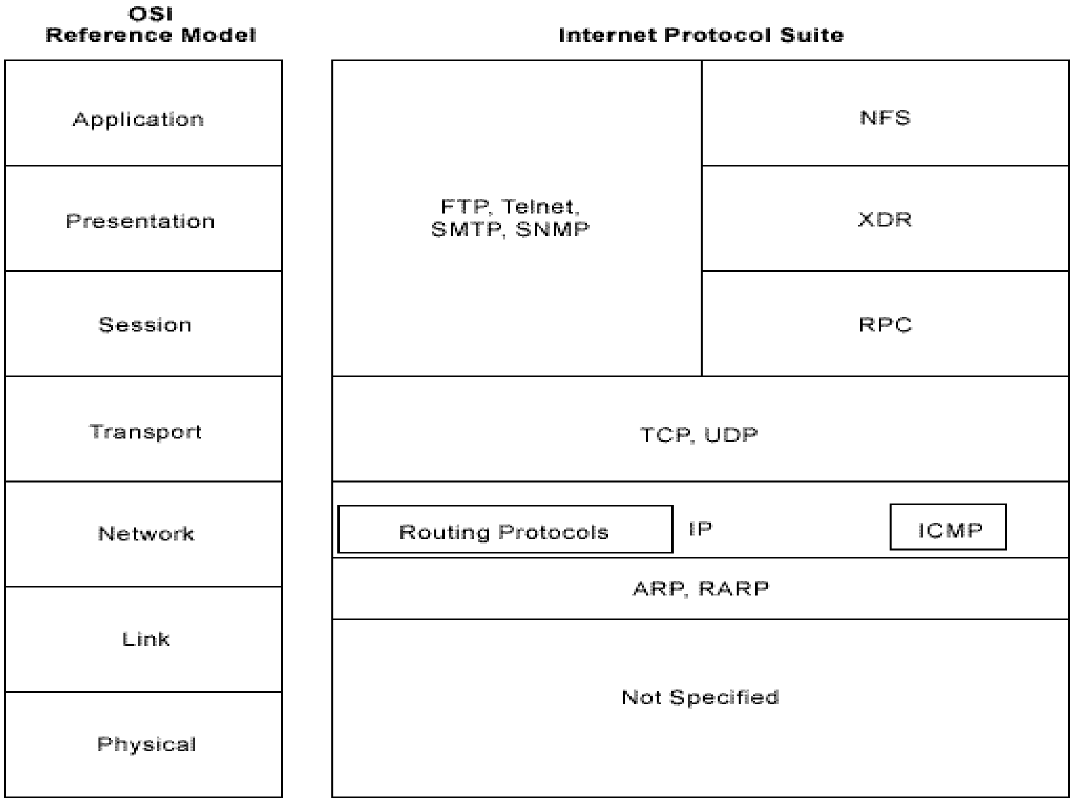
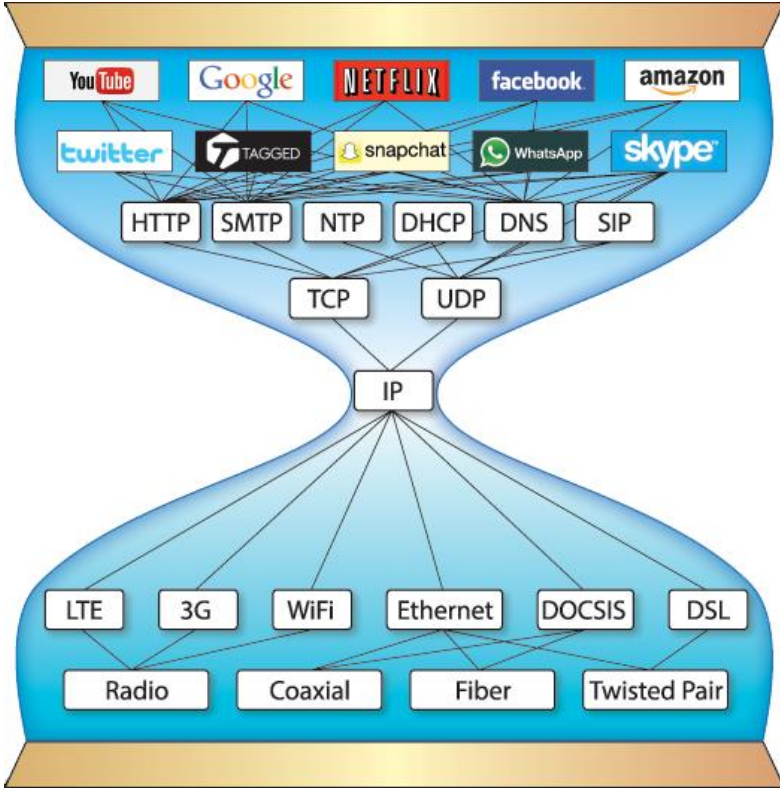
Networking Technologies and Applications

