Anycast DNS

ISC Webinar, October 14, 2015

Logistics

- Webinar is 1 hour long
- This session will be recorded and posted at http://www.isc.org/webinars
- Participants are muted to improve audio quality for everyone.
- We want questions! Please enter into the WebEx Q&A tab
  - The presenter may defer some questions until the end of the presentation

Presenter

Jason Lomonaco,
Sr. Network Engineer

Agenda

- Define Anycast
- Examine use cases
- Explore the impact on Internet protocols
- Explore Anycast and DNS
- Share ISC's operational experience
- Answer questions

ANYCAST

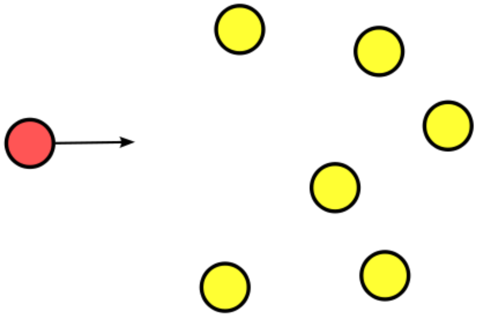
Define

What is Anycast?

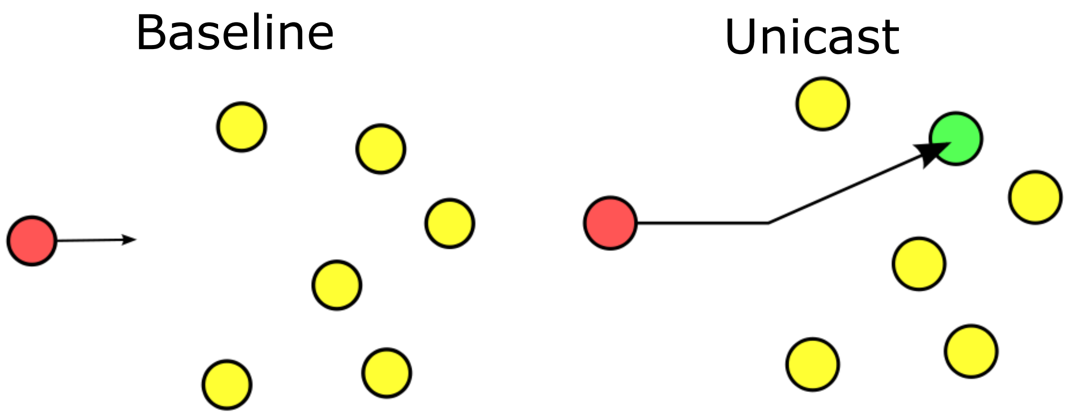
- Anycast describes a method of using the same IP address on multiple servers
- Fundamentally, Anycast is a *routing scheme*
- Anycast is more about the configuration of routers and routing than servers
  - Server admins have to understand what's going on in order to properly operate the service

