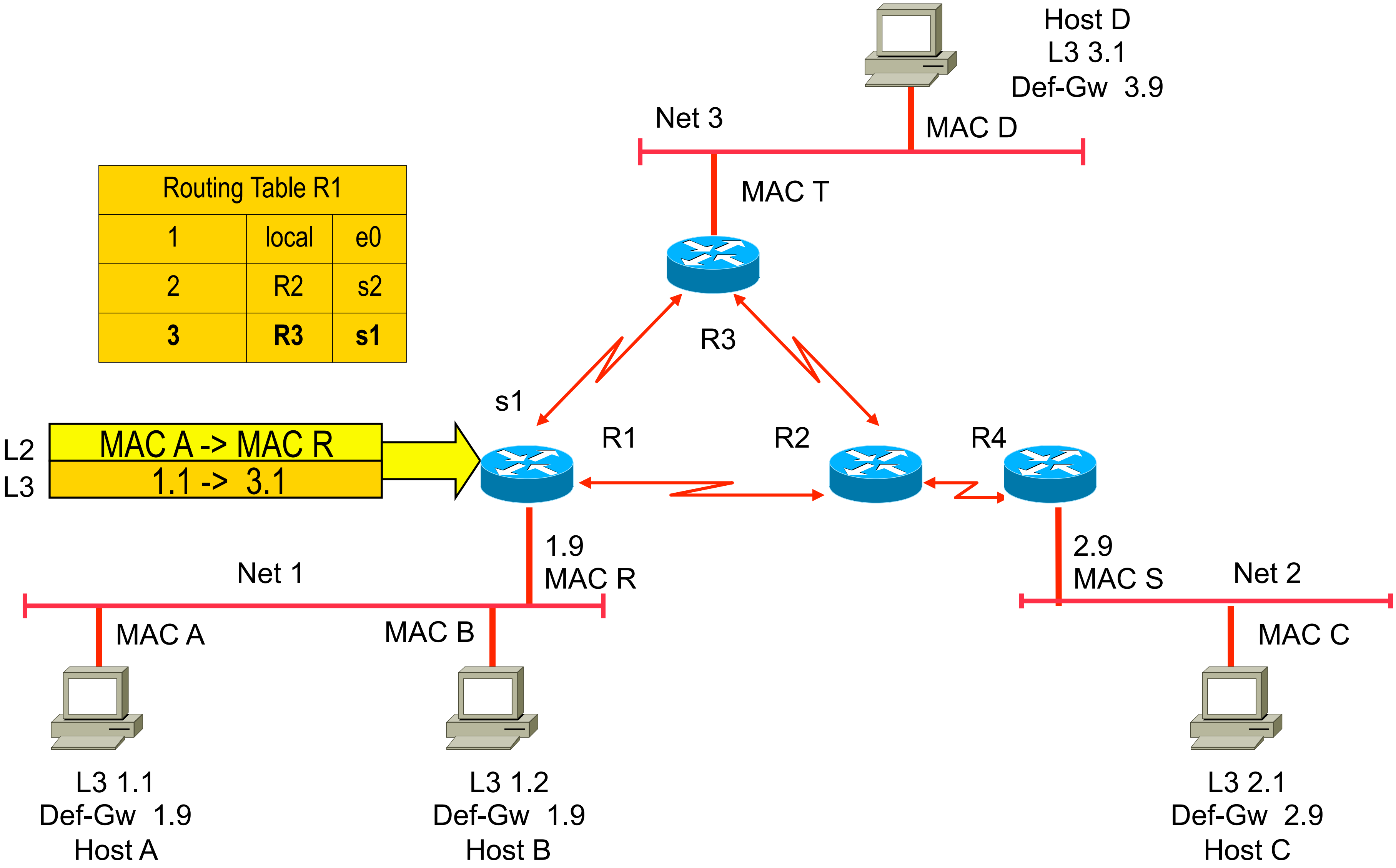
Routing Table R1		
1	local	e0
2	R2	s2
3	R3	s1

Routing Table R2		
1	R1	s1
2	R4	s3
3	R3	s2

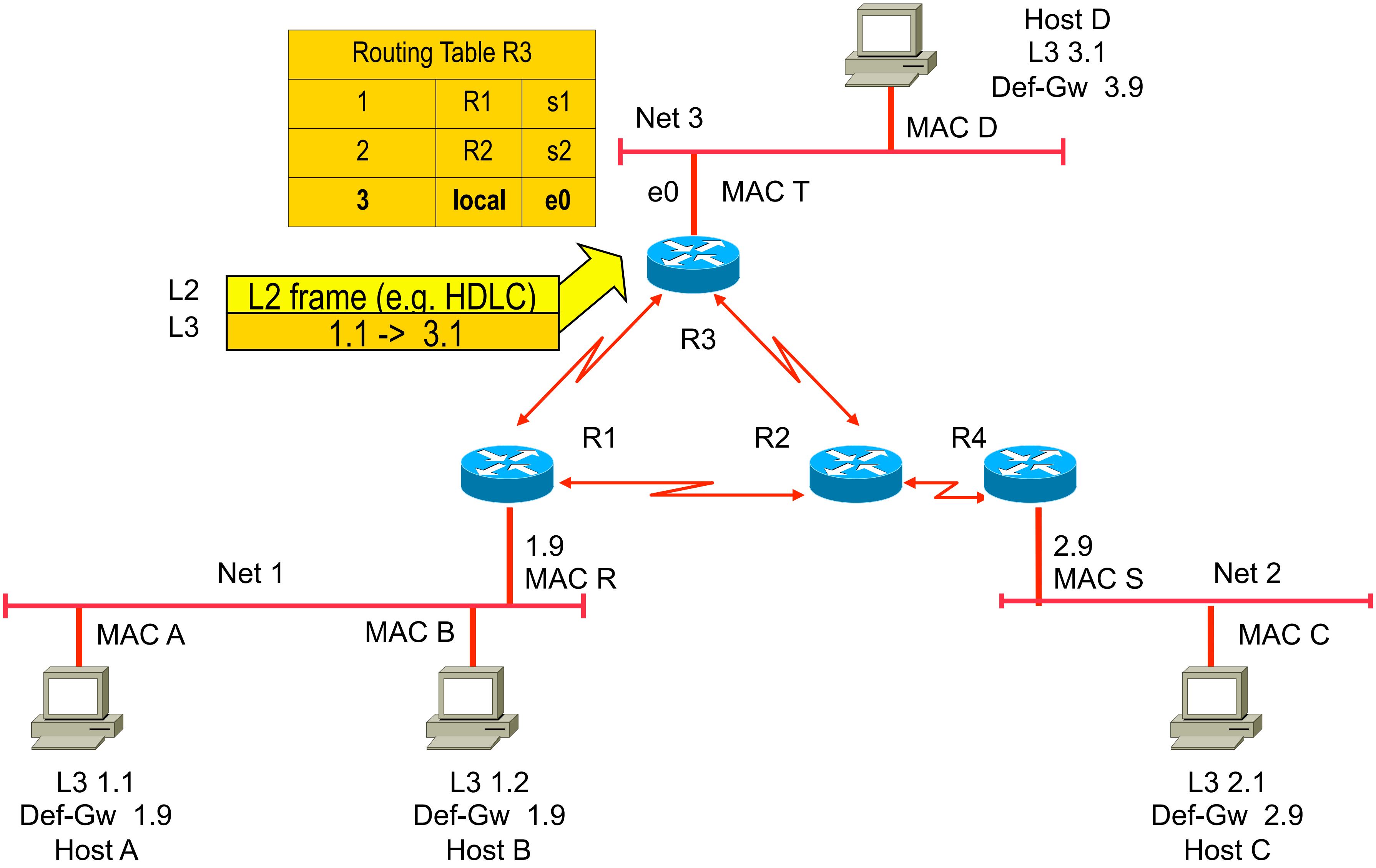
Routing Table R4		
1	R2	s1
2	local	e0
3	R2	s2

Packet 1.1 to 3.1 (1)

Routing Table R1		
1	local	e0
2	R2	s2
3	**R3**	**s1**



Packet 1.1 to 3.1 (2)

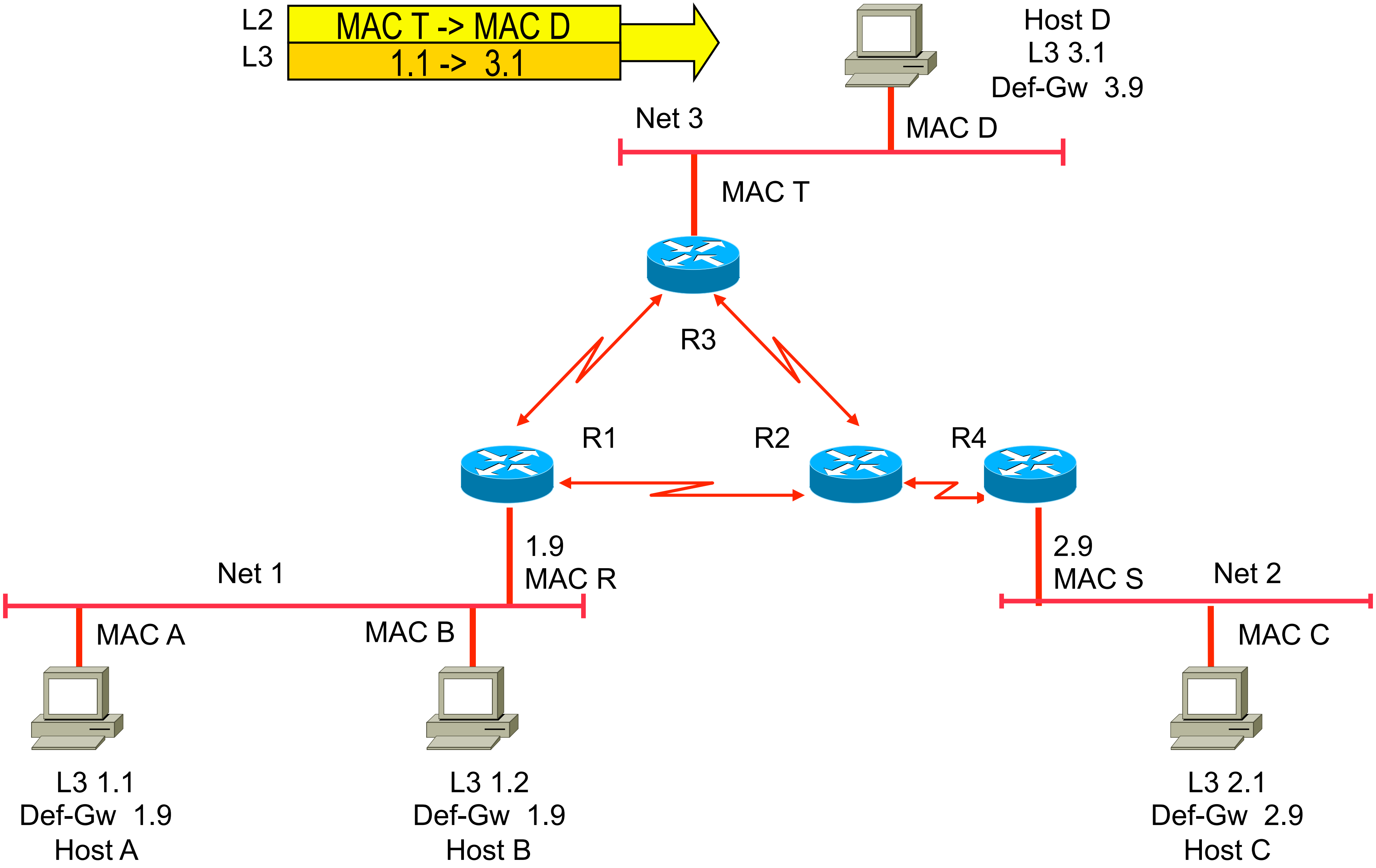


Routing Table R3		
1	R1	s1
2	R2	s2
3	**local**	**e0**

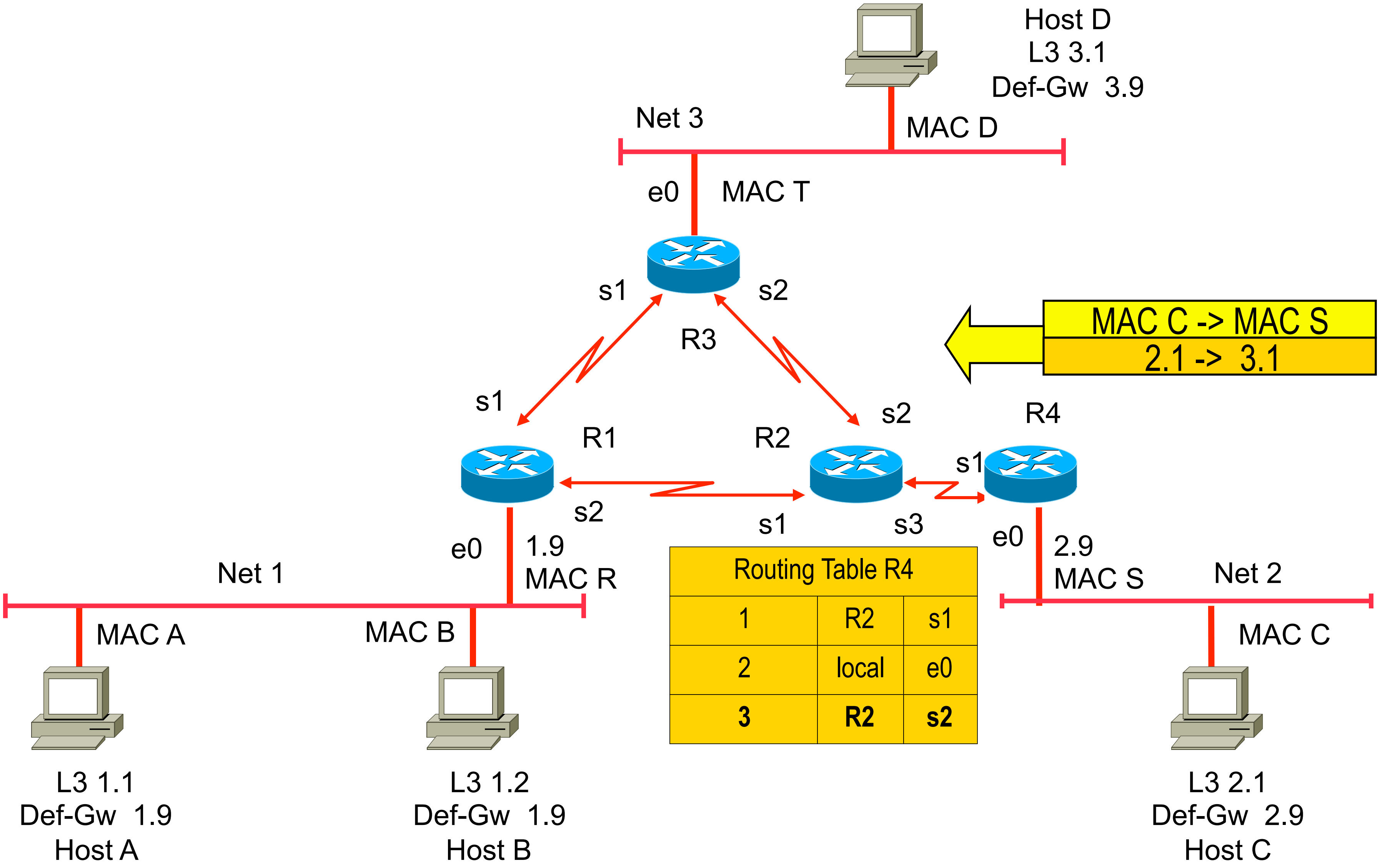
L2 L2 frame (e.g. HDLC)

L3 1.1 -> 3.1

Packet 1.1 to 3.1 (3)



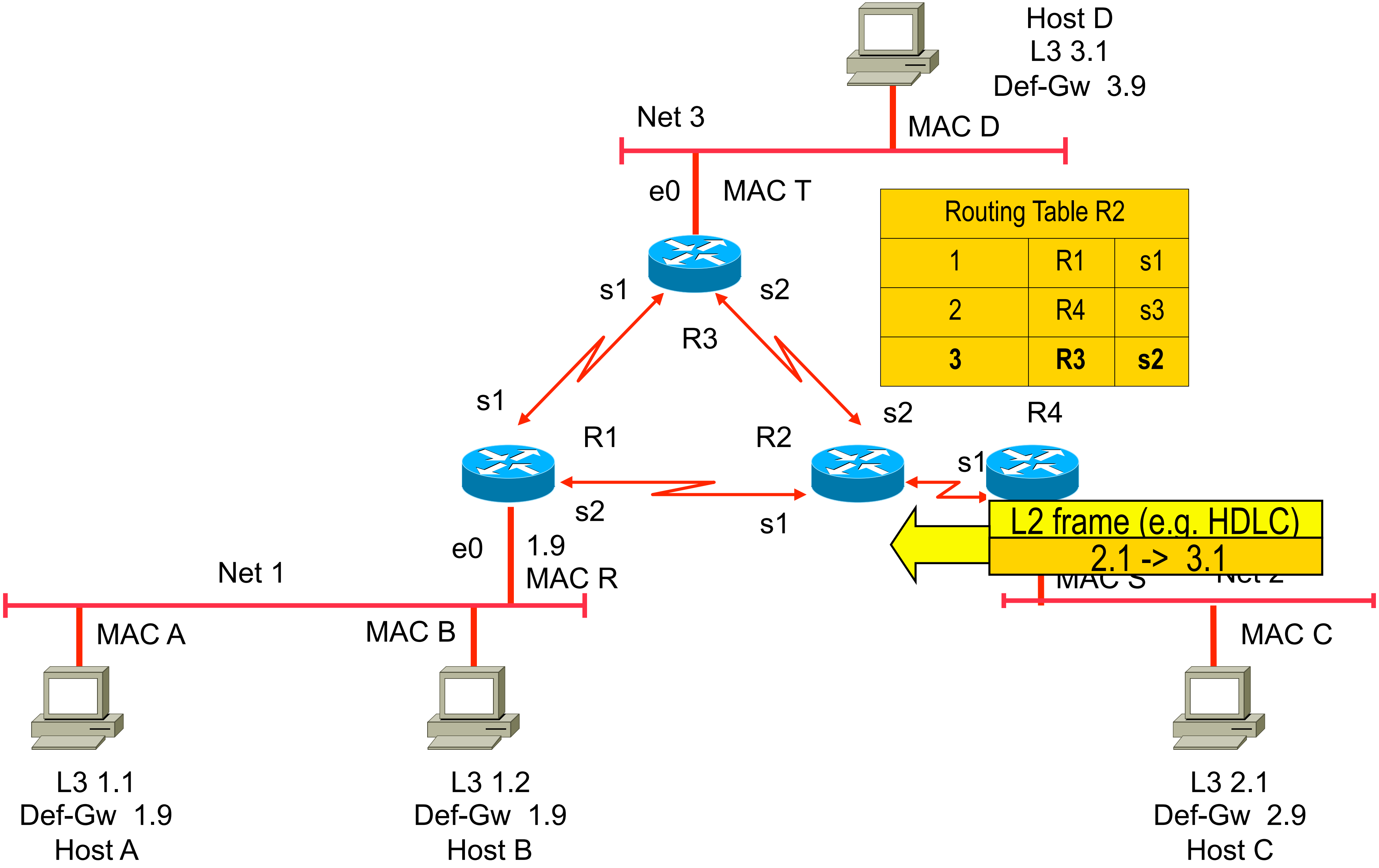
Packet 2.1 to 3.1 (1)



Routing Table R4

1	R2	s1
2	local	e0
3	**R2**	**s2**

Packet 2.1 to 3.1 (2)



Routing Table R2		
1	R1	s1
2	R4	s3
3	**R3**	**s2**

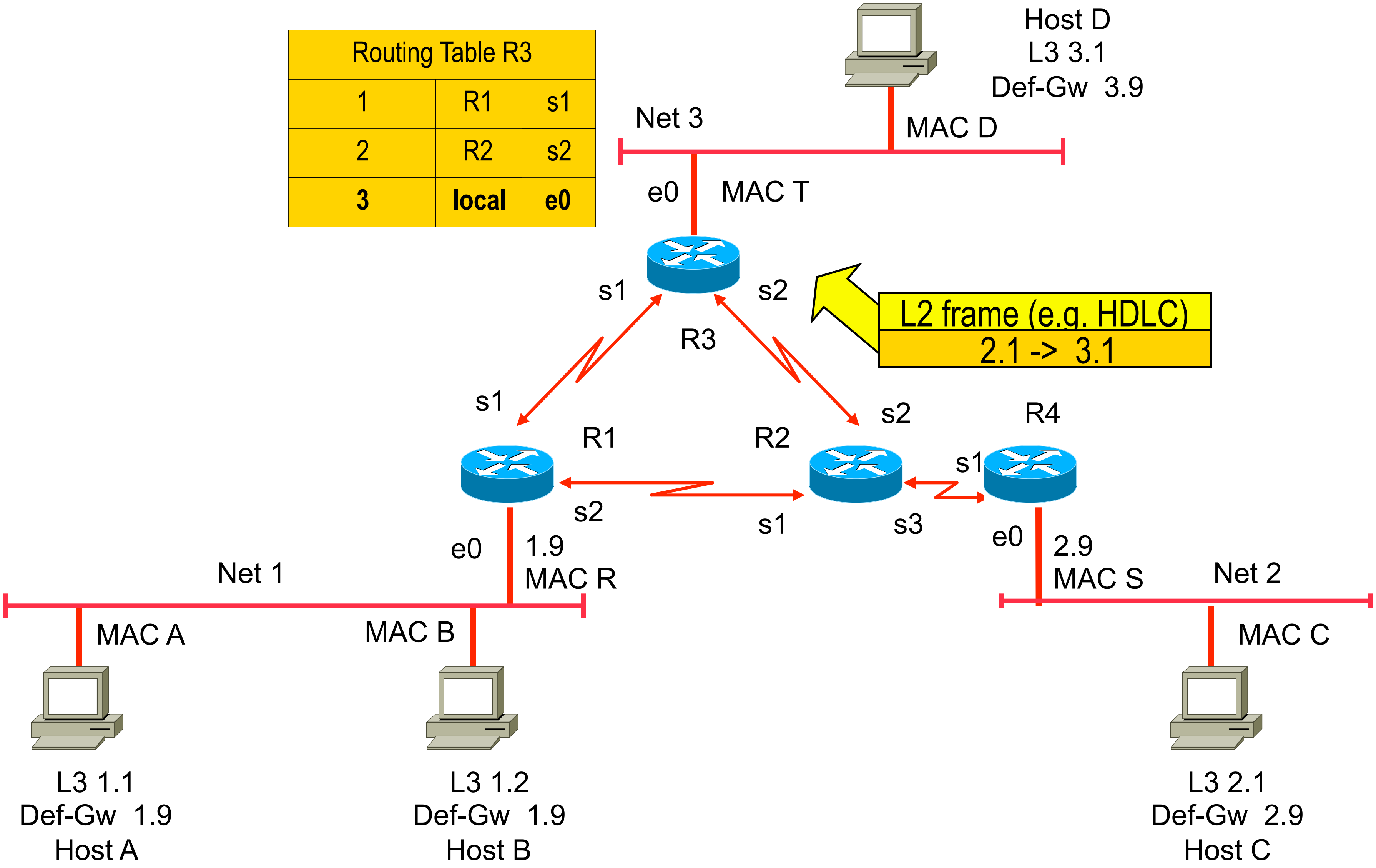
L2 frame (e.g. HDLC)

2.1 -> 3.1

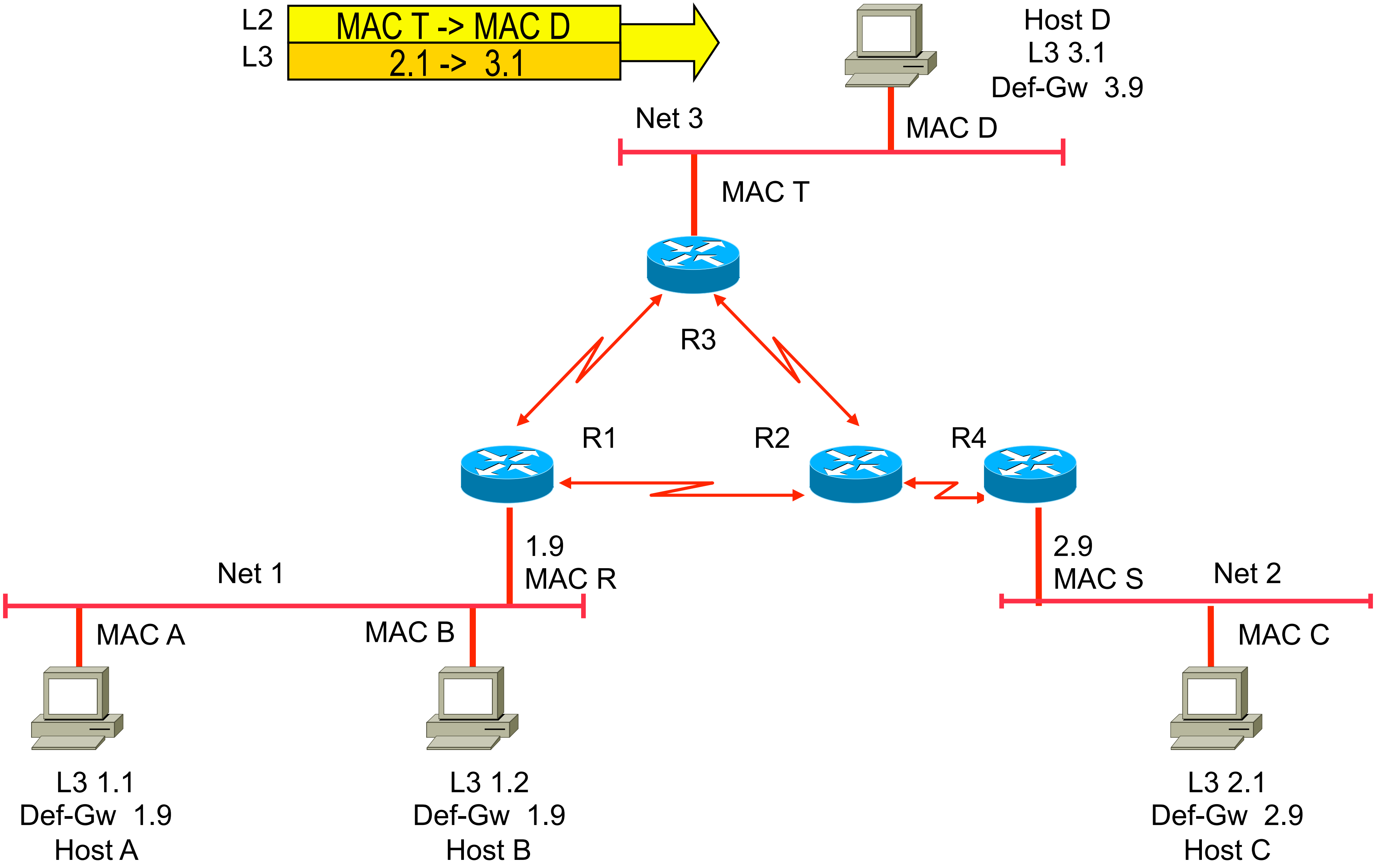
Packet 2.1 to 3.1 (3)

Takes Different Path as Packet from 1.1 to 1.3 -> Load Distribution

Routing Table R3		
1	R1	s1
2	R2	s2
3	**local**	**e0**



Packet 2.1 to 3.1 (3)



Bridging versus Routing

	Bridging		Routing
⊕	Depends on MAC addresses only	⊖	Requires structured addresses (must be configured)
⊕	Invisible for end-systems; transparent for higher layers	⊖	End system must know its default-router
⊖	Bridge must process every frame	⊕	Router processes only packets addressed to it
⊖	Number of table-entries = number of all devices in the whole network	⊕	Number of table-entries = number of IP networks (Net-IDs) only
⊖	Spanning Tree eliminates redundant lines; no load balance is possible	⊕	Redundant lines and load balance are possible
⊖	No flow control (may be changed by usage of MAC Pause command)	⊖	Flow control is possible in theory (router is seen by end systems) but ICMP source quench is not the right way

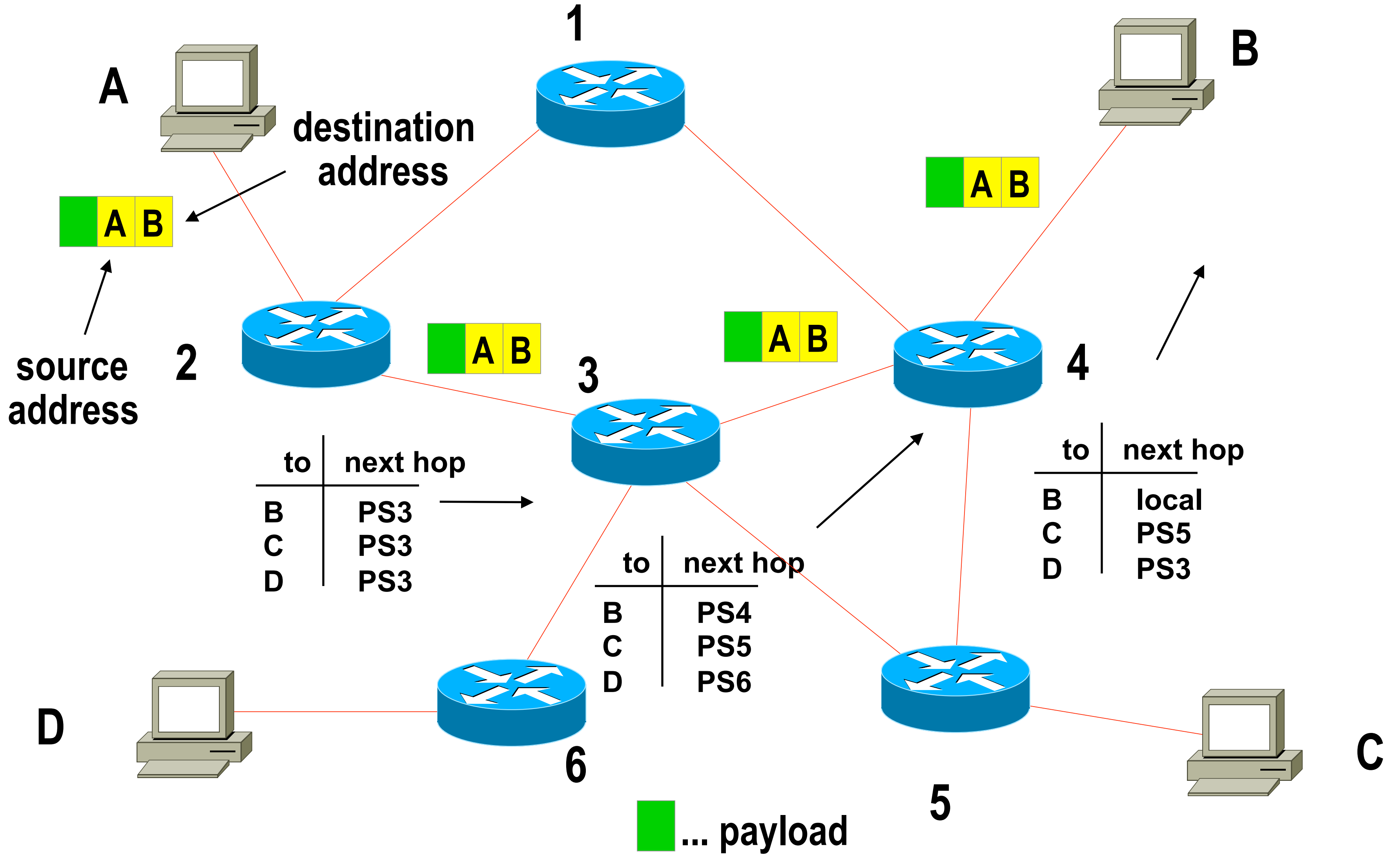
Bridging versus Routing

Bridging	Routing
− No LAN/WAN coupling because of high traffic (broadcast domain!)	+ Does not stress WAN with subnet's L2 broad- or multicasts; commonly used as "gateway"
− Paths selected by STP may not match communication behavior/needs of end systems	+ Router knows best way for every destination a packet is sent for
+ Faster, because implemented in HW; no address resolution	− Slower, because usually implemented in SW; address resolution (ARP) necessary; hardware-optimization overcomes this nowadays
+ Location change of an end-system does not require updating any addresses	− Location change of an end-system requires adjustment of layer 3 address
− Spanning tree necessary against endless circling of frames and broadcast storms, STP lacks from a global view of the network topology	− Routing-protocols necessary to determine network topology, modern link-state routing has network topology database in router and hence a global view

Datagram Service Principles

- **Connectionless service**
  - Packets can be sent without establishing a logical connection between end systems in advance
  - Packets have no sequence numbers
  - They are called **"Datagrams"**
- **Every incoming datagram**
  - Is processed independently regarding to all other datagrams by the packet switches
- **The forwarding decision for incoming packets**
  - Depends on the current state of the routing table
- **Each packet contains**
  - Complete address information (source and destination)

Packet forwarding (CL) is based on routing tables only
(Datagram Service, Best-Effort Service)

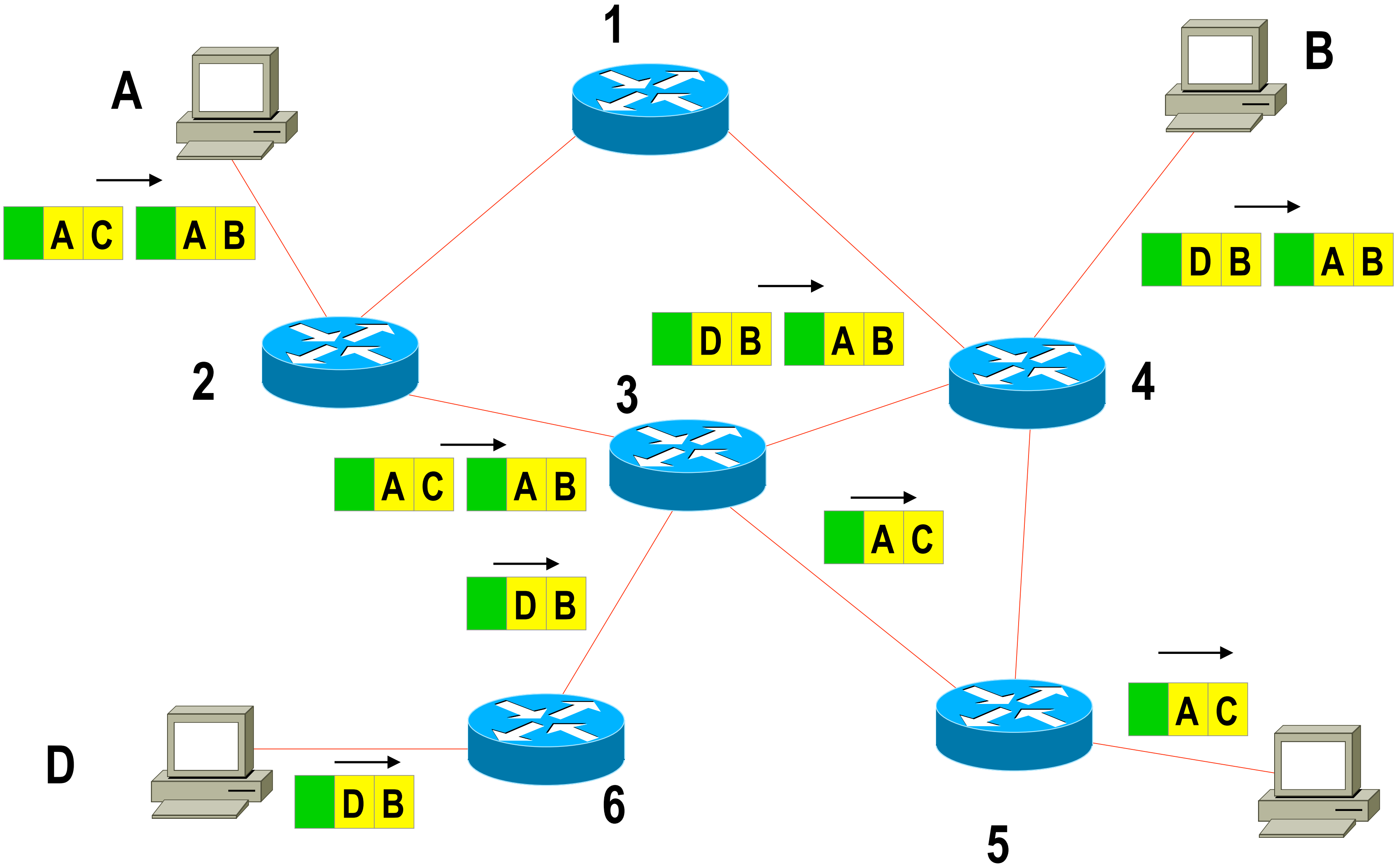


to	next hop
B	PS3
C	PS3
D	PS3

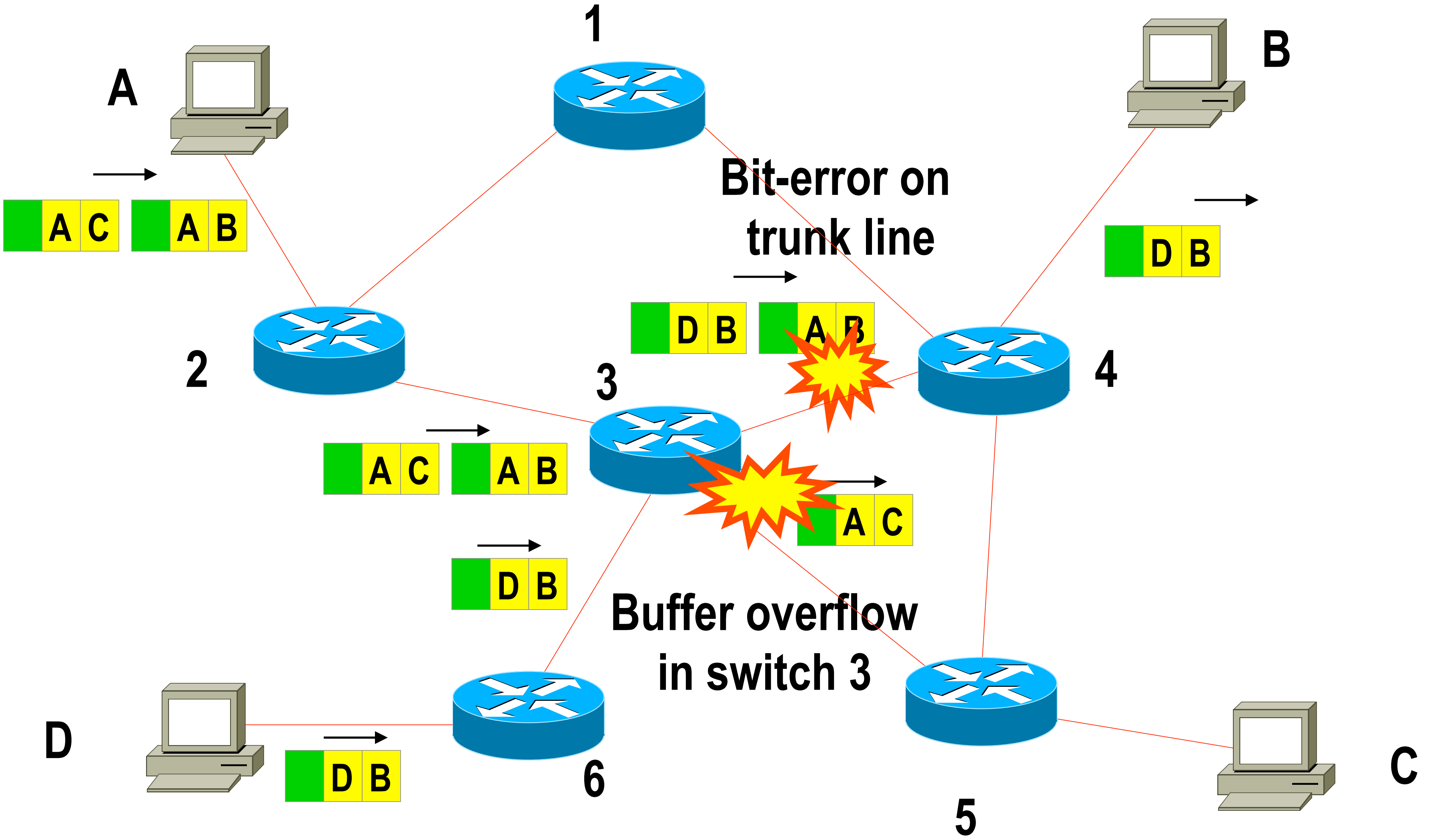
to	next hop
B	PS4
C	PS5
D	PS6

to	next hop
B	local
C	PS5
D	PS3

Datagrams are forwarded completely independent from each other based on current state of routing tables



Best-Effort Service



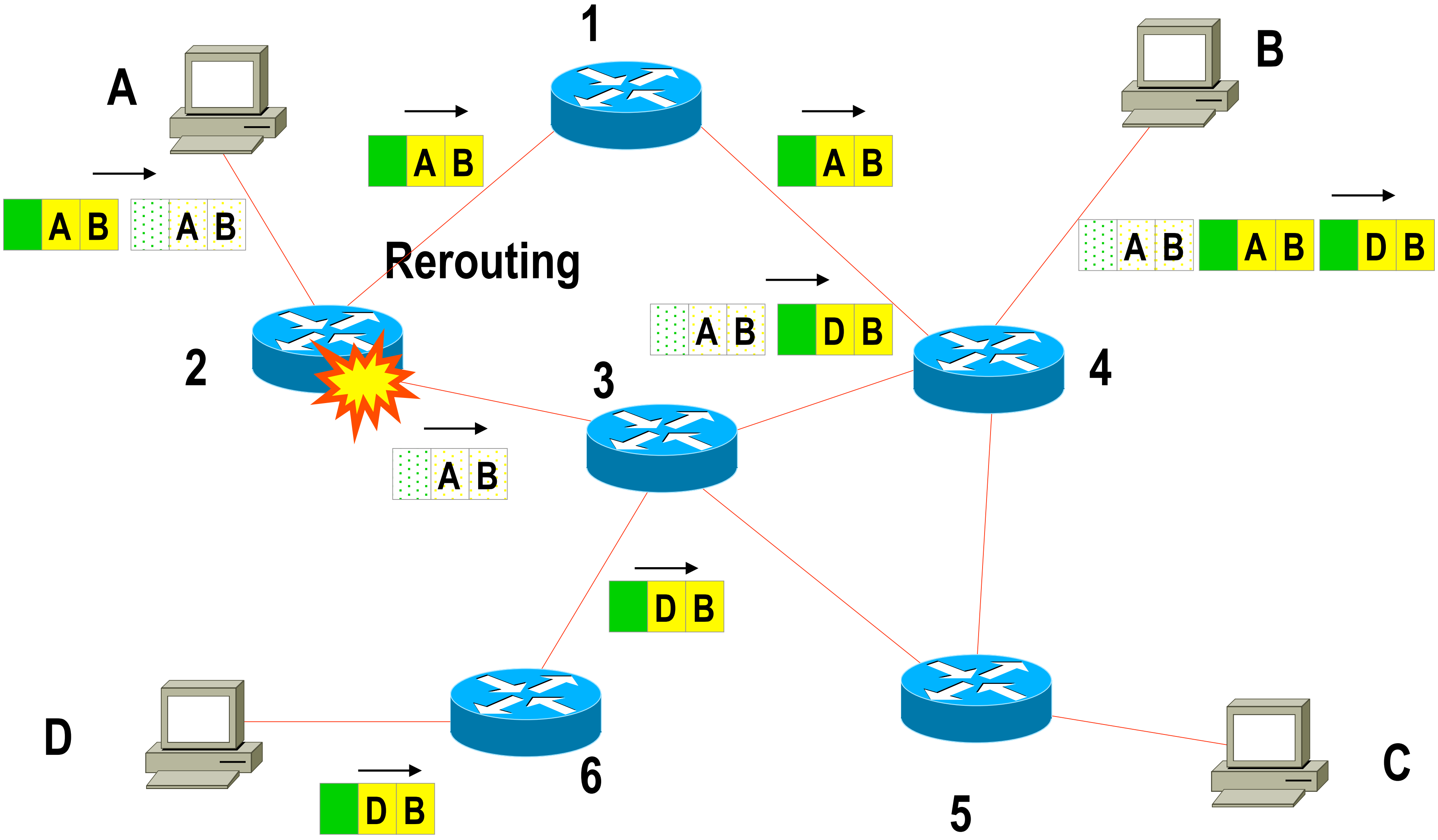
Bit-error on trunk line

Buffer overflow in switch 3

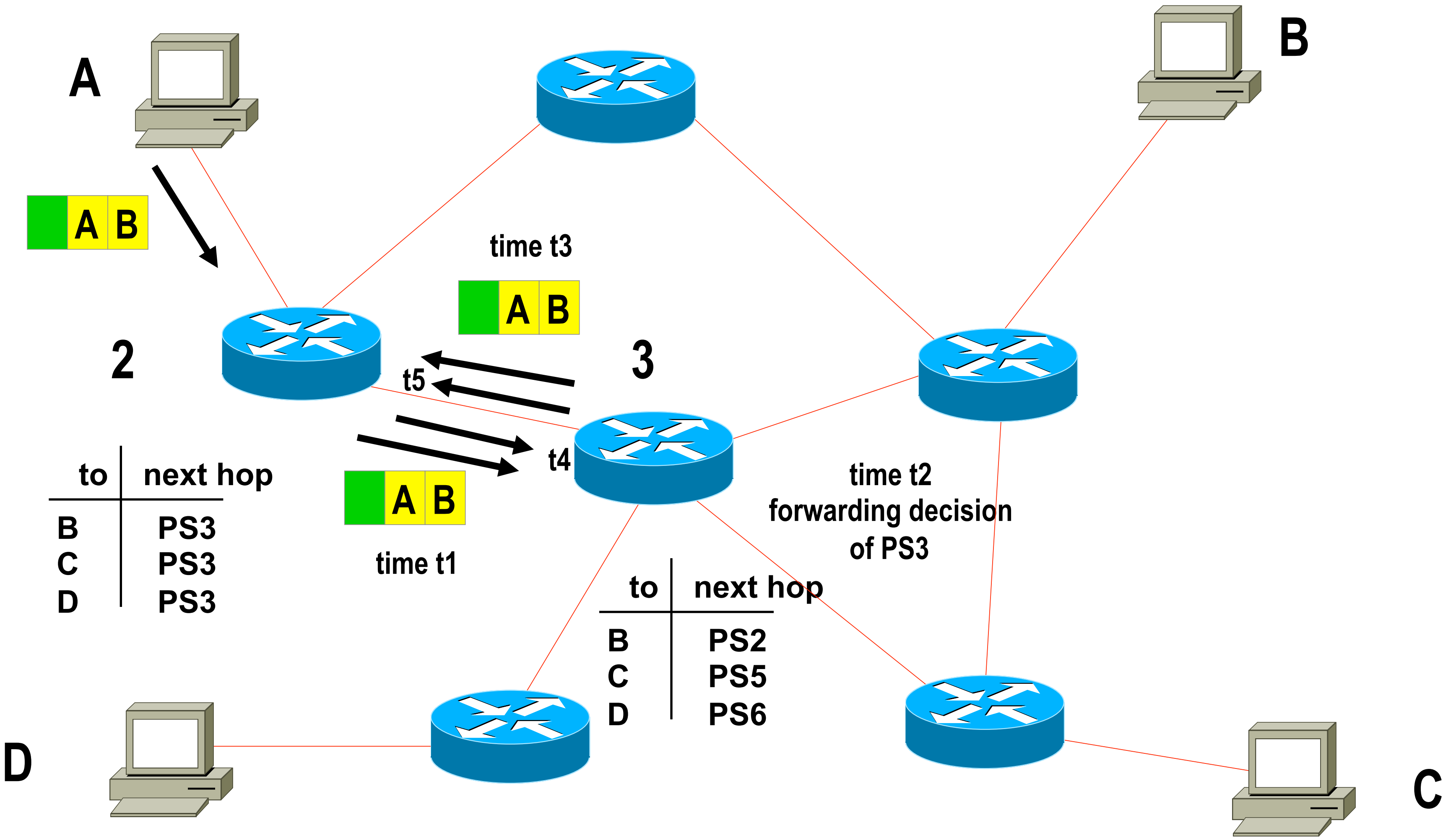
Datagram Service Facts (1)

- **Packets may be discarded / dropped by packet switches**
  - In case of network congestion
  - In case of transmission errors
- **Best effort service**
  - Transport of packets depends on available resources
- **The end systems may take responsibility**
  - For error recovery (retransmission of dropped or corrupted packets)
  - For sequencing and handling of duplicates
- **Reliable data transport requires good transport layer**
  - "Dumb network, smart hosts"

Rerouting – Sequencing Not Guaranteed !



Datagram forwarding needs a kill-mechanism to overcome inconsistent routing tables



to	next hop
B	PS3
C	PS3
D	PS3

to	next hop
B	PS2
C	PS5
D	PS6

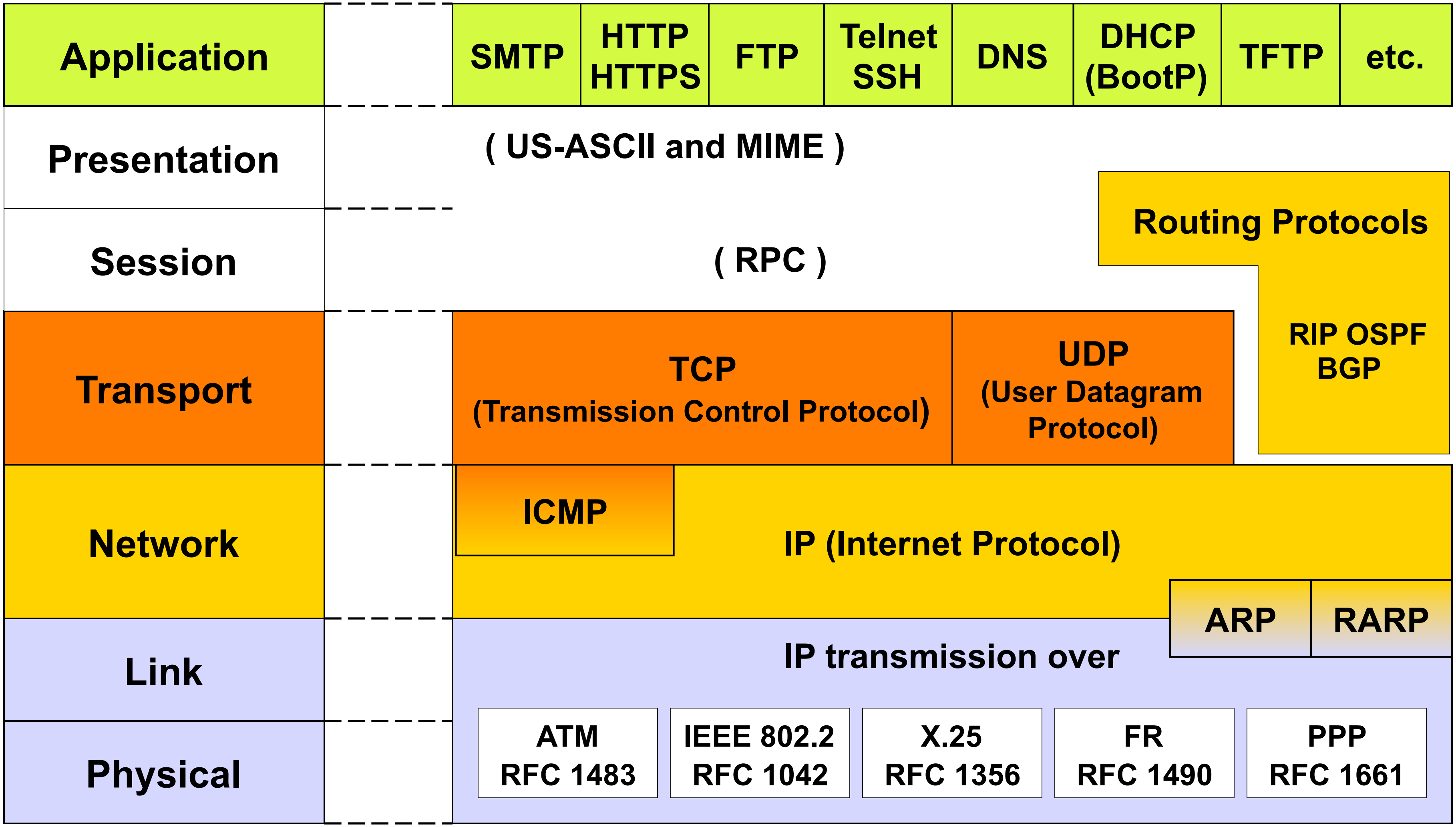
🟩 ... Packet payload

A B ... Source Address / Destination Address

Datagram Service Facts (2)

- **Rerouting in case of topology changes or load balancing means**
  - Packets with the same address information can take different paths to destination
  - Packets may arrive out of sequence
- **Sequence not guaranteed**
  - Rerouting on topology change
  - Load sharing on redundant paths
  - End stations must care
  - Delivery of packets is not guaranteed by the network, must be handled by end systems using higher layer protocol

TCP/IP Protocol Suite



Application

Presentation

Session

Transport

Network

Link

Physical

SMTP

HTTP
HTTPS

FTP

Telnet
SSH

DNS

DHCP
(BootP)

TFTP

etc.

(US-ASCII and MIME)

(RPC)

Routing Protocols

RIP OSPF
BGP

TCP
(Transmission Control Protocol)

UDP
(User Datagram Protocol)

ICMP

IP (Internet Protocol)

ARP

RARP

IP transmission over

ATM
RFC 1483

IEEE 802.2
RFC 1042

X.25
RFC 1356

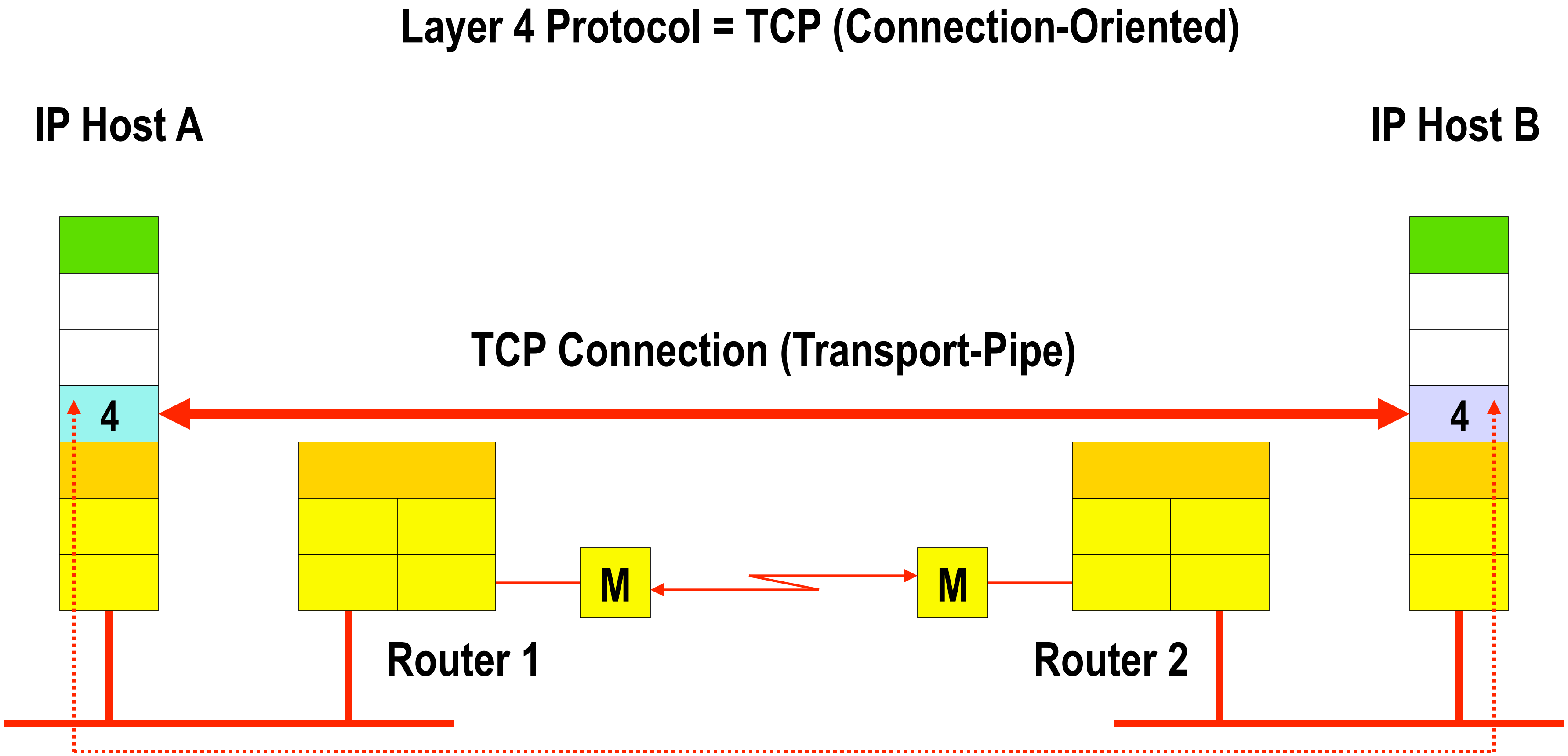
FR
RFC 1490

PPP
RFC 1661

TCP/IP Technology

- **Shared responsibility between network and end systems**
  - Routers responsible for delivering datagrams to remote networks based on structured IP address
  - IP hosts responsible for end-to-end control
- **End to end control**
  - Is implemented in upper layers of IP hosts
  - TCP (Transmission Control Protocol)
    - Connection oriented
    - Sequencing, windowing
    - Error recovery by retransmission
    - Flow control between end systems

TCP and OSI Transport Layer 4

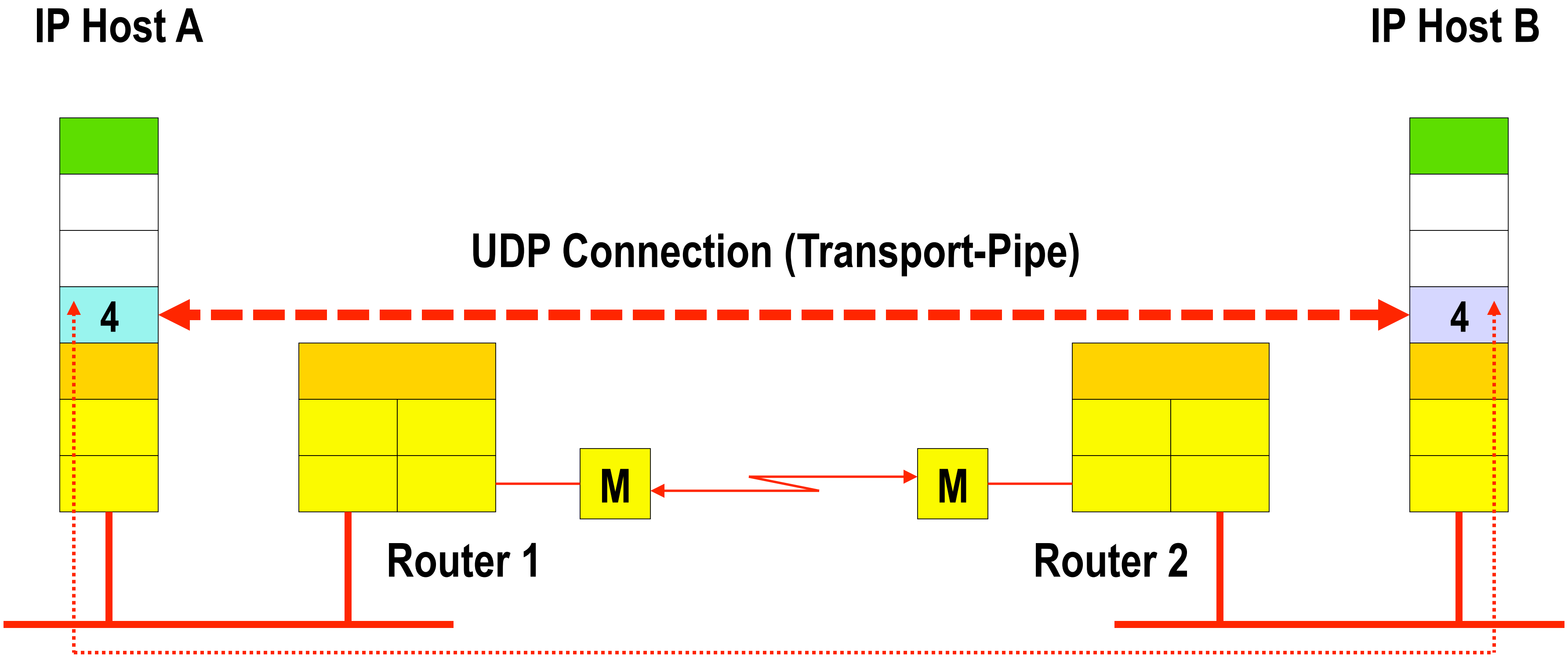

Layer 4 Protocol = TCP (Connection-Oriented)
IP Host A
IP Host B
TCP Connection (Transport-Pipe)
4
4
M
M
Router 1
Router 2

UDP (User Datagram Protocol)

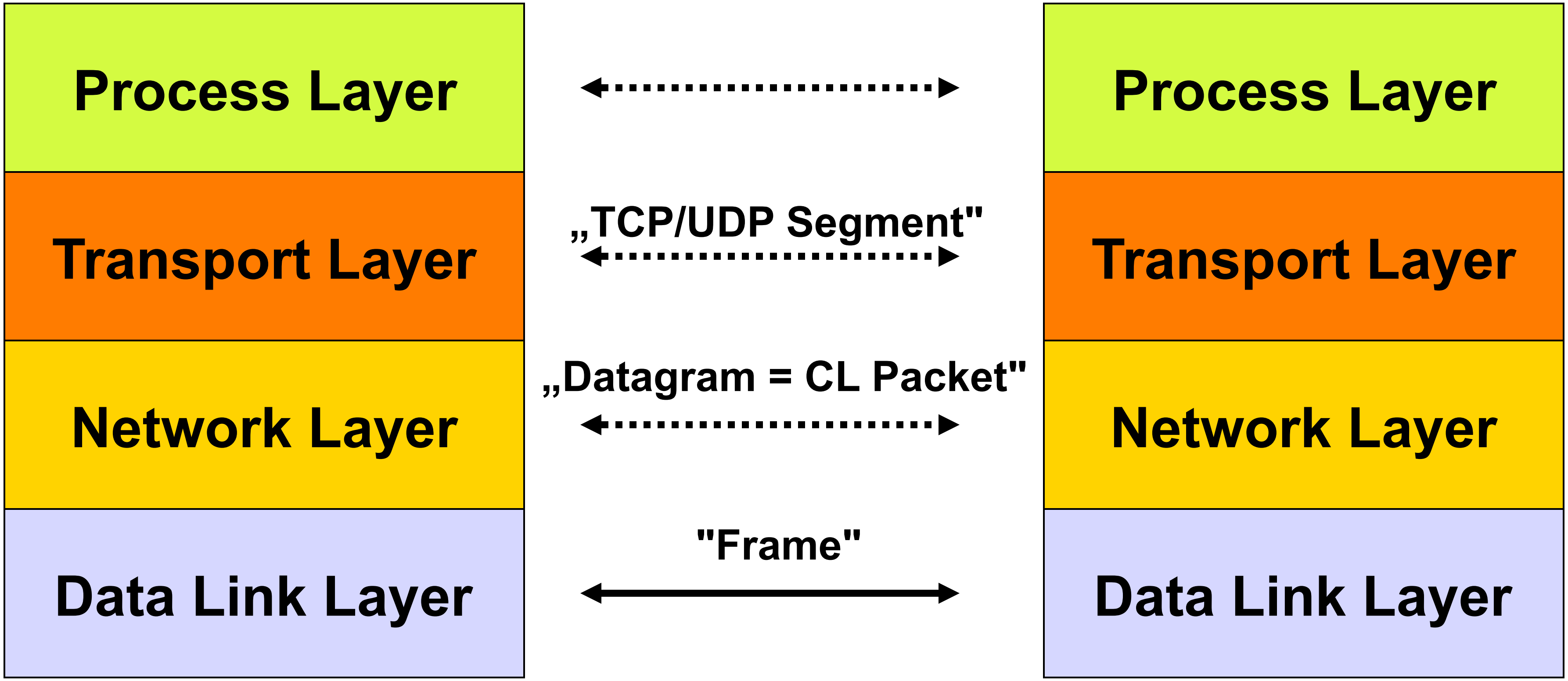
- UDP is a connectionless layer 4 service (datagram service)
- Layer 3 Functions are extended by port addressing and a checksum to ensure integrity
- UDP uses the same port numbers as TCP (if applicable)
- Less complex than TCP, easier to implement

