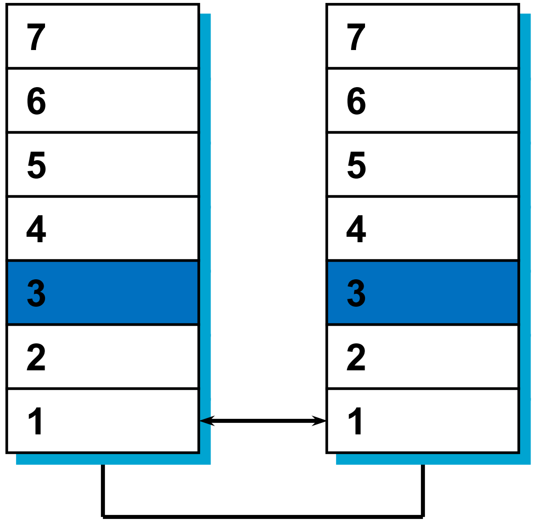
Operating Systems & Computer Networks
11. Internet Layer - Network Layer

Dr. Larissa Groth
Computer Systems & Telematics (CST)



Roadmap

8. Networked Computer & Internet
9. Network Access Layer I – Physical Layer
10. Network Access Layer II – Data Link Layer
11. **Internet Layer – Network Layer**
12. Transport Layer
13. Applications

Lernziele I

- Sie nennen:
  - welche Geräte auf welchen Schichten des TCP/IP-Protokollstacks arbeiten
  - was unter einer MAC-Adresse zu verstehen ist
  - was unter einer IP-Adresse zu verstehen ist
  - was unter einem "Autonomen System" zu verstehen ist
  - die wesentlichen Aufgaben dieser Schicht und die Ein- und Ausgabe über die Service Access Points
  - für welche Spezialfälle bestimmte IP-Adressen reserviert sind

- Sie stellen dar:
  - wie Repeater digitale Signale auf Physical Layer regenerieren und verstärken
  - wie ein Switch durch Backward Learning-Verfahren die Zuordnung zwischen Ports und MAC-Adressen lernt
  - warum die Unterscheidung in Inter- und Intra-Domain-Routing notwendig ist
  - wie ein Router durch Open Shortest Path First mittels Algorithmus von Dijkstra zum Finden kürzester Wege die Weiterleitung von Paketen innerhalb eines Autonomen Systems durchführt
  - wie ein Router durch Border Gateway Protocol die Weiterleitung von Paketen zwischen Autonomen Systemen durchführt
  - wie durch das Address Resolution Protocol die Ziel-MAC-Adresse für die Weiterleitung eines Pakets anhand einer Ziel-IP-Adresse gefunden wird
  - wie mittels Classless Inter Domain Routing der Netzwerk- und Host-Teil spezifiziert werden kann

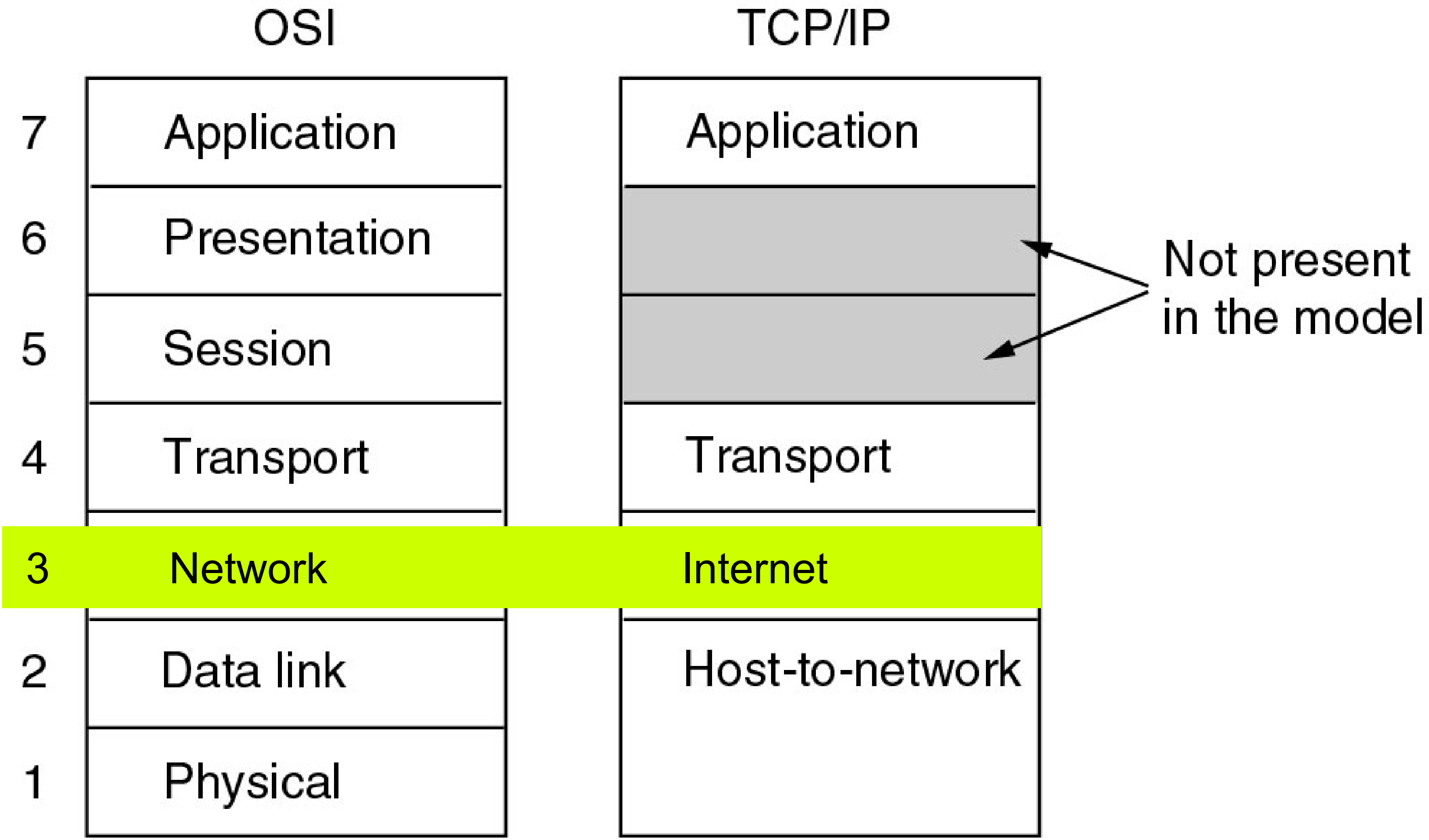
Lernziele II

- Sie wenden Verfahren auf konkrete Eingaben an:
  - Backward Learning-Algorithmus zur Weiterleitung von Paketen durch einen Switch anhand von MAC-Adressen
  - Open Shortest Path First-Algorithmus zur Weiterleitung von Paketen durch einen Router in einem Autonomen System anhand von IP-Adressen
  - Finden der Ziel-MAC-Adresse mittels Address Resolution Protocol für die Weiterleitung von IP-Paketen

- Sie argumentieren:
  - warum das weltweite Routing nicht auf Basis von MAC-Adressen erfolgen kann

- Sie untersuchen:
  - zu welchen Netzen in Classless Inter Domain Routing-Notation gegebene IP-Adressen gehören

Physical Layer



Reasons for Multiple Networks

Limited number of users/throughput in a single network

Historical reasons:

- Different groups started out individually setting up networks
- Usually heterogeneous

Geographic distribution of different groups over different buildings, campus, …

- Impractical/impossible to use a single network because of distance
  - Most MAC protocols set maximum segment length for medium access, e.g., CSMA/CD
- Long round-trip delay will negatively influence performance

Reliability

- Don’t put all your eggs into one basket
- “Babbling idiot” problem (isolation of errors)

