



The Internet Protocol (IP)

The Blood of the Internet



"Information Superhighway is really an acronym for 'Interactive Network For Organizing, Retrieving, Manipulating, Accessing And Transferring Information On National Systems, Unleashing Practically Every Rebellious Human Intelligence, Gratifying Hackers, Wiseacres, And Yahoos'."



Keven Kwaku

The Internet Protocol (IP)

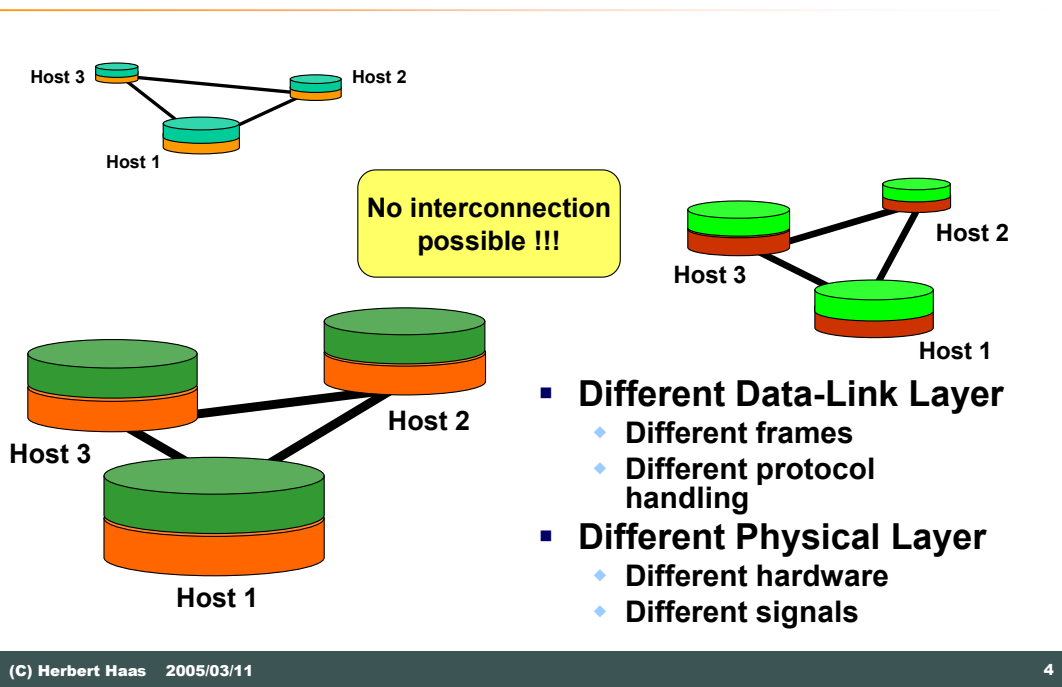


- **Introduction**
- **IP Addressing**
 - ♦ IP Header
 - ♦ IP Address Format
- **Address Classes**
 - ♦ Class A - E
- **Subnetting, VLSM**
- **IP Fragmentation**

In this chapter we talk about the **Internet Protocol (IP)**, especially about IP Version 4. IPv4 was standardized in September 1981 in RFC 791.