István Moldován
BME TMIT

October 16, 2019

Layering and hourglass model





IP (Internet Protocol)

- Allows any to nodes to communicate over the Internet
- The goal is to deliver a packet to the destination – no guarantees (**best effort**)
  - No guarantees for the delivery
  - No guarantees for the ordering
- The packet crosses several routers, gateways
  - Routing protocols needed
  - Packets sent towards the same destination can follow different paths
    - Packet switching vs. Circuit switching

IPv4 header

<table>
<tr><th>Octet</th><th colspan="8">0</th><th colspan="8">1</th><th colspan="8">2</th><th colspan="8">3</th></tr>
<tr><th>Bit</th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th><th>16</th><th>17</th><th>18</th><th>19</th><th>20</th><th>21</th><th>22</th><th>23</th><th>24</th><th>25</th><th>26</th><th>27</th><th>28</th><th>29</th><th>30</th><th>31</th></tr>
<tr><td>0</td><td colspan="4">Version</td><td colspan="4">IHL</td><td colspan="6">DSCP</td><td colspan="2">ECN</td><td colspan="16">Total Length</td></tr>
<tr><td>32</td><td colspan="16">Identification</td><td colspan="3">Flags</td><td colspan="13">Fragment Offset</td></tr>
<tr><td>64</td><td colspan="8">Time To Live</td><td colspan="8">Protocol</td><td colspan="16">Header Checksum</td></tr>
<tr><td>96</td><td colspan="32">Source IP Address</td></tr>
<tr><td>128</td><td colspan="32">Destination IP Address</td></tr>
<tr><td>160</td><td colspan="32" rowspan="4">Options (if IHL > 5)</td></tr>
<tr><td>192</td></tr>
<tr><td>224</td></tr>
<tr><td>256</td></tr>
</table>

- **Version** – 4 (IPv4)
- **IHL** - Internet Header Length (32 bit words)
- **DSCP** – Differentiated Services Code Point
  - Support for QoS – **Best Effort (BE), Expedited Forwarding (EF), Assured Forwarding (AF)**

IPv4 header

<table>
<tr><th>Octet</th><th colspan="8">0</th><th colspan="8">1</th><th colspan="8">2</th><th colspan="8">3</th></tr>
<tr><th>Bit</th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th><th>16</th><th>17</th><th>18</th><th>19</th><th>20</th><th>21</th><th>22</th><th>23</th><th>24</th><th>25</th><th>26</th><th>27</th><th>28</th><th>29</th><th>30</th><th>31</th></tr>
<tr><th>0</th><td colspan="4">Version</td><td colspan="4">IHL</td><td colspan="6">DSCP</td><td colspan="2">ECN</td><td colspan="16">Total Length</td></tr>
<tr><th>32</th><td colspan="16">Identification</td><td colspan="3">Flags</td><td colspan="13">Fragment Offset</td></tr>
<tr><th>64</th><td colspan="8">Time To Live</td><td colspan="8">Protocol</td><td colspan="16">Header Checksum</td></tr>
<tr><th>96</th><td colspan="32">Source IP Address</td></tr>
<tr><th>128</th><td colspan="32">Destination IP Address</td></tr>
<tr><th>160</th><td colspan="32" rowspan="4">Options (if IHL > 5)</td></tr>
<tr><th>192</th></tr>
<tr><th>224</th></tr>
<tr><th>256</th></tr>
</table>

- **ECN** – Explicit Congestion Notification
  - Packets are not dropped in case of congestion, just marked
  - The destination tells to the source to lower its sending rate
- **Total Length –** in bytes
  - Maximum packet 65.535 byte

IP fragmentation

- The packet crosses several networks during its transmission
  - Lower MTU (Maximum Transmission Unit) -> fragmentation
  - The IP header contains the fragment number
  - Reassembly of the fragments is also done by IP
- Fragmentation can be avoided
  - “Path MTU discovery”- minimum MTU on the path
  - The source sends small packets than the Path MTU

The IPv4 header

Octet	0								1								2								3							
Bit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	Version				IHL				DSCP						ECN		Total Length															
32	Identification																Flags			Fragment Offset												
64	Time To Live								Protocol								Header Checksum															
96	Source IP Address																															
128	Destination IP Address																															
160	Options (if IHL > 5)																															
192																																
224																																
256																																

