IP Technology

Introduction, IP Protocol Details
IP Addressing and IP Forwarding
ARP, ICMP, PPP, HSRP, VRRP

Agenda

- **Introduction**
  - Short History of the Internet (not part of the exam!)
  - Basic Principles
- **IP**
  - IP Protocol
  - Addressing
  - Classful versus Classless (not part of the exam!)
- **IP Forwarding**
  - Principles
  - ARP
  - ICMP
  - PPP
- **First Hop Redundancy**
  - Proxy ARP, IDRP
  - HSRP
  - VRRP (not part of the exam!)

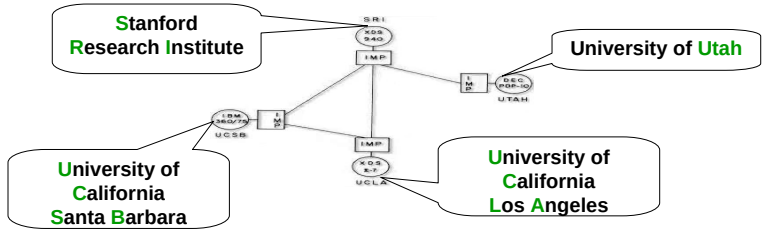
Before Arpanet

- 1957 - USSR launches Sputnik
- 1958 - US Congress Funds the Advanced Research Projects Agency (ARPA) for Space and Computer Research
- 1958 - ARPA Placed Under DOD
- 1958 - Space Research is Spun off to Separate Organization, NASA
- 1958 - ARPANET Design Discussions Started

In 1957 the USSR launched Sputnik, the first artificial earth satellite. In response, the United States formed the Advanced Research Projects Agency (ARPA) within the Department of Defense (DoD) to establish US lead in science and technology applicable to the military. The Cold War with his atomic menace lead the military to new technologies. The electronic communication was one of the important technologies. But there was one big problem with this kind of communication, if one communication-point went down, the whole communication stops. The design of the ARPANET began.

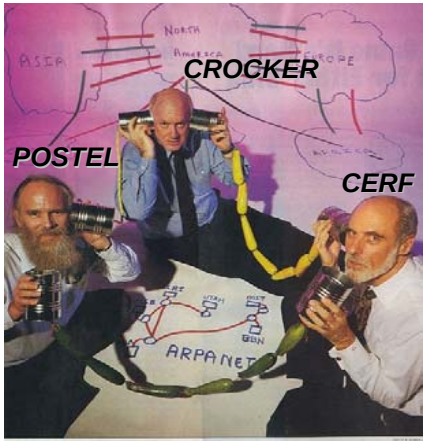
Birth Of The Arpanet

- Birth of the Arpanet: 1. Sept. 1969
- Predecessors of "Routers": Interface Message Processors (IMPs)
- First packet-switched network
- Connected UCLA, SRI, UCSB, UTAH



The physical network was constructed in 1969, linking four nodes: University of California at Los Angeles, SRI (in Stanford), University of California at Santa Barbara, and University of Utah. The network was wired together via 50 Kbps circuits. The first IMP's based on computers of the types Honeywell DDP-516 and Forts. The IMP's are the predecessors of "routers" and serve for the connection between the different computers. The first hosts where computers from IBM, DEC and SDS, and all running different operation systems.

Birth Of The Arpanet



Newsweek, Aug 8,1994

Jon Postel, Steve Crocker and Vinton Cerf working for the UCLA and developed the packet switching principle of the ARPANET.

Birth Of The Arpanet

- **1970 - Arpanet use the Network Control Protocol (NCP)**
- **1971 - Arpanet connects 15 sites including universities and research organizations**
  - Birth of TELNET and FTP
- **1972 - Ray Tomlinson created first email program**
  - ALOHAnet connected to the Arpanet

In 1969 Steve Crocker writes the first RFC to establish a way to document and to discuss the new upcoming technologies.

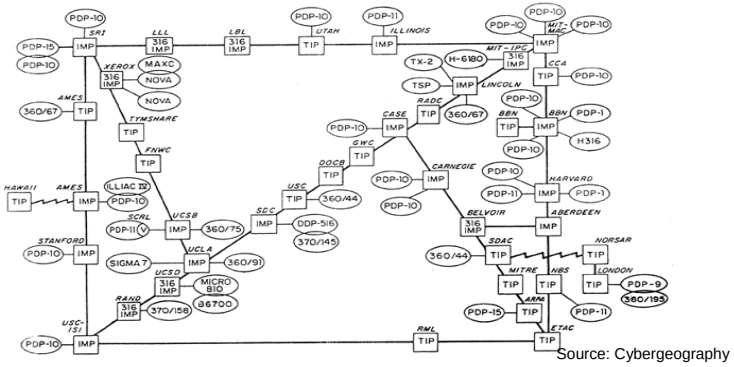
The NCP was the first protocol which connect all hosts in the ARPANET and in July 1970 the NCP protocol was standardize with the help of the RFCs. The ARPANET began to grow and connects 15 hosts:

University of California at Los Angeles (UCLA)
Stanford Research Institute (SRI)
University of California at Santa Barbara (UCSB)
University of Utah
Bolt Beranek and Newman (BBN)
Massachusetts Institute of Technology (MIT)
RAND Corporation
SDC
Harvard
Lincoln Labs
Stanford
University of Illinois at Urbana Champaign (UIUC)
Case Western Reserve University (CWRU)
Carnegie Mellon University (CMU)
NASA-Ames

The development was going on and new protocols and systems were created (TELNET, FTP, EMAIL).

Birth Of The Arpanet

- 1973 - Arpanet comprises 35+ hosts



Source: Cybergeography

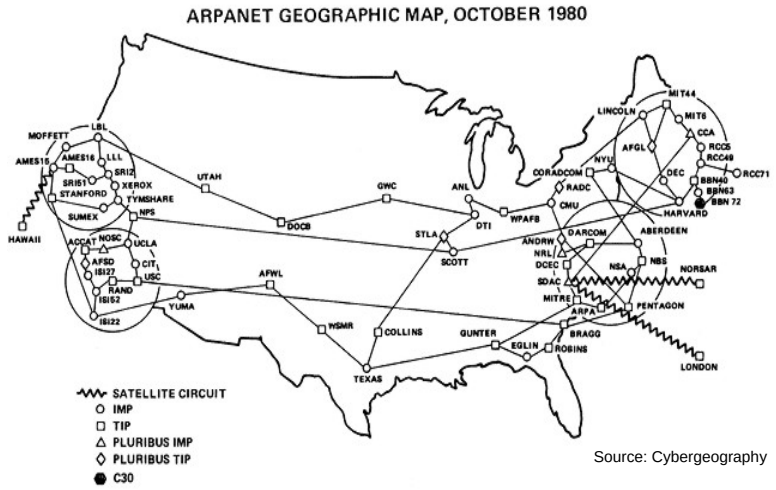
In 1973 a multiplication of the existing IMPs began. The old DDP-516 computer changes to a Honeywell 316 and got the name “Terminal IMP“ (TIP).

The Arpanet Problem - Birth of TCP/IP and the Internet

- Arpanet communicate with NCP but other networks use different protocols
- 1974 - Transmission Control Protocol (TCP) specification published
- TCP enabled the expansion from the Arpanet to a worldwide Internet !
- The Winner: TCP/IP
- 1978 - TCP Split into TCP and IP
- 1983 - Arpanet converts to TCP/IP
  - UNIX (v4.2 BSD) released with TCP/IP
  - DARPA switched from Arpanet architecture to Internet architecture with TCP/IP as base protocols

Arpanet uses NCP (Network Control Protocol) for communication between packet switches. In order to connect other networks to the ARPANET a new problem occurred, because every network uses a different protocol. Robert Khan and Vinton Cerf started to design a network overlay protocol. In Mai 1974 a workgroup was founded for solving the problem (workgroup name “A Protocol for Packet Network Intercommunication“) and in December the first RFC (RFC 675) “Specification of Internet Transmission Control Protocol” was edited. The number of hosts in the Arpanet reached 62. With the creation of TCP the expansion to a worldwide Internet was enabled. In 1978 first TCP was split into IP and TCP. After 5 years of TCP/IP development the protocol technology switched from a experimental to a operational protocol. To push the development on the TCP/IP protocol the Internet Configuration Control Board (ICCB) was created. On 1rst January 1983 NCP changed completely to TCP/IP.

Getting Bigger And Bigger



Source: Cybergeography

1986 the ARPANET enfolded over whole USA and parts of Europe.

Development Going On

- 1983 - Arpanet Splits into Arpanet and MILNET
- 1983 - Internet Activities Board (IAB)
- 1984 - Domain Name System (DNS)
- 1985 - Symbolics.com first registered domain.

The ARPANET works fine, but in 1983 the Department of Defense decided to create an own military network, called MILNET.

To push the development on the Internet the Internet Configuration Control Board (ICCB) was chanced to the Internet Activities Board (IAB). Now the IAB has the function to control and edit the RFCs.

Because of the fast grow of the Internet (around 1000 hosts) the Domain Name System (DNS) was created. It took 2 years until all hosts were connected to the DNS.

Who Is Who?

J.C.R. Licklider (MIT, ARPA/IPTO)
- Memos about a global, distributed network and addressing
- Vision of a universal network had a powerful influence

Robert Taylor (ARPA)
- Designed the ALTO workstation

Larry Roberts, Barry Wessler (ARPA)
- Roberts: the principal architect of the Arpanet
- Wessler: ARPA administrator

Wesley Clark (Washington University)

Frank Heart, Robert Kahn, Dave Walden, Willy Crowther, Severo Ornstein et al (BBN)
- First packet switch
- First router
- First person-to-person network email

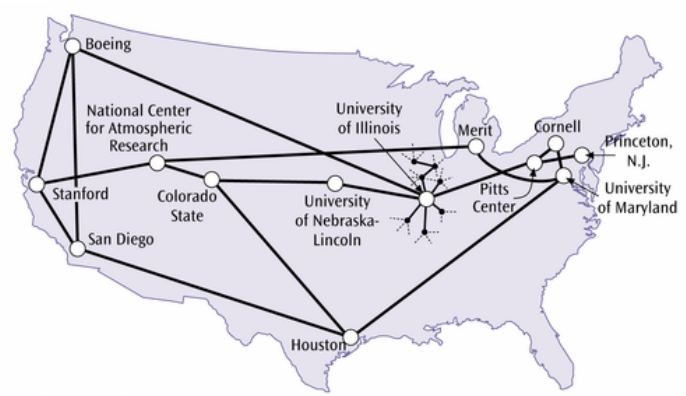
Leonard Kleinrock (UCLA) and Crocker, Postel, Kline, Braden, Cerf et al
- Principles of packet switching

NSFNET Backbone

- National Science Foundation (NFS) creates the NSFNET Backbone 1986
- It connects Cornell, Princeton, UC-SD, Pitt and UI-UC with 56k Lines
- Dramatic growth of hosts
  - 1986: February 2000, November 5000.
- Backbone is upgraded to T1 (1.544Mb/s) - 1988

In 1986 the NFS created the NSFNET backbone to which each of the local networks could be attached. With this step the diffusion to a worldwide Internet began. In 1987 the number of hosts raised above 28000 and the 1000st RFC was published. The number of hosts grew and so the backbone was upgraded to a T1 connection in 1988 (Merit Network Inc., IBM and MCI were working on that update).

NSFNET Backbone



Source: Cybergeography

The NSFNET Backbone spanned over whole USA.

The Internet

- 1989 - Number of hosts: 100,000 !
  - Reseaux IP Europeens (RIPE) founded
- 1990 - Arpanet Decommissioned, Now officially called "Internet“
- 1990 - First Internet provider, “The World” comes online

In 1989 two new organization were founded which should boost the development on TCP/IP and the Internet. The Internet Engineering Task Force (IETF) and the Internet Research Task Force (IRTF).

The Internet

- 1991 - World Wide Web (WWW) Created by Tim Berners-Lee at CERN, http://www.cern.ch/
- 1991 - Backbone is upgraded to T3 (44.736Mbps)
- 1992 - Internet Society (ISOC) is chartered
- 1992 - Number of hosts: 1,000,000

The invention of the World Wide Web (WWW) had the most important impact to the Internet growth and development. WWW had been created by Tim Berners-Lee, a MIT graduate, working for the CERN in Switzerland. The first browser he wrote was called "Nexus" and was also capable to display inline graphics already.

Backbone networks from many different organizations had been created, such as General Atomics (CERFNet) and Performance Systems International (PSINet) and UUNET Technologies (AlterNet).

Upgrades of the initial modem-speed lines to T3 and more were made during the early 1990s.

With the help of the ISOC the development was going on and many new protocols, for example: Multipurpose Internet Mail Extensions (MIME), were created.

The Internet

- 1992 - Term: “Surfing the Internet” coined by Jean Armour Polly
- 1993 - Mosaic introduced first graphical Web browser
- 1993 - WWW is 0.1% of NSFNET Traffic

In the following years the WWW significantly influenced the development of the Internet. Although only a few users could utilize this new service, many journalists paid great attention to WWW, and soon everybody wanted to "surf in the Internet". Also 1993 was a milestone in the history of the WWW, as NCSA released the first fully featured graphical web browser called "Mosaic". Later the famous Netscape Navigator was created upon this code.

Network Access Points

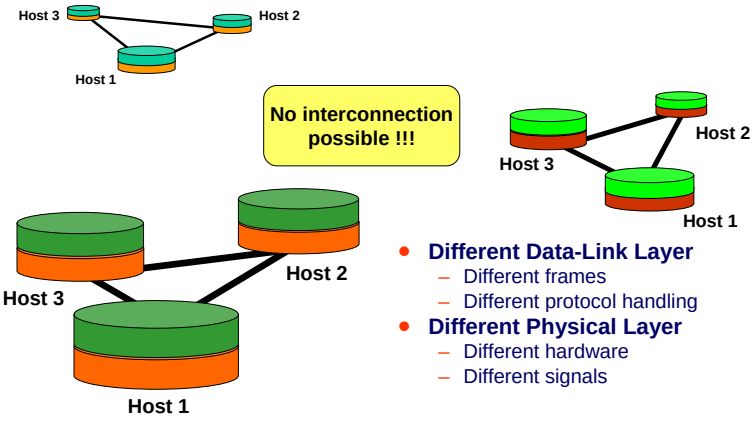
- 1993 - NSF specifies creations of Network Access Points (NAPs)
  - Privatize the Internet – Replace Government funded
  - NSFNET backbone with (many) commercial Internet backbones
  - Central points to Interconnect Commercial Internet Backbones
  - Allow anyone to access the Internet via Internet
  - Service Providers (ISPs) – Connected to Backbones
- 1994 - Four NAPs Created
  - San Francisco, Chicago, Washington D.C., New Jersey
- 1995 - NSFNET Backbone is decommissioned

The NSP plan to chance the backbone structure of the Internet and choose to leave the Internet. Instead of the backbone structure many independent Network Access Points should be created. On this NAP’s regional networks can connect. The growth of the Internet was around 100% per year and so there where around 2 millions host and over 16000 networks connected each other.

Agenda

- **Introduction**
  - Short History of the Internet (not part of the exam!)
  - Basic Principles
- **IP**
  - IP Protocol
  - Addressing
  - Classful versus Classless (not part of the exam!)
- **IP Forwarding**
  - Principles
  - ARP
  - ICMP
  - PPP
- **First Hop Redundancy**
  - Proxy ARP, IDRP
  - HSRP
  - VRRP (not part of the exam!)

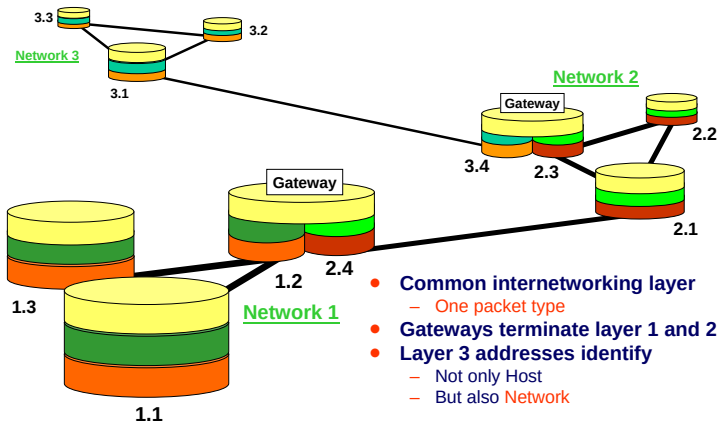
Need of an Inter-Net Protocol (1)



- **Different Data-Link Layer**
  - Different frames
  - Different protocol handling
- **Different Physical Layer**
  - Different hardware
  - Different signals

Why do we need an Inter-Net Protocol? Different networks have different Data-Link Layer. Every Network runs a different protocol. Some networks use proprietary link layer protocols or X.25, other networks have Ethernet or HDLC. You see, every network has its own hardware, signals and frames. As long as they do not want to communicate with each other, there is no problem...

Need of an Inter-Net Protocol (2)



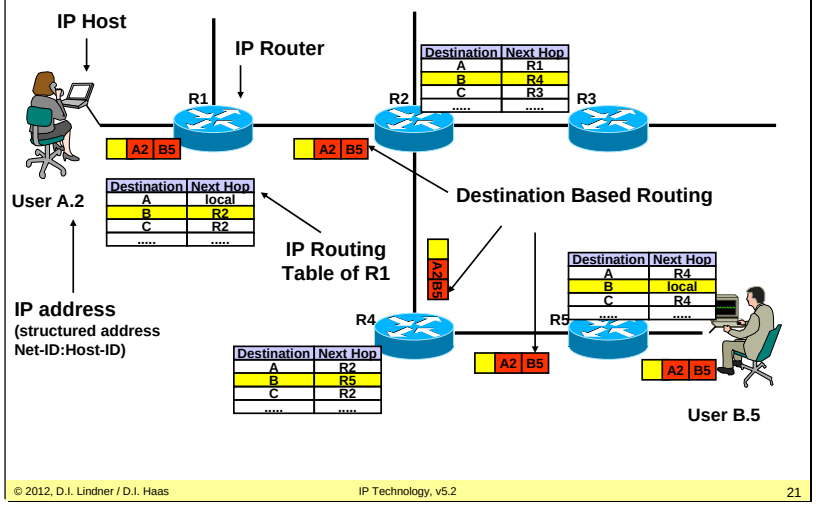
- **Common internetworking layer**
  - One packet type
- **Gateways terminate layer 1 and 2**
- **Layer 3 addresses identify**
  - Not only Host
  - But also Network

If we want to interconnect these networks we would need a common internetworking layer. Network interconnections are realized with dedicated hosts called "Gateways" which include at least two different network interface cards (NIC) – each with an appropriate physical and link layer. These gateways transport the common Inter-Net protocol (encapsulated in layer 2) and terminate layer 1 and layer 2 on each side. In the late 1970's the IP protocol was widely used as Inter-Net protocol. It works on Layer 3 and identifies the host and the network using dedicated addresses.

IP Technology

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

IP Datagram Service

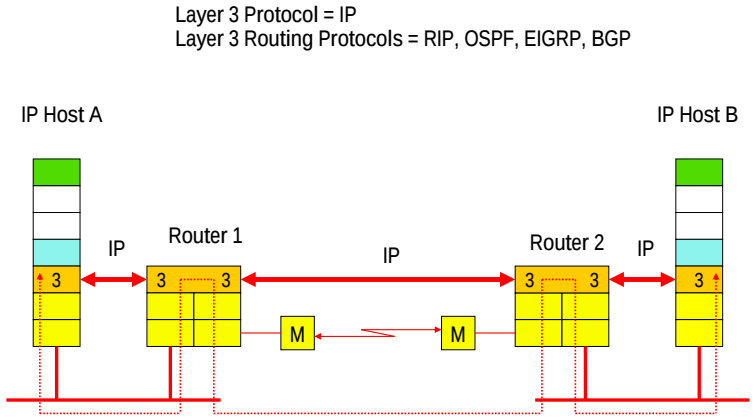


In the Datagram technology user A.2 sends out data packets destined for the user B.5. Each single datagram holds the information about sender and receiver address.

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.

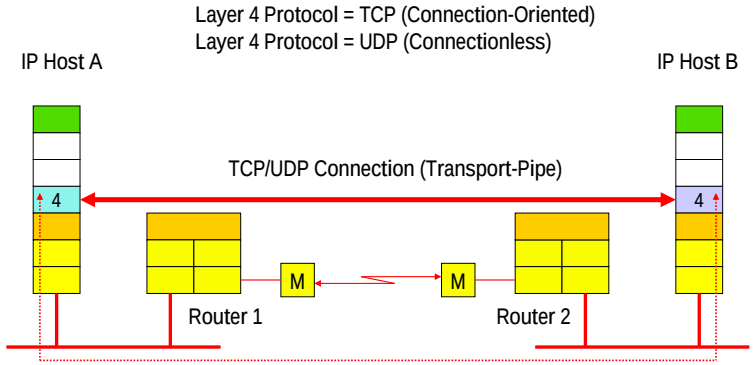
IP and OSI Network Layer 3



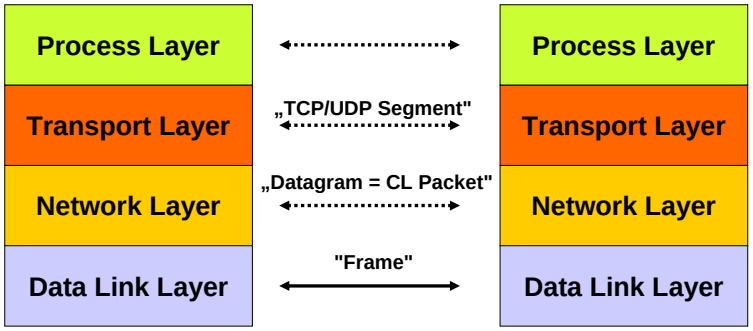
TCP Technology

- **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/UDP and OSI Transport Layer 4

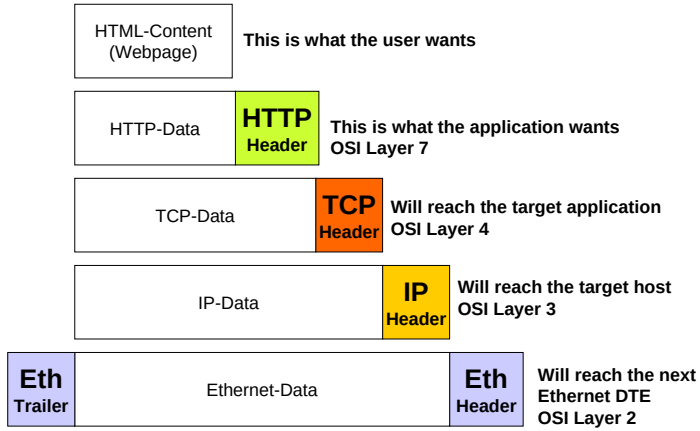


DoD 4-Layer Model (Internet)



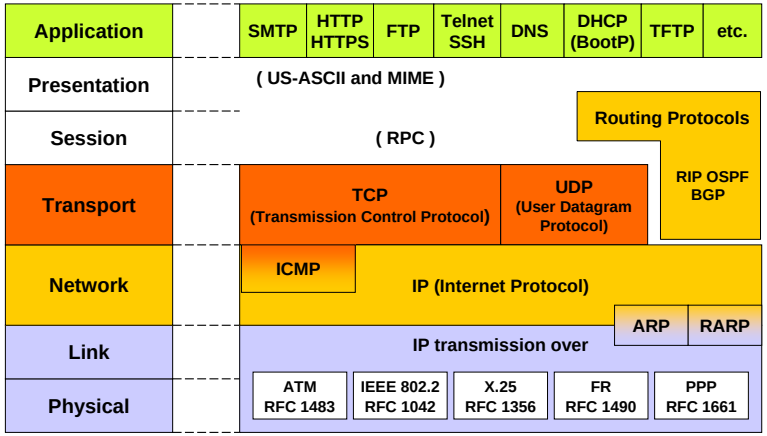
The picture above shows the W. Stevens 4 layer model which is used also in the Internet. The Internet layer model is also called "Department of Defense" (DoD) model.

Internet Encapsulation



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.

TCP/IP Protocol Suite



 IP Technology, v5.2 27

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 Story of Success

- **IP over everything**
  - Overlay technique
- **End-to-end principle**
  - Network could be stupid simple
  - End systems do the sophisticated tasks like TCP
- **TCP**
  - Best implementation of a transport protocol nowadays
- **WWW**
  - Killer application in the 1990´s
- **Standardization**
  - Standardization of running code

 IP Technology, v5.2 28

One reason for IP's success is its ability to adapt to all types of layer 2 technologies. On one hand, the IP developers were very quick to design convergence ("helper") protocols, for example to resolve L2/L3 addresses on multipoint connections or encapsulation headers for delineation on dialup or serial links, such as PPP. On the other hand, IP is a relative simple protocol. Because of this it had been integrated in many different operating systems, most importantly UNIX.

IP over everything means that layering a unique IP protocol on top of various network technologies is technology-independent. Just a definition is necessary how to transfer IP datagrams using a given transmission- or network-technology. Hence it is easy to adopt to new network technologies.

Note: IP's simplicity is based on the end-to-end philosophy. That is, the network itself does not care for reliable transmission; only the end-systems care for error recovery. This way, the network can be kept simple.

End-to-end principle avoids sophisticated tasks to be performed by network infrastructure (routers). The IP host takes care if reliability of information transport is necessary. Routers can be held dump, IP hosts are the smart ones.

TCP is tolerant and adaptive to network operational conditions, robust against network failures, adapts to varying network delays and varying network load.

Right functionality partition between IP and TCP: IP knows nothing about end systems applications, makes best effort to route packets through the network, it only cares about networks and host-addresses. TCP takes care of end-to-end issues (error recovery, flow control, sequencing,…). hence end systems need to know nothing about network internals (Note: that might change with the need for QoS in the IP world). TCP carries the Port-Number. The Port-Number is necessary for the host. With the Port-number he knows which datagram belongs to which application.

WWW was invented 1991, world take first notice in 1993. WWW (the web browser) was the killer application allowing normal people to use technology for information gathering, communication and fun.

