How The Internet Works

or

What You Need To Know From Computer Networking

Objective

- A different hack night - not so focused on security
- Answer "what happens when you go to google.com"
- Only concerned about the network parts
    - Moving packets from A to B

- I'll have breaks for questions
    - Lots of interdependent knowledge
- Do ask me for clarifications while I'm talking though

Layers

Layers

- Network protocols are built on top of each other
- Commonly grouped into 7 layers - the OSI model

- Physical (1)
- Data Link
- Network
- Transport
- Session
- Presentation
- Application (7)

Layer 1 - Physical

- Physical medium used to transport bits (symbols)
- How those bits are encoded

- Copper
- Coax
- Fiber

- Hubs

Layer 2 - Data Link

- Framing
- Chunks of data going from A to B
- “Physical addresses”

- Ethernet (802.3)
- PPP

- Switches

Layer 3 - Network

- Packets
- Moving packets between different physical layers
- “Logical addresses”

- IP

- Routers

Layer 4 - Transport

- Datagrams
- Multiplexing (ports), delivery guarantees (ensuring packets arrive, that they arrive in order, etc.)

- TCP
- UDP

Layer 5-7

- Abstractions within computers to make processing data easier
- Sockets
- Encryption
- HTTP, FTP, etc.

- This talk is staying at 2 <= layer <= 4

Questions?

Things We Care About

MAC Addresses (Layer 2)

- Physical address tied to your network card
- How other layer 2 devices on the same subnet talk to you
- (In theory) globally unique
- Written as 6 hex bytes (48 bits) separated with colons: 6c:40:08:b4:0a:b0

IP Addresses (Layer 3)

- Logical address for your computer
- Unique within your network
- 2 types
  - IPv4
    - 32 bits
    - 4 decimal bytes separated with dots: 1.2.3.4
  - IPv6
    - 128 bits
    - 8 groups of 2 hex bytes separated with colons: fe80:0:0:0:852:1d5c:9804:e376

IP Addresses (Layer 3)

- Some reserved ranges:
  - Loopback: 127.0.0.0/8
  - Private networks (RFC1918):
    - 10.0.0.0/8
    - 172.16.0.0/12
    - 192.168.0.0/16
  - Multicast: 224.0.0.0/4
  - And more!
- These can not be routed on the public internet

IP Networks

- Let's pick apart this **subnet**: 192.168.0.0/24
- 192.168.0.0 is the first IP in the network
- /24 is the subnet mask in CIDR notation
  - 24 is the number of 1 bits set if you write out the mask in binary
  - 11111111 11111111 11111111 00000000
  - Subnet mask may also be written in dot decimal: 255.255.255.0
- What does this do for us?
  - Sending a packet from IP A to IP B
  - If A AND subnet_mask == B AND subnet_mask, A and B are on the same network
  - The packet doesn't need to be routed

IP Networks

- Subnet: 192.168.0.0/25
- 192.168.0.3 needs to send a packet to 192.168.0.120

```
● 192.168.0.3:      11000000.10101000.00000000.00000011
● 255.255.255.128:  11111111.11111111.11111111.10000000
● Result:           11000000.10101000.00000000.00000000 = 192.168.0.0

● 192.168.0.120:    11000000.10101000.00000000.01111000
● 255.255.255.128:  11111111.11111111.11111111.10000000
● Result:           11000000.10101000.00000000.00000000 = 192.168.0.0
```

IP Networks

- Subnet: 192.168.0.0/25
- 192.168.0.3 needs to send a packet to 192.168.0.129

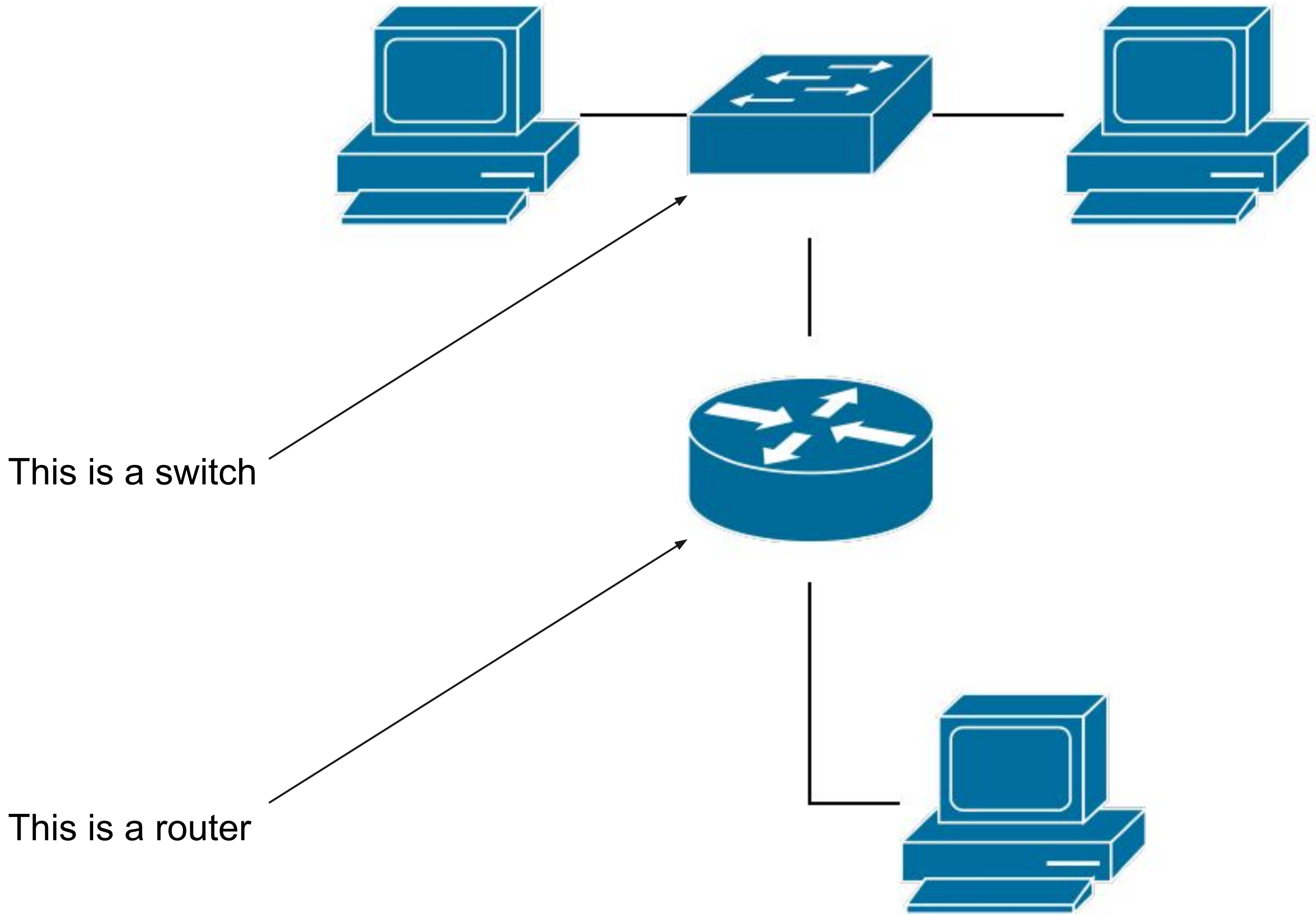
```
● 192.168.0.3:     11000000.10101000.00000000.00000011
● 255.255.255.128: 11111111.11111111.11111111.10000000
● Result:          11000000.10101000.00000000.00000000 = 192.168.0.0

● 192.168.0.129:   11000000.10101000.00000000.10000001
● 255.255.255.128: 11111111.11111111.11111111.10000000
● Result:          11000000.10101000.00000000.10000000 = 192.168.0.128
```

Questions?

How Traffic Flows

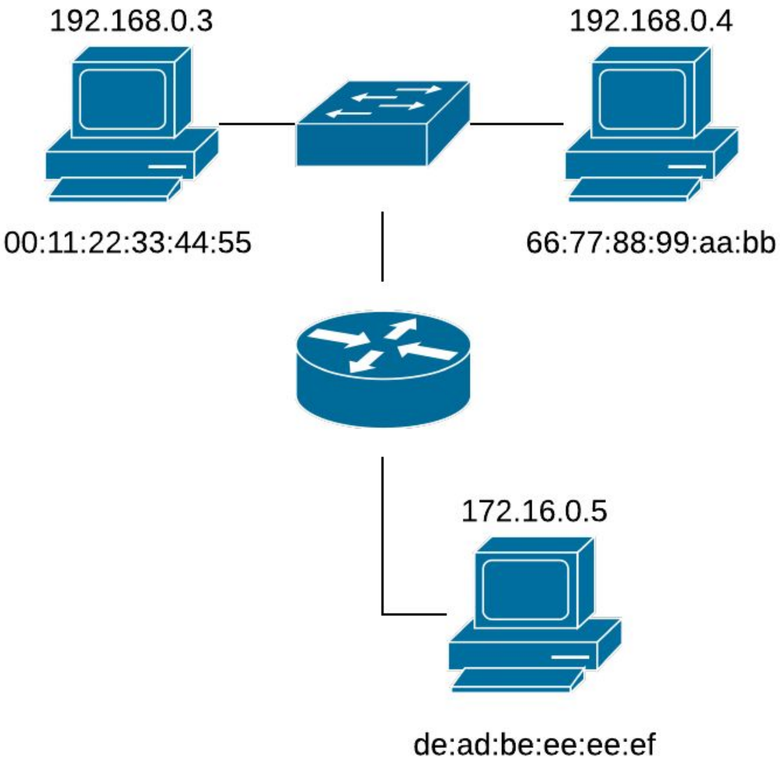
Layer 2

- Packets are delivered to devices on the subnet based on MAC address
- Cannot hop subnets without being routed