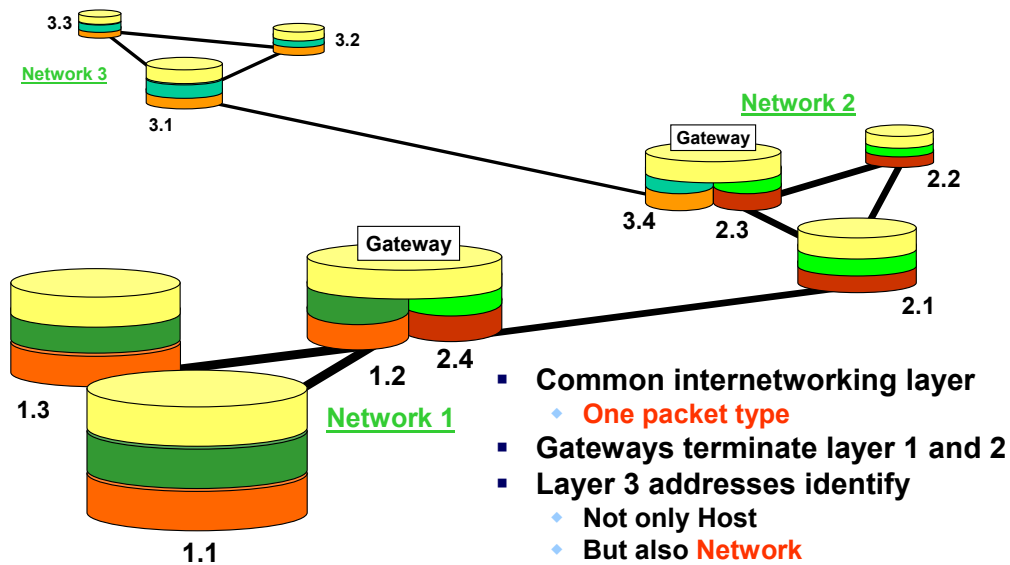
IP is a packet-switching technology on OSI layer 3. IP is connectionless and an overlay technique. In this module we discuss fundamental questions around the IP protocol, such as: What other (helper) protocols are necessary? What is an IP-Address? What is Subnetting and VLSM?

Need of an Inter-Net Protocol (1)



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)



(C) Herbert Haas 2005/03/11

5

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 Introduction (1)



- **Packet switching technology**
 - ♦ Packet switch = router = "**gateway**" (IETF terminology)
 - ♦ End system is called **IP host**
 - ♦ Layer 3 address (Structured)
- **Datagram Service**
 - ♦ Connectionless
 - ♦ Best effort delivery

IP can be described by mentioning two facts: First, IP is "just another" packet switching technology on layer 3 and the most important thing here is the structured IP address, identifying the network and the host. Second, the type of packet-switching is connectionless, that is there is no need to establish a connection prior of sending packets. We call this a "best effort" or "datagram delivery". There is no guarantee, that all packets are delivered reliably.

IP Introduction (2)

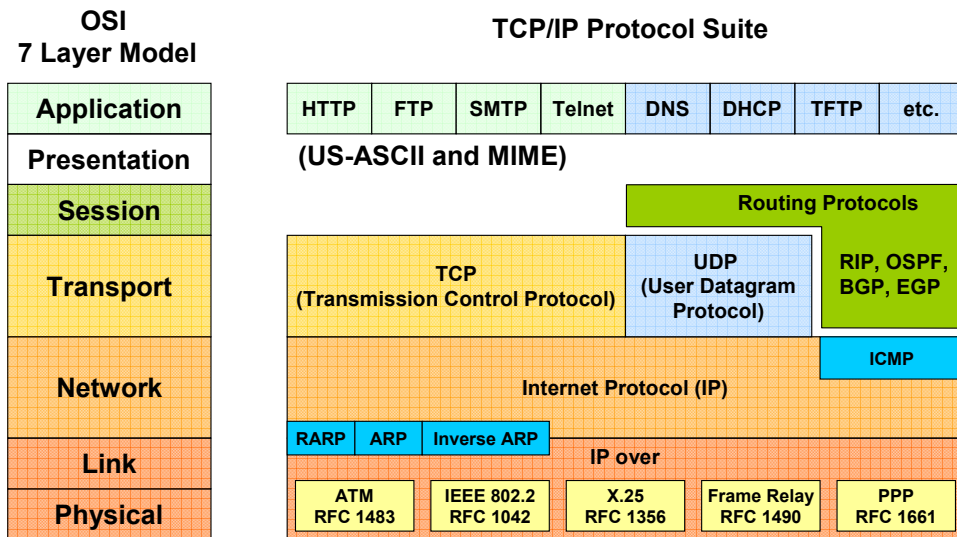


- **Shared responsibility**
 - ◆ Both network and hosts must take care for delivery (!)
 - ◆ Routers deliver datagrams to remote hosts based on IP address
 - ◆ Hosts responsible for end-to-end control
- **End-to-end control relies on TCP**
 - ◆ Layer 4

The End-to-end control is implemented in the upper Layers of the IP host, by TCP (Transmission Control Protocol - Layer 4 Protocol).