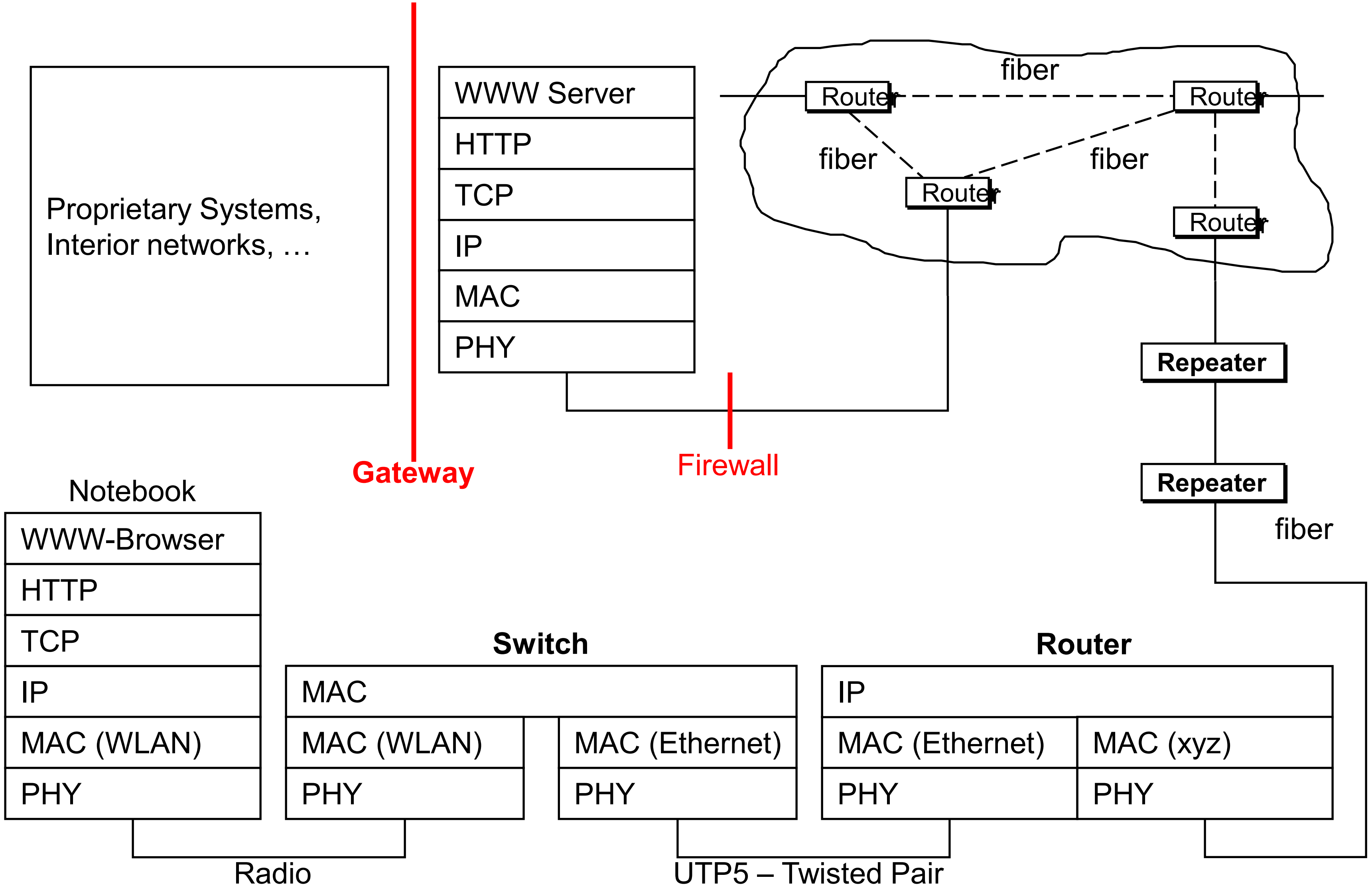
Security

- Contain possible damage caused by promiscuous operation

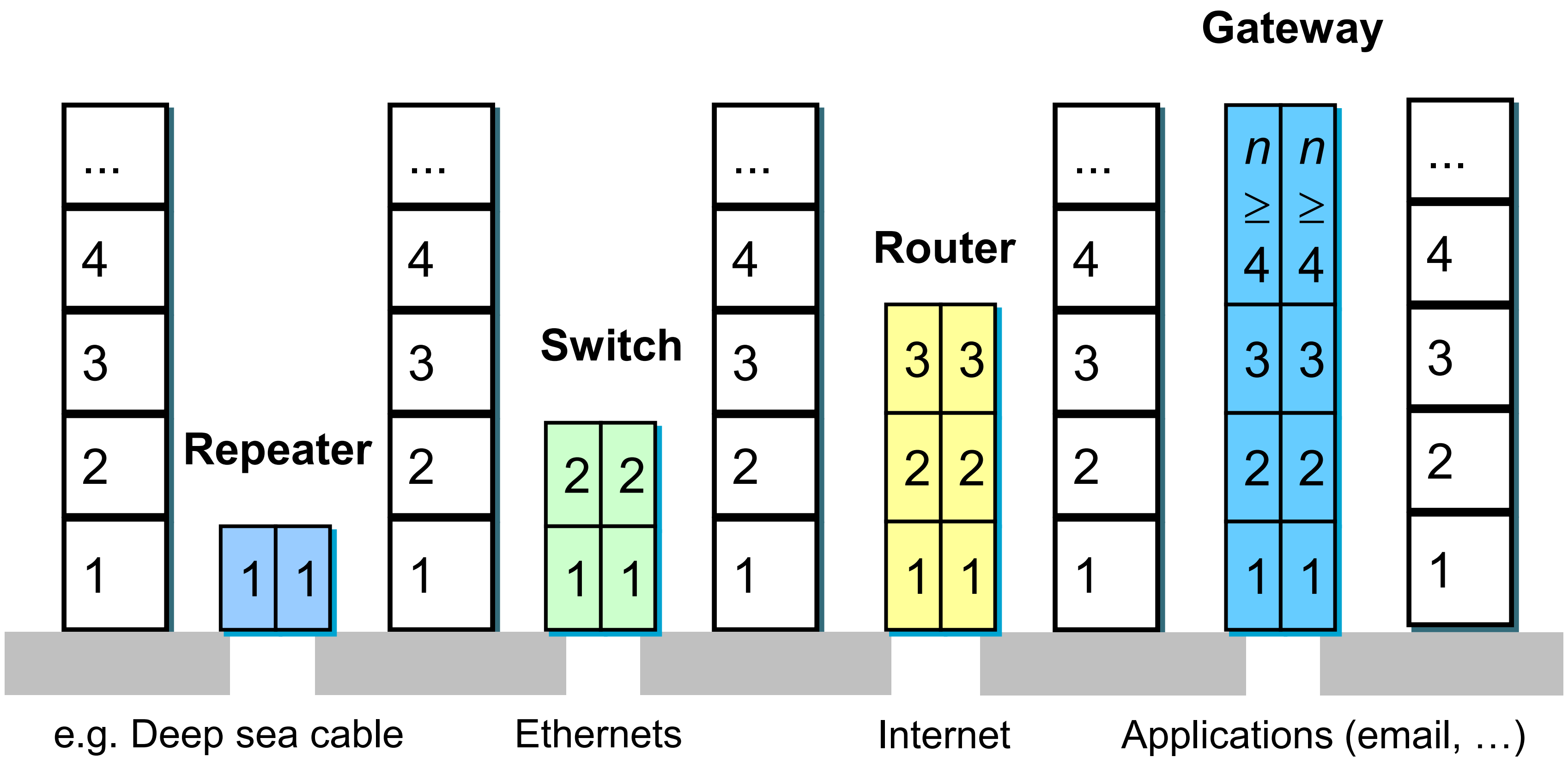
Political / business reasons

- Different authorities, policies, laws, levels of trust, …

Internetworking Units



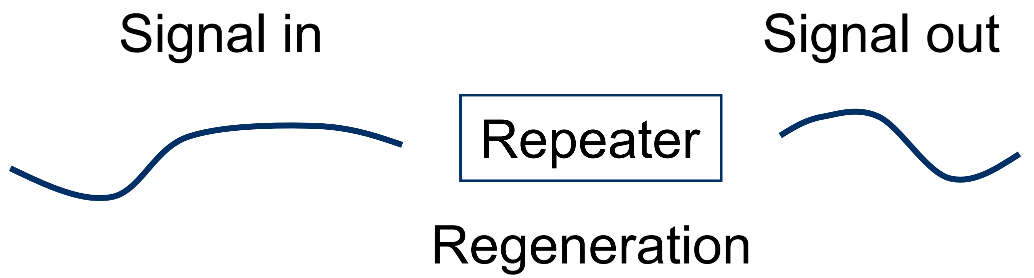
Internetworking Units

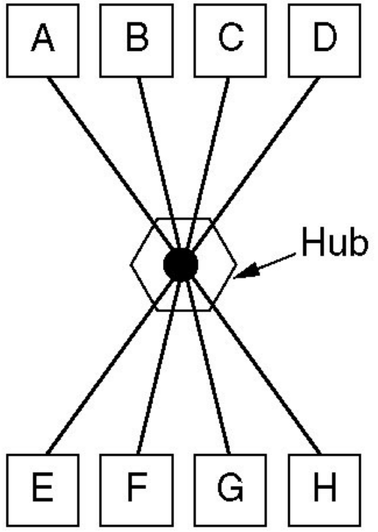


Repeater / Hub

Simplest option: Repeater

- Physical layer device, connected to two or more cables
- Amplifies/regenerates arriving signal, puts on other cables
  - Combats attenuation
  - ➢ Signal encodes data (represented by bits)
    - Can be regenerated
    - Opposed to only amplified (which would also amplify noise)
    - ➢ Analog vs. digital transmission
- Neither understands nor cares about *content (bits)* of packets





Problems of Physical Layer Solutions

Physical layer devices, e.g. repeater or hub, do not solve the more interesting problems

- E.g. no mechanism for handling load, scalability, ...