- **Identification** – identifier of a fragmented IP packet
- **Fragment Offset** – the offset of the fragment, compared to the beginning of the large packet (0 for the first fragment)
- **Flags –** 3 bits to control fragmentation
  - First bit set to 0 (reserved for future use)
  - **DF – Don't Fragment bit** – if larger than the path MTU, just drop it (e.g., for Path MTU Discovery)
  - **MF – More Fragments bit** – more fragments will come (1 if the last fragment, otherwise 0)

IPv4 header

Octet	0								1								2								3							
Bit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	Version				IHL				DSCP						ECN		Total Length															
32	Identification																Flags			Fragment Offset												
64	Time To Live								Protocol								Header Checksum															
96	Source IP Address																															
128	Destination IP Address																															
160	Options (if IHL > 5)																															
192																																
224																																
256																																

- **Time To Live** – limits the spreading of the packet
  - Each router decreases it with 1, before forwarding. If it reaches 0, the packet is dropped
- **Protocol** – Which protocol generated the payload
  - ICMP (1), IGMP (2), TCP (6), EGP (8), IGP (9), UDP (17), IPv6 (41), RSVP (46), OSPF (89)

IPv4 header

Octet	0								1								2								3							
Bit	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	Version				IHL				DSCP						ECN		Total Length															
32	Identification																Flags			Fragment Offset												
64	Time To Live								Protocol								Header Checksum															
96	Source IP Address																															
128	Destination IP Address																															
160	Options (if IHL > 5)																															
192																																
224																																
256																																

- **Header Checksum** – controls only if the header is correct
  - If an error in the payload, that should be handled by the encapsulated protocol
  - As the TTL is decreased, each router should recalculate the checksum, and refresh this field accordingly
- **Options** – rarely used (as opposed to IPv6)

What is an IP Address?

- An IP address is a unique global address for a network interface

- An IP(v4) address:

  - is a **32 bit long** identifier

  - encodes a network number (**network prefix**) and a **host number**

Dotted Decimal Notation

- IP addresses are written in a so-called ***dotted decimal* notation**
- Each byte is identified by a decimal number in the range [0..255]:

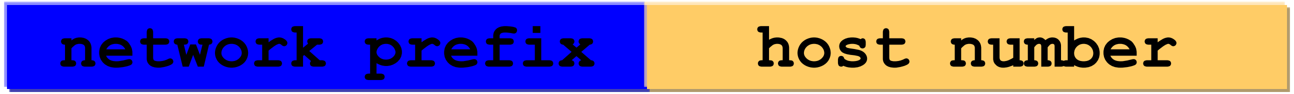
- **Example:**

10000000	10001111	10001001	10010000
1st Byte	2nd Byte	3rd Byte	4th Byte
= 128	**= 143**	**= 137**	**= 144**

128.143.137.144

Network prefix and Host number

- The network prefix identifies a network and the host number identifies a specific host (actually, interface on the network).



- **How do we know how long the network prefix is?**
  - The network prefix **used** to be implicitly defined (**class-based addressing, A,B,C,D…**)
  - The network prefix now is flexible and is indicated by a **prefix**/**netmask (classless).**

Example

Example: argon.cs.virginia.edu

- IP address is 128.143.137.144
  - Is that enough info to route datagram??? -> No, need netmask or prefix at every IP device (host and router)
- Using Prefix notation IP address is: **`128.143.137.144/16`**
  - Network prefix is 16 bits long

