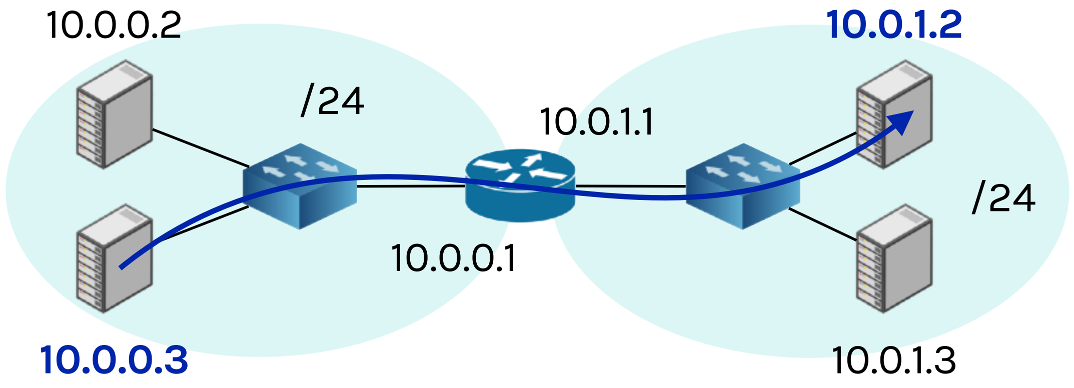
Computer Networks (WS23/24)

L5: The Network Layer - Part 1

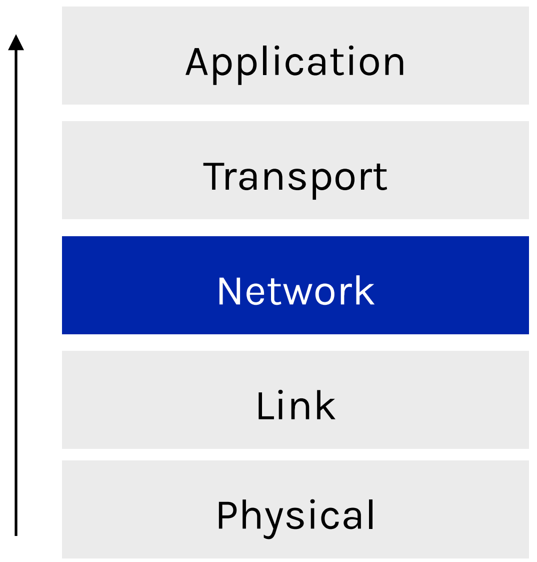
Prof. Dr. Lin Wang
Computer Networks Group (PBNet)
Department of Computer Science
Paderborn University

Materials inspired by Shyam Gollakota and Laurent Vanbever

Learning objectives



What happens when a packet travels across networks?

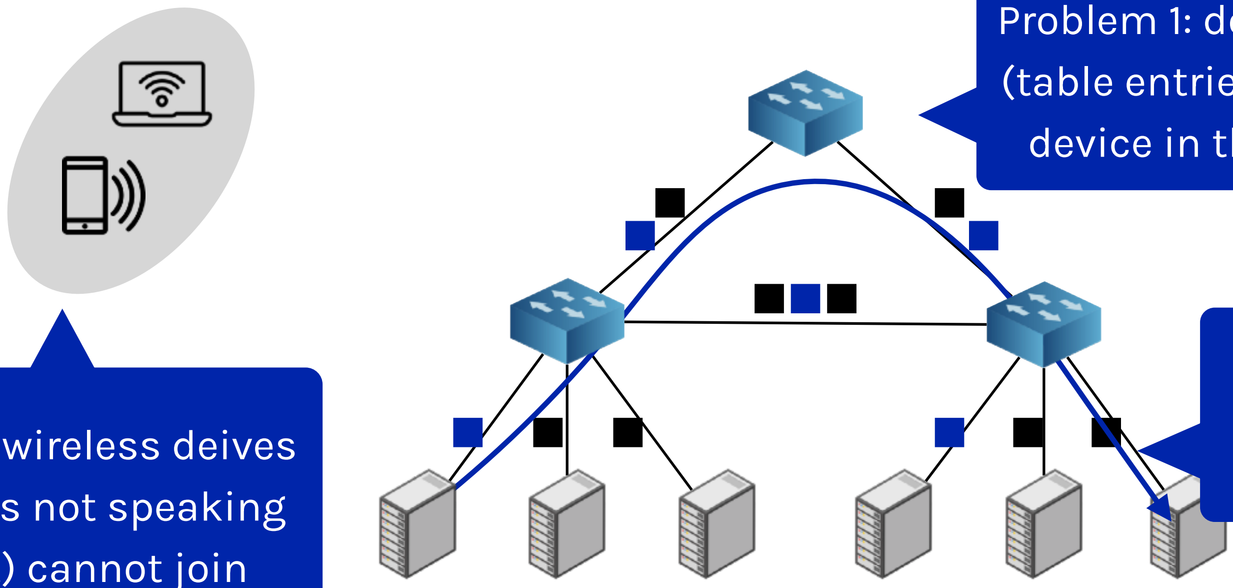


Part 1

- Inter-networking
- IP forwarding
- Network Address Translation (NAT)
- Helper protocols: ARP, DHCP, ICMP

Inter-networking

Internet based on switched Ethernet

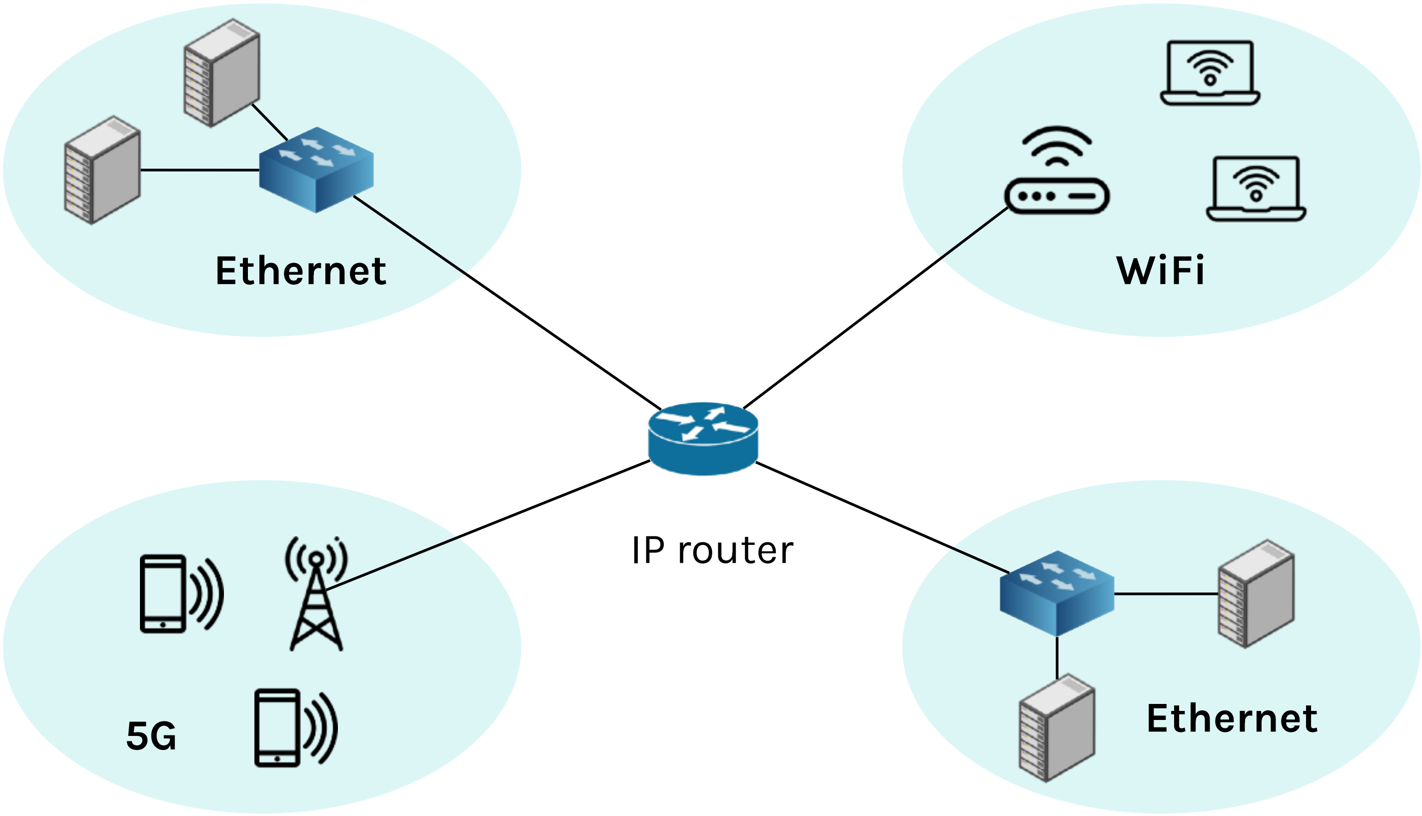


Problem 1: do not scale (table entries for every device in the world)

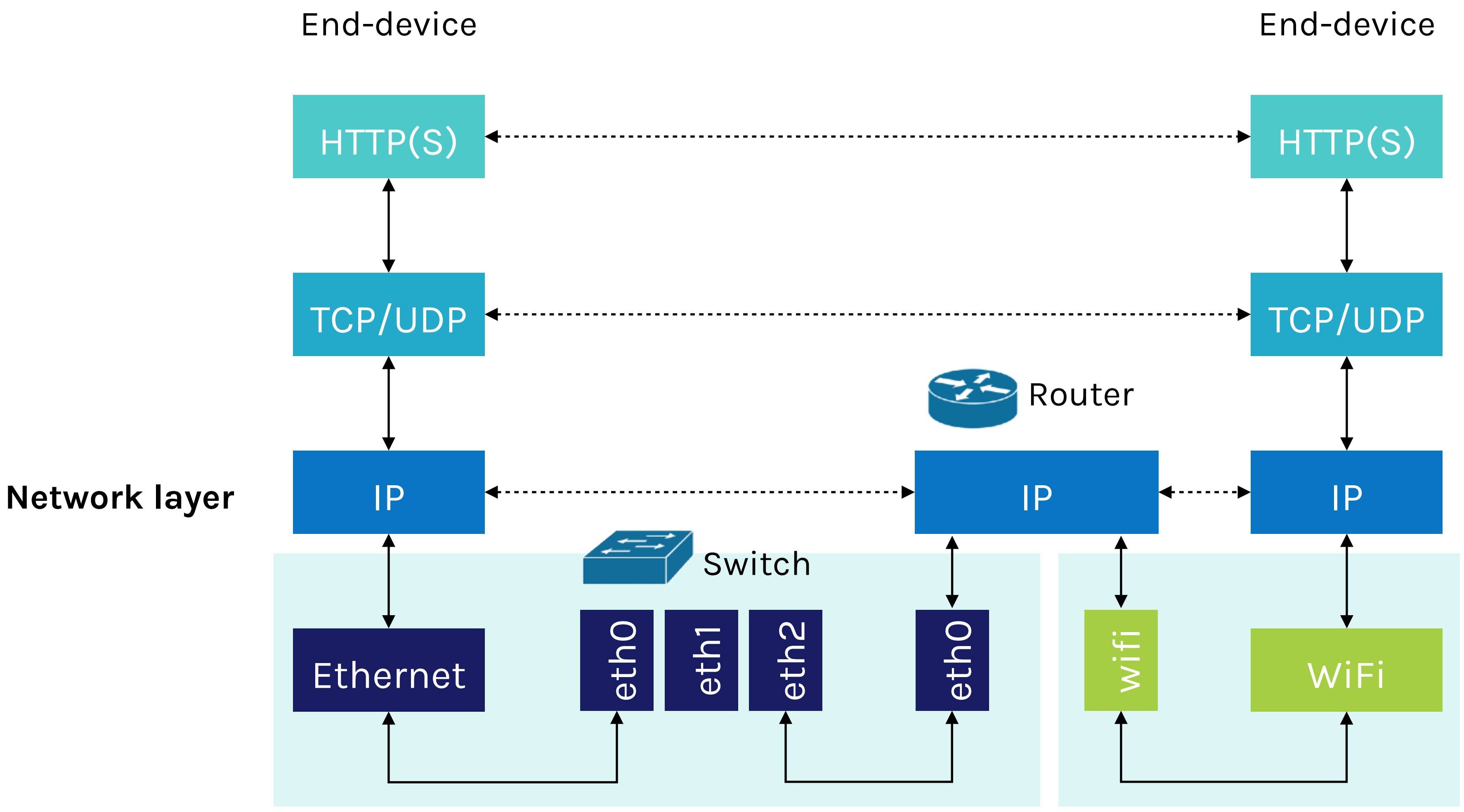
Problem 2: wireless deives (and others not speaking Ethernet) cannot join

Problem 3: lack of traffic control (bounded by STP)

Inter-networking



The network layer



People behind it

Pioneers: Cerf and Kahn

- Fathers of the Interent
- In 1974, later led to TCP/IP

Tackled the problem of interconnecting networks

- Instead of mandating a single network technology

Vinton Cerf

Bob Kahn

ACM Turing Award 2004

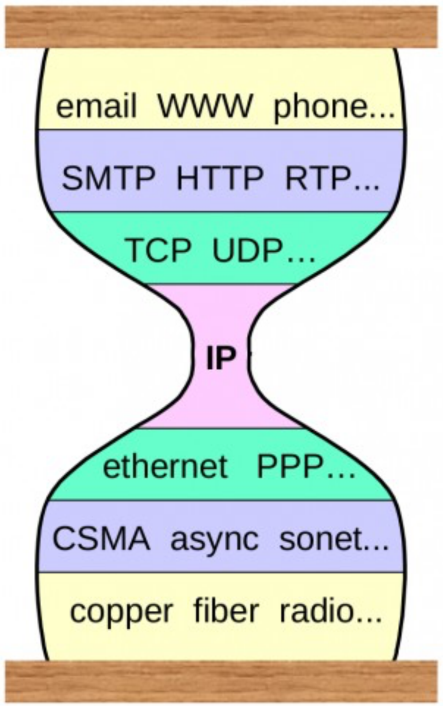
Internet reference model

Internet Protocol (IP) is the narrow waist

- Supports many different links below: Ethernet, WiFi, 4G/LTE, 5G
- Supports many application above

IP as a lowest common denominator

- Asks little of lower-layer networks
- Gives little as a higher-layer service



IPv4 address

Unique 32-bit numbers associated to a network interface (on a host or a router…)

Usually written using dotted-quad notation

82 . 130 . 102 . 10

01010010 10000010 01100110 00001010

IPv6 address

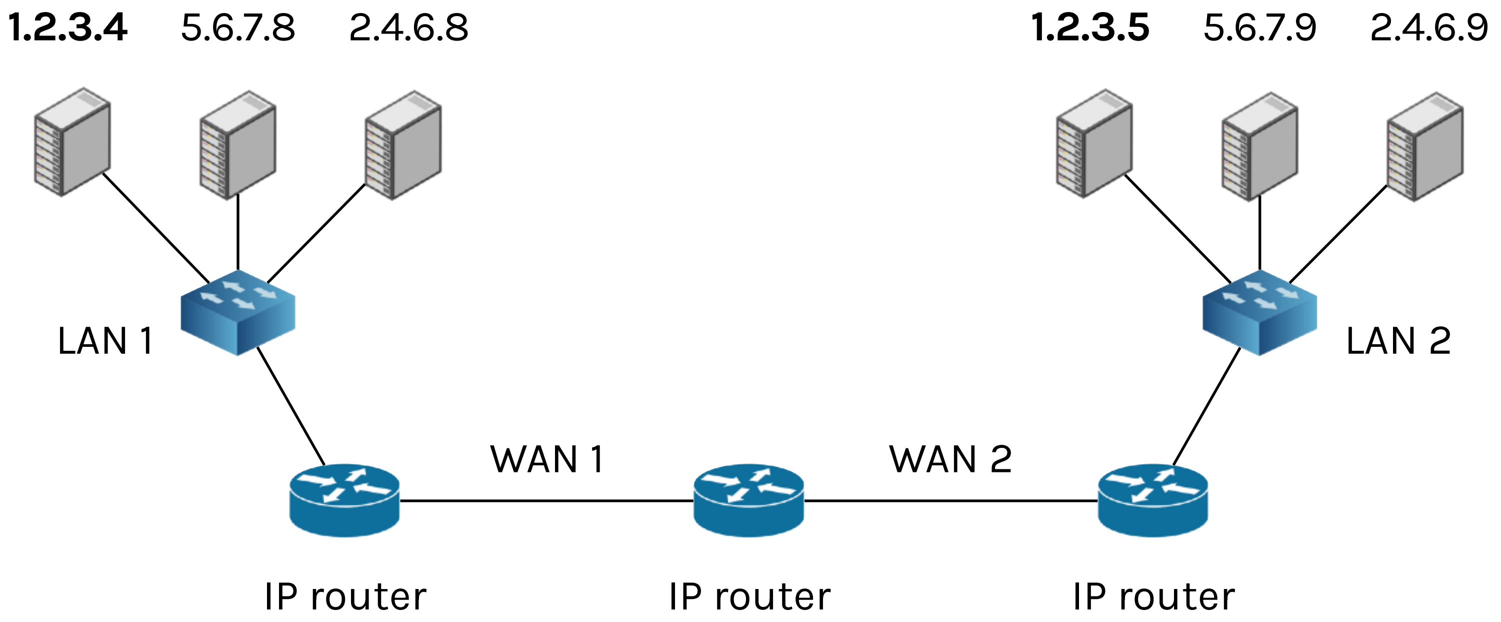
Notation

- 8 groups of 16 bits, each separated by colons (:)
- Leading zeros in any groups are removed
- One section of zeros is replaced by a double colon (::)

Examples

- 1080:0:0:0:8:800:200C:417A → 1080::8:800:200C:417A
- FF01:0:0:0:0:0:0:0101 → FF01::101
- 0:0:0:0:0:0:0:1 → ::1

IP address assignment



1.2.3.4 ←
1.2.3.5 →
...

