Computer Networks

IP – The Internet Protocol

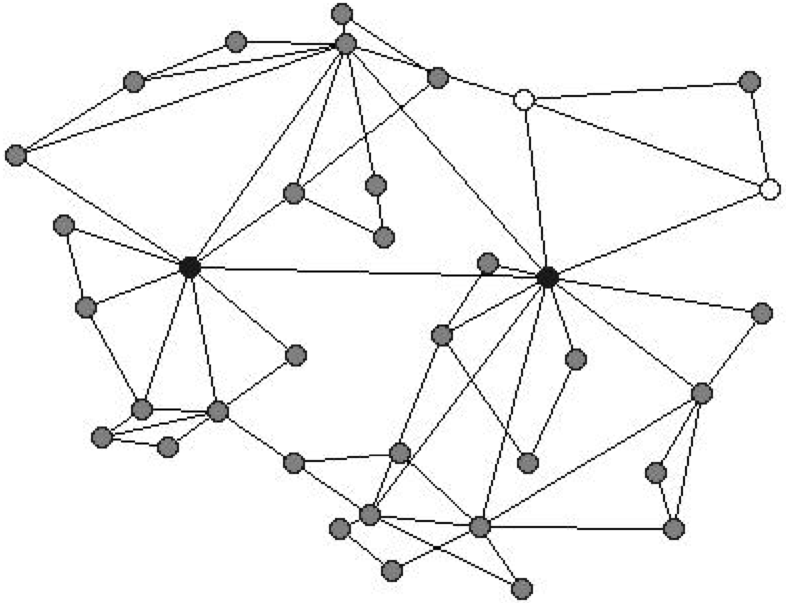
Internet Protocol

- The most widespread OSI Network Layer (3rd layer) protocol
- Datagram is the formal name for this layer data portion to be transmitted
  - commonly referred as packet, no mistake here
- unreliable
  - no datagram tracking, indexation, retransmissions whatsoever
- connectionless
  - no IP connection is being established
  - no pre-established relations prior transmission

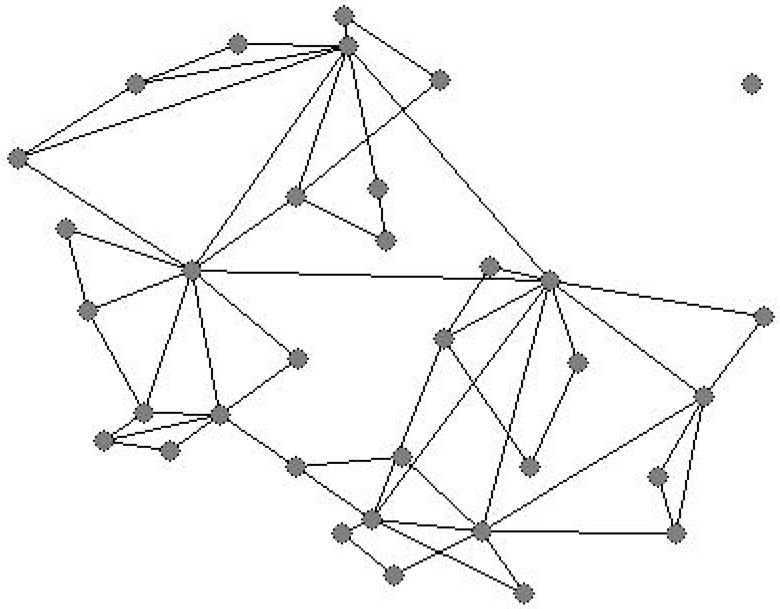
Internet Protocol – the DoD model

- The assumptions behind ARPANet project formulated by the Department of Defense – the DoD
    - distributed processing of information
    - loss of part of the network do not influence functioning of the remaining part, as long as no partitioning occurs

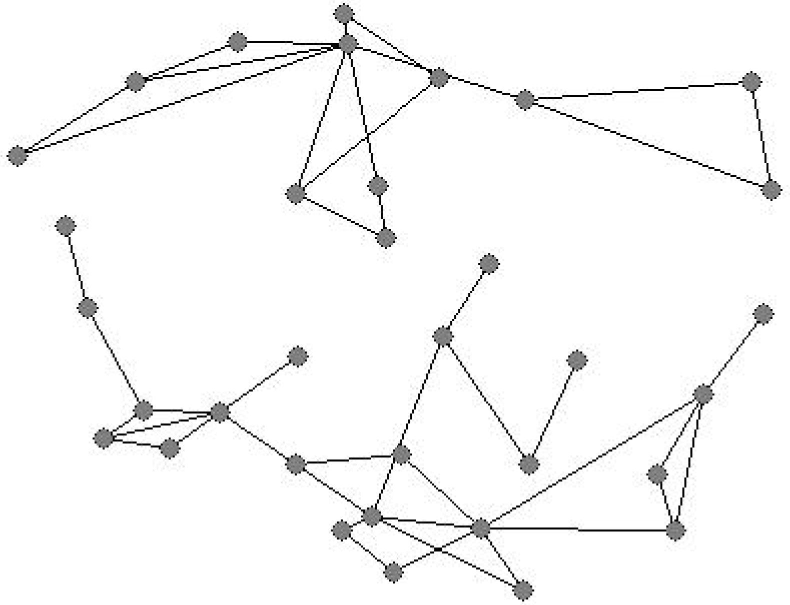
Exemplary network topology



Exemplary network topology



Exemplary network topology



IP datagram reordering

- Internet Protocol uses an opportunistic path to the destination
- due to the IP protocol design (in order to cope with node failures) no datagram order may be keeped unlike in LANs
    - datagram may come in wrong order
- no reliability is provided by the IP protocol
    - if reliability is a must during transmission, upper layer protocols should cope with it

The IP world – Inter-net transmission

- Inter-net – meaning between at least two networks
- The need to distinct networks and hosts inside a network

IPv4 address factors

- Network part
  - first bits of an address
  - example: 192.168.1.15
- Host part
  - last bits of an address
  - example 192.168.1.15
- In classless IPv4 addressing just the address itself stated strict distinction between the host and network address parts
- Along with CIDR – Classless Inter-Domain Routing the term mask was introduced allowing to more flexible subnetting

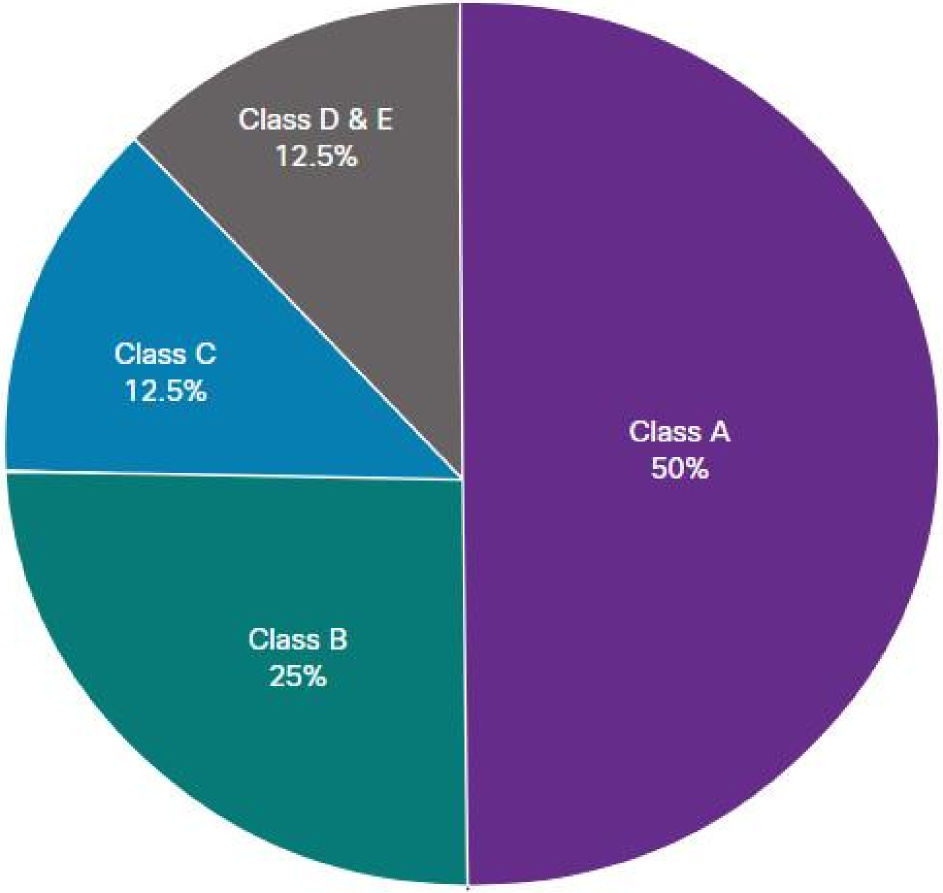
IPv4 address classes

- Class A: first byte in binary 0xxx xxxx
  - 0.0.0.0 – 127.255.255.255 address range
  - 2^7 such networks
  - equivalent network mask 255.0.0.0
- Class B: first byte in binary 10xx xxxx
  - 128.0.0.0 – 191.255.255.255 address range
  - 2^{14} such networks
  - equivalent network mask 255.255.0.0

IPv4 address classes

- Class C: first byte in binary 110x xxxx
  - 192.0.0.0 – 223.255.255.255
  - 2^{21} such networks
  - equivalent network mask 255.255.255.0
- Class D: first byte in binary 1110 xxxx
  - 224.0.0.0 – 239.255.255.255
  - number of networks undefined, although 2^{28} total addresses in address space
- Class E: reserved

IPv4 address class share



Class A

Total Networks: 128
Total Hosts/Net: 16,777,214

Class B

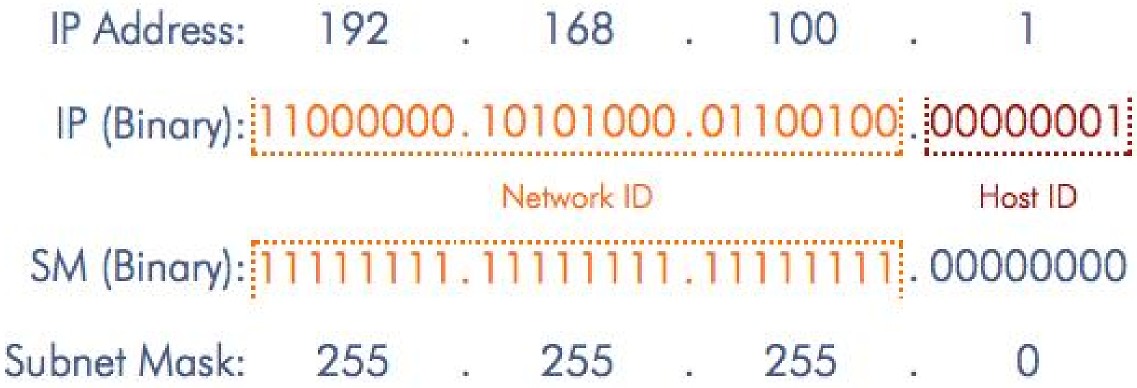
Total Networks: 16,384
Total Hosts/Net: 65,534

Class C

Total Networks: 2,097,152
Total Hosts/Net: 254

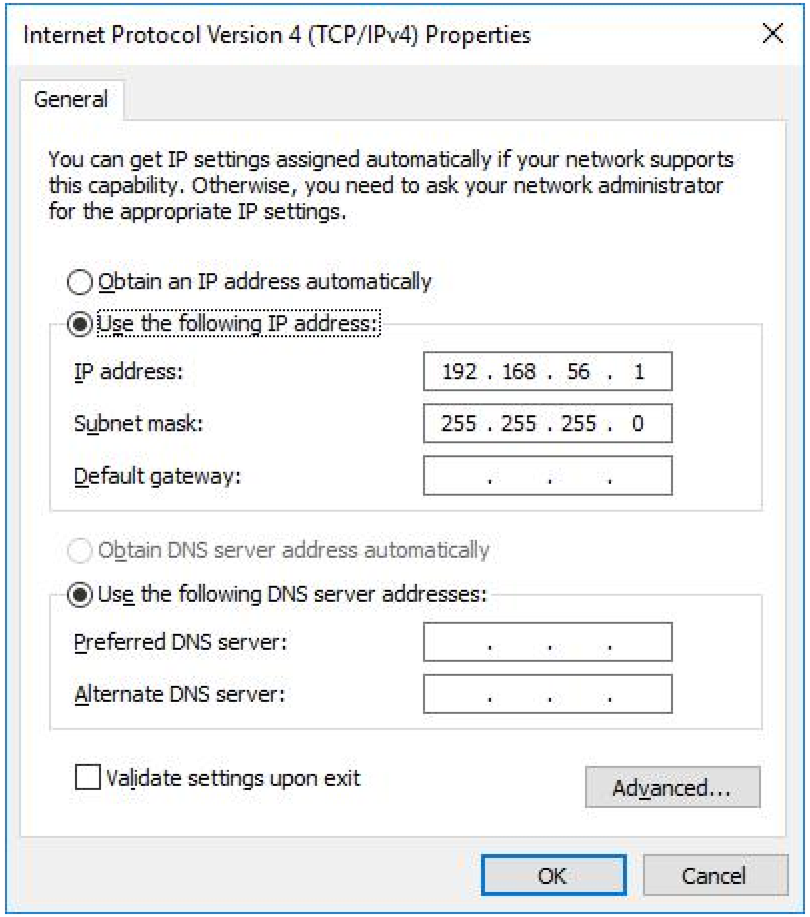
IPv4 CIDR address masks

- class A addreses relatively large
  - three bytes for host addressing – 2^{24} hosts = 16777216 hosts
  - no LAN would cope with so large broadcast domain
  - the mask – a solution for dividing large address spaces into smaller ones – subnets
- the name mask comes from binary AND operation on IP address
  - host part gets „masked out”



IP Address:	192	168	100	1
IP (Binary):	11000000	10101000	01100100	00000001
	Network ID			Host ID
SM (Binary):	11111111	11111111	11111111	00000000
Subnet Mask:	255	255	255	0

Example IP/mask setting



Internet Protocol Version 4 (TCP/IPv4) Properties

You can get IP settings assigned automatically if your network supports this capability. Otherwise, you need to ask your network administrator for the appropriate IP settings.

○ Obtain an IP address automatically

● Use the following IP address:

IP address: 192 . 168 . 56 . 1

Subnet mask: 255 . 255 . 255 . 0

Default gateway: . . .

○ Obtain DNS server address automatically

● Use the following DNS server addresses:

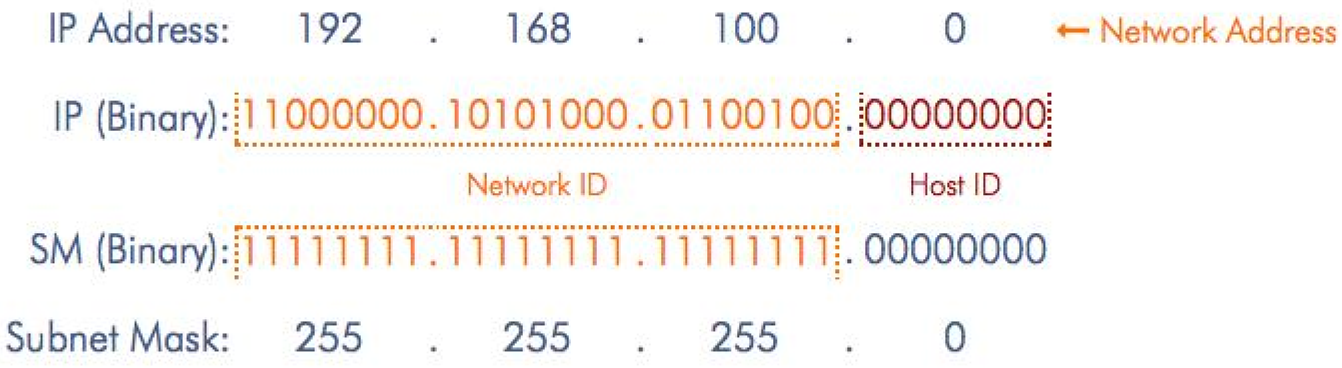
Preferred DNS server: . . .

Alternate DNS server: . . .