- Network mask is: `255.255.0.0` or hex format: `ffff0000`

-----> Network id (IP address **AND** Netmask) is: `128.143.0.0`

-----> Host number (IP address **AND** inverse of Netmask) is: `137.144`

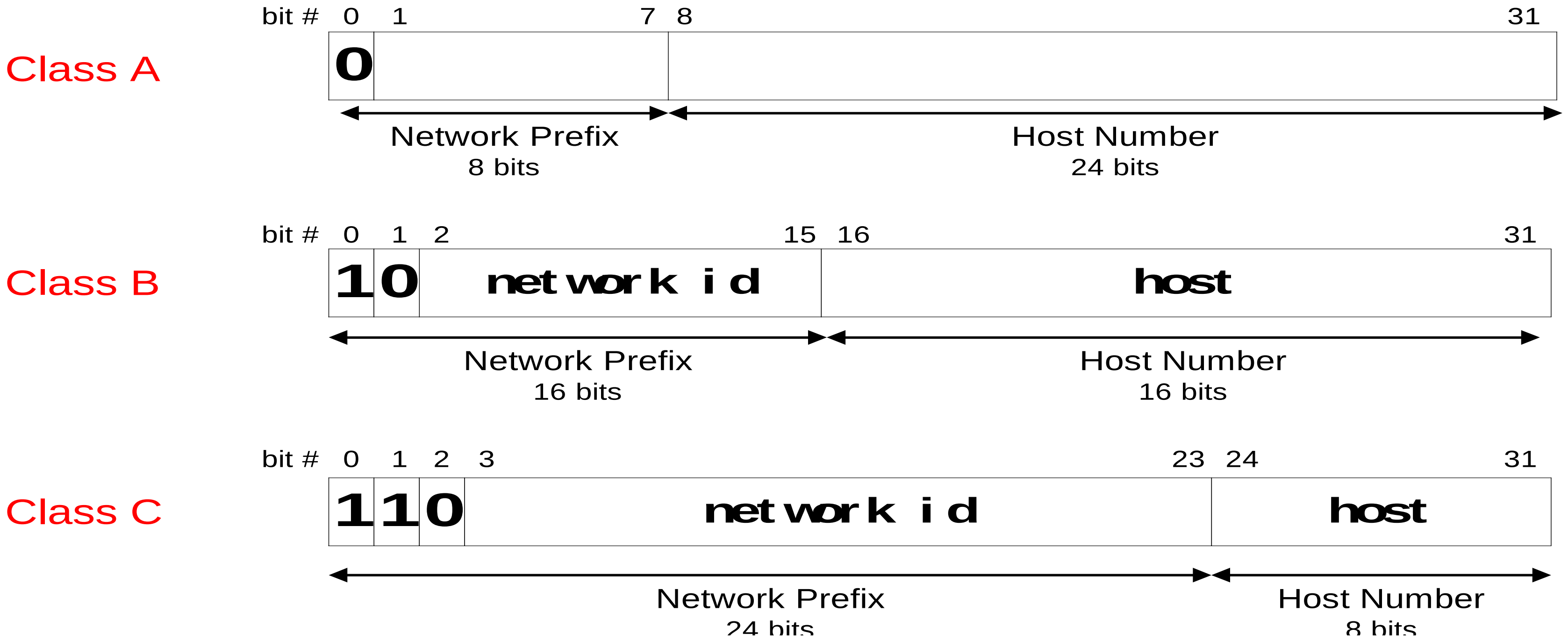
128.143	137.144

The old way: Classful IP Adresses

- When Internet addresses were standardized (early 1980s), the Internet address space was divided up into classes:
  - **Class A:** Network prefix is 8 bits long
  - **Class B:** Network prefix is 16 bits long
  - **Class C:** Network prefix is 24 bits long
- Each IP address contained a key which identifies the class:
  - **Class A:** IP address starts with "0"
  - **Class B:** IP address starts with "10"
  - **Class C:** IP address starts with "110"

	Number of networks	**Maximum nr. of hosts on a network**	**Value of first byte**
Class A	126	16,777,214	1 – 126
Class B	16,384	65,534	128 – 191
Class C	2,097,152	254	192 - 223

The old way: Internet Address Classes



The old way: Internet Address Classes

Class D

bit #	0	1	2	3	4 … 31
	1	1	1	0	multicast group id

Class E

bit #	0	1	2	3	4	5 … 31
	1	1	1	1	0	(reserved for future use)

- We will learn about multicast addresses later in this course.