This is a switch
This is a router

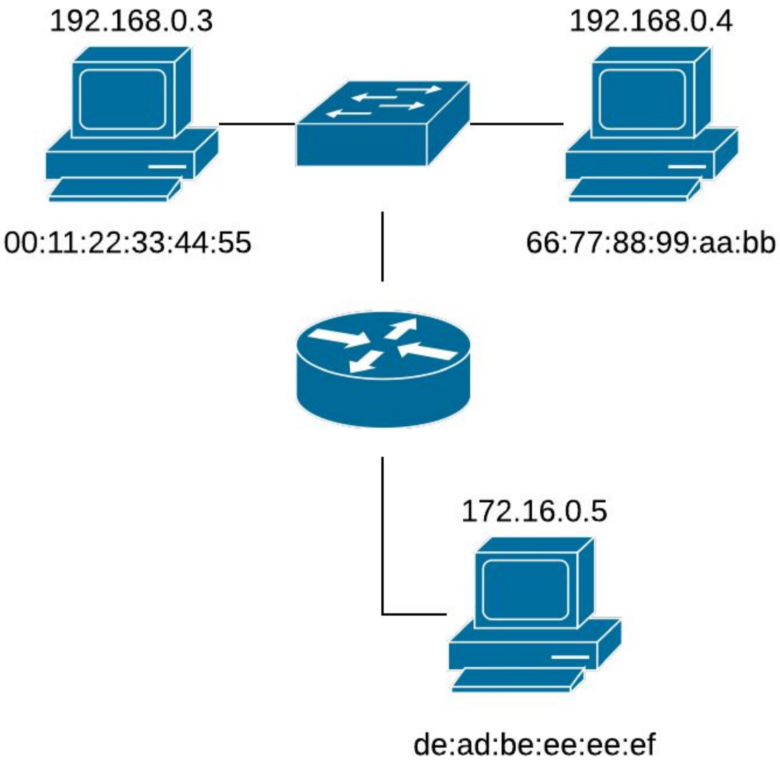
Layer 2

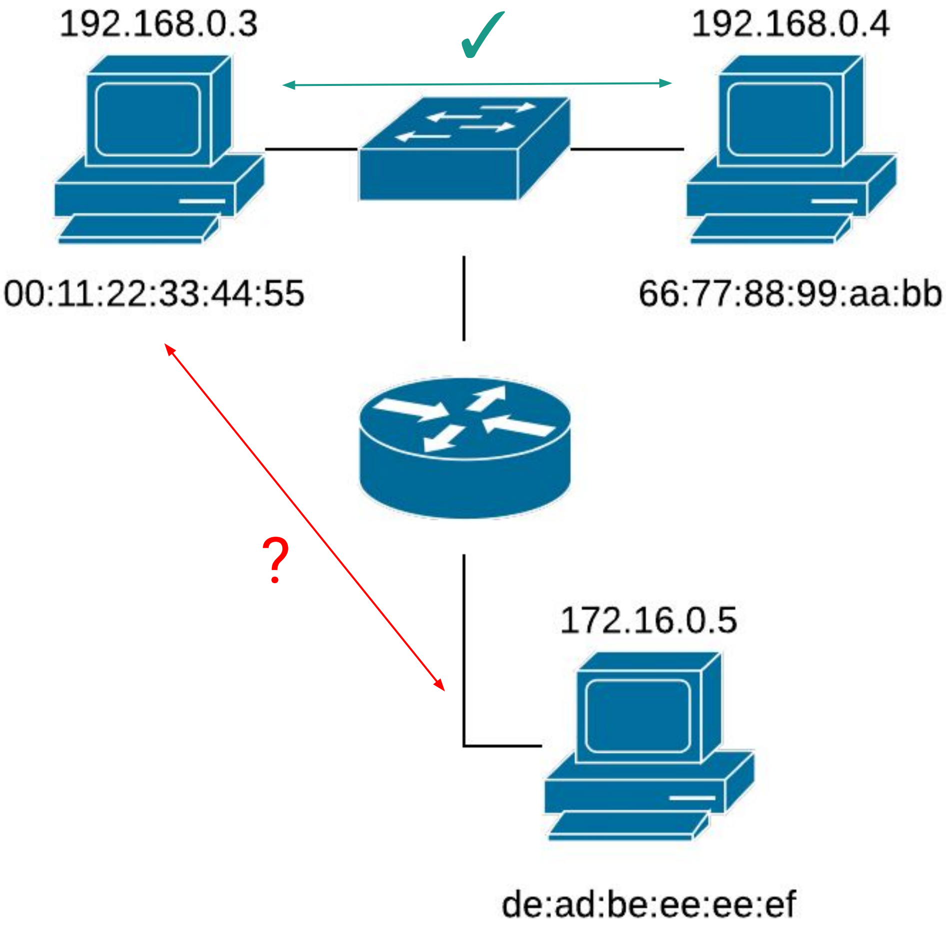
- 00:11:22:33:44:55 can communicate with 66:77:88:99:aa:bb
- But not with de:ad:be:ee:ee:ef
- Either way, we don't think about MAC addresses in code - we think about IPs
- How are IPs resolved to MAC addresses?



ARP

- Address Resolution Protocol
- 00: wants to talk with 66:
- 00: sends an ARP packet asking “who has 192.168.0.4” to the magic ff:ff:ff:ff:ff:ff MAC address
- The switch sends the ARP packet out all ports
- 66: sees the packet, knows it is 192.168.0.4, responds “I am 192.168.0.4”
- 00: receives the response, now knows what MAC address belongs to 192.168.0.4




192.168.0.3
✓
192.168.0.4
00:11:22:33:44:55
66:77:88:99:aa:bb
?
172.16.0.5
de:ad:be:ee:ee:ef

Layer 3

- How does a packet get directed between different networks?
- Every device has a **routing table**
- For each subnet, where should the device send the packet
- May go out a link directly
- May be **forwarded** to a router
- May be dropped
    - DoS prevention

Routing

Destination	Gateway	Netif
default	10.23.0.1	en0
10.23/21	0.0.0.0	en0
10.23.0.1/32	0.0.0.0	en0
127	127.0.0.1	lo0
127.0.0.1	127.0.0.1	lo0
169.254	0.0.0.0	en0
224.0.0/4	0.0.0.0	en0
255.255.255.255/32	0.0.0.0	en0

Routing

Destination	Gateway	Netif
default	10.23.0.1	en0
10.23/21	0.0.0.0	en0
10.23.0.1/32	0.0.0.0	en0
127	127.0.0.1	lo0
127.0.0.1	127.0.0.1	lo0
169.254	0.0.0.0	en0
224.0.0/4	0.0.0.0	en0
255.255.255.255/32	0.0.0.0	en0

Default Route

- Sometimes written in the CIDR notation: 0.0.0.0/0
    - All IPs
- “Catch-all” route
- Usually points to the **default gateway**

Routing

Destination	Gateway	Netif
default	10.23.0.1	en0
10.23/21	0.0.0.0	en0
10.23.0.1/32	0.0.0.0	en0
127	127.0.0.1	lo0
127.0.0.1	127.0.0.1	lo0
169.254	0.0.0.0	en0
224.0.0/4	0.0.0.0	en0
255.255.255.255/32	0.0.0.0	en0

Most Specific Prefix

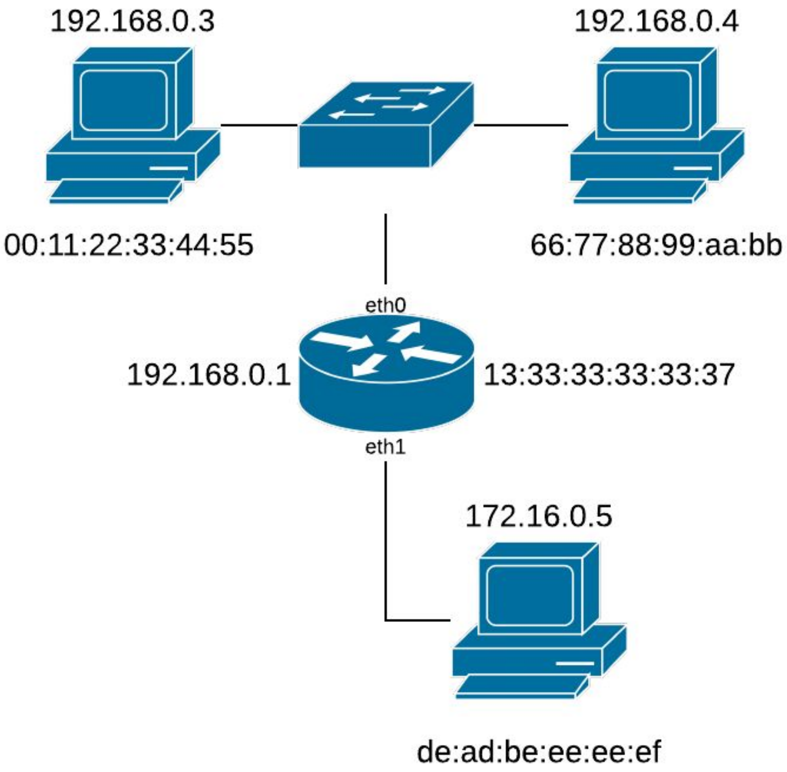
- What happens when subnets overlap?
- The **most specific** route is chosen
  - The one with the fewest number of 0’s in the mask
  - Or the largest number in CIDR form