□ Validate settings upon exit

IPv4 addressing specifics

- IP address with host part equal 0 is called network address



IP Address: 192 . 168 . 100 . 0 ← Network Address

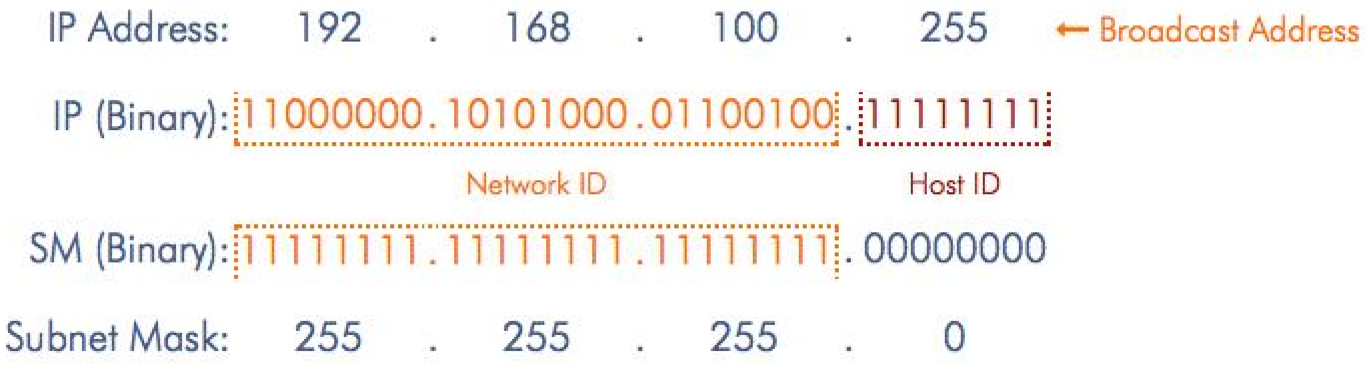
IP (Binary): 11000000.10101000.01100100.00000000

Network ID Host ID

SM (Binary): 11111111.11111111.11111111.00000000

Subnet Mask: 255 . 255 . 255 . 0

- IP address with host part all ones in binary is called broadcast address
  - broadcast IP addresses are mapped to broadcast MAC addresses upon sending



IP Address: 192 . 168 . 100 . 255 ← Broadcast Address

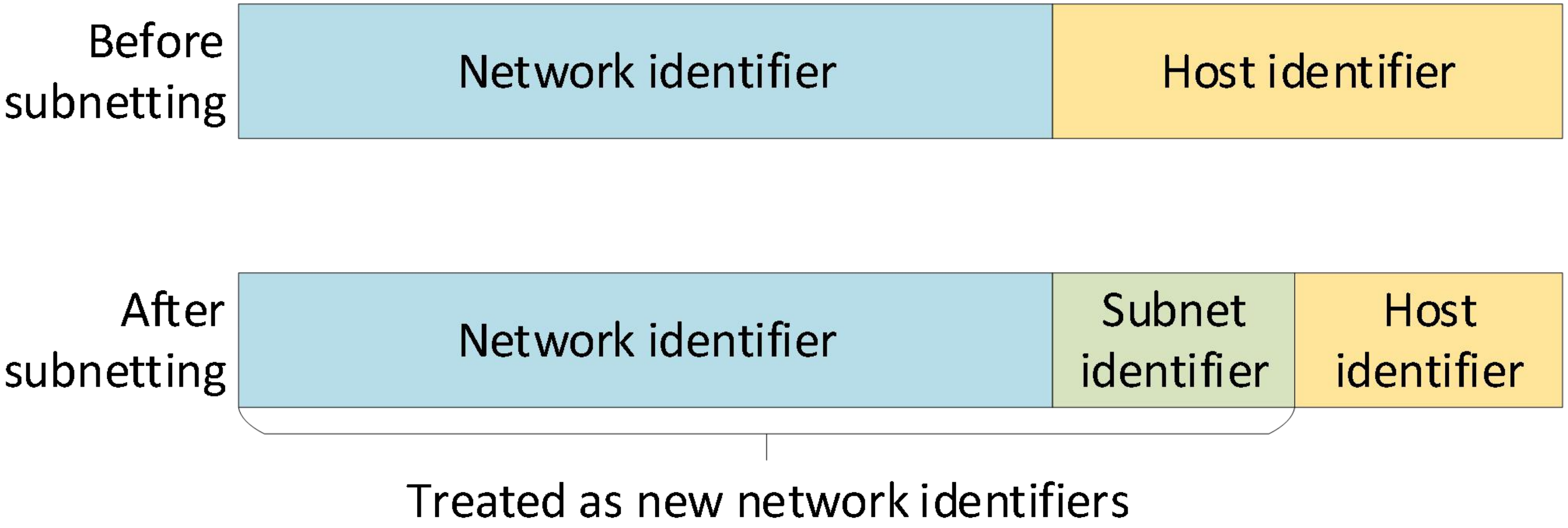
IP (Binary): 11000000.10101000.01100100.11111111

Network ID Host ID

SM (Binary): 11111111.11111111.11111111.00000000

Subnet Mask: 255 . 255 . 255 . 0

Subnetting



Before subnetting

Network identifier

Host identifier

After subnetting

Network identifier

Subnet identifier

Host identifier

Treated as new network identifiers

Subnetting

- mask has to have consecutive bits lit
- common masks:
  - 255.0.0.0 = 8 bits 1111 1111 0000 0000 0000 0000 0000 0000
  - 255.255.0.0 = 16 bits 1111 1111 1111 1111 0000 0000 0000 0000
  - 255.255.255.0 = 24 bits 1111 1111 1111 1111 1111 1111 0000 0000
  - 255.255.255.192 = 26 bits 1111 1111 1111 1111 1111 1111 1100 0000
  - 255.255.255.252 = 30 bits 1111 1111 1111 1111 1111 1111 1111 1100
- dotted-decimal notation may be shorted to just the number of bits lit notation:
  - 153.19.55.100/26 = 153.19.55.100 mask 255.255.255.192

IPv4 – globally unique addresses

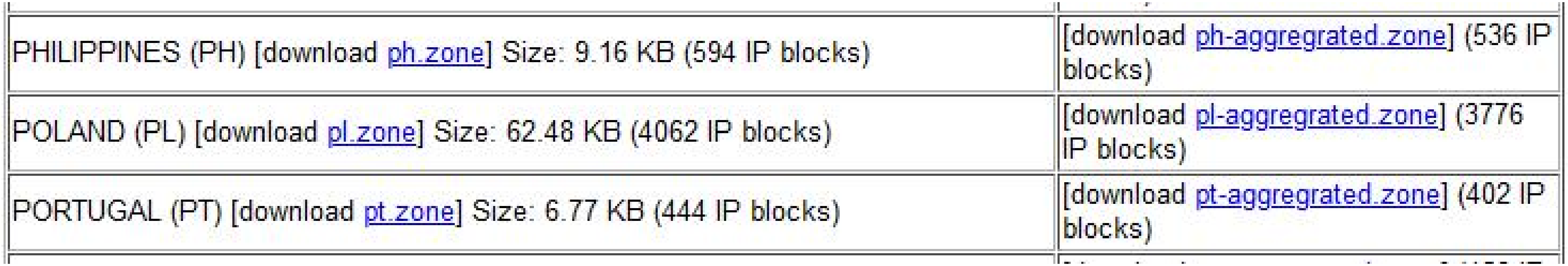
- similarily to MAC addresses, most of IPv4 addresses ale so-called global addresses
- in case of IP address duplication routing issues arise
    - no solution is provided to resolve such issues
    - in closed (!) environments the addressing may be totally arbitrary
- non-unique addresses provided for closed environments
    - may be re-used around the world, co-caled private addressing
    - one class A private network – 10.0.0.0/8
    - 16 class B private networks – 172.16.0.0/16 – 172.31.0.0/16
    - the beloved 192.168.0.0/24 networks – 256 of them (from 0 to 255)

IPv4 – address properties

- one can buy an IP address only from ISPs operating in this specific region
- each globally unique network is attributed geographically
- rather general geolocalization possible using only the network address
- more precise IP address localization requires additional attribution
  IP -> e.g. city
- many ISPs provide us with so-called dynamic IP addresses, which may change in time
  - this is mainly in order to forbid placing servers in non-commercial servicing regime
  - not only it is related to cheaper prices, but mostly to SLAs – Service Level Agreements

IP address space in Poland

- Prefix lists for specific countries can be downloaded from:
  - https://www.ip2location.com/free/visitor-blocker
  - https://www.ipdeny.com/ipblocks/
- For Poland we have…



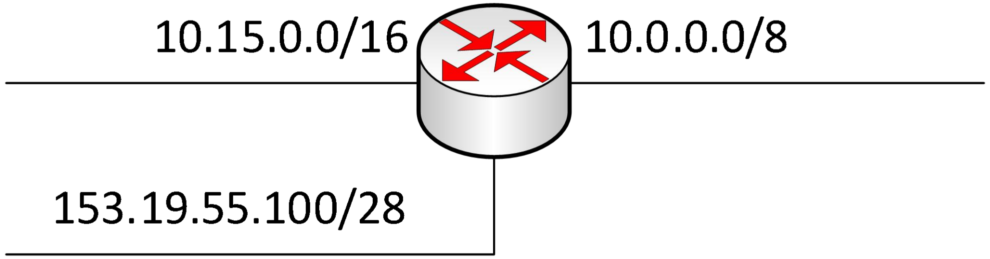
PHILIPPINES (PH) [download ph.zone] Size: 9.16 KB (594 IP blocks)	[download ph-aggregrated.zone] (536 IP blocks)
POLAND (PL) [download pl.zone] Size: 62.48 KB (4062 IP blocks)	[download pl-aggregrated.zone] (3776 IP blocks)
PORTUGAL (PT) [download pt.zone] Size: 6.77 KB (444 IP blocks)	[download pt-aggregrated.zone] (402 IP blocks)

Polish IPv4 adress space sample

- 194.126.207.0/24
- 194.126.210.0/24
- 194.126.216.0/24
- 194.126.221.0/24
- 194.126.222.0/24
- 194.126.229.0/24
- 194.126.232.0/24
- 194.126.238.0/24
- 194.126.254.0/24
- 194.127.136.0/24
- 194.127.137.0/24
- 194.140.233.0/24
- 194.140.241.0/24
- 194.140.250.0/24
- 194.140.255.0/24

Routing mode of action

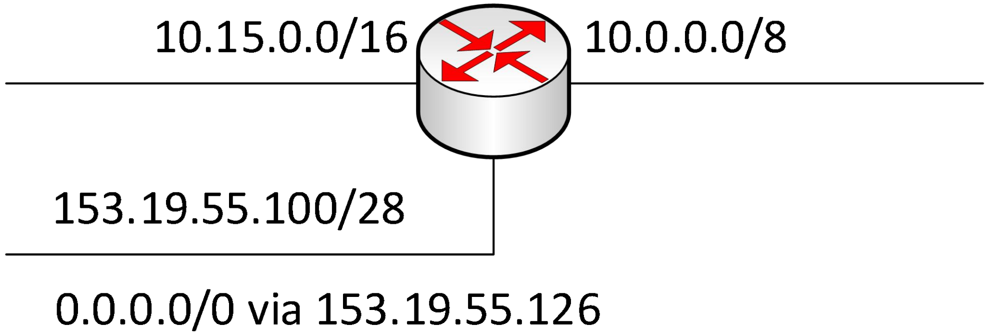
- router interfaces are configured with IP addresses and masks
  - example addressing: 153.19.55.100/28
- router masks every destination IP address to determine the network
- output interface is the one which matches the longest part of network



- last resort route: 0.0.0.0/0
  - mask length equal 0 – no network will ever match
  - so-called default gateway

The default gateway

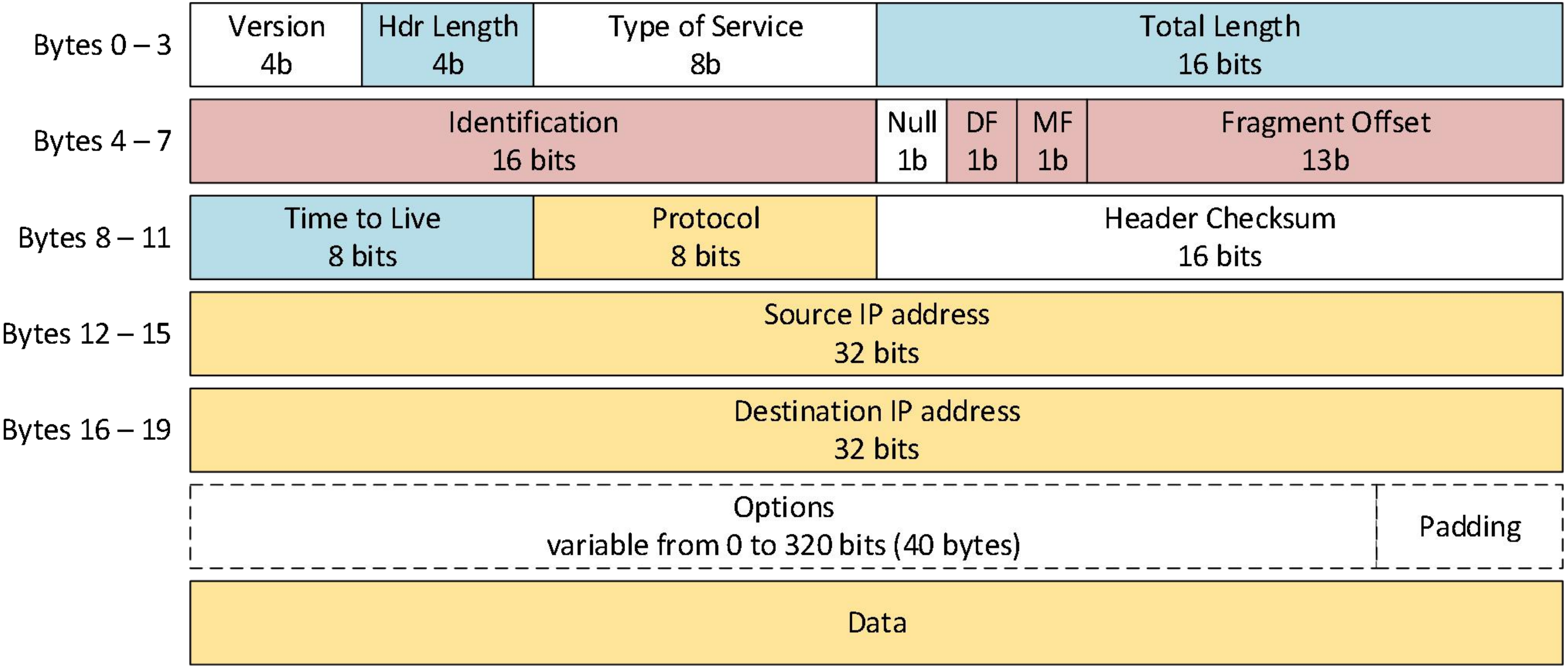
- route pointing towards 0.0.0.0/0
- routing table contain specific, single IP address which will take care of the routing
- the Gateway must be reachable via this interface – according to the subnet
- the simpliest and most common example of *static-routing*
- all traffic which do not match other routing table entries gets routed via the default gateway



Special purpose addresses

- 127.0.0.1 – loopback address – every host itself
- 0.0.0.0/0 – the default route – packets are being sent there if no other network on interfaces matches
- 169.254.0.0/16 – stateless IPv4 addressing – if the interface has neither static IP nor dynamic obtained from DHCP server
  - initially registered by Microsoft
  - recently Linux machines started making use of it
  - host part is being randomized minimizing probability of IP address conflicts
  - if no router exists in a network, all the hosts configured the same way can communicate

IPv4 header indepth



Bytes 0 – 3

Version 4b | Hdr Length 4b | Type of Service 8b | Total Length 16 bits

Bytes 4 – 7

Identification 16 bits | Null 1b | DF 1b | MF 1b | Fragment Offset 13b

Bytes 8 – 11

Time to Live 8 bits | Protocol 8 bits | Header Checksum 16 bits

Bytes 12 – 15

Source IP address 32 bits

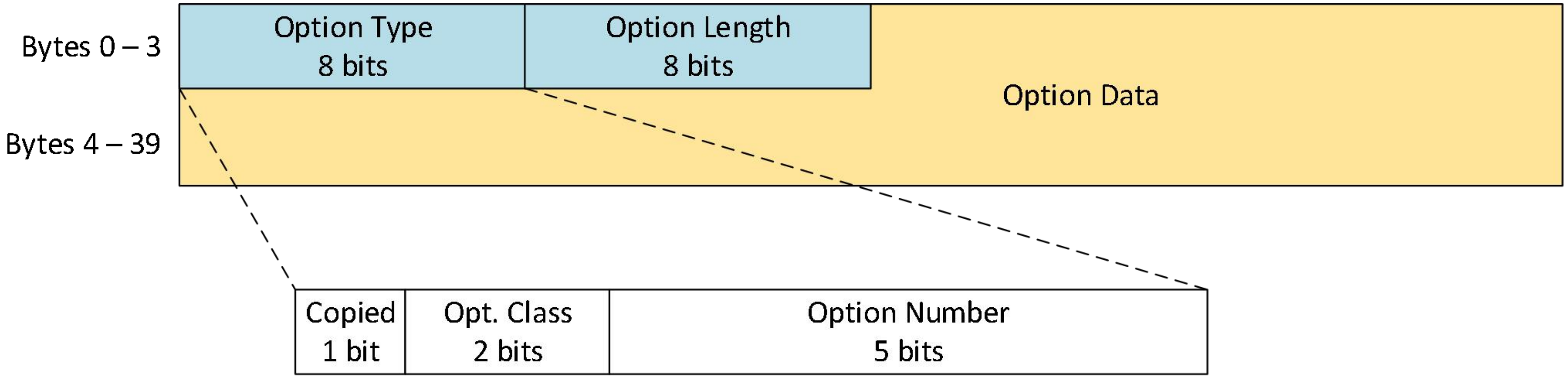
Bytes 16 – 19

Destination IP address 32 bits

Options variable from 0 to 320 bits (40 bytes) | Padding

Data

IPv4 Header Options



- Copied – should the option be copied to all the fragmengs of a datagram
- Option Class – 0: Control, 2: Debugging and Measurement
- Option Number – what option is carried in the header

IPv4 Header Options

Option Class	Option Number	Length (bytes)	Option Name	Description
0	0	1	End Of Options List	An option containing just a single zero byte, used to mark the end of a list of options.
0	1	1	No Operation	A “dummy option” used as “internal padding” to align certain options on a 32-bit boundary when required.
0	2	11	Security	An option provided for the military to indicate the security classification of IP datagrams.
0	3	Variable	Loose Source Route	One of two options for source routing of IP datagrams. See below for an explanation.
0	7	Variable	Record Route	This option allows the route used by a datagram to be recorded within the header for the datagram itself. If a source device sends a datagram with this option in it, each router that “handles” the datagram adds its IP address to this option. The recipient can then extract the list of IP addresses to see the route taken by the datagram. Note that the length of this option is set by the originating device. It cannot be enlarged as the datagram is routed, and if it “fills up” before it arrives at its destination, only a partial route will be recorded.
0	9	Variable	Strict Source Route	One of two options for source routing of IP datagrams. See below for an explanation.
2	4	Variable	Timestamp	This option is similar to the Record Route option. However, instead of each device that handles the datagram inserting its IP address into the option, it puts in a timestamp, so the recipient can see how long it took for the datagram to travel between routers. As with the Record Route option, the length of this option is set by the originating device and cannot be enlarged by intermediate devices.
2	18	12	Traceroute	Used in the enhanced implementation of the traceroute utility, as described in RFC 1393.

