How The Internet Works

The Life and Times of an HTTP Request

Noah Kantrowitz
SF Python

It's not a big truck.
It's a series of tubes.

Ted Stevens

We never, ever in the history of mankind have had access to so much information so quickly and so easily.

Vint Cerf

Google

Google

https://www.google.com

Google

Google Search I'm Feeling Lucky

https://www.google.com/

https://www.google.com/

DNS

DNS

- Map names to IP addresses.
- gethostbyname()
- RFC 1034 & 1035.
- "What is the A for www.google.com?"

DNS Header

ID							
QR	Opcode	AA	TC	RD	RA	Z	RCODE
QDCOUNT							
ANCOUNT							
NSCOUNT							
ARCOUNT							

DNS Header

ID							
QR	Opcode	AA	TC	RD	RA	Z	RCODE
QDCOUNT							
ANCOUNT							
NSCOUNT							
ARCOUNT							

DNS Header

ID							
QR	Opcode	AA	TC	RD	RA	Z	RCODE
QDCOUNT							
ANCOUNT							
NSCOUNT							
ARCOUNT							

DNS Header

ID							
QR	Opcode	AA	TC	RD	RA	Z	RCODE
QDCOUNT							
ANCOUNT							
NSCOUNT							
ARCOUNT							

DNS Header

ID							
QR	Opcode	AA	TC	RD	RA	Z	RCODE
QDCOUNT							
ANCOUNT							
NSCOUNT							
ARCOUNT							

DNS Question

QNAME
QTYPE
QCLASS

DNS Question

QNAME
QTYPE
QCLASS

DNS Question

QNAME
QTYPE
QCLASS

DNS Question

QNAME

QTYPE

QCLASS

DNS Message

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

DNS Message

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

DNS Message

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

DNS Message

```
0001010000010000
0000000003777777
06676F6F676C6503
636F6D0000010001
```

Network

Show All

Wi-Fi

Wi-Fi | TCP/IP | DNS | WINS | 802.1X | Proxies | Hardware

DNS Servers:

208.201.224.11
208.201.224.33

Search Domains:

+ − IPv4 or IPv6 addresses

+ −

?

Cancel OK

IP and UDP

IP and UDP

- Address and port.
- 208.201.224.11:53
- Wrapped in order.
- DNS inside UDP inside IP.
- RFC 791 & 768.

UDP Header

Source Port	Destination Port
Length	Checksum
Data	

UDP Header

Source Port	Destination Port
Length	Checksum
Data	

IP Header

Version	IHL	Type of Service	Total Length	
Identification			Flags	Fragment Offset
Time to Live	Protocol	Header Checksum		
Source Address				
Destination Address				
Options				Padding

IP Header

Version	IHL	Type of Service	Total Length	
Identification			Flags	Fragment Offset
Time to Live	Protocol	Header Checksum		
Source Address				
Destination Address				
Options				Padding

IP Header

Version	IHL	Type of Service	Total Length	
Identification			Flags	Fragment Offset
Time to Live	Protocol	Header Checksum		
Source Address				
Destination Address				
Options				Padding