Forwarding table

For each IP address, we need an entry in the table

LAN: Local Area Network

WAN: Wide Area Network

Two universal tricks in Computer Science

When you need…	more **flexibility**
You add…	a layer of **indirection**

When you need…	more **scalability**
You add…	a **hierarchical** structure

Hierarchical postal addresses

City	Paderborn
Zip	33098
Street	Pohlweg
Building	O
Room (in building)	O3-158
Name	Lin Wang

Nobody maintains where every single building is in the Deutsch Post system

Hierarchical forwarding

Step 1	Deliver the letter to the post office responsible for the city and zip code
Step 2	Assign letter to the mail person covering the street
Step 3	Drop letter into the mailbox attached to the building
Step 4	Hand in the letter to the addressed person

IP addresses are hierarchical

32 bits

01010010.10000010.01100110.00001010

Prefix
(Identifying the **network**)

Suffix
(Identifying the **hosts** in the network)

IP prefix

82 . 130 . 102 . 0 /24

Prefix length (in bits)

Prefix part	Suffix part	IP address
01010010.10000010.01100110.	00000000	82.130.102.0
		Identifies the network itself
01010010.10000010.01100110.	00000001	82.130.102.1
01010010.10000010.01100110.	00000010	82.130.102.2
01010010.10000010.01100110.	00000011	82.130.102.3
......		
01010010.10000010.01100110.	11111110	82.130.102.254
01010010.10000010.01100110.	11111111	82.130.102.255
		Broadcast address

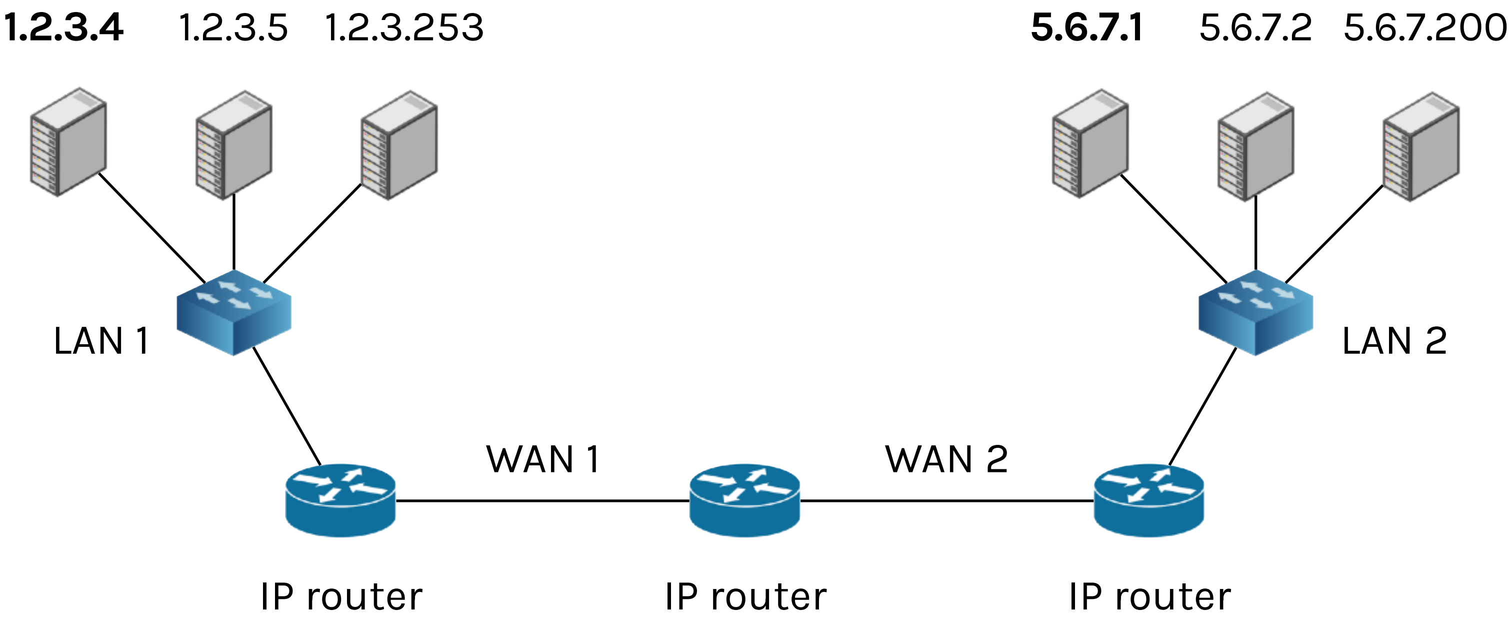
Only 254 valid addresses to allocate to hosts for /24

IP prefix with masks

Address	**82.130.102.0**
	01010010.10000010.01100110.00000000
Mask	**255.255.255.0**
	11111111.11111111.11111111.00000000

ANDing the address and the mask gives you the IP prefix

Scalable forwarding



1.2.3.0/24 ←
5.6.7.0/24 →

Forwarding table

Only two entries needed with prefix matching

Legacy classful networking

	Leading bits	Prefix length	#hosts	Start addr.	End addr.
Class A	`0`	8	2^{24}	`0.0.0.0`	`127.255.255.255`
Class B	`10`	16	2^{16}	`128.0.0.0`	`191.255.255.255`
Class C	`110`	24	2^{8}	`192.0.0.0`	`223.255.255.255`
Class D (Multicast)	`1110`			`224.0.0.0`	`239.255.255.255`
Class E (Reserved)	`1111`			`240.0.0.0`	`255.255.255.255`

Class C was too small ⇒ everyone requested class B (too big, a lot of waste)

Classless Inter-Domain Routing (CIDR)

Enables flexible division between network and hosts addresses

CIDR must specify both the address and the mask

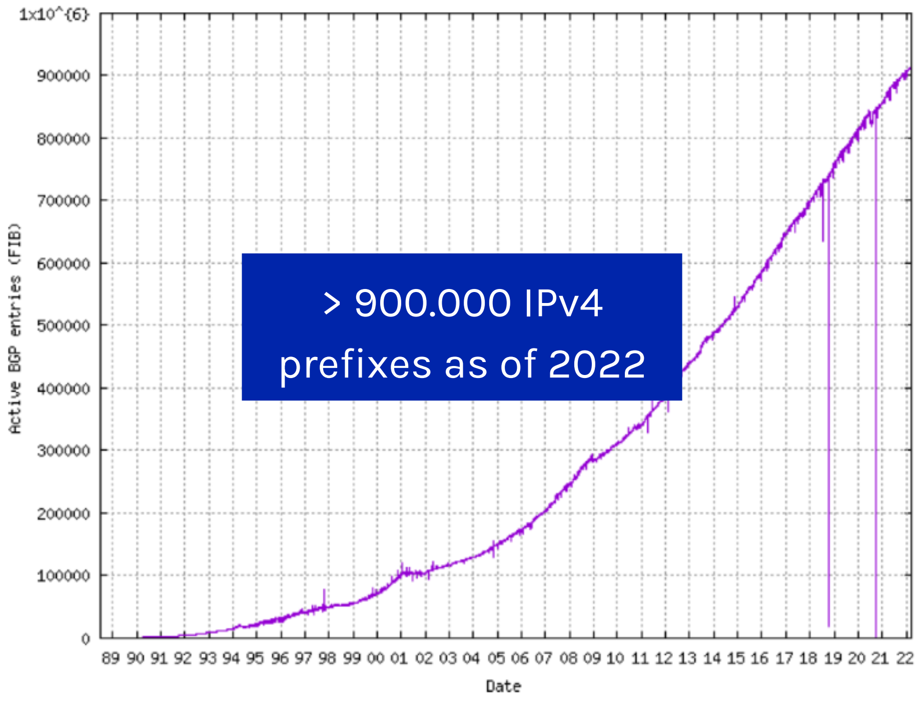
- Mask in the classful address is implicit in the first address bits
- Mask in CIDR is carried by the routing algorithm, not implicitly in the address

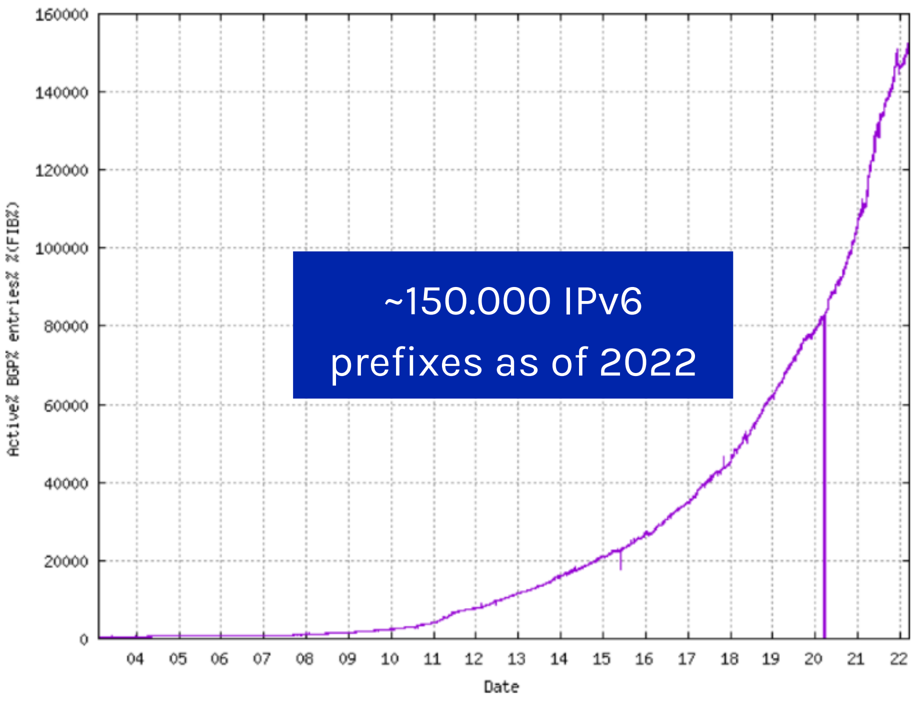
Example: an organization needs 500 addresses

Type	Allocation	Utilization
Classful	Class B (/16)	1%
CIDR	/23 (2 Class C's)	98%

With CIDR, the address utilization is always higher than 50%. Why?

IP prefixes on the Internet





http://www.cidr-report.org/
https://www.cidr-report.org/v6/as2.0/

Hierarchical IP address allocation

ICANN

Internet Corporation for Assigned Names and Numbers

Reginal Internet Registries (RIRs)

ARIN	lacnic	RIPE NCC RIPE NETWORK COORDINATION CENTRE	APNIC	AFRINIC The Internet Numbers Registry for Africa
America	Latin America	Europe	Asia-Pacific	Africa

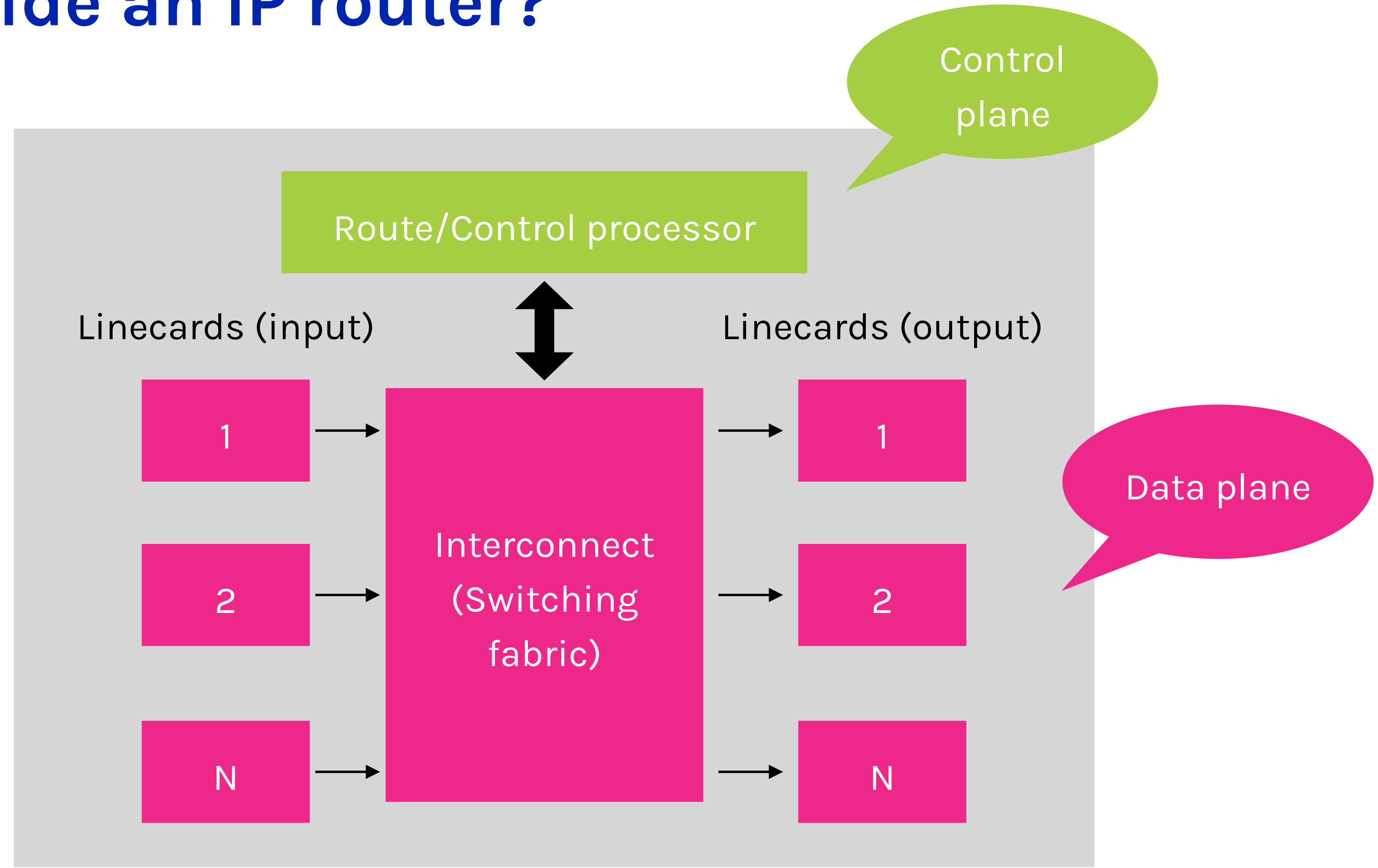
ISPs / Institutions

UNIVERSITÄT PADERBORN

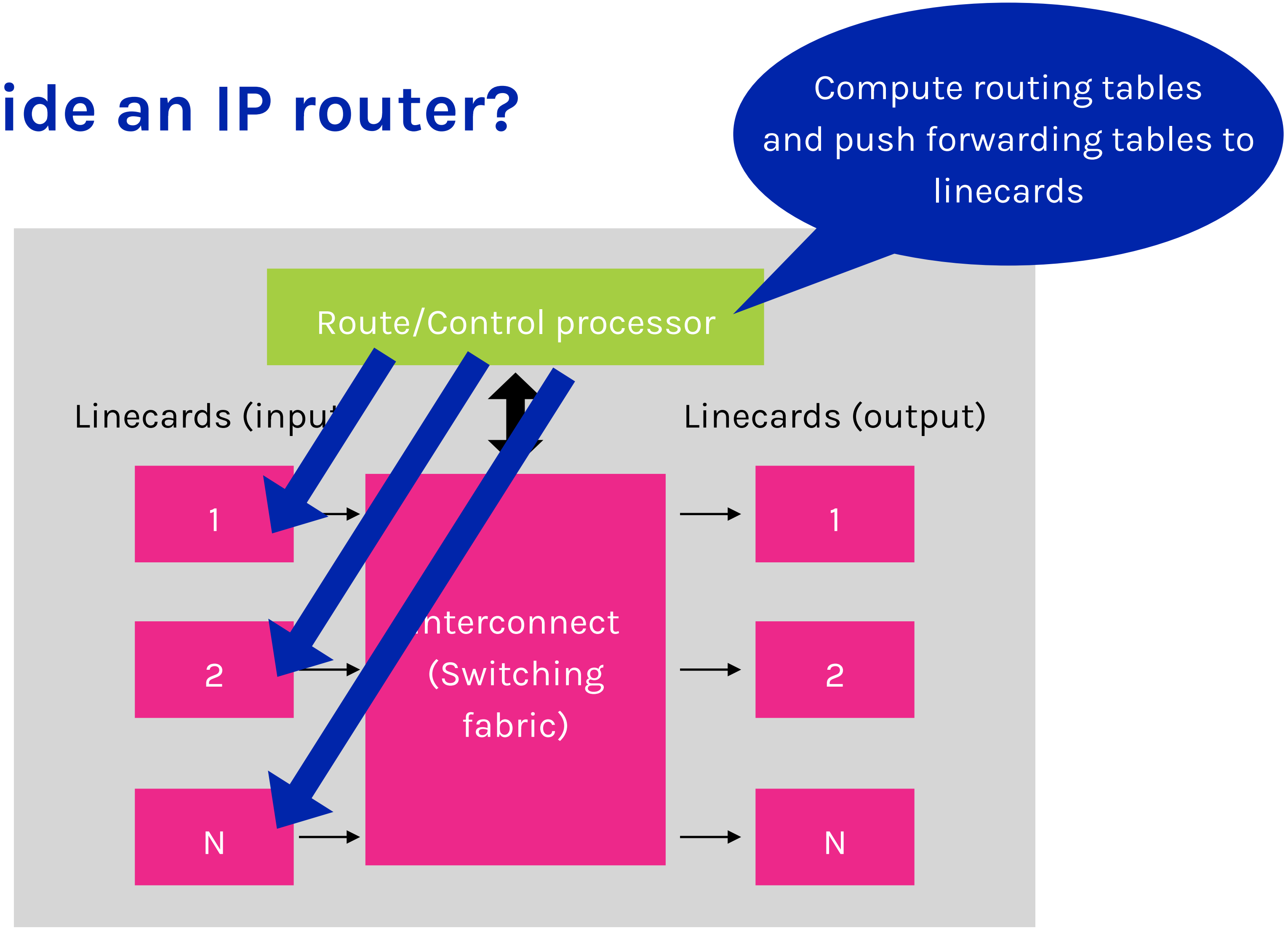
TECHNISCHE UNIVERSITÄT DARMSTADT

IP Forwarding

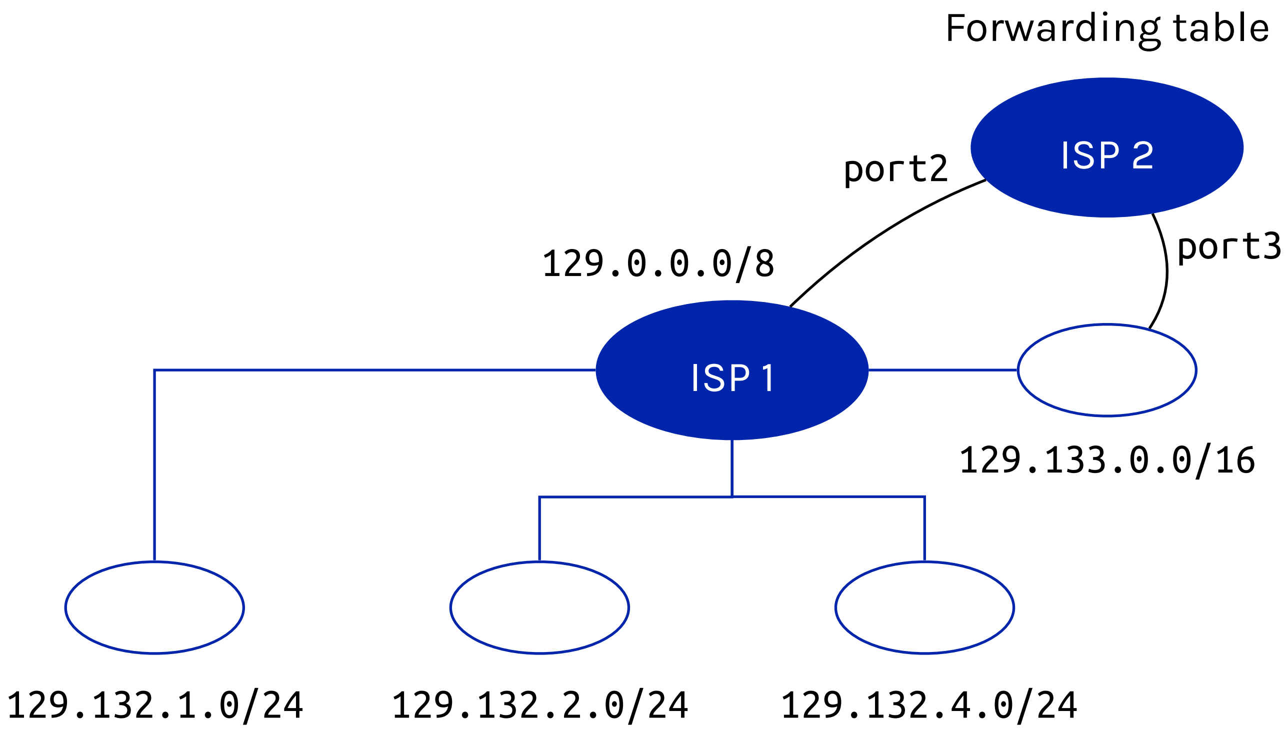
What's inside an IP router?