Combining L2 and L3 operations

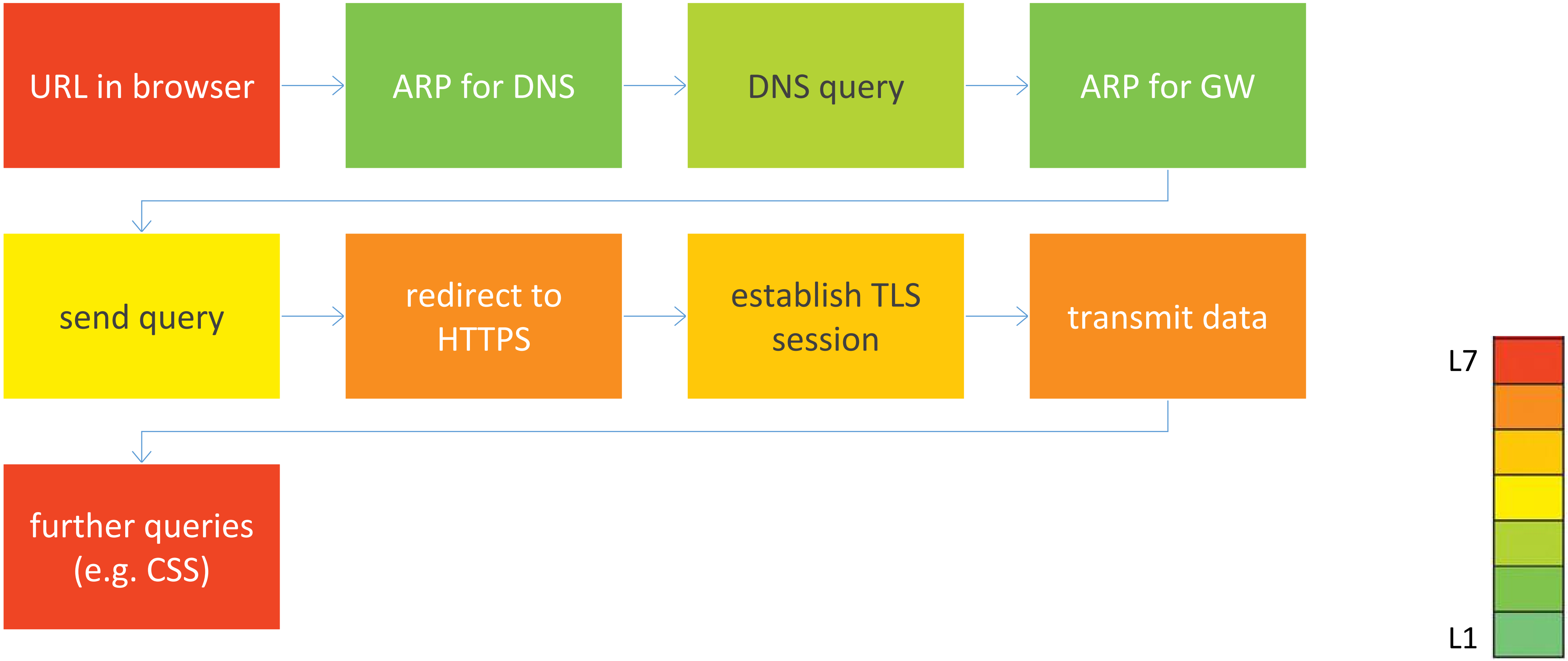
The ARP protocol

- ARP – Address Resolution Protocol
  - provides the table of IP -> MAC mapping
  - uses broadcast packets (sent as broadcast frames in shared media)
- Generally three ARP packets are in use:
  - the query – „who-has” messages
  - the answer – „is-at” messages
  - gratuitous ARP messages – „I am here”
- Upon query an entry is added to ARP table
  - lifetime of an entry equals typically 2 minutes
  - after expiry the query must be repeated

ARP table example: IPs -> MACs

```
Internet Address      Physical Address      Type
 192.168.1.1          6c-b3-e5-26-e5-89     dynamic
 192.168.1.11         ac-f1-df-de-83-70     dynamic
 192.168.1.18         5c-08-b6-5a-f3-1b     dynamic
 192.168.1.25         20-71-42-34-89-63     dynamic
 192.168.1.28         1c-fb-ce-09-1b-ff     dynamic
 192.168.1.255        ff-ff-ff-ff-ff-ff     static
 224.0.0.22           01-00-5e-00-00-16     static
 224.0.0.251          01-00-5e-00-00-fb     static
 224.0.0.252          01-00-5e-00-00-fc     static
 239.255.255.250      01-00-5e-7f-ff-fa     static
 255.255.255.255      ff-ff-ff-ff-ff-ff     static
```

Putting it all together – typical Web session

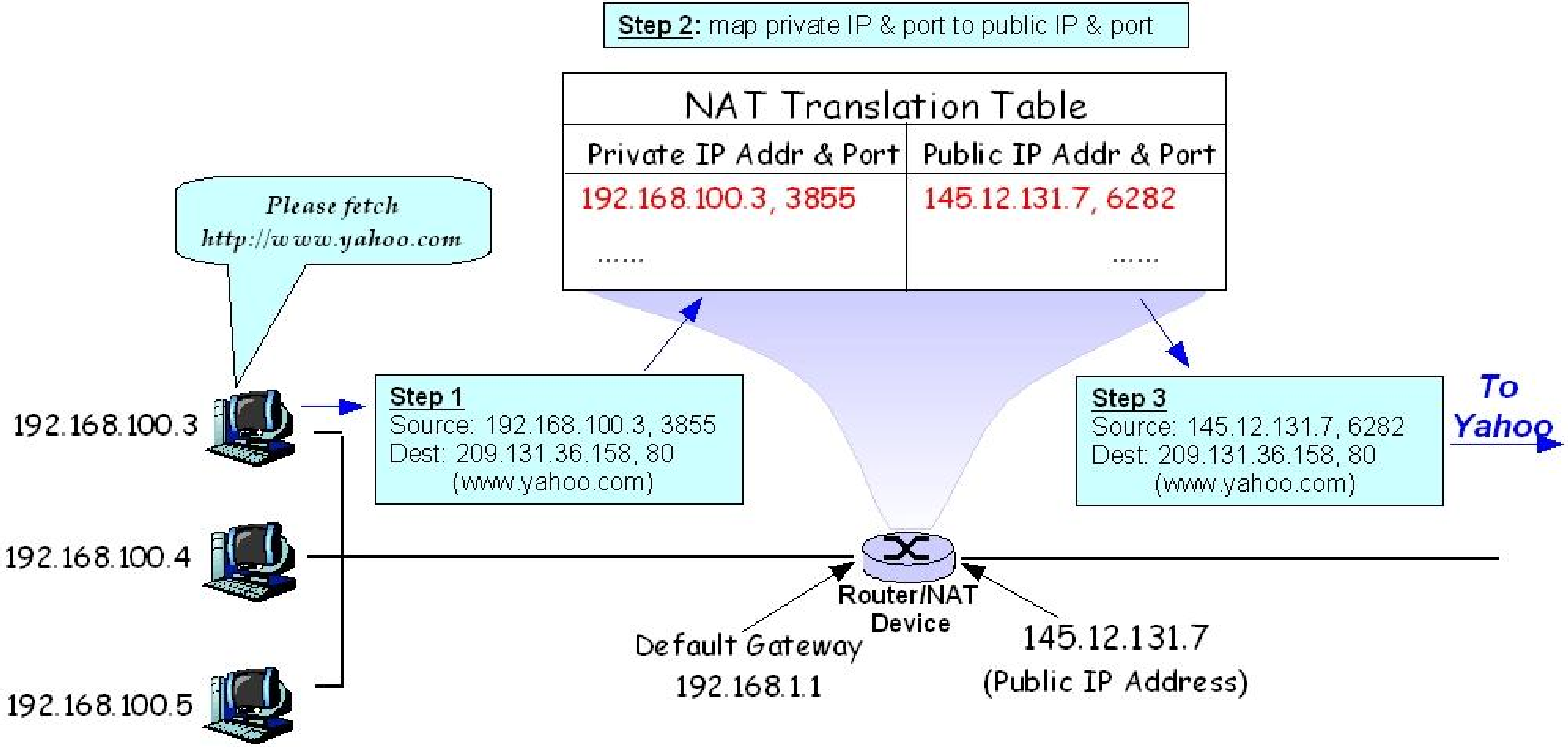


URL in browser → ARP for DNS → DNS query → ARP for GW → send query → redirect to HTTPS → establish TLS session → transmit data → further queries (e.g. CSS)

NAT – Network Address Translation

- One of mechanisms for IPv4 address exhaustion
- Assumes that multiple private (aka non-routable) addresses may communicate under one (or multiple) public addresses
- Also known as SNAT – Source NAT
  - source IP address is changed upon sending by router
  - router changes its source address from privte to its own public
  - to avoid ambiguity (non-deterministic matching) source ports for TCP and UDP protocols are also being changed

NAT



Step 2: map private IP & port to public IP & port

NAT Translation Table

Private IP Addr & Port	Public IP Addr & Port
192.168.100.3, 3855	145.12.131.7, 6282
……	……

Please fetch http://www.yahoo.com

192.168.100.3

192.168.100.4

192.168.100.5

Step 1
Source: 192.168.100.3, 3855
Dest: 209.131.36.158, 80
(www.yahoo.com)

Step 3
Source: 145.12.131.7, 6282
Dest: 209.131.36.158, 80
(www.yahoo.com)

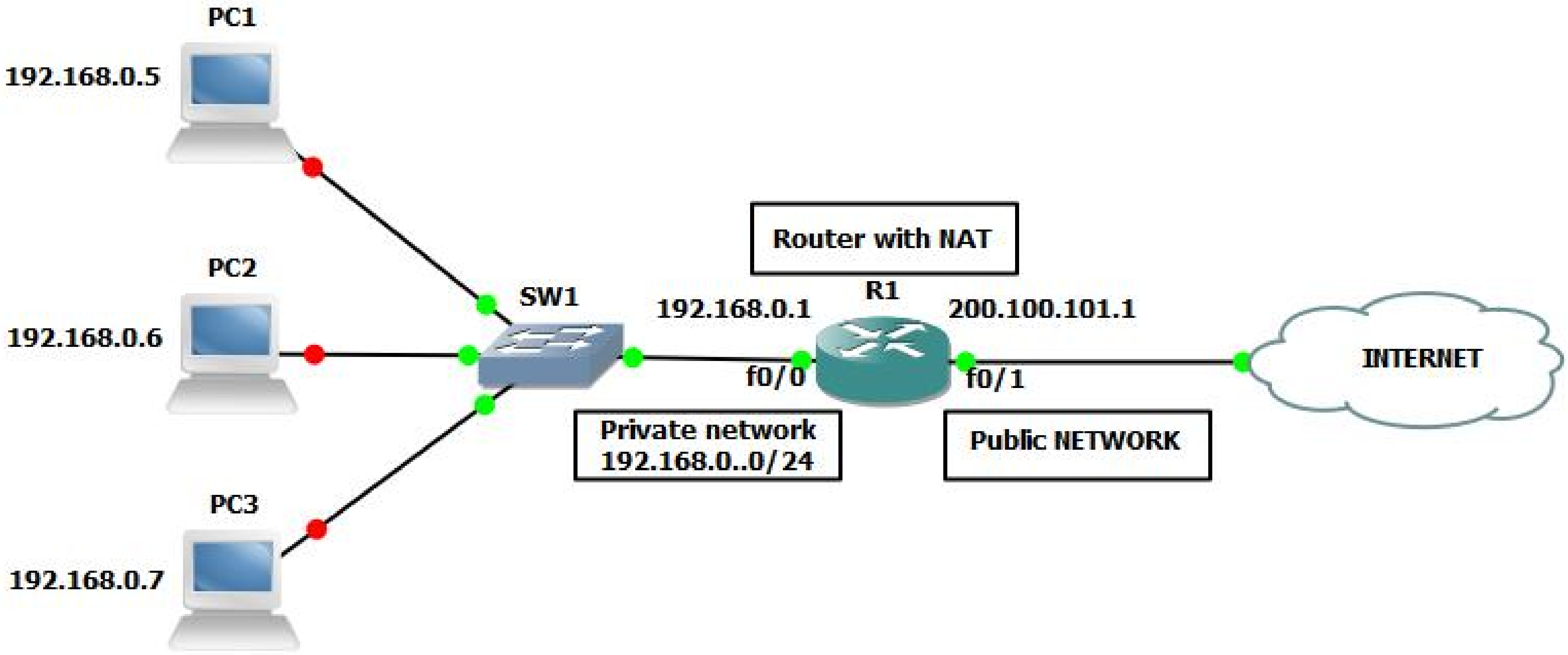
To Yahoo

Router/NAT
Device

Default Gateway
192.168.1.1

145.12.131.7
(Public IP Address)

NAT


PC1
192.168.0.5
PC2
192.168.0.6
PC3
192.168.0.7
SW1
192.168.0.1
Router with NAT
R1
200.100.101.1
f0/0
f0/1
Private network
192.168.0..0/24
Public NETWORK
INTERNET

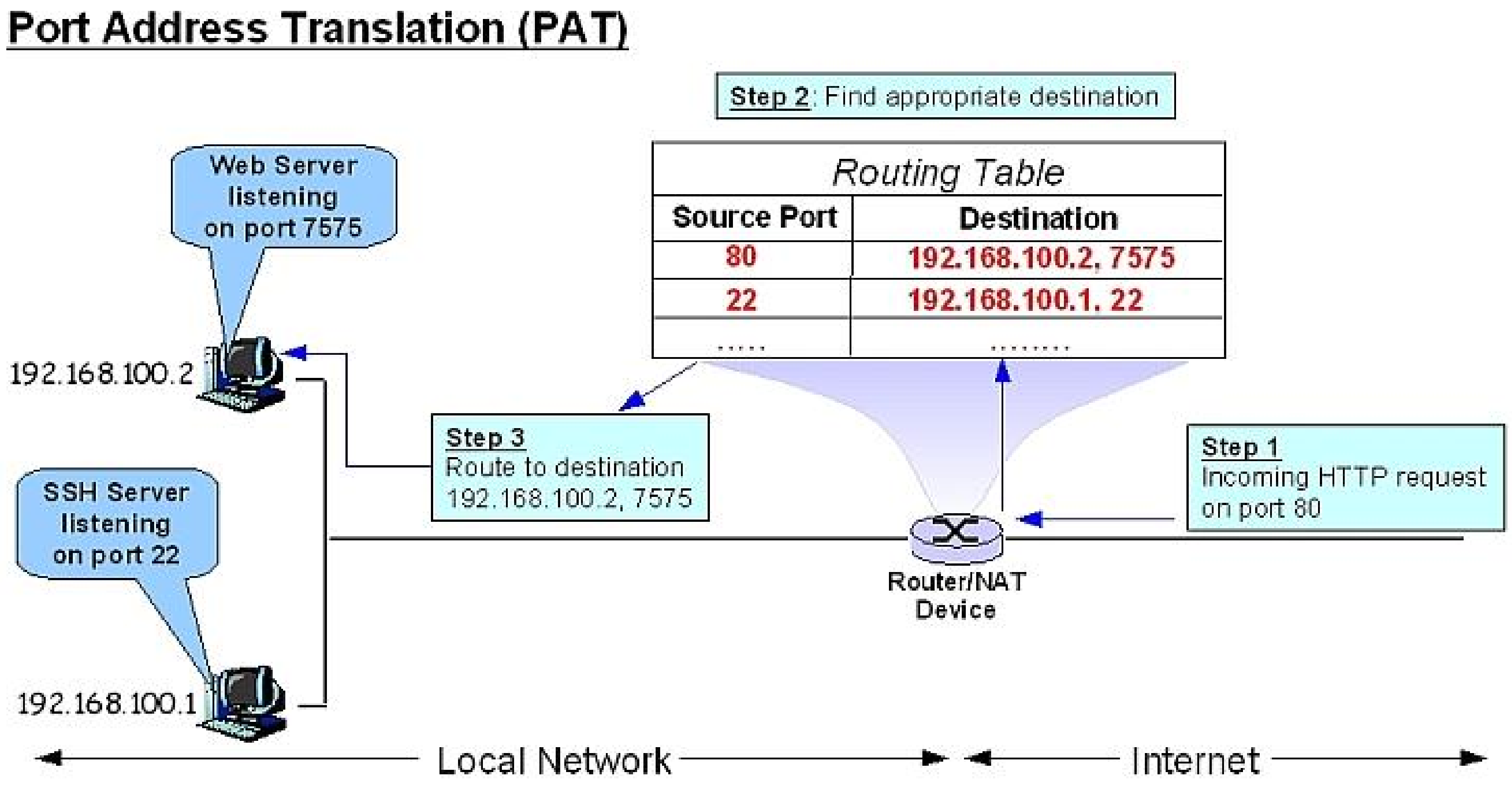
SNAT options

- static NAT
  - one-to-one: one private address mapped to one public
- dynamic NAT
  - many-to-one: many private IPs mapped to one public IP
  - many-to-many:

Destination NAT

- the opposite of NAT
- allows to redirect network traffic (typically TCP/UDP) which is addressed to a public IP to some host behind NAT, which uses private addressing

DNAT (PAT)



Port Address Translation (PAT)

Step 2: Find appropriate destination

Routing Table

Source Port	Destination
80	192.168.100.2, 7575
22	192.168.100.1. 22
.....	....

Step 3
Route to destination
192.168.100.2, 7575

Step 1
Incoming HTTP request
on port 80

Web Server listening on port 7575
192.168.100.2
SSH Server listening on port 22
192.168.100.1
Router/NAT Device
Local Network
Internet

Routing in the Internet

Routing – general terms

- Routing is responsible for proper path selection for forwarded packets
- In order to provide bi-directional communication channel, routing in both directions is required
- The easiest to observe is static routing
  - it may be observed on almost every station connected to the Internet
  - the default gateway – the „last resort” routing 0.0.0.0/0
  - if no other route matches to any of the interface, use the default one
- Routes may be added in dynamic way due to routing protocols
- Network- and broadcast-addressed packets are discarded
  - one can not „ping the whole world” although broadcast ping is possible in broadcast domain, bounded by router

General routing classification

- Static
  - networks configured on interfaces, static network entries, default gateway
- Dynamic
  - Intra-domain
    - distance-vector: RIP
    - link-state: OSPF, EIGRP
  - Inter-domain
    - path-vector: BGP

Dynamic routing protocols properties

- Adding routes to routing table
  - network, netmask -> gateway, interface, metrics
  - route metrics can be perceived as a cost to reach the network – the higher the further network is
  - metrics is taken under consideration last – only if two interfaces lead to exactly the same destination network
- Removing routes from the table
  - if a route unreachability was detected (e.g. interface lost signal)
  - if a timer exceeded and no update was received
- Automatic removal and addition is a remedy to link/node failures
  - if a new route has been found it can take place of the missing, removed previously entity

Autonomous Systems

- AS – Autonomous System, mostly related to Inter-domain BGP routing
- A set of (public, routable) networks under supervision of a single entity
- The entity fully controlls how routing is done inside the AS
- Initially defined for IPv4 (RFC 1771), changed in RFC 1930
- Initially using two-byte identifier numbers
  - 64512 to 65535 reserved
  - RFC 4893 changed the size to four bytes
- Currently around 40000 AS numbers in use
- Allows also for the distinction of inter- and intra-domain routing protocols