UDP and OSI Transport Layer 4

Layer 4 Protocol = UDP (Connectionless)



DoD 4-Layer Model (Internet)



Process Layer

Transport Layer

Network Layer

Data Link Layer

„TCP/UDP Segment"

„Datagram = CL Packet"

"Frame"

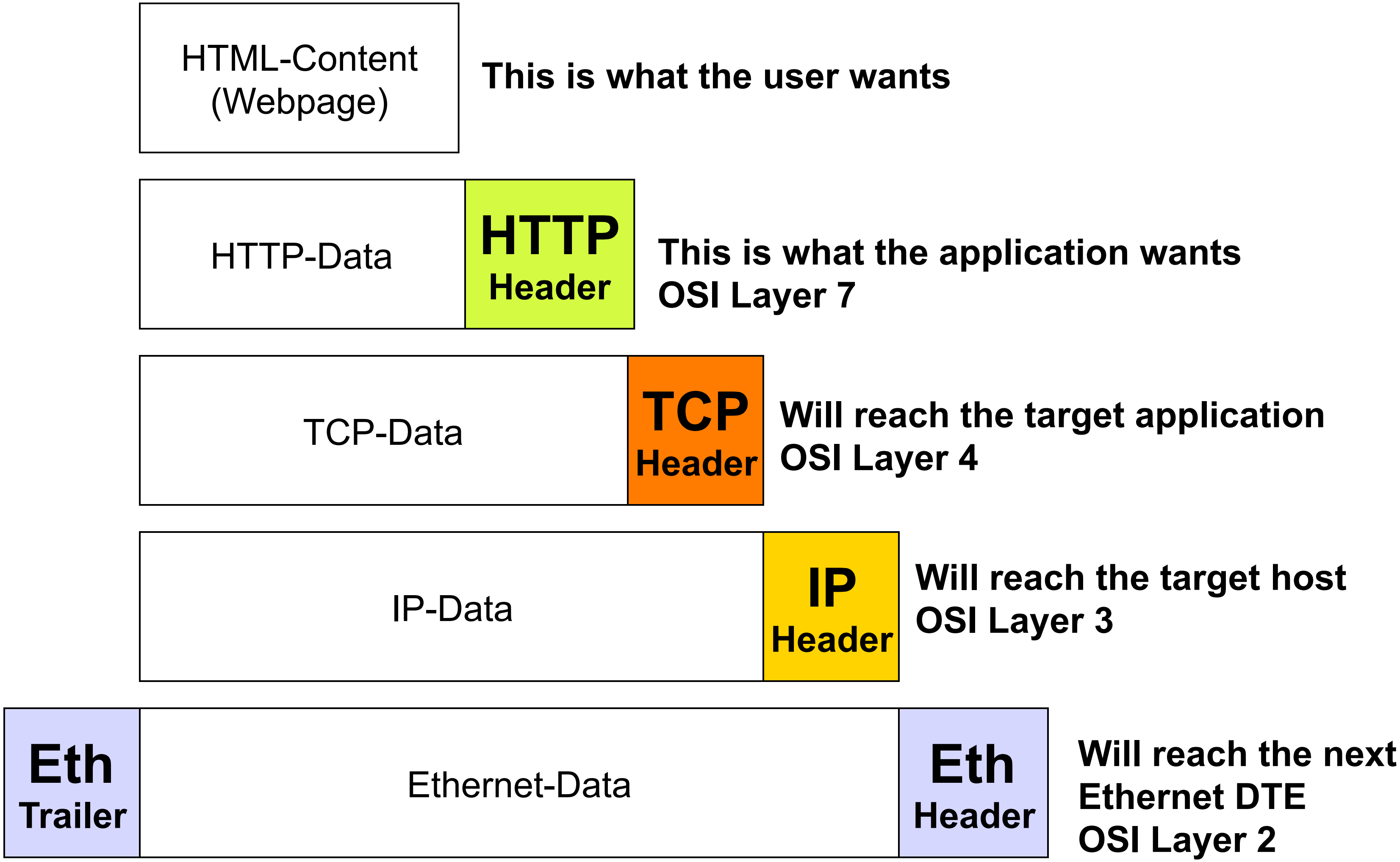
Process Layer

Transport Layer

Network Layer

Data Link Layer

Internet Encapsulation



HTML-Content (Webpage)

This is what the user wants

HTTP-Data | HTTP Header

This is what the application wants
OSI Layer 7

TCP-Data | TCP Header

Will reach the target application
OSI Layer 4

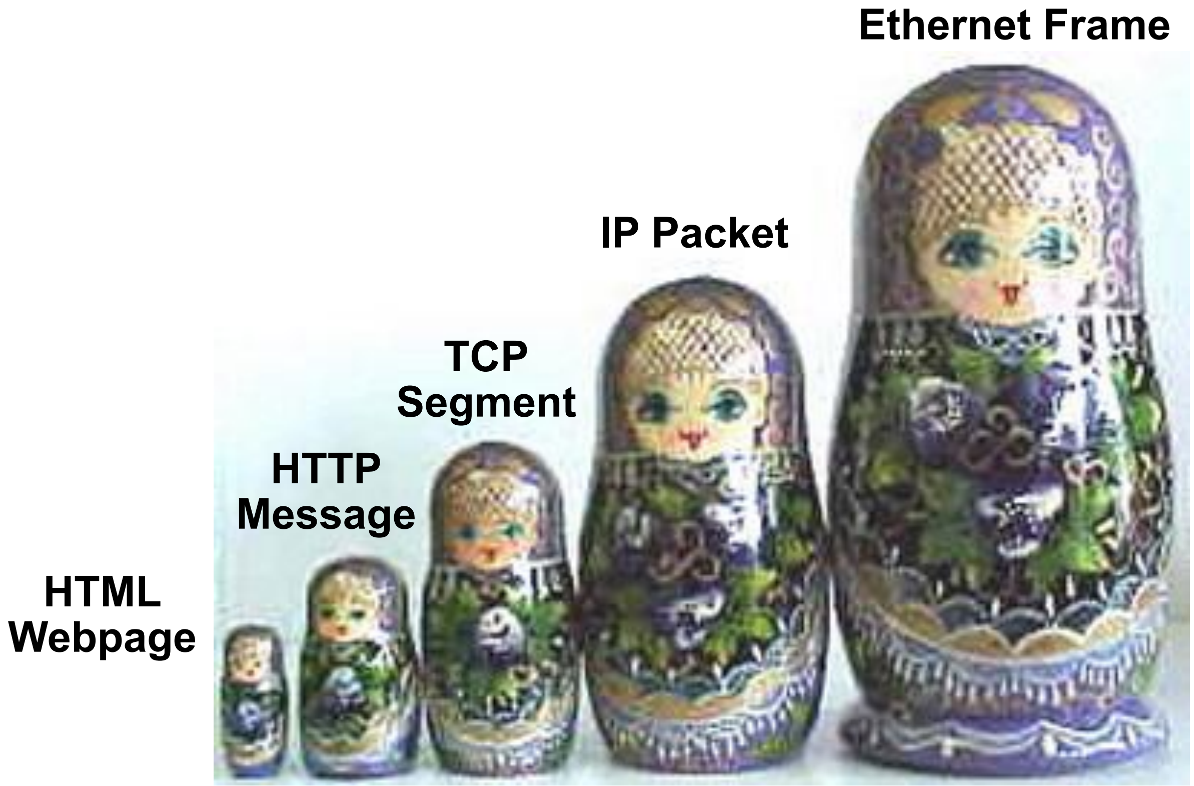
IP-Data | IP Header

Will reach the target host
OSI Layer 3

Eth Trailer | Ethernet-Data | Eth Header

Will reach the next
Ethernet DTE
OSI Layer 2

Practical Encapsulation



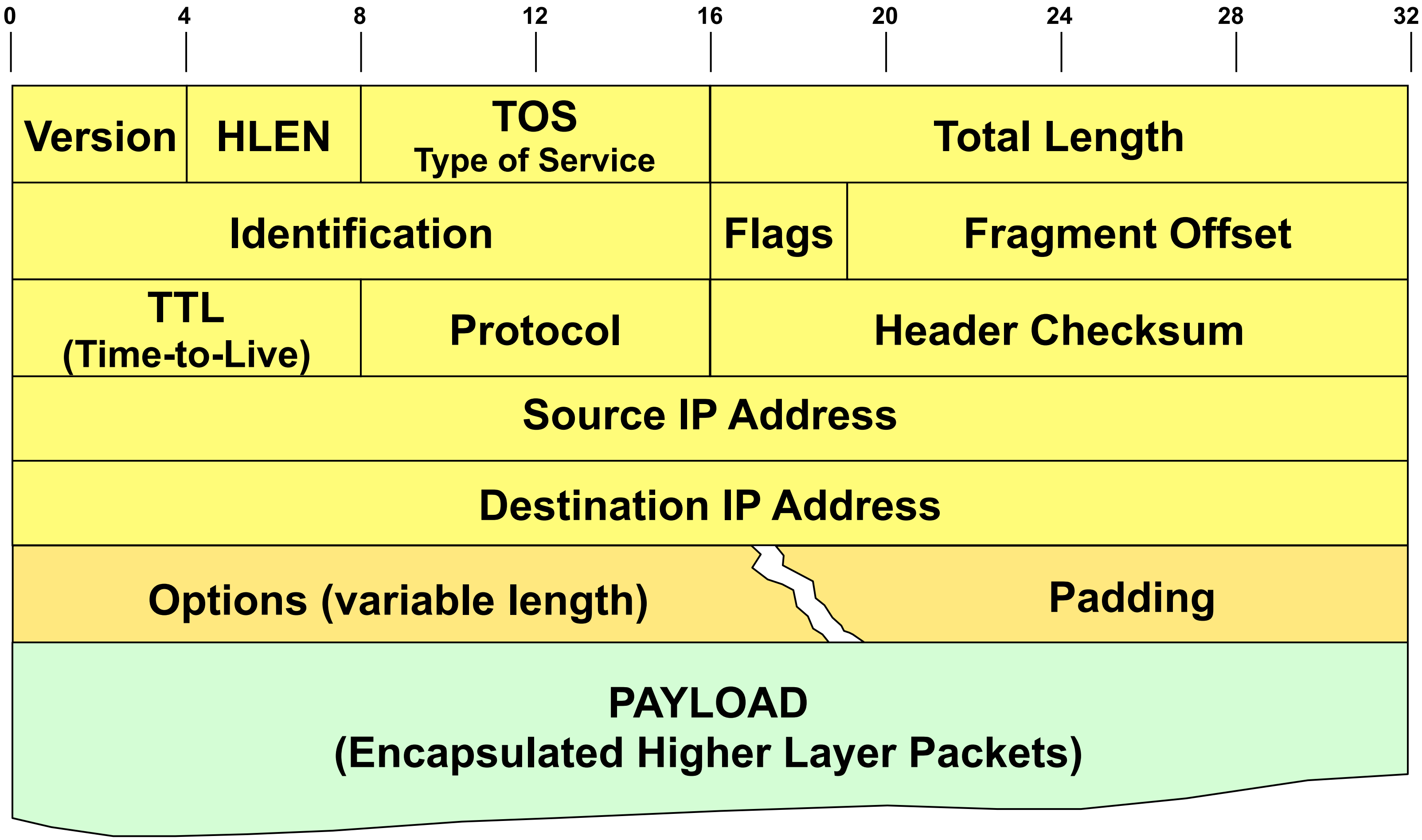
Agenda

- L2 versus L3 Switching
- <u>IP Protocol, IP Addressing</u>
- IP Forwarding
- ARP and ICMP
- IP Routing
- First Hop Redundancy

IP Protocol Functions

- **Packet forwarding**
  - Based on network addressing (Net-IDs)
- **Error detection**
  - Packet header only
- **Fragmentation and reassembly**
  - Necessary, if a datagram has to pass a network with a smaller maximum frame size
  - MTU (Maximum Transmission Unit)
  - Reassembly is done at the receiver
- **Mechanisms to limit the lifetime of a datagram**
  - To omit an endless circulating of datagrams if routing loops occur in the network

The IP Header



0 4 8 12 16 20 24 28 32

Version | HLEN | TOS
Type of Service | Total Length

Identification | Flags | Fragment Offset

TTL
(Time-to-Live) | Protocol | Header Checksum

Source IP Address

Destination IP Address

Options (variable length) | Padding

PAYLOAD
(Encapsulated Higher Layer Packets)

The IP Address

- **Identifies access to a network (network interface)**
- **Two level hierarchy:**
  - Network number (Net-ID)
  - Host number (Host-ID)
- **Dotted Decimal Notation**

Binary IP Address: 11000000101010000000000100000001

Decimal Value: 3232235777

Decimal Representation *per byte*:

1	1	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1

192	168	1	1

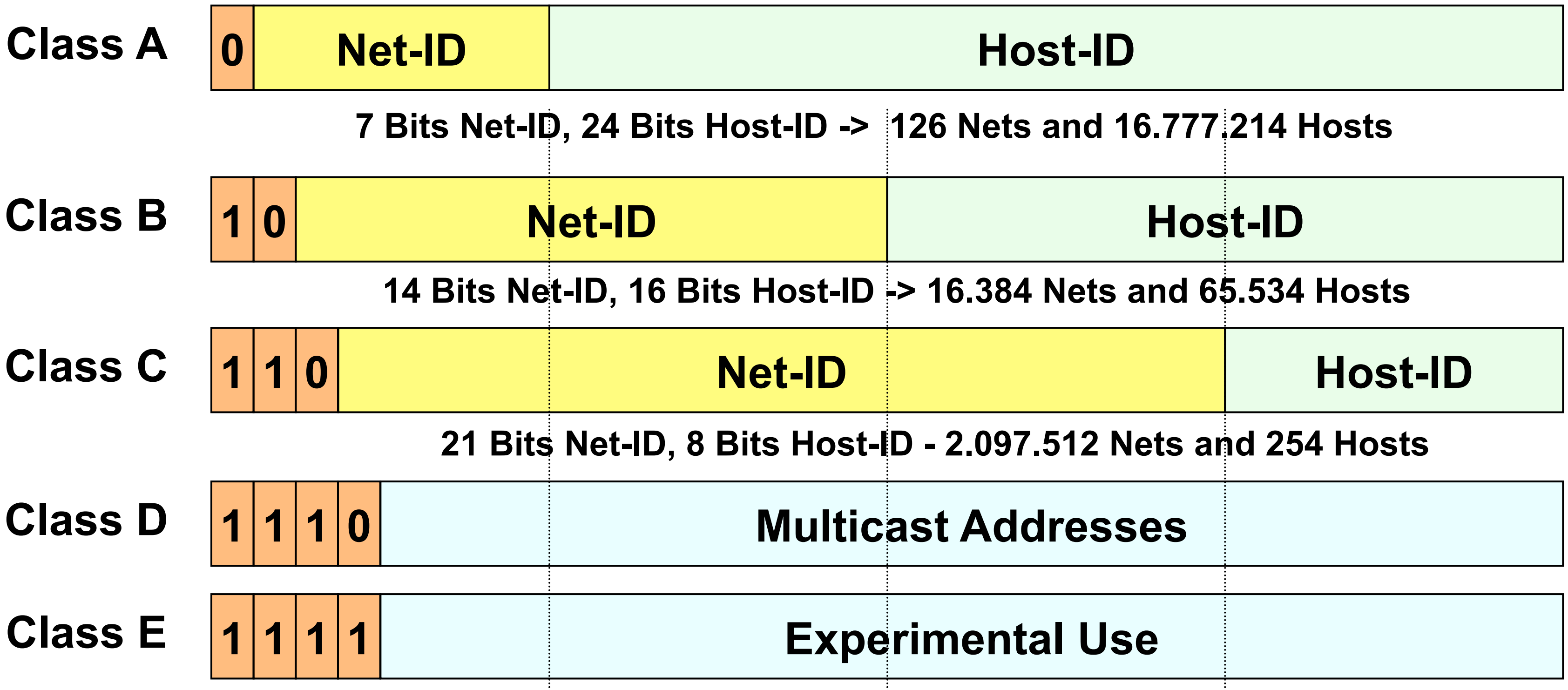
→ **192 . 168 . 1 . 1**

Binary versus Decimal Notation

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0	
1	0	0	0	0	0	0	0	128
0	1	0	0	0	0	0	0	64
0	0	1	0	0	0	0	0	32
0	0	0	1	0	0	0	0	16
0	0	0	0	1	0	0	0	8
0	0	0	0	0	1	0	0	4
0	0	0	0	0	0	1	0	2
0	0	0	0	0	0	0	1	1
1	1	1	1	1	1	1	1	255

IP Address Classes

Originally border between Net-ID and Host-ID was identified by ranges within the IP address room -> address classes -> *„First Octet Rule“*



7 Bits Net-ID, 24 Bits Host-ID -> 126 Nets and 16.777.214 Hosts

14 Bits Net-ID, 16 Bits Host-ID -> 16.384 Nets and 65.534 Hosts

21 Bits Net-ID, 8 Bits Host-ID - 2.097.512 Nets and 254 Hosts

First octet rule:
A (1-126), B (128-191), C (192-223)
D (224-239, Multicast) E (240-254, Experimental)

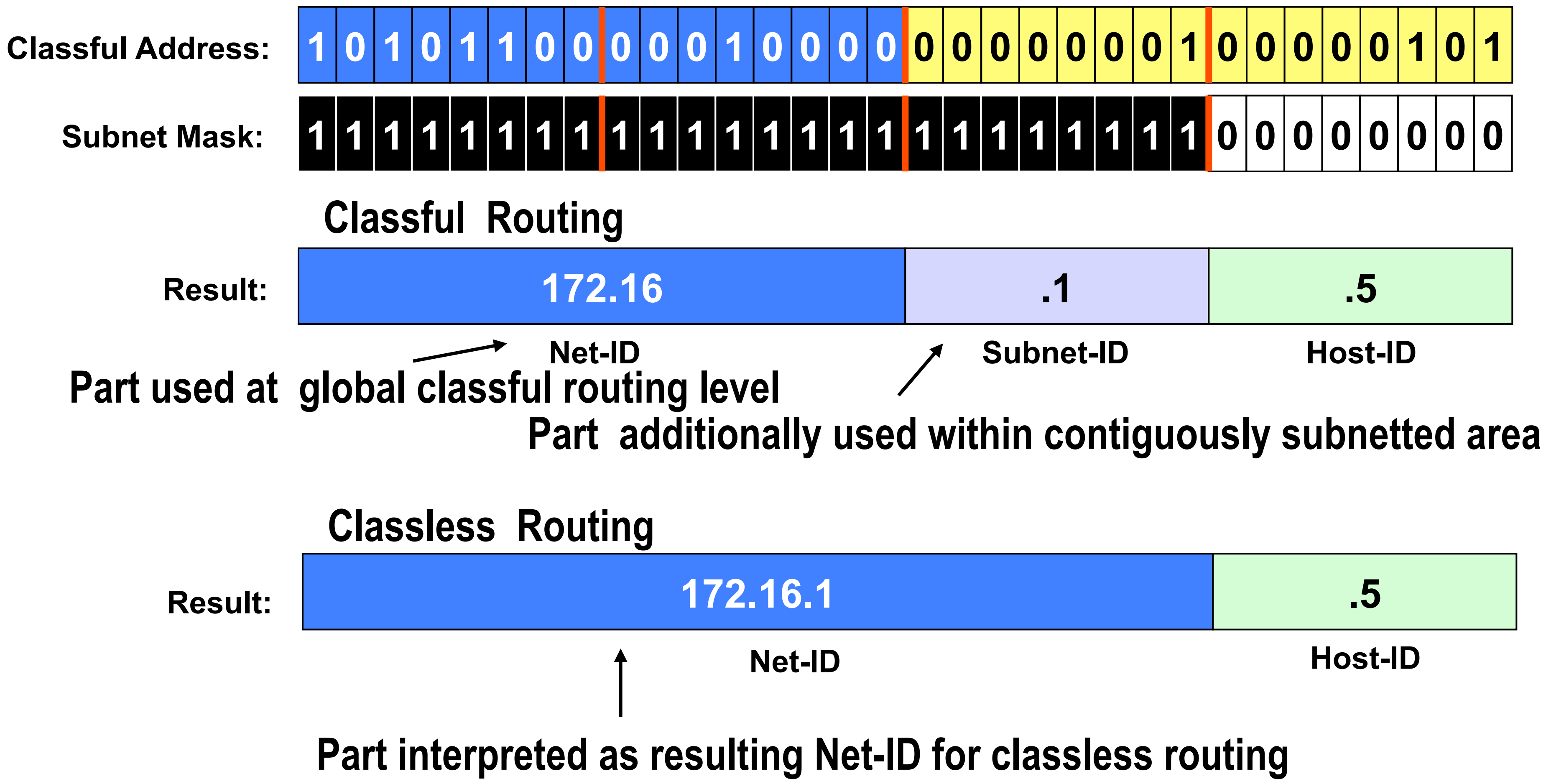
Nowadays

- **Border between Net-ID and Host-ID of an IP address is identified**
  - by Subnetmask
- **Subnetmask**
  - is either written in IP address style e.g. 255.255.0.0
  - or given by prefix / length notation e.g. 10.3.0.0 / 16
- **Classless Routing**
  - No interpretation of old IP address classes A, B, C
  - Modern IP routing protocols can carry subnetmask
    - hence no classless routing limitations anymore
  - VLSM (Variable Length Subnet Mask)
  - Address room can be managed in the most flexible way

Subnetting Example

Class B Address: 172.16.1.5, Subnet Mask: 255.255.255.0

Alternative (prefix/length) notation: 172.16.1.5 / 24



Classful Address:	1	0	1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1
Subnet Mask:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0

Classful Routing

Result:	172.16	.1	.5
	Net-ID	Subnet-ID	Host-ID

Part used at global classful routing level

Part additionally used within contiguously subnetted area

Classless Routing

Result:	172.16.1	.5
	Net-ID	Host-ID

Part interpreted as resulting Net-ID for classless routing

Possible Subnet Mask Values

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0	
1	0	0	0	0	0	0	0	128
1	1	0	0	0	0	0	0	192
1	1	1	0	0	0	0	0	224
1	1	1	1	0	0	0	0	240
1	1	1	1	1	0	0	0	248
1	1	1	1	1	1	0	0	252
1	1	1	1	1	1	1	0	254
1	1	1	1	1	1	1	1	255

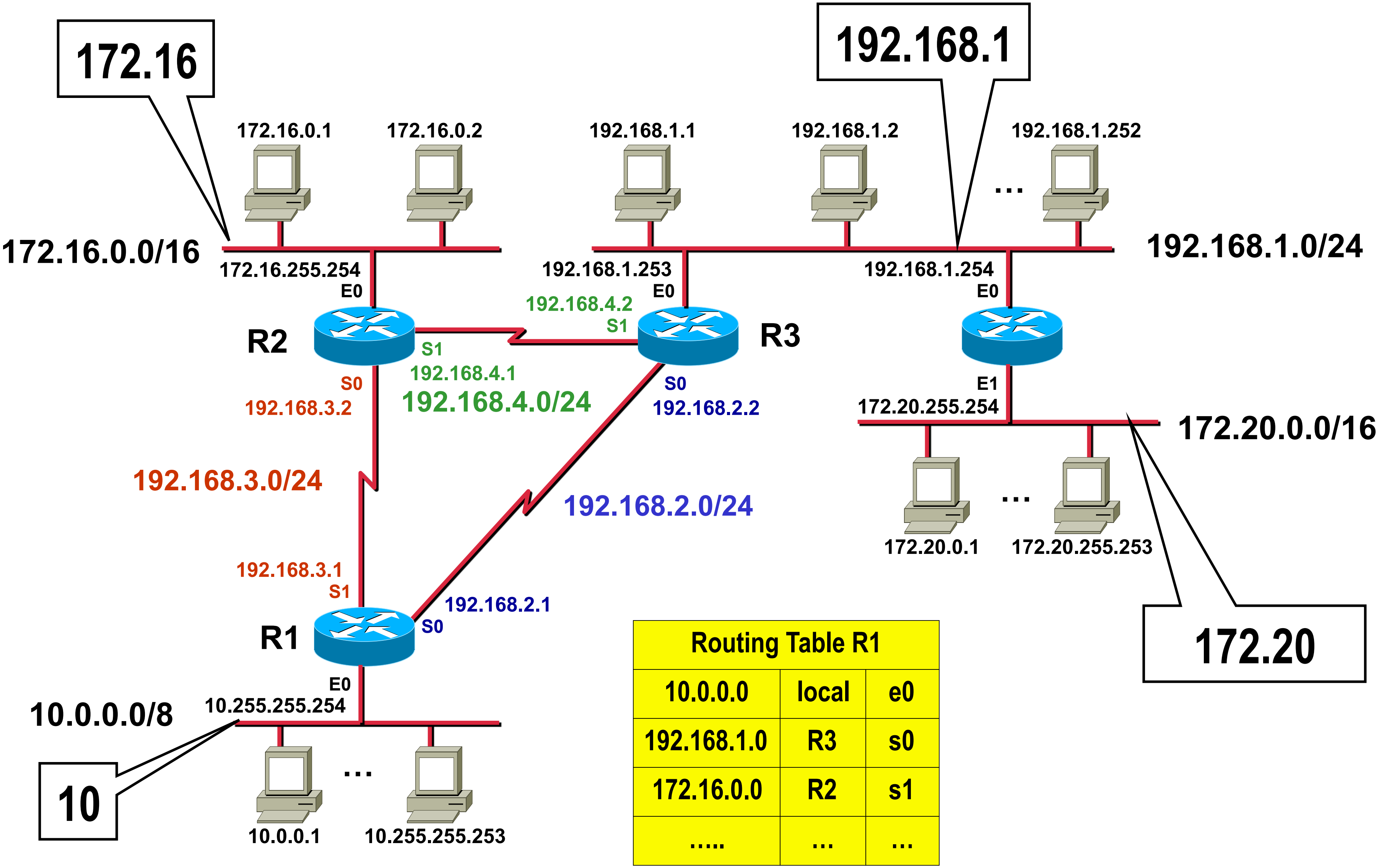
Special Addresses

- All ones in the Host-ID represents „IP Directed-Broadcast“ (10.255.255.255)
- All ones in the Net-ID and Host-ID represents „IP Limited Broadcast” (255.255.255.255)
- All zeros in the Host-ID represents the „Network-Address“ (10.0.0.0)
- Network 127.x.x.x is reserved for "Loopback“
- All zeros in the Net-ID and Host-ID means
  - This host on this network (0.0.0.0)
  - Used during initialization phase (e.g. DHCP)
    - Host uses IP for communication with DHCP server but has no IP address assigned so far

Private Addresses / NAT

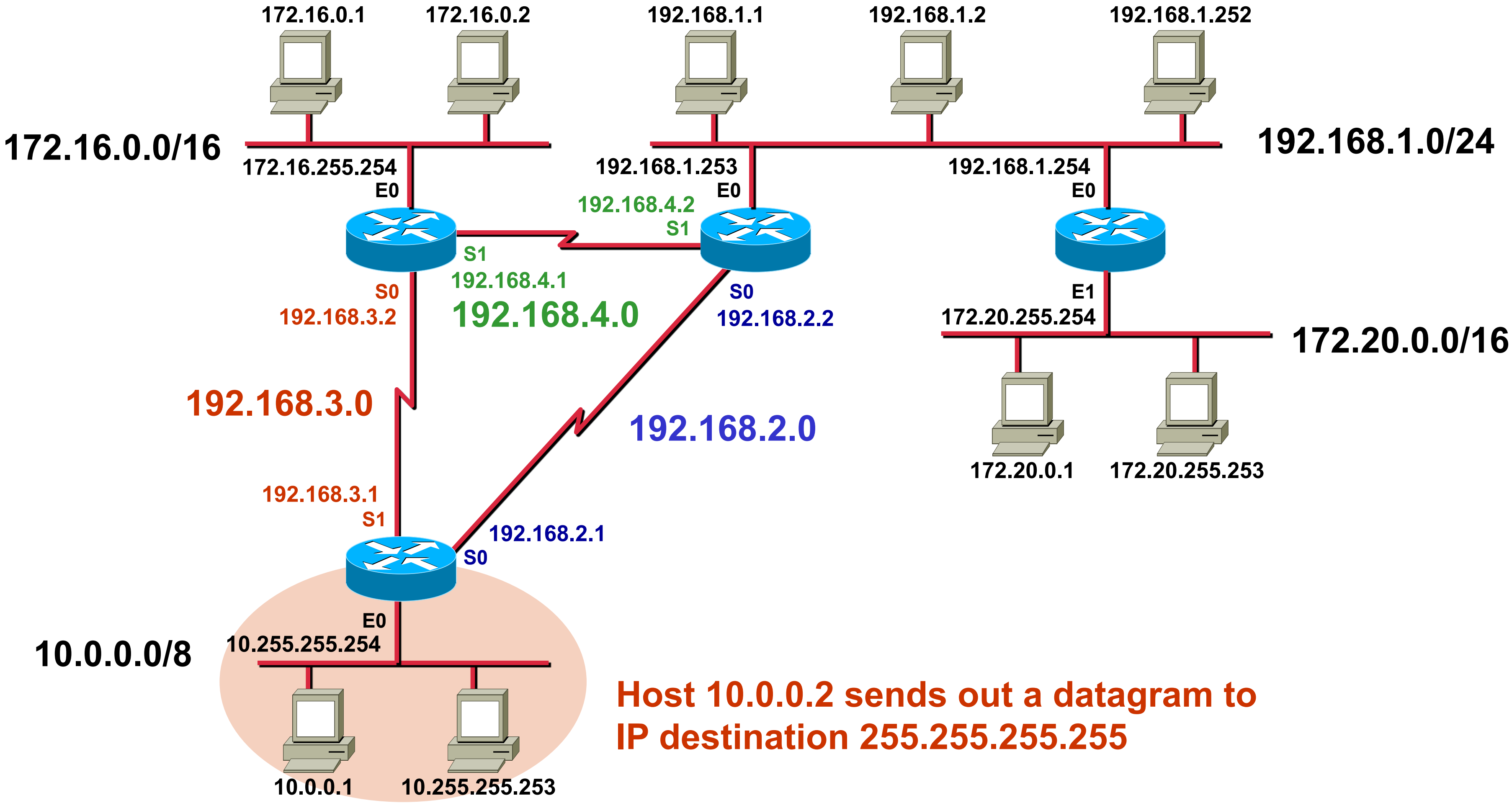
- **Address range for private use**
  - 10.0.0.0 - 10.255.255.255
  - 172.16.0.0 - 172.31.255.255
  - 192.168.0.0 - 192.168.255.255
  - RFC 1918
- **NAT (Network Address Translation)**
  - Is necessary to connect IP hosts with private addresses via NAT Gateway to Internet which needs official IP addresses
  - NAT static 1:1 mapping
  - NAT dynamic n:1 mapping with PAT
    - (UDP/TCP) port address translation
    - 1 official (global routable) IP address may be shared by many internal private stations

Net-ID Addressing Example



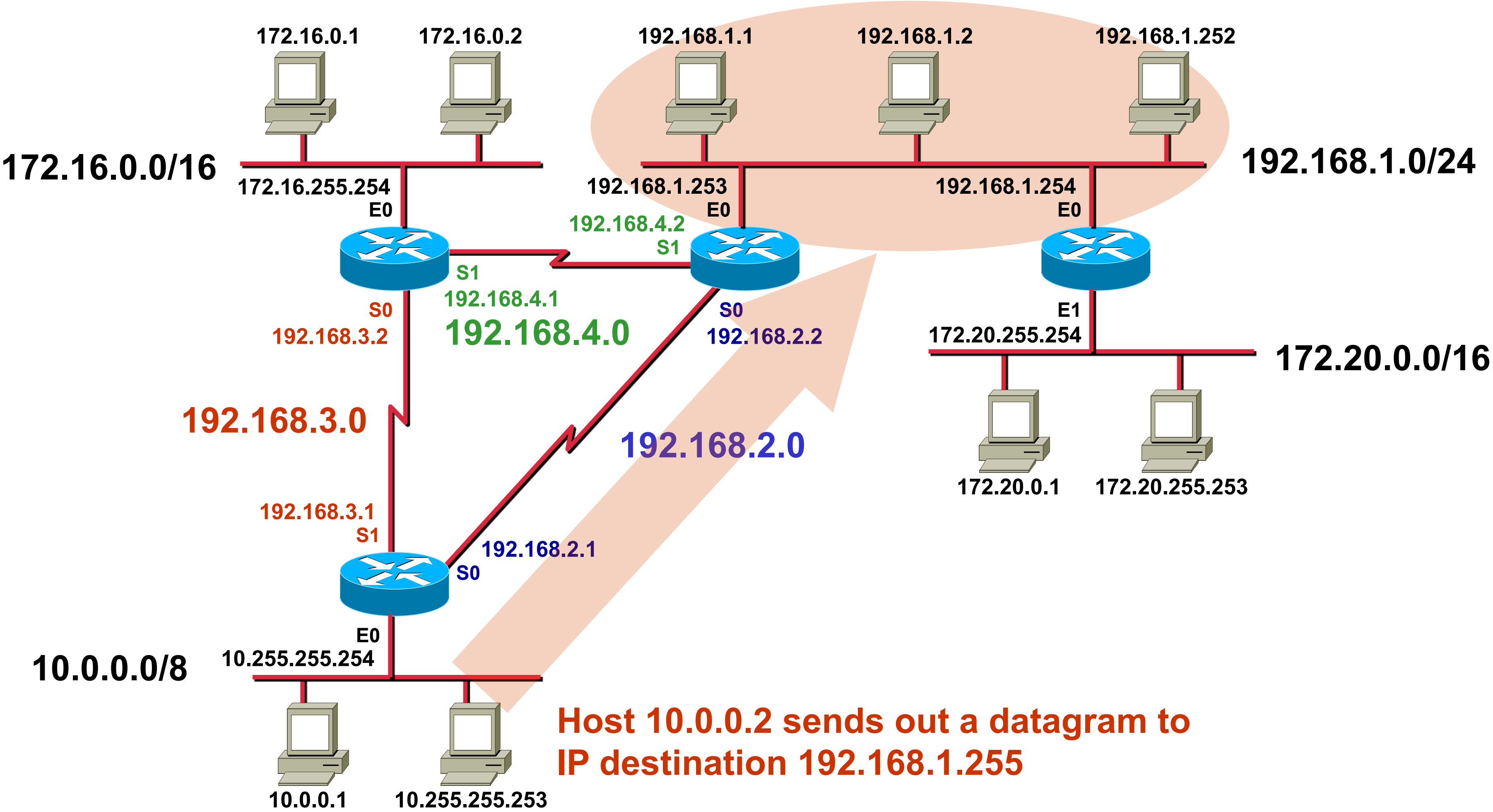
Routing Table R1		
10.0.0.0	local	e0
192.168.1.0	R3	s0
172.16.0.0	R2	s1
…..	…	…

IP Limited Broadcast



Host 10.0.0.2 sends out a datagram to
IP destination 255.255.255.255

IP Directed Broadcast



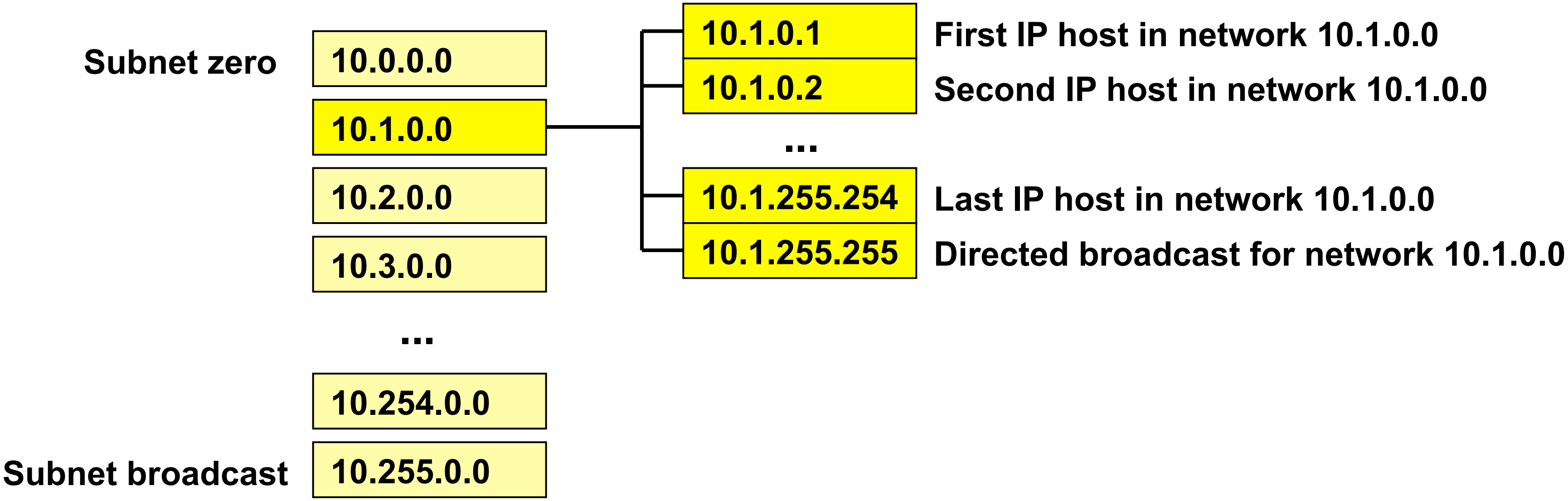
Host 10.0.0.2 sends out a datagram to IP destination 192.168.1.255

Subnet Example 1

"Use the class A network 10.0.0.0 and 8 bit subnetting"

1) That is: 10.0.0.0 with 255.255.0.0 (pseudo class B) or 10.0.0.0/16

2) Resulting subnetworks:


Subnet zero
10.0.0.0
10.1.0.0
10.2.0.0
10.3.0.0
...
10.254.0.0
Subnet broadcast
10.255.0.0
10.1.0.1
First IP host in network 10.1.0.0
10.1.0.2
Second IP host in network 10.1.0.0
...
10.1.255.254
Last IP host in network 10.1.0.0
10.1.255.255
Directed broadcast for network 10.1.0.0

Subnet Mask -> Exam 1

- **Class A address**

Subnet mask	255.255.0.0
IP- Address	10.3.49.45

? Net-ID, ? Host-ID

Net-ID	**=**	**10.3.0.0**
Host-ID	=	0.0.49.45

65534 IP hosts

range: 10.3.0.1 -> 10.3.255.254

10.3.0.0 -> network itself

10.3.255.255 -> directed broadcast for this network

Subnet Mask -> Exam 2

- **Class B address**

Subnet mask 255.255.255.192

IP- Address 172.16.3.144

? Net-ID, ? Host-ID

```
address binary      10101110 . 00010000 . 00000011 . 10010000
mask (binary)       11111111 . 11111111 . 11111111.  11000000
-----------------------------------------------------------------
logical AND (bit by bit)
net-id              10101100 . 00010000 . 00000011 . 10000000
```

Net-ID = 172.16.3.128

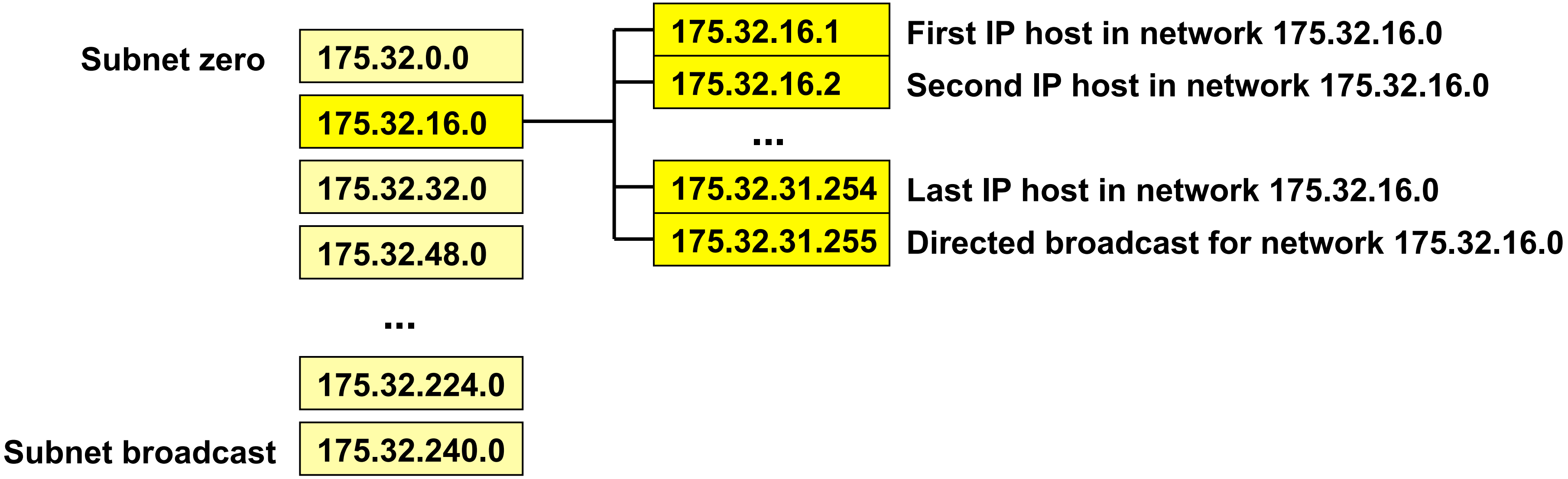
Host-ID = 0.0.0.16

Subnet Example 2

"Use the class B network 175.32.0.0 and 4 bit subnetting"

1) That is: 175.32.0.0 with 255.255.240.0 or 175.32.0.0/20

2) Resulting subnetworks:



Subnet zero 175.32.0.0

175.32.16.0

175.32.32.0

175.32.48.0

...

175.32.224.0

Subnet broadcast 175.32.240.0

175.32.16.1 First IP host in network 175.32.16.0

175.32.16.2 Second IP host in network 175.32.16.0

...

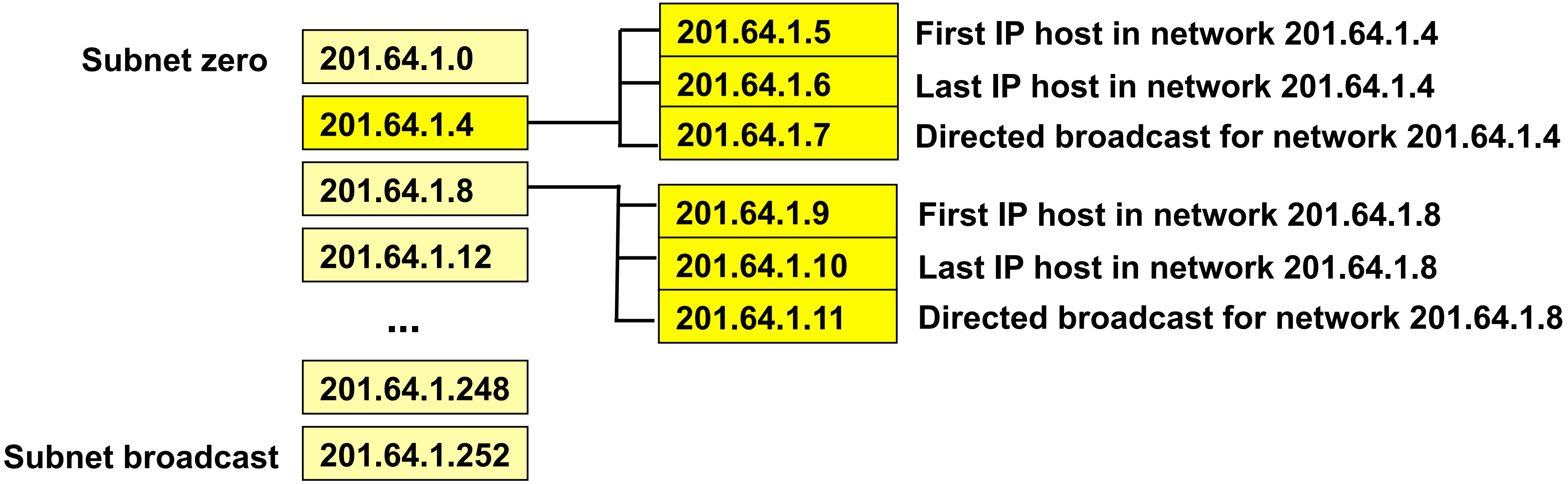
175.32.31.254 Last IP host in network 175.32.16.0

175.32.31.255 Directed broadcast for network 175.32.16.0

Subnet Example 3

"Use the class C network 201.64.1.0 and 6 bit subnetting"

1) That is: 201.64.1.0 with 255.255.255.252 or 201.64.1.0/30
2) Resulting subnetworks:



Subnet zero 201.64.1.0

201.64.1.4

201.64.1.8

201.64.1.12

...

201.64.1.248

Subnet broadcast 201.64.1.252

201.64.1.5 First IP host in network 201.64.1.4

201.64.1.6 Last IP host in network 201.64.1.4

201.64.1.7 Directed broadcast for network 201.64.1.4

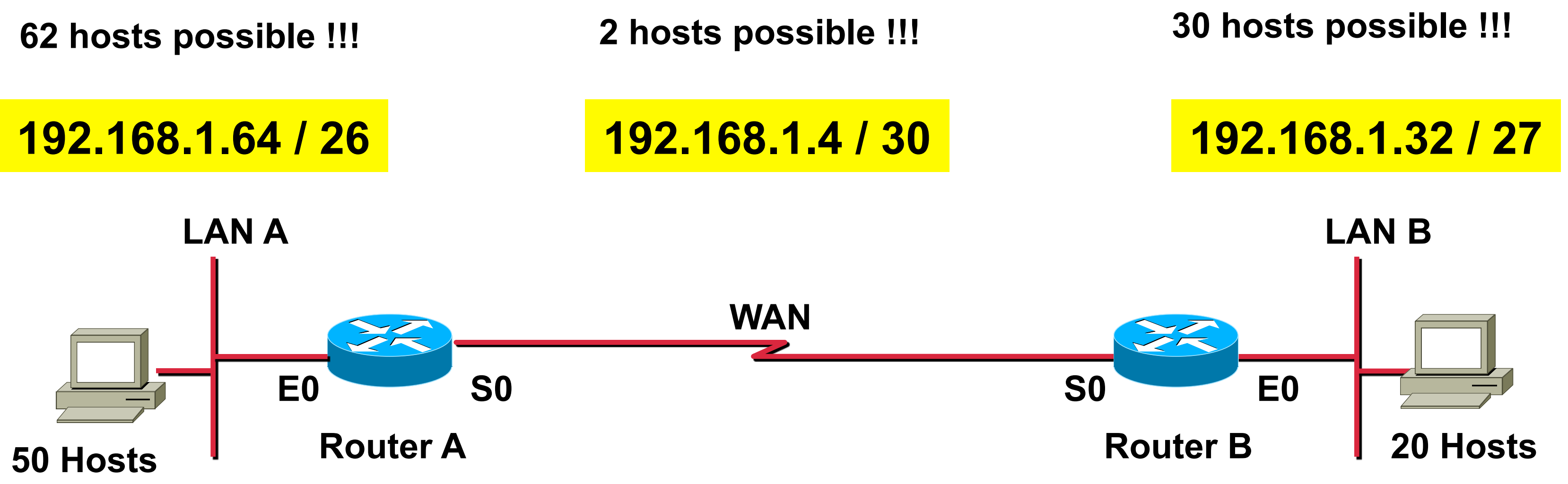
201.64.1.9 First IP host in network 201.64.1.8

201.64.1.10 Last IP host in network 201.64.1.8

201.64.1.11 Directed broadcast for network 201.64.1.8

Variable Length Subnetting (VLSM)

- **Remember:**
  - IP-routing is only possible between different "IP-Networks = Net-IDs”
  - Every link must have an IP net-ID
- **Today IP addresses are rare!**
- **The assignment of IP-Addresses must be as efficient as possible!**



62 hosts possible !!!

192.168.1.64 / 26

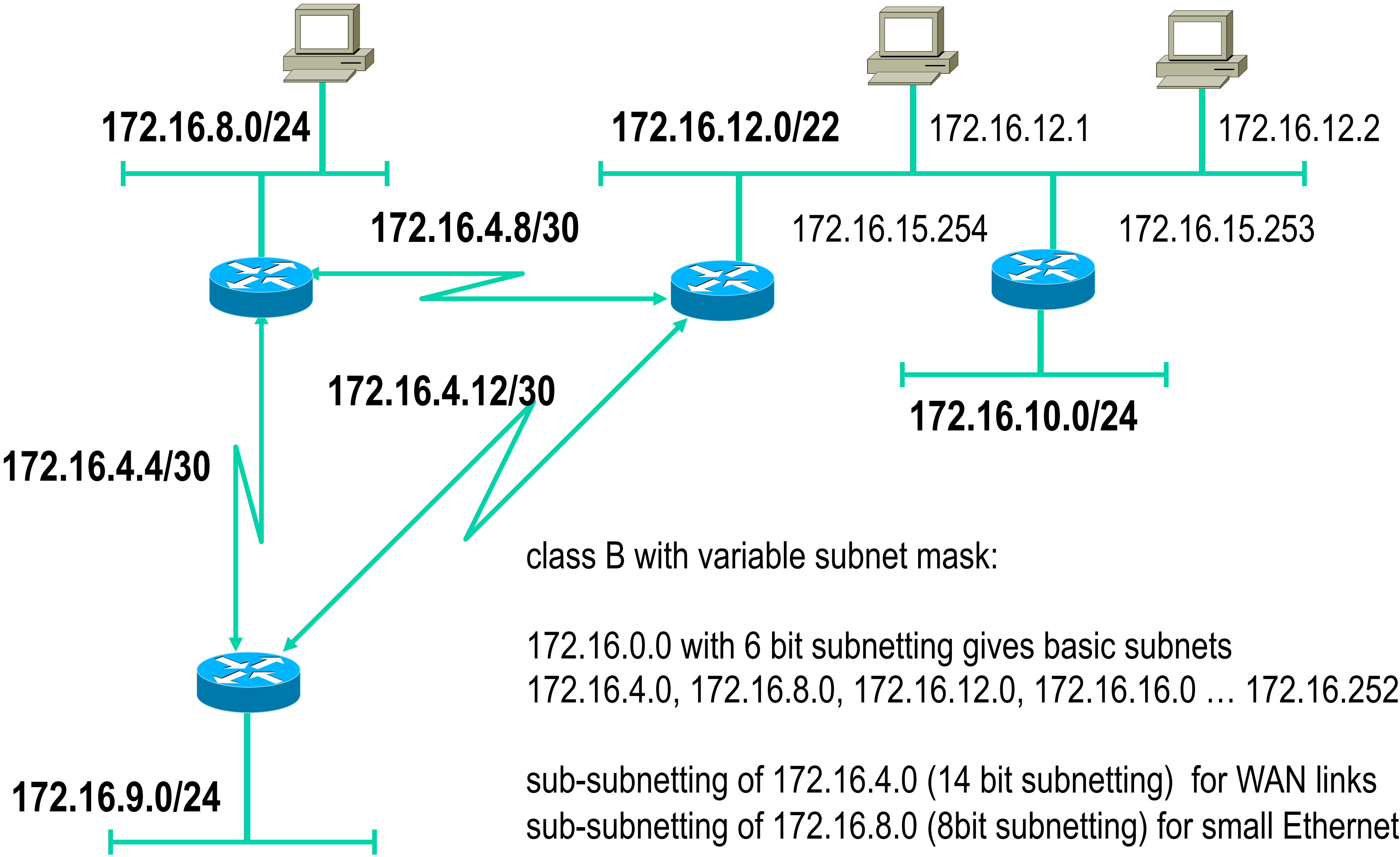
2 hosts possible !!!

192.168.1.4 / 30

30 hosts possible !!!

192.168.1.32 / 27

Example VLSM



class B with variable subnet mask:

172.16.0.0 with 6 bit subnetting gives basic subnets
172.16.4.0, 172.16.8.0, 172.16.12.0, 172.16.16.0 … 172.16.252.0

sub-subnetting of 172.16.4.0 (14 bit subnetting) for WAN links
sub-subnetting of 172.16.8.0 (8bit subnetting) for small Ethernets

VLSM Example (1)

- **First step 6 bit subnetting of 172.16.0.0**
  - 172.16.0.0 with 255.255.252.0 (172.16.0.0 / 22)
  - Subnetworks:
    - 172.16.0.0
    - 172.16.4.0
    - 172.16.8.0
    - 172.16.12.0
    - 172.16.16.0

      ……….
    - 172.16.248.0
    - 172.16.252.0
  - Subnetworks are capable of addressing 1022 IP systems

VLSM Example (2)

- **Next step sub-subnetting**
  - Basic subnet 172.16.4.0 255.255.252.0 (172.16.4.0 / 22)
  - Sub-subnetworks with mask 255.255.255.252 (/ 30):
    - 172.16.4.0 / 30
    - 172.16.4.4 / 30
      - 172.16.4.4 net-ID
      - 172.16.4.5 first IP host of subnet 172.16.4.4
      - 172.16.4.6 last IP host of subnet 172.16.4.4
      - 172.16.4.7 directed broadcast of subnet 172.16.4.4
    - 172.16.4.8 / 30
    - 172.16.4.12 / 30
    - ……….
    - 172.16.4.252 / 30
  - Sub-subnetworks capable of addressing 2 IP systems

VLSM Example (3)

- **Next step sub-subnetting**
  - Basic subnet 172.16.8.0 255.255.252.0 (172.16.8.0 / 22)
  - Sub-subnetworks with mask 255.255.255.0 (/ 24):
    - 172.16.8.0 / 24
    - 172.16.9.0 / 24
      - 172.16.9.0 net-ID
      - 172.16.9.1 first IP host of subnet 172.16.9.0

      - 172.16.9.254 last IP host of subnet 172.16.9.0
      - 172.16.9.255 directed broadcast of subnet 172.16.9.0
    - 172.16.10.0 / 24
    - 172.16.11.0 / 24
  - Sub-subnetworks capable of addressing 254 IP systems

VLSM Example (4)

- **No sub-subnetting for basic subnet 172.16.12.0**
  - 172.16.12.0 with 255.255.252.0 (172.16.12.0 / 22)
    - 172.16.12.0 net-ID
    - 172.16.12.1 first IP host of subnet 172.16.12.0

      -------------
    - 172.16.15.254 last IP host of subnet 172.16.12.0
    - 172.16.15.255 directed broadcast of subnet 172.16.12.0
  - Subnetwork capable of addressing 1022 IP systems

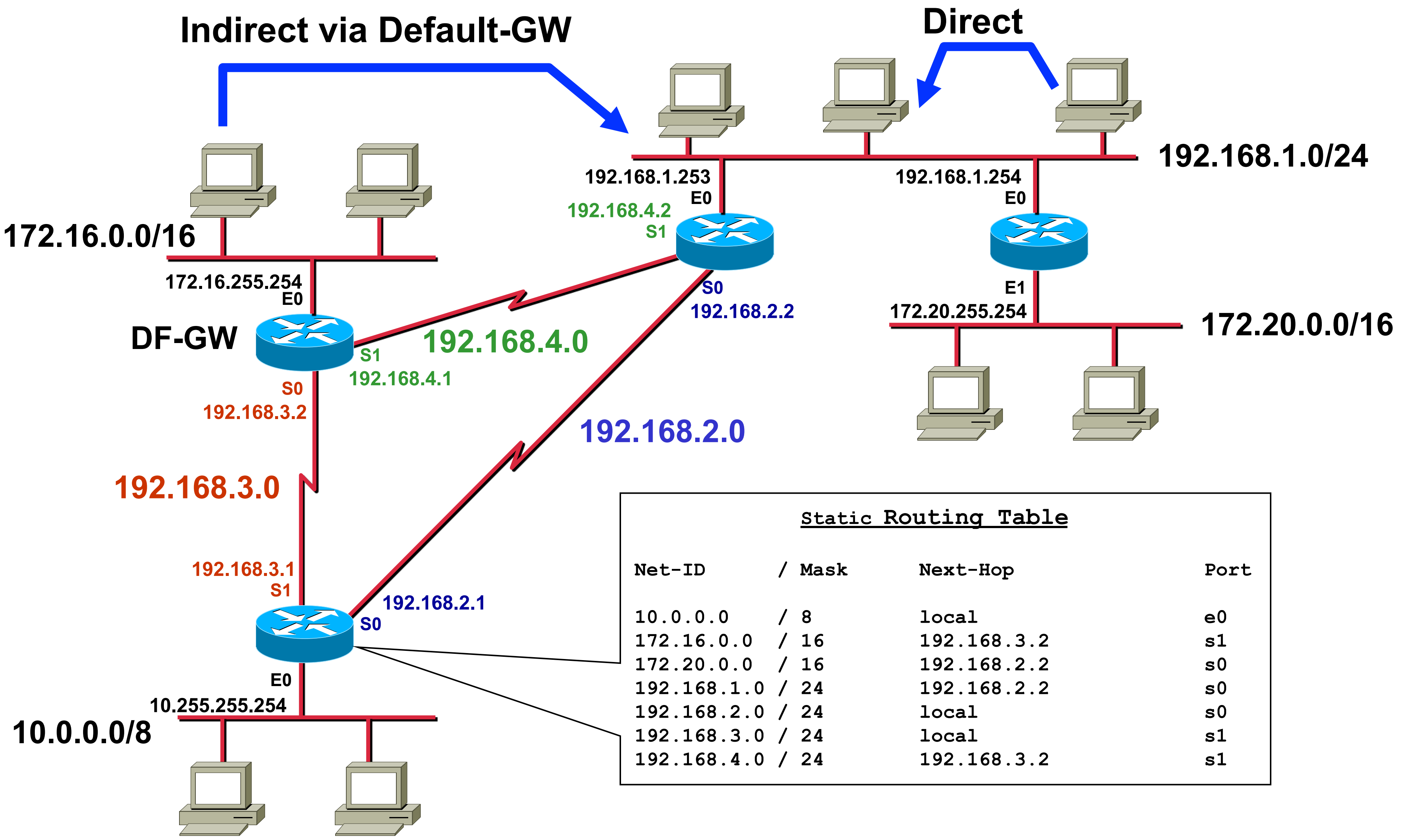
Agenda

- L2 versus L3 Switching
- IP Protocol, IP Addressing
- <u>IP Forwarding</u>
- ARP and ICMP
- IP Routing
- First Hop Redundancy

IP Forwarding Responsibilities

- **IP hosts and IP routers take part in this process**
  - IP hosts responsible for direct delivery of IP datagram's
  - IP routers responsible for selecting the best path in a meshed network in case of indirect delivery of IP datagram's
    - Decision based on current state of routing table
- **Direct versus indirect delivery**
  - Depends on destination net-ID
    - Net-ID equal source net-ID -> direct delivery
    - Net-ID unequal source net-ID -> indirect delivery
- **IP hosts know about default router aka "Default Gateway"**
  - As next hop in case of indirect delivery of IP datagrams

Direct versus Indirect Delivery Default Gateway / Routing Table



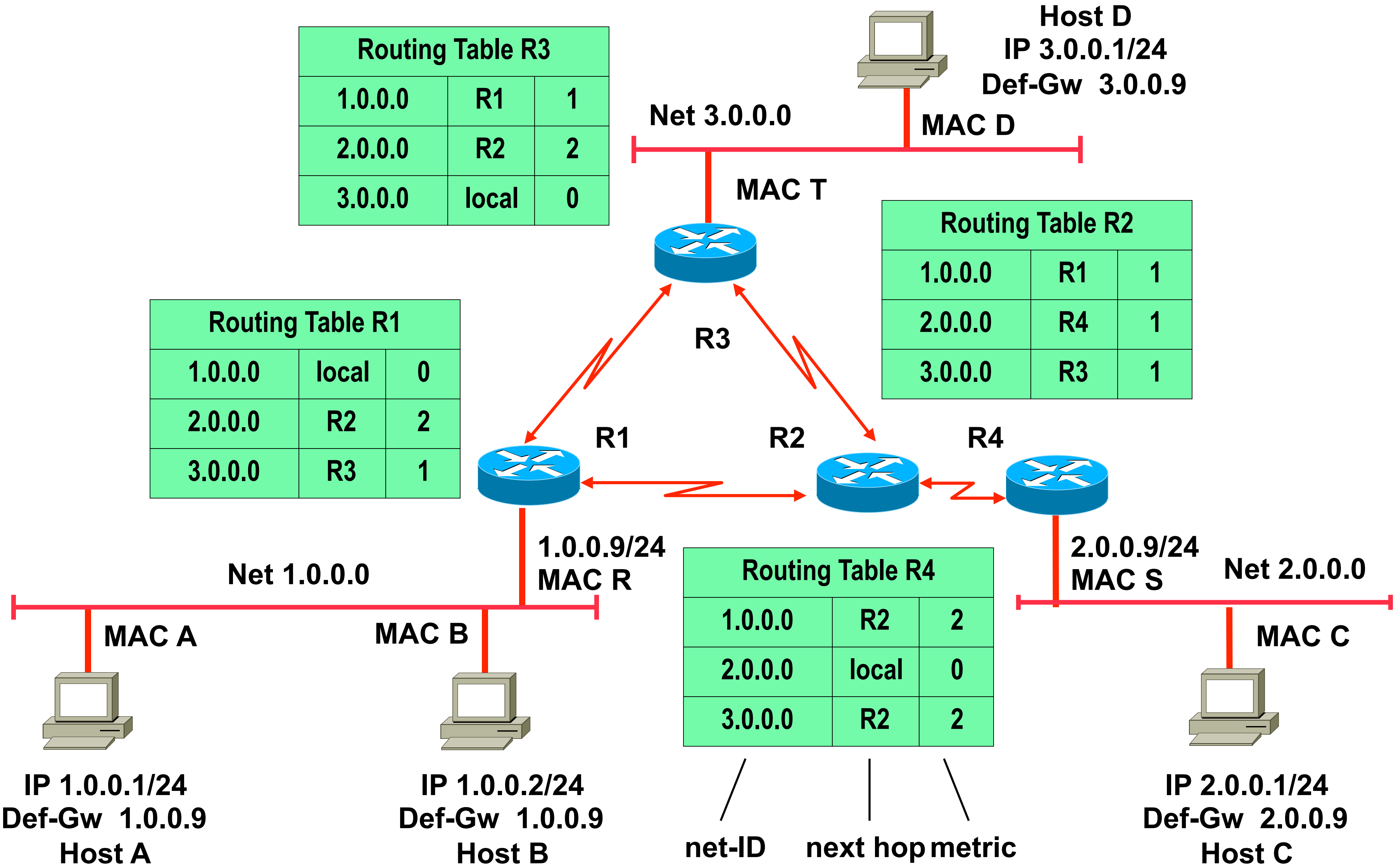
Static Routing Table

Net-ID	/ Mask	Next-Hop	Port
10.0.0.0	/ 8	local	e0
172.16.0.0	/ 16	192.168.3.2	s1
172.20.0.0	/ 16	192.168.2.2	s0
192.168.1.0	/ 24	192.168.2.2	s0
192.168.2.0	/ 24	local	s0
192.168.3.0	/ 24	local	s1
192.168.4.0	/ 24	192.168.3.2	s1