Internet Standardization - RFC

- **Requests for Comments (RFC)**
  - "Give me your input to my ideas I have already implemented"
- **Today's process is best described by**
  - RFC-2026 (The Internet Standards Process Revision3)
  - Draft -> IETF decision if new RFC -> RFC number
- **Status April 2012:**
  - RFC 6607
- **Attention:**
  - Not every RFC is an Internet Standard
  - Categories:
    - Informational, Experimental, Historic
    - Proposed Standard
    - Draft Standard
    - Standard
- **Where to find:**
  - http://www.rfc-editor.org/index.html

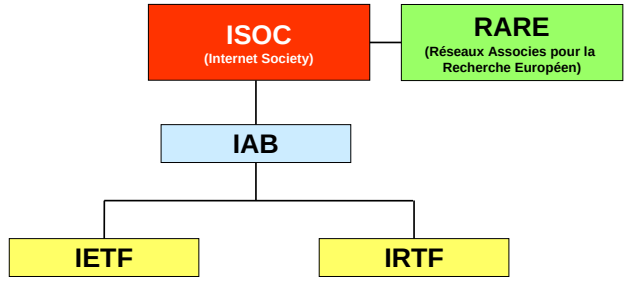
 IP Technology, v5.2 29

All documentation, standards, proposals for new protocols and enhancements for the Internet are published as RFCs which are accessible by everyone for free.

RFCs were the initial approach of engineers to discuss questions, suggestions via e-mail in order to speed up development compared to the slow processes known by other standardization organization such as ISO and ITU.

Nowadays a RFC starts as a draft document with a version number. A draft can be written by everyone who likes it. The IETF (Internet Engineering Task Force) decides if the draft is something which is "good" for the Internet technology or not. If not or if the draft is seen to be not complete the draft will remain for six months at the IETF server and will be removed after six months. The draft owner can create an adapted draft with a new version number and the game starts again. If finally a draft is seen as something which is worth to be considered, it will get a RFC number. RFCs are numbered in sequence of publishing hence adopted enhancements or changes to a protocol will result in a new RFC number.

Internet Organizations



 IP Technology, v5.2 30

The Internet Society (ISOC) provides leadership in addressing issues that confront the future of the Internet, and is the organization home for the groups responsible for Internet infrastructure standards, including the Internet Engineering Task Force (IETF) and the Internet Architecture Board (IAB).

The Réseaux Associes pour la Recherche Européen (RARE) was founded in 1986 to build and maintain a European high speed data network infrastructure. RARE is also a member of ISOC and ETSI (European Telecommunications Standards Institute). EBONE was initiated by RARA and RARA is a close cooperation with RIPE (Réseaux IP Européen).

The Internet Architecture Board (IAB) is responsible for technical directions, coordination and standardization of the TCP/IP technology. It was formerly known as Internet Activity Board and is the highest authority and controls the IETF and IRTF.

The Internet Engineering Task Force (IETF) is "actually" the most important technical organization for the Internet working groups and is organized in several areas. Area manager and IETF chairman form the IESG (Internet Engineering Steering Group). The IETF is also responsible to maintain the RFCs.

The Internet Research Task Force (IRTF) coordinates and prioritize research groups that are controlled by the IRSG (Internet Research Steering Group).

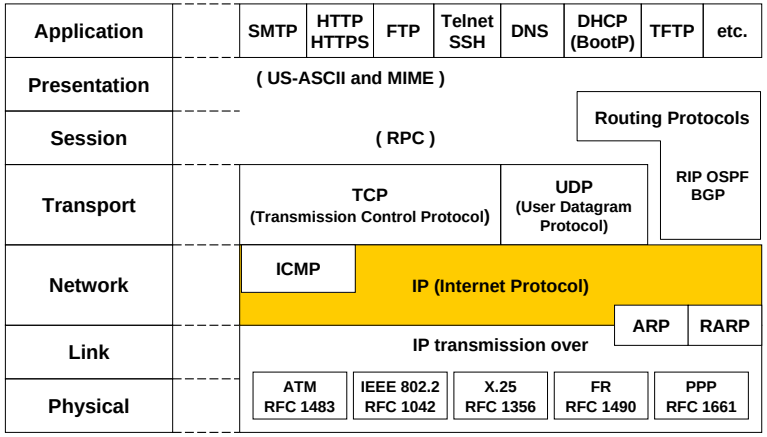
Internet in Europe

- **RIPE NCC (Réseaux IP Européens Network Coordination Center)**
  - Internet Registry
    - Assigning IP addresses
    - Assigning AS numbers
  - Routing Registry
    - Coordinating policies between Internet Service Providers (ISP)
  - How to contact?
    - RIPE NCC
    - Singel 258
    - 1016 AB Amsterdam
    - The Netherlands
    - Phone: +31 20 535 4444 , Fax: +31 20 535 4445
    - E-Mail: <ncc@ripe.net>, WWW: <http://www.ripe.net>

Agenda

- **Introduction**
  - Short History of the Internet (not part of the exam!)
  - Basic Principles
- **IP**
  - IP Protocol
  - Addressing
  - Classful versus Classless (not part of the exam!)
- **IP Forwarding**
  - Principles
  - ARP
  - ICMP
  - PPP
- **First Hop Redundancy**
  - Proxy ARP, IDRP
  - HSRP
  - VRRP (not part of the exam!)

TCP/IP Protocol Suite



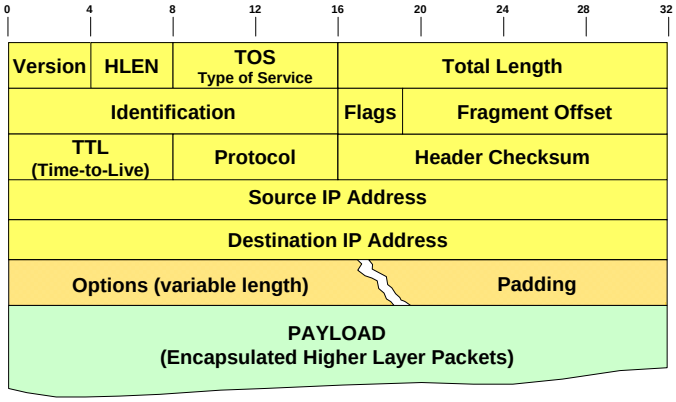
IP Internet Protocol (RFC 791)

- **OSI layer 3 protocol**
  - With datagram service (unreliable connectionless service, "best effort service")
- **Transports packets (datagrams) from a sender to a receiver**
  - Through one or more networks
- **Doesn't guarantee**
  - Delivery or correct sequence of packets (-> task of higher layers)
- **IP datagrams are encapsulated in layer 2 frames**
  - Encapsulation is a key feature of the TCP/IP suite: It provides versatility and independence from the physical network

IP Protocol Functions

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

The IP Header



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

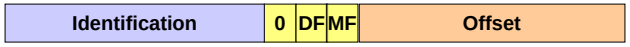
Destination IP Address: 32 bit IP address of the receiver (destination) of a datagram

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.

Fragmentation Fields

- **Identification:**
  - All fragments of a datagram have the same unique identification
  - Necessary for reassembling fragments **at the destination**
  - In praxis a hidden sequence number although not really used because of the connectionless best-effort delivery behavior of IP
- **Fragment Offset:**
  - Indicates the position of a fragment in relation to the beginning of the original datagram
  - Offset is measured in multiples of 8 bytes (64 bits)
- **Flags:**
  - DF (Don't Fragment)
    - Can be used for Path MTU discovery
  - MF (More Fragments)
    - More fragments of the same original datagram will follow



 IP Technology, v5.2 37

As already mentioned fragmentation is necessary, if a datagram has to pass a network with a smaller maximum frame size / MTU size than the current length of the given datagram.

Some details for fragment offset: The first fragment and non-fragmented packets have an offset of 0. Fragments (except the last) must be a multiple of 8 bytes. Fragments with the same combination of source address / destination address / protocol / identification will be reassembled to the original datagram at the receiver.

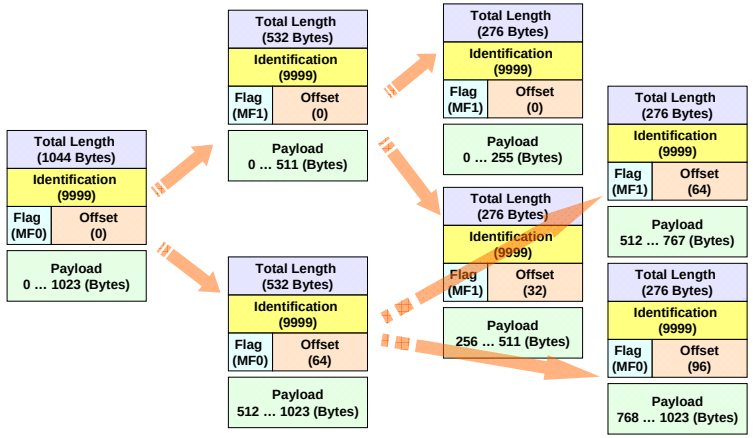
Flags:

DF (Don't fragment): If set fragmentation is not allowed and the corresponding datagram has to be discarded by router if MTU (maximum transmission unit) size of next link is too small. This can be used for Path-MTU discovery where an IP host will probe which is the best datagram size without experiencing fragmentation in the network. Fragmentation has performance constraints: It is bad for the router performance because of the fragmentation process and also bad for the IP host performance because of reassembling. Because of this, packets are typically sent with the lowest MTU size that may occur somewhere in the network. An (older) RFC recommendation specifies 576 Bytes to be used as minimum MTU but in the age of Ethernet most people use 1500 Bytes to gain more efficiency. IP version 6 does not fragment anymore but uses Path MTU discovery instead.

MF (More fragment): If set more fragments will follow. The last fragment of a given datagram will have MF set to 0.

Reassembling: Is done at the destination, because fragments can take different paths. Buffer space has to be provided at the receiver. Some fragments of a datagram may not arrive because of the unreliable nature of IP. If a datagram can not be reconstructed because of missing fragments in order to free buffers a reassembly timer is used. The first arriving fragment of an IP datagram (with MF=1 or MF=0 with unequal 0) starts the timer. The TTL of this fragment is used a s timeout in seconds. If the timer expires before the datagram was reconstructed, all fragments stored in the buffer so far will be discarded.

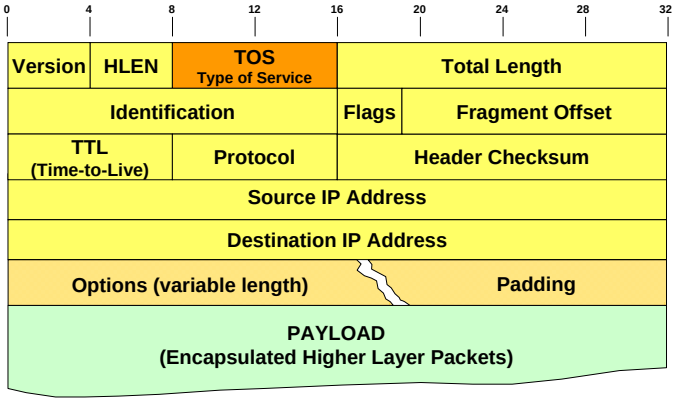
IP Fragmentation in Action



 IP Technology, v5.2 38

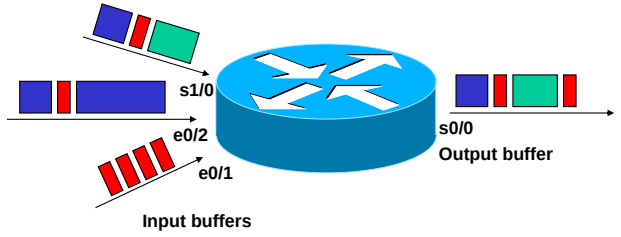
The example above shows how an IP packet (left) is fragmented into two smaller fragments (middle) by a router and further fragmented into a total number of four fragments by a another router (right).

The Way to IP QoS (0):



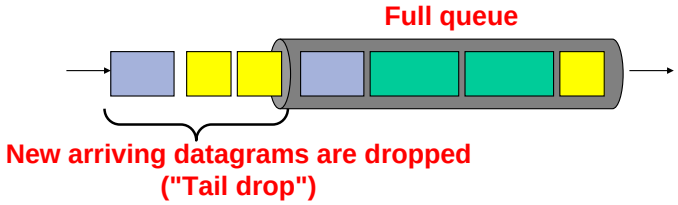
The Way to IP QoS (1): Need for Queuing

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



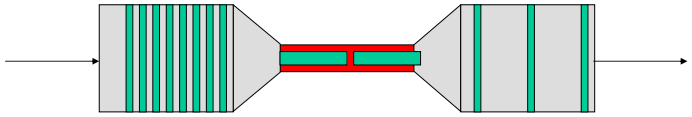
The Way to IP QoS (2): No QoS with FIFO Queuing

- ***Tail-drop queuing*** **is the standard dropping behavior in FIFO queues**
  - If queue is full all subsequent datagrams are dropped



The Way to IP QoS (3): Bottleneck and Traffic Bursts

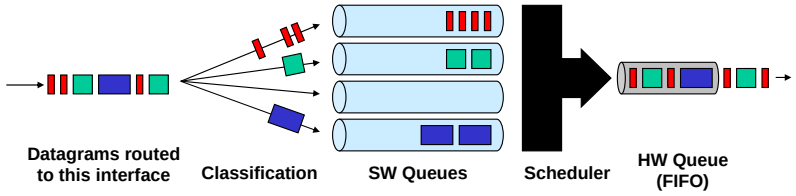
- Problem (buffer overflows) appears at bottleneck links



Pipe model of a network path: Big fat pipes (high data rates) outside, a bottleneck link in the middle. The green packets are sent at the maximum achievable rate so that the inter packet delay is almost zero at the bottleneck link; however there is a significant inter packet gap in the fat pipes. s

The Way to IP QoS (4): QoS with SW Queues

- **Queuing actually encompasses two parts: SW and HW queues!**
- **HW queuing is typically only FIFO**
- **SW queuing is typically more sophisticated**
  - WRR, CBWFQ, LLQ, etc
- **SW queue only needed if HW-queue full**
  - Otherwise datagrams bypasses the SW-queues



The Way to IP QoS (5)

- **TOS (Type Of Service)**
  - Old meaning (RFC 791 and RFC 1349)
  - Priority (precedence) of a datagram in relation to other datagrams queued up in the router
  - Preferred network characteristics to be expected by that datagram
  - **Precedence bits:**
    - Allow router to queue datagrams in different output queues in case of congestion
    - Allow router to schedule datagrams of different queues according to a QOS (Quality Of Service) policy (e.g. round robin, priority)
  - **D, T, R and C bits**:
    - low **D**elay, high **T**hroughput, high **R**eliability, low monetary **C**ost
    - Can be used to forward a datagram according to a routing table which corresponds to the preferred network characteristics for that destination
      - Needs routing tables per network characteristic

Both things were not really useable in IP networks. Why? If people know that they will get better performance by setting these bits they will do it. Without any control between IP hosts and the IP network (who is allowed to set the bits and who not) a QOS policy can not be implemented in a network. So in the past a router set all bypassing user IP datagrams to precedence 0 and just use precedence 7 for prioritizing own or received routing messages . The idea having different routing tables according different network characteristics (e.g. differentiating between long-delay satellite links versus small delay terrastic links) in a router failed because there was no dynamic routing protocol supporting different routing tables in a router for a long period (OSPF was the first routing protocol to allow different metrics for different characteristics, but is was to late; the last version of OSPF removed that support!).

TOS Field (RFC 791, 1349)

Precedence	D	T	R	C	"0"

Precedence (Priority):	DTRC bits:			
111 Network Control	0 0 0 0	. .	. . . .	normal service
110 Internetwork Control	1 0 0 0	D	Delay	min. delay
101 Critic/ECP	0 1 0 0	T	Throughput	max. throughput
100 Flash Override	0 0 1 0	R	Reliability	max. reliability
011 Flash	0 0 0 1	C	Cost	min. cost
010 Immediate				
001 Priority	No other values are defined but have to be accepted (ignored) by a router or host.			
000 Routine				

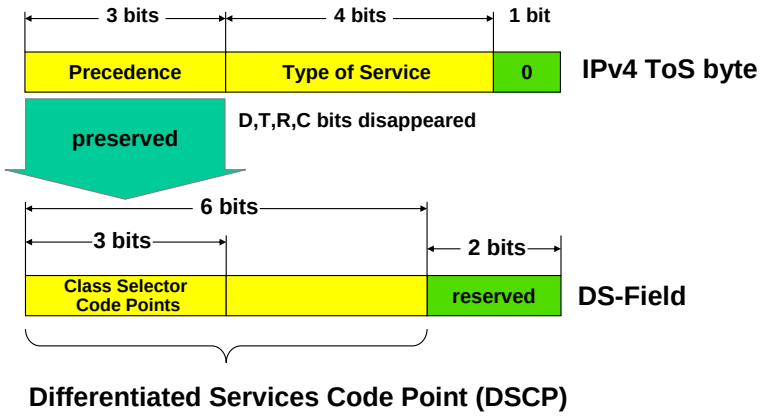
The Way to IP QoS (6)

- **Two models for IP QoS:**
  - Integrated Services Model
    - Flow based with RSVP (Resource ReserVation Protocol) and dynamic QoS like ATM QoS
    - Failed because of scalability
  - Differentiated Services Model
    - Based on differentiation of traffic classes and a QoS customer - QoS provider relationship with static traffic contract
    - Precedence idea of old TOS recycled !!!
    - It is the current technique to have something like QoS in the IP world
      - But still not comparable with ATM QoS !!!
    - TOS was redefined by the IETF to become the **"Differentiated Service Code Point (DSCP)"**

Remember IP is a best-effort service, therefore not suited for interactive real-time traffic like voice and video. ATM was designed for supporting QoS in the most perfect way. During the 1990s there was a battle between ATM and IP world. IP lost its simplicity by dealing now with QoS. Two flavours: Integrated services model and differentiated services model were borne by the IETF.

See RFC 2474: "Definition of the Differentiated Service Field in the IPv4 and IPv6 Headers" and RFC 2475: "An Architecture for Differentiated Services"

IPv4 TOS Recycling (RFC 2474)

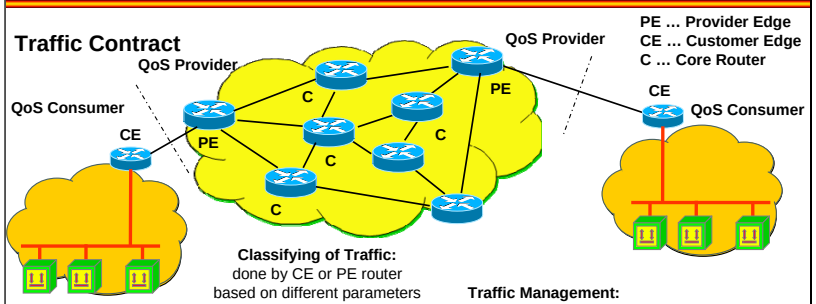


With 6 bits we can now differentiate 64 traffic classes.

The Way to IP QoS (7)

- **DSCP Usage:**
  - Is used to tag (label) the traffic class of a datagram
  - All labeled datagrams of a traffic class will receive a defined PHB (Per Hop behavior)
- **Typical scenario:**
  - QoS service provider <-> QoS customer
  - IP datagrams are classified and can be labeled (marked) at the border of IP QoS domain
  - Border has to perform traffic policing according to the static traffic contract
  - Customer may shape traffic to obey the traffic contract
  - Traffic class will receive their PHP handling within in IP QoS Domain
    - e.g. Limited delay, Guaranteed throughput

Differentiated Services Model: Elements



Traffic Shaping:
done by CE router

Classifying of Traffic:
done by CE or PE router
based on different parameters
(e.g. interface, IP, TCP header)

Marking of Traffic:
done by CE or PE router
by specifying DSCP
(service class)

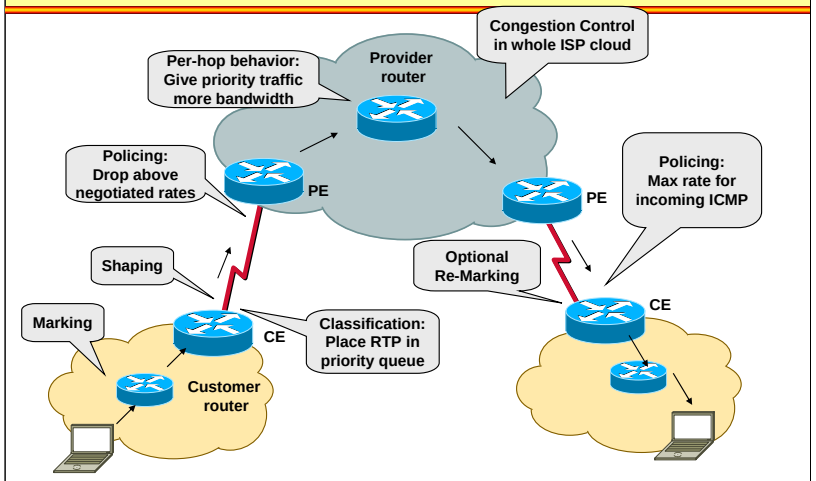
Traffic Policing:
done at PE router by CAR
(Committed Access Rate)

Traffic Management:
queuing per service class,
done by every core router C

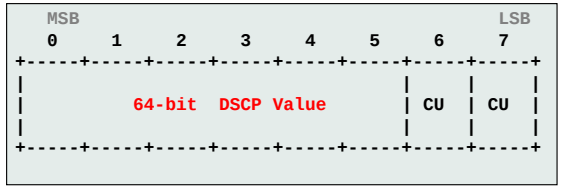
Call Admission Control:
done by provider by
provisioning network
resources for service classes

Signaling:
not necessary because
of static approach

Differentiated Services Model: In Action

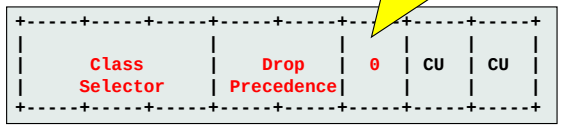


DSCP Details



Practically only 5 bits are used:

Only zero if bits 3,4 should be interpreted as drop precedence



DSCP Values Overview

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

14 Recommended Code Points

- **Expedited Forwarding (EF)**
  - DSCP 46 = 101 110 binary
  - For low delay, low loss, and low jitter
  - Defined in RFC 3246
- **Assured Forwarding (AF)**
  - 12 codepoints: 4 classes and 3 drop precedence each
  - Defined in RFC 2597
- **Best Effort (BE)**
  - 000000 binary
- **The legacy IP Precedence values (0-7) are preserved**
  - Can be directly mapped into the three Class Selector bits (0,1,2) *with the three LSBs (3,4,5) set to zero*
  - This results in the seven CSx values
    - CS0 = DSCP 00 = 000000
      ...
    - CS7 = DSCP 56 = 111000

Assured Forwarding (AF)

- **Guarantees a certain bandwidth to a traffic class**
  - If the traffic exceeds the committed bandwidth the drop probability is raised according to the specified drop precedence
- **There are 12 different AF behavior code points**
  - Consisting of 4 classes (AF1y to AF4y)
  - And 3 drop probabilities (AFx1 to AFx3) for each class (low/med/hi)

Drop:	Class 1			Class 2			Class 3			Class 4		
Low	AF11	10	001010	AF21	18	010010	AF31	26	011010	AF41	34	100010
Medium	AF12	12	001100	AF22	20	010100	AF32	28	011100	AF42	36	100100
High	AF13	14	001110	AF23	22	010110	AF33	30	011110	AF43	38	100110

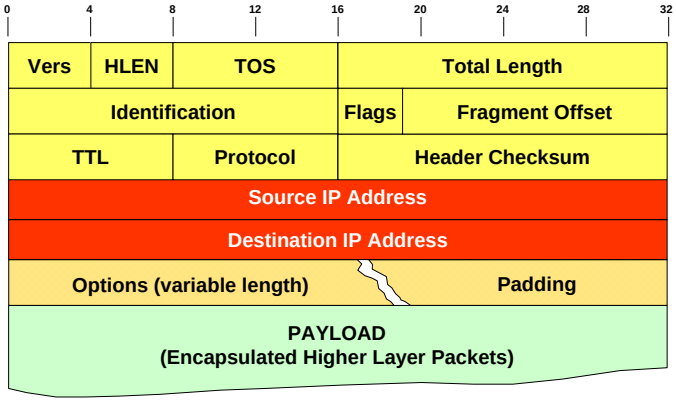
decimal | binary

AF has been defined in RFC 2597.

Agenda

- **Introduction**
  - Short History of the Internet (not part of the exam!)
  - Basic Principles
- **IP**
  - IP Protocol
  - Addressing
  - Classful versus Classless (not part of the exam!)
- **IP Forwarding**
  - Principles
  - ARP
  - ICMP
  - PPP
- **First Hop Redundancy**
  - Proxy ARP, IDRP
  - HSRP
  - VRRP (not part of the exam!)

IP Header - The IP Addresses



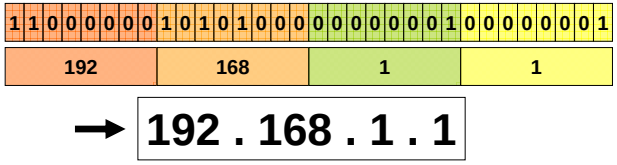
The IP Address

- **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*:



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 Address Classes

- **Net-ID? Host-ID?**
- **Originally 5 Classes defined!**
  - A (1-127)
  - B (128-191)
  - C (192-223)
  - D (224-239, Multicast)
  - E (240-254, Experimental)
  - "First octet rule"
- **Classes define number of address-bits for Net-ID**

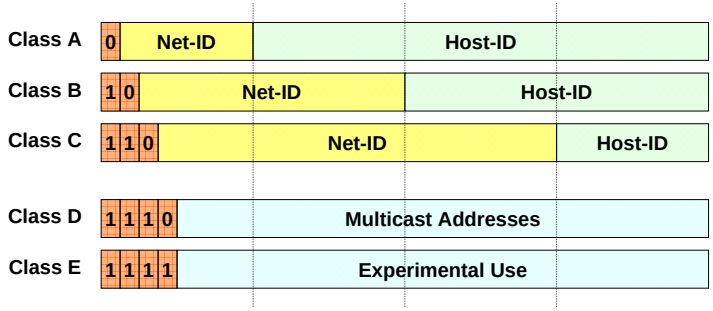
In the beginning of the Internet, five address classes had been defined. Classes A, B, and C had been created to provide different network addresses ranges. Additionally Class D is the range of IP multicast addresses, that is they have no topological structure. Finally, class E had been reserved for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

The 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. The usage of classes has a long tradition in the Internet and was a main reason for IP address depletion.