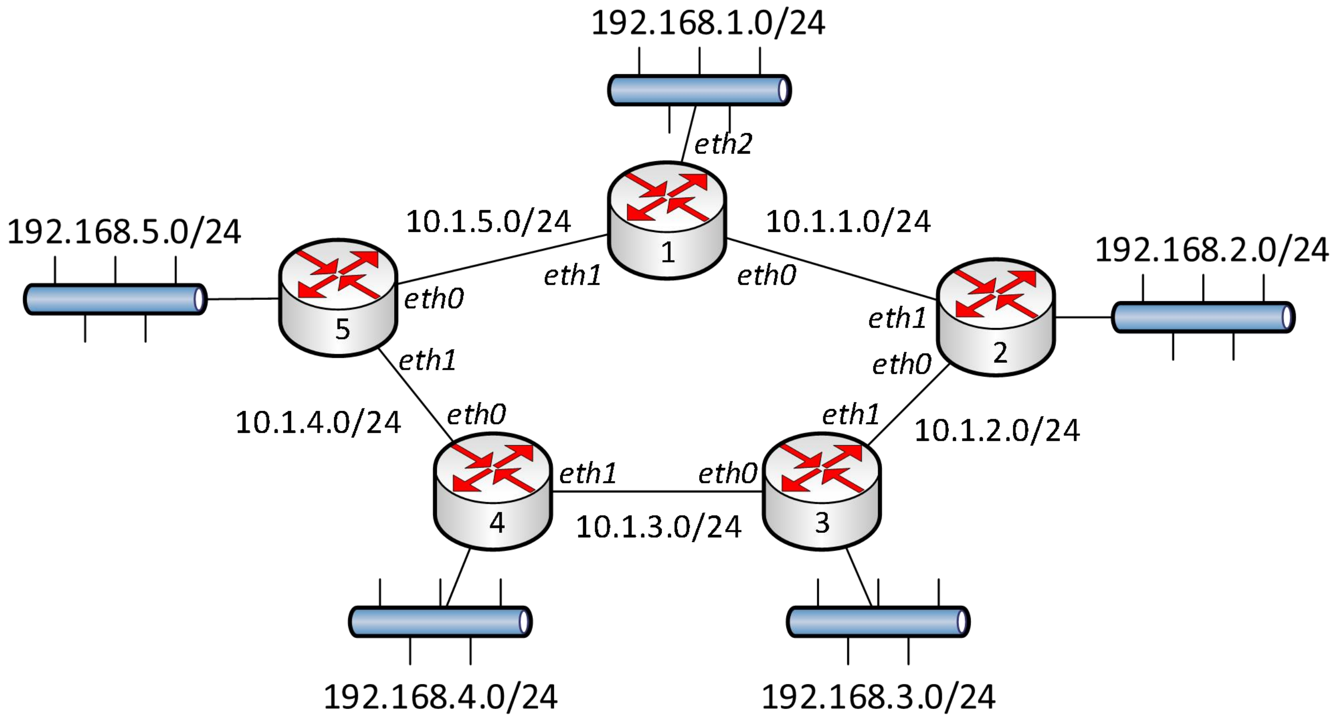
Distance vector – Link state comparison

Basis for comparison	Distance vector routing	Link state routing
Algorithm	Bellman-Ford	Dijsktra
Network view	Topology information from the neighbour point of view	Complete information on the network topology
Best path calculation	Based on the least number of hops	Based on the cost
Updates	Full routing table	Link state updates
Updates frequency	Periodic updates	Triggered updates
CPU and memory	Low utilisation	Intensive
Simplicity	Relatively simple	Relatively complicated
Convergence time	Moderate	Short
Hierarchical structure	No	Yes
Intermediate Nodes	No	Yes

RIP – Routing Information Protocol

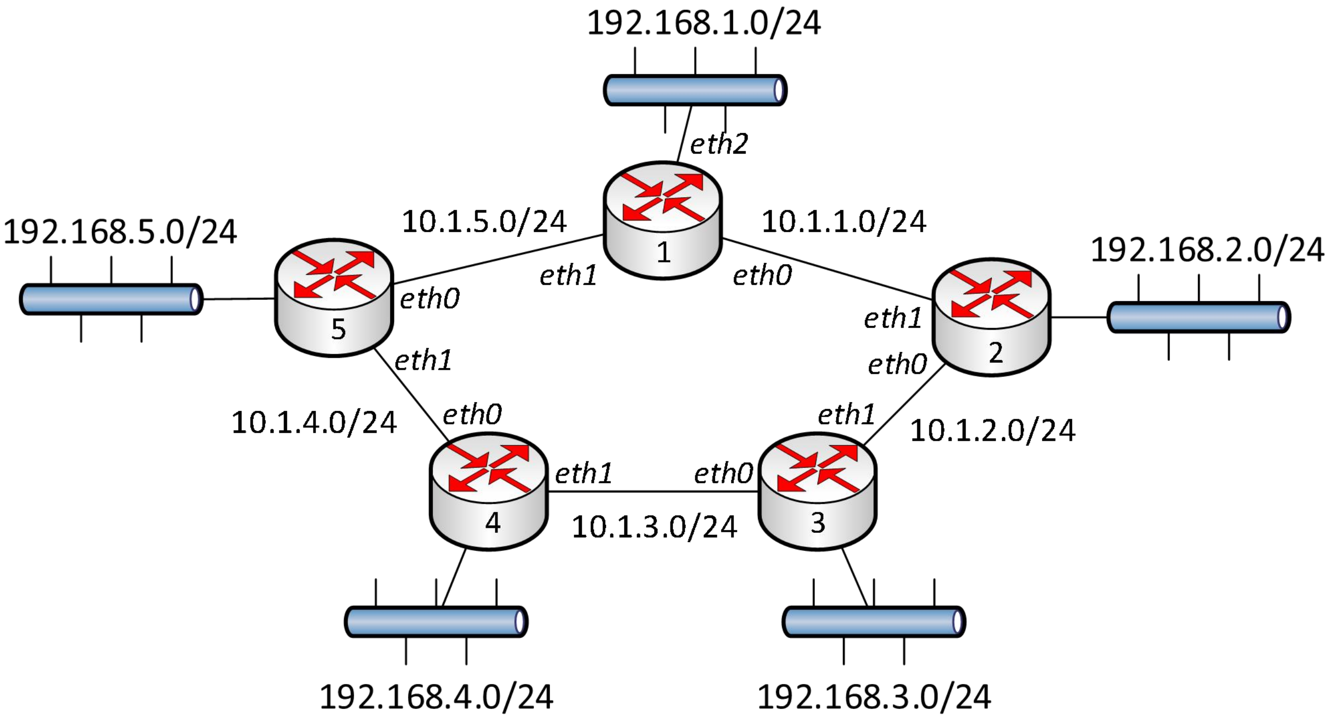
- One of the simplest routing protocols
- Routes are transmitted periodically
  - full routing table is being sent
- Initially classful (no masks, no subnets)
  - RIPv2 introduced classless routing
  - RIPng introduced IPv6 support
- Number of hops is the metrics
  - host reachable in 15 or more hops considered unreachable
  - does not take under consideration link speeds (hence preference of links)

RIP protocol – an example



- Initial router 1 routing table:
  - 10.1.1.0/24 directly eth0
  - 10.1.5.0/24 directly eth1
  - 192.168.1.0/24 directly eth2
- After the reception from 2:
  - 10.1.1.0/24 directly eth0
  - 10.1.5.0/24 directly eth1
  - 192.168.1.0/24 directly eth2
  - 10.1.2.0/24 -> 10.1.1.2 eth0
  - 192.168.2.0/24 -> 10.1.1.2 eth0
- and from 5 (new routes only):
  - 10.1.4.0/24 -> 10.1.5.5 eth1
  - 192.168.5.0/24 -> 10.1.5.5 eth1

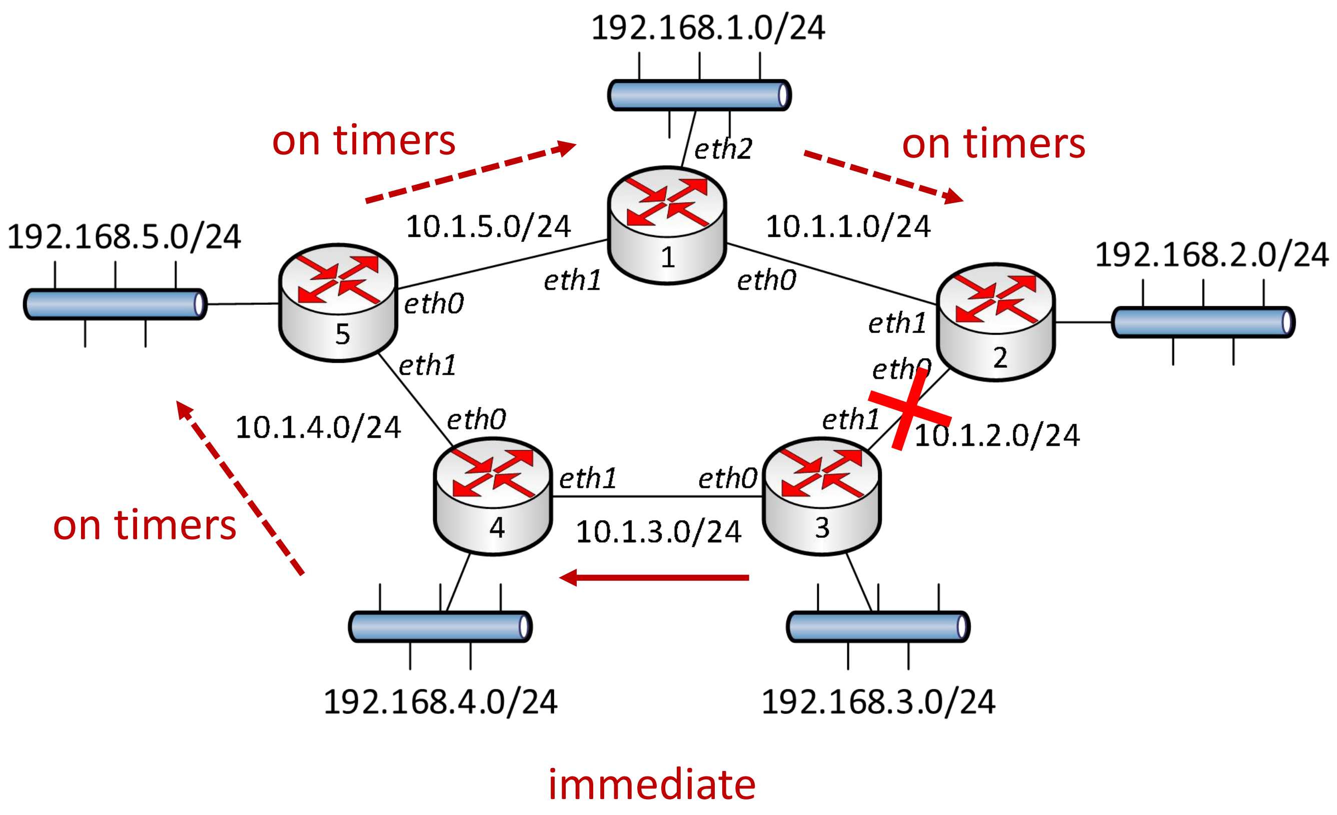
RIP protocol – an example



- Finally on router 1:
  - 10.1.1.0/24 directly eth0
  - 10.1.2.0/24 -> 10.1.1.2 eth0, 1 hop
  - *10.1.3.0/24 -> 10.1.1.2 eth0, 2 hops*

    or -> 10.1.5.5 eth1, 2 hops
  - 10.1.4.0/24 -> 10.1.5.5 eth1, 1 hop
  - 10.1.5.0/24 directly eth1
  - 192.168.1.0/24 directly eth2
  - 192.168.2.0/24 -> 10.1.1.2 eth0, 1 hop
  - 192.168.3.0/24 -> 10.1.1.2 eth0, 2 hops
  - 192.168.4.0/24 -> 10.1.5.5 eth1, 2 hop
  - 192.168.5.0/24 -> 10.1.5.5 eth1, 1 hop

RIP protocol – after a link failure



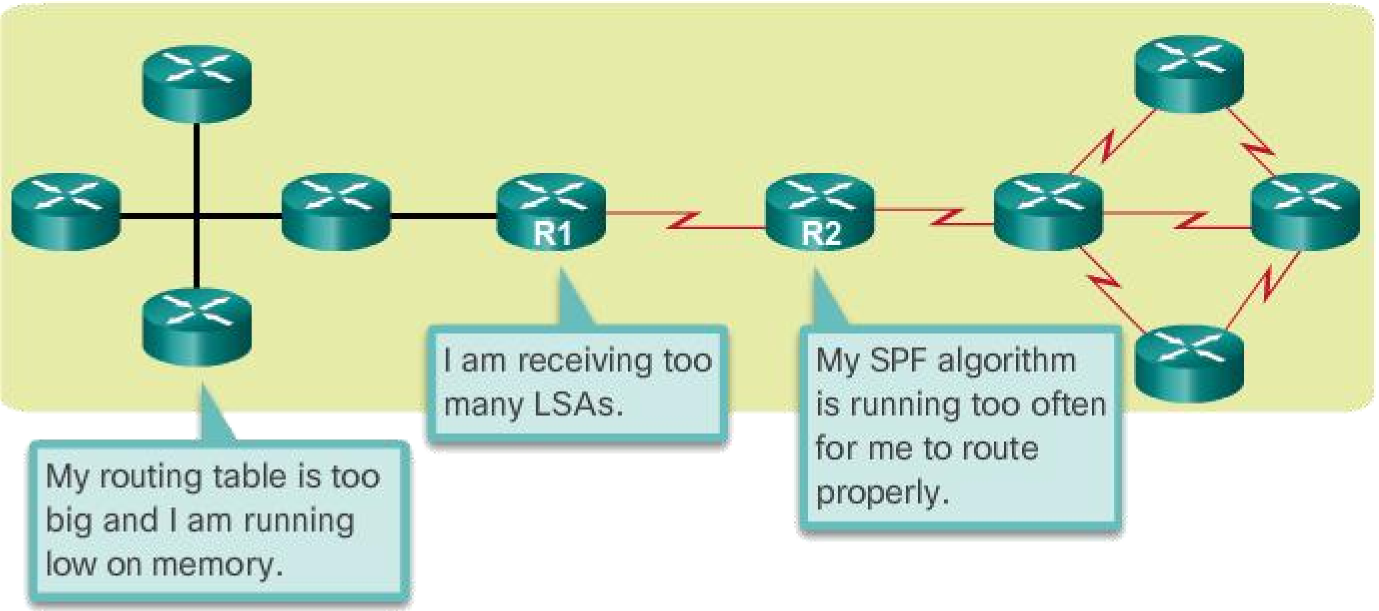
- Finally on router 1:
  - 10.1.1.0/24 directly eth0
  - 10.1.2.0/24 -> 10.1.1.2 eth0, 1 hop
  - 10.1.3.0/24 -> 10.1.5.5 eth1, 2 hops

  - 10.1.4.0/24 -> 10.1.5.5 eth1, 1 hop
  - 10.1.5.0/24 directly eth1
  - 192.168.1.0/24 directly eth2
  - 192.168.2.0/24 -> 10.1.1.2 eth0, 1 hop
  - 192.168.3.0/24 -> 10.1.1.2 eth1, 2 hops
  - 192.168.4.0/24 -> 10.1.5.5 eth1, 2 hop
  - 192.168.5.0/24 -> 10.1.5.5 eth1, 1 hop

OSPF – Open Shortest Path First protocol

- Intra-domain link-state routing protocol
- „Open" as in „Open Source"
- Link-state
  - only topology changes are being sent
  - every router has to have full topology overview
  - the full topology allows to autonomously react upon link state update (e.g. failure)
- Distinctive routing messages used among others to:
  - control adjacency of nodes (Hello packet)
  - build the topology itself (DBD – Database Descriptor)
  - send link updates (LSU – Link-State Updates)
- Due to so-called area definitions allows for hierarchical topology usage
  - privileged role of so-called backbone area (Area 0 – devoted to default routing)

Single- vs multi-area OSPF configurations

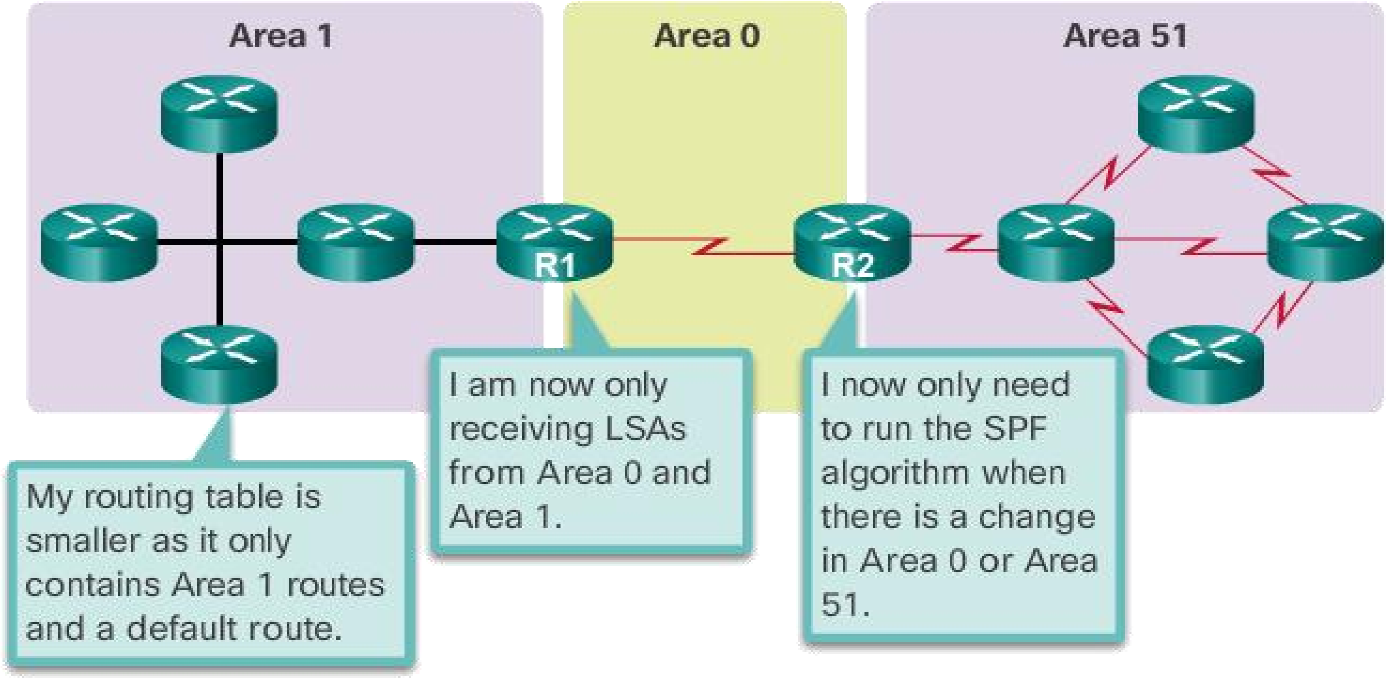


My routing table is too big and I am running low on memory.

I am receiving too many LSAs.

My SPF algorithm is running too often for me to route properly.

Single- vs multi-area OSPF configurations

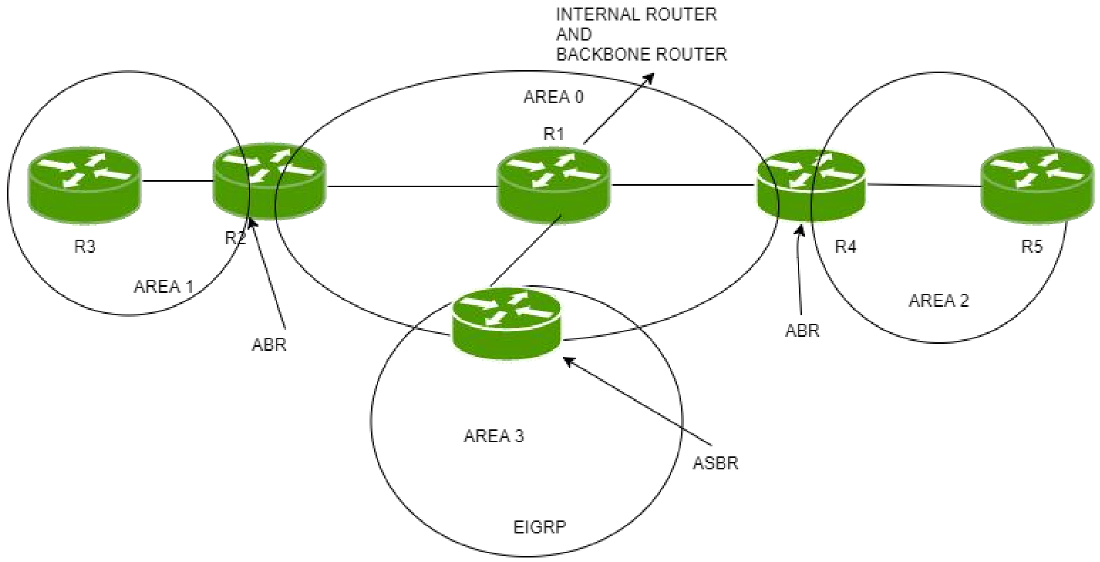


My routing table is smaller as it only contains Area 1 routes and a default route.