Routing Schemes Compared

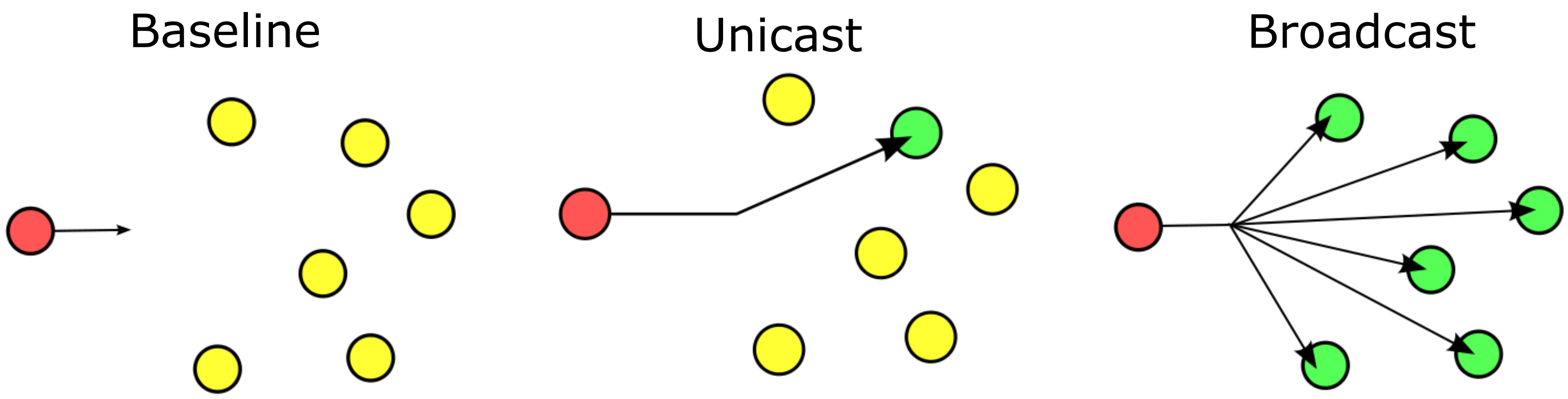
Baseline



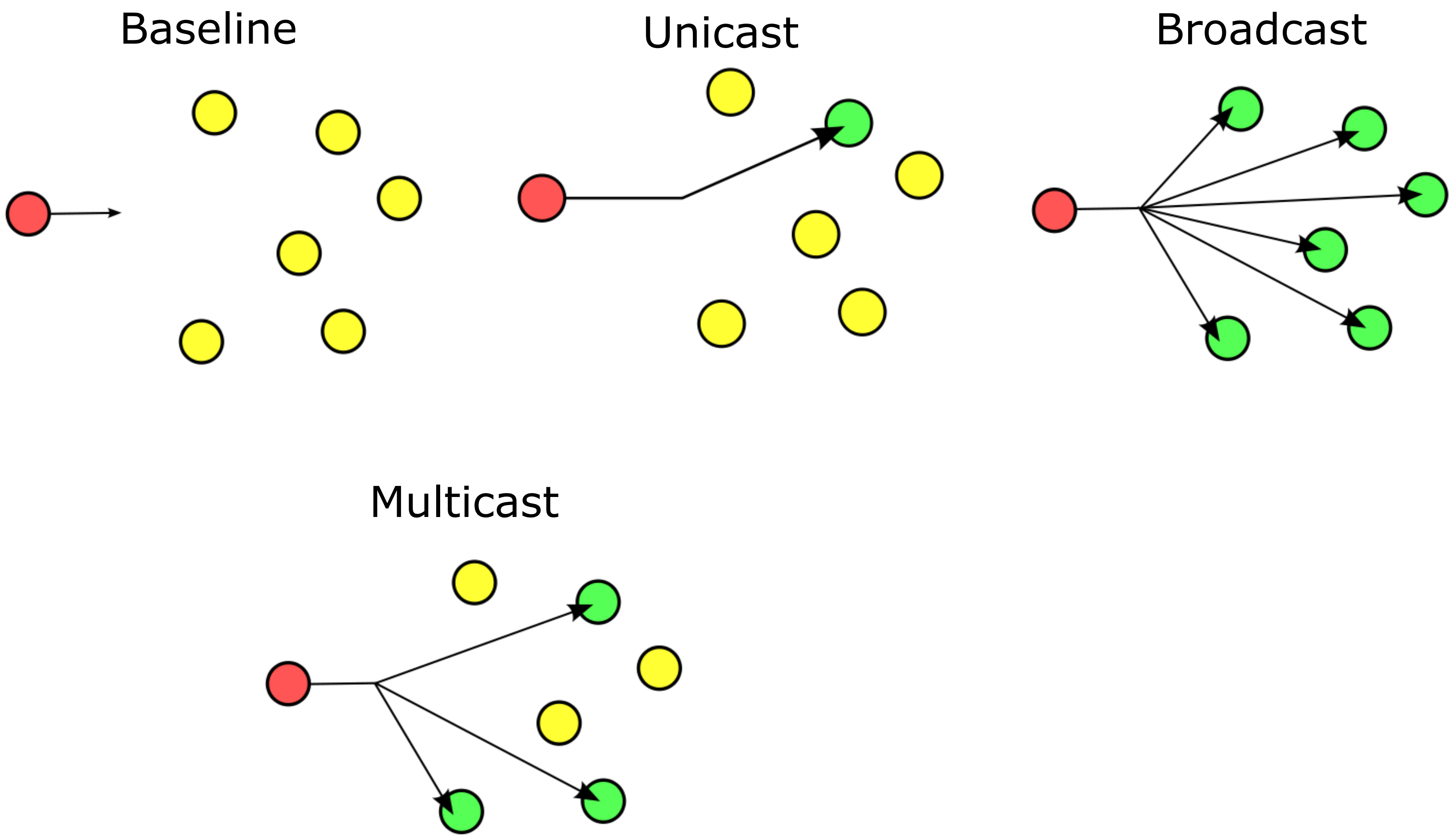
Routing Schemes Compared



Routing Schemes Compared



Routing Schemes Compared



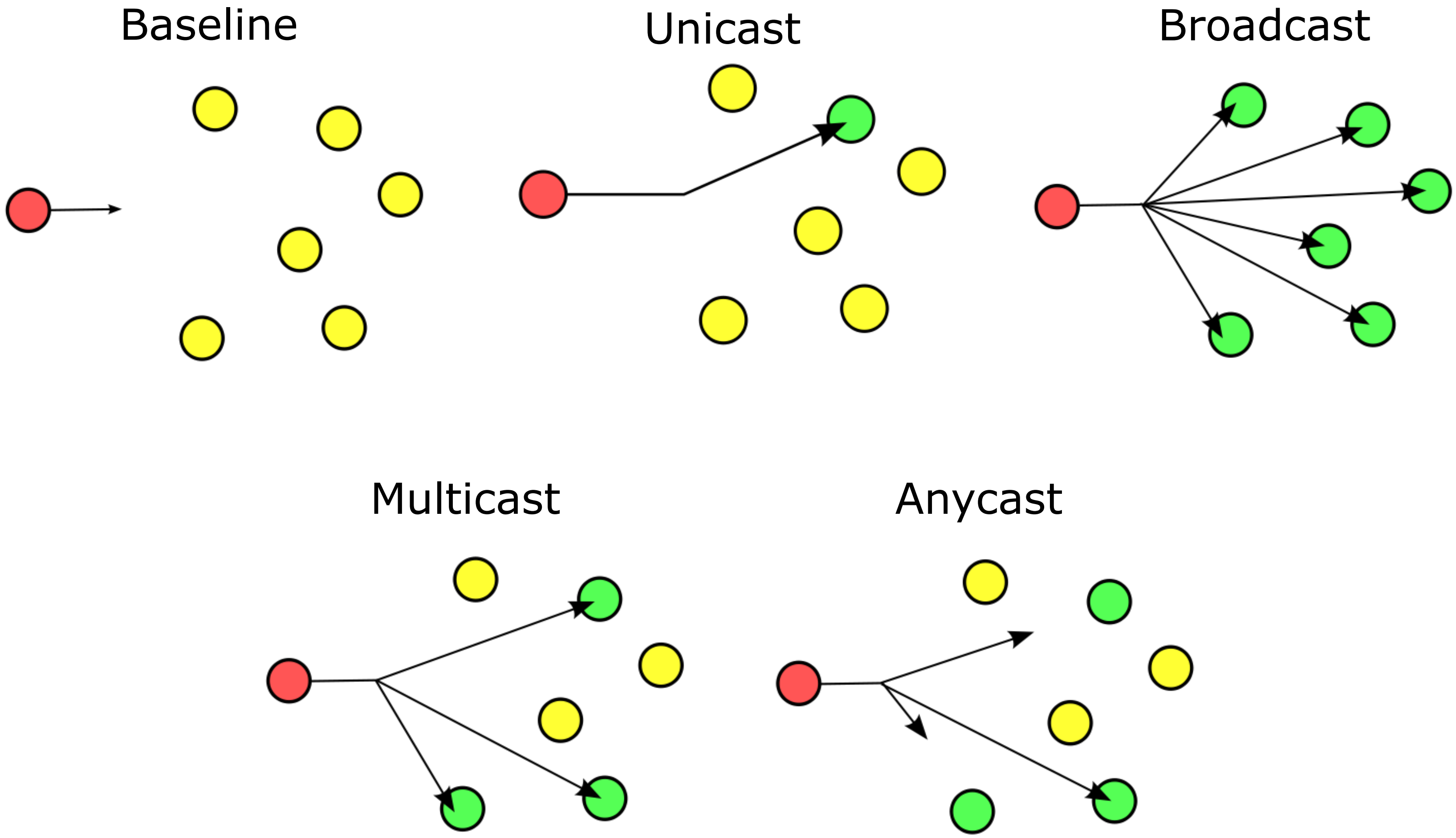
Baseline

Unicast

Broadcast