Principle

- **IP Forwarding is done by routers in case of indirect routing**
  - Based on the destination address of a given IP datagram
  - Following the path to the destination hop by hop
- **Routing tables**
  - Have information about which next hop router a given destination network can be reached
- **L2 header must be changed hop by hop**
  - If LAN then physical L2 address (MAC addresses) must be adapted for direct communication on LAN
- **Mapping between IP and L2 address on LAN**
  - Is done by Address Resolution Protocol (ARP)

Example Topology



Routing Table R3

1.0.0.0	R1	1
2.0.0.0	R2	2
3.0.0.0	local	0

Routing Table R2

1.0.0.0	R1	1
2.0.0.0	R4	1
3.0.0.0	R3	1

Routing Table R1

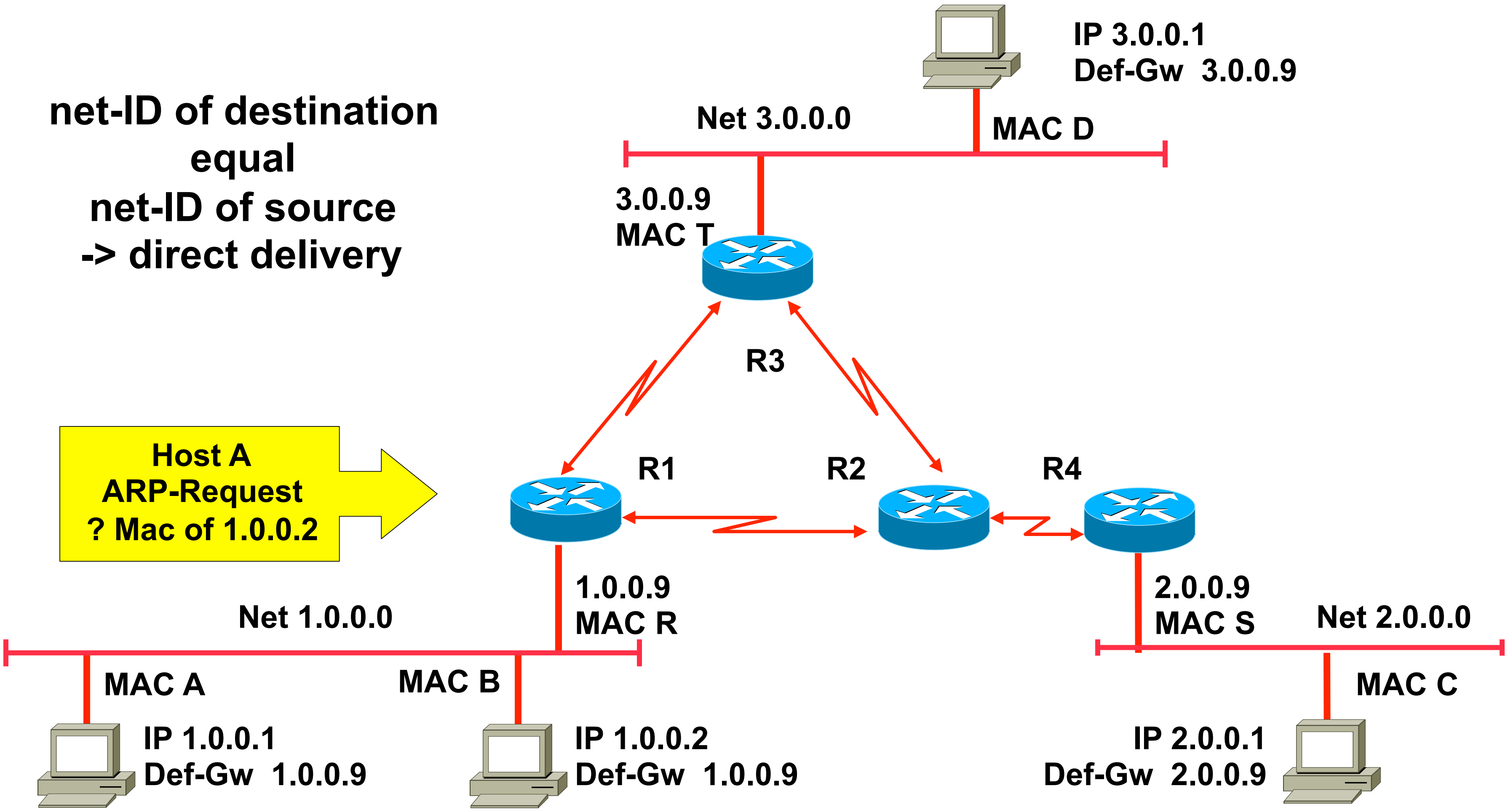
1.0.0.0	local	0
2.0.0.0	R2	2
3.0.0.0	R3	1

Routing Table R4

1.0.0.0	R2	2
2.0.0.0	local	0
3.0.0.0	R2	2

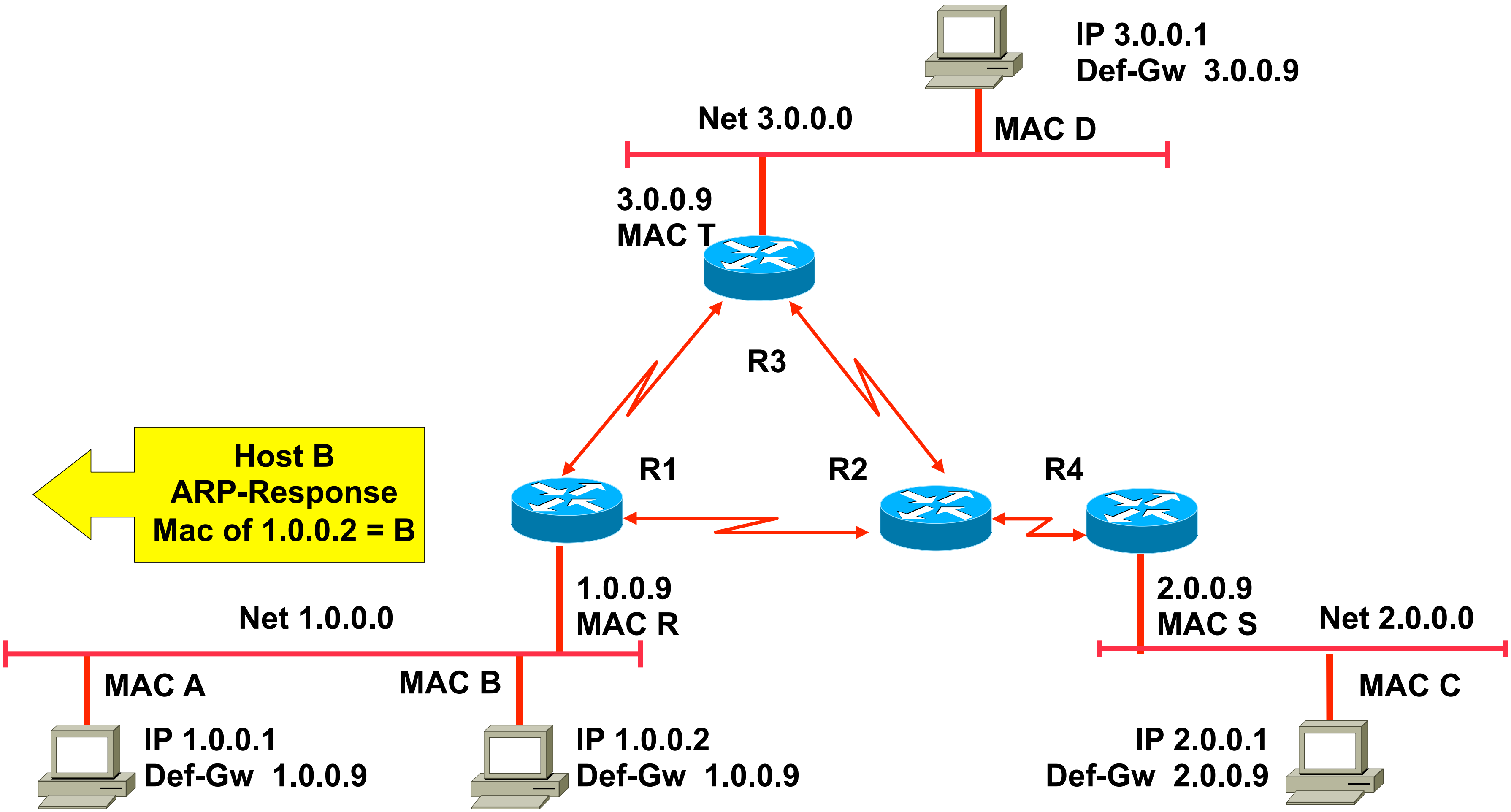
Direct Delivery 1.0.0.1 - > 1.0.0.2

net-ID of destination
equal
net-ID of source
-> direct delivery



ARP ... Address Resolution Protocol

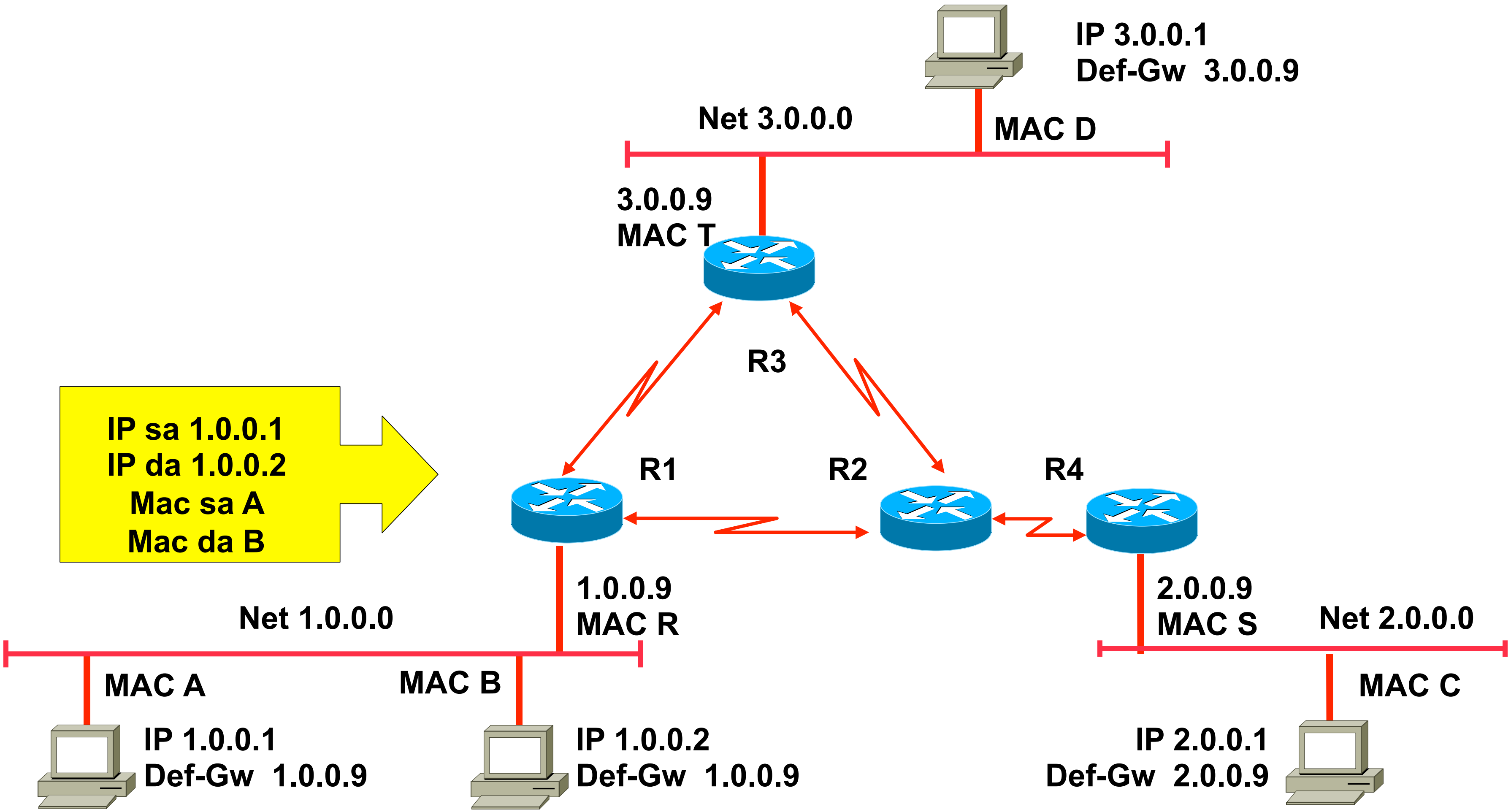
Direct Delivery 1.0.0.1 - > 1.0.0.2



Host B
ARP-Response
Mac of 1.0.0.2 = B

ARP-Cache Host A	
1.0.0.2	MAC B

Direct Delivery 1.0.0.1 - > 1.0.0.2



IP 3.0.0.1
Def-Gw 3.0.0.9

Net 3.0.0.0

MAC D

3.0.0.9
MAC T

R3

IP sa 1.0.0.1
IP da 1.0.0.2
Mac sa A
Mac da B

R1

R2

R4

1.0.0.9
MAC R

2.0.0.9
MAC S

Net 1.0.0.0

Net 2.0.0.0

MAC A

MAC B

MAC C

IP 1.0.0.1
Def-Gw 1.0.0.9

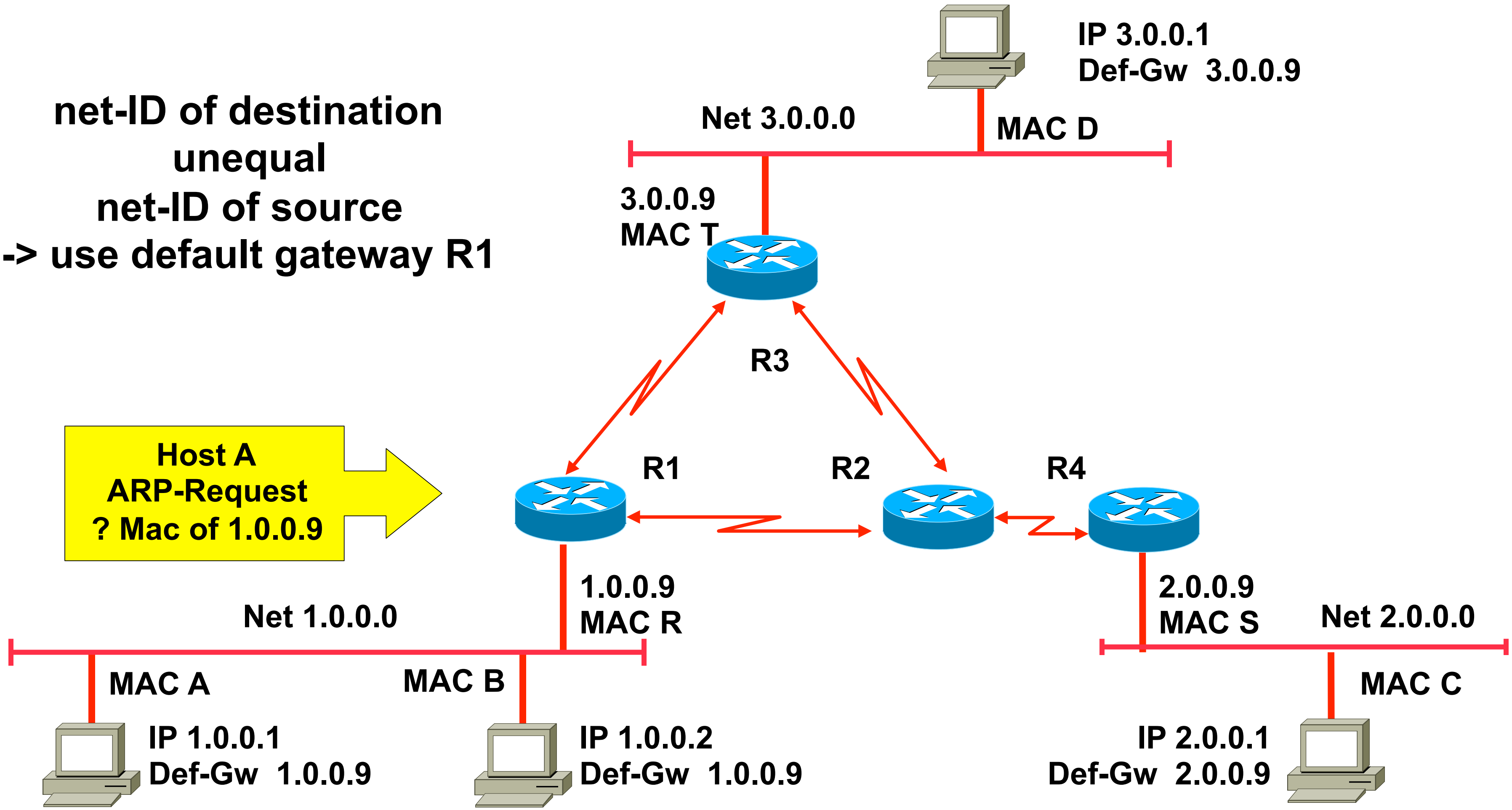
IP 1.0.0.2
Def-Gw 1.0.0.9

IP 2.0.0.1
Def-Gw 2.0.0.9

ARP-Cache Host A	
1.0.0.2	MAC B

Indirect Delivery 1.0.0.1 - > 2.0.0.1

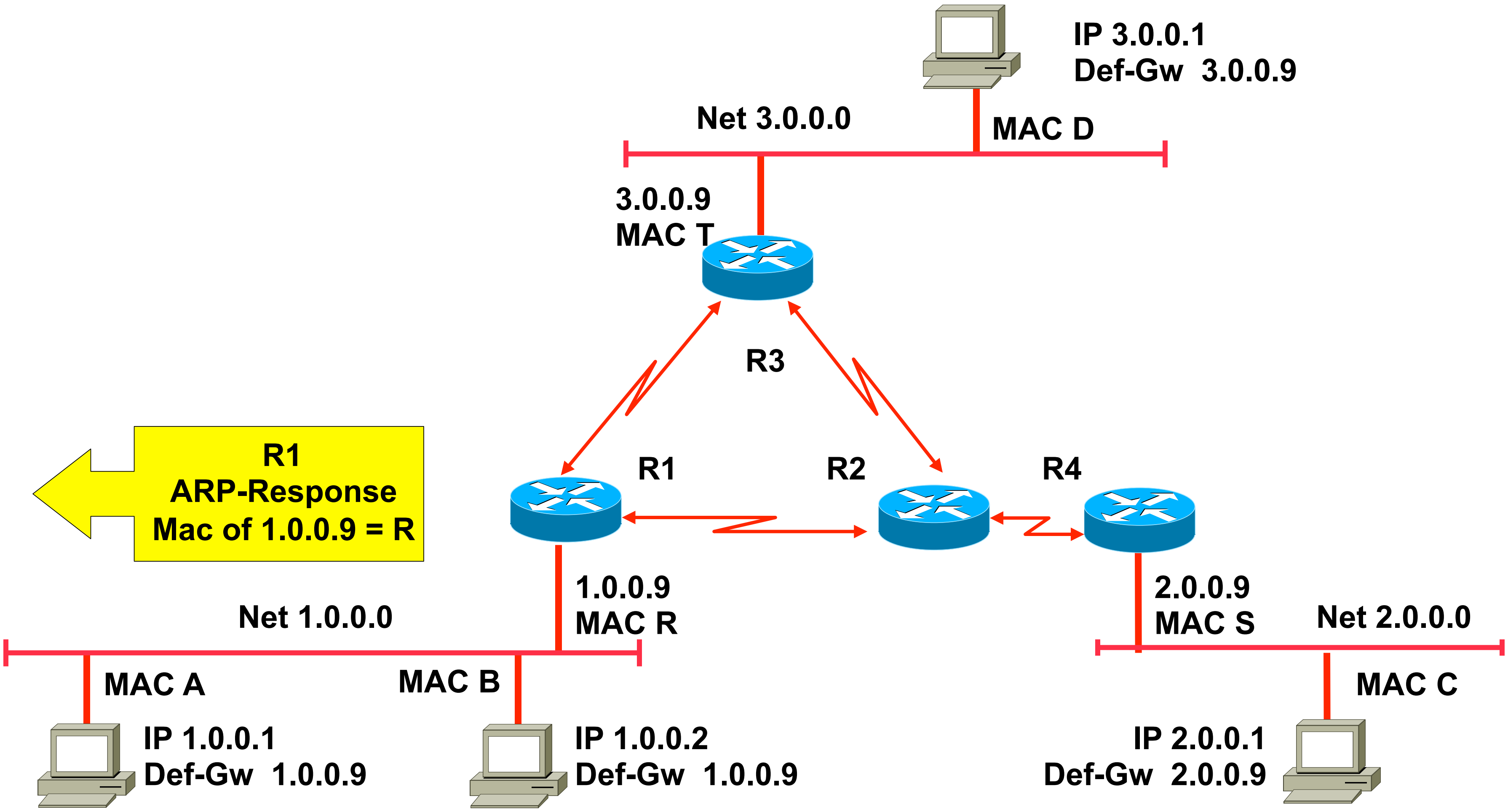
net-ID of destination
unequal
net-ID of source
-> use default gateway R1



Host A
ARP-Request
? Mac of 1.0.0.9

ARP-Cache Host A	
1.0.0.2	MAC B

Indirect Delivery 1.0.0.1 - > 2.0.0.1



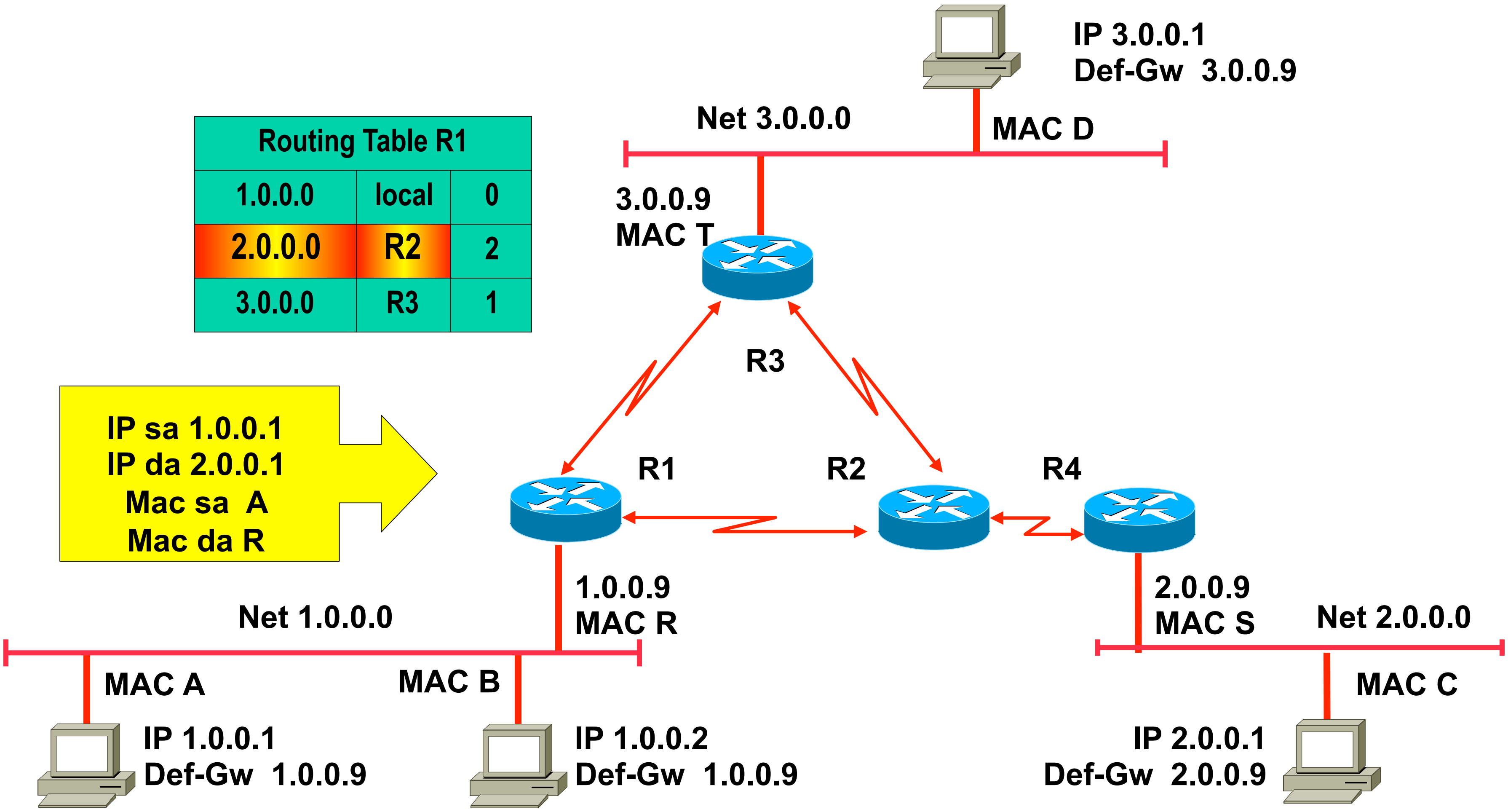
R1
ARP-Response
Mac of 1.0.0.9 = R

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

Indirect Delivery 1.0.0.1 - > 2.0.0.1

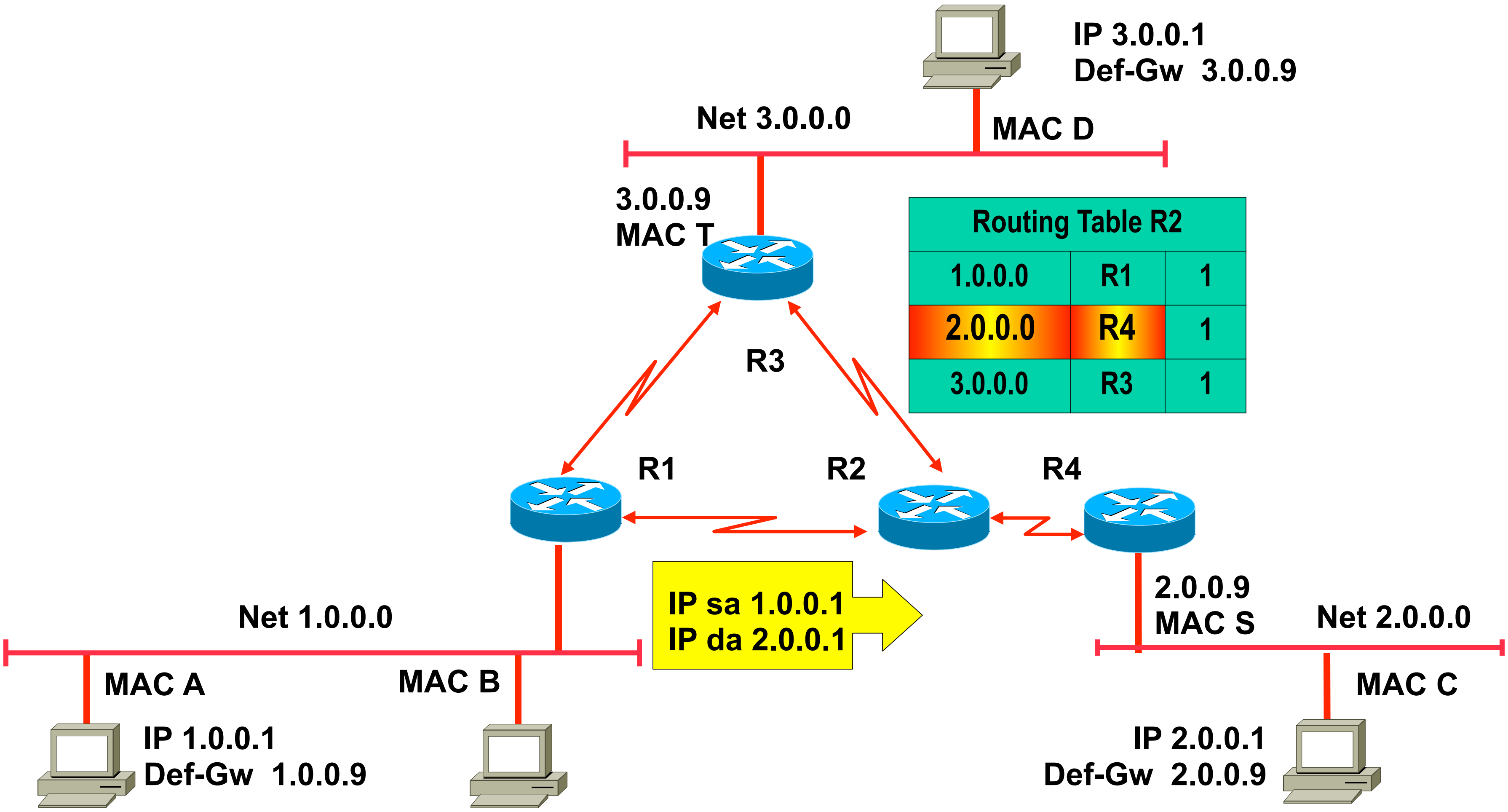
Routing Table R1		
1.0.0.0	local	0
2.0.0.0	R2	2
3.0.0.0	R3	1

IP sa 1.0.0.1
IP da 2.0.0.1
Mac sa A
Mac da R



ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

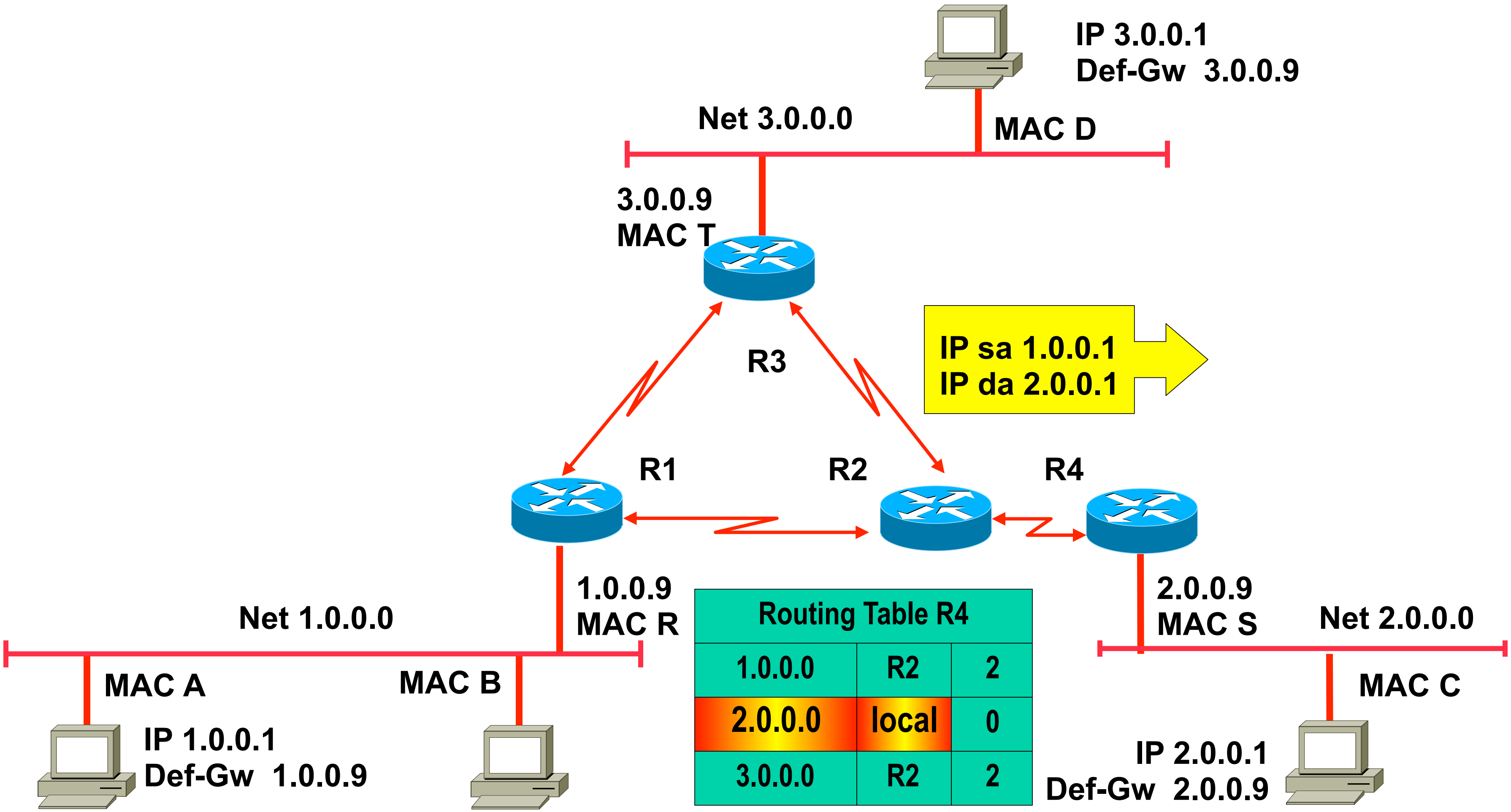
Indirect Delivery 1.0.0.1 - > 2.0.0.1



Routing Table R2		
1.0.0.0	R1	1
2.0.0.0	R4	1
3.0.0.0	R3	1

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

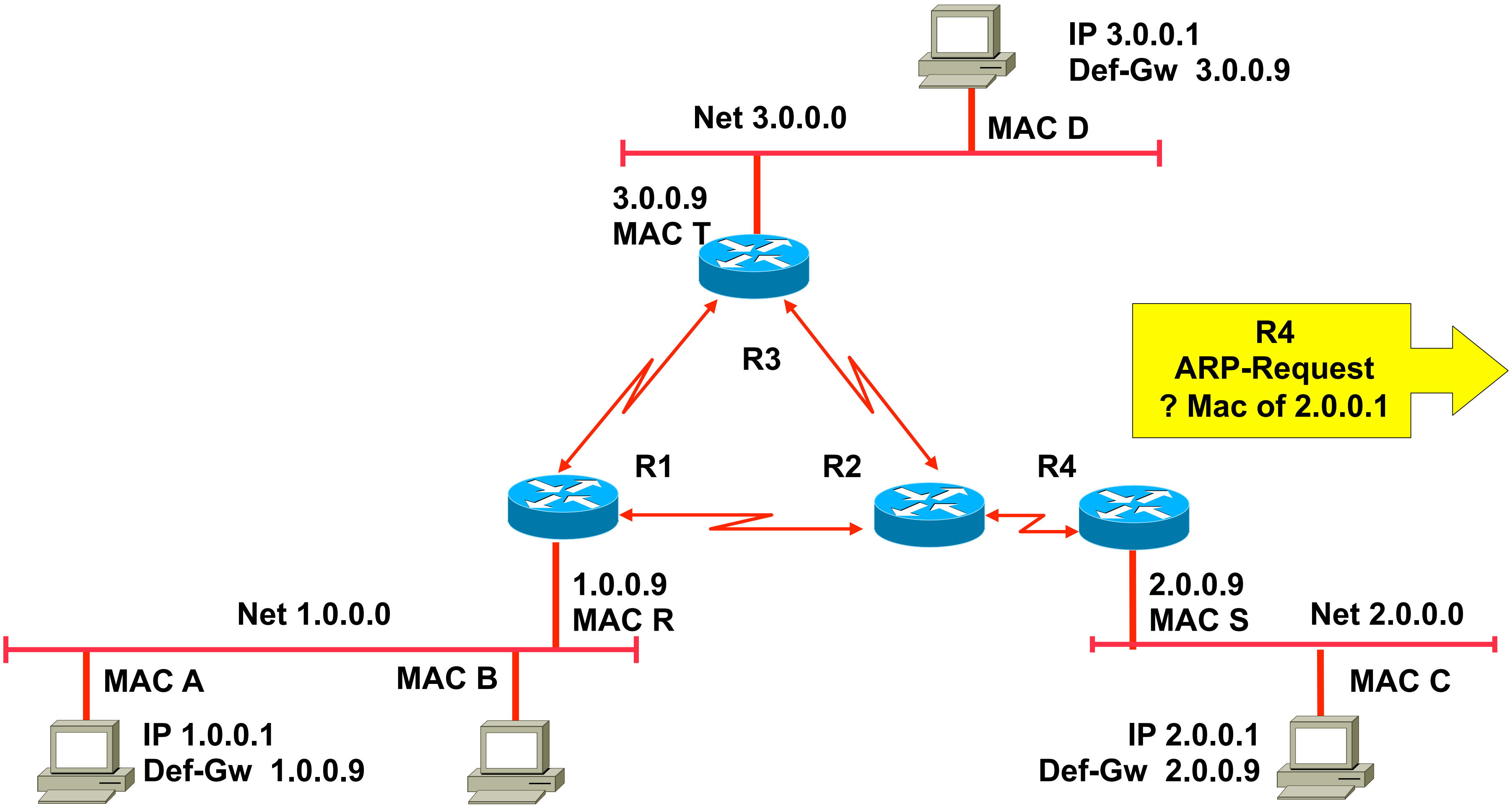
Indirect Delivery 1.0.0.1 - > 2.0.0.1



Routing Table R4		
1.0.0.0	R2	2
2.0.0.0	local	0
3.0.0.0	R2	2

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

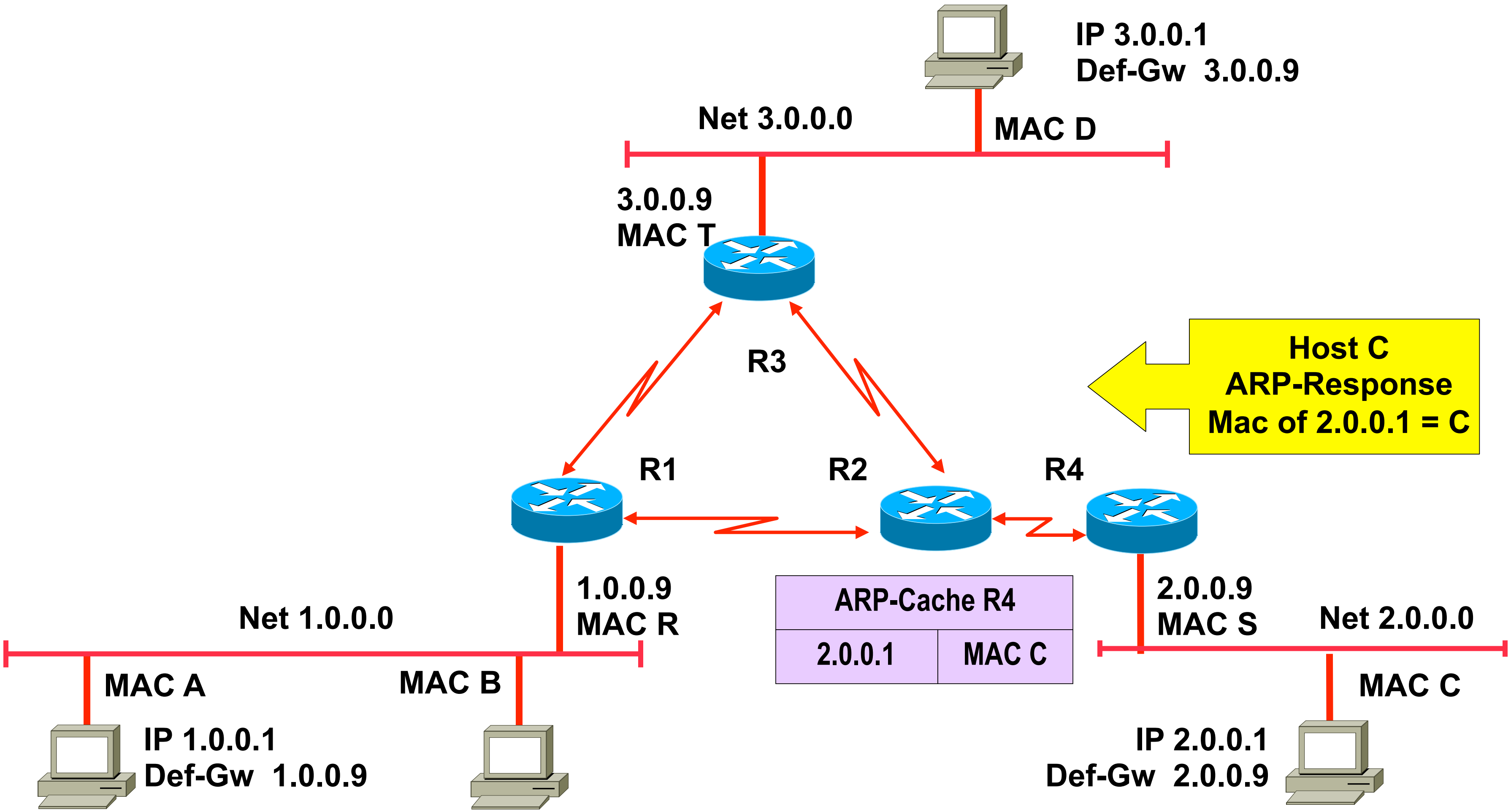
Indirect Delivery 1.0.0.1 - > 2.0.0.1



R4
ARP-Request
? Mac of 2.0.0.1

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

Indirect Delivery 1.0.0.1 - > 2.0.0.1

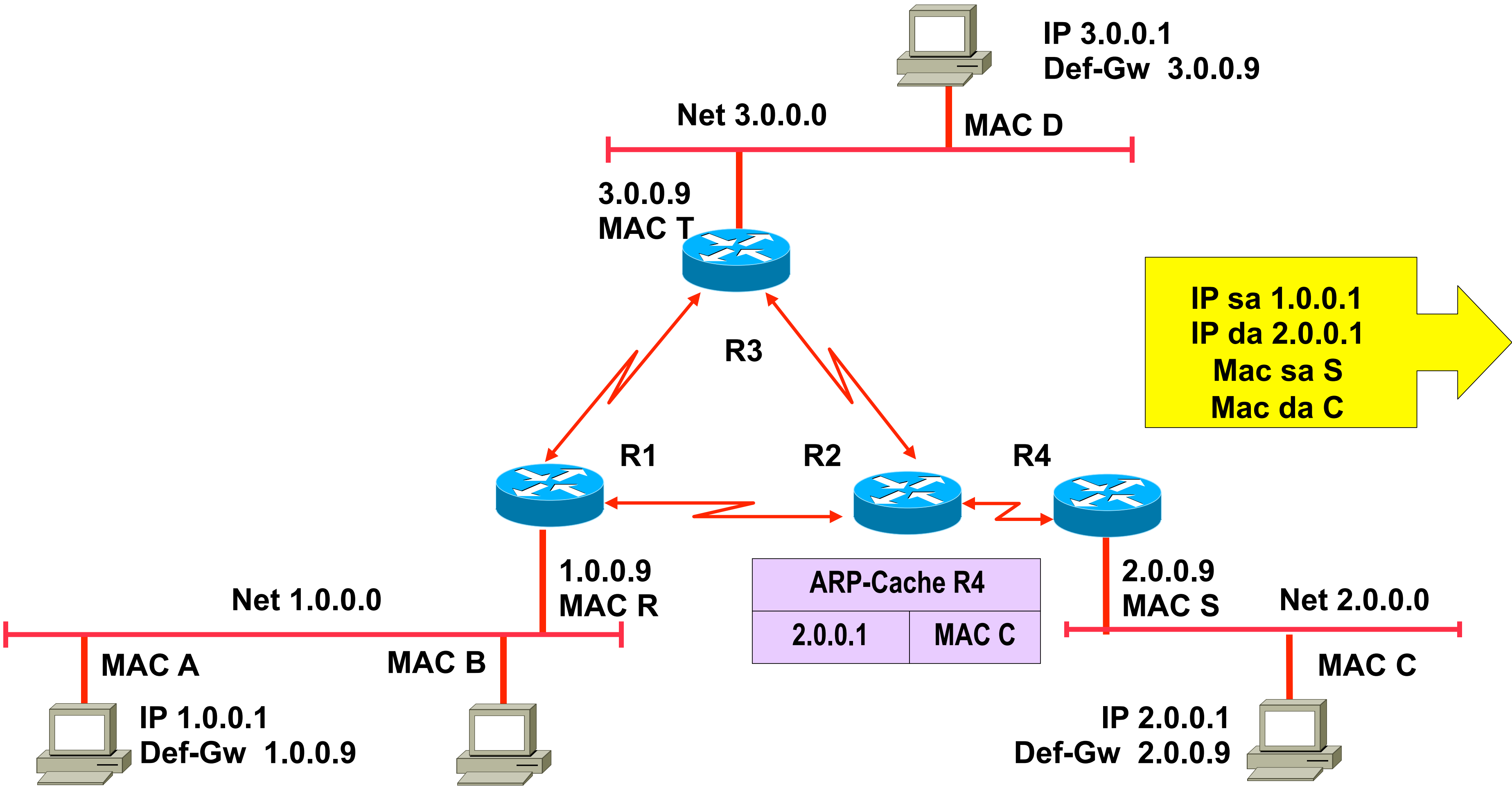


Host C
ARP-Response
Mac of 2.0.0.1 = C

ARP-Cache R4	
2.0.0.1	MAC C

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

Indirect Delivery 1.0.0.1 - > 2.0.0.1

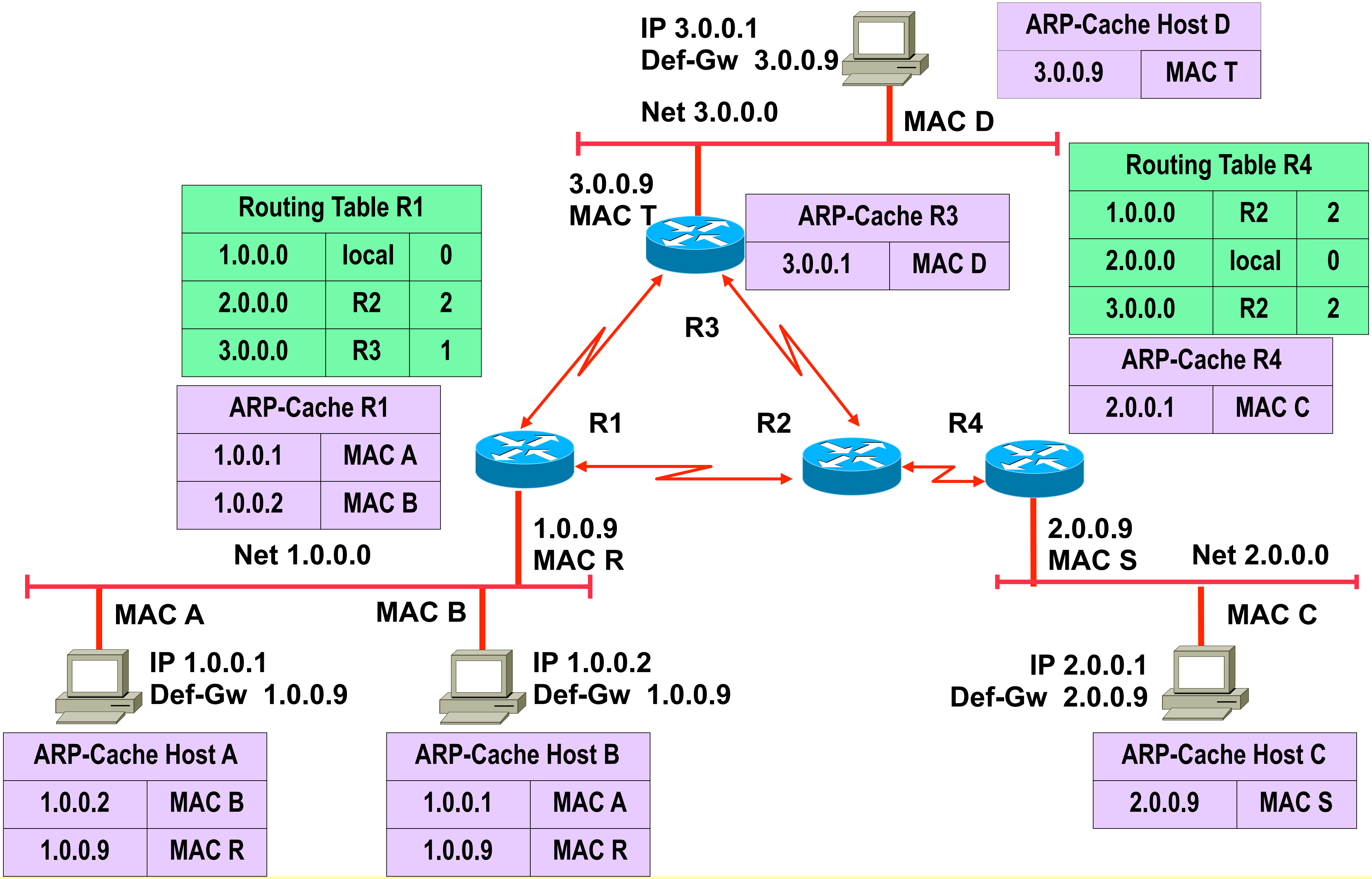


IP sa 1.0.0.1
IP da 2.0.0.1
Mac sa S
Mac da C

ARP-Cache R4	
2.0.0.1	MAC C

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

ARP Cache - Final Picture



ARP-Cache Host D	
3.0.0.9	MAC T

Routing Table R1		
1.0.0.0	local	0
2.0.0.0	R2	2
3.0.0.0	R3	1

ARP-Cache R3	
3.0.0.1	MAC D

Routing Table R4		
1.0.0.0	R2	2
2.0.0.0	local	0
3.0.0.0	R2	2

ARP-Cache R4	
2.0.0.1	MAC C

ARP-Cache R1	
1.0.0.1	MAC A
1.0.0.2	MAC B

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

ARP-Cache Host B	
1.0.0.1	MAC A
1.0.0.9	MAC R

ARP-Cache Host C	
2.0.0.9	MAC S

Agenda

- L2 versus L3 Switching
- IP Protocol, IP Addressing
- IP Forwarding
- ARP and ICMP
- IP Routing
- First Hop Redundancy

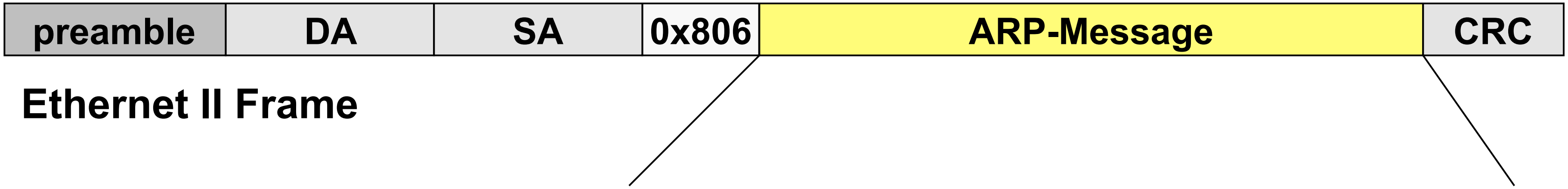
IP Address versus L2 Address

- **IP address**
  - Identifies the access to a network (interface)
- **If the physical network is of point-to-point link to another IP system**
  - This IP system can be reached without any further addressing on layer 2
- **On a shared media or multipoint-network**
  - Layer 2 addresses are necessary to deliver packets to a specific station using the corresponding L2 technology (LAN, Frame-Relay, ATM ...)
- **Hence a mapping between IP address and L2 address is needed**

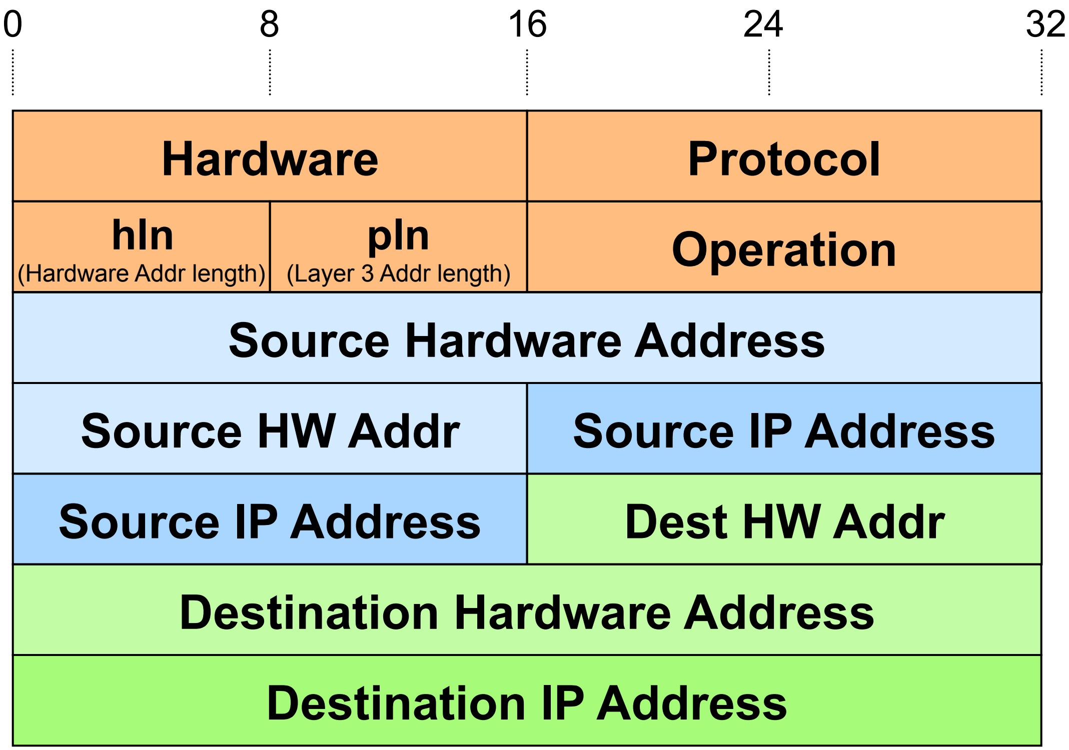
ARP (Address Resolution Protocol)

- **In case of LAN**
  - The mapping is between MAC- and IP-addresses
- **Mapping can be static or dynamic**
- **ARP protocol is used in case of dynamic mapping**
  - RFC 826
  - Defines procedure to request a mapping for a given IP address and stores the result in the so called ARP cache memory
  - ARP cache will be checked first before new requests are sent
  - ARP cache can be refreshed or times out

ARP Format



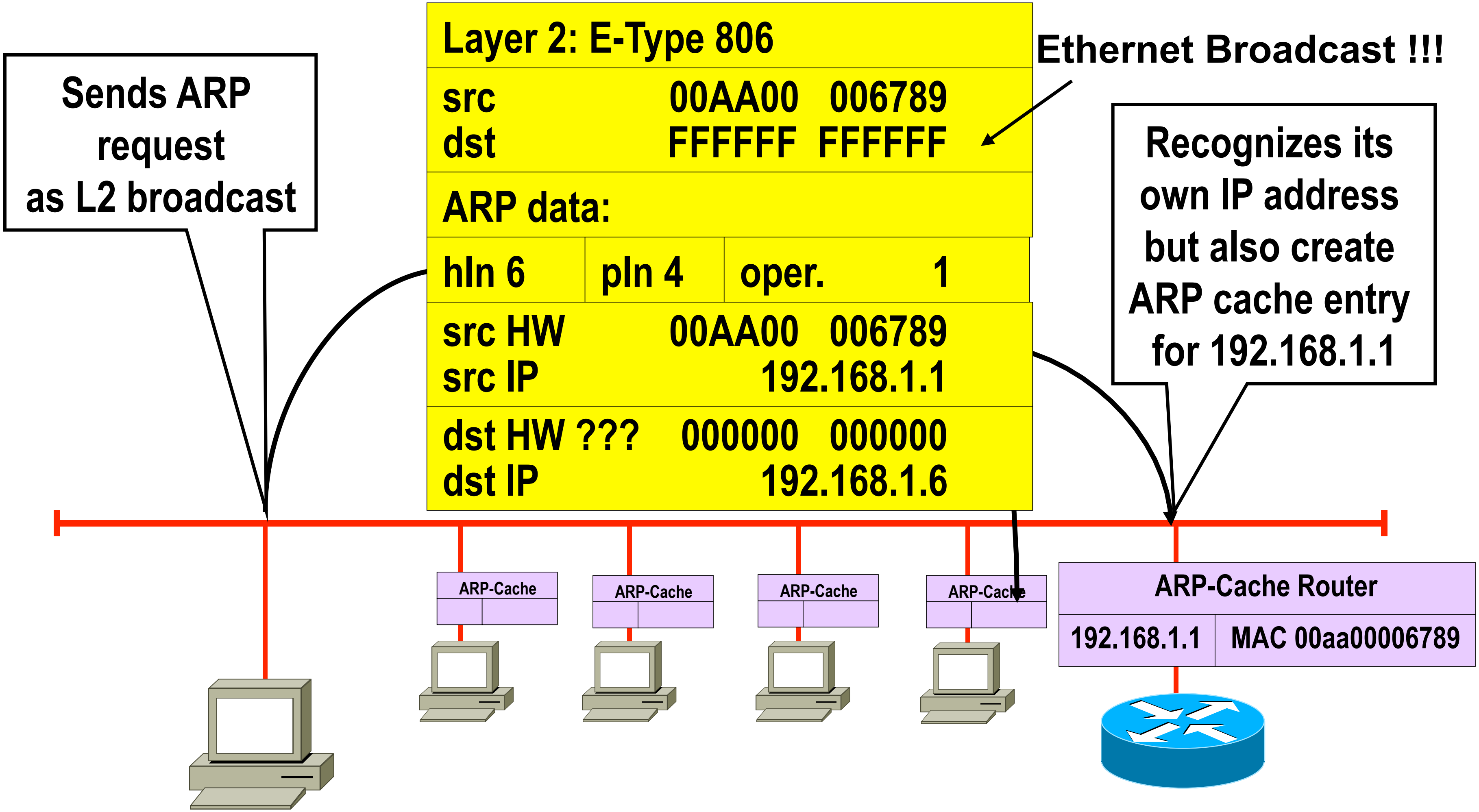
Ethernet II Frame



Example ARP Request (Ethernet / IP):

Hardware: 6 (IEEE802.x)
Protocol: 0x0800 (IP)
hln: 6 (MAC Address in Bytes)
pln: 4 (IP Address in Bytes)
Operation: 1 (ARP Request)
Source HW Addr: hex: 00 60 97 bc 88 f1
Source IP Addr: 192.168.1.1
Dest HW Addr: hex: ff ff ff ff ff ff
Dest IP Addr: 192.168.1.254

ARP Request



Sends ARP request as L2 broadcast

Layer 2: E-Type 806

src	00AA00	006789
dst	FFFFFF	FFFFFF

Ethernet Broadcast !!!

ARP data:

hln 6	pln 4	oper. 1

src HW	00AA00	006789
src IP		192.168.1.1
dst HW ???	000000	000000
dst IP		192.168.1.6

Recognizes its own IP address but also create ARP cache entry for 192.168.1.1

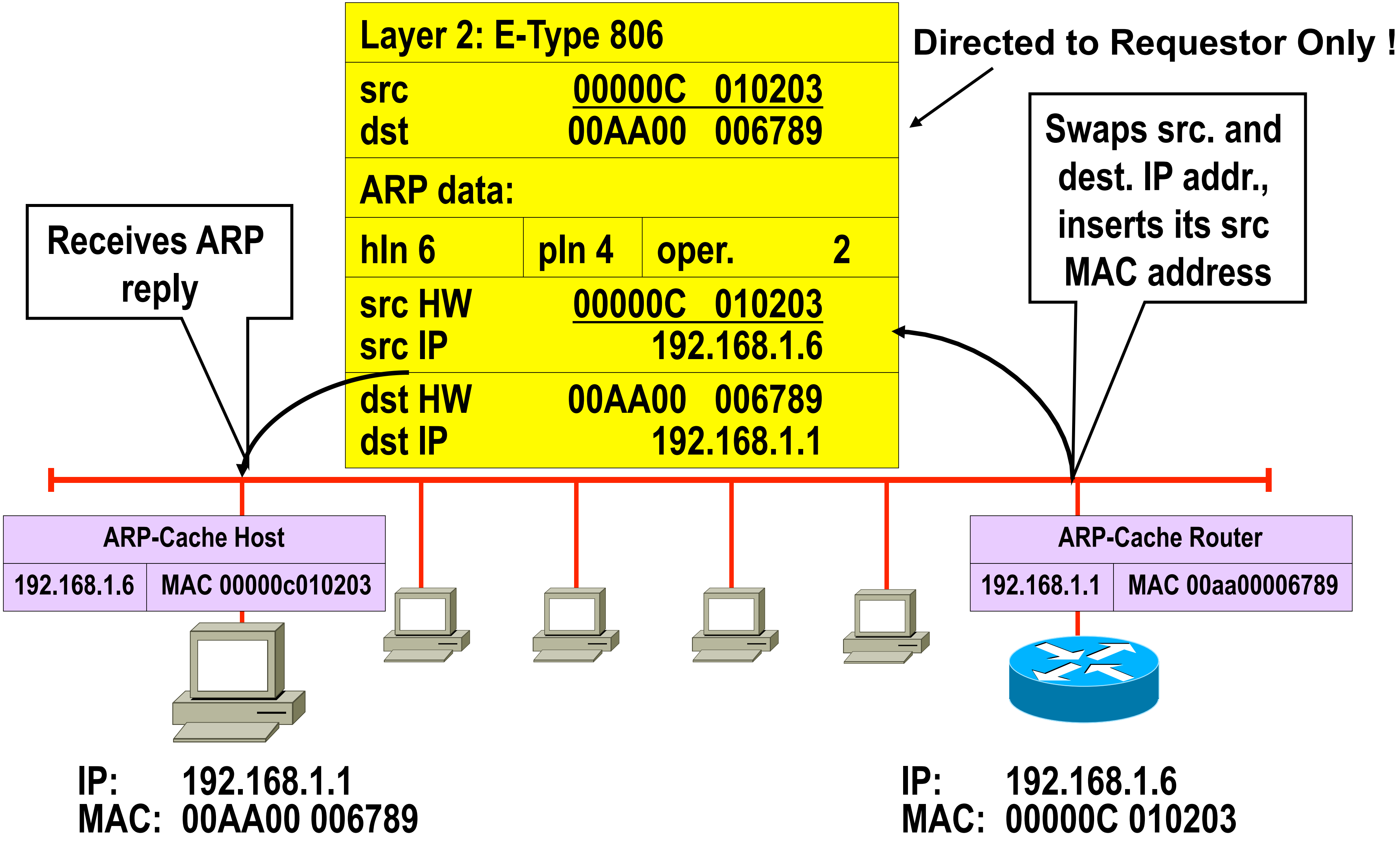
ARP-Cache Router

192.168.1.1	MAC 00aa00006789

IP: 192.168.1.1
MAC: 00AA00 006789

IP: 192.168.1.6
MAC: 00000C 010203

ARP Reply



Layer 2: E-Type 806

src 00000C 010203
dst 00AA00 006789

ARP data:

hln 6	pln 4	oper. 2

src HW 00000C 010203
src IP 192.168.1.6

dst HW 00AA00 006789
dst IP 192.168.1.1

Directed to Requestor Only !

