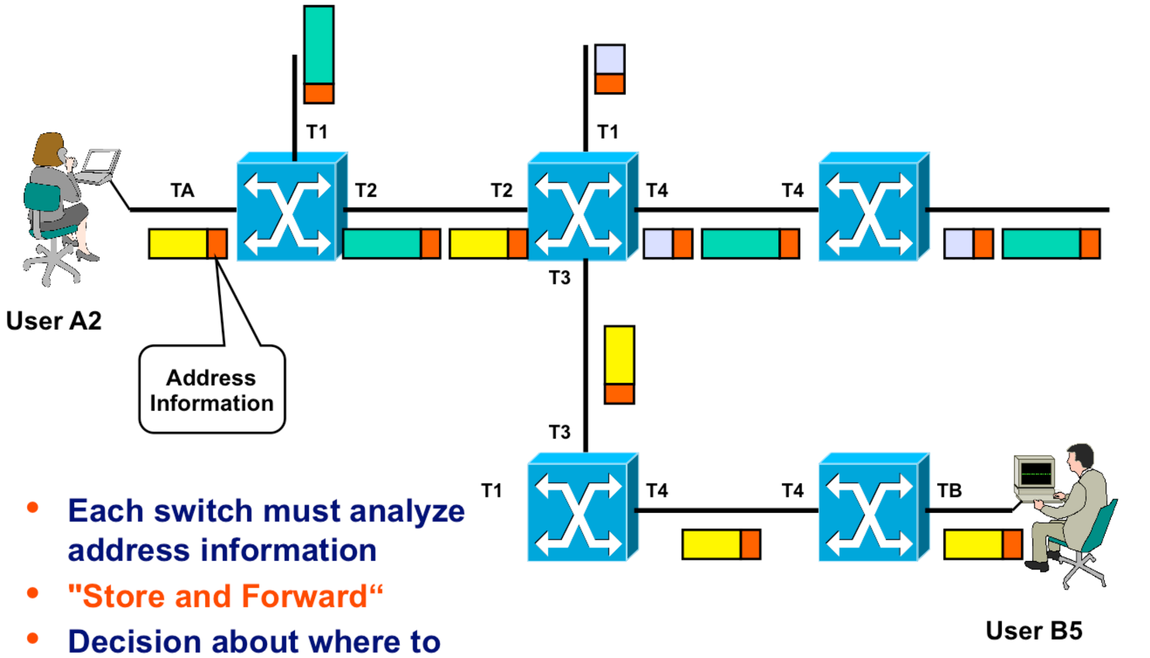
Primer IP Technology

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

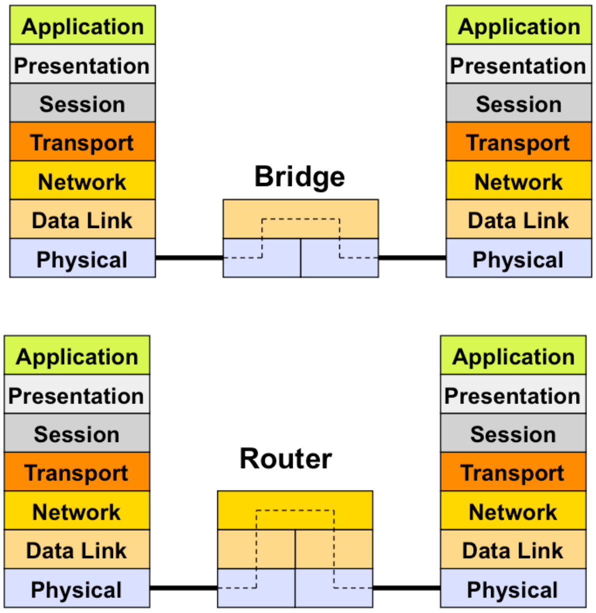
 IP Technology Primer, v6.4 3

In packet switching technology which is based on statistical time division multiplexing addresses are needed, remember there is no correlation between timeslot and destination.

Each switch must analyze the destination address of every data packet to be able to forward it according to some forwarding table.

In our example user A2 communicates with user B2 by the help of addresses.

Bridging (Ethernet Switching) versus IP Routing

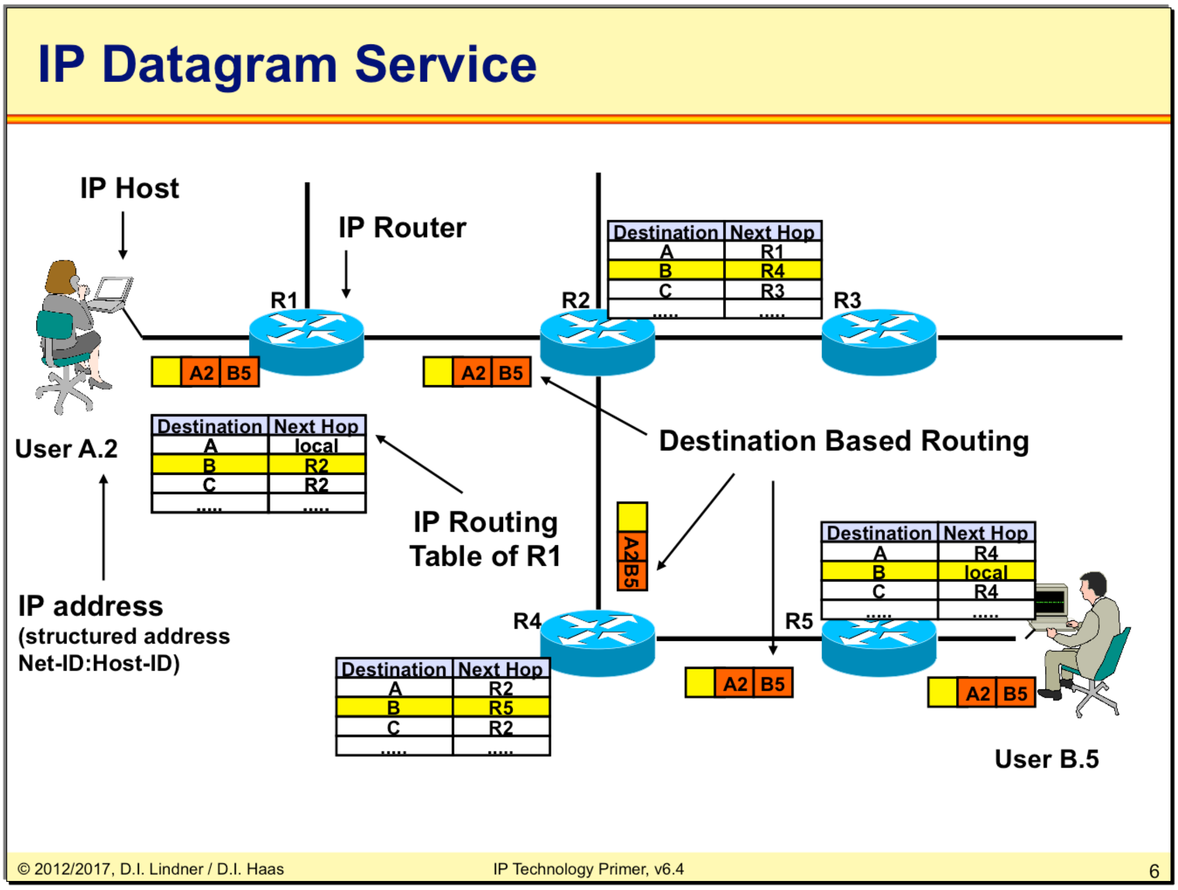


- **Bridging is**
  - Connectionless packet switching on OSI layer 2 using unique but unstructured MAC addresses without any topology information
  - Signpost in the MAC address table
- **Routing is**
  - Connectionless packet switching on OSI layer 3 using unique and structured addresses which contain 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 Technology (v6.4)



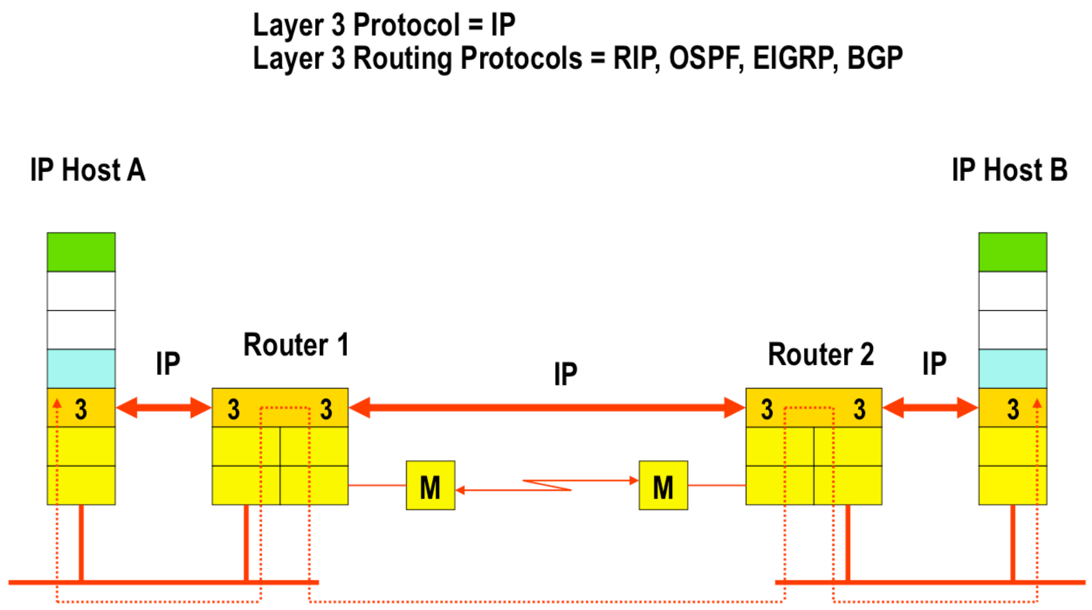
In the Datagram technology user A.2 sends out data packets destined for the user B.5. Each single datagram holds the information about sender and receiver address.

The datagram forwarding devices in our example routers hold a routing table in memory. In the routing table we find a correlation between the destination address of a data packet and the corresponding outgoing interface as well as the next hop router. So data packets are forwarded through the network on a hop by hop basis.

The routing tables can be set up either by manual configuration of the administrator or by the help of dynamic routing protocols like RIP, OSPF, IS-IS, etc. The use of dynamic routing protocols may lead to rerouting decisions in case of network failure and so packet overtaking may happen in these systems.

IP Technology (v6.4)

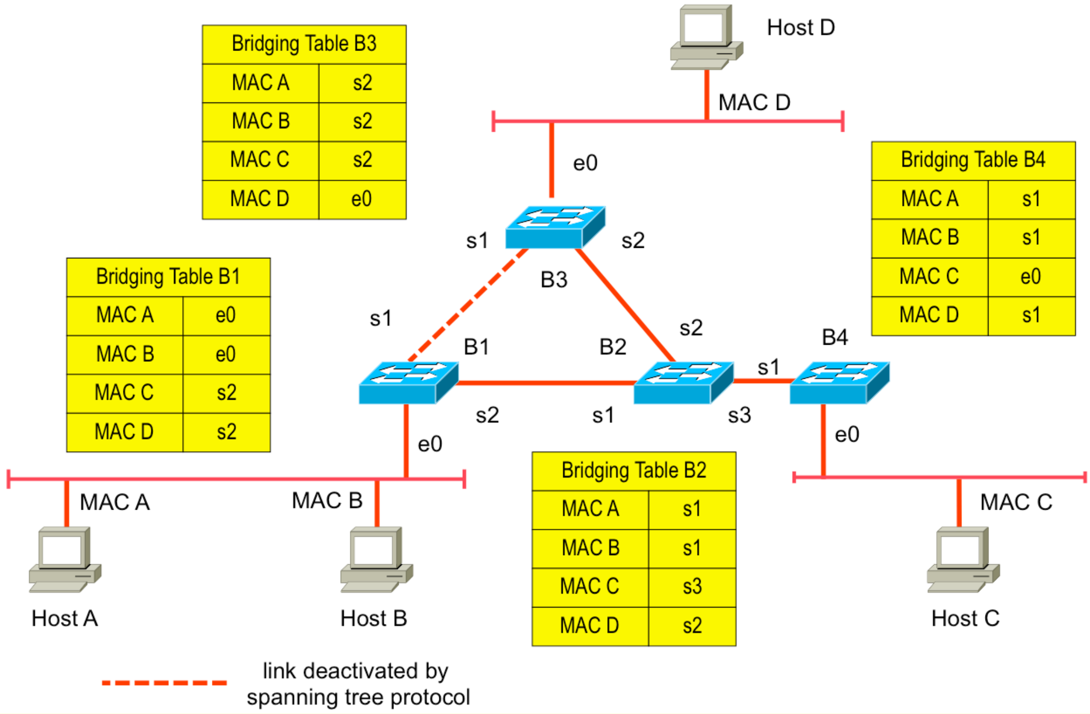
IP and OSI Network Layer 3



 IP Technology Primer, v6.4 7

IP Technology (v6.4)

Example Topology for Review Bridging



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

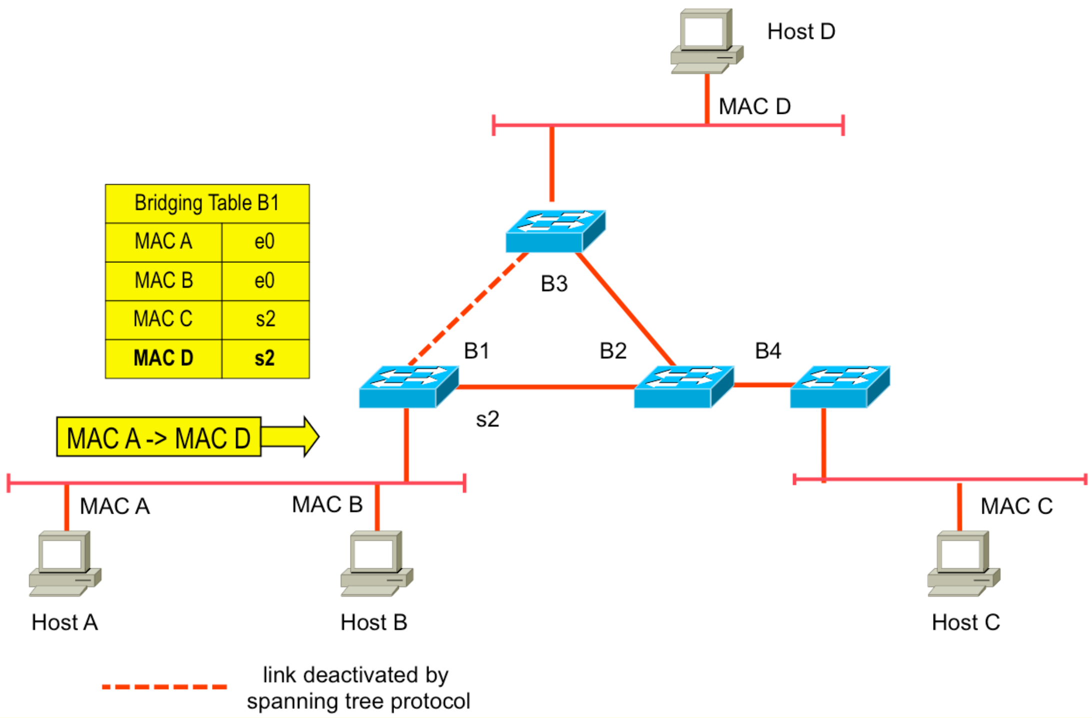
Bridging Table B4	
MAC A	s1
MAC B	s1
MAC C	e0
MAC D	s1

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

 IP Technology Primer, v6.4 8

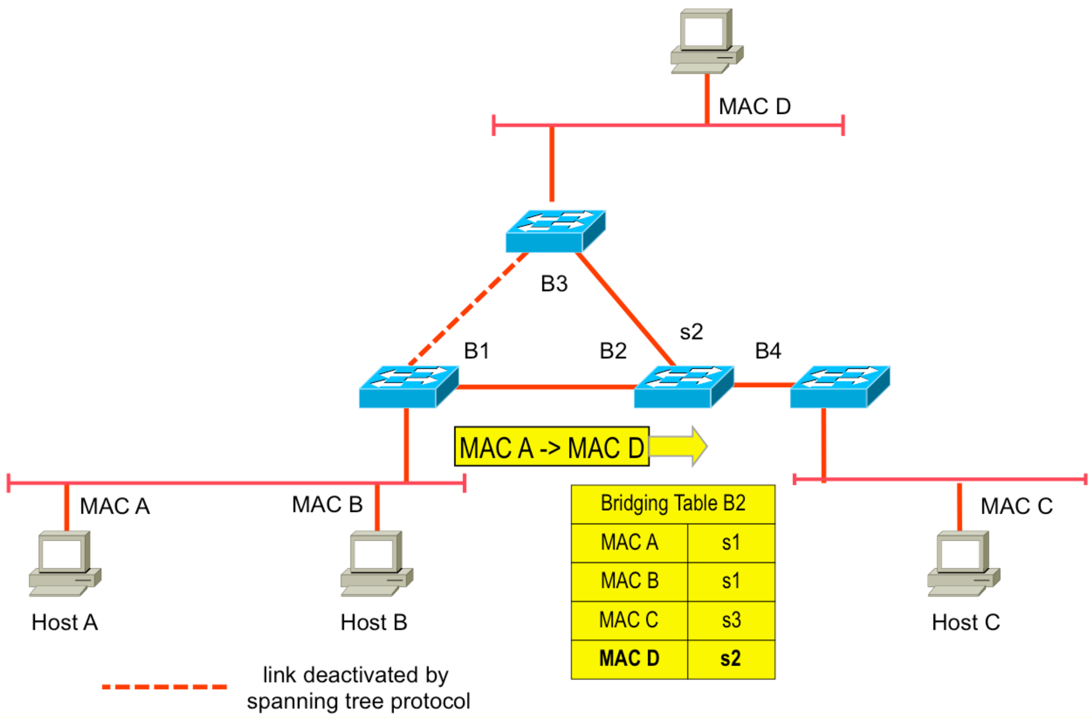
Frame MAC A to MAC D (1)



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

IP Technology (v6.4)

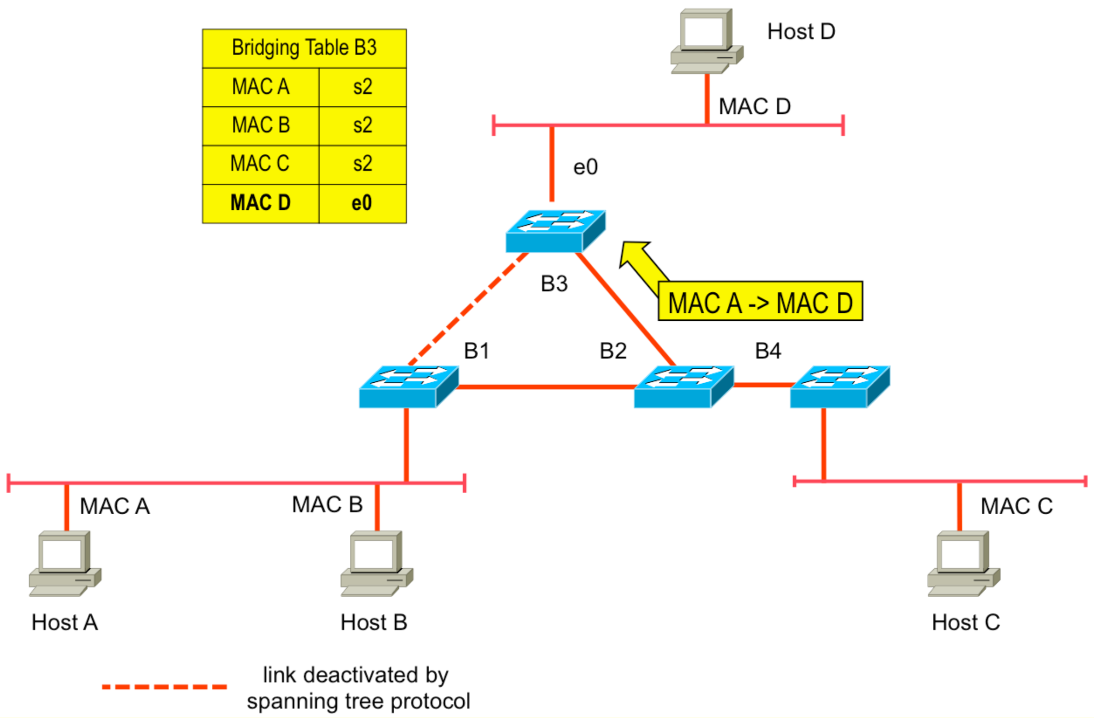
Frame MAC A to MAC D (2)



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

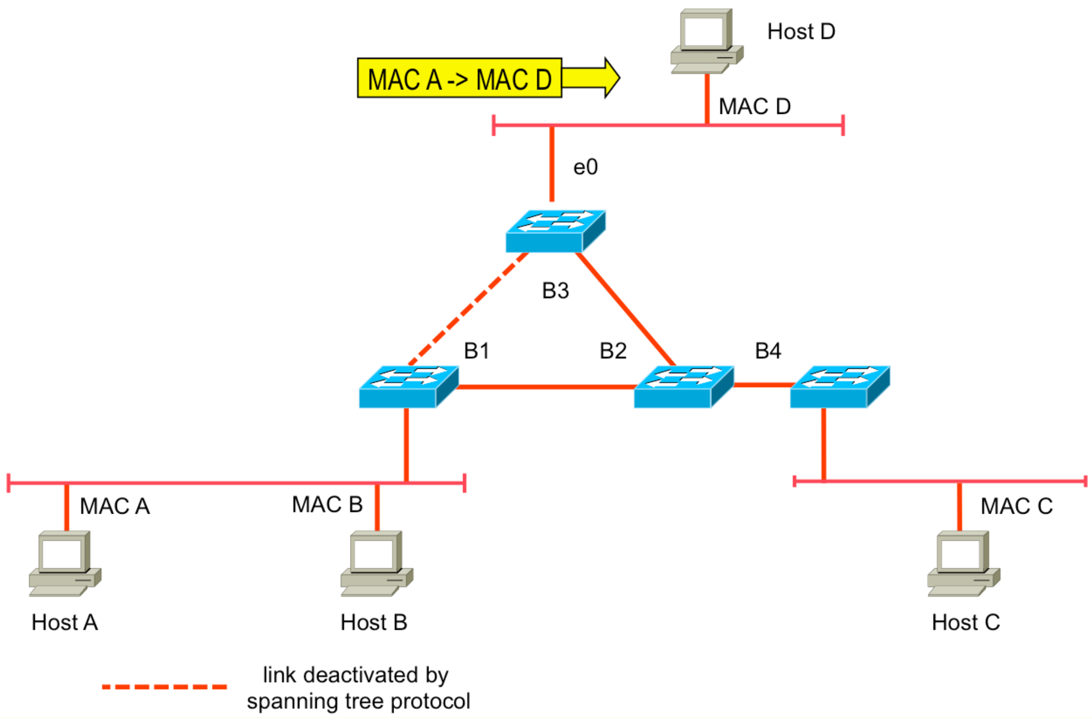
Frame MAC A to MAC D (3)

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



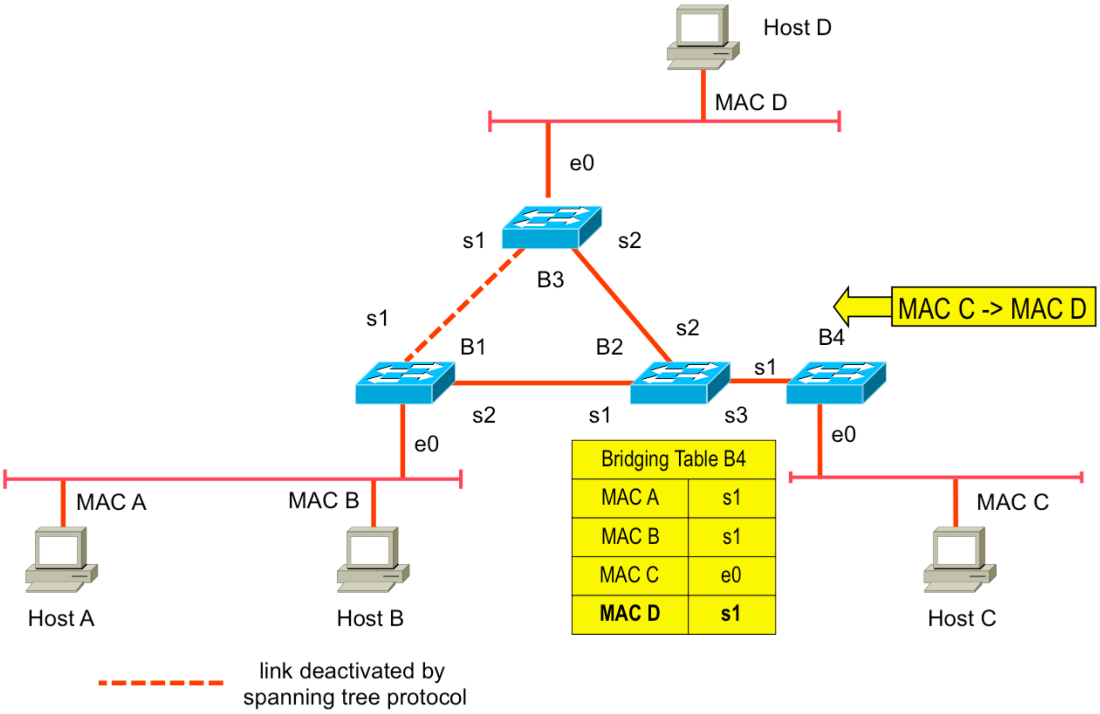
IP Technology (v6.4)

Frame MAC A to MAC D (4)



IP Technology (v6.4)

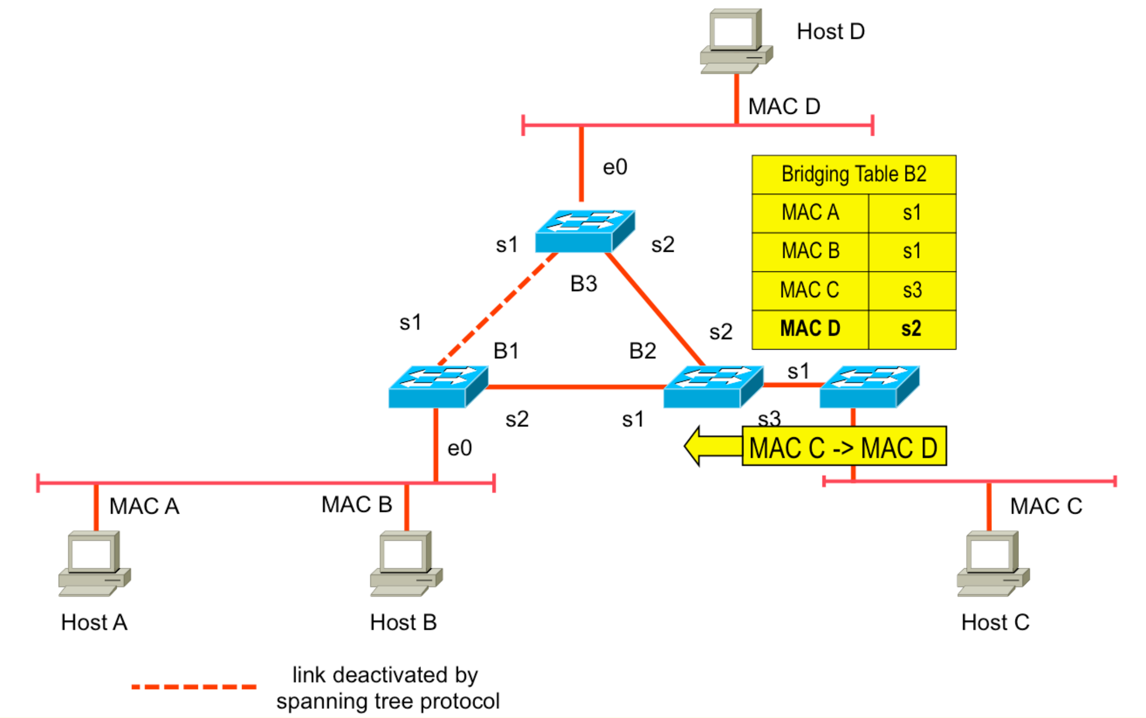
Frame MAC C to MAC D (1)



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

 IP Technology Primer, v6.4 13

Frame MAC C to MAC D (2)

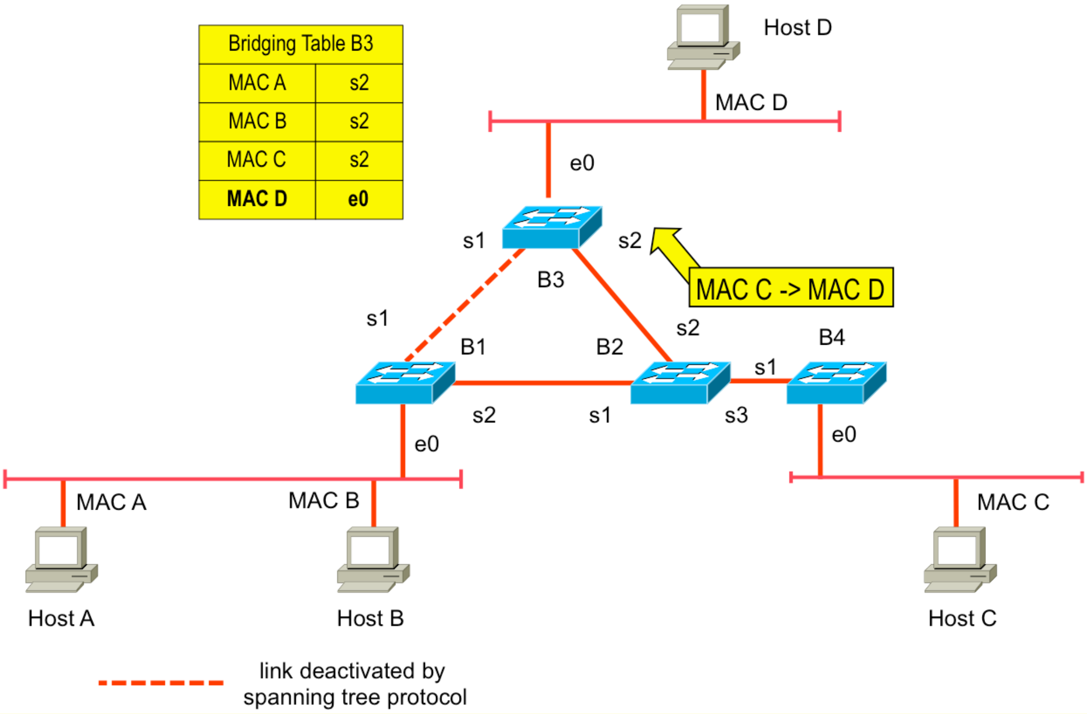


 IP Technology Primer, v6.4 14

Frame MAC C to MAC D (3)

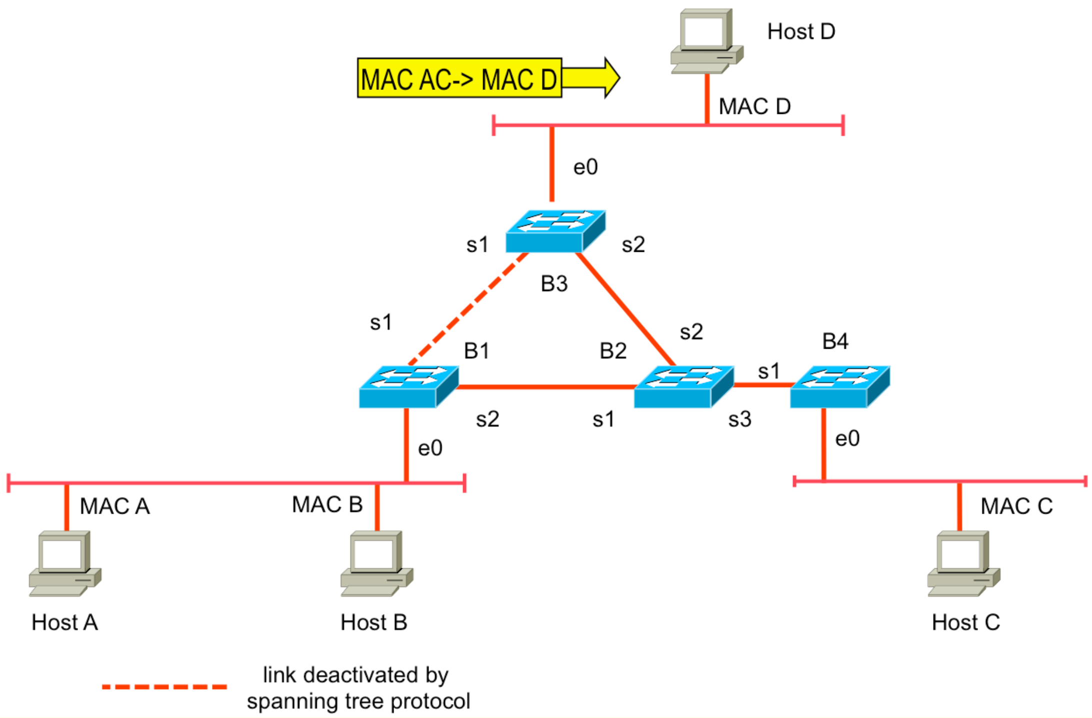
Takes Same Path as Frame from MAC A to MAC D -> No Load Distribution

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



IP Technology (v6.4)

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

© 2012/2017, D.I. Lindner / D.I. Haas

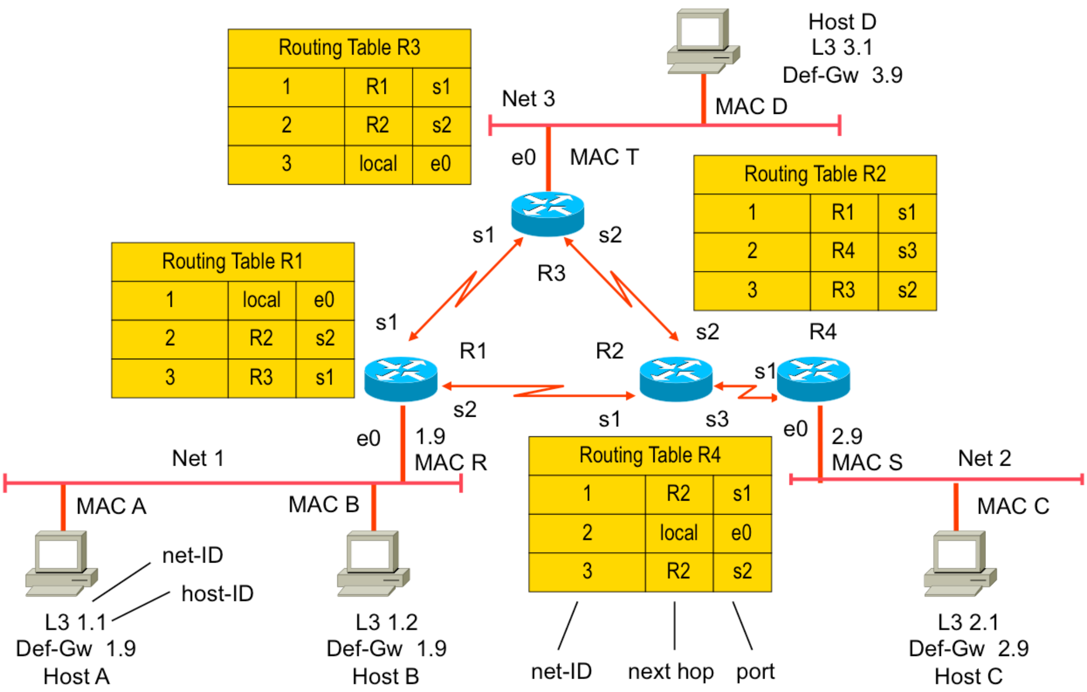
Routing Facts 1

- **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 R2		
1	R1	s1
2	R4	s3
3	R3	s2

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

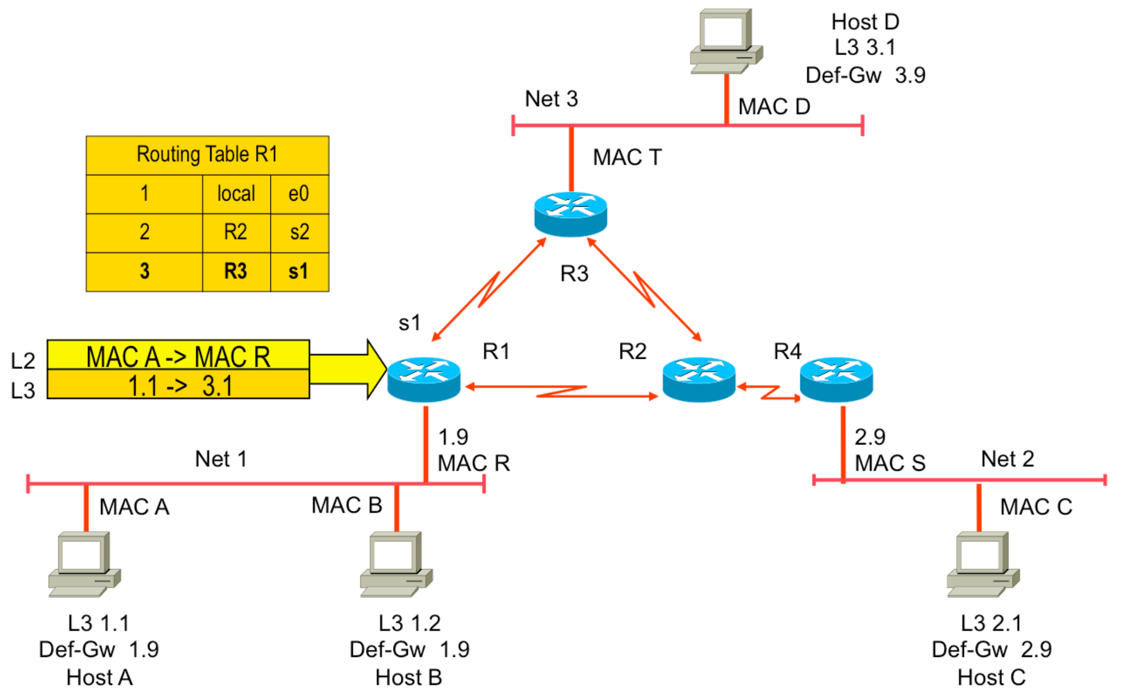
Routing Table R4		
1	R2	s1
2	local	e0
3	R2	s2

 IP Technology Primer, v6.4 20

IP Technology (v6.4)

Packet 1.1 to 3.1 (1)

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

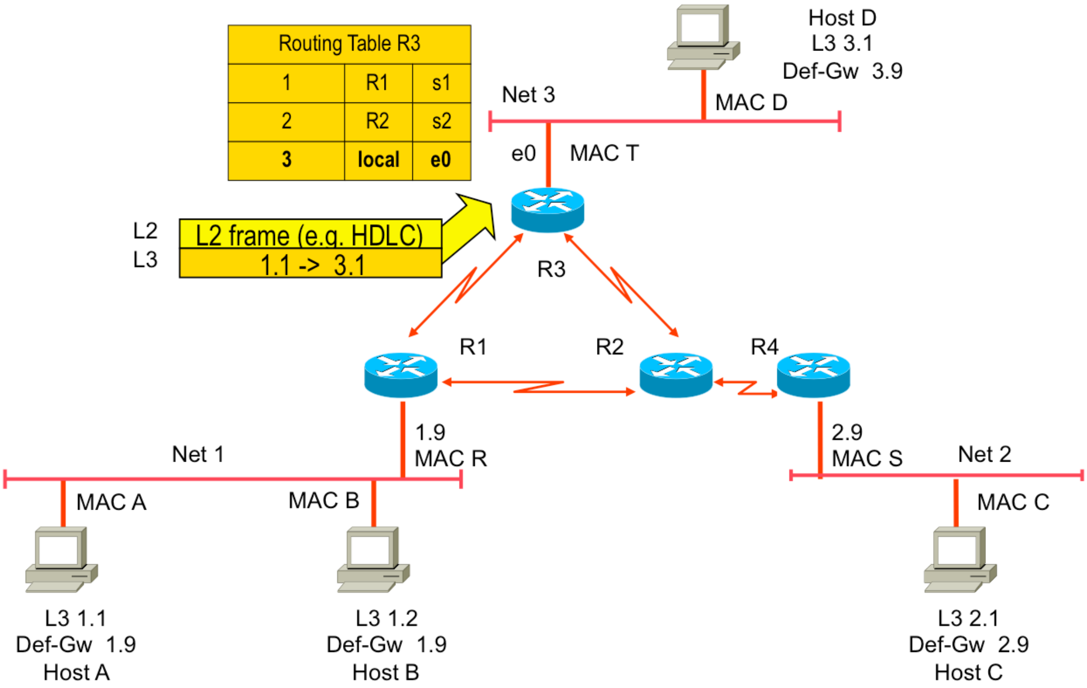


 IP Technology Primer, v6.4 21

IP Technology (v6.4)

Packet 1.1 to 3.1 (2)

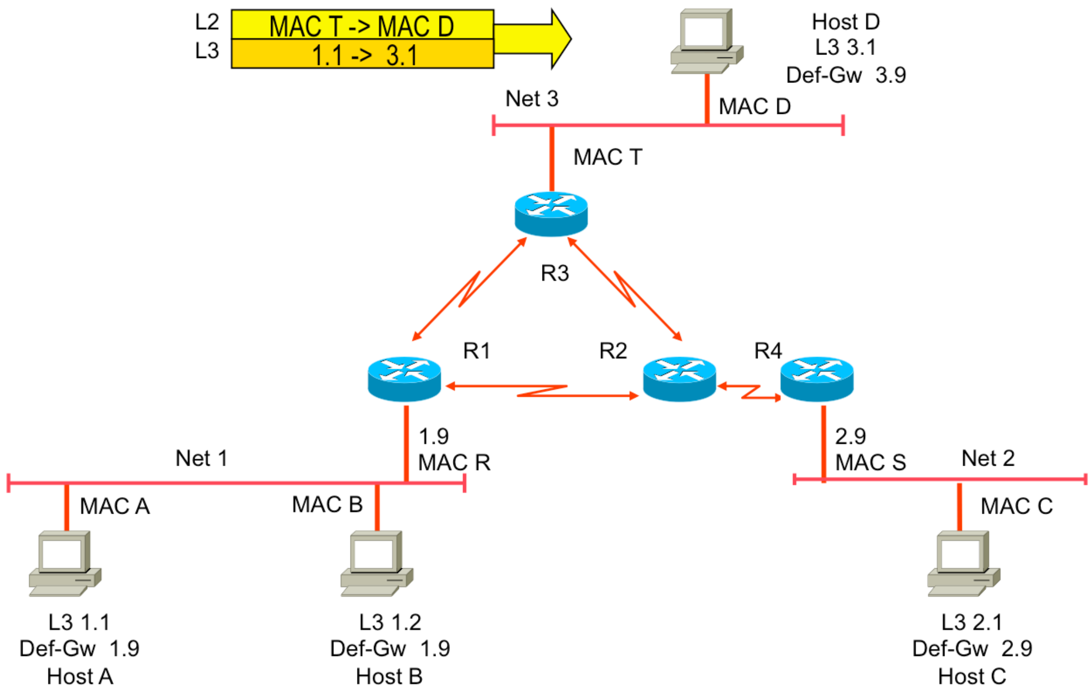
Routing Table R3		
1	R1	s1
2	R2	s2
3	**local**	**e0**