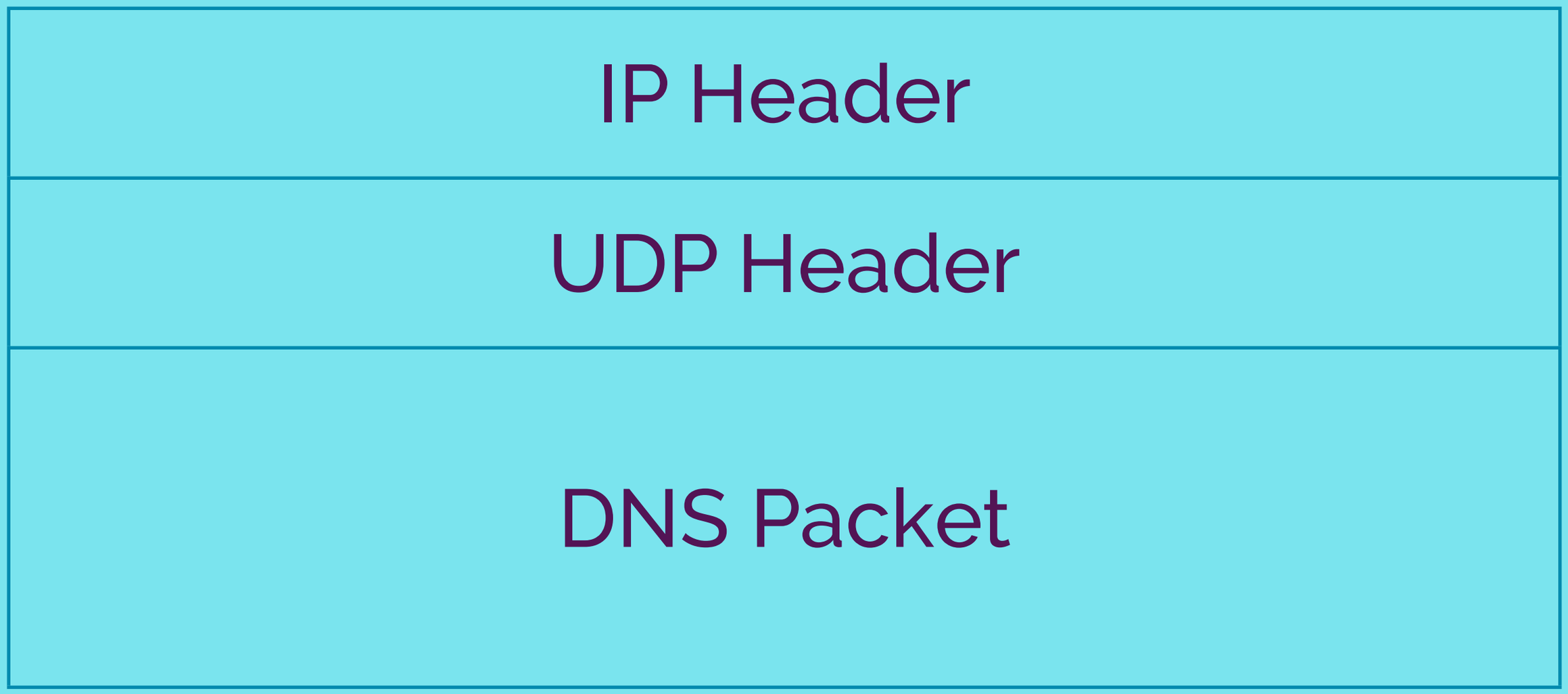
IP Header

Version	IHL	Type of Service	Total Length	
Identification			Flags	Fragment Offset
Time to Live	Protocol	Header Checksum		
Source Address				
Destination Address				
Options				Padding

IP Header

Version	IHL	Type of Service	Total Length	
Identification			Flags	Fragment Offset
Time to Live	Protocol	Header Checksum		
Source Address				
Destination Address				
Options				Padding

IP Packet



Local Network

Local Network

- Computer
- Switch
- Router
- Modem

Network

Show All

Wi-Fi

Wi-Fi | TCP/IP | DNS | WINS | 802.1X | Proxies | Hardware

Configure IPv4: Using DHCP

IPv4 Address: 192.168.1.3

Renew DHCP Lease

Subnet Mask: 255.255.255.0

DHCP Client ID:

(If required)

Router: 192.168.1.1

Configure IPv6: Automatically

Router:

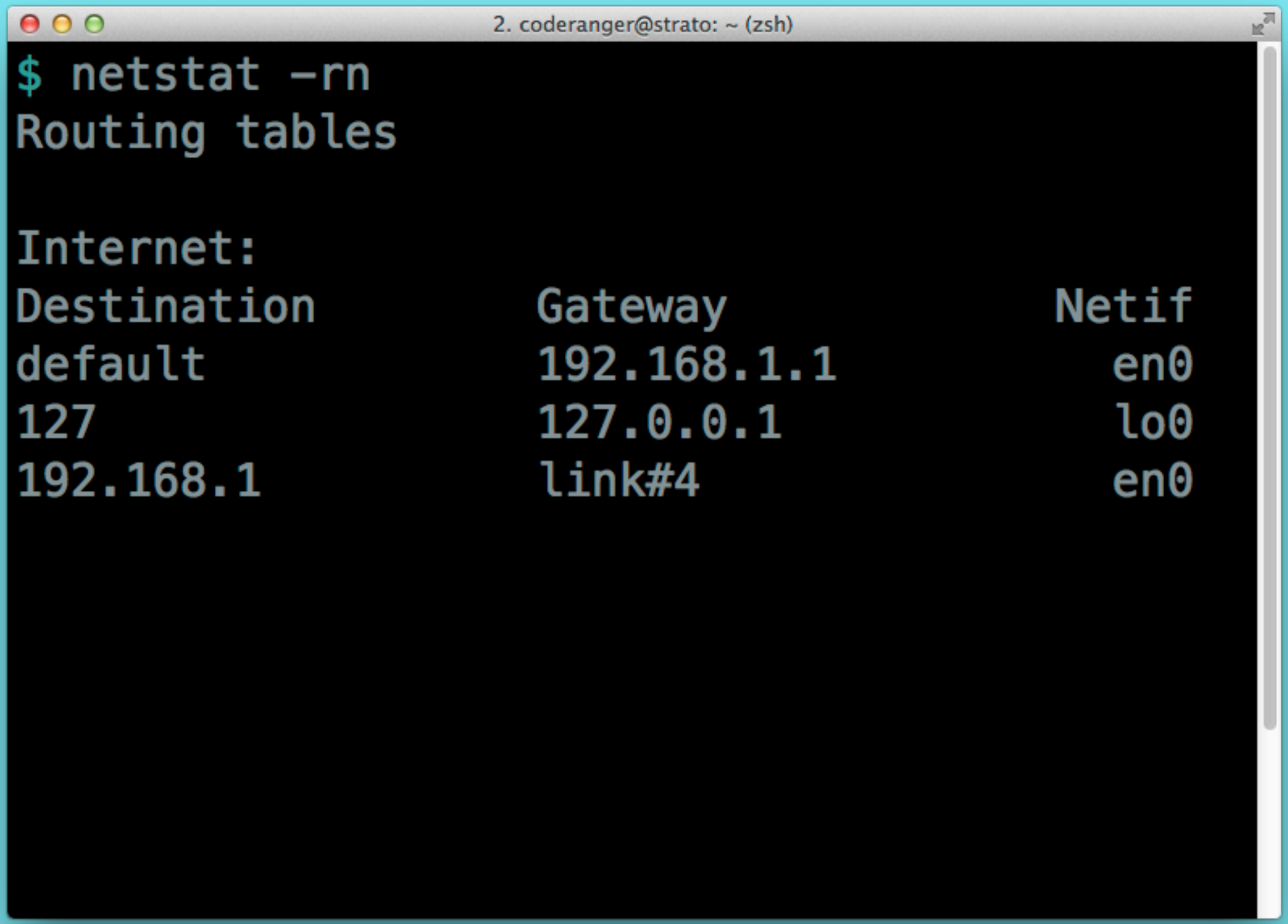
IPv6 Address:

Prefix Length:

?