Swaps src. and dest. IP addr., inserts its src MAC address

Receives ARP reply

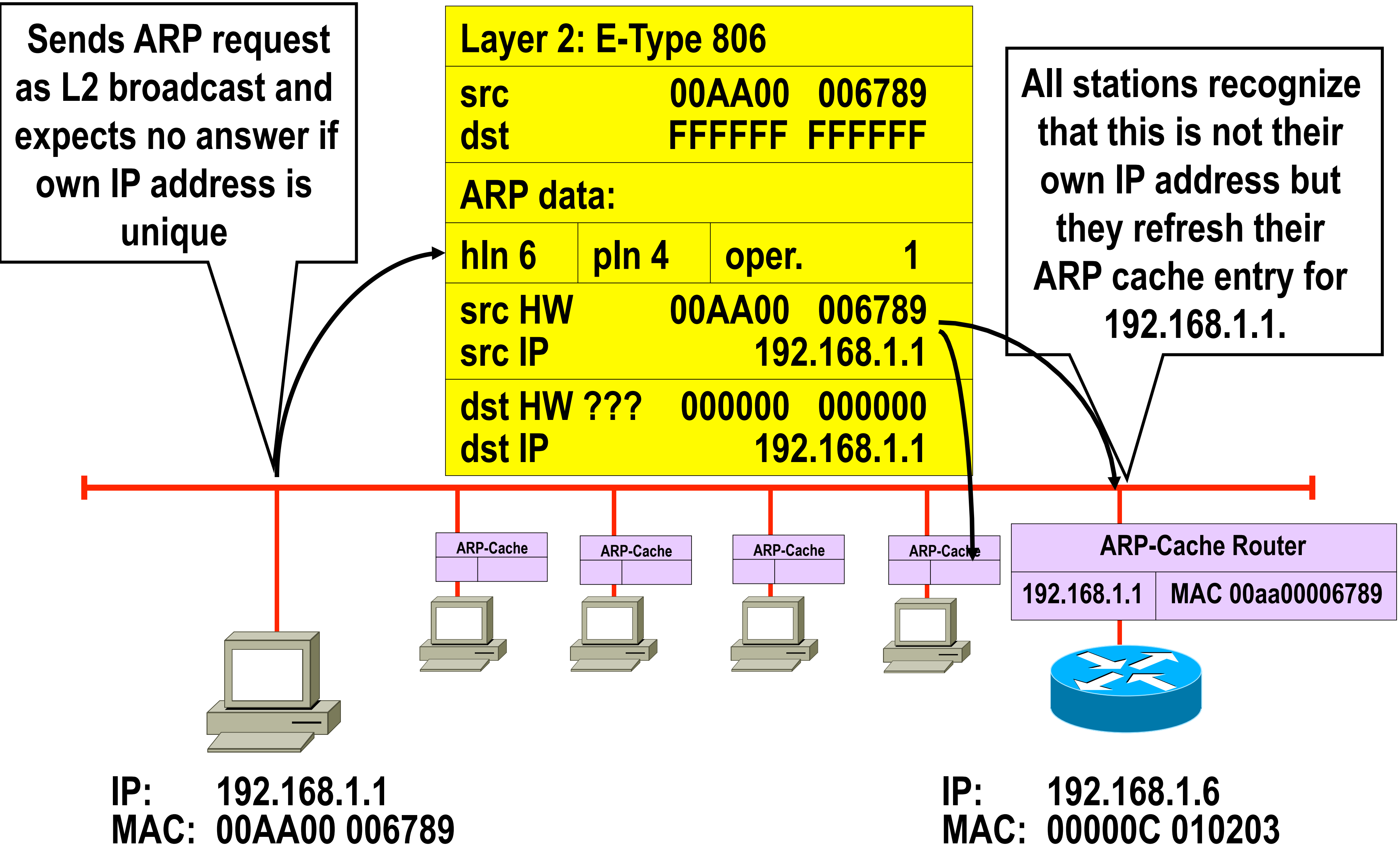
ARP-Cache Host	
192.168.1.6	MAC 00000c010203

ARP-Cache Router	
192.168.1.1	MAC 00aa00006789

IP: 192.168.1.1
MAC: 00AA00 006789

IP: 192.168.1.6
MAC: 00000C 010203

Gratuitous ARP for Duplicate Address Check and ARP Cache Refresh



Sends ARP request as L2 broadcast and expects no answer if own IP address is unique

Layer 2: E-Type 806

src 00AA00 006789
dst FFFFFF FFFFFF

ARP data:

hln 6 | pln 4 | oper. 1

src HW 00AA00 006789
src IP 192.168.1.1

dst HW ??? 000000 000000
dst IP 192.168.1.1

All stations recognize that this is not their own IP address but they refresh their ARP cache entry for 192.168.1.1.

ARP-Cache Router	
192.168.1.1	MAC 00aa00006789

IP: 192.168.1.1
MAC: 00AA00 006789

IP: 192.168.1.6
MAC: 00000C 010203

ICMP (RFC 792)

- **Datagram service of IP**
  - Best effort -> IP datagrams can be lost
  - If network cannot deliver packets the sender must be informed somehow !
    - Reasons: no route, TTL expired, ...
- **ICMP (Internet Control Message Protocol)**
  - Enhances network reliability and performance by carrying error and diagnostic messages
- **ICMP must be supported by every IP station**
  - Implementation differences!
- **Analysis of ICMP messages**
  - Network management systems or can give valuable hints for the network administrator

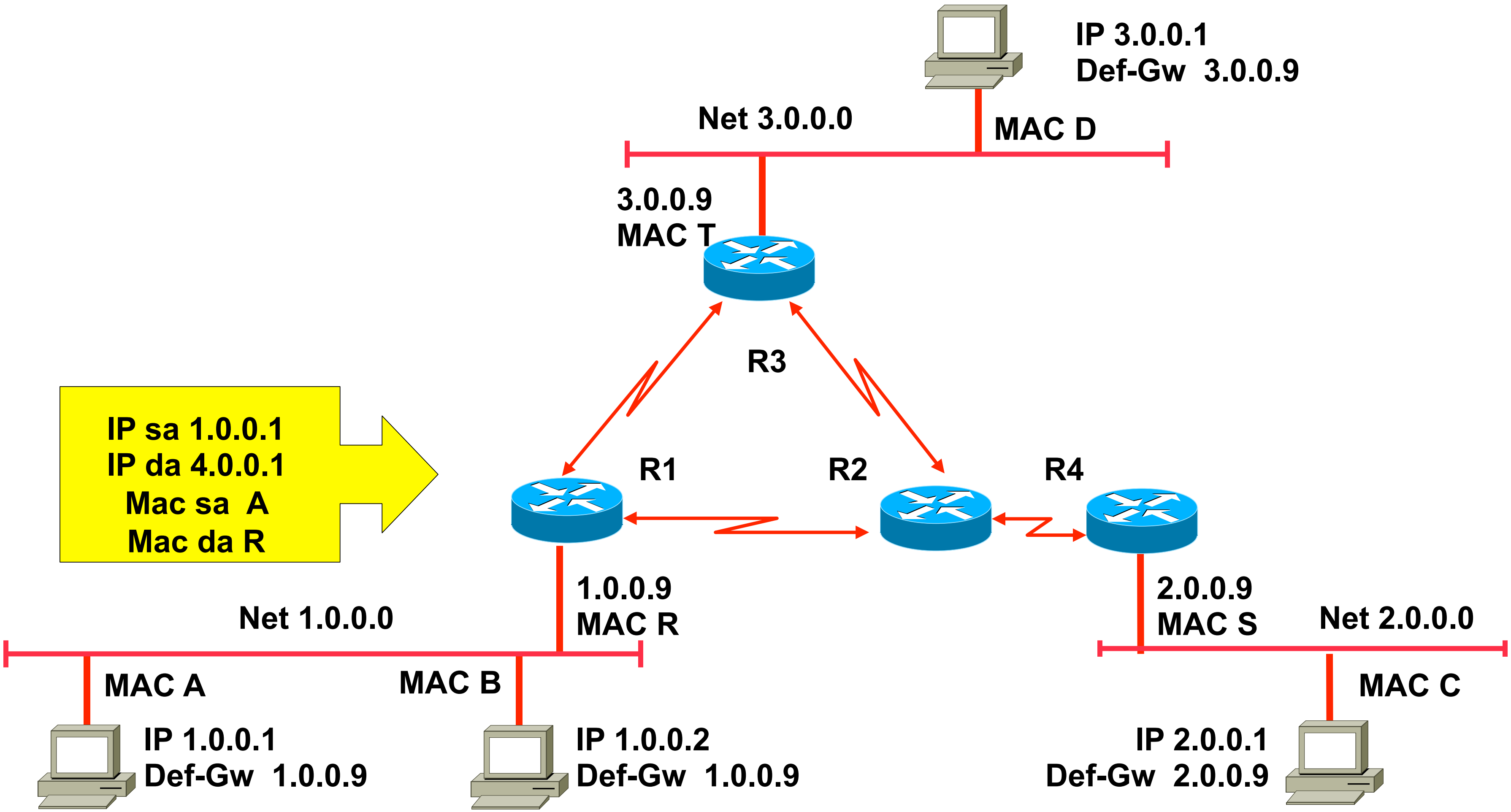
ICMP

- **Principle of ICMP operation**
  - IP station (router or destination), which detects any transmission problems, generates an ICMP message
  - ICMP message is addressed to the originating station (sender of the original IP packet)
- **ICMP messages are sent as IP packets**
    - Protocol field = 1, ICMP header and code in the IP data area
- **If an IP datagram carrying an ICMP message cannot be delivered**
  - No additional ICMP error message is generated to avoid an ICMP avalanche
  - "ICMP must not invoke ICMP"
    - Exception: PING command (Echo request and echo response)

ICMP Message Types

0	Echo Reply ("Ping Response")
3	Destination Unreachable
	Reason specified in Code field of ICMP message
4	Source Quench (decrease data rate of sender)
	Theoretical Flow Control Possibility of IP
5	Redirect (use different router)
	More information in Code field of ICMP message
8	Echo Request ("Ping Request")
11	Time Exceeded
	code = 0 time to live exceeded in transit
	code = 1 reassembly timer expired
12	Parameter Problem (IP header)
13/14	Time Stamp Request / Time Stamp Reply
15/16	Information Request / Reply
	e.g. finding the Net-ID of the network
17/18	Address Mask Request / Reply

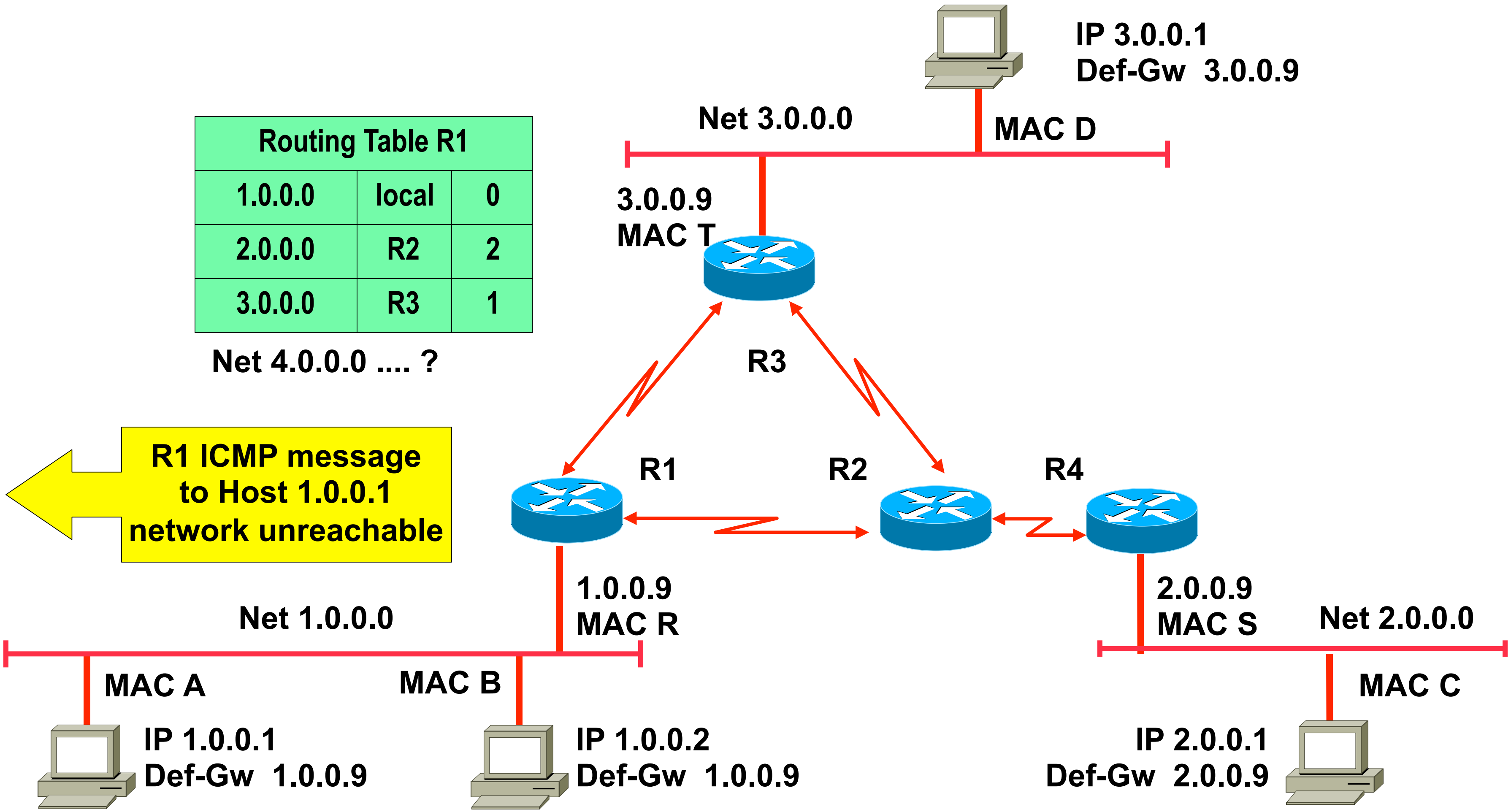
Delivery 1.0.0.1 - > 4.0.0.1



IP sa 1.0.0.1
IP da 4.0.0.1
Mac sa A
Mac da R

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

ICMP Destination Unreachable (code: network unreachable)



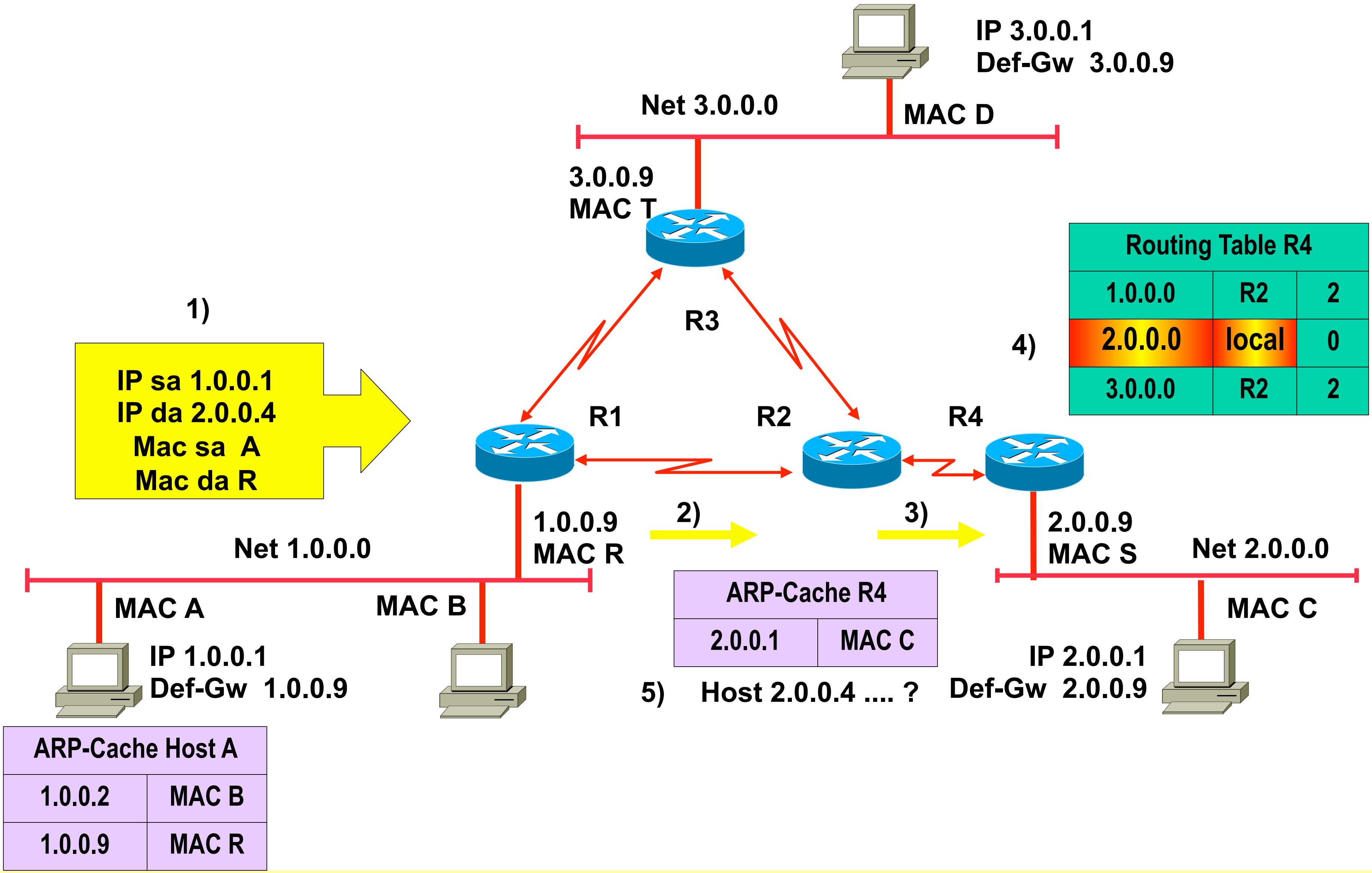
Routing Table R1		
1.0.0.0	local	0
2.0.0.0	R2	2
3.0.0.0	R3	1

Net 4.0.0.0 .... ?

R1 ICMP message to Host 1.0.0.1 network unreachable

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

Delivery 1.0.0.1 - > 2.0.0.4



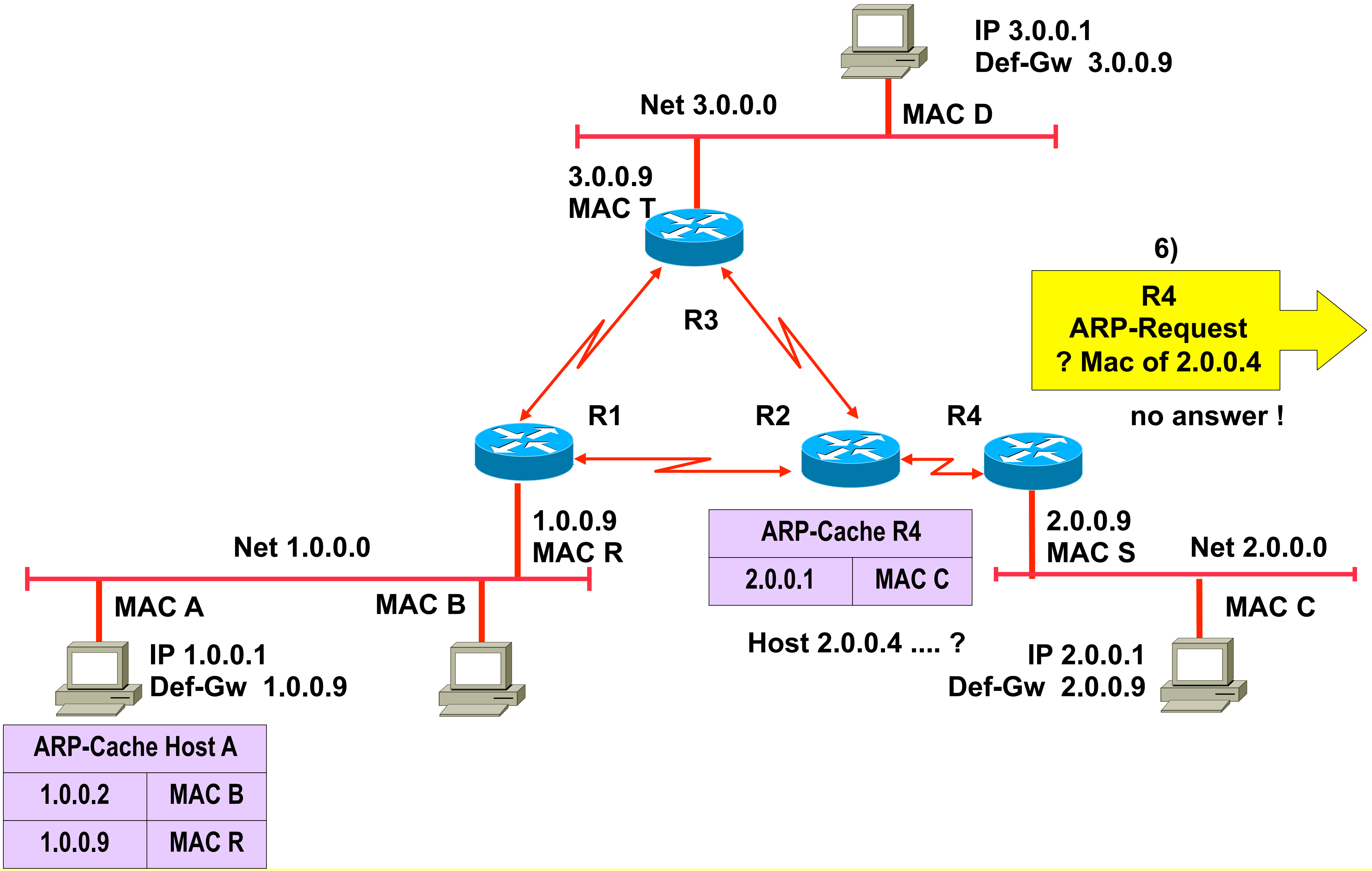
IP sa 1.0.0.1
IP da 2.0.0.4
Mac sa A
Mac da R

Routing Table R4		
1.0.0.0	R2	2
2.0.0.0	local	0
3.0.0.0	R2	2

ARP-Cache R4	
2.0.0.1	MAC C

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

Delivery 1.0.0.1 - > 2.0.0.4



6)

R4
ARP-Request
? Mac of 2.0.0.4

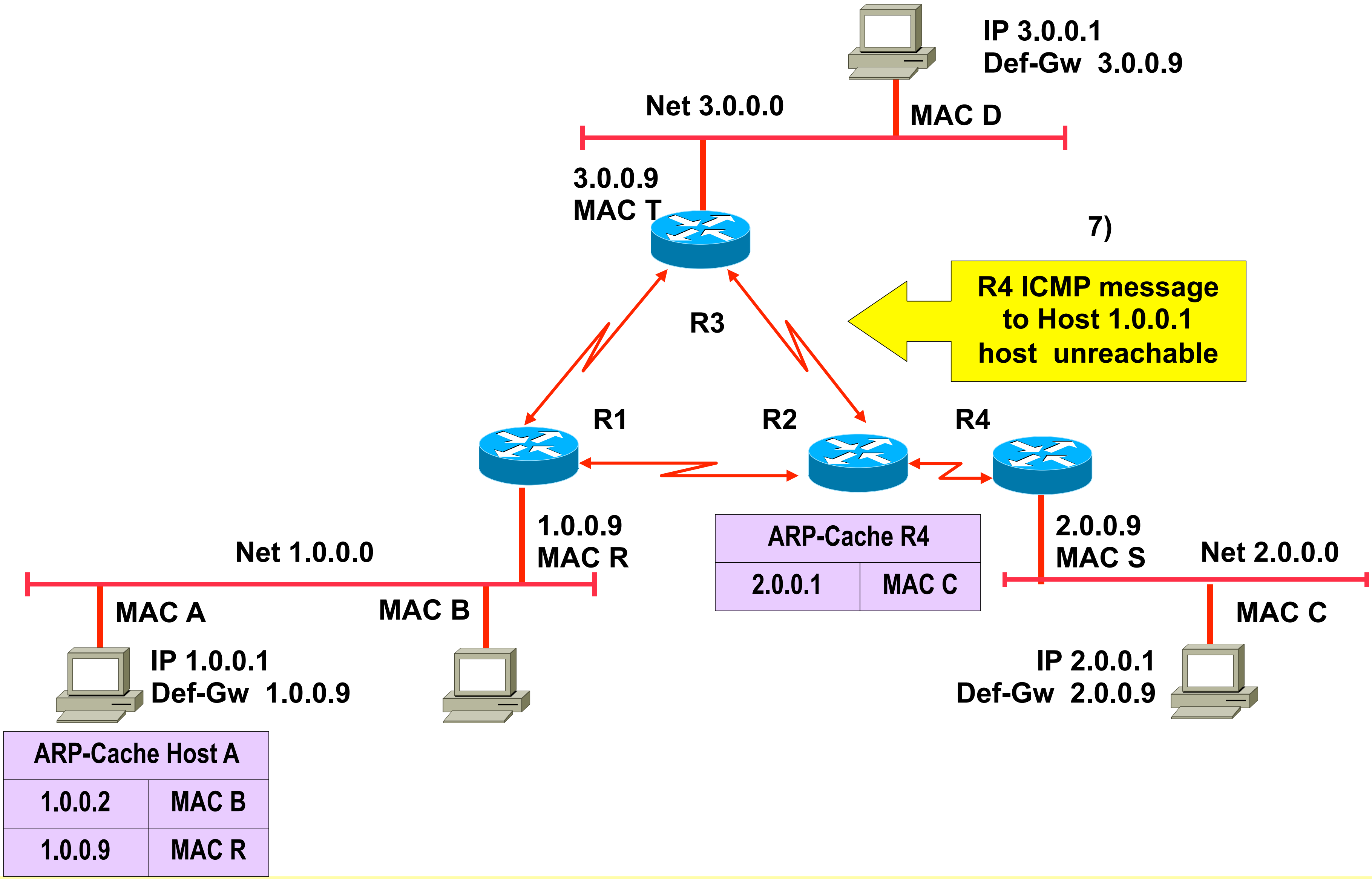
no answer !

ARP-Cache R4	
2.0.0.1	MAC C

Host 2.0.0.4 .... ?

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

ICMP Destination Unreachable (code: host unreachable)

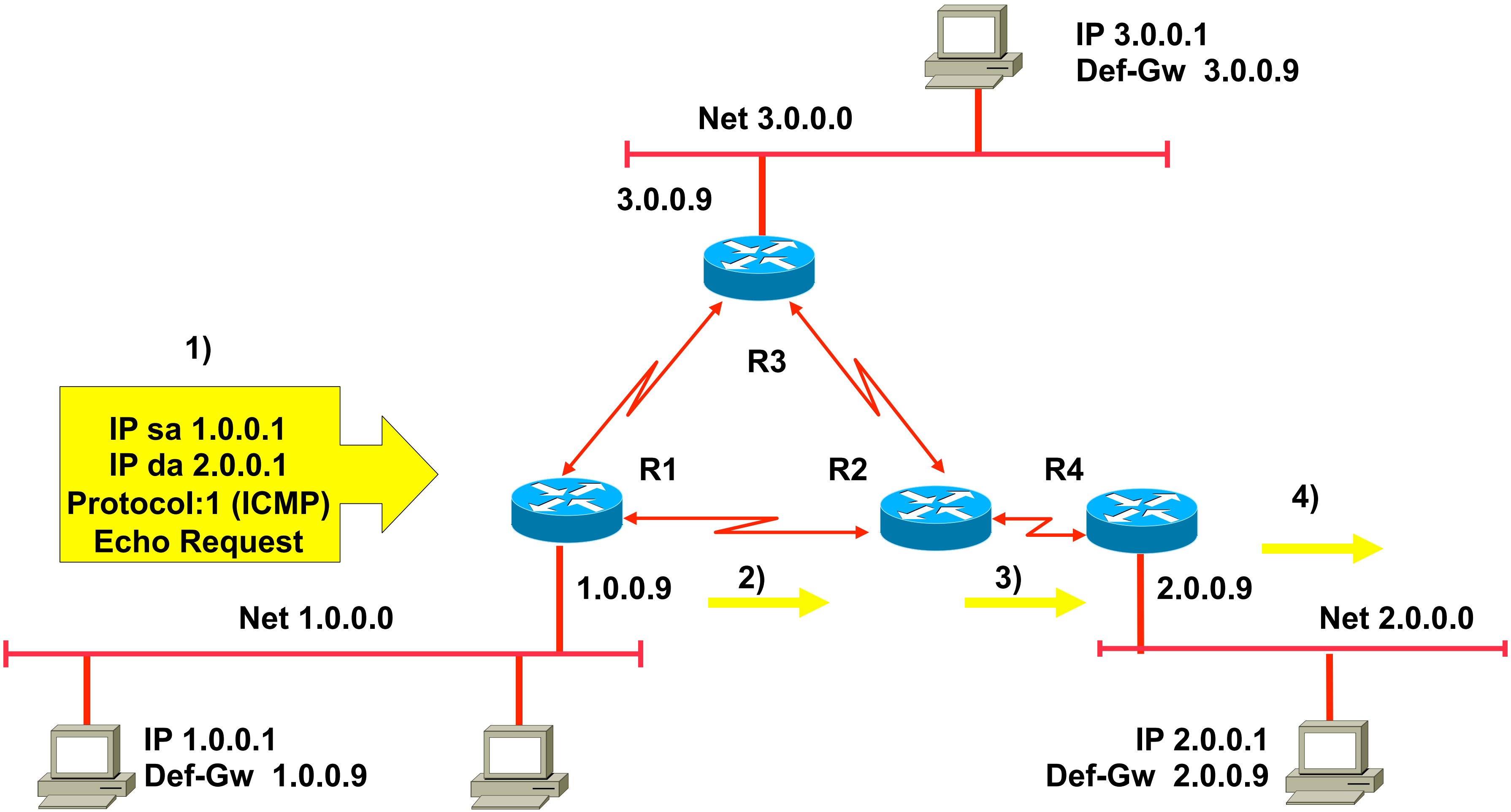


R4 ICMP message
to Host 1.0.0.1
host unreachable

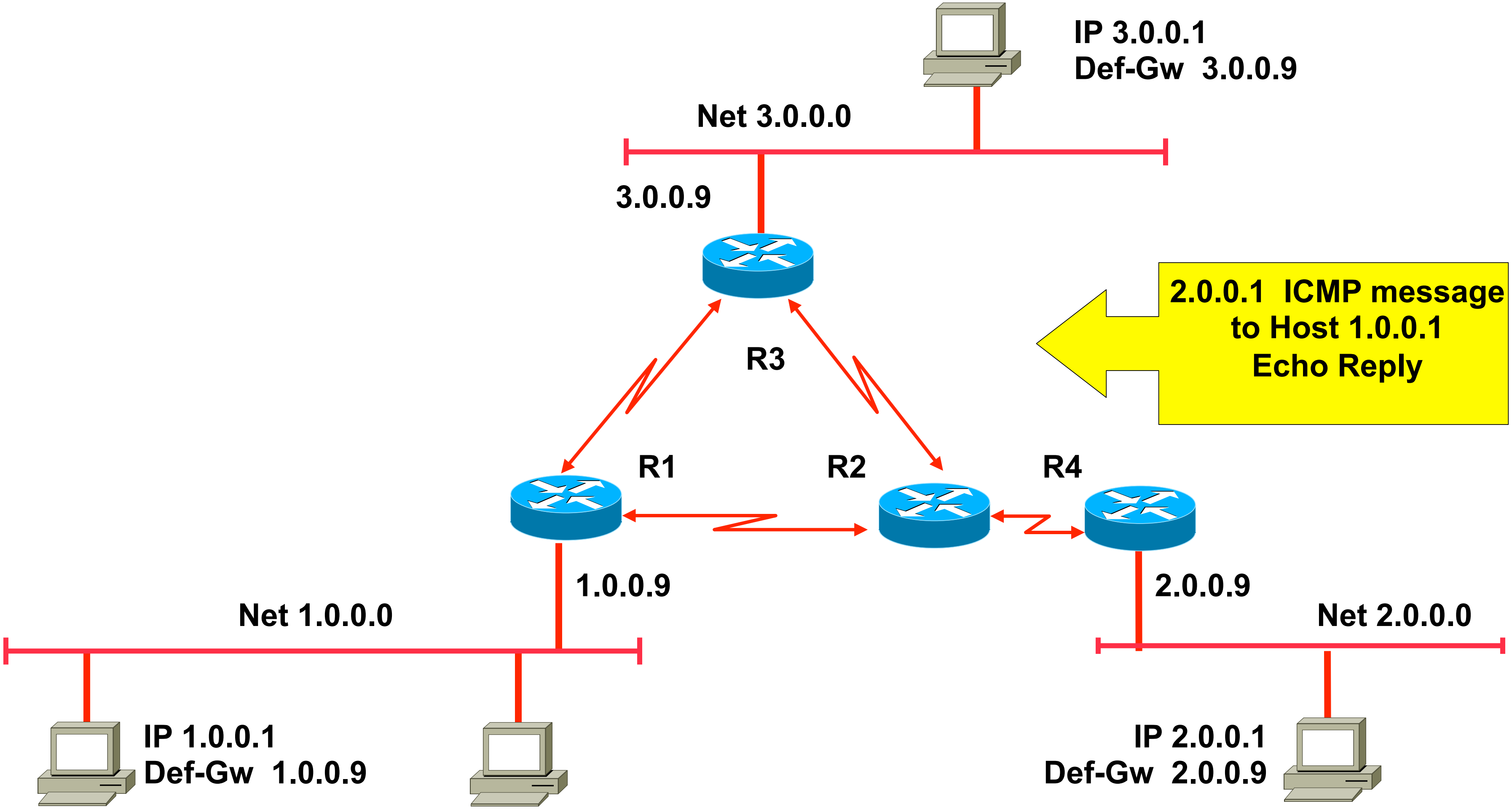
ARP-Cache R4	
2.0.0.1	MAC C

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

Ping 1.0.0.1 - > 2.0.0.1

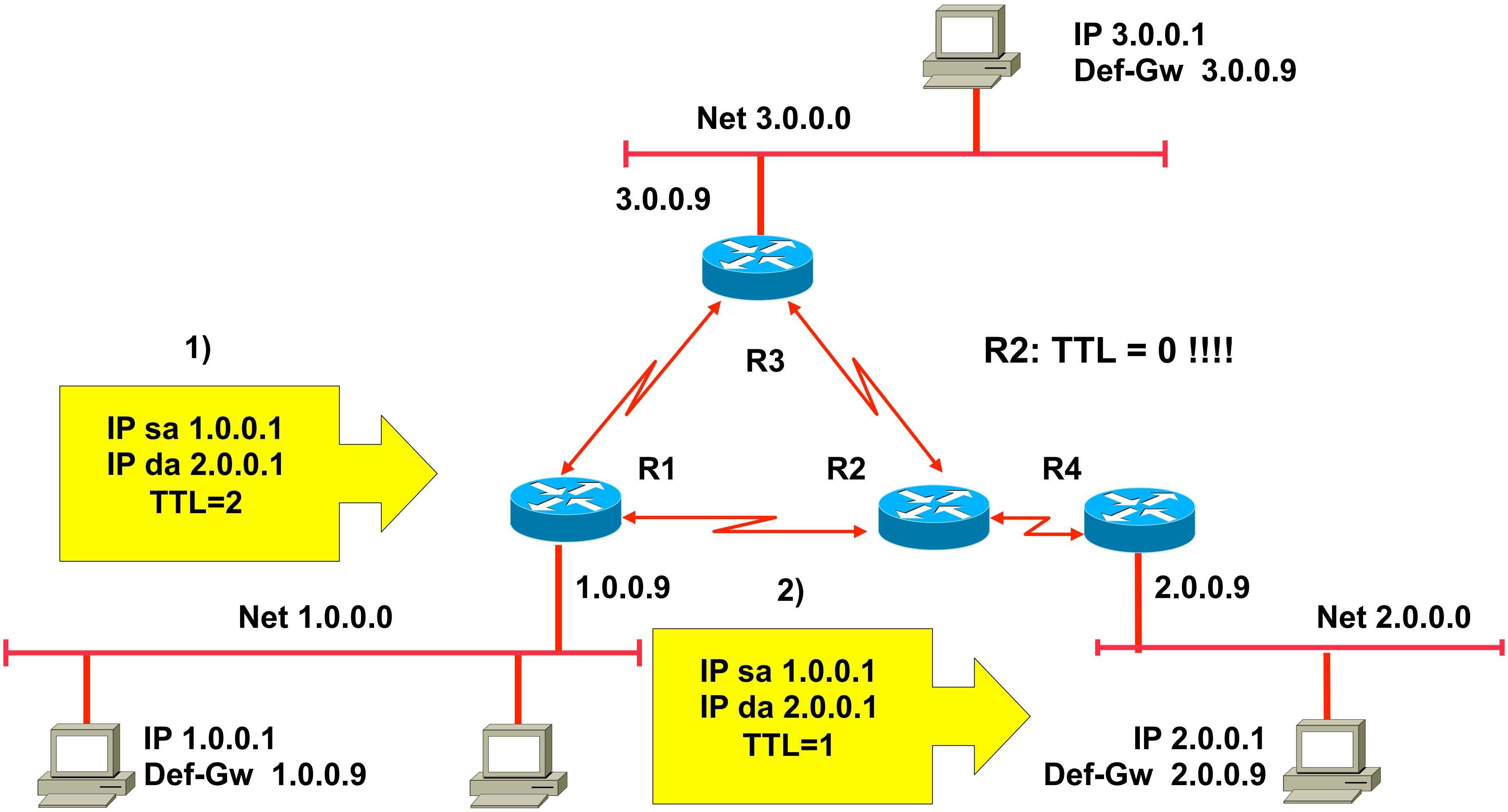


Ping Echo 2.0.0.1 - > 1.0.0.1



2.0.0.1 ICMP message
to Host 1.0.0.1
Echo Reply

Delivery 1.0.0.1 - > 2.0.0.1 (TTL=2)



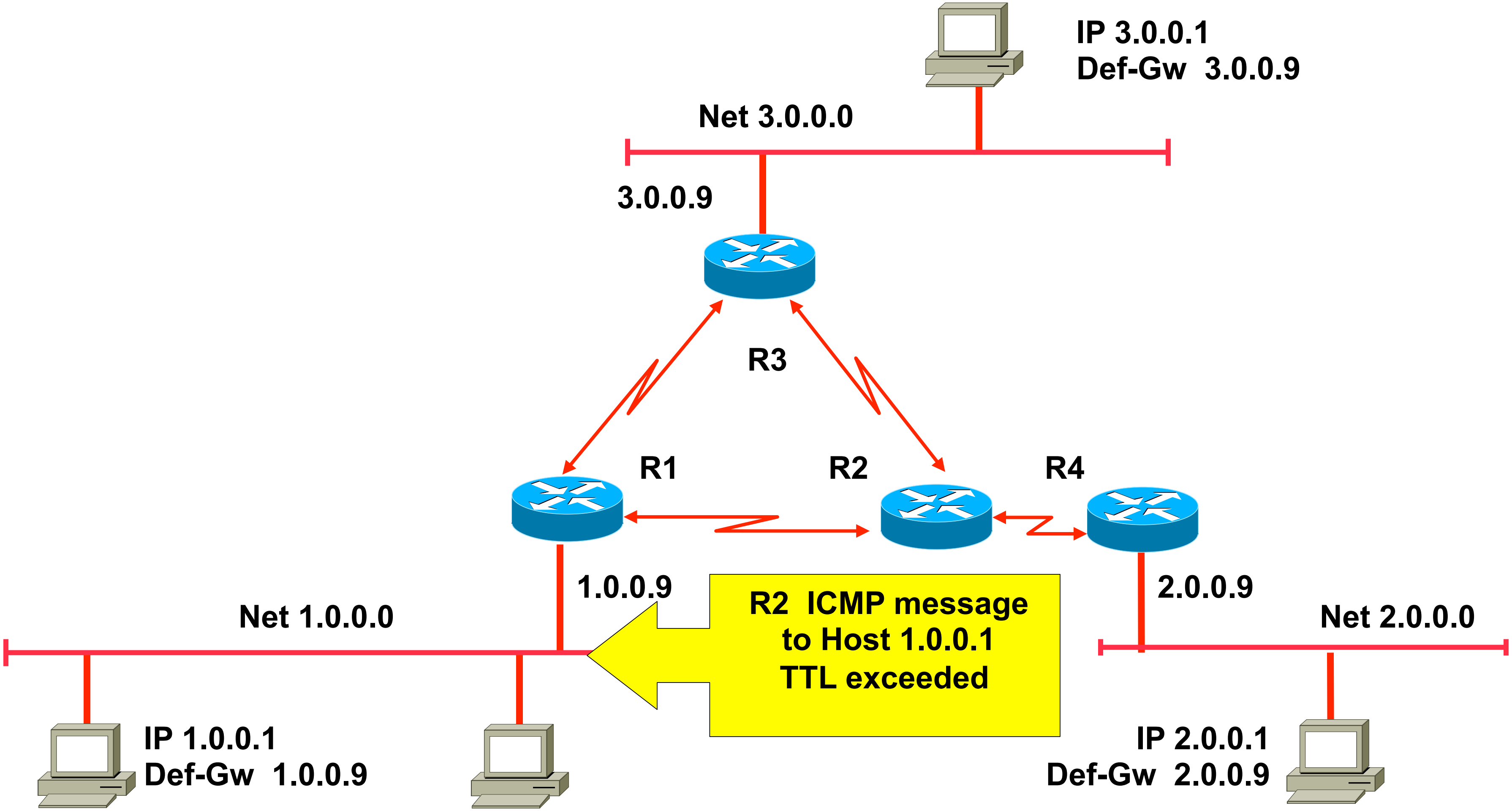
1)

IP sa 1.0.0.1
IP da 2.0.0.1
TTL=2

2)

IP sa 1.0.0.1
IP da 2.0.0.1
TTL=1

ICMP TTL exceeded



R2 ICMP message to Host 1.0.0.1 TTL exceeded

Traceroute

- **Using ICMP TTL exceed messages**
  - The current route, a datagram will take through the network, can be find
- **Just generate IP messages**
  - With increasing values for TTL
- **You will find the route**
  - Hop by hop
- **Two types of messages generated by of trace route CLI commands:**
  - ICMP-Echo
  - UDP

Agenda

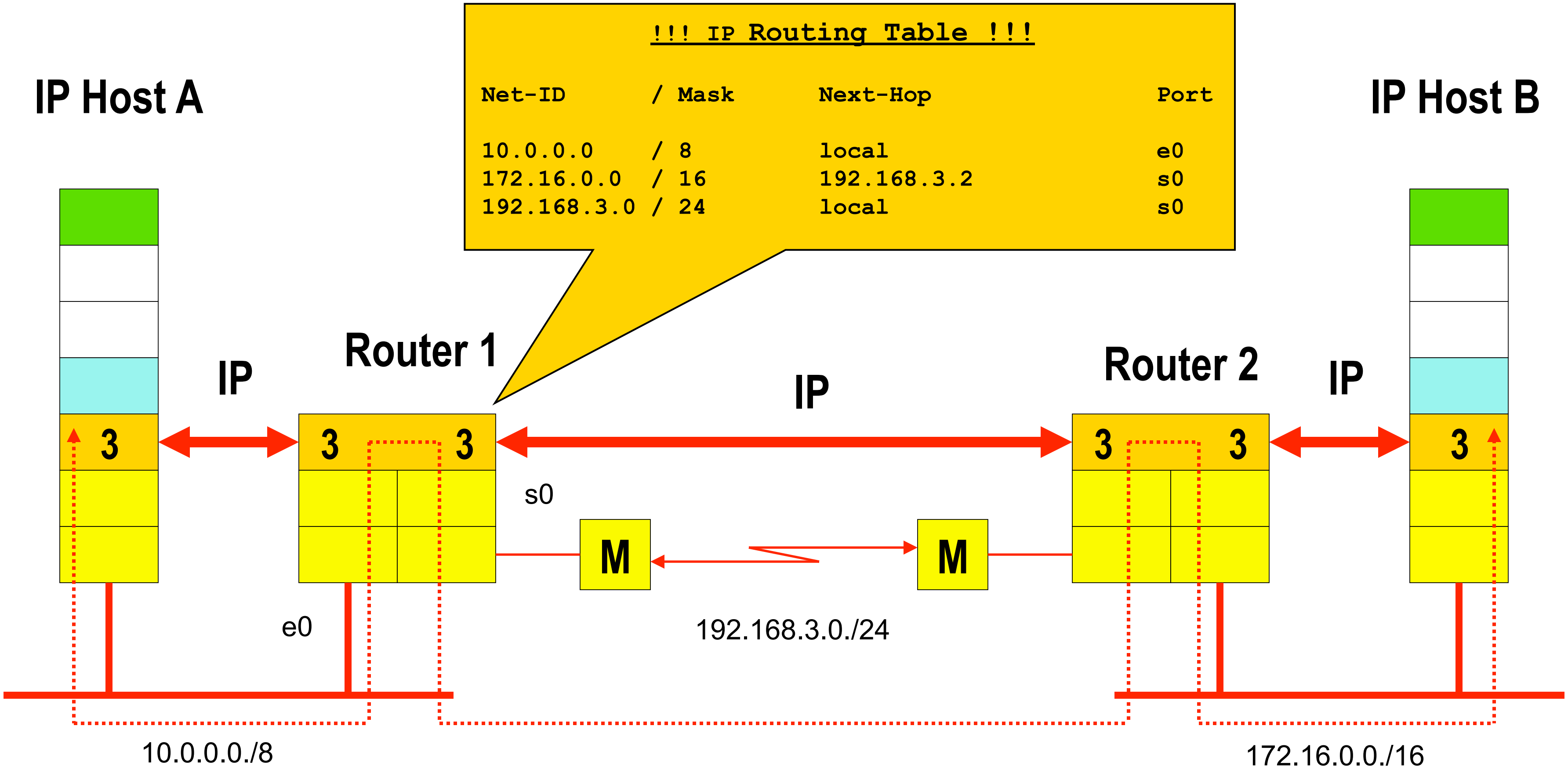
- L2 versus L3 Switching
- IP Protocol, IP Addressing
- IP Forwarding
- ARP and ICMP
- IP Routing
  - Introduction
  - OSPF Basics
  - OSPF Communication Procedures (Router LSA)
  - LSA Broadcast Handling (Flooding)
  - OSPF Splitted Area
  - Broadcast Networks (Network LSA)
- First Hop Redundancy

What is Routing?

- ***Finding / choosing a path to a destination address***
- **Direct delivery performed by IP host**
  - Destination network = local network
- **Indirect delivery performed by router**
  - Destination network ≠ local network
  - Datagram is forwarded to default gateway
  - Passed on by the router based on routing table
- **Routing table**
  - Database of known destinations
  - Signposts leading to next hop

IP, IP Routing Protocol, IP Routing Table

Layer 3 Protocol = IP
Layer 3 Routing Protocols = RIP, OSPF, EIGRP, BGP



!!! IP Routing Table !!!

Net-ID	/ Mask	Next-Hop	Port
10.0.0.0	/ 8	local	e0
172.16.0.0	/ 16	192.168.3.2	s0
192.168.3.0	/ 24	local	s0

IP Routing Paradigm

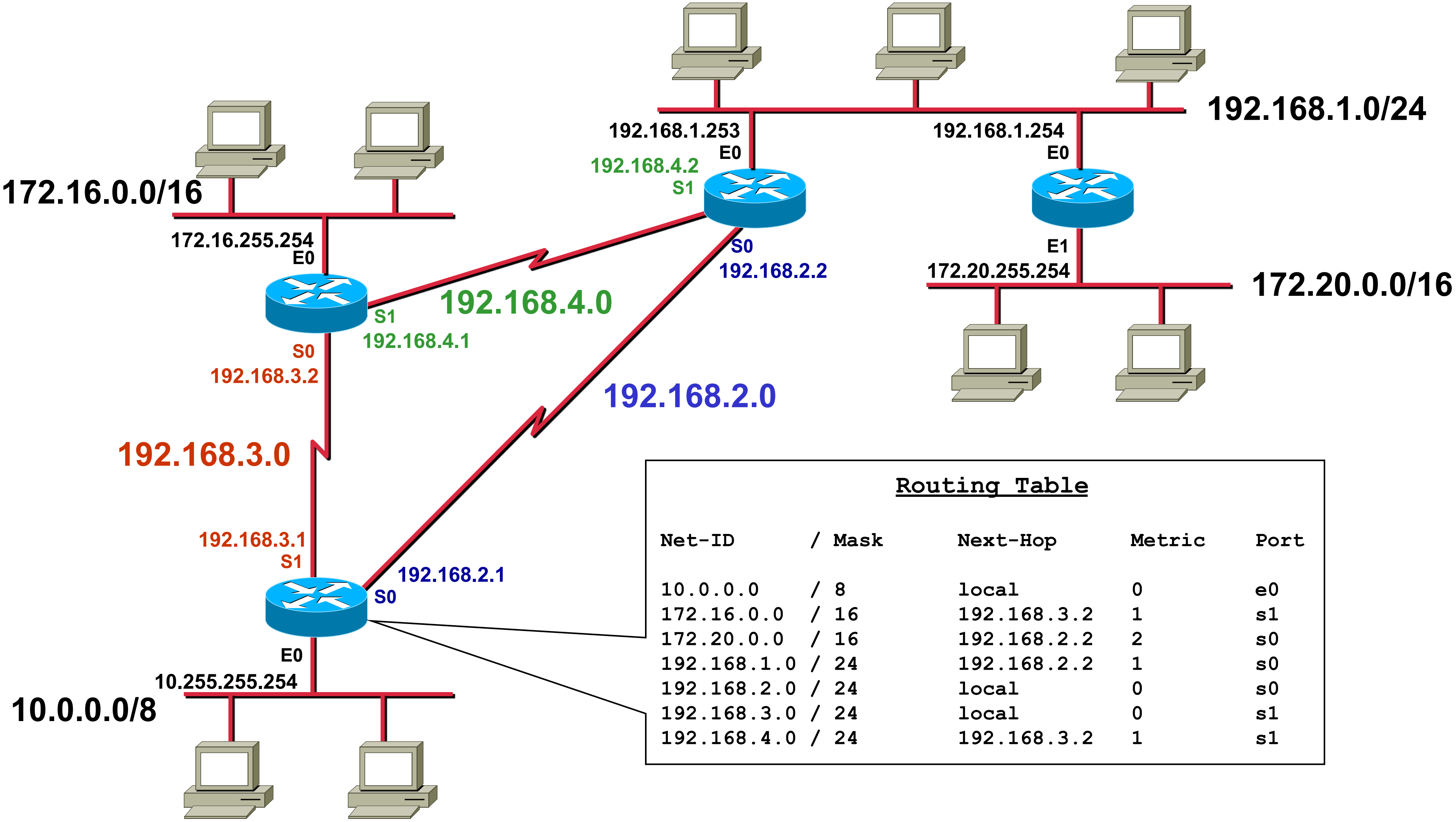
- **Destination Based Routing**
  - Source address is not taken into account for the forward decision
- **Hop by Hop Routing**
  - IP datagrams follow the path (signpost) given by the current state of routing table entries
- **Least Cost Routing**
  - Typically only the best path is considered for forwarding of IP datagrams
  - Alternate paths will not be used in order to reach a given destination
    - Note: Some methods allow load balancing if paths are equal

Static versus Dynamic Routing

- **Static**
  - Routing tables are preconfigured by network administrator
  - Non-responsive to topology changes
  - Can be labor intensive to set up and modify in complex networks
  - No overhead concerning CPU time and traffic

- **Dynamic**
  - Routing tables are dynamically updated with information received from other routers
  - Responsive to topology changes
  - Low maintenance labor cost
  - Communication between routers is done by routing protocols using routing messages for their communication
  - Routing messages need a certain percentage of bandwidth
  - Dynamic routing need a certain percentage of CPU time of the router
  - That means overhead

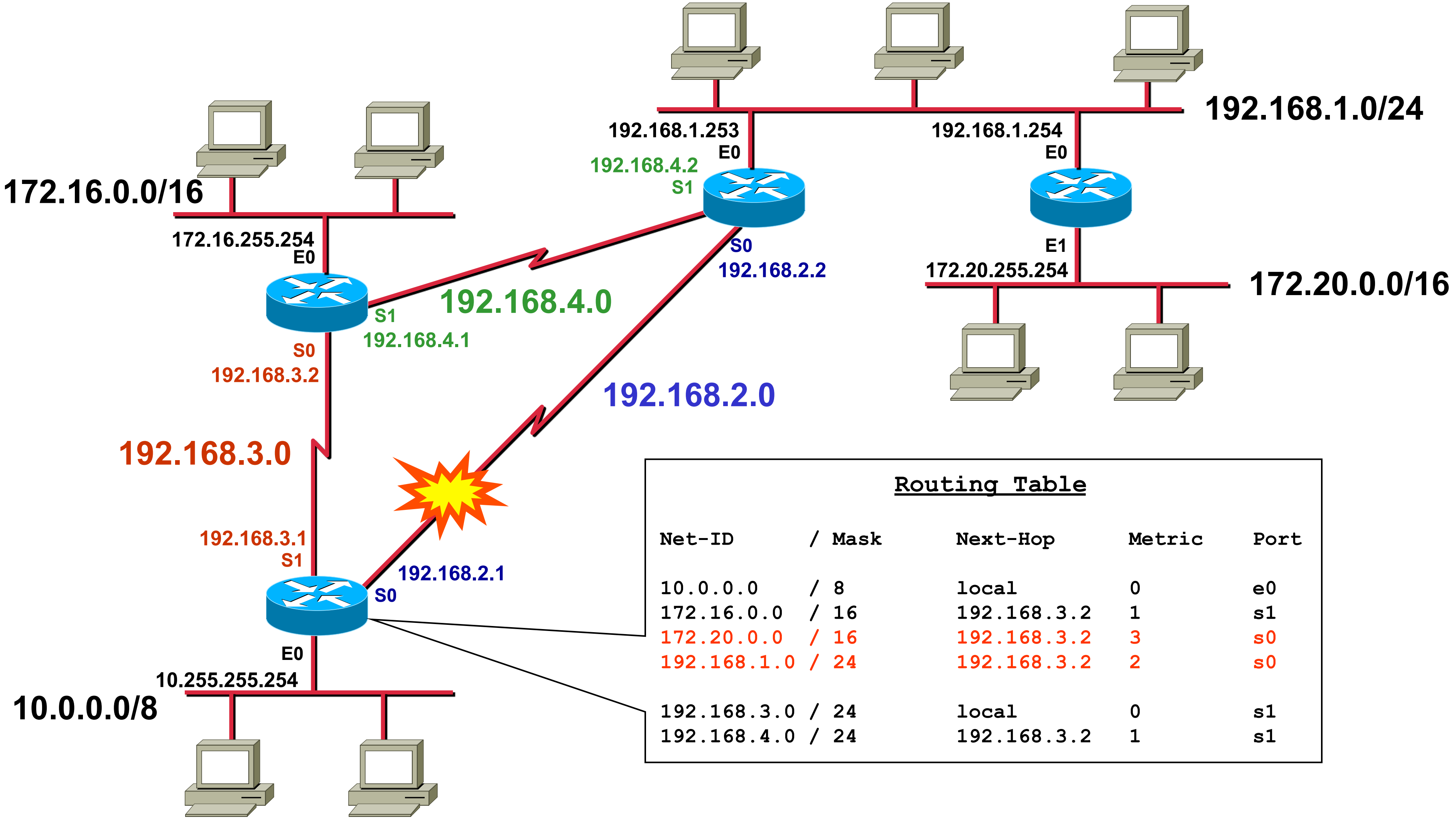
Routing Table - Dynamic Routing (1)



Routing Table

Net-ID	/ Mask	Next-Hop	Metric	Port
10.0.0.0	/ 8	local	0	e0
172.16.0.0	/ 16	192.168.3.2	1	s1
172.20.0.0	/ 16	192.168.2.2	2	s0
192.168.1.0	/ 24	192.168.2.2	1	s0
192.168.2.0	/ 24	local	0	s0
192.168.3.0	/ 24	local	0	s1
192.168.4.0	/ 24	192.168.3.2	1	s1

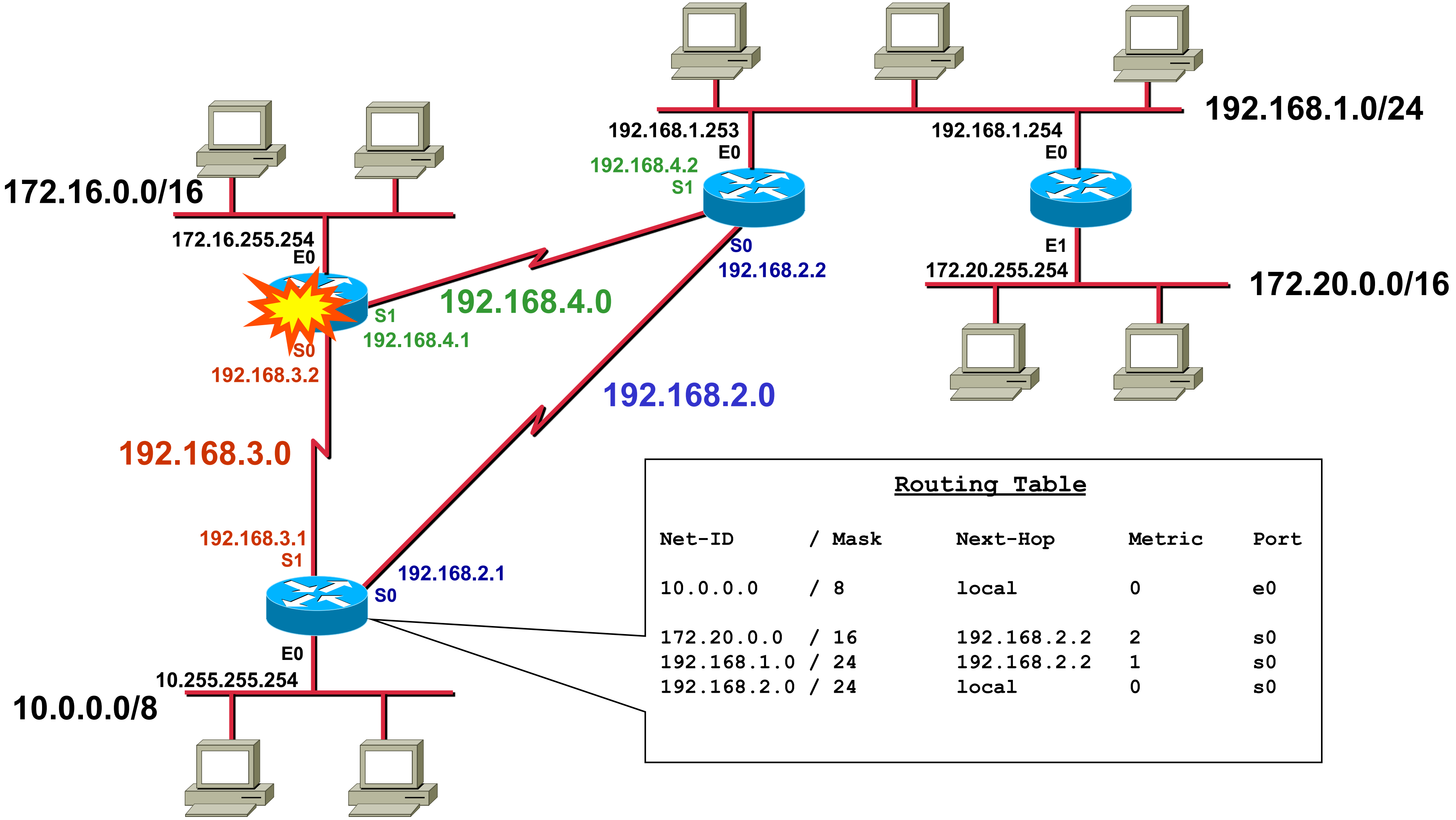
Routing Table - Dynamic Routing (2)



Routing Table

Net-ID / Mask	Next-Hop	Metric	Port
10.0.0.0 / 8	local	0	e0
172.16.0.0 / 16	192.168.3.2	1	s1
172.20.0.0 / 16	192.168.3.2	3	s0
192.168.1.0 / 24	192.168.3.2	2	s0
192.168.3.0 / 24	local	0	s1
192.168.4.0 / 24	192.168.3.2	1	s1

Routing Table - Dynamic Routing (3)



Routing Table

Net-ID	/ Mask	Next-Hop	Metric	Port
10.0.0.0	/ 8	local	0	e0
172.20.0.0	/ 16	192.168.2.2	2	s0
192.168.1.0	/ 24	192.168.2.2	1	s0
192.168.2.0	/ 24	local	0	s0

Dynamic Routing

- **Basic principle**
  - Routing tables are dynamically updated with information from other routers exchanged by routing protocols
  - Routing protocol
    - Discovers current network topology
    - Determines the best path to every reachable network
    - Stores information about best paths in the routing table
  - Metric information is necessary for best path decision
    - In most cases summarization of <u>static</u> preconfigured values along the given path
      - Hops, interface cost, interface bandwidth, interface delay, etc.
  - Two basic technologies
    - Distance vector, Link state

Routing Metric

- **Routing protocols typically find out more than one route to the destination**
- **Metric help to decide which path to use**
  - – Static values
    - Hop count, distance (RIP)
    - Cost like reciprocal value of bandwidth (OSPF)
    - Bandwidth (EIGRP), Delay (EIGRP), MTU
  - – Variable or dynamic values
    - Load (EIGRP)
    - Reliability (EIGRP)
    - Very seldom used
      - – Cisco citation:

        "If you do not know what you are doing do not even think using or touching them!"