 IP Technology Primer, v6.4 22

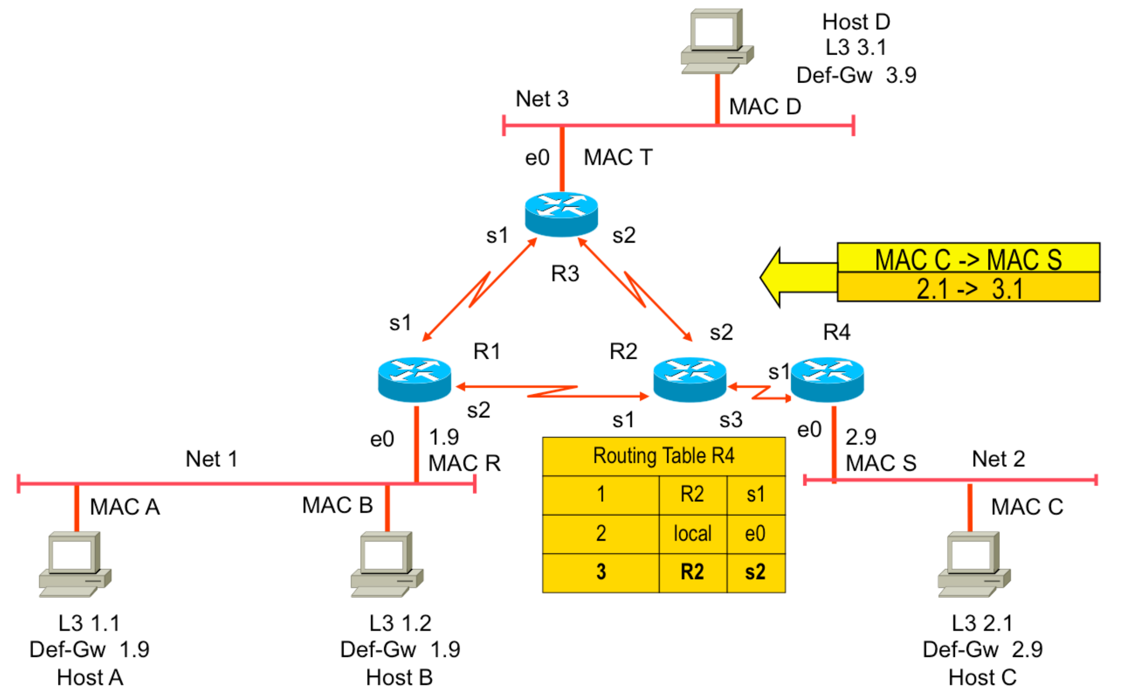
IP Technology (v6.4)

Packet 1.1 to 3.1 (3)



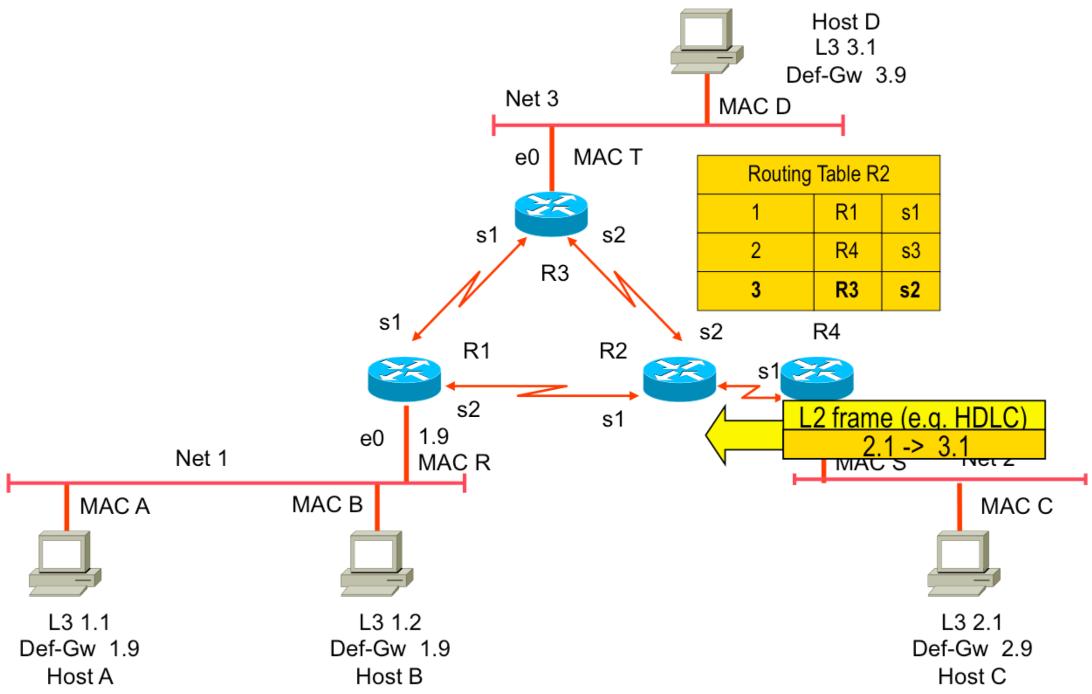
 IP Technology Primer, v6.4 23

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**

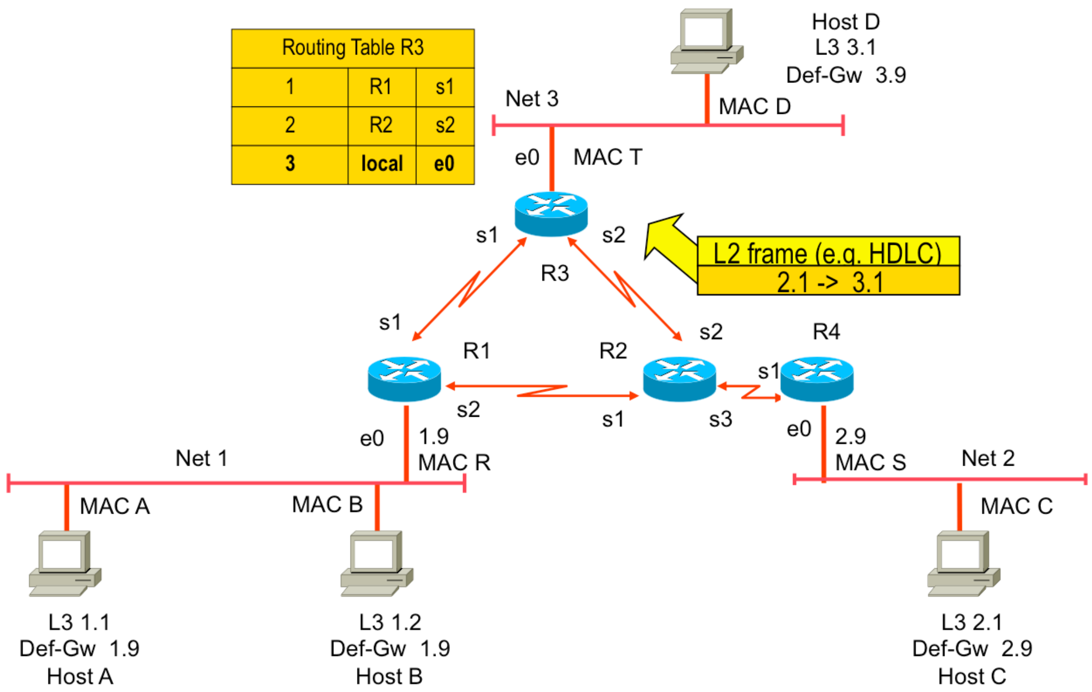
 IP Technology Primer, v6.4

IP Technology (v6.4)

Packet 2.1 to 3.1 (3)

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

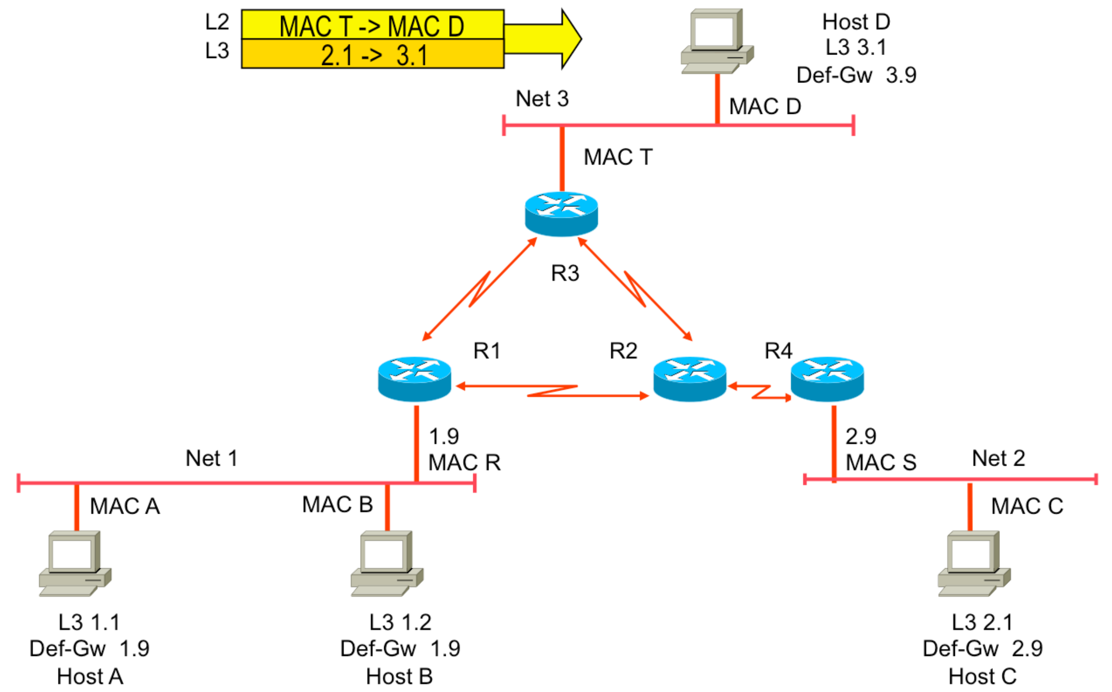
Routing Table R3		
1	R1	s1
2	R2	s2
3	**local**	**e0**



 IP Technology Primer, v6.4 26

IP Technology (v6.4)

Packet 2.1 to 3.1 (3)

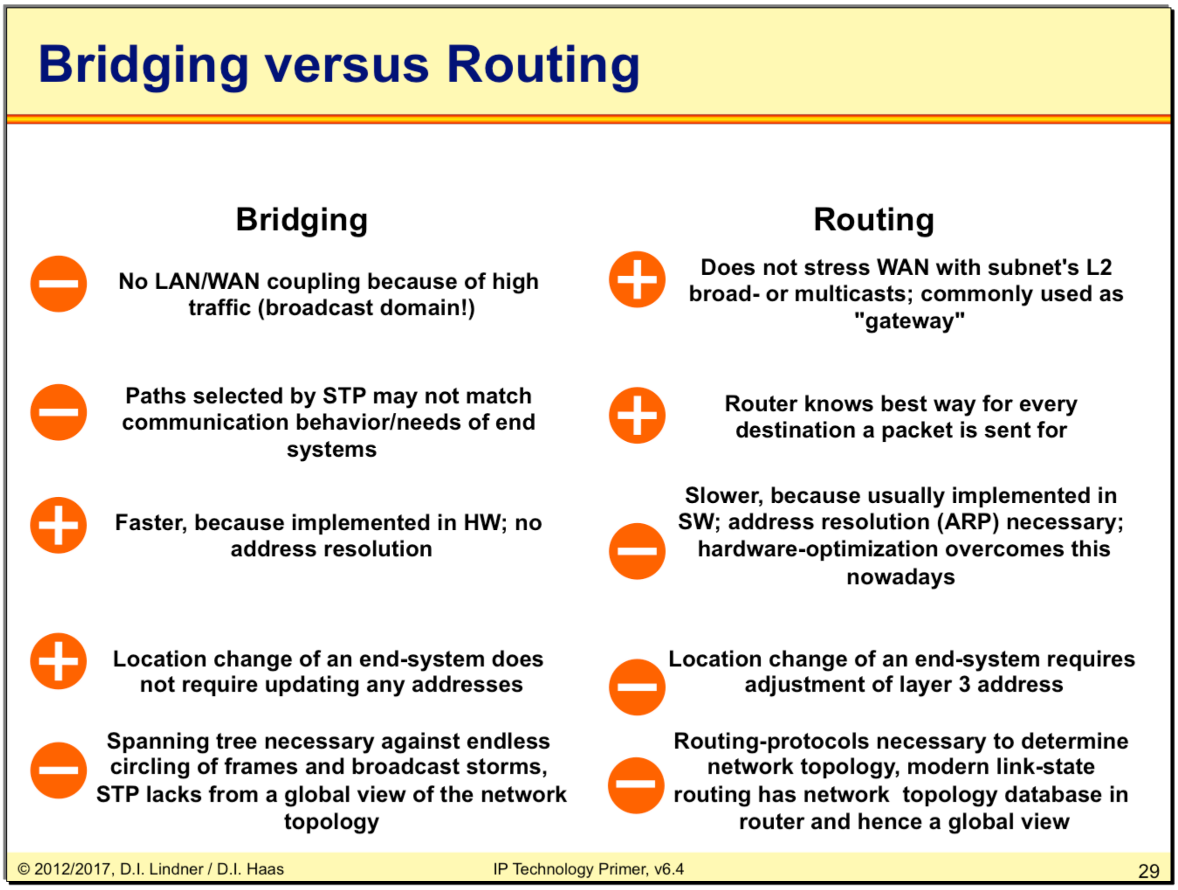


 IP Technology Primer, v6.4 27

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

The list shown above summaries all pro and cons of bridging (switching) and routing.



The list shown above summaries all pro and cons of bridging (switching) and routing (continued from previous slide).

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)

 IP Technology Primer, v6.4 30

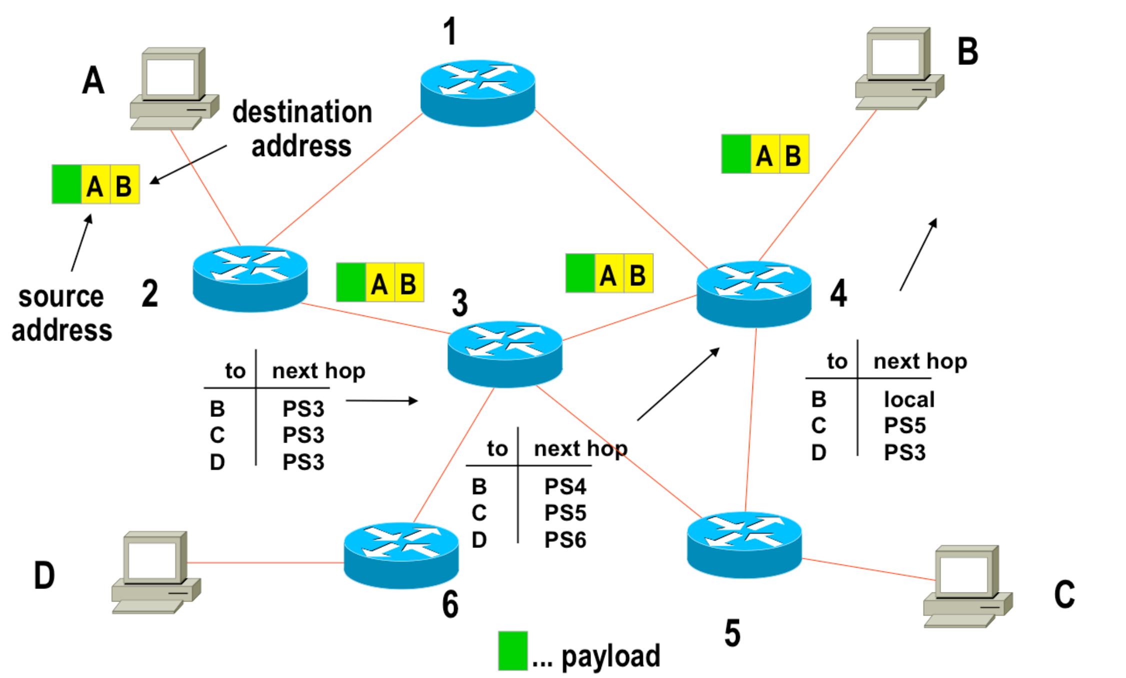
The addresses used in datagram service technologies need to be globally unique and structured. They contain topological information. Structured means a part of the address is reserved for the user identification while another part of the address is used for topology information (describes network where the user is located).

As already mentioned routing table can be based on a static configuration or on dynamic routing protocols.

Networks which are build on the datagram service technology typically need two different types of protocols: routed protocols which are used by the end user and routing protocols between routers to build up the routing tables.

IP Technology (v6.4)

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



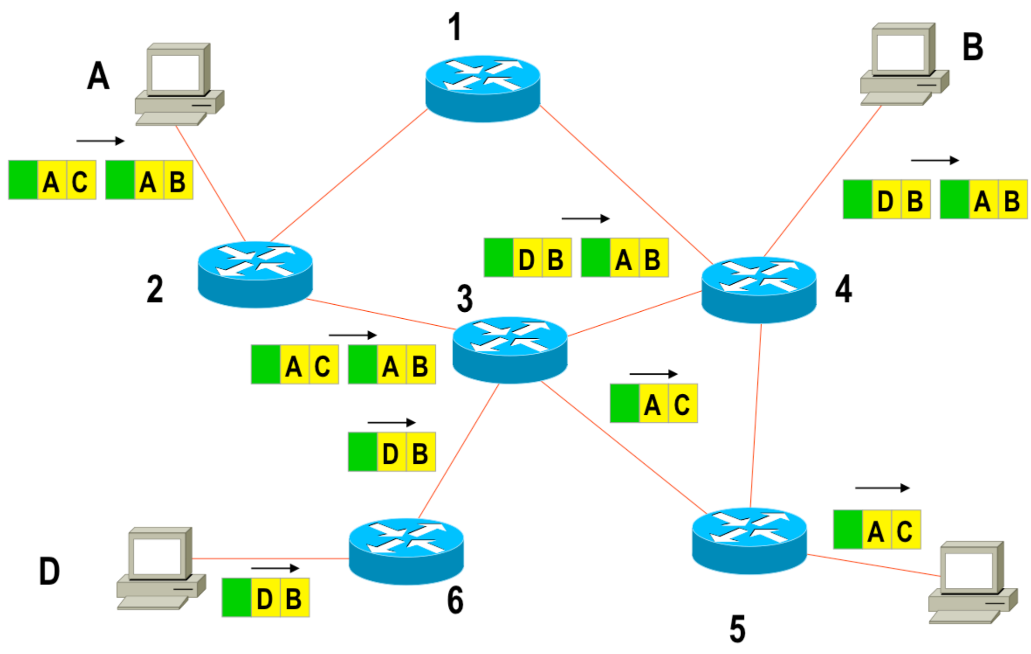
 IP Technology Primer, v6.4 31

In the datagram technology device A sends out data packets destined for the device B. Each single datagram holds the information about sender and receiver address.

The datagram forwarding devices hold a routing table in memory. In the routing table we find a correlation between the destination address of a data packet and the corresponding outgoing interface as well as the next hop. So data packets are forwarded through the network on a hop by hop basis.

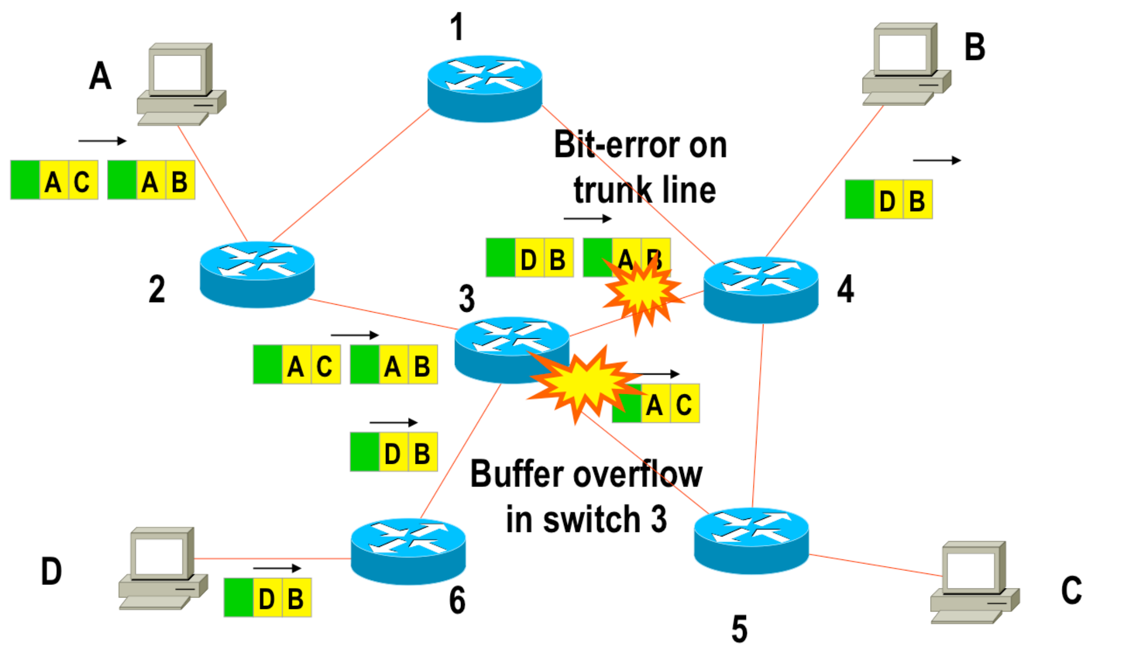
IP Technology (v6.4)

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



The routing tables can be set up either by manual configuration of the administrator or by the help of dynamic routing protocols (in case of IP that are protocols like RIP, OSPF, IS-IS, etc). The use of dynamic routing protocols may lead to rerouting decisions in case of network failure and so packet overtaking may happen in these systems.

Best-Effort Service



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"

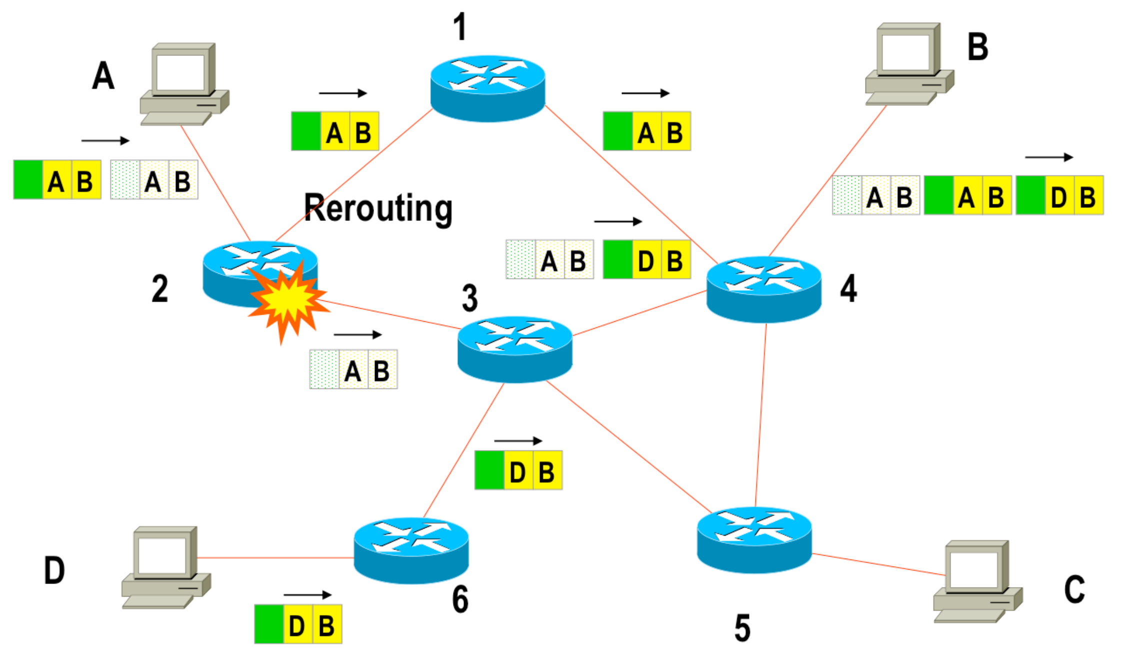
 IP Technology Primer, v6.4 34

Networks based on datagram technology support only best effort service, this means as good as it gets.

Routers that drop data packets because of buffer overflow or other problems don t care about error recovery. Error recovery is a task that needs to be performed by the end stations of a network. They have to take care for retransmissions in case of packet loss or transmission errors. This is typically done by layer 4 protocols like TCP which uses an connection-oriented mode.

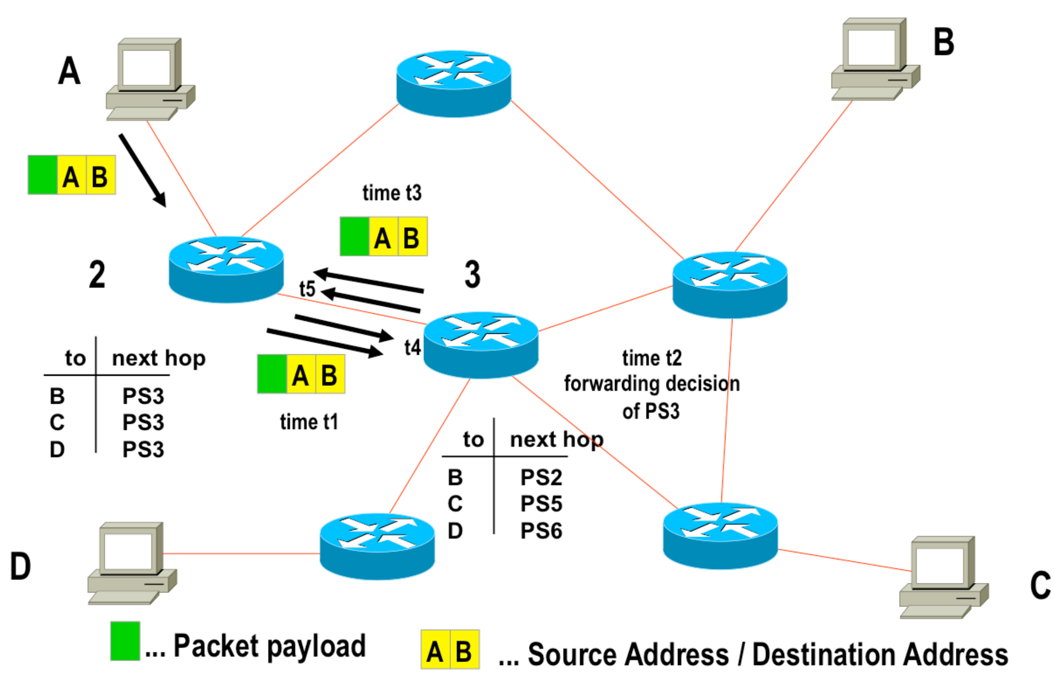
IP Technology (v6.4)

Rerouting – Sequencing Not Guaranteed !



 IP Technology Primer, v6.4 35

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



 IP Technology Primer, v6.4 36

In case of inconsistent information held in routing tables routing loops may occur which would lead to endless circling packets. Endless circulation means blocking of buffer memory in a packet switch. If there are two many endless circling packets in a network then all the buffers will be used up and hence other well-behaving traffic will be discarded because of lack of buffers. Special methods (kill mechanism) are necessary for avoiding or dampening that situation. Some protocols like IP use a maximum Time to Live field in their header to get rid of the endless cycling data packets.

That is a very important issue for all packet switching networks relying on forwarding of packets based on routing tables only.

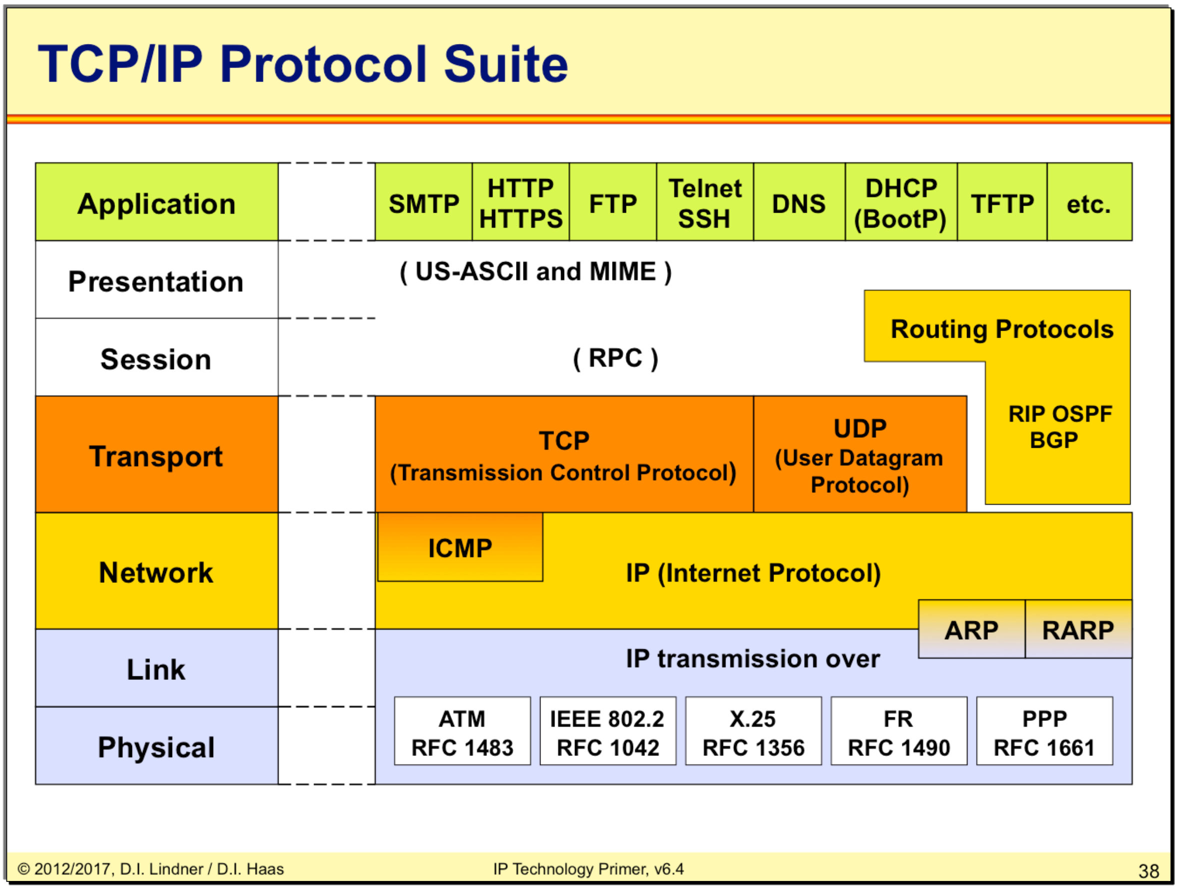
Datagram Service Facts (2)

- **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

 IP Technology Primer, v6.4 37

Topology changes cause rerouting when dynamic routing protocols are used and load sharing is practiced in the case of two or more paths with identical distance towards the destination. Rerouting and load balancing may also lead to packet overtaking, so the correct order of data packet arrival is not guaranteed.

IP Technology (v6.4)



IP is the connectionless layer 3 protocol. Datagram transport, fragmentation, addressing, all this is done by IP. ICMP (IP Control Message Protocol) is also seen as part of layer 3 providing error signaling to IP stations. It is carried in IP. most famous ICMP messages are those used for the PING-application. On the Transport Layer (Layer 4) you can see TCP and UDP. TCP protects the transmission of a “segment” and takes care for reliable delivery. UDP passes on just the connectionless service (best-effort-service) of IP to the higher layers (applications). ARP (Address Resolution Protocol) maps addresses between IP and L2 in case of a shared media (like LAN). In case of dynamic routing -> routing protocols are needed. RIP (Routing Information Protocol), OSPF (Open Shortest Path First protocol) are used within a limited area (so called autonomous system) of the Internet (such as within an ISP (Internet Service Provider) or within company or organization) whereas BGP is used for Internet routing. RIP is carried in UDP segments, OSPF is carried in IP datagrams and BGP is carried in TCP segments.

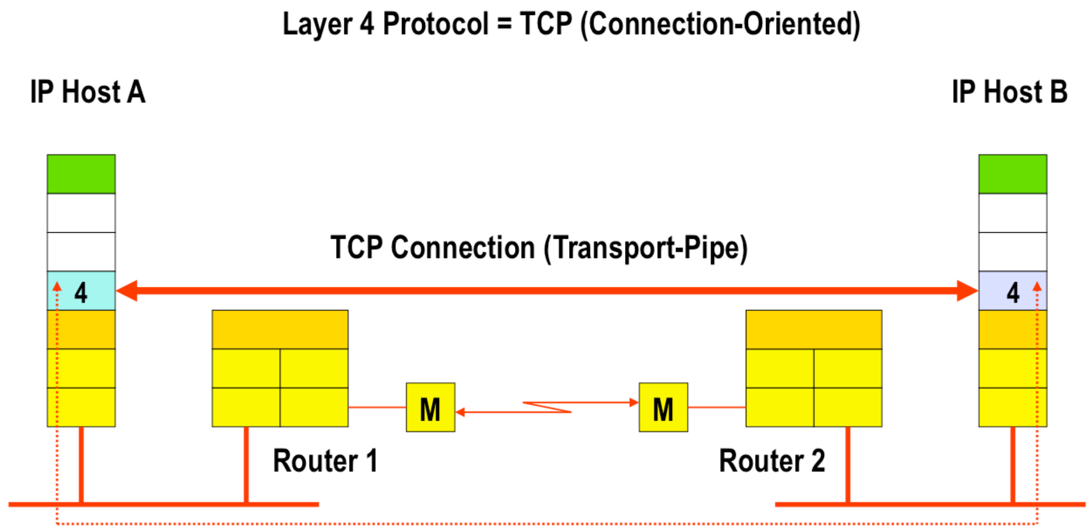
Some popular applications are shown: SMTP (Simple Mail Transport Protocol) for delivering emails, HTTP (HyperText Transfer Protocol) for WEB (HTTPS for secure/encrypted HTTP), FTP (File Transfer Protocol) for file transport, Telnet for remote login / virtual terminal, (SSH Secure Shell - > encrypted Telnet), DNS (Domain Name System) for resolving symbolic names to IP addresses, DHCP (Dynamic Host Configuration Protocol) for assigning IP addresses to IP hosts, TFTP (Trivial File Transport Protocol) as Idle-RQ technique for delivering files with small implementation overhead (e.g. needed for booting of a system). Of course there are lot of other important applications - which are not shown in the picture - like SNMP (Simple Network Management Protocol), SIP (Session Initiation Protocol) and RTP (Realtime Transport Protocol) used for VOIP (Voice Over IP).

TCP/IP seems to lack from OSI layer 5 and 6. That is not really true: Often parts of the presentation layer is covered in the application themselves in a very pragmatic way (like using US-ASCII as the base coding of email content (SMTP) or file content (FTP) or character set for terminal (Telnet)) or the content could be described and structured using MIME (Multipurpose Internet Mail Extensions). The later is also used for WEB and allows to carry nearly everything using HTTP. Pragmatic means, that no negotiation takes place about type of content to be delivered, e.g. a binary file containing a program is supposed to be usable/ readable for the receiving system. There is nothing which converts a MS PowerPoint presentation to an Apple keynote presentation during the transfer over a network. Also often parts of the session layer are included in the applications, sometimes the session layer is covered by a piece of software in a system like the RPC (Remote Procedure Call).

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



UDP (User Datagram Protocol)

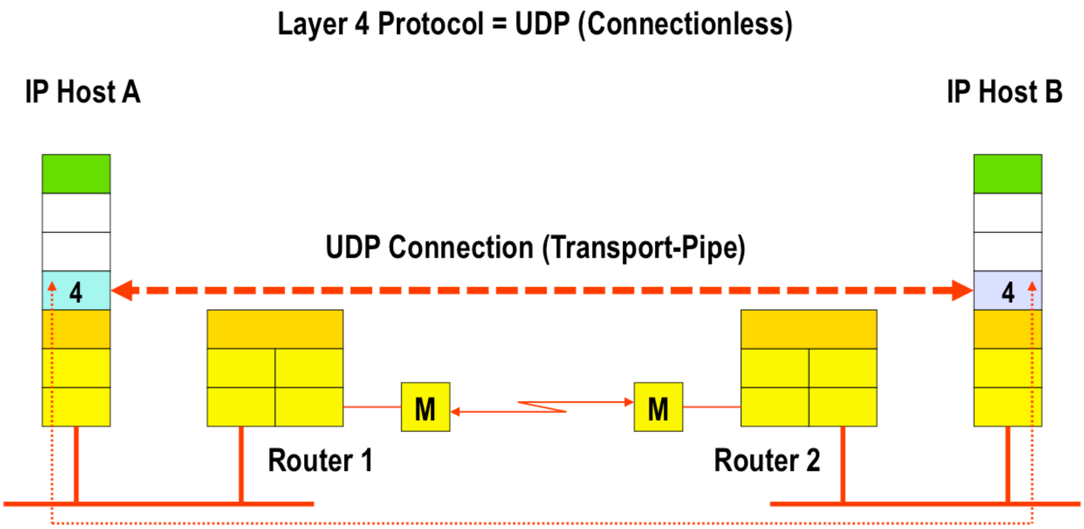
- UDP is a connectionless layer 4 service (datagram service)
- Layer 3 Functions are extended by port addressing and a checksum to ensure integrity
- UDP uses the same port numbers as TCP (if applicable)
- Less complex than TCP, easier to implement

UDP is connectionless and supports no error recovery or flow control. Therefore an UDP-stack is extremely lightweight compared to TCP.

Typically applications that do not require error recovery but rely on speed use UDP, such as multimedia protocols.

IP Technology (v6.4)

UDP and OSI Transport Layer 4

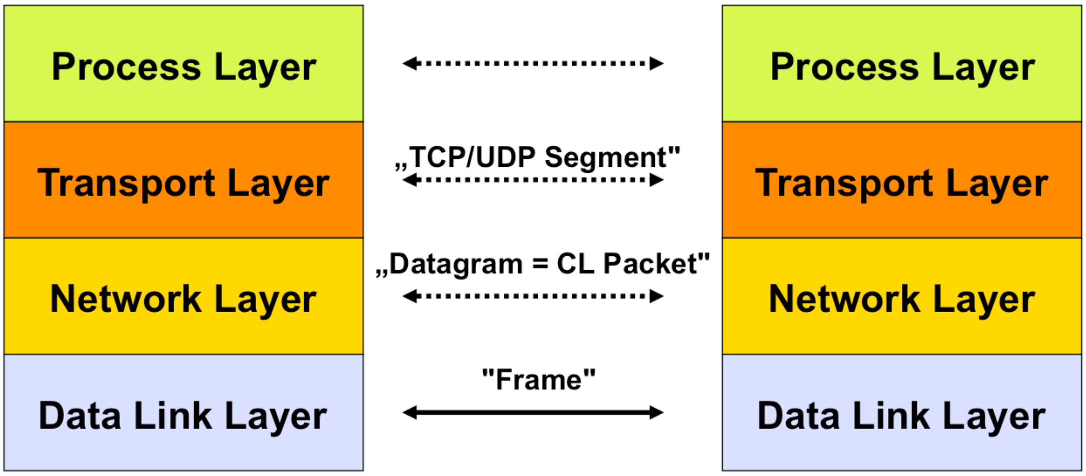


 IP Technology Primer, v6.4 42

Recognizes that even the IP hosts see a transport pipe, this pipe is unreliable.

IP Technology (v6.4)

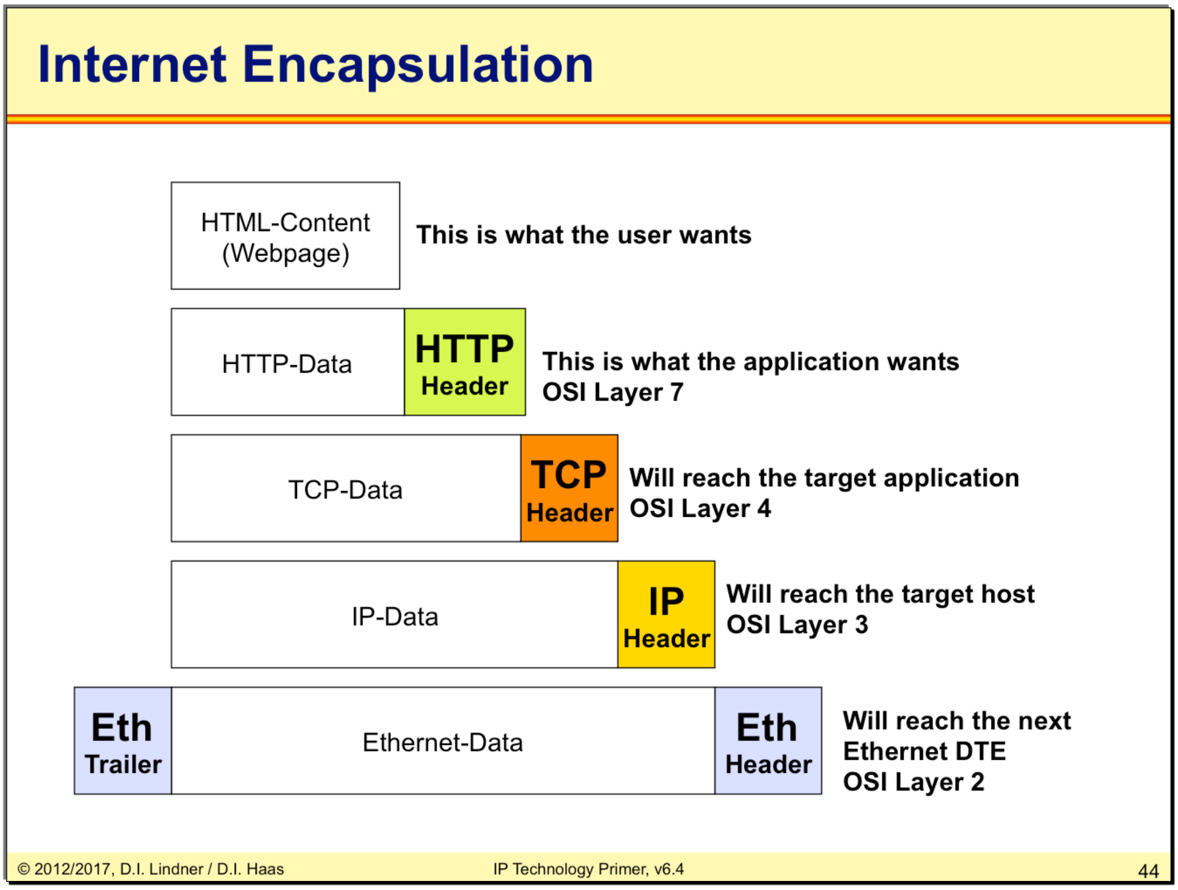
DoD 4-Layer Model (Internet)



 IP Technology Primer, v6.4 43

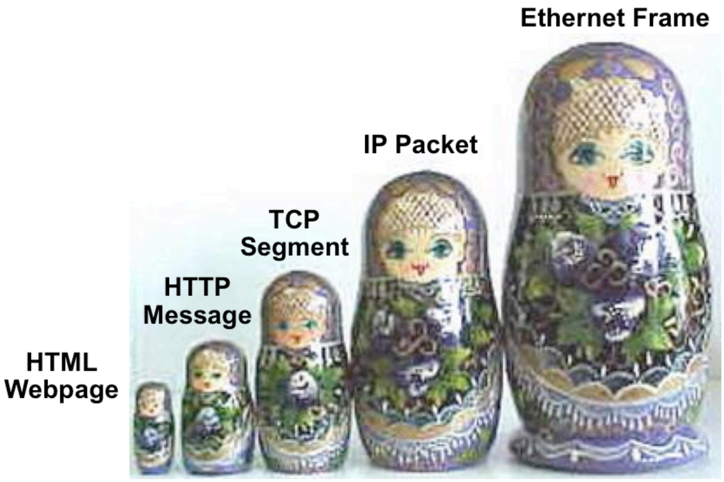
The picture above shows the W. Stevens 4 layer model which is used also in the Internet. The Internet layer model is also called "Department of Defense" (DoD) model.