Some knowledge of data link layer structure is necessary

- Ability to understand/inspect content of packets/frames and do something with that knowledge

➢ Link-layer devices:

- Switch: Interconnect several terminals
- Bridge: Interconnect several networks (of different type)

➢ Nowadays terms sometimes used interchangeably

IEEE 802.3/Ethernet Frame Format

Common frame format:

[byte]

PR	SD	DA	SA	VLAN	Type	Data	PAD	FCS
7	1	6	6	4	2	≤ 1500	≤ 46	4

64 - 1518 (1522 VLAN)

PR: Preamble for synchronization

SD: Start-of-frame Delimiter

DA: Destination MAC Address

SA: Source MAC Address

VLAN: VLAN tag (if present), 0x8100, 3 bit priority, 12 bit ID (cf. chapter 12)

Type: Protocol type of payload (length if ≤ 0x0600), e.g. 0x0800 for IPv4, 0x86DD for IPv6

Data: Payload, max. 1500 byte

PAD: Padding, required for short frames

FCS: Frame Check Sequence, CRC32

Switch

Used to connect several terminals or networks

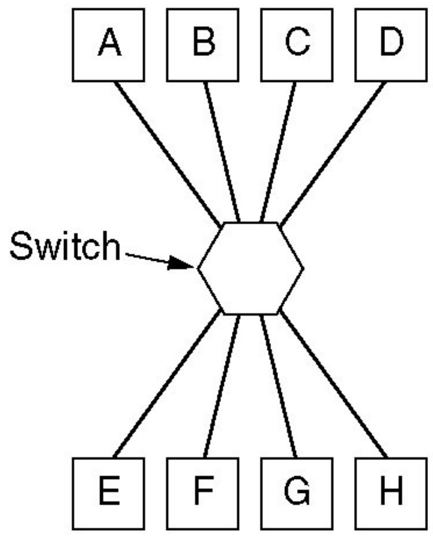
Switch inspects arriving packet’s MAC addresses and forwards it *only* on correct cable/port

- Does not bother other terminals
- Requires data buffer and knowledge *on which* port which terminal is connected
  - Mapping function of MAC address to port

➢How to obtain knowledge about network topology?

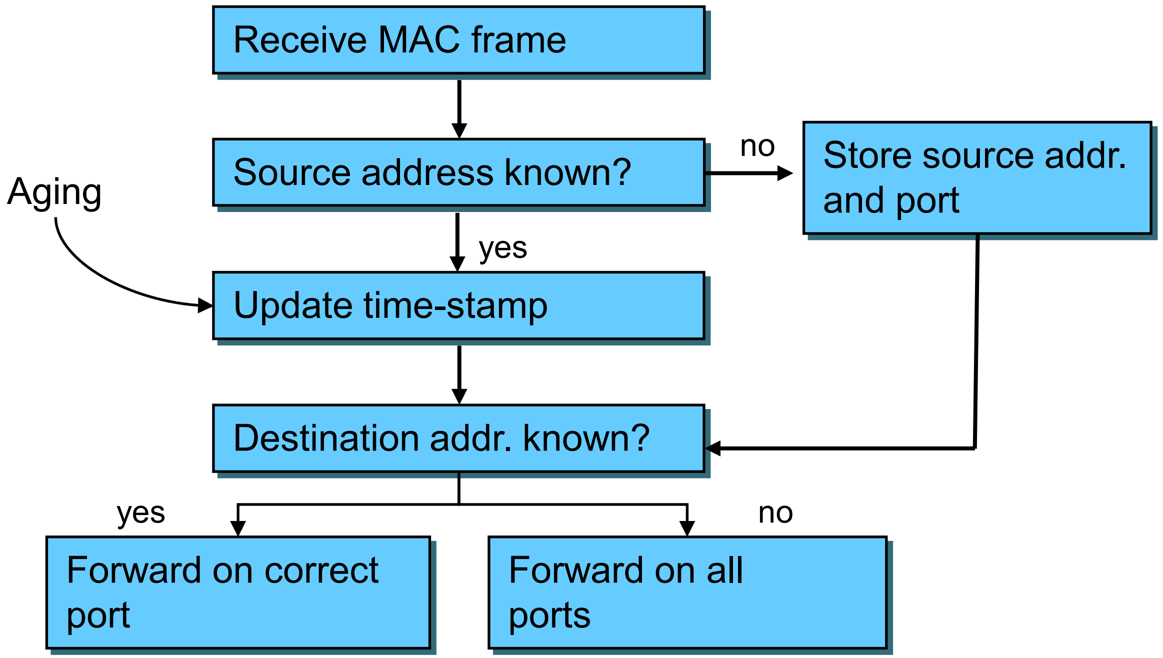
- Observe *from* where packets come to decide how to reach sending terminal

➢ *Backward learning*



Backward Learning – Algorithm

1. Learn address/port mapping from incoming packets
   - Remove expired entries (aging)
2. Forward based on knowledge about destination address
   1. Destination address is known
      Forward on correct port
   2. Destination address is unknown
      Forward on all ports
      - ➢ Only correct receiver will process frame, others will drop it



Routers

All devices so far either ignored addresses (repeaters, hubs) or worked on MAC-layer addresses (switches, bridges)

For interconnection outside a single LAN or connection of LANs, these simple addresses are insufficient

- Unstructured, “flat” addresses do not scale
  - All forwarding devices would need a list of *all* addresses
- Structured network topologies do not scale
  - World-wide spanning tree is unfeasible

➢Need more sophisticated addressing structure and devices that operate on it

- Routers and routing
- E.g. based on Internet Protocol (IP) addresses

Example: Route to NASA

```
Z:\>tracert www.nasa.gov

Tracing route to www.nasa.gov.speedera.net [213.61.6.3]
over a maximum of 30 hops:

  1    <1 ms    <1 ms    <1 ms  router-114.inf.fu-berlin.de [160.45.114.1]
  2    <1 ms    <1 ms    <1 ms  zedat.router.fu-berlin.de [160.45.252.181]
  3     1 ms    <1 ms    <1 ms  ice.spine.fu-berlin.de [130.133.98.2]
  4     1 ms    <1 ms    <1 ms  ar-fuberlin1.g-win.dfn.de [188.1.33.33]
  5     1 ms    <1 ms    <1 ms  cr-berlin1-po5-0.g-win.dfn.de [188.1.20.5]
  6     9 ms     9 ms     9 ms  cr-frankfurt1-po9-2.g-win.dfn.de [188.1.18.185]
  7    10 ms     9 ms     9 ms  ir-frankfurt2-po3-0.g-win.dfn.de [188.1.80.38]
  8    10 ms     9 ms     9 ms  DECIX.fe0-0-guy-smiley.FFM.router.COLT.net
                                           [80.81.192.61]
  9    10 ms     9 ms     9 ms  ir1.fra.de.colt.net [213.61.46.70]
 10    11 ms    10 ms     9 ms  ge2-2.ar06.fra.DE.COLT-ISC.NET [213.61.63.8]
 11    11 ms    10 ms    10 ms  213.61.4.141
 12    11 ms    10 ms    10 ms  h-213.61.6.3.host.de.colt.net [213.61.6.3]

Trace complete.
```

Not all addresses can be resolved to names (see DNS)

Some requests are redirected to Content Delivery Networks

Some nodes simply don’t answer…

Example: Route to NASA (redone)