Routing

- What does it mean to **route**?
- Send the layer 3 (IP) packet to the gateway's MAC address
- Router will receive the packet, figure out where to send it
  - Presumably a more complex routing table
- May route/forward again, or directly deliver

Layer 3

- .3 wants to talk with .5
- .3 determines .5 is on a different subnet
- .3 looks up the route to get to .5 in its routing table
- Finds the default gateway .1
- .3 sends encapsulates the layer 3 IP packet **addressed to .5** in a layer 2 frame going to the MAC address of the gateway (13:)

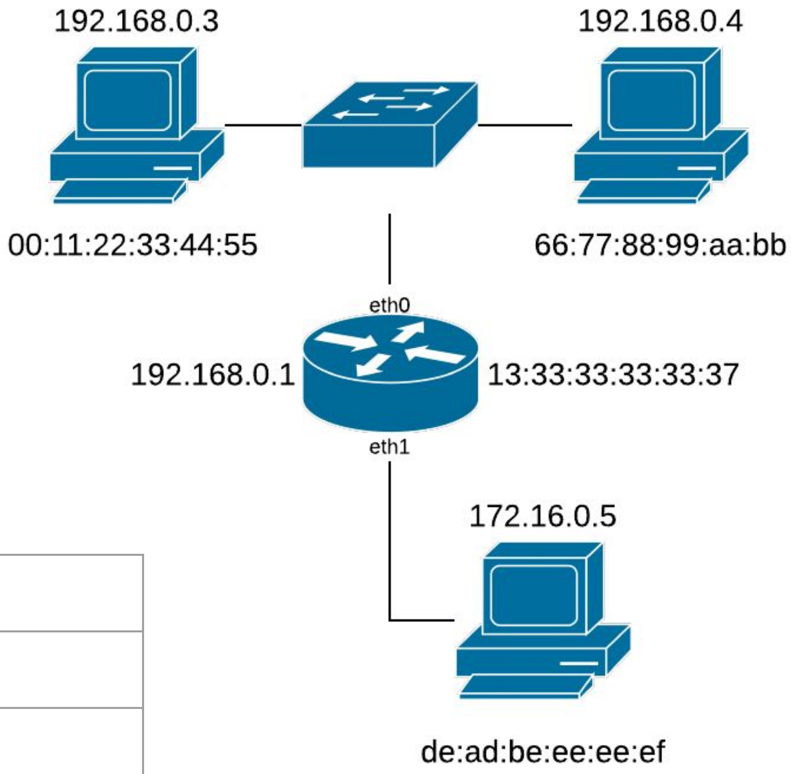


Layer 3

- .1 receives the packet, sees it's going to .5
- Does a routing table lookup, finds .5 is on a subnet attached to it

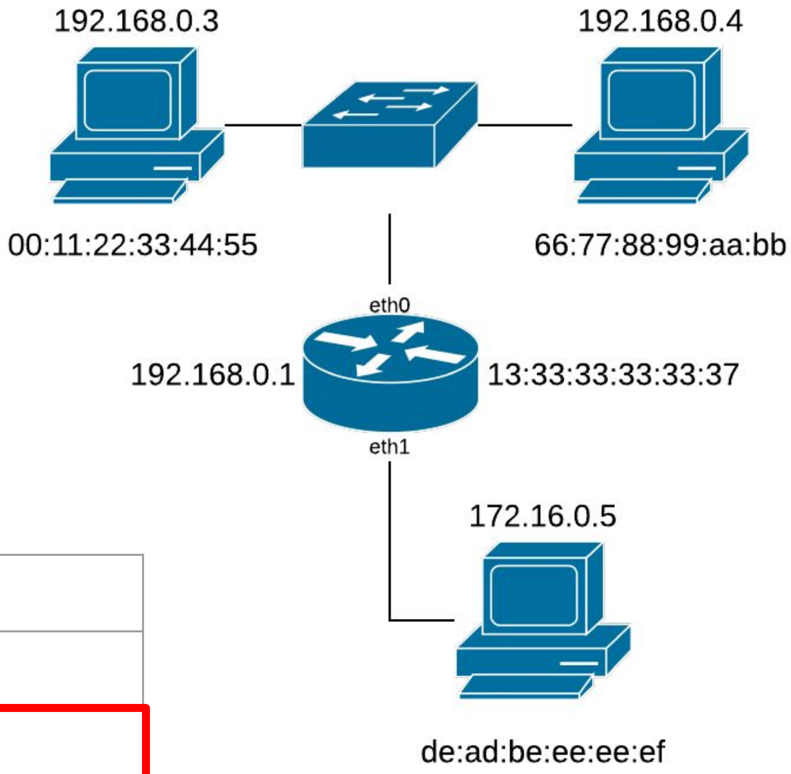
.1's Routing Table

Destination	Gateway	Netif
192.168.0.0/24	0.0.0.0	eth0
172.16.0.0/24	0.0.0.0	eth1



Layer 3

- .1 receives the packet, sees it's going to .5
- Does a routing table lookup, finds .5 is on a subnet attached to it

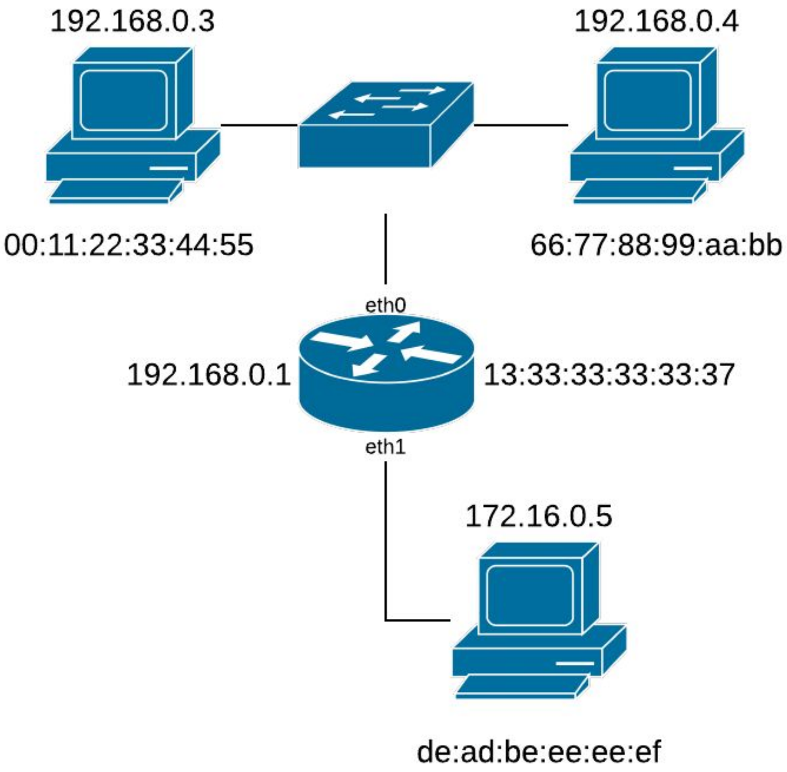


.1's Routing Table

Destination	Gateway	Netif
192.168.0.0/24	0.0.0.0	eth0
172.16.0.0/24	0.0.0.0	eth1

Layer 3

- Puts the **unmodified** IP packet in a layer 2 frame with a destination MAC of de: and sends it out eth1
- .5 receives the packet



Questions?

Briefly:
How IPs Are Assigned

IP Assignment

- Can be statically set or dynamically assigned (DHCP)
- Static use cases
    - Static/unmoving network hardware
    - Critical hardware which DHCP relies on
- Dynamic use cases
    - Everything else

DHCP

- Dynamic Host Configuration Protocol
- Clients send a “discovery” message and are offered an IP
- Can also set routes, DNS servers, search domains, etc.