Cancel

OK


2. coderanger@strato: ~ (zsh)
$ netstat -rn
Routing tables
Internet:
Destination Gateway Netif
default 192.168.1.1 en0
127 127.0.0.1 lo0
192.168.1 link#4 en0

208.201.224.11

Destination	Gateway	Interface
127.0.0.0/8	*	lo
192.168.1.0/24	*	eth0
default	192.168.1.1	eth0

208.201.224.11

Destination	Gateway	Interface
127.0.0.0/8	*	lo
192.168.1.0/24	*	eth0
default	192.168.1.1	eth0

208.201.224.11

Destination	Gateway	Interface
127.0.0.0/8	*	lo
192.168.1.0/24	*	eth0
default	192.168.1.1	eth0

208.201.224.11

Destination	Gateway	Interface
127.0.0.0/8	*	lo
192.168.1.0/24	*	eth0
default	192.168.1.1	eth0

Ethernet

- Wire framing.
- MAC address.
- Segments.
- IEEE 802.3

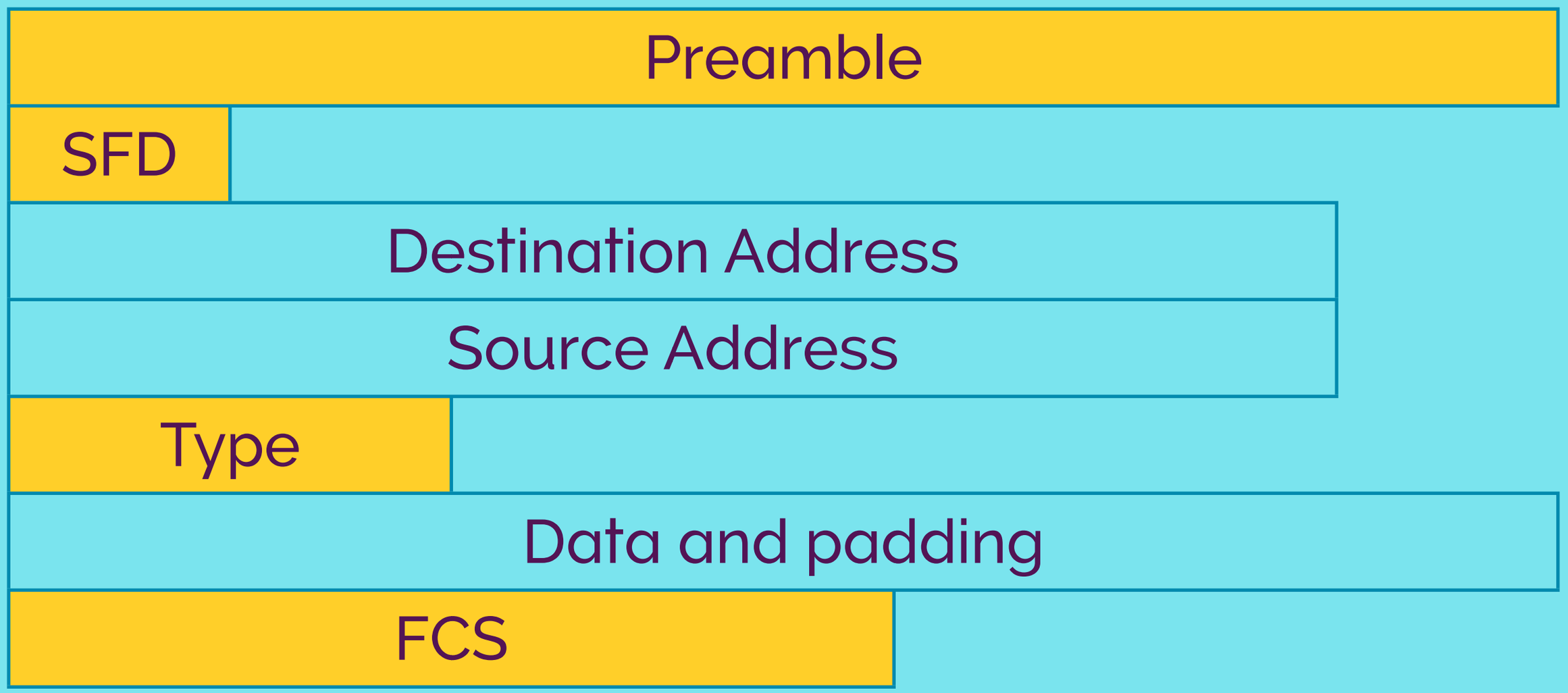
Segments

- Electrical broadcast.
- Thicknet.
- 10BASE-T and hubs.
- Switches.