Addressing rules

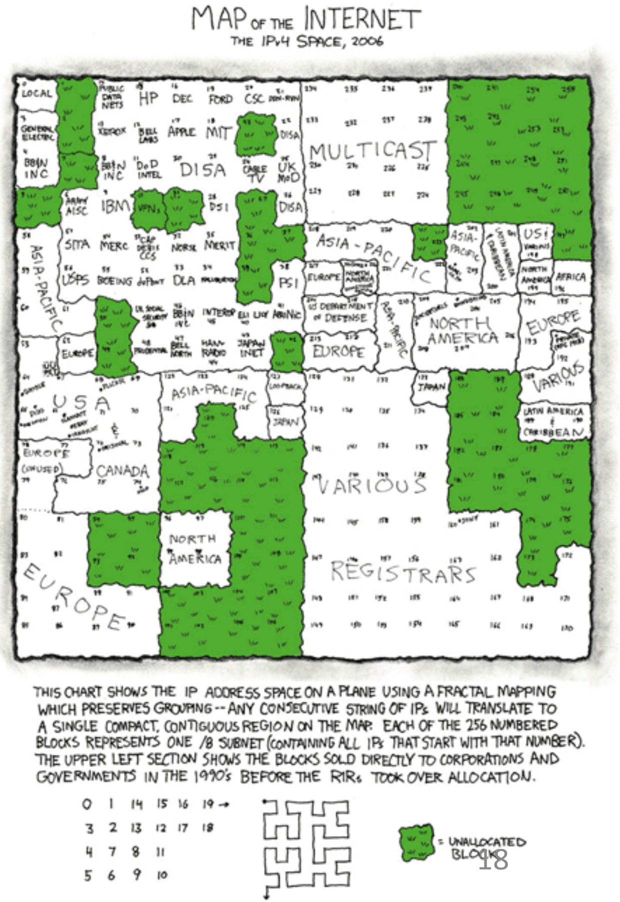
- The Network ID cannot be 127
  - Reserved for the loop-back interface
- The host ID cannot be 255
  - 255 a broadcast address
- The host ID cannot be 0
  - 0 means „this network”
- The host ID has to unique on the given network

Problems with Classful IP Addresses

- The original classful address scheme had a number of problems

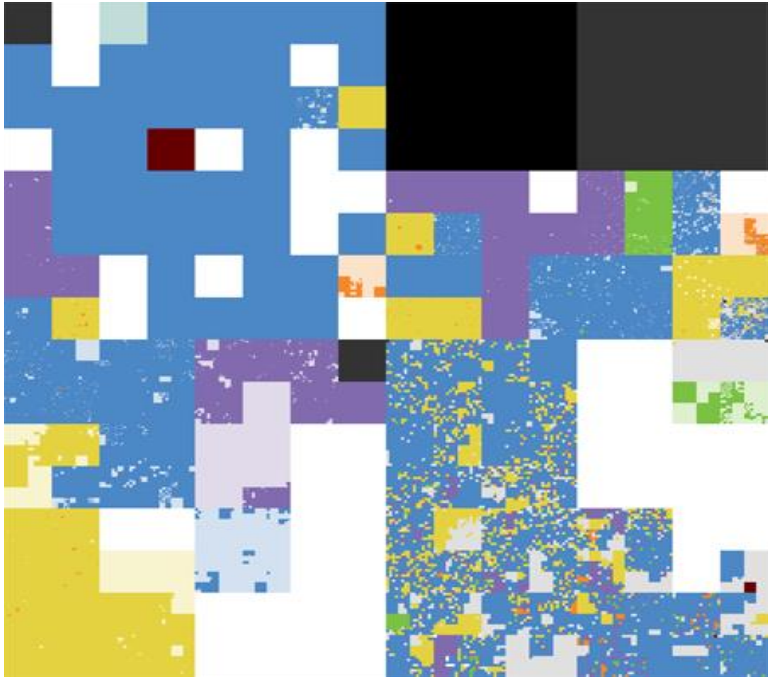
Problem 1. Too few network addresses for large networks

- Class A and Class B addresses are gone
- Initially given to institutions
  - Upper left corner
  - HP, Apple, MIT, IBM, Ford, etc
- Later RIRs are created
  - Regional Internet Registrar



IPv4 addresses (2006)

- Blue: ARIN – North America
- Yellow: RIPE NCC – Europe
- Magenta: APNIC – Asia-Pacific
- Green: LACNIC – Latin-America
- Orange: AfriNIC – Africa
- Black: Multicast
- Grey: Special addresses
  - Loopback, private, class E, etc.
- White: free



Problems with Classful IP Addresses

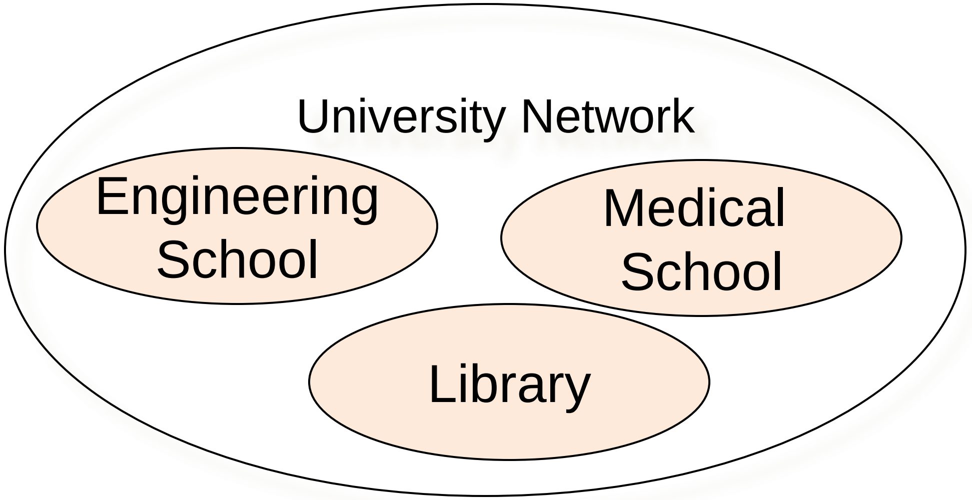
- The original classful address scheme had a number of problems