Advanced Routing

Route Distribution

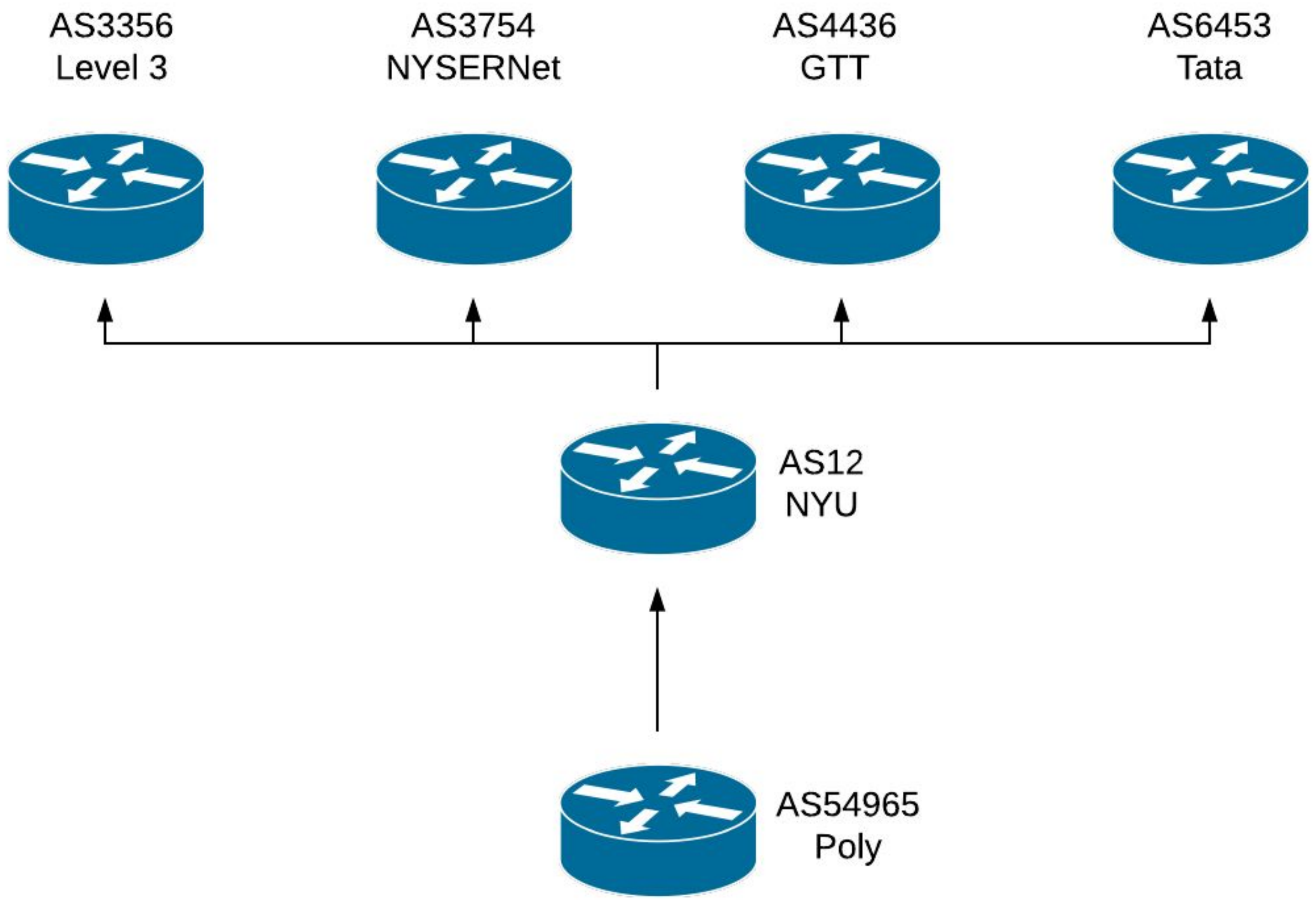
- Routing rules can be statically set or dynamically added
- Many ways to distribute routes
  - DHCP
  - Open Shortest Path First (OSPF)
  - Enhanced Interior Gateway Routing Protocol (EIGRP)
  - Border Gateway Protocol (BGP)
  - *Tons* more