I am now only receiving LSAs from Area 0 and Area 1.

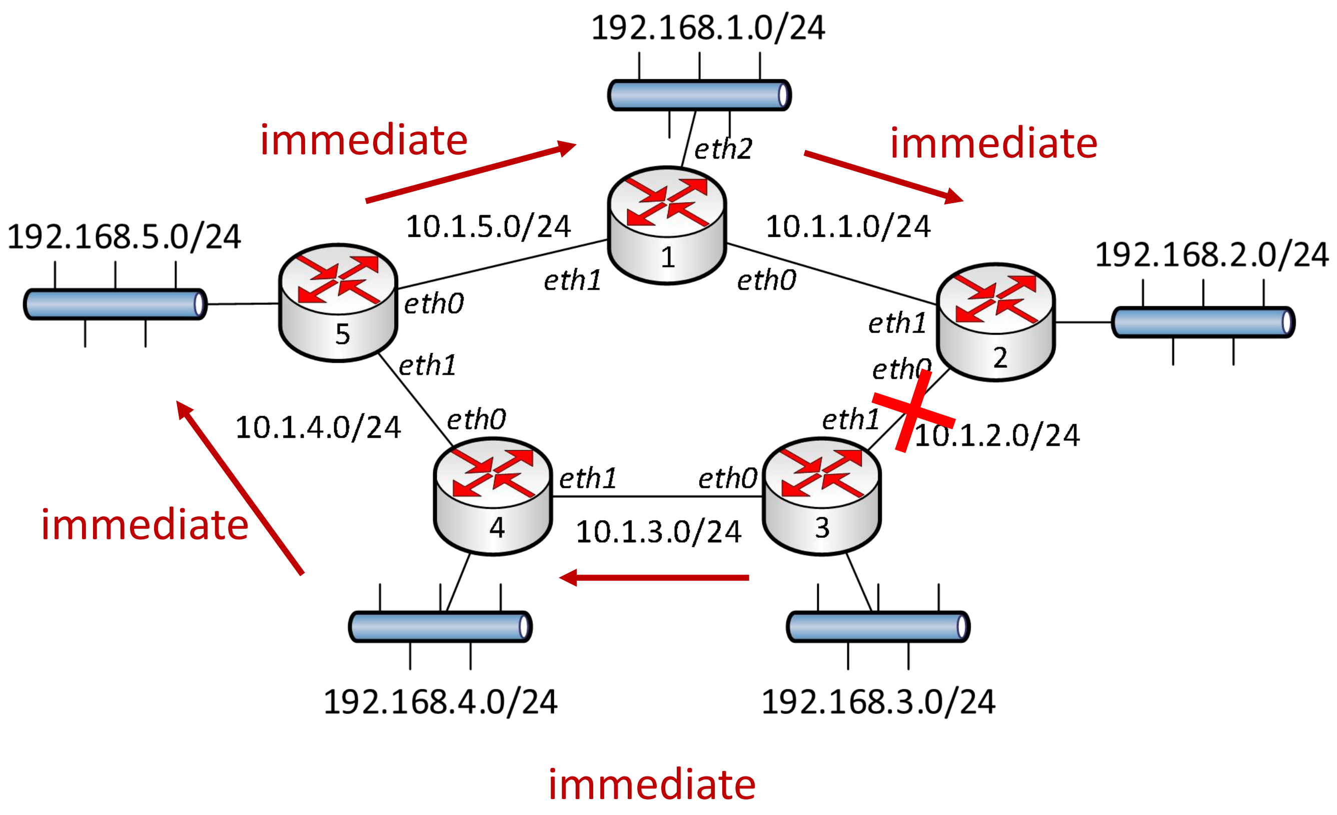
I now only need to run the SPF algorithm when there is a change in Area 0 or Area 51.

OSPF router roles



- Backbone router
  - the one belonging to Area 0
- Internal router
  - all of its interfaces belong to the same area
- Area Boundary Router – ABR
  - the one connecting Area 0 and some other
- Autonomous System Border Router – ASBR
  - the one neighboring with some other AS

OSPF protocol – after a link failure



- Finally on router 1:
  - 10.1.1.0/24 directly eth0
  - 10.1.2.0/24 -> 10.1.1.2 eth0, 1 hop
  - 10.1.3.0/24 -> 10.1.5.5 eth1, 2 hops
  - 10.1.4.0/24 -> 10.1.5.5 eth1, 1 hop
  - 10.1.5.0/24 directly eth1
  - 192.168.1.0/24 directly eth2
  - 192.168.2.0/24 -> 10.1.1.2 eth0, 1 hop
  - 192.168.3.0/24 -> 10.1.1.2 eth1, 2 hops
  - 192.168.4.0/24 -> 10.1.5.5 eth1, 2 hop
  - 192.168.5.0/24 -> 10.1.5.5 eth1, 1 hop

OSPF Route Metrics

- Depending on link speeds:
  - up to vendors, an example: 10Mbps – 10, 100Mbps – 1
- Depending on priority:
  - inter-area – highest priority
  - intra-area
  - external 1
  - external 2 – lowest priority
- Priority preempts metrics

BGP routing

- Inter-domain routing protocol
  - allows for routes exchange between Autonomous Systems – AS-es
- Working as the Internet-building protocol since 1994
  - replaced obsolete now EGP protocol
  - originally classful, currently supports classless routing
  - version 4 is being used nowadays
- The only one employing TCP connections to transmit data and keep track of changes
- Very well scaling
  - known to support both Inter- (eBGP) and Intra-domain (iBGP) routing

BGP-related RFC documents (!)

Selective Route Refresh for BGP, IETF draft

RFC 1772, Application of the Border Gateway Protocol in the Internet Protocol (BGP-4) using SMIv2

RFC 2439, BGP Route Flap Damping

RFC 2918, Route Refresh Capability for BGP-4

RFC 3765, NOPEER Community for Border Gateway Protocol (BGP) Route Scope Control

RFC 4271, A Border Gateway Protocol 4 (BGP-4)

RFC 4272, BGP Security Vulnerabilities Analysis

RFC 4273, Definitions of Managed Objects for BGP-4

RFC 4274, BGP-4 Protocol Analysis

RFC 4275, BGP-4 MIB Implementation Survey

RFC 4276, BGP-4 Implementation Report

RFC 4277, Experience with the BGP-4 Protocol

RFC 4278, Standards Maturity Variance Regarding the TCP MD5 Signature Option (RFC 2385) and the BGP-4 Specification

RFC 4456, BGP Route Reflection – An Alternative to Full Mesh Internal BGP (iBGP)

RFC 4724, Graceful Restart Mechanism for BGP

RFC 4760, Multiprotocol Extensions for BGP-4

RFC 4893, BGP Support for Four-octet AS Number Space

RFC 5065, Autonomous System Confederations for BGP

RFC 5492, Capabilities Advertisement with BGP-4

RFC 5575, Dissemination of Flow Specification Rules

RFC 7752, North-Bound Distribution of Link-State and Traffic Engineering Information Using BGP

RFC 7911, Advertisement of Multiple Paths in BGP

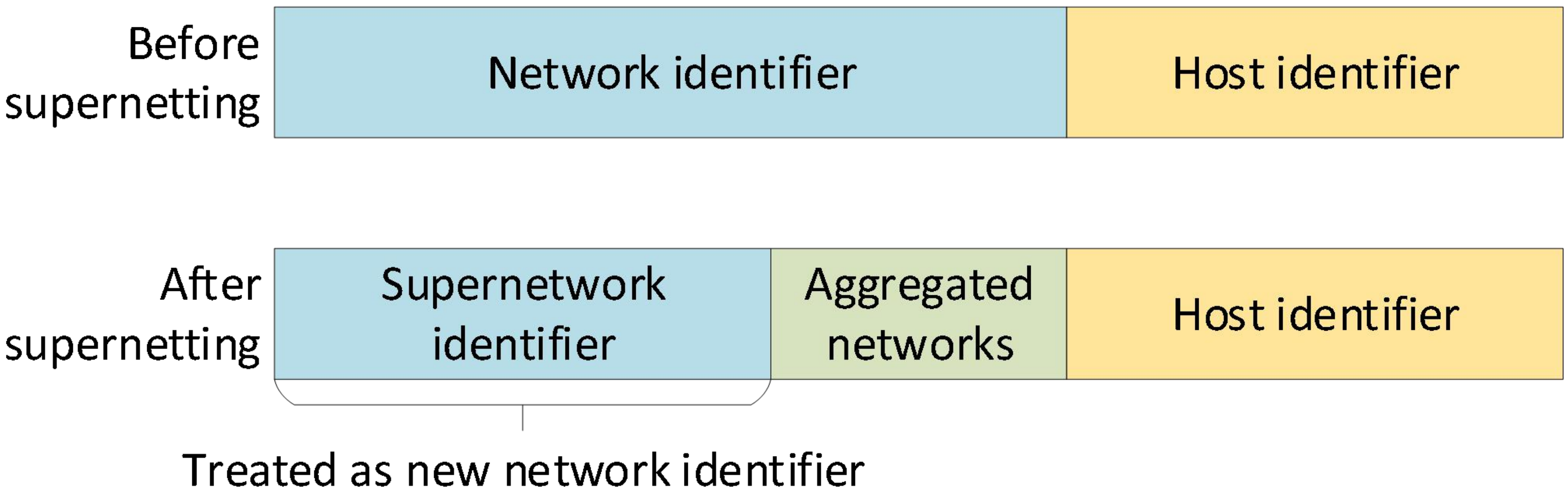
draft-ietf-idr-custom-decision-08 – BGP Custom Decision Process, Feb 3, 2017

RFC 3392, Obsolete – Capabilities Advertisement with BGP-4

RFC 2796, Obsolete – BGP Route Reflection – An Alternative to Full Mesh iBGP

Supernetting

- Supernetting – aka route aggregation



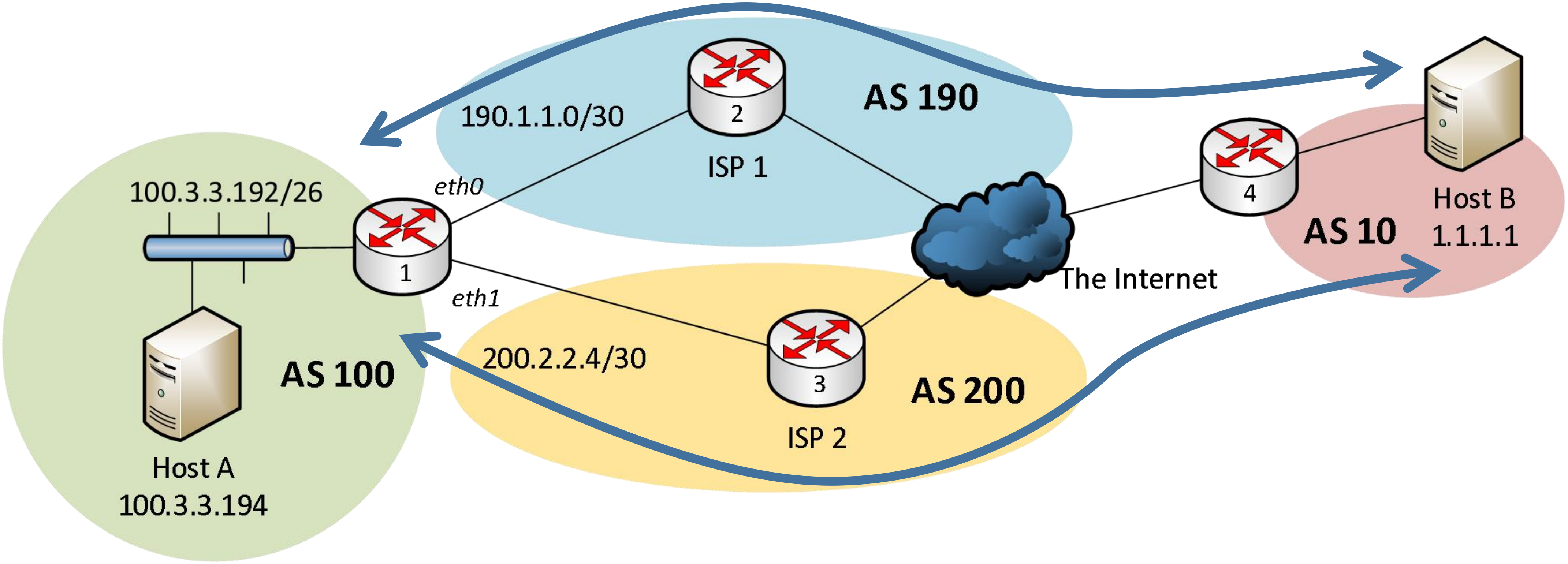
Supernetting example

- Two (2^1) networks example:
  192.168.2.0/24 -> third byte in binary: 0000 0010, mask 1111 1111
  192.168.3.0/24 -> third byte in binary: 0000 0011, mask 1111 1111
- aggegated to a supernet with one bit shorter mask:
  192.168.2.0/23 -> third byte in binary: 0000 0010, mask 1111 1110

- Similarly: eight (2^3) consecutive networks from:
  192.168.64.0/24 (0100 0000 mask 1111 1111) to:
  192.168.71.0/24 (0100 0111 mask 1111 1111)
- can be aggregated to a single network:
  192.168.64.0/21 (0100 0000 mask 1111 1000 – three bits shorter mask)

Alternative BGP applications

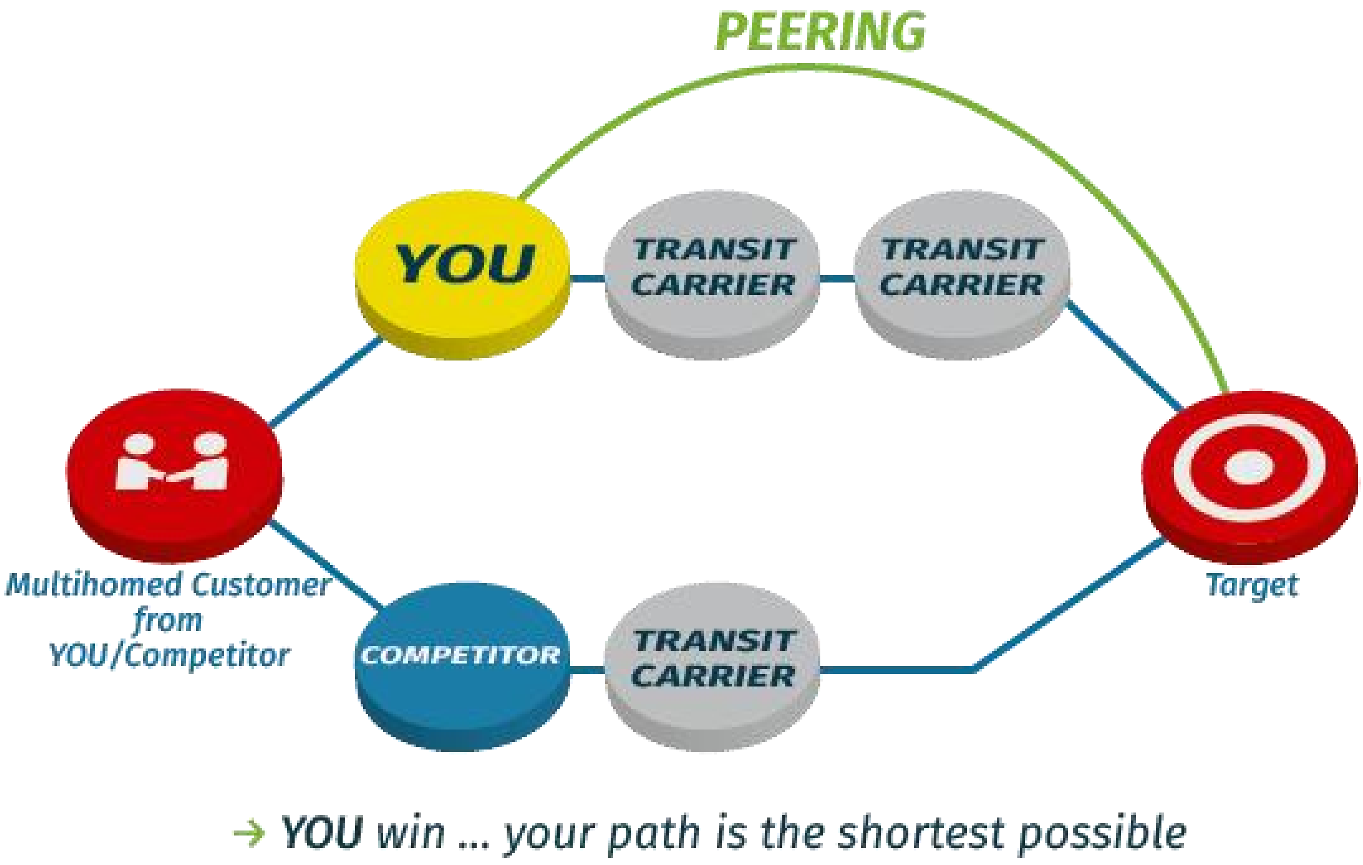
- Supports IPv4, IPv6 and other addressing protocols like MPLS
- Protocol of choice for multi-homed AS connections



Alternative BGP application – peering

- BGP needs bi-directional communication channel to establish TCP stream, hence needs a bit of routing
- There exist some Internet nodes which act as so-called Content Providers
- If a content provider buys so-called peering with an IXP (Internet eXchange Point), communication typically occurs in L2 using:
  - direct fiber-optic links
  - L2 technologies like Q-in-Q or MPLS (sometimes called L2.5)
- As no L3 is involved, customers perceive CP closer in the number of IP hops

BGP Peering example


PEERING
YOU
TRANSIT CARRIER
TRANSIT CARRIER
Multihomed Customer from YOU/Competitor
Target
COMPETITOR
TRANSIT CARRIER
→ YOU win ... your path is the shortest possible

https://www.de-cix.net/en/de-cix-service-world/globepeer/ten-reasons-why-you-should-peer