Problem 2. Two-layer hierarchy is not appropriate for large networks with Class A and Class B addresses.

  - **Fix #1: Subnetting**

Subnetting

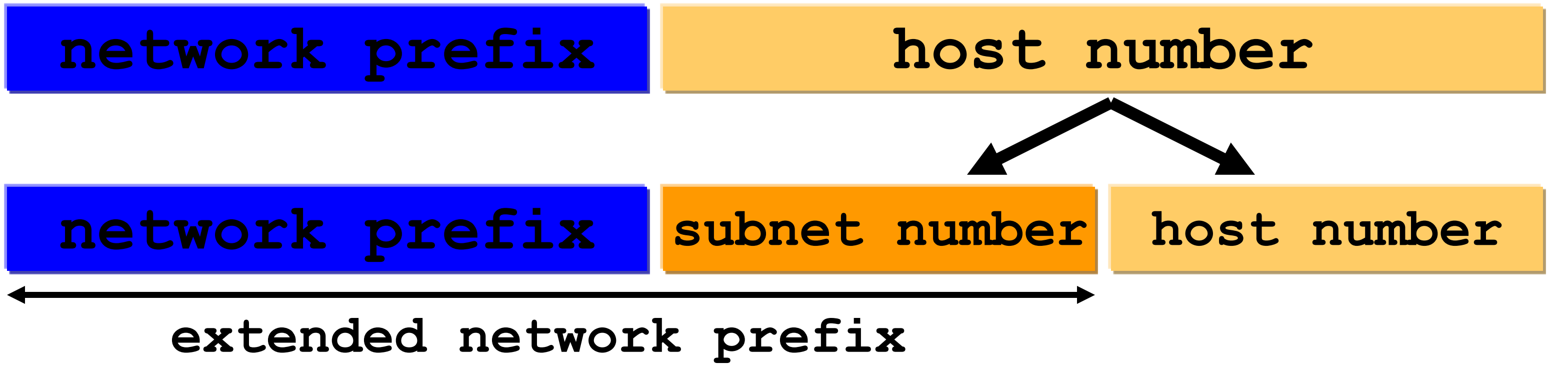
- **Problem**: Organizations have multiple networks which are independently managed
  - **Solution 1:** Allocate an address for each network
    - Difficult to manage
    - From the outside of the organization, each network must be addressable, must have an identifiable address.
  - **Solution 2:** Add another level of hierarchy to the IP addressing structure



→ **Subnetting**

Basic Idea of Subnetting

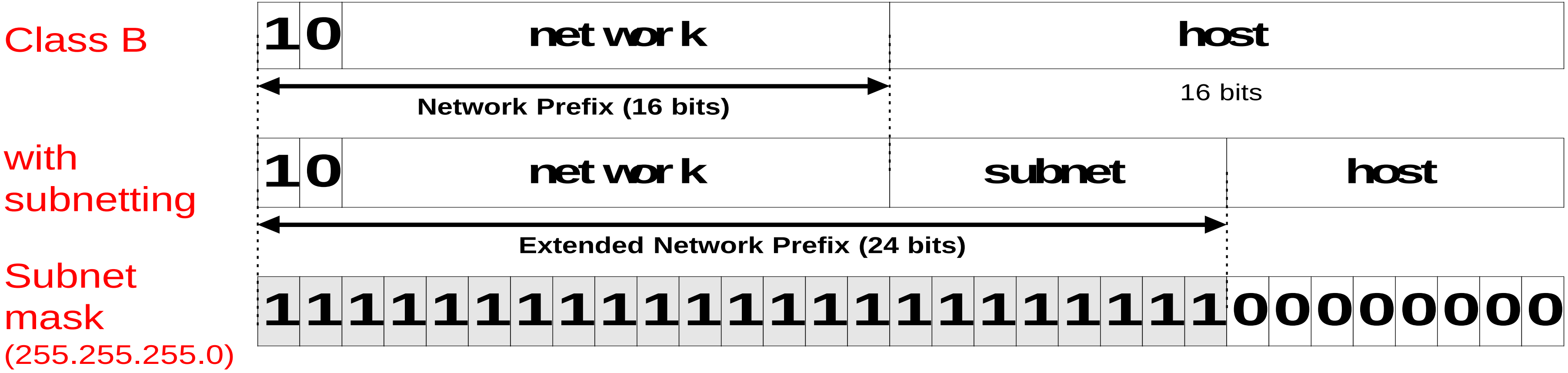
- Split the host number portion of an IP address into a **subnet number** and a (smaller) **host number**.
- Result is a 3-layer hierarchy