TCP is a connection oriented protocol. It takes care about flow-control, sequencing, windowing and error recovery.

IP Introduction (3)



(C) Herbert Haas 2005/03/11

8

IP is a Network-Layer (Layer 3) protocol. Packet transport, fragmentation, addressing, all this is done by IP. ICMP, also a Layer 3 Protocol, which is carried in IP is used for the PING-application. On the Transport Layer (Layer 4) you see TCP. TCP protects the IP header and takes care for reliable delivery.

IP Introduction (4)



- **IP over anything: Overlay Technique**
 - ◆ IP can be easily integrated upon layer 2 technologies
 - ◆ Open development quickly adapts to new transport and switching methods
- **End-to-end principle**
 - ◆ Only hosts must be intelligent (TCP)
 - ◆ Routers remain simple

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. And on the other hand, IP is a relative simple protocol and because of this it had been integrated in many different operating systems, most importantly UNIX.

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.

IP Introduction (5)



- **TCP cares for reliability**
 - ◆ Connection oriented
 - ◆ Error recovery
 - ◆ Flow control
 - ◆ Sequencing
- **IP is the router's language**
 - ◆ No idea about applications
 - ◆ Best effort delivery

IP knows nothing about the end system applications, it only cares about networks and host-addresses. 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. TCP also takes care of the end-to-end issues (error recovery, flow control, sequencing,...).

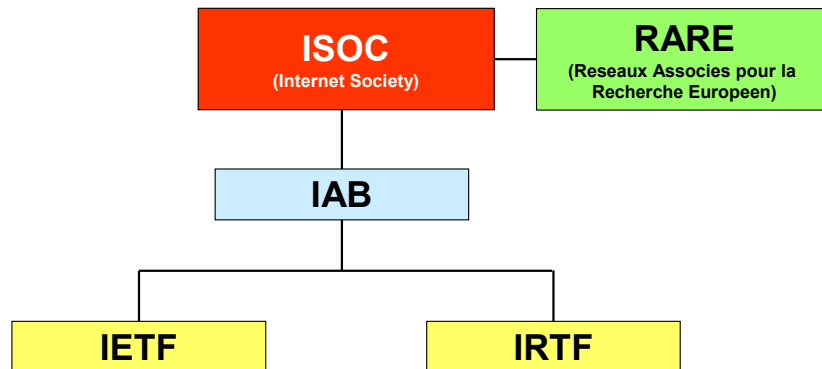
IP Introduction (6)



- **Request for Comments (RFCs)**
 - ◆ **De facto standards for the Internet**
 - ◆ **Initially posted by snail mail**
 - ◆ **IETF (Internet Engineering Task Force) reviews and confirms them**
 - ◆ **RFCs are numbered in sequence of publishing**
 - ◆ **Everybody may write an RFC (!)**

All ideas and standards of the Internet developers are maintained in so-called "Request for Comments" (RFCs) documents. The RFCs are freely available and can be downloaded from several sites, for example <http://www.ietf.org> or <http://www.rfc-editor.org>. Of course, they can also be ordered by the Network Information Center (NIC).

Internet Organizations



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).

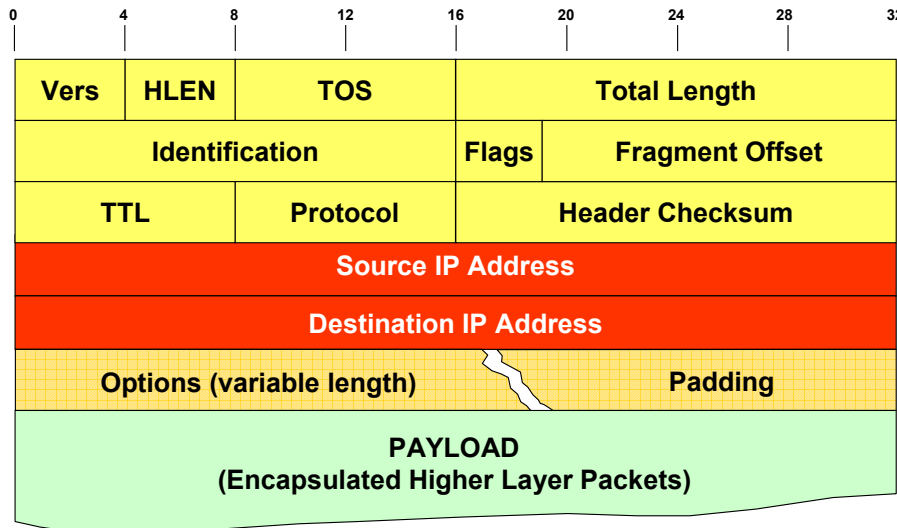
The **Reseaux Associes pour la Recherche Europeen (RARE)** was founded in 1986 to build and maintain a European high speed data network infrastructure. RARE is also a member of ISOC and ETSI (European Telecommunications Standards Institute). EBONE was initiated by RARA and RARA is a close cooperation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