Top Polish IXPs

	EPIX	Equinix (PLIX)*	Thinx	TPIX
Full name	EPIX Internet Exchange	Equinix Internet Exchange	Thinx IX	TPIX
Owner	Stowarzyszenie e-Południe	Equinix	ATM S.A. (Atman)	Orange Polska S.A.
The resources				
Participants	780	350	170	254
Average traffic	1800 Gb/s	1500 Gb/s*	500 Gb/s	800 Gb/s
Foreign IXPs	DE-CIX, NIX, AMS-IX	–	DE-CIX, Giganet.ua	b.d
Polish IXPs	Thinx, Equinix/PLIX, TPIX, POZIX	Thinx, EPIX	EPIX (Polmix), Equinix/PLIX	EPIX, Equinix/PLIX
Tier 1/2	RETN, GTT, Telia, CenturyLink, Liberty Global, Hurricane Electric, Cogent	Operatorzy dostępni w LIM	Tata Communications, GTT	Telia Sonera

https://www.atman.pl/blog-post/5-rzeczy-ktore-chciales-wiedziec-o-punktach-wymiany-ruchu-ale-bales-sie-zapytac/

Top Polish IXPs

	EPIX	Equinix (PLIX)*	Thinx	TPIX
Facebook	yes	yes	yes	-
Google Global Cache	yes	yes	yes	yes
Akamai	yes	yes	yes	-
Atende (redCDN)	yes	yes	yes	yes
CloudFlare	yes	yes	yes	yes
NetFlix	yes	yes	yes	yes
OVH	-	yes	yes	yes

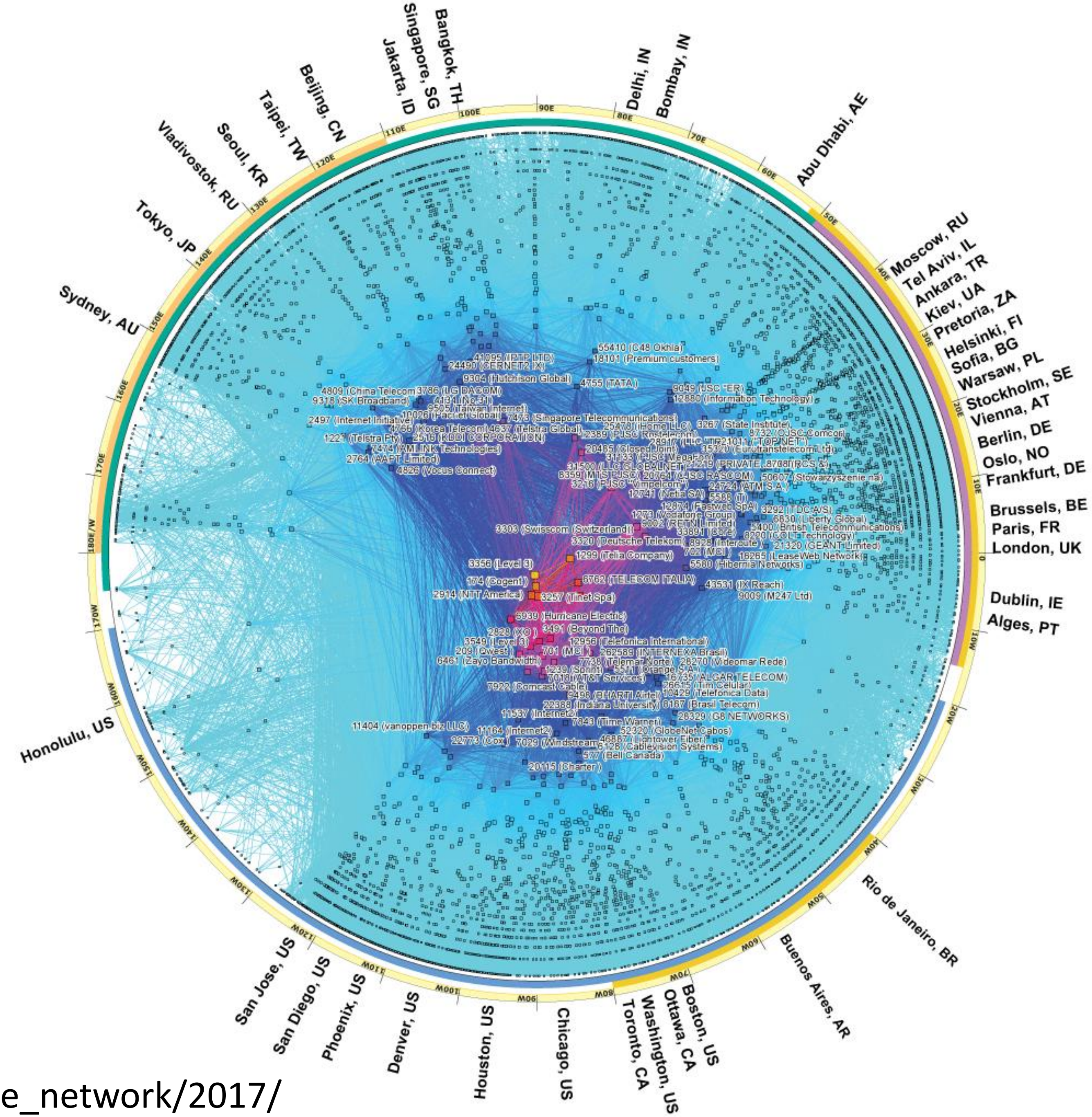
https://www.atman.pl/blog-post/5-rzeczy-ktore-chciales-wiedziec-o-punktach-wymiany-ruchu-ale-bales-sie-zapytac/

The Internet

Tiers in the Internet

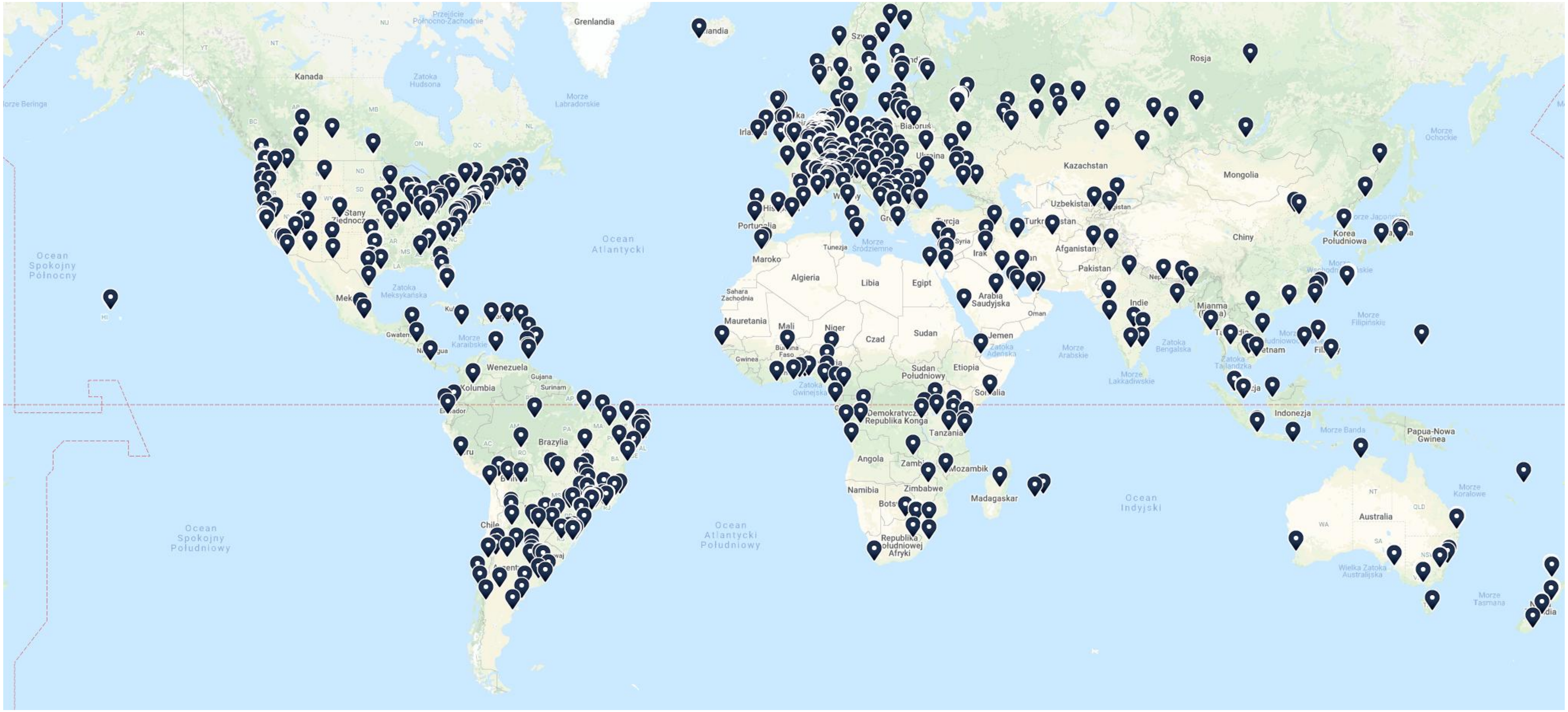
- There are three so-called tiers in the Internet
- Tiers allow to classify ISPs – Internet Service Providers from the closest (Tier 1 ISP) to the most distant (Tier 3 ISP) from the Internet core
- The most common ISP we typically interact with is a Tier 3 ISP – they sell Internet to customers and buy so-called peering
- In general Tier 2 ISPs do barter transactions – volume of traffic entering them should equal the volume of traffic leaving them
- Tier 1 ISPs actually form the Internet core
  - only 16 of them
- A Tier 1 ISP is an ISP that has access to the entire Internet Region solely via its free and reciprocal peering agreements

Example visualization



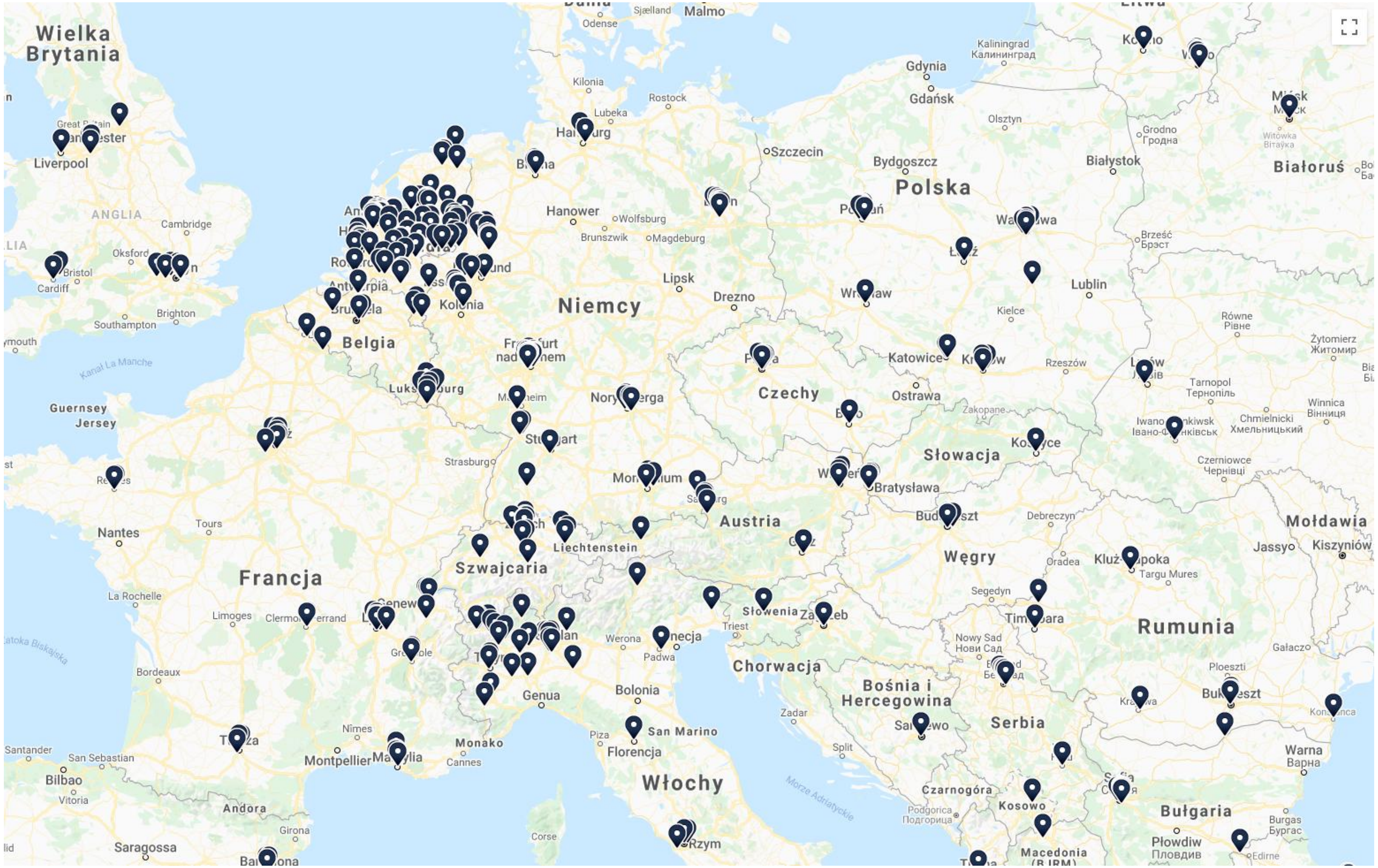
https://www.caida.org/research/topology/as_core_network/2017/

IXP map



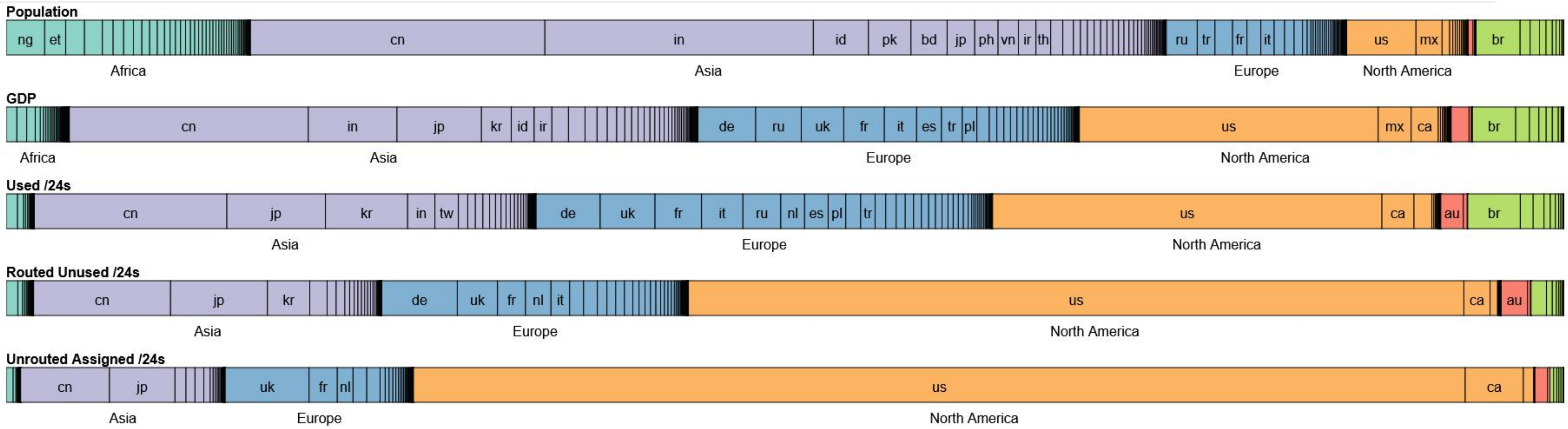
Grenlandia
landia
Kanada
Zatoka Hudsona
Morze Labradorskie
Przejście Północno-Zachodnie
Morze Beringa
AK
YT
NT
NU
BC
AB
MB
ON
QC
NL
ND
SD
WY
NC
HI
Ocean Spokojny Północny
Stany Zjednoczone
Zatoka Meksykańska
Mek
Ku
Gwaten
Morze Karaibskie
Wenezuela
Gujana
Surinam
Kolumbia
Ekwador
Peru
Brazylia
Boliwia
Chile
Argent
Ocean Spokojny Południowy
Ocean Atlantycki
Ocean Atlantycki Południowy
Irla
Portugalia
His
Maroko
Algieria
Libia
Egipt
Tunezja
Morze Śródziemne
Sahara Zachodnia
Mauretania
Mali
Niger
Czad
Sudan
Burkina Faso
Gwinea
Zatoka Gwinejska
Sudan Południowy
Etiopia
Somalia
Demokratyczna Republika Konga
Tanzania
Angola
Zambia
Mozambik
Namibia
Zimbabwe
Bots
Madagaskar
Republika Południowej Afryki
Ocean Indyjski
Białoruś
Ukraina
Turcja
Syria
Irak
Arabia Saudyjska
Oman
Jemen
Zatoka Adeńska
Morze Arabskie
Rosja
Kazachstan
Uzbekistan
Turkmenistan
Afganistan
Pakistan
Indie
Nepal
Mongolia
Chiny
Korea Południowa
Morze Japońskie
Morze Ochockie
Mianma
Wietnam
Filipiny
Morze Filipińskie
Zatoka Bengalska
Morze Lakkadiwskie
Zatoka Tajlandzka
Indonezja
Morze Banda
Papua-Nowa Gwinea
Australia
NT
QLD
WA
SA
NSW
Morze Koralowe
Wielka Zatoka Australijska
Morze Tasmana

IXP map



Wielka Brytania
Liverpool
ANGLIA
Cambridge
Oksford
Bristol
Cardiff
Brighton
Southampton
Kanał La Manche
Guernsey
Jersey
Nantes
Tours
La Rochelle
Limoges
Francja
Bordeaux
Santander
San Sebastian
Bilbao
Vitoria
Andora
Girona
Saragossa
Montpellier
Nîmes
Monako
Cannes
Belgia
Strasburg
Szwajcaria
Liechtenstein
Odense
Sjælland
Malmo
Kilonia
Lubeka
Rostock
Hanower
Wolfsburg
Brunszwik
Magdeburg
Lipsk
Drezno
Niemcy
Szczecin
Gdynia
Gdańsk
Kaliningrad
Калининград
Olsztyn
Bydgoszcz
Polska
Białystok
Grodno
Гродна
Brześć
Брэст
Lublin
Kielce
Katowice
Rzeszów
Ostrawa
Zakopane
Czechy
Słowacja
Bratysława
Austria
Węgry
Debreczyn
Oradea
Targu Mures
Segedyn
Nowy Sad
Нови Сад
Rumunia
Ploeszti
Gałacz
Mołdawia
Jassy
Kiszyniów
Białoruś
Witówka
Вітаўка
Równe
Рівне
Żytomierz
Житомир
Tarnopol
Тернопіль
Winnica
Вінниця
Chmielnicki
Хмельницький
Czerniowce
Чернівці
Werona
Padwa
Triest
Słowenia
Chorwacja
Bośnia i Hercegowina
Zadar
Split
Serbia
Czarnogóra
Podgorica
Подгорица
Kosowo
Macedonia
Bułgaria
Płowdiw
Пловдив
Burgas
Бургас
Warna
Варна
Edirne
Genua
Bolonia
San Marino
Piza
Florencja
Włochy
Corse
Morze Adriatyckie