What's inside an IP router?



Forwarding table on IP routers

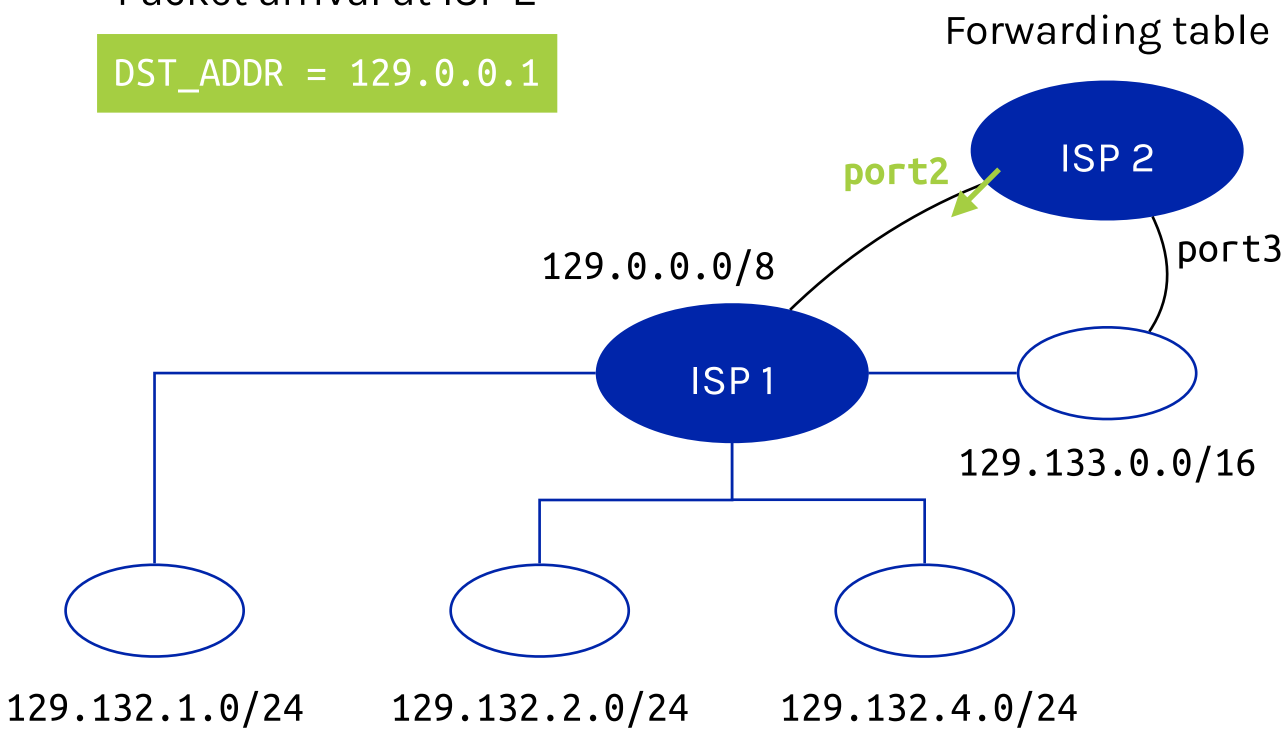


IP prefix	Output
129.0.0.0/8	port2
129.132.1.0/24	port2
129.132.2.0/24	port2
129.133.0.0/16	port3

Forwarding based on prefix matching

Packet arrival at ISP 2

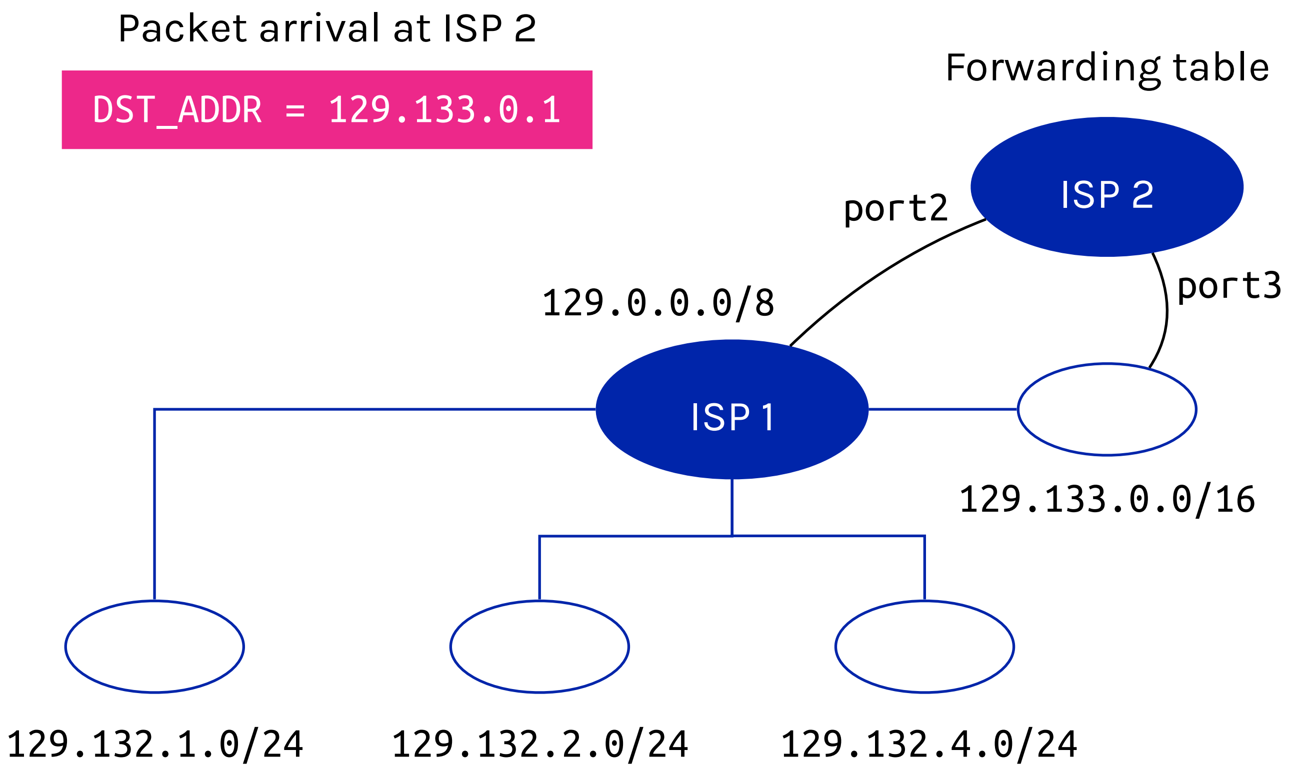
DST_ADDR = 129.0.0.1



IP prefix	Output
129.0.0.0/8	port2
129.132.1.0/24	port2
129.132.2.0/24	port2
129.133.0.0/16	port3

Router at ISP 2 performs an IP lookup to find the matching prefix → forward the packet to port2

Forwarding based on prefix matching



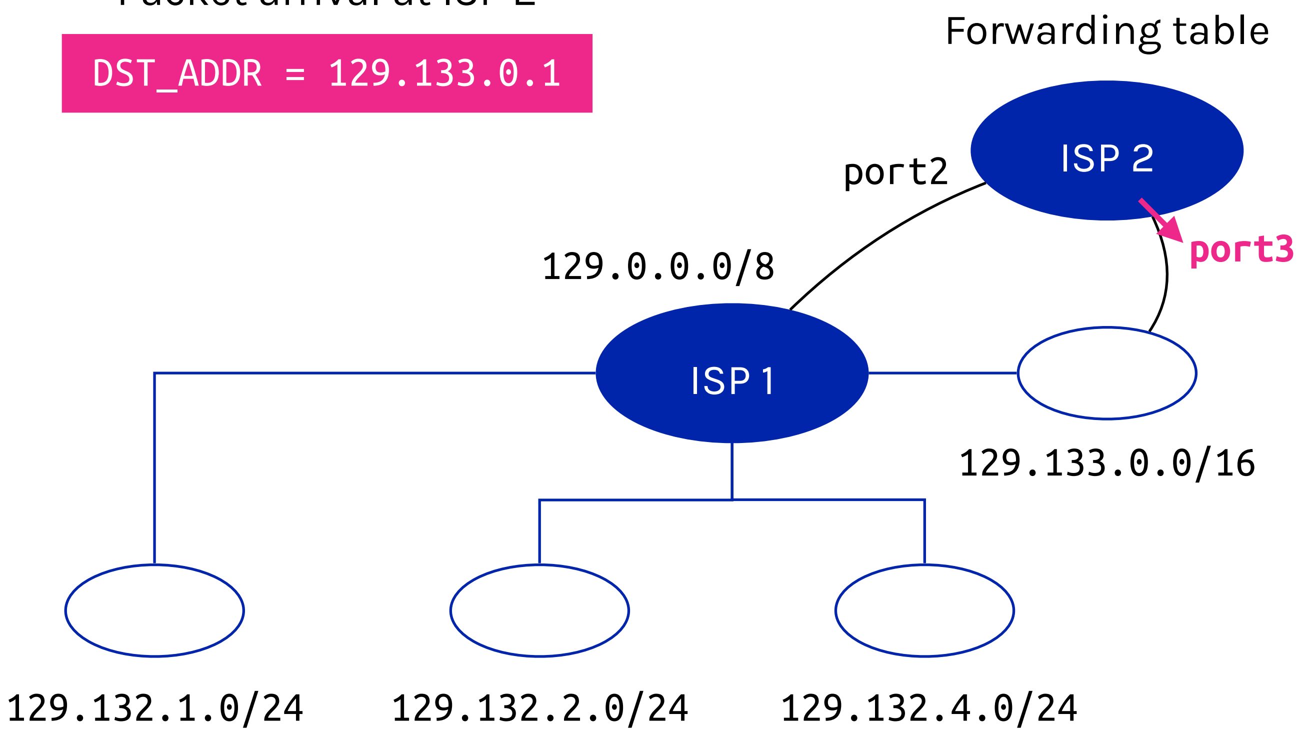
IP prefix	Output
129.0.0.0/8	**port2**
129.132.1.0/24	port2
129.132.2.0/24	port2
129.133.0.0/16	**port3**

Where the packet should be forwarded?

Longest prefix matching

Packet arrival at ISP 2

DST_ADDR = 129.133.0.1

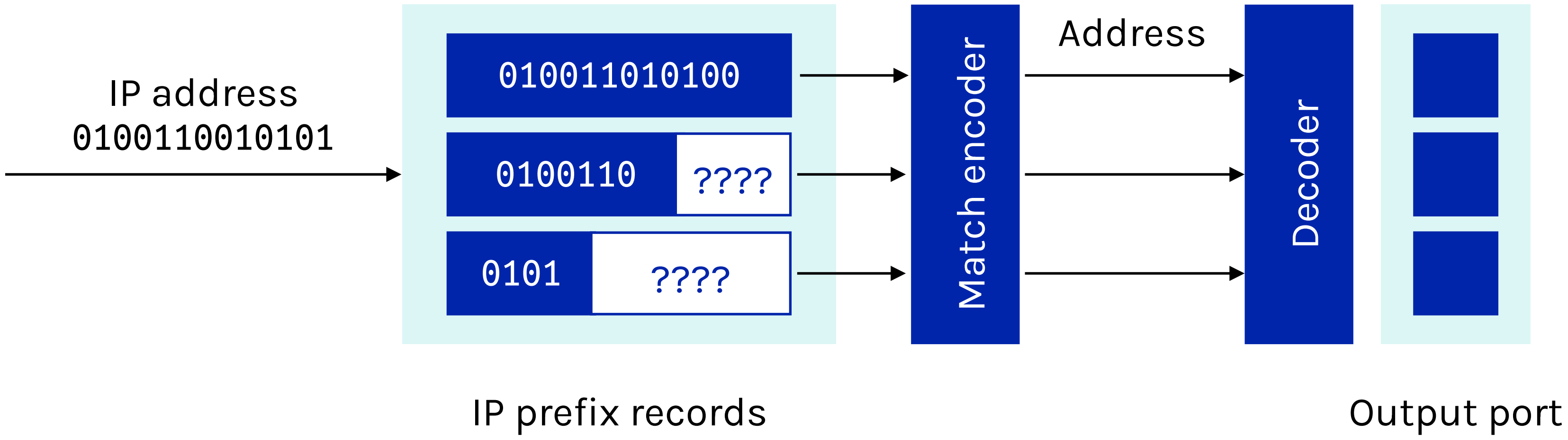


IP prefix	Output
129.0.0.0/8	port2
129.132.1.0/24	port2
129.132.2.0/24	port2
129.133.0.0/16	**port3**

Router at ISP 2 matches on the longest prefix → forward the packet to port3

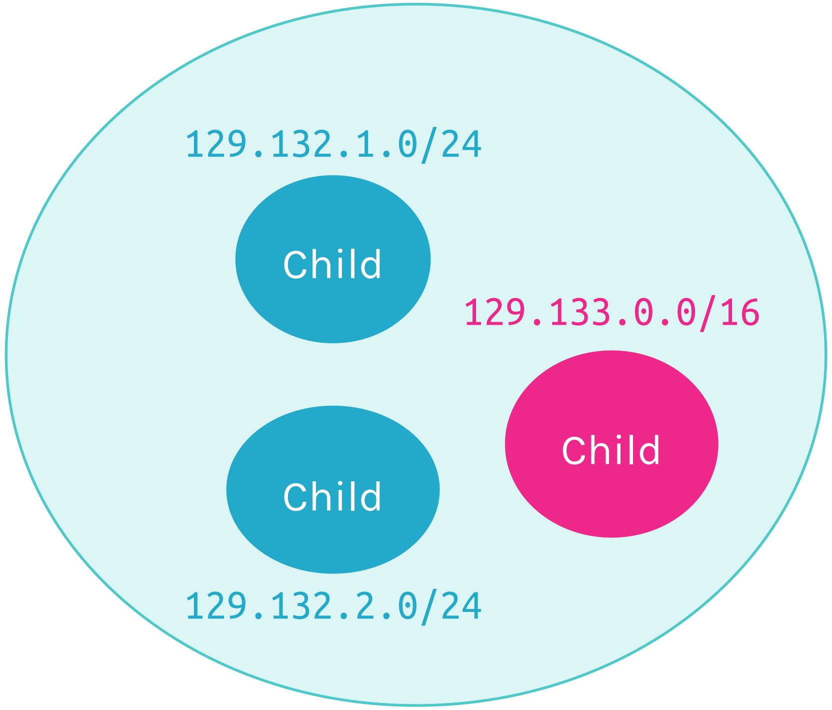
Hardware support for longest prefix matching

Ternary Content Addressable Memory (TCAM)



Simplifying the forwarding table

A child prefix can be filtered from the table if it shares the same output as its parent