IP Technology (v6.4)



In our example let's suppose a webserver sends a webpage (HTML code) to a client. The webpage is carried via the Hyper Text Transfer Protocol (HTTP) which provides for error and status messages, encoding styles and other things. The HTTP header and body is carried via TCP segments, which are sent via IP packets. On some links in-between, the IP packets might be carried inside Ethernet frames.

Practical Encapsulation



 IP Technology Primer, v6.4 45

The idea of encapsulation is fundamental in the data communication world. Adjacent layers encapsulate or decapsulate information by adding/removing additional "overheads" or "headers" in order to implement layer-specific functionalities. The whole process can be regarded as Matroschka-puppet principle.

In our example let's suppose a web-server sends a webpage (HTML code) to a client. The webpage is carried via the Hyper Text Transfer Protocol (HTTP) which provides for error and status messages, encoding styles and other things. The HTTP header and body is carried via TCP segments, which are sent via IP packets. On some links in-between, the IP packets might be carried inside Ethernet frames.

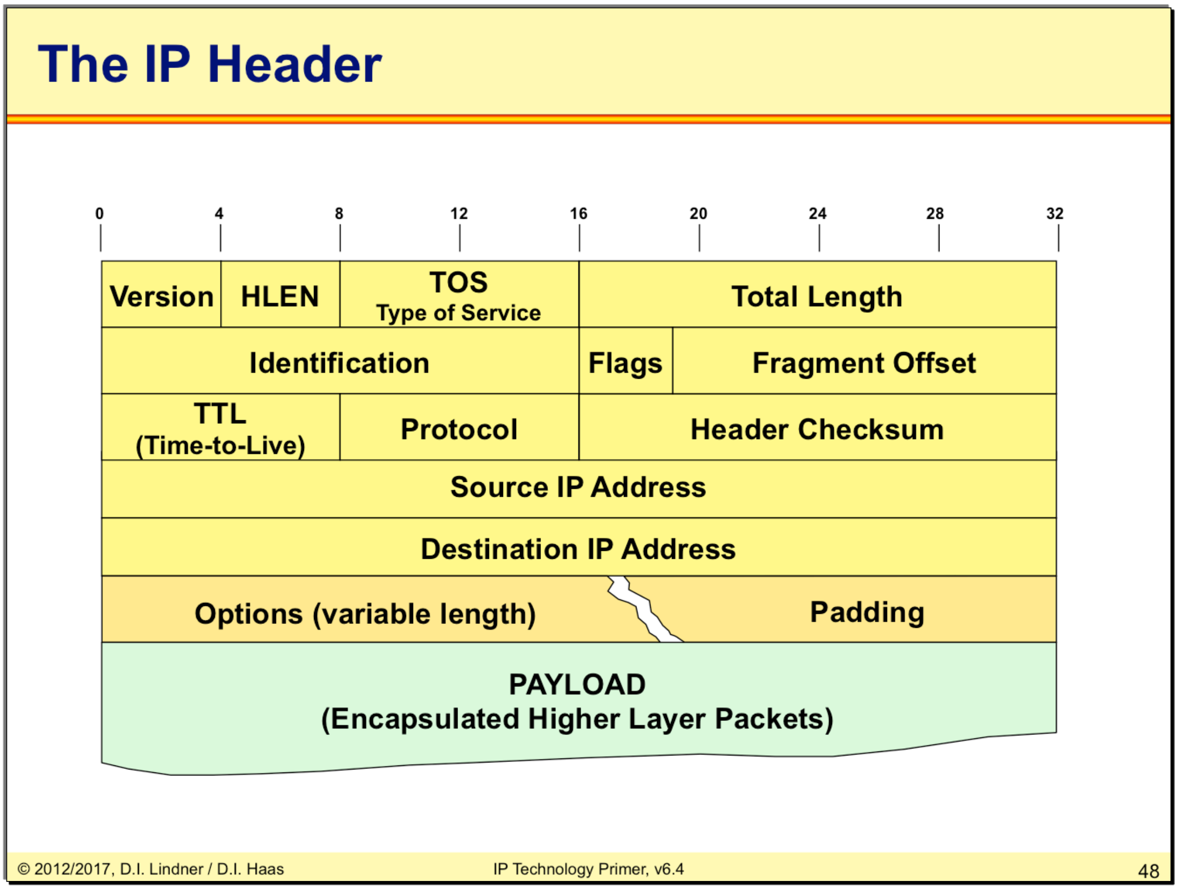
Agenda

- L2 versus L3 Switching
- IP Protocol, IP Addressing
- 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

IP Technology (v6.4)



Version: Version of the IP protocol. Current version is 4. Useful for testing or for migration to a new version, e.g. IPv6.

HLEN: Length of the header in 32 bit words. Header without options (HLEN 5 = 20 bytes).

TOS: Type of service -> covered by following slides.

Total Length: The length of the datagram including header and data. If fragmented -> length of fragment. Maximum datagram size = 65535 octets.

Identification, Flags (3 bits) **and Fragment Offset** (13 bits) -> covered by following slides.

TTL: This field indicates the maximum lifetime the datagram is allowed to remain in the system/network. The datagram must be destroyed, if the field contains the value zero. Units are seconds, range 0-255. It is set by the source to a starting value. 32 to 64 are common values. Every router decrements the TTL by the processing/waiting time of a datagram is to be forwarded. If the time is less than one second, TTL is just decremented by one. Therefore nowadays TTL is just a hop count. If TTL reaches 0, the datagram or fragment is discarded. An end system use the remaining TTL value of the first arriving fragment to set the reassembly timer.

Attention: Because of decrementing TTL for each datagram a router has to recompute the header checksum too. That is one of the reasons while IP routing (L3 switching) is still slower than Ethernet switching (L2 switching).

Protocol: Describes what protocol is used in the next level e.g. 1 (ICMP), 6 (TCP), 8 (EGP), 17 (UDP), 89 (OSPF), etc… Over 100 different IP protocol types are registered so far.

Header Checksum: A Checksum for the header only -> modulo 2 sum of the individual bytes computed byte by byte.

Source IP Address: 32 bit IP address of the source (sender) of a datagram

Destination IP Address: 32 bit IP address of the receiver (destination) of a datagram

Padding: "0"-bytes to fill the header to a 32 bit boundary in case of options.

IP Options: Options were used for timestamps, security and special routing aspects. Record Route option: Records the route of a packet through the network. Each router, which forwards the packet, enters its IP address into the provided space. Loose Source Route option: A datagram or fragment has to pass the routers in the sequence provided in the list. Other intermediate routers not listed may also be passed. Strict Source Route option: A datagram or fragment has to pass the routers in the sequence listed in the source route. No other router are allowed to pass. Today most IP Options are blocked by firewalls because of inherent security flaws e.g. source routing could divert an IP stream to a hackers network station.

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**

The IP Address is a 32 bit value in the IP header. The address identifies the access to a network. Always keep in mind that IP addresses are basically simple numbers only. There is no natural structure in it.

It is widely common to write down an IP address in the so-called "dotted decimal notation", where each byte is represented by a decimal number (0-255) and those numbers are separated by dots.

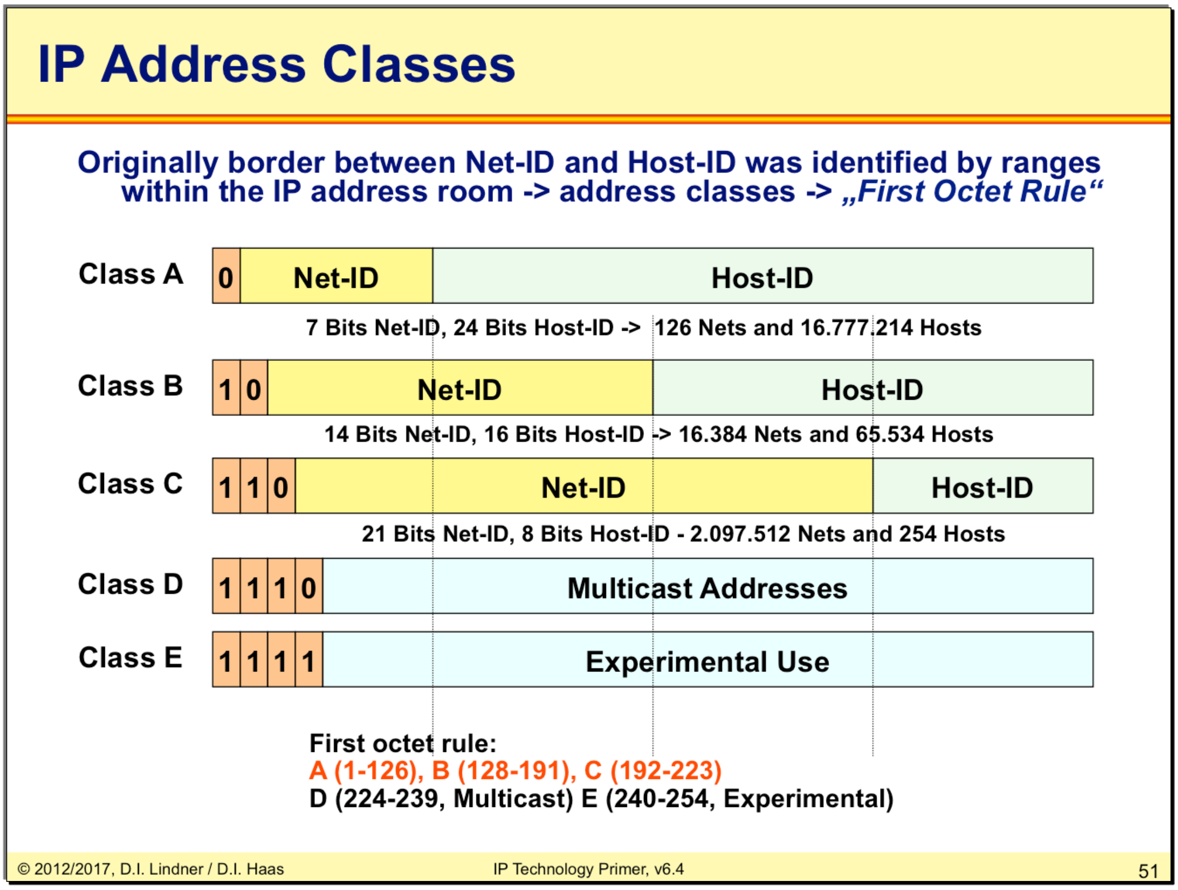
In order to make an address routable we need topological information on it. Therefore, the address is split into two parts: the network number (or "Net-ID") and the host number (or "Host-ID"). The Net-ID must be unique for each IP network connected to the Internet and is maintained by RIPE ("Internet Registry") in Europe. The Host-ID can be arbitrarily assigned by each local network manager.

You can compare the structure of an IP address with the following picture: The Net-ID is like the street name and the Host-ID is like the house number of a building connected to this street. The Net-ID contains the topology information in the network map and must be unique. The Host-ID has only local meaning. So the same Host-ID can be used on different streets.

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 Technology (v6.4)



In the beginning of the Internet, five address classes had been defined in order to identity the border between Net-ID and host-ID and a fixed way. The idea of classes helps a router to decide how many bits of a given IP address identify a network number and how many bits are therefore available for host numbering.

Classes A, B, and C had been created to provide different network addresses ranges. Additionally Class D is the range of IP multicast addresses, that is they have no topological structure. Finally, class E had been reserved for research experiments and are not used in the Internet.

The usage of classes has a long tradition in the Internet and was a main reason for IP address depletion which first was overcome by classless routing and NAT and finally by IPv6.

The first byte (or "octet") of an IP address identifies the class. For example the address 205.176.253.5 is a class C address.

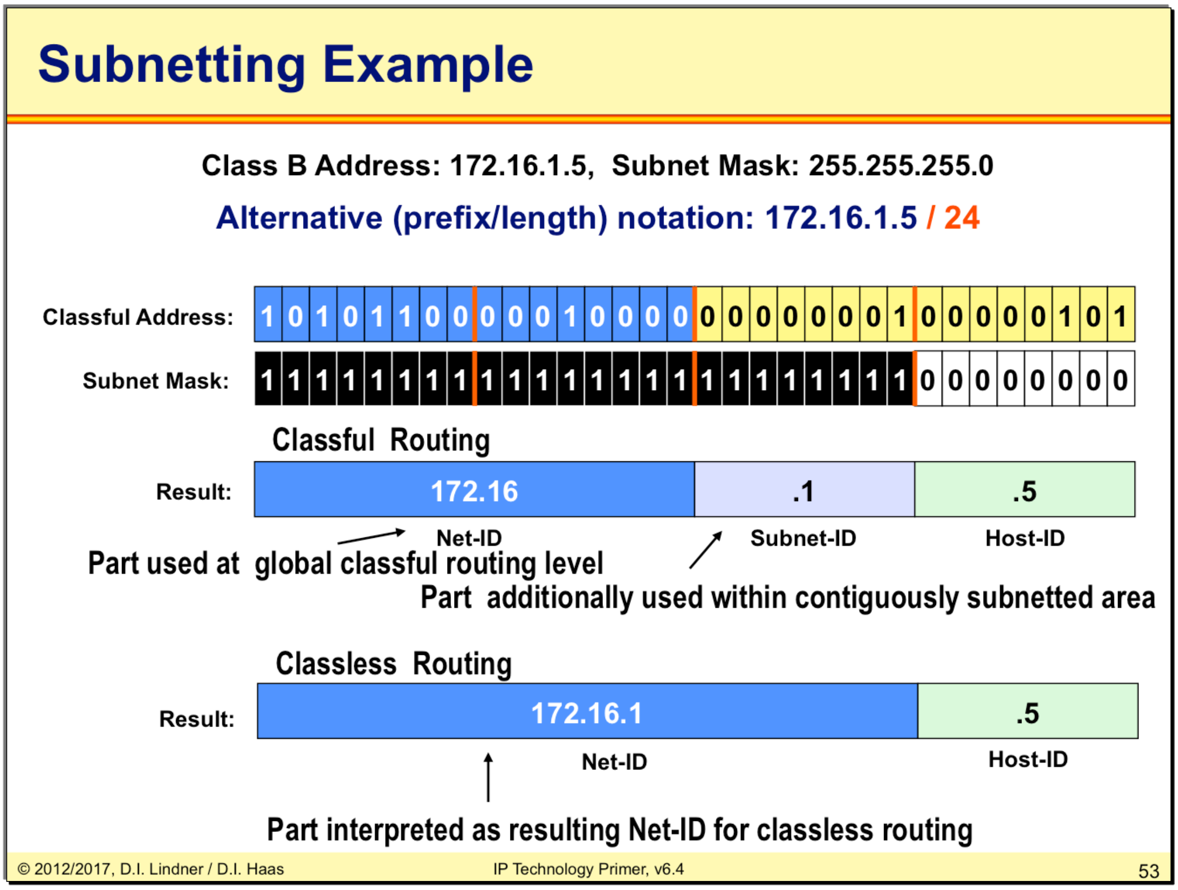
The "**classful**" method of identifying network-IDs based on the given IP address class is inflexible and lead to address space depletion. Class C networks are too small for most organizations but class A and B are too large. A waste of the IP address space happened by giving class B or class A address space to customers which do not need the entire space. LANs were getting bigger and bigger and a logical separation of an organizations network (e. g. of a class A network number) would be a great help. Even a class A address would not help in that case because with a single class A Net-ID only one physical flat network can be addressed (even if 16.777.214 hosts are possible on this flat network. Another problem which was

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

 IP Technology Primer, v6.4 52

In 1985, RFC 950 defined a standard procedure to support **subnetting** of a single Class A, B or C network number into smaller pieces. Now organizations can deploy additional subnets without needing to obtain a new network number from the Internet. Instead of the classful two-level hierarchy, subnetting provides a **three-level** hierarchy. The idea of subnetting is, to divide the standard host-number field into two parts, the subnet-number and the host-number *on that subnet*. The subnet structure of a network is never visible outside of a the organizations private network. The route from the Internet to any subnet of a given IP address is the same, no matter which subnet the destination host is on. This is because all subnets of a given network number use the same network-prefix but different subnet numbers.



Number of bits to be used for Net-ID and Subnet-ID are specified by subnet mask (also written in dotted decimal notation):

Ones portion represents network part.

Zeros portion represent the host part.

Note: A subnet mask must always consist of a contiguous series of "1". For example, these are not valid subnet masks: 254.255.0.0, 255.127.255.0, 255.255.255.195

There are two notations:

The old but still commonly used notation is to write the subnet mask like an IP address. Examples: 255.255.0.0, 255.255.255.0, 255.255.192.0.

The new notation is much simpler and identifies the subnet mask by a simple number, that is the number of "1"-bits. Examples: /16, /24, or /18. Thus a network can be specified as 172.16.128.0/18 or shorter as 172.16.128/18 (prefix notation).

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

© 2012/2017, D.I. Lindner / D.I. Haas

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

 IP Technology Primer, v6.4 55

A network broadcast is used to send a broadcast packet to a dedicated network. The IETF strongly discourages the use of IP directed broadcast and it is not defined for IPv6.

If a destination IP address consists of "all 1", which can be represented by decimal numbers as "255.255.255.255", then this is recognized as "local" or "limited" broadcast. A limited broadcast is never forwarded by routers, otherwise the whole Internet would be congested by "broadcast storms". Note that broadcast addresses must not be used for source addresses.

A network is described using the "network address", which is simply its IP address with host part set to zero. Network addresses are used in routing entries and routing protocols, since a router only deals with networks and doesn't care for host addresses.

Each operating system provides a virtual IP interface, called the loopback interface. Per default the IP addresses 127.x.x.x are reserved for this reason. Initially, the idea came from the UNIX world as IP is only one of several means to achieve inter-process communication upon a UNIX workstation. Other methods are named/unnamed pipes, shared memories, or message queues for example.

When using IP for inter-process-communication, the involved client/server processes can be distributed upon different servers across a network—without any modification of the source codes!

By default, a modern operating system assigns the IP address 127.0.0.1 to the local loopback interface.

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

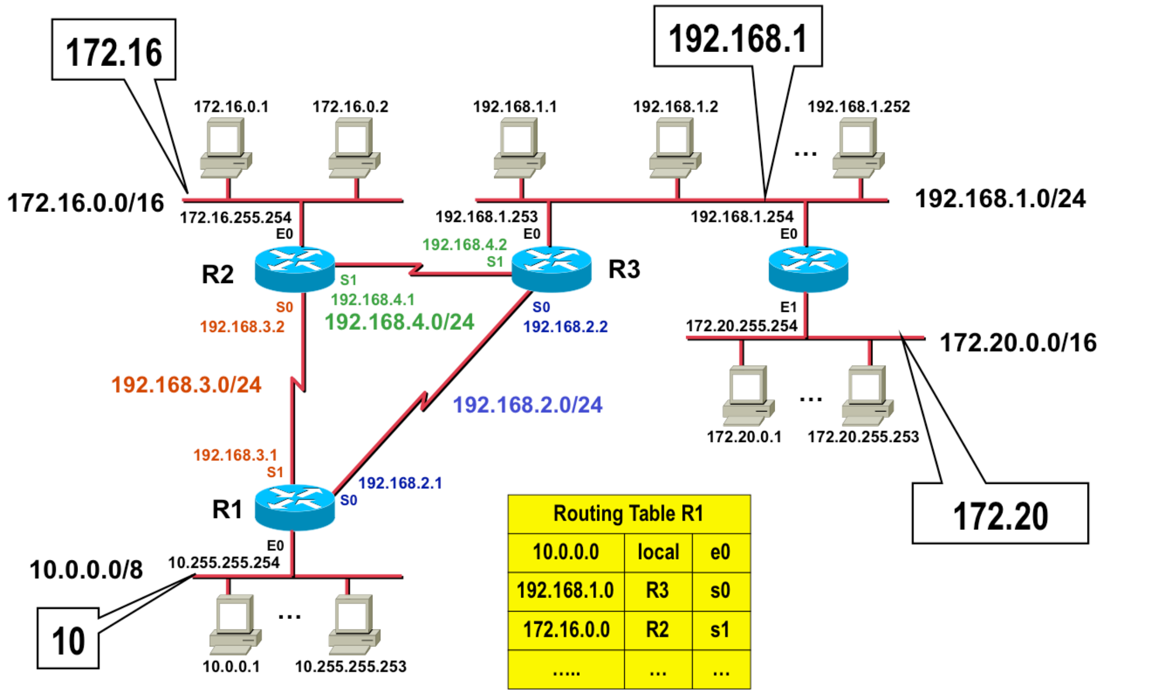
 IP Technology Primer, v6.4 56

So-called RFC 1918 addresses are class A, B, and C address blocks which can be used for internal purposes. Such addresses must not be used in the Internet. All gateways connected to the Internet should filter packets that contain these private addresses. Furthermore these addresses must not be used in Internet routing updates.

Because of those rigid filter policies, it is relatively safe to utilize RFC 1918 addresses in local networks—everybody in the Internet knows which addresses must be filtered.

IP Technology (v6.4)

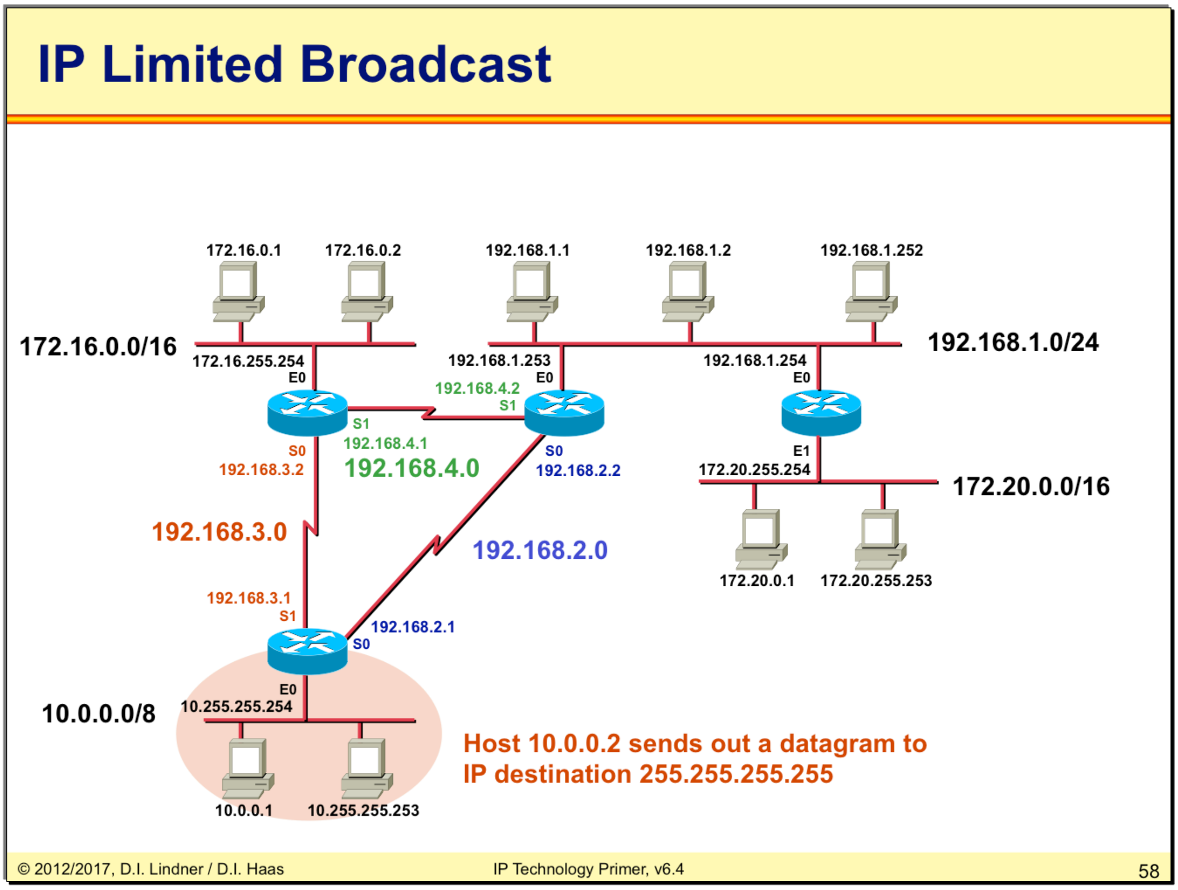
Net-ID Addressing Example



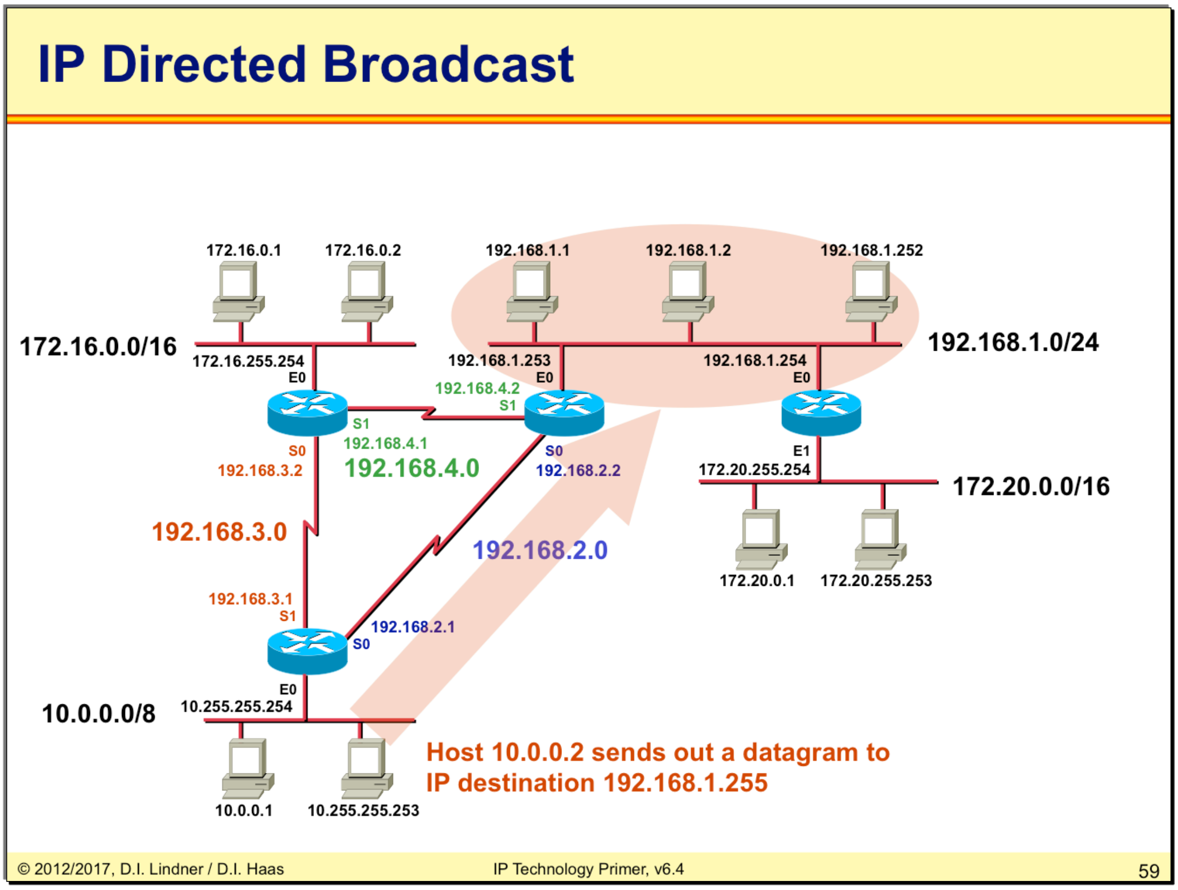
Routing Table R1		
10.0.0.0	local	e0
192.168.1.0	R3	s0
172.16.0.0	R2	s1
.....	...	...

 IP Technology Primer, v6.4 57

IP Technology (v6.4)



The example above shows a “limited broadcast” (all ones in net-part and host-part). Only the hosts in Net 10 receive this datagram.



In this example a datagram to the Network 192.168.1.0 is sent but the host-ID is set to "all-ones". As routers do not care about the host IDs, this datagram is forwarded according its destination network number, and only the last router is responsible for direct delivery.

When the last router examines the (destination-) host-ID of the datagram, it notices that this is a broadcast address and transforms the whole address into a limited broadcast address (255.255.255.255). Finally the router can send this datagram into the local network without issuing an ARP request.

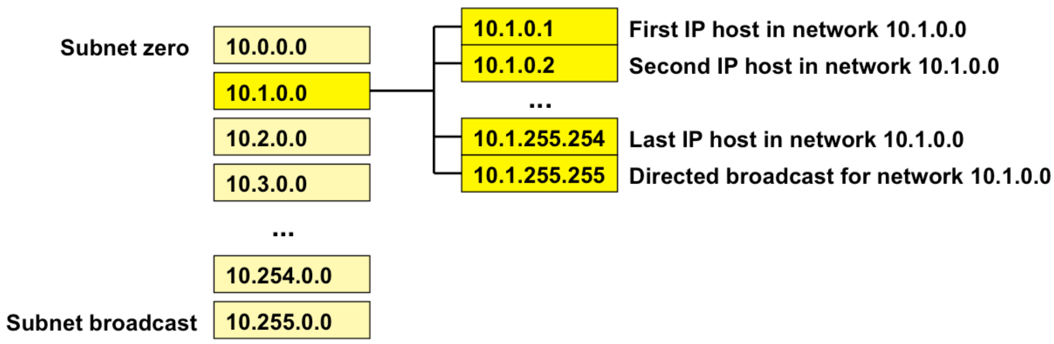
Note that directed broadcasts are not recommended anymore as they can be abused for denial-of-service (DoS) attacks. Typically, directed broadcasts are filtered by the firewall. IPv6 does not provide broadcasts at all!

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:**



The example above shows how to subnet a class A network—in our case network 10. Here we use a 16-bit subnet mask allowing us to define 2^8 – 2 subnets, because the natural subnet mask of a class A network is 8 bits in length.

The diagram above shows the total range of subnetworks including the "forbidden" ones, that is subnet zero and the subnet broadcast.

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"

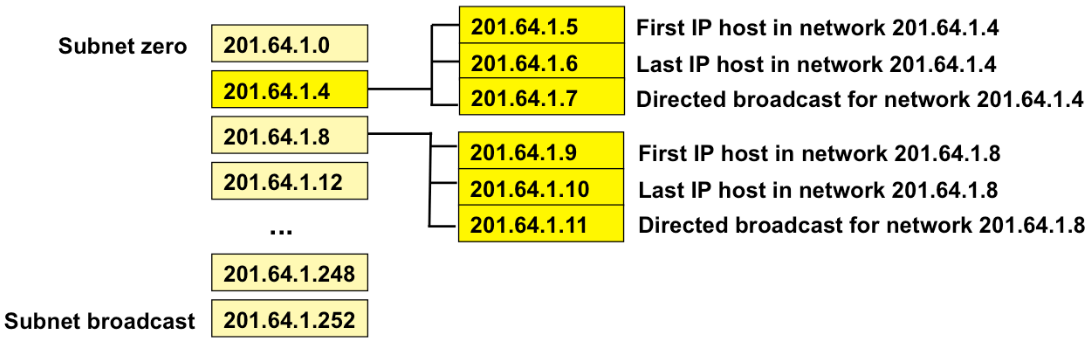
1) **That is: 175.32.0.0 with 255.255.240.0 or 175.32.0.0/20**
2) **Resulting subnetworks:**

Subnet zero	175.32.0.0			
	175.32.16.0	→	175.32.16.1	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
	175.32.32.0			
	175.32.48.0			
	...			
	175.32.224.0			
Subnet broadcast	175.32.240.0			

Subnet Example 3

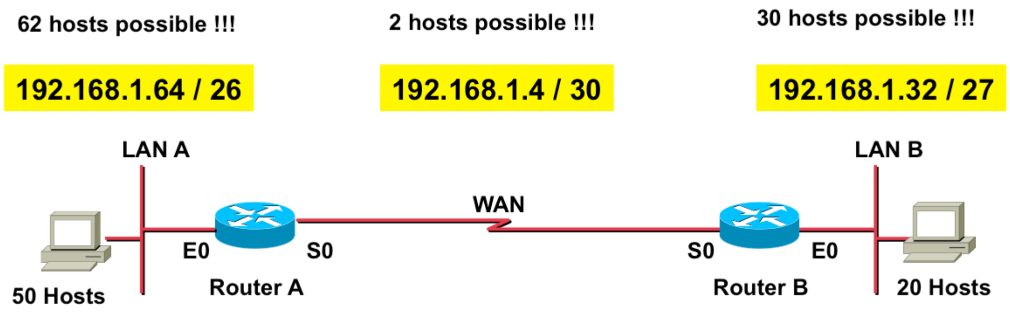
"Use the class C network 201.64.1.0 and 6 bit subnetting"

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



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!**

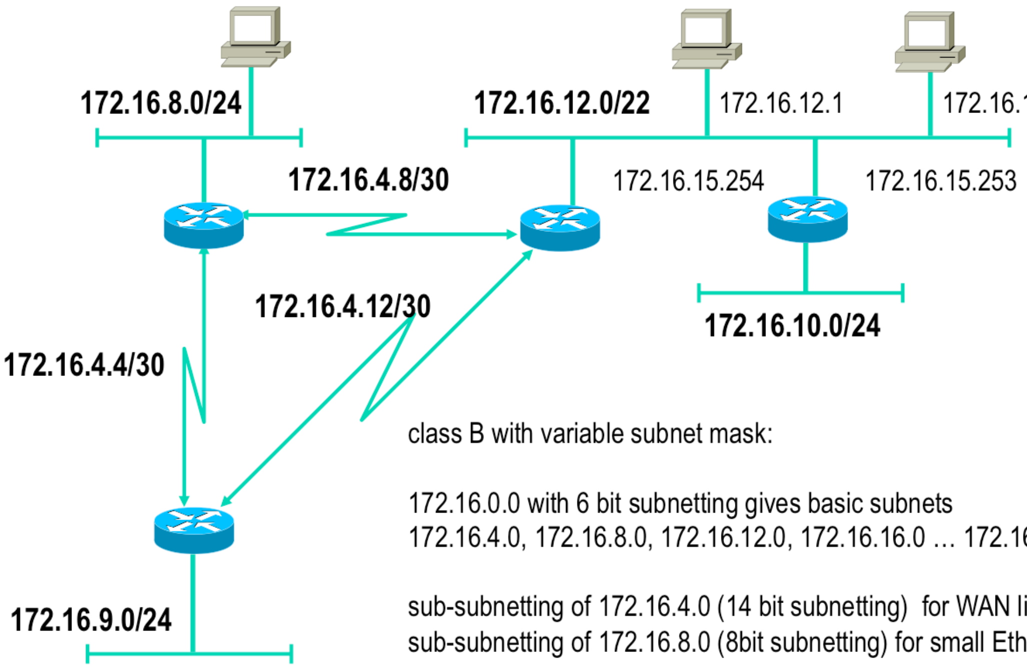


With earlier limitation, an organization is locked into a fixed number of fixed subnets. That is called classful routing. VLSM supports more efficient use of an organization's IP address space. VLSM was created in 1987. RFC 1009 defined how a subnetted network could use more than one subnet mask.

A short address design history:

1980	Classful Addressing (RFC 791)
1985	Subnetting (RFC 950)
1987	VLSM (RFC 1009)
1993	CIDR (Classless Interdomain Routing, RFC 1517 – 1520)

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

 IP Technology Primer, v6.4 66

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

© 2012/2017, D.I. Lindner / D.I. Haas

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

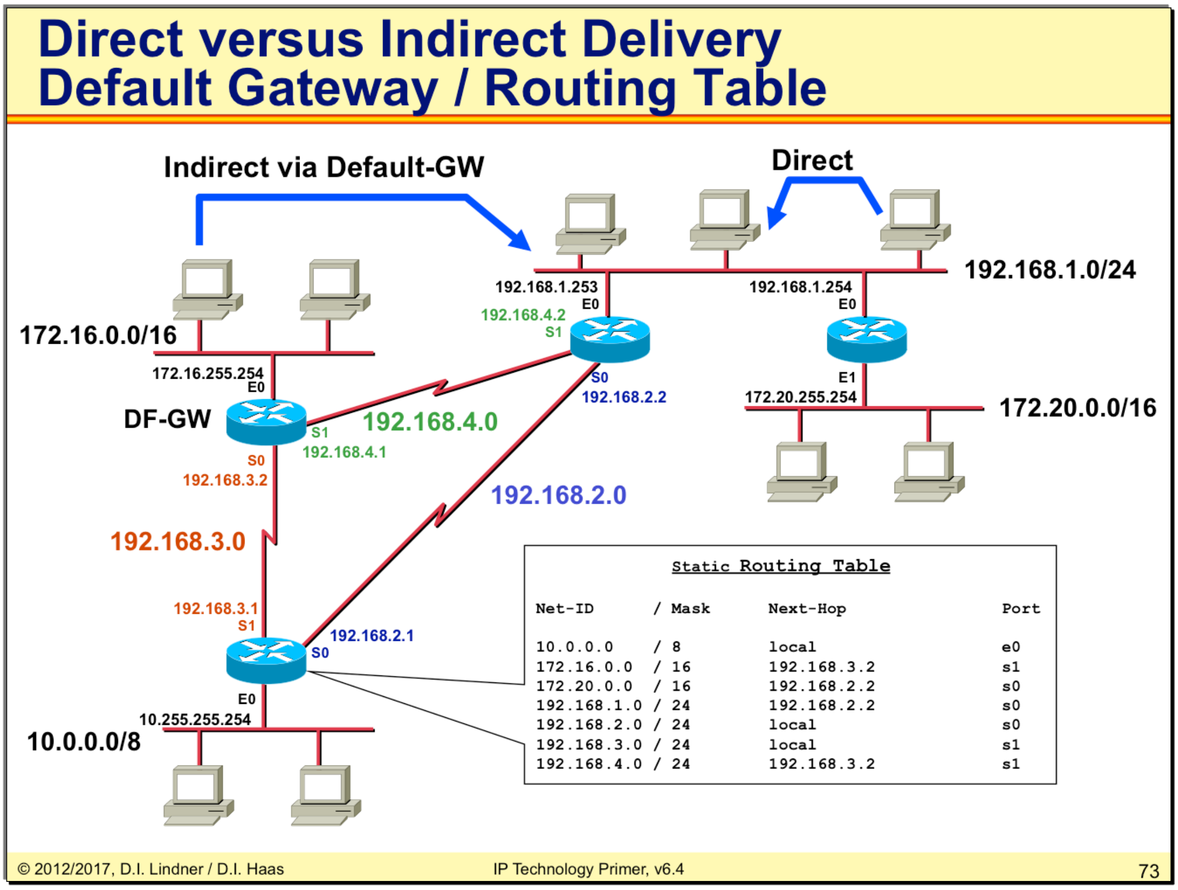
Agenda

- L2 versus L3 Switching
- IP Protocol, IP Addressing
- IP Forwarding
- 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

IP Technology (v6.4)



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

Routing table contains signpost as for every known (or specified) destination network:

net-ID / subnet-mask

next hop router (and next hop MAC address in case of LAN)

outgoing port

In the picture above there is small network, and a good example of a routing table. For example a host in network 10 want to send a datagram to a user in network 192.168.1. The destination address ≠ local address so the router must do a forward decision. The router compare the destination address with his routing table and found the right match (192.168.1.0/24 192.168.2.2 1 s0). Now he sends out the datagram via port s0 to the next hop, the router with the IP-Address of 192.168.2.2. This router is directly connected to the network 192.168.1.0. After an ARP-request the datagram is delivered to the right user.

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)

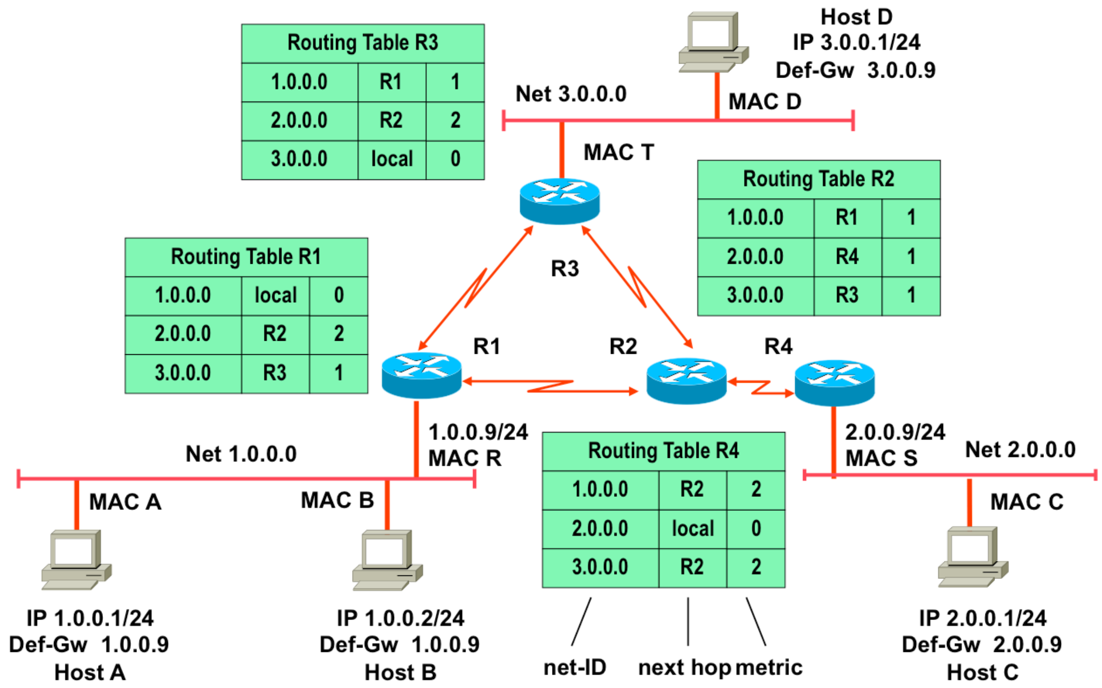
 IP Technology Primer, v6.4 74

Note that also simple workstations and PCs maintain routing tables—but not for routing pass-through packets, rather locally originated datagrams should be routed to the most reasonable next hop. Typically, the routing table consists only of a single entry, which is the default gateway for this host. But also additional entries can be made, indicating other gateways for some dedicated routes.

Additionally, an ARP cache must be maintained by a host. The ARP cache stores layer-2 MAC addresses and associated IP addresses of interfaces to which communication had occurred recently. Any ARP result is stored in this cache, thus subsequent packets to the same destination do not invoke the ARP each time. Per default the ARP cache is flushed after 20 minutes. Of course this value can be configured individually—even by DHCP.