Multicast

Routing Schemes Compared



Diagrams from http://en.wikipedia.org/wiki/Anycast, and are public domain.

Properties of Anycast

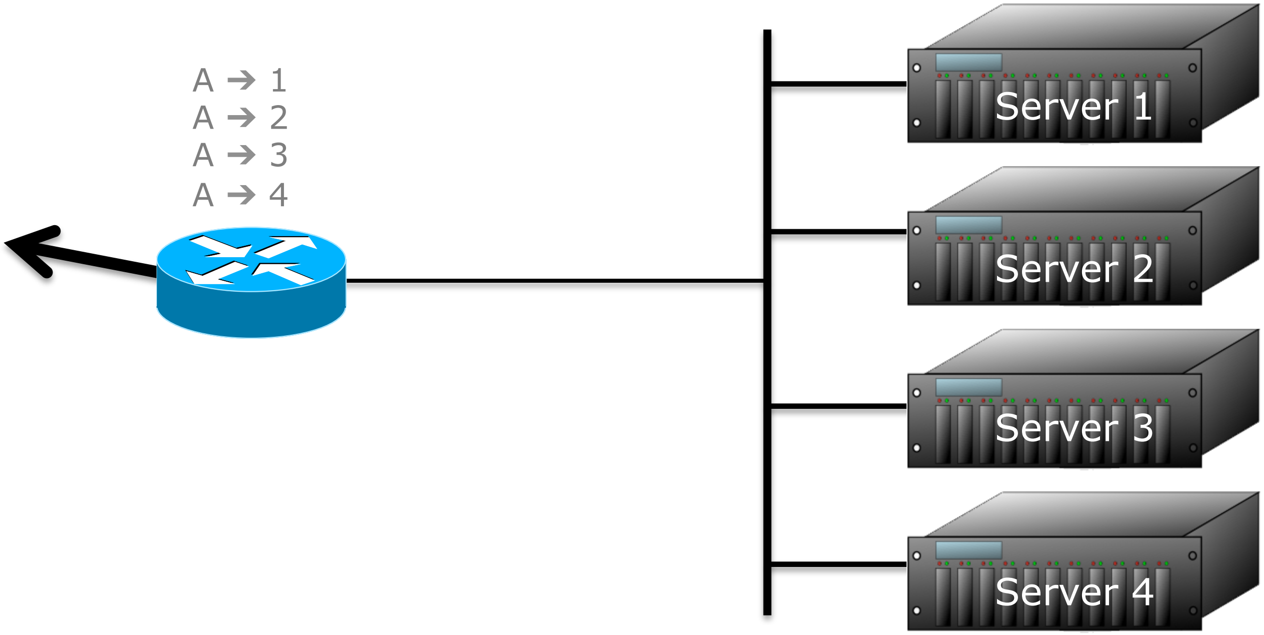
- Each packet sent to an Anycasted IP address may reach a different server
- Packets are routed to the IP address with the best ***network metric***
  - This is often the nearest server, but not always. Metrics could be set based on other factors, such as bandwidth, cost, load or reliability
- Servers with an Anycast address must also have a Unicast IP address