Dynamic Routing

- **Each router can run one or more routing protocols**
- **Routing protocols**
  - Are information sources to create routing table
  - Announce network reachability information
    - By doing this a router declares that traffic destined to a certain network can be sent to him
    - Network reachability information flows in the opposite direction to the traffic destined to a network
- **Routing protocols differ in**
  - Convergence time, loop avoidance, maximum network size, reliability and complexity

Routing Protocol Comparison

Routing Protocol	Complexity	Max. Size	Convergence Time	Reliability	Protocol Traffic
RIP	very simple	16 Hops	High (minutes)	Not absolutely loop-safe	High
RIPv2	very simple	16 Hops	High (minutes)	Not absolutely loop-safe	High
IGRP	simple	x	High (minutes)	Medium	High
EIGRP	complex	x	Fast (seconds)	High	Medium
OSPF	very complex	Thousands of Routers	Fast (seconds)	High	Low
IS-IS	complex	Thousands of Routers	Fast (seconds)	High	Low
BGP-4	very complex	more than 100,000 networks	Middle	Very High	Low

Distance Vector Protocols (1)

- **After powering-up each router only knows about directly attached networks**
- **Routing table is sent periodically to all neighbor-routers**
- **Received updates are examined, changes are adopted in own routing table**
  - Changes announced by next periodic routing update
- **Metric information is based on hops (distance between hops)**
  - Hop count metric is a special case for the more generic distance value between two routers
  - Hop count means distance = 1 between any two neighboring routers
- **"Bellman-Ford" algorithm**

Distance Vector Protocols (2)

- **Limited view of topology**
  - Next hop is always originating router
  - Topology behind next hop unknown
  - Signpost principle
- **Loops can occur!**
- **Additional mechanisms needed**
  - Maximum hop count
  - Split horizon (with poison reverse)
  - Triggered update
  - Hold down
  - Route Poisoning

Distance Vector Protocols (3)

- **Examples**
  - RIP, RIPv2 (Routing Information Protocol)
  - IGRP (Cisco, Interior Gateway Routing Protocol)
  - IPX RIP (Novell)
  - AppleTalk RTMP (Routing Table Maintenance Protocol)

Link State Protocols (1)

- **Each two neighbored routers establish adjacency**
- **Routers learn real topology information**
  - Through "Link State Advertisements (LSAs)"
  - Stored in database (Roadmap principle)
- **Routers have a global view of network topology**
  - Exact knowledge about all routers, links and their costs (metric) of a network
- **Updates only upon topology changes**
  - Propagated by *flooding* of LSAs
    (very fast convergence)

Link State Protocols (2)

- **Routing table entries are calculated by applying the Shortest Path First (SPF) algorithm on the database**
  - Loop-safe
  - Only the lowest cost path is stored in routing table
  - But alternative paths are immediately known
  - Could be CPU and memory greedy
    - Mainly a concern in the past
- **Large networks can be split into areas**

Link State Protocols (3)

- **With the lack of topology changes**
  - Local hello messages are used to supervise local links (to test reachability of immediate-neighboring routers)
  - Therefore less routing overhead concerning link bandwidth than periodic updates of distance vector protocols
- **But more network load is caused by such a routing protocol**
  - During connection of former separated parts of a network
  - During topology database synchronization

Link State Protocols (4)

- **Examples**
  - OSPF (Open Shortest Path First)
  - Integrated IS-IS (IP world)
    - note: Integrated IS-IS takes another approach to handle large networks (topic outside the scope of this course)
  - IS-IS (OSI world)
  - PNNI (in the ATM world)
  - APPN (IBM world),
  - NLSP (Novell world)

Agenda

- L2 versus L3 Switching
- IP Protocol, IP Addressing
- IP Forwarding
- ARP and ICMP
- IP Routing
  - Introduction
  - OSPF Basics
  - OSPF Communication Procedures (Router LSA)
  - LSA Broadcast Handling (Flooding)
  - OSPF Splitted Area
  - Broadcast Networks (Network LSA)
- First Hop Redundancy

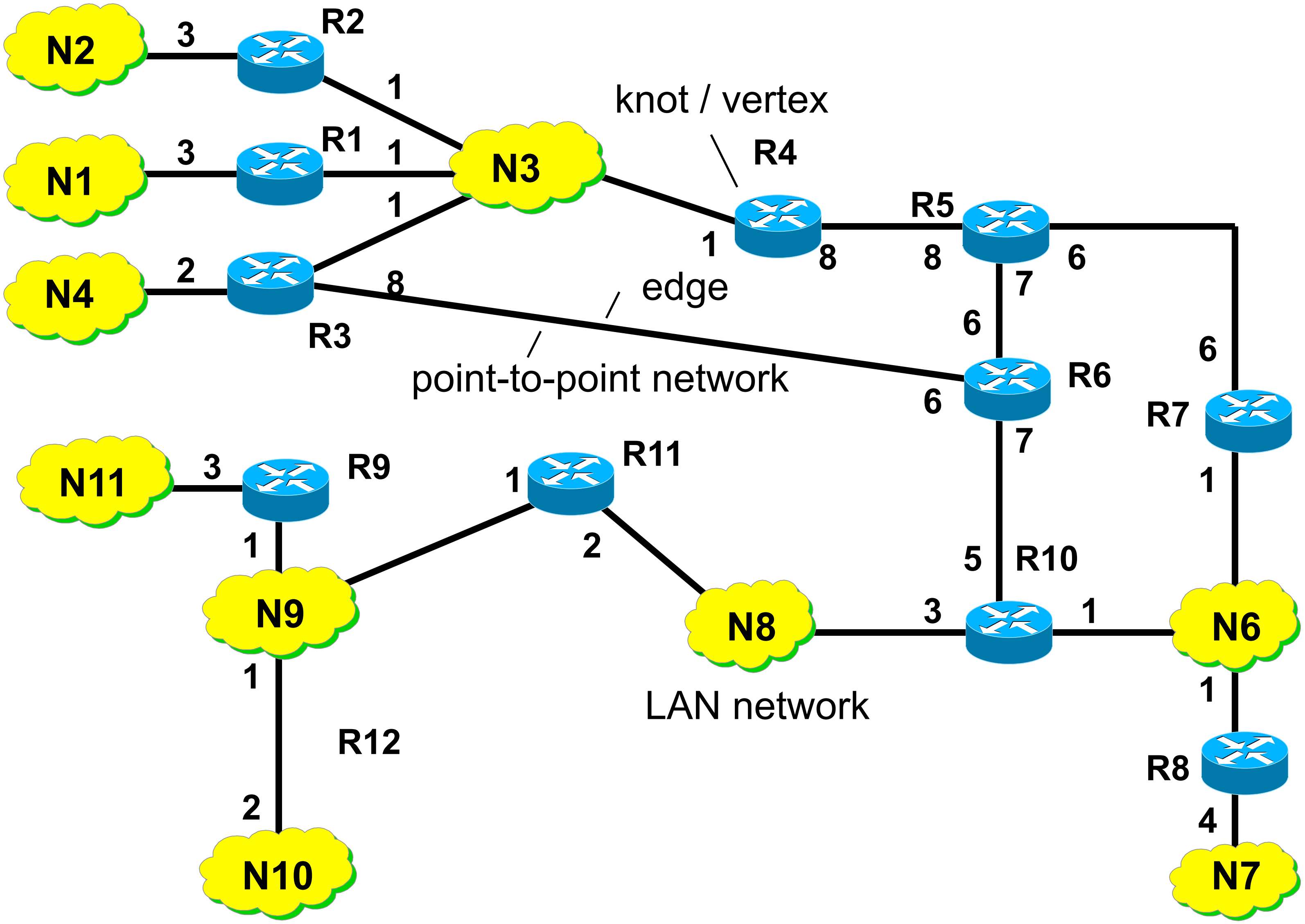
“OSPF (Open Shortest Path First)

- **OSPF is a link-state routing protocol**
  - Inherently fast convergence
  - Designed for large networks
  - Designed to be reliable
- **Basic ideas:**
  - Every router knows topology of the whole network including subnets and routers
    - “Roadmap”
  - Topology (roadmap) stored in router’s OSPF database
  - Shortest Path First (SPF) algorithm applied to find the best path
    - Invented by E. W. Dijkstra
    - Creates a (loop-free) tree with local router as source
    - Is used to find the best path by calculating very efficiently all paths to all destinations at once; best path is entered into the routing table
  - Changes are flooded over the network to update the OSPF database
    - Like traffic announcements used by car navigation systems
    - LSA (Link State Advertisements)

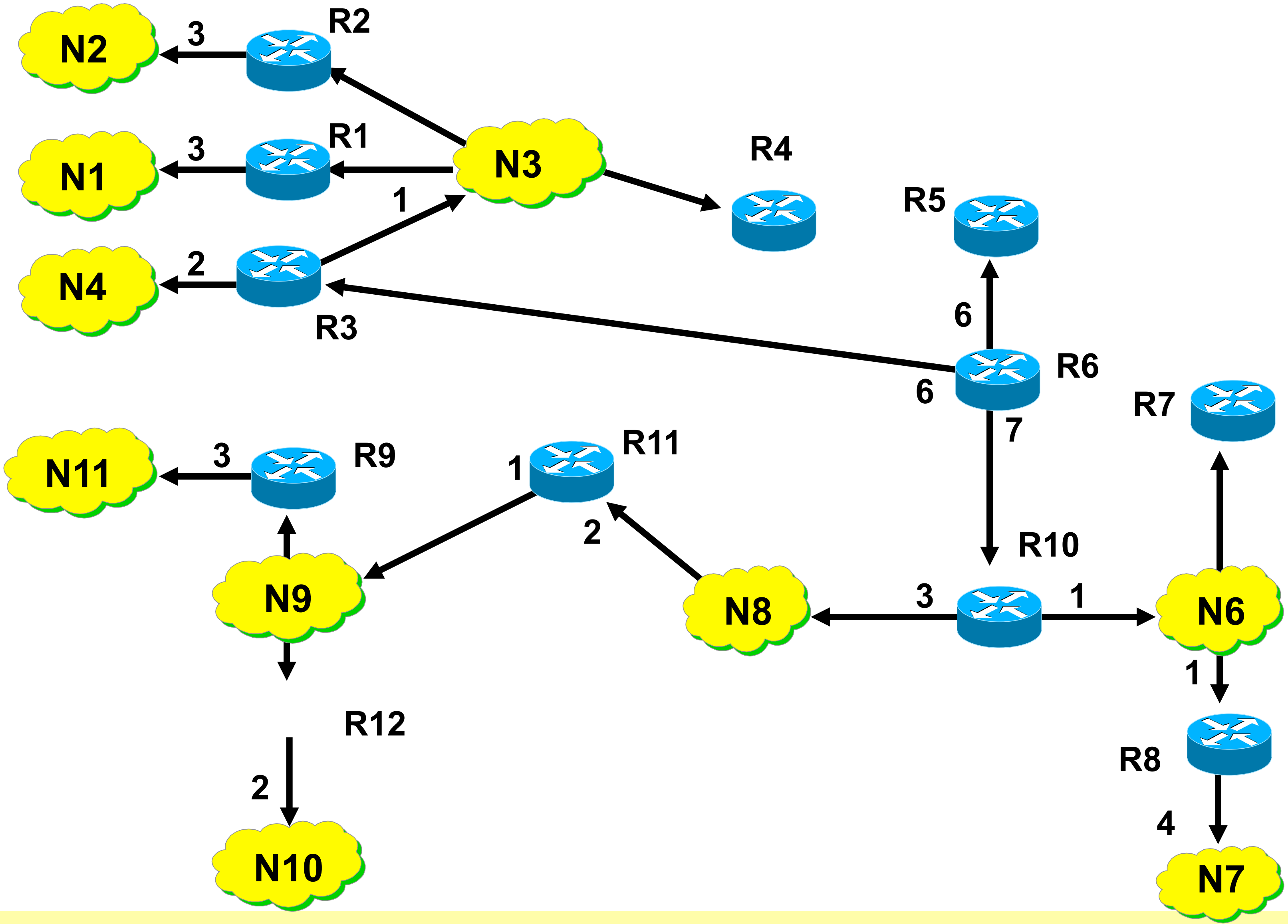
OSPF Topology Database

- **Every router maintains a topology database**
  - Like a "network roadmap"
  - Describes the whole network !!
    - Note: RIP provides only "signposts"
- **Database is based on a graph**
  - Where each knot (vertex) stands for a router
  - Where each edge stands for a subnet
    - Connecting the routers
    - Path-costs are assigned to the edges
- **Router uses the graph**
  - To calculate shortest paths to all subnets
    - Router itself is the root of the shortest path

OSPF Domain


N2
3
R2
1
N1
3
R1
1
N3
1
N4
2
R3
8
knot / vertex
R4
1
8
R5
8
6
7
6
edge
point-to-point network
6
R6
6
R7
7
N11
3
R9
1
R11
2
1
N9
1
R12
2
N10
N8
3
5
R10
1
N6
1
LAN network
1
R8
4
N7

Shortest Paths regarding Router R6

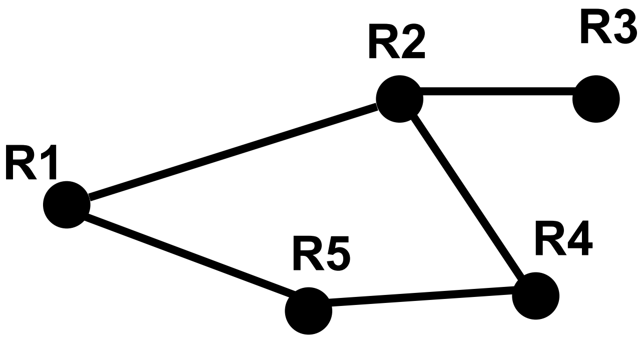


Routing Table Router 6

NET-ID	NEXT HOP	DISTANCE
N1	R3	10
N2	R3	10
N3	R3	7
N4	R3	8
N6	R10	8
N7	R10	12
N8	R10	10
N9	R10	11
N10	R10	13
N11	R10	14

What is Topology Information?

- The smallest topological unit is simply the information element ROUTER-LINK-ROUTER
- So the question is: Which router is linked to which other routers?
- Link-state



=

Link Database:

R1– R2
R1– R5
R2– R3
R2– R4
R4– R5

The Link Database exactly describes the roadmap

Agenda

- L2 versus L3 Switching
- IP Protocol, IP Addressing
- IP Forwarding
- ARP and ICMP
- IP Routing
  - – Introduction
  - – OSPF Basics
  - – OSPF Communication Procedures (Router LSA)
  - – LSA Broadcast Handling (Flooding)
  - – OSPF Splitted Area
  - – Broadcast Networks (Network LSA)
- First Hop Redundancy

Creating the Database

- **The basic means for creating and maintaining the database are the so-called**

  Link States

- **A link state stands for an intact (synchronized) local neighbourhood between two routers**
  - The link state is created by these two routers
  - Other routers are notified about this link state via a special broadcast-mechanism ("traffic-news")
    - Flooding together with sequence numbers stored in topology database
  - Link states are verified continuously

How are Link States used?

- **Adjacent routers declare themselves as neighbours by setting the link state up (or down otherwise)**
  - The link-state can be checked with hello messages
    - Note: Link state down is no explicitly expressed, it is just the absence of the link to the former neighbour in the LSA announcement
- **Every link state change is published to all routers of the OSPF domain using Link State Advertisements (LSAs)**
  - Is a broadcast mechanism
  - Whole topology map relies on correct generation and delivery of LSAs
    - Synchronization of a distributed database !!!

OSPF Communication Principle 1

- **OSPF messages are transported by IP**
  - ip protocol number 89
- **During initialization a router sends hello-messages to all directly reachable routers**
  - To determine its neighbourhood
  - Can be done automatically in broadcast networks and point-to-point connections by using the IP multicast-address 224.0.0.5 (all OSPF routers)
- **This router also receives hello-messages from other routers**

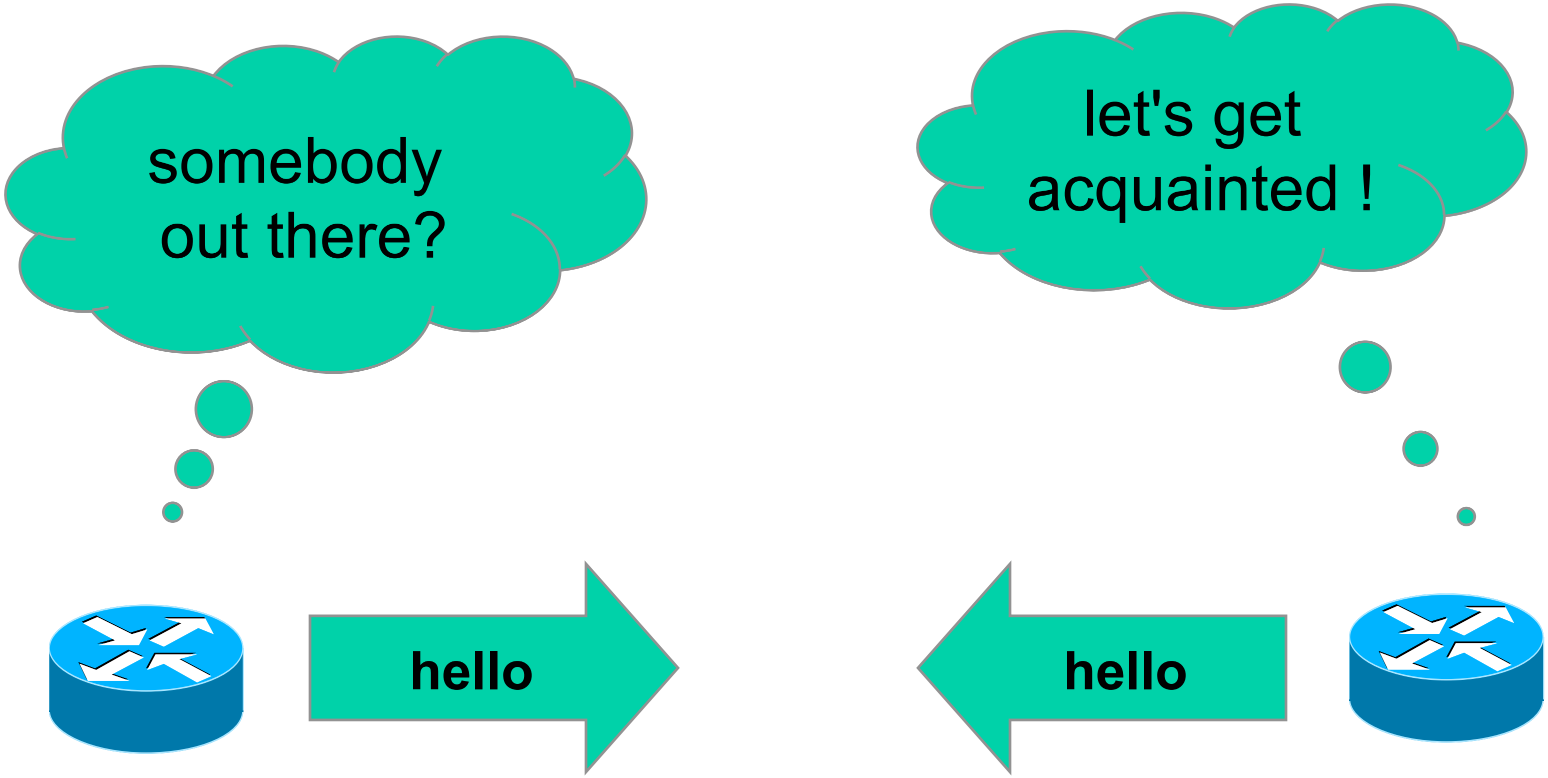
OSPF Communication Principle 2

- **Each two acquainted routers send database description messages to each other, in order to publish their topology database**
- **Unknown or old entries are updated via link state request and link state update messages**
  - Which synchronizes the topology databases
- **After successful synchronization both routers declare their neighbourhood (adjacency) via router LSAs (using link state update messages)**
  - Distributed across the whole network

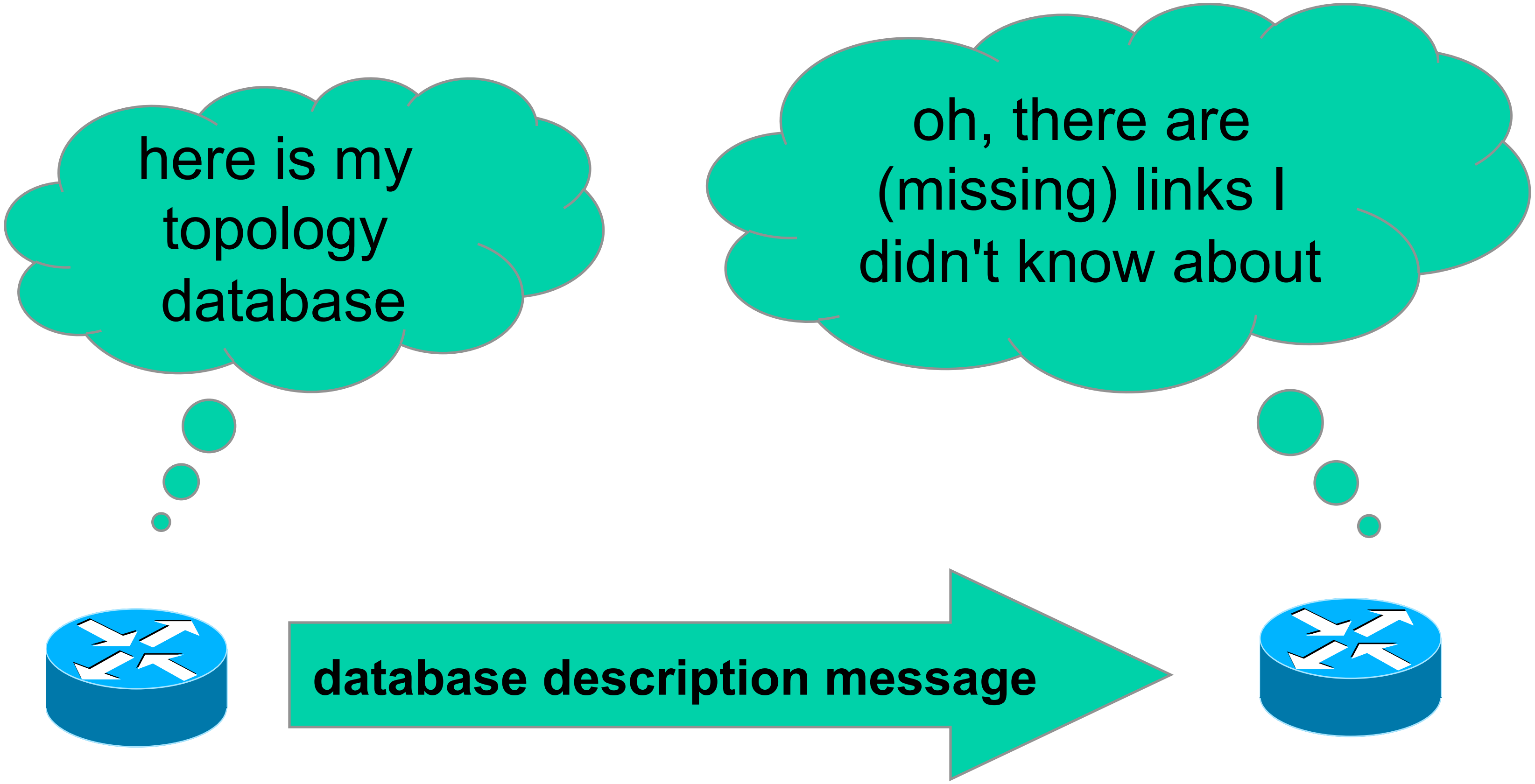
OSPF Communication Principle 3

- **Periodically, every router verifies its link state to its adjacent neighbours using hello messages**
- **From now only changes of link states are distributed**
  - Using link state update messages (LSA broadcast-mechanism)
- **If neighbourhood situation remains unchanged, the periodic hello messages represents the only routing overhead**
  - Note: additionally all Link States are refreshed every 30 minutes with LSA broadcast mechanism

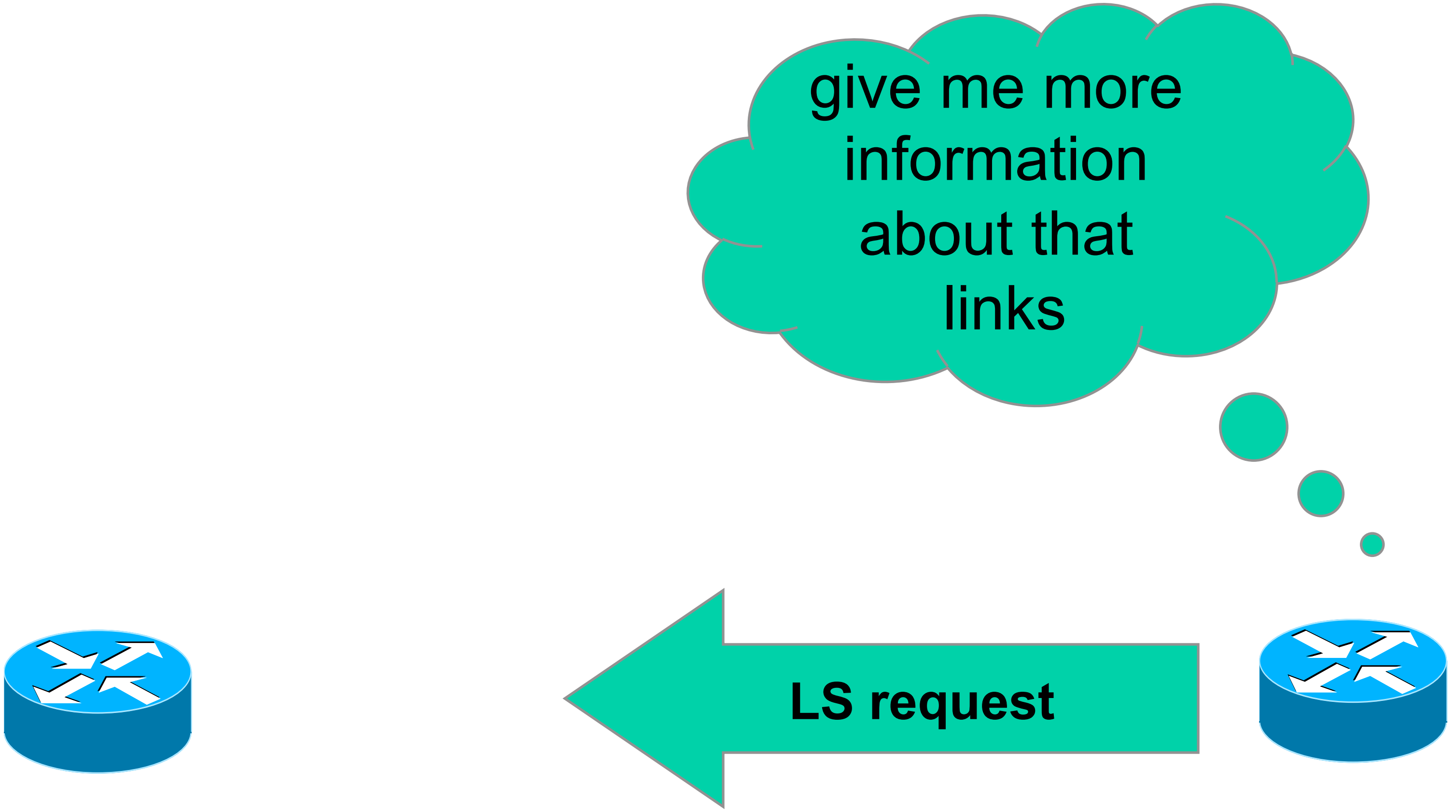
OSPF Communications Summary 1



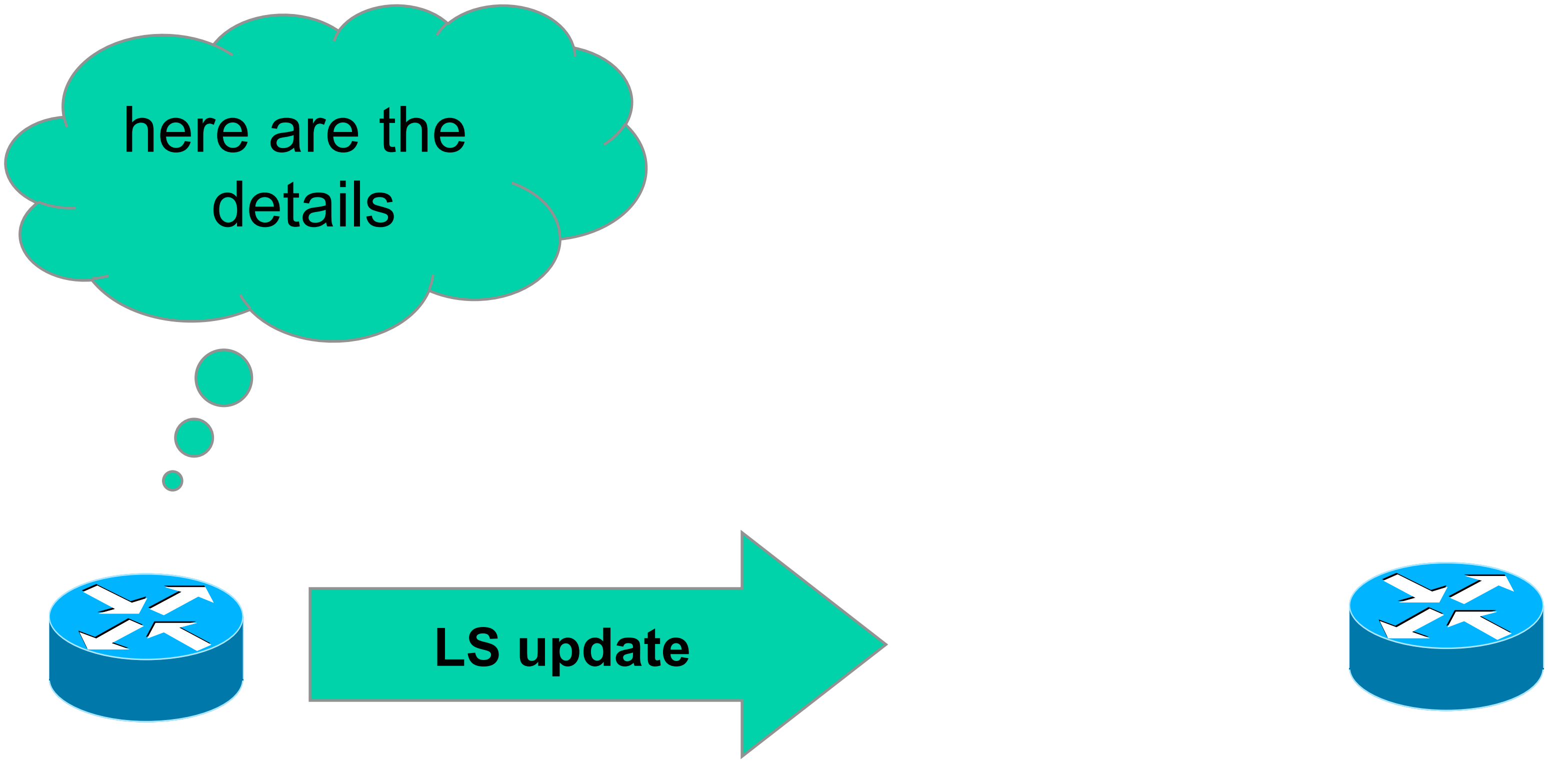
OSPF Communications Summary 2


here is my topology database
oh, there are (missing) links I didn't know about
database description message

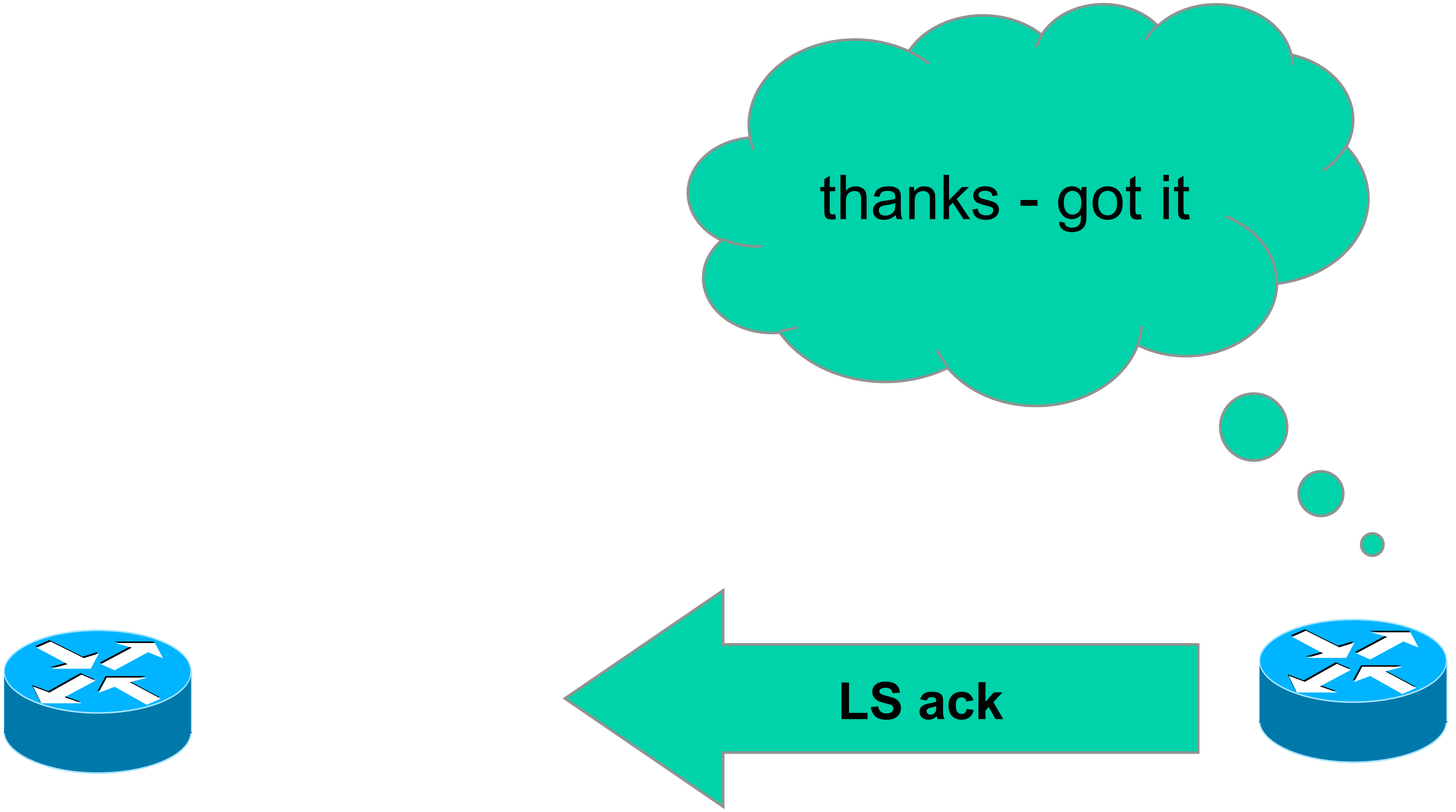
OSPF Communications Summary 3



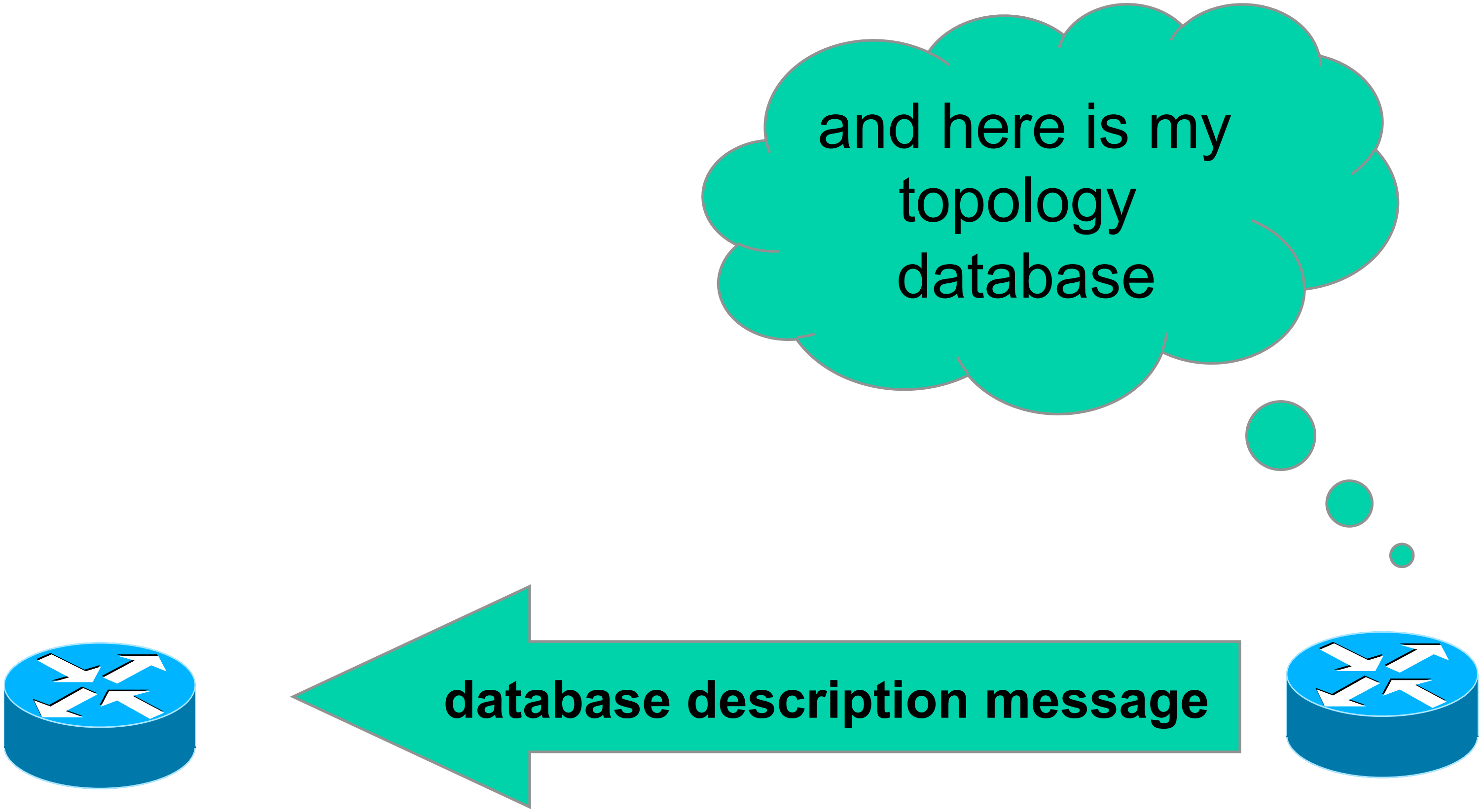
OSPF Communications Summary 4



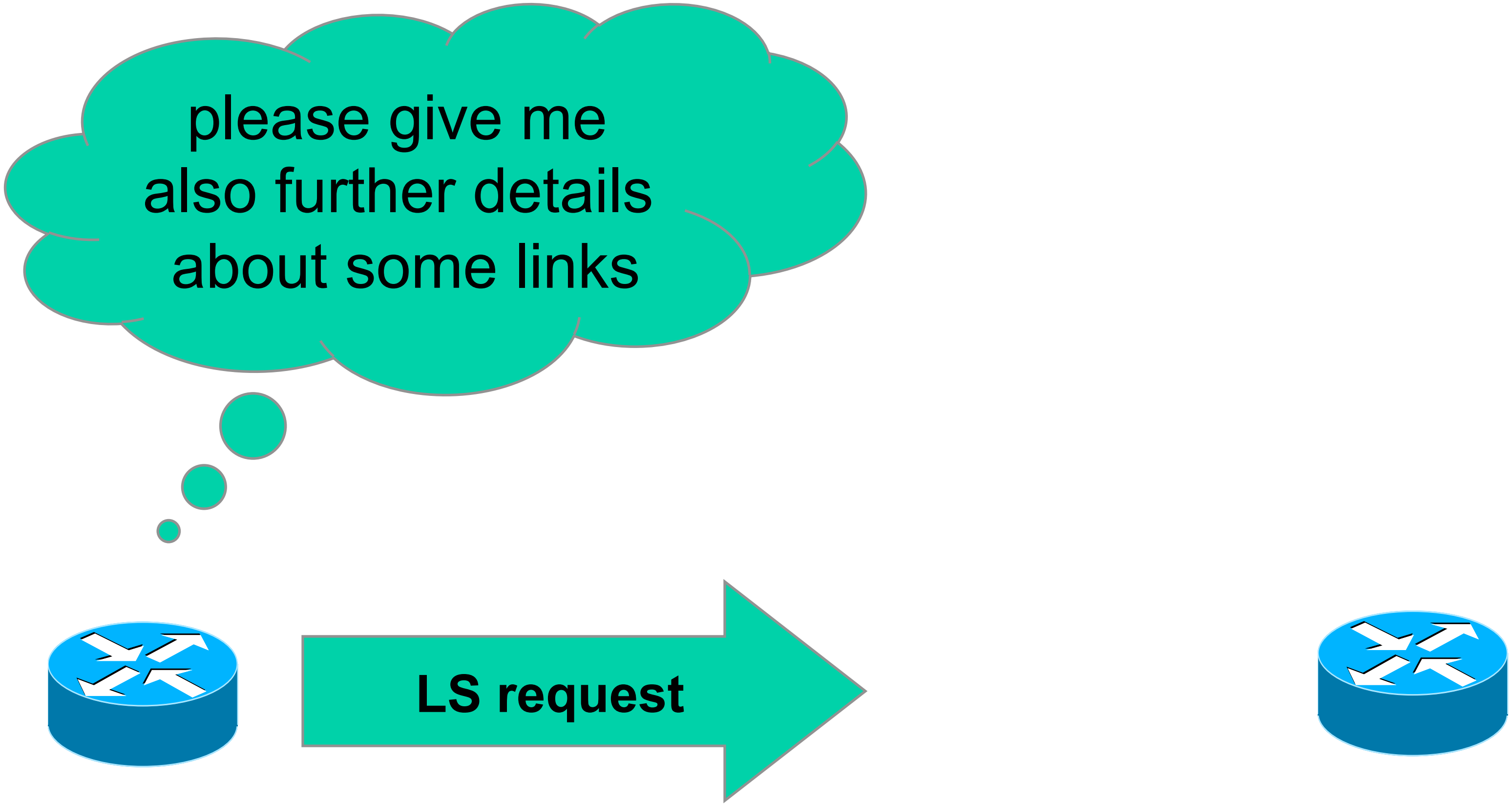
OSPF Communications Summary 5



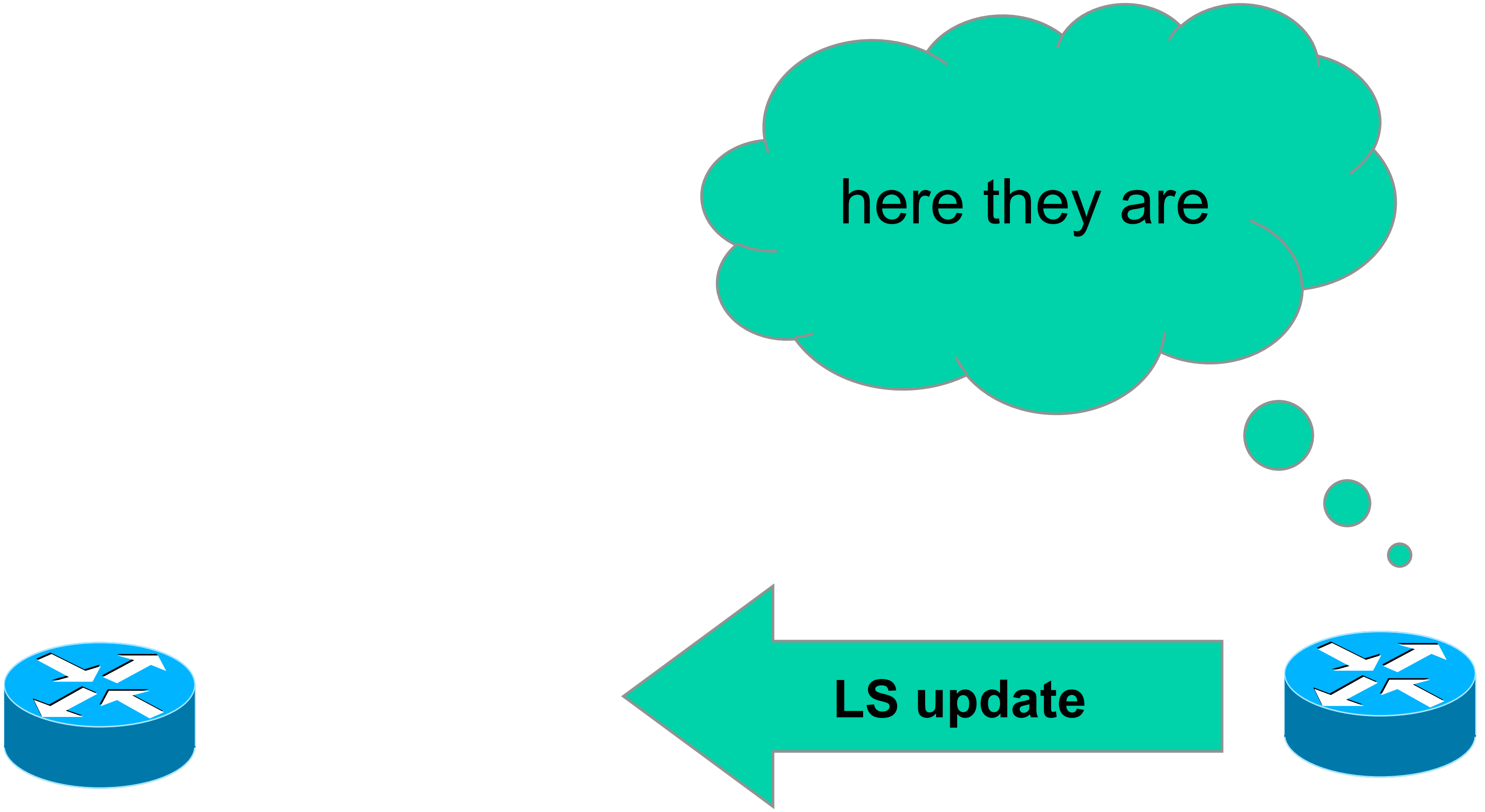
OSPF Communications Summary 6


and here is my topology database
database description message

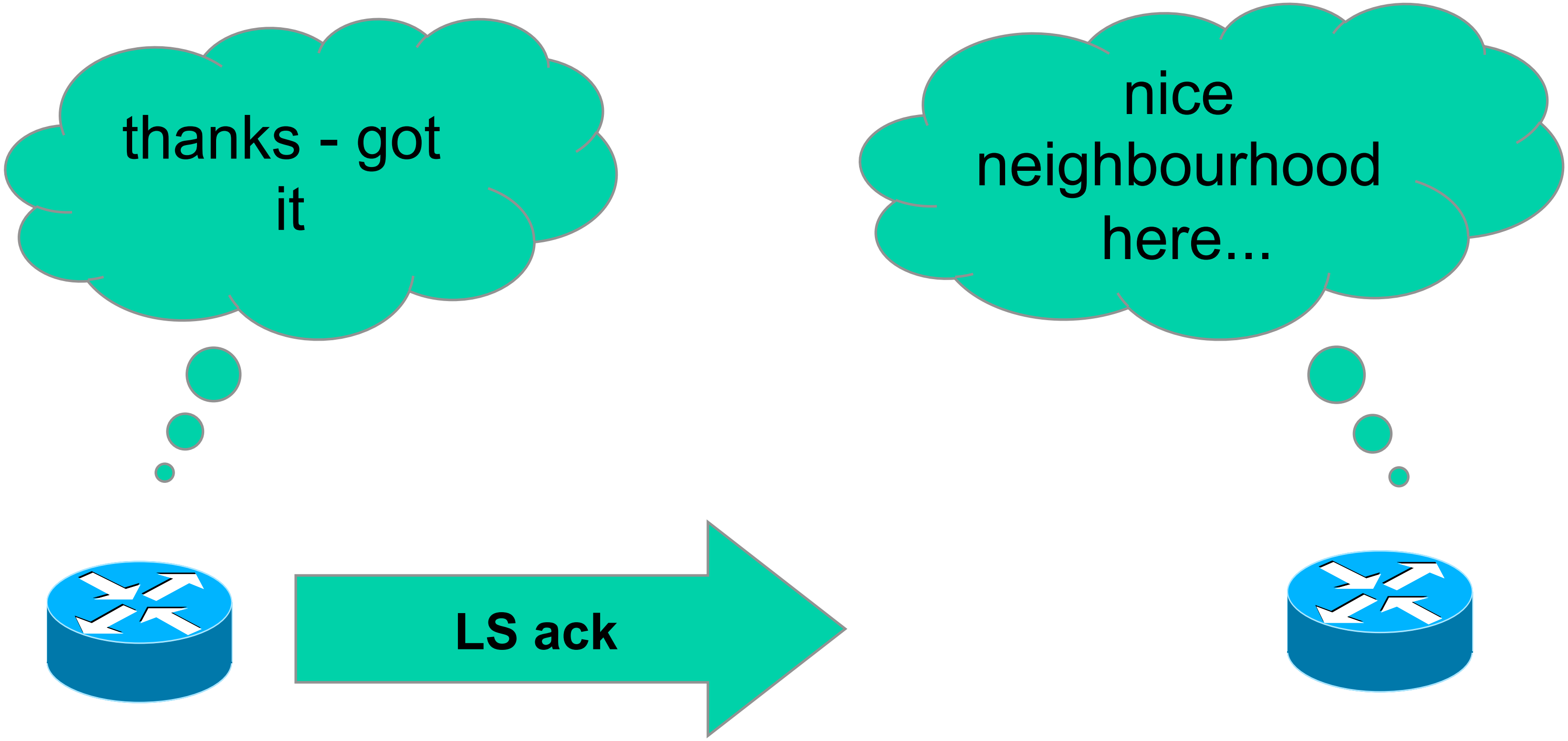
OSPF Communications Summary 7



OSPF Communications Summary 8



OSPF Communications Summary 9

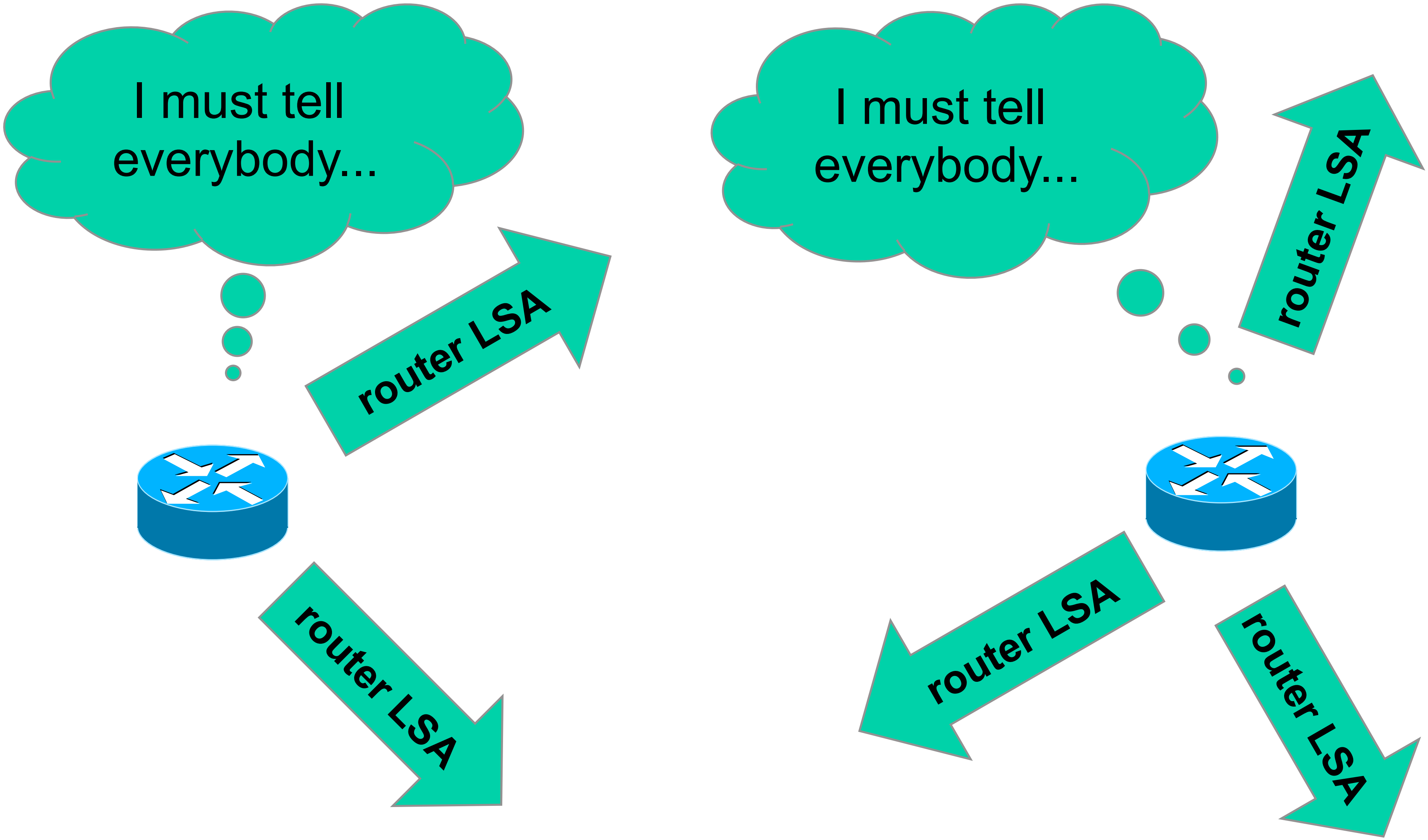


thanks - got it

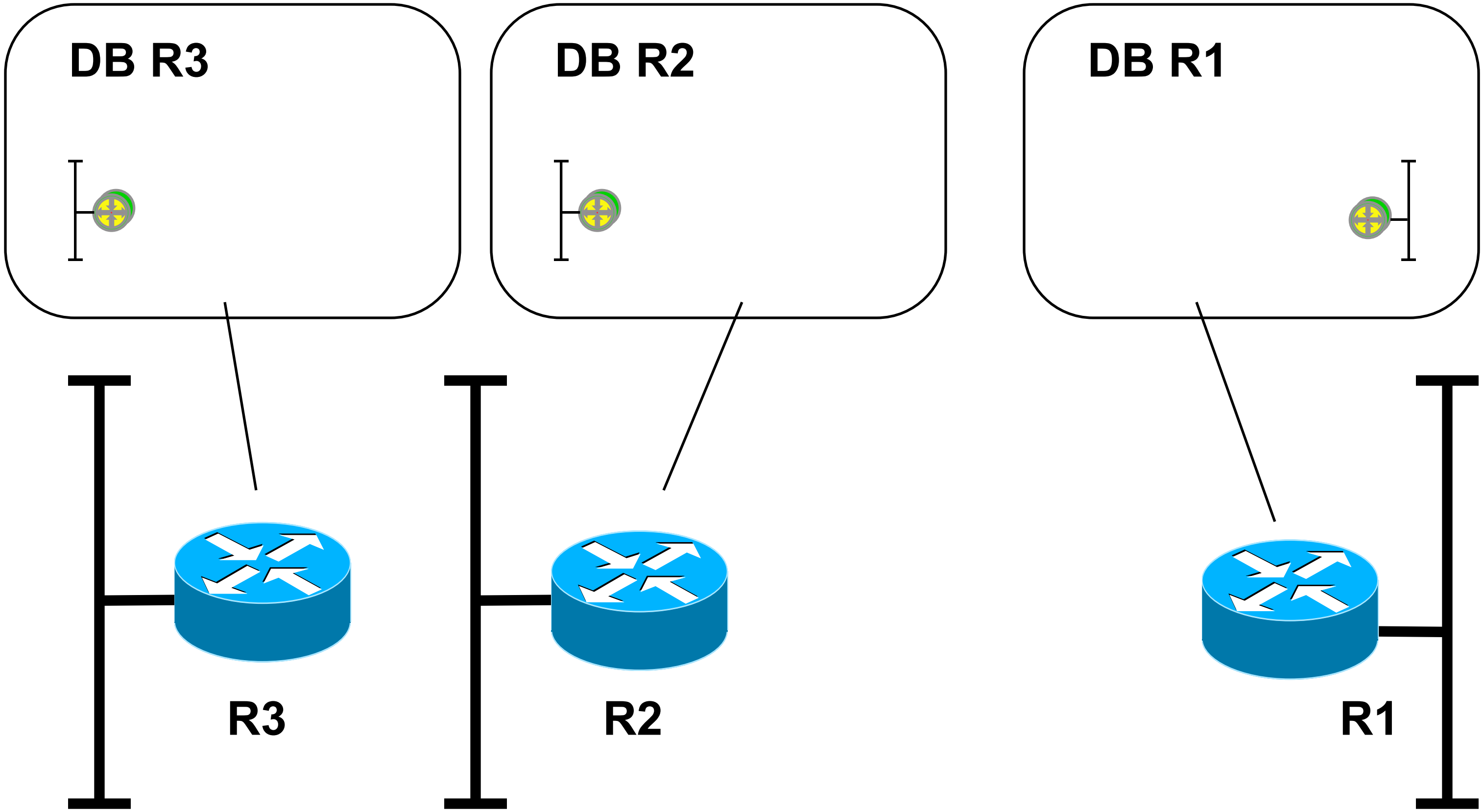
nice neighbourhood here...