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).

The IP Header



(C) Herbert Haas 2005/03/11

13

The IP header consists of the following fields:

Version (Vers): 4 Bits. This Header describe IP Version 4.

Header Length (HLEN): 4 Bits.

Type of Service (TOS): 8 Bits. TOS is a parameter, who describe the quality of transmission of the datagram through a particular network.

Total Length: 16 Bits. Is the length of the datagram including header and data.

Identification: 16 Bits. See Page 37.

Flags: 3 Bits. See Page 37.

Fragment Offset: 13 Bits. See Page 38.

Time to Live (TTL): 8 Bits. This field indicates the maximum time the datagram is allowed to remain in the system. The datagram must be destroyed, if the field contains the value zero.

Protocol: 8 Bits. Describe what protocol is used in the next level.

Header Checksum: 16 Bits. A Checksum for the Header only.

Source IP Address: 32 Bits.

Destination IP Address: 32 Bits.

Options: Variable length. The Option field can be used for security or routing information's.

Padding: Variable length. Its only used that the Internet header ends on a 32 Bit boundary.

The IP Address



■ Dotted Decimal Notation

Binary IP Address: 1100000010101000000000100000001

Decimal Value: 3232235777

Decimal Representation *per byte*:

1	1	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
192								168								1				1							

→ **192 . 168 . 1 . 1**

The IP Address is a 32 bit value in the IP header. The address identifies the access to a network.

Always keep in mind that IP addresses are basically simple numbers only. There is no natural structure in it.

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

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

The host-ID can be arbitrarily assigned by each local network manager.

IP Address Classes



- **Net-ID? Host-ID?**
- **5 Classes defined!**
 - ♦ **A (1-127)**
 - ♦ **B (128-191)**
 - ♦ **C (192-223)**
 - ♦ **D (224-239, Multicast)**
 - ♦ **E (240-254, Experimental)**
- **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.

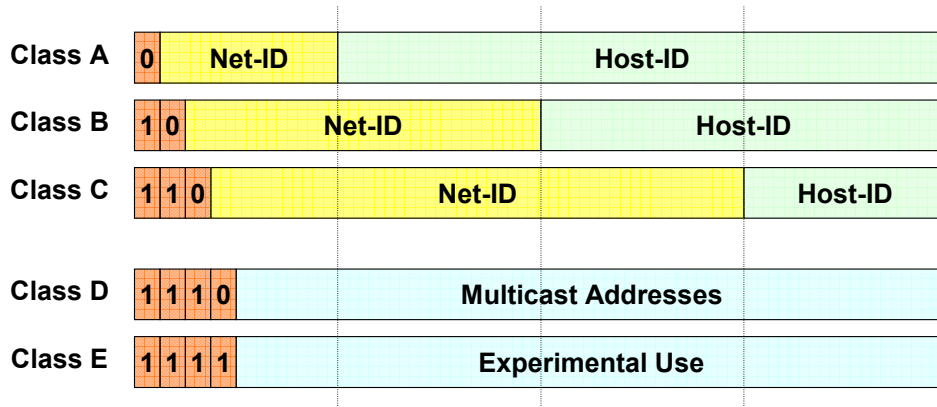
The idea of classes helps a router to decide how many bits of a given IP address identify a network number and how many bits are therefore available for host numbering. The usage of classes has a long tradition in the Internet and was a main reason for IP address depletion.