USE CASES

Examine

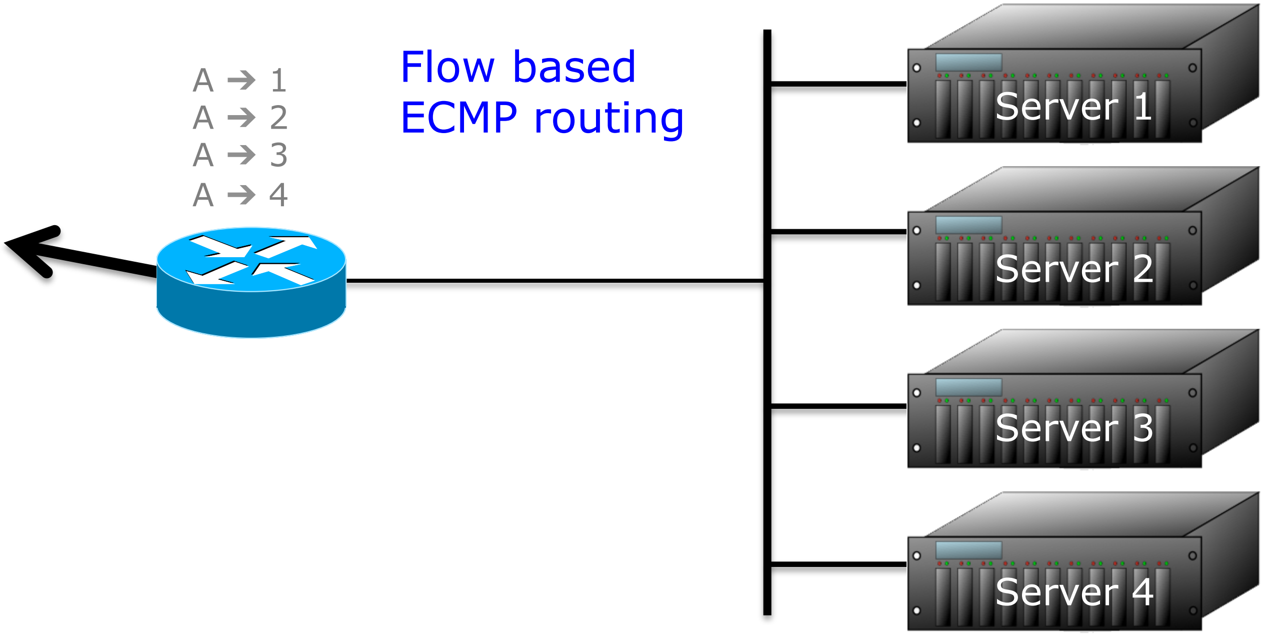
Use Cases

- Local Anycast
  - Distributes load across multiple servers on same subnet
  - Eliminates need for load balancer by making the network (router) distribute traffic