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.

IP Address Classes

Classes are defined by *„first octet rule"*



The first part of the address identifies the Network on which the host resides, the second part identifies the host on the given network.

Class A: 7 Bits Net-ID, 24 Bits Host-ID - 126 Nets and 16.777.214 Hosts

Class B: 14 Bits Net-ID, 16 Bits Host-ID - 16.384 Nets and 65.534 Hosts

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

Special Addresses

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

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.

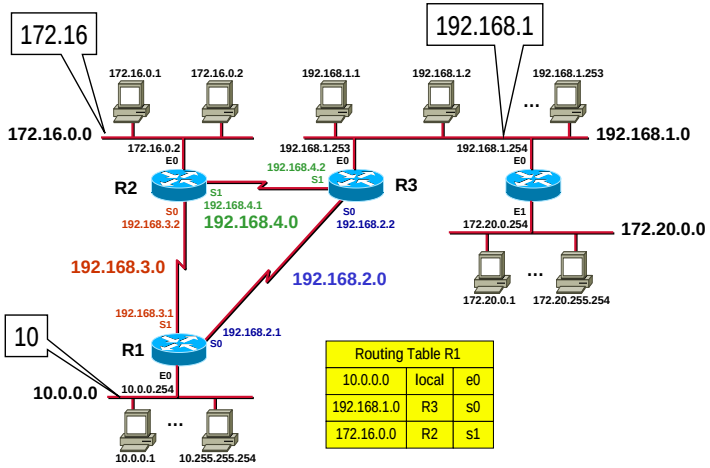
Private Addresses / NAT

- **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
  - Either static 1:1 mapping
  - Or dynamic n:1 mapping with port address translation
    - 1 official IP address may be shared by many internal private stations

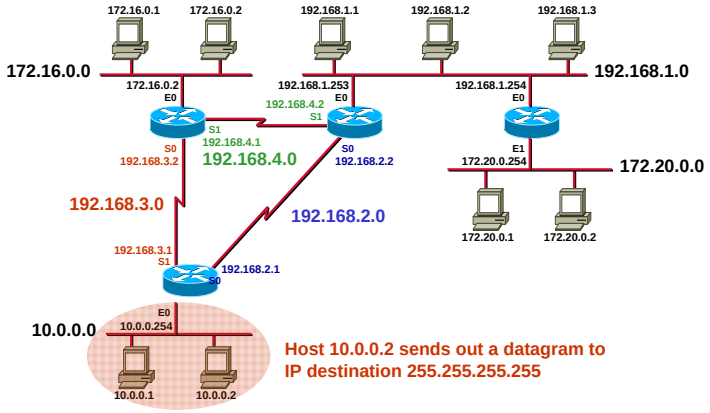
So-called RFC 1918 addresses are class A, B, and C address blocks which can be used for internal purposes. Such addresses must not be used in the Internet. All gateways connected to the Internet should filter packets that contain these private addresses. Furthermore these addresses must not be used in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

Addressing Example (Net-ID)

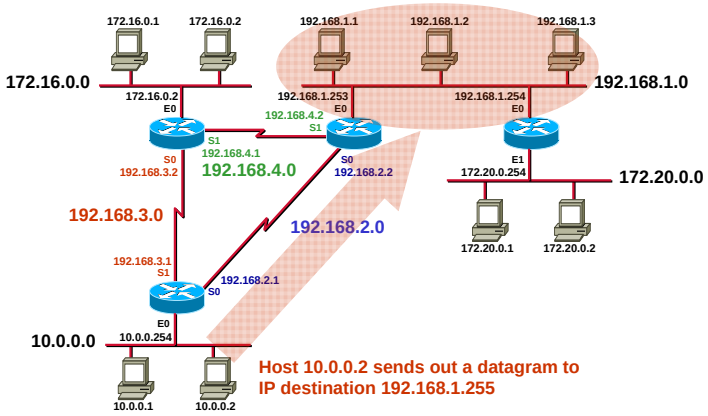


IP Limited Broadcast



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

IP Directed Broadcast



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!

Subnetting

- **Two level hierarchy of classful addressing was sufficient in the early days of the Internet**
  - Later that lead to waste of the address space especially with the appearance of LANs in organizations
- **Subnetting**
  - Allows a additional (third) level of hierarchy
  - Some bits of the Host-ID can be used as Subnet-ID
  - Subnet-ID extends the classful Net-ID meaning
    - Subnet-ID bits are only locally interpreted inside the subnetted area
    - Net-ID bits are still globally seen outside the subnetted area

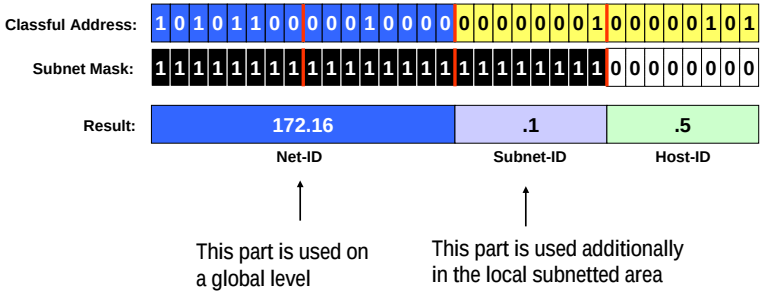
The "**classful**" method of identifying network-IDs based on the given IP address class is inflexible and lead to address space depletion. Class C networks are too small for most organizations but class A and B are too large. A waste of the IP address space happened by giving class B or class A address 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 introduced by classful addressing was exponential growing of the Internet routing tables by giving multiple class C addresses to customers in order to support their addressing needs.

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.

Subnetting Example

Alternative (newer) notation: 172.16.1.5 /24

Class B Address: 172.16.1.5, Subnet Mask: 255.255.255.0



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

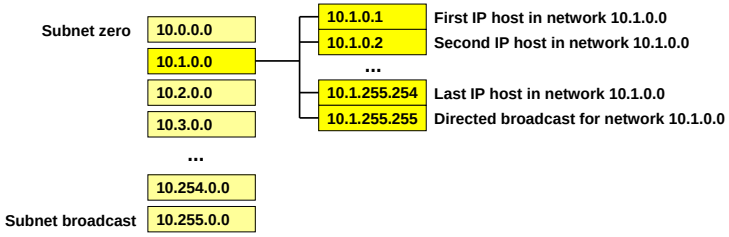
2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0	
1	0	0	0	0	0	0	0	128
1	1	0	0	0	0	0	0	192
1	1	1	0	0	0	0	0	224
1	1	1	1	0	0	0	0	240
1	1	1	1	1	0	0	0	248
1	1	1	1	1	1	0	0	252
1	1	1	1	1	1	1	0	254
1	1	1	1	1	1	1	1	255

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 Zero / Subnet Broadcast

- **Consider network 10.0.0.0**
  - Is it a class A net "10" ?
  - Or do we have a subnet "10.0" ?
- **Consider broadcast 10.255.255.255**
  - Is it a directed broadcast for the whole net 10 ?
  - Or only for the subnet 10.255 ?
- **Subnet zero and subnet broadcast can be ambiguous!**

The older routing protocols, such as RIPv1 or IGRP, specifies routes in routing updates as a single 32-bit address with no information about subnet mask. The class of an address defines what is NET-ID and Host-ID. A simple convention was then followed. If the host field contained all 0 bits, then the address was a network route that matched every address within that classful network, the equivalent of a /8, /16, or /24 prefix, depending on the address class. Any 1 bits in the host field caused it to be interpreted as a host route, matching only the exact address specified, the equivalent of /32 prefix. This is why the all-zeros address is reserved - it was used by the routing protocols to match the entire classful network.

With the advent of subnetting this schema was undermined , but the designers of subnetting decided against any changes to the format of the routing protocols. This meant that there was still only a single 32-bit address to work with, though its interpretation became much more complex. Addresses in foreign networks (classful networks not directly attached to the router processing the information) were interpreted as before. Addresses in local networks were processed using the subnet mask programmed into the router. The address was first split into its three fields. If both subnet and host fields were all 0s, it was a network route, as before. An address with 1 bits in the subnet field, but all 0 bits in the host field was a subnet route, matching all addresses within that subnet. Finally, addresses with 1 bits in the host field were interpreted as host routes, as before. This lead to more reserved addresses - both the all-0s subnet and the all-0s host in each subnet were reserved.

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

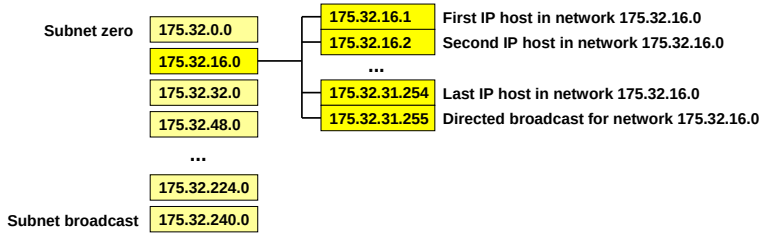
  address binary 1010 1100 . 0001 0000 . 0000 0011 . 1001 0000
  mask (binary) 1111 1111 . 1111 1111 . 1111 1111 . 1100 0000
  ---
  logical AND (bit by bit)
  net-id 1010 1100 . 0001 0000 . 0000 0011 . 1000 0000

  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"

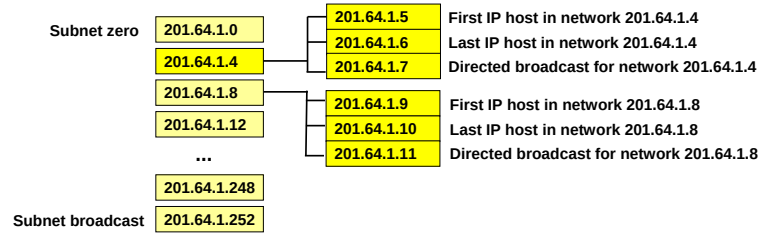
1) That is: 175.32.0.0 with 255.255.240.0 or 175.32.0.0/20
2) Resulting subnetworks:



Subnet Example 3

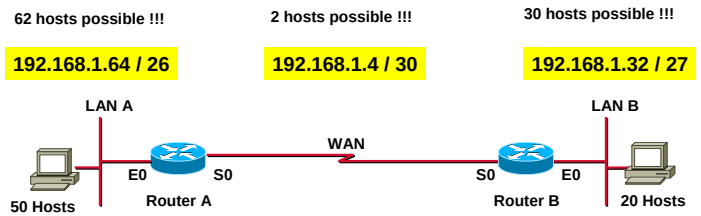
"Use the class C network 201.64.1.0 and 6 bit subnetting"

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"
  - 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.

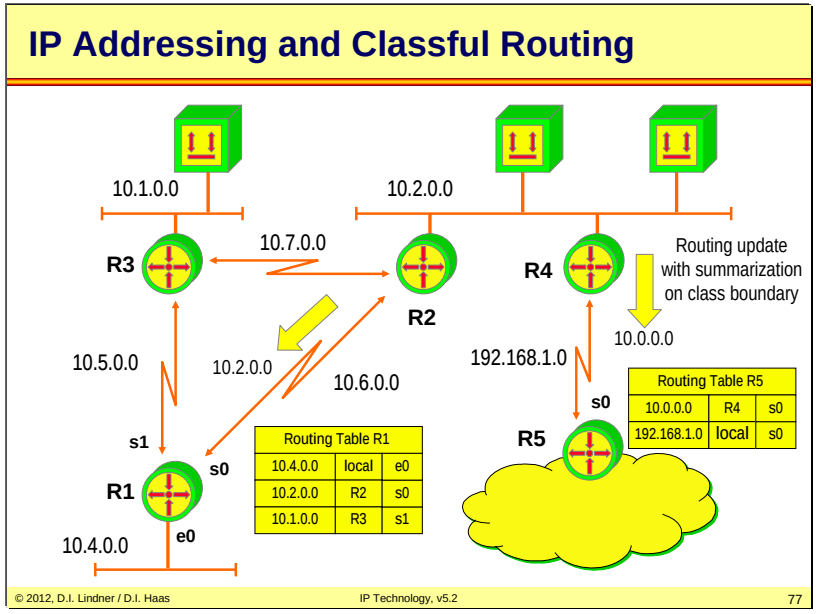
A short address design history:

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

If you have to understand IP addressing issues from the scratch please study the next chapter about "Classful versus Classless" issues. This chapter is not part of the exam !!!

Agenda

- **Introduction**
  - Short History of the Internet (not part of the exam!)
  - Basic Principles
- **IP**
  - IP Protocol
  - Addressing
  - Classful versus Classless (not part of the exam!)
- **IP Forwarding**
  - Principles
  - ARP
  - ICMP
  - PPP
- **First Hop Redundancy**
  - Proxy ARP, IDRP
  - HSRP
  - VRRP (not part of the exam!)



Routing protocols like RIPv2, IGRP can not carry subnetmask information in routing updates. This has several consequences.

1. If a given class A, B or C address is subnetted the subnetmask must be constant in the whole subnetted area (no variable length subnet mask (VLSM) can be used).

2. If a routing update is sent to an interface with an network number different to the subnetted network only the major class A, B or C network number will be announced. So called route summarization will be performed on class boundaries hence a subnetted area must be contiguous.

This behavior is called classful routing.

The routing table lookup in classful networks is done is such a way (assumption: an IP datagram with a given IP address is received by a classful router):

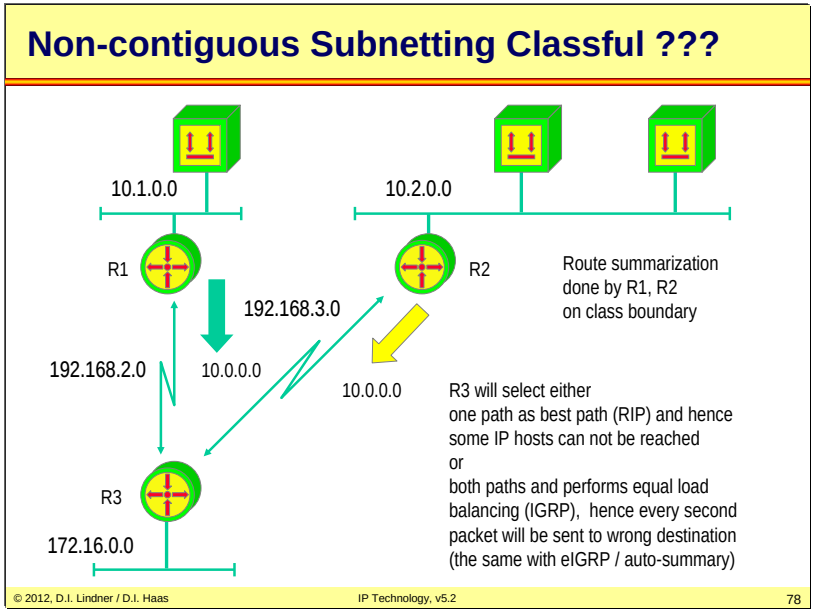
1. IP address is interpreted as class A, B or C and the major net is determined

2. Next the lookup for the major net in the routing table is performed. If there is no entry the IP datagram will be discarded.

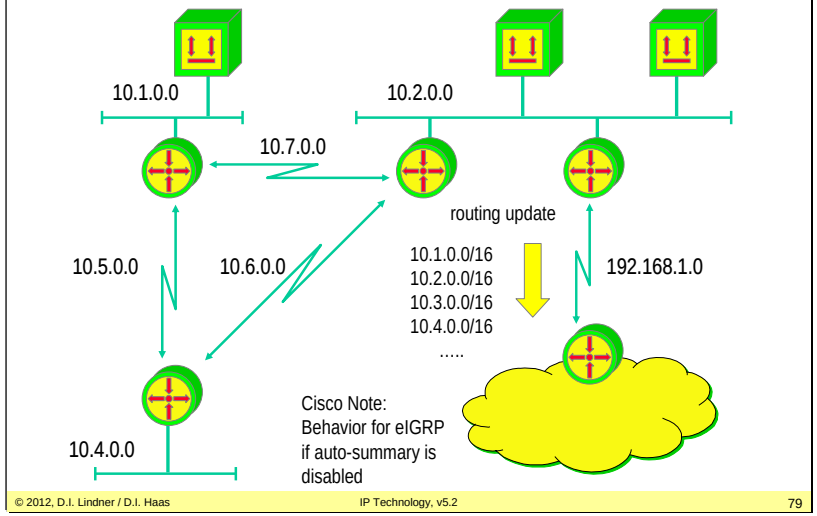
3. If there is a match the IP address is compared to every known subnet of this major network. If there is no such subnet the IP datagram will be discarded.

Hence a problem may arise with default routing: If the major network is known by the router, but the subnet does not exist, the IP datagram will be discarded even if a default route exists. Therefore subnetted area must be contiguous –> all subnets of a given major net must be reachable using only paths (networks) with these subnet-IDs.

Remark: Cisco's configuration command ***ip classless*** will change such an behavior in case of default routing to the behavior of classless routing even if classful routing is used.



Classless Routing



Routing protocols like RIPv2, OSPF, eIGRP can carry subnet mask information in routing updates. This has several advantages:

1. Variable length subnet mask (VLSM) can be used and subnetting of a given address can be done according to the number of hosts required on a certain subnet. More efficient use of address space.

2. Route summarization can be performed on any address boundary and not only on class boundaries. A routing update contains prefix (relevant part of IP address) and length (number of ones used in subnetmask) and allows supernetting (actual subnetmask is smaller than natural subnetmask of given class).

This behavior is called classless routing.

The routing table lookup in classless networks is done is such a way (assumption: an IP datagram with a given IP address is received by a classless router):

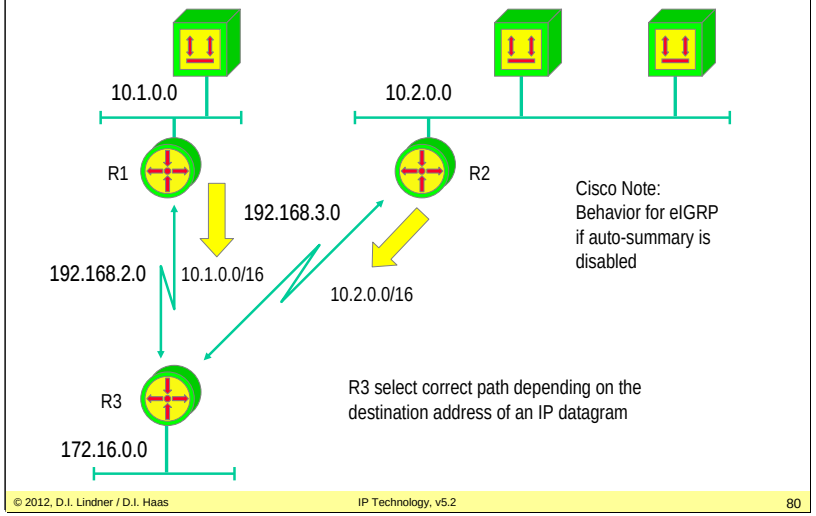
1. IP address is not interpreted as class A, B or C

2. A lookup in the routing table for the best match for this IP address is performed. IP prefixes of the routing table are compared with the given IP address bit by bit from left to right.

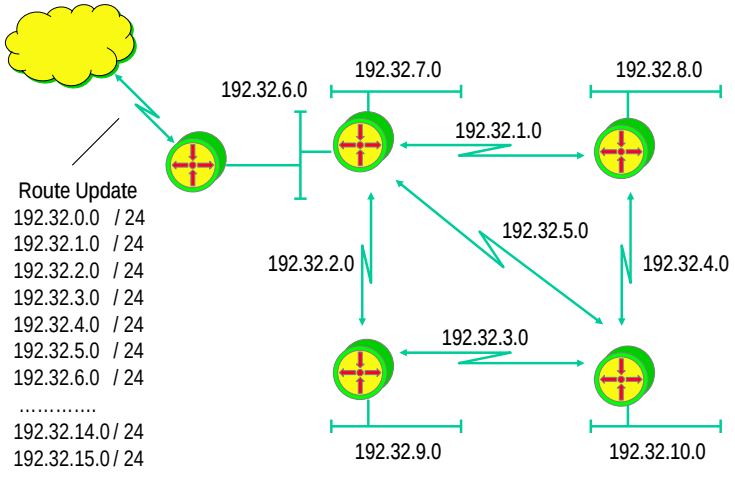
3. IP datagram is passed on to the network which matches best -> “Longest Match Routing Rule”

Result: IP addresses with any kind of subnetting can independently be used from the underlying network topology without any constrains concerning non-contiguous of subnetted area.

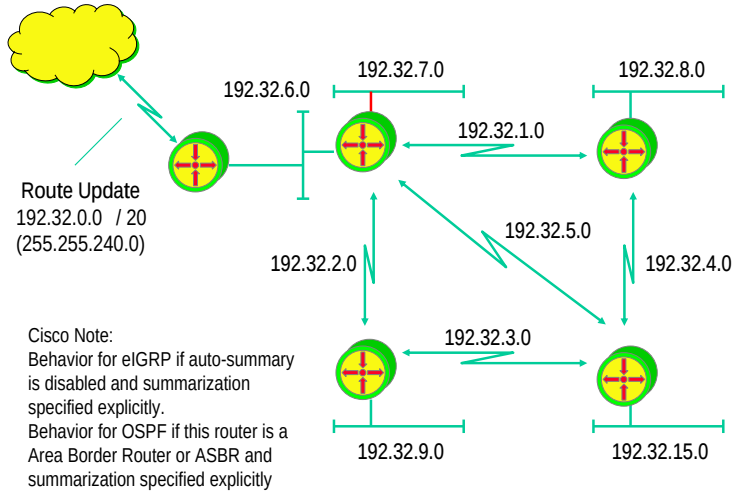
Non-contiguous Subnetting Classless !!!



Routing Updates without Supernetting



Route Summarization with Supernetting



VLSM Example (1)

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

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

VLSM Example (2)

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

VLSM Example (3)

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

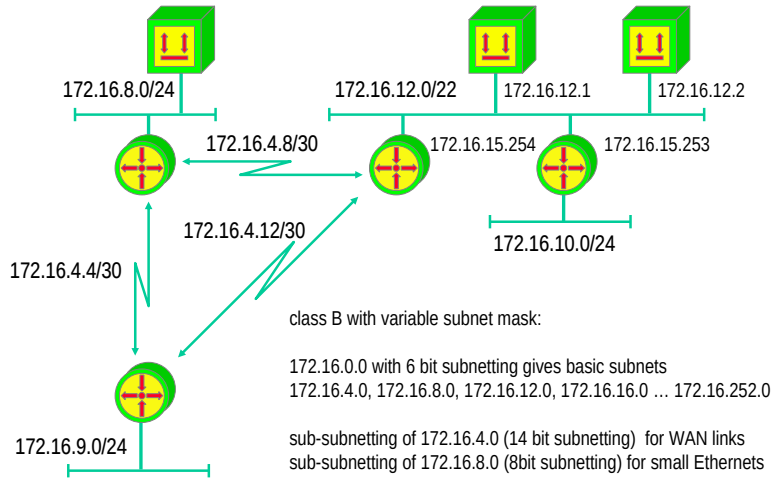
      - 172.16.9.254 last IP host of subnet 172.16.9.0
      - 172.16.9.255 directed broadcast of subnet 172.16.9.0
    - 172.16.10.0 / 24
    - 172.16.11.0 / 24
  - Sub-subnetworks capable of addressing 254 IP systems

VLSM Example (4)

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

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

Example VLSM



IP Address Space Depletion

- **The growing demand of IP addresses**
  - Has put a strain on the classful model
  - Class B exhaustion
  - Class C are to small for most organization
  - Many class C addresses given to a certain organization leads to explosion of routing table entries in the Internet core routers
- **Measures to handle these problems**
  - Creative IP address allocation
  - CIDR
  - Private IP addresses and network address translation (NAT)
  - IPv6

CIDR

- **Classless Interdomain Routing (CIDR)**
  - Address assignment and aggregation (route summarization) strategy
  - Temporary solution to overcome depletion of IP address space and explosion of routing tables in the Internet core routers
- **Basic ideas**
  - Classless routing (prefix, length)
  - Supernetting
  - Coordinated address allocation
    - until 1992 IP addresses had no relation at all to the networks topology

CIDR

- **CIDR address allocation**
  - Addressing plan for class C addresses by continents
    - 192.0.0.0 - 193.255.255.255 … Multiregional
    - 194.0.0.0 - 195.255.255.255 … Europe
    - 198.0.0.0 - 199.255.255.255 … North America
    - 200.0.0.0 - 201.255.255.255 … Central/South America
  - Provider addressing strategy
    - Internet Service Providers (ISP) are given contiguous blocks of class C addresses which in turn are granted to their customers
    - Consequence: change of provider means renumbering
  - Class C network numbers are allocated in such a way that route summarization (or sometimes called route aggregation) into supernets is possible

CIDR

- **Definitions of terms often used interchangeably**
  - CIDR block
    - is the <prefix, length> notation
  - Supernets
    - have a prefix length shorter than the networks natural mask
  - Aggregates
    - indicate any summary route
- **In order to implement CIDR**
  - Classless routing protocols between routing domains must be used
    - BGP-4 as interdomain routing protocol
  - Classless routing within a routing domain
    - RIPv2, OSPF, eIGRP

Private Address Range - RFC 1918

- **Three blocks of address ranges are reserved for addressing of private networks**
  - 10.0.0.0 - 10.255.255.255 (10/8 prefix)
  - 172.16.0.0 - 172.31.255.255 (172.16/12 prefix)
  - 192.168.0.0 - 192.168.255.255 (192.168/16 prefix)
  - Note:
    - In pre-CIDR notation the first block is nothing but a single class A network number, while the second block is a set of 16 contiguous class B network numbers, and third block is a set of 256 contiguous class C network numbers.
- **Translation between private addresses and globally unique addresses -> NAT**

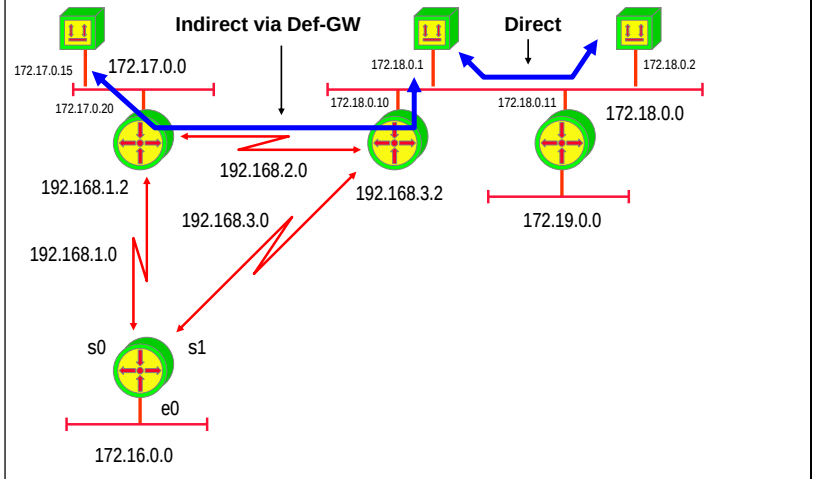
Agenda

- **Introduction**
  - Short History of the Internet (not part of the exam!)
  - Basic Principles
- **IP**
  - IP Protocol
  - Addressing
  - Classful versus Classless (not part of the exam!)
- **IP Forwarding**
  - Principles
  - ARP
  - ICMP
  - PPP
- **First Hop Redundancy**
  - Proxy ARP, IDRP
  - HSRP
  - VRRP (not part of the exam!)