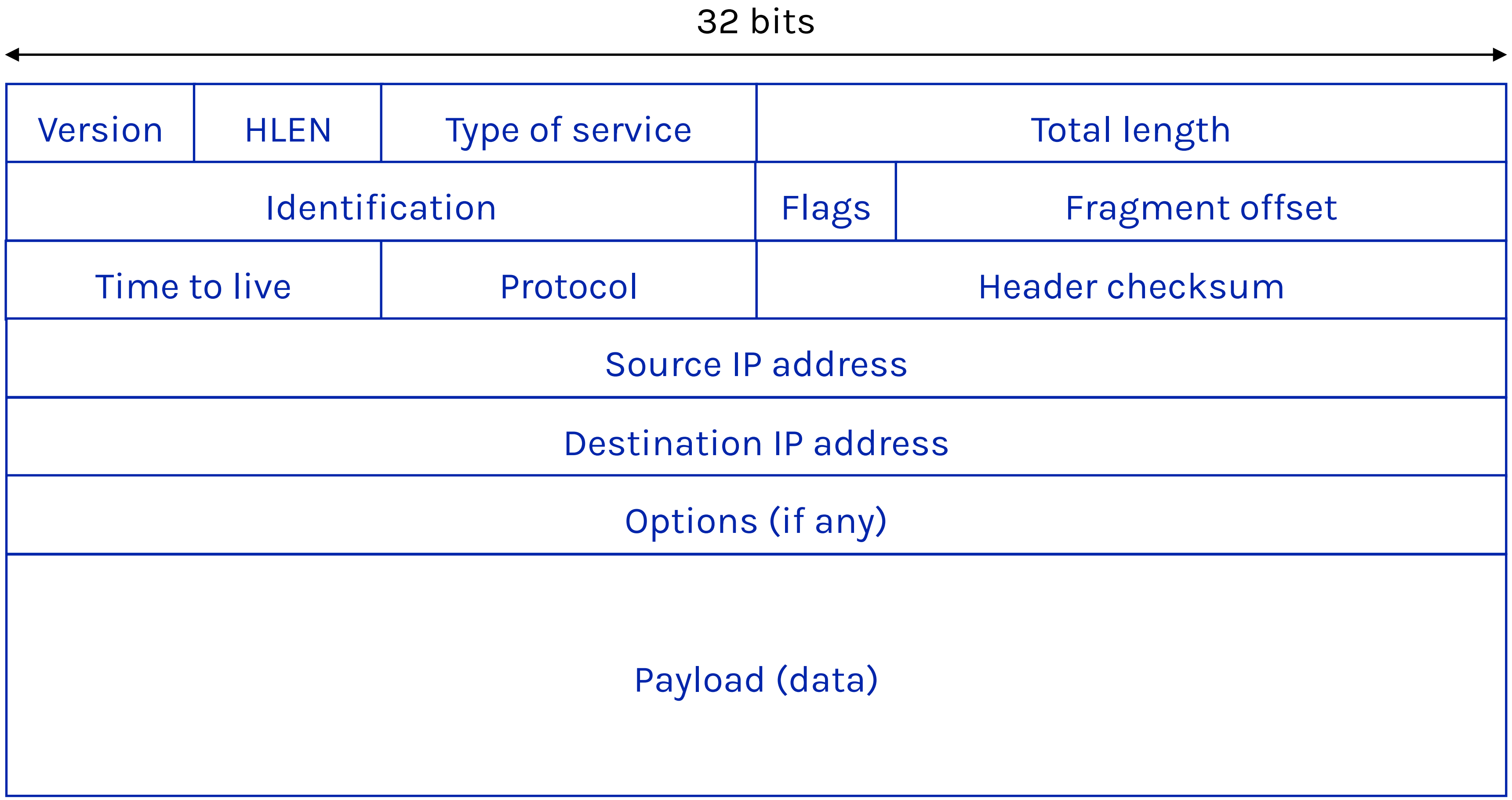


Parent: 129.0.0.0/8

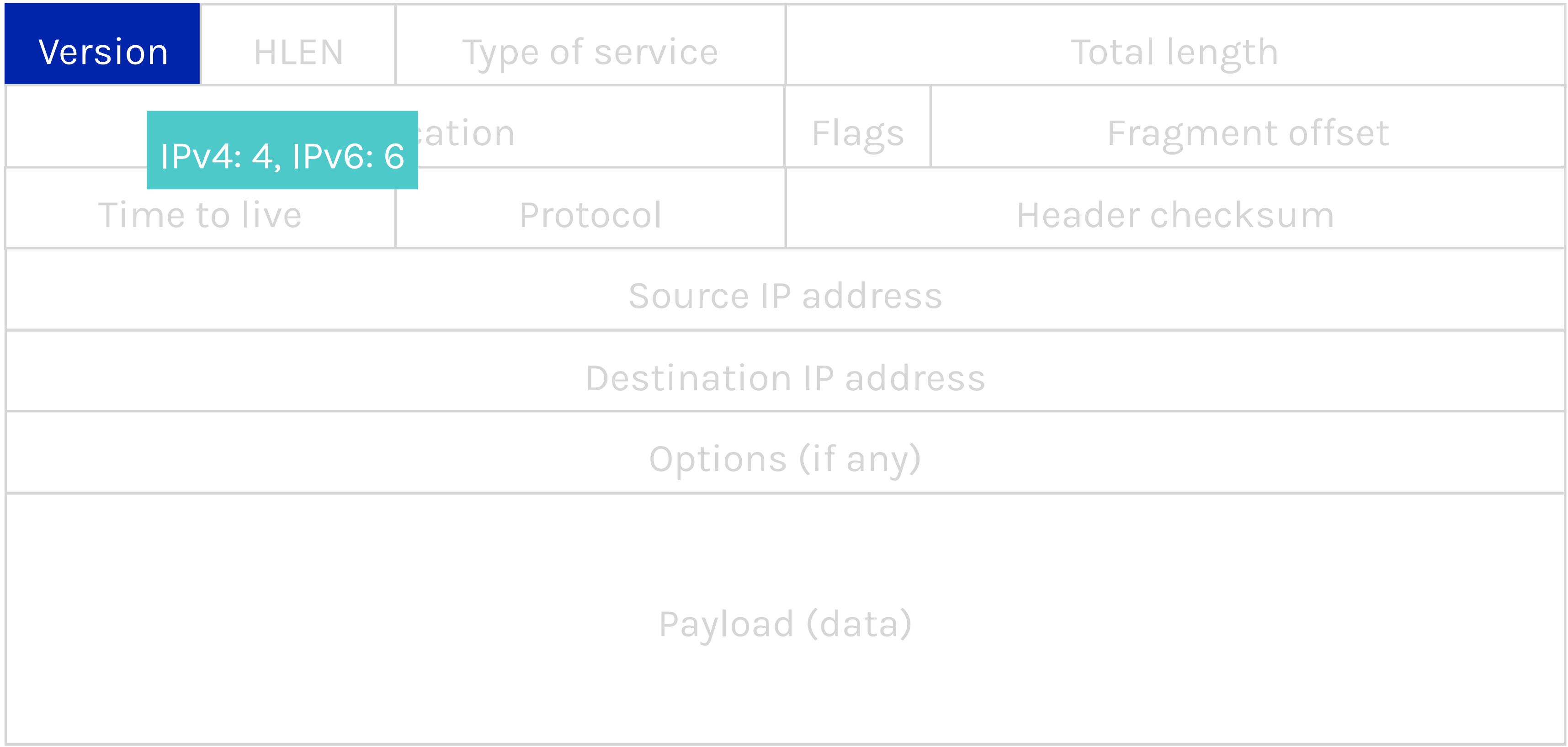
IP prefix	Output	
129.0.0.0/8	port2	
129.132.1.0/24	port2	✖
129.132.2.0/24	port2	✖
129.133.0.0/16	port3	

IP prefix	Output
129.0.0.0/8	port2
129.133.0.0/16	port3

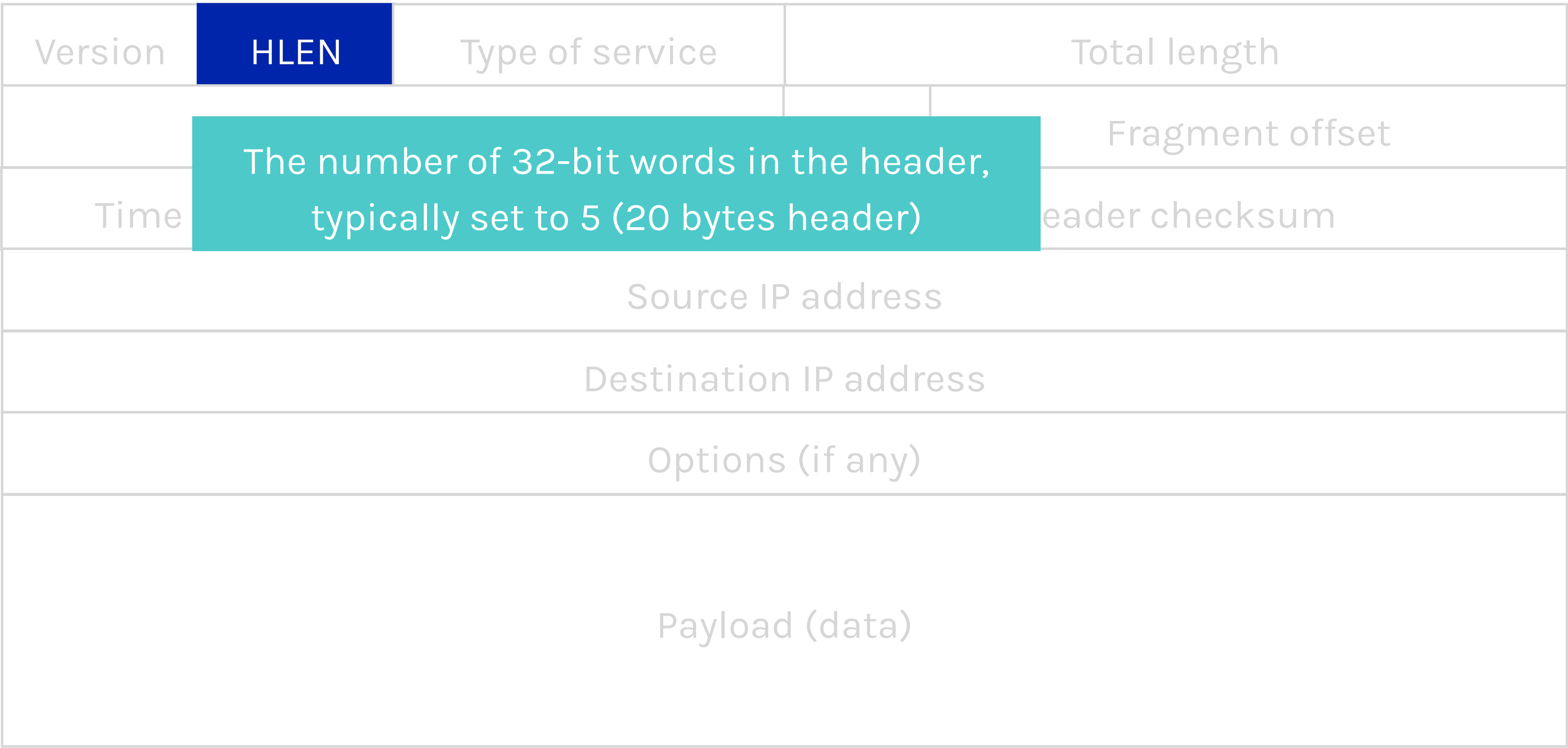
IPv4 packet format



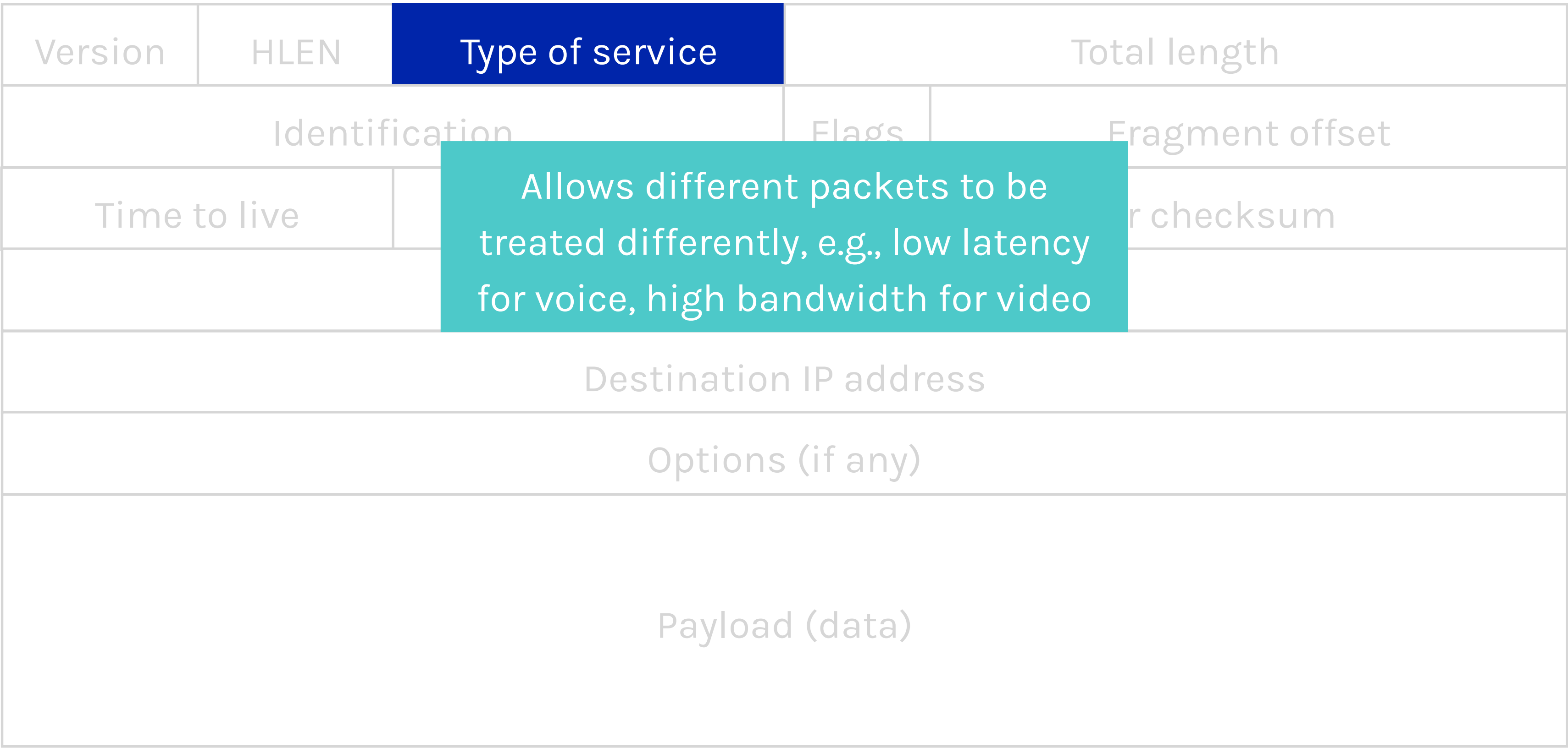
IPv4 packet format



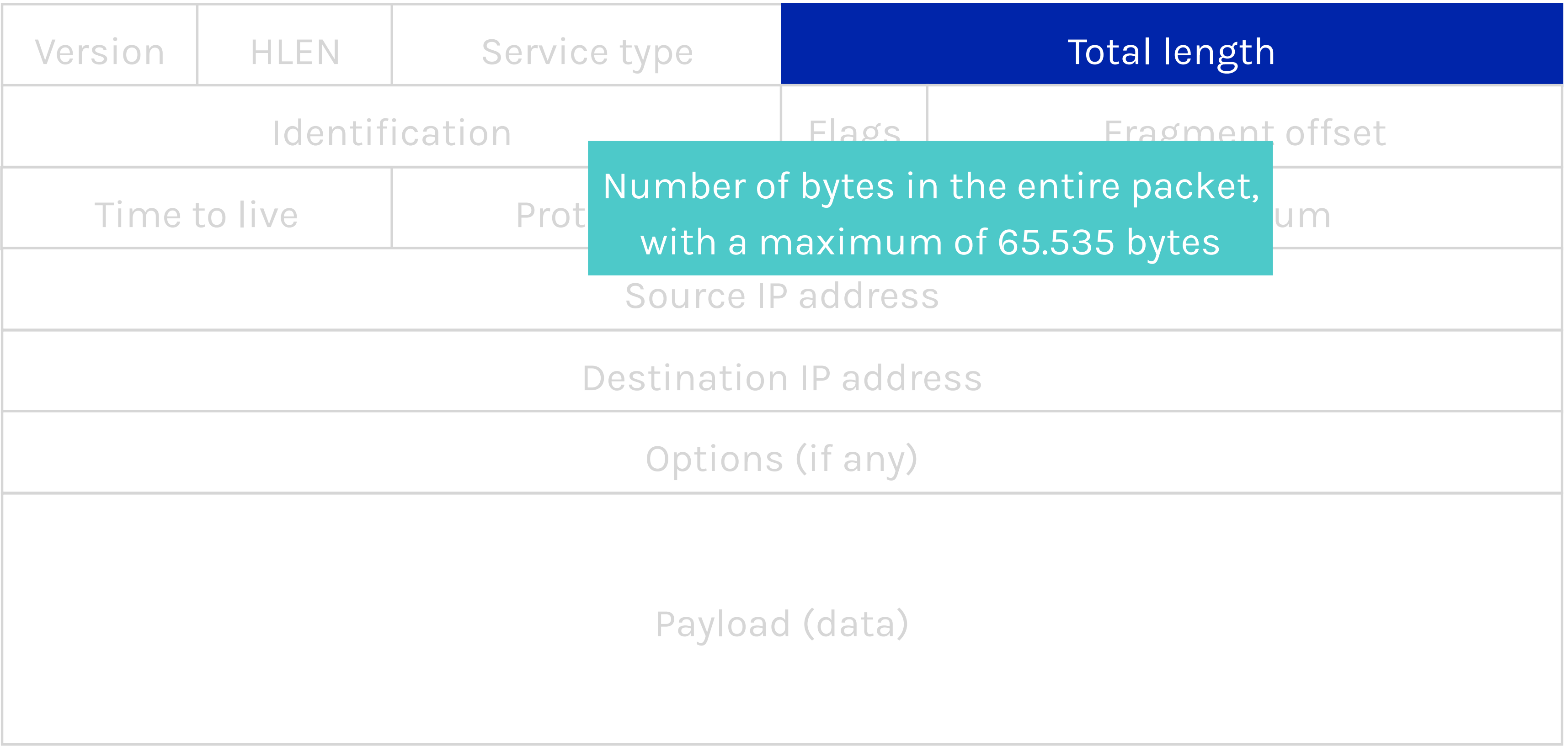
IPv4 packet format



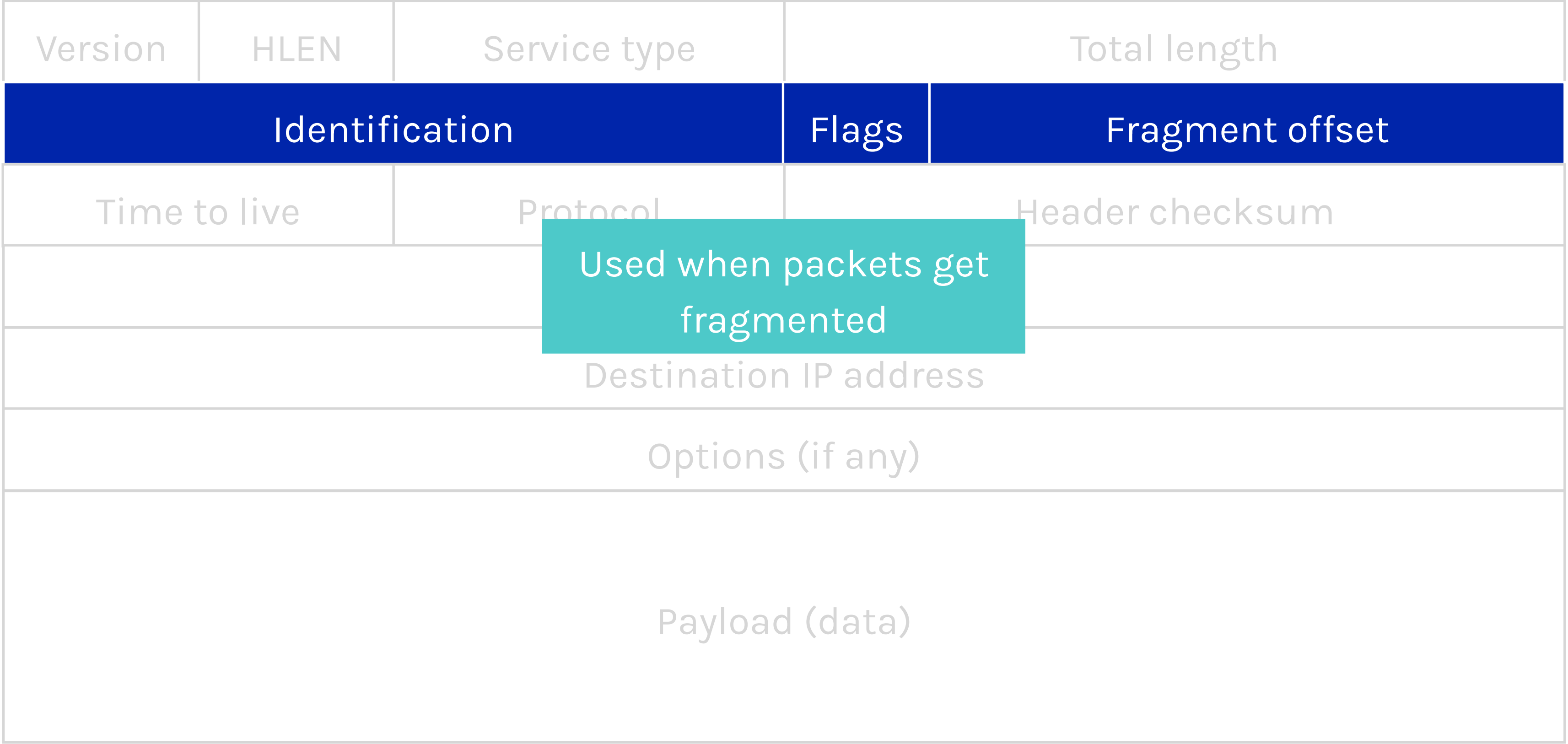
IPv4 packet format



IPv4 packet format



IPv4 packet format



Maximum transmission unit (MTU)