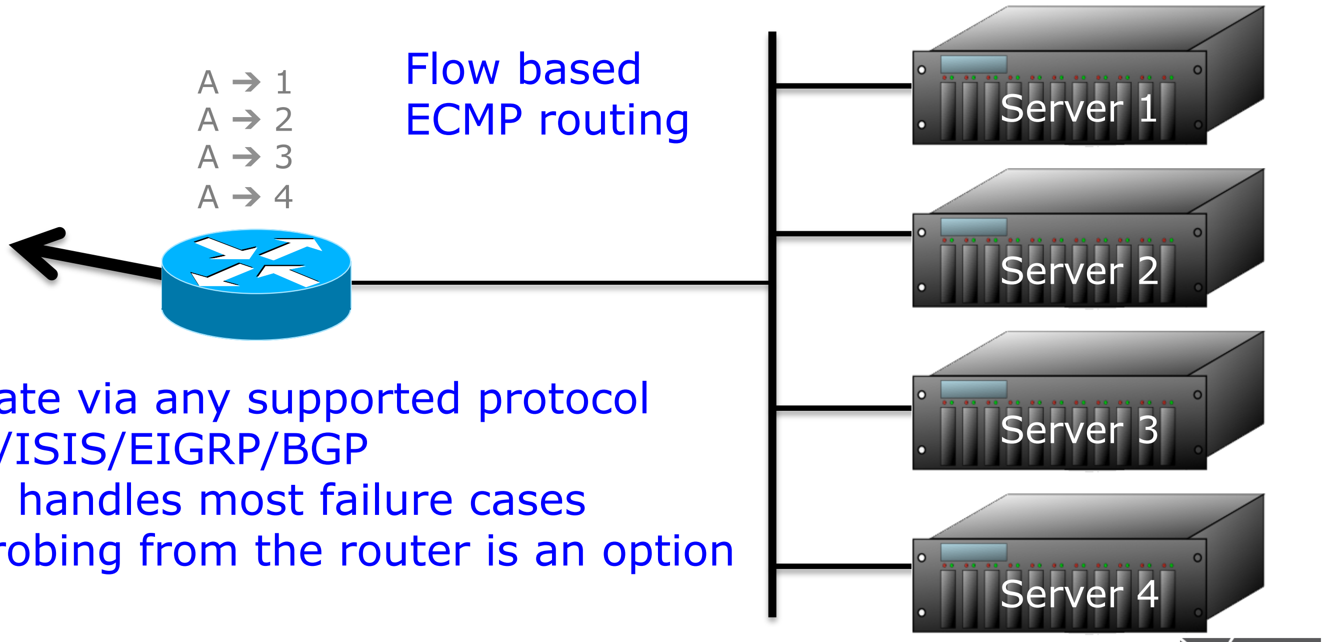
Use Cases

- Local Anycast
  - Distributes load across multiple servers on same subnet
  - Eliminates need for load balancer by making the network (router) distribute traffic



Use Cases

- Local Anycast
  - Distributes load across multiple servers on same subnet
  - Eliminates need for load balancer by making the network (router) distribute traffic

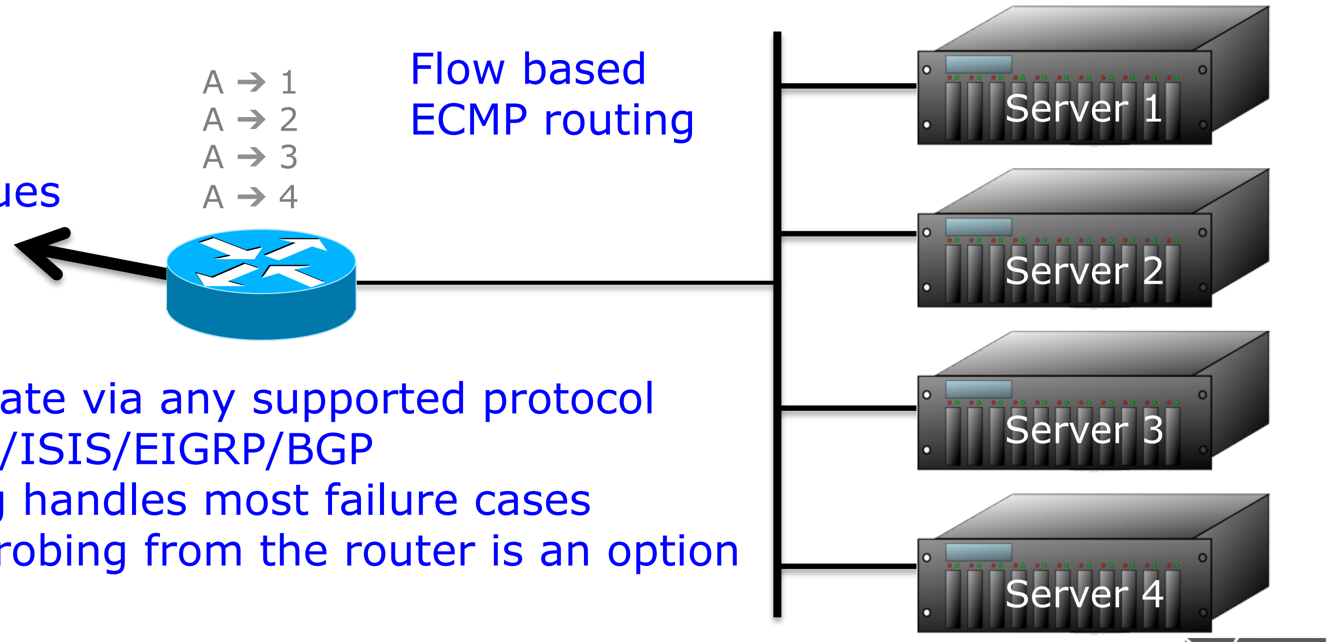


Flow based
ECMP routing

Routes may originate via any supported protocol
- static/RIP/OSPF/ISIS/EIGRP/BGP
- dynamic routing handles most failure cases
- active service probing from the router is an option