- Then:
  - Subnets can be freely assigned within the organization
  - Internally, subnets are treated as separate networks
  - Subnet structure is not visible outside the organization

Subnet Masks

- Routers and hosts use an **extended network prefix** (**subnet mask)** to identify the start of the host numbers



* There are different ways of subnetting. Commonly used netmasks for university networks with /16 prefix (Class B) are 255.255.255.0 and 255.255.0.0

Advantages of Subnetting

- With subnetting, IP addresses use a 3-layer hierarchy:
  - Network
  - Subnet
  - Host
- Improves efficiency of IP addresses by not consuming an entire address space for each physical network.
- Reduces router complexity. Since external routers do not know about subnetting, the complexity of routing tables at external routers is reduced.

- Note: Length of the subnet mask need not be identical at all subnetworks.

Subnetting Example

- An organization with 4 departements has the following IP address space: 10.2.22.0/23. As the systems manager, you are required to create subnets to accommodate the IT needs of 4 departments. The subnets have to support to 200, 61, 55, and 41 hosts respectively. What are the 4 **subnet network** numbers?

- Solution:
  - 10.2.22.0/24 (256 addresses > 200)
  - 10.2.23.0/26 (64 addresses >61)
  - 10.2.23.64/26 (64 addresses > 55)
  - 10.2.23.128/26 (64 addresses > 41)

Problems with Classful IP Addresses

Problem 3. Inflexible. Assume a company requires 2,000 addresses

- Class A and B addresses are overkill
- Class C address is insufficient (requires 8 Class C addresses)

Problem 4: Exploding Routing Tables: Routing on the backbone Internet needs to have an entry for each network address. In 1993, the size of the routing tables started to outgrow the capacity of routers.

Fix #2 (to both of these problems): Classless Interdomain Routing (CIDR)

CIDR - Classless Interdomain Routing

- **Goals:**
  - Restructure IP address assignments to increase efficiency
  - Hierarchical routing aggregation to minimize route table entries

Key Concept: The length of the network id (prefix) in IP addresses is arbitrary/flexible and is defined by the network hierarchy.

- Consequence:
  - Routers use the IP address **and** the length of the prefix for forwarding.
  - All advertised IP addresses must include a prefix

CIDR Example

- CIDR notation of a network address:

  192.0.2.0/18

  - "18" says that the first 18 bits are the network part of the address
- The network part is called the network prefix
- Example:
  - Assume that a site requires an IP network domain that can support 1000 IP host addresses
  - With CIDR, the network is assigned a continuous block of 1024 = 2^{10} (>1000) addresses with a 32-10 = 22-bit long prefix

CIDR: Prefix Size vs. Host Space

CIDR Block Prefix	# of Host Addresses
/27	32 hosts
/26	64 hosts
/25	128 hosts
/24	256 hosts
/23	512 hosts
/22	1,024 hosts
/21	2,048 hosts
/20	4,096 hosts
/19	8,192 hosts
/18	16,384 hosts
/17	32,768 hosts
/16	65,536 hosts
/15	131,072 hosts
/14	262,144 hosts
/13	524,288 hosts