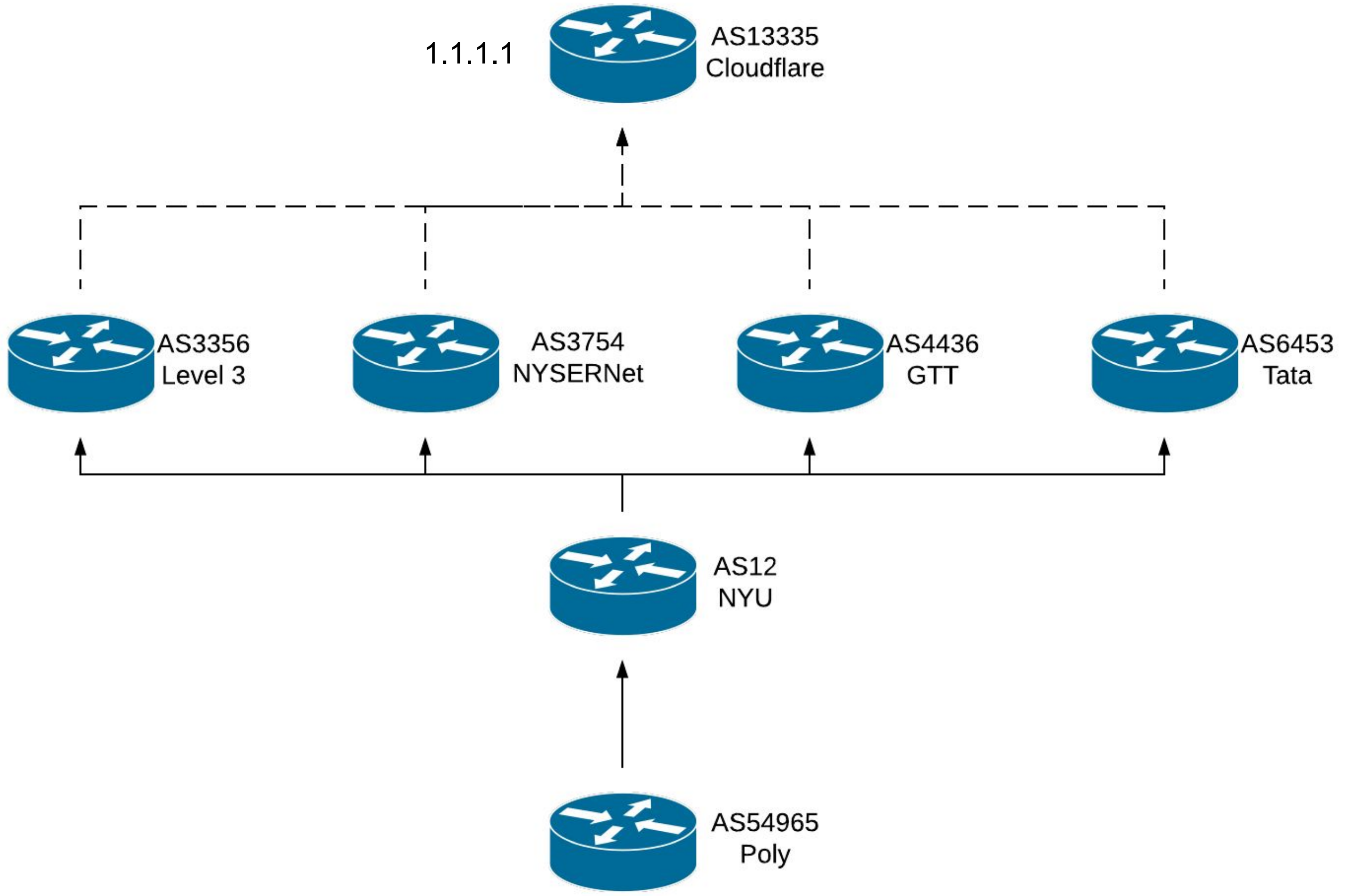
Route Distribution

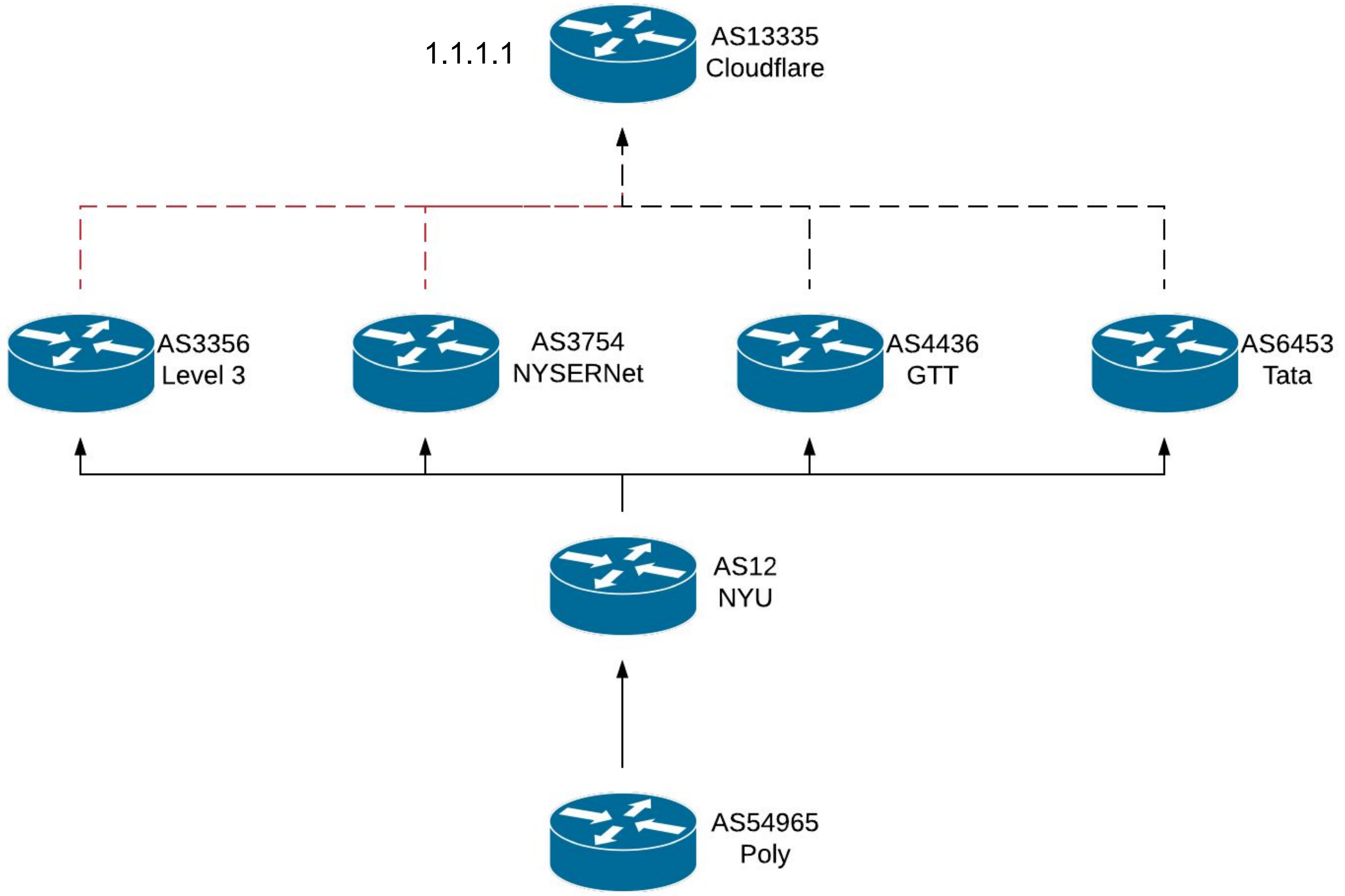
- Why distribute routes automatically?
- The public internet has >700,000 routes
- NetEng's aren't paid enough to add all of those by hand
- Protocols also deal with
  - Path optimization (shortest path, cheapest, etc.)
  - Reconvergence after hardware failure (assuming redundancy in the network)

BGP

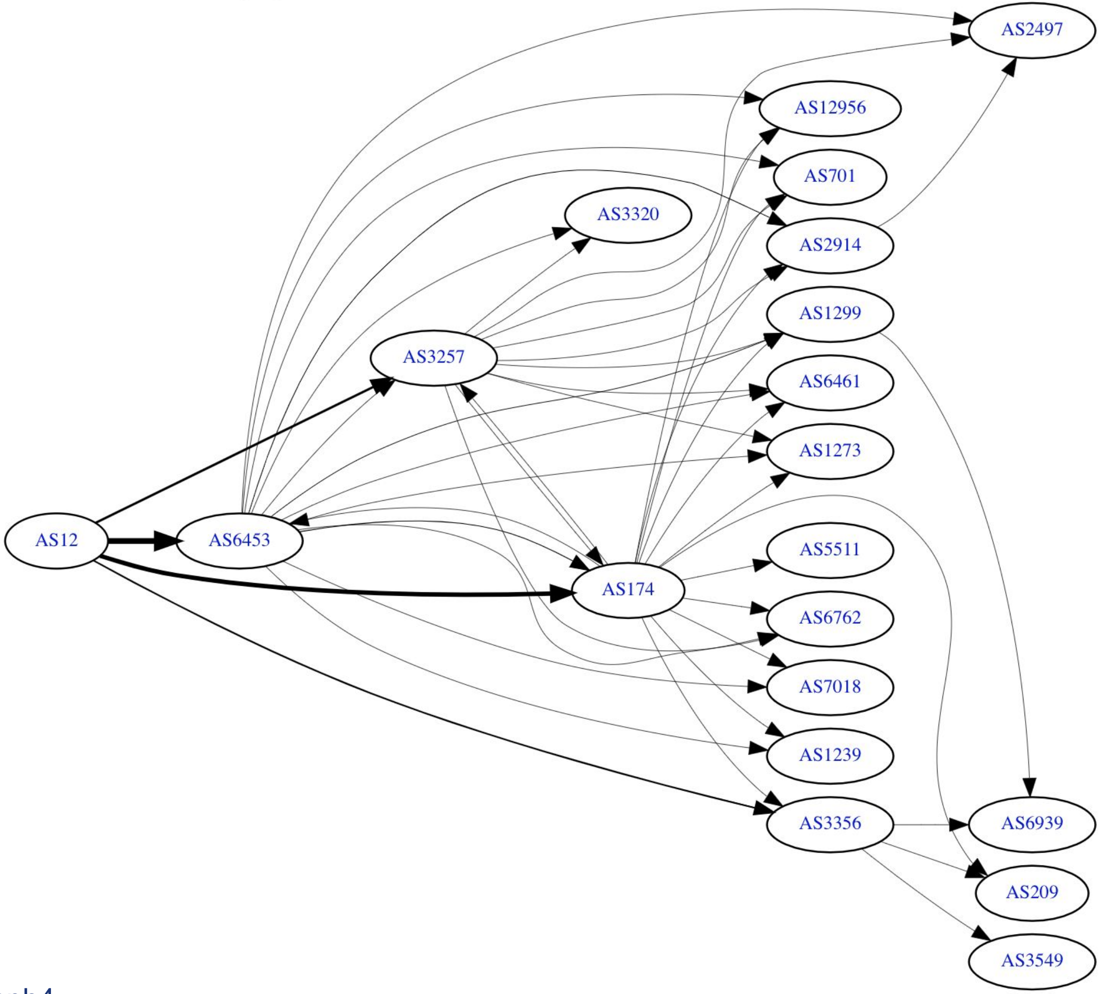
- Border Gateway Protocol
- Logical organizations are assigned Autonomous System Numbers (ASNs)
    - NYU has an ASN
    - Amazon has multiple ASNs
    - Google has even more ASNs
- Each AS **peers** with other ASes
    - Directly (either as a paid upstream like an ISP or cost free for large ASes)
    - As part of an Internet Exchange Point (IX/IXP)
- The peers advertise routes for IP space that they can get to
    - NYU advertises IPs that they own
    - NYU’s ISPs advertise the rest of the internet to NYU