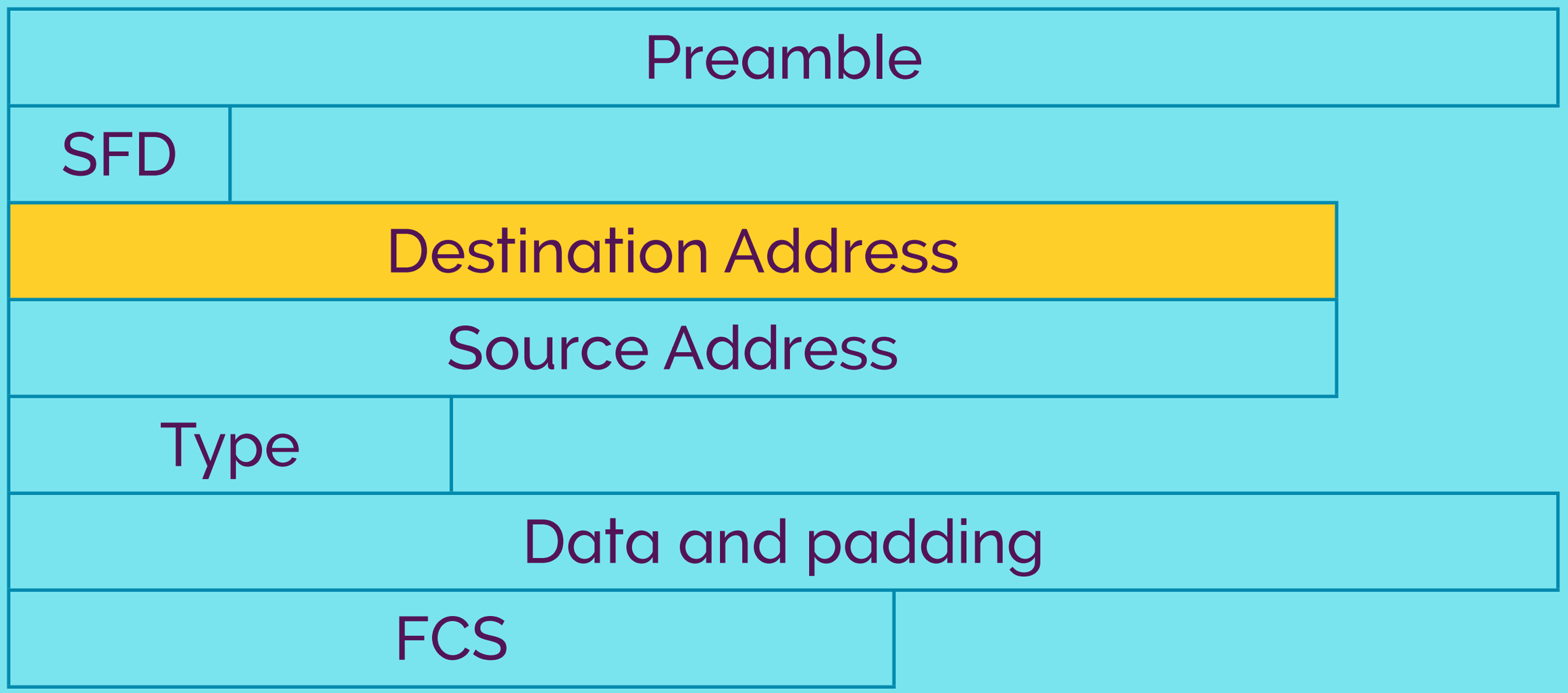
Ethernet Frame

Preamble	
SFD	
Destination Address	
Source Address	
Type	
Data and padding	
FCS	

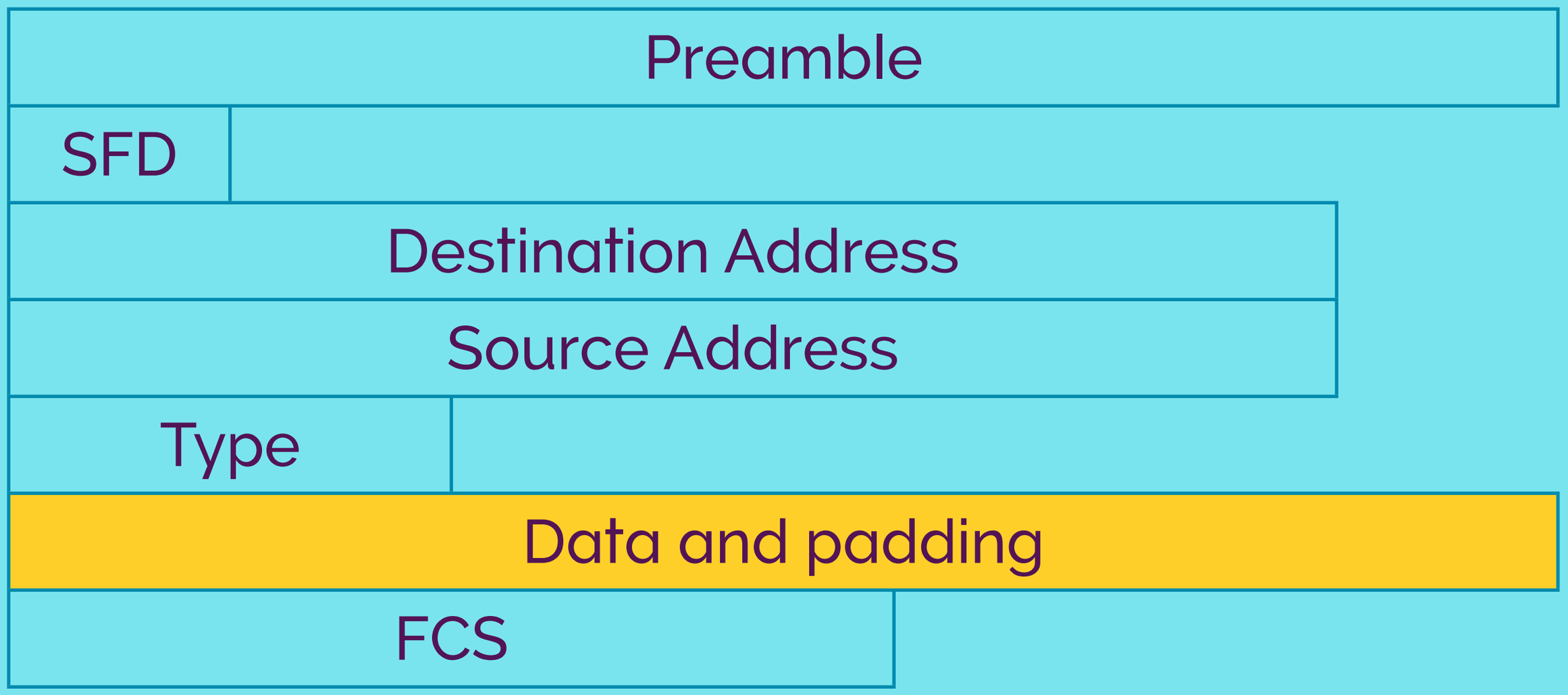
Ethernet Frame



Ethernet Frame



Ethernet Frame



1000-BASET

- Electrical signaling.
- IEEE 802.3ab

Ethernet Frame

Preamble	
SFD	
Destination Address	
Source Address	
Type	
Data and padding	
FCS	

ARP

ARP

- Bridge between IP and Ethernet.
- NDP for IPv6.
- "Who has IP address 1.2.3.4?"