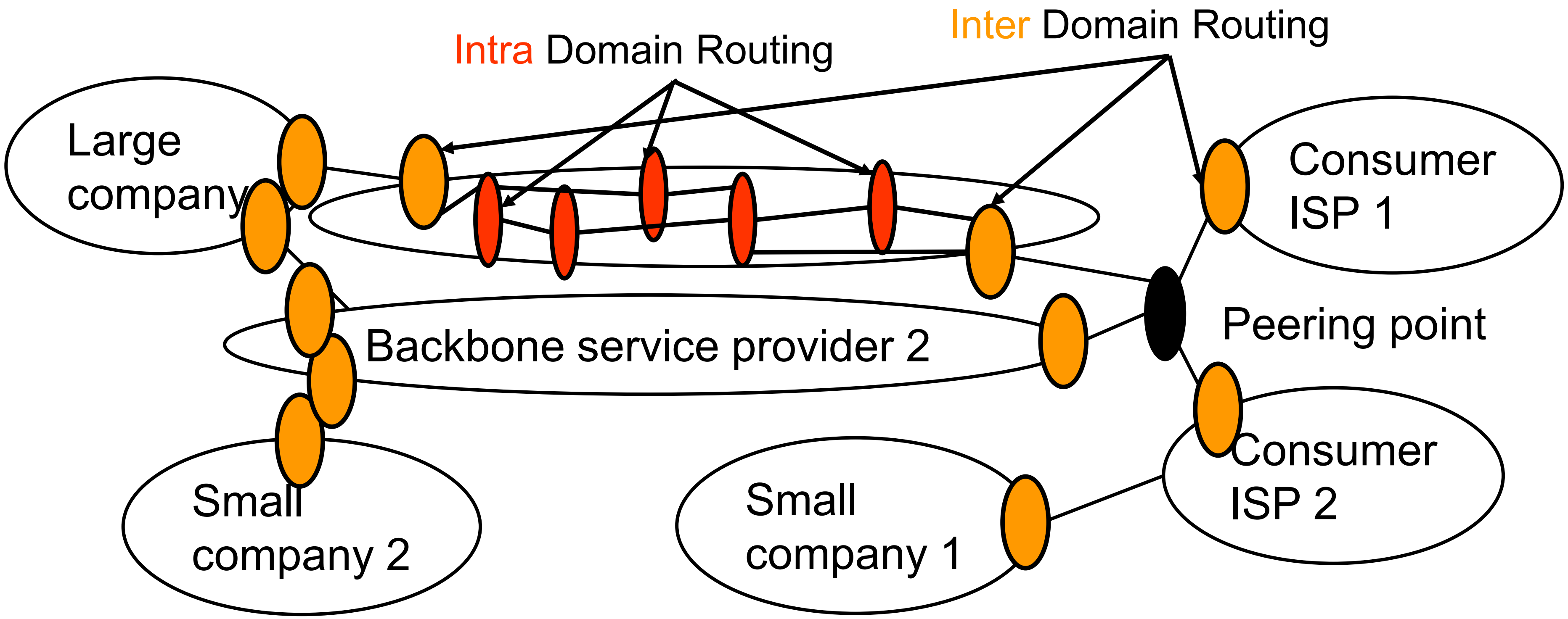
```
C:\>tracert www.nasa.gov

Tracing route to iznasa.hs.llnwd.net [2a02:3d0:623:a000::8008]
over a maximum of 30 hops:

  1    <1 ms    <1 ms    <1 ms  router-714.imp.fu-berlin.de
                                 [2001:638:80a:105::1]
  2    <1 ms    <1 ms    <1 ms  2001:638:80a:1::1
  3     1 ms     1 ms    <1 ms  2001:638:80a:3::1
  4     *        *        *     Request timed out.
  5    10 ms    10 ms    11 ms  2001:7f8:8::5926:0:1
  6    17 ms    17 ms    17 ms  tge1-4.fr5.dus1.ipv6.llnw.net
                                 [2a02:3d0:622:6c::2]
  7    12 ms    47 ms    12 ms  tge3-4.fr4.fra1.ipv6.llnw.net
                                 [2607:f4e8:1:c6::1]
  8    12 ms    12 ms    12 ms  2a02:3d0:623:6d::2
  9    15 ms    12 ms    12 ms
         https-2a02-3d0-623-a000--8008.fra.ipv6.llnw.net
         [2a02:3d0:623:a000::8008]

Trace complete.
```

What happened here?

The Idea of Internet Routing

Routing comprises:

- Updating of routing tables according to routing algorithm
- Exchange of routing information using routing protocol
- Forwarding of data based on routing tables and addresses



Autonomous Systems in the IP World

Large organizations can own multiple networks that are under single administrative control

➢ Forming *autonomous system* or *routing domain*

Autonomous systems form yet another level of aggregating routing information

➢ Give raise to *inter-* and *intra-domain routing*

Inter-domain routing is hard

- One organization might not be interested in carrying a competitor’s traffic
- Routing metrics of different domains cannot be compared
  - ➢ Only *reachability* can be expressed
- Scalability: Currently, inter-domain routers have to know about 200,000 – 400,000 networks

Intra-domain Routing: OSPF

The Internet's most prevalent intra-domain (= interior gateway) routing protocol: *Open Shortest Path First* (OSPF)

Main properties:

- Open, variety of routing distances, dynamic algorithm
- Routing based on traffic type (e.g. real-time traffic uses different paths)
- Load balancing: Also put some packets on the 2nd, 3rd best path
- Hierarchical routing, some security in place, support tunneled routers in transit networks

Essential operation: Compute shortest paths on graph abstraction of autonomous system

➢ Link state algorithm

Basic Ideas of Link State Routing

Distributed, adaptive routing

Algorithm:

1. Discovery of new neighbors
   - HELLO packet
2. Measurement of delay / cost to all neighbors
   - ECHO packet measures round trip time
3. Creation of link state packets containing all learned data
   - Sender and list of neighbors (including delay, age, ...)
   - Periodic or event triggered update (e.g. upon detecting new neighbors, line failure, ...)
4. Flooding of packet to all neighbors
   - Flooding, but with enhancements: Duplicate removal, deletion of old packets, ...
5. Shortest path calculation to all other routers (e.g. Dijkstra)
   - Computing intensive, optimizations exist

Inter-domain Routing: BGPv4

Routing between domains: *Border Gateway Protocols* (BGP)

BGP's perspective: Only autonomous systems and their connections
- Routing complicated by politics, e.g. only route packets for paying customers, …
- Legal constraints, e.g. traffic originating and ending in Canada must not leave Canada while in transit

Basic operation: Distance vector protocol
- Propagate information about reachable networks and distances one hop at a time
  - Each router learns only next step to destination
- Optimizations in BGP:
  - Not only keep track of cost via a given neighbor, but store entire paths to destination ASs
    - > Path vector protocol
  - More robust, solves problems like count to infinity, i.e. can handle disconnected networks efficiently

Conclusion: Interconnections

Single LANs are insufficient to provide communication for all but the simplest installations

Interconnection of LANs necessary

- Interconnect on purely physical layer: Repeater, hub
- Interconnect on data link layer: Bridges, switches
- Interconnect on network layer: Router
- Interconnect on higher layer: Gateway

Problems:

- Redundant bridges can cause traffic floods; need spanning tree algorithm
- Simple addresses do not scale; need routers