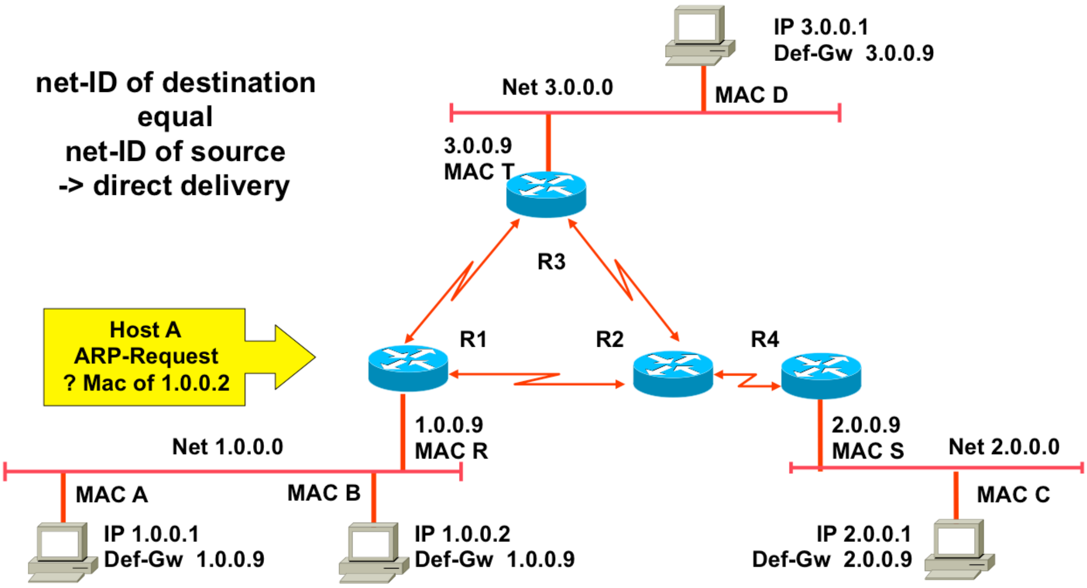
IP Technology (v6.4)

Example Topology



Direct Delivery 1.0.0.1 - > 1.0.0.2

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

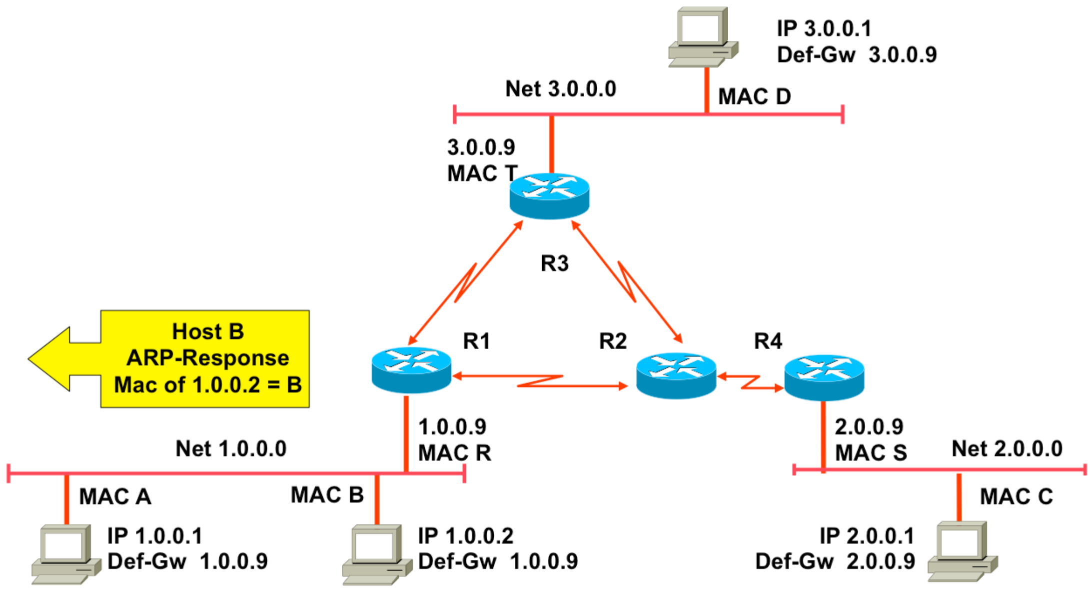


ARP ... Address Resolution Protocol

 IP Technology Primer, v6.4 76

IP Technology (v6.4)

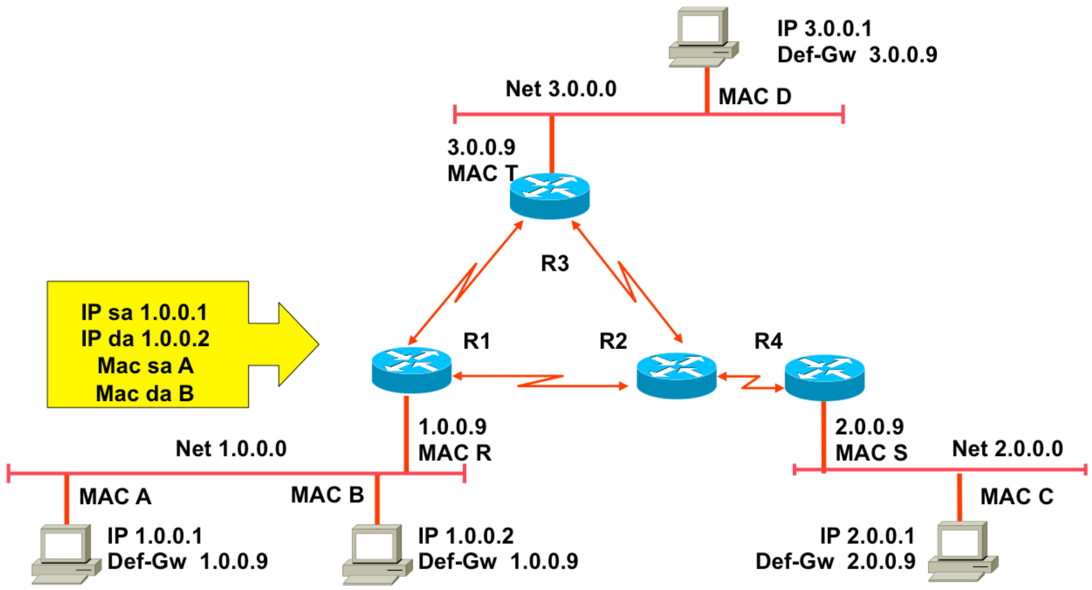
Direct Delivery 1.0.0.1 - > 1.0.0.2



ARP-Cache Host A	
1.0.0.2	MAC B

IP Technology (v6.4)

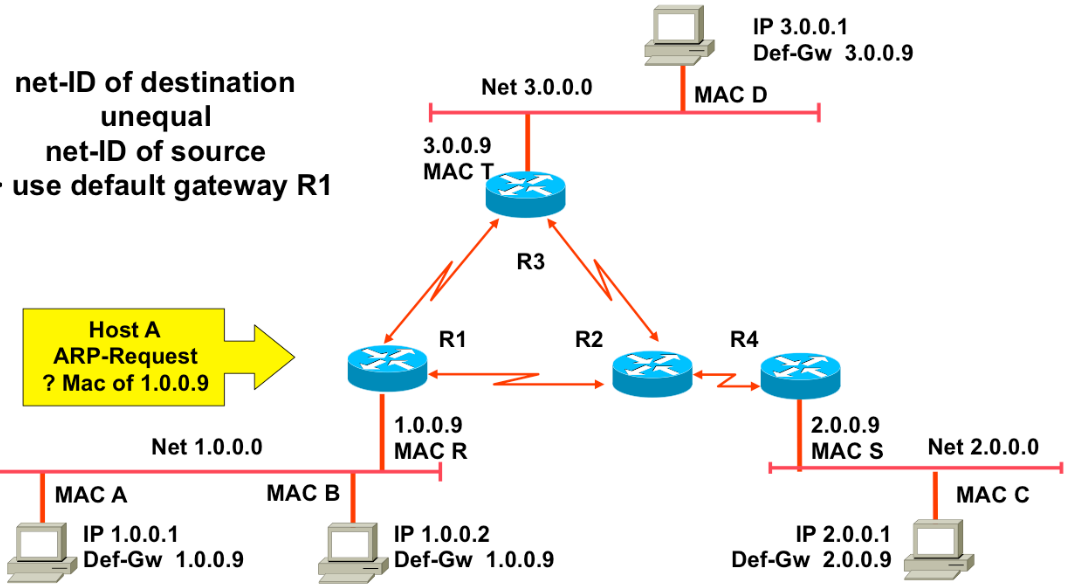
Direct Delivery 1.0.0.1 - > 1.0.0.2



ARP-Cache Host A	
1.0.0.2	MAC B

Indirect Delivery 1.0.0.1 - > 2.0.0.1

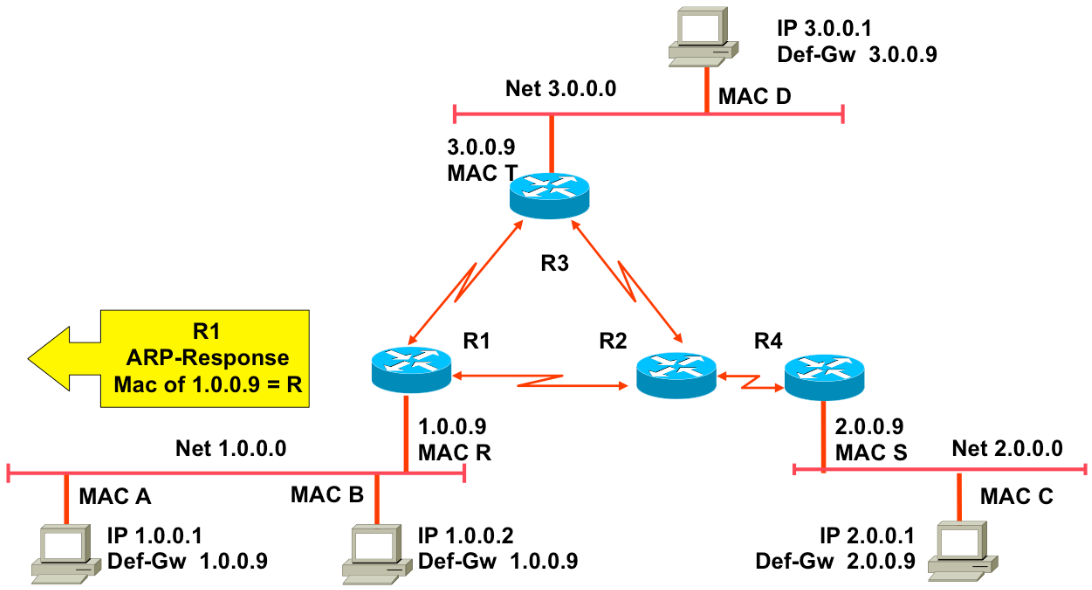
net-ID of destination
unequal
net-ID of source
-> use default gateway R1



ARP-Cache Host A	
1.0.0.2	MAC B

IP Technology (v6.4)

Indirect Delivery 1.0.0.1 - > 2.0.0.1

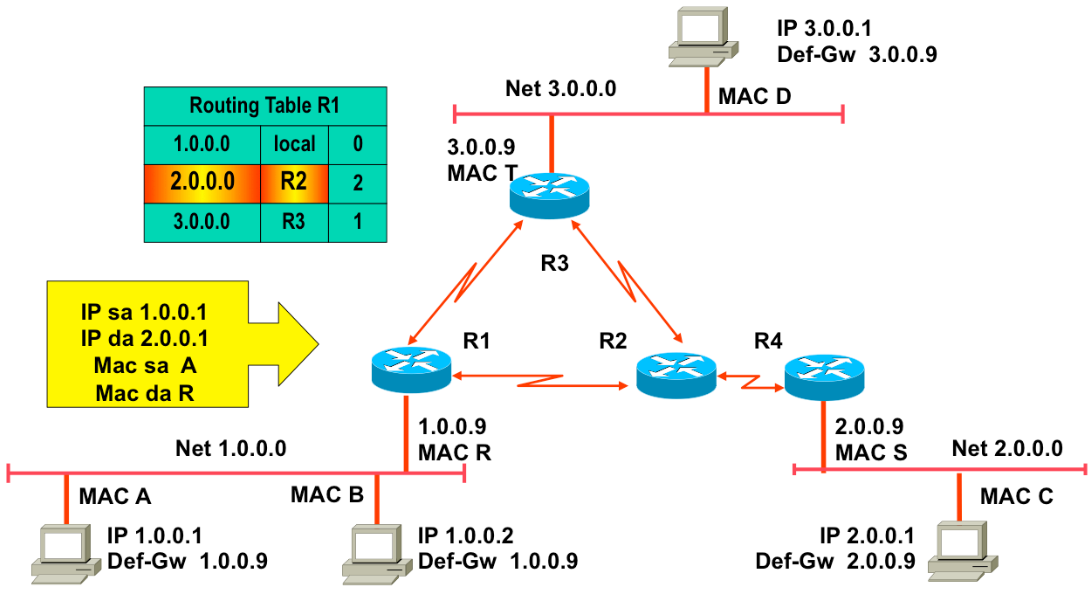


ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

 IP Technology Primer, v6.4 80

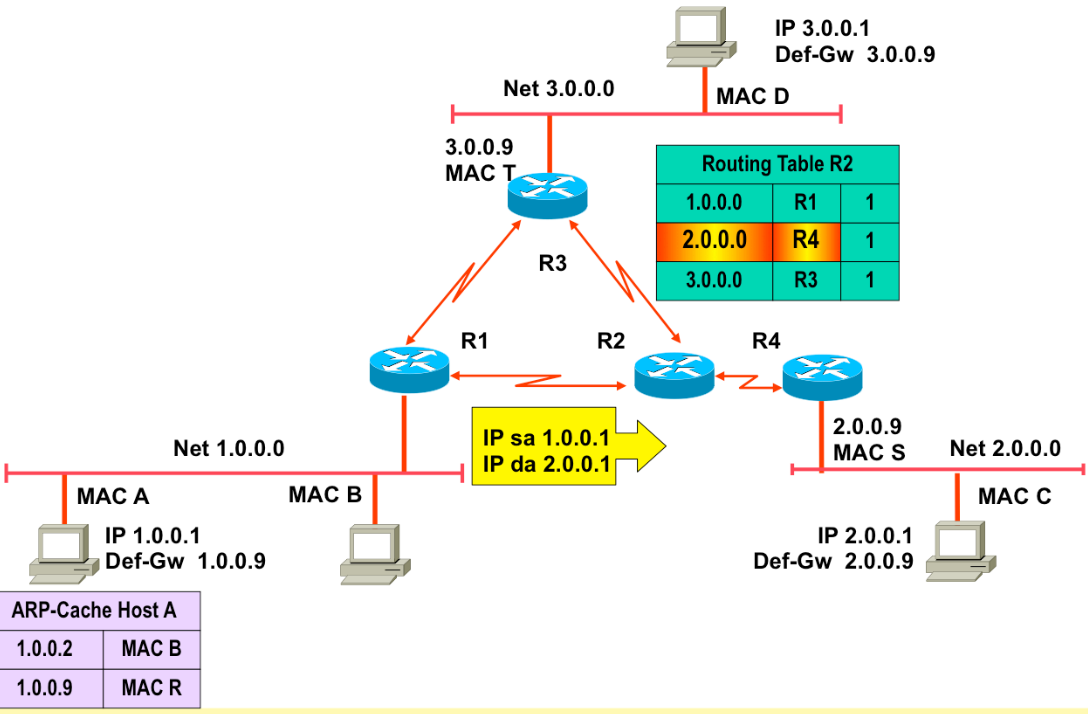
Indirect Delivery 1.0.0.1 - > 2.0.0.1

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



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

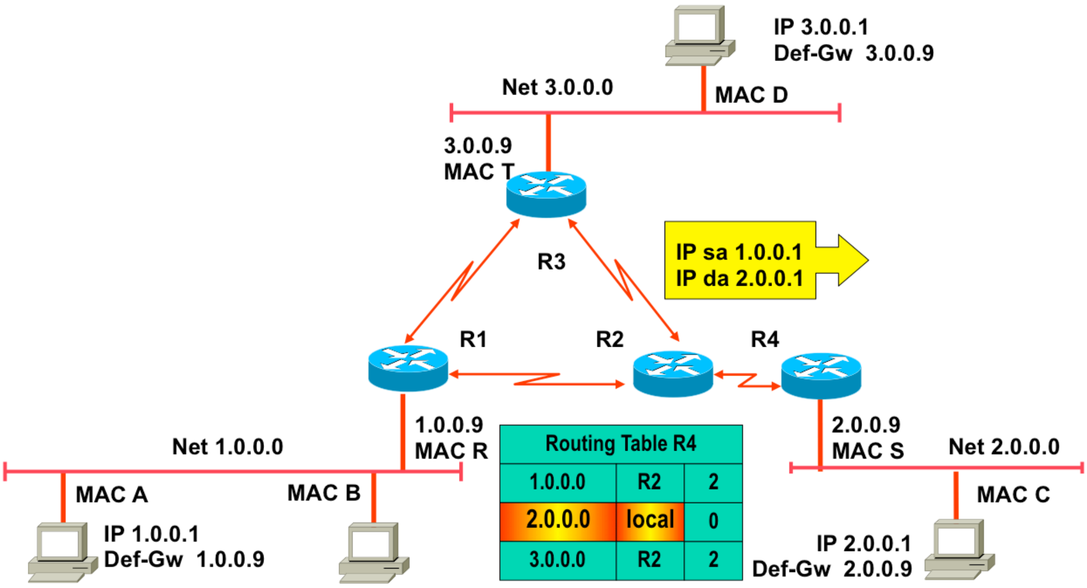


Routing Table R2		
1.0.0.0	R1	1
2.0.0.0	R4	1
3.0.0.0	R3	1

ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

IP Technology (v6.4)

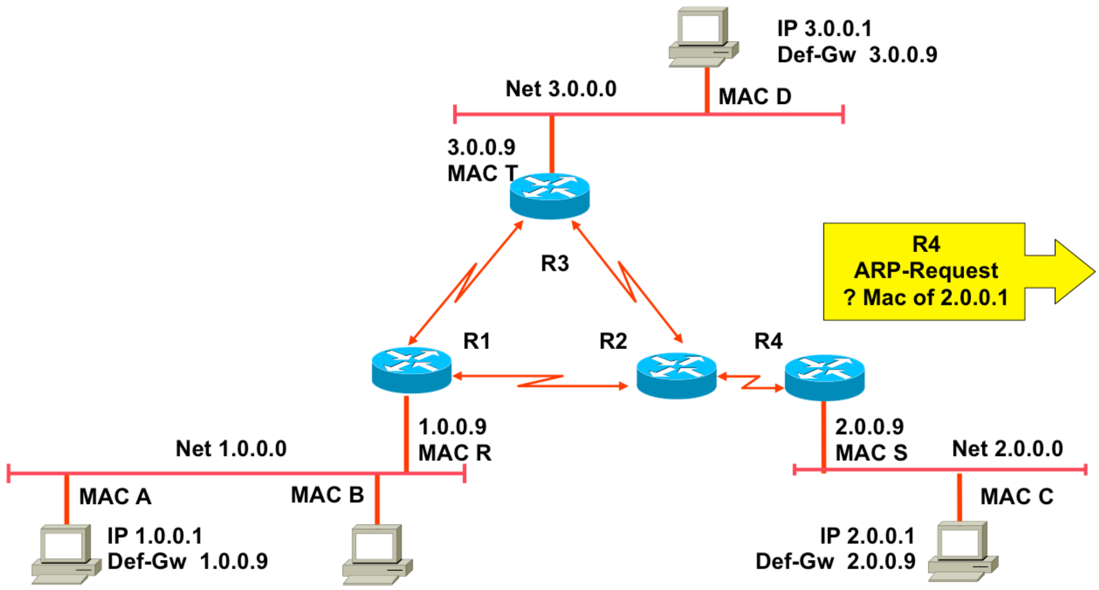
Indirect Delivery 1.0.0.1 - > 2.0.0.1



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

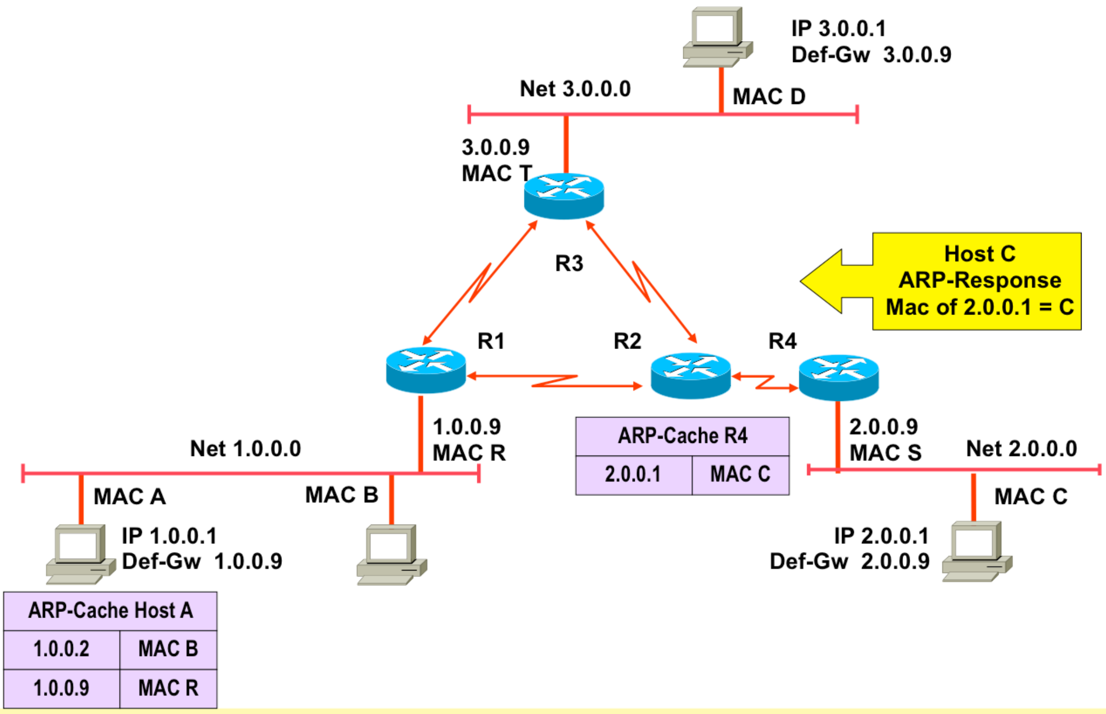
Indirect Delivery 1.0.0.1 - > 2.0.0.1



ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

IP Technology (v6.4)

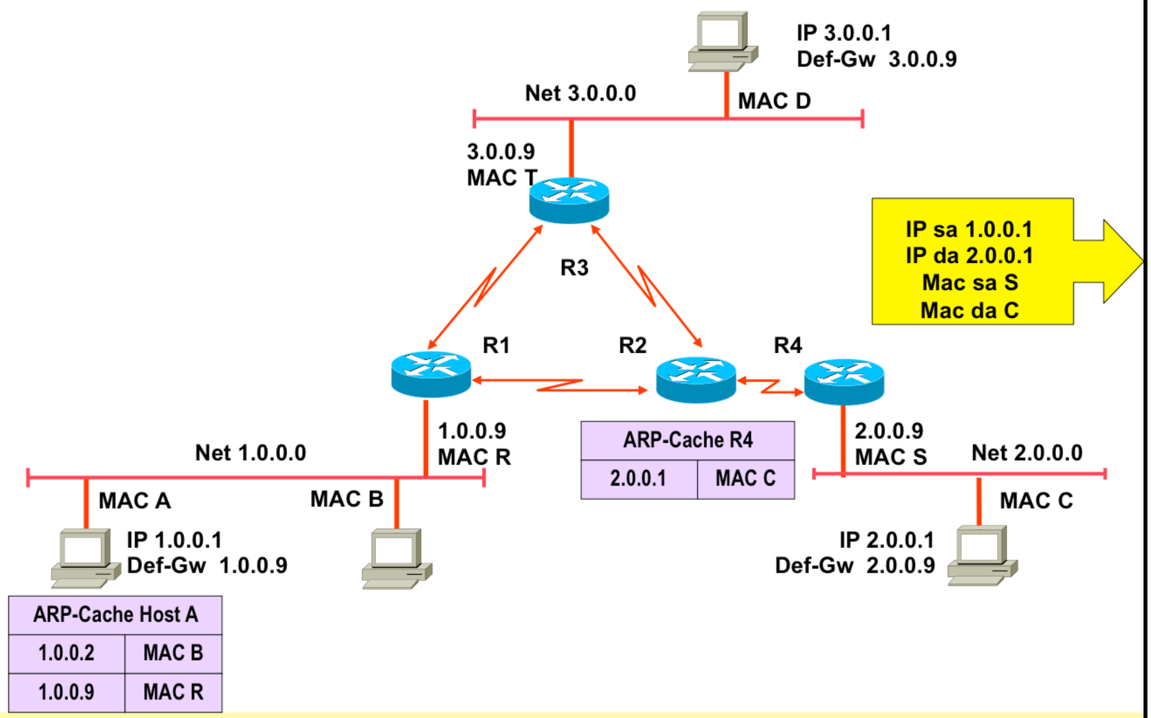
Indirect Delivery 1.0.0.1 - > 2.0.0.1



© 2012/2017, D.I. Lindner / D.I. Haas   IP Technology Primer, v6.4   85

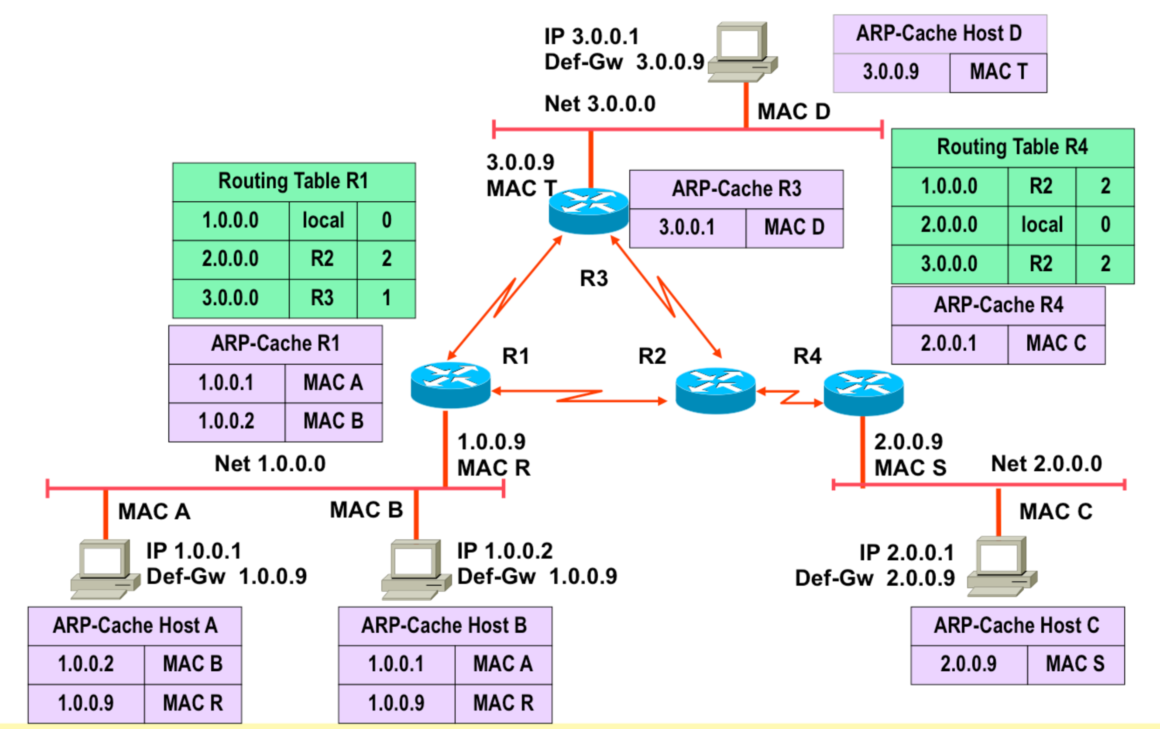
IP Technology (v6.4)

Indirect Delivery 1.0.0.1 - > 2.0.0.1



 IP Technology Primer, v6.4 86

ARP Cache - Final Picture



Agenda

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

IP Address versus L2 Address

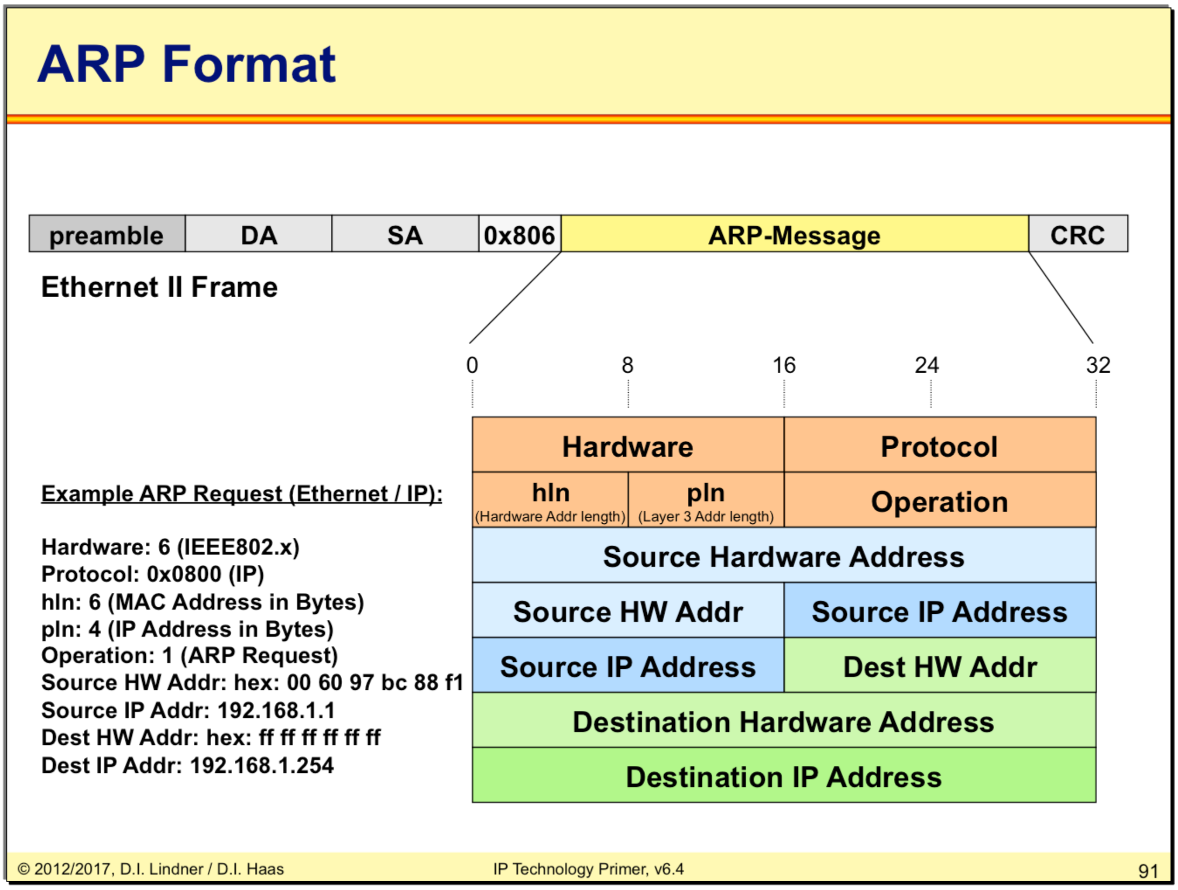
- **IP address**
  - Identifies the access to a network (interface)
- **If the physical network is of point-to-point link to another IP system**
  - This IP system can be reached without any further addressing on layer 2
- **On a shared media or multipoint-network**
  - Layer 2 addresses are necessary 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**

 IP Technology Primer, v6.4 89

On a multipoint network every station needs a layer-2 address. When IP packets should be sent to a local destination the sender must first determine the corresponding layer-2 address. A multipoint network is also known as a shared medium. It could be a broadcast domain (like Ethernet) or not (like Frame-Relay or ATM). Therefore the layer-2 address could be a MAC address, a DLCI (Frame-Relay) or similar. In this chapter we only focus on Ethernet only.

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 messages are carried within Ethernet II frames or SNAP encapsulation using type field 0x806. ARP has been designed to support different layer 3 protocols (IP is just one of them).

Hardware: Defines the type of network hardware, e.g.:

1	Ethernet DIX
6	802.x-LAN
7	ARCNET
11	LocalTalk

Protocol: Identifies the layer 3 protocol (same values as for Ethertype, e.g. 0x800 for IP)

hln: Length of hardware address in bytes

pln: Length of layer 3 address in bytes

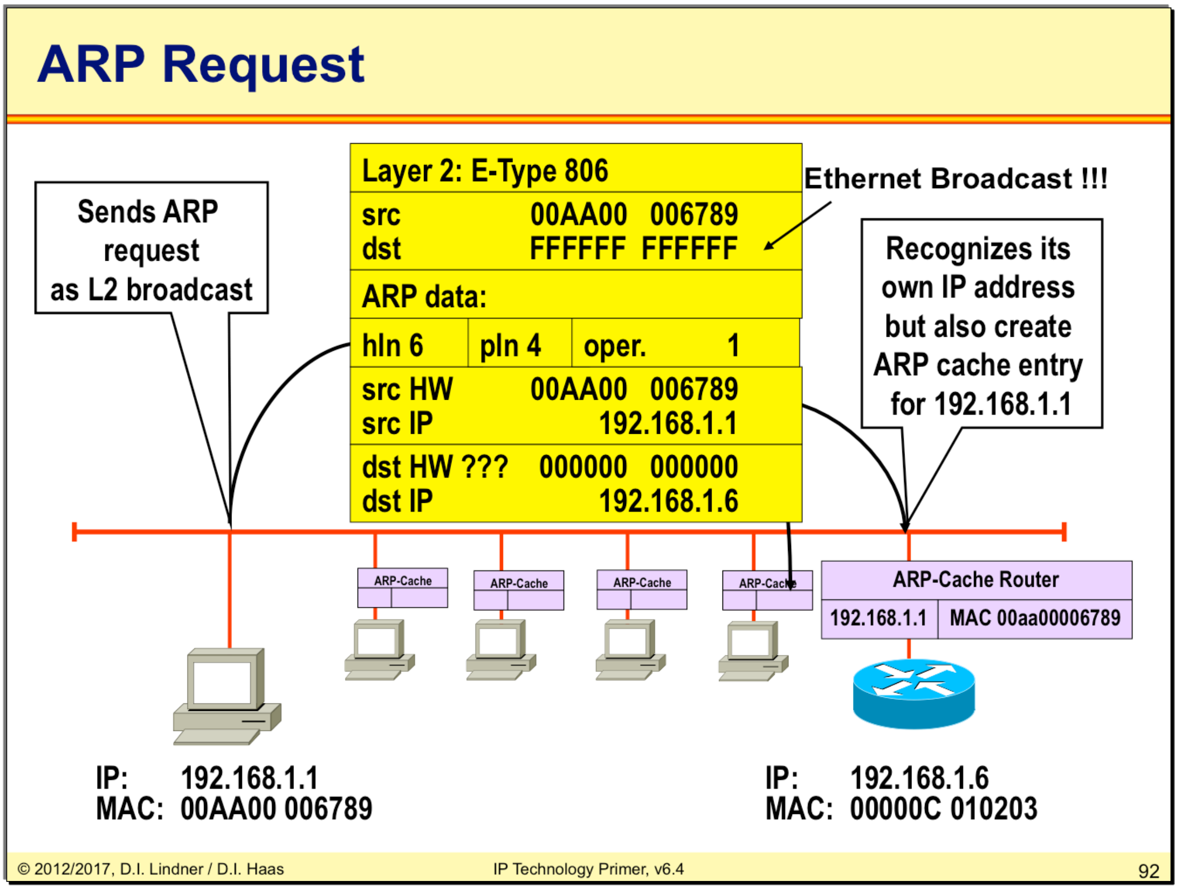
Operation:

1 .... ARP Request
2 .... ARP Response
3 .... RARP Request
4 .... RARP Response

Addresses:

Hardware addresses: MAC addresses (source and destination).
IP addresses: layer 3 addresses (source and destination).

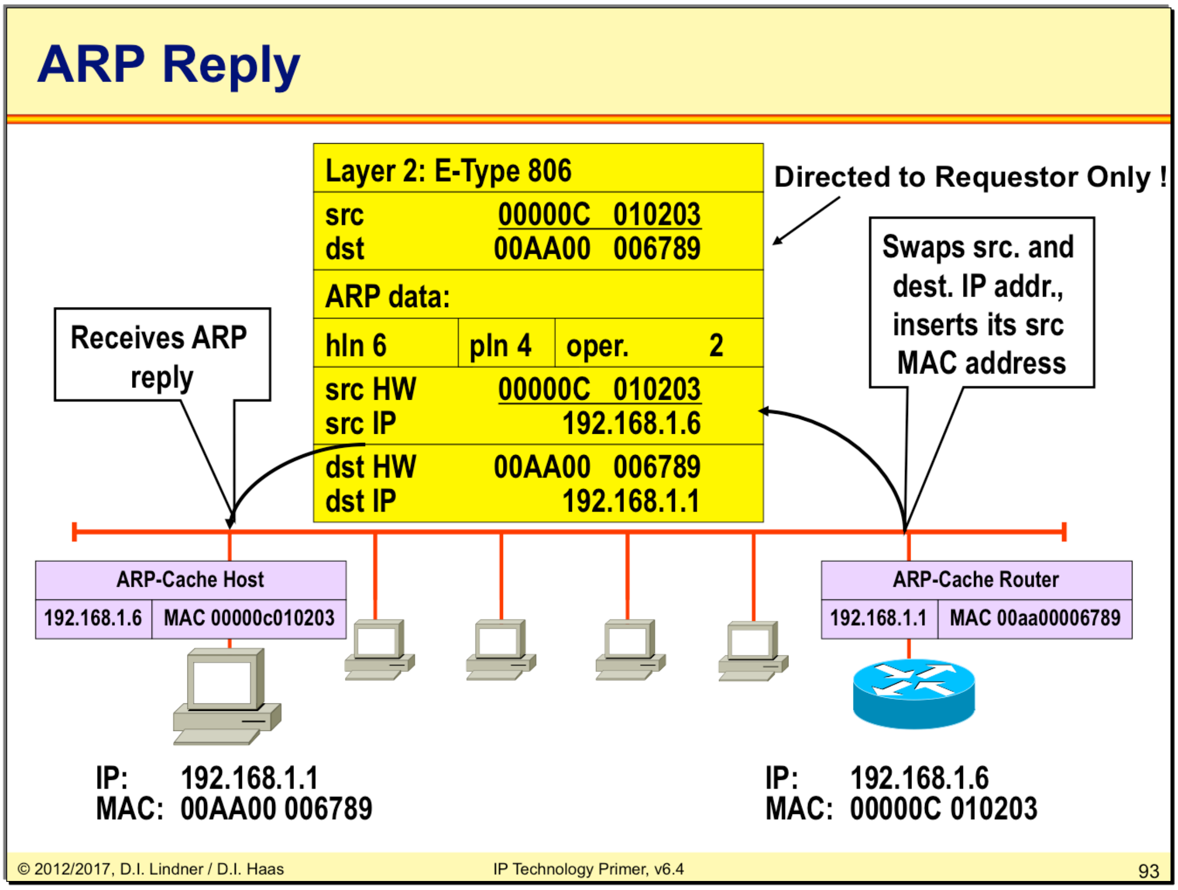
ARP request and responses are not forwarded by routers (only L2 messages)



Operation of ARP:

Station A (192.168.1.1) wants to send an IP datagram to station B (192.168.1.6) but doesn't know the MAC address (both are connected to the same LAN). A sends an ARP request in form of a MAC broadcast (destination = FF, source = Mac_A), ARP request holds IP address of B. Station B and all other stations connected to the LAN see the ARP request with its IP address: B and all other stations store the newly learned mapping (source MAC- and IP-address of A) into their ARP caches.

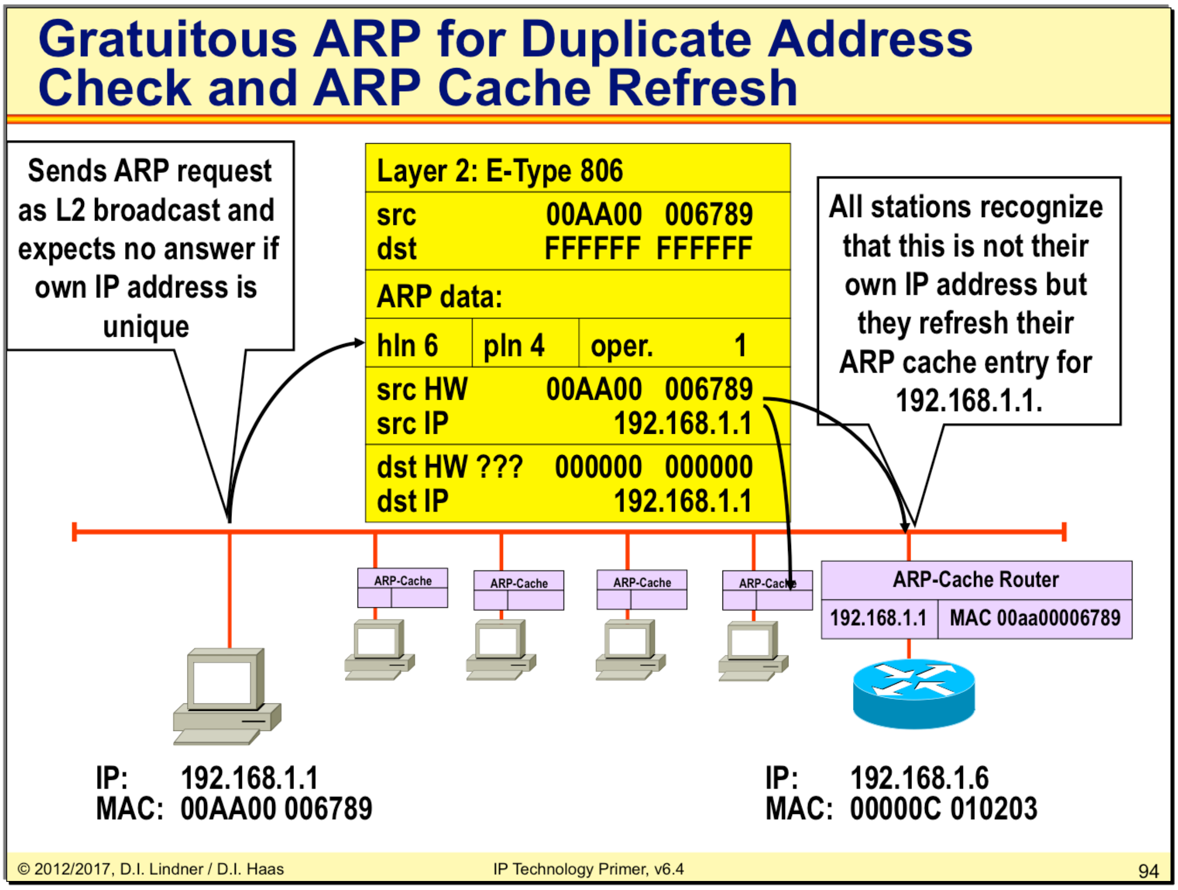
IP Technology (v6.4)



Now station B sees sends an ARP response as a directed MAC frame (SA=Mac_B, DA=Mac_A).The ARP response holds MAC address of station B. A stores the MAC- / IP-address mapping for station B in its ARP cache.

For subsequent IP datagrams from A to B or from B to A the MAC addresses are taken from the ARP cache (no further ARP request / response are necessary).

Entries in the ARP cache are deleted if they aren't used for a defined period (usually 20 minutes), this aging mechanism allows for changes in the network and saves table space.



Gratuitous ARP is an ARP request where an IP station asks for address resolution of its own IP address.