AS3356
Level 3
AS3754
NYSERNet
AS4436
GTT
AS6453
Tata
AS12
NYU
AS54965
Poly

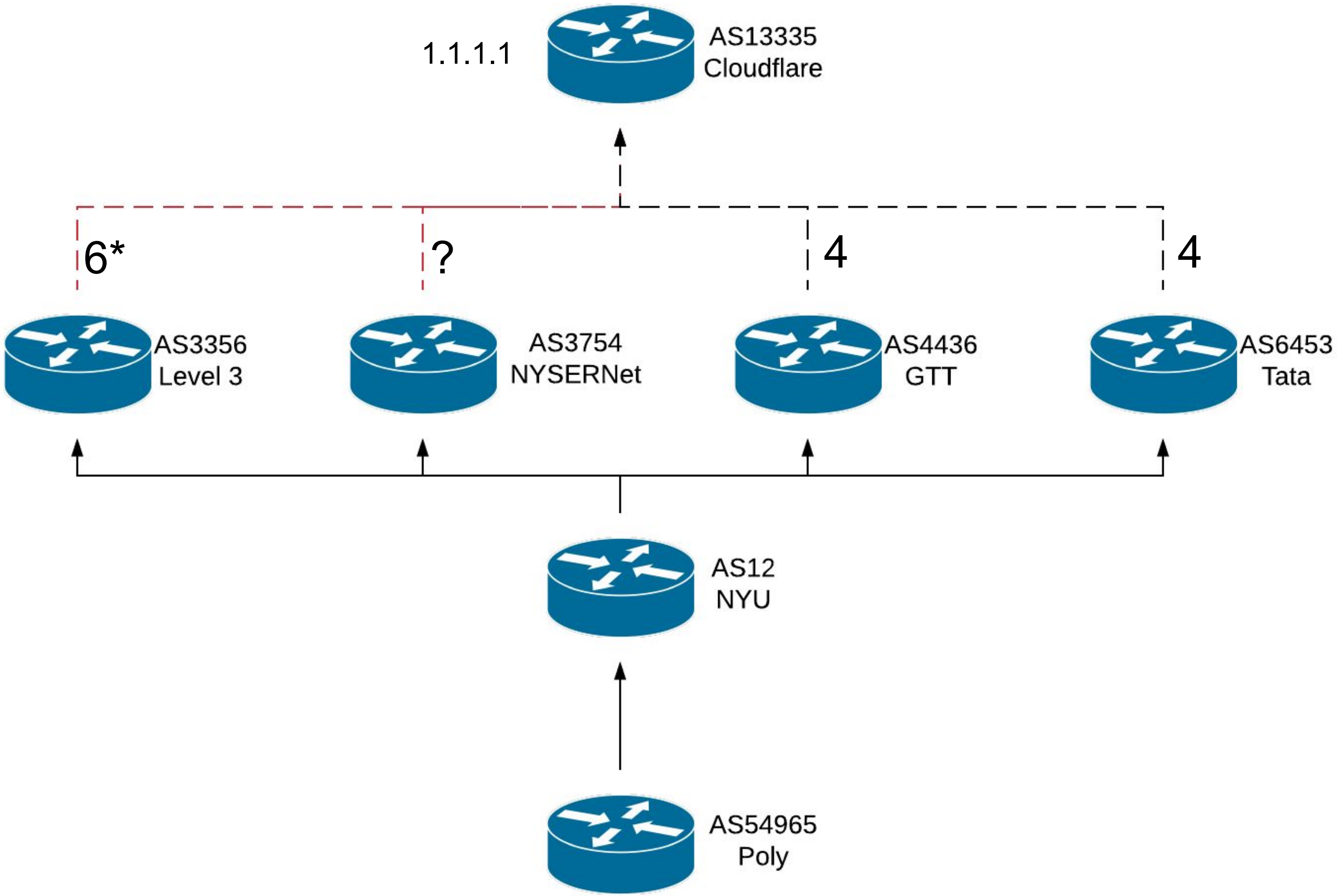

1.1.1.1
AS13335
Cloudflare
AS3356
Level 3
AS3754
NYSERNet
AS4436
GTT
AS6453
Tata
AS12
NYU
AS54965
Poly


1.1.1.1
AS13335
Cloudflare
AS3356
Level 3
AS3754
NYSERNet
AS4436
GTT
AS6453
Tata
AS12
NYU
AS54965
Poly

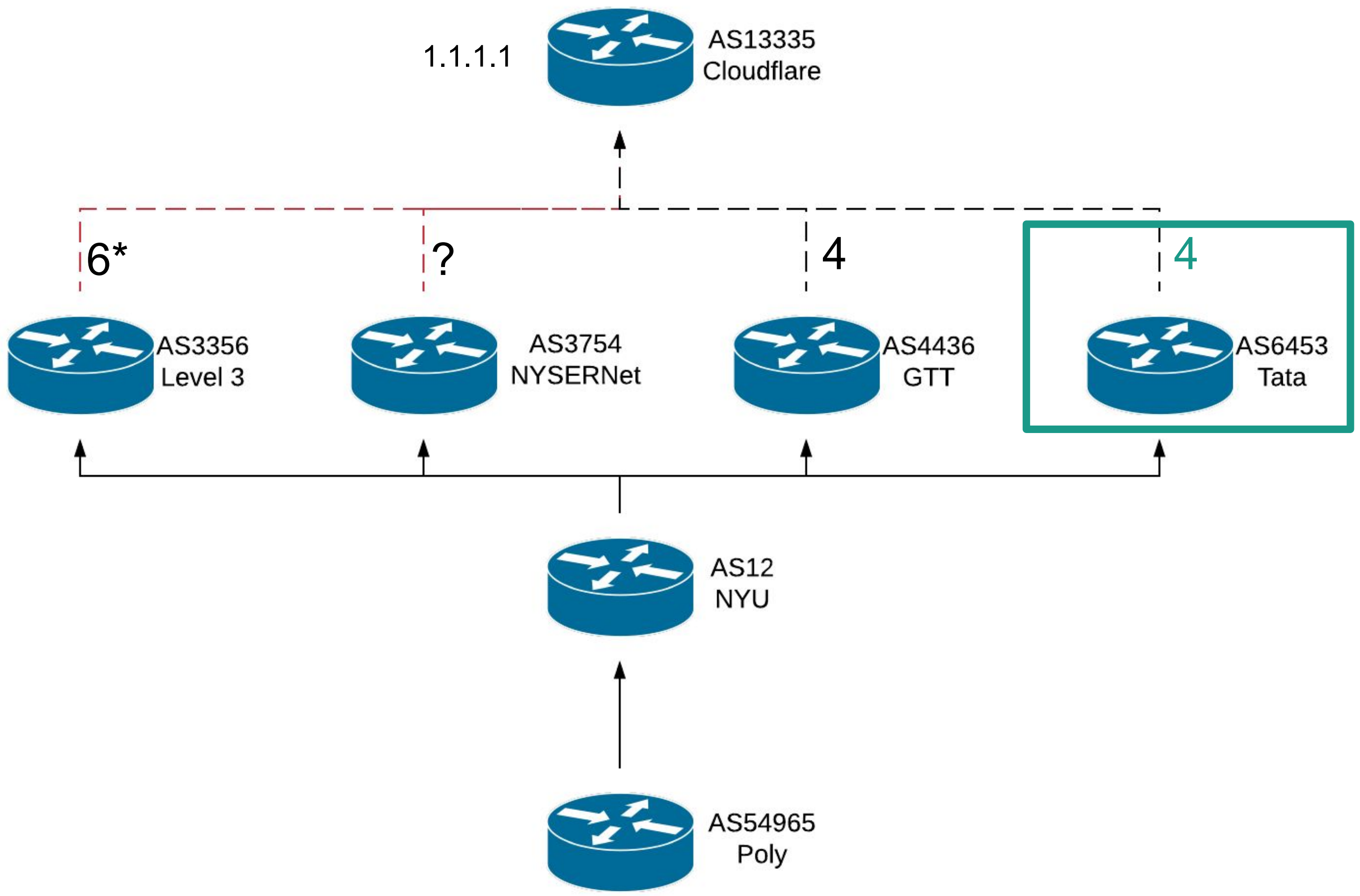
AS12 IPv4 Route Propagation

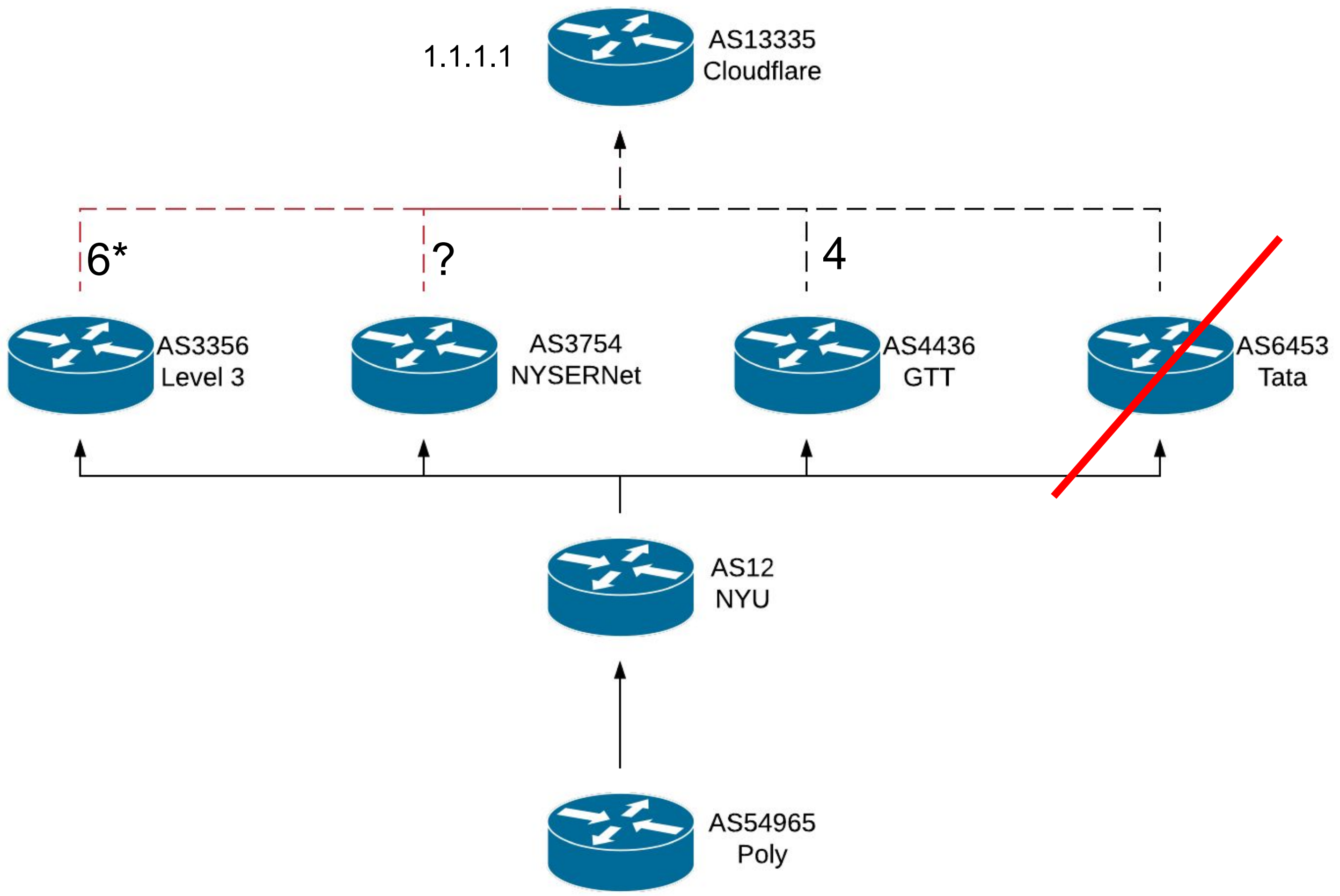


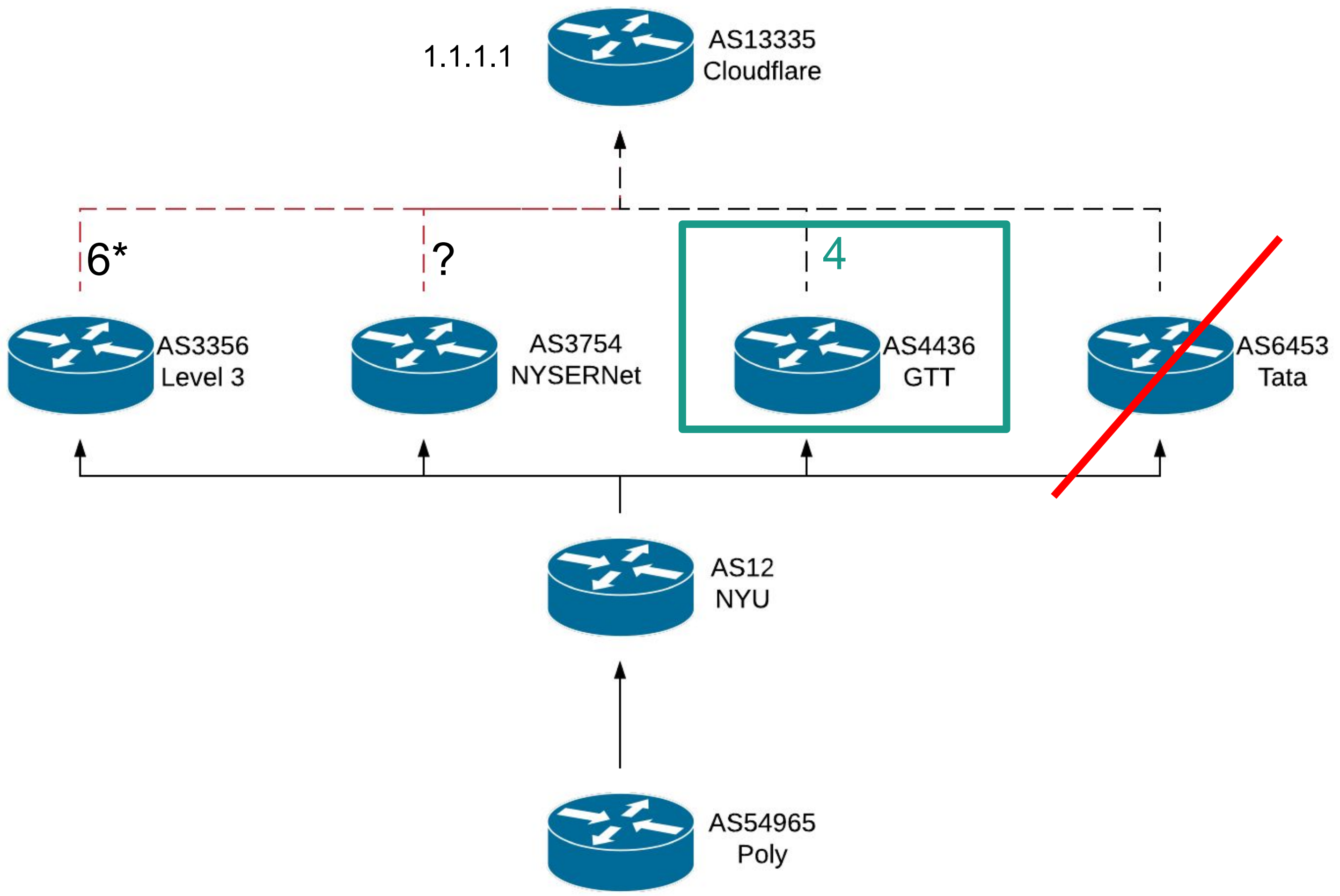
https://bgp.he.net/AS12#_graph4



* Level 3 connects to Cloudflare through Tata


1.1.1.1
AS13335
Cloudflare
6*
?
4
4
AS3356
Level 3
AS3754
NYSERNet
AS4436
GTT
AS6453
Tata
AS12
NYU
AS54965
Poly


1.1.1.1
AS13335
Cloudflare
6*
?
4
AS3356
Level 3
AS3754
NYSERNet
AS4436
GTT
AS6453
Tata
AS12
NYU
AS54965
Poly


1.1.1.1
AS13335
Cloudflare
6*
?
4
AS3356
Level 3
AS3754
NYSERNet
AS4436
GTT
AS6453
Tata
AS12
NYU
AS54965
Poly

Weird stuff can still happen

```
 1. 128.238.66.1
 2. 128.238.60.2
 3. ???
 4. 216.165.101.1
 5. dmzgwa-ptp-nyugwa-vl3081.net.nyu.edu
 6. EXTGWD-PTP-DMZGWA.NET.NYU.EDU
 7. via-111-8th-st-nyc001jp01.yipes.com
 8. xe-2-0-0.cr8-nyc3.ip4.gtt.net
    xe-3-0-0.cr8-nyc3.ip4.gtt.net
 9. as6453.nyc30.ip4.gtt.net
10. if-ae-8-2.tcore1.nto-new-york.as6453.net
11. if-ae-0-2.tcore1.sqn-san-jose.as6453.net
12. if-ae-18-2.tcore2.sv1-santa-clara.as6453.net
```

Useful Utilities

- mtr
- ISP looking glasses