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

Broadcasts and Networks



- All ones in the host-part represents „network-broadcast“ (**10.255.255.255**)
- All ones in the net-part and host-part represents „limited broadcast in this network“ (**255.255.255.255**)
- All zeros in the host-part represents the „network-address“ (**10.0.0.0**)

A network broadcast is used to send a broadcast packet to a dedicated network. The IETF strongly discourages the use of network 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.

Reserved Addresses



- **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**
- **Network 127.x.x.x is reserved for "Loopback"**

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.

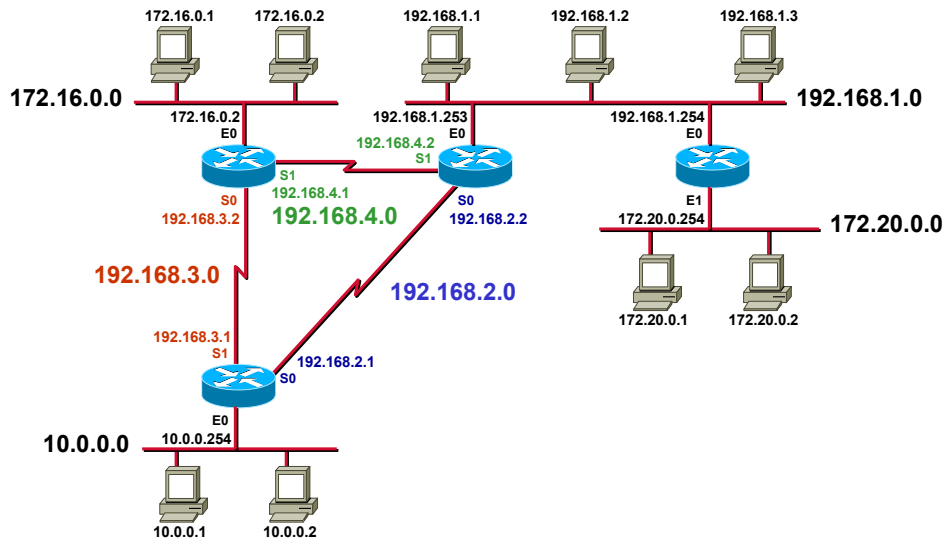
Because of those rigid filter policies, it is relatively safe to utilize RFC 1918 addresses in local networks—everybody in the Internet knows which addresses must be filtered.

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.

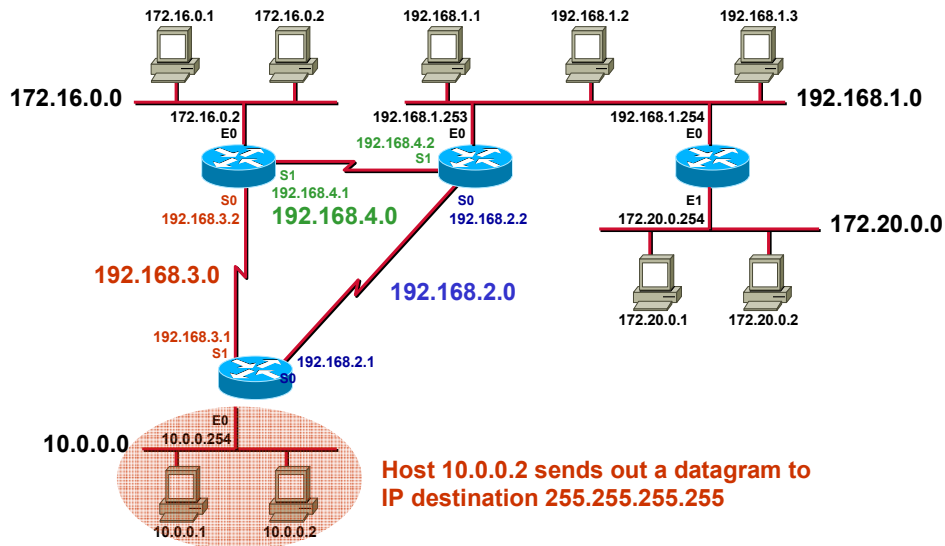
When using IP for inter-process-communication, the involved client/server processes can be distributed upon different servers across a network—without any modification of the source codes!