This is typically used:

1. For detecting duplicate IP addresses on the connected LAN.
2. For refreshing the ARP caches of the other IP systems before the ARP caches times out.
3. For actualizing the ARP caches of the other IP systems in case the IP systems has changed the MAC address (e.g. change of Ethernet card).

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

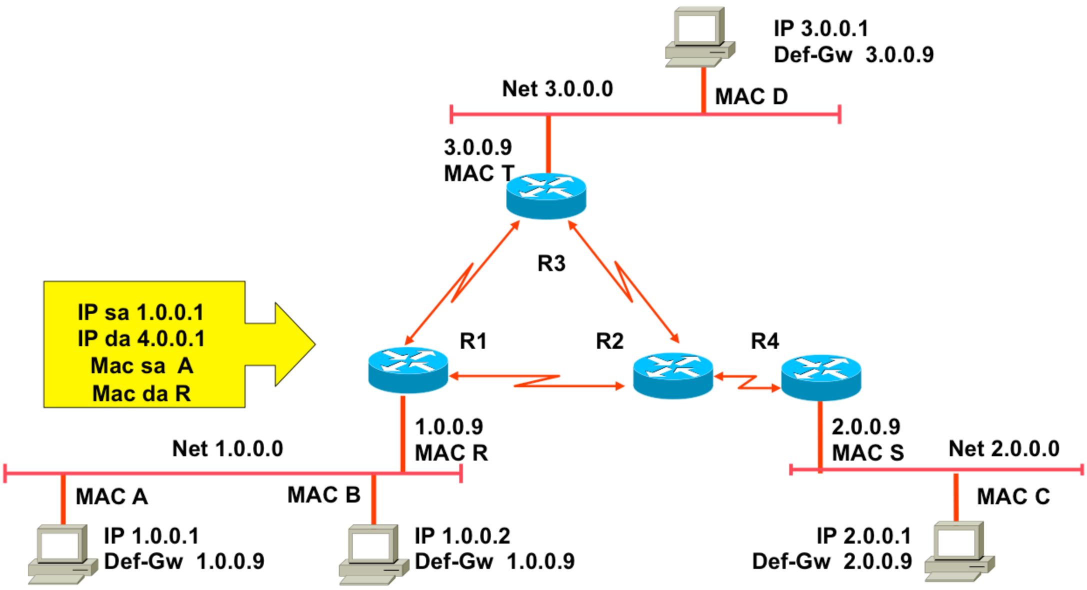
 IP Technology Primer, v6.4 97

Using ICMP Types:

0, 8	"PING" testing whether an IP station (router or end system) can be reached and is operational
3, 11, 12	Signaling errors concerning reachability, TTL / reassembly timeouts and errors in the IP header
4	Flow control (only possibility to signal a possible buffer overflow)
5	Signaling of alternative (shorter) routes to a target
13 - 18	Diagnosis or management

IP Technology (v6.4)

Delivery 1.0.0.1 - > 4.0.0.1

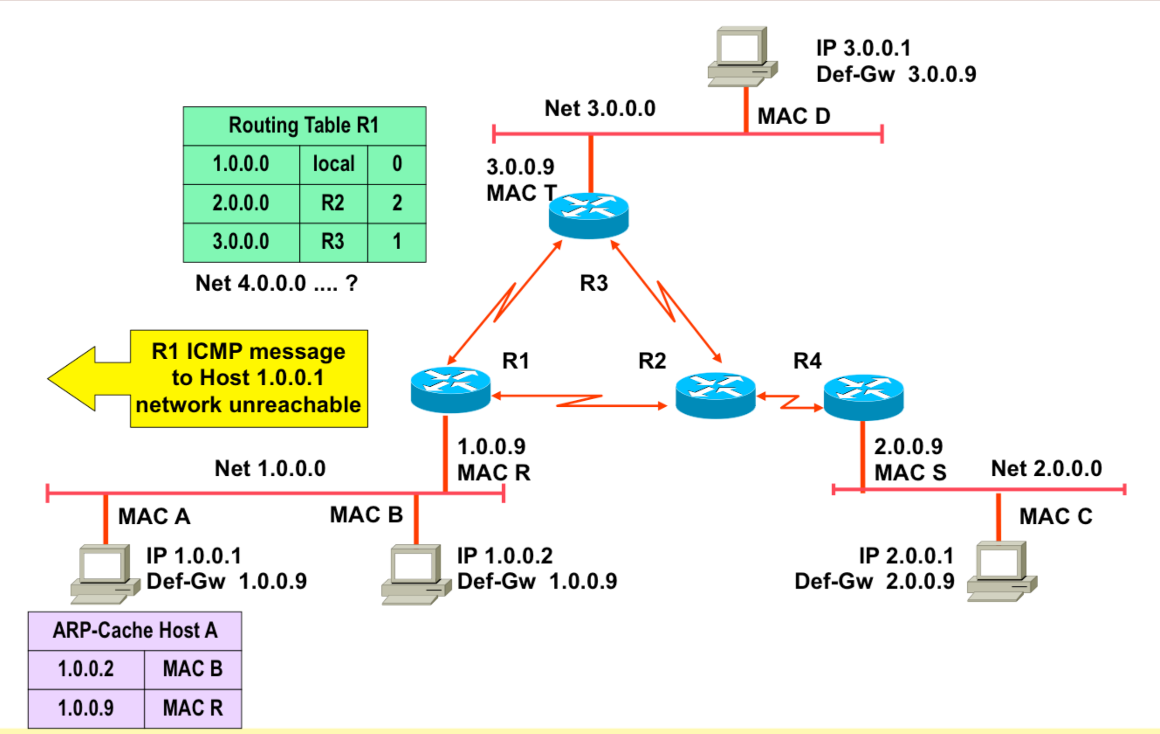


ARP-Cache Host A	
1.0.0.2	MAC B
1.0.0.9	MAC R

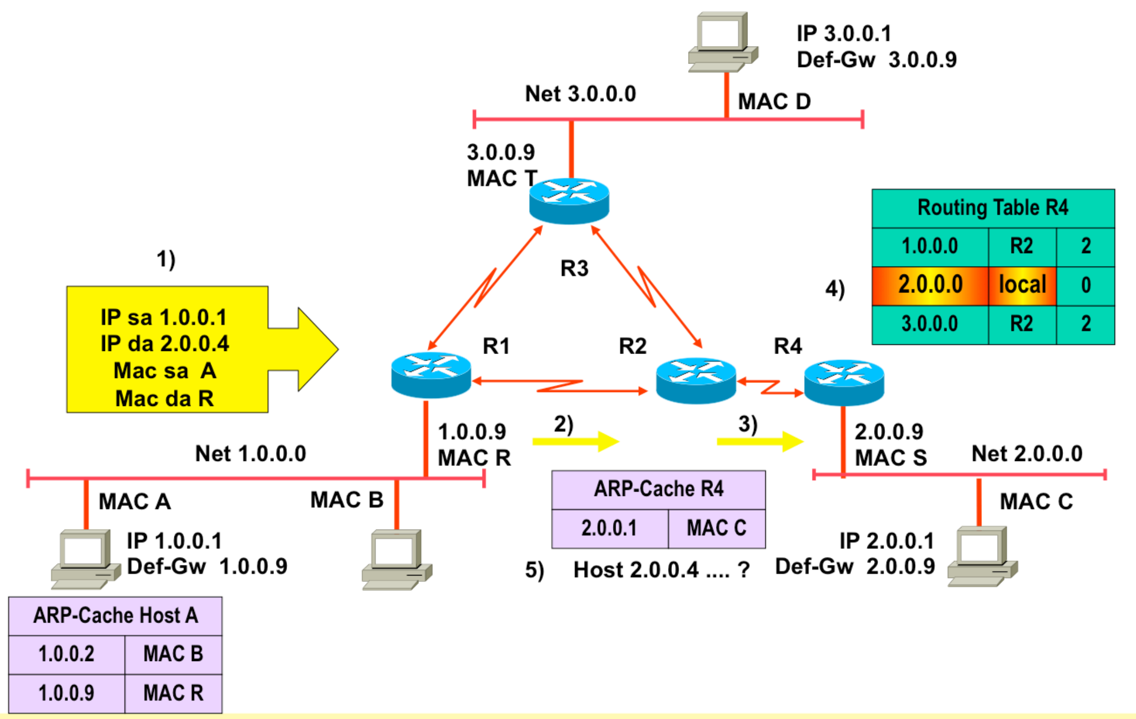
 IP Technology Primer, v6.4 98

IP Technology (v6.4)

ICMP Destination Unreachable (code: network unreachable)

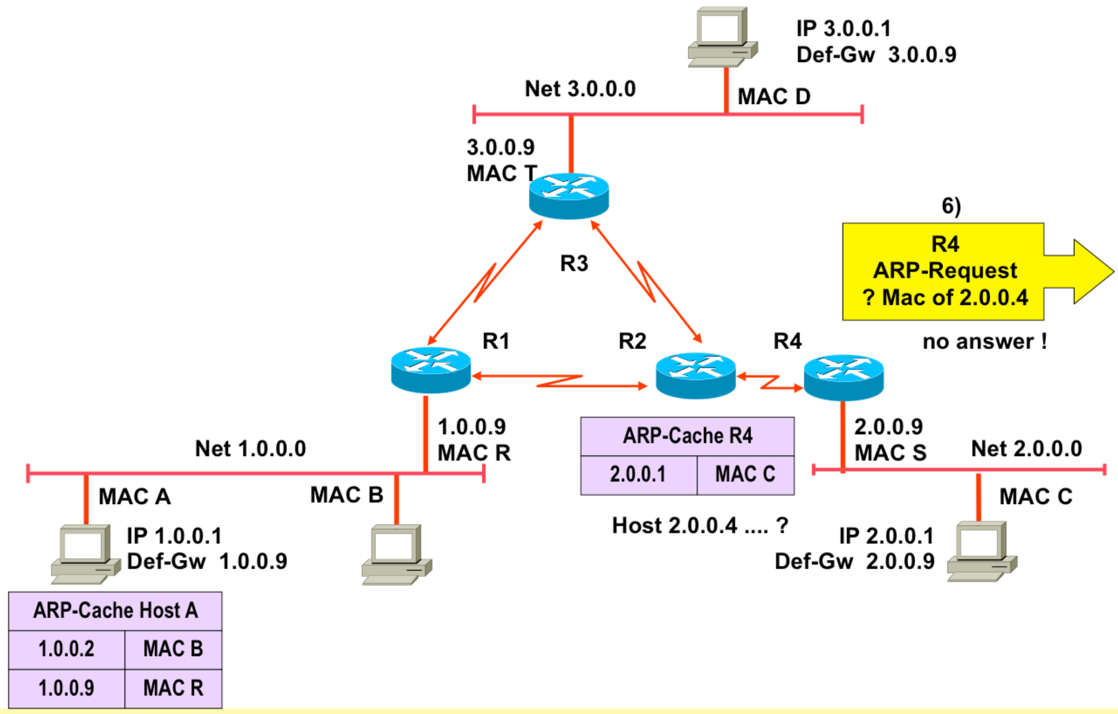


Delivery 1.0.0.1 - > 2.0.0.4



IP Technology (v6.4)

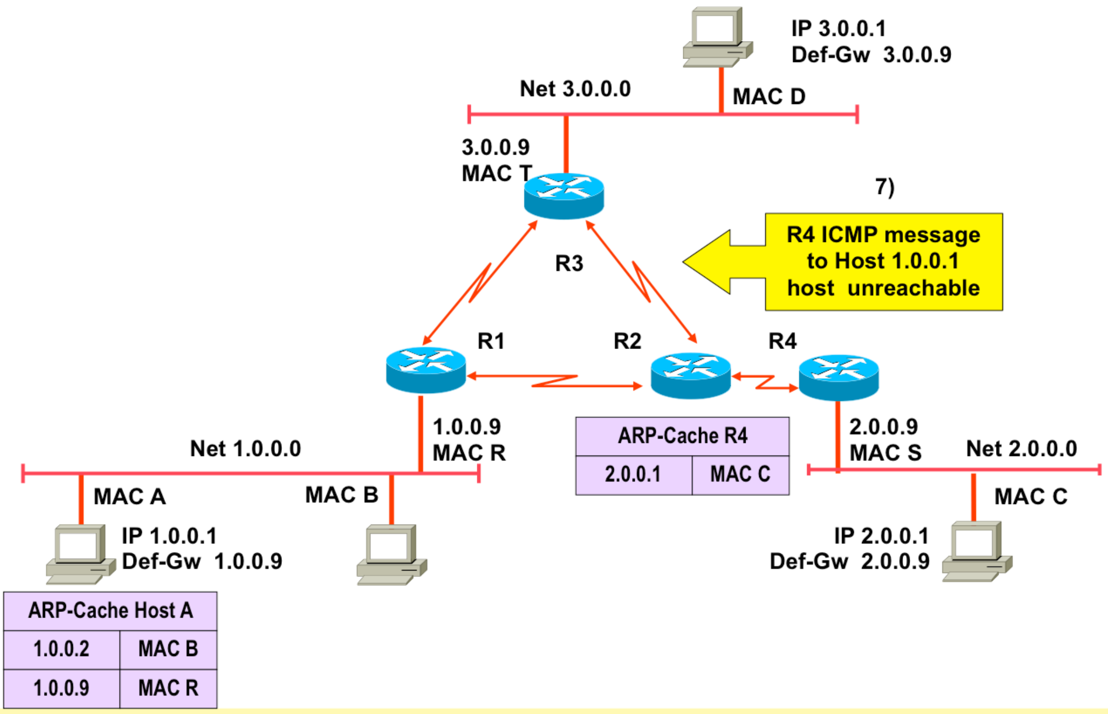
Delivery 1.0.0.1 - > 2.0.0.4



 IP Technology Primer, v6.4

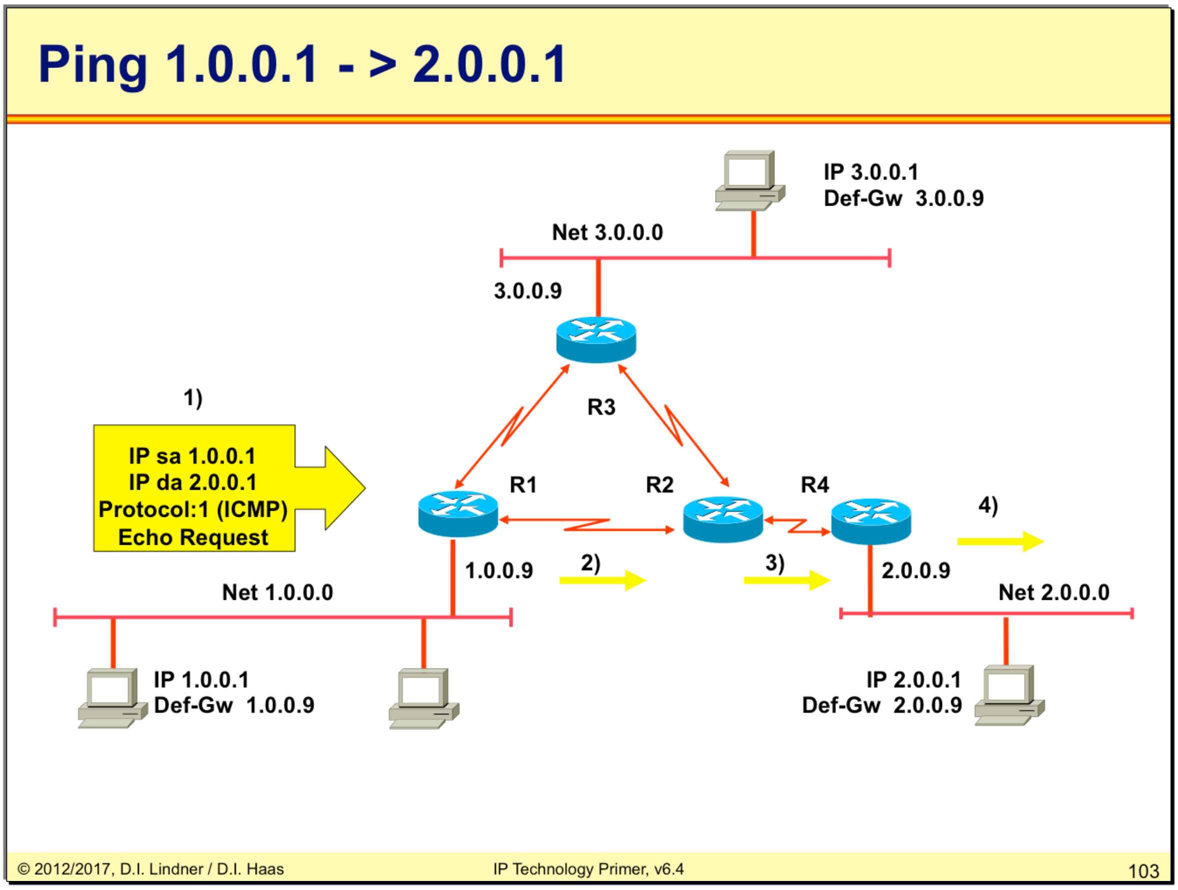
IP Technology (v6.4)

ICMP Destination Unreachable (code: host unreachable)



 IP Technology Primer, v6.4 102

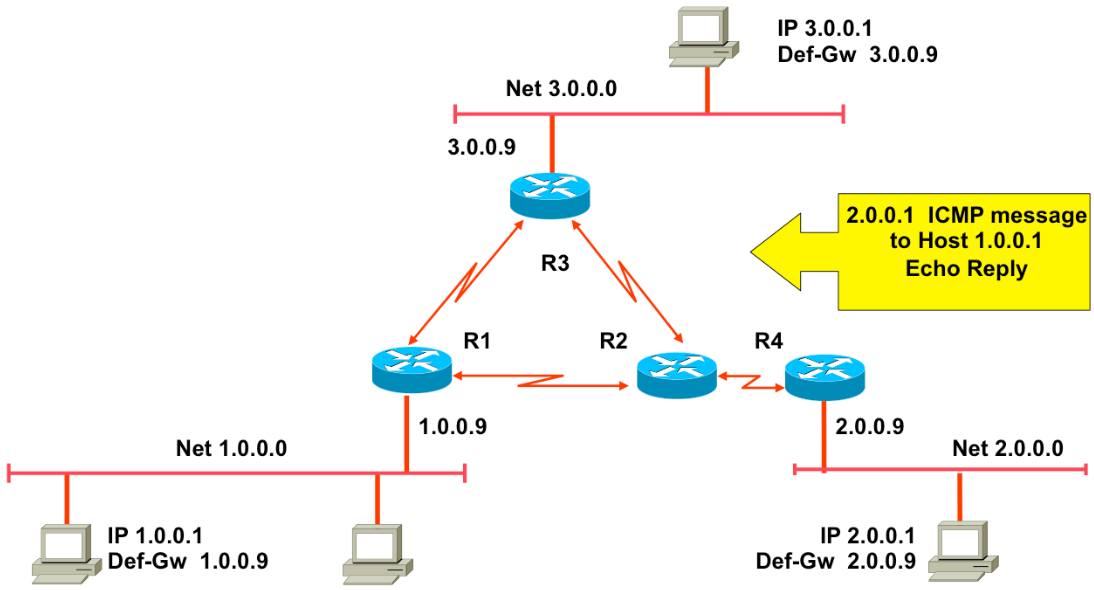
IP Technology (v6.4)



PING - Packet Internet Groper:

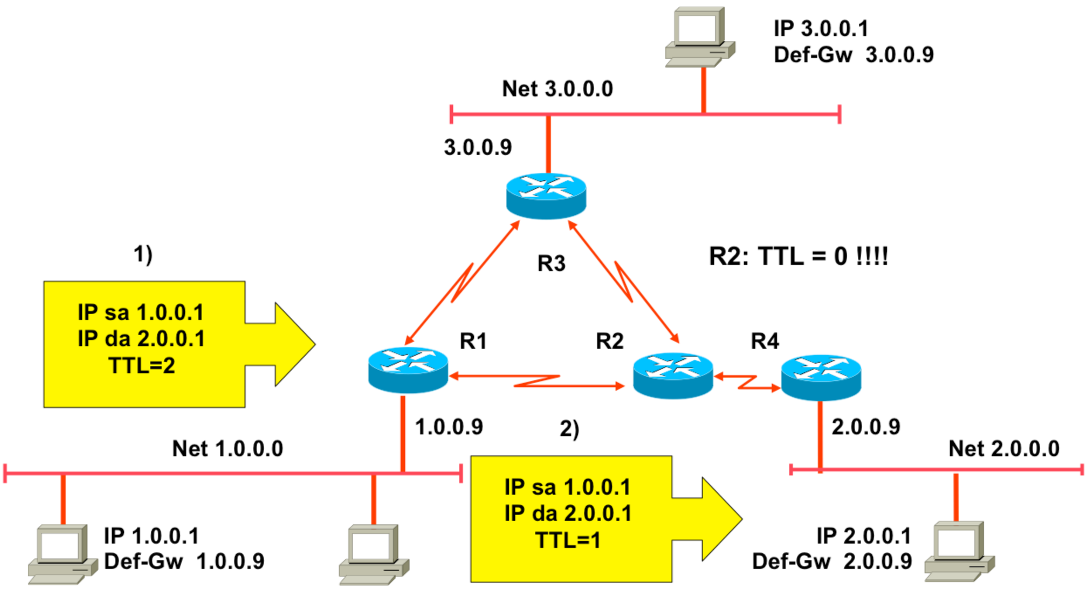
Checks the reachability of an IP station several times in a sequence and measures answer time for each trial. In case the station is reachable you get an indication about the round-trip-delay in the network. If station is not reachable the trial times out after e.g. two seconds.

Ping Echo 2.0.0.1 - > 1.0.0.1



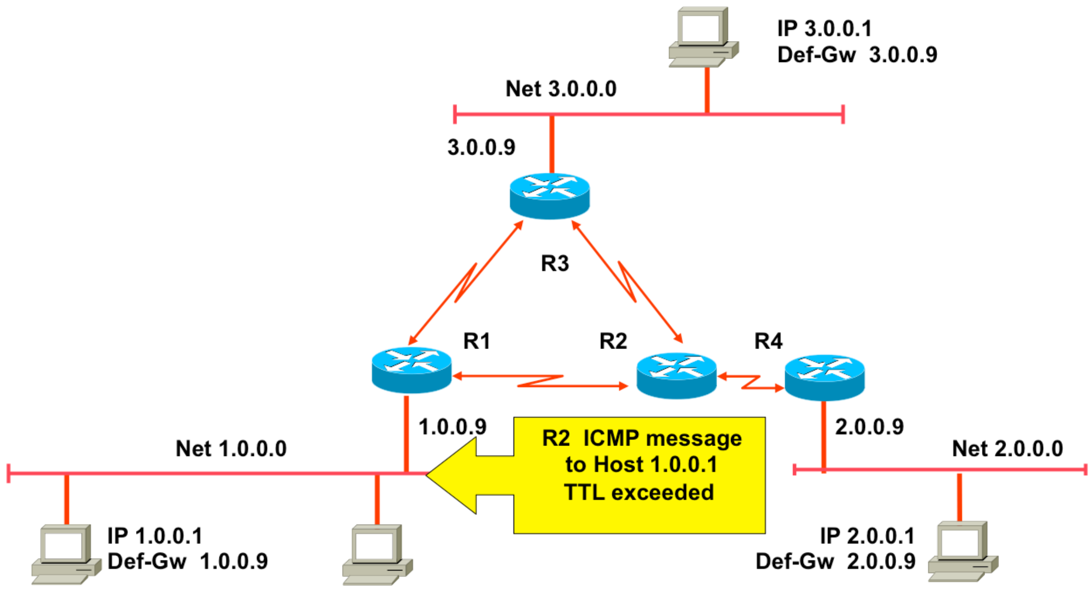
IP Technology (v6.4)

Delivery 1.0.0.1 - > 2.0.0.1 (TTL=2)



IP Technology (v6.4)

ICMP 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

 IP Technology Primer, v6.4 107

UDP segment and manipulation of the TTL field (time to live) of the corresponding IP header is used to generate ICMP error messages TTL exceeded or UDP port not reachable. UDP segments with undefined port numbers (> 30000) are used. A simple ICMP Echo requests with TTL manipulation may not work because either after reaching the final IP host no TTL exceeded message will be generated by the destination host (this is done by routers only) or it might be blocked by the host firewall of the destination.

Traceroute operation example:

UDP datagram with TTL=1 is sent for three times

UDP datagram with TTL=2 is sent for three times

....

The routers in the path generate ICMP time exceeded messages because TTL reaches 0.

If the UDP datagram arrives at the destination, an ICMP port unreachable message is generated.

From the source addresses (= router address) of the ICMP error messages the path can be reconstructed.

The IP addresses are resolved to names by using DNS.

tracert 140.252.13.65

1 ny-providerx-int-99 (139.128.3.99)	20ms	10ms	10ms
2 sf-providery-int-23 (172.252.12.21)	20ms	10ms	10ms
2 www.example.com (140.252.13.65)	*	120ms	120ms

Output of "*", if no answer arrives within 5 seconds.

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?

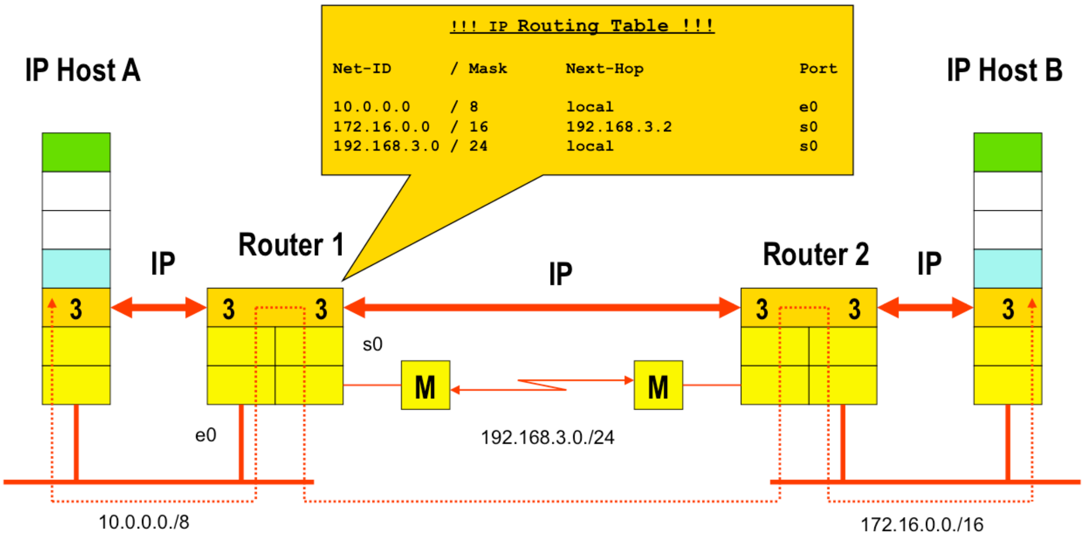
- ***Finding / choosing a path to a destination address***
- **Direct delivery performed by IP host**
  - Destination network = local network
- **Indirect delivery performed by router**
  - Destination network ≠ local network
  - Datagram is forwarded to default gateway
  - Passed on by the router based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

 IP Technology Primer, v6.4 109

Routing is the process of choosing a path over which to send IP datagrams destined to a given destination address. There are 2 ways to delivery a packet. The direct delivery and the indirect delivery. IP hosts are responsible for direct delivery of IP datagrams whereas routers are responsible for selecting the best path in a meshed network in case of indirect delivery of IP datagrams. IP hosts are further responsible for choosing a default router ("default gateway") as next hop in case of indirect delivery of IP datagrams. When there is a direct delivery (destination network = local network) the host makes for example an ARP-request (Ethernet) and then delivery the datagram to the right host. If there is a indirect delivery (destination network ≠ local network) the IP host forwards the datagram to its default gateway.

IP, IP Routing Protocol, IP Routing Table

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



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 Technology Primer, v6.4 110

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

 IP Technology Primer, v6.4 111

The IP routing paradigm is fundamental in IP routing. Firstly, IP routing is "destination based routing", that means the source IP address is never examined during the routing process. Secondly, IP routing is "hop-by-hop", which emphasizes the difference to virtual circuit principles. The routing table in every router within the autonomous system must be both accurate and up to date (consistent and loop-free) so that datagrams can be directed across the network to their destination.

In IP the path of a packet is not pre-defined and not connection oriented, rather each single router performs a routing decision for each datagram. Thirdly, IP routing is "least cost" in that only that path with the lowest metric is selected in case of multiple redundant paths to the same destination.

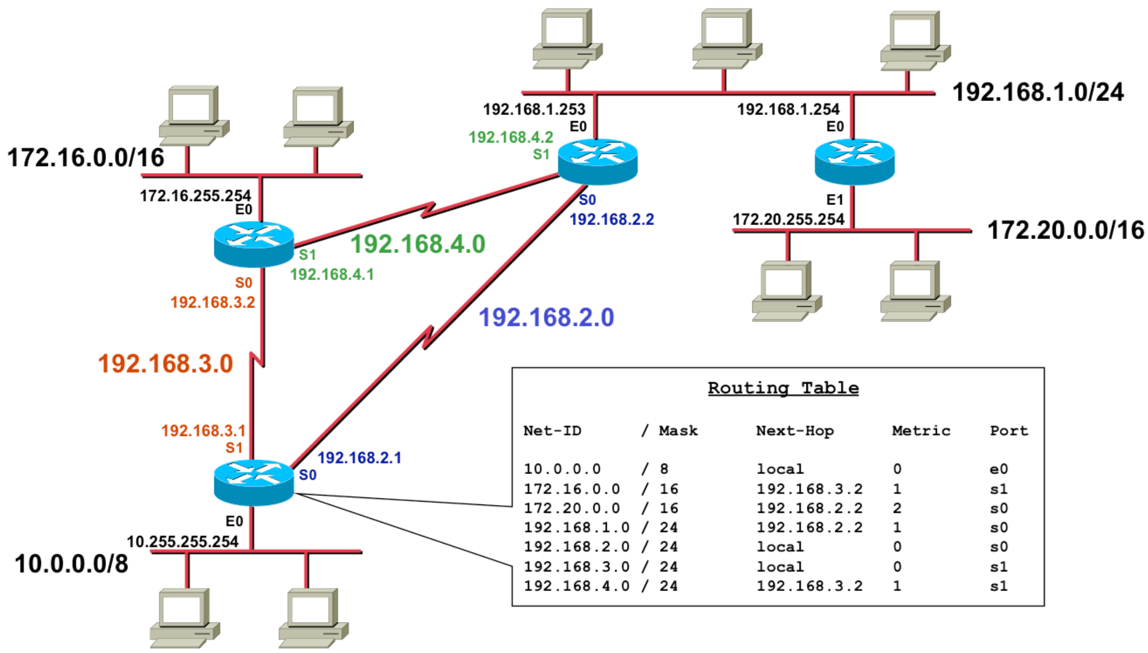
Note that several vendors extend these rules by providing additional features, but the routing paradigm generally holds for most of the routers in the Internet, at least for the basic routing processes.

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

 IP Technology Primer, v6.4 113

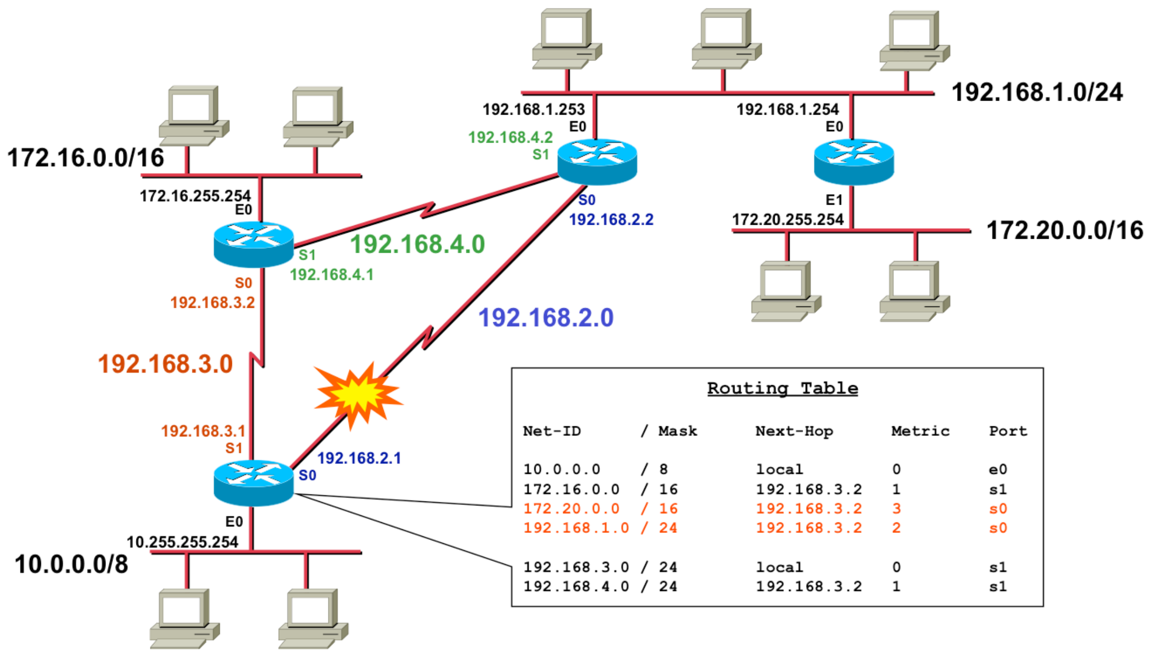
Now we see some additional fields in the a routing table bulit by a dynamic routing protocol (in our case RIP with hop counts is assumed):

Routing table contains signpost as for every known (or specified) destination network:

- net-ID / subnet-mask
- next hop router (and next hop MAC address in case of LAN)
- outgoing port
- metric (information how far away is a certain destination network) -> hop counts in our picture
- time reference (information about the age of the table entry)

IP Technology (v6.4)

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

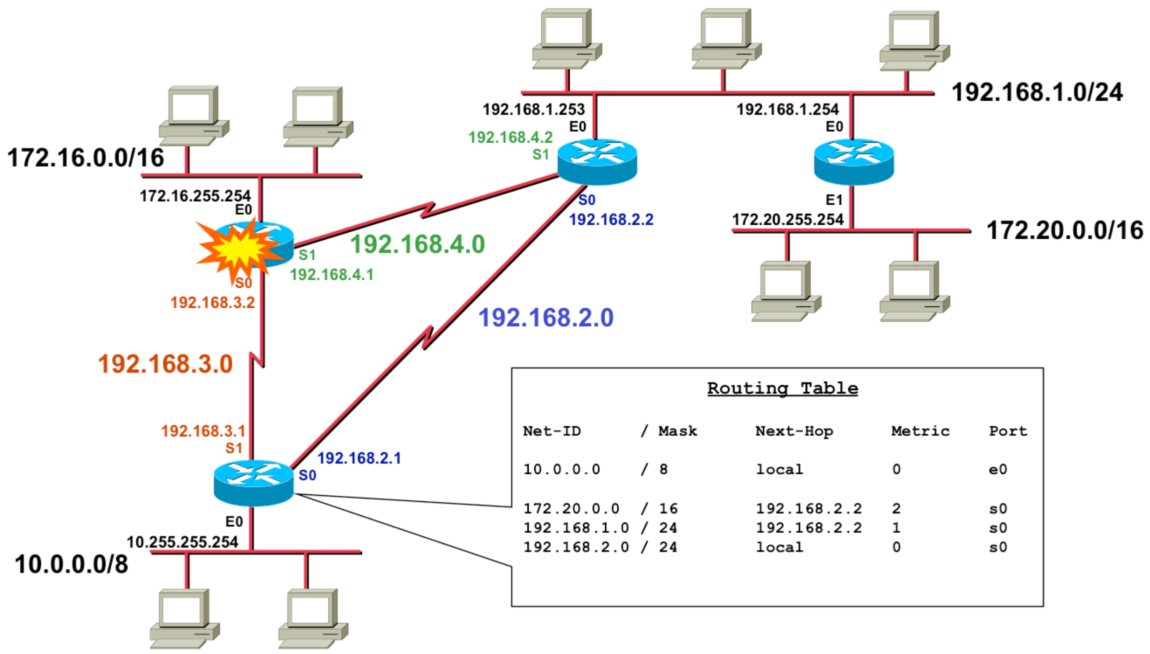
 IP Technology Primer, v6.4 114

What can a dynamic routing protocol detect?

Loss of a link between any two directly connected routers
Loss of a router connected in a meshed network
Loss of network directly connected to a router

In our example loss of link 192.168.2.0 causes adaption of the routing table hence traffic from 10.0.0.0 to 192.168.1.0 or 172.20.0.0 will take the alternate = only remaining path via 192.168.3.2. Hop count to these networks has risen by one. If link 192.168.2.0 comes back the dynamic routing will adapt back to picture of last slide.

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

 IP Technology Primer, v6.4 115

In our example loss of left router causes adaption of the routing table networks 172.16.0.0, 192.168.3.0 and 192.168.4.0 are not longer seen in the routing table If left router comes back the dynamic routing will learn about these network again, hence we can see the automatic appearance of networks in a routing table in case of power on.

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 static preconfigured values along the given path
      - Hops, interface cost, interface bandwidth, interface delay, etc.
  - Two basic technologies
    - Distance vector, Link state

What can a dynamic routing protocol detect? Basically only loss of links and loss of routers. In case of redundancy an alternate route will be stored in the routing table.

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!"

 IP Technology Primer, v6.4 117

Often router find more than one path to forward a packet to a given destination. The metric helps router find the "best" way. Note that there are several types of metrics used in modern routing protocols. Typically they cannot be compared with each other. For example a simple hop-count is no measure for speed (bandwidth).

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

 IP Technology Primer, v6.4 118

In contrast to static routing where every route must be configured manually, dynamic routing works with one or more routing protocols. These protocols inform the router and create the routing table automatically. Widely used in the Internet. Convergence time is the time until all routers will have the same consistent view of the network after a topology change. Until that temporary routing loops are possible, if entries in routing tables point to each other or lead to circles.

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

 IP Technology Primer, v6.4 119

The table above gives a rough comparison of the most important routing protocols used today. Note that some values can not easily determined and are left blank for this reason.

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 works with the Signpost principle. A Part of the own routing table is sent periodically to all neighbor routers (e.g.: RIP: every 30 seconds).

A signpost carries the Destination network, the Hop Count (metric, "distance") and the Next Hop.

After a router receives a update, he extracts new information's. Known routes with worse metric are ignored.

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

 IP Technology Primer, v6.4 121

Routers view is based on its routing table only. There is an exact view how to reach local neighbors but the network topology behind neighbors is hidden. Therefore such a router has only a limited view of the network topology which causes several problems. Additional mechanism are necessary first to solve problems like count to infinity and routing loops and second to reduce convergence time. That is the time to reach consistent routing tables in all routers after a topology change.

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)

 IP Technology Primer, v6.4 123

Topology changes (link up or down, link state) are recognized by routers responsible for supervising those links and are flooded by responsible routers to the whole network again by using (**L**ink **S**tate **A**dvertisements, LSAs).

Flooding is a controlled multicast procedure to guarantee that every router gets corresponding LSA information as fast as possible but with avoiding a LSA broacast storm in case of redundancy.

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**

 IP Technology Primer, v6.4 124

Applying the SPF algorithm on the link state database, each router can create routing table entries by its own.

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)

IP Technology (v6.4)

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**

 IP Technology Primer, v6.4 127

"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)

 IP Technology Primer, v6.4 128

Distance vector protocols like RIP have several dramatic disadvantages. Examples are slow adaptation in case of network topology changes, size of routing update is proportional to network size and son on.

This led to the development of link-state protocols.

OSPF is the important implementation of link-state technique for IP routing.

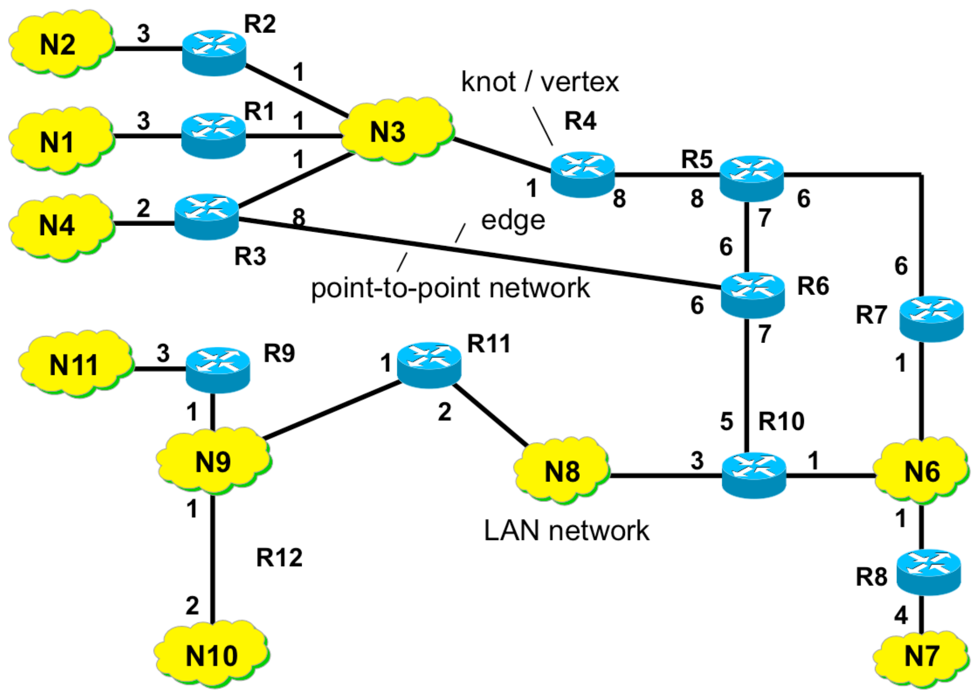
OSPF was developed by IETF to replace RIP. In general link-state routing protocols have some advantages over distance vector, like faster convergence, support for lager networks.

Some other features of OSPF include the usage of areas, which makes possible a hierarchical network topologies, classless behavior, there are no such a problem like in RIP with discontigous subnets. OSPF also supports VLSM and authentication.

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



 IP Technology Primer, v6.4 130

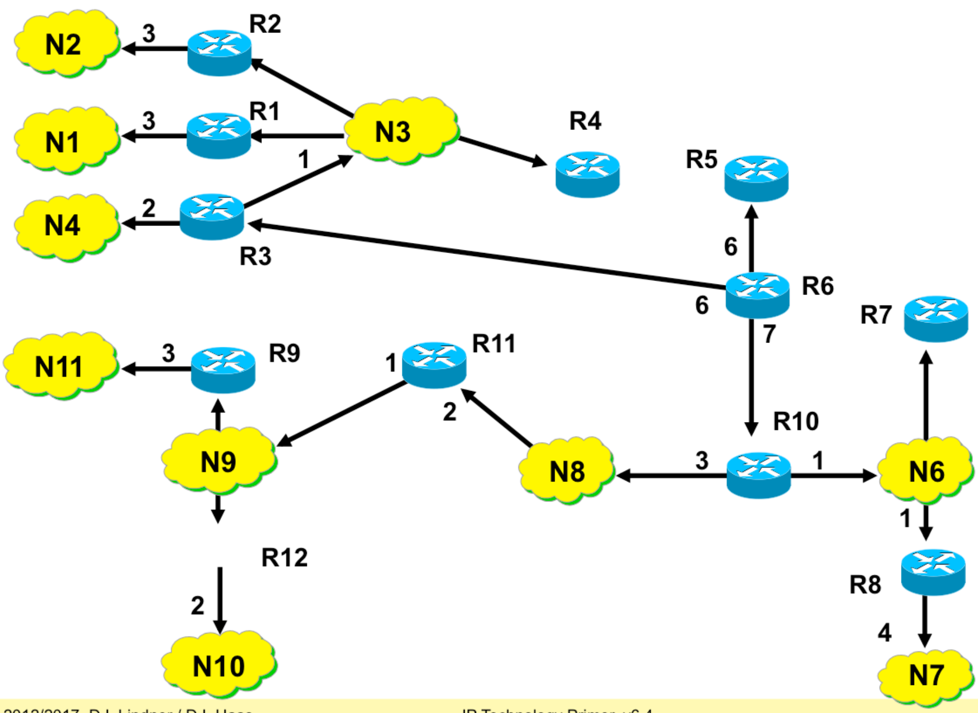
With this topology-database a router can calculate the best path to all destination-networks by applying Dijkstra's SPF (Shortest Path First) algorithms.