- In an Ethernet frame.
- Broadcast FF:FF:FF:FF:FF:FF.

ARP Packet

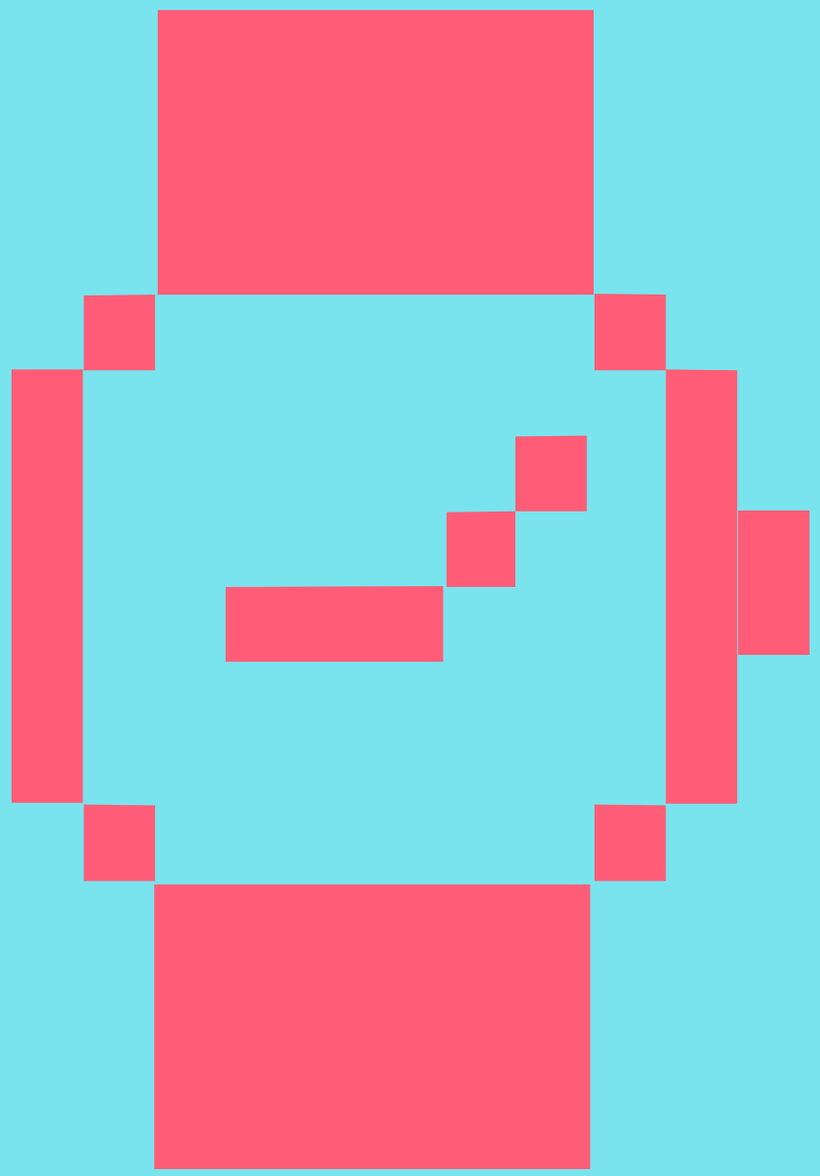
HTYPE	PTYPE	HLEN	PLEN	OPER
SHA				SPA
THA				TPA

ARP Packet

HTYPE	PTYPE	HLEN	PLEN	OPER
SHA				SPA
THA				TPA

ARP Packet

HTYPE	PTYPE	HLEN	PLEN	OPER
SHA				SPA
THA				TPA



Sending

- Ethernet – 40:4A:3:ED:D2:1C
- IP – 208.201.224.11
- UDP – 53
- DNS – www.google.com A

Local Router

Local Router

- Static routing (again).
- NAT.