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

Addressing Example

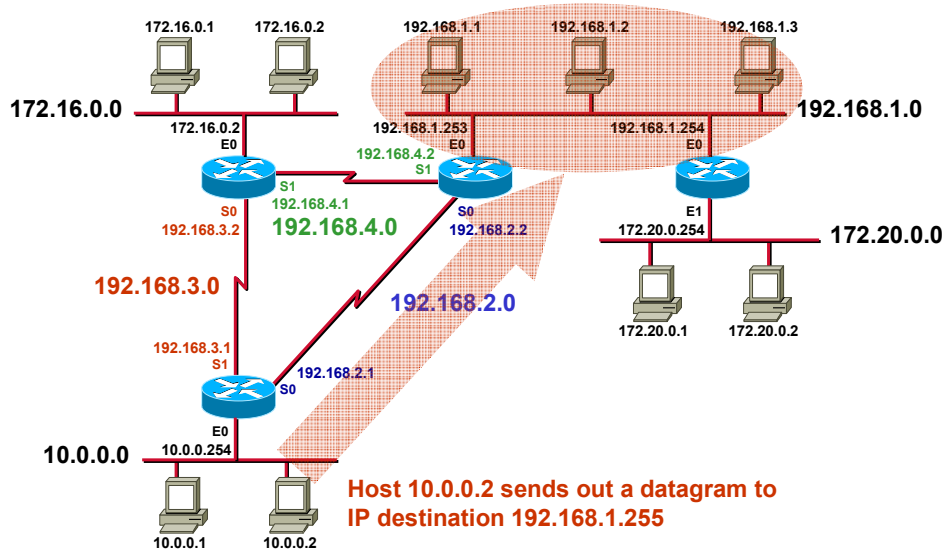


IP Limited Broadcast



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



(C) Herbert Haas 2005/03/11

21

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! (IPv6 is not discussed in this chapter)

Classful Address Waste



	Total	Allocated	Allocated %
Class A	126	48	54%
Class B	16383	7006	43%
Class C	2097151	40724	2%

Network Number Statistics, April 1992 (Source: RFC 1335)

- Two-level hierarchy was sufficient in the early days of the Internet
- The growing sizes of LANs demanded for a third hierarchical level
- "Subnetting" allows to identify some bits of the host-ID to be interpreted as "Subnet"

The "**classful**" method of identifying network-IDs of a given IP address is inflexible and lead to address space depletion. The table above shows how the total address space had been allocated by April 1992, according to RFC 1335. Note that only 2% of more than 2 million Class C addresses had been assigned. Class C networks are too small for most organizations but class A and B are too large. Of course many companies tried to grab a class A network number because of the huge address space—they would never need another IP network number anymore.

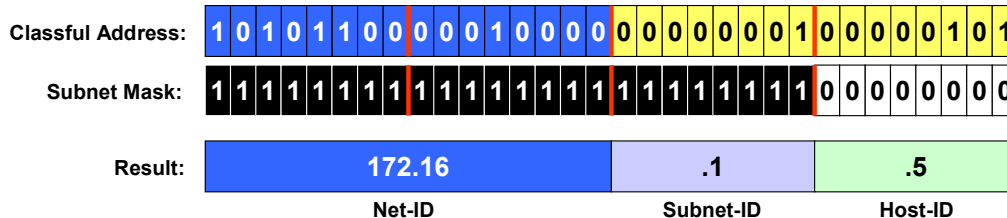
LANs were getting bigger and bigger and a logical separation of an organization's network (e. g. of a class A network number) would be a great help. Until now, multiple network numbers had been assigned to single companies, which caused two problems: waste of IP address space and growing Internet routing tables.

Even 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.

Subnetting Example



Class B Address: 172.16.1.5, Subnet Mask: 255.255.255.0



Alternative (newer) notation: 172.16.1.5 /24

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 organization's 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.

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).

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.

Subnetting has been introduced in 1985 by RFC 950.

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 RIP, relayed routes as a single 32-bit address. The high-order bits allowed each address to split into its network and host fields.

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.

The advent of subnetting undermined this scheme, 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 led to more reserved addresses - both the all-0s subnet and the all-0s host in each subnet were reserved.