LS ack

OSPF Communications Summary 10

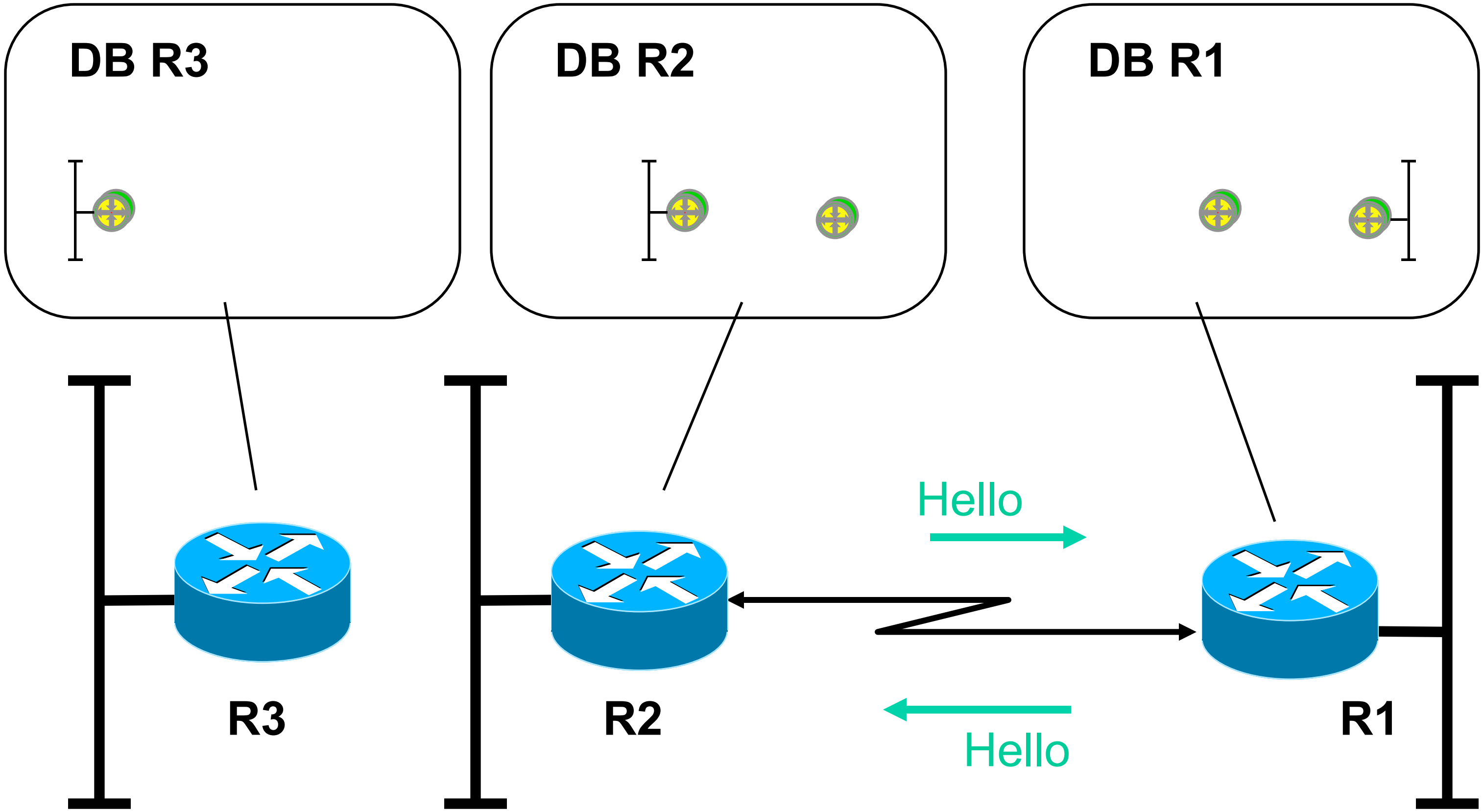


OSPF Start-up



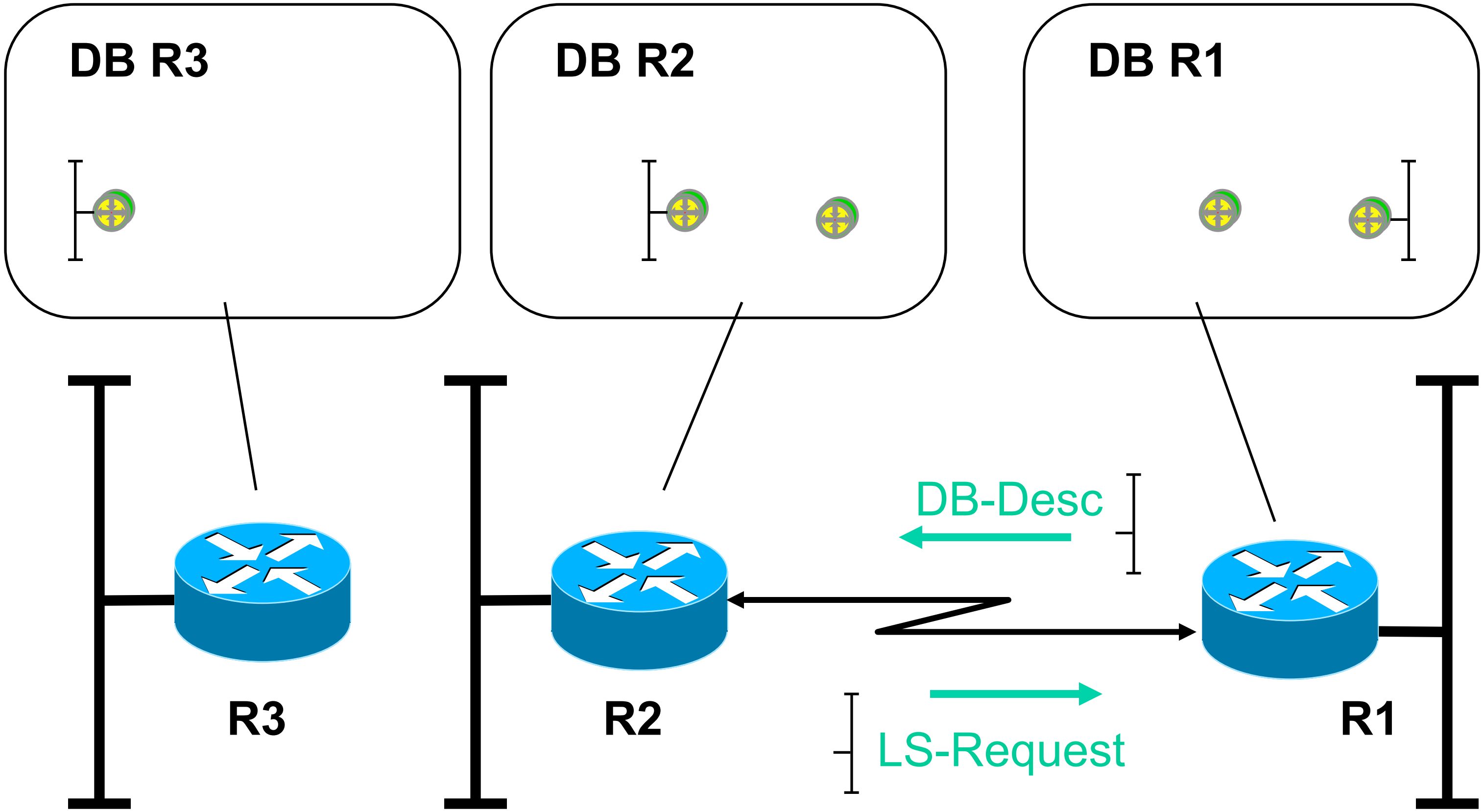
**starting position: all routers initialized,
no connection between R1-R2 or R2-R3**

OSPF Hello R1 - R2



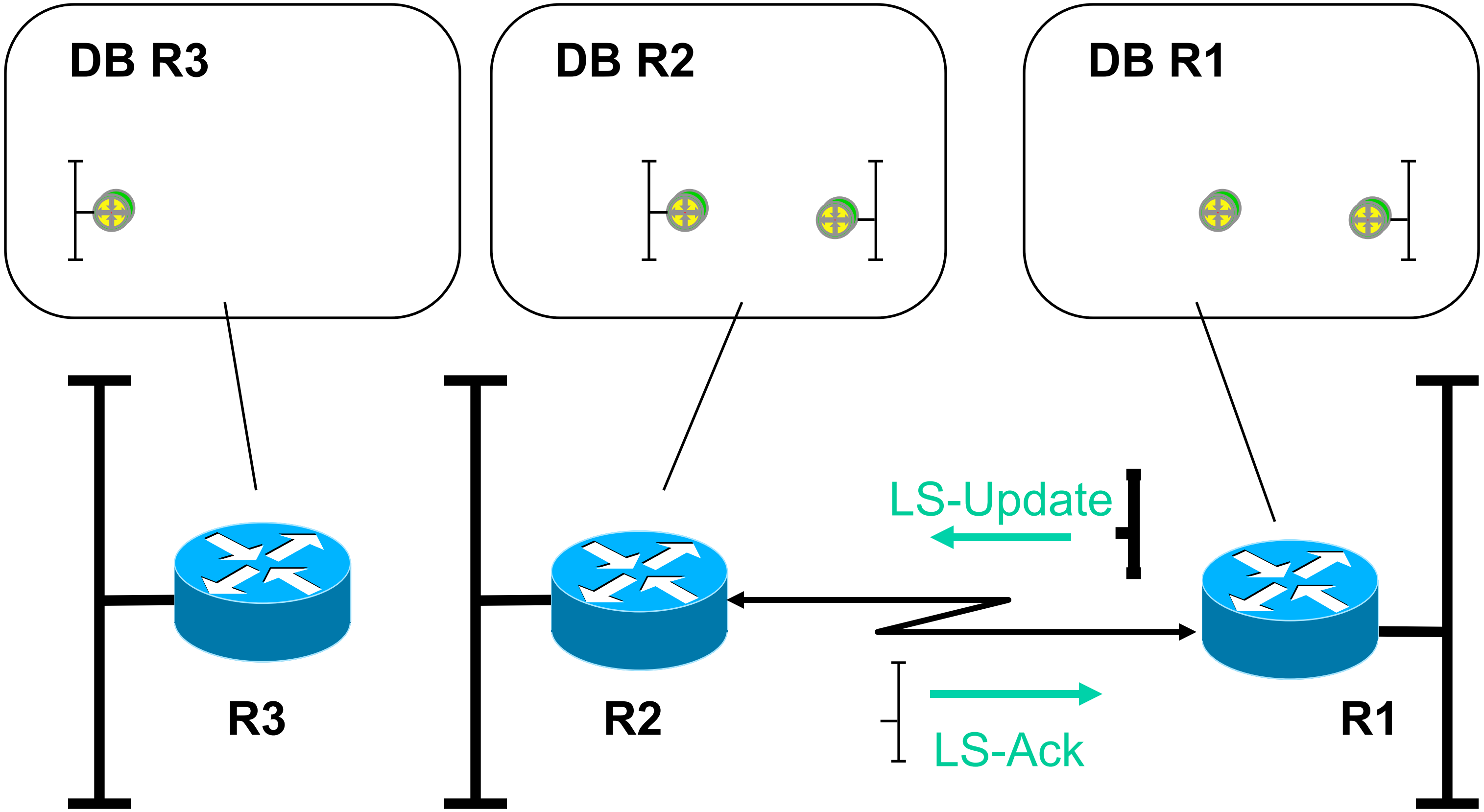
link between R1-R2 activated: get acquainted using hello messages

OSPF Data Base Description R1 -> R2



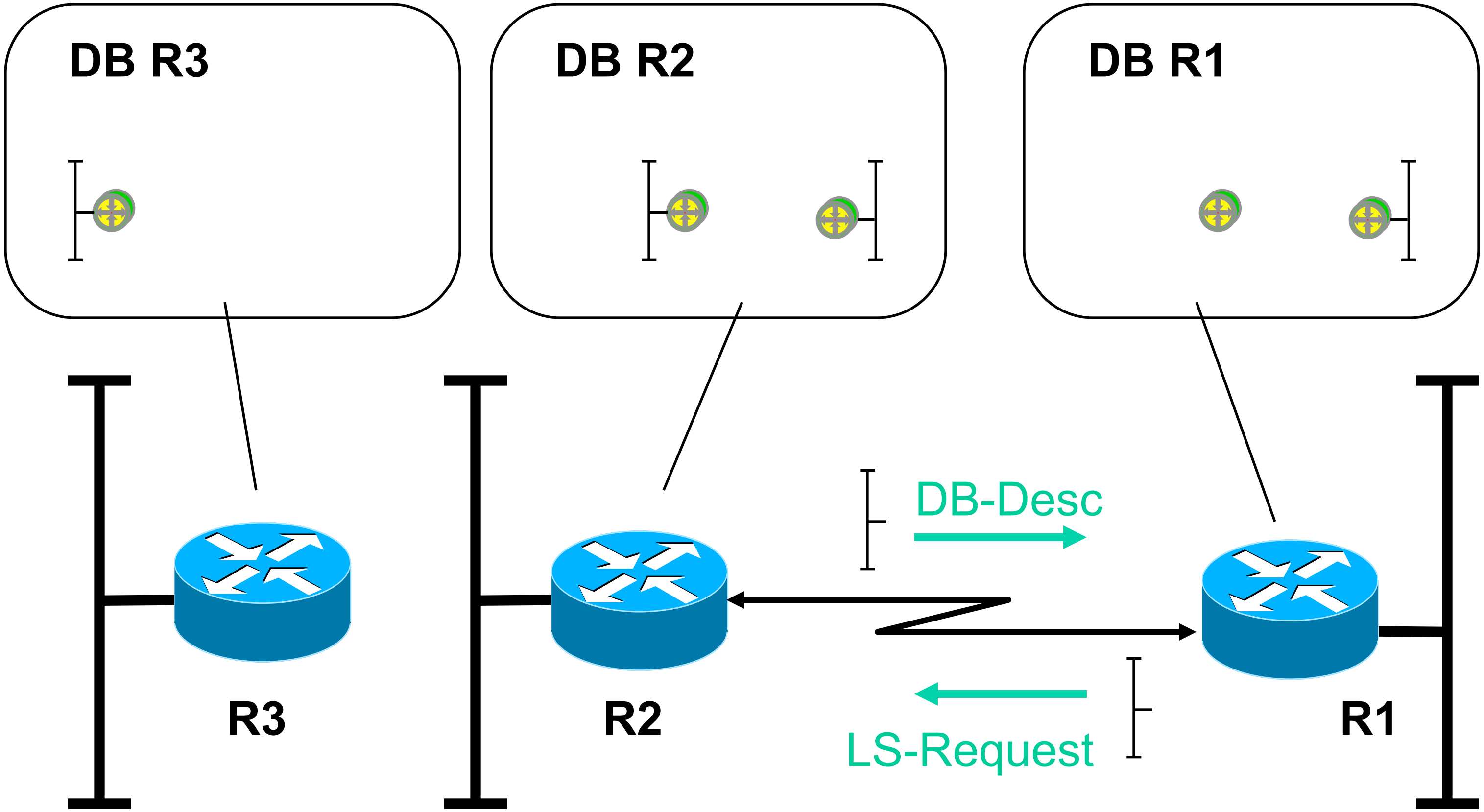
database synchronization: R1 master sends Database-Description, R2 slave sends Link-State Request

OSPF Data Base Update R1 -> R2



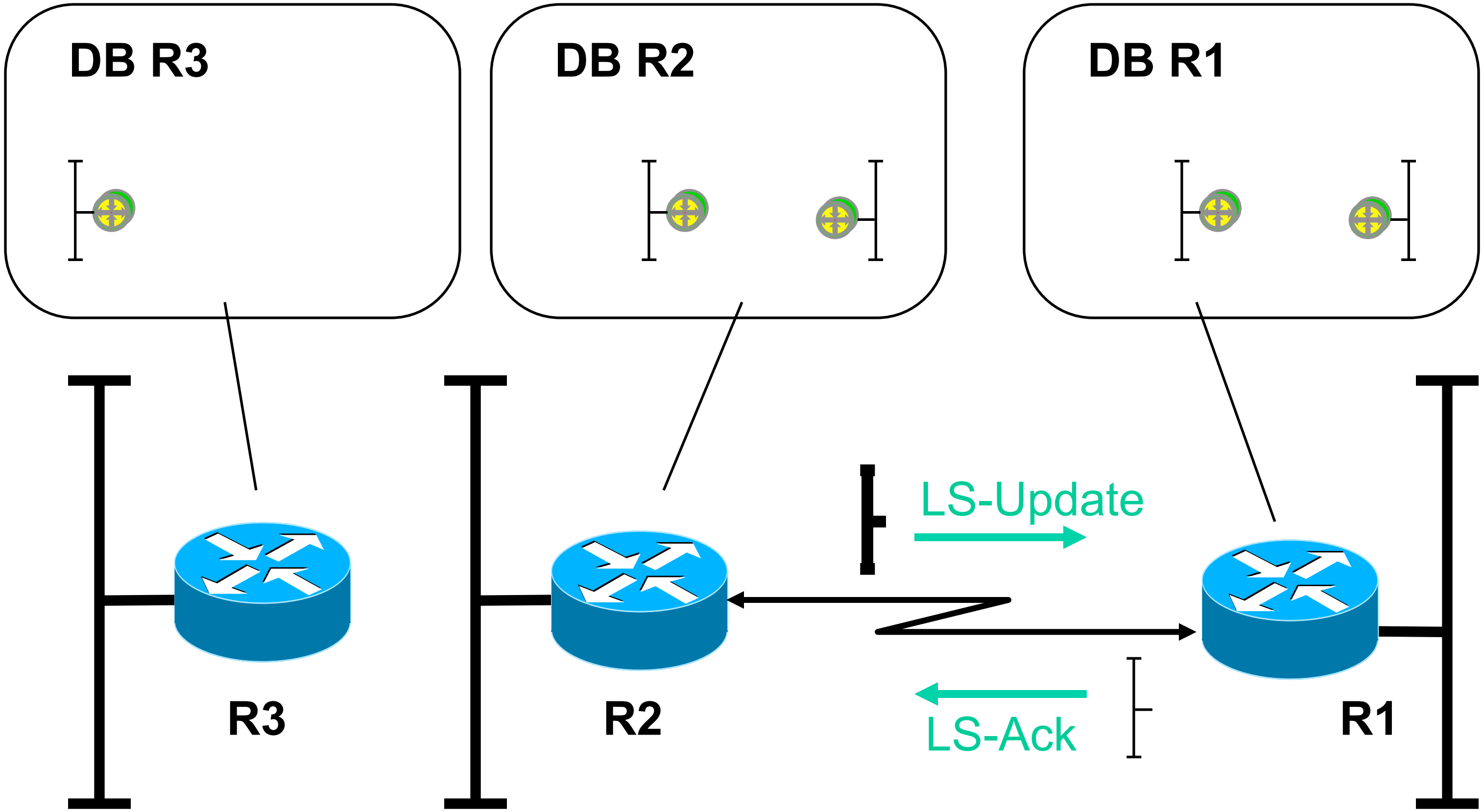
database synchronization: R1 master sends Link-State Update, R2 slave sends Link-State Acknowledgement

OSPF Data Base Description R2 -> R1



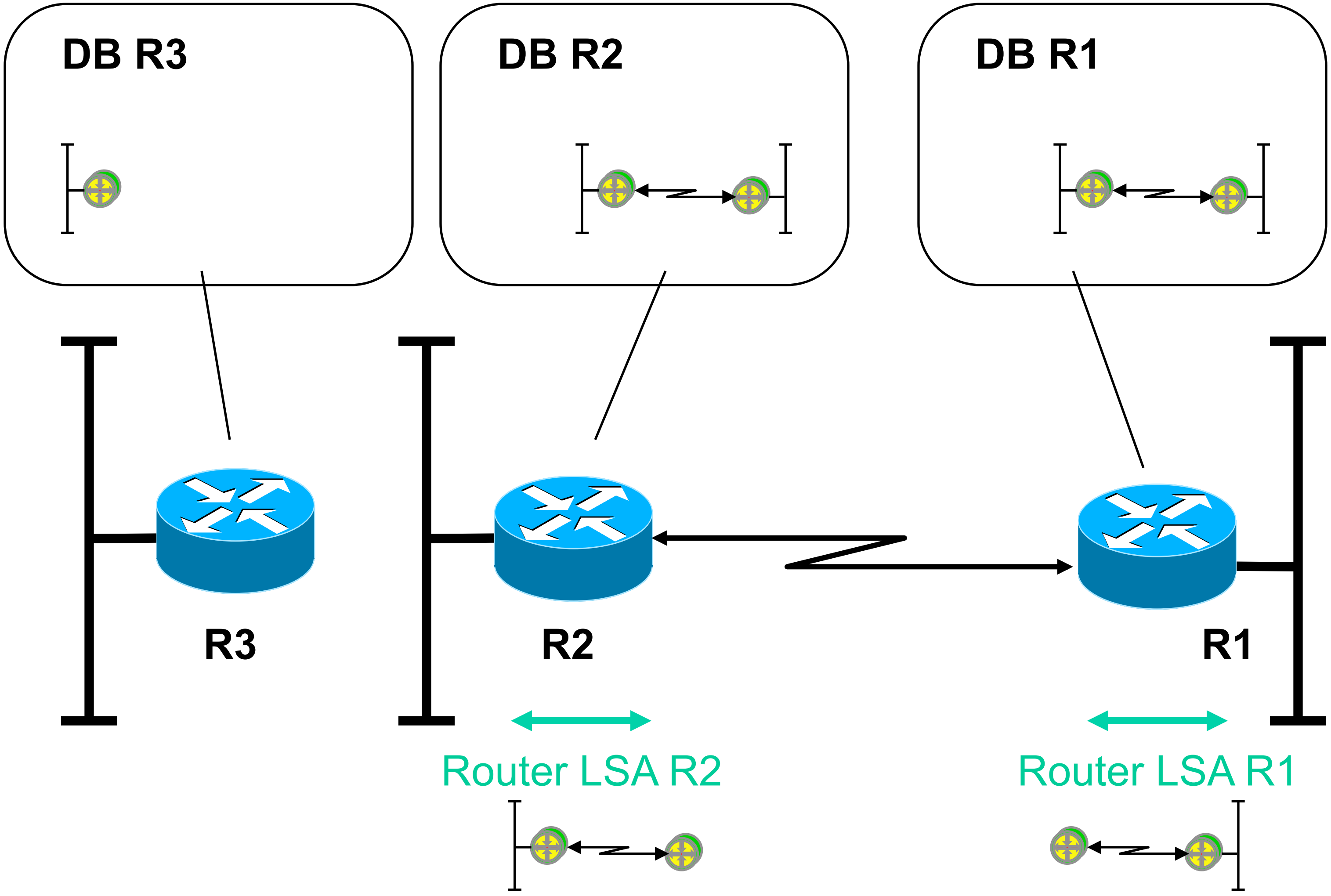
database synchronization: R2 master sends Database-Description, R1 slave sends Link-State Request

OSPF Data Base Update R2 -> R1



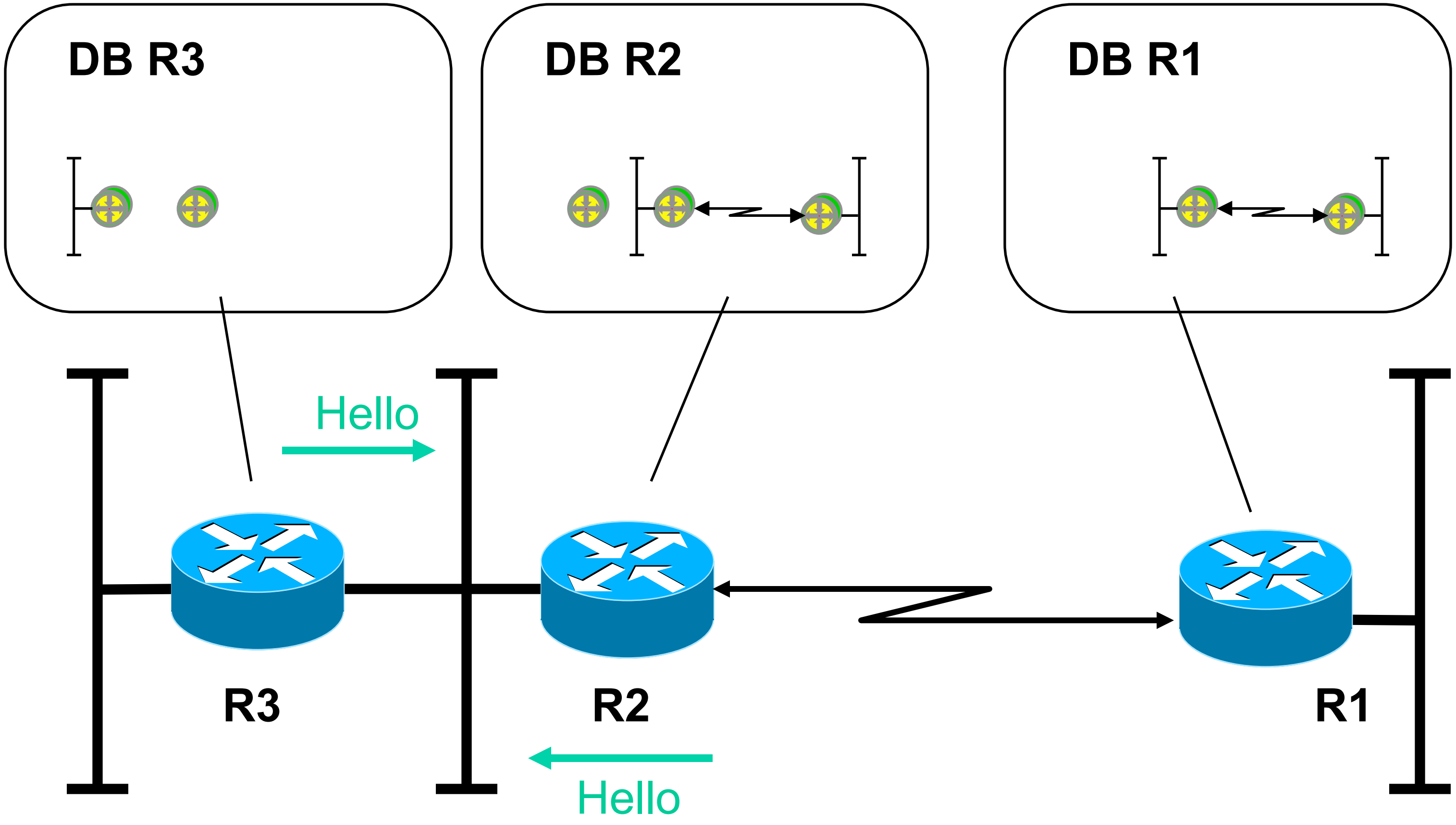
database synchronization: R2 master sends Link-State Update, R1 slave sends Link-State Acknowledgement

OSPF Router LSA Emission



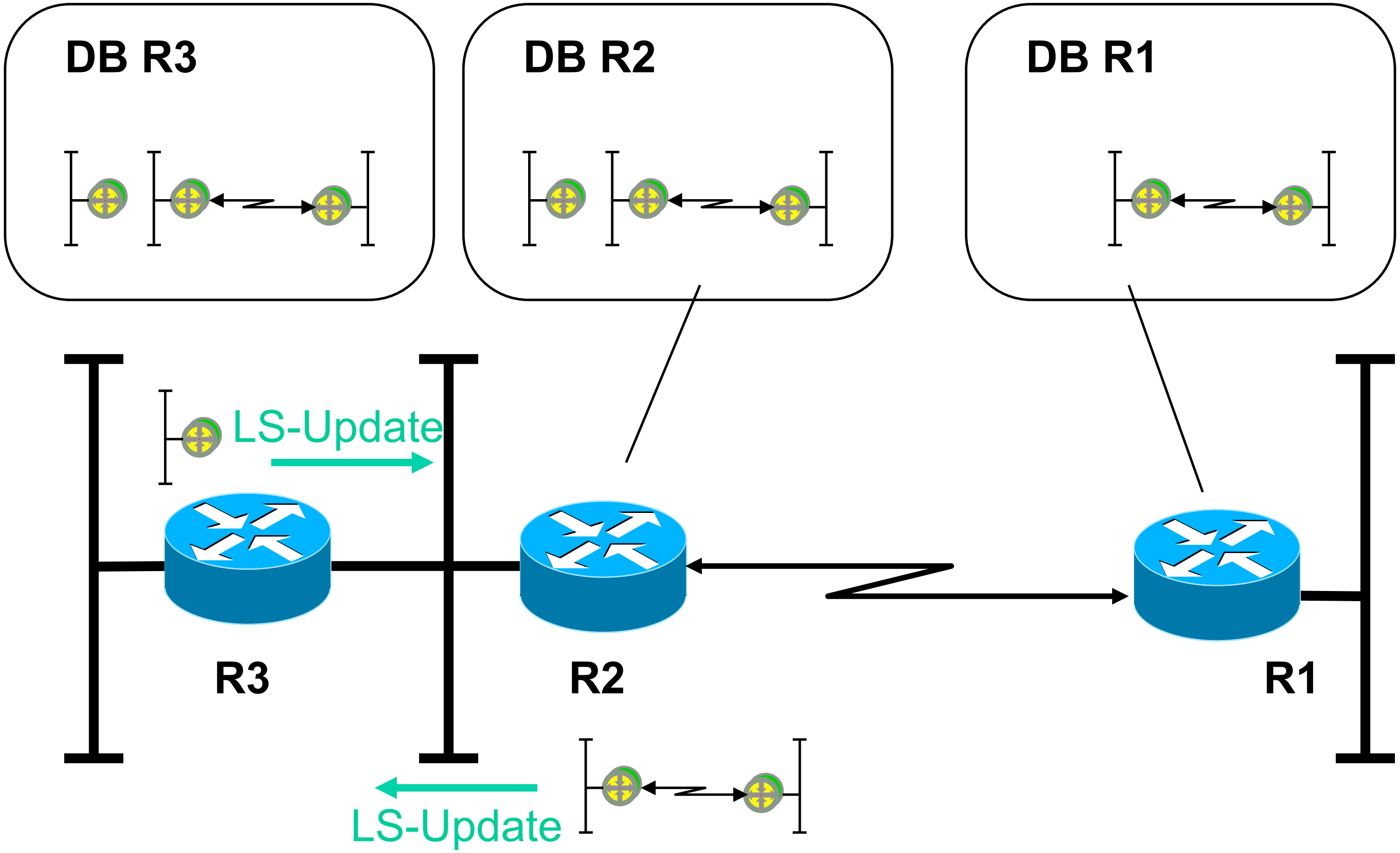
R1 and R2 have synchronized their database completely and notify other nodes about their links

OSPF Hello R2 - R3



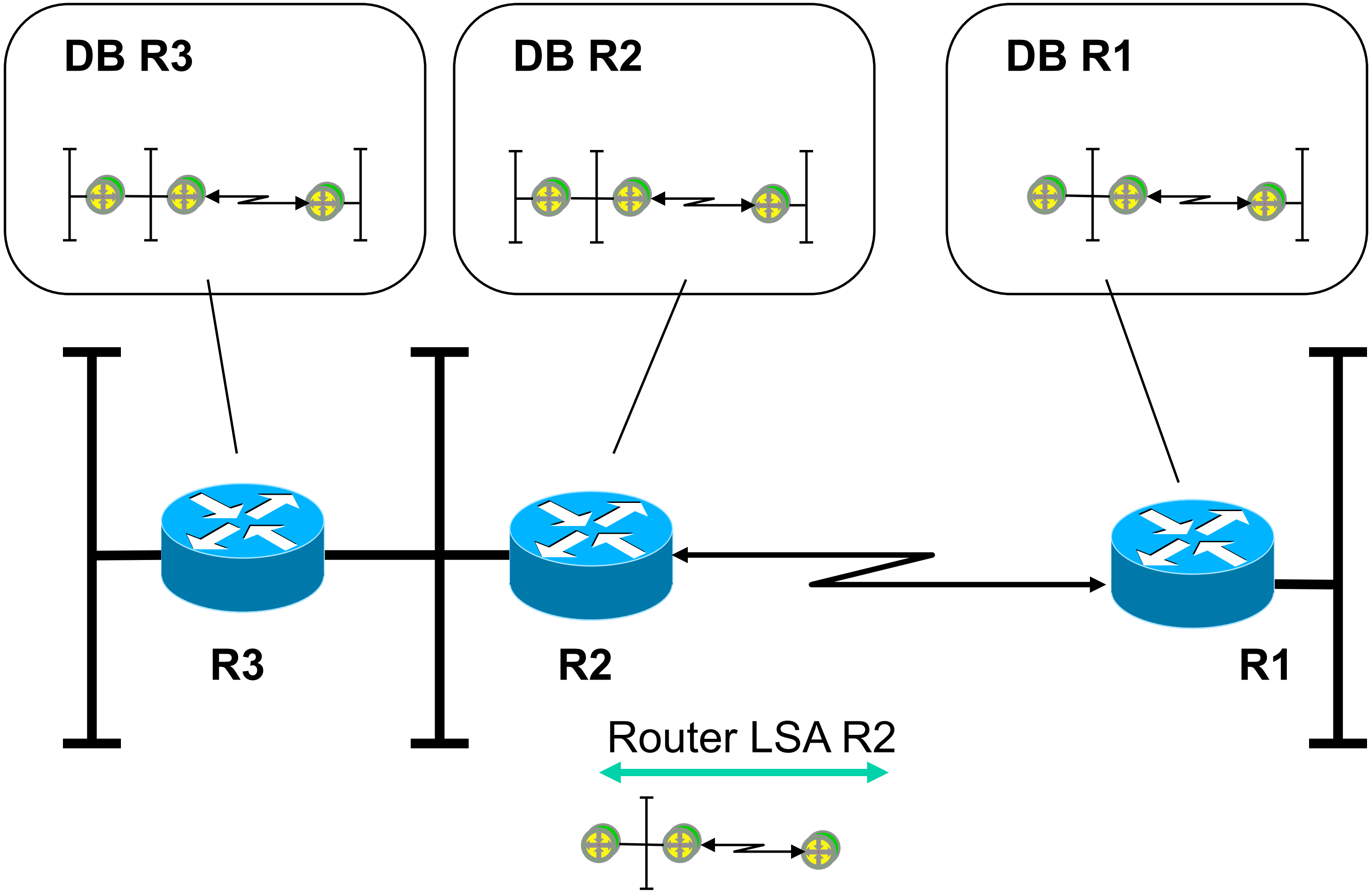
link between R2-R3 activated: get acquainted using Hello, determination of designated router

OSPF Database Update



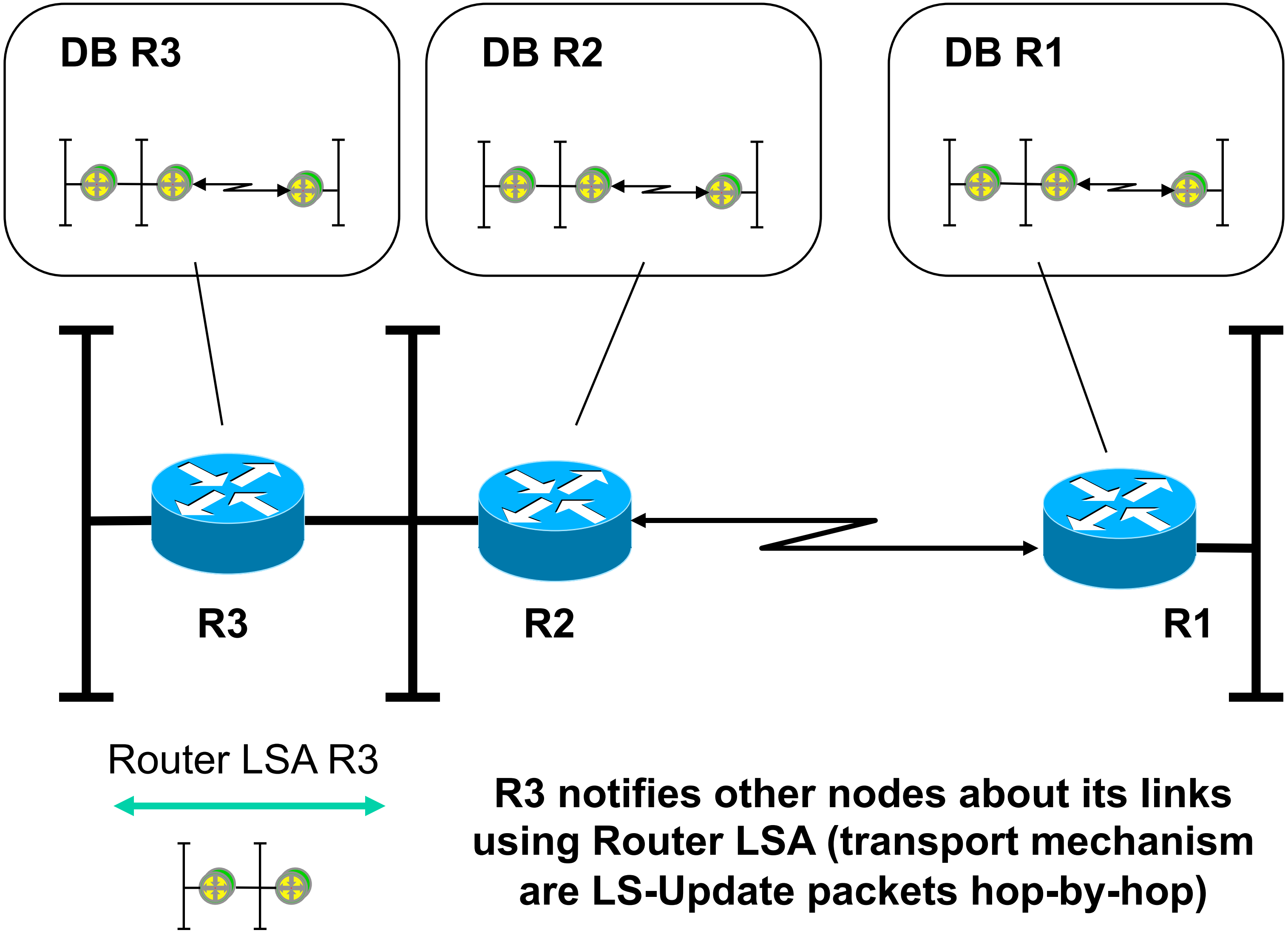
R2 and R3 synchronize their databases (DB-Des., LS-Req.,LS-Upd., LS-Ack.)

OSPF Router LSA Emission R2



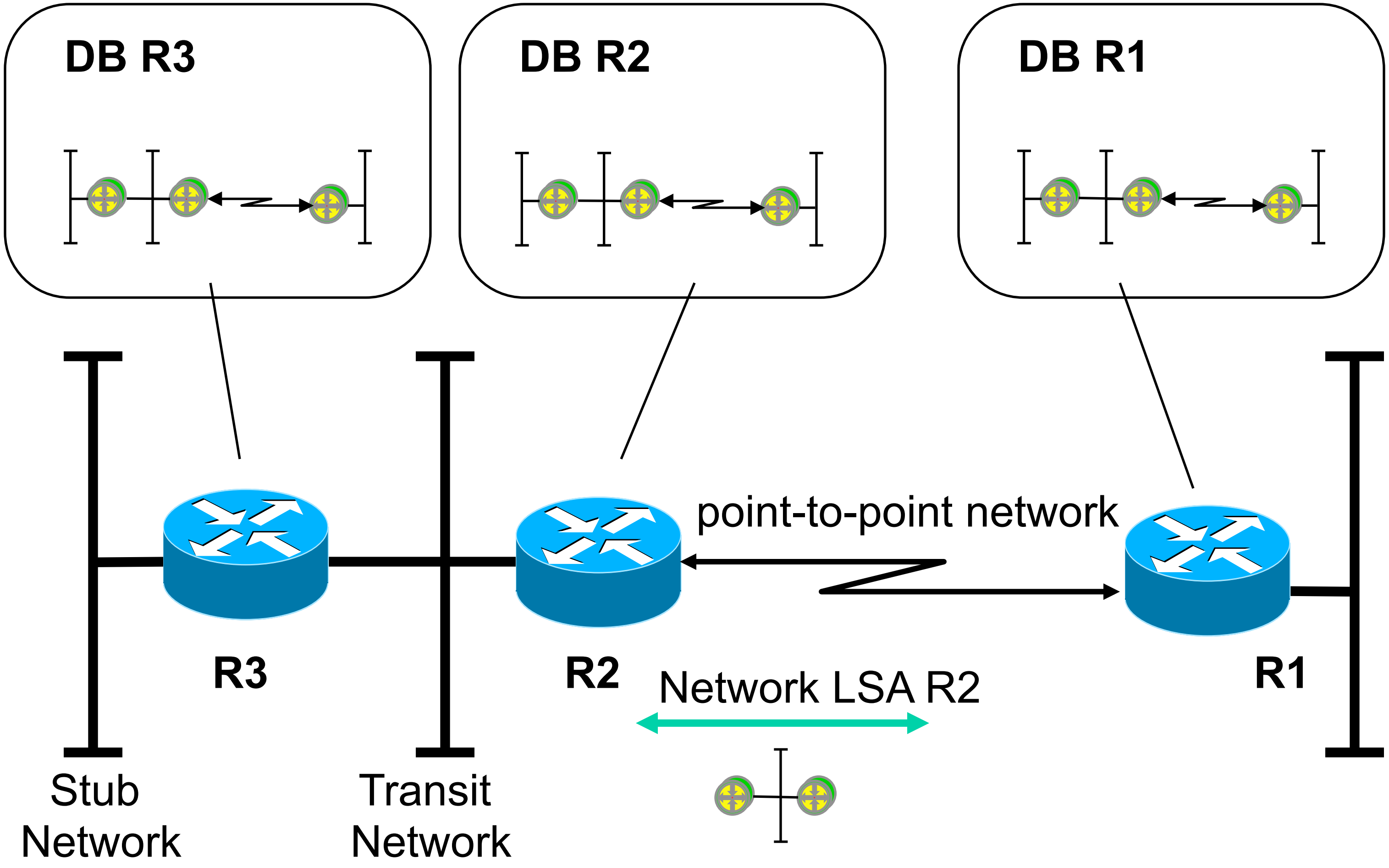
R2 notifies other nodes about its links using Router LSA, (transport mechanism are LS-Update packets hop-by-hop)

OSPF Router LSA Emission R3



R3 notifies other nodes about its links using Router LSA (transport mechanism are LS-Update packets hop-by-hop)

OSPF Network LSA R2



Designated Router R2 notifies other nodes about the multi-access network using Network-LSA (transport mechanism are LS-Update packets hop-by-hop

Agenda

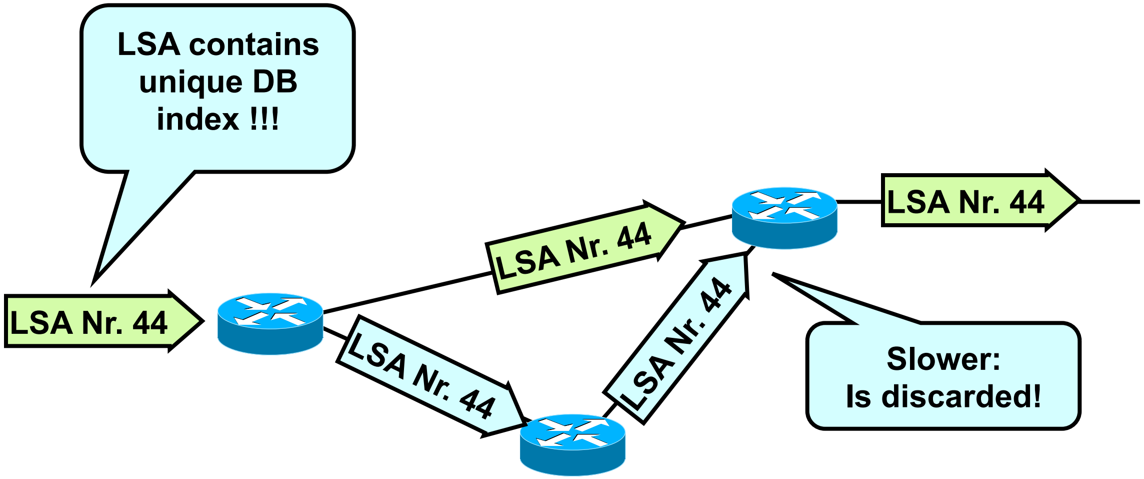
- **L2 versus L3 Switching**
- **IP Protocol, IP Addressing**
- **IP Forwarding**
- **ARP and ICMP**
- **IP Routing**
  - Introduction
  - OSPF Basics
  - OSPF Communication Procedures (Router LSA)
  - LSA Broadcast Handling (Flooding)
  - OSPF Splitted Area
  - Broadcast Networks (Network LSA)
- **First Hop Redundancy**

LSA Broadcast Mechanism (1)

- **Flooding mechanism**
  - Receive of LSA on incoming interface
  - Forwarding of LSA on all other interfaces except incoming interface
  - Well known principle to reach all parts of a meshed network
    - Remember: Transparent bridging – Ethernet switching for unknown destination MAC address
  - "Hot-Potato" method
- **Avoidance of broadcast storm:**
  - With the help of LSA sequence numbers carried in LSA packets and unique indexes of topology database
    - Remember: In case of Ethernet switching we had STP to avoid the broadcast storm
    - În our case we want to establish topology database so we do not have any STP information; SPF information and hence routing tables will result from existence of consistent topology databases
    - "Chicken-Egg" problem

LSA Sequence Number

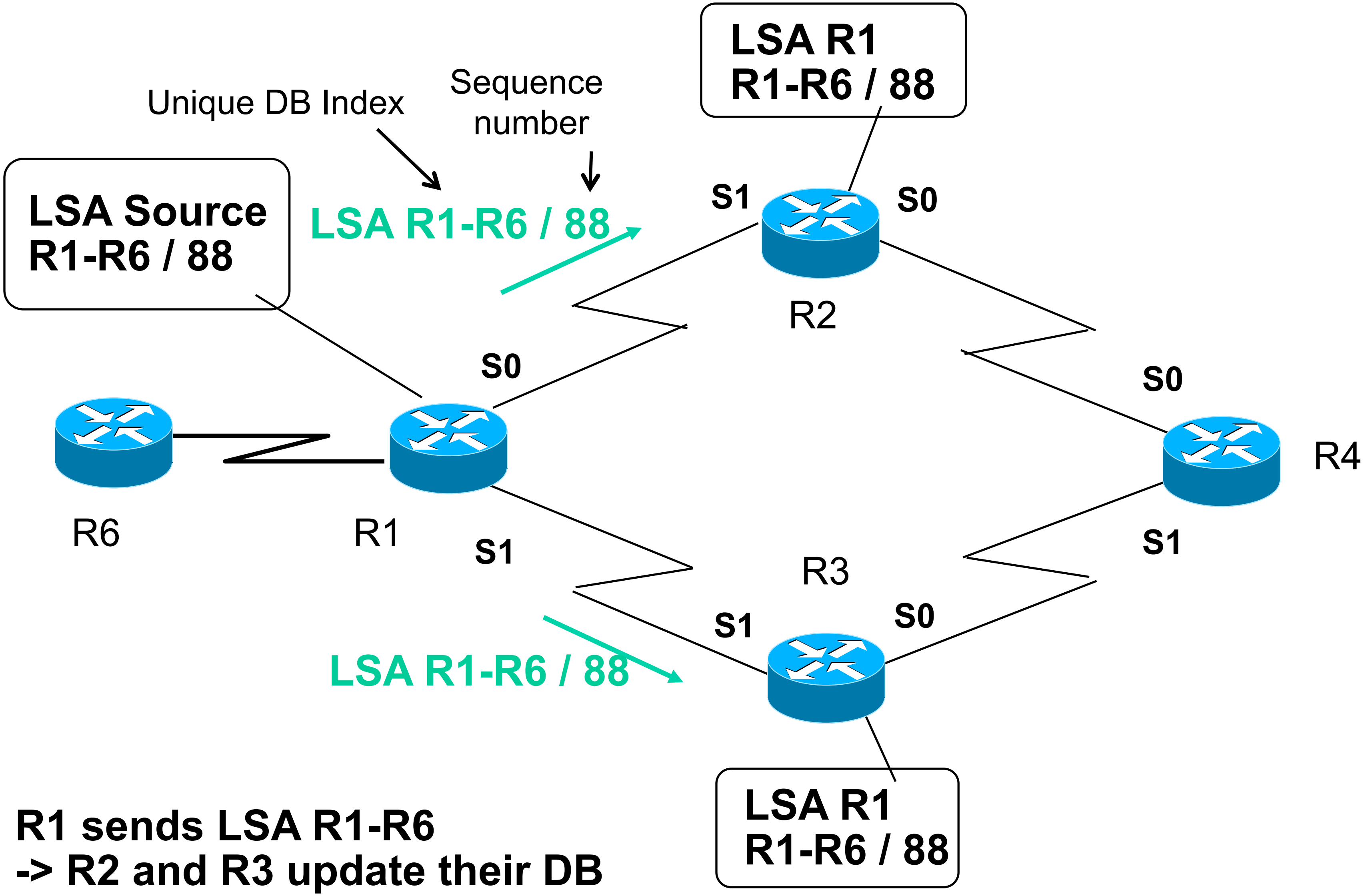
- **In order to stop flooding, each LSA carries a sequence number**
- **Only increased if LSA has changed**
  - So each router can check if a particular LSA had already been forwarded
  - To avoid LSA storms
- **32 bit number**



LSA Broadcast Mechanism (2)

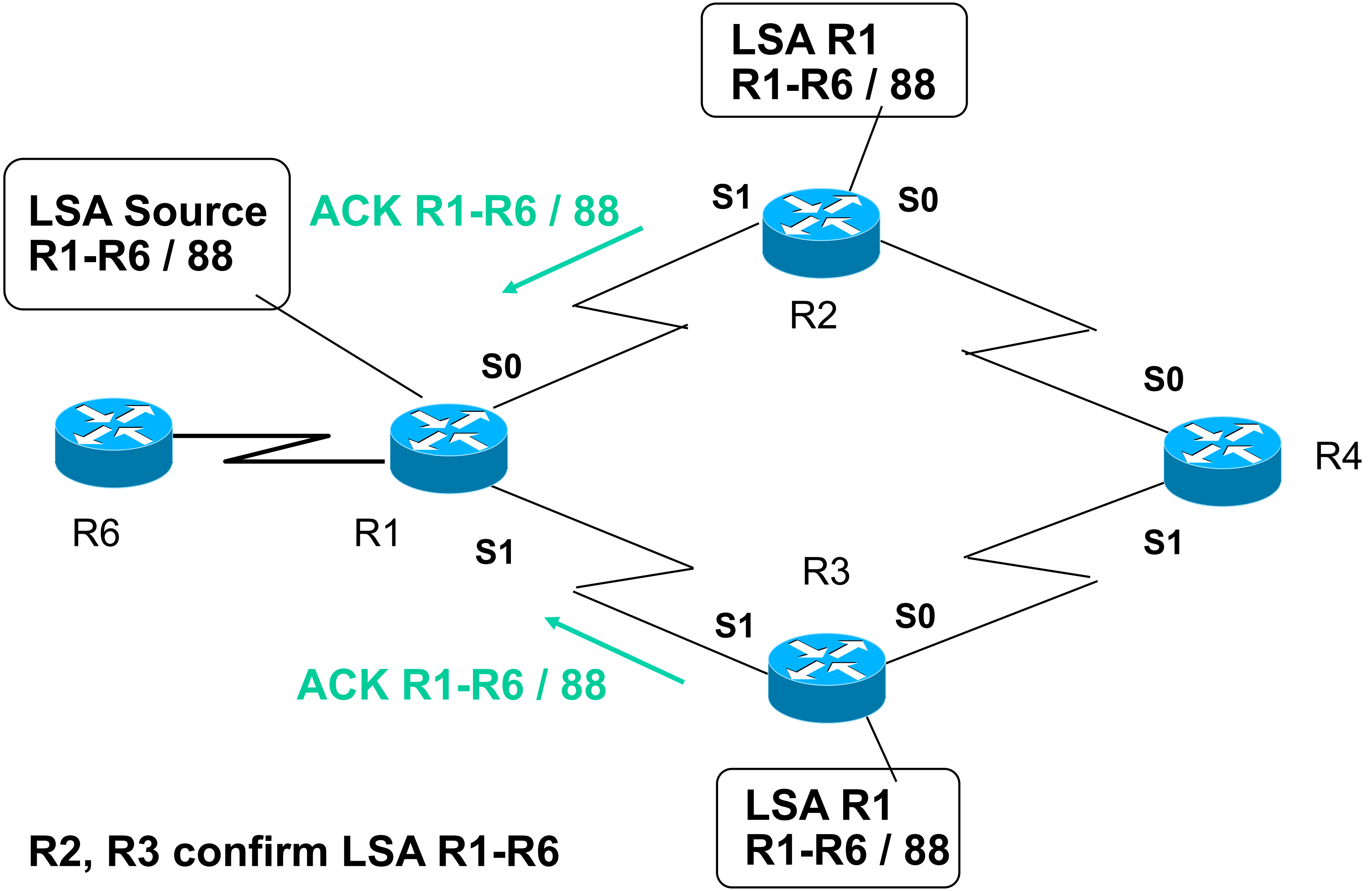
- **LSA must be safely distributed to all routers within an area (domain)**
  - Consistency of the topology-database depends on it
  - Every LS-Update is acknowledged explicitly (using LS-ACK) by the neighbor router
  - If a LS-ACK stays out, the LS-Update is repeated (timeout)
  - If the LS-ACK fails after several trials, the adjacency-relation (the link state between the routers) is cleared

LSA Broadcast Example (1)



R1 sends LSA R1-R6
-> R2 and R3 update their DB

LSA Broadcast Example (2)



LSA R1
R1-R6 / 88

LSA Source
R1-R6 / 88

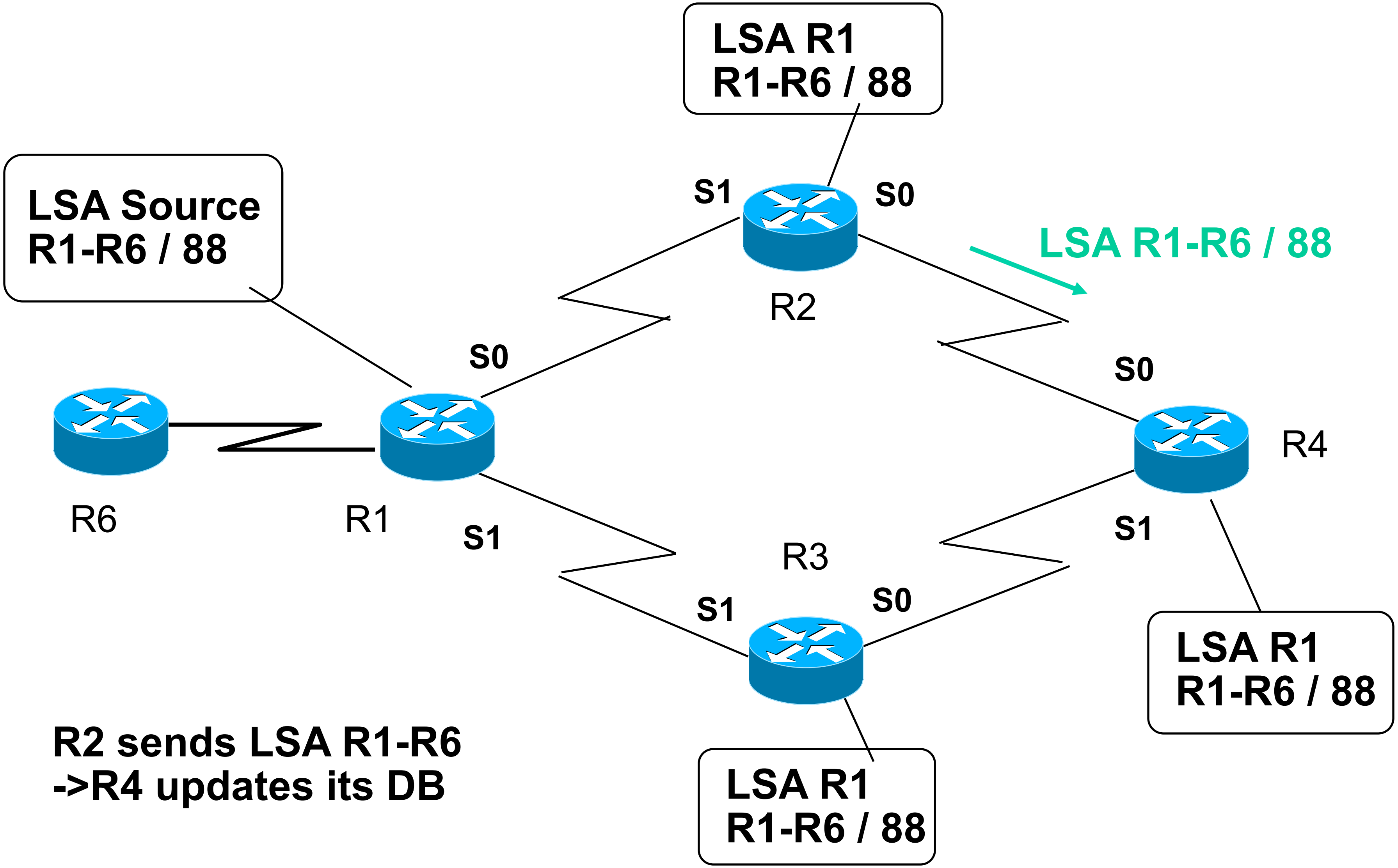
ACK R1-R6 / 88

ACK R1-R6 / 88

LSA R1
R1-R6 / 88

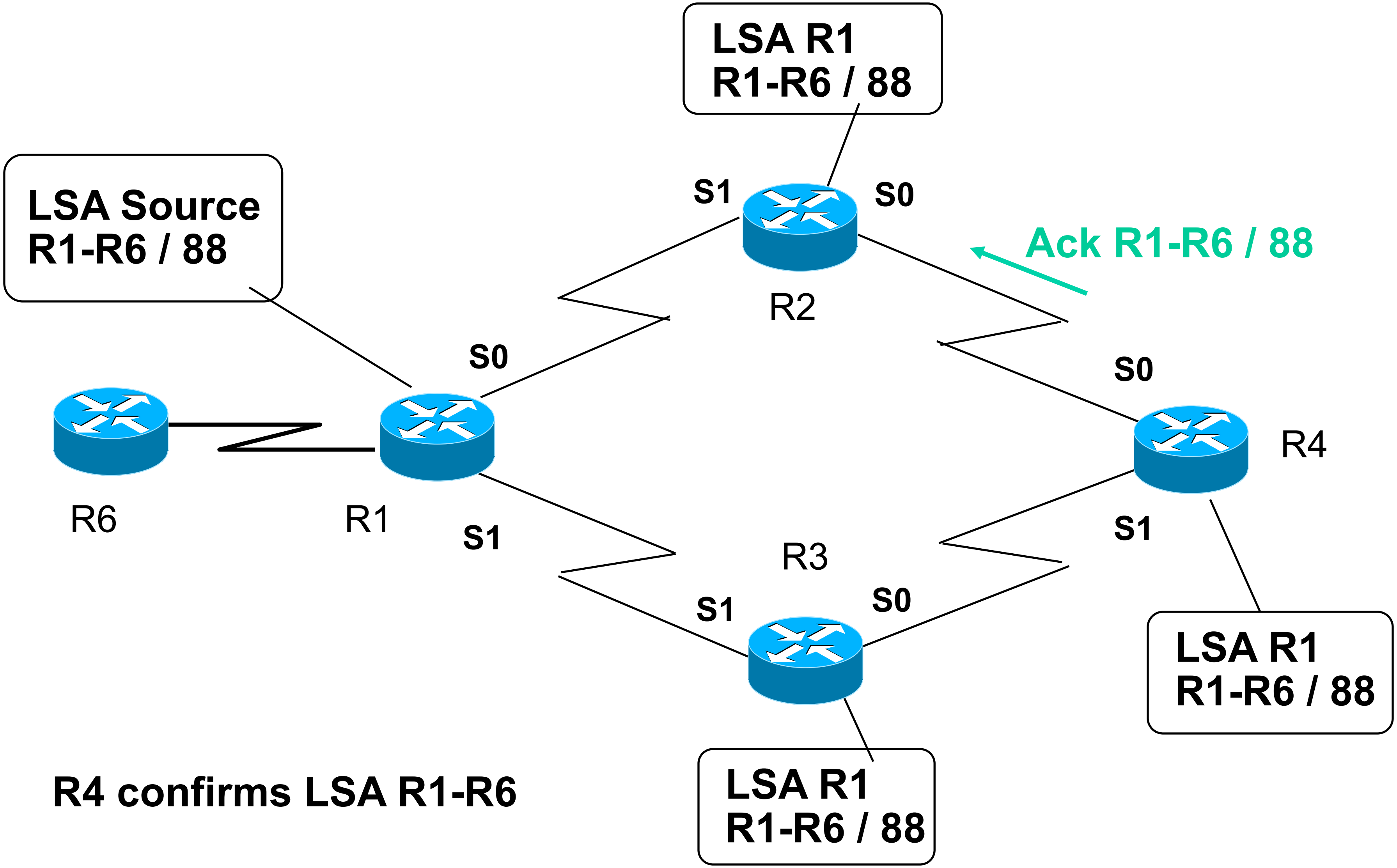
R2, R3 confirm LSA R1-R6

LSA Broadcast Example (3)



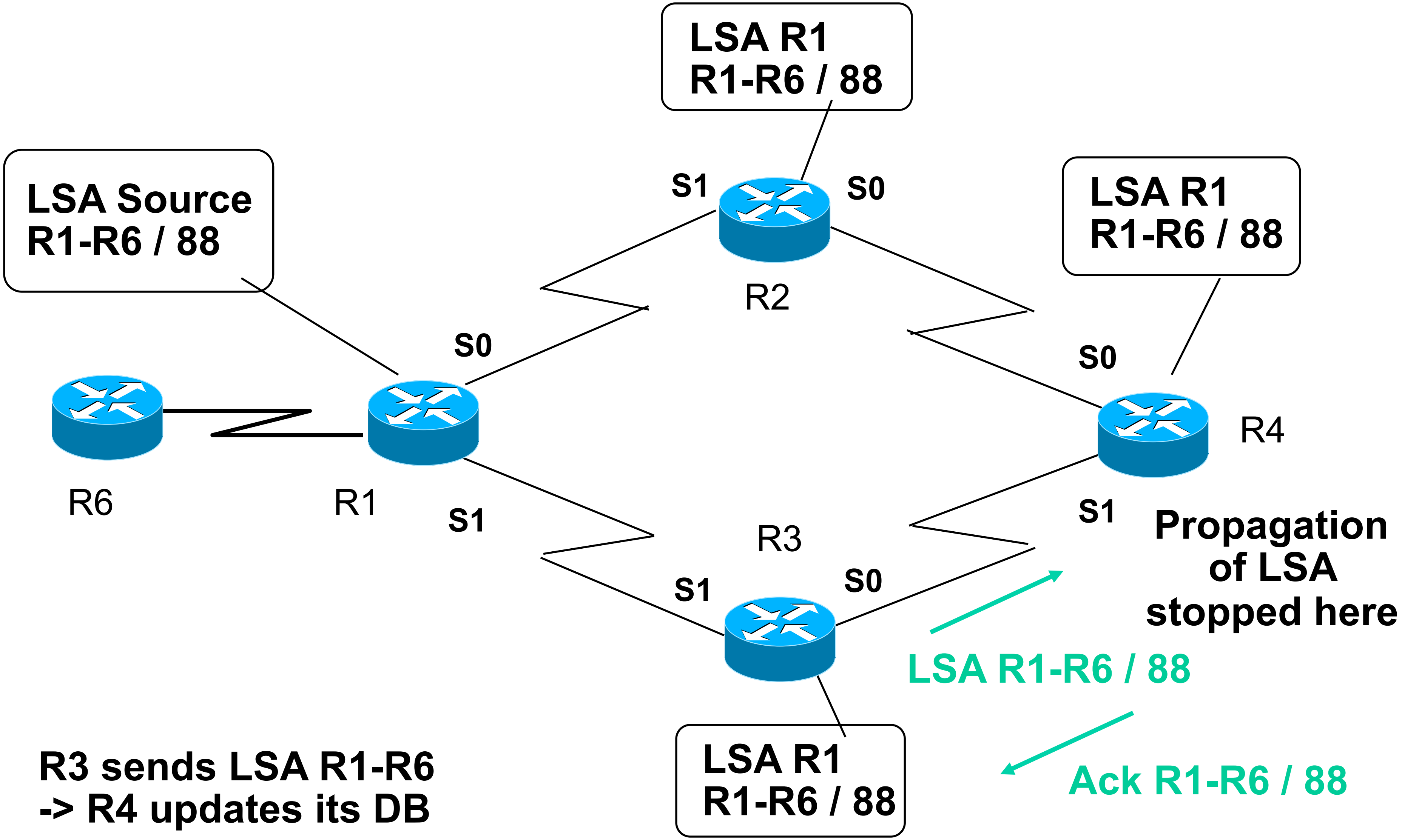
R2 sends LSA R1-R6
->R4 updates its DB

LSA Broadcast Example (4)



R4 confirms LSA R1-R6

LSA Broadcast Example (5)



Propagation of LSA stopped here

R3 sends LSA R1-R6
-> R4 updates its DB

LSA Usage

- **Additionally, link states are repeated every 30 minutes to refresh the databases**
  - Link states – if not refreshed - become obsolete after 60 minutes and are removed from the databases
- **Reasons:**
  - Automatic correction of unnoticed topology-mistakes (e.g. happened during distribution or some router internal failures in the memory)
  - Combining two separated parts of an OSPF area (here OSPF also assures database consistency without intervention of an administrator)

How are LSA unique?

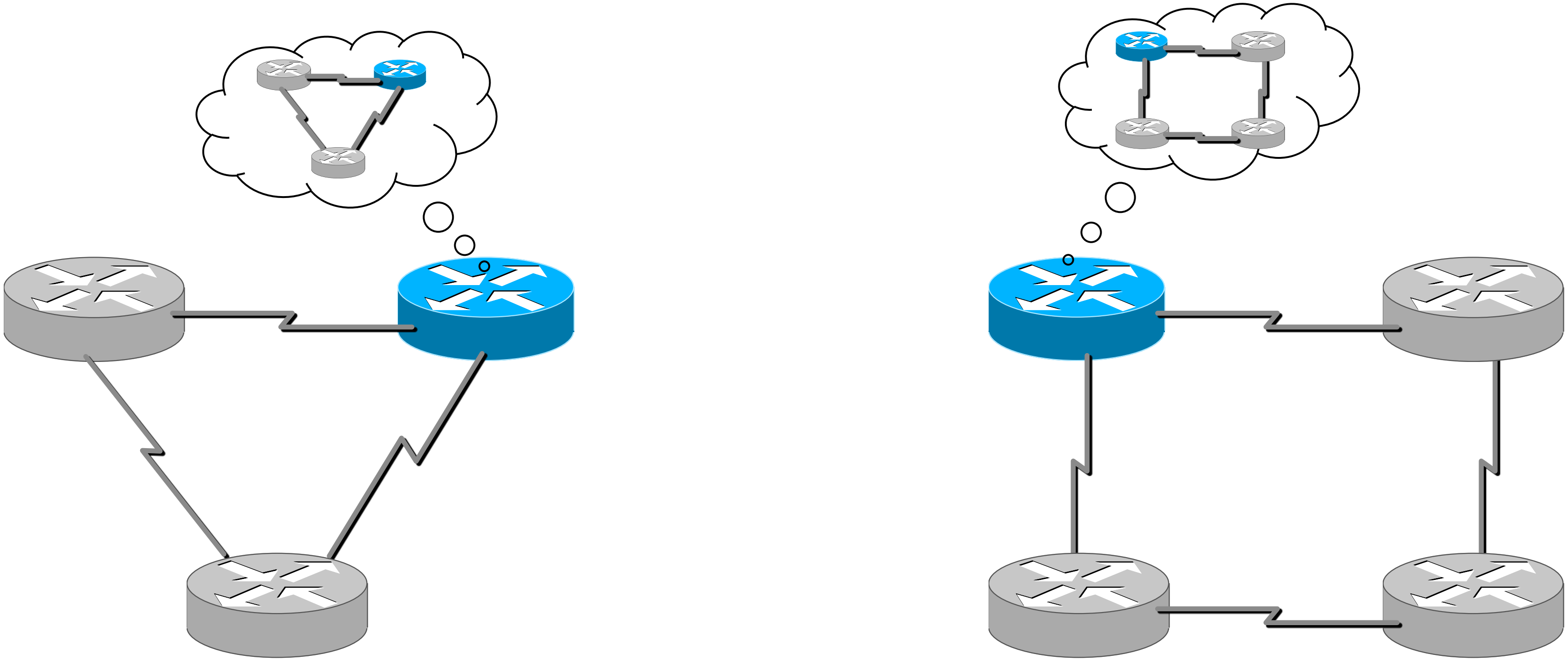
- **Each router as a node in the graph (link state topology database)**
  - Is identified by a unique Router-ID
  - Note: automatically selected on Cisco routers
    - Either numerically highest IP address of all loopback interfaces
    - Or if no loopback interfaces then highest IP address of physical interfaces
- **Every link and hence LS between two routers**
  - Can be identified by the combination of the corresponding Router-IDs
  - Note:
    - If there are several parallel physical links between two routers the Port-ID will act as tie-breaker

Agenda

- L2 versus L3 Switching
- IP Protocol, IP Addressing
- IP Forwarding
- ARP and ICMP
- IP Routing
  - Introduction
  - OSPF Basics
  - OSPF Communication Procedures (Router LSA)
  - LSA Broadcast Handling (Flooding)
  - OSPF Splitted Area
  - Broadcast Networks (Network LSA)
- First Hop Redundancy

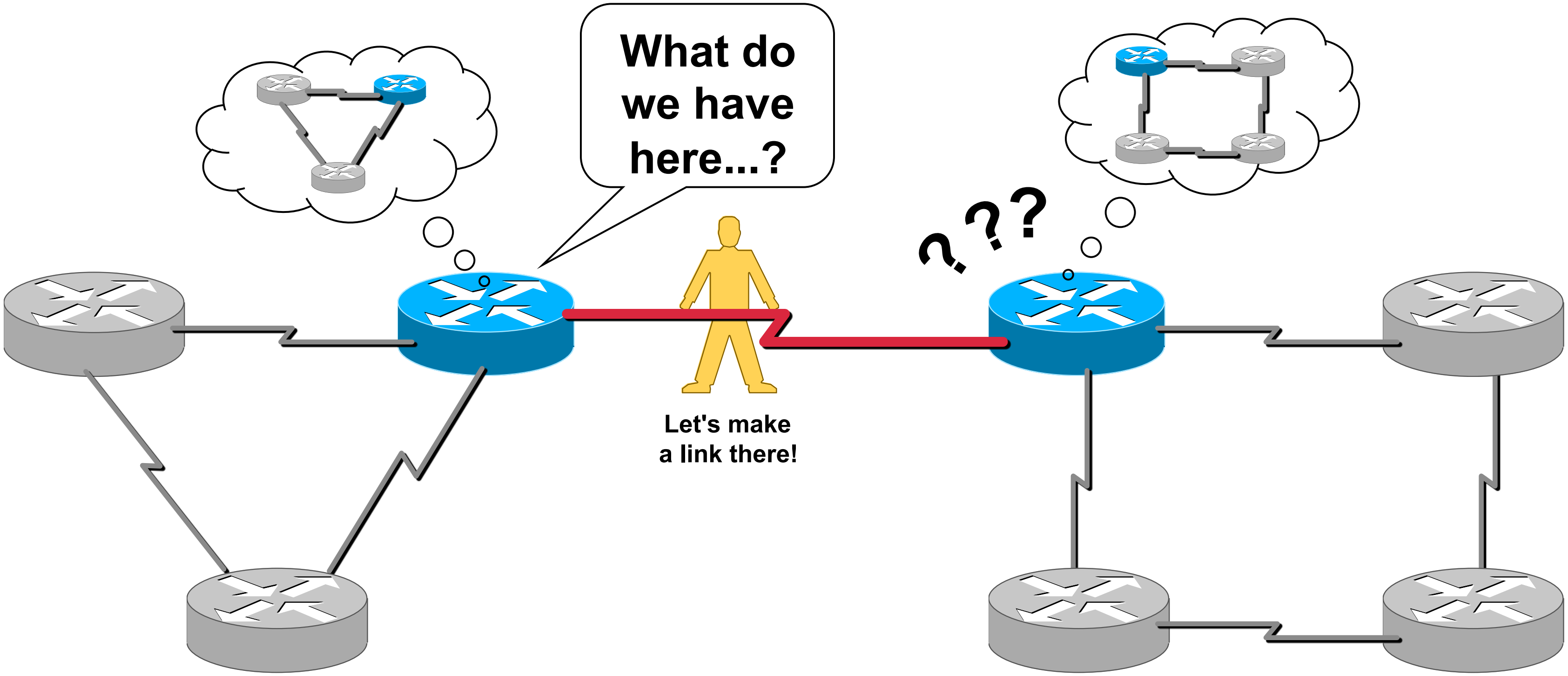
Basic Principle (1)

- Consider two routers, lucky integrated in their own networks...