Internet Protocol

IP and Supporting Protocols

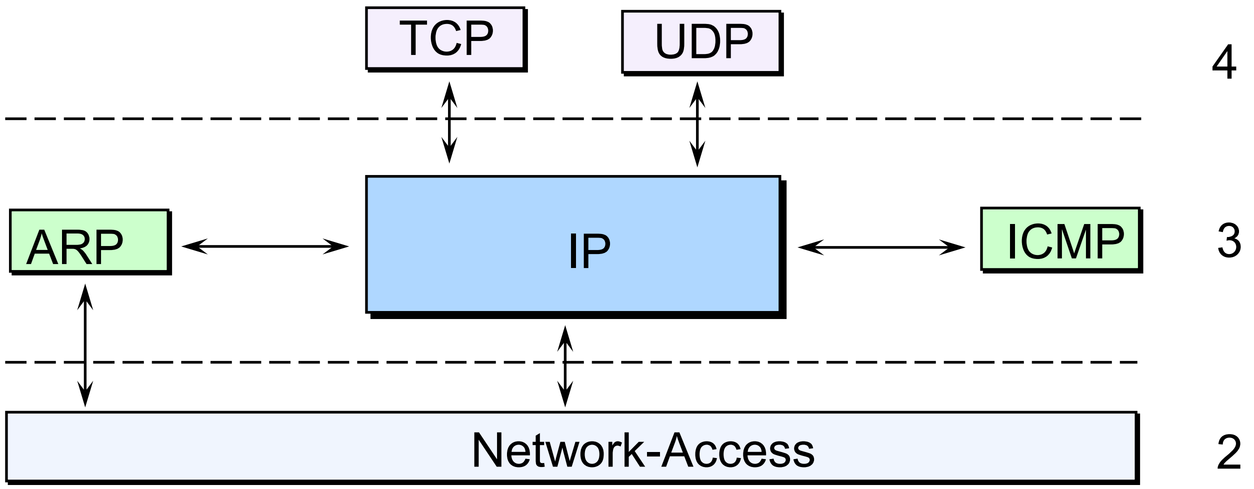
Transport protocols (Layer 4, TCP or UDP) hand over data together with IP address of receiver to Internet Protocol (IP)

IP may need to ask Address Resolution Protocol (ARP) for MAC address (Layer 2)

IP hands over data together with MAC address to Layer 2

IP forwards data to higher layers (TCP or UDP)

Internet Control Message Protocol (ICMP) can signal problems during transmission



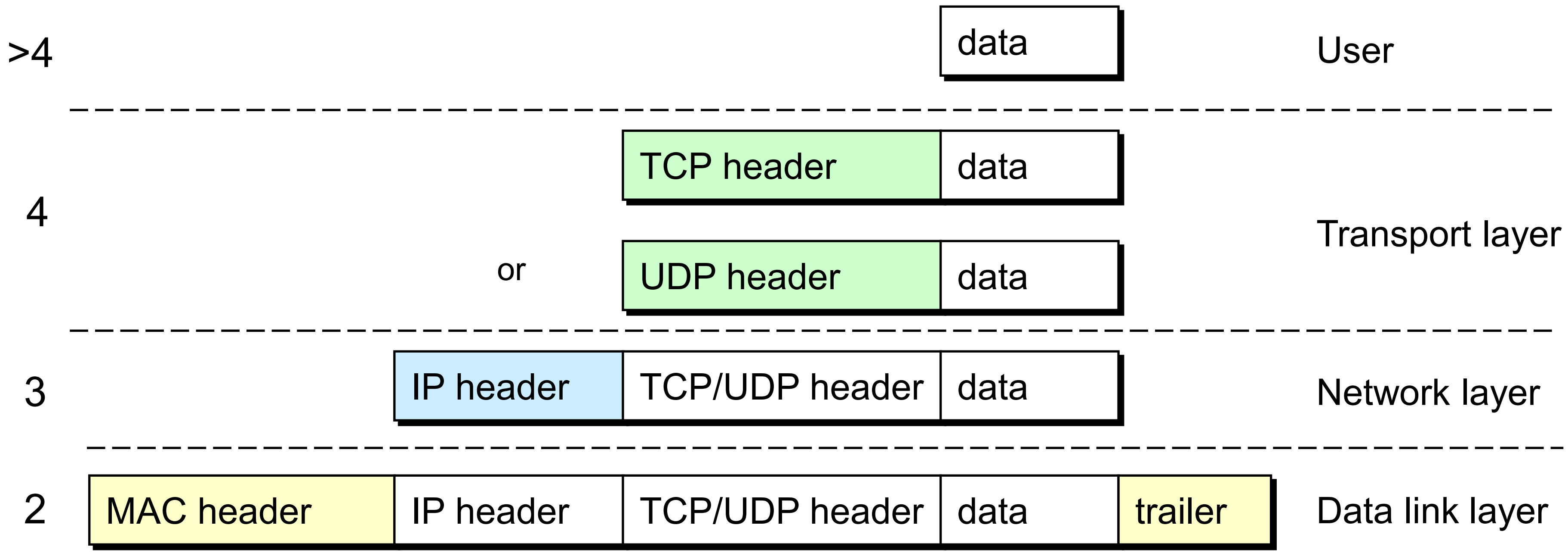
Data Encapsulation / Decapsulation

IP forwards data packets through network to receiver

TCP/UDP add ports (dynamic addresses of processes)

TCP offers reliable data transmission

Packets (PDU, protocol data unit) are encapsulated



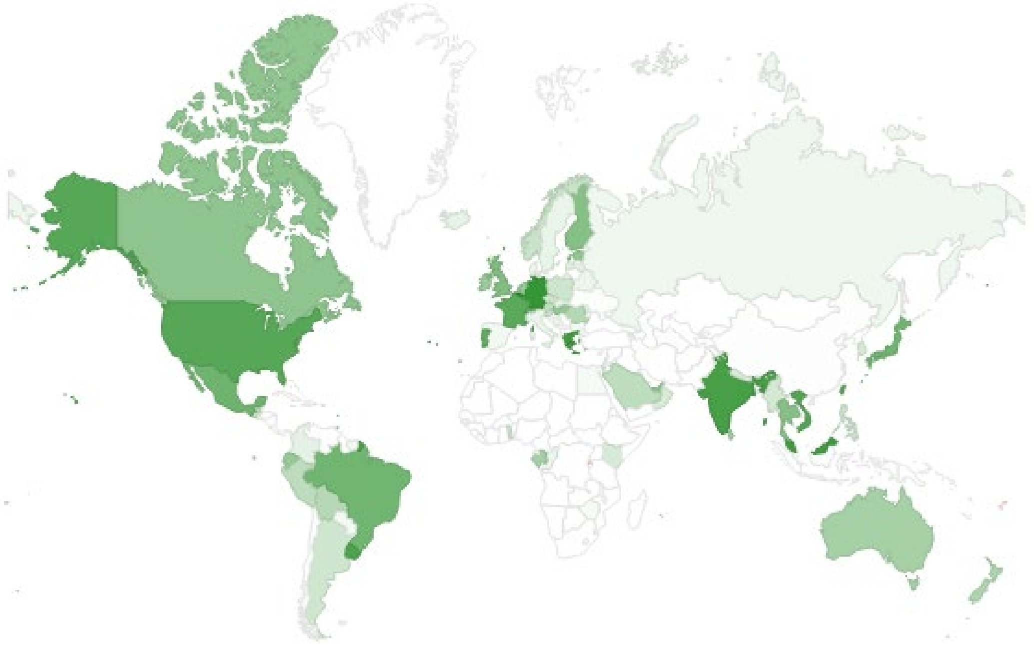
Internet Protocol (IP)

History
- Original development with support of US Department of Defense
- Already used back in 1969 in APANET

Tasks
- Routing support using structured addresses
- Checking of packet lifetime to avoid routing loops
- Fragmentation and reassembly
- Network diagnostics support

Development
- Today IP (version 4) is still most widely used layer 3 protocol
- Further development started back in the 80s/90s
  - Project IPng (IP next generation) of the IETF (Internet Engineering Task Force)
- Result in mid 90s: IPv6, still not as widely used as expected
- Today widely used, but could be more…
  - E.g., 2020: about 32% access Google via IPv6 (Germany 50%, USA 41%, Sweden 6%)

Per country IPv6 adoption as seen by Google



Source: www.google.com

Properties of IP

Packet oriented

Connectionless (datagram service)