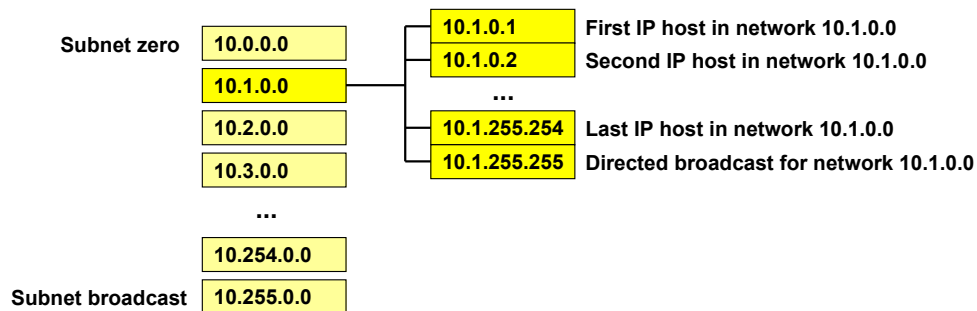
Subnet Example 1



"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.

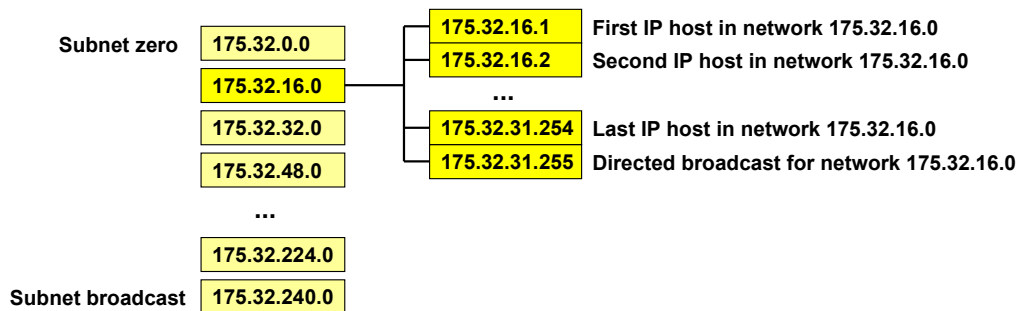
The diagram above shows the total range of subnetworks including the "forbidden" ones, that is subnet zero and the subnet broadcast.

Subnet 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:



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!**

192.168.1.64 / 26

192.168.1.4 / 30

192.168.1.32 / 27



(C) Herbert Haas 2005/03/11

27

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

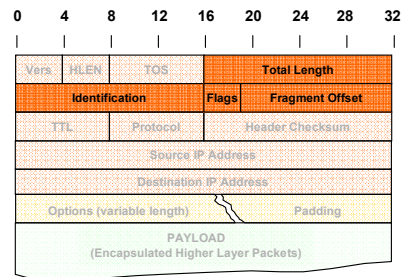
A short address design history:

1980	Classful Addressing	RFC 791
1985	Subnetting	RFC 950
1987	VLSM	RFC 1009
1993	CIDR	RFC 1517 - 1520

IP Fragmentation (1)



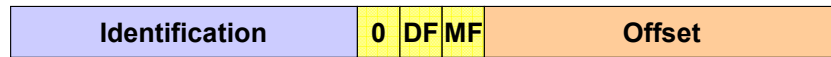
- Typical task of a Network Layer
- Used when packet length > link MTU
- 4 IP header fields are used
 - ◆ Identification
 - ◆ Flag "DF"
 - ◆ Flag "MF"
 - ◆ Fragment Offset



Fragmentation is one of the most important things of IPv4. Four fields in the IP Header are only for fragmentation. Fragmentation is performed by any router that is forwarding packets upon links with a too small MTU size. For example, if packets arriving on the Ethernet interface having MTU size of 1500 Bytes should be forwarded on a PPP link having MTU size of 576 Bytes must be fragmented when the packets are larger than the PPP MTU.

For example if IP packets are sent over different routes, fragmentation takes care that the same packet deliver correct by the host.