Route Table

Destination	Gateway	Interface
127.0.0.0/8	*	lo
192.168.1.0/24	*	eth0
173.228.34.0/24	*	eth1
default	173.228.34.1	eth1

208.201.224.11

Destination	Gateway	Interface
127.0.0.0/8	*	lo
192.168.1.0/24	*	eth0
173.228.34.0/24	*	eth1
default	173.228.34.1	eth1

208.201.224.11

Destination	Gateway	Interface
127.0.0.0/8	*	lo
192.168.1.0/24	*	eth0
173.228.34.0/24	*	eth1
default	173.228.34.1	eth1

208.201.224.11

Destination	Gateway	Interface
127.0.0.0/8	*	lo
192.168.1.0/24	*	eth0
173.228.34.0/24	*	eth1
default	173.228.34.1	eth1

208.201.224.11

Destination	Gateway	Interface
127.0.0.0/8	*	lo
192.168.1.0/24	*	eth0
173.228.34.0/24	*	eth1
default	173.228.34.1	eth1

NAT

- Rewrite IP and UDP header.
  - Source address and port.
  - Share one public address.

Sending, Round Two

- ARP lookup (or cached).
- Ethernet frame.
- Send to ISP border router.

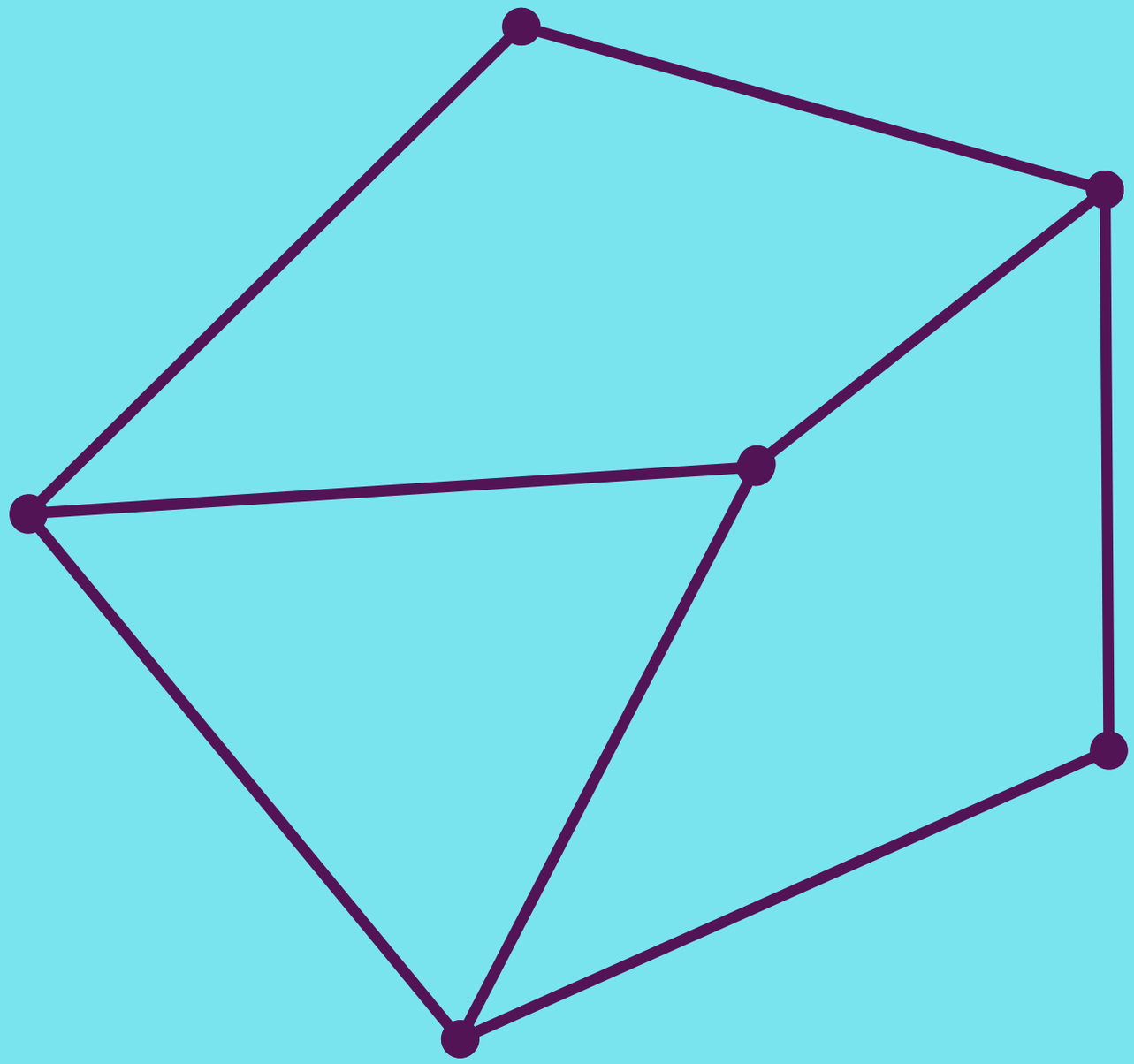
The Internet

A Mesh of Trees

- Tree-like at the edges.
- Partial mesh in middle.
- ~Full mesh in the core.