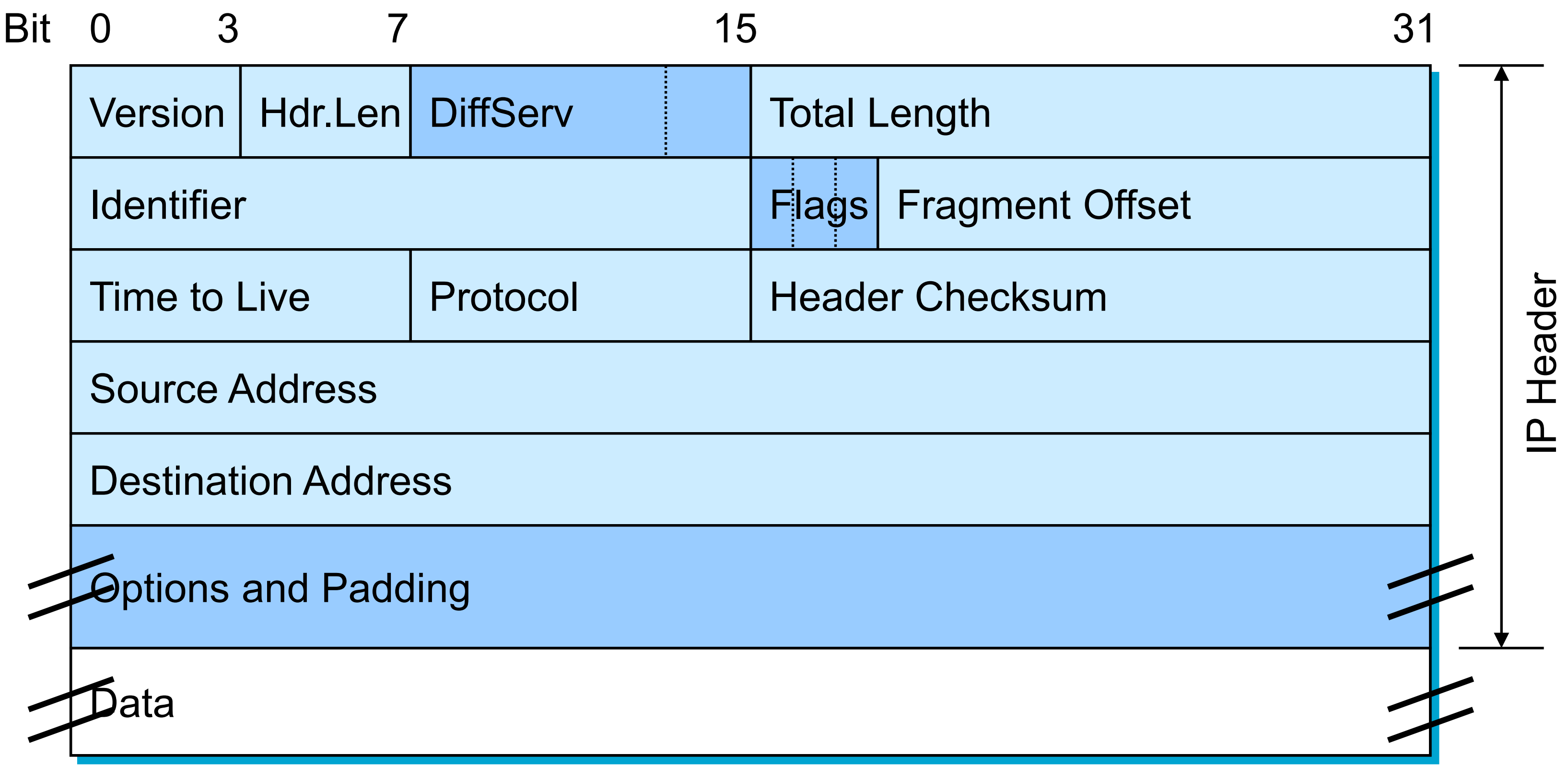
Unreliable transmission
- Datagrams can be lost
- Datagrams can be duplicated
- Datagrams can be reordered
- Datagrams can circle, but solved by Time to Live (TTL) field
- IP cannot handle Layer 2 errors
- At least there is ICMP to signal errors

Routing support via structured addresses

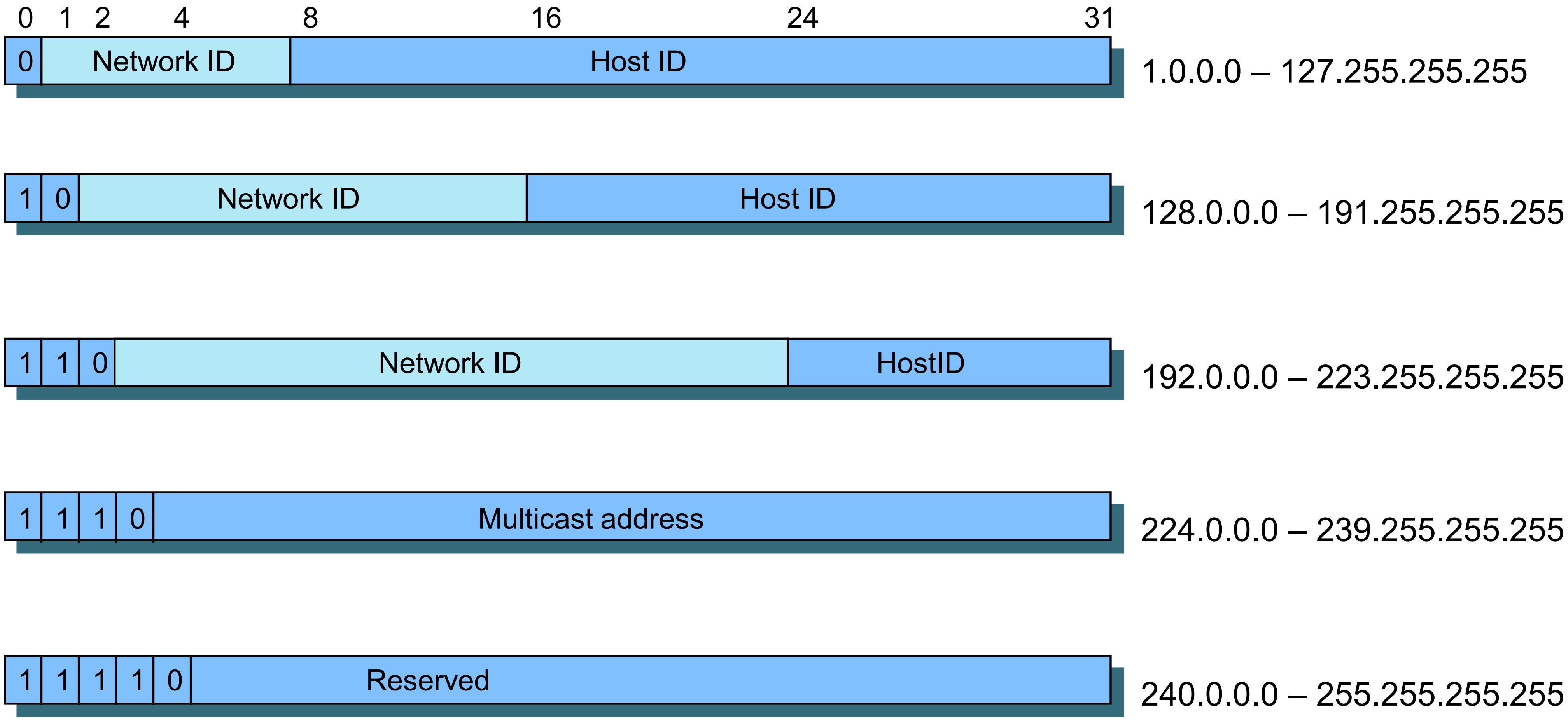
No flow control (yet, first steps taken)

Used in private and public networks

IPv4 Datagram

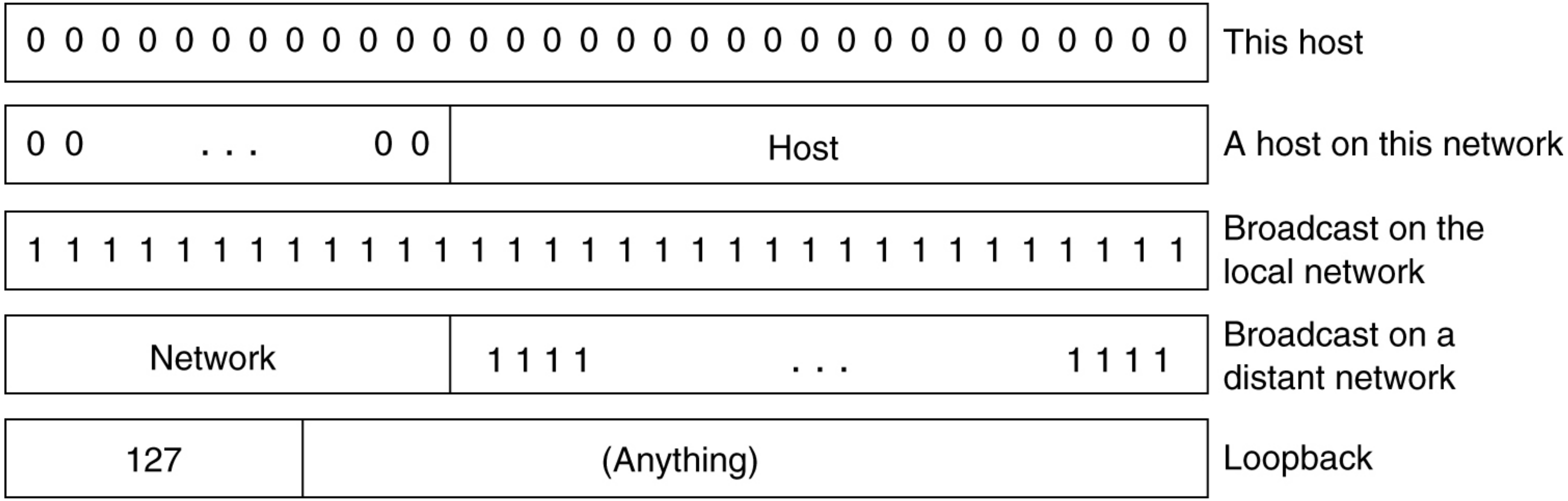


Structured IP Addresses and Address Classes (Classical View)