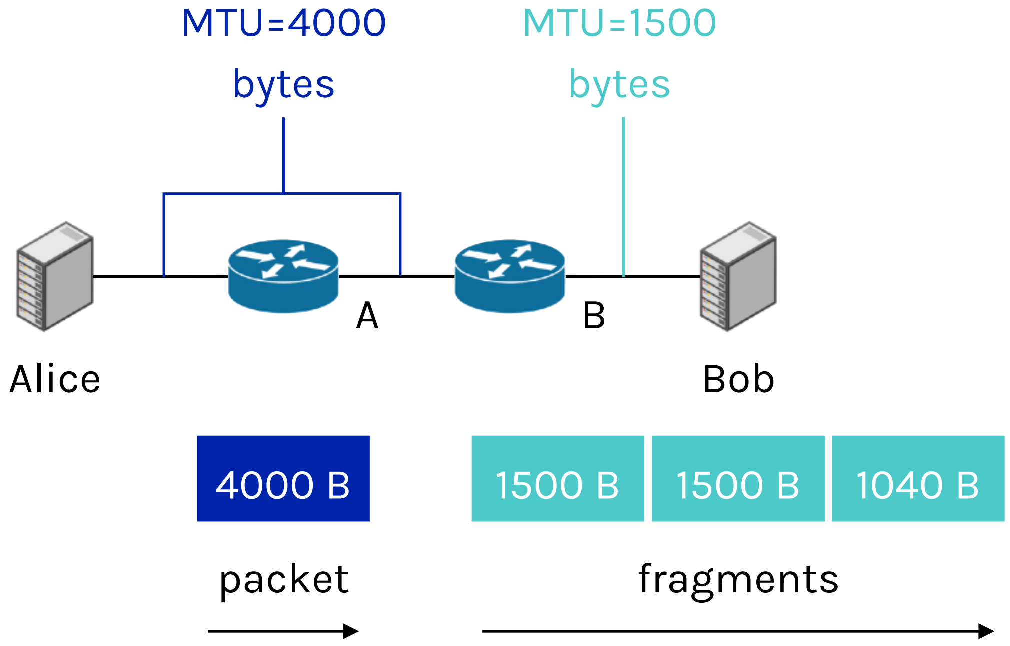
MTU is the maximum number of bytes a link can carry as one unit

- 1500 bytes for normal Ethernet, 9000 for Jumbo frames

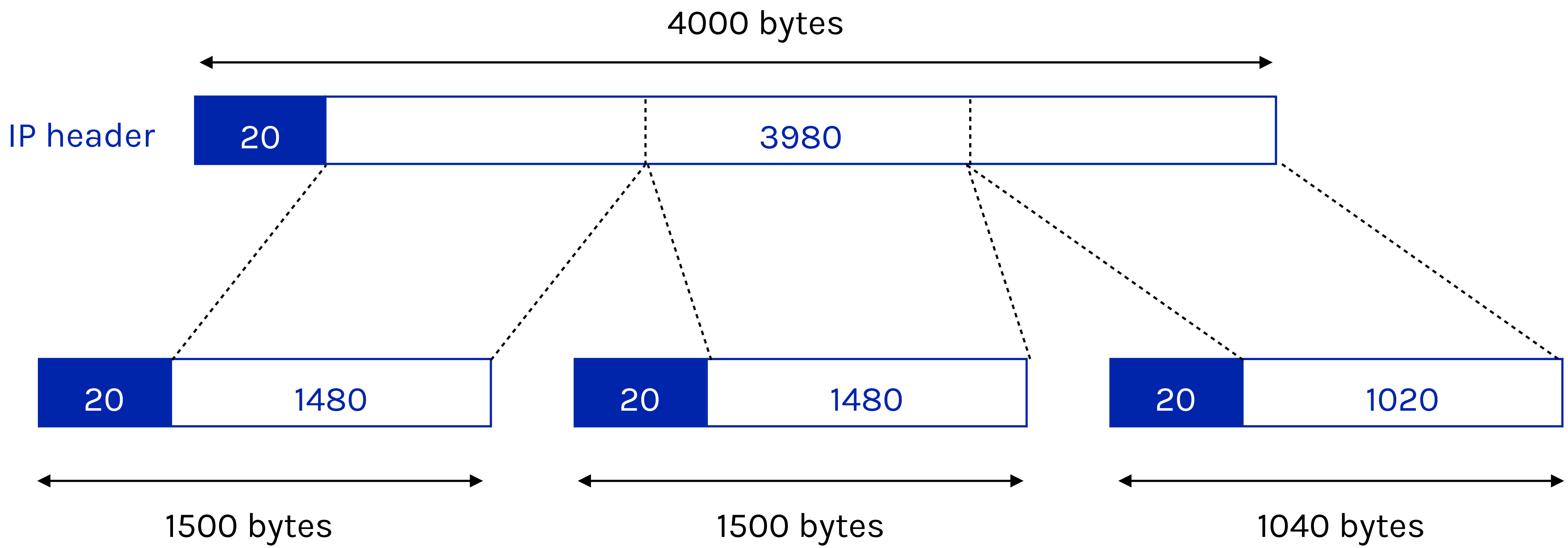
A router fragments a packet if outgoing link MTU < packet size

Fragmented packets are recomposed at the destination

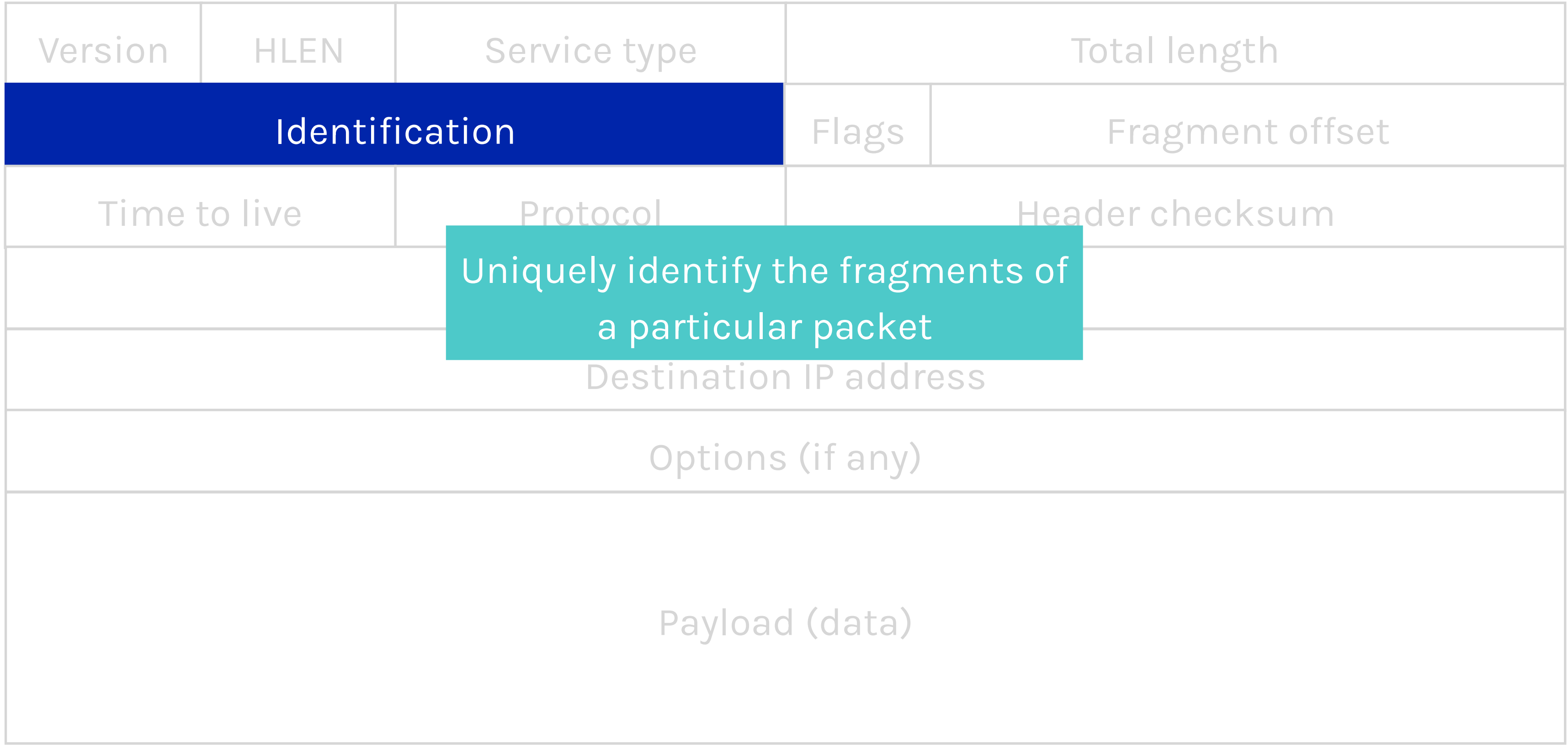
- Why not directly in the network?



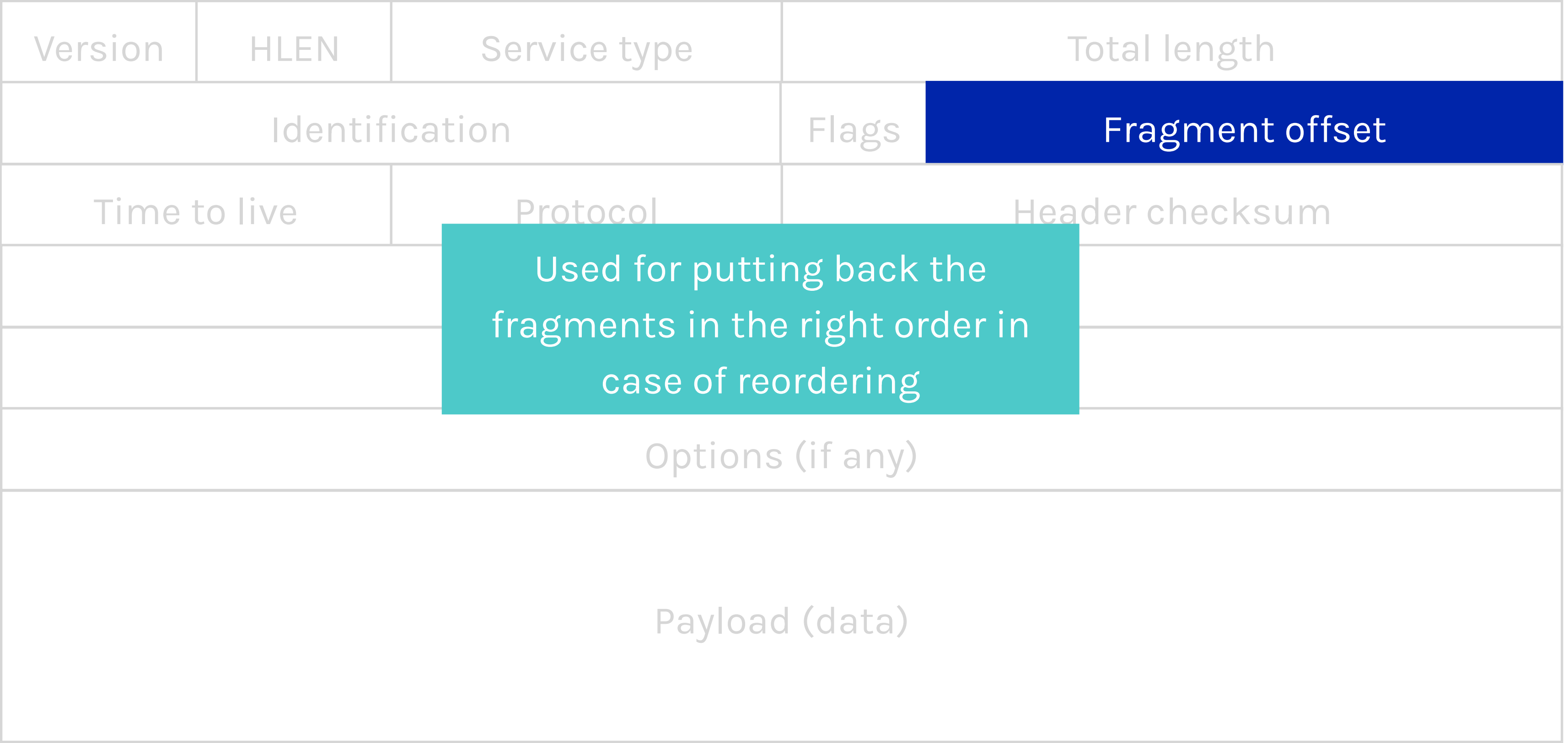
IP fragmentation



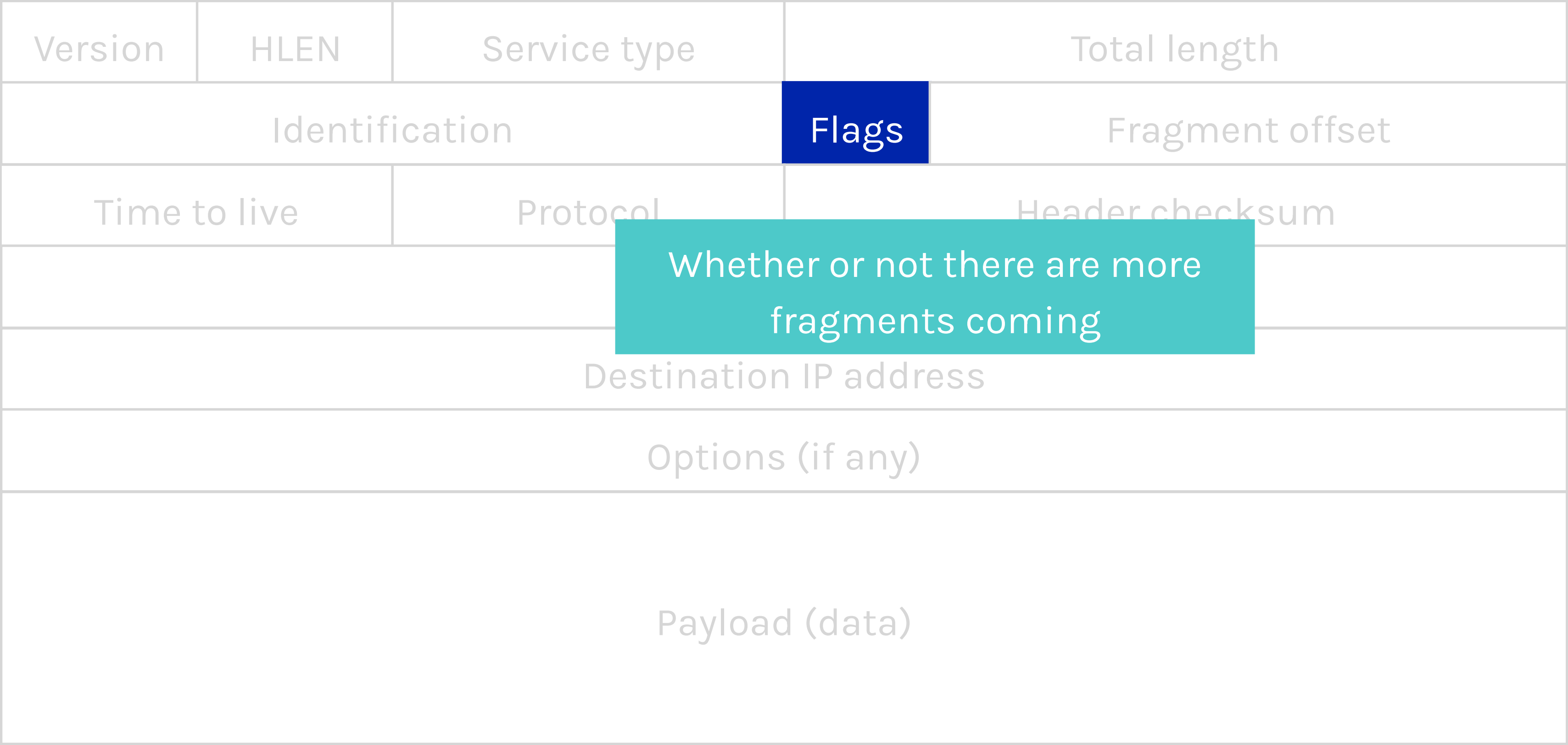
IPv4 packet format



IPv4 packet format



IPv4 packet format



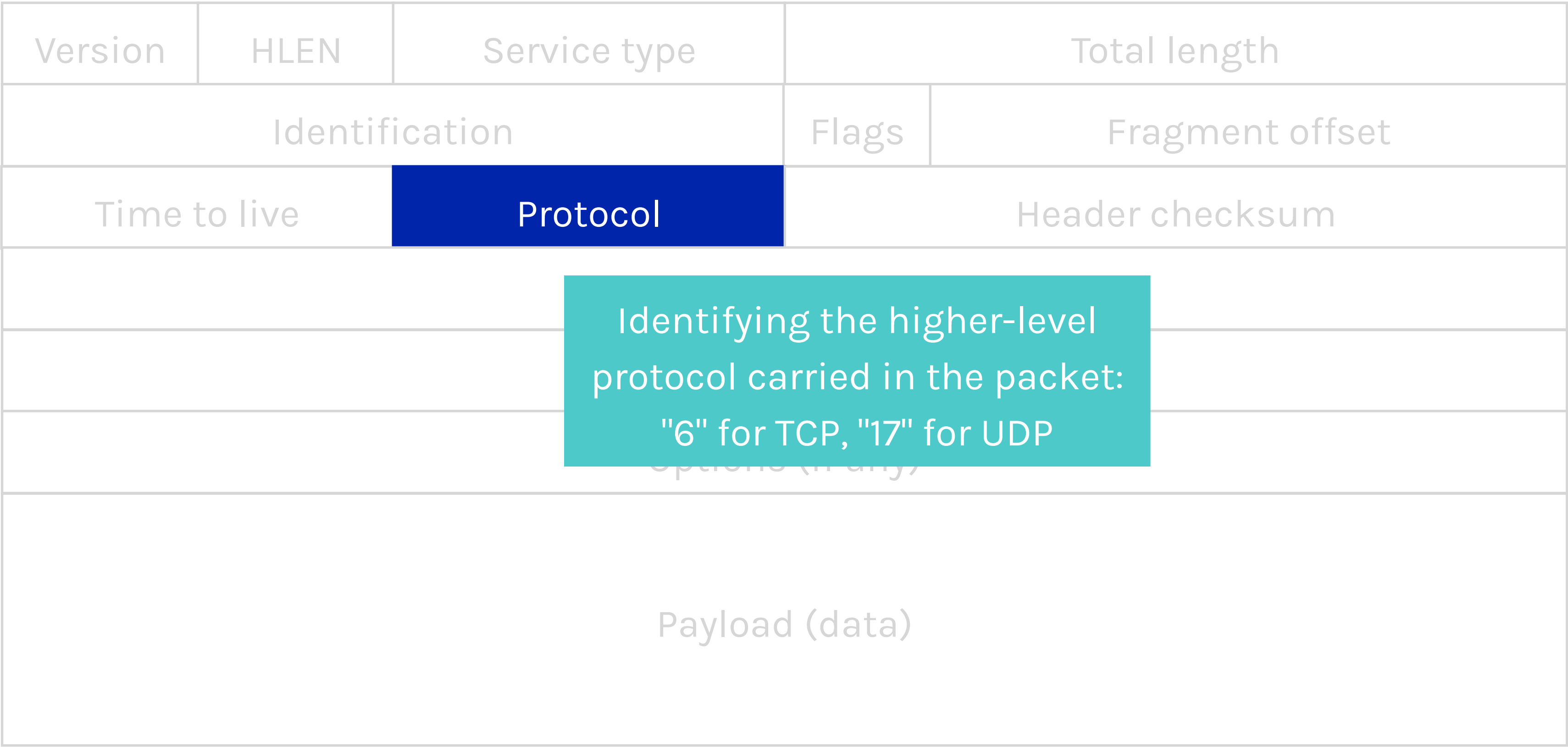
IPv4 packet format

Version	HLEN	Service type	Total length
Identification		Flags	Fragment offset
Time to live		Protocol	Header checksum