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 choose a “default” router aka “Default Gateway”**
  - As next hop in case of indirect delivery of IP datagrams

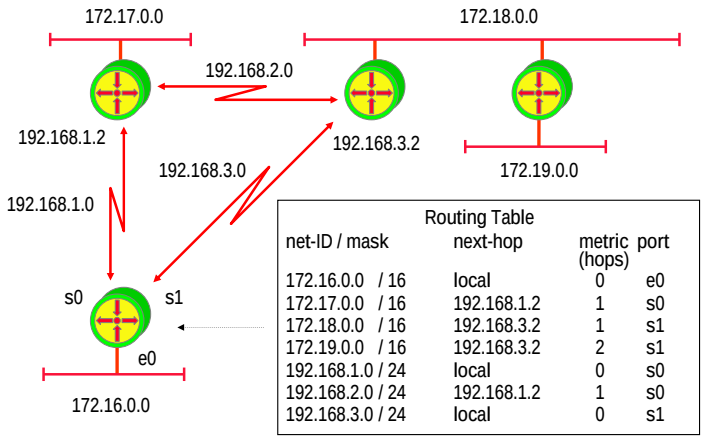
Direct versus Indirect Delivery



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)

Routing Table Example



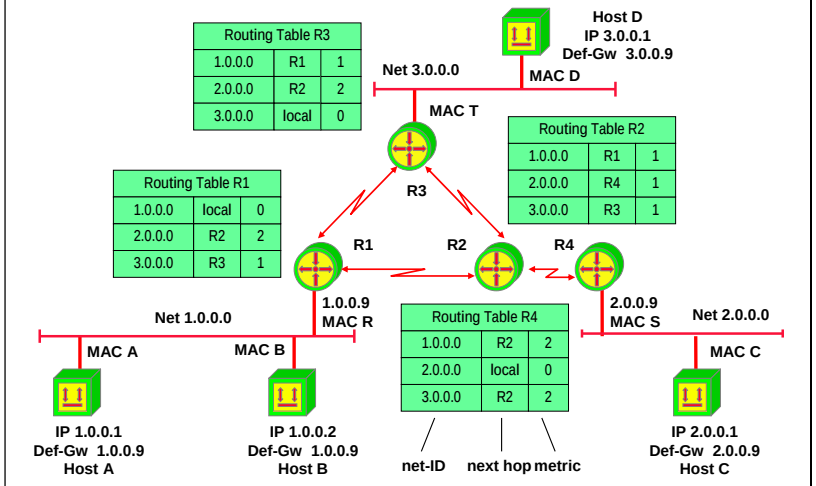
Routing Table

net-ID / mask	next-hop	metric (hops)	port
172.16.0.0 / 16	local	0	e0
172.17.0.0 / 16	192.168.1.2	1	s0
172.18.0.0 / 16	192.168.3.2	1	s1
172.19.0.0 / 16	192.168.3.2	2	s1
192.168.1.0 / 24	local	0	s0
192.168.2.0 / 24	192.168.1.2	1	s0
192.168.3.0 / 24	local	0	s1

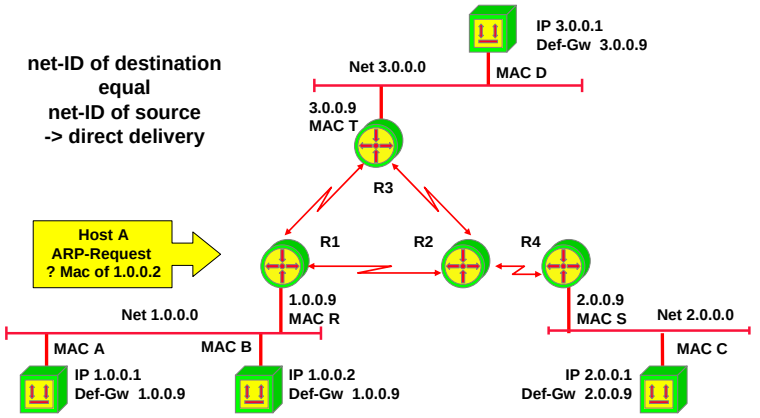
IP Routing Paradigm

- **Destination Based Routing**
  - Source address is not taken into account for the forward decision
- **Hop by Hop Routing**
  - IP datagrams follow the path, which is pointed by the current state of the routing tables
- **Least Cost Routing**
  - Normally only the best path is considered for forwarding of IP datagrams
  - Alternate paths will not be used in order to reach a given destination

Example Topology

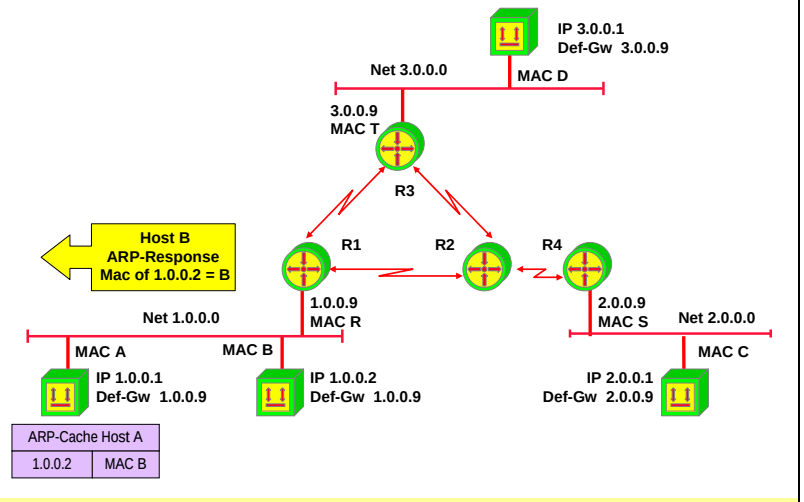


Direct Delivery 1.0.0.1 - > 1.0.0.2

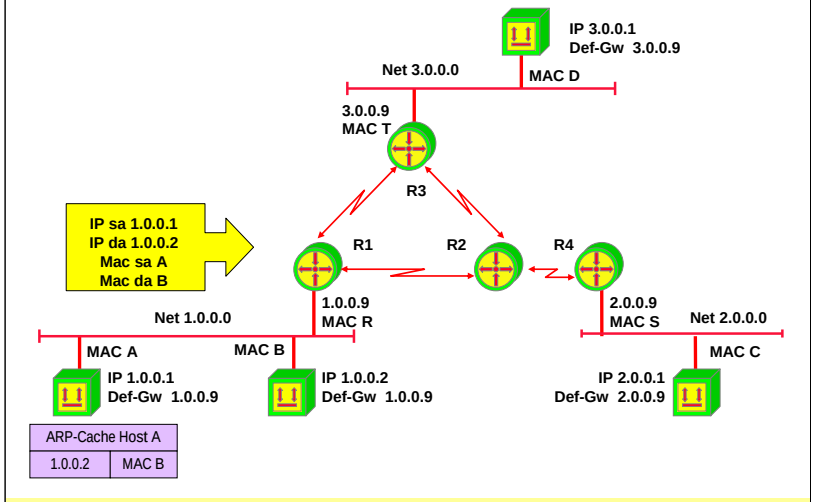


ARP ... Address Resolution Protocol

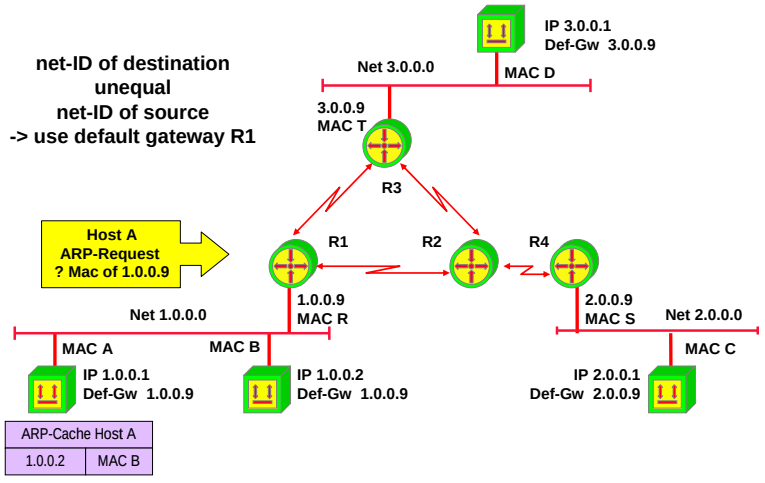
Direct Delivery 1.0.0.1 - > 1.0.0.2



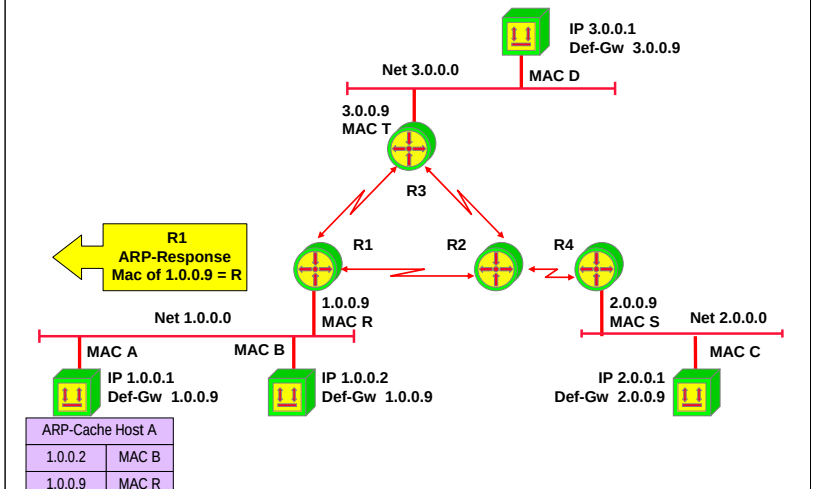
Direct Delivery 1.0.0.1 - > 1.0.0.2



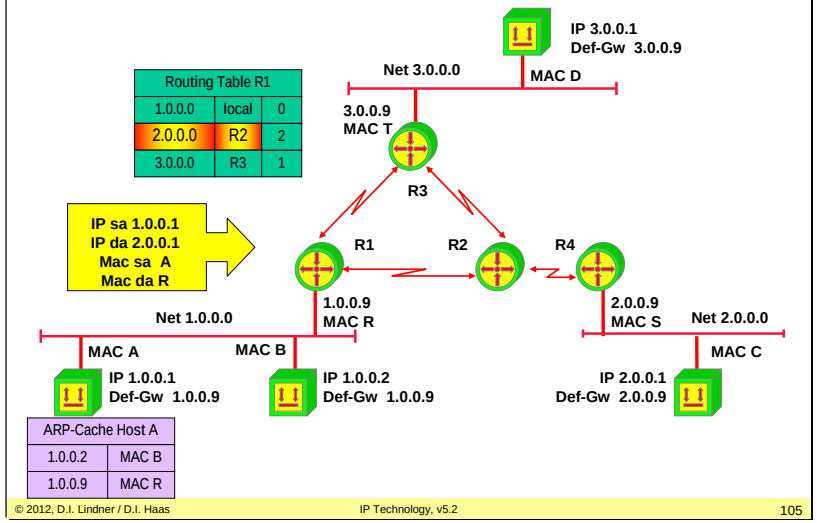
Indirect Delivery 1.0.0.1 - > 2.0.0.1



Indirect Delivery 1.0.0.1 - > 2.0.0.1

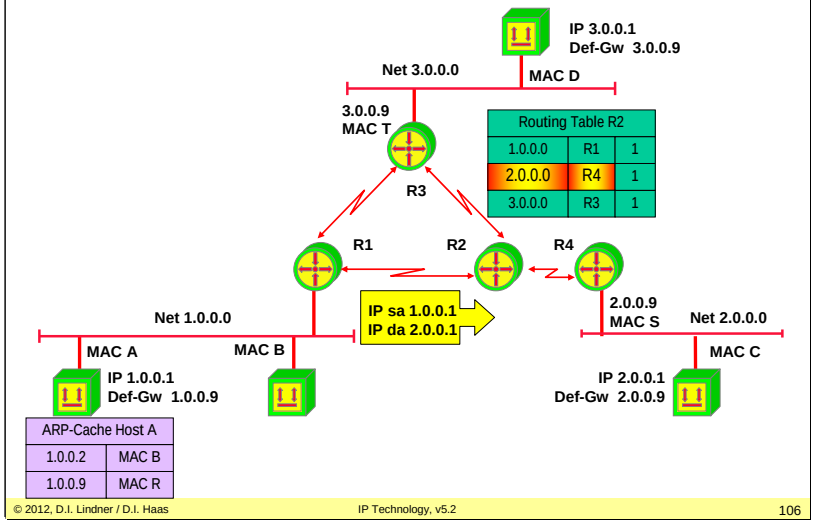


Indirect Delivery 1.0.0.1 - > 2.0.0.1



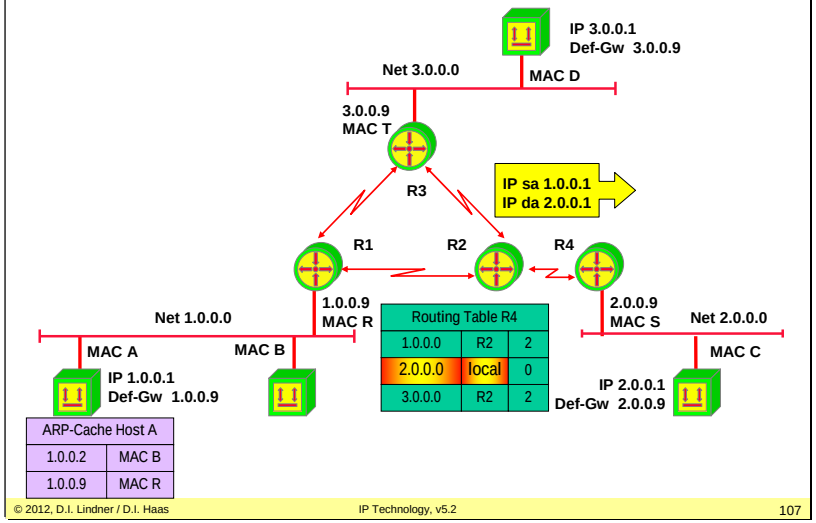
 IP Technology, v5.2 105

Indirect Delivery 1.0.0.1 - > 2.0.0.1



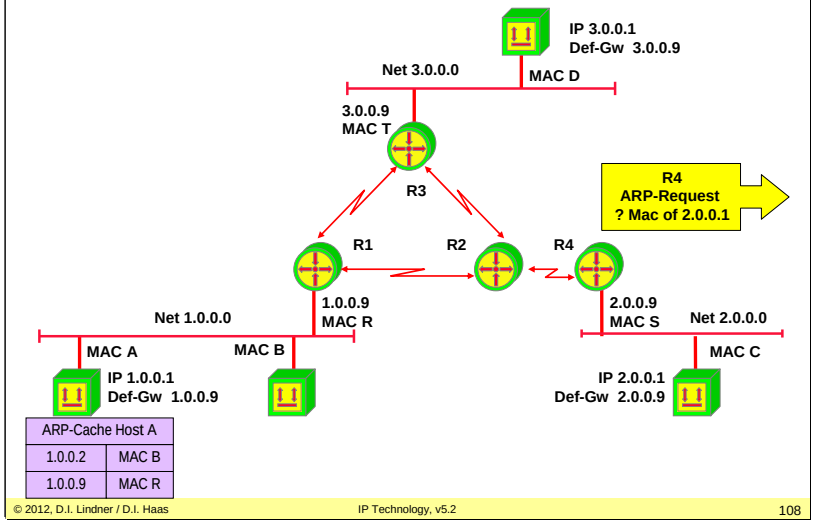
 IP Technology, v5.2 106

Indirect Delivery 1.0.0.1 - > 2.0.0.1



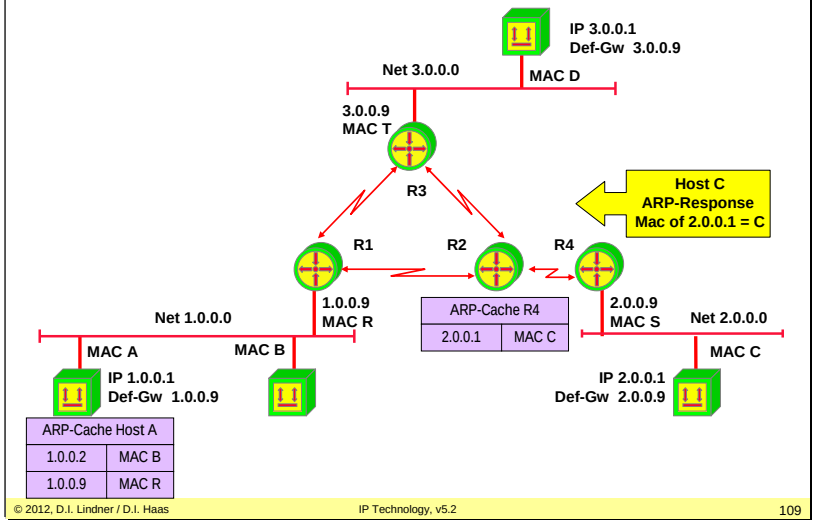
 IP Technology, v5.2 107

Indirect Delivery 1.0.0.1 - > 2.0.0.1

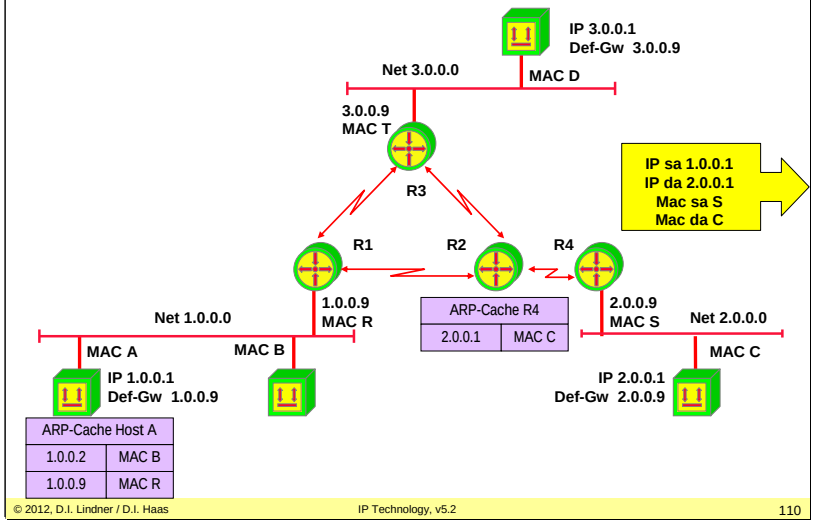


 IP Technology, v5.2 108

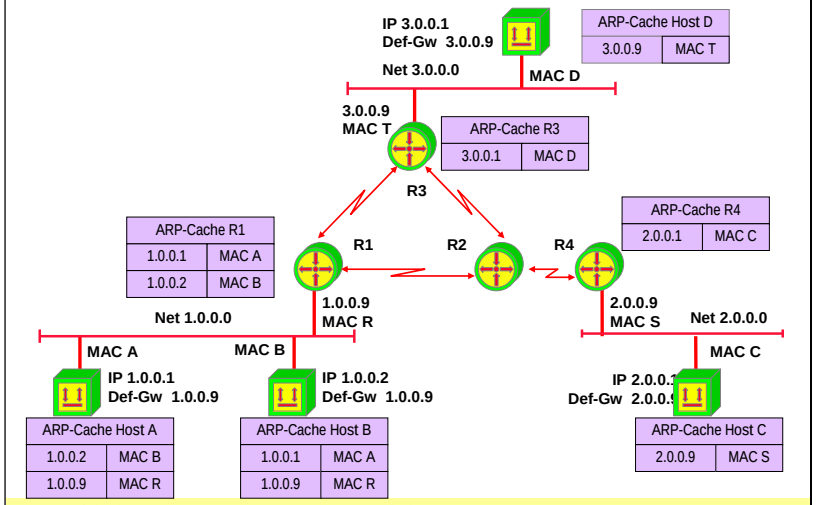
Indirect Delivery 1.0.0.1 - > 2.0.0.1



Indirect Delivery 1.0.0.1 - > 2.0.0.1



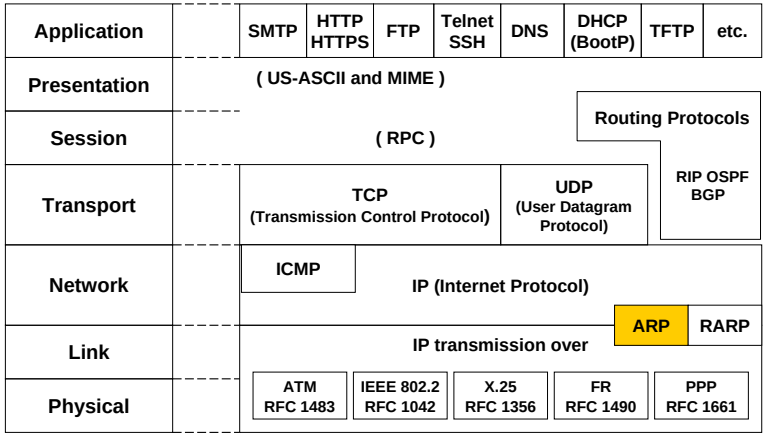
ARP Cache - Final Picture



Agenda

- **Introduction**
  - Short History of the Internet (not part of the exam!)
  - Basic Principles
- **IP**
  - IP Protocol
  - Addressing
  - Classful versus Classless (not part of the exam!)
- **IP Forwarding**
  - Principles
  - ARP
  - ICMP
  - PPP
- **First Hop Redundancy**
  - Proxy ARP, IDRP
  - HSRP
  - VRRP (not part of the exam!)

TCP/IP Protocol Suite



IP Address versus L2 Address

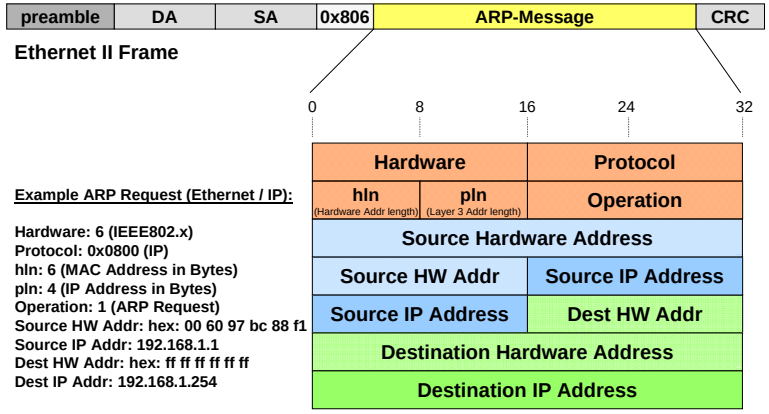
- **IP address**
  - Identifies the access to a network (interface)
- **If the physical network is of point-to-point link to another IP system**
  - This IP system can be reached without any further addressing on layer 2
- **On a shared media or multipoint-network**
  - Layer 2 addresses are necessary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

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 Format



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

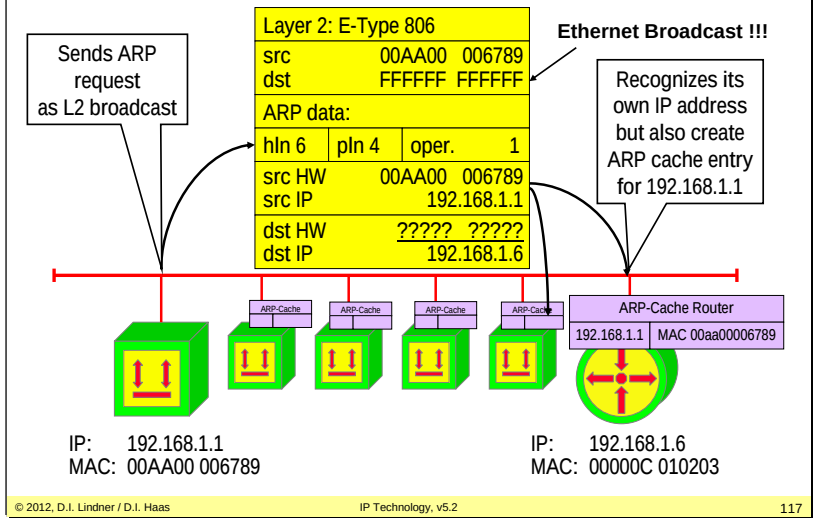
pln: Length of layer 3 address in bytes

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)

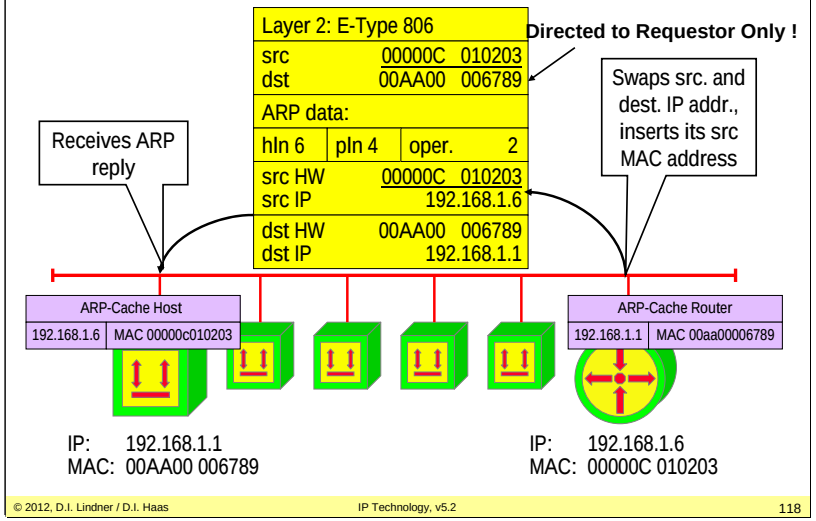
ARP Request



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.