Special IP Addresses

Some IP addresses are reserved for special uses:



Not all of the network/host combinations are available

➢So-called “private” IP addresses

- Used for internal networks (addresses not routable)
- Example: 10.0.0.1, 192.168.0.1

Bridging Addressing Gap: ARP

➢What happens once a packet arrives at its destination network / LAN?

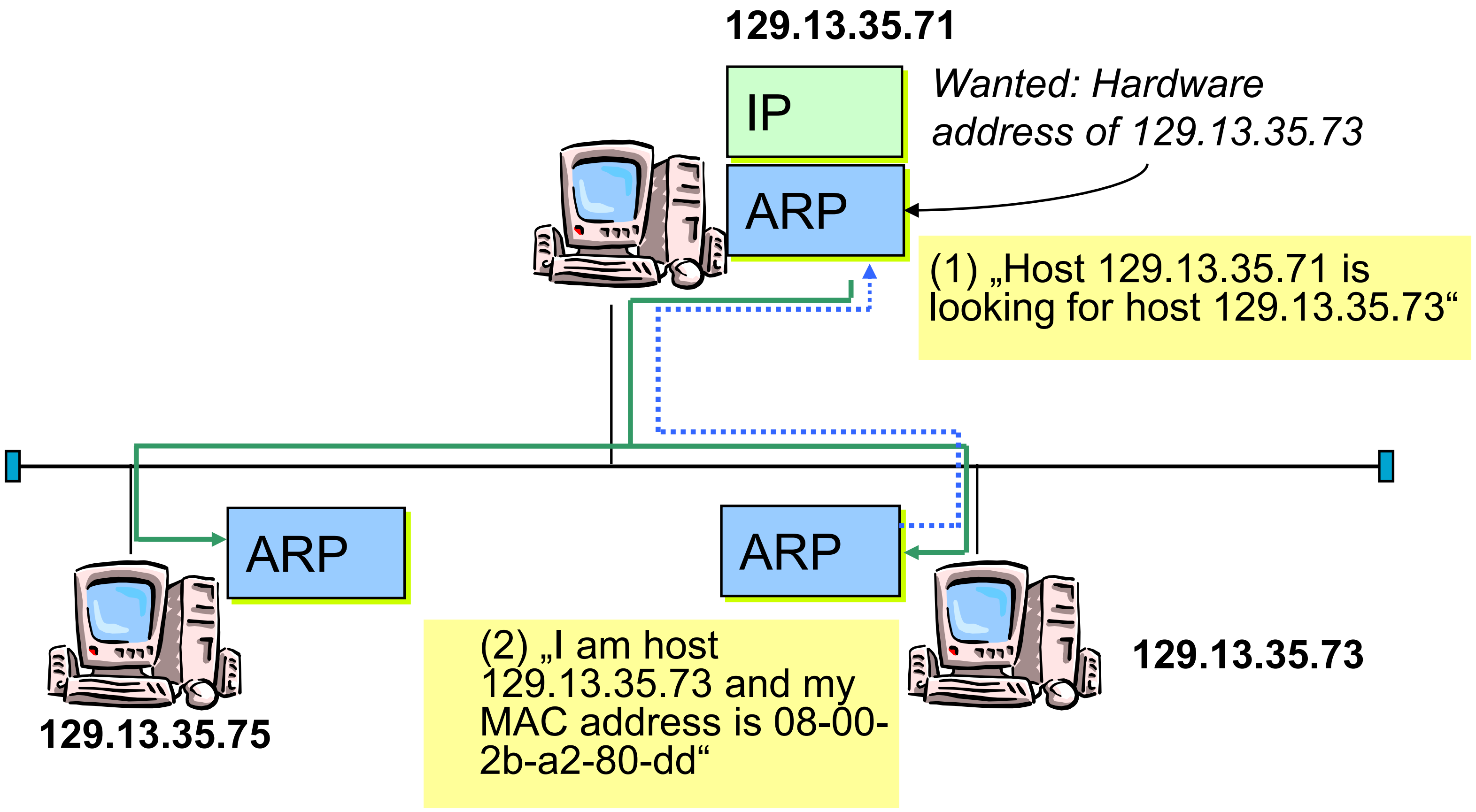
- IP address (which is all that is known about destination) needs to be translated into a MAC address that corresponds to the IP address

Simple solution: Broadcast

- Broadcast on LAN, asking which node has requested IP address
- Node answers with its MAC address
- Router can then forward packet to that MAC address

➢*Address Resolution Protocol* (ARP)

Example: ARP


129.13.35.71
IP
ARP
Wanted: Hardware address of 129.13.35.73
(1) „Host 129.13.35.71 is looking for host 129.13.35.73“
ARP
ARP
(2) „I am host 129.13.35.73 and my MAC address is 08-00-2b-a2-80-dd“
129.13.35.73
129.13.35.75

Scalability Problems of IP

Class A and B networks can contain *many* hosts

- Too many for a router to easily deal with
- Additionally, administrative problems in larger networks

➢ Solution: Subnetting, i.e. a network is subdivided into several smaller networks by breaking up the address space

Network classes waste a lot of addresses

- Example: Organization with 2000 hosts requires a class B address, wasting 64K-2K ≈ 62.000 host addresses

➢ Solution: **Classless addressing ➔ Classless Inter Domain Routing (CIDR)**

  - Dynamic boundaries between host/network part of IP address
  - Aggregation on routers to reduce size of global routing table

Classless Inter Domain Routing (CIDR)

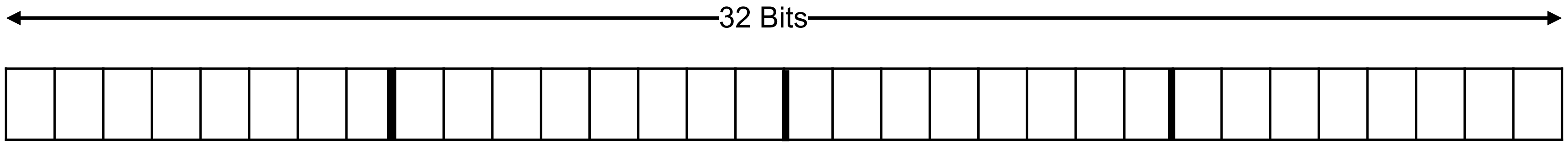
Example 1:

Does the host with the address 192.168.1.255 belong to the network that is specified as 192.168.1.0/24?

Classless Inter Domain Routing (CIDR)

Example 1:

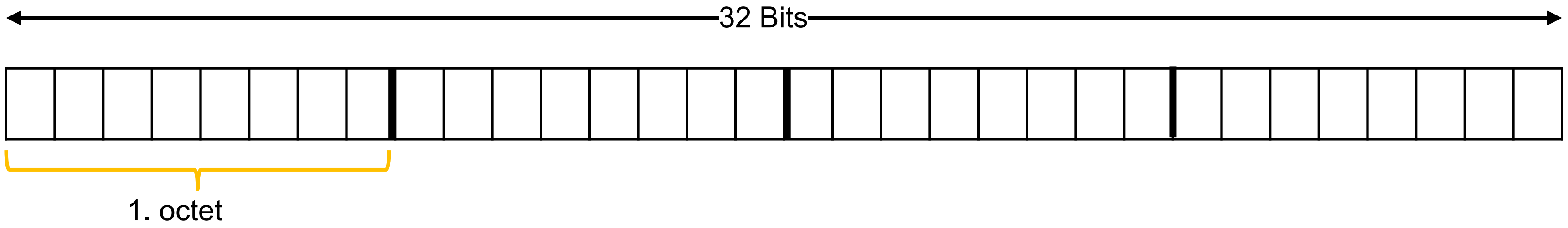
Does the host with the address 192.168.1.255 belong to the network that is specified as 192.168.1.0/24?