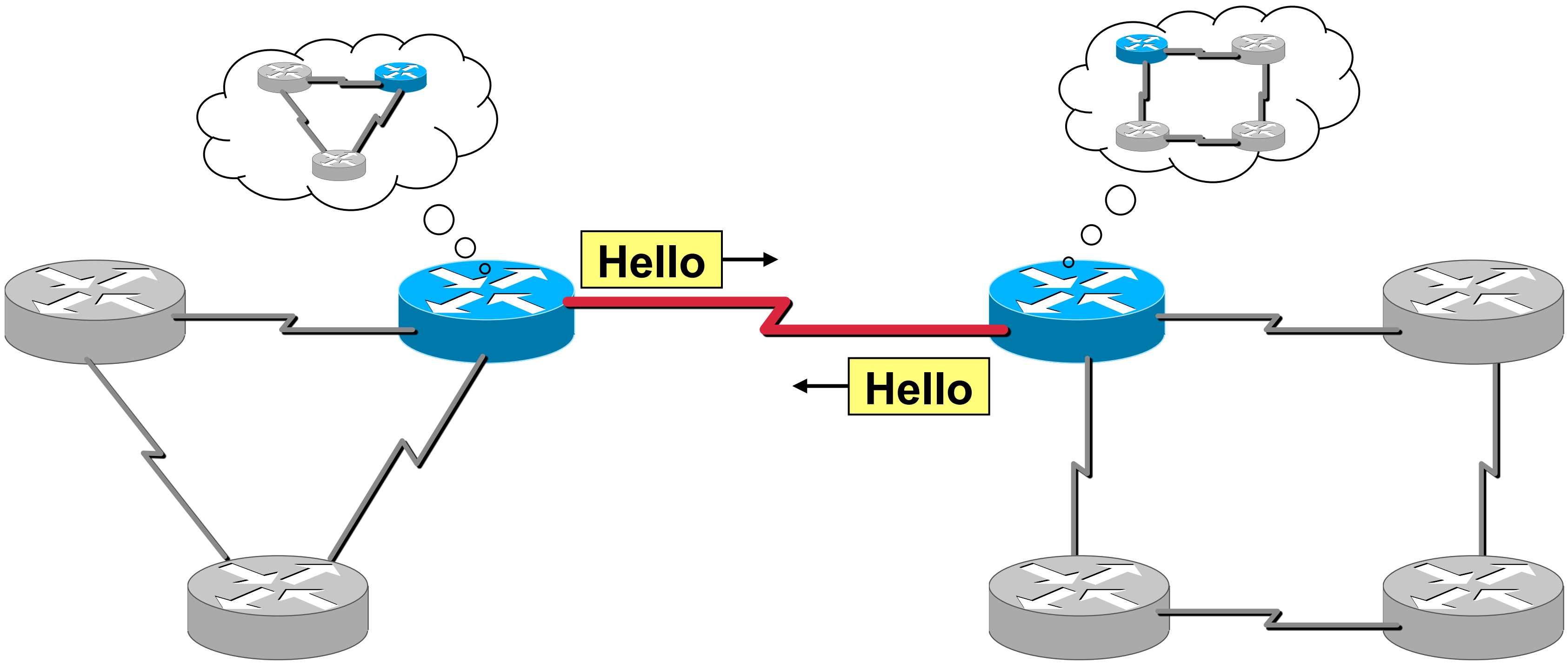
Basic Principle (2)

- Suddenly, some brave administrator connects them via a serial cable...
- Both interfaces are still in the "Down state"



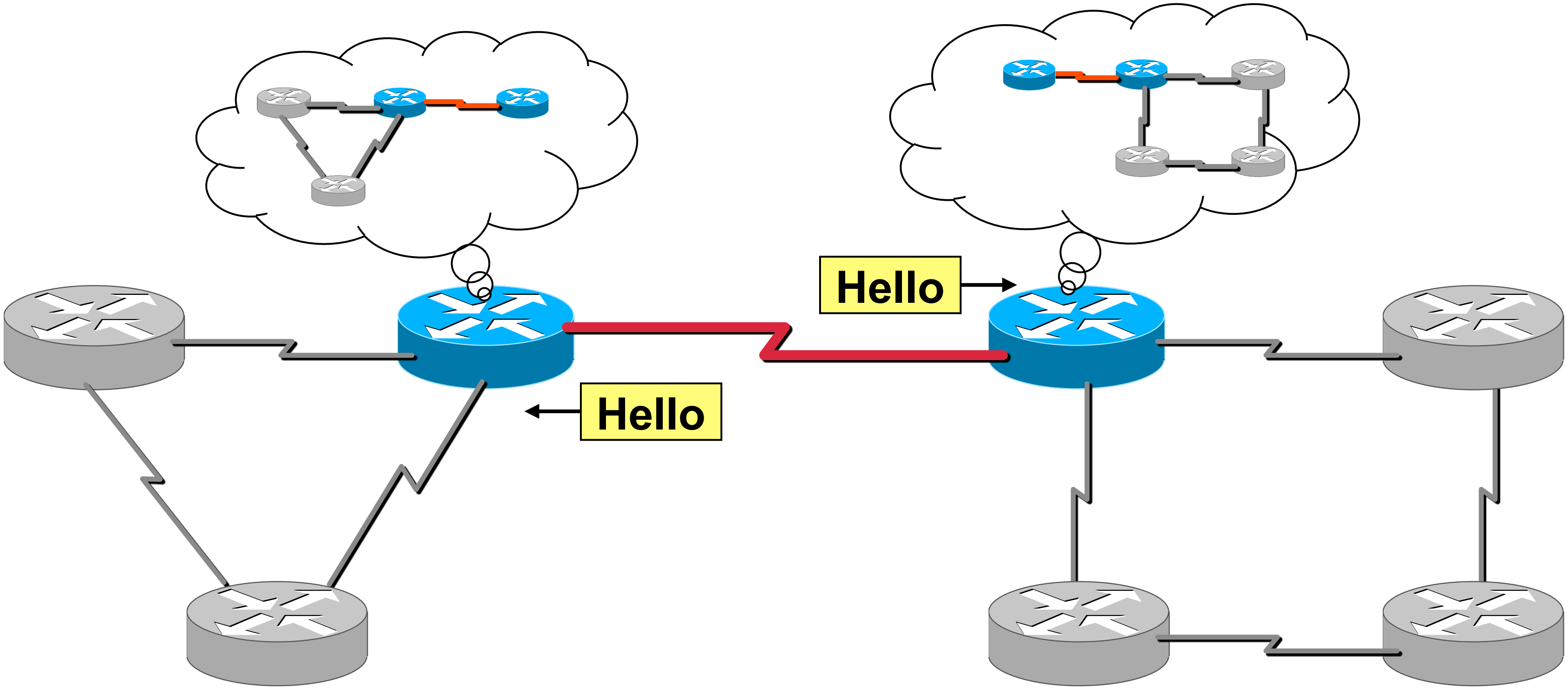
Basic Principle (3)

- **Init state:**
  - Friendly as routers are, they welcome each other using the "Hello protocol"…



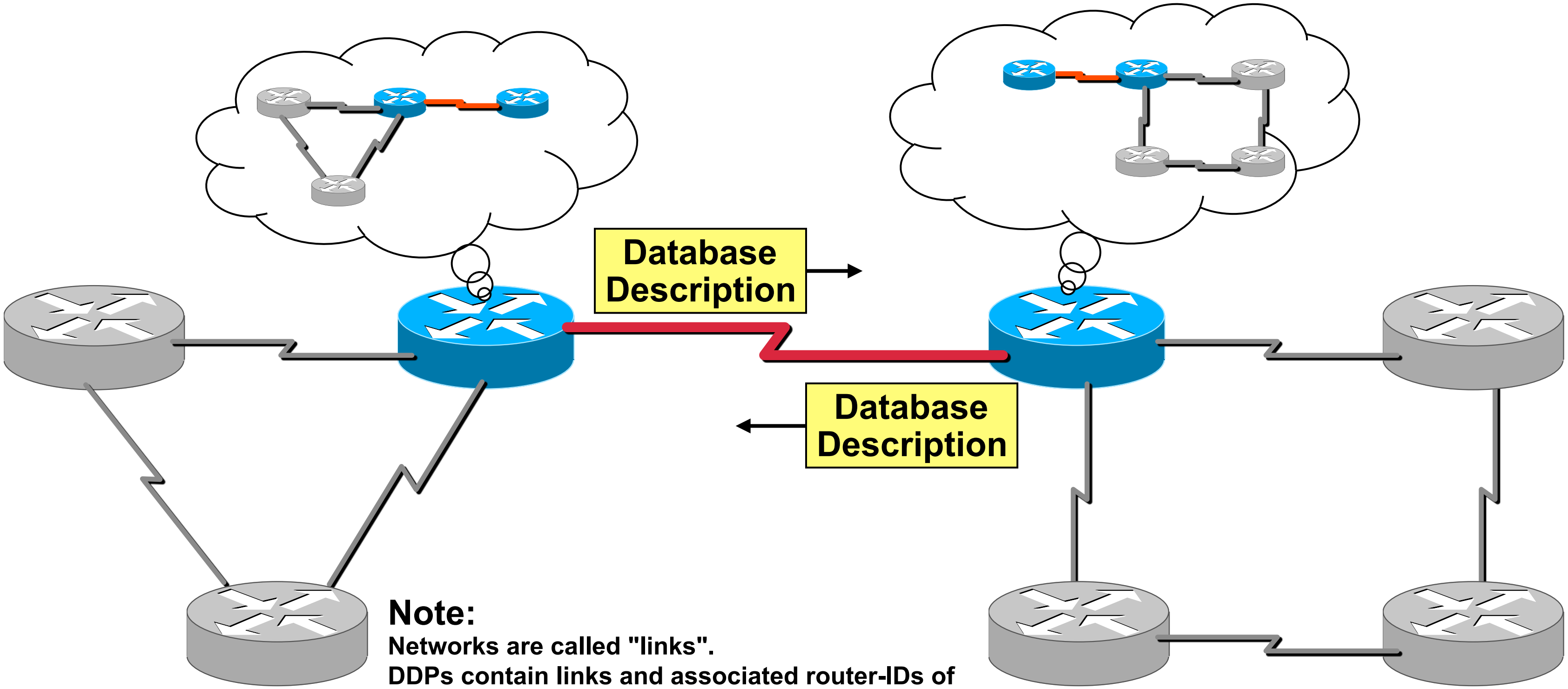
Basic Principle (4)

- **Two-way state:**
  - Each Hello packet contains a list of all neighbors (IDs)
  - Even the two routers themselves are now listed (=> 2-way state condition)
  - Both routers are going to establish the new link in their database...



Basic Principle (5)

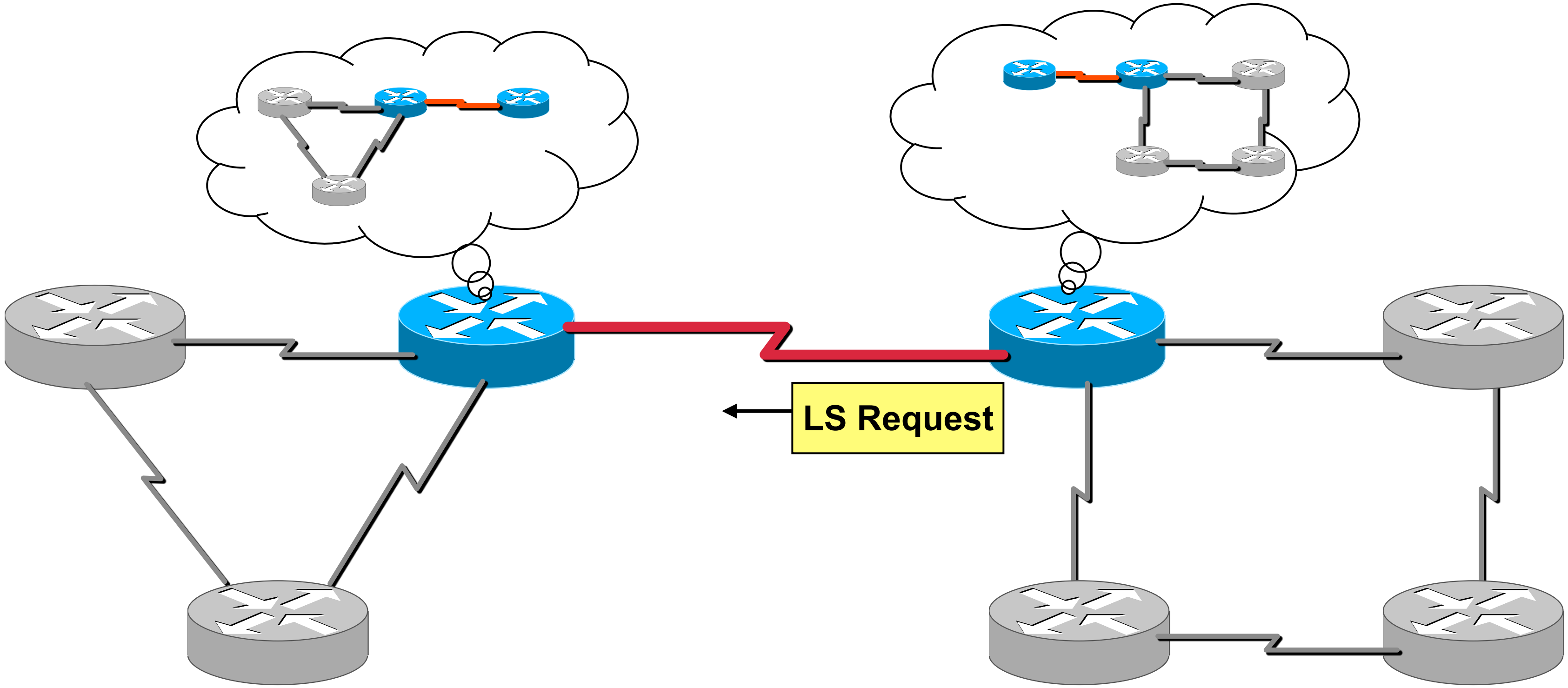
- **Exstart state:**
  - Determination of master (highest IP address) and slave
  - Needed for loading state later
- **Exchange state:**
  - Both router start to offer a short version of their own roadmap, using "Database Description Packets" (DDPs)
  - DDPs contain partial LSAs, which summarize the links of every router in the neighbor’s topology table.



Note:
Networks are called "links".
DDPs contain links and associated router-IDs of the originators of the corresponding LSAs.

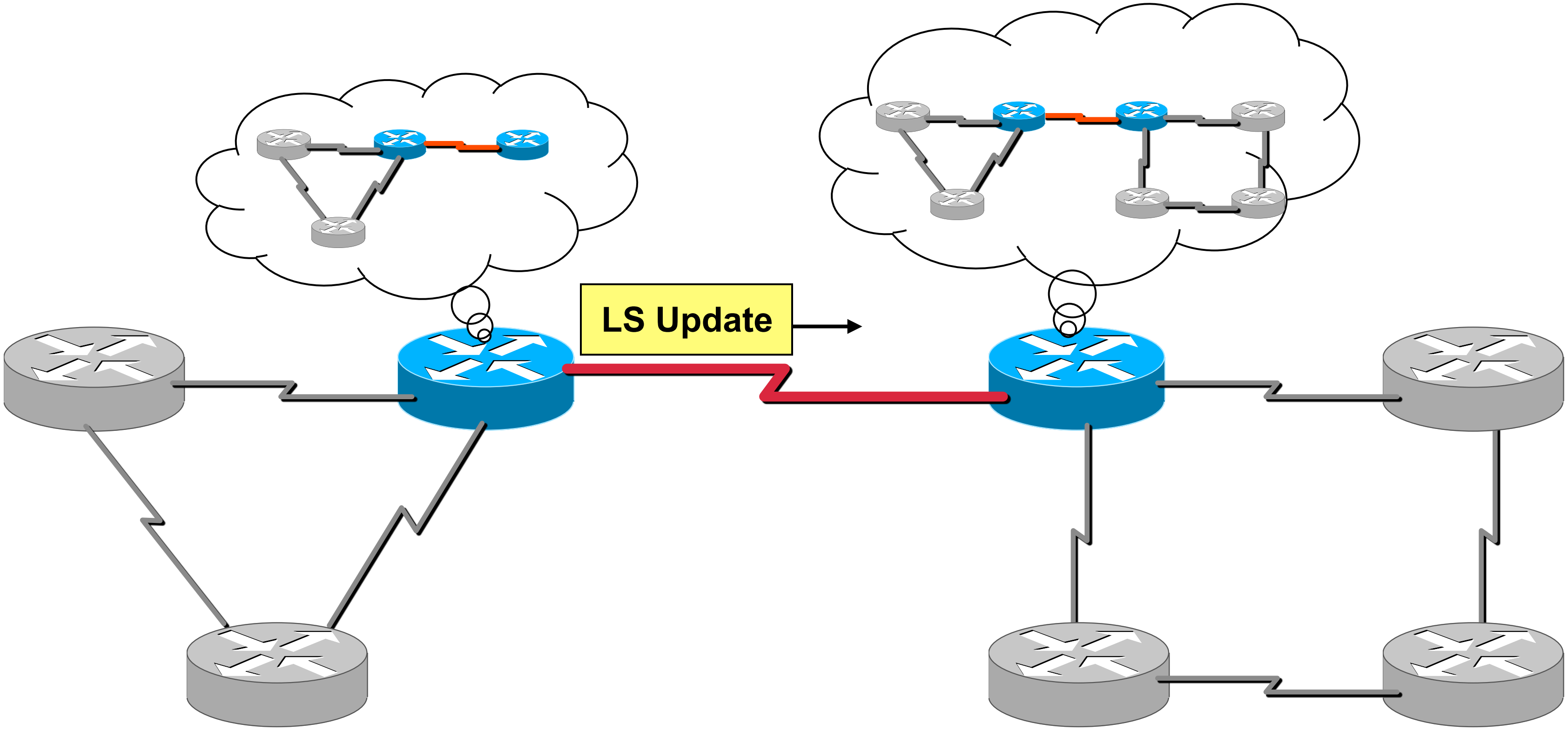
Basic Principle (6)

- **Loading State:**
  - One router (here the right one) recognizes some missing links and asks for detailed information using a "Link State Request" (LSR) packet...



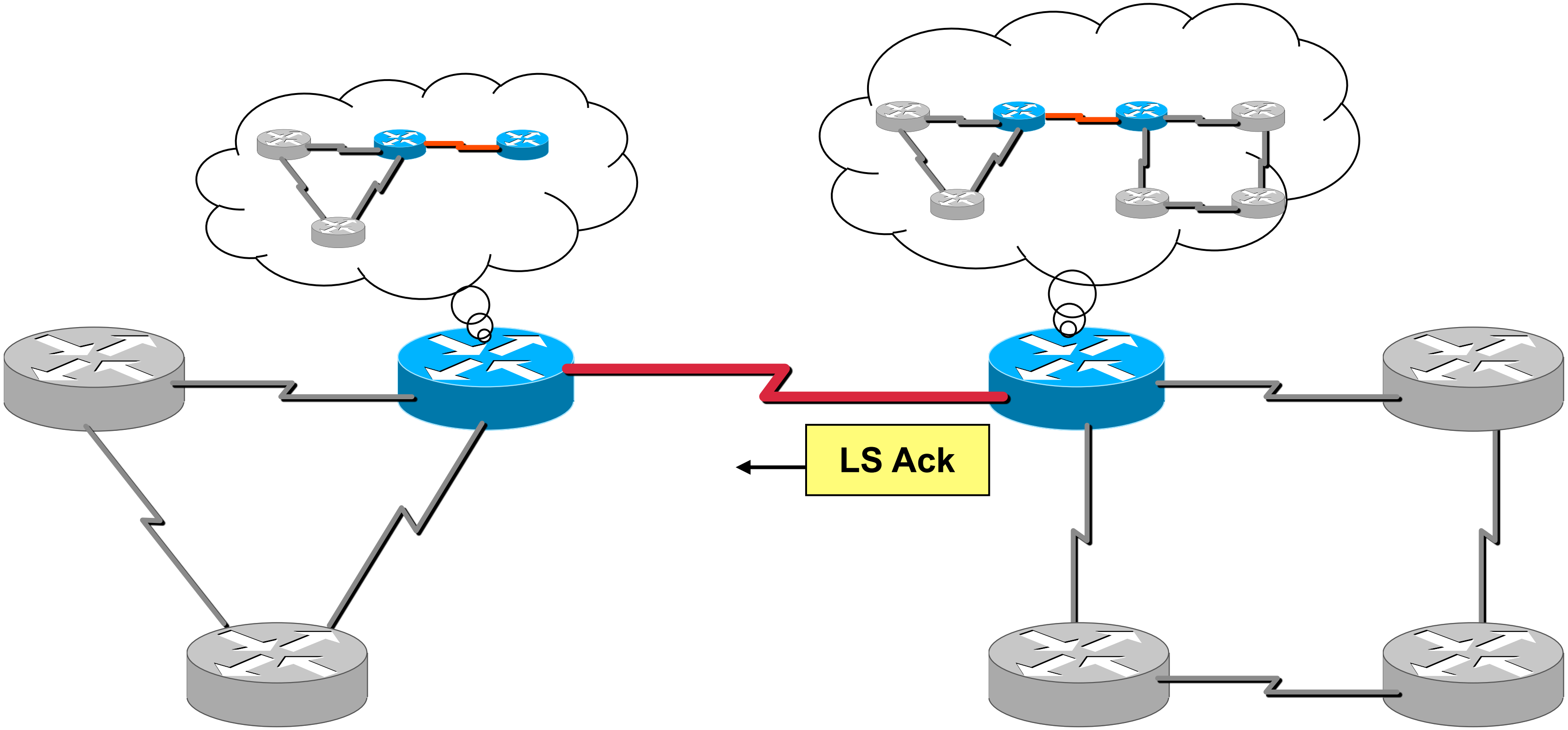
Basic Principle (7)

- The left router replies immediately with the requested link information, using a
  "Link State Update" (LSU) packet ...



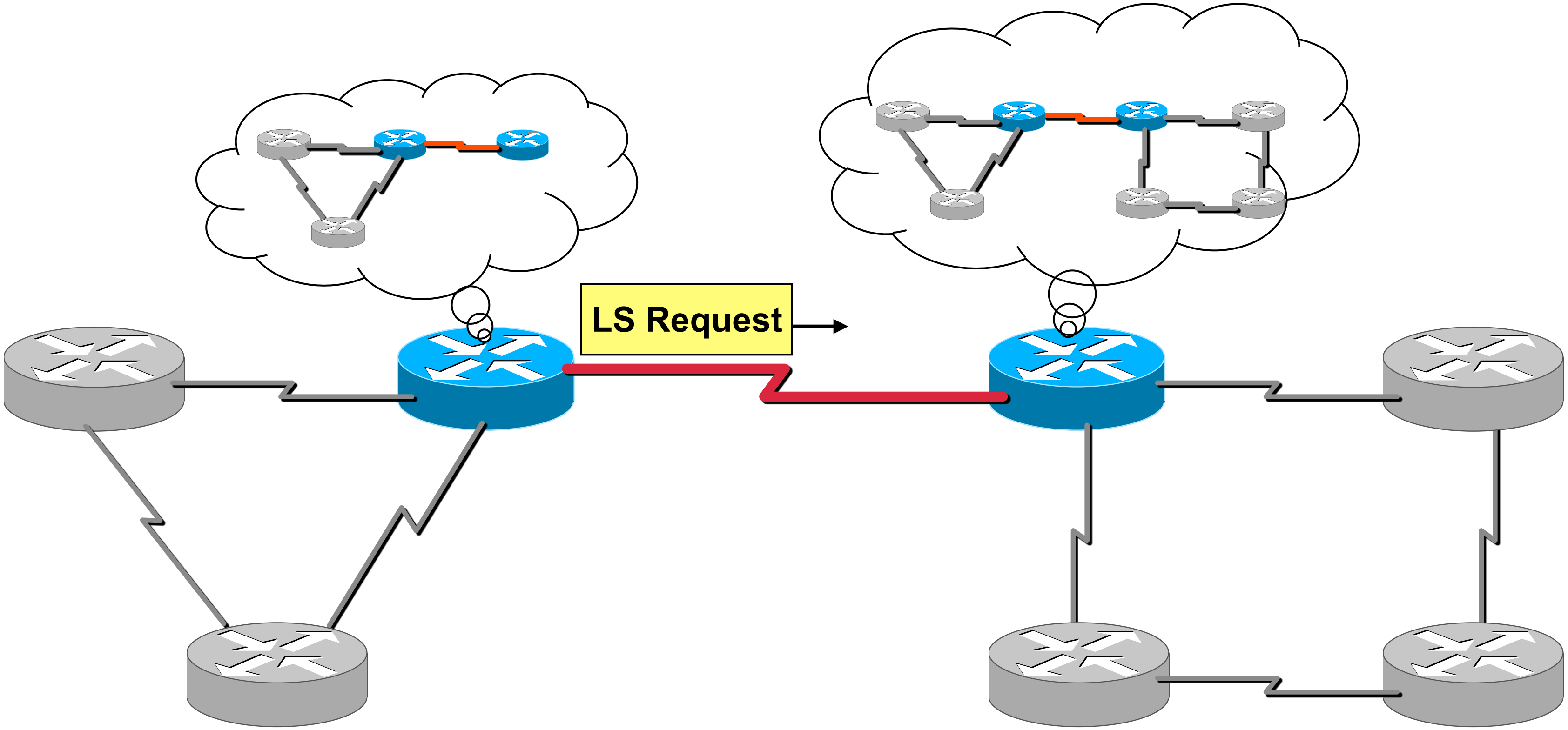
Basic Principle (8)

- The right router is very thankful, and returns a "Link State Acknowledgement"...



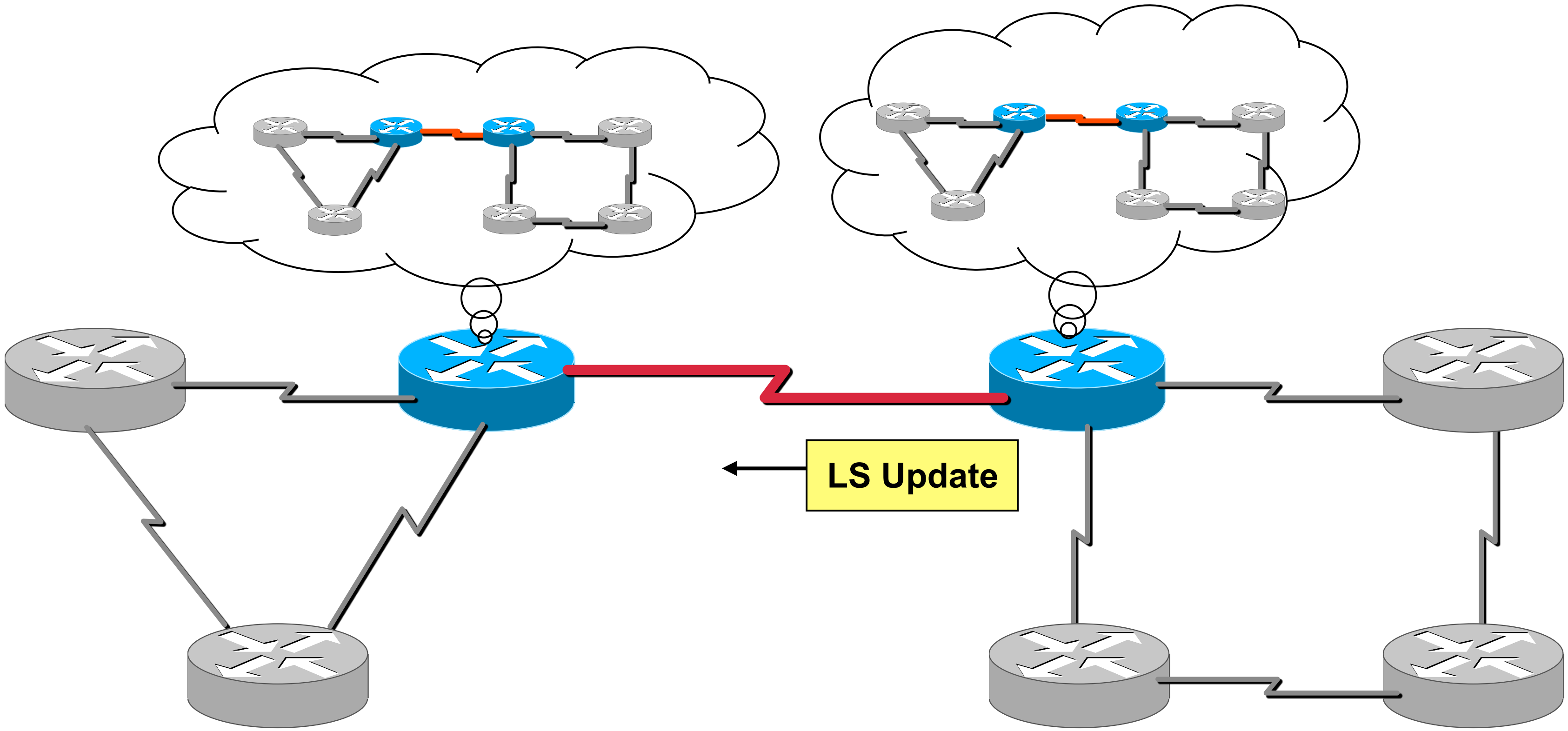
Basic Principle (9)

- **Then the left router recognizes some unknown links and asks for further details...**



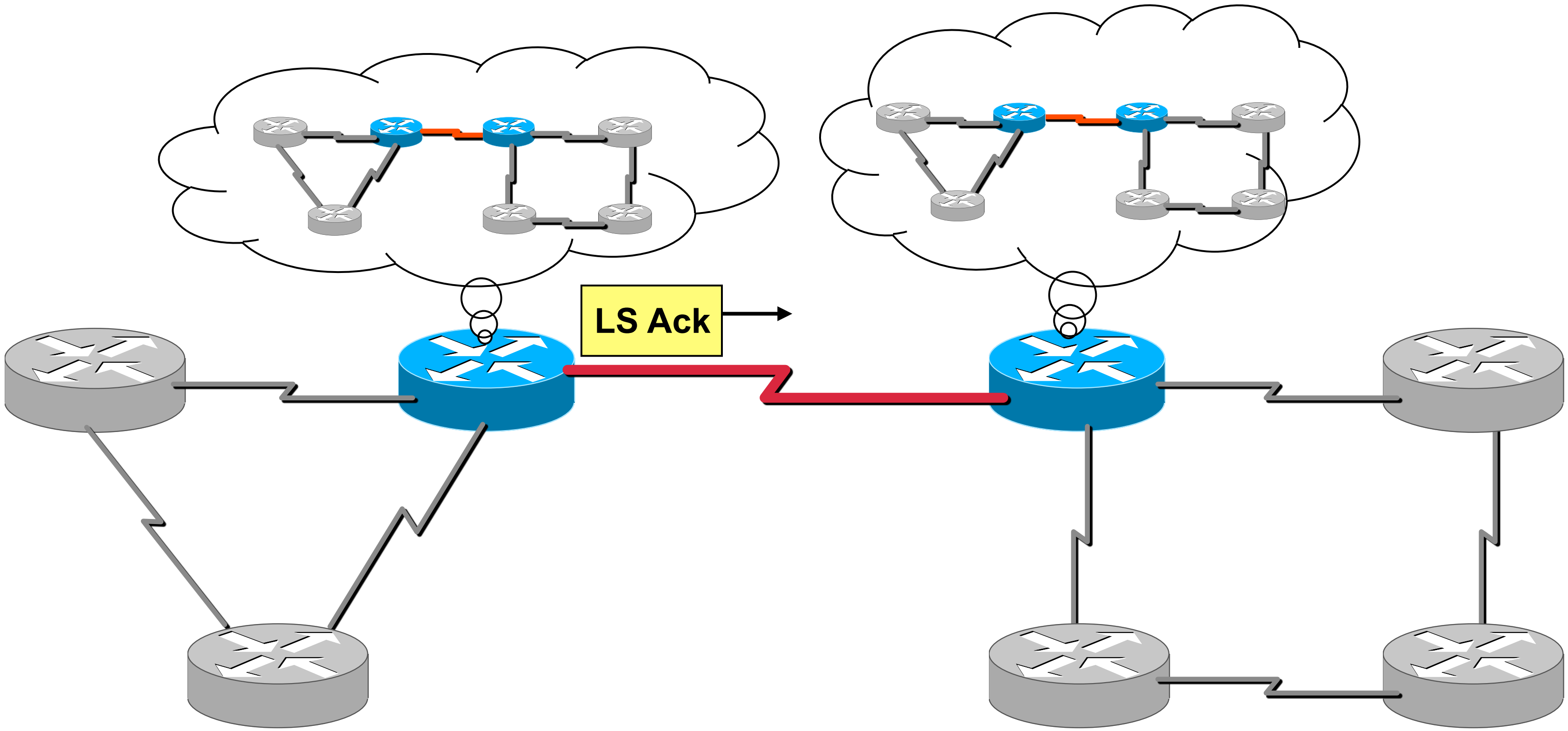
Basic Principle (10)

- **The right router sends detailed information for the requested unknown links...**



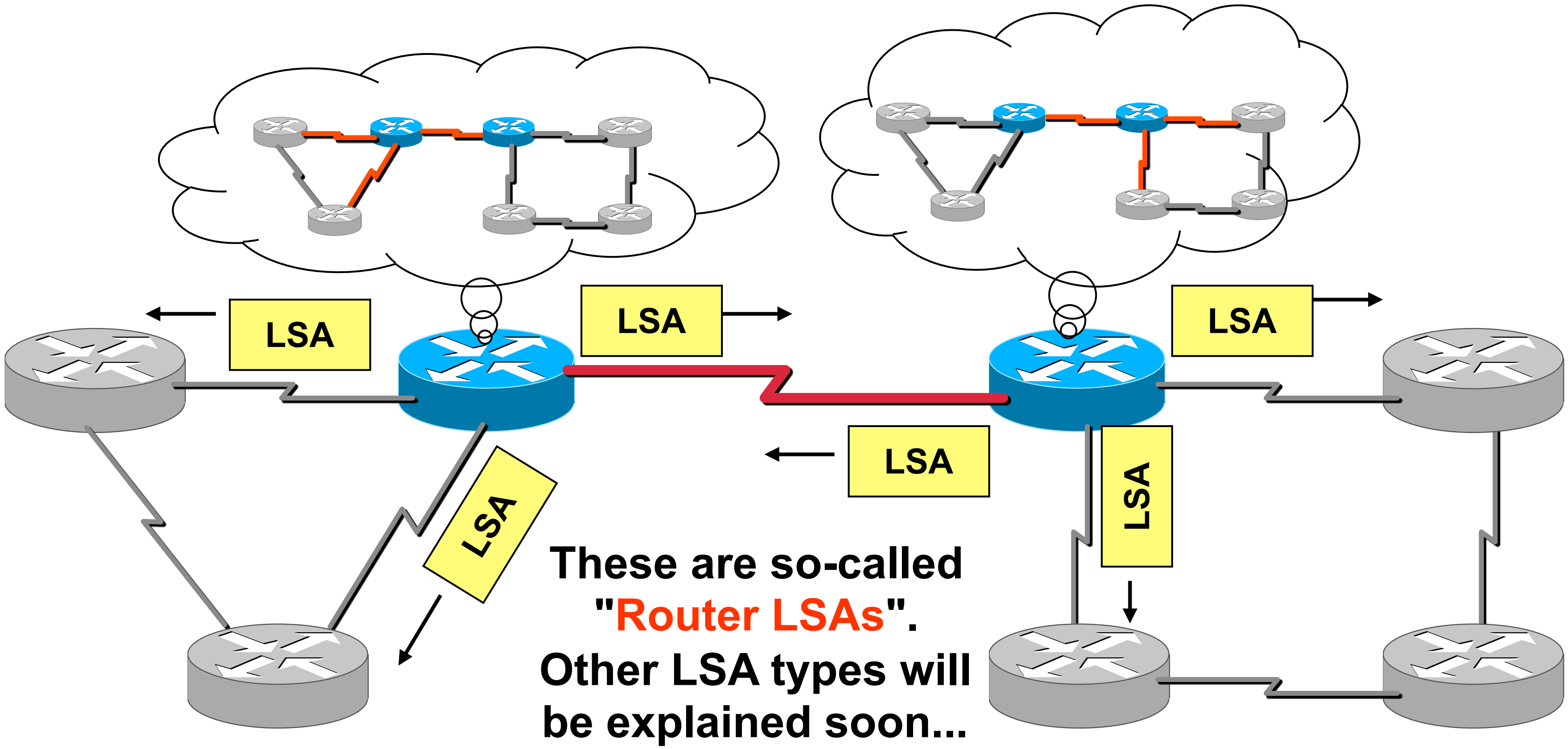
Basic Principle (11)

- **The left router replies with a link state acknowledgement – a new adjacency has been established...**
  - Neighbors are "fully adjacent" and reached the "full state"



Basic Principle (12)

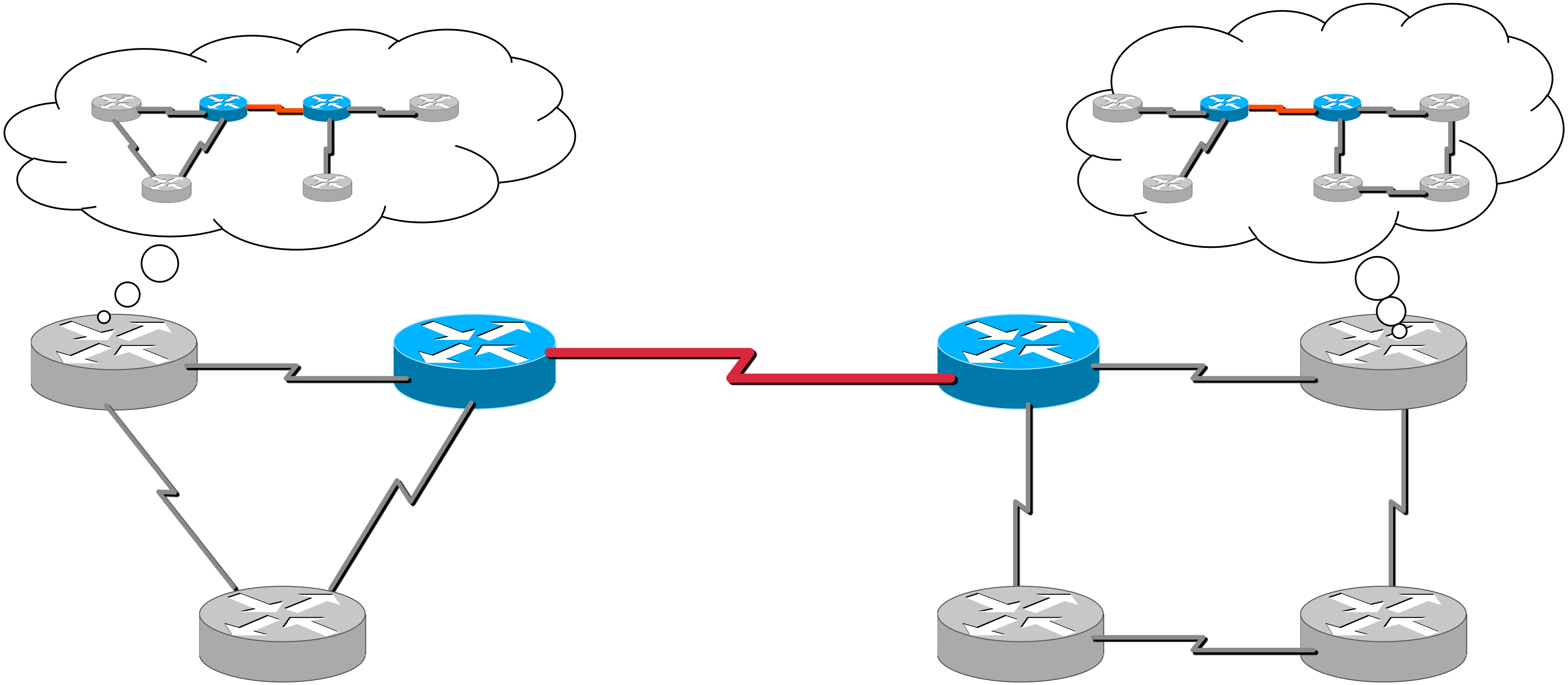
- Both routers tell all other routers about all local adjacencies by flooding link state advertisements (LSAs)
- Both routers now see their own IDs listed in the periodically sent Hello packets



These are so-called "Router LSAs". Other LSA types will be explained soon...

Database Inconsistency

- When connecting two networks, LSA flooding only distributes information of the local links of the involved neighbors (!)



Healing Inconsistency

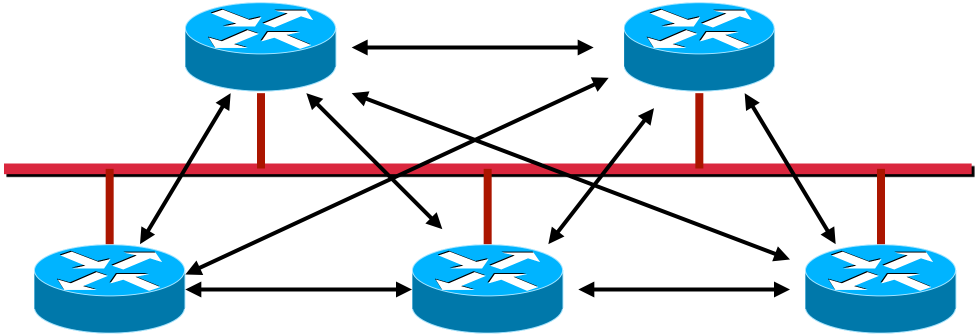
- **Every router sends its LSAs every 30 minutes (!)**
  - Heals but long time of routing table / topology table inconsistence when combining a former split area of a OSPF domain
- **Triggering database synchronization between any two routers in the network**
  - In order to avoid long time of inconsistence
  - So whenever a router is informed by a Router-LSA about some changes in the network this router additionally will do a database synchronization with the router from which the Router-LSA was received
  - Database description packets will help to reduce traffic to the necessary minimum

Agenda

- L2 versus L3 Switching
- IP Protocol, IP Addressing
- IP Forwarding
- ARP and ICMP
- <u>IP Routing</u>
  - Introduction
  - OSPF Basics
  - OSPF Communication Procedures (Router LSA)
  - LSA Broadcast Handling (Flooding)
  - OSPF Splitted Area
  - <u>Broadcast Networks (Network LSA)</u>
- First Hop Redundancy

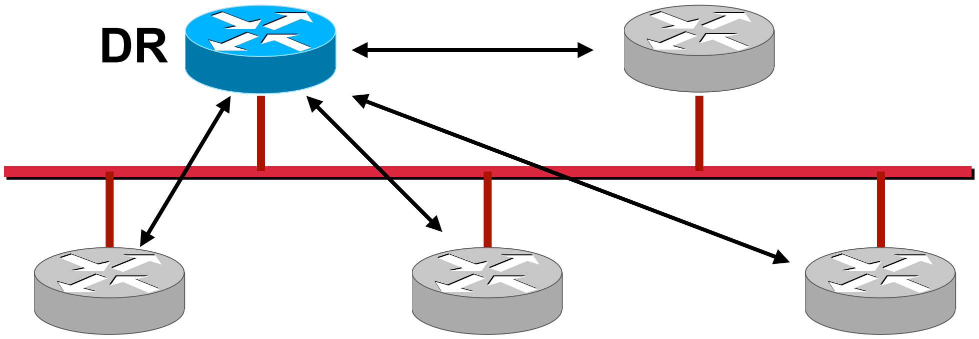
Broadcast Multi-Access Media (1)

- **When several OSPF routers have access to the same Ethernet segment they would create n(n-1)/2 adjacencies**
- **Furthermore, SPF algorithm requires to represent a fully meshed network as tree**



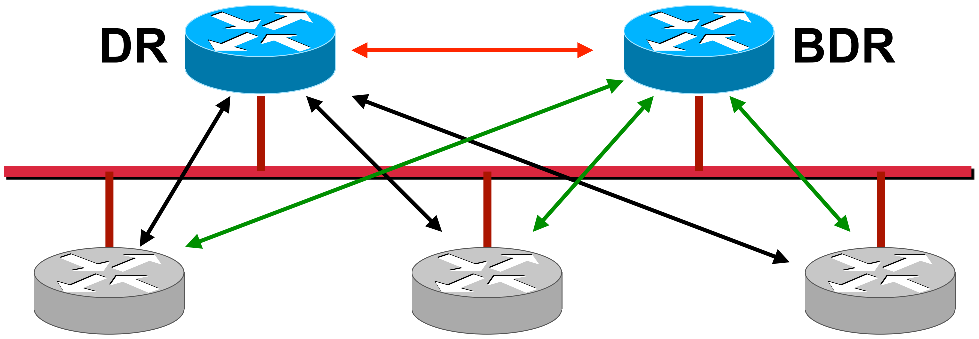
Broadcast Multi-Access Media (2)

- **Solution: Elect one "Designated Router" (DR) to represent the whole LAN segment**
  - Election uses the Hello protocol
- **DR sends Network LSA**
  - List of all local routers
  - Ensures that every router on the link has the same topology database
  - Also contains subnet mask (!)
- **Each other router establishes an adjacency only to the DR**
  - Using "All DR" multicast address 224.0.0.6



Broadcast Multi-Access Media (3)

- **Only the DR will send LSAs to the rest of the network**
- **For backup purposes also a Backup DR is elected (BDR)**
  - All routers also establish adjacencies to the BDR
  - BDR itself also establishes adjacency to DR



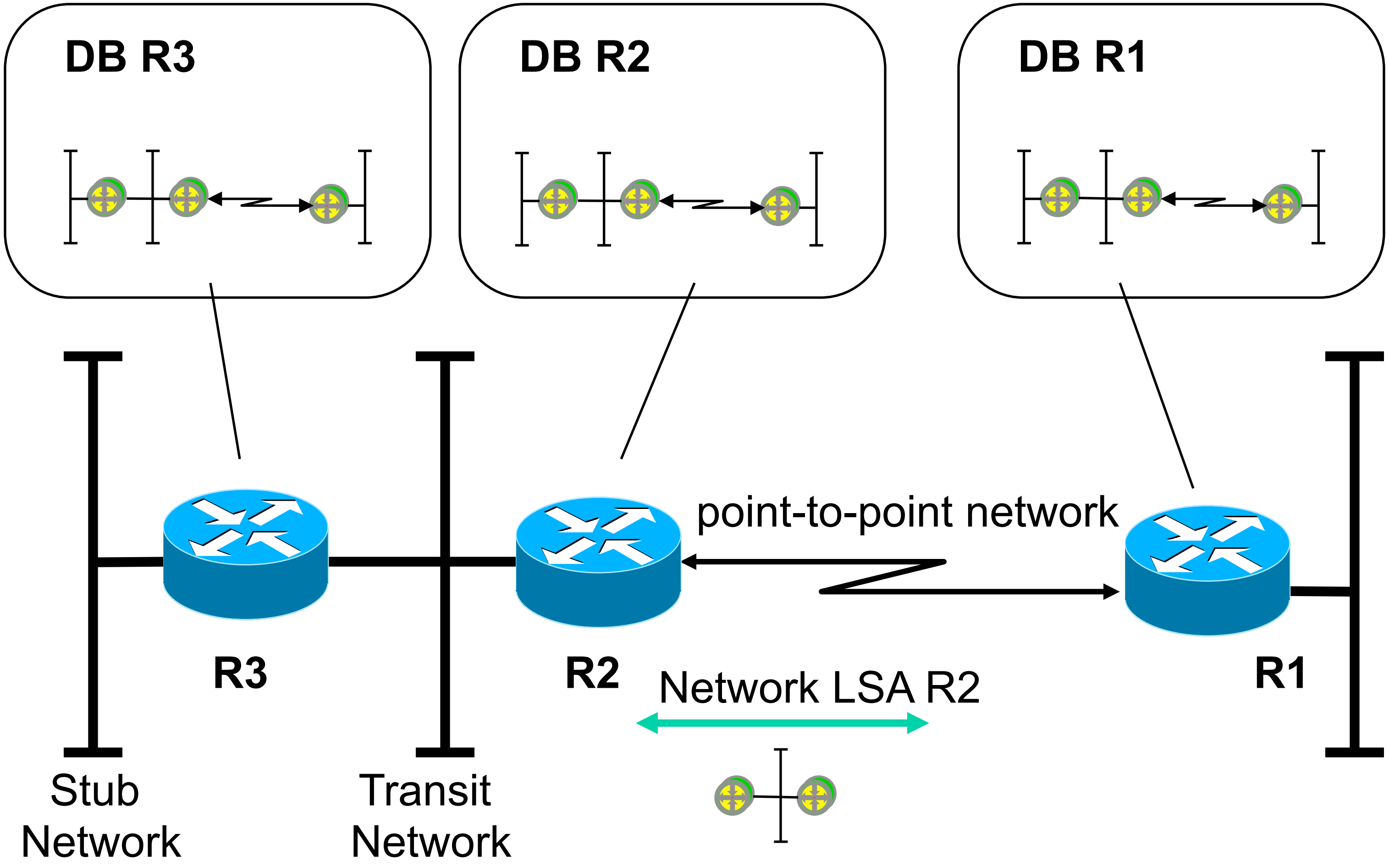
DR/BDR Election Process

- **Election process starts if no DR/BDR listed in the hello packets during the init state (i. e. when two routers begin to establish an adjacency)**
  - Note: if already one DR/BDR chosen, any new router in the LAN would not change anything!
  - Therefore, the power-on order of routers is critical !!!
- **Always configure loopback interface in order to "name" your routers**
  - Loopback interface never goes down
  - Ensures stability
  - Simple to manage

DR, Router LSA, Network LSA

- **Designated Router (DR) is responsible**
  - For maintaining neighbourhood relationship via virtual point-to-point links using the already known mechanism
    - DB-Description, LS-Request LS-Update, LS-Acknowledgement, Hello, etc.
- **Router-LSA implicitly describes**
  - These virtual point-to-point links by specifying such a network as transit-network
    - Remark: Stub-network is a LAN network where no OSPF router is behind
- **To inform all other routers of domain about such a special topology situation**
  - DR is additionally responsible for emitting Network LSAs
- **Network LSA describes**
  - Which routers are members of the corresponding broadcast network

OSPF Network LSA R2



Designated Router R2 notifies other nodes about the multi-access network using Network-LSA (transport mechanism are LS-Update packets hop-by-hop

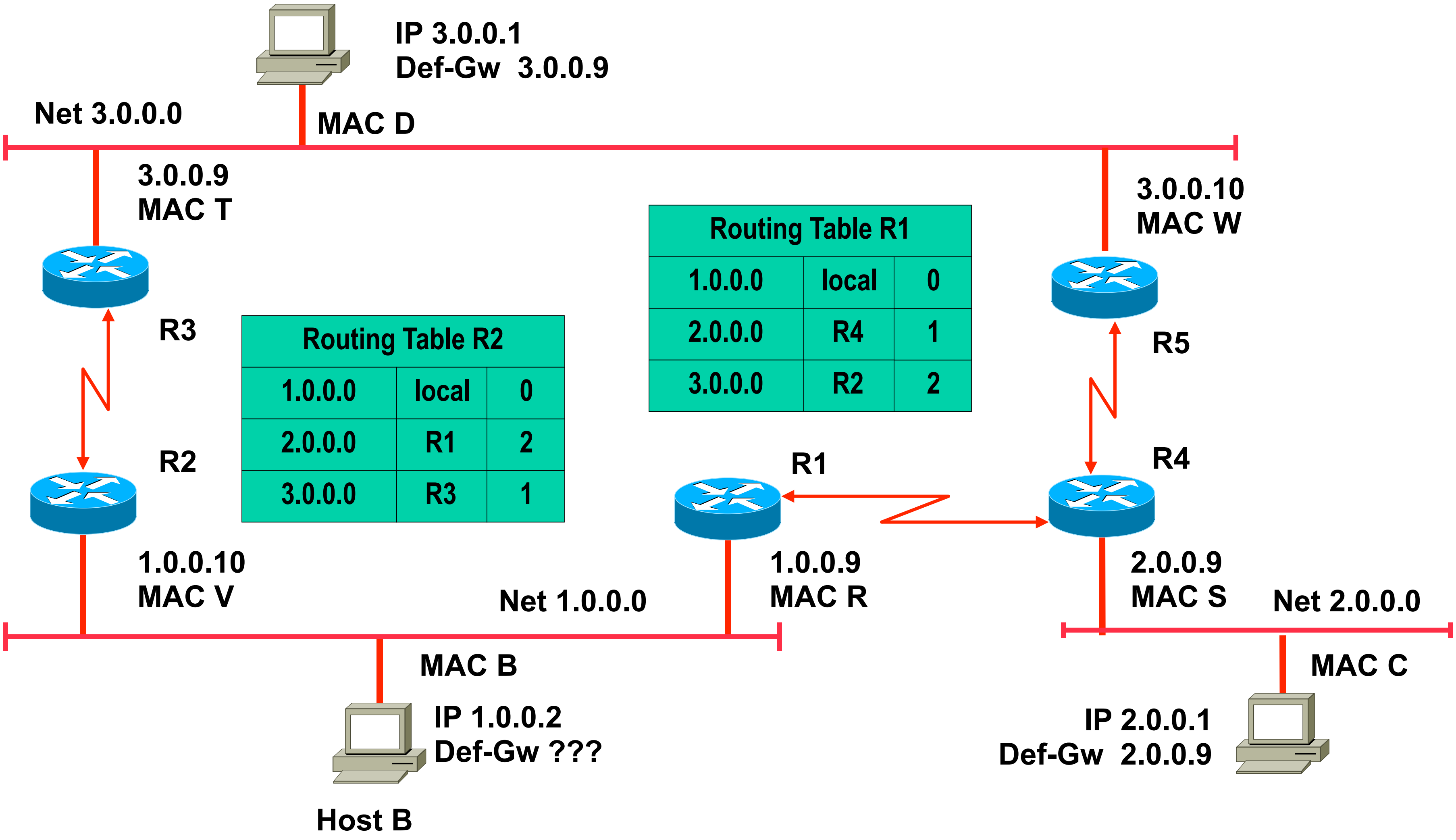
Details: OSPF Multicast Usage

- **OSPF uses dedicated IP multicast addresses for exchanging routing messages**
  - 224.0.0.5 (“All OSPF Routers")
  - 224.0.0.6 (“All Designated Routers")
- **224.0.0.5 is used as destination address**
  - By all routers for Hello-messages
    - DR and BR determination at start-up
    - link state supervision
  - By DR router for messages towards all non-DR routers
    - LS-Update, LS-Acknowledgement
- **224.0.0.6 is used as destination address**
  - By all non-DR routers for messages towards the DR
    - LS-Update, LS-Request, LS-Acknowledgement and database description messages

Agenda

- L2 versus L3 Switching
- IP Protocol, IP Addressing
- IP Forwarding
- ARP and ICMP
- IP Routing
- First Hop Redundancy

First L3 Hop?



IP 3.0.0.1
Def-Gw 3.0.0.9

Net 3.0.0.0

MAC D

3.0.0.9
MAC T

3.0.0.10
MAC W

R3

R5

Routing Table R1		
1.0.0.0	local	0
2.0.0.0	R4	1
3.0.0.0	R2	2

Routing Table R2		
1.0.0.0	local	0
2.0.0.0	R1	2
3.0.0.0	R3	1

R2

R1

R4

1.0.0.10
MAC V

Net 1.0.0.0

1.0.0.9
MAC R

2.0.0.9
MAC S

Net 2.0.0.0

MAC B

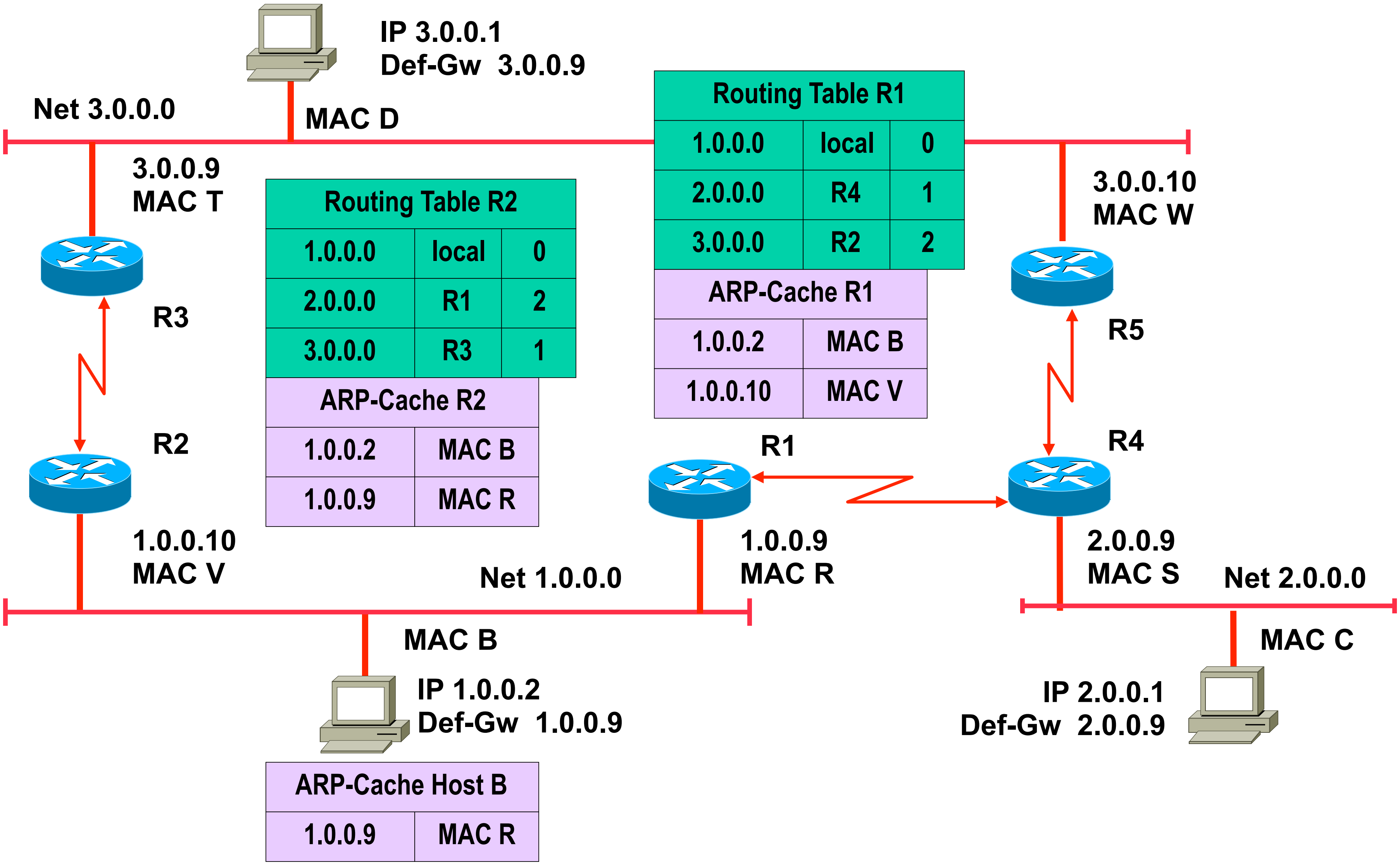
MAC C

IP 1.0.0.2
Def-Gw ???