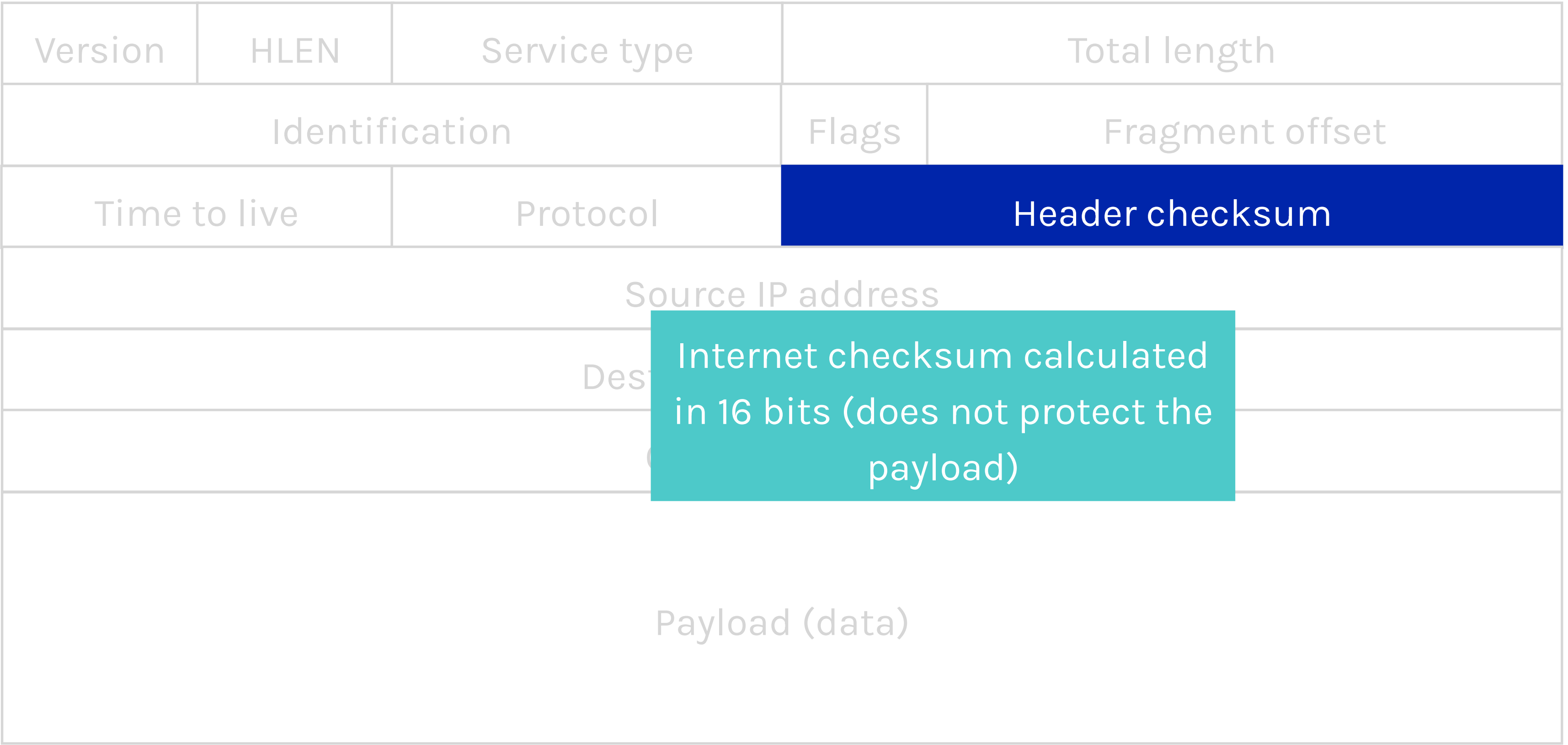
Time-to-live (TTL) is decremented by 1 at each router and the packet is discarded if TTL reaches 0

Default TTL values:
Windows: 128
Linux/Mac: 64
(Can be used for OS fingerprinting)

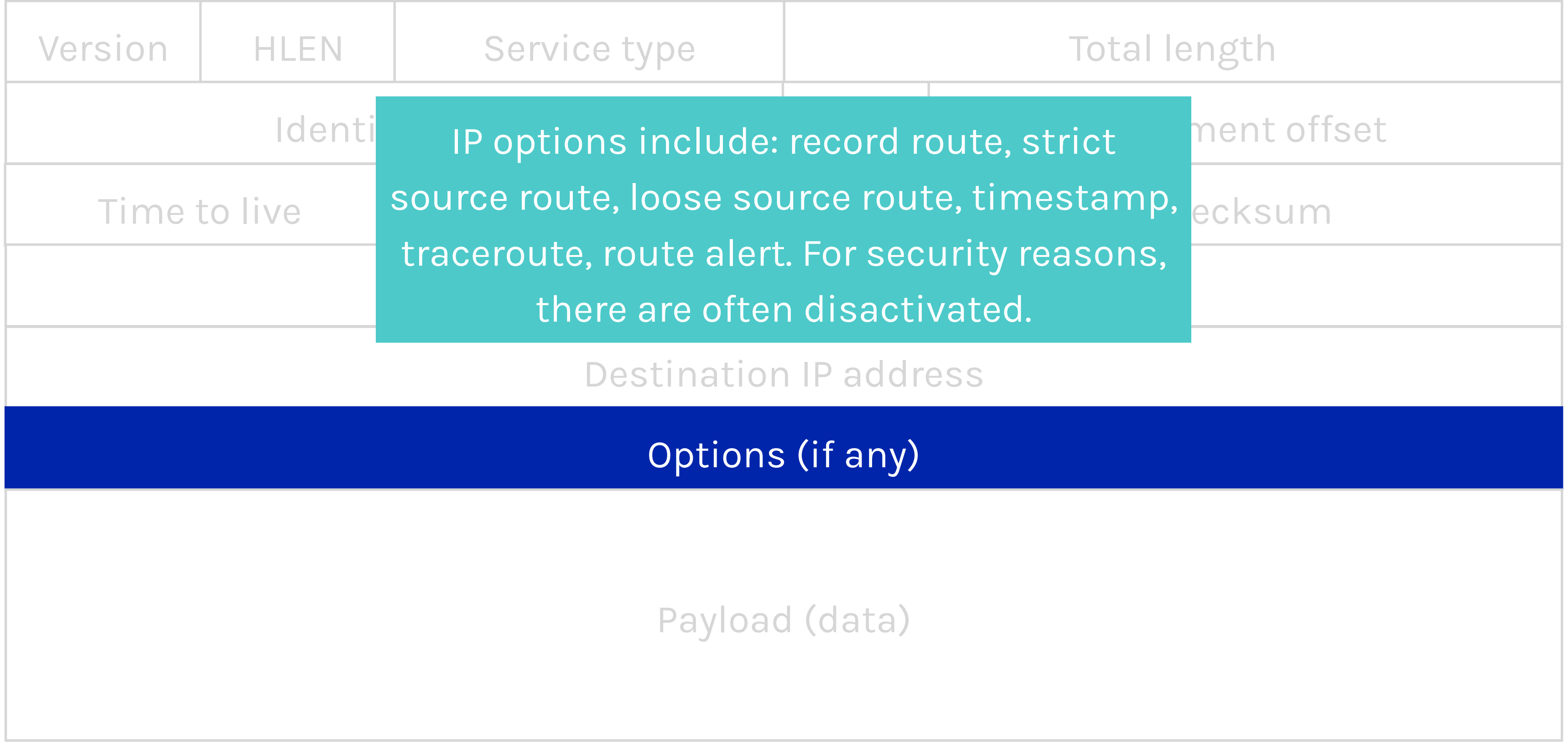
IPv4 packet format



IPv4 packet format



IPv4 packet format



http://www.tcpipguide.com/free/t_IPDatagramOptionsandOptionFormat-2.htm

IPv4 addresses have been exhausted, but they still account for most of the Internet traffic

IPv6 is simpler than IPv4

Removed

- Fragmentation
- Checksum
- Header length

Fragmentation, Checksum → Leave problems to the end-host

Header length → Simplify handling

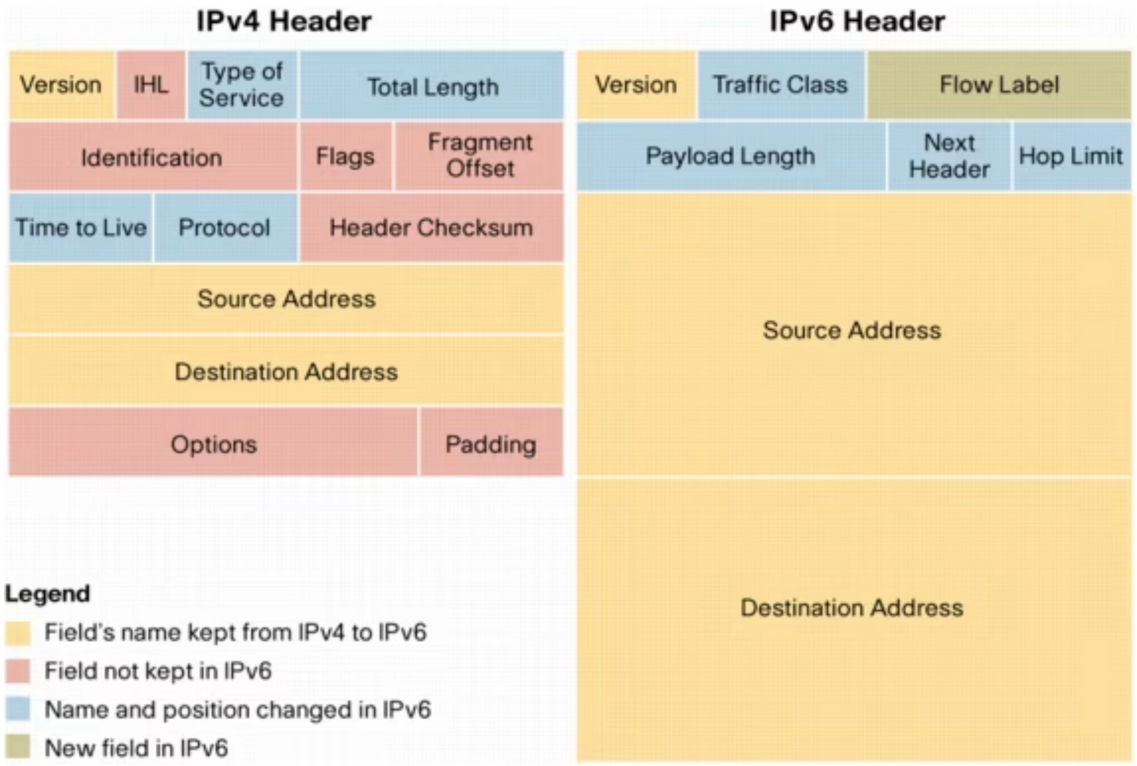
Added

- New options mechanism
- Expanded addresses
- Flow label

New options mechanism → Simplify handling

Flow label → Flexibility

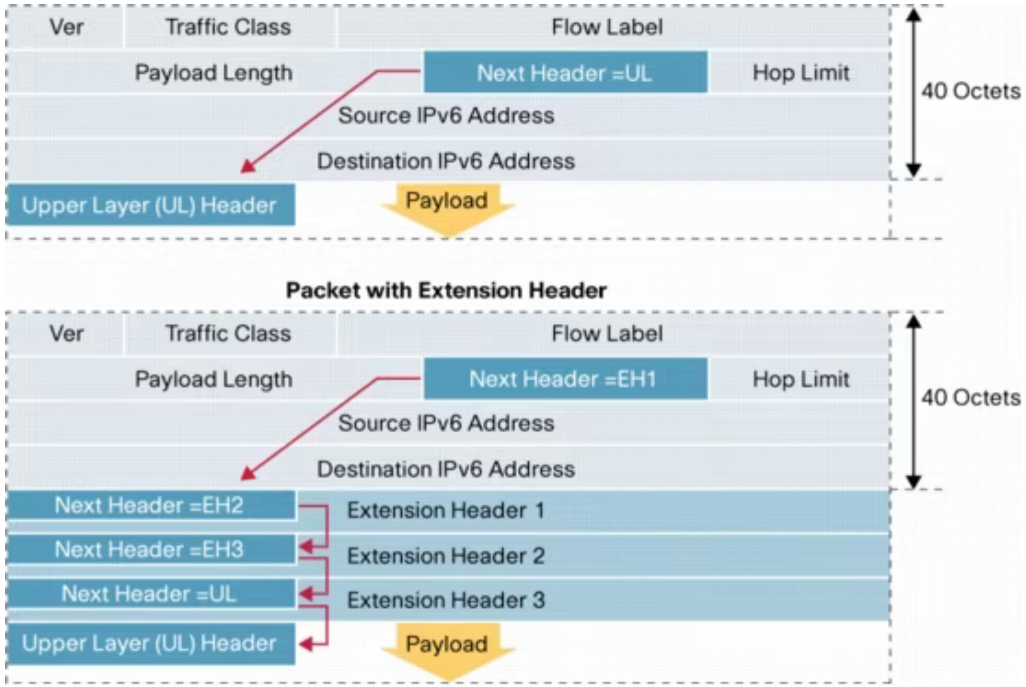
IPv4 vs. IPv6



https://www.cisco.com/en/US/technologies/tk648/tk872/technologies_white_paper0900aecd8054d37d.html

IPv6 options

Enables to insert arbitrary options in the header (see RFC 2460)



https://www.cisco.com/en/US/technologies/tk648/tk872/technologies_white_paper0900aecd8054d37d.html

IPv6 unicast address

Hierarchically allocated, similar to global IPv4 addresses

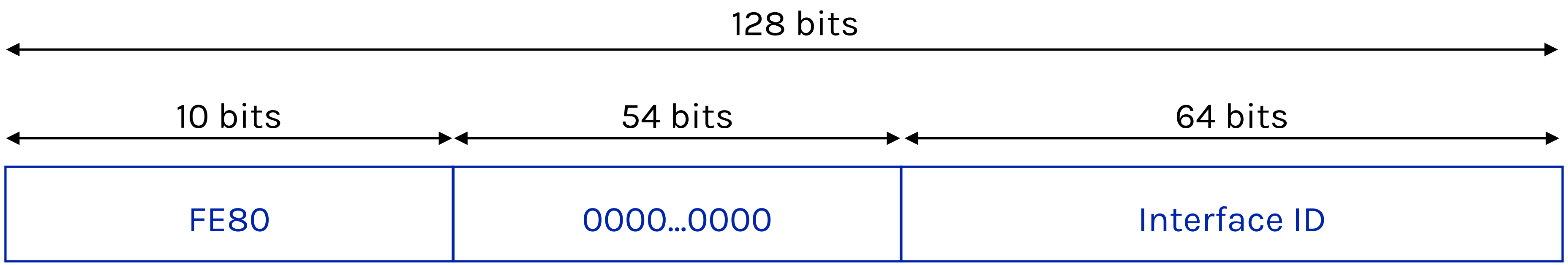
128 bits

N bits	M bits	128-N-M bits
Global routing prefix	Subnet ID	Interface ID
Identifies the ISP responsible for the address	A subnet or a customer of this ISP	Usually 64 bits, based on the MAC address (EUI-64, deprecated), obtained from a server

Currently, only 2000::/3 is used for global unicast; all addresses are in the range of 2000 to 3FFFF

IPv6 link-local address

Same as private IPv4 addresses

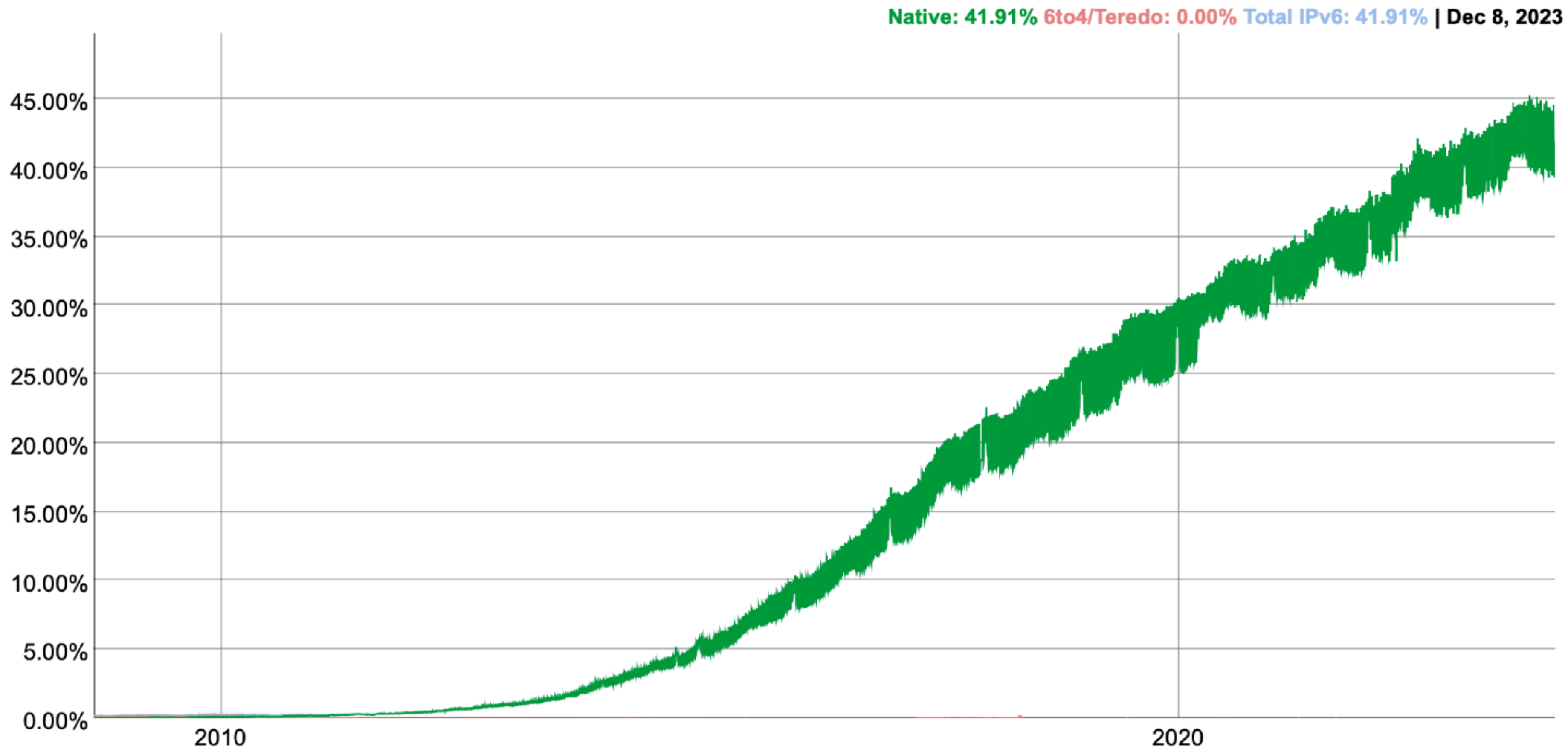


```
en0: flags=8863<UP,BROADCAST,SMART,RUNNING,SIMPLEX,MULTICAST> mtu 1500
        options=400<CHANNEL_IO>
        ether 6c:7e:67:d7:f5:71
        inet6 fe80::c8c:a9ae:f3b1:8b9d%en0 prefixlen 64 secured scopeid 0xf
        inet 192.168.2.104 netmask 0xffffff00 broadcast 192.168.2.255
        inet6 2003:d0:271b:2f58:859:fcea:5d83:f3bb prefixlen 64 autoconf secured
        inet6 2003:d0:271b:2f58:f082:f7a3:7b05:eb99 prefixlen 64 autoconf temporary
        inet6 2003:d0:271b:2fbf:82b:fb86:b0ba:9a79 prefixlen 64 autoconf secured
        inet6 2003:d0:271b:2fbf:8588:7d45:e7cc:6c51 prefixlen 64 autoconf temporary
        inet6 2003:d0:271b:2fb8:40d:e3fb:2690:5fb9 prefixlen 64 autoconf secured
        inet6 2003:d0:271b:2fb8:38d2:1256:1047:aa04 prefixlen 64 autoconf temporary
        nd6 options=201<PERFORMNUD,DAD>
        media: autoselect
        status: active
```