The topology-database describes all other possible paths too. So in critical situations (failures) the router can independently calculate an alternative path.

There is no waiting for rumors of other routers anymore which was the reason for several RIP problems.

IP Technology (v6.4)

Shortest Paths regarding Router R6



 IP Technology Primer, v6.4 131

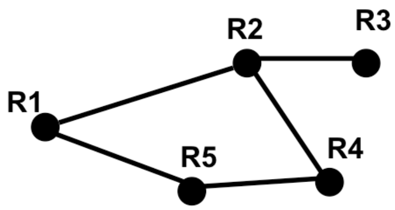
After calculating the shortest path the routing table is constructed by just adding next hop and summary metric taken from the shortest path tree for every network.

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

 IP Technology Primer, v6.4 133

Obviously the dots are routers and the links between the routers are actually networks. The basic idea of OSPF and the topology table is that simple.

OSPF is actually much more complicated. There are 5 types of networks defined in OSPF: point-to-point networks, broadcast networks, non-broadcast multi-access networks, point-to-multipoint networks, and virtual links. Furthermore it is reasonable to divide the topology into multiple "areas" to increase performance ("divide and conquer"). These are the reasons why OSPF is a rather complex protocol. This is explained later.

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**

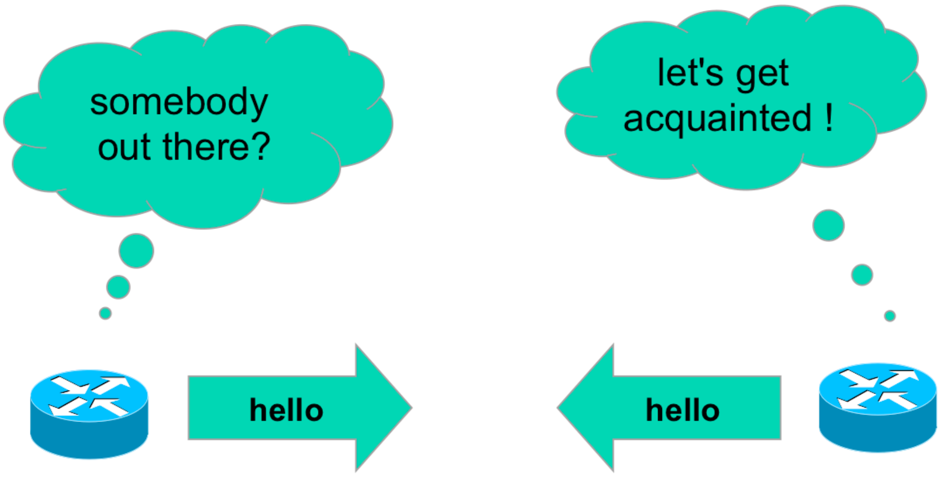
OSPF Communication Principle 2

- **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

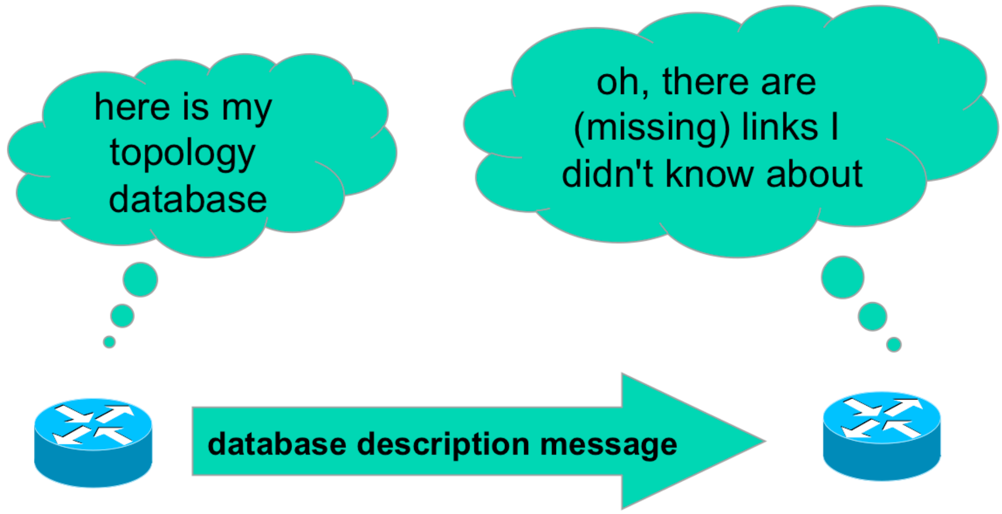
OSPF Communication Principle 3

- **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

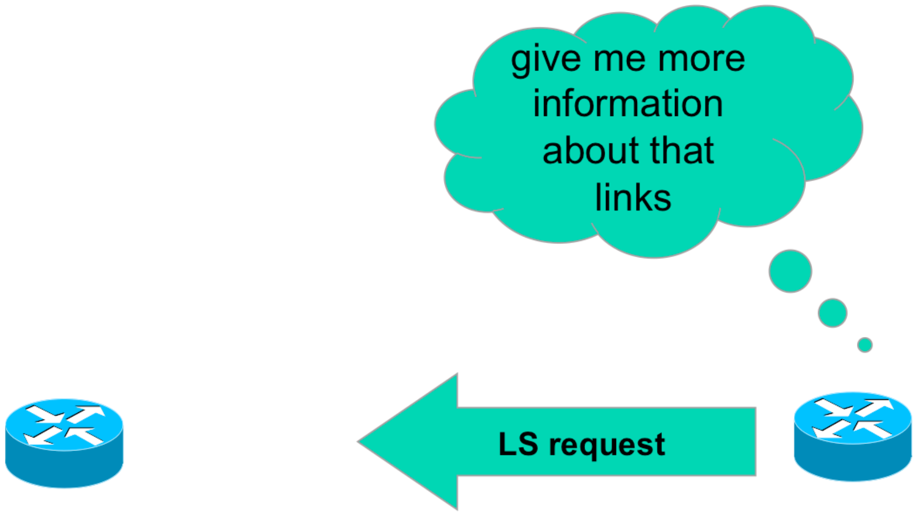
OSPF Communications Summary 1



OSPF Communications Summary 2

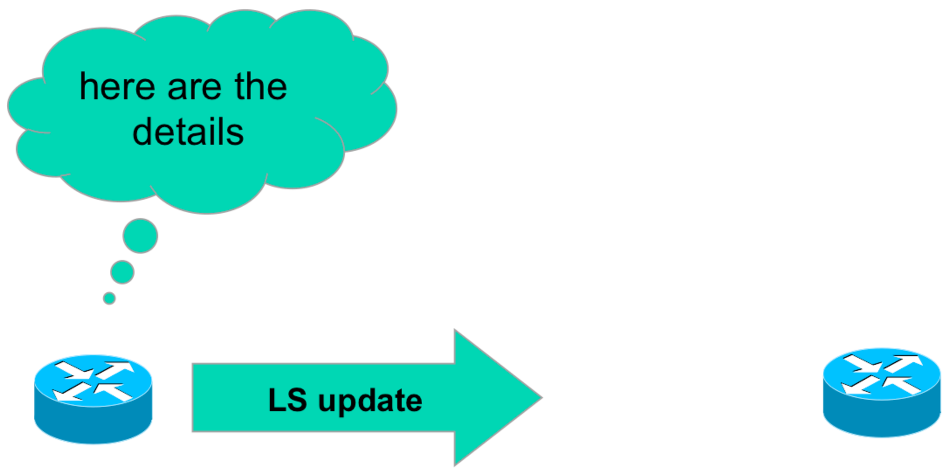


OSPF Communications Summary 3

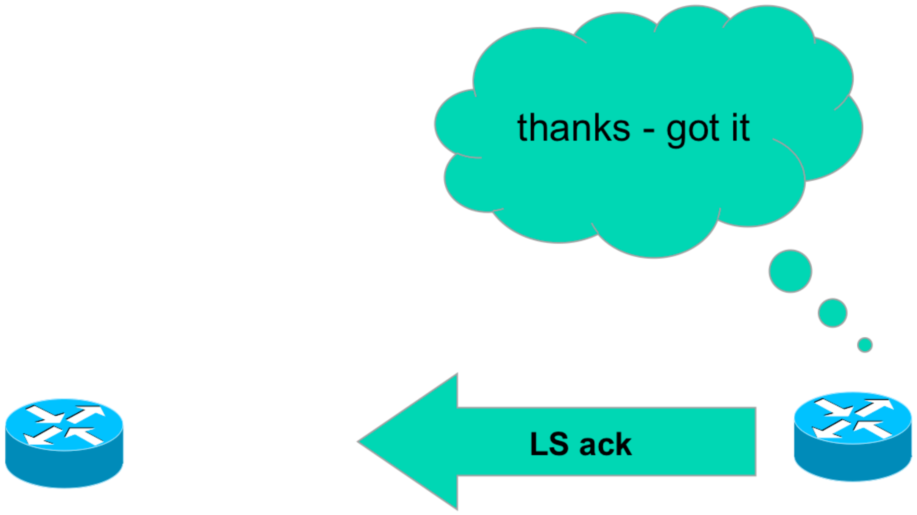


IP Technology (v6.4)

OSPF Communications Summary 4

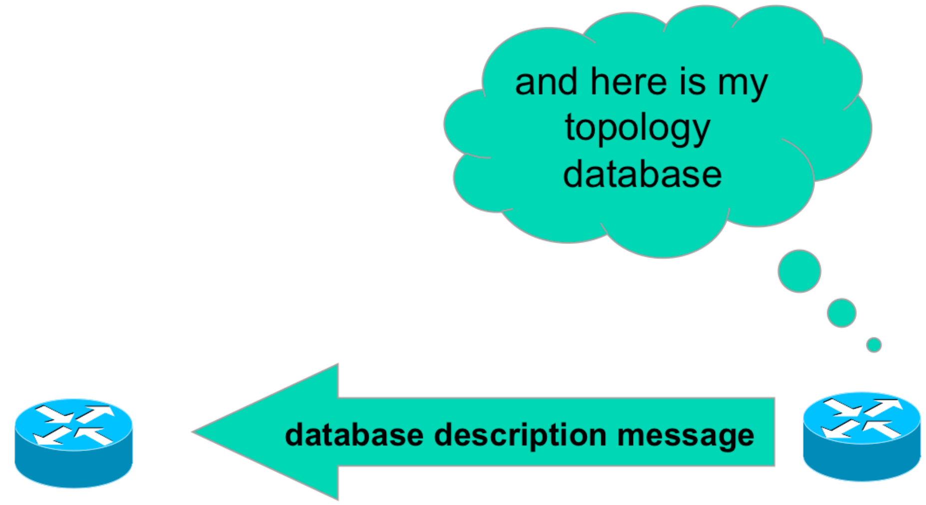


OSPF Communications Summary 5



 IP Technology Primer, v6.4 144

OSPF Communications Summary 6

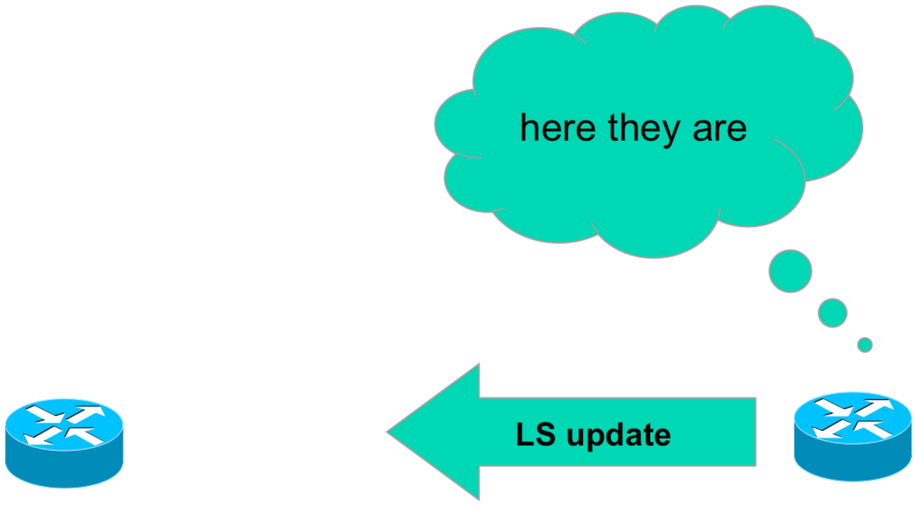


 IP Technology Primer, v6.4 145

IP Technology (v6.4)

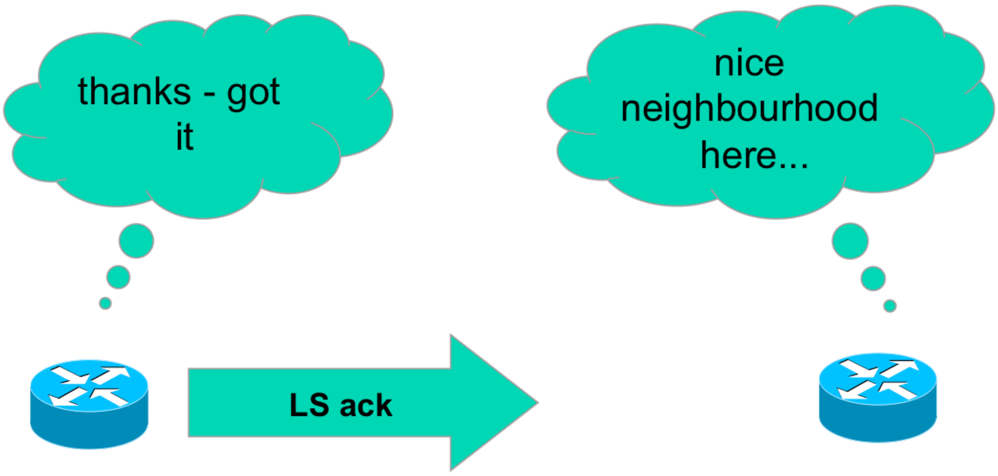
OSPF Communications Summary 7

OSPF Communications Summary 8

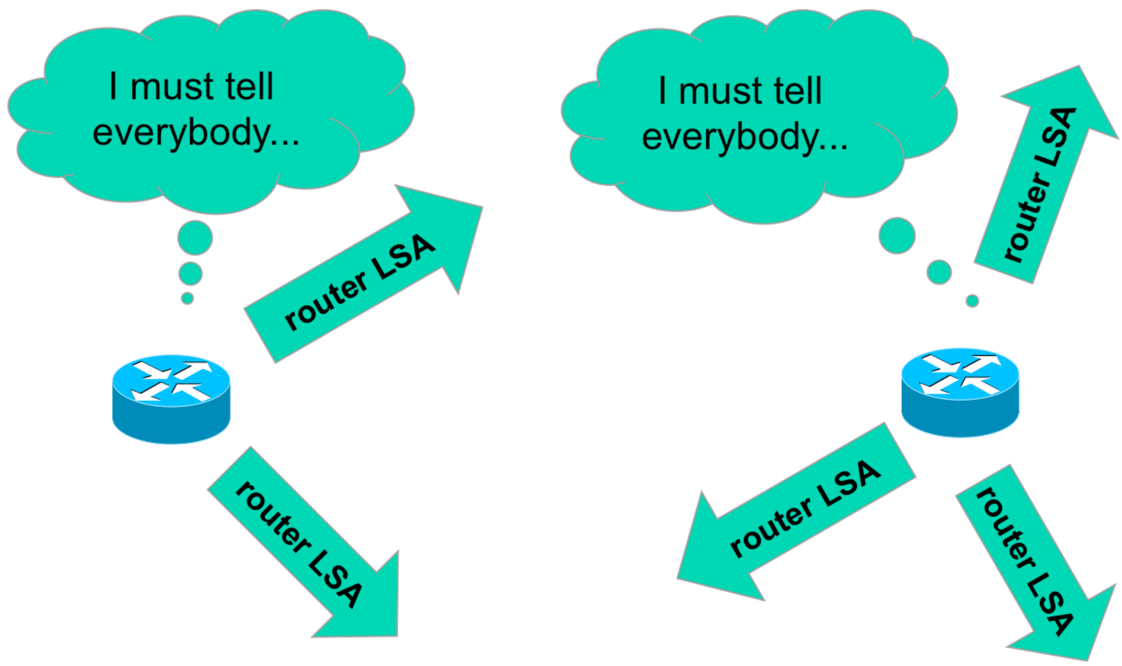


 IP Technology Primer, v6.4 147

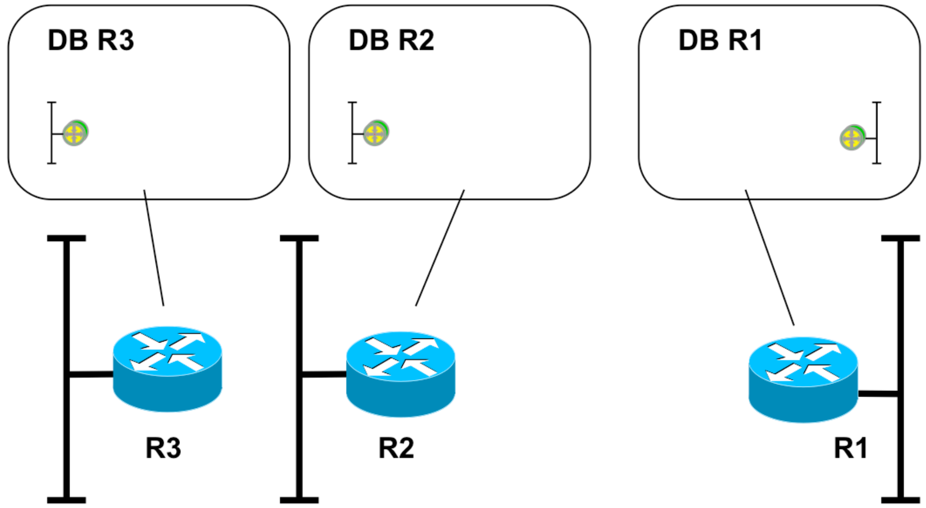
OSPF Communications Summary 9



OSPF Communications Summary 10

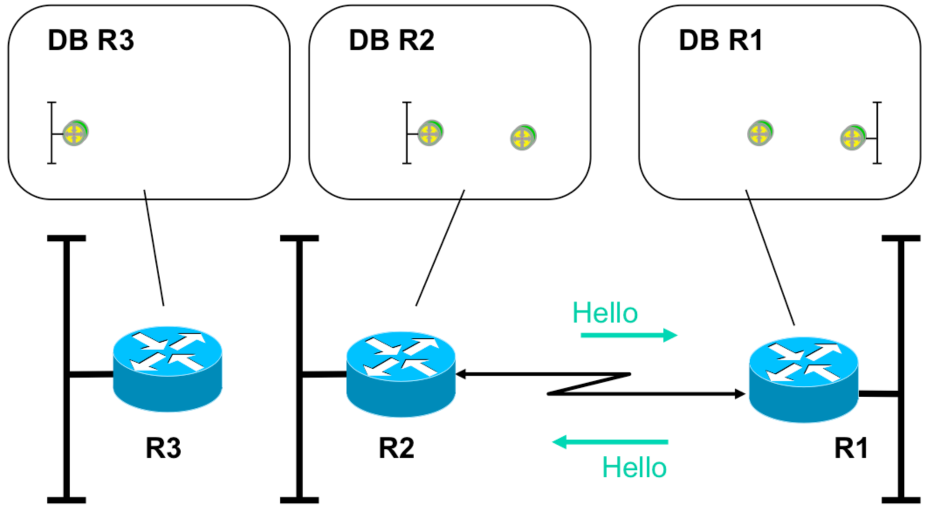


OSPF Start-up



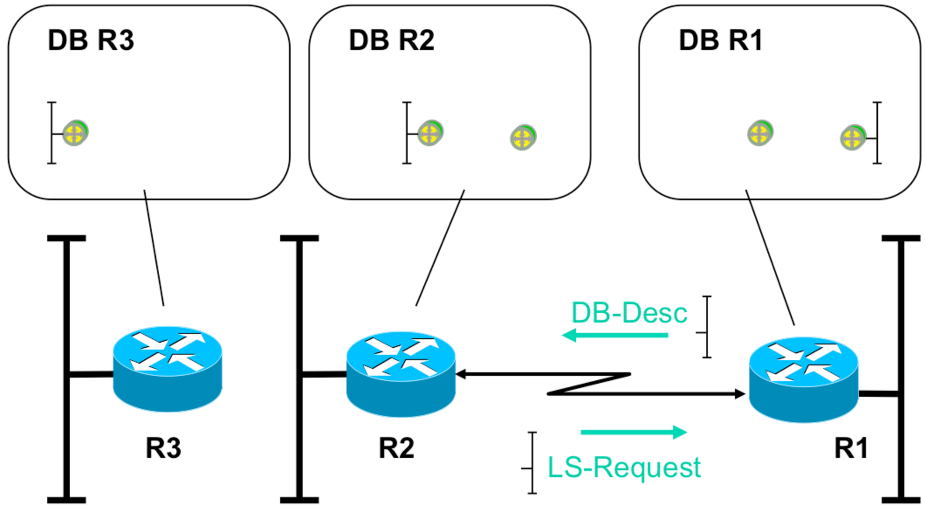
starting position: all routers initialized,
no connection between R1-R2 or R2-R3

OSPF Hello R1 - R2



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

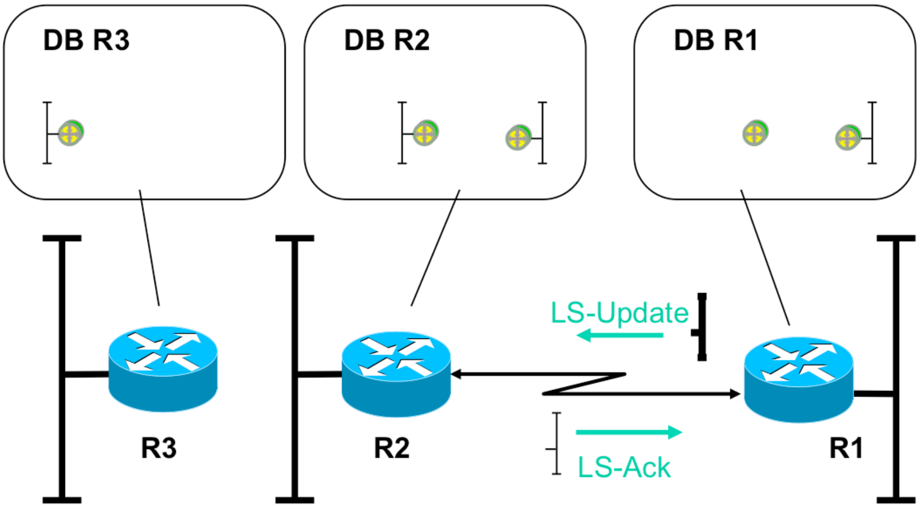
OSPF Data Base Description R1 -> R2



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

IP Technology (v6.4)

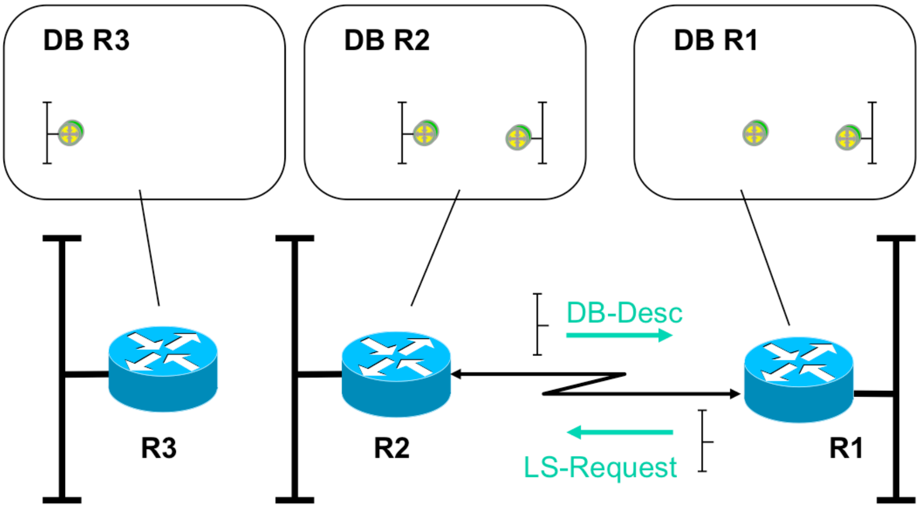
OSPF Data Base Update R1 -> R2



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

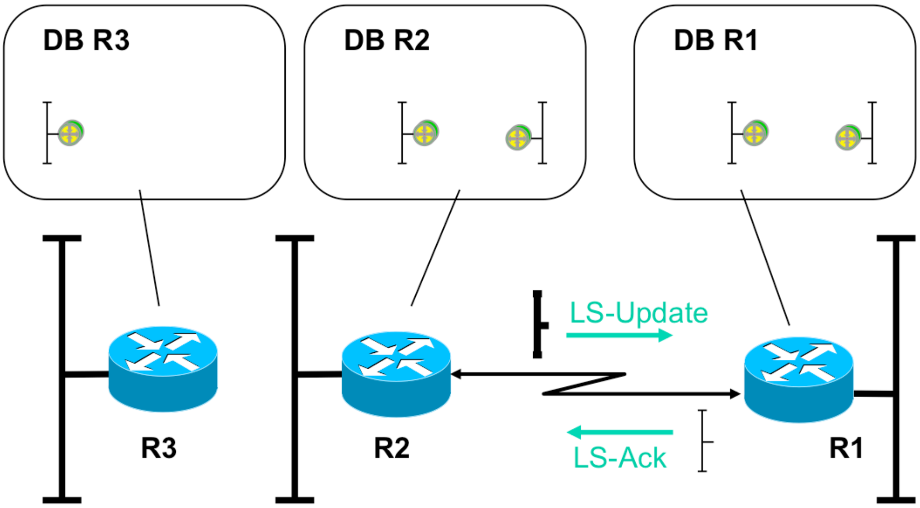
 IP Technology Primer, v6.4 153

OSPF Data Base Description R2 -> R1



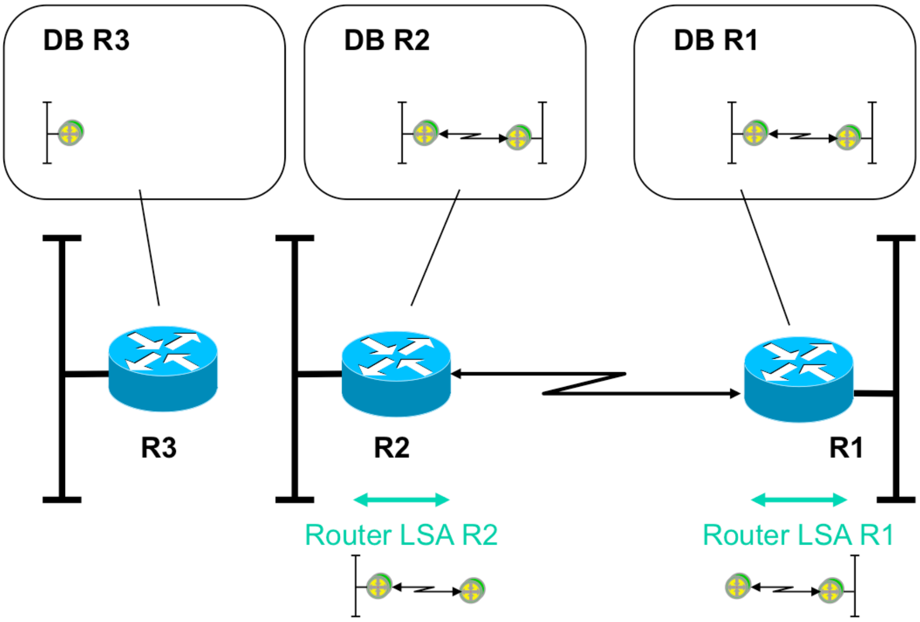
database synchronization: R2 master sends Database-Description, R1 slave sends Link-State Request

OSPF Data Base Update R2 -> R1



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

OSPF Router LSA Emission

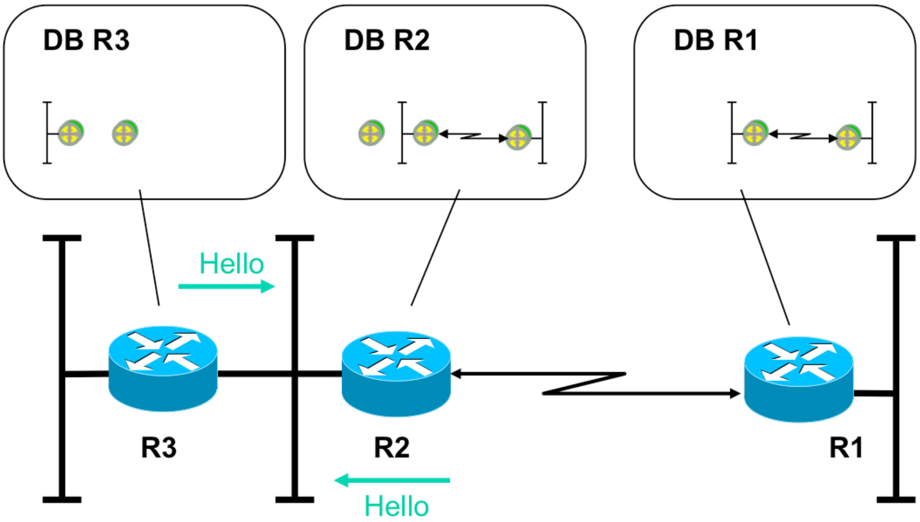


R1 and R2 have synchronized their database completely and notify other nodes about their links

 IP Technology Primer, v6.4 156

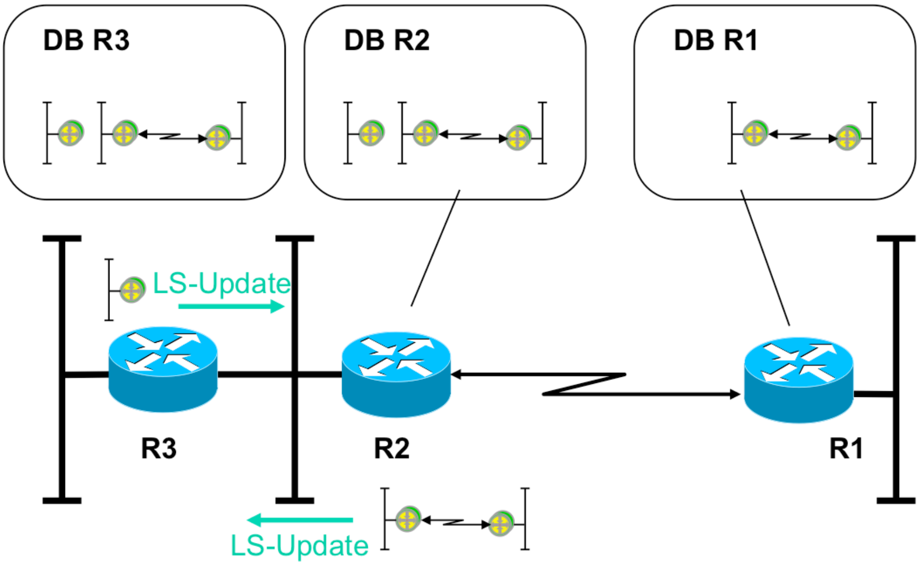
IP Technology (v6.4)

OSPF Hello R2 - R3



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

OSPF Database Update

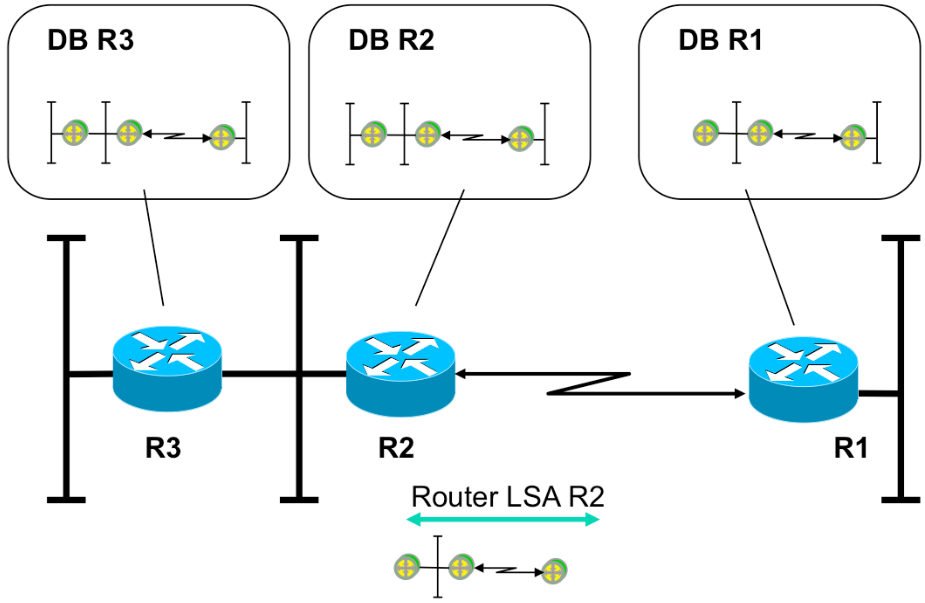


R2 and R3 synchronize their databases (DB-Des., LS-Req.,LS-Upd., LS-Ack.)

 IP Technology Primer, v6.4 158

IP Technology (v6.4)

OSPF Router LSA Emission R2

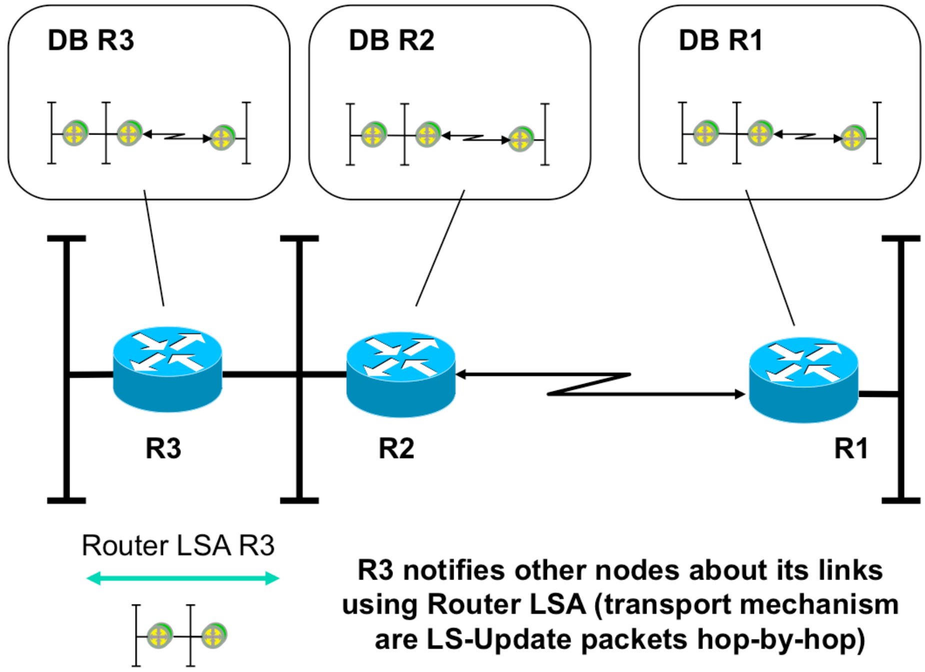


R2 notifies other nodes about its links using Router LSA, (transport mechanism are LS-Update packets hop-by-hop)

 IP Technology Primer, v6.4 159

IP Technology (v6.4)

OSPF Router LSA Emission R3

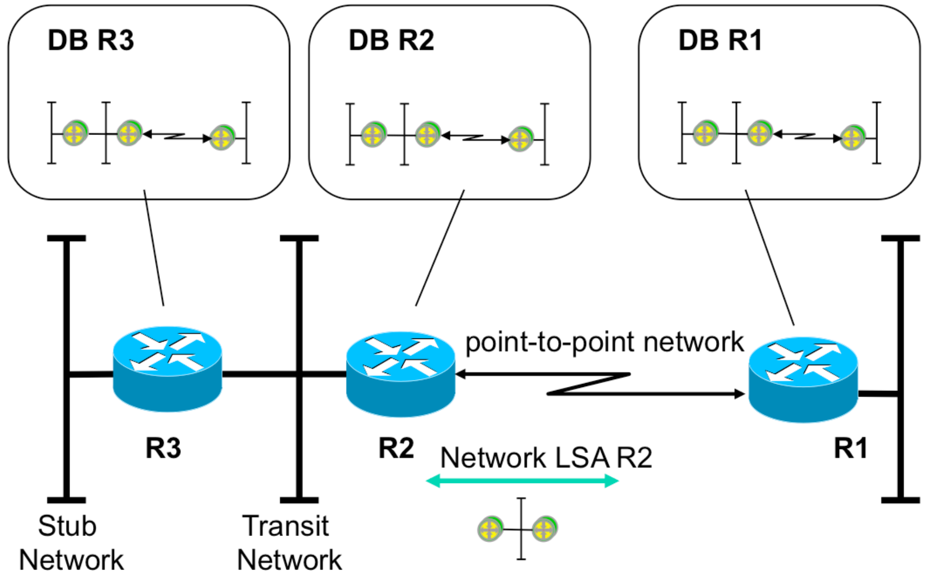


R3 notifies other nodes about its links using Router LSA (transport mechanism are LS-Update packets hop-by-hop)

 IP Technology Primer, v6.4 160

IP Technology (v6.4)

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

 IP Technology Primer, v6.4 161

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**

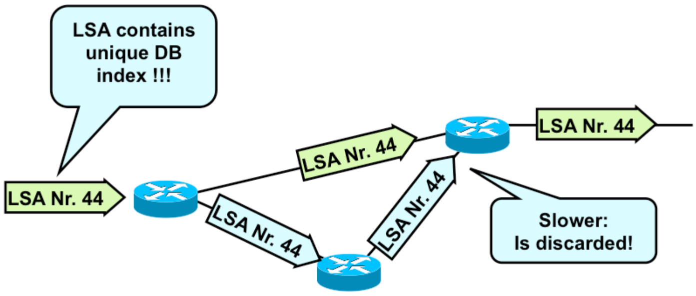
 IP Technology Primer, v6.4 162

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**



 IP Technology Primer, v6.4 164

When reaching the end of the 32 bit sequence number the associated router will wait for an hour so that this LSA ages out in each link state database. Then the router resets the sequence number (lowest negative number i. e. MSB=1, 80000001) and continues to flood this LSA.

Each LSA carries also a 16 bit age value, which is set to zero when originated and increased by every router during flooding. LSAs are also aged as they are held in each router's database. If sequence numbers are the same, the router compares the ages the younger the better but only if the age difference between the recently received LSA is greater than MaxAgeDiff; otherwise both LSAs are considered to be identical.

Note:

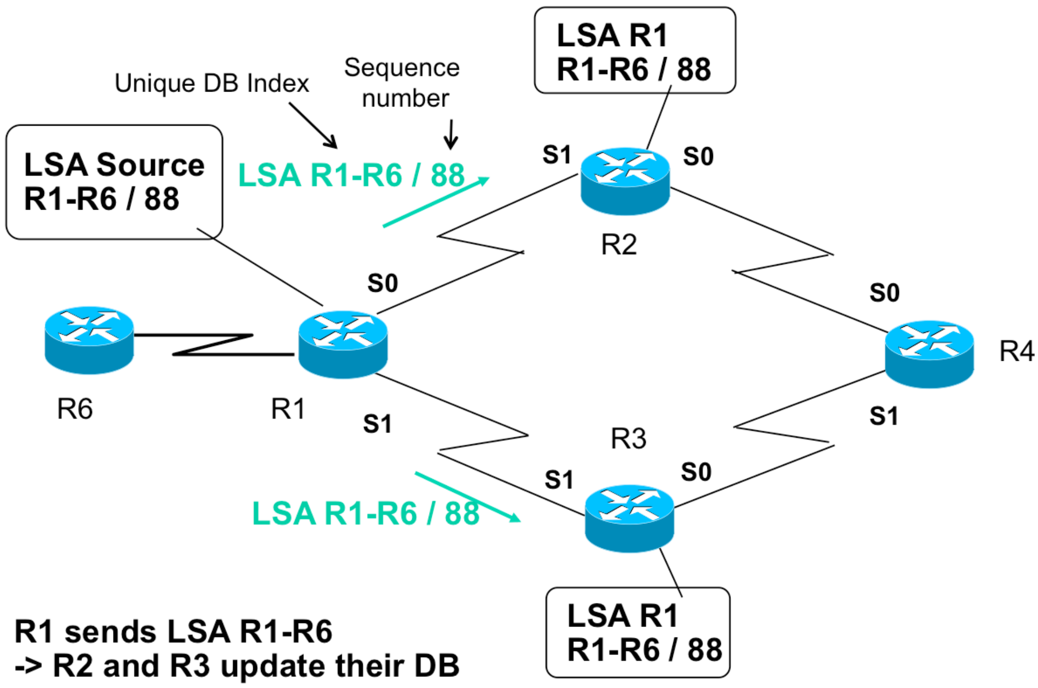
Radia Perlman proposed a "Lollipop" sequence number space but today a linear space is used as described above.

Since signed integers are used to describe sequence numbers, 8000001 represents the most-negative number in a hexadecimal format. To verify this, the 2-complement of this number must be calculated. This can be done in two steps. First calculate the 1-complement by simply inverting the binary number, that is the most significant byte "0x80" which is "1000 000" is transformed to "0111 111", the least significant byte "0x01" which is "0000 0001" is transformed to "1111 1110" and all other bytes in between are now "1111 1111". Secondly, in order to receive the 2-complement, "1" must be added. Then the final result is "0111 1111 1111 1111 1111 1111 1111 1111", which is the absolute number (without sign).