ARP Reply

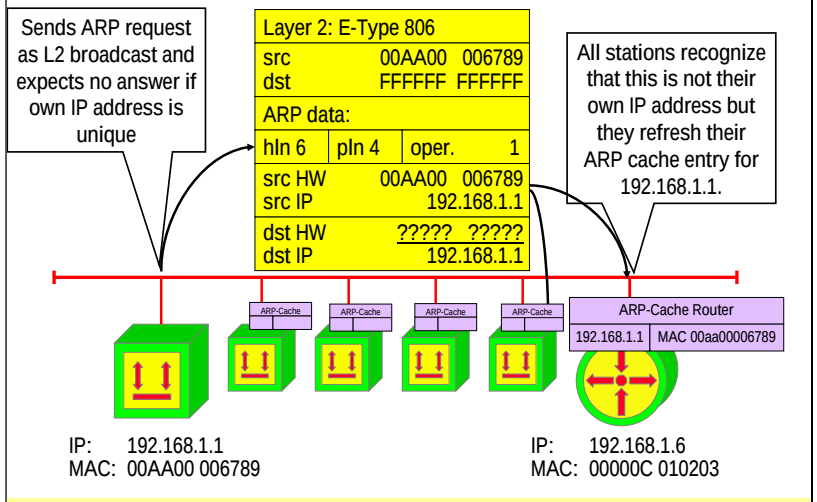


Now station B sees sends an ARP response as a directed MAC frame (SA=Mac_B, DA=Mac_A).The ARP response holds MAC address of station B. A stores the MAC- / IP-address mapping for station B in its ARP cache.

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 for Duplicate Address Check and ARP Cache Refresh



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

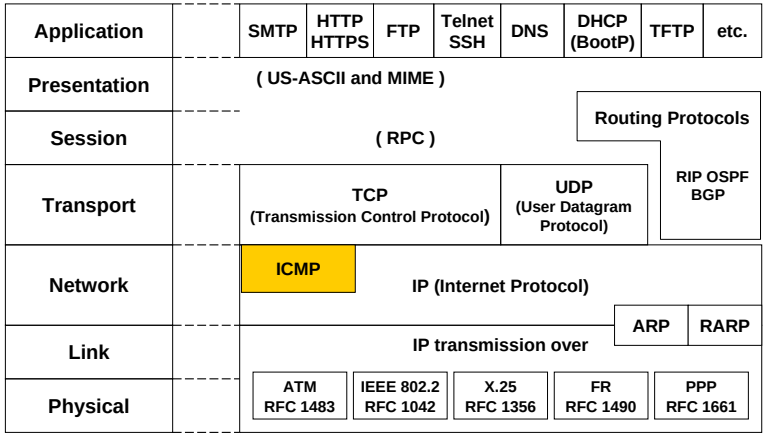
This is typically used:

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).

Agenda

- **Introduction**
  - Short History of the Internet (not part of the exam!)
  - Basic Principles
- **IP**
  - IP Protocol
  - Addressing
  - Classful versus Classless (not part of the exam!)
- **IP Forwarding**
  - Principles
  - ARP
  - ICMP
  - PPP
- **First Hop Redundancy**
  - Proxy ARP, IDRP
  - HSRP
  - VRRP (not part of the exam!)

TCP/IP Protocol Suite



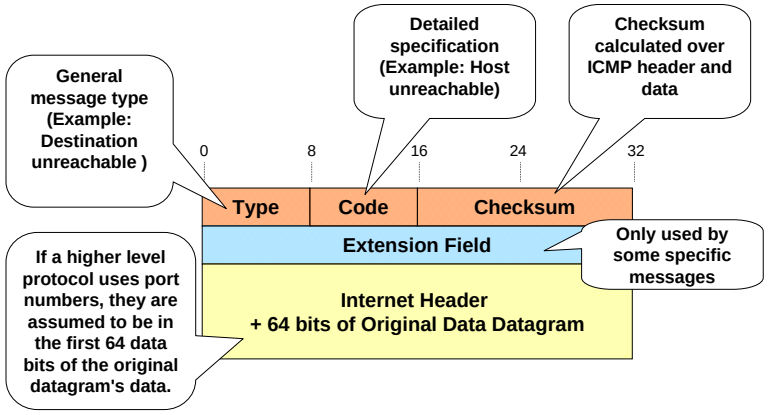
ICMP (RFC 792)

- **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 Format



Type Field

0	Echo reply ("Ping")
3	Destination Unreachable
	Reason specified in Code
4	Source Quench (decrease data rate of sender)
	Theoretical Flow Control Possibility of IP
5	Redirect (use different router)
	More information in Code
8	Echo Request ("PING")
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
	(finding the Net-ID of the network; e.g. SLIP)
17/18	Address Mask Request / Reply

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

Code Field for Type 3 (destination unreachable)

- 0 ... Network unreachable: no path to network known or network down; generated by intermediate or far-end router
- 1 ... Host unreachable: Host-ID can't be resolved or host not responding; generated by far-end router
- 2 ... Protocol unreachable: protocol specified in IP header not available; generated by end system
- 3 ... Port unreachable: port (service) specified in layer 4 not available; generated by end system
- 4 ... Fragmentation needed and do not fragment bit set: DF bit =1 but the packet is too big for the network (MTU); generated by router
- 5 ... Source route failed: Path in IP Options couldn't be followed; generated by intermediate or far-end router

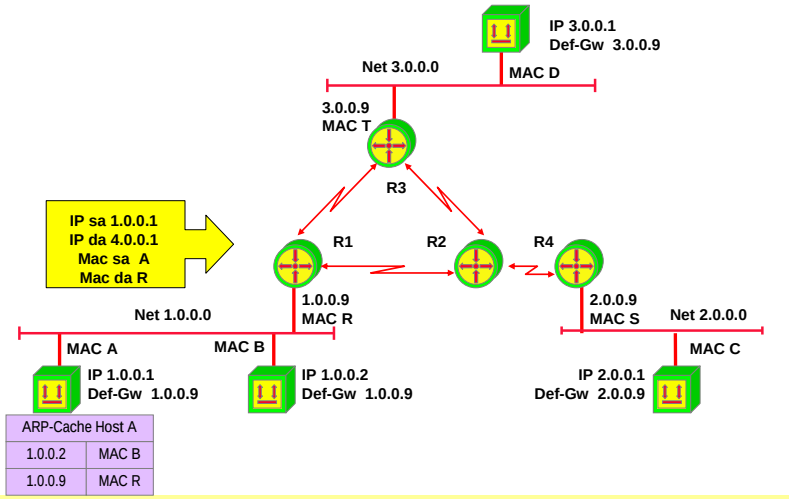
Code Field for Type 3 (destination unreachable)

The following additional codes are defined in RFC1122 (Host Requirements) page 38:

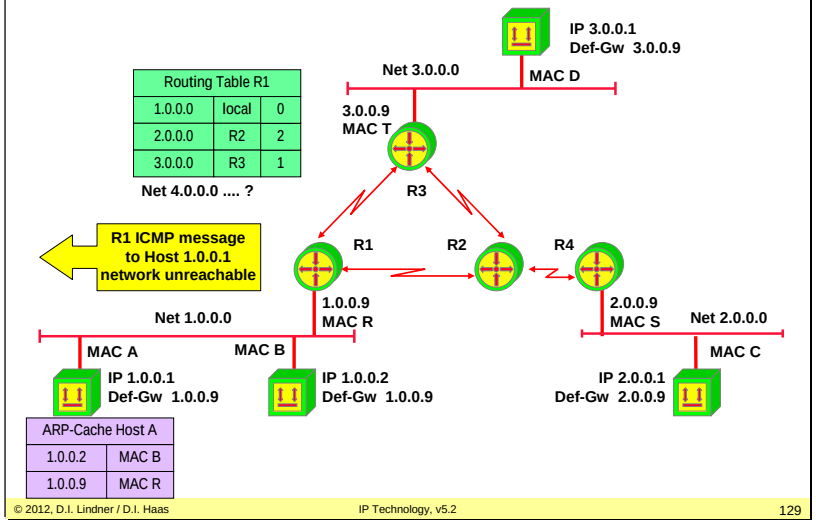
6 ... Destination network unknown
7 ... Destination host unknown
8 ... Source host isolated
9 ... Communication with destination network administratively prohibited
10 ... Communication with destination host administratively prohibited
11 ... Network unreachable for type of service
12 ... Host unreachable for type of service

Nowadays most of those messages are blocked by host firewalls (e.g. Microsoft Windows7 firewall) in order not to give to much information to an attacker.

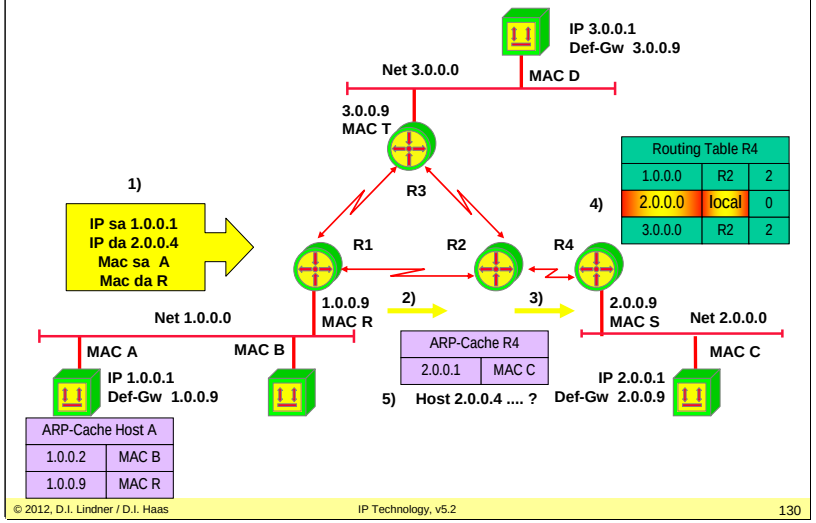
Delivery 1.0.0.1 - > 4.0.0.1



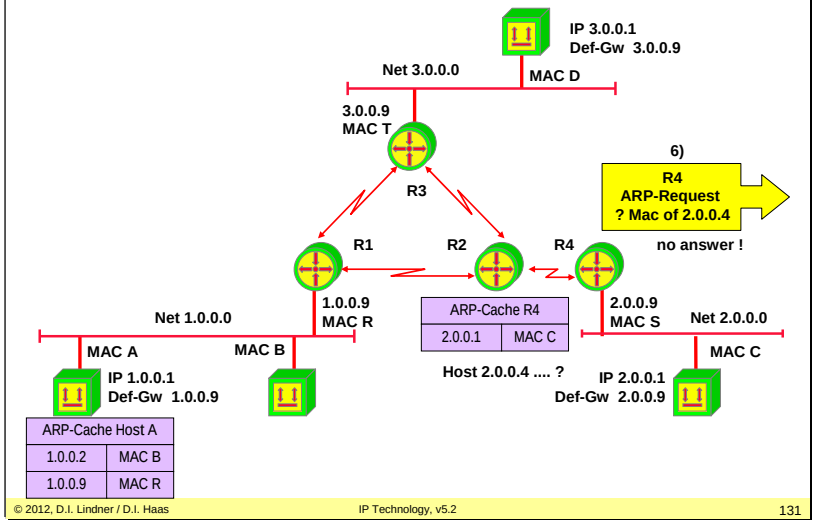
ICMP network unreachable



Delivery 1.0.0.1 - > 2.0.0.4

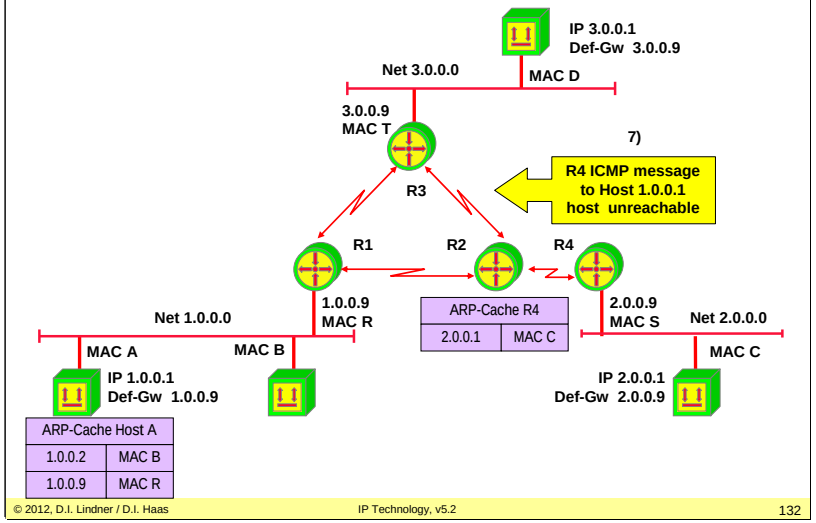


Delivery 1.0.0.1 - > 2.0.0.4



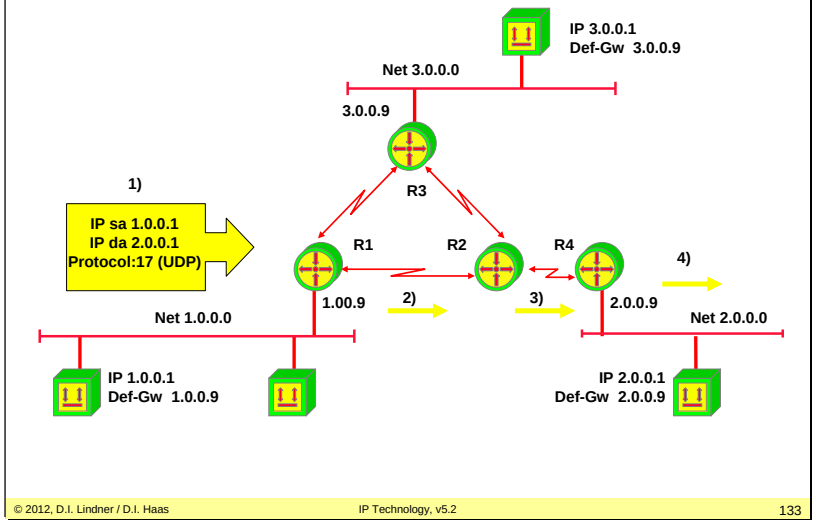
 IP Technology, v5.2 131

ICMP host unreachable

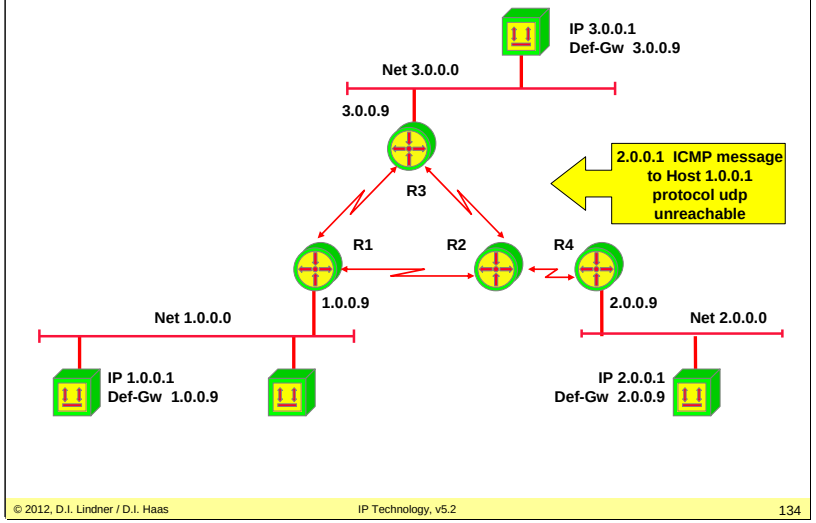


 IP Technology, v5.2 132

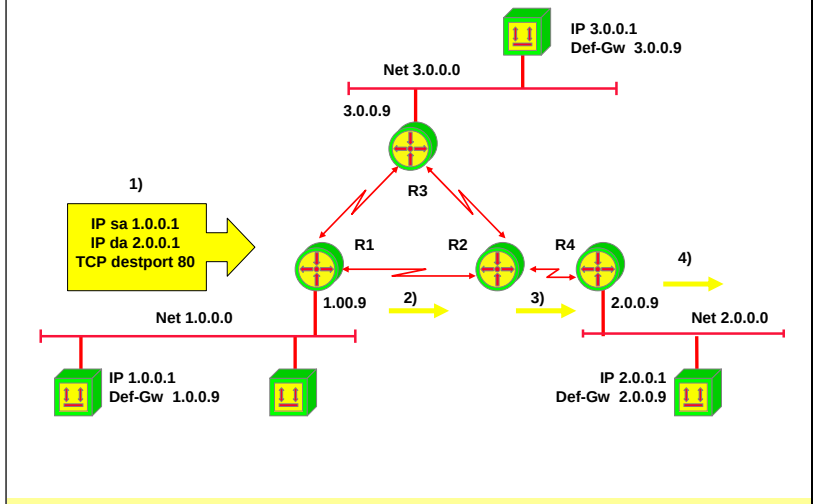
Delivery 1.0.0.1 - > 2.0.0.1 (protocol udp)



ICMP protocol unreachable



Delivery 1.0.0.1 - > 2.0.0.1 (http_server_proc)



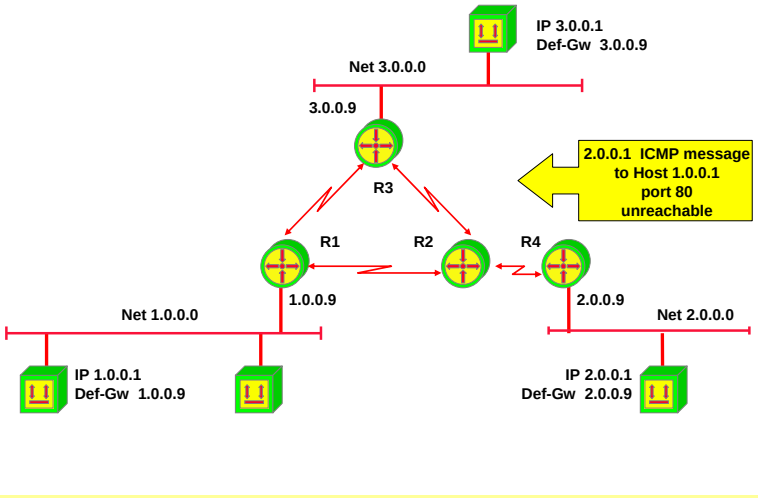
 IP Technology, v5.2 135

Remark:

Usually such an request will usually lead to an TCP RESET/ACK response if server is not listening !!!!

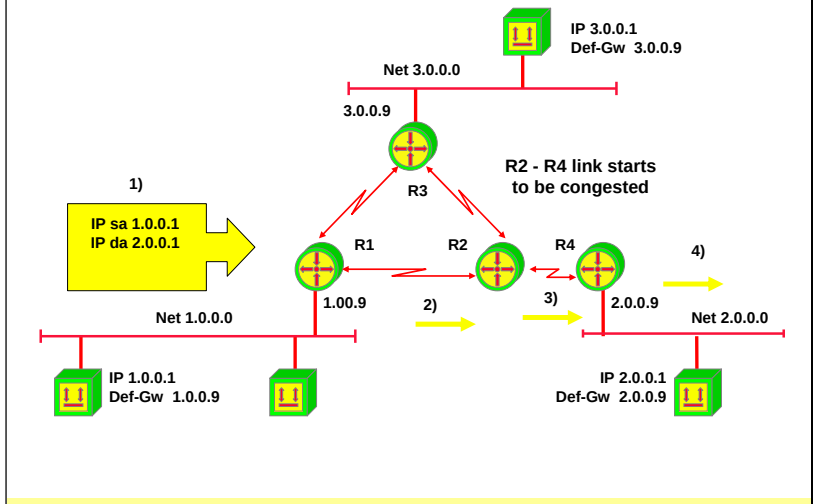
But look to RFC 1122 page 38!!!

ICMP port unreachable (no http_server_proc)



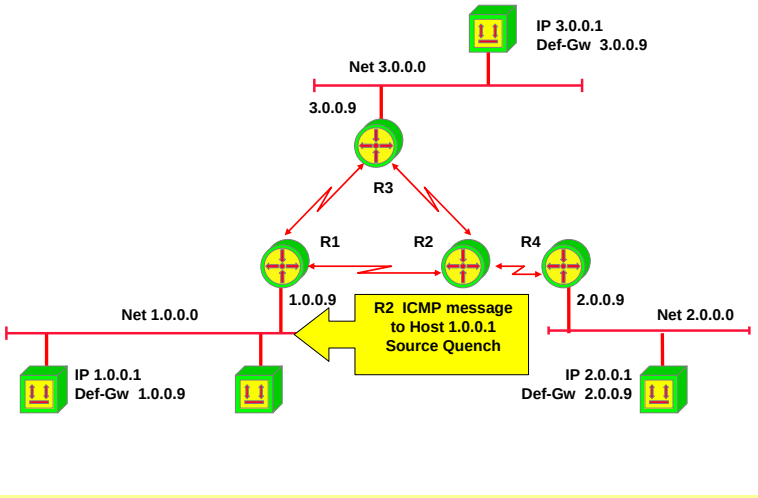
 IP Technology, v5.2 136

R2 -> R4 Link Congested



 IP Technology, v5.2 137

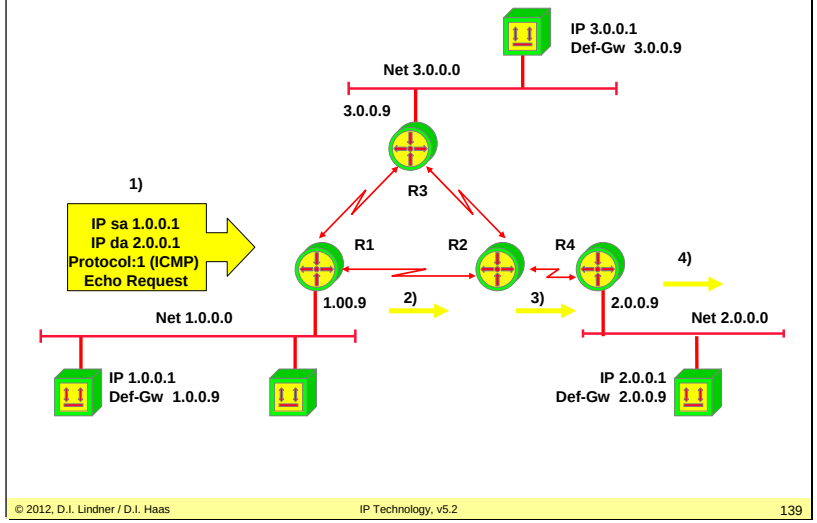
ICMP Source Quench (Flow Control STOP?)



 IP Technology, v5.2 138

Think about stations which are good Internet citizens reducing their traffic load and others which do not care about a source quench message. Guess who will get more performance?

Ping 1.0.0.1 - > 2.0.0.1

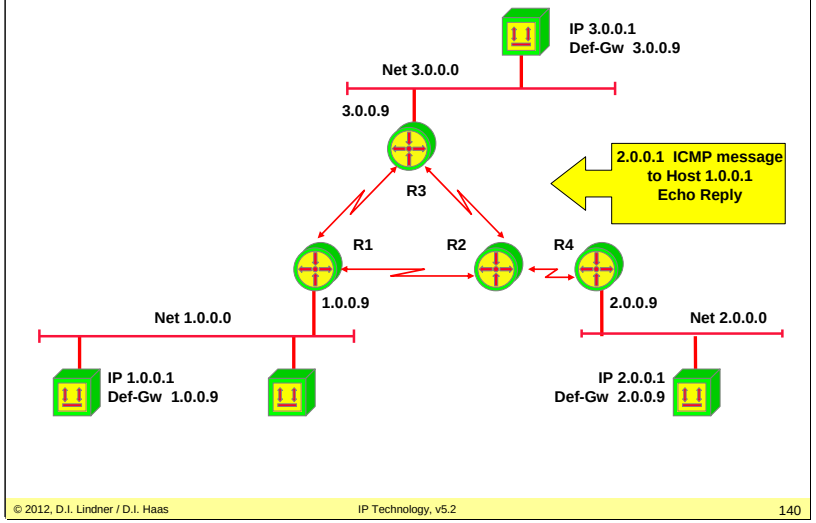


 IP Technology, v5.2 139

PING - Packet Internet Groper:

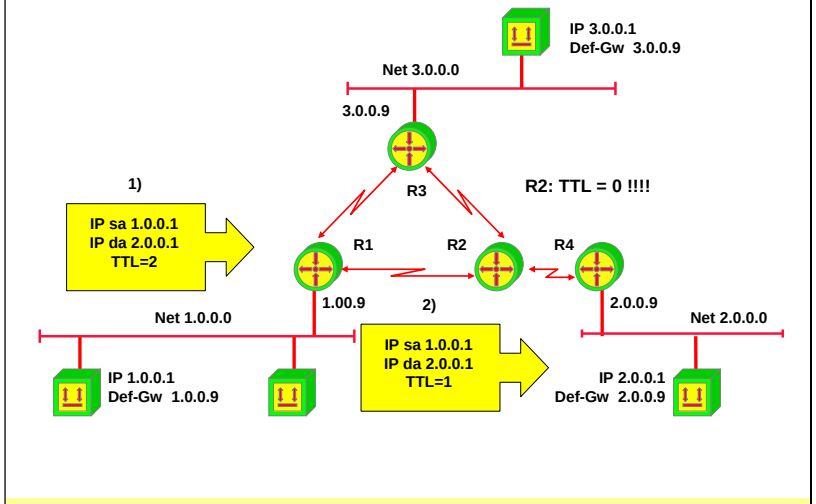
Checks the reachability of an IP station several times in a sequence and measures answer time for each trial. In case the station is reachable you get an indication about the round-trip-delay in the network. If station is not reachable the trial times out after e.g. two seconds.