Next Hops

- Regional routers.
- ISP backbone.



BGP

Routing Protocol

- Distribute routes.
- Update over time.
- Find optimal paths.

BGP

- Gossip based.
- Prefix based.
- Share best routes.
- RFC 4271.

BGP

- Gossip based.
- Prefix based.
- Share best routes.

Autonomous Systems

- 1-2^{32} (née 2^{16})
- ~51000 so far.
- AS7065

IANA

- ICANN department
- Internet Assigned Numbers Authority
- 5 Regional Internet Registries
- AfriNIC, ARIN, APNIC
- LACNIC, RIPE NCC

BGP Algorithm

1 Weight

2 Local Pref

3 AS Path

4 Origin

5 MED

6 Metric

7 First

8 Tie Breaker

BGP Algorithm

1 Weight

2 Local Pref

3 AS Path

4 Origin

5 MED

6 Metric

7 First

8 Tie Breaker

BGP Algorithm

1 Weight

2 Local Pref

3 AS Path

4 Origin

5 MED

6 Metric

7 First

8 Tie Breaker

BGP Algorithm

1 Weight

2 Local Pref

3 AS Path

4 Origin

5 MED

6 Metric

7 First

8 Tie Breaker

BGP Algorithm

1 Weight

2 Local Pref

3 AS Path

4 Origin

5 MED

6 Metric

7 First

8 Tie Breaker

BGP Algorithm

1 Weight

2 Local Pref

3 AS Path

4 Origin

5 MED

6 Metric

7 First

8 Tie Breaker

BGP Algorithm

1 Weight

2 Local Pref

3 AS Path

4 Origin

5 MED

6 Metric

7 First

8 Tie Breaker

BGP Algorithm

1 Weight

2 Local Pref

3 AS Path

4 Origin

5 MED

6 Metric

7 First

8 Tie Breaker

BGP Algorithm

1 Weight

2 Local Pref

3 AS Path

4 Origin

5 MED

6 Metric

7 First

8 Tie Breaker

Security

- BGP Hijacking
- Ingress Filtering
- Rarely used signatures

Backbone Router

- AS46375 to AS7065

DNS Server

Receiving

- Decode and match IP.
- Decode UDP port.
- Deliver to process.

DNS Server

- Decode question.
- Check local cache.
- Recursion?

DNS Recursion

Root Servers

- 13 DNS servers.
- Fixed IP addresses.
- Maps TLDs to DNS servers.
- [a-m].root-servers.net

Recursion Round 1

- DNS query to 198.41.0.4
- com. IN NS a.gtld-servers.net
- com. IN NS b.gtld-servers.net.
- a.gtld-servers.net. IN A 192.5.6.30
- b.gtld-servers.net. IN A 192.33.14.30

Recursion Round 2

- DNS query to 192.5.6.30
- google.com. IN NS ns1.google.com.
- google.com. IN NS ns2.google.com.
- ns1.google.com. IN A 216.239.32.10
- ns2.google.com. IN A 216.239.34.10

Recursion Round 3

- DNS query to 216.239.32.10
- www.google.com.
  IN A 216.58.192.36

DNS Reply

- DNS message.
- Headers & answer section.
- Sent back over the wires.
- Un-NAT.

TCP

TCP

- Reliable ACKs.
- Three-way handshake.
- Congestion control.
- RFC 675, 793, ...

TCP Headers

<table>
<tr><td colspan="8">Source Port</td><td colspan="8">Destination Port</td></tr>
<tr><td colspan="16">Sequence Number</td></tr>
<tr><td colspan="16">Acknowledgement Number</td></tr>
<tr><td>Data Offset</td><td colspan="5">Reserved</td><td>URG</td><td>ACK</td><td>PSH</td><td>RST</td><td>SYN</td><td>FIN</td><td colspan="8">Window</td></tr>
<tr><td colspan="8">Checksum</td><td colspan="8">Urgent Pointer</td></tr>
<tr><td colspan="12">Options</td><td colspan="4">Padding</td></tr>
</table>

TCP Headers

<table>
<tr><td colspan="8">Source Port</td><td colspan="8">Destination Port</td></tr>
<tr><td colspan="16">Sequence Number</td></tr>
<tr><td colspan="16">Acknowledgement Number</td></tr>
<tr><td colspan="3">Data Offset</td><td colspan="4">Reserved</td><td>URG</td><td>ACK</td><td>PSH</td><td>RST</td><td>SYN</td><td>FIN</td><td colspan="6">Window</td></tr>
<tr><td colspan="8">Checksum</td><td colspan="8">Urgent Pointer</td></tr>
<tr><td colspan="12">Options</td><td colspan="4">Padding</td></tr>
</table>