Host B

IP 2.0.0.1
Def-Gw 2.0.0.9

Delivery 1.0.0.2 -> 3.0.0.1



Routing Table R1		
1.0.0.0	local	0
2.0.0.0	R4	1
3.0.0.0	R2	2

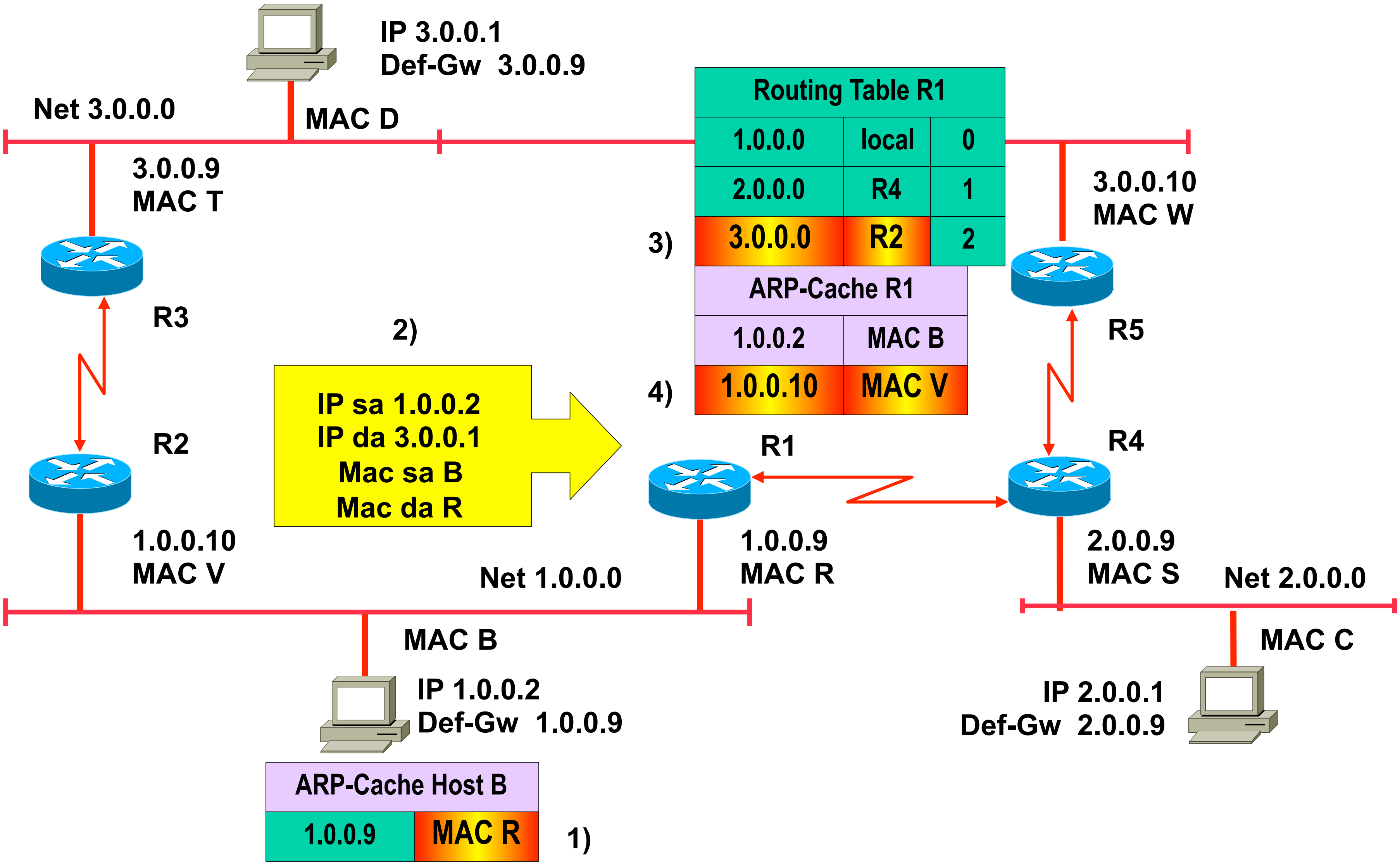
ARP-Cache R1	
1.0.0.2	MAC B
1.0.0.10	MAC V

Routing Table R2		
1.0.0.0	local	0
2.0.0.0	R1	2
3.0.0.0	R3	1

ARP-Cache R2	
1.0.0.2	MAC B
1.0.0.9	MAC R

ARP-Cache Host B	
1.0.0.9	MAC R

Delivery 1.0.0.2 -> 3.0.0.1



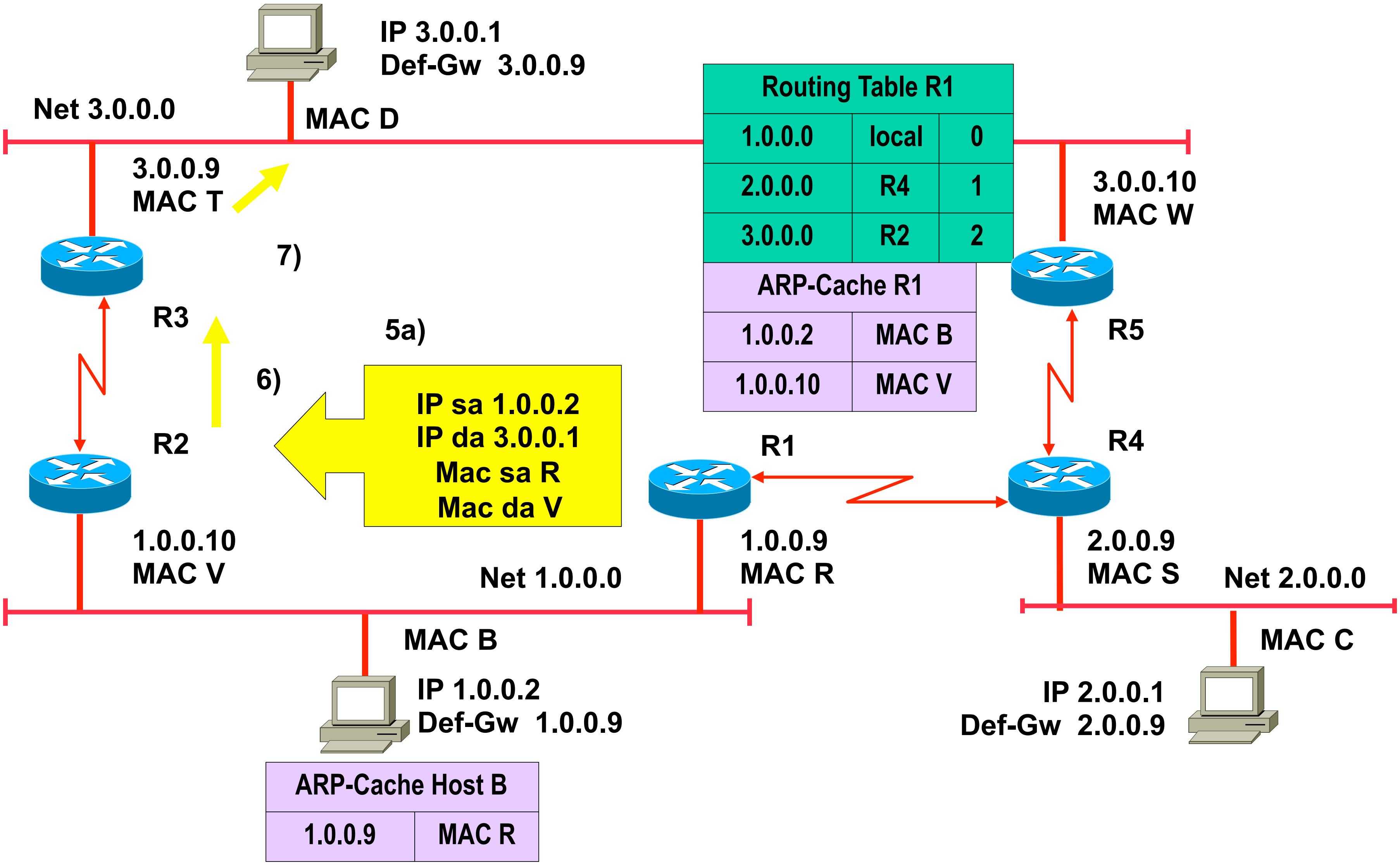
IP sa 1.0.0.2
IP da 3.0.0.1
Mac sa B
Mac da R

Routing Table R1		
1.0.0.0	local	0
2.0.0.0	R4	1
3.0.0.0	R2	2

ARP-Cache R1	
1.0.0.2	MAC B
1.0.0.10	MAC V

ARP-Cache Host B	
1.0.0.9	MAC R

Delivery 1.0.0.2 -> 3.0.0.1



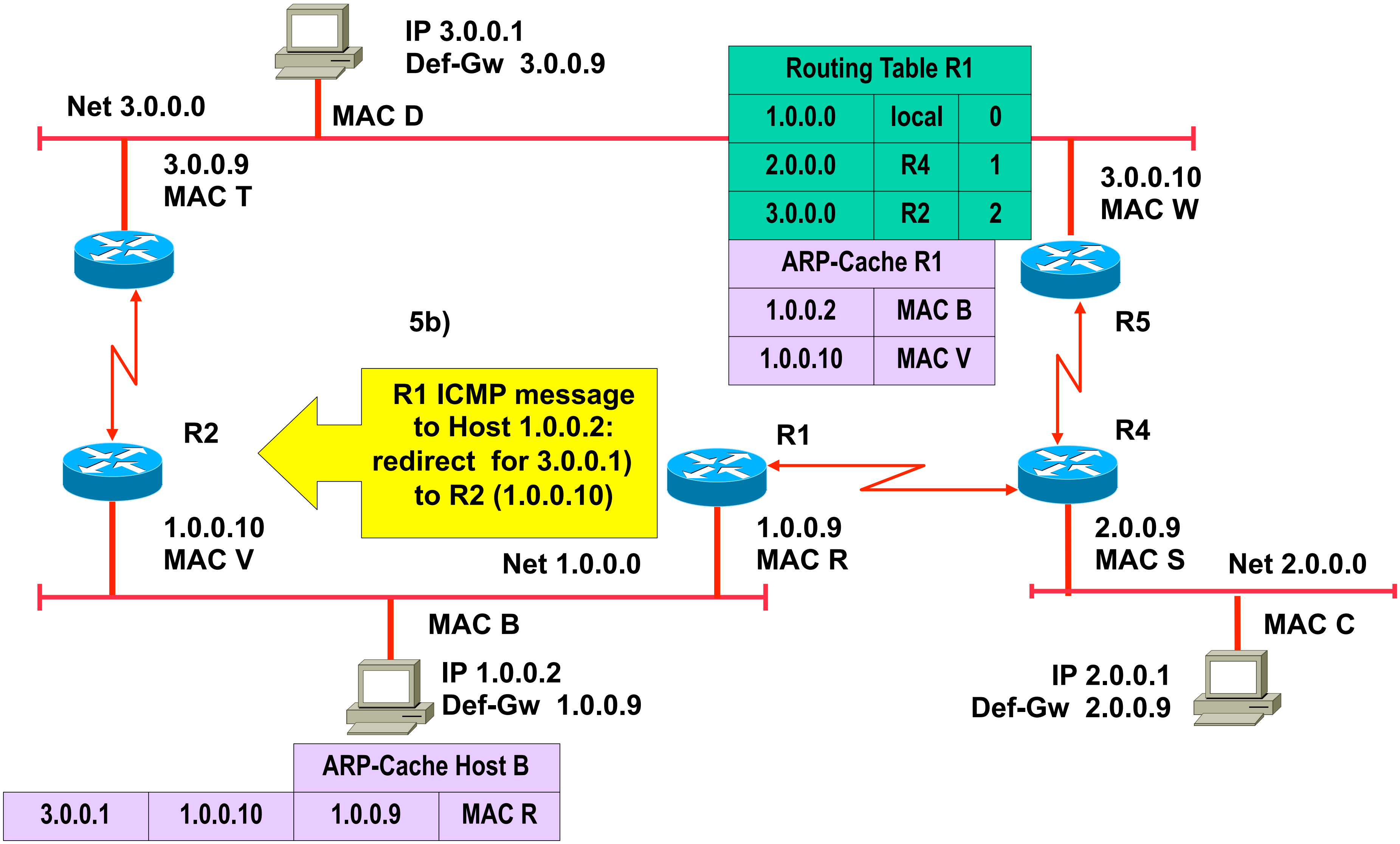
IP sa 1.0.0.2
IP da 3.0.0.1
Mac sa R
Mac da V

Routing Table R1		
1.0.0.0	local	0
2.0.0.0	R4	1
3.0.0.0	R2	2

ARP-Cache R1	
1.0.0.2	MAC B
1.0.0.10	MAC V

ARP-Cache Host B	
1.0.0.9	MAC R

ICMP redirect



IP 3.0.0.1
Def-Gw 3.0.0.9

MAC D

Net 3.0.0.0

3.0.0.9
MAC T

Routing Table R1		
1.0.0.0	local	0
2.0.0.0	R4	1
3.0.0.0	R2	2

ARP-Cache R1	
1.0.0.2	MAC B
1.0.0.10	MAC V

3.0.0.10
MAC W

R5

5b)

R2

R1 ICMP message
to Host 1.0.0.2:
redirect for 3.0.0.1)
to R2 (1.0.0.10)

R1

R4

1.0.0.10
MAC V

Net 1.0.0.0

1.0.0.9
MAC R

2.0.0.9
MAC S

Net 2.0.0.0

MAC B

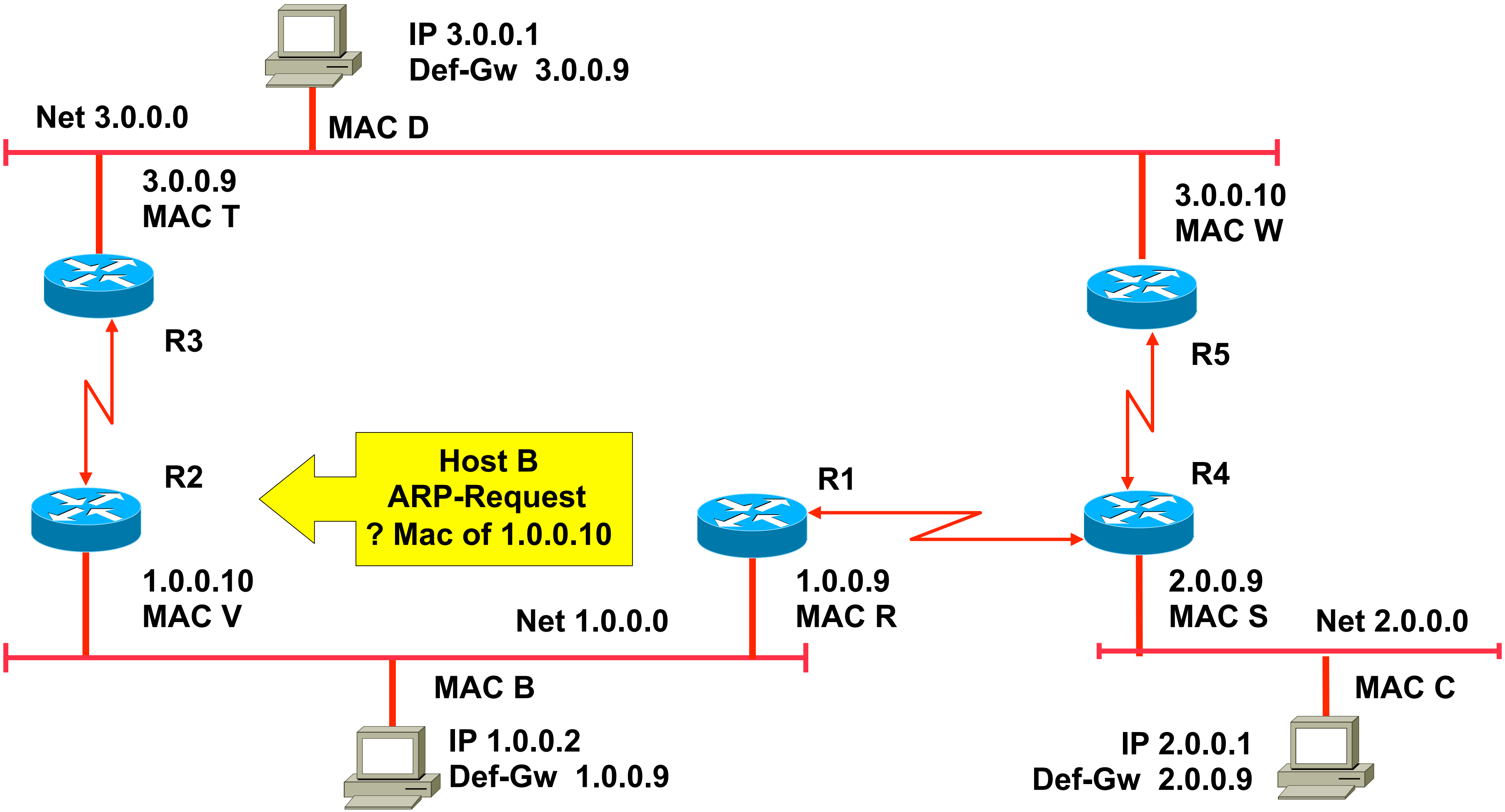
IP 1.0.0.2
Def-Gw 1.0.0.9

MAC C

IP 2.0.0.1
Def-Gw 2.0.0.9

		ARP-Cache Host B	
3.0.0.1	1.0.0.10	1.0.0.9	MAC R

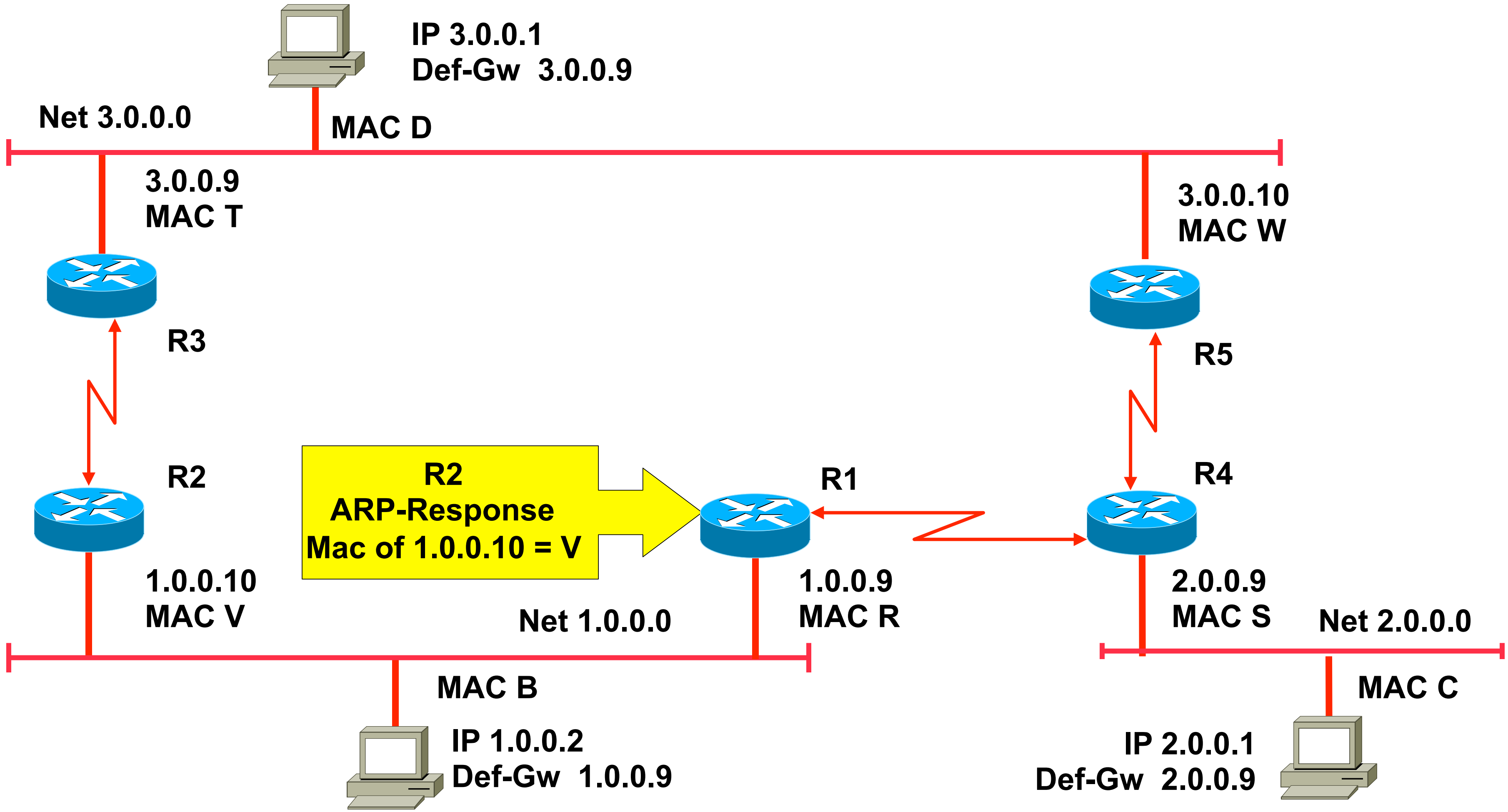
Delivery 1.0.0.2 -> 3.0.0.1



Host B
ARP-Request
? Mac of 1.0.0.10

		ARP-Cache Host B	
3.0.0.1	1.0.0.10	1.0.0.9	MAC R

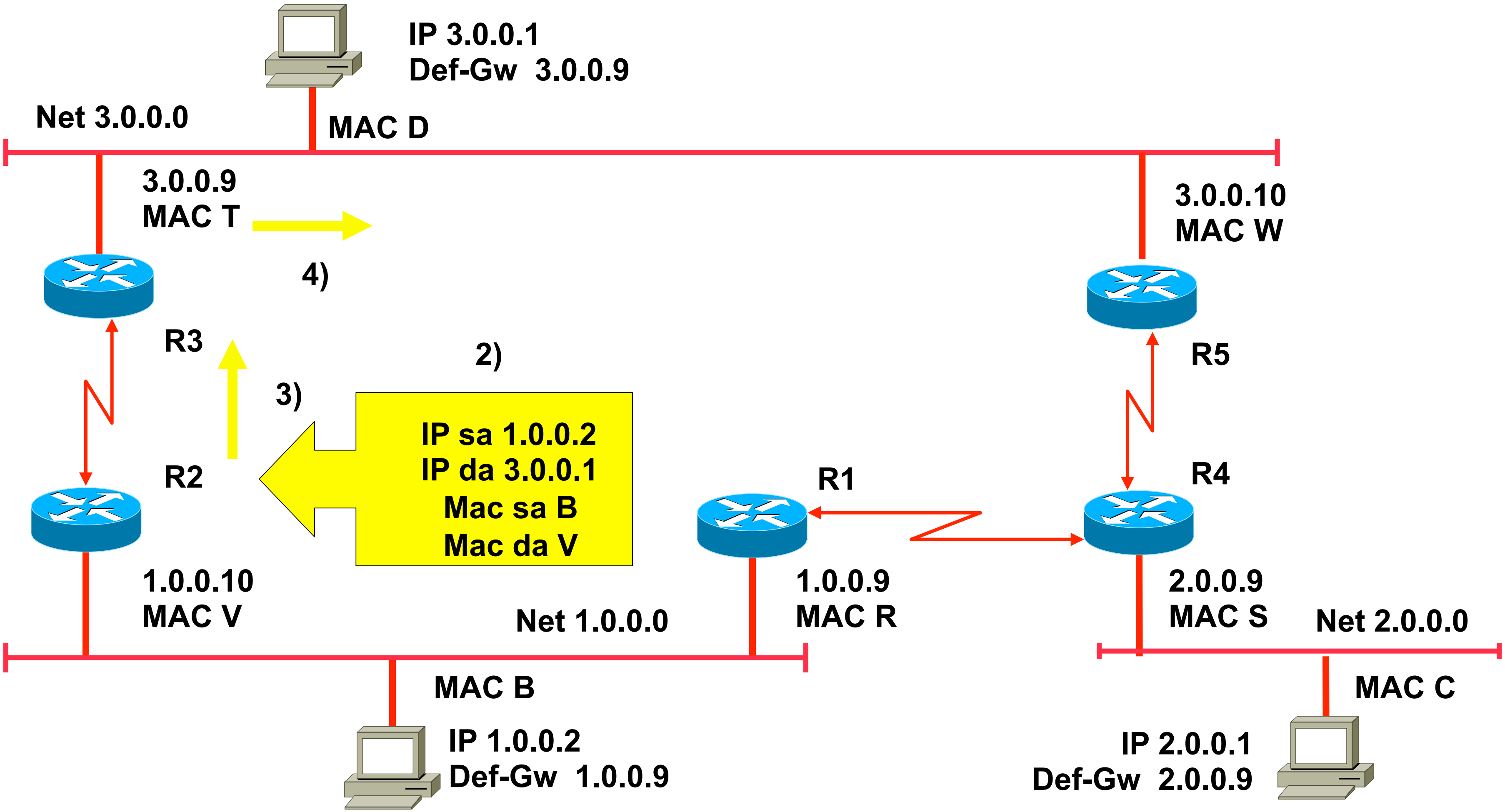
Delivery 1.0.0.2 -> 3.0.0.1



R2
ARP-Response
Mac of 1.0.0.10 = V

		ARP-Cache Host B	
3.0.0.1	1.0.0.10	1.0.0.9	MAC R
		1.0.0.10	MAC V

Next Packet 1.0.0.2 -> 3.0.0.1

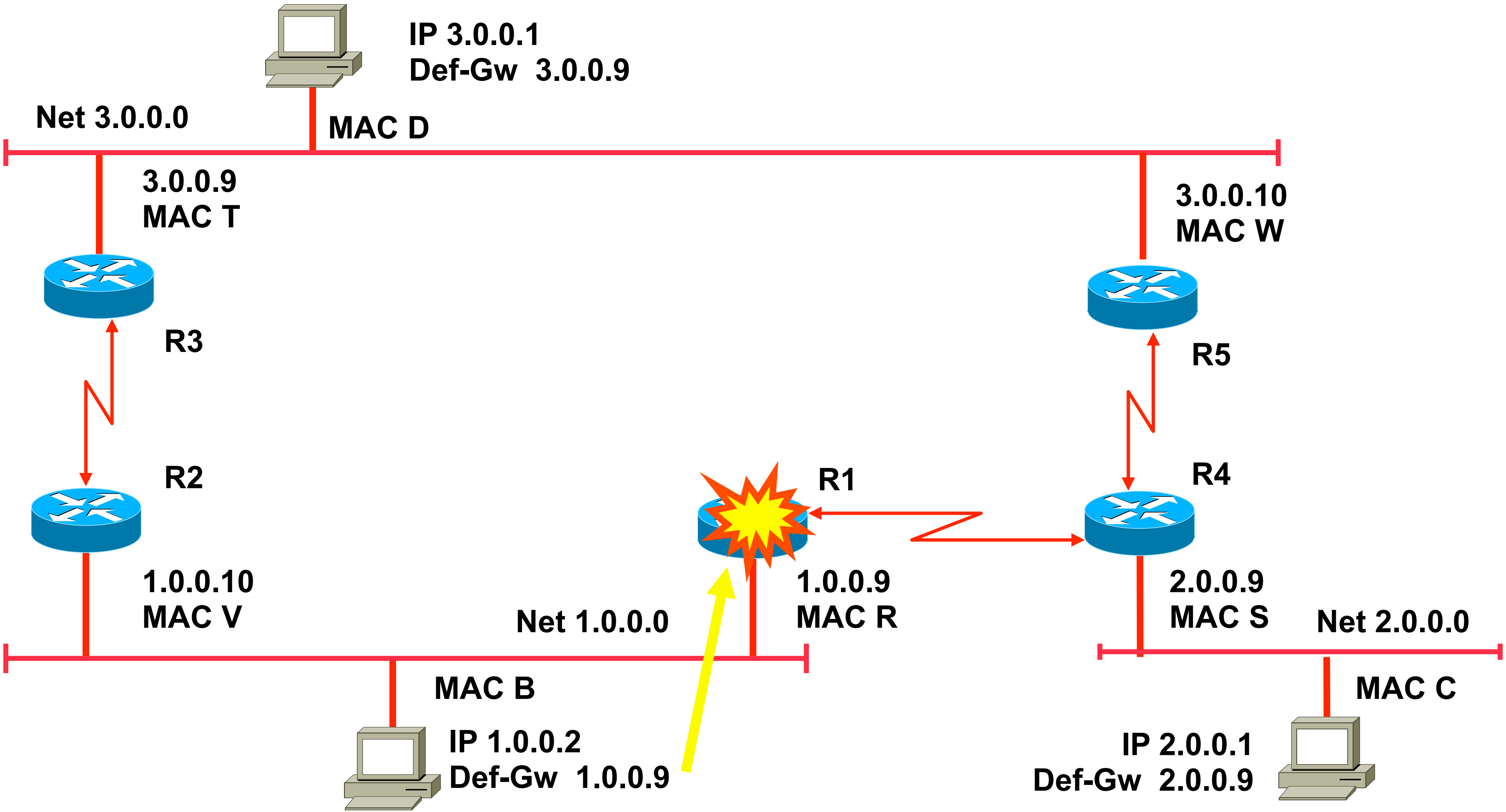


IP sa 1.0.0.2
IP da 3.0.0.1
Mac sa B
Mac da V

		ARP-Cache Host B	
3.0.0.1	1.0.0.10	1.0.0.9	MAC R
		1.0.0.10	MAC V

1)

Unavailability of R1:
System B losses connectivity to Net 3.0.0.0 or 2.0.0.0



		ARP-Cache Host B	
3.0.0.1	1.0.0.10	1.0.0.9	MAC R

Default gateway at
system B still points
to router 1.0.0.9 !!!

First Hop Redundancy 1

- **The problem:**
  - How can local routers be recognized by IP hosts?
  - Note: Normally IP host has limited view of topology
    - IP host knows to which IP subnet connected (own Net-ID)
    - IP host knows one “Default Gateway” to reach other IP networks
  - Static configuration of “Default Gateway” means:
    - Loss of the default router results in a catastrophic event, isolating all end-hosts that are unable to detect any alternate path that might be available
- **Two design philosophies:**
  - Solve the problem at the IP host level
    - OS of the IP host has to support an appropriate functionality
  - Solve the problem at the IP router level
    - OS of the IP host has to support the basic functionality only
      - That is static configuration of one “Default Gateway”
    - Appropriate functionality needed at the router

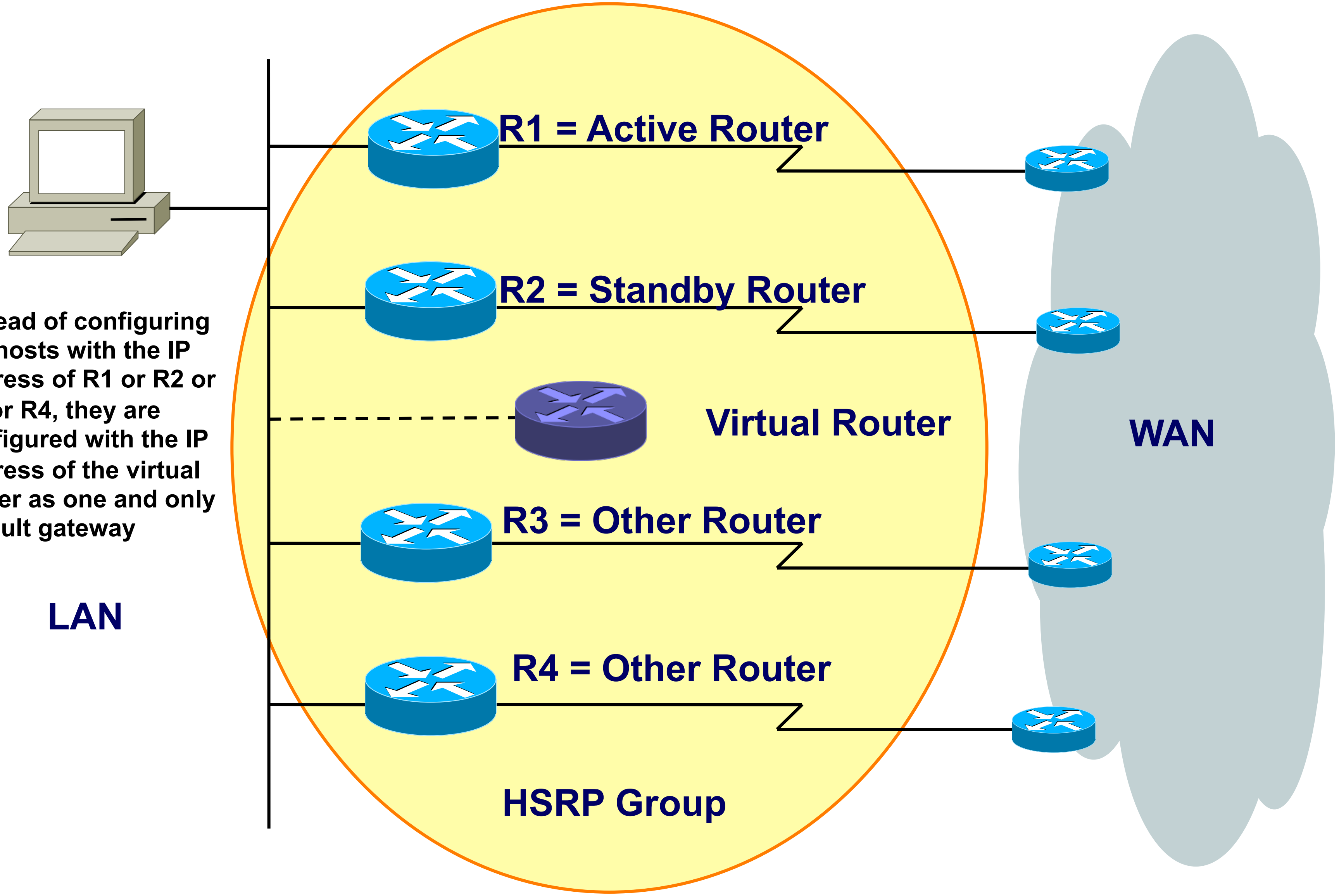
First Hop Redundancy 2

- **Methods for solving it at the IP host level:**
  - Proxy ARP
  - DHCP (Dynamic Host Configuration Protocol)
- **Methods for solving it at the IP router level:**
  - HSRP (Hot Standby Router Protocol)
    - Cisco proprietary
  - VRRP (Virtual Router Redundancy Protocol)
    - Same as HSRP but open RFC

HSRP – Hot Standby Router Protocol

- **HSRP (Hot Standby Router Protocol)**
  - Proprietary protocol invented by Cisco
  - RFC 2281 (Informational)
- **Basic idea: a set of routers pretend a single (virtual) router to the IP hosts on a LAN**
  - Active router
    - One router is responsible for forwarding the datagrams that hosts send to the virtual router
  - Standby router
    - If active router fails, the standby takes over the datagram forwarding duties of the active router
  - Conspiring routers form a so called HSRP group

HSRP Overview



Instead of configuring the hosts with the IP address of R1 or R2 or R3 or R4, they are configured with the IP address of the virtual router as one and only default gateway

HSRP Principles (1)

- **Basics:**
  - A group of routers forms a HSRP group
  - The group is represented by a virtual router
    - With a virtual IP address and virtual MAC address for that group
  - IP hosts are configured with the virtual IP address as default gateway
  - One router is elected by HSRP as the active router, one router is elected as the standby router of that group
    - HSRP messages are UDP messages to port 1985, addressed to IP multicast 224.0.0.2 using Ethernet multicast frames
      - Note HSRP version 1
  - Active router responds to ARP request directed to the virtual IP address with the virtual MAC address
  - Standby router supervises if the active router is alive
    - By listening to HSRP messages sent by the active

HSRP Principles (2)

- **Roles:**
  - Active router
    - Is responsible for the virtual IP address hence attracts any IP traffic which should leave the subnet
  - Standby router
    - Takes over the role of the active router in case the active router fails for the subnet
  - Additional HSRP member routers - Other
    - Other routers are neither active nor standby. They just monitor the messages of the current active and standby routers and transition into one of those roles if the current router fails for the subnet
  - Virtual router
    - The virtual router is not an actual router
    - Rather, it is a concept of the entire HSRP group acting as one virtual router for the IP hosts of the given subnet

HSRP Principles (3)

- **Roles (cont.):**
  - Active, Standby, Other defined by HSRP priority
  - Priority value can be configured
    - Default value is 100
  - The higher the better
    - Will become the active router after initialization
    - If priority is equal than the higher IP address decides
  - Preemption allows to give up the role of the active router
    - When a router with higher priority is reported by HSRP messages
  - Preemption happens
    - Either when the failed router comes back, a better router is activated or object tracking has changed priority

HSRP Principles (4)

- **Two basic failover scenarios:**
  - 1) Active router is not reachable via LAN
    - Standby router will take over active role
    - A new standby router is elected from the remaining routers of a HSRP group
    - Timing depends on HSRP hello message interval and hold-time
      - Default hello-time = 3 seconds, default hold-time = 10 seconds
      - Note HSRP version 1
  - 2) Active router losses connectivity either to a WAN interface or losses connectivity to a given IP route
    - Tracking will lower the priority of the active router
    - If preemption is configured on all routers the standby router will take over
    - Remember: Preemption allows another router to take over the role of the active router even if the current active router does not fail

HSRP Protocol Fields

- **Standby protocol runs on top of UDP (port 1985)**
  - IP packets are sent to IP multicast address 224.0.0.2 (HSRPv1) or 224.0.0.102 (HSRPv2) with a IP TTL = 1

0 4	8	16	31
Version	**Op Code**	**State**	**Hellotime**
Holdtime	**Priority**	**Group**	**Reserved**
Authentication Data			
Authentication Data			
Virtual IP Address			

- **Version**: Version of the HSRP messages
- **Op code**: 4 types

  Hello: Indicates that a router is running and is capable of becoming the active or standby router

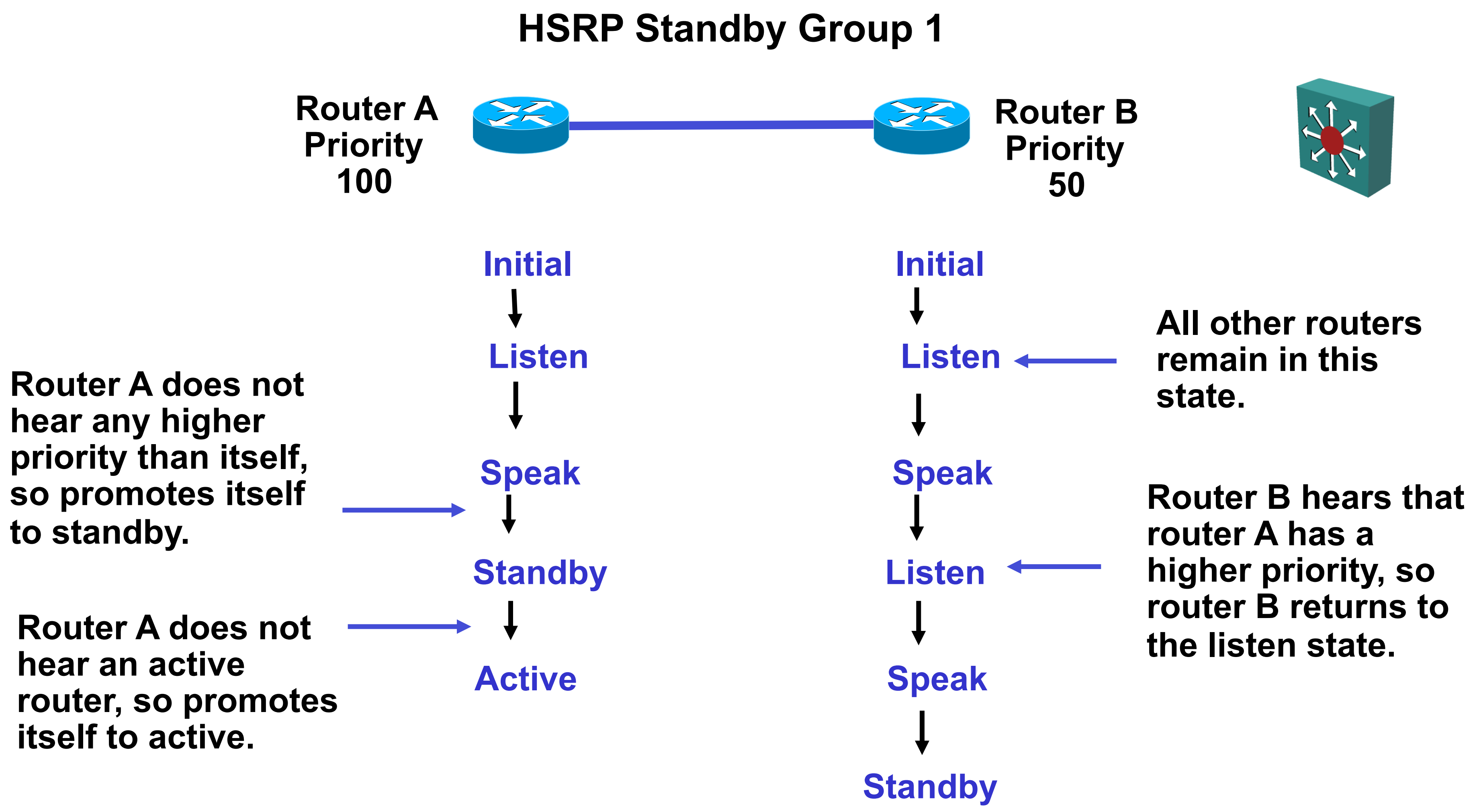
  Coup: When a router wishes to become the active router

  Resign: When a router no longer wishes to be the active router

  Advertise: Announce state of own HSRP interface

- **States**: Initial, learn, listen, speak, standby, active
- **Hellotime**: Contains the period between the hello messages that the router sends
- **Holdtime**: Amount of time the current hello message is valid
- **Priority**: Compares priorities of 2 different routers
- **Group**: Identifies standby group (0...255)
- **Authentication data**: Cleartext or MD5 signed hash

HSRP States Details



HSRP Standby Group 1

Router A
Priority
100

Router B
Priority
50

Initial

Listen

Speak

Standby

Active

Router A does not hear any higher priority than itself, so promotes itself to standby.

Router A does not hear an active router, so promotes itself to active.

Initial

Listen

Speak

Listen

Speak

Standby

All other routers remain in this state.

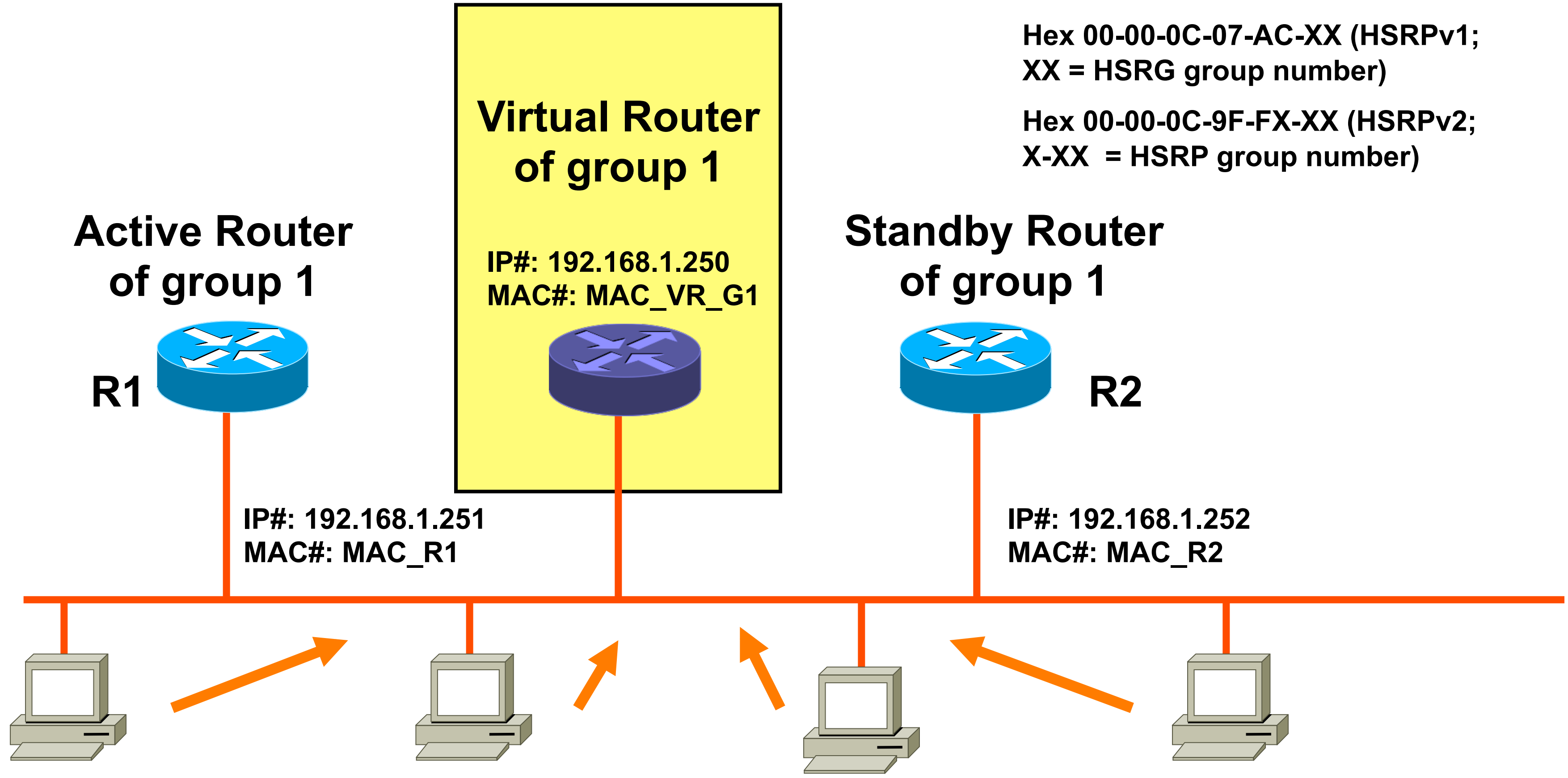
Router B hears that router A has a higher priority, so router B returns to the listen state.

HSRP: Real and Virtual IP Addresses / MAC Addresses

Default values on Cisco routers for virtual MAC address:

Hex 00-00-0C-07-AC-XX (HSRPv1; XX = HSRG group number)

Hex 00-00-0C-9F-FX-XX (HSRPv2; X-XX = HSRP group number)



Virtual Router of group 1

IP#: 192.168.1.250
MAC#: MAC_VR_G1

Active Router of group 1

R1

IP#: 192.168.1.251
MAC#: MAC_R1

Standby Router of group 1

R2

IP#: 192.168.1.252
MAC#: MAC_R2

Def-GW: 192.168.1.250

Def-GW: 192.168.1.250

Def-GW: 192.168.1.250

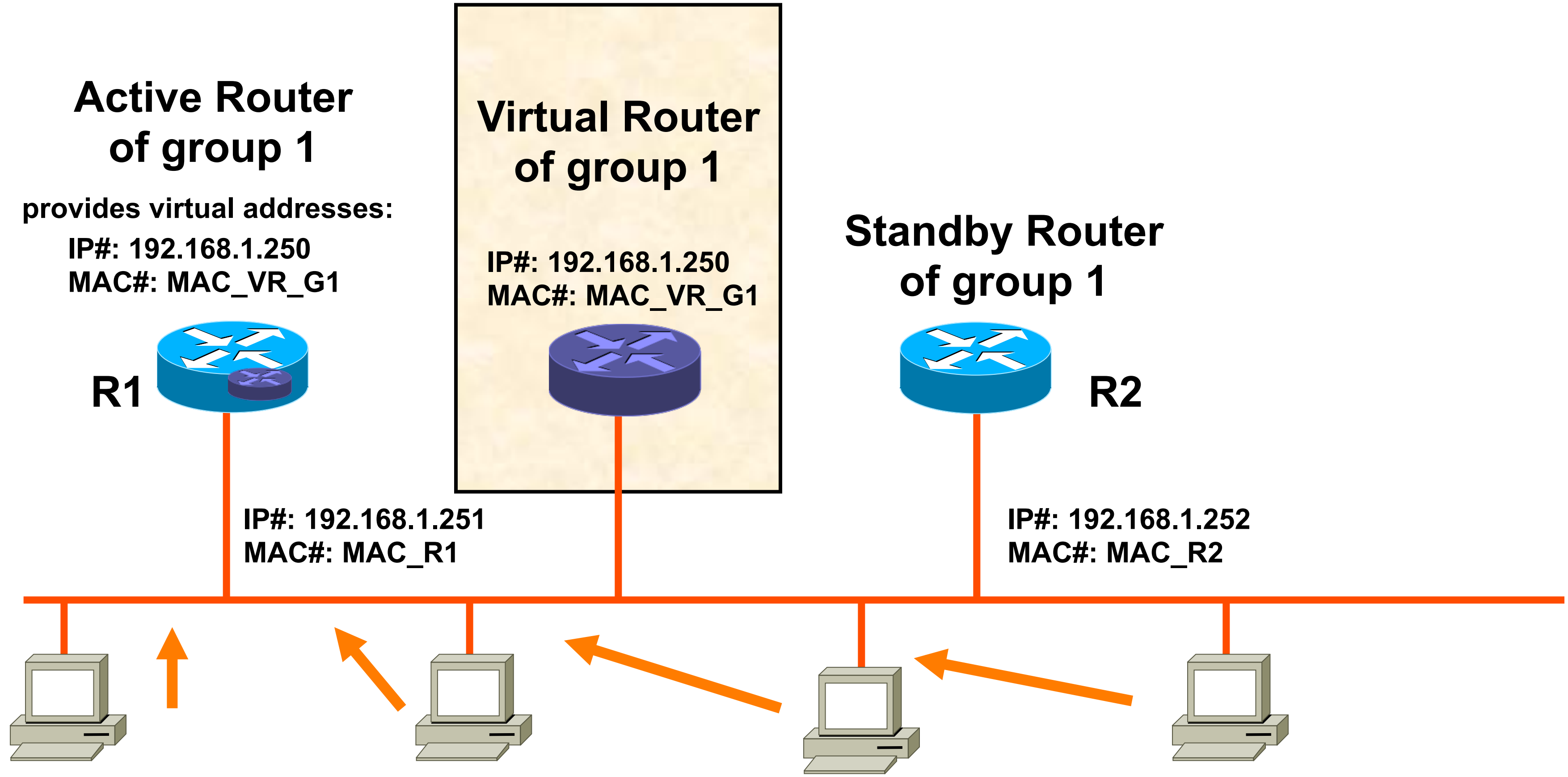
Def-GW: 192.168.1.250

IP address of virtual router of HSRP group1

ARP-Cache PCs	
192.168.1.250	MAC_VR_G1

MAC address of virtual router of HSRP group1

HSRP in Action (1)



Active Router of group 1

provides virtual addresses:
IP#: 192.168.1.250
MAC#: MAC_VR_G1

R1

Virtual Router of group 1

IP#: 192.168.1.250
MAC#: MAC_VR_G1

Standby Router of group 1

R2

IP#: 192.168.1.251
MAC#: MAC_R1

IP#: 192.168.1.252
MAC#: MAC_R2

Def-GW: 192.168.1.250

Def-GW: 192.168.1.250

Def-GW: 192.168.1.250

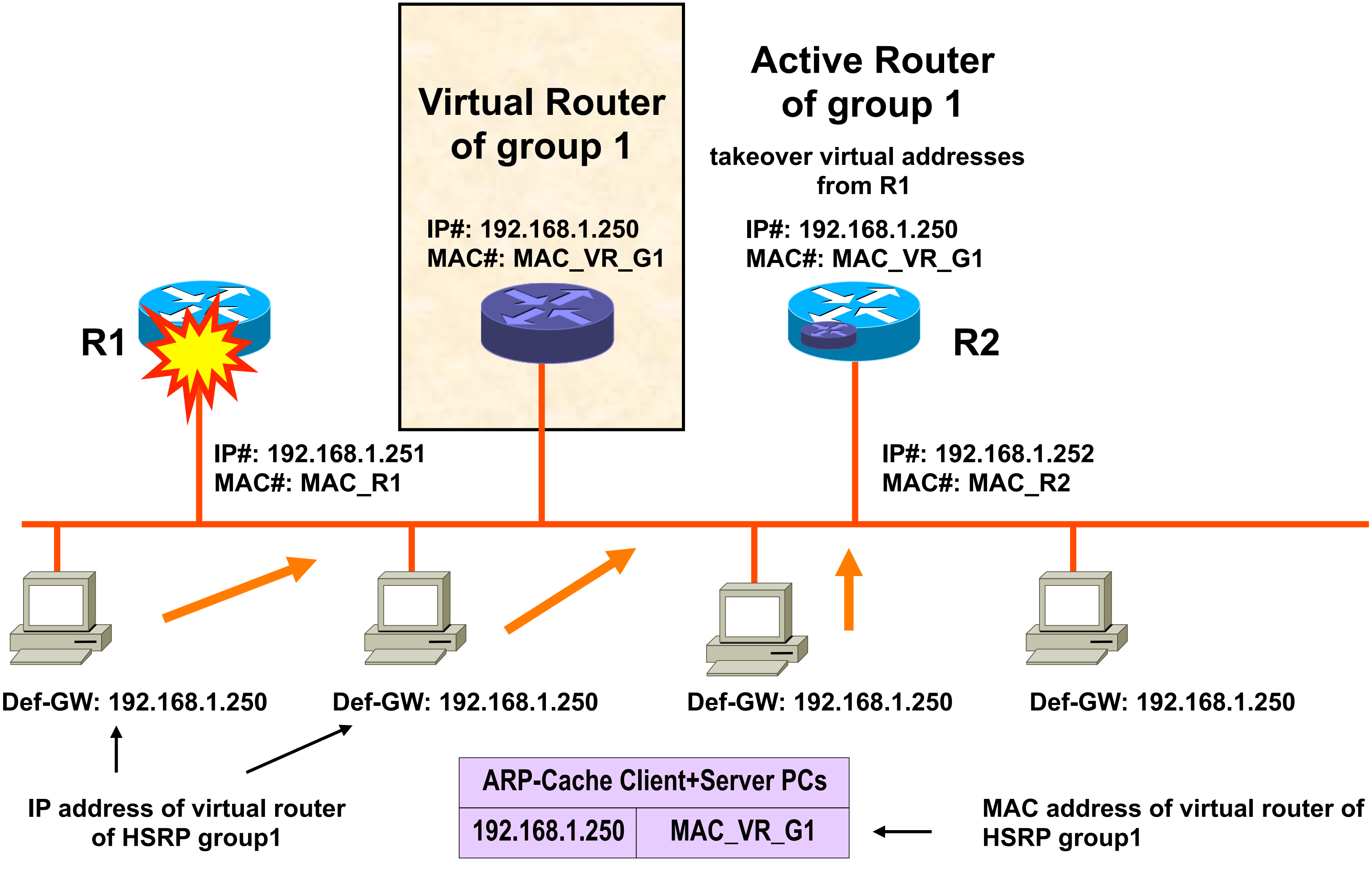
Def-GW: 192.168.1.250

IP address of virtual router of HSRP group1

ARP-Cache PCs	
192.168.1.250	MAC_VR_G1

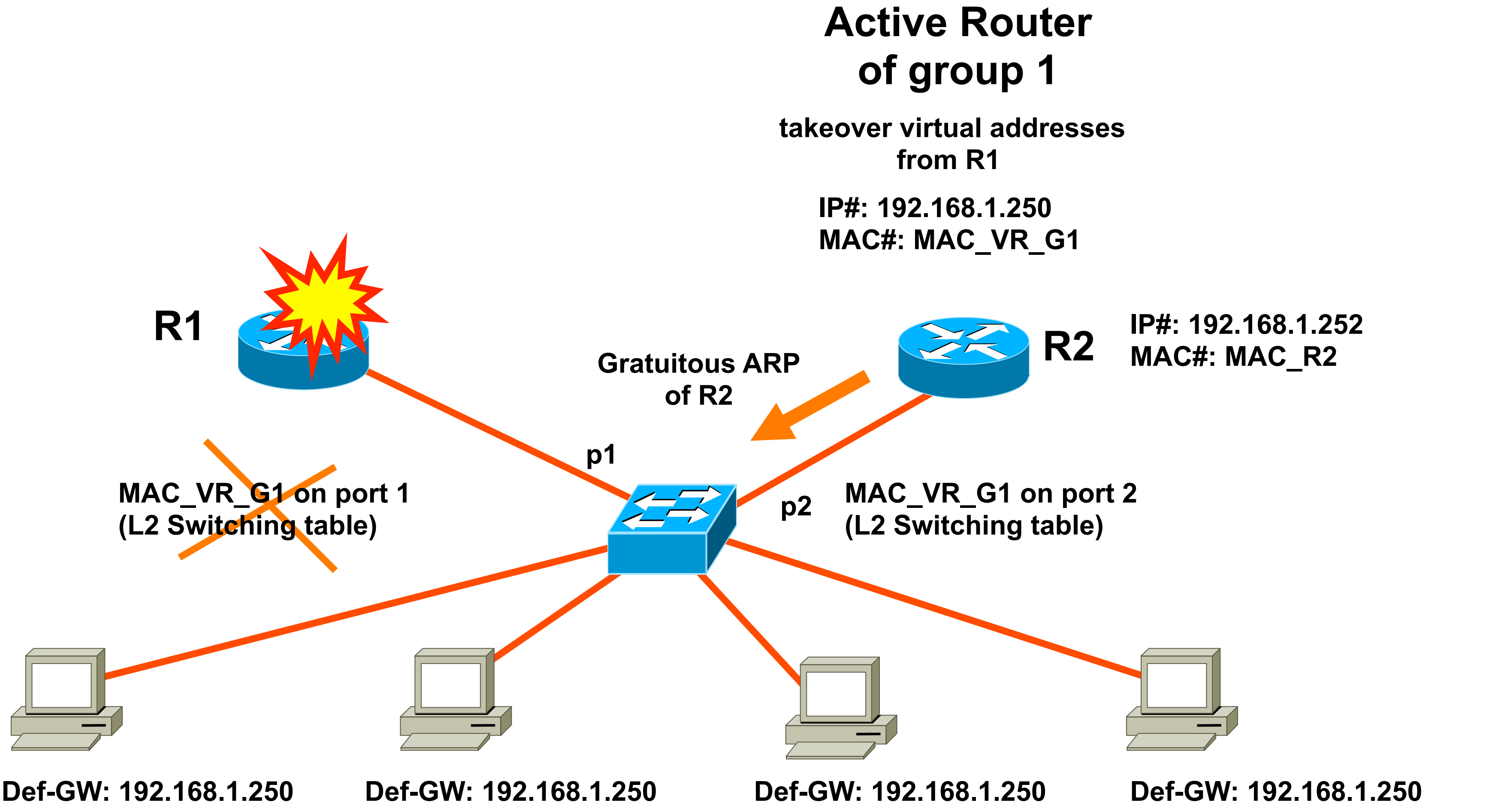
MAC address of virtual router of HSRP group1

HSRP in Action (2)



ARP-Cache Client+Server PCs	
192.168.1.250	MAC_VR_G1

HSRP – Gratuitous ARP



Active Router
of group 1

takeover virtual addresses
from R1

IP#: 192.168.1.250
MAC#: MAC_VR_G1

R1

R2

IP#: 192.168.1.252
MAC#: MAC_R2

Gratuitous ARP
of R2

p1

p2

~~MAC_VR_G1 on port 1~~
~~(L2 Switching table)~~

MAC_VR_G1 on port 2
(L2 Switching table)

Def-GW: 192.168.1.250

Def-GW: 192.168.1.250

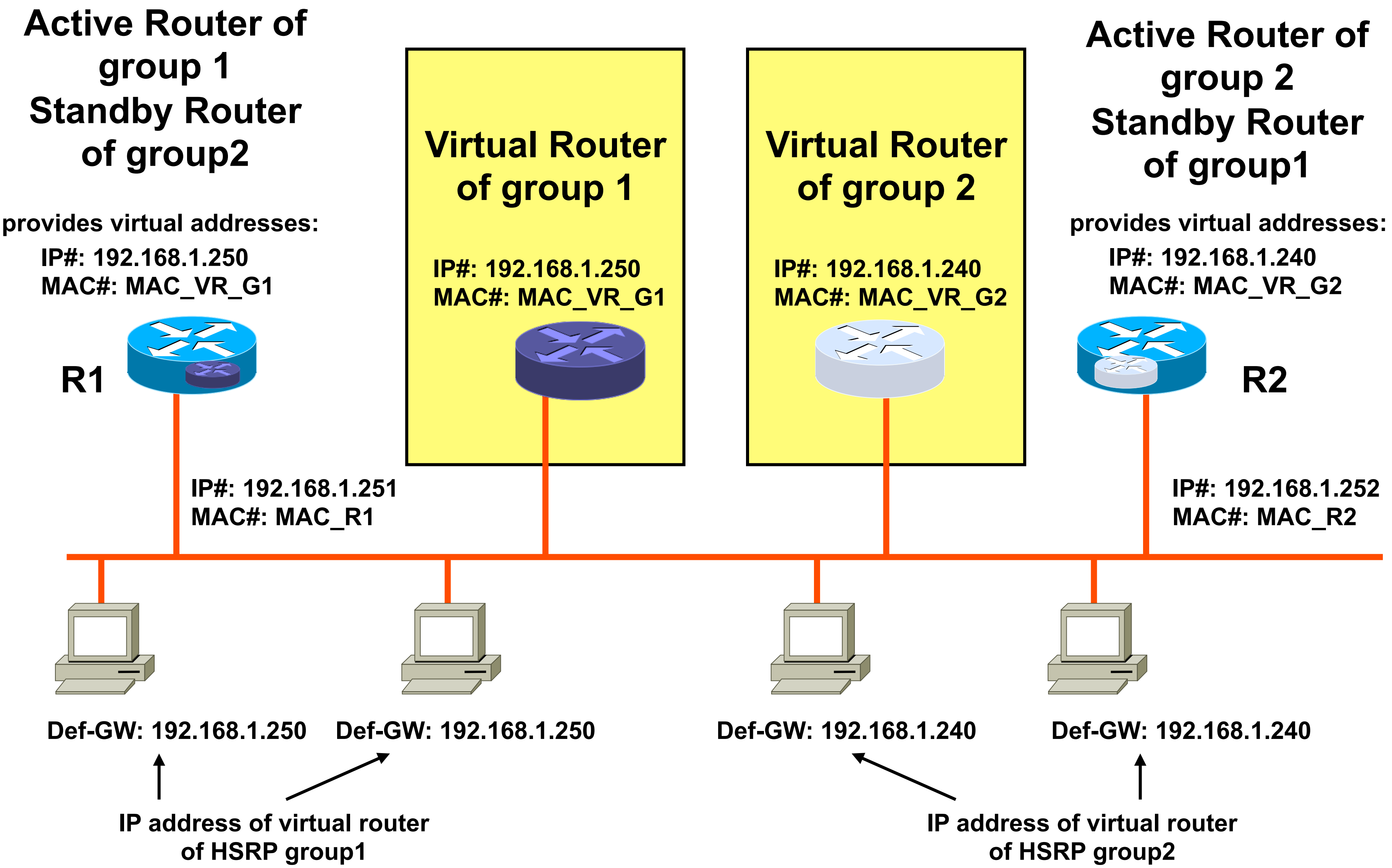
Def-GW: 192.168.1.250

Def-GW: 192.168.1.250

ARP-Cache Client+Server PCs	
192.168.1.250	MAC_VR_G1

← MAC address of virtual router of HSRP group1

HSRP Load Balancing



Active Router of group 1
Standby Router of group2

provides virtual addresses:
IP#: 192.168.1.250
MAC#: MAC_VR_G1

R1

IP#: 192.168.1.251
MAC#: MAC_R1

Virtual Router of group 1

IP#: 192.168.1.250
MAC#: MAC_VR_G1

Virtual Router of group 2

IP#: 192.168.1.240
MAC#: MAC_VR_G2

Active Router of group 2
Standby Router of group1

provides virtual addresses:
IP#: 192.168.1.240
MAC#: MAC_VR_G2

R2

IP#: 192.168.1.252
MAC#: MAC_R2

Def-GW: 192.168.1.250

Def-GW: 192.168.1.250

Def-GW: 192.168.1.240

Def-GW: 192.168.1.240

IP address of virtual router of HSRP group1

IP address of virtual router of HSRP group2