Ping Echo 2.0.0.1 - > 1.0.0.1

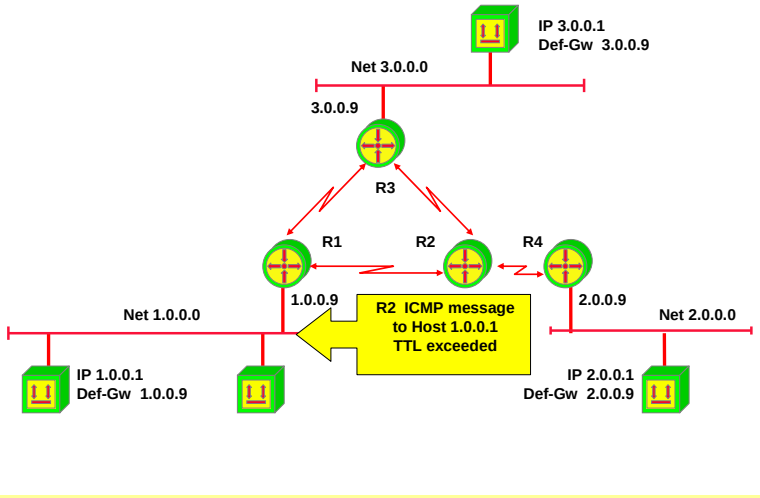


 IP Technology, v5.2 140

Delivery 1.0.0.1 - > 2.0.0.1 (TTL=2)



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

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.

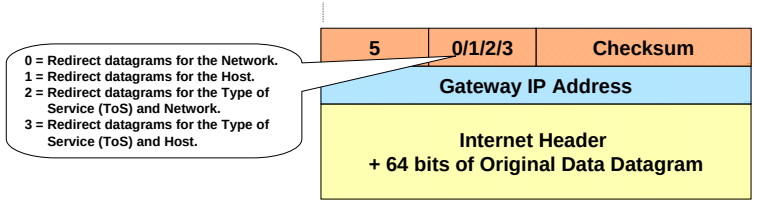
tracert 140.252.13.65

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.

Code Field for Type 5 (Redirect)

- **If a router knows of a better (faster, shorter) path to a target then it will notify the sender through ICMP redirect**
  - In any case the router will still forward the packets on the inefficient path
  - Datagrams will be sent twice through a LAN, if the sender ignores the redirect message



Rules:

The interface on which the datagram comes into the router is the same interface on which the same datagram gets routed out.

The subnet/network of the source IP address is the same subnet/network of the next-hop IP address of the routed packet.

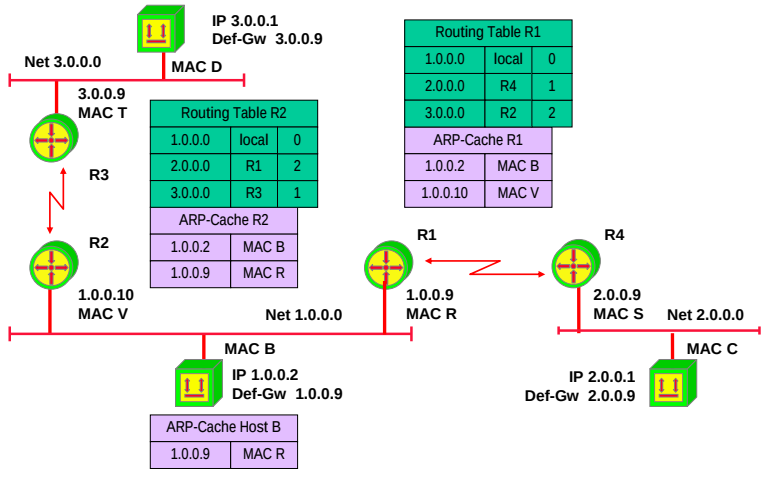
The datagram is not source-routed.

The kernel is configured to send redirects.

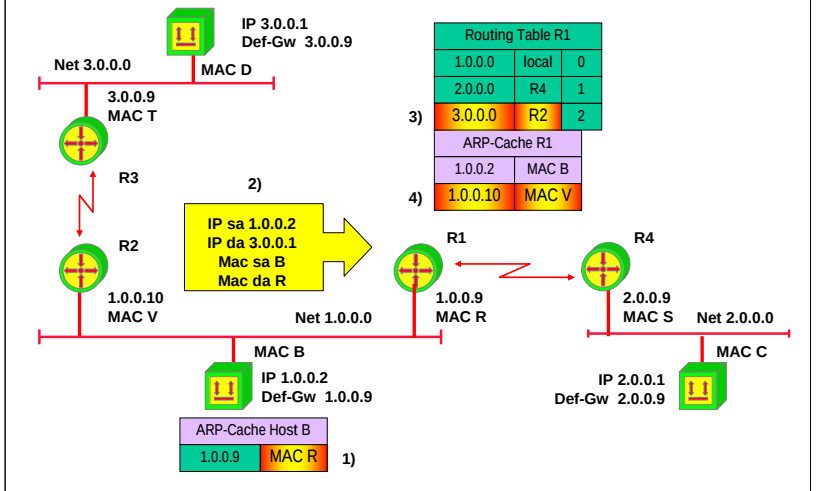
By default, Cisco routers send ICMP redirects. It can be disabled by the interface subcommand "no ip redirects".

It might be dangerous to listen and react to ICMP redirect messages in an Internet cafe. It could be a Man-in-the Middle attack.

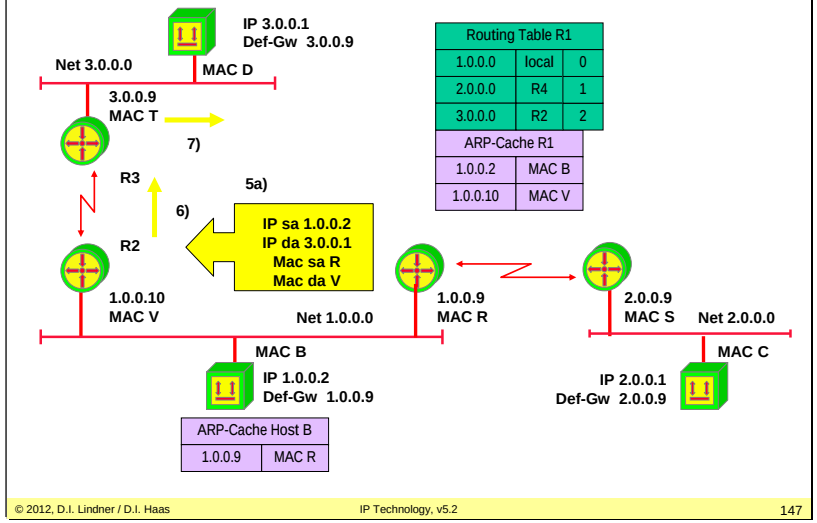
Delivery 1.0.0.2 -> 3.0.0.1



Delivery 1.0.0.2 -> 3.0.0.1

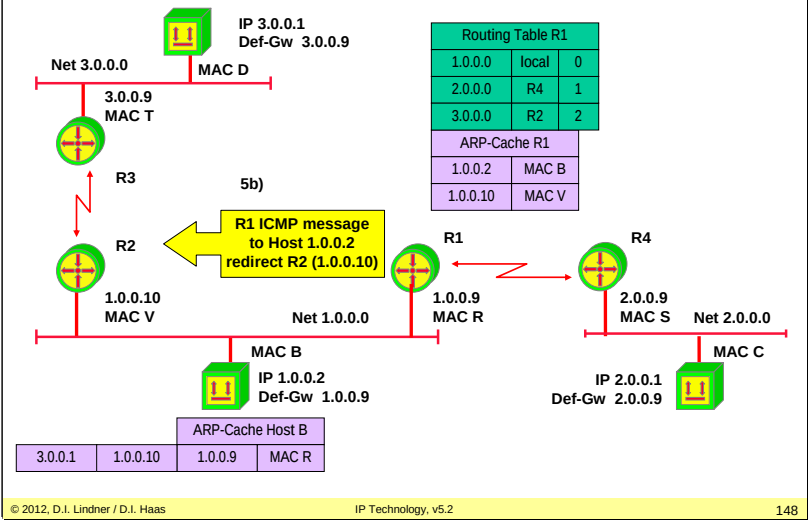


Delivery 1.0.0.2 -> 3.0.0.1



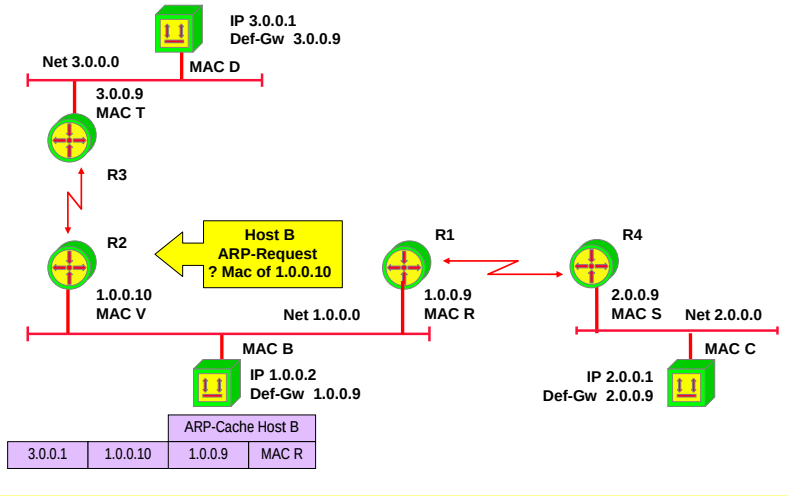
IP datagram is forwarded on the same interface as it was received -> redirect would be nice to avoid sending this datagram twice on net 1.0.0.0.

ICMP redirect

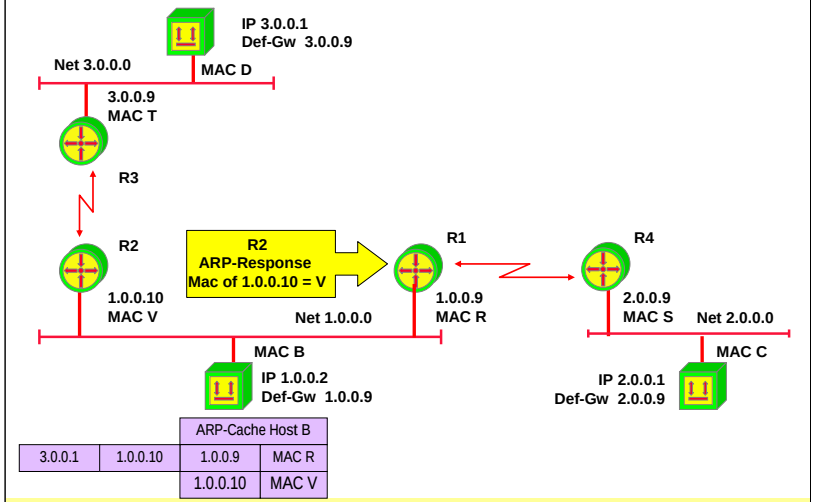


Message 5b is sent to IP 1.0.0.2 !!!

Delivery 1.0.0.2 -> 3.0.0.1

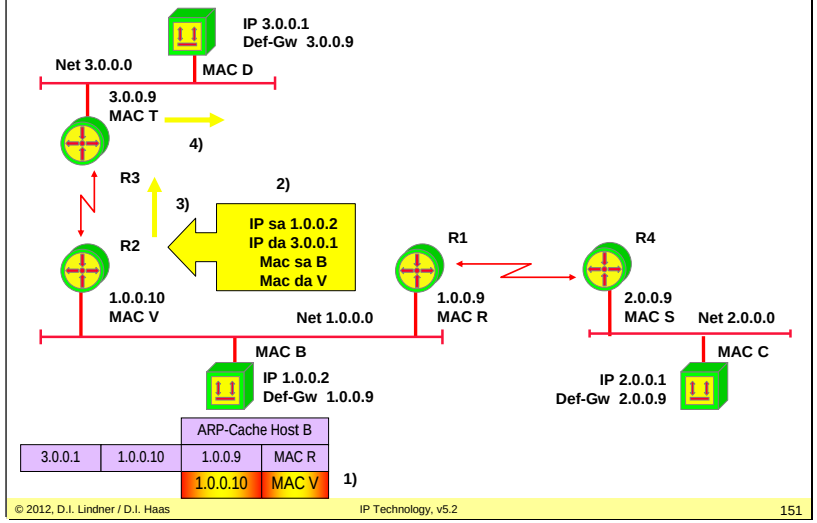


Delivery 1.0.0.2 -> 3.0.0.1



ARP response is sent to IP 1.0.0.2

Next Packet 1.0.0.2 -> 3.0.0.1

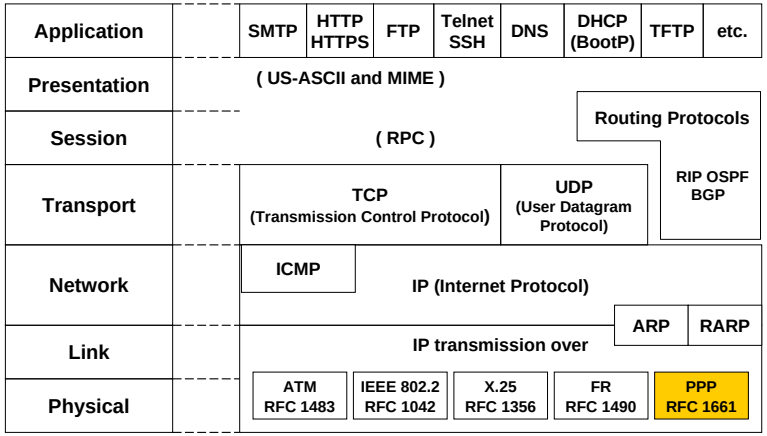


Next datagram of 1.0.0.0 is now sent to the correct (nearer) router.

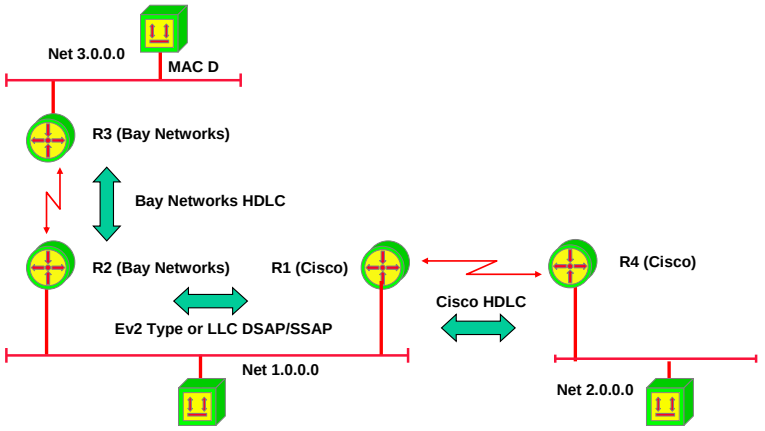
Agenda

- **Introduction**
  - Short History of the Internet (not part of the exam!)
  - Basic Principles
- **IP**
  - IP Protocol
  - Addressing
  - Classful versus Classless (not part of the exam!)
- **IP Forwarding**
  - Principles
  - ARP
  - ICMP
  - PPP
- **First Hop Redundancy**
  - Proxy ARP, IDRP
  - HSRP
  - VRRP (not part of the exam!)

TCP/IP Protocol Suite



Interoperability without PPP



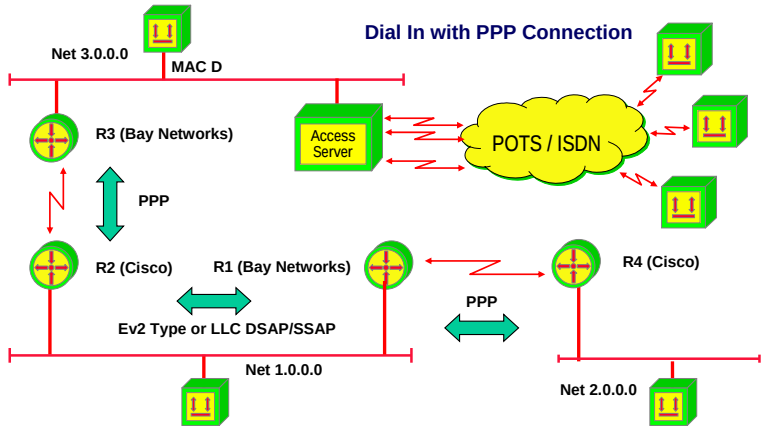
Reasons for PPP (Point-to-Point Protocol)

Communication between router of different vendors on a LAN was possible from the very beginning. Remember: Ethernet V2 Protocol Type field or LLC-DSAP/SSAP fields carry information about the protocol stack (e.g. IP or IPX or SNA or NetBEUI or AppleTalk).

Communication between router of different vendors on a serial line was not possible because of the proprietary “kind of HDLC” encapsulation method used by different vendors.

PPP standardizes multiprotocol encapsulation on a serial line. Interoperability was the one main focus at the first stage.

Interoperability with PPP



 IP Technology, v5.2 155

After the interoperability issue was solved PPP focuses on providing dial-in connectivity for IP systems. PPP connections between IP hosts (PCs) and access-servers allow working from remote in the same way as if the IP host would be directly connected to a LAN.

PPP became a standardized dial-in method for all kind of access-technology:

First modems and POTS (Plain Old Telephone Network) were used in order to establishing a PPP connection between IP host (PC) and an access server, later ISDN with PPP over transparent B-channel was introduced.

Nowadays ADSL (Asymmetric Digital Subscriber Line) or VDSL technology uses variants of PPP in order to connect your home network to your ISP (Internet service provider). PPPoE (PPP over Ethernet) and PPPoA (PPP over ATM) are these variants allowing your home network to be bridged or tunneled over ATM to the access server of the ISP.

In Dial-In VPN technology developed by Microsoft and Cisco we can find PPP tunneling over IP networks allowing a kind of Virtual Private Network (VPN) functionality to be established over the non-trusted Internet. So with Microsoft PPTP (Point-to-Point Tunneling Protocol), Cisco L2F (L2 Forwarding Protocol) and L2TP (Layer2 Tunneling Protocol, IETF-RFC) you will can see these PPP tunneling techniques in action.

PPP Overview

- **Data link protocol (L2)**
- **Used to encapsulate network layer datagrams or bridged packets (multiprotocol traffic)**
  - Over serial communication links in a well defined manner
- **Connectionless service**
  - Although we speak about a PPP connection, details are provided later
- **Symmetric point-to-point protocol**
- **Industry standard for dial-in service**
  - Used for interoperability, even over leased lines
- **Supports the simultaneous use of network protocols**

 IP Technology, v5.2 156

The Point-to-Point Protocol (PPP) provides a standard method for transporting multi-protocol datagrams over point-to-point links.

PPP Components

- **HDLC framing and encapsulation (RFC 1662)**
  - Bitstuffing for synchronous serial lines
  - Modified bytestuffing for asynchronous serial
  - Only connectionless service used (UI frame)
- **Link Control Protocol (LCP, RFC 1661)**
  - Establishes and closes the PPP connection / PPP link
  - Tests the link for quality of service features
  - Negotiation of parameters
  - Configures the PPP connection / PPP link
- **Family of Network Control Protocols (NCPs, div. RFCs)**
  - Configures and maintains network layer protocols
  - NCPs exist for IP, OSI, DECnet, AppleTalk, Novell
  - NCPs are started after PPP link establishment through LCP

PPP consists of three main components:

1. A method for encapsulating multi-protocol datagrams taken from good old HDLC in connectionless mode.
2. A Link Control Protocol (LCP) for establishing, configuring, and testing the data-link connection.
3. A family of Network Control Protocols (NCPs) for establishing and configuring different network-layer protocols.

Some more details …

HDLC is basis for encapsulation but only framing and error detection are necessary; hence simple unnumbered information frames (UI) are sufficient.

PPP supports full-duplex links only.

PPP Frame = IP Datagram + 2-8 bytes extra header (extra header consists of HDLC header and PPP header)

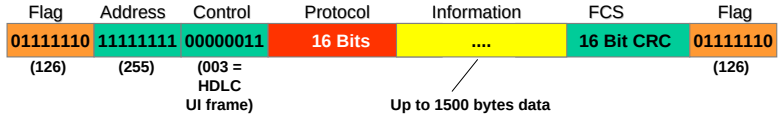
Overhead:

Only 8 additional octets are necessary to form the encapsulation when used with the default HDLC framing. In environments where bandwidth is at an issue, the encapsulation and framing may be shortened to 2 or 4 octets.

Bytestuffing on asynchronous lines:

If the flag byte (126) occurs in the data field it has to be escaped using the escape byte 125, while byte 126 is transmitted as a two byte sequence (125, 94) and the escape byte itself is transmitted as (125, 93). Hence bytestuffing is data dependent overhead!

PPP Frame Format



- **Some protocol fields (hex values)**

0021	Internet Protocol	0027	DECnet Phase 4
0029	AppleTalk	002B	Novell IPX
8021	IP Control Protocol	8027	DECnet Control Protocol
8029	AppleTalk Control Prot.	802B	IPX Control Protocol
C021	Link Control Protocol	C023	Authentication Protocol
C223	Authentication CHAP		

Flag: HDLC flag

Address: 11111111 means all stations. PPP does not assign individual station addresses on L2.

Protocol: The True PPP Field

The most important field is the protocol field, which has two octets and its value identifies the datagram encapsulated in the information field of the packet.

PPP Header Compression:

If protocol field compression is enabled, the protocol field is reduced from 2 to 1 byte. Since the first two bytes are always constant, that is the address byte (always 255) and the control byte (always 003), PPP also supports address-and-control-field-compression, which omits these bytes.

Protocol Field

0xxx – 3xxx	L3 protocol type
4xxx – 7xxx	L3 protocol type without associated NCPs
8xxx – bxxx	Associated NCPs for protocols in range 0xxx – 3xxx
cxxx – fxxx	LCP, PAP, CHAP, ...

0021	IP
002b	Novell IPX
002d	Van Jacobson Compressed TCP/IP
002f	Van Jacobson Uncompressed TCP/IP
8021	IP-NCP (IPCP)
802b	IPX-NCP (IPXCP)

Important Examples

c021	Link Control Protocol (LCP)
c023	Password Auth. Protocol (PAP)
c025	Link Quality Report
c223	Challenge Handshake Auth. Protocol (CHAP)

Protocol Field Values:

Protocol field values in the "0***" to "3***" range identify the network-layer protocol of specific packets, and values in the "8***" to "b***" range identify packets belonging to the associated Network Control Protocols (NCPs), if any. Protocol field values in the "4***" to "7***" range are used for protocols with low volume traffic which have no associated NCP. Protocol field values in the "c***" to "f***" range identify packets as link-layer Control Protocols (such as LCP).

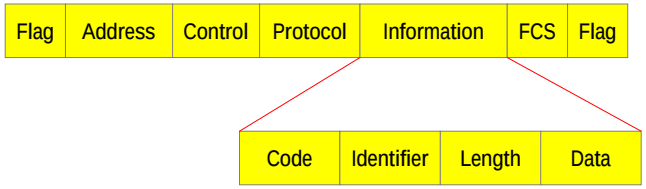
All these numbers are controlled by the IANA (see RFC-1060).

LCP Tasks

- **Establishment of PPP connection**
  - Setup, configure, test and terminate PPP connection
  - Supports various environments
  - Allows certain configuration options to be negotiated
- **Negotiation of options**
  - Encapsulation format options
  - Maximal packet sizes
  - Identification and authentication of peers (!)
  - Determination of proper link functionality

In order to be sufficiently versatile to be portable to a wide variety of environments, PPP provides a Link Control Protocol (LCP). The LCP is used to automatically agree upon the encapsulation format options, handle varying limits on sizes of packets, authenticate the identity of its peer on the link, determine when a link is functioning properly and when it is defunct, detect a looped-back link and other common misconfiguration errors, and terminate the link.

LCP Frame Format



- **Carried in PPP information field**
  - Protocol field has to be 0xC021
  - Code field indicates type of LCP packet
  - Identifier field is used to match requests and replies
  - Data field values are determined by the code field (e.g. contains options to be negotiated)

Types of LCP Packets

- **There are three classes of LCP packets:**
  - Class 1: Link Configuration packets used to establish and configure a PPP link
    - Configure-Request (code 1, details in option field), Configure-Ack (code 2), Configure-Nak (code 3, not supported option) and Configure-Reject (code 4, not supported option)
  - Class 2: Link Termination packets used to terminate a link
    - Terminate-Request (code 5) and Terminate-Ack (code 6)
  - Class 3: Link Maintenance packets used to manage and debug a PPP link
    - Code-Reject (code 7, unknown LCP code field), Protocol-Reject (code 8, unknown PPP protocol field), Echo-Request (code 9), Echo-Reply (code 10) and Discard-Request (code 11)

PPP Connection

- **PPP connection is established in four phases**
  - Phase 1: Link establishment and configuration negotiation
    - Done by LCP (note: deals only with link operations, does not negotiate the implementation of network layer protocols)
  - Phase 2: Optional procedures that were agreed during negotiation of phase 1 (e.g. CHAP authentication or compression)
  - Phase 3: Network layer protocol configuration negotiation done by corresponding NCP´s
    - E.g. IPCP, IPXCP, …
    - Actual PPP usage for configured protocols after phase 3
  - Phase 4: Link termination

PPP Phases

- **Task of phase 1**
  - LCP is used to automatically
    - Agree upon the encapsulation format options
    - Handle varying limits on sizes of packets
    - Detect a looped-back link and other common configuration errors (magic number for loopback detection)
  - Options which may be negotiated
    - Maximum receive unit
    - Usage of an authentication protocol
    - Quality protocol
    - Protocol-Field-Compression
    - Address-and-Control-Field-Compression
    - These options are described in RFC 1661 (except authentication protocols)

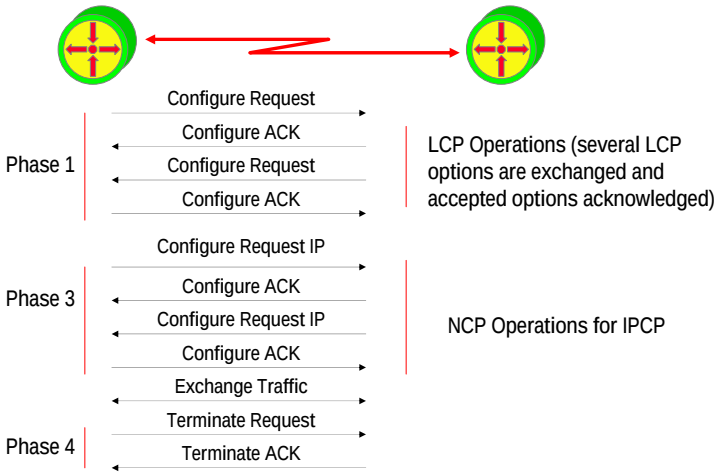
PPP Phases (cont.)