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

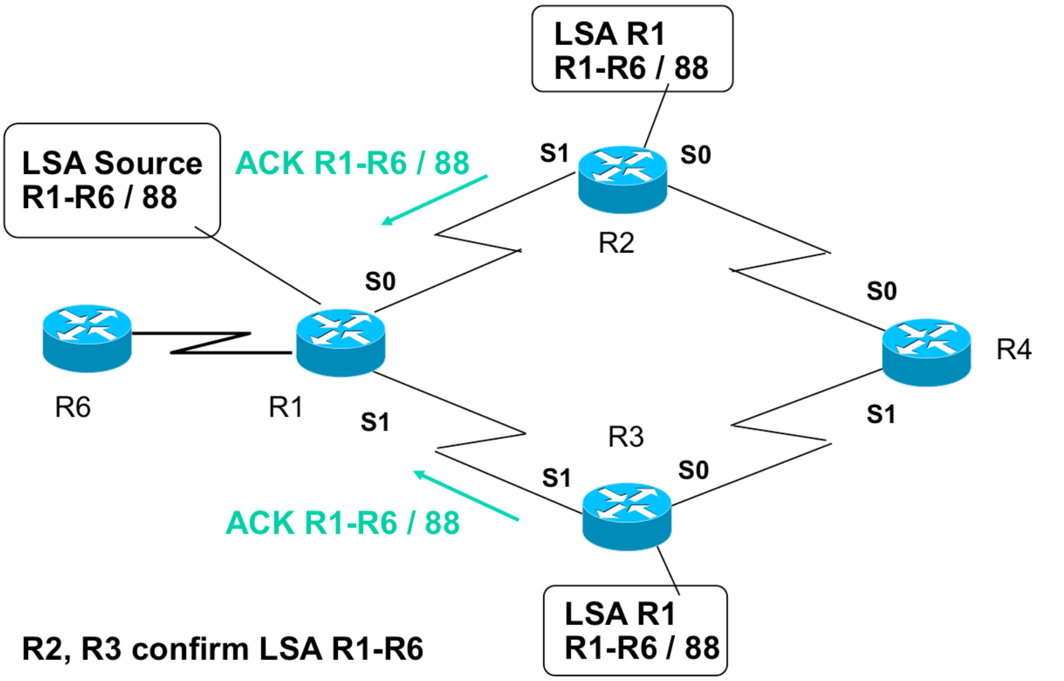
LSA Broadcast Example (1)



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

IP Technology (v6.4)

LSA Broadcast Example (2)

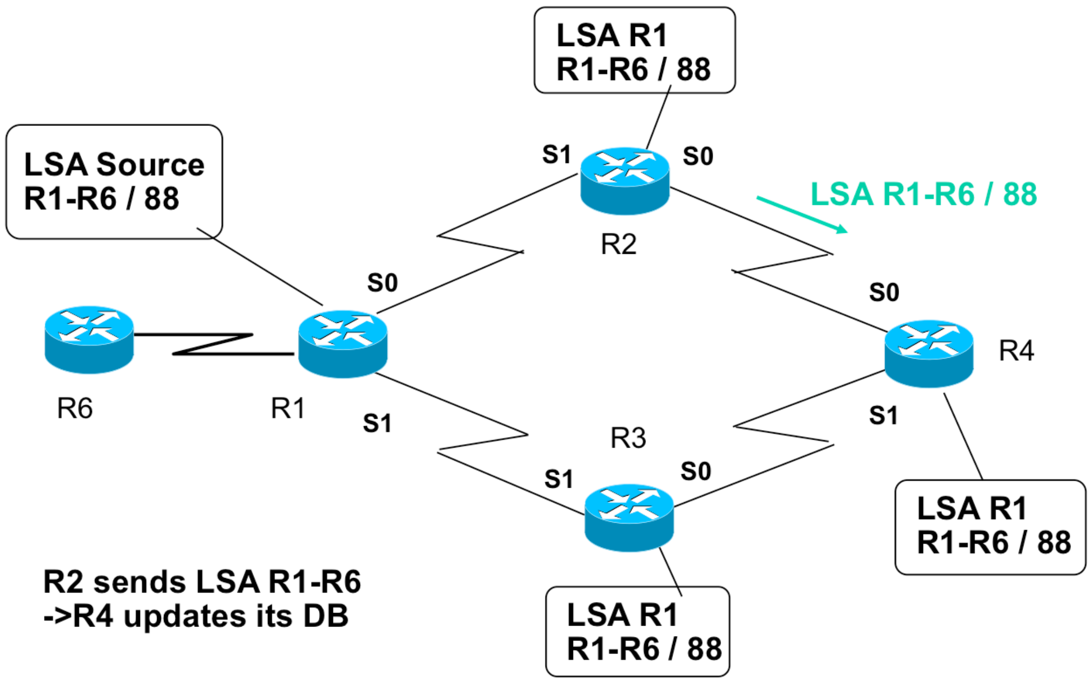


R2, R3 confirm LSA R1-R6

 IP Technology Primer, v6.4 167

IP Technology (v6.4)

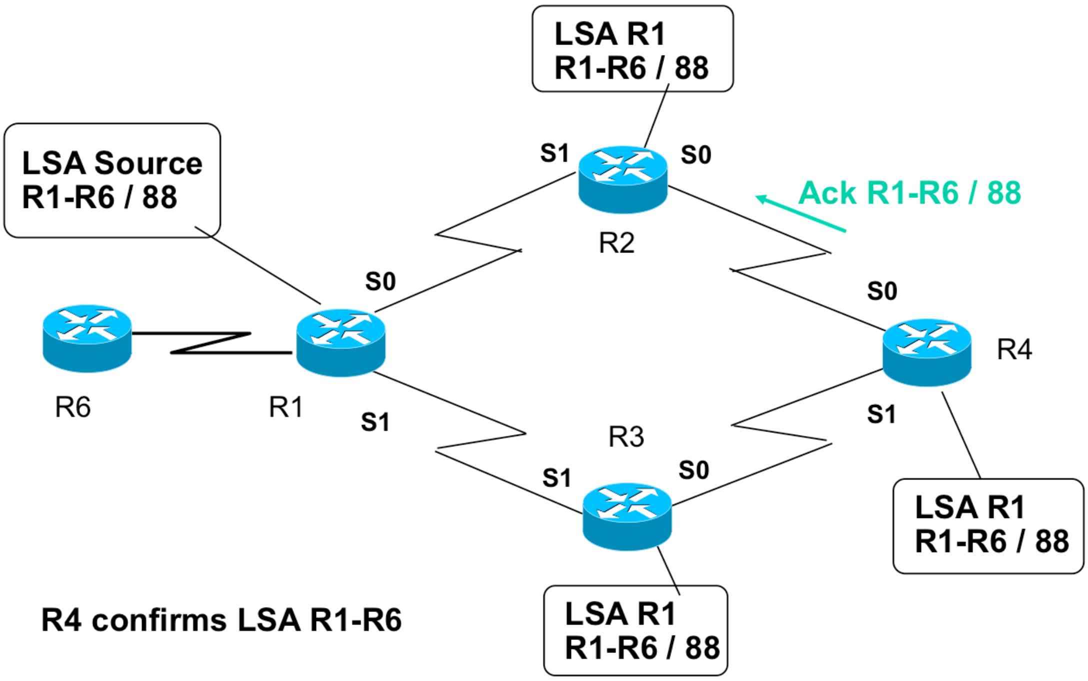
LSA Broadcast Example (3)



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

 IP Technology Primer, v6.4 168

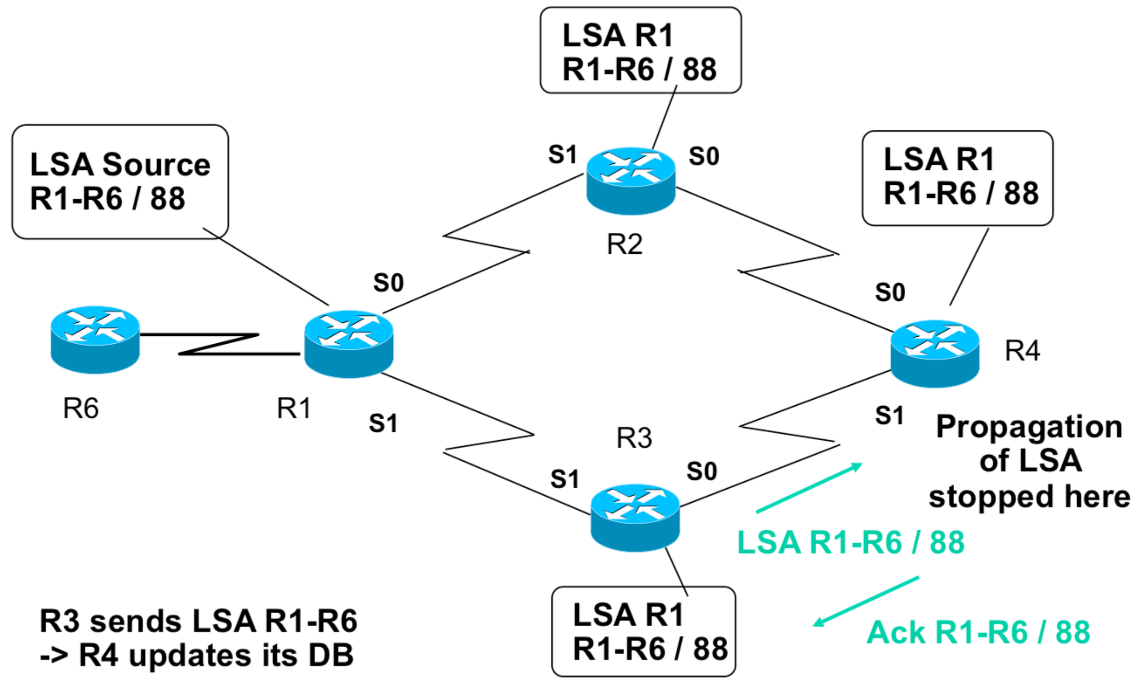
LSA Broadcast Example (4)



R4 confirms LSA R1-R6

IP Technology (v6.4)

LSA Broadcast Example (5)



 IP Technology Primer, v6.4 170

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

 IP Technology Primer, v6.4 172

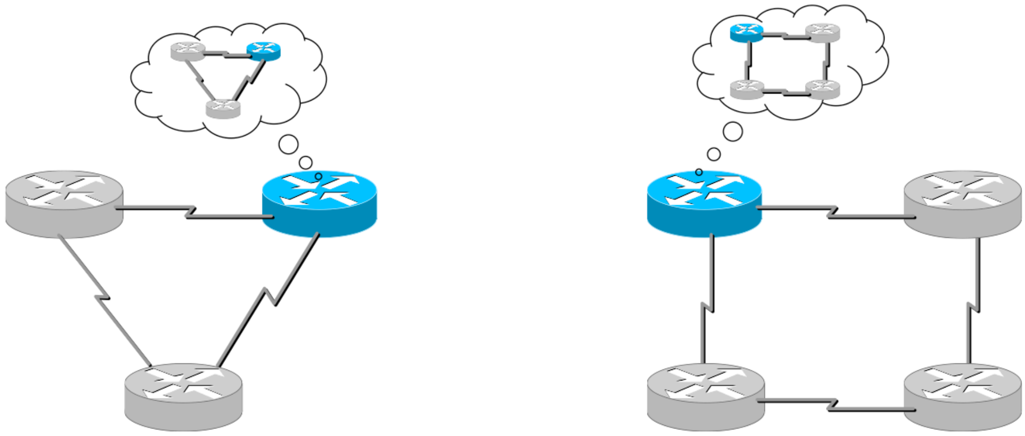
Note that loopback interfaces are more stable than any physical interface. Furthermore it's easier for an administrator to manage the network using loopback addresses for Router-IDs.

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...**

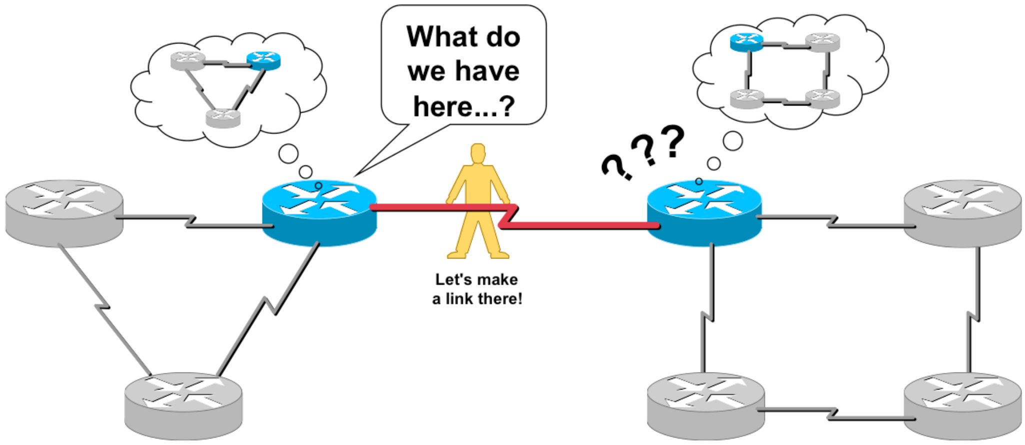


 IP Technology Primer, v6.4 174

The routers on the slide have 2 stable networks, there are no periodic link state updates, just hello messages.

Basic Principle (2)

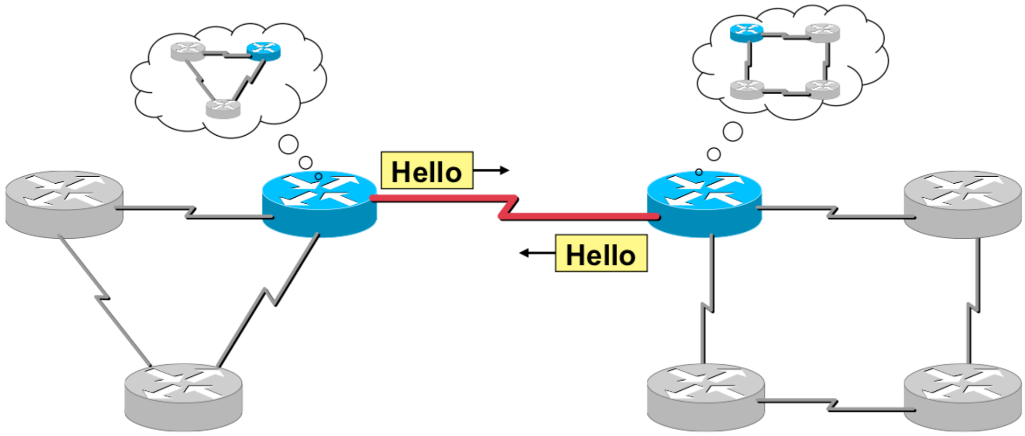
- Suddenly, some brave administrator connects them via a serial cable...
- Both interfaces are still in the "Down state"



After the link is connected, the routers detect a new network (OSPF is configured on the interface and interfaces are enabled).

Basic Principle (3)

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

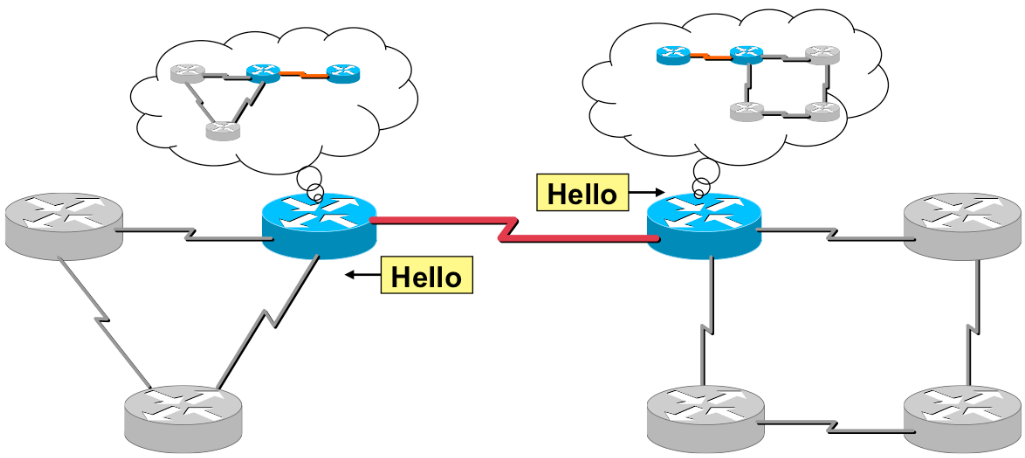


 IP Technology Primer, v6.4 176

OSPF routers send Hello packets out all OSPF enabled interfaces on a multicast address 224.0.0.5. Then the router waits for a reply (another hello from the other side) which must arrive within 4 x hello interval, otherwise the router falls back to the down state again. That is, the init state lasts only up to 4 times the hello interval.

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...

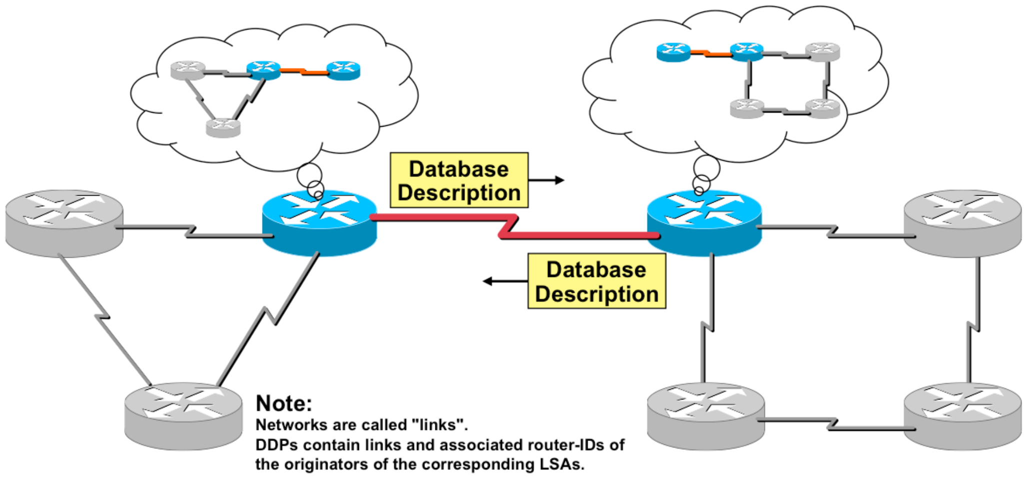


 IP Technology Primer, v6.4 177

If two routers sharing a common link and they agree on a certain parameters in their respective Hello packets, they will become neighbors.

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.



After neighborship is established, the routers enter the "exstart state" and determine who of them is master and who is slave. This will be needed later as the master will begin to send LS-Request packets. The rule is simple: the router with the highest IP address (of the two involved interfaces on that link) is master.

Then, both routers enter the exchange state and exchange database description packets (DDPs), which contain partial LSAs and therefore can be regarded as a summary of their topology database.

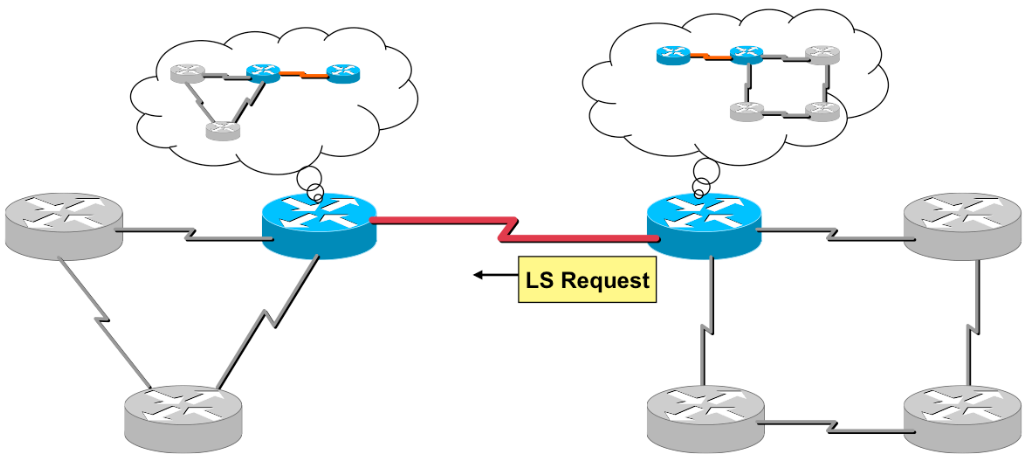
Note: typically a series of DDPs are sent from each side. Each advertised link is identified by a OSPF router ID, which represents the originator of that information.

Both routers send out a series of database description packets containing the networks held in the topology database. These networks are referred to as links. Most of the information about the links has been received from other routers (via LSAs). The router ID refers to the source of the link information.

Each link will have an interface ID for the outgoing interface, a link ID, and a metric to state the value of the path. The database description packet will not contain all the necessary information, but just a summary (enough for the receiving router to determine whether more information is required or whether it already contains that entry in its database).

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...

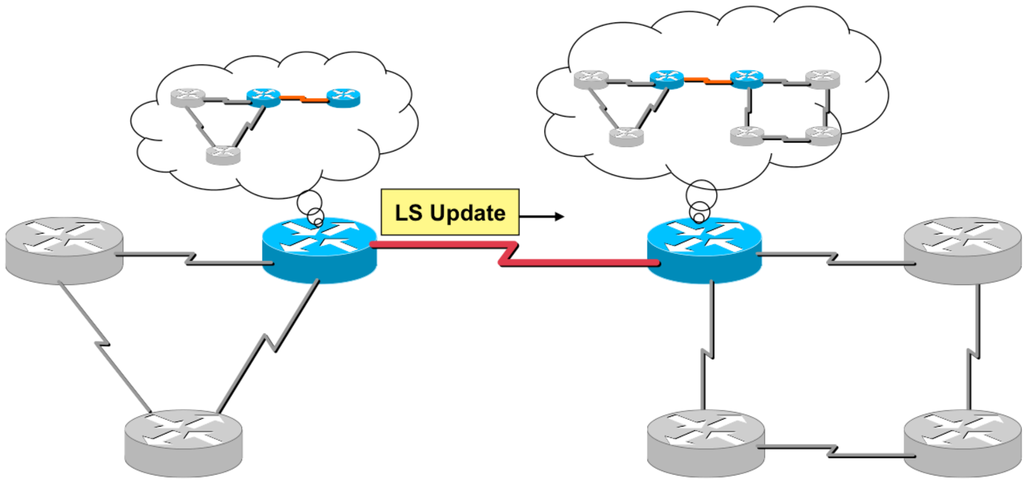


 IP Technology Primer, v6.4 179

The reciever checks its database, sees it is a new information and requests a detailed information with Link State Request packet LSR.

Basic Principle (7)

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

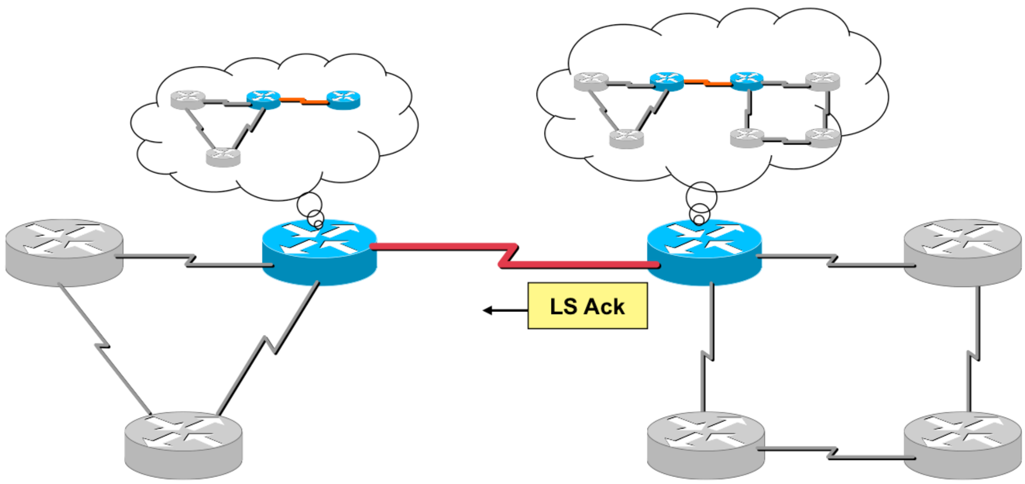


 IP Technology Primer, v6.4 180

As a reply the left router sends a Link State Update packet LSU which contains detailed information about requested links.

Basic Principle (8)

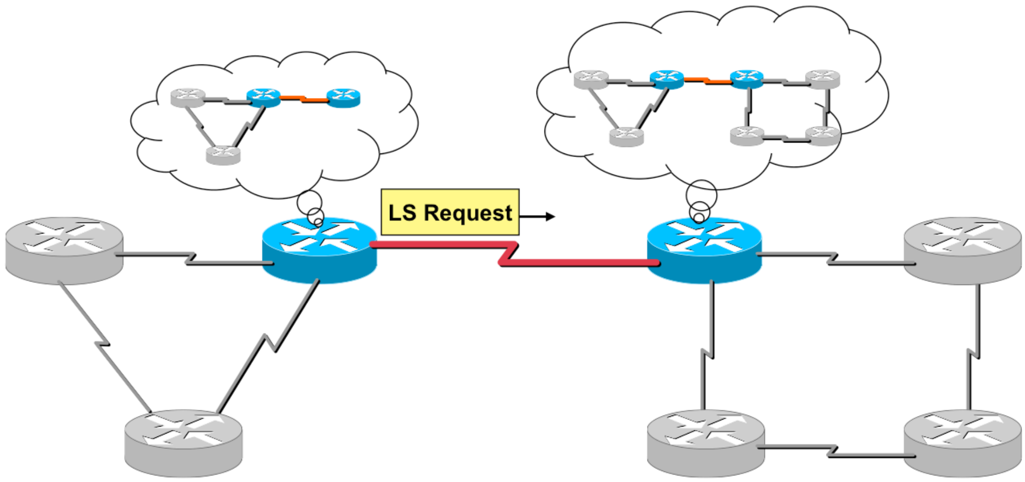
- **The right router is very thankful, and returns a "Link State Acknowledgement"...**



Link State Acknowledgement LSAck is used to make sure that the information is received.

Basic Principle (9)

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

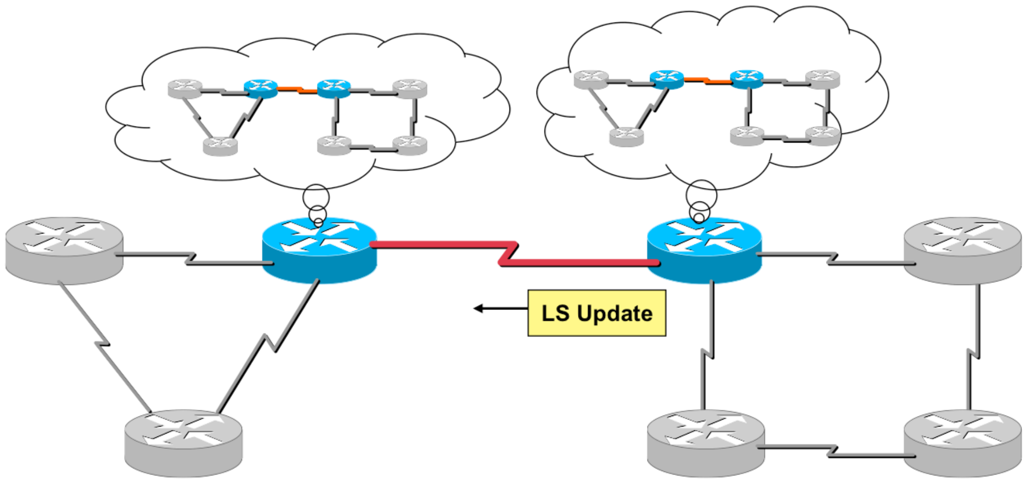


 IP Technology Primer, v6.4 182

LSR is sent in the other direction asking for detailed information.

Basic Principle (10)

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

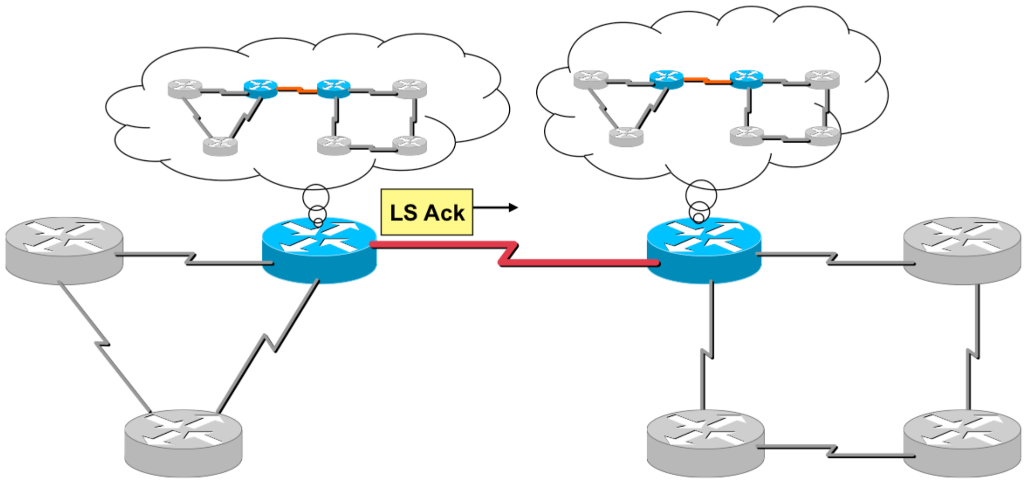


 IP Technology Primer, v6.4 183

Then a LSU is sent back.

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"

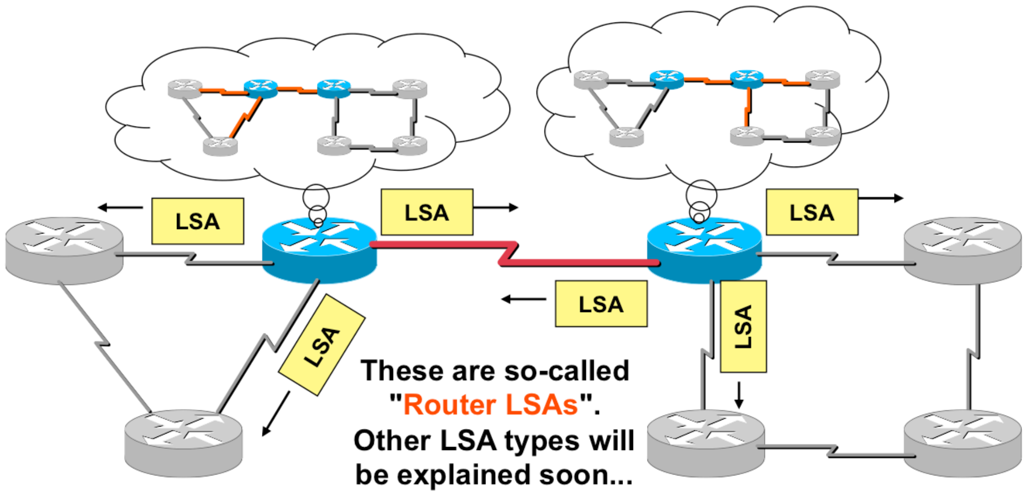


 IP Technology Primer, v6.4 184

LSAck – saying thanks for info.

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

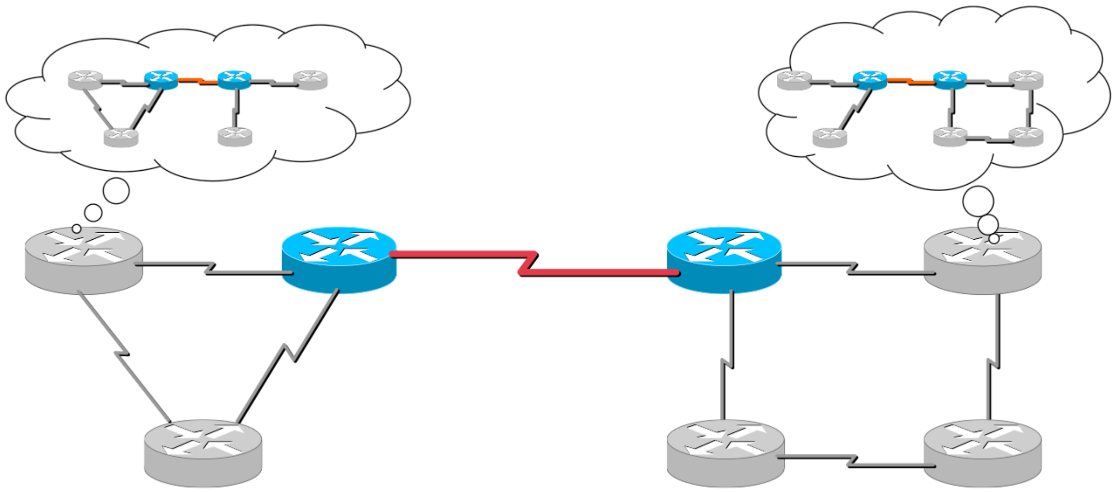


 IP Technology Primer, v6.4 185

Now the both routers have a new information in their databases. This information is flooded to all other adjacent routers as a router LSA or LSA type 1 in which the router sends information about its own links.

Database Inconsistency

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



It might happen if you connect two existing networks together. As you can see some routers may miss a new information.

Healing Inconsistency

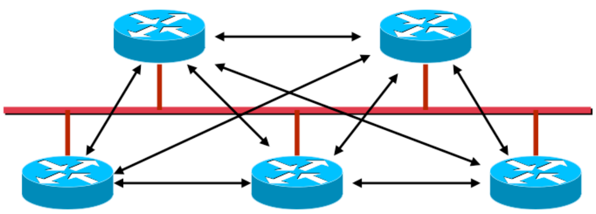
- **Every router sends its LSAs every 30 minutes (!)**
  - Heals but long time of routing table / topology table inconsistence when combining a former split area of a OSPF domain
- **Triggering database synchronization between any two routers in the network**
  - In order to avoid long time of inconsistence
  - 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**
- **IP Routing**
  - Introduction
  - OSPF Basics
  - OSPF Communication Procedures (Router LSA)
  - LSA Broadcast Handling (Flooding)
  - OSPF Splitted Area
  - Broadcast Networks (Network LSA)
- **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**



 IP Technology Primer, v6.4 189

Basic concept of link state requires point-to-point relationships. That fits best for point-to-point networks like serial lines but that causes a problem with shared media multi-access networks (e.g. LANs or with networks running in a so called NBMA-mode (Non Broadcast Multi Access) like X.25, Frame Relay, ATM. Hello, database description and LSA updates between each of these routers can cause huge network traffic and CPU load.

Consider the flooding process after establishment of each adjacency!!! The formation of an adjacency between every attached router would create a lot of unnecessary LSAs. A router would flood an LSA to all its adjacent neighbors, creating many copies of the same LSA on the same network.

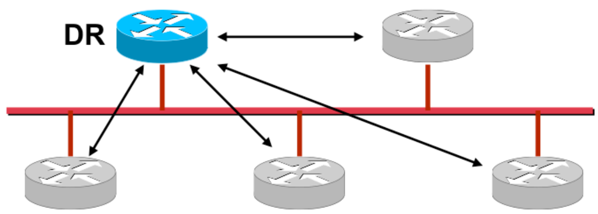
Information about all possible neighbourhood-relations seems to be redundant. The well known concept of virtual (network) node (or virtual router) is introduced to solve the problem.

Only the virtual node needs to maintain N-1 point-to-point relationship to the other nodes and hence any-to-any is not necessary.

In OSPF the virtual node is called Designated Router (DR).

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



To prevent the problems described in the previous slide, a Designated Router (DR) is elected on a multi-access network. DR is responsible for representation of the multi-access network and all the routers on it to the rest of network and management of flooding process on a multi-access network. The network itself becomes a "pseudonode" on the graph. The pseudonode is represented by the DR.

All other routers peer with the DR, which informs them of any changes on the segment.

Note: For LAN segments, the Router LSA does NOT contain the subnet mask. The subnet mask for this LAN segment is also carried inside the Network LSA.

In case of a failure the Designated Router would be single point of failure.

Therefore a Backup Designated Router (BR) is elected, too.

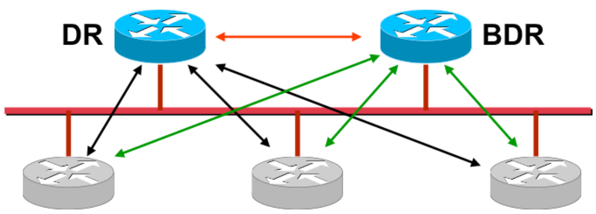
DR and BR are elected by exchanging hello-messages at start-up.

Attention !!!

The concept of DR/BDR influences only how routing information is exchanged among those routers. There is no influence on actual IP forwarding which is based on routing tables.

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



 IP Technology Primer, v6.4 191

Each multi-access interface has a "Router Priority" ranging from 0 to 255 (default 1). Routers with a priority of 0 cannot become DR or BDR. The election process is performed with Hello packets which carry the priority. If some routers have the same priority, the one with the highest numerical Router ID wins. If a DR fails the BDR becomes active immediately (Hello stays out) and a new election for the BDR is started.

Note: After election of DR and BDR, adding a new router with higher priority will not replace them. The first two routers immediately become DR and BDR. The only way to control the election is to set the priority for all other routers ("DROTHER") to zero, so they cannot become DR or BDR.

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

It is recommended in OSPF to use the loopback interfaces for router ID. You should configure a loopback interface first and then start the OSPF process, otherwise the highest ip address from a physical interface will be taken.

Designated and Backup Designated Router are determined using the router-priority field of the Hello message. On a DR failure, the Backup Router (BDR) continues the service.

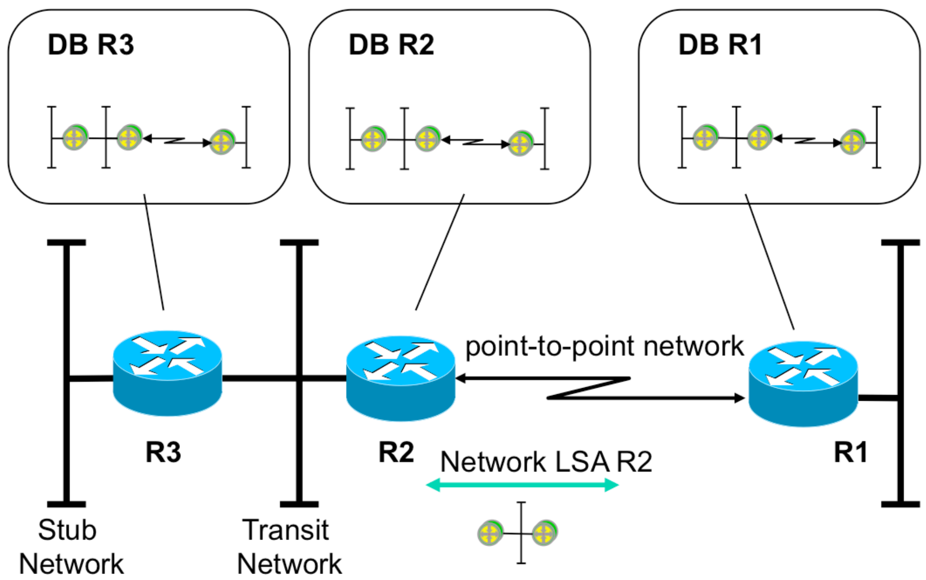
BDR listens to the traffic on the virtual point-to-point links between all routers and the DR. Multicast addresses are used for ease that newtork sniffing. BDR recognizes a DR failure through missing acknowledge messages. Remember: Every LS-Update message requires a LS-Acknowledgement message.

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

IP Technology (v6.4)

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

 IP Technology Primer, v6.4 194

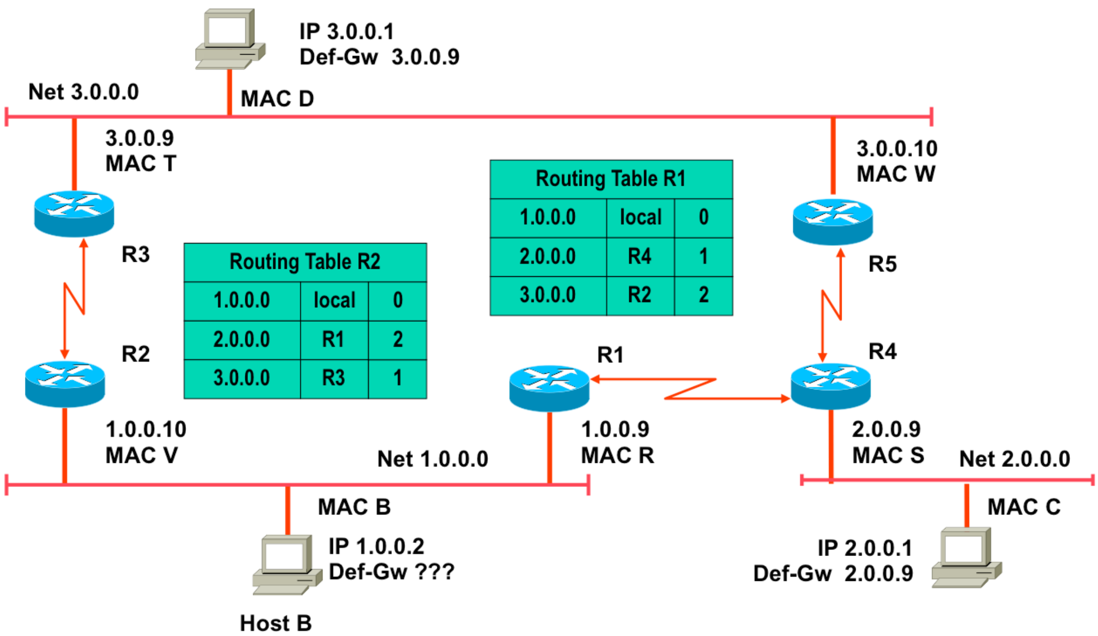
Details: OSPF Multicast Usage

- **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?



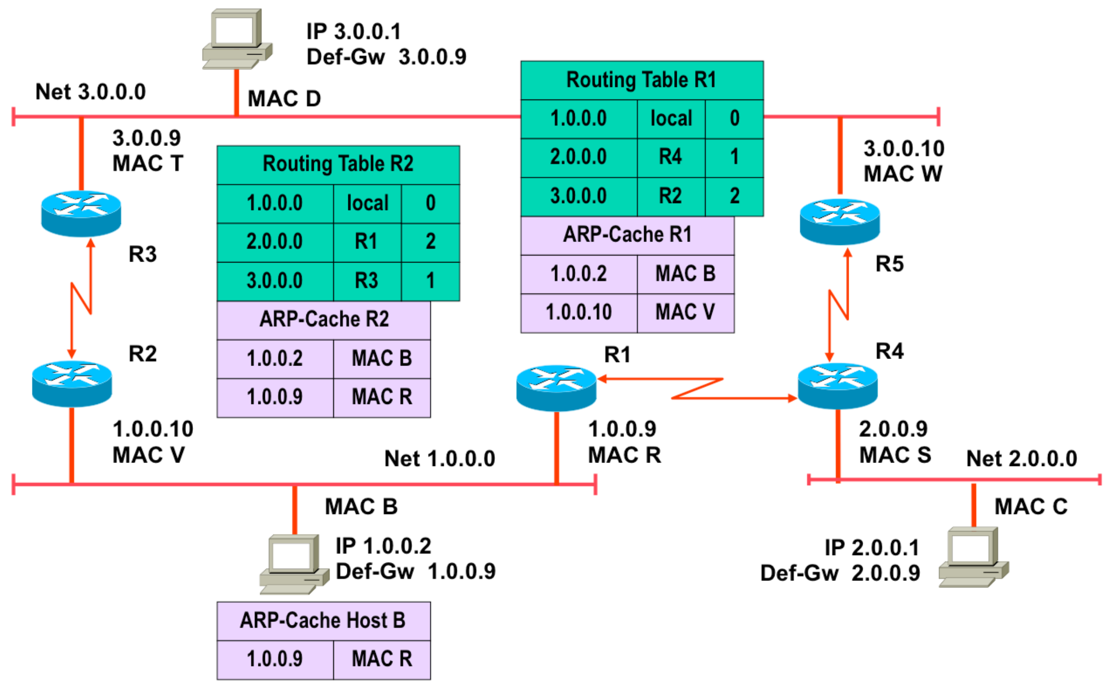
Routing Table R2		
1.0.0.0	local	0
2.0.0.0	R1	2
3.0.0.0	R3	1

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

 IP Technology Primer, v6.4 197

The drawing shall outline the basic problem in case of redundancy of local routers. If only the IP address one default gateway is configurable in the end system B, which one should be configured? As long as both default gateways R1 and R2 are available there is no problem when host B takes the wrong (more far away) default gateway in order to reach a destination network. Remember that in such a case a router will forward the IP datagram to the other router and will sent a ICMP redirect message to host B. But what if the router which is configured as default-gateway is not any longer powered-on? Then host B can not reach foreign networks in case of indirect delivery.