TCP Headers

Source Port								Destination Port
Sequence Number								
Acknowledgement Number								
Data Offset	Reserved	URG	ACK	PSH	RST	SYN	FIN	Window
Checksum								Urgent Pointer
Options								Padding

TCP Headers

Source Port								Destination Port
Sequence Number								
Acknowledgement Number								
Data Offset	Reserved		URG	ACK	PSH	RST	SYN	FIN	Window
Checksum								Urgent Pointer
Options								Padding

TCP Headers

Source Port							Destination Port
Sequence Number							
Acknowledgement Number							
Data Offset	Reserved	URG	ACK	PSH	RST	SYN	FIN	Window
Checksum							Urgent Pointer
Options							Padding

Three Way Handshake

- SYN
- SYN-ACK
- ACK

Acknowledgements

- Send 1 … 10.
- ACK 10.
- Send 11 … 20.
- ACK 15.
- Send 15 … 24.

Extras

- Slow-start
- Avoidance
- Fast resend
- Karn's
- Nagle's
- SACK
- Scaling
- CUBIC

TLS

TLS (aka SSL)

- Stream encryption.
- Mutual authentication.
- RFC 5246.


Crypto

TLS Handshake

- ClientHello
- SeverHello
- ChangeCipherSpec
- Finished

Client Round 1

- ClientHello

Server Round 1

- ServerHello
- Certificate
- ServerKeyExchange
- ServerHelloDone

Client Round 2

- ClientKeyExchange
- ChangeCipherSpec
- Finished

Server Round 2

- ChangeCipherSpec
- Finished

Application Data

- Data wrapper.
- Transparent.

HTTP

HTTP

- Request and response.
- Verbs, paths, codes.
- RFC 2616.

HTTP Request

- GET / HTTP/1.1
- Host: www.google.com

Sending

- TLS
- TCP
- IP
- Ethernet

HTTP Response

- HTTP/1.1 200 OK
- Content-Length: 17914
- \r\n
- <!doctype html><html ...

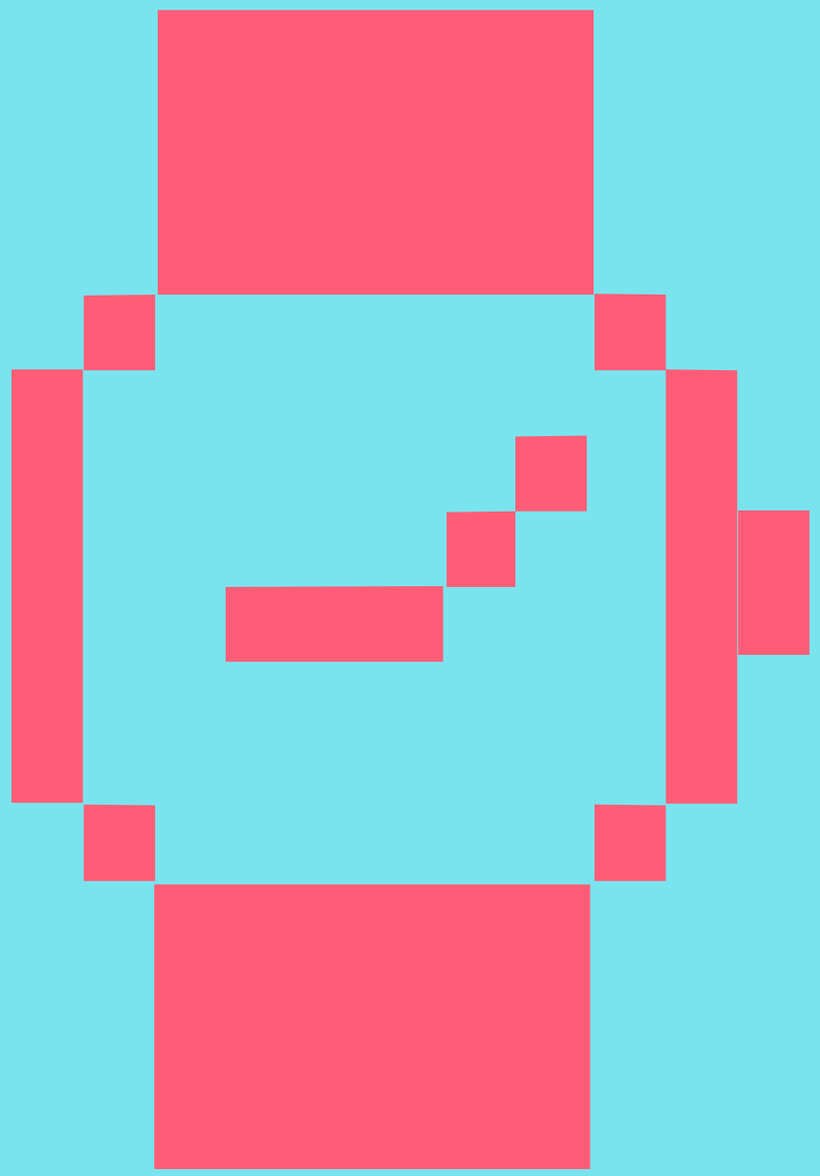
Google

Google

https://www.google.com

Google

Google Search I'm Feeling Lucky



How does the Internet work?

Surprisingly well!

Questions?