- **Task of phase 1 (cont.)**
  - Options which may be negotiated but implementations are specified in other RFCs
    - PPP link quality protocol (RFC 1989)
    - PPP compression control protocol (RFC 1962)
    - PPP compression STAC (RFC 1974)
    - PPP compression PREDICTOR (RFC 1978)
    - PPP multilink (RFC 1990)
    - PPP callback (draft-ietf-pppext-callback-ds-01.txt)
    - PPP authentication CHAP (RFC 1994)
    - PPP authentication PAP (RFC 1334)
    - PPP Extensible Authentication Protocol (EAP), RFC 2284

PPP Phases (cont.)

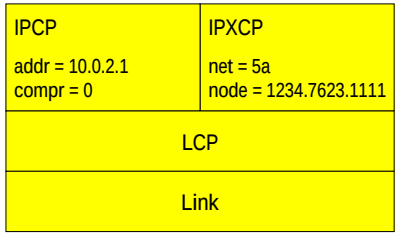
- **Task of phase 2**
  - Providing of optional facilities
    - Authentication, compression initialization, multilink, etc.
- **Task of phase 3**
  - Network layer protocol configuration negotiation
    - After link establishment, stations negotiate/configure the protocols that will be used at the network layer; performed by the appropriate network control protocol
    - Particular protocol used depends on which family of NCPs is implemented
- **Task of phase 4**
  - Link termination
    - Responsibility of LCP, usually triggered by an upper layer protocol of a specific event

PPP Link Operation Example



Network Control Protocol

- One per upper layer protocol (IP, IPX…)
- Each NCP negotiates parameters appropriate for that protocol
- **NCP for IP (IPCP)**
  - **Provides similar functionality as DHCP for LAN**
    - IP address, Default Gateway, DNS Server, TTL, TCP header compression can be negotiated or assigned



Point-to-Point links tend to exacerbate many problems with the current family of network protocols. For instance, assignment and management of IP addresses, which is a problem even in LAN environments, is especially difficult over circuit-switched point-to-point links (such as dial-up access servers). These problems are handled by a family of Network Control Protocols (NCPs), which each manage the specific needs required by their respective network-layer protocols.

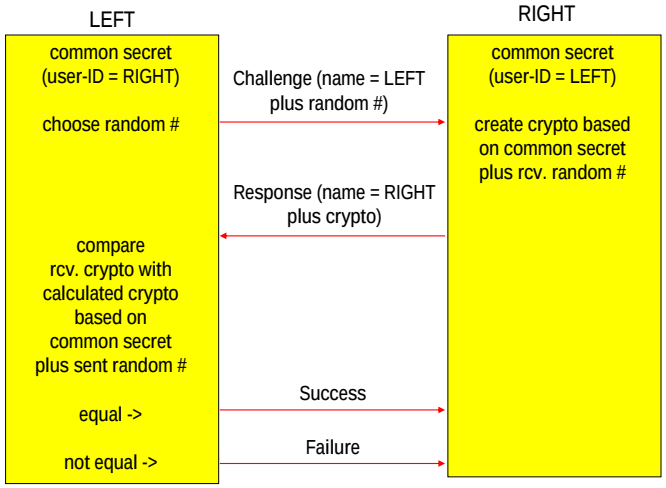
NCPs have been developed for all important network layer protocols such as IP, which uses the IP Control Proocol (IPCP).

There are also NCPs designed to enable compression and authentication.

CHAP Authentication RFC 1994

- **Challenge Authentication Protocol**
  - Follows establishment of LCP
  - Identifies user
  - Three way handshake procedure
  - One way authentication only
  - Station which starts the three way handshake proofs authentication of other station
  - Cryptographic hash function (e.g. keyed MD5) is applied to random numbers used (hopefully) only once
    - Network snooping does not reveal any passwords
    - Offline dictionary attacks are possible
  - Overcomes weaknesses of PAP (Password Authentication Protocol) which used transmission of cleartext passwords (!!!)
- **Three way handshake have to be performed in both directions**
  - If two way authentication is necessary

CHAP Authentication Procedure

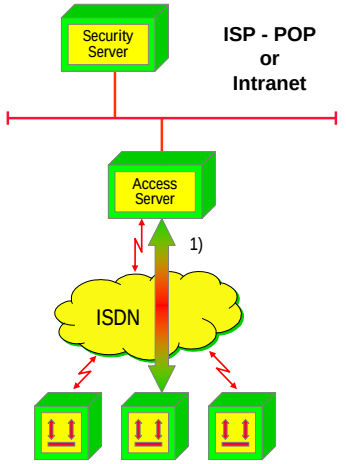


After PPP link successfully installed by LCP the local station sends a challenge message to remote station. The challenge contain random number and own user-id. Remote station replies with value using one way hash function (e.g. MD5) based on crypto negotiated (pre-shared secret configured already) for this user-id. Response is compared with LEFT stations own calculation of random number with same crypto. If equal a success messages is sent to remote station (if unequal a failure message is sent). Thee way Handshake is complete. Now LEFT has verified that RIGHT knows the secret hence RIGHT is successfully identified (one-way authentication). For the other direction the same procedure takes place in the other direction. After additional three messages we can reach two-way authentication (LEFT is successfully identified by right).

PPP as Dial-In Technology

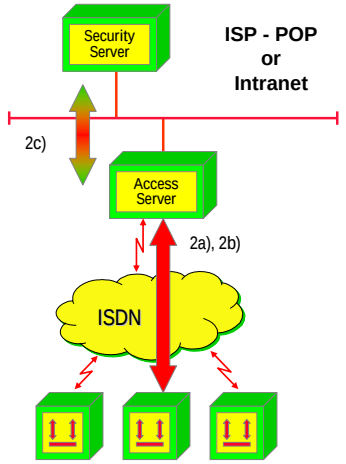
- **Dial-In:**
  - Into a corporate network (Intranet) of a company
    - Here the term RAS (remote access server) is commonly used to describe the point for accessing the dial-in service
  - Into the Internet by having an dial-in account with an Internet Service Provider (ISP)
    - Here the term POP (point-of-presence) is used to describe the point for accessing the service

RAS Operation 1



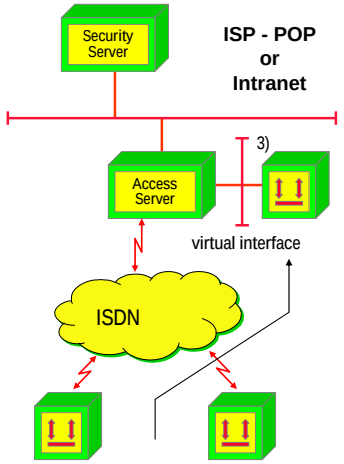
- **remote PC places ISDN call to access server, ISDN link is established (1)**

RAS Operation 2



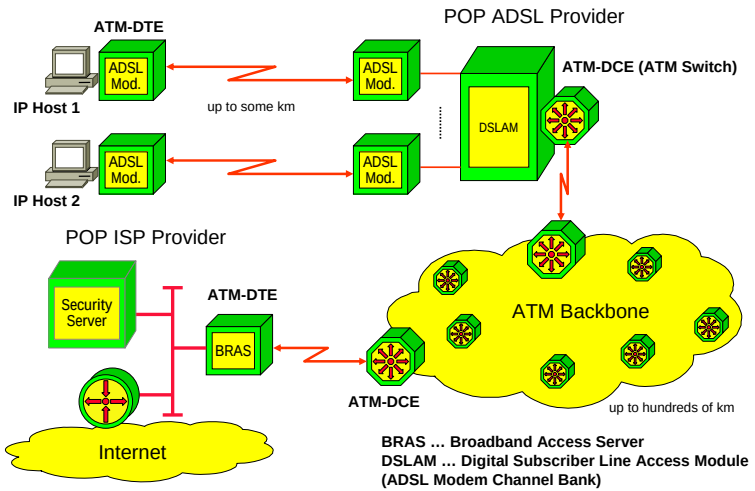
- **PPP link (multiprotocol over serial line) is established**
  - LCP Link Control Protocol (2a)
    - Establishes PPP link plus negotiates parameters like authentication CHAP
  - Authentication
    - CHAP Challenge Authentication Protocol to transport passwords (2b)
    - Verification maybe done by central security server (2c) -> Radius, TACACS, TACACS+

RAS Operation 3

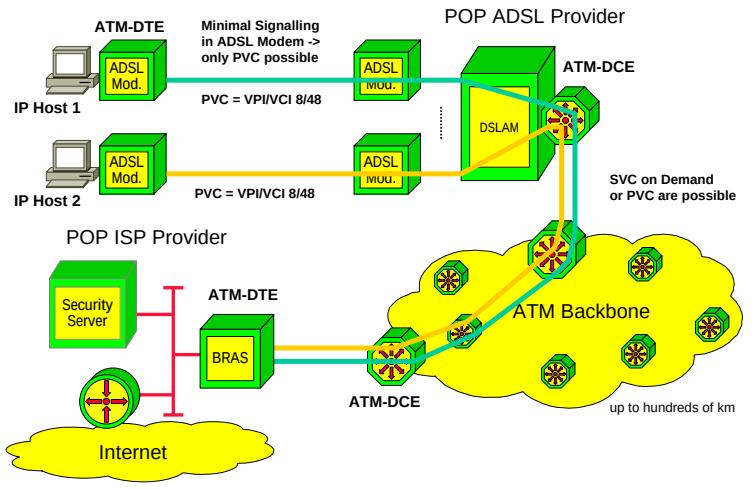


- **PPP NCP (Network Control Protocol) IPCP**
  - Assigns IP address, Def. GW, DNS to remote PC
- **Remote PC appears as**
  - Device reachable via virtual interface (3), IP host Route
- **Optionally**
  - Filter could be established on that virtual interface
    - Authorization
  - Accounting may be performed
    - Actually done by security server (AAA server)
    - TACACS, Radius

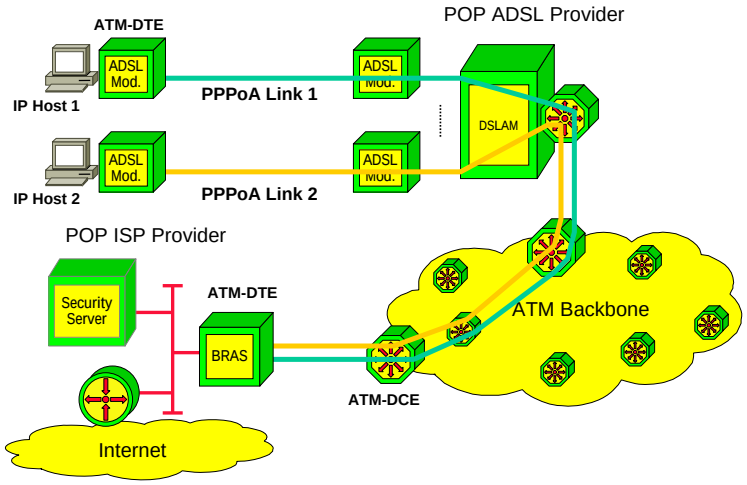
ADSL: Physical Topology



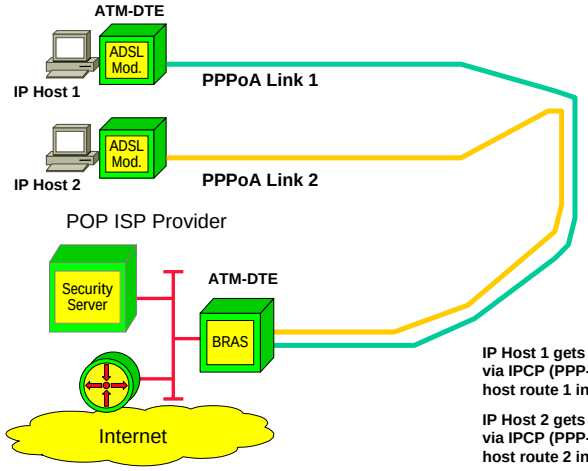
ADSL: ATM Virtual Circuits



ADSL: PPP over ATM (PPPoA)



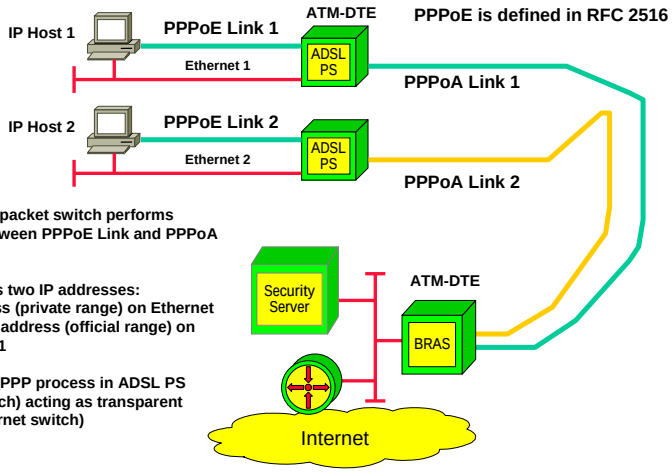
ADSL: PPP over ATM (PPPoA), IPCP



IP Host 1 gets global IP address via IPCP (PPP-NCP), appears as host route 1 in BRAS

IP Host 2 gets global IP address via IPCP (PPP-NCP), appears as host route 2 in BRAS

ADSL: PPP over Ethernet (PPPoE)



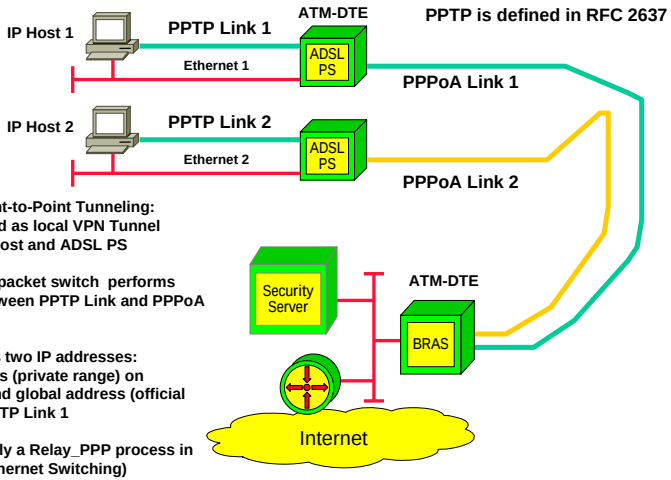
ADSL PS as packet switch performs mapping between PPPoE Link and PPPoA Link

IP Host 1 has two IP addresses: Local address (private range) on Ethernet 1 and global address (official range) on PPPoE Link 1

Note: Relay_PPP process in ADSL PS (Packet Switch) acting as transparent bridge (Ethernet switch)

 IP Technology, v5.2 179

ADSL: PPTP over Ethernet (Microsoft VPN)



PPTP ... Point-to-Point Tunneling: Protocol used as local VPN Tunnel between IP Host and ADSL PS

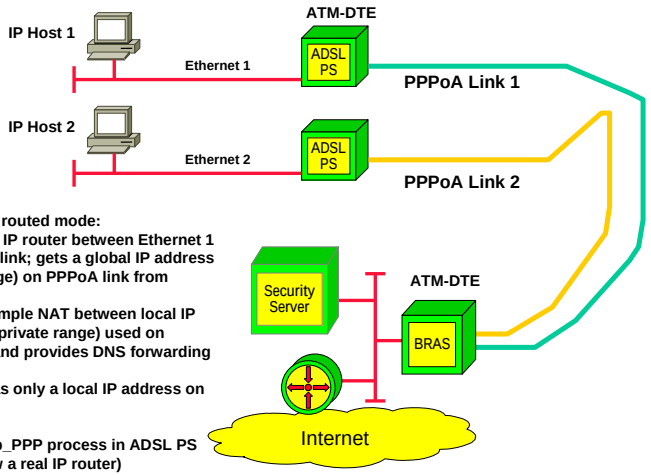
ADSL PS as packet switch performs mapping between PPTP Link and PPPoA Link

IP Host 1 has two IP addresses: Local address (private range) on Ethernet 1 and global address (official range) on PPTP Link 1

Note: Still only a Relay_PPP process in ADSL PS (Ethernet Switching)

 IP Technology, v5.2 180

ADSL: Routed PPPoA



ADSL PS in routed mode:
Acts as real IP router between Ethernet 1 and PPPoA link; gets a global IP address (official range) on PPPoA link from provider.
Performs simple NAT between local IP addresses (private range) used on Ethernet 1 and provides DNS forwarding

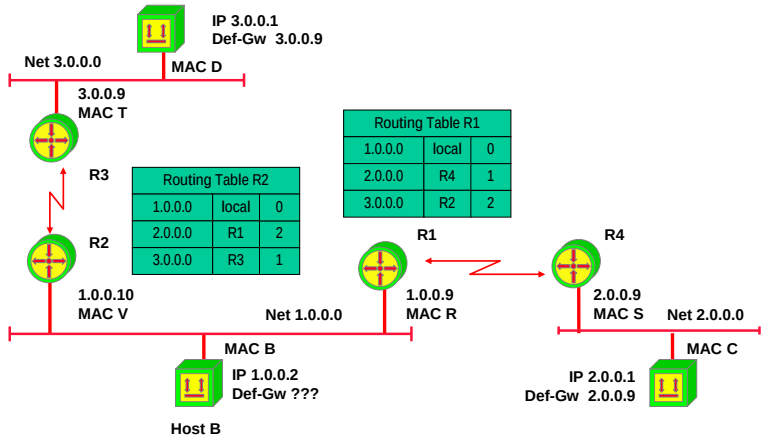
IP Host 1 has only a local IP address on Ethernet 1

Note: Dialup_PPP process in ADSL PS (PS is a now a real IP router)

Agenda

- **Introduction**
  - Short History of the Internet (not part of the exam!)
  - Basic Principles
- **IP**
  - IP Protocol
  - Addressing
  - Classful versus Classless (not part of the exam!)
- **IP Forwarding**
  - Principles
  - ARP
  - ICMP
  - PPP
- **First Hop Redundancy**
  - Proxy ARP, IDRP
  - HSRP
  - VRRP (not part of the exam!)

First L3 Hop?



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.

First Hop Redundancy (Layer 3) 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
    - IP host knows one "Default Gateway" to reach other IP networks
  - Static configuration of "Default Gateway" means:
    - Loss of the default router results in a catastrophic event, isolating all end-hosts that are unable to detect any alternate path that might be available
- **Two design philosophies:**
  - Solve the problem at the IP host level
    - OS of the IP host has to support an appropriate functionality
  - Solve the problem at the IP router level
    - OS of the IP host has to support the basic functionality only
      - That is static configuration of one "Default Gateway"
    - Appropriate functionality needed at the router

First Hop Redundancy (Layer 3) 2

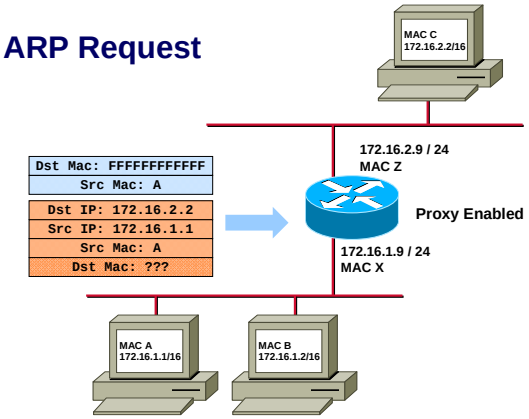
- **Methods for solving it at the IP host level:**
  - Proxy ARP
  - IRDP (ICMP Router Discovery Protocol)
  - DHCP (Dynamic Host Configuration Protocol)
  - IP Routing (RIPv2, OSPF)
- **Methods for solving it at the IP router level:**
  - HSRP (Hot Standby Router Protocol)
    - Cisco proprietary
  - VRRP (Virtual Router Redundancy Protocol)
    - Same as HSRP but open RFC
  - GLBP (Gateway Load Balancing Protocol)
    - Cisco proprietary
    - Not handled in this lecture

Old Proxy ARP Usage

- **Old method for migration from transparent bridging to IP routing**
  - Two LANs connected by a transparent bridge (=broadcast domain) using a given IP Net-ID should be decoupled by a router
  - IP address were already assigned to the LAN segments in such a way that IP subnets can be built by the replacing router
  - Now by enabling proxy ARP gateway functionality on the router the host can still use their old subnet mask in order to communicate with all other stations
  - The proxy ARP gateway of the router will answer ARP requests
  - Term “proxy” means “instead of”
    - Some system is doing some function instead of the expected system
- **Replaced nowadays by usage of IP subnetting**
  - on all systems

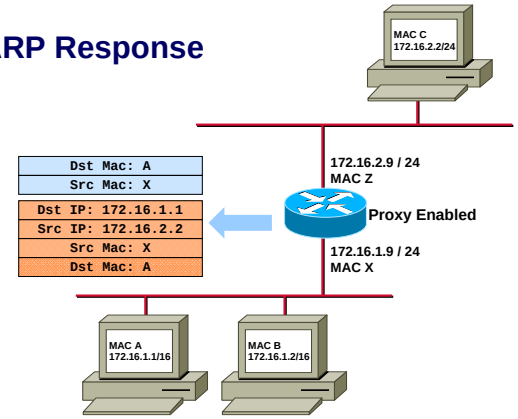
Old Proxy in Action (1)

Proxy ARP Request



Old Proxy in Action (2)

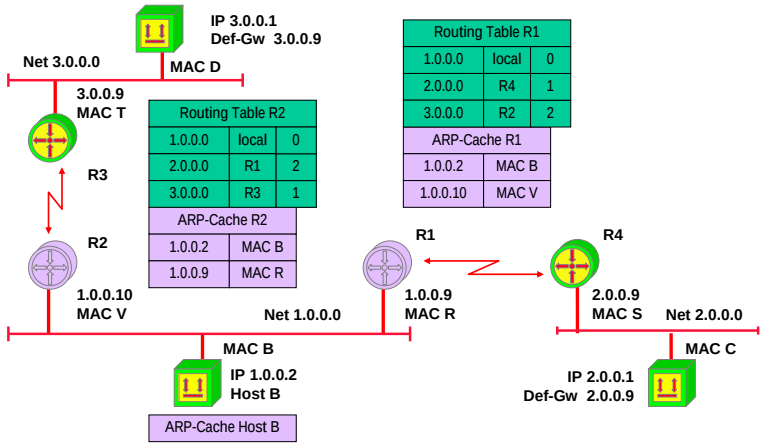
Proxy ARP Response



Proxy ARP Usage Nowadays

- **Proxy ARP is can be used if an IP host didn't know the address of the default gateway or want to find it dynamically:**
  - Normally in an IP host a static entry will tell the IP address of the router
    - If an IP datagram has to be sent to a non-local Net-ID, an ARP request will find the MAC address of the default gateway
  - With proxy ARP extensions in the IP host and with proxy ARP support enabled in the router
    - The MAC address of the router can be found without knowing the routers IP address
    - An ARP request will be sent for IP hosts with NET-IDs different from the local Net-ID and the router will respond
  - Unix stations or Windows NT/XP:
    - Proxy ARP extensions are triggered by setting the default gateway to the systems IP address itself

1.0.0.2 -> 3.0.0.1 / 2.0.0.1 with proxy ARP (1)

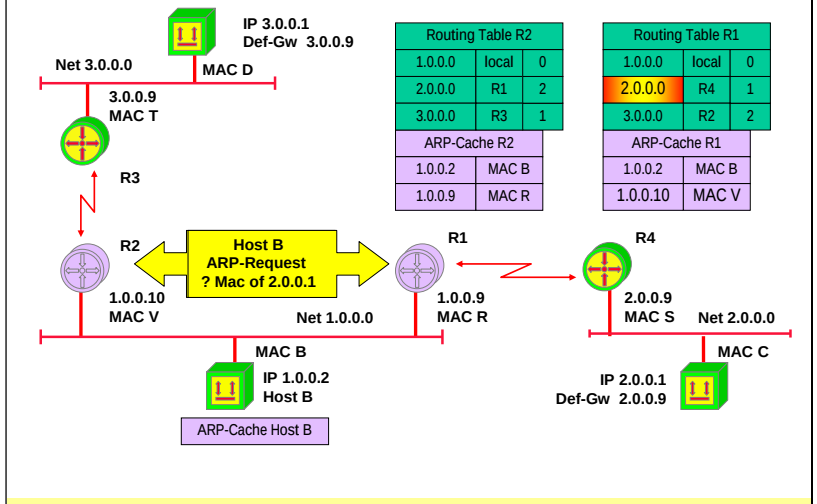


R1 and R2 proxy ARP enabled; Host B sends ARP also for net-ID unequal own net-ID

Router R1 and R2 are configured to support proxy ARP (acting as a proxy ARP gateway). Host B is configured to use proxy ARP extension by pointing to its own IP address as default gateway.