IPv4 address usage

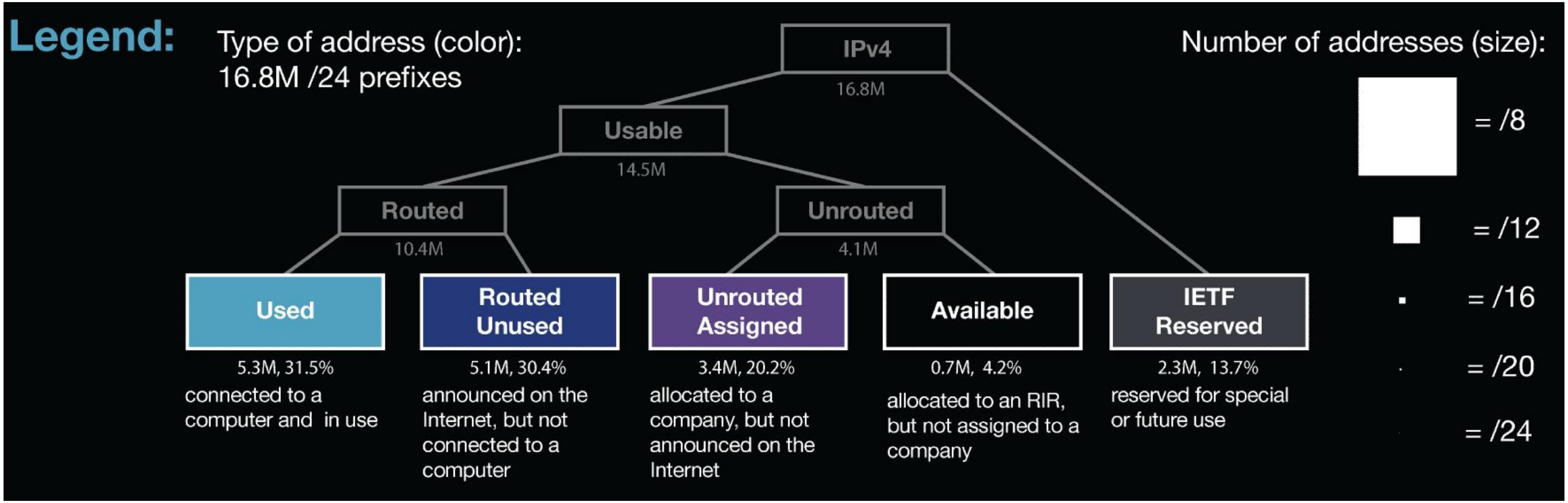


- Routed unused – subject to study using passive and active probing (explained in publication)
- Unrouted Assigned – sold, but not present in BGP routing tables at all

  based on publicatin „Lost in Space: Improving Inference of IPv4 Address Space Utilization" from 2014

  Interactive map:
  https://www.caida.org/publications/papers/2014/lost_in_space/supplemental/country_inequality/

Registered IPv4 addresses heatmap – legend



Legend:

Type of address (color):
16.8M /24 prefixes

IPv4
16.8M

Usable
14.5M

Routed
10.4M

Unrouted
4.1M

Used
5.3M, 31.5%
connected to a computer and in use

Routed Unused
5.1M, 30.4%
announced on the Internet, but not connected to a computer

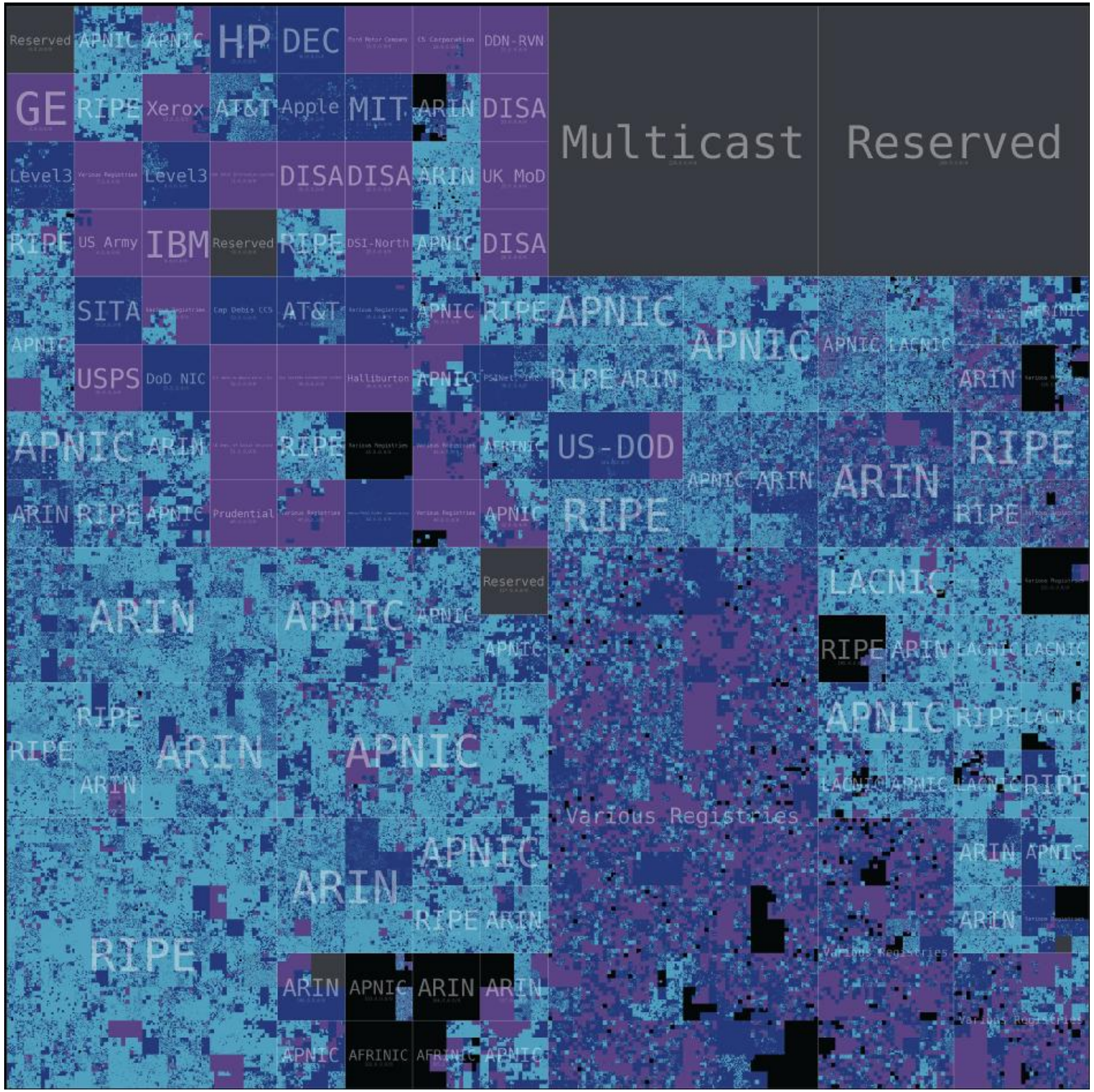
Unrouted Assigned
3.4M, 20.2%
allocated to a company, but not announced on the Internet

Available
0.7M, 4.2%
allocated to an RIR, but not assigned to a company

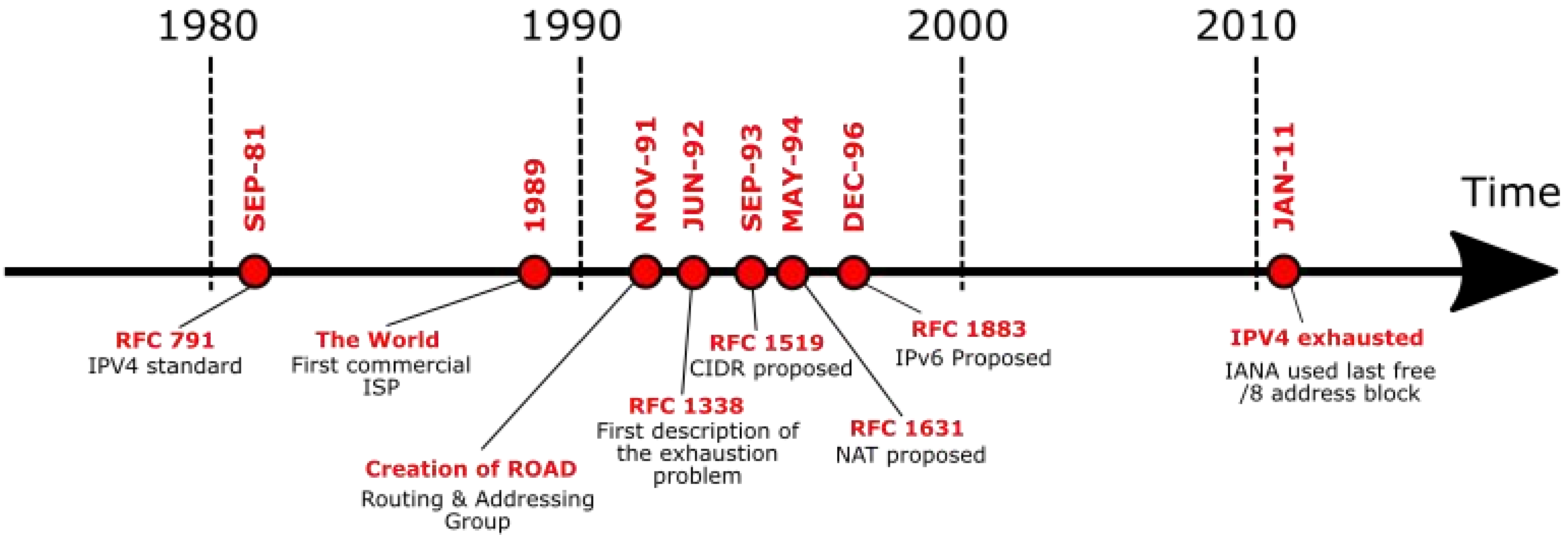
IETF Reserved
2.3M, 13.7%
reserved for special or future use

Number of addresses (size):

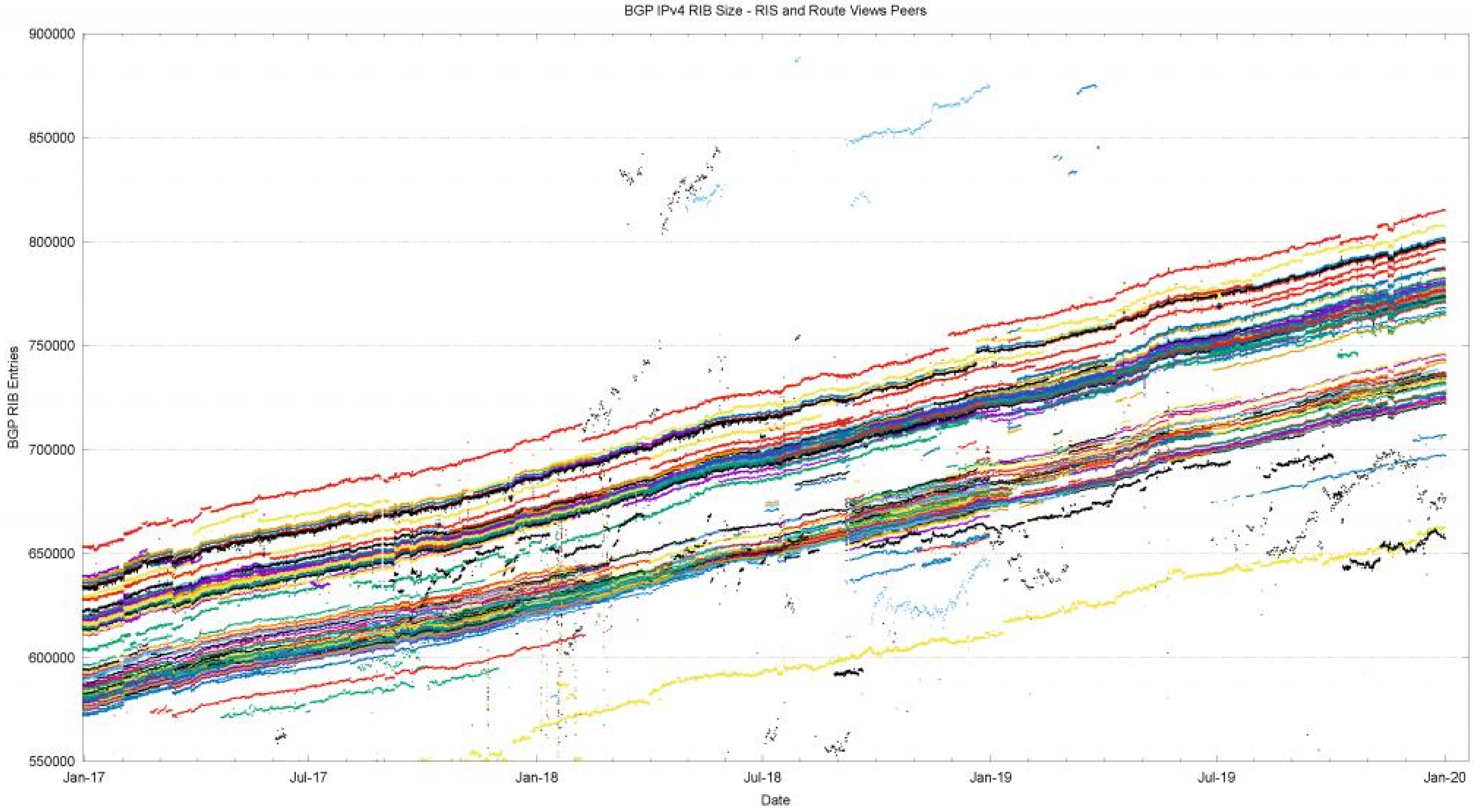
IPv4 addresses heatmap



IPv4 Address Exhaustion – timeline



BGP Rounting table sizes

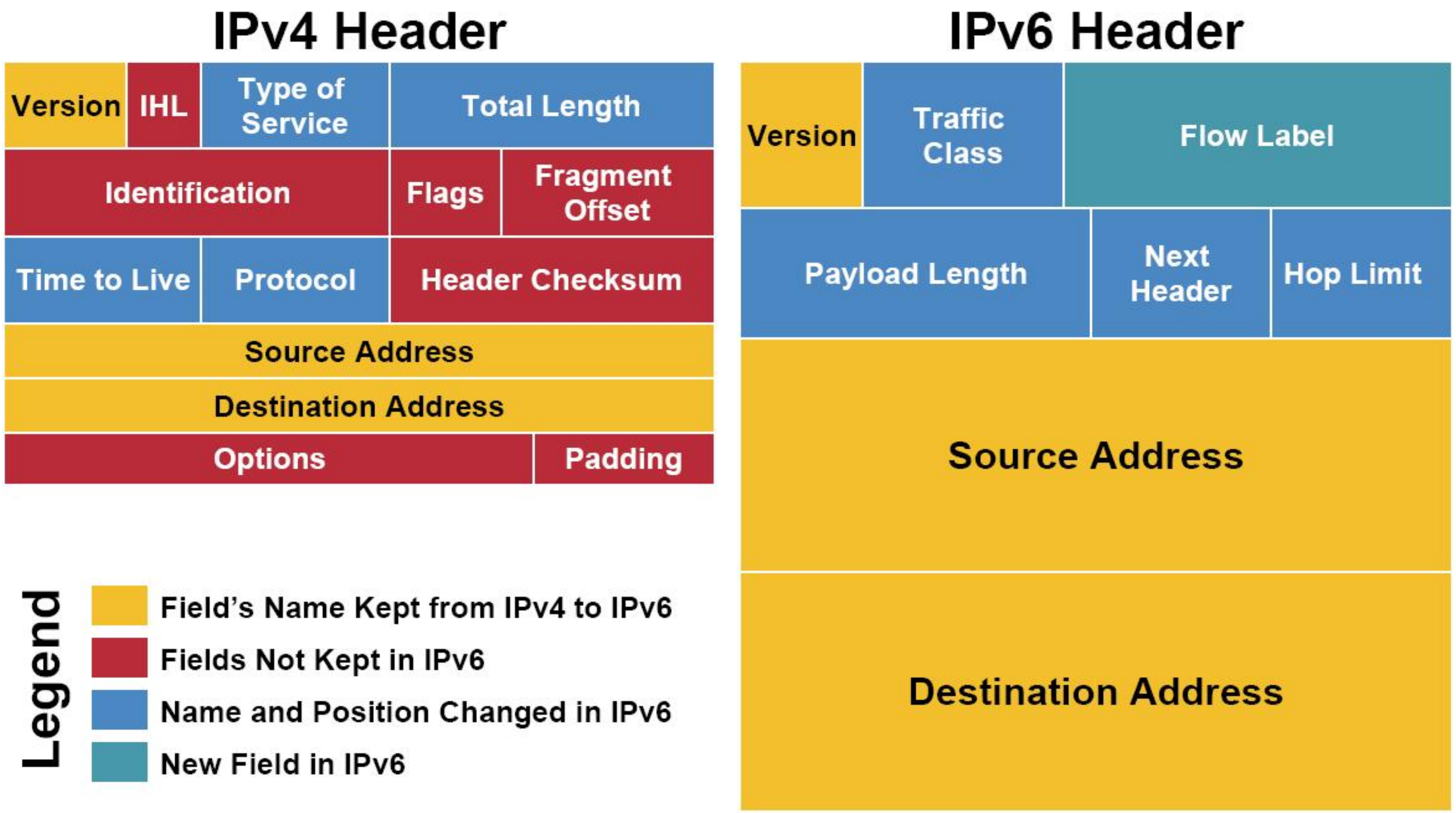


The IPv6 Protocol

Motivation behind IPv6 introduction

- First of all – the exhausting IPv4 address space
- Efficiency of routing
- Simplification of header processing
- The artificial broadcast limiting by routers