Each host/router must generate a link-local address for each of its interfaces

An interface can have multiple IPv6 addresses

IPv6 adoption



https://www.google.com/intl/en/ipv6/statistics.html

IPv6 deployment challenges

Requires every device to support it (all routers, end-hosts, middleboxes, applications…)

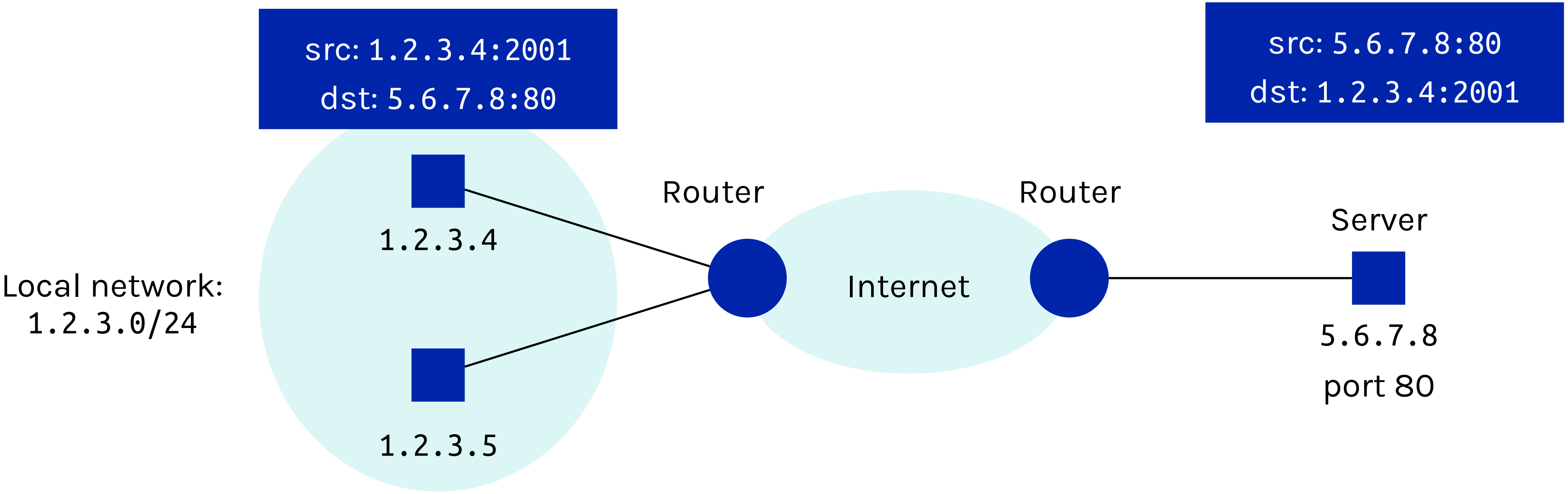
Most of IPv6 features were back-ported to IPv4 (no obvious advantages in using IPv6)

Network Address Translation (NAT) is working well

Network Address Translation

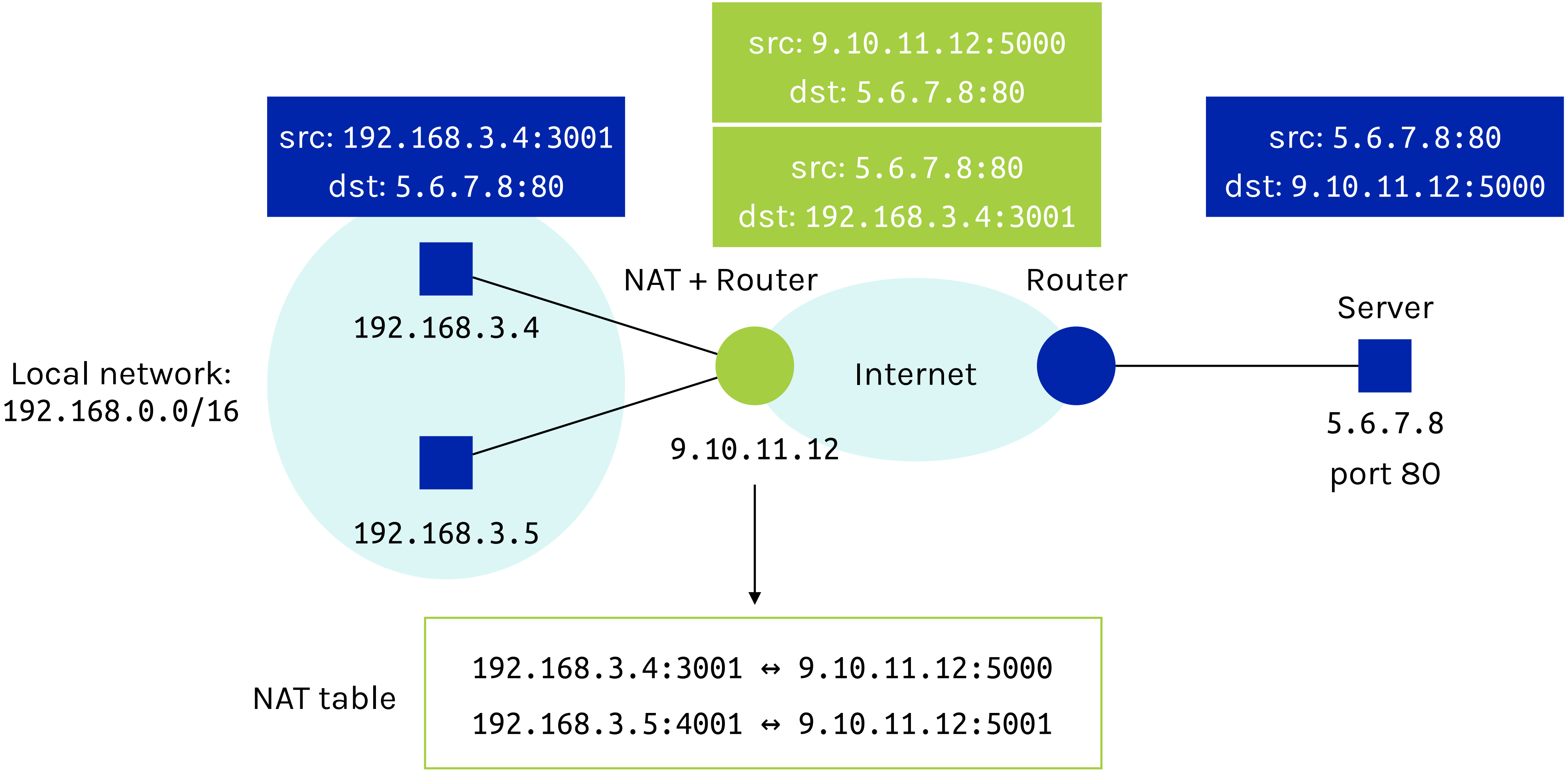
The Internet before NAT

Every machine connected to the Internet had a unique IP



The Internet before NAT

Every machine connected to the Internet had a unique IP



NAT (dis)advantages

Better privacy/anonymization

- All hosts in one network get the same public IP
- But hosts may still be identified by cookies, browser version,...

Better security

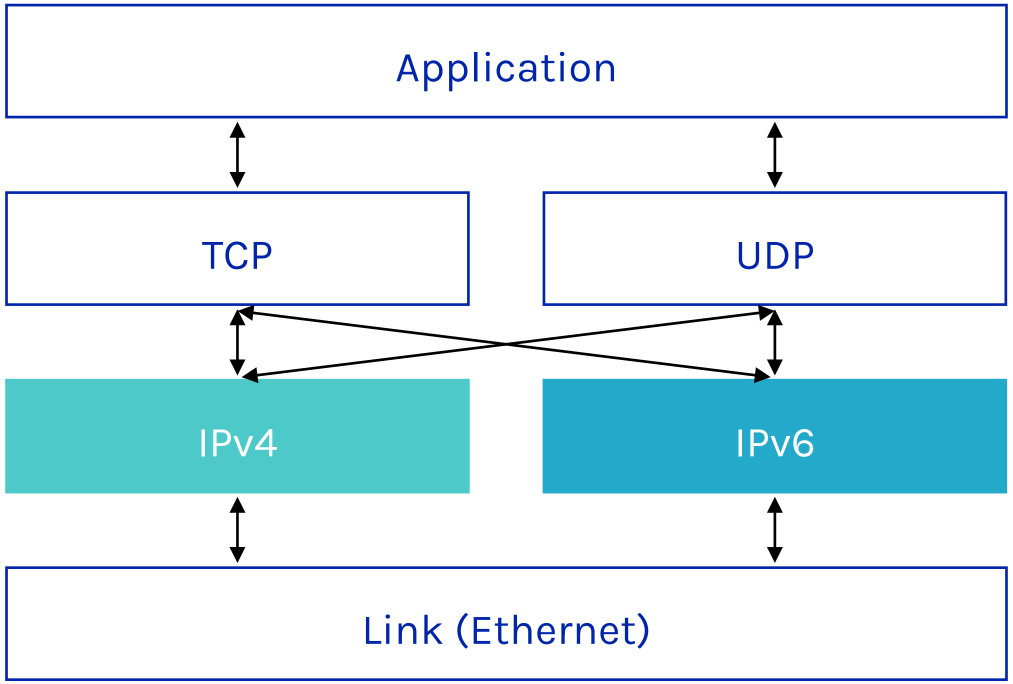
- From the outside you cannot directly reach the hosts
- Problematic for applications like online gaming

Limited scalability (limited mapping table)

- Example: WiFi access problems in public places often due to full NAT table

IPv4 to IPv6 transition

Duo stack approach used in many OSes and applications



IPv4 to IPv6 transition mechanisms

6in4

6to4

Teredo

SIIT

6rd

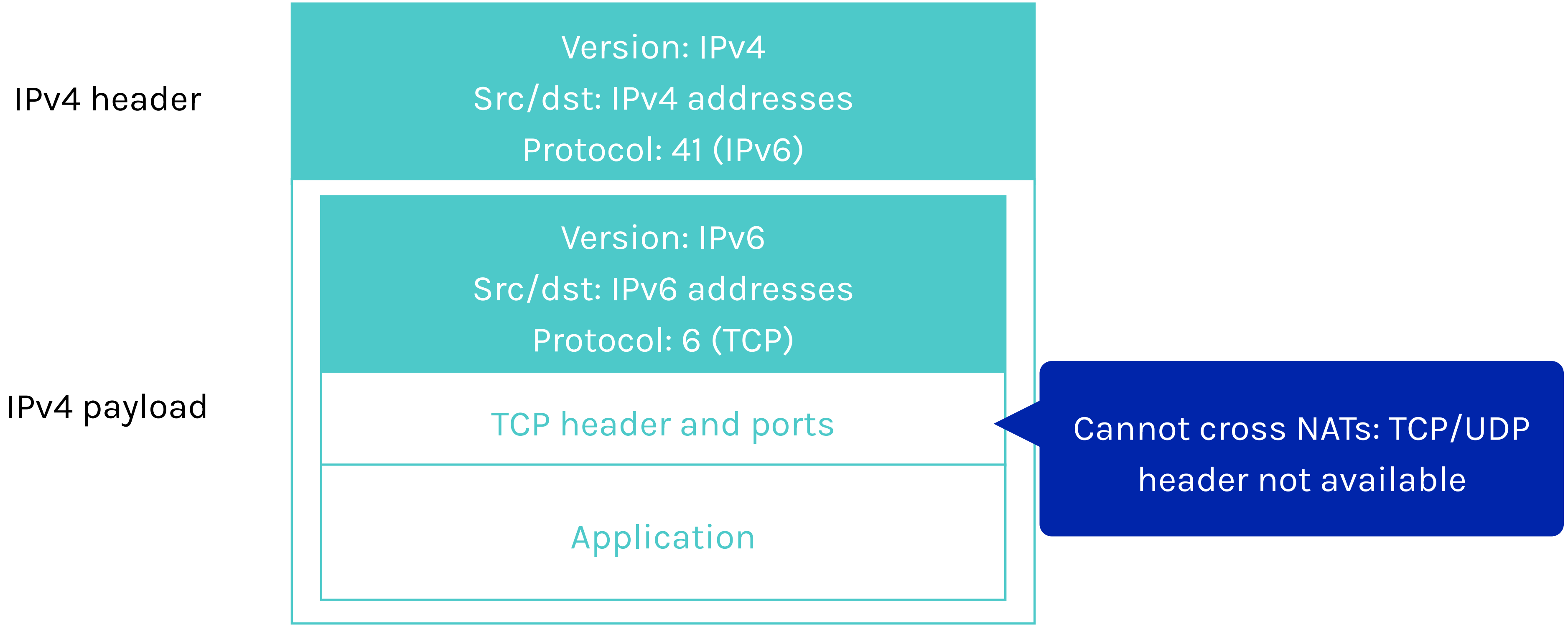
GRE

AYiYA

...

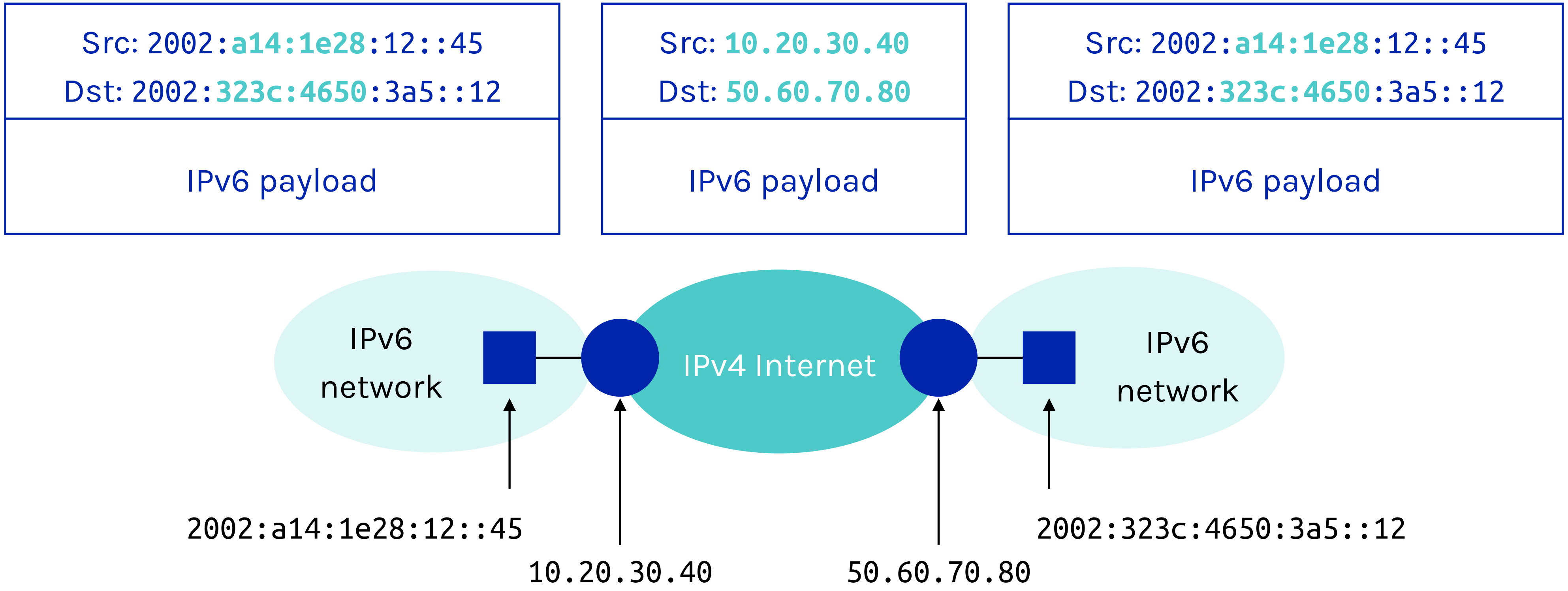
6in4

Transmits IPv6 packets over statically configured IPv4 tunnels

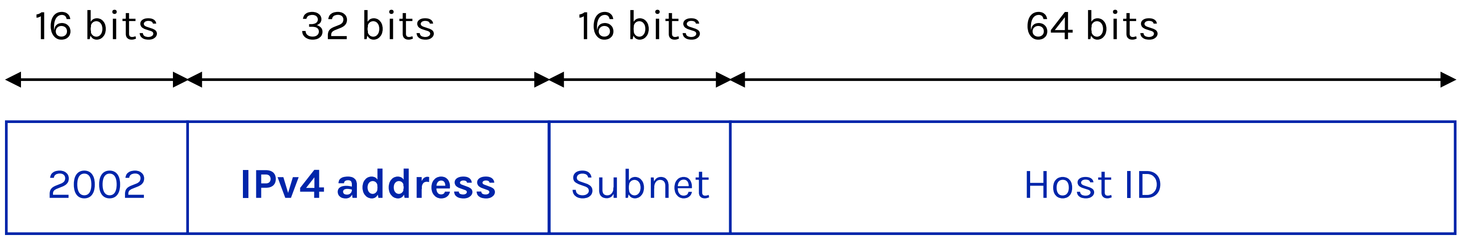


6to4

Transmits IPv6 packets over IPv4 networks without explicit tunnels



Special IPv6 addresses in 6to4

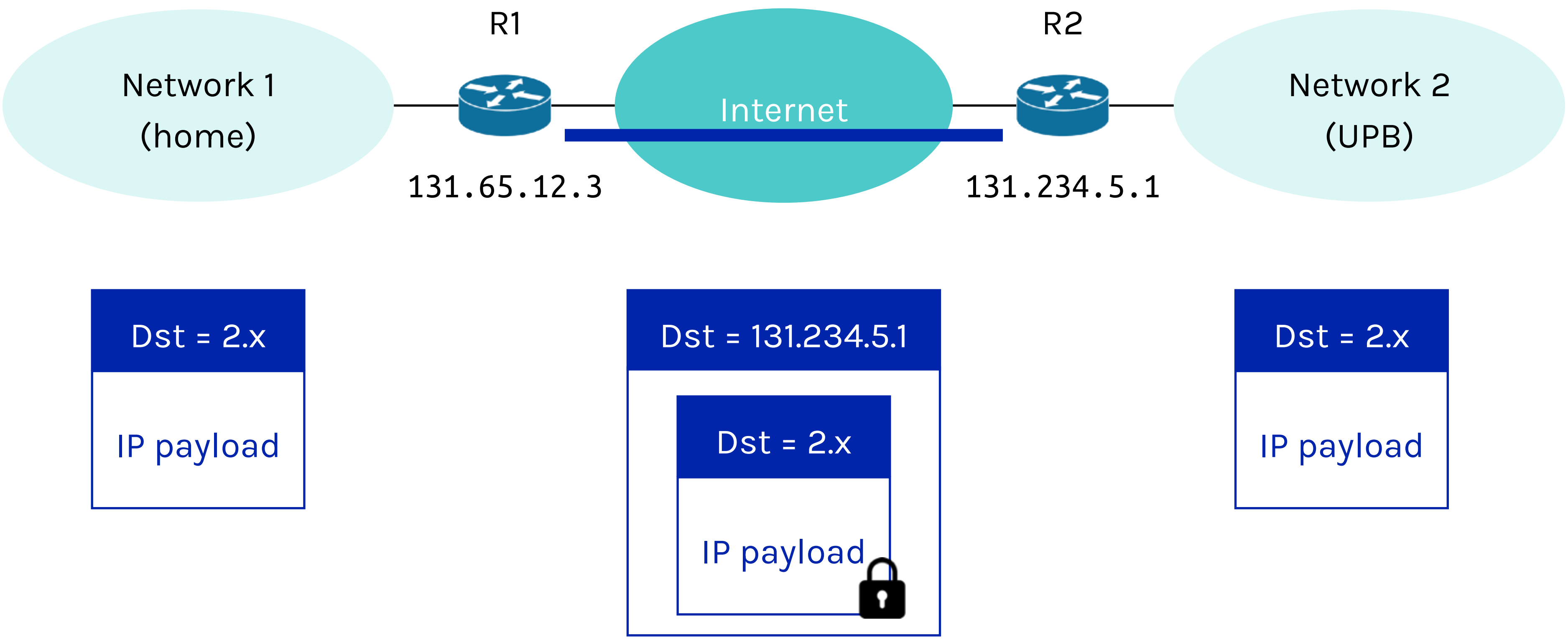


IPv4 `192.15.3.73`

`c0.0f.03.49`

IPv6 `2002:`**`c00f:0349`**`::/49`

Virtual Private Network (VPN)