- https://bgp.he.net

BGP Security

- BGP is basically unsecured
- Anybody who can control peering routers can dictate where IP traffic flows
- Recall: most specific routes take precedence
- An attacker with access to a well-connected router could advertise very specific /24s for all of Google (for example)
- All traffic destined to Google that determined the attacker’s path to be optimal would go to the attacker
- Lets them do basically anything
- Get SSL certs, control DNS, etc. etc.

BGP Security

- Mitigations:
  - Route filtering (limiting the routes that will be accepted to what's expected)
  - (Soon) RPKI (RFC6480)

Distributing Traffic

- Many (large) sites have Points of Presence (PoPs) in many cities (metros)
- These sites need a way to route traffic to the nearest PoP
- Two ways to accomplish
  - DNS
  - Anycast
- DNS only works with hostnames
- Can we distribute 1.1.1.1, 8.8.8.8, 9.9.9.9, etc.?

Anycast

- Many PoPs can advertise the same IP(s)
- Shortest path selection naturally routes people to the closest PoP
    - For some definition of close

Questions?

Additional Stuff

Limited IPs

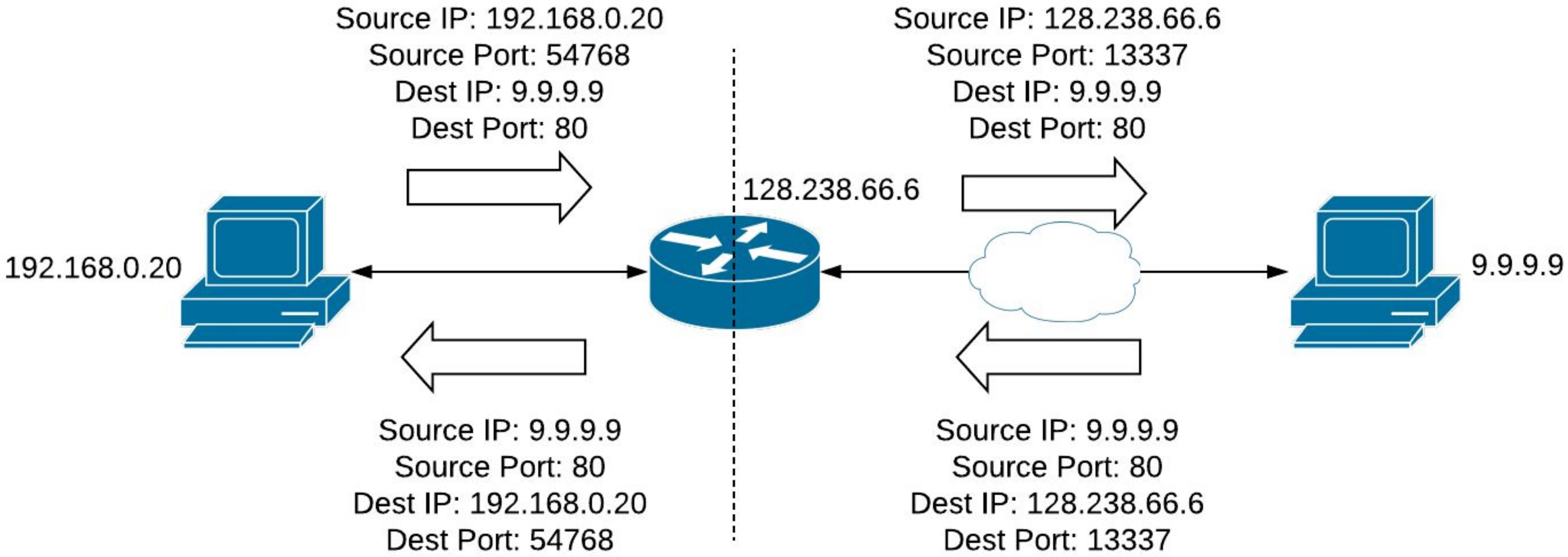
- Recall: IPv4 addresses are 32 bits
- Only 4.2 billion IPv4 addresses
  - Minus 224.0.0.0/4
  - Minus 240.0.0.0/4
  - Minus 0/8
  - Minus 10/8
  - ...
- *Many* more than 4.2 billion networked devices
- How can we give them all access?