Use Cases

- Local Anycast
  - Distributes load across multiple servers on same subnet
  - Eliminates need for load balancer by making the network (router) distribute traffic



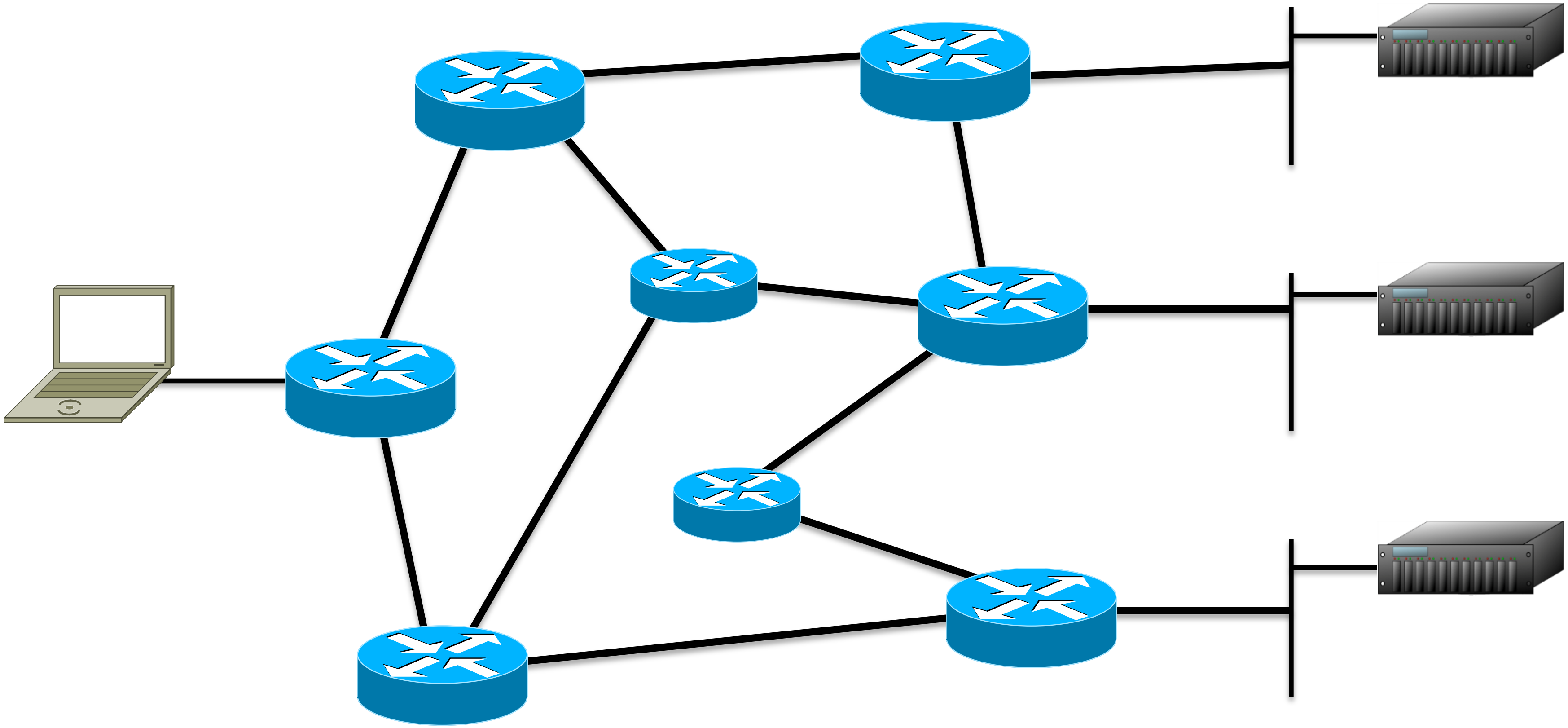
ONE ROUTE!
Reduces routing issues

Flow based
ECMP routing

Routes may originate via any supported protocol
- static/RIP/OSPF/ISIS/EIGRP/BGP
- dynamic routing handles most failure cases
- active service probing from the router is an option