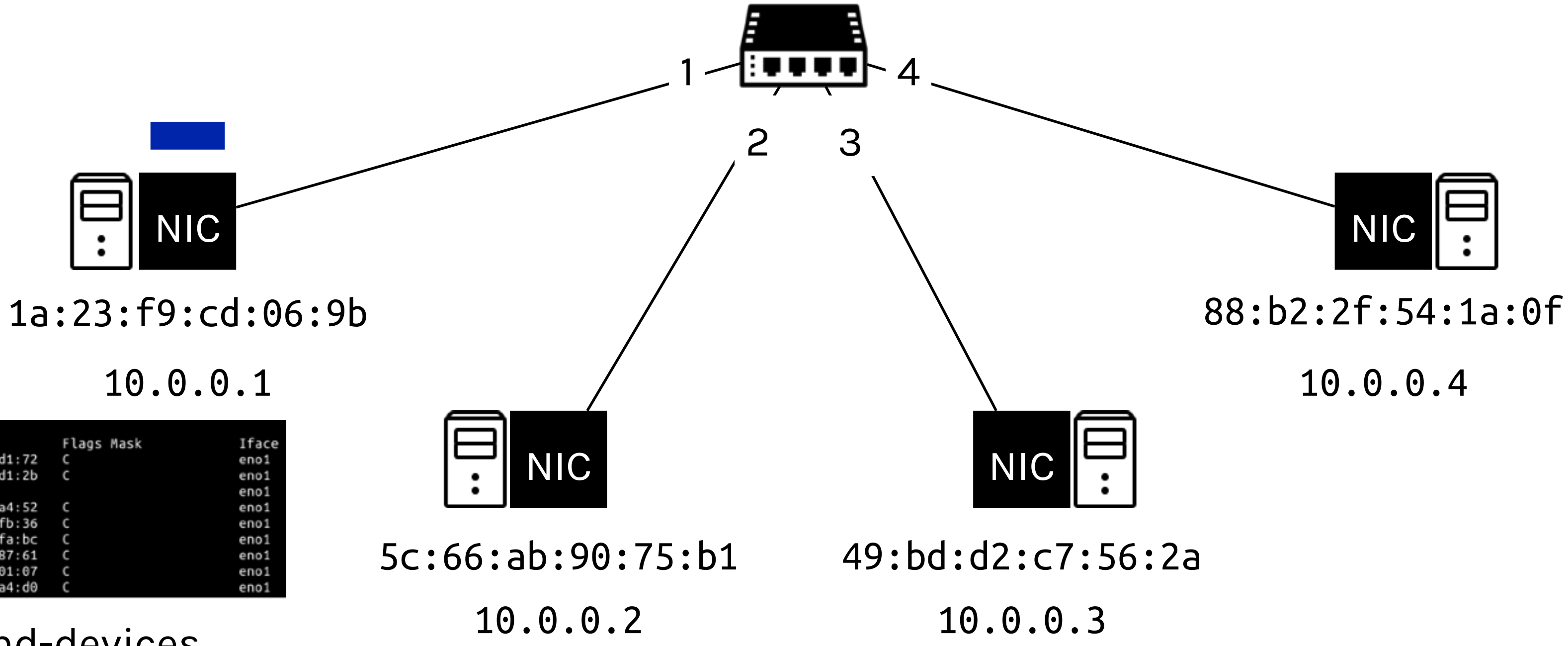
https://en.wikipedia.org/wiki/Virtual_private_network

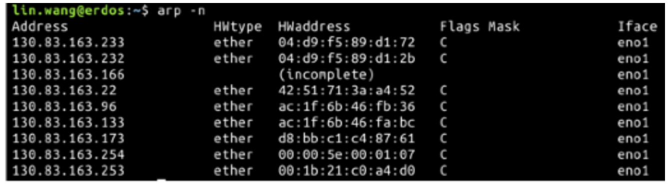
Helper Protocols

Address Resolution Protocol (ARP)

ARP query Whoever has the IP of `10.0.0.4`, please respond to me

ARP response That is me, and here is my MAC address `88:b2:2f:54:1a:0f`





```
lin.wang@erdos:~$ arp -n
Address                  HWtype  HWaddress           Flags Mask            Iface
130.83.163.233           ether   04:d9:f5:89:d1:72   C                     eno1
130.83.163.232           ether   04:d9:f5:89:d1:2b   C                     eno1
130.83.163.166                   (incomplete)                              eno1
130.83.163.22            ether   42:51:71:3a:a4:52   C                     eno1
130.83.163.96            ether   ac:1f:6b:46:fb:36   C                     eno1
130.83.163.133           ether   ac:1f:6b:46:fa:bc   C                     eno1
130.83.163.173           ether   d8:bb:c1:c4:87:61   C                     eno1
130.83.163.254           ether   00:00:5e:00:01:07   C                     eno1
130.83.163.253           ether   00:1b:21:c0:a4:d0   C                     eno1
```

ARP table on end-devices

https://datatracker.ietf.org/doc/html/rfc826

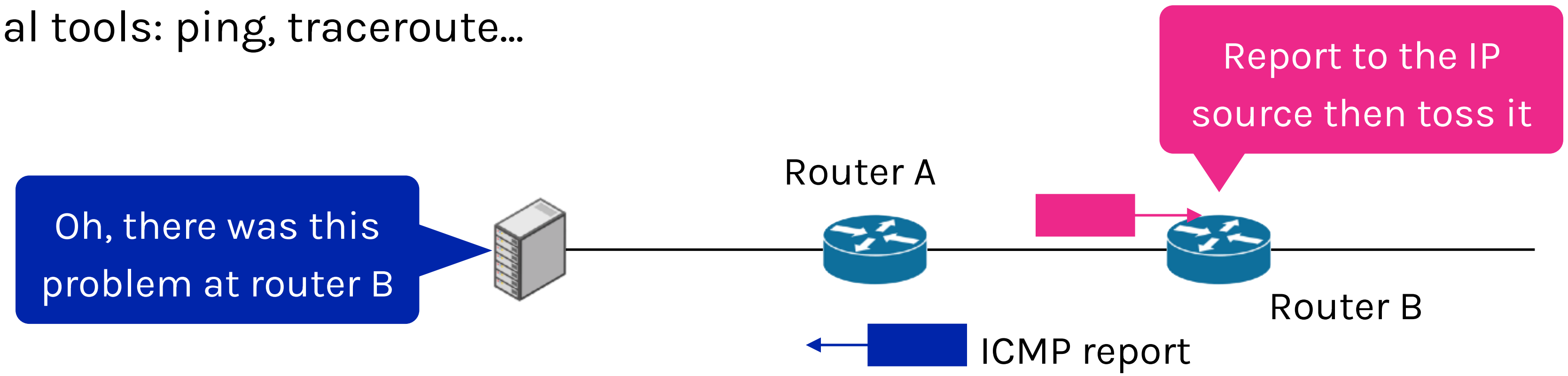
Internet Control Message Protocol (ICMP)

ICMP is a companion protocol to IP

- Sits on top of IP (IP protocol = 1)

Provides error report and testing

- Error is at router while forwarding
- Usual tools: ping, traceroute…



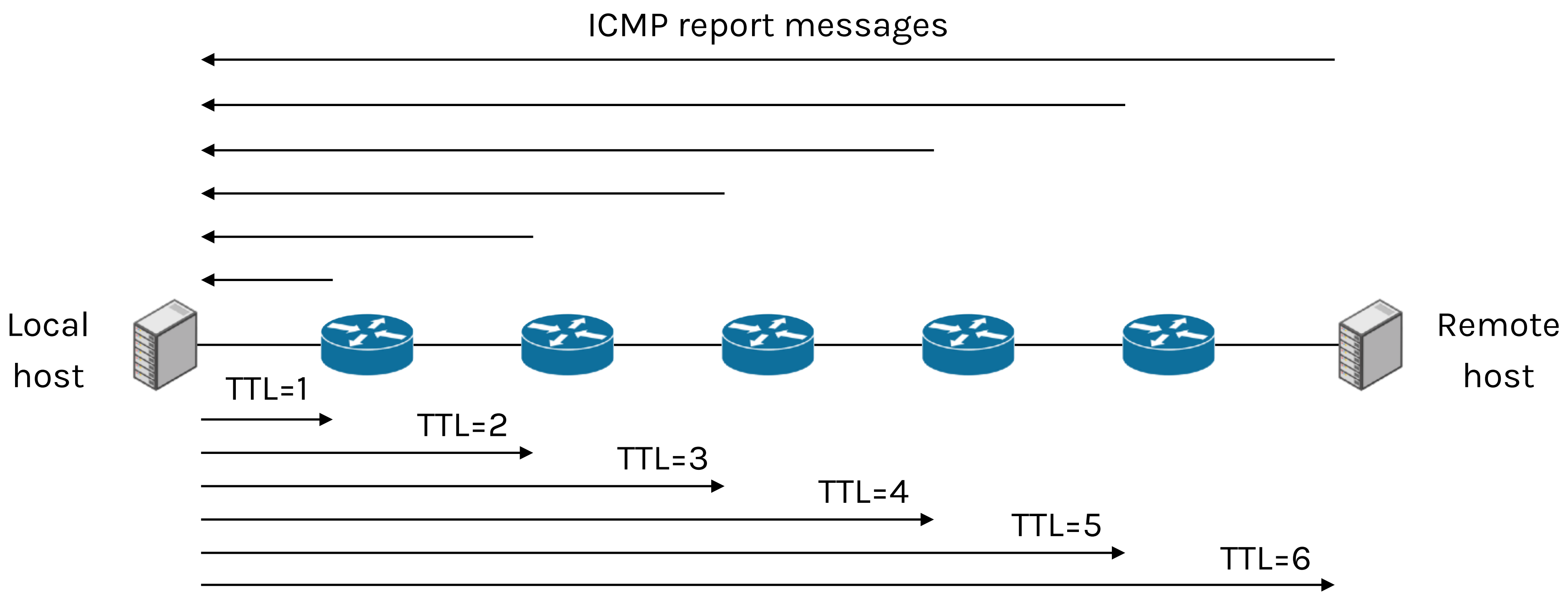
ICMP message format

Portion of offending packet,
starting with its IP header

Src=router, Dst=A Protocol=1	Type=X, Code=Y	Src=A, Dst=B XXXXXXXX
IP header	ICMP header	ICMP data

Type / code	Name	Usage
3 / 0 or 1	Dest. unreachable (net or host)	Lack of connectivity
3 / 4	Dest. unreachable (fragment)	Path MTU discovery
11 / 0	Time exceeded (transit)	Traceroute
8 or 0 / 0	Echo request or reply	Ping (testing, not error)

Traceroute


ICMP report messages
Local
host
Remote
host
TTL=1
TTL=2
TTL=3
TTL=4
TTL=5
TTL=6

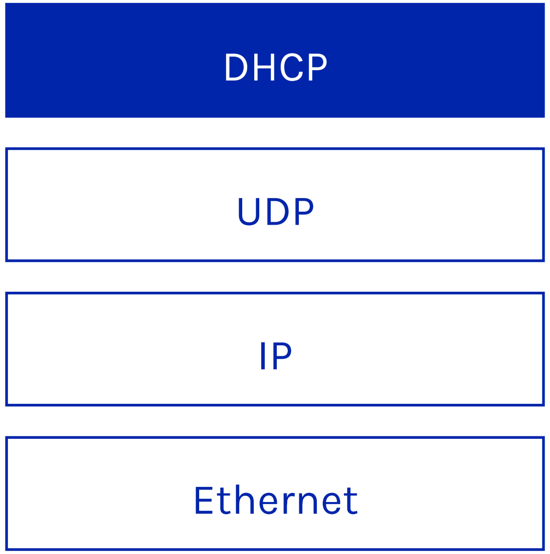
Dynamic Host Configuration Protocol (DHCP)

Leases IP addresses to nodes and provides other parameters

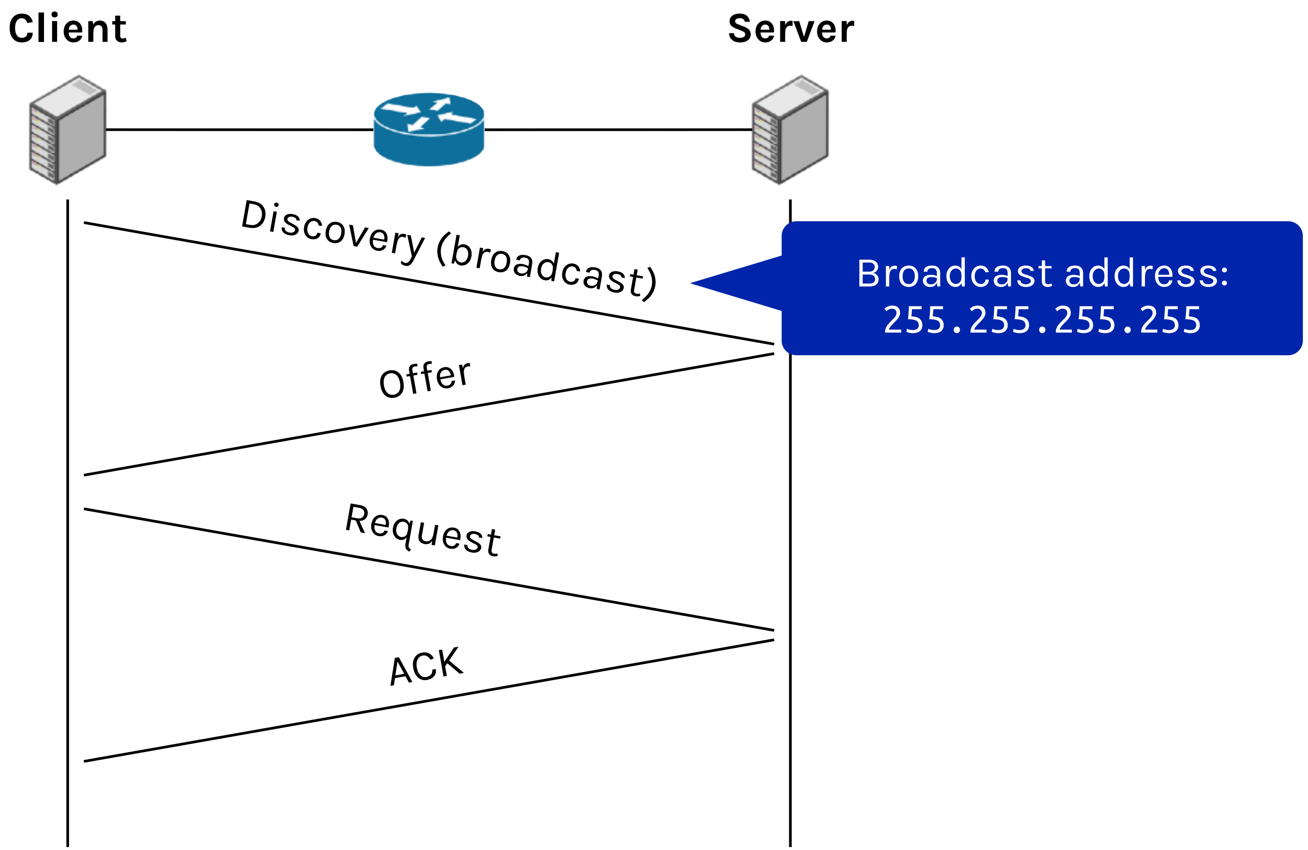
- Network prefix
- Address of local router (gateway)
- DNS server, time server,...

DHCP is a client-server application

- Uses UDP ports 67, 68



DHCP messages



The server may be replicated for reliability

Summary

Inter-networking

- Internet narrow waist
- IP address and prefix

IP forwarding

- Router architecture
- Prefix matching
- IPv4 packet format
- IPv6

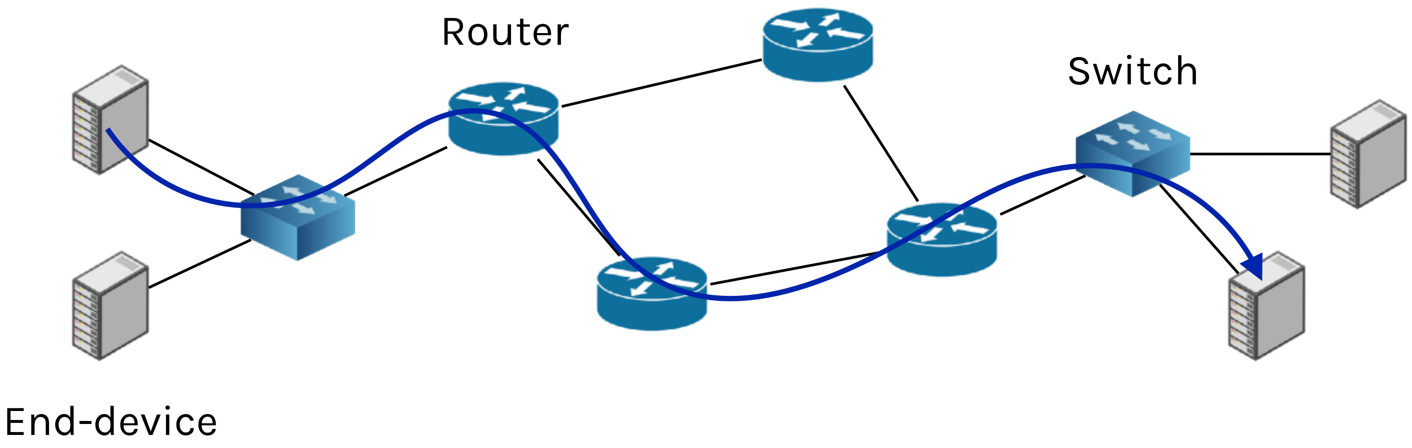
Network Address Translation

- NAT ideas
- IPv4 to IPv6 transition

Helper protocols

- ARP
- ICMP and traceroute
- DHCP

Next time: network layer



How to construct the routing path and navigate through the Internet?

Further reading material

Andrew S. Tanenbaum, David J. Wetherall. Computer Networks (5th edition).

- Section 5.5: Internetworking

Larry Peterson, Bruce Davie. Computer Networks: A Systems Approach.

- Section 3.3 Internet (IP)