IP Fragmentation (2)



- **Identification**
 - ♦ Each fragment of a IP datagram must carry the same identification number
 - ♦ Necessary for reassembly
- **Flags**
 - ♦ DF (don't fragment)
 - ♦ MF (more fragments)
- **Fragment Offset**
 - ♦ Indicates the position of a fragment in the original datagram
 - ♦ Multiple of 8 octets

Identification:

No sequence number! Must be unique for the combination source / destination / protocol during the lifetime of a datagram (see also time to live).

DF:

If set: fragmentation is not allowed (except intranet fragmentation). Packets must be discarded by router if max frame size is too small.

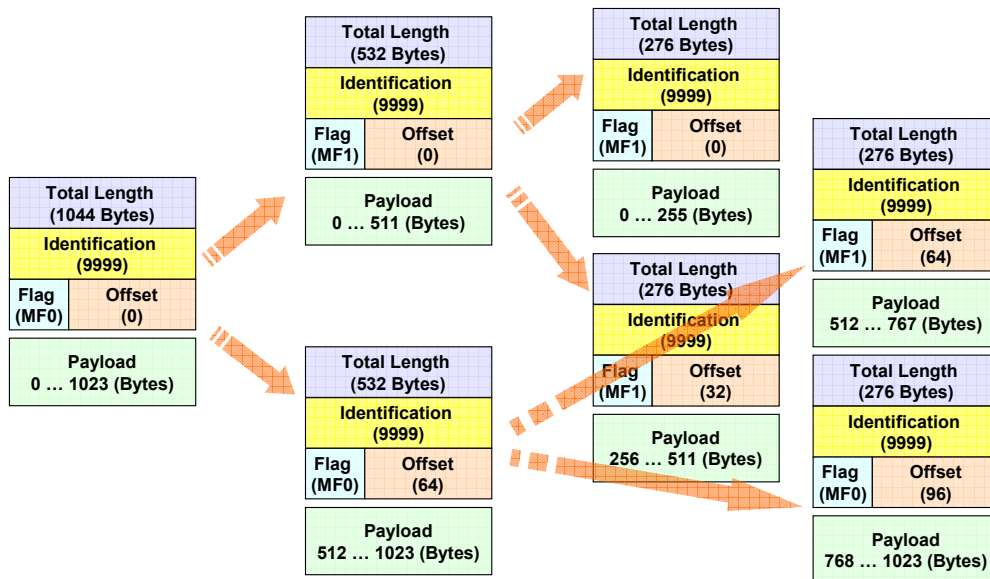
MF:

If set: more fragments of datagram follow.

Offset is measured in multiples of 8 octets (64 bits). The first fragment and unfragmented packets have an offset of 0.

Fragments with the same combination of source address, destination address, protocol number, and identification will be reassembled to a datagram.

IP Fragmentation (3)



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).

IP Fragmentation (4)



- **Reassembly is done at the destination**
 - ◆ Buffer space has to be provided at the receiver
- **The first arriving fragment issues a reassembly timer**
 - ◆ Provided that MF=1 and/or Offset \neq 0
- **The reassembly timer limits the lifetime of an incomplete datagram and allows better use of buffer resources**

Because fragments can take different paths, reassembly is done at the destination. If the reassembly timer expires before the packet was reconstructed, all fragments will be discarded and the buffer is set free.

That is: Fragmentation might be a resource- and time-consuming matter. 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.

Summary



- **The Internet Protocol**
 - ◆ Is an "open" (RFC defined) standard
- **An IP Address is a 32 bit value but structured**
- **To define net-ID and host-ID**
 - ◆ Classes A, B, C
 - ◆ Subnetting and VLSM allows to utilize the address-space much more efficient

Quiz



- **Why is there also a source address in the IP header?**
- **Why is there no field for the subnet-mask in the IP Header?**
- **Is Subnet-Zero used in "Real Life"?**
- **Do Routers today really care about IP-Classes?**
- **Is VLSM still important? (why / why not)**

Hints



- Q1:to define a way back to sender (TCP)
- Q2:hosts perform XOR calculation with own configured subnetmask
- Q3:Yes – it would mean a waste of addresses to not use it
- Q4: Not Really – they only perform XOR calculation (classless routing)
- Q5: Not Really – Private Addresses together with NAT are being used (sometimes it's used on WAN Links)