Classless Inter Domain Routing (CIDR)

Example 1:

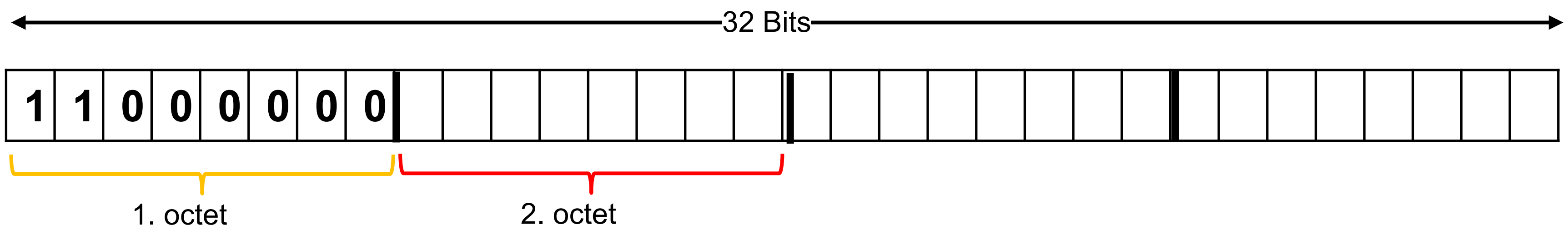
Does the host with the address 192.168.1.255 belong to the network that is specified as 192.168.1.0/24?



Classless Inter Domain Routing (CIDR)

Example 1:

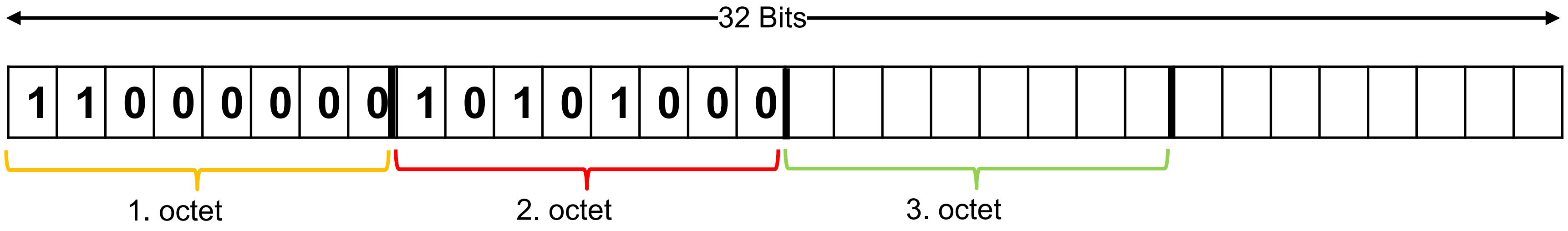
Does the host with the address 192.168.1.255 belong to the network that is specified as 192.168.1.0/24?



Classless Inter Domain Routing (CIDR)

Example 1:

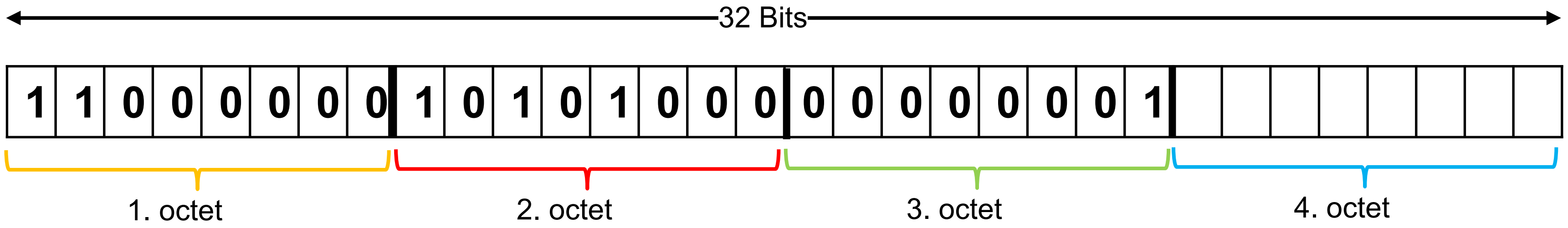
Does the host with the address 192.168.1.255 belong to the network that is specified as 192.168.1.0/24?



Classless Inter Domain Routing (CIDR)

Example 1:

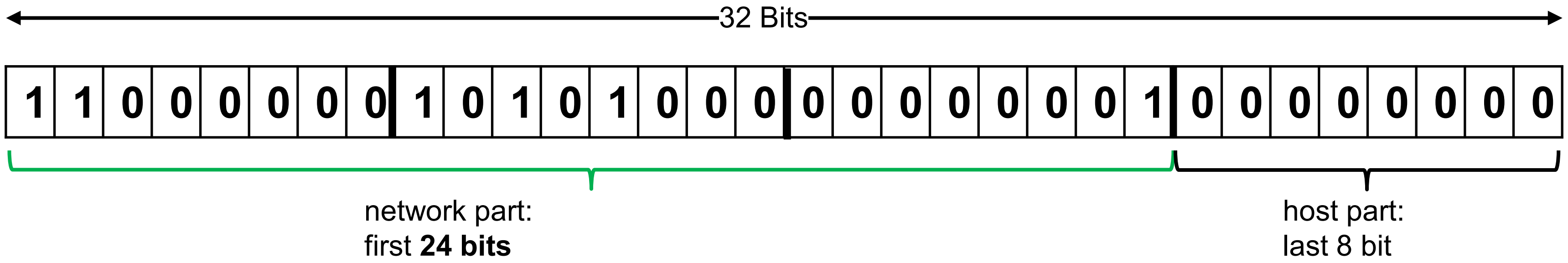
Does the host with the address 192.168.1.255 belong to the network that is specified as 192.168.1.0/24?



Classless Inter Domain Routing (CIDR)

Example 1:

Does the host with the address 192.168.1.255 belong to the network that is specified as 192.168.1.0/24?



Classless Inter Domain Routing (CIDR)

Example 1:

Does the host with the address 192.168.1.255 belong to the network that is specified as 192.168.1.0/24?

32 Bits

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

network part:
first **24 bits**

host part:
last 8 bit

host address 192.168.1.255 :

- first 3 octets (=24 bits) identical to network part ✅

host address

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

network part:
first **24 bits**

host part:
last 8 bit

Classless Inter Domain Routing (CIDR)

Example 2:

Does the host with the address 192.168.1.255 belong to the network that is specified as 192.168.1.64/26?

Classless Inter Domain Routing (CIDR)

Example 2:

Does the host with the address 192.168.1.255 belong to the network that is specified as 192.168.1.64/26?

32 Bits

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

network part:
first **26 bits**

host part:
last 6 bit

Classless Inter Domain Routing (CIDR)

Example 2:

Does the host with the address 192.168.1.255 belong to the network that is specified as 192.168.1.64/26?

←32 Bits→

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

network part:
first **26 bits**

host part:
last 6 bit

host address 192.168.1.255 :

- first 26 bits **NOT** identical to network part!

host address

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

network part:
first **26 bits**

host part:
last 6 bit

Classless Inter Domain Routing (CIDR)

Example 2:

Does the host with the address 192.168.1.255 belong to the network that is specified as 192.168.1.64/26?

32 Bits

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

network part:
first **26 bits**

host part:
last 6 bit

host address 192.168.1.255 :

- first 26 bits **NOT** identical to network part! ❌

host address

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

network part:
first **26 bits**

host part:
last 6 bit

Conclusion: Internet Protocol

Unreliable datagram transfer

Needs supporting protocols
- ARP for mapping IP to MAC address
- ICMP for error signaling

Classical addressing wastes addresses
 - Classless addressing, CIDR

Version 4 dominant, version 6 coming (since years…)
- **Much** more in Telematics

Roadmap

8. Networked Computer & Internet
9. Network Access Layer I – Physical Layer
10. Network Access Layer II – Data Link Layer
11. **Internet Layer – Network Layer**
12. Transport Layer
13. Applications