Delivery 1.0.0.2 -> 3.0.0.1



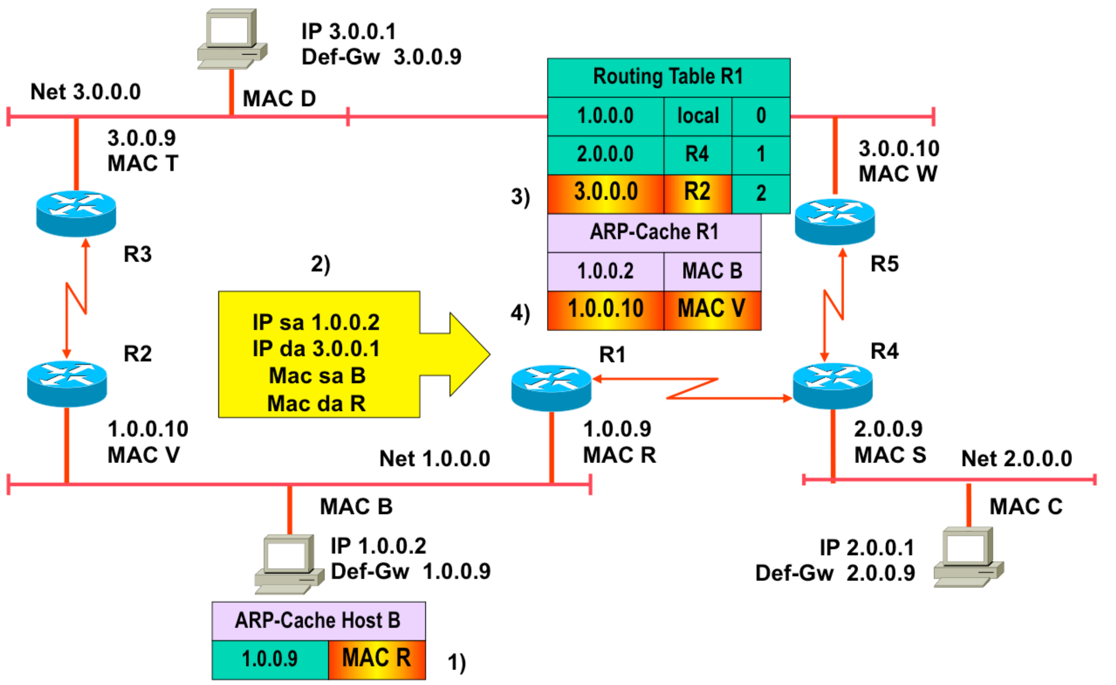
 IP Technology Primer, v6.4 198

Assume IP host 1.0.0.2 selects router R1 1.0.0.9 as one and only default-router.

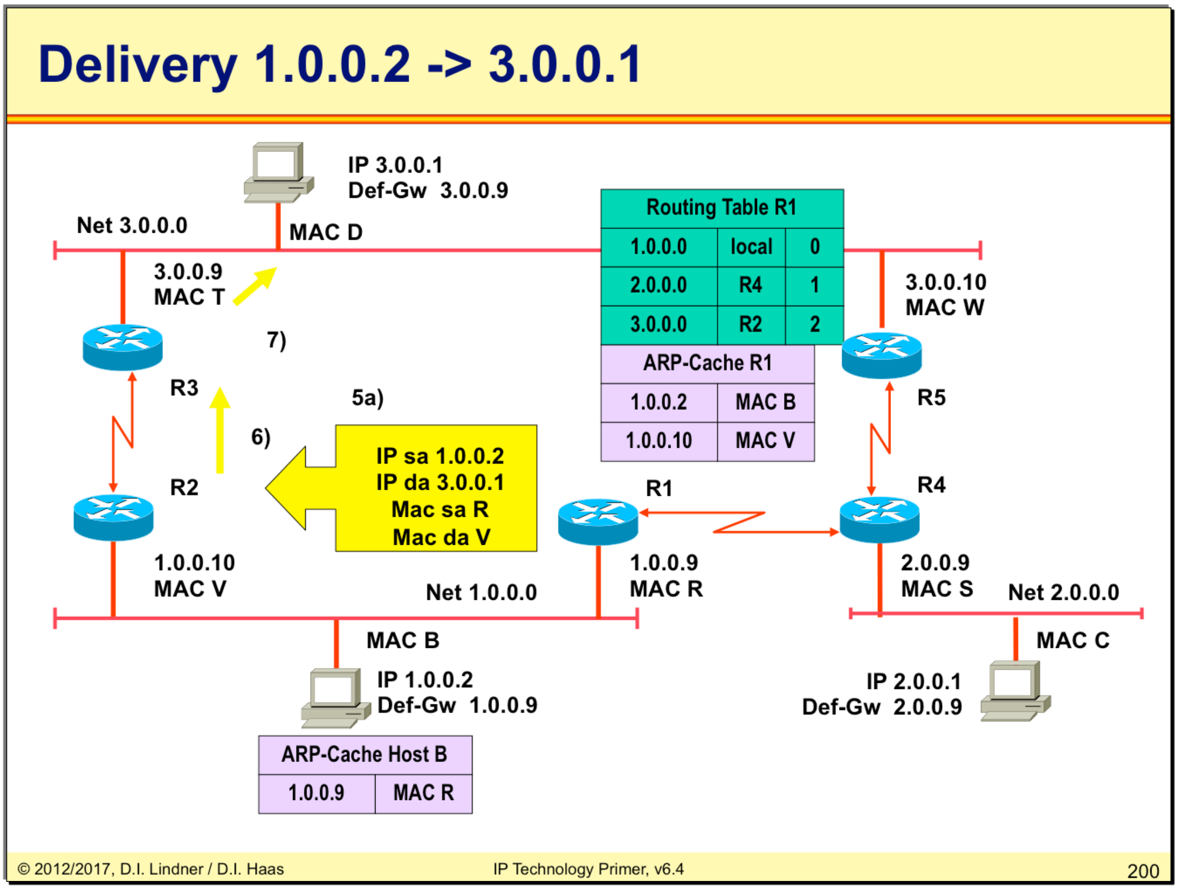
Picture shows that ARP caches are already filled with appropriate mappings of L2 and L3 addresses.

IP Technology (v6.4)

Delivery 1.0.0.2 -> 3.0.0.1

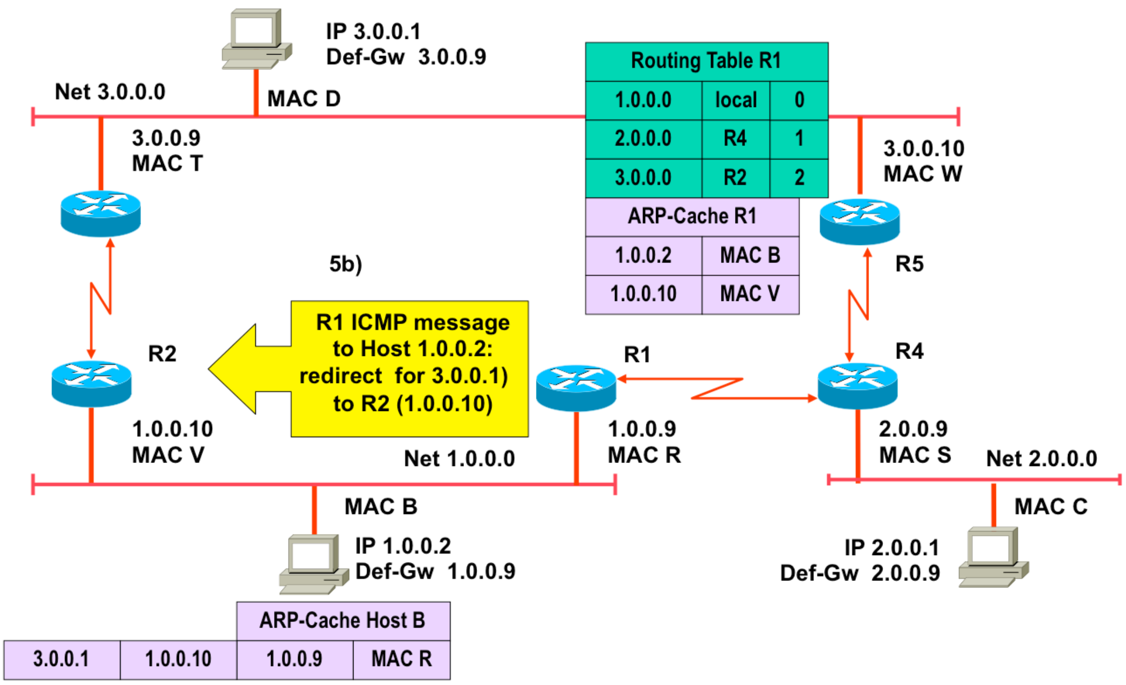


IP Technology (v6.4)



At router R1 the IP datagram is forwarded on the same interface as it was received -> redirect would be nice to avoid sending this datagram twice on net 1.0.0.0.

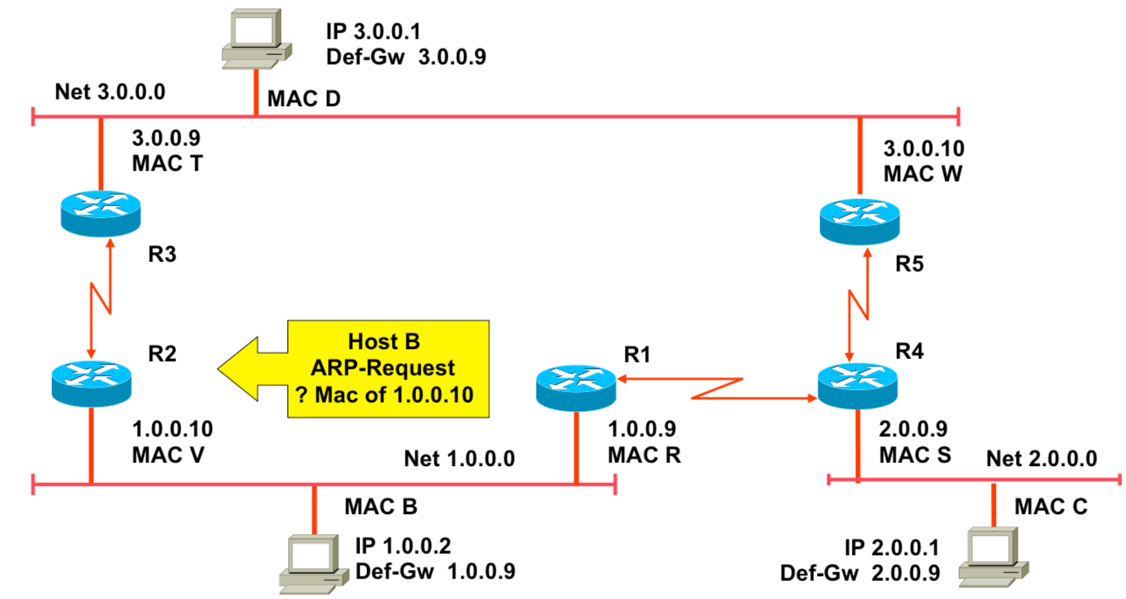
ICMP redirect



Message 5b is sent to IP 1.0.0.2 !!!

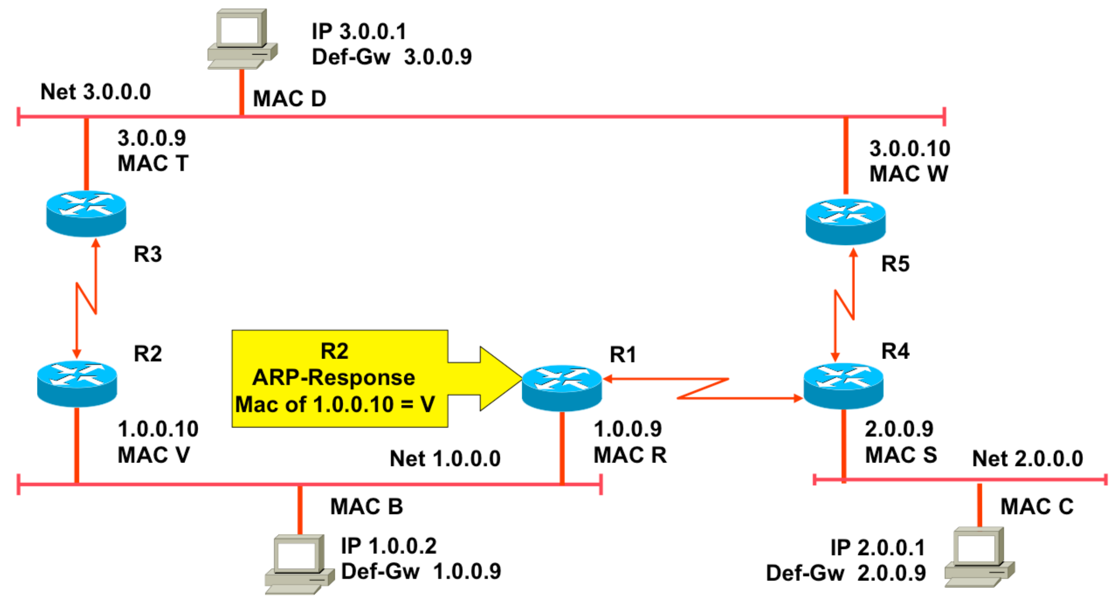
IP Technology (v6.4)

Delivery 1.0.0.2 -> 3.0.0.1



		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

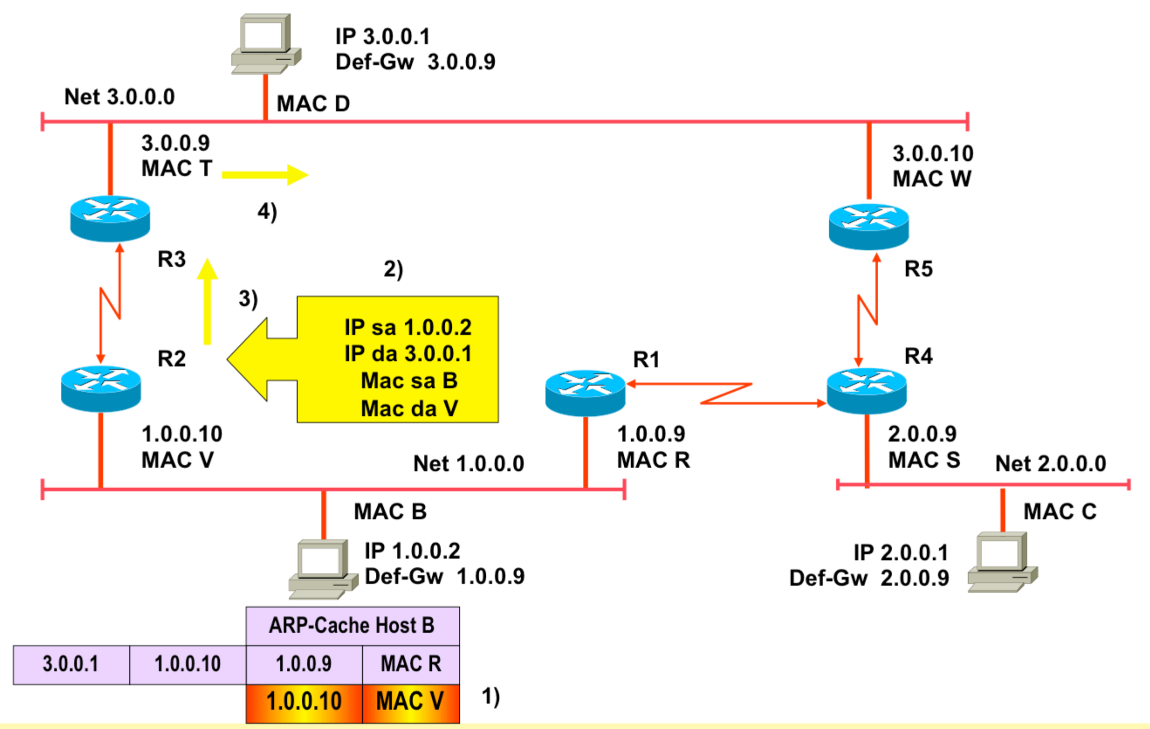


		ARP-Cache Host B	
3.0.0.1	1.0.0.10	1.0.0.9	MAC R
		1.0.0.10	MAC V

 IP Technology Primer, v6.4 203

ARP response is sent to IP 1.0.0.2

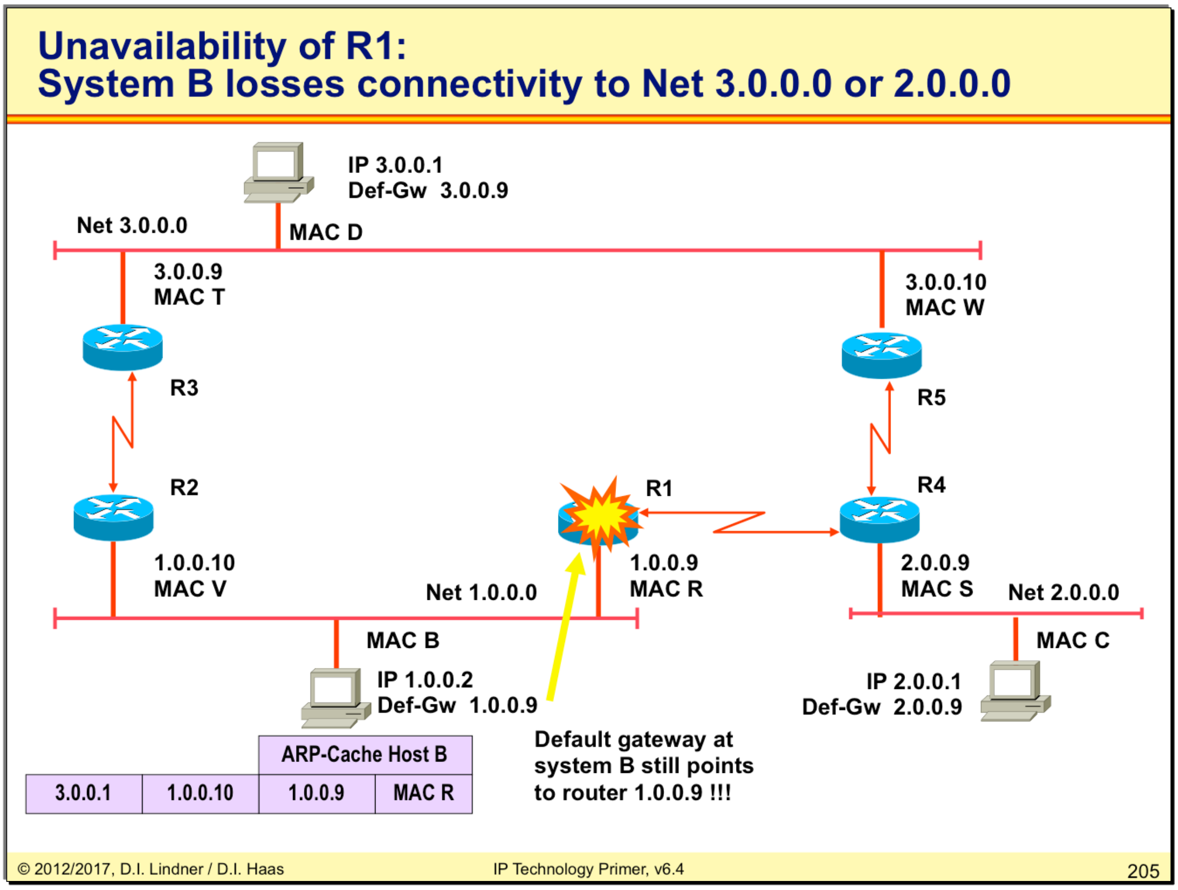
Next Packet 1.0.0.2 -> 3.0.0.1



 IP Technology Primer, v6.4 204

Next datagram of 1.0.0.0 is now sent to the correct (nearer) router.

IP Technology (v6.4)



Imagine what happens when R1 is lost. IP host 1.0.0.2 can nor reach any foreign IP network anymore, because it still points to MAC R of 1.0.0.9. We have a black-hole problem. Even after ARP cache times out there is still no automatic possibility for IP host 1.0.0.2 to switch to router R2.

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

 IP Technology Primer, v6.4 206

First Hop Redundancy 2

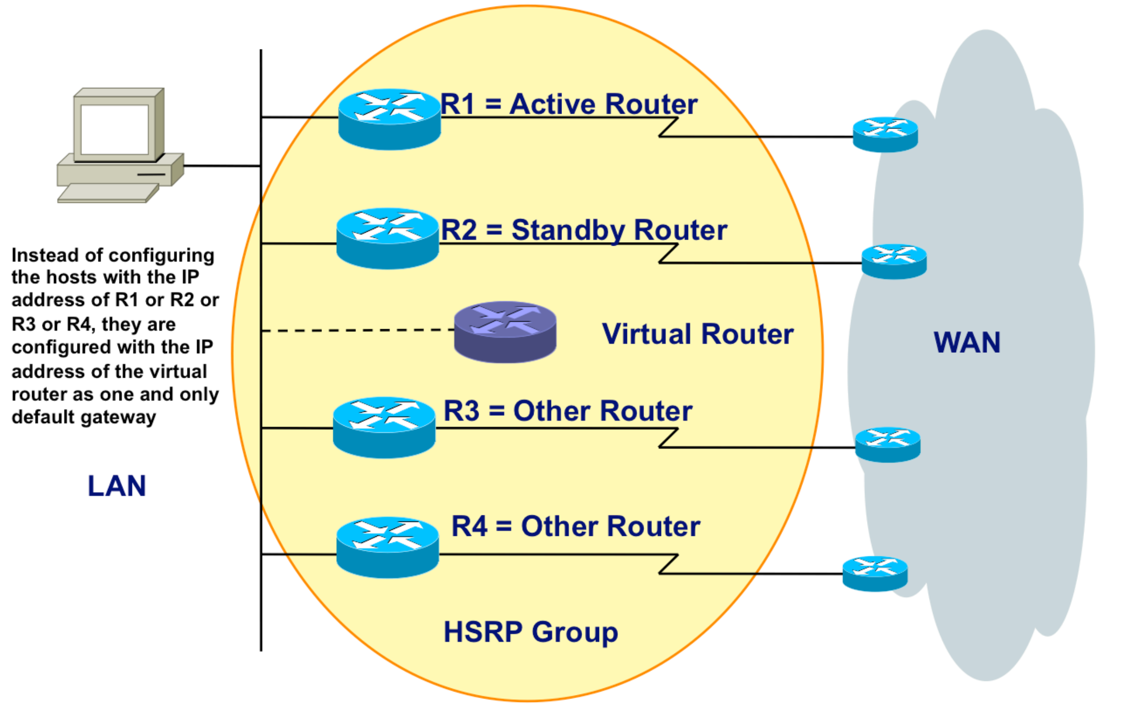
- **Methods for solving it at the IP host level:**
  - Proxy ARP
  - DHCP (Dynamic Host Configuration Protocol)
- **Methods for solving it at the IP router level:**
  - HSRP (Hot Standby Router Protocol)
    - Cisco proprietary
  - VRRP (Virtual Router Redundancy Protocol)
    - Same as HSRP but open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

IP Technology (v6.4)

HSRP Overview



 IP Technology Primer, v6.4 209

Router 1 is configured as the active router. It is configured with the IP address and the MAC address of the virtual router and listens to both virtual addresses (IP and MAC). The standby router, R2 is also configured with the IP address and MAC address of the virtual router (IP and MAC). If for any reason Router 1 stops, the HSRP routing protocol converges, and Router 2 assumes the duties of Router A and becomes the active router. Router 2 is now listening to the virtual IP address and the virtual MAC address. Additionally one of the other routers is elected to be the new standby router.

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

 IP Technology Primer, v6.4 210

Note: Routers must be able to support more than one unicast MAC address on an Ethernet interface. The active router has to listen to its own MAC address and the MAC address of the virtual router, it represents. That is not the normal behavior of an Ethernet network card. Therefore new network hardware was necessary for routers in order to support HSRP.

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

Tracking options have to be configured – otherwise only failover scenario 1 will be supported by HSRP.

Connectivity loss to a WAN interface is detected by Cisco IOS basic tracking options, Connectivity loss to an IP route is detected by Cisco IOS enhanced tracking options. The presence of enhanced tracking options depends on IOS version.

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

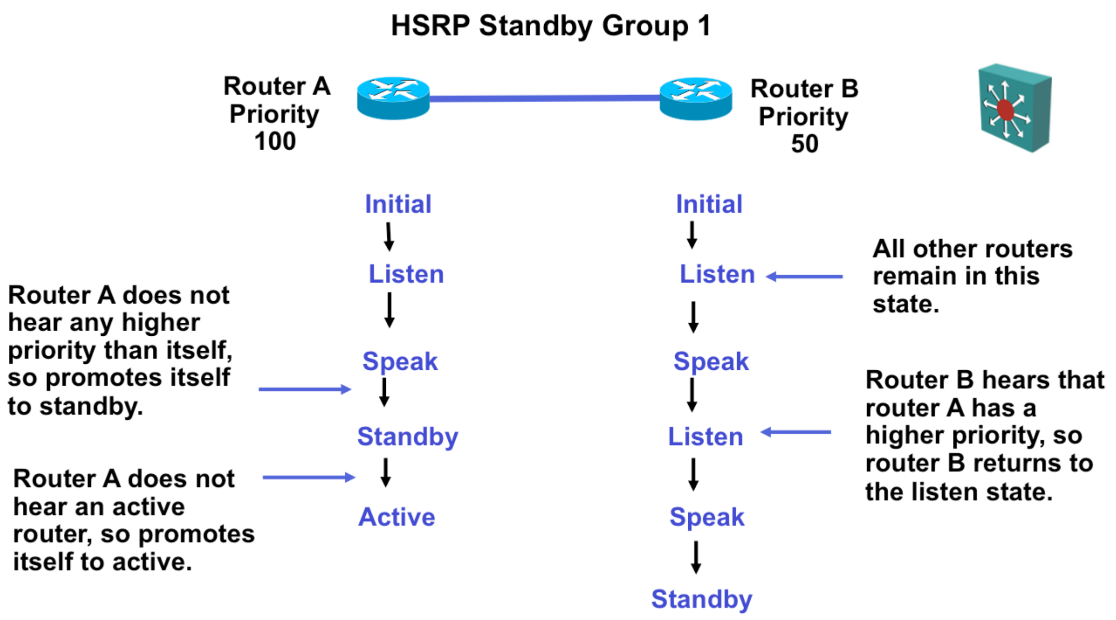
 IP Technology Primer, v6.4 214

HSRP Versions:

HSRP version 1:
Second timers
256 groups (0 – 255)
Virtual Mac Address: 00-00-0C-07-AC-XX (XX value = HSRP group number)
IP multicast 224.0.0.2

HSRP version 2:
Millisecond timers
Hello-time 15 - 999 milliseconds
Hold-time - 3000 millseconds
4096 groups (0-4095) –> Allow a HSRP group number to match the extended VLAN-ID
Virtual Mac Address: 00-00-0C-9F-FX-XX (X-XX value = HSRP group number)
IP multicast 224.0.0.102 -> To avoid conflicts with CGMP (Cisco Group Management Protocol, which uses 224.0.0.2)

HSRP States Details



© 2012/2017, D.I. Lindner / D.I. Haas IP Technology Primer, v6.4 215

Initial state— All routers begin in the initial state. This state is entered via a configuration change or when an interface is initiated.

Learn state— The router **has not determined the virtual IP address**, and has **not yet seen a hello message from the active router**. In this state, the router is still waiting to hear from the active router.

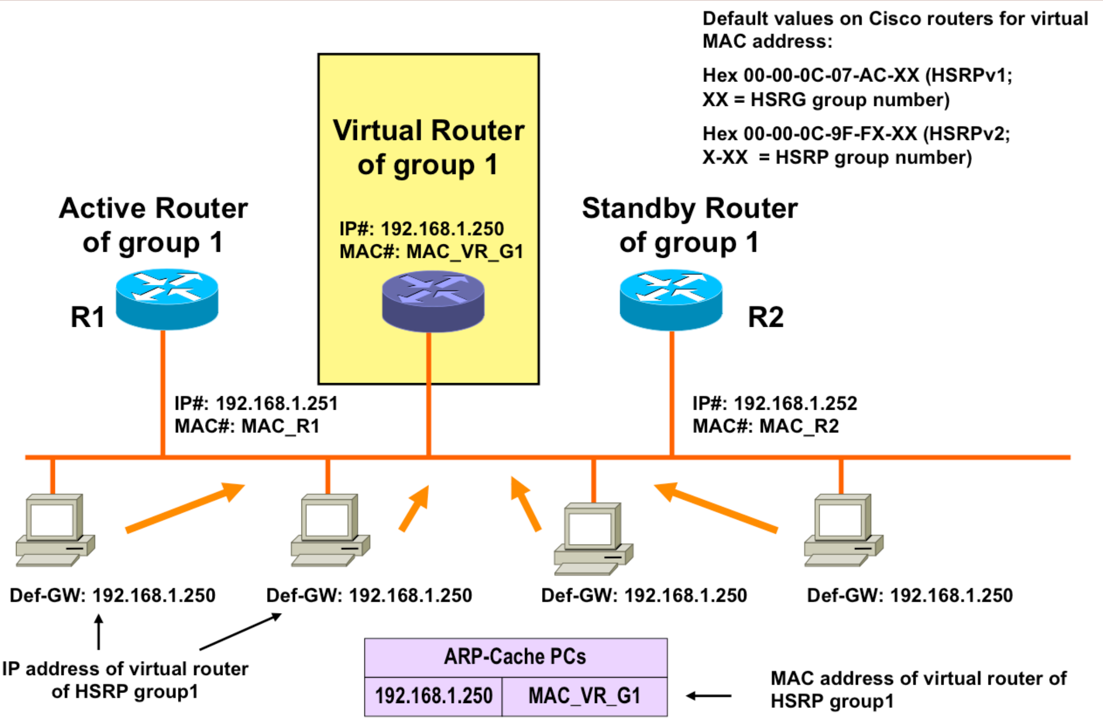
Listen state— The router **knows the virtual IP address**, **but is neither the active router nor the standby router**. All other routers participating in the HSRP group besides the active or standby routers reside in this state.

Speak state— HSRP routers in the speak state **send periodic hello messages and actively participate in the election of the active or standby router**. The router remains in the speak state unless it becomes an active or standby router.

Standby state— In the standby state, the HSRP router is a **candidate to become the next active router** and sends periodic hello messages. There must be at least one standby router in the HSRP group.

Active state— In the active state, the router is **currently forwarding packets** that are sent to the virtual MAC and IP address of the HSRP group. The active router also sends periodic hello messages.

HSRP: Real and Virtual IP Addresses / MAC Addresses



 IP Technology Primer, v6.4 216

Some more HSRP details:

The active router assumes and maintains its active role through the transmission of hello messages (default 3 seconds, HSRP version 1.

The hello interval time defines the interval between successive HSRP hello messages sent by active and standby routers.

The router with the highest standby priority in the group becomes the active router.

The default priority for an HSRP router is 100.

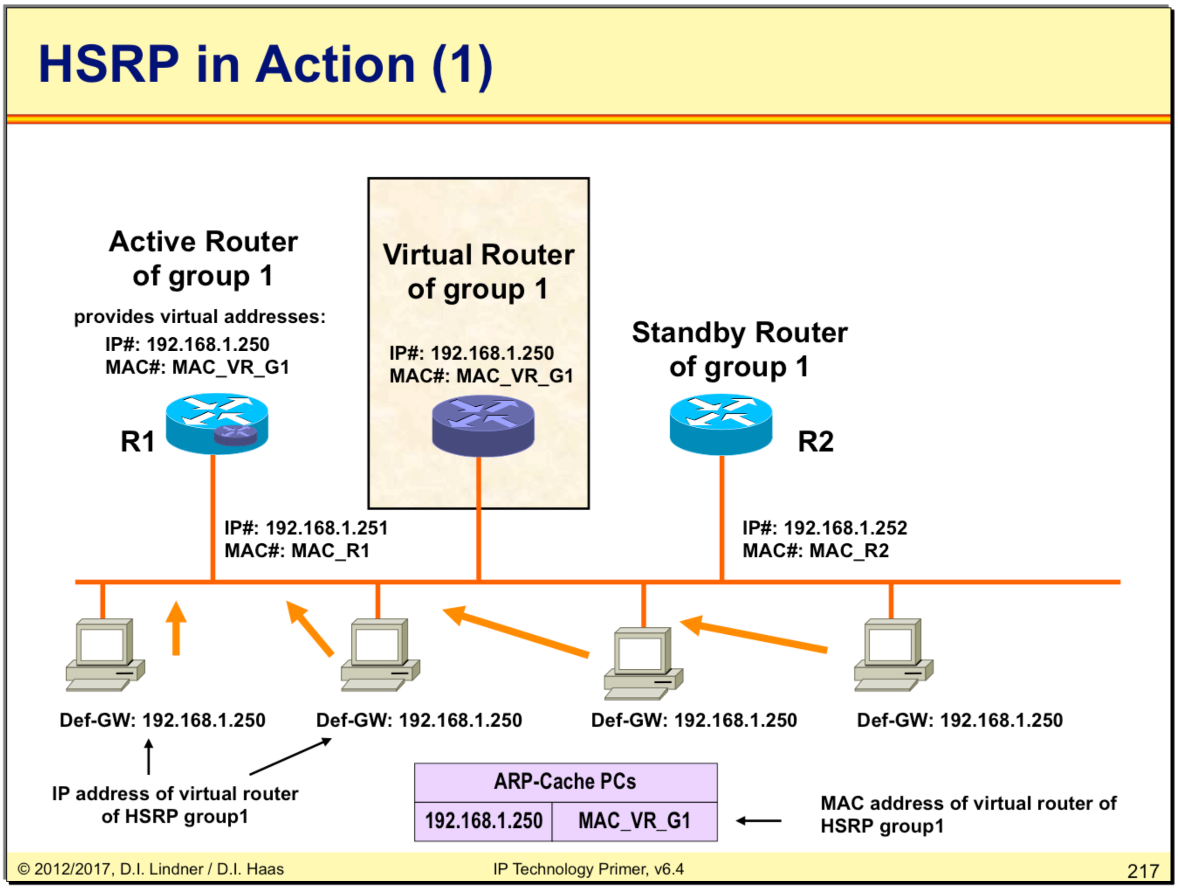
When the preempt option is not configured, the first router to initialize HSRP becomes the active router.

The second router in the HSRP group to initialize or second highest priority is elected as the standby router.

The function of the standby router is to monitor the operational status of the HSRP group and to quickly assume datagram-forwarding responsibility if the active router becomes inoperable.

The standby router also transmits hello messages to inform all other routers in the group of its standby router role and status.

IP Technology (v6.4)



Some more HSRP details:

The virtual router presents a consistent available router (default gateway) to the hosts .

The virtual router is assigned its own IP address and virtual MAC address. However, the active router acting as the virtual router actually forwards the packets.

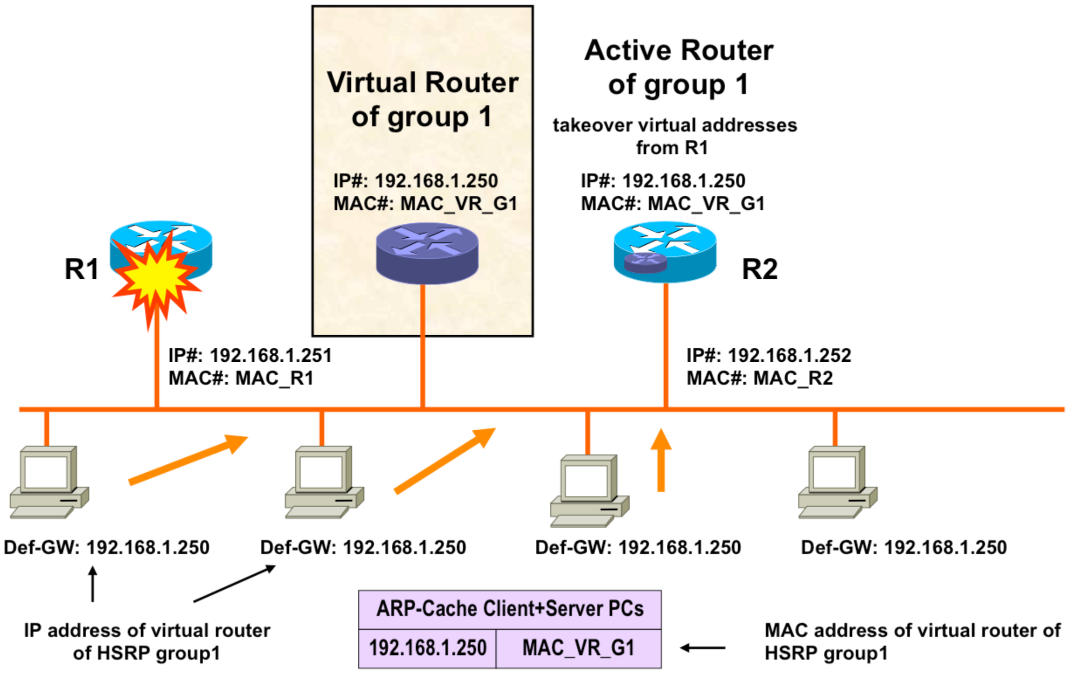
Additional HSRP member routers - other routers :

These routers in listen state monitor the hello messages but do not respond.

They forward any packets addressed to their own IP addresses.

They do not forward packets destined for the virtual router because they are not the active router.

HSRP in Action (2)



 IP Technology Primer, v6.4 218

Some more HSRP details:

When the active router fails, the HSRP routers stop receiving hello messages from the active and the standby router assumes the role of the active router.

This occurs when the *holdtime* expires (default 10 seconds, HSRP version 1).

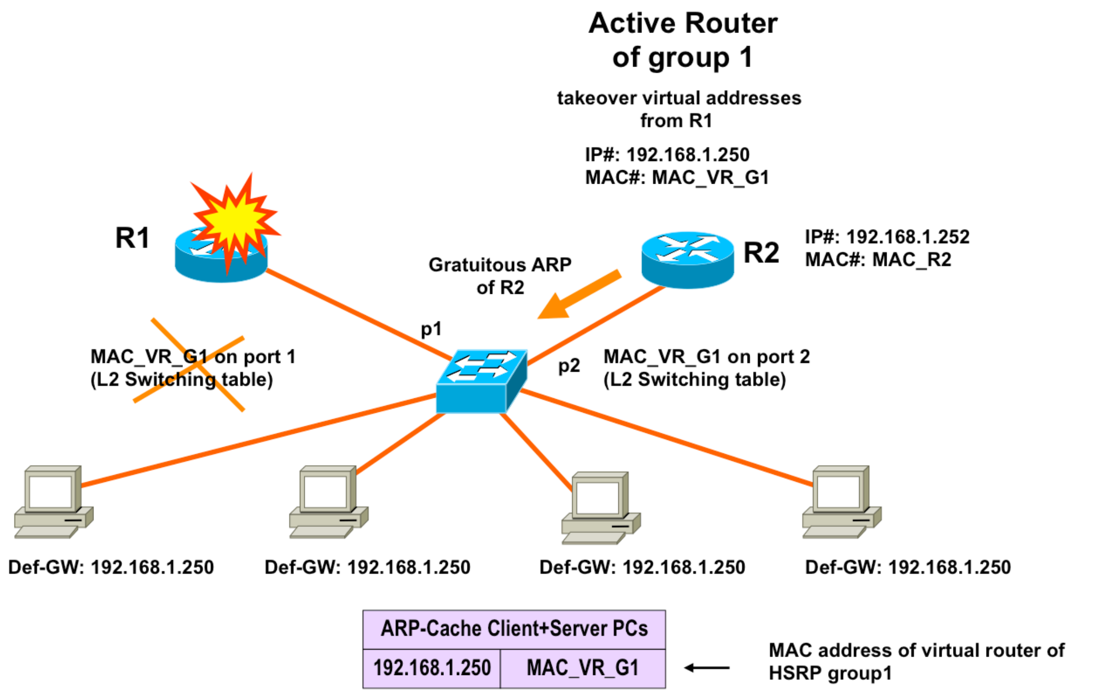
If there are other routers participating in the group, those routers then contend to be the new standby router.

Because the new active router assumes both the IP address and virtual MAC address of the virtual router, the end stations see no disruption in service.

The end-user stations continue to send packets to the virtual router's virtual MAC address and IP address where the new active router delivers the packets to the destination.

IP Technology (v6.4)

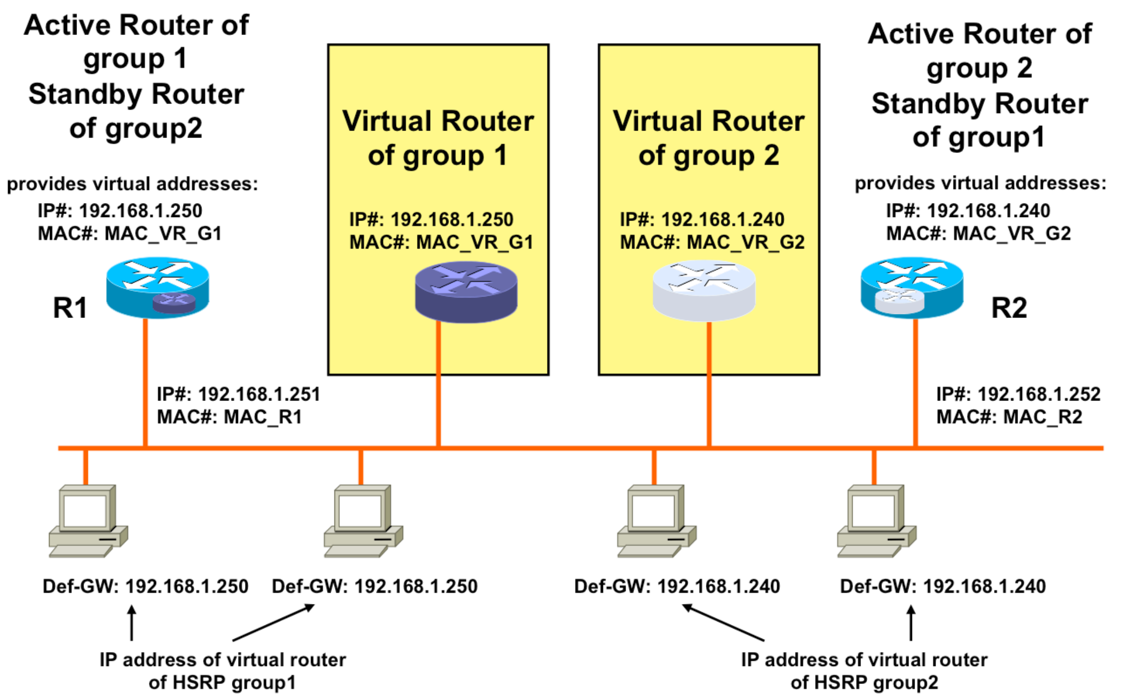
HSRP – Gratuitous ARP



 IP Technology Primer, v6.4 219

Gratuitous ARP has to be sent by router R2 in order to actualize the MAC address table of the underlying L2 Ethernet switches. Now port p2 points to the virtual Mac address MAC_VR_G1.

HSRP Load Balancing



 IP Technology Primer, v6.4 220