Header complexity comparison



IPv4 Header

Version | IHL | Type of Service | Total Length

Identification | Flags | Fragment Offset

Time to Live | Protocol | Header Checksum

Source Address

Destination Address

Options | Padding

IPv6 Header

Version | Traffic Class | Flow Label

Payload Length | Next Header | Hop Limit

Source Address

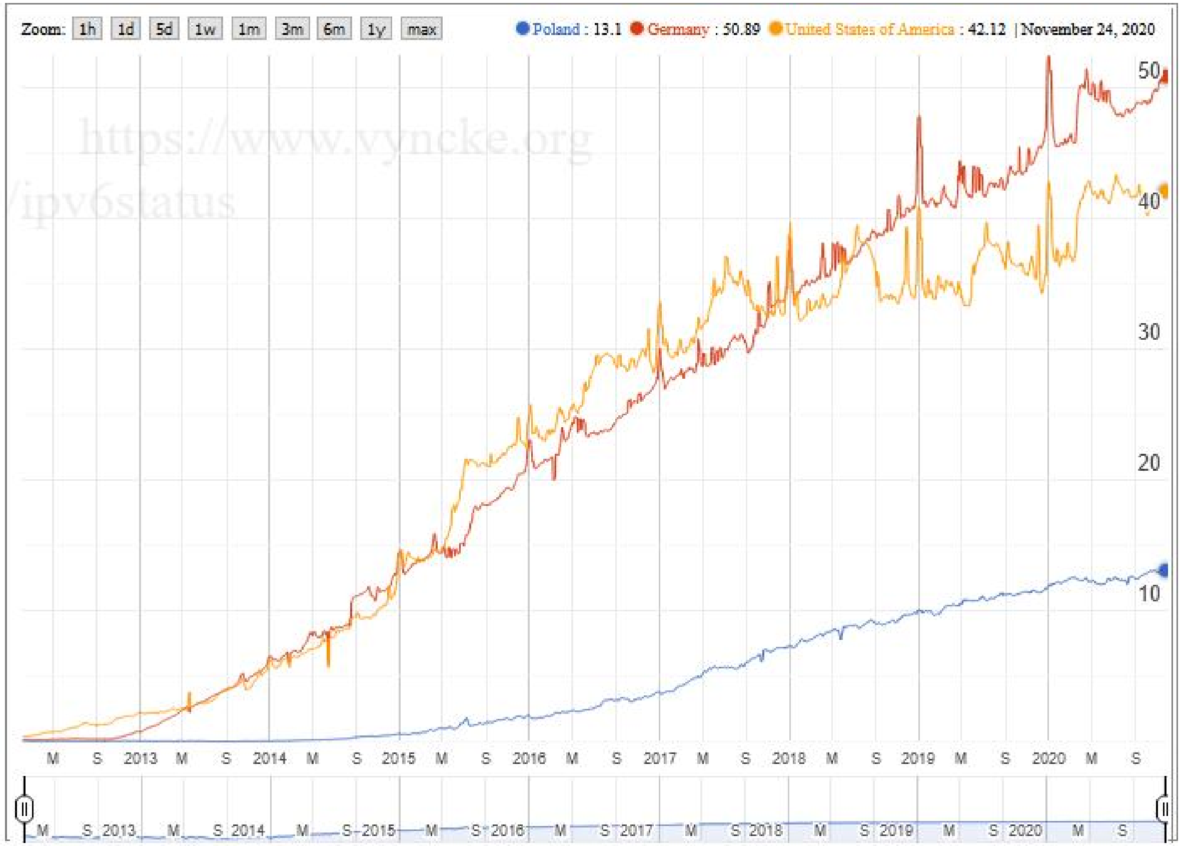
Destination Address

Legend

- 🟨 Field’s Name Kept from IPv4 to IPv6
- 🟥 Fields Not Kept in IPv6
- 🟦 Name and Position Changed in IPv6
- 🟩 New Field in IPv6

IPv6 adoption in different countries

- Interactive portal:
  - https://www.vyncke.org/ipv6status



https://www.vyncke.org/ipv6status/compare.php?metric=p&countries=pl,us,de

IPv6 address notations

- IPv6 addresses length
    - 128 bits = 16 bytes
    - 4 x 4 bytes = 4 x IPv4 address length
- Noted in hexadecimal instead of decimal in groups of two bytes
    - 2 B = four hexadecimal digits
- Colons instead of dots
    - example IPv6 address: 2001:4070:0011:0500:0000:0000:0000:0100

IPv6 address notations

- Fully noted IPv6 address (quite rare in practice):

  2001:4070:0011:0500:0000:0000:0000:0100
- The first applied shortening in notation – omit non-significant zeros in four-digit blocks:

  2001:4070:11:500:0:0:0:100
- Second applied shortening in notation – omit single consecutive block of zeros and put two colons instead:

  2001:4070:11:500::100
- Due to special meaning of colons square brackets are used in some cases:

  http://[2001:4070:11:500::100]/

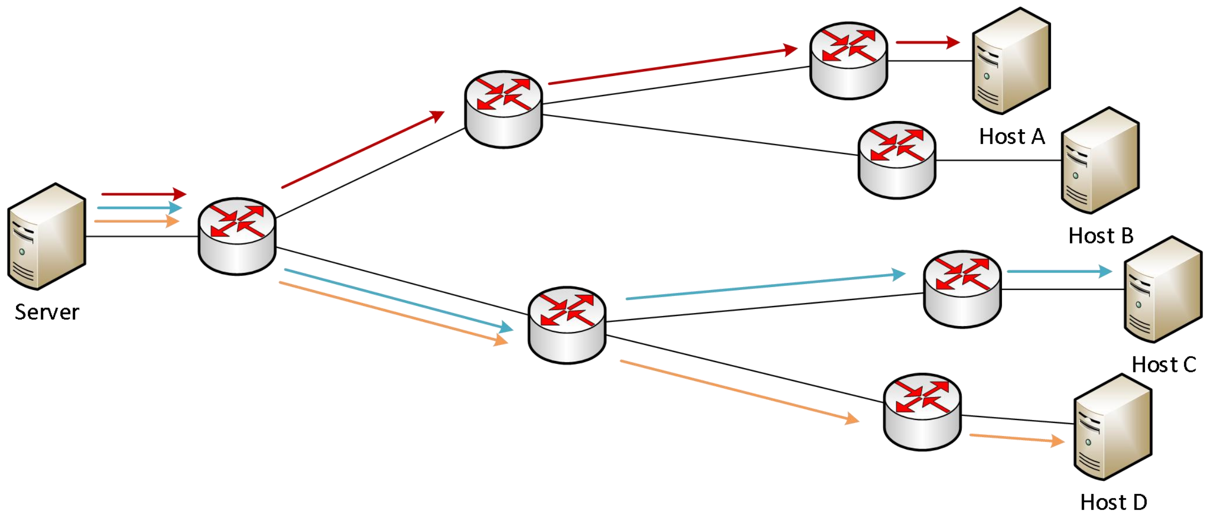
Datagram fragmentation in IPv6

- No fragmentation possible by intermediate routers on the path
- All needed fragmentation done at the source
  - no fragmentation fields in IPv6 header whatsoever
  - mandatory MTU discovery at the source needed

Address types in IPv6

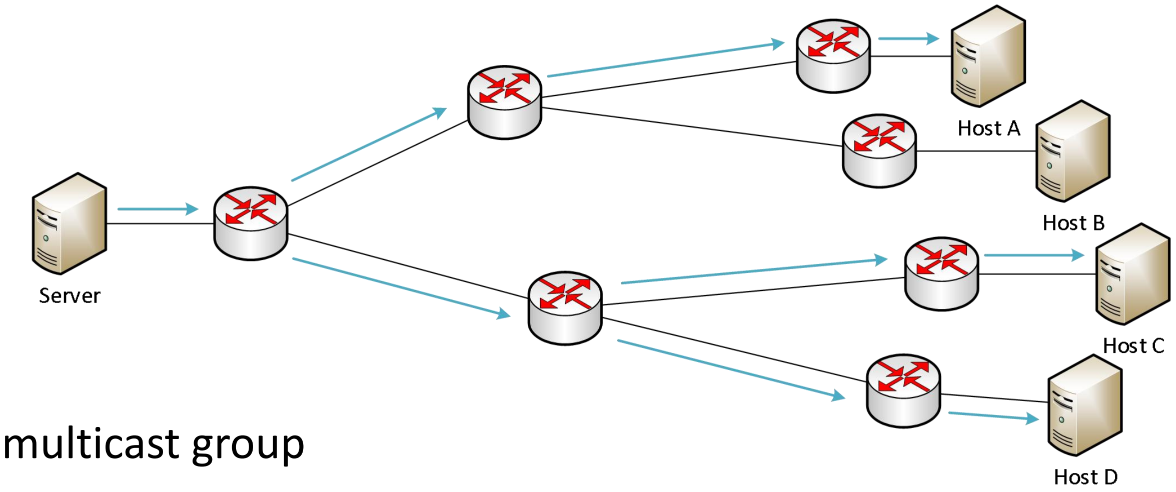
- Unicast
  - typical IPv6 addresses used for one-to-one host communication
- Multicast
  - addresses used in one-to-many communication
  - mapped to multicast MAC addresses in broadcast media like Ethernet or WiFi
- Anycast
  - addresses used in one-to-”one of many”
  - non-unique addresses across the globe
  - routers route traffic to the closest one
  - used in geographically-spread services to balance the load
- No broadcast addresses at all

Unicast and multicast comparison



Unicast

host B not receiving unicast stream



Multicast

host B not participating in multicast group

Typical IPv6 address scopes

- Global
  - most of IPv6 addresses
  - the whole 2000::/3 address space
- Link-local
  - used in communication to the nearest router only (single broadcast domain in broadcast media like Ethernet or WiFi)
  - fe80::/10 (fe80:: to febf:ffff:ffff:ffff:ffff:ffff:ffff:ffff)
- Multicast
  - ff00::/8 (ff00:: to ffff:ffff:ffff:ffff:ffff:ffff:ffff:ffff)
- Private
  - fc00::/7 (fc00:: to fdff:ffff:ffff:ffff:ffff:ffff:ffff:ffff)

ICMPv6 introduced functionalities

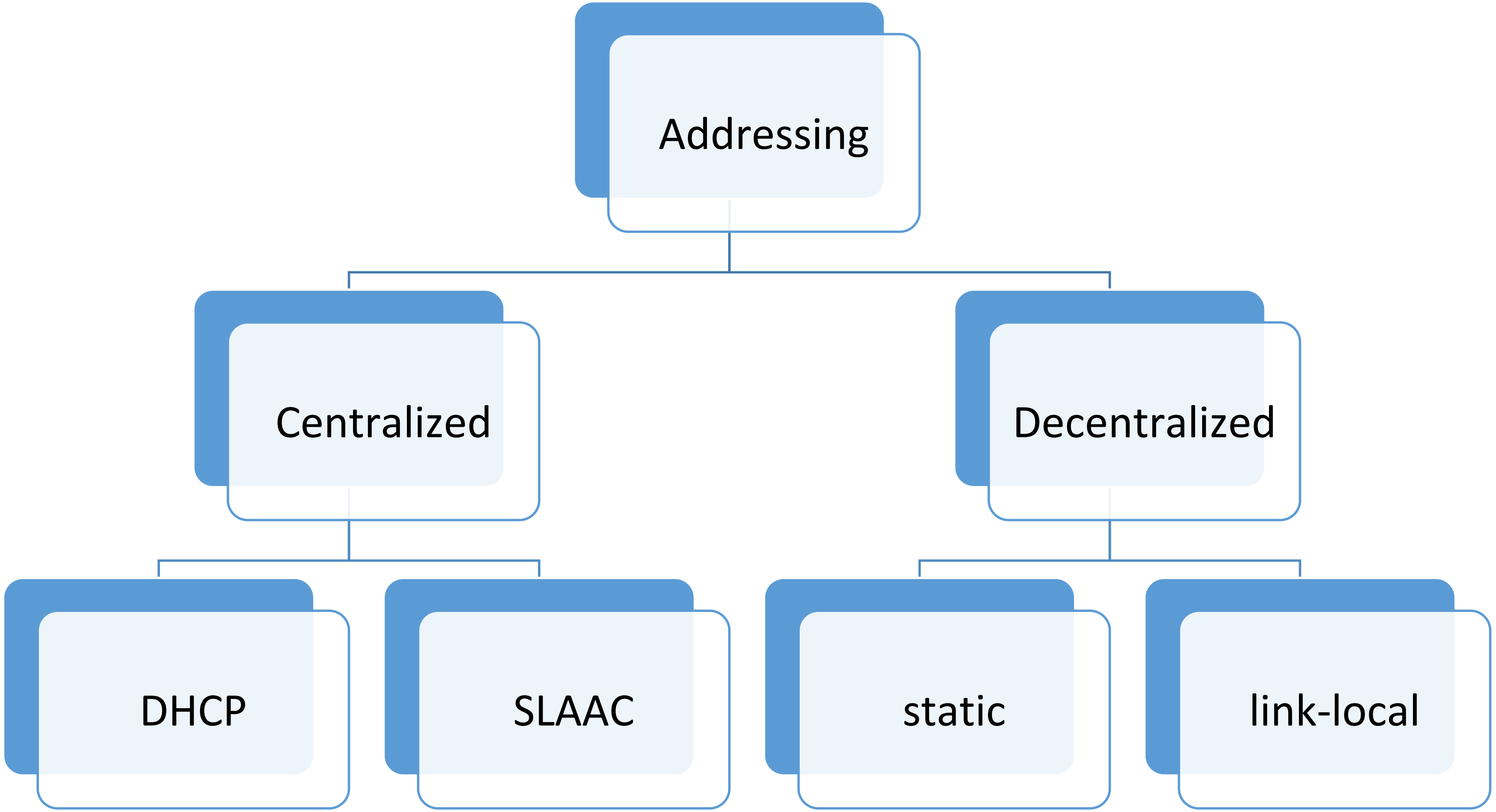
- ND – Neighbor Discovery
  - functionally replaces ARP known from IPv4
- MLD – Multicast Listener Discovery
  - allows for management of multicast listener groups
  - important for mapping to L2 switching mechanisms
- SLAAC – Stateless Address Autoconfiguration
  - automatic decentralized IPv6 address configuration on nodes
  - possible to employ in both link-local and global IPv6 address configuration
- RR – Router Renumbering
  - prefix management and IPv6 router autoconfiguration

IPv6 multicast addresses

Address	Scope	Purpose
FF01::1	Node-Local	All nodes
FF01::2	Node-Local	All routers
FF02::1	Link-Local	All nodes
FF02::2	Link-Local	All routers
FF02::5	Link-Local	OSPFv3 Routers
FF02::6	Link-Local	OSPFv3 Designated Routers

- In general the third byte is divided into two 4b fields:
  - flags (always 0 for well-known addresses)
  - scope
    - 1 – Node-Local
    - 2 – Link-Local

IPv6 node addressing possibilities



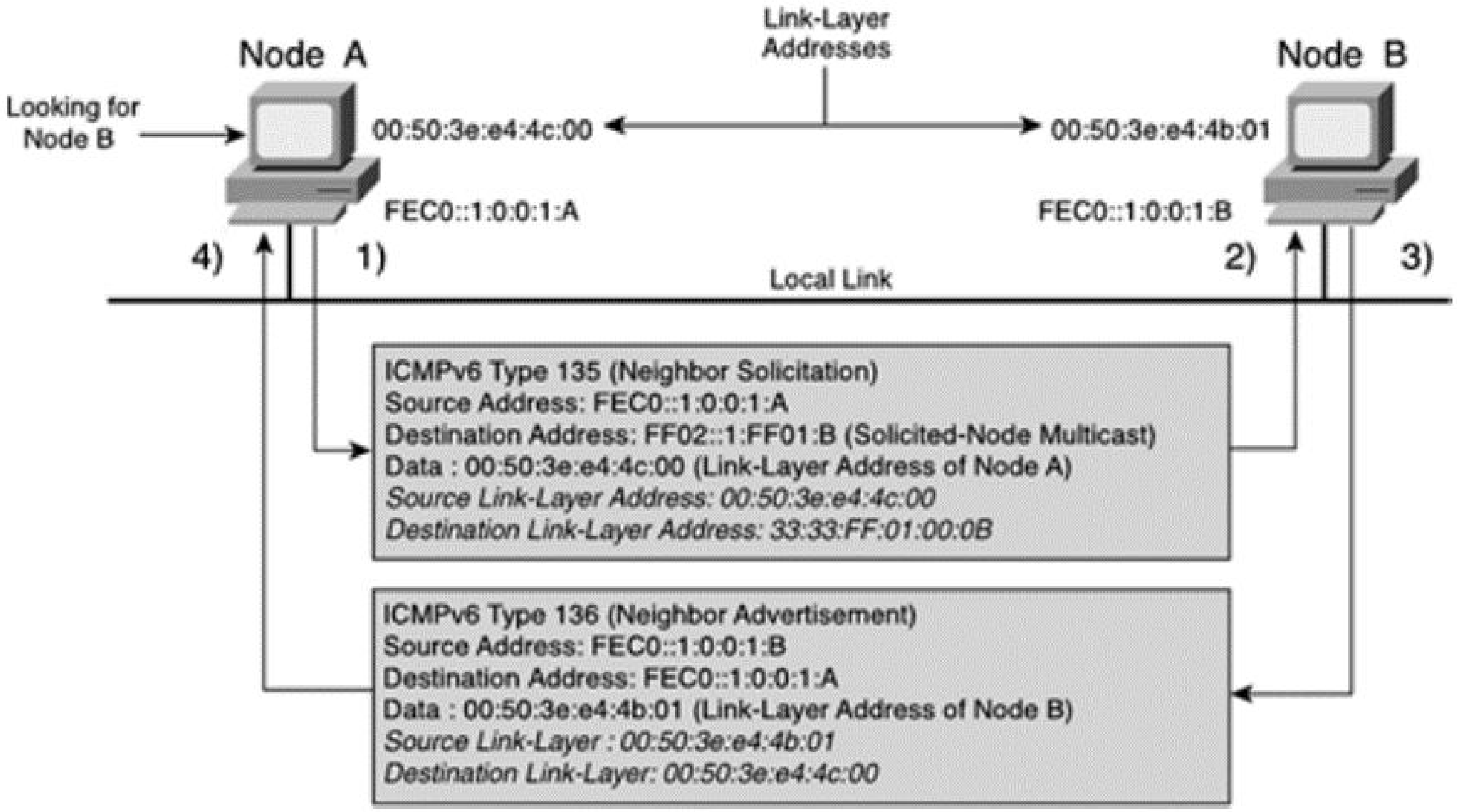
IPv6 link-local addresses

- Rely on MAC addresses, which get first converted to EUI-64 form
  - MAC adress (48 bits)
    00:01:de:ad:be:ef
  - EUI-64 (64 bits)
    00:01:de:ff:fe:ad:be:ef
- Allow for communication regarding just the broadcast domain (link scope)
  - IPv6 link-local address
    fe80::0001:deff:fead:beef

SLAAC – Stateless Address Autoconfiguration

- Rely on prefixes advertised by routers
- Privacy extensions
  - improve address randomization

The Neighbor Discovery Protocol



ICMPv6 Type 135 (Neighbor Solicitation)
Source Address: FEC0::1:0:0:1:A
Destination Address: FF02::1:FF01:B (Solicited-Node Multicast)
Data : 00:50:3e:e4:4c:00 (Link-Layer Address of Node A)
Source Link-Layer Address: 00:50:3e:e4:4c:00
Destination Link-Layer Address: 33:33:FF:01:00:0B

ICMPv6 Type 136 (Neighbor Advertisement)
Source Address: FEC0::1:0:0:1:B
Destination Address: FEC0::1:0:0:1:A
Data : 00:50:3e:e4:4b:01 (Link-Layer Address of Node B)
Source Link-Layer : 00:50:3e:e4:4b:01
Destination Link-Layer: 00:50:3e:e4:4c:00