CIDR and Address assignments

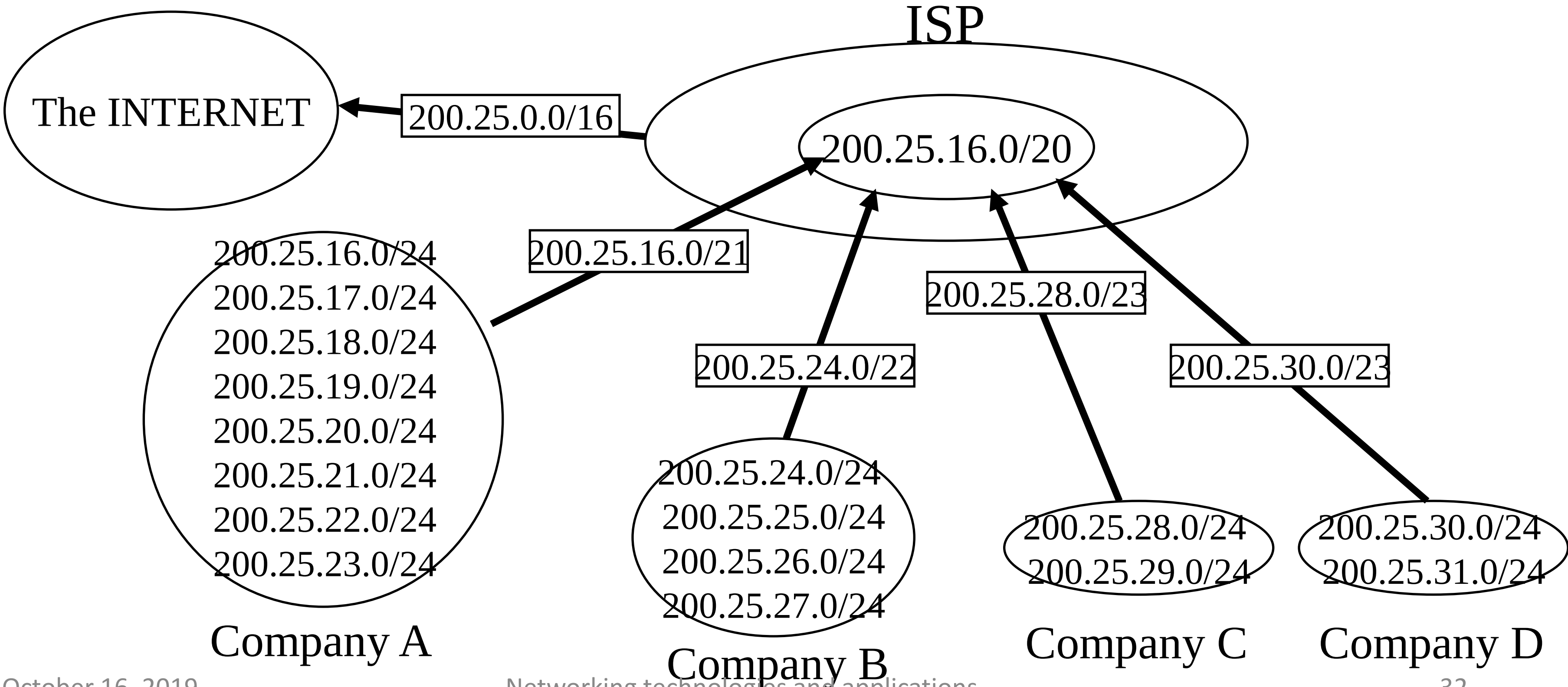
- IANA – Internet Assigned Numbers Authority
  - The RIRs get short prefix CIDR blocks
    - Regional Internet Registries
  - E.g., 62.0.0.0/8 assigned to RIPE NCC
    - Réseaux IP Européens Network Coordination Centre
- RIRs fragment and redistribute parts of the address space
  - Backbone ISPs obtain large blocks of IP address space and then reallocate portions of their address blocks to their customers.

Example:

- Assume that an ISP owns the address block 206.0.64.0/18, which represents 16,384 ($2^{32-18}=2^{14}$) IP host addresses
- Suppose a client requires 800 host addresses
  - ➢ $512=2^9<800<1024=2^{10}$ -> 32-10 = 22,
  - ➢ Assigning a /22 block, i.e., 206.0.68.0/22 -> gives a block of 1,024 (2^{10}) IP addresses to client.

01000100

CIDR example



CIDR and Routing

- **Aggregation** of routing table entries:
  - 128.143.0.0/16 and 128.142.0.0/16 can be represented as 128.142.0.0/15 at a router.
    - 143 = 128.10001111.0.0 142 = 128.10001110.0.0
- **Longest prefix match**: Routing table lookup finds the routing entry that matches the longest prefix
  - Why?

E.g., What is the outgoing interface for destination IP address: 128.143.137.0?

Prefix	Interface/outgoing link
128.143.128.0/17	interface #1
128.128.0.0/9	interface #2
128.0.0.0/4	interface #5

Routing table

Problems with Classful IP Addresses

Problem 5. The Internet is going to outgrow the 32-bit addresses

- **Fix #3: IP Version 6**