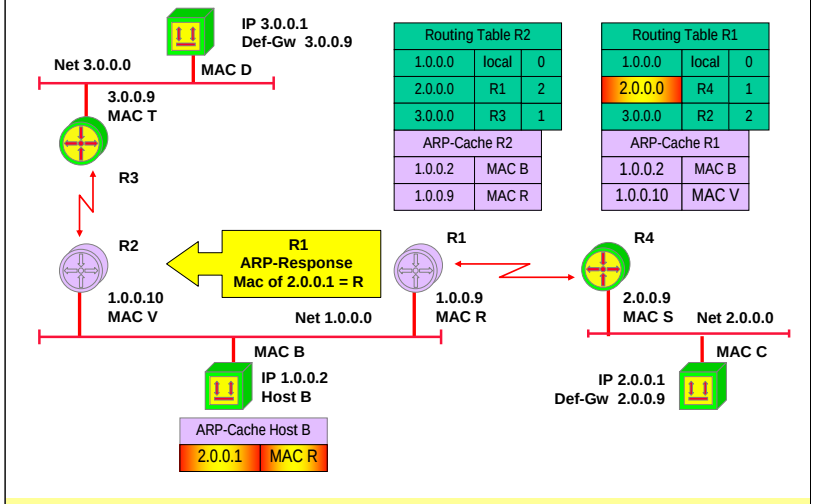
Cisco routers have proxy ARP gateway functionality enabled by default You have to turn it off, if you do not want it.

1.0.0.2 -> 3.0.0.1 / 2.0.0.1 with proxy ARP (2)



Request is sent to all systems on the given LAN (ARP request uses L2 broadcast addressing)!

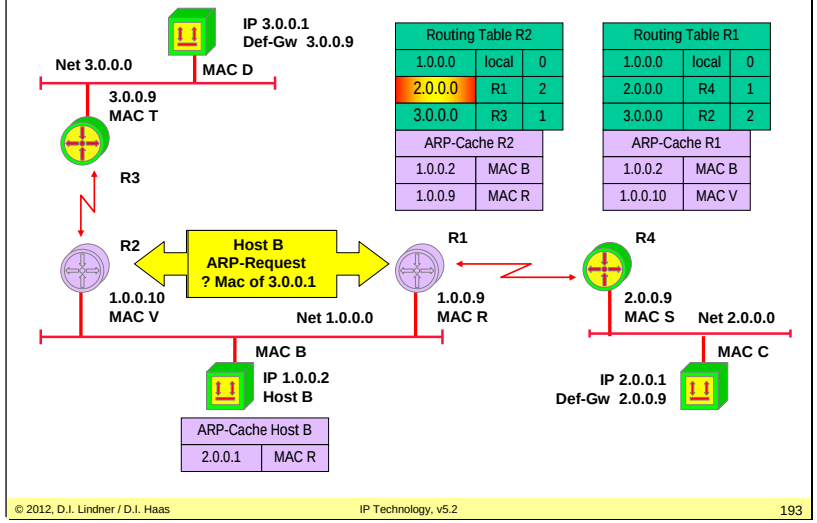
1.0.0.2 -> 3.0.0.1 / 2.0.0.1 with proxy ARP (3)



Response of R1 sent to Host B only (ARP reply uses L2 directed addressing)!

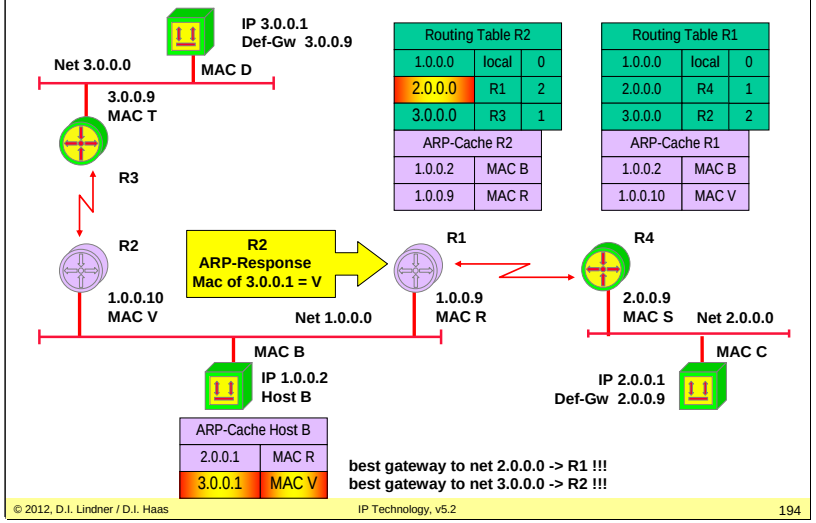
R2 will not answer the ARP request because a proxy ARP GW must not reply if the destination is reachable through the same interface. Either the destination is in same segment or another proxy ARP GW will reply, knowing a better route.

1.0.0.2 -> 3.0.0.1 / 2.0.0.1 with proxy ARP (4)



Request is sent to all systems on the given LAN (ARP request uses L2 broadcast addressing)!

1.0.0.2 -> 3.0.0.1 / 2.0.0.1 with proxy ARP (5)



Response of R2 sent to Host B only (ARP reply uses directed addressing)!

By the way: Think about security. What will happen if ARP replies are spoofed by another machine on the LAN network - wanting to become Man-In-The-Middle. That was not in the design. Instead the standard says that if there are multiple proxy ARP GWs in the same subnet the requesting host should use the first ARP response it receives. The reason for that approach was the implementation of a simple load balancing service.

Other Techniques to Solve the Problem 1

- **IRDP**
  - ICMP Router Discovery Messages (RFC 1256)
  - Routers periodically advertise their IP address on a shared media together with an preference value and a lifetime
    - ICMP Router Advertisement Message
  - Hosts may listen to these messages to find out all possible Default Gateways
    - Or may ask by sending an ICMP Router Solicitation Message
- **DHCP**
  - Dynamic Host Configuration Protocol (RFC 2131)
  - More than one Default Gateway can be specified
  - Every Default Gateway has a preference value

Other Techniques to Solve the Problem 2

- **With IDRP and DHCP**
  - You still depend on OS functionality in order to trigger switchover between redundant local routers
    - How often the currently selected router will be tested for reachability? What is if the currently selected router is reachable via LAN but networks behind are not reachable?
- **Therefore running a classical IP routing protocol on the IP host would be optimal**
  - RIPv2
    - But slow convergence if the currently selected router fails, no hello messages hence 180 seconds for recognizing that event
  - OSPF
    - Fast convergence because of hello messages, the best but the most complex solution
  - But IP routing on an IP host for that reason is seldom done

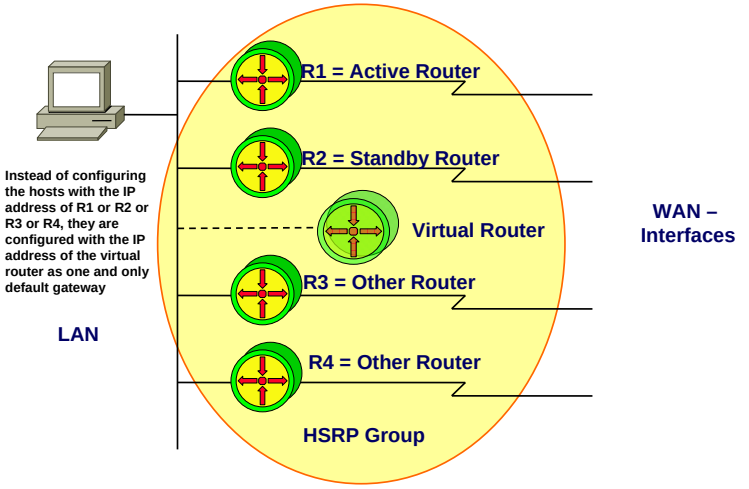
Agenda

- **Introduction**
  - Short History of the Internet (not part of the exam!)
  - Basic Principles
- **IP**
  - IP Protocol
  - Addressing
  - Classful versus Classless (not part of the exam!)
- **IP Forwarding**
  - Principles
  - ARP
  - ICMP
  - PPP
- **First Hop Redundancy**
  - Proxy ARP, IDRP
  - HSRP
  - VRRP (not part of the exam!)

HSRP – Hot Standby Router Protocol

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

HSRP Overview



 IP Technology, v5.2 199

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, v5.2 200

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)

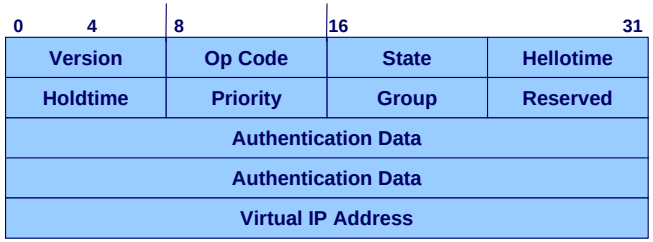
- **Two basic failover scenarios:**
  - 1) Active router is not reachable via LAN
    - Standby router will take over active role
    - A new standby router is elected from the remaining routers of a HSRP group
    - Timing depends on HSRP hello message interval and hold-time
      - Default hello-time = 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 or 224.0.0.102 with a IP TTL = 1

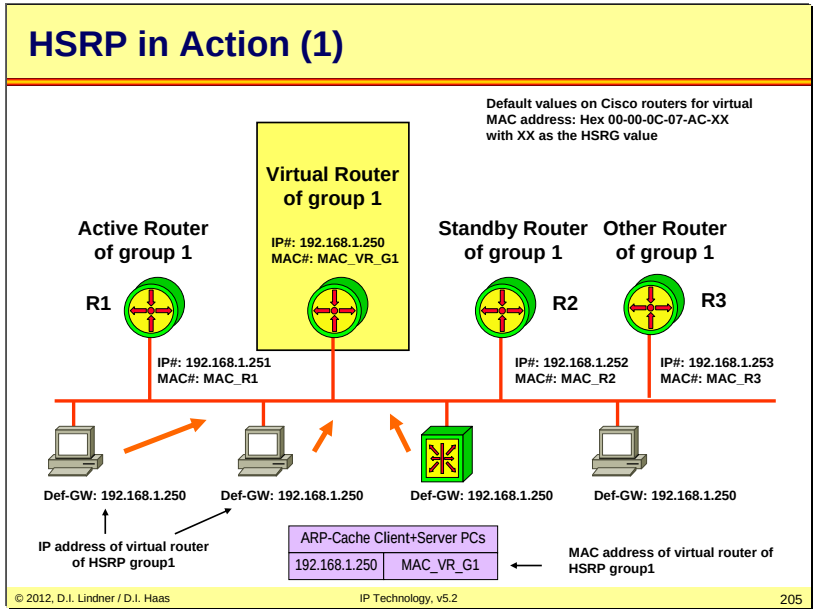


- **Version**: Version of the HSRP messages
- **Op code**: 3 types
  - **Hello**: Indicates that a router is running and is capable of becoming the active or standby router
  - **Coup**: When a router wishes to become the active router
  - **Resign**: When a router no longer wishes to be the active router
- **States**: Initial, learn, listen, speak, standby, active
- **Hellotime**: Contains the period between the hello messages that the router sends
- **Holdtime**: Amount of time the current hello message is valid
- **Priority**: Compares priorities of 2 different routers
- **Group**: Identifies standby group (0...255)
- **Authentication data**: Cleartext or MD5 signed hash

HSRP 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)



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.

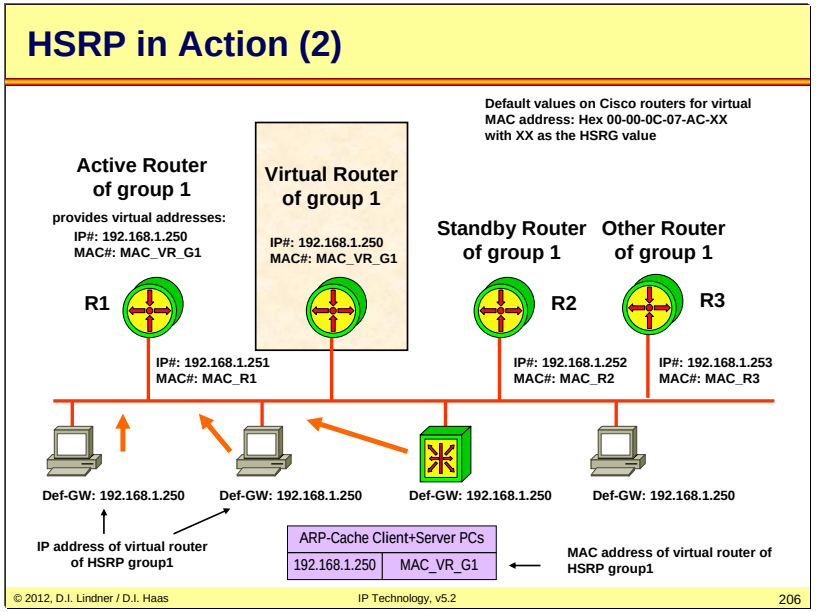
The default priority for an HSRP router is 100.

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.



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.

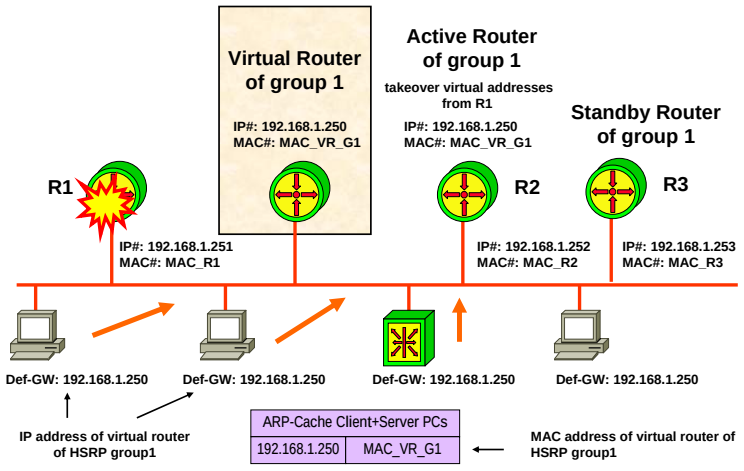
Additional HSRP member routers - other routers :

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 (3)



 IP Technology, v5.2 207

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.

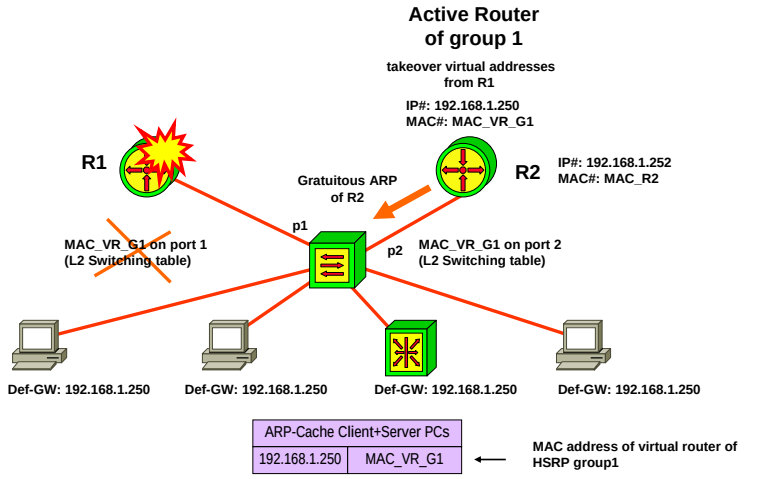
The end-user stations continue to send packets to the virtual router's virtual MAC address and IP address where the new active router delivers the packets to the destination.

HSRP Additional Aspects

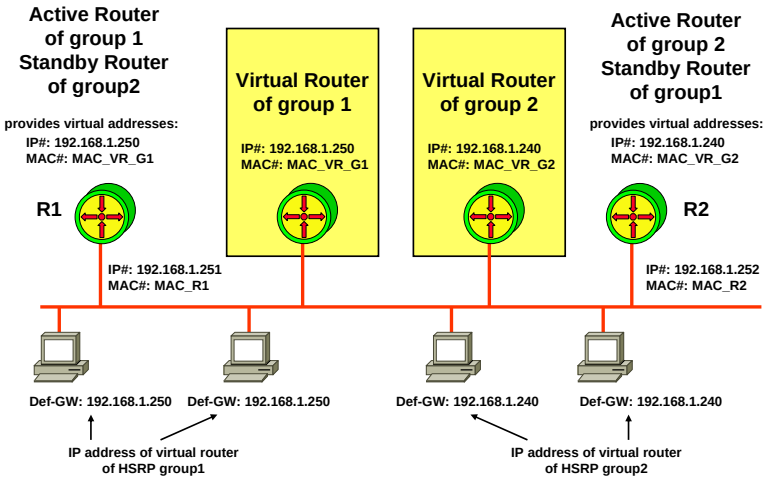
- **L2 Ethernet Switching Table Refresh:**
  - Done by gratuitous ARP in case of switchover
- **Load Balancing:**
  - You can achieve this by specifying at least two different HSRP groups with complementary roles
- **HSRP Security**
  - Authentication of messages by generation of fingerprints and checking this fingerprints
    - Based on keyed MD5
  - Against HSRP spoofing

 IP Technology, v5.2 208

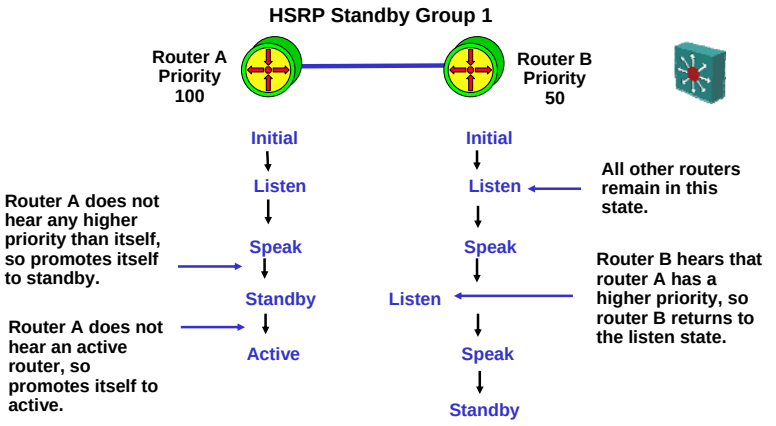
HSRP – Gratuitous ARP



HSRP Load Balancing



HSRP States Details



 IP Technology, v5.2 211

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.

Agenda

- **Introduction**
  - Short History of the Internet (not part of the exam!)
  - Basic Principles
- **IP**
  - IP Protocol
  - Addressing
  - Classful versus Classless (not part of the exam!)
- **IP Forwarding**
  - Principles
  - ARP
  - ICMP
  - PPP
- **First Hop Redundancy**
  - Proxy ARP, IDRP
  - HSRP
  - VRRP (not part of the exam!)

 IP Technology, v5.2 212

VRRP Operation 1

- **VRRP (Virtual Router Redundancy Protocol)**
  - RFC 2338 (Standards Track)
- **Principle:**
  - A group of routers forms a VRRP group
  - The group is represented by a virtual router
    - With is identified by a VRID (Virtual Router ID) and a virtual MAC address
  - One router is elected as the **virtual router master**, all other routers get the role of **virtual router backup** routers
  - The real IP address of the virtual router master become the IP address of the virtual router for a given VRRP group
    - IP address owner
  - Default Gateway of IP hosts is configured with the IP address of the virtual router for a given VRRP group
  - Virtual router master responds to ARP request directed to the IP address of the virtual router with the virtual MAC address
  - Backup routers supervise if master router is alive and take over the role of the master in case of failure
    - VRRP protocol using IP protocol number 112, IP multicast 224.0.0.18, and Ethernet multicast as destination address
  - Router must be able to support more than one unicast MAC address on an Ethernet interface

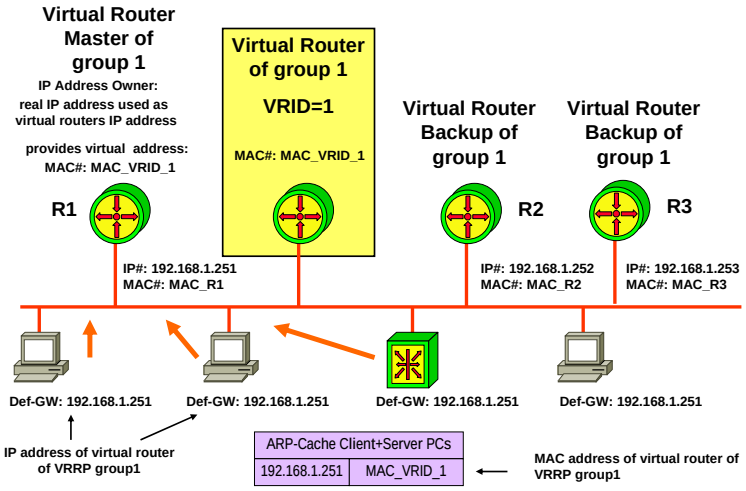
VRRP Operation 2

- **Roles of router:**
  - Virtual router master, virtual router backup defined by VRRP priority
  - Priority value can be configured
    - Default value is 100
  - The higher the better
    - Will become the master after initialization
    - If priority is equal than the higher IP address decides
  - Preempt allows to give up the role of the master router when a router with higher priority is activated or reported
    - e.g. a failed router comes back or tracking has changed priority
- **Load Balancing:**
  - Specify at least two different VRRP groups with complementary roles
- **VRRP authentication:**
  - Based on keyed MD5
  - Against VRRP spoofing

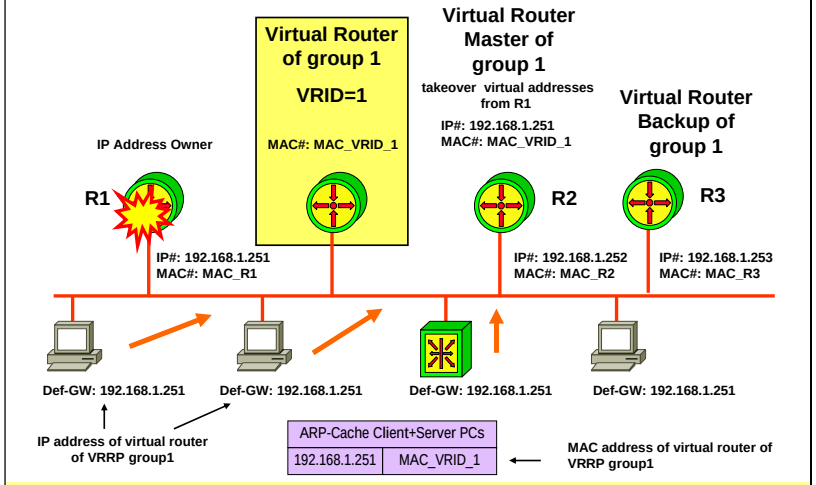
VRRP Operation 3

- **Failover scenarios:**
  - Master router not reachable via LAN
    - Backup router with highest priority will take over master role
    - Timing depends on VRRP advertisements interval and master down interval
      - Default advert-interval = 1 seconds
      - Default master-down-interval = 3 * advert-interval + skew-time
  - Master router losses connectivity to a WAN interface (basic tracking options) or losses connectivity to an IP route (enhanced tracking options)
    - If tracking and preempt is configured backup router will take over
      - Tracking will lower the priority
      - Preempt allows another router to take over the role of the master router even if the current master router does not fail
  - Enhanced tracking options depend on IOS version

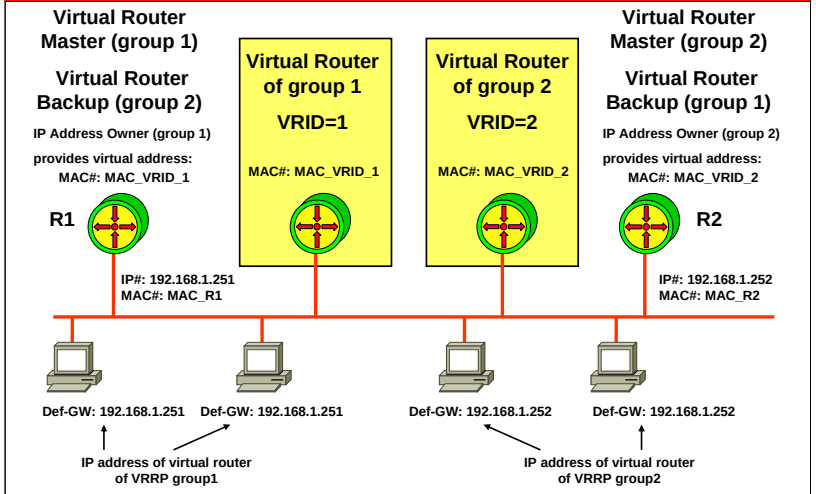
VRRP 1



VRRP 2



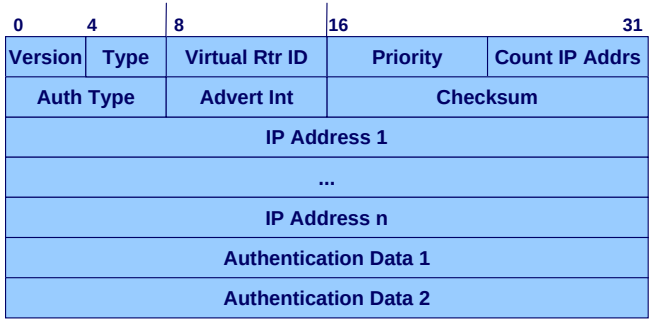
VRRP Load Balancing



Some VRRP Details

- **VRRP:**
  - Second or milliseconds timers
  - VRID range
    - 1 – 255
    - Maximum 255 groups
  - Virtual Mac Address: 00-00-5E-00-01-VRID
    - VRID value = group number
  - IP multicast 224.0.0.18

VRRP Protocol Fields



- Version - This version is version 2.
- Type - The only packet type defined in this version of the protocol is: 1 ADVERTISEMENT.
- Virtual Rtr ID - The Virtual Router Identifier (VRID) field identifies the virtual router this packet is reporting status for.
- Priority - VRRP routers backing up a virtual router MUST use priority values between 1-254 (decimal).
- Count IP Addresses -The number of IP addresses
- Auth Type - Identifies the authentication method being utilized.
- Advertisement Interval - Indicates the time interval (in seconds) between advertisements.
- Checksum - used to detect data corruption
- IP Address(es) - One or more IP addresses that are associated with the virtual router.
- Authentication Data - The authentication string is currently only utilized for simple text authentication