NAT

- Network Address Translation
- Two types: Source NAT (SNAT), Destination NAT (DNAT)
    - NAT commonly refers to DNAT
- At a (sub)net edge, translate/rewrite IPs
- Simple example
    - Rewrite traffic destined to 8.8.8.8 to go to 9.9.9.9
    - The opposite translation (9.9.9.9 => 8.8.8.8) happens when the received traffic goes back the other way

NAT

- More commonly used to translate connections from internal IPs (10/8, 172.16/12, 192.168/16) to IPs on the public internet
- Rewrites traffic to come from the NAT device itself
- Maps the 4-tuple (int_src_ip, int_src_port, dst_ip, dst_port) to (nat_src_ip, nat_src_port, dst_ip, dst_port)


Source IP: 192.168.0.20
Source Port: 54768
Dest IP: 9.9.9.9
Dest Port: 80
Source IP: 128.238.66.6
Source Port: 13337
Dest IP: 9.9.9.9
Dest Port: 80
128.238.66.6
192.168.0.20
9.9.9.9
Source IP: 9.9.9.9
Source Port: 80
Dest IP: 192.168.0.20
Dest Port: 54768
Source IP: 9.9.9.9
Source Port: 80
Dest IP: 128.238.66.6
Dest Port: 13337

VPNs

- Virtual Private Networks
- Two main types:
  - Site-to-site: connect entire networks over a secure tunnel
  - Remote access: connect your computer to a remote network securely
- Basic idea: encapsulate and encrypt raw packets in another packet to the VPN server
- The VPN server receives the encapsulated packet, decrypts it, and drops it on the network

Questions?

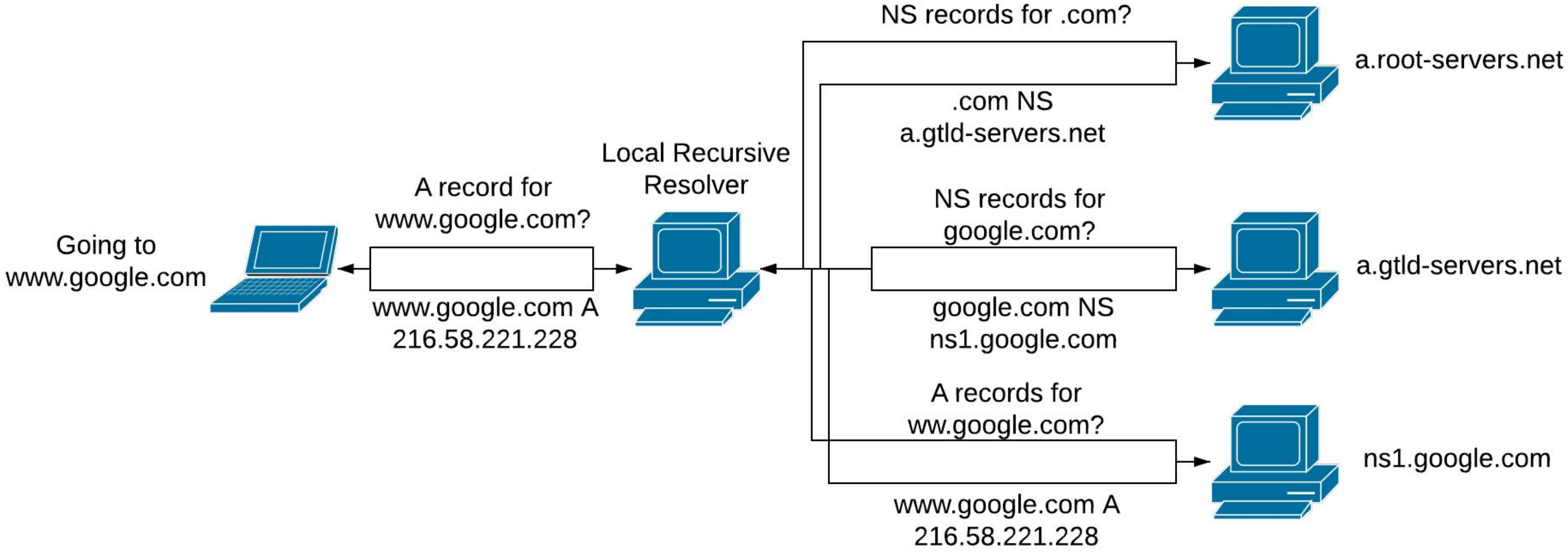
DNS

DNS

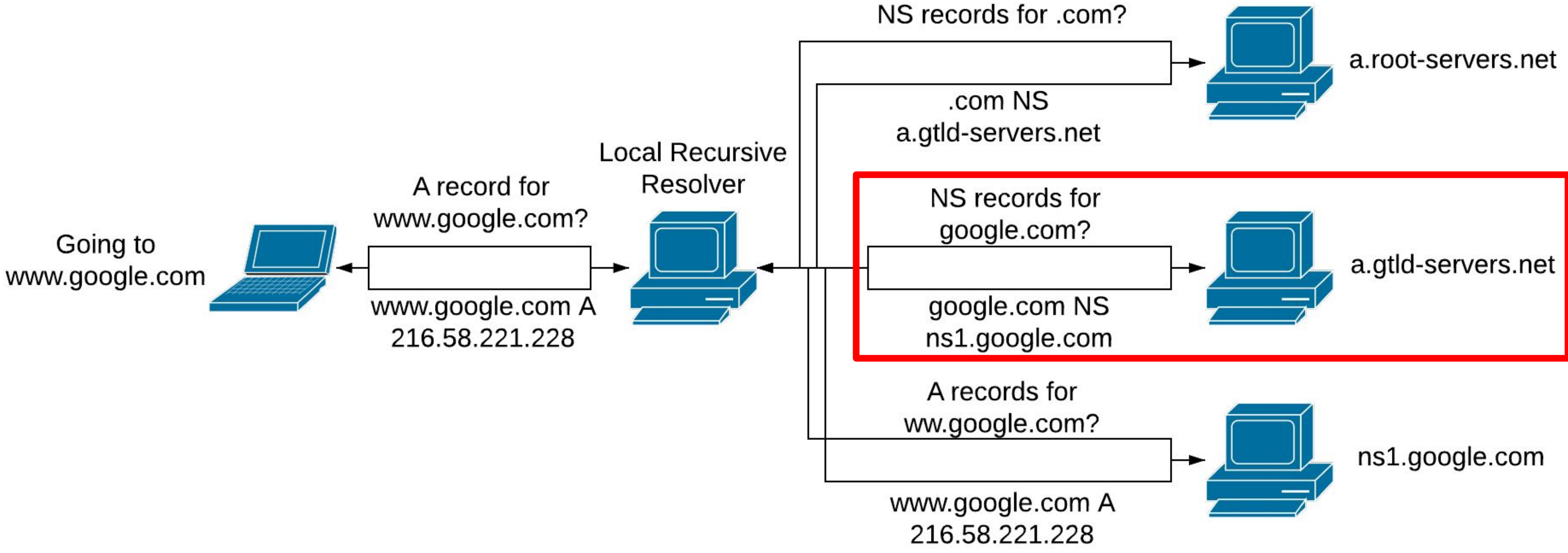
- Domain Name Server
- Transforms domain names into IPs (google.com -> 172.217.12.206)
  - Or other data…
- Many record types: A, AAAA, CNAME, MX, PTR, TXT, etc.
- Hierarchical
  - Root servers know where TLD nameservers are
  - TLD nameservers know where domain nameservers are
  - Domain nameservers can resolve records (or delegate further)

DNS

- Two main types of servers: recursive, and authoritative
- Recursive servers
  - “Normal” servers you’re familiar with
  - On home networks
  - 1.1.1.1, 8.8.8.8, 9.9.9.9, etc.
- Authoritative servers
  - Holds the actual information for a domain



Going to
www.google.com
A record for
www.google.com?
www.google.com A
216.58.221.228
Local Recursive
Resolver
NS records for .com?
.com NS
a.gtld-servers.net
a.root-servers.net
NS records for
google.com?
google.com NS
ns1.google.com
a.gtld-servers.net
A records for
ww.google.com?
www.google.com A
216.58.221.228
ns1.google.com


Going to
www.google.com
A record for
www.google.com?
www.google.com A
216.58.221.228
Local Recursive
Resolver
NS records for .com?
.com NS
a.gtld-servers.net
a.root-servers.net
NS records for
google.com?
google.com NS
ns1.google.com
a.gtld-servers.net
A records for
ww.google.com?
www.google.com A
216.58.221.228
ns1.google.com

DNS Security

- DNS is not encrypted and not authenticated
- People on the network can observe the sites you go to by looking at DNS traffic
- DNSSEC is an extension to DNS which allows authentication
    - Chain of trust like normal SSL/TLS
- Work being done to encrypt: DoH, DNS over TLS
- SSL/TLS cert verification removes spoofing ability

DNS Security

- Other DNS issues:
  - Dangling records
  - Cache poisoning
  - Reflected DoS (UDP)

Horus

Q&A / AMA

Thanks!