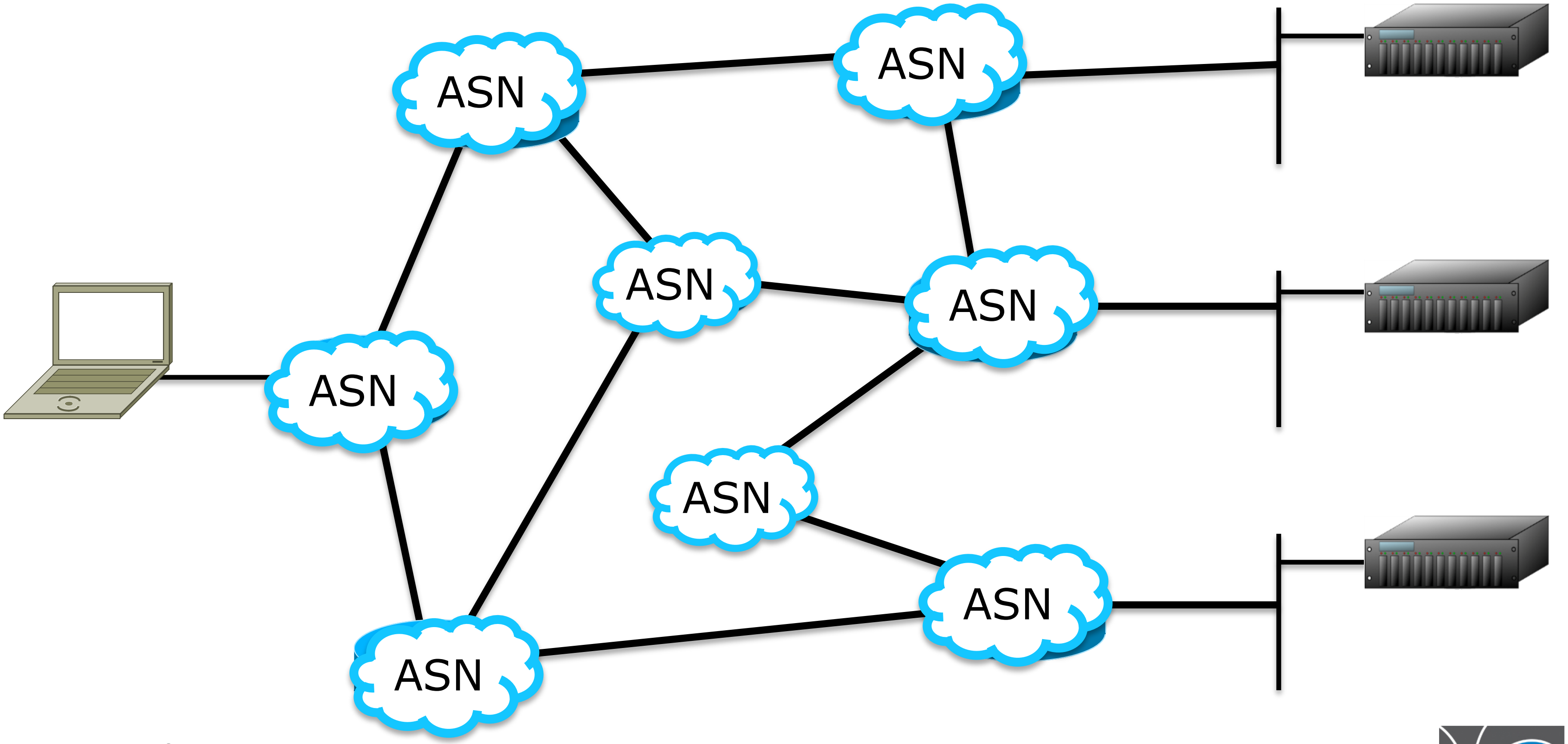
Global Anycast

- Distributes load across multiple locations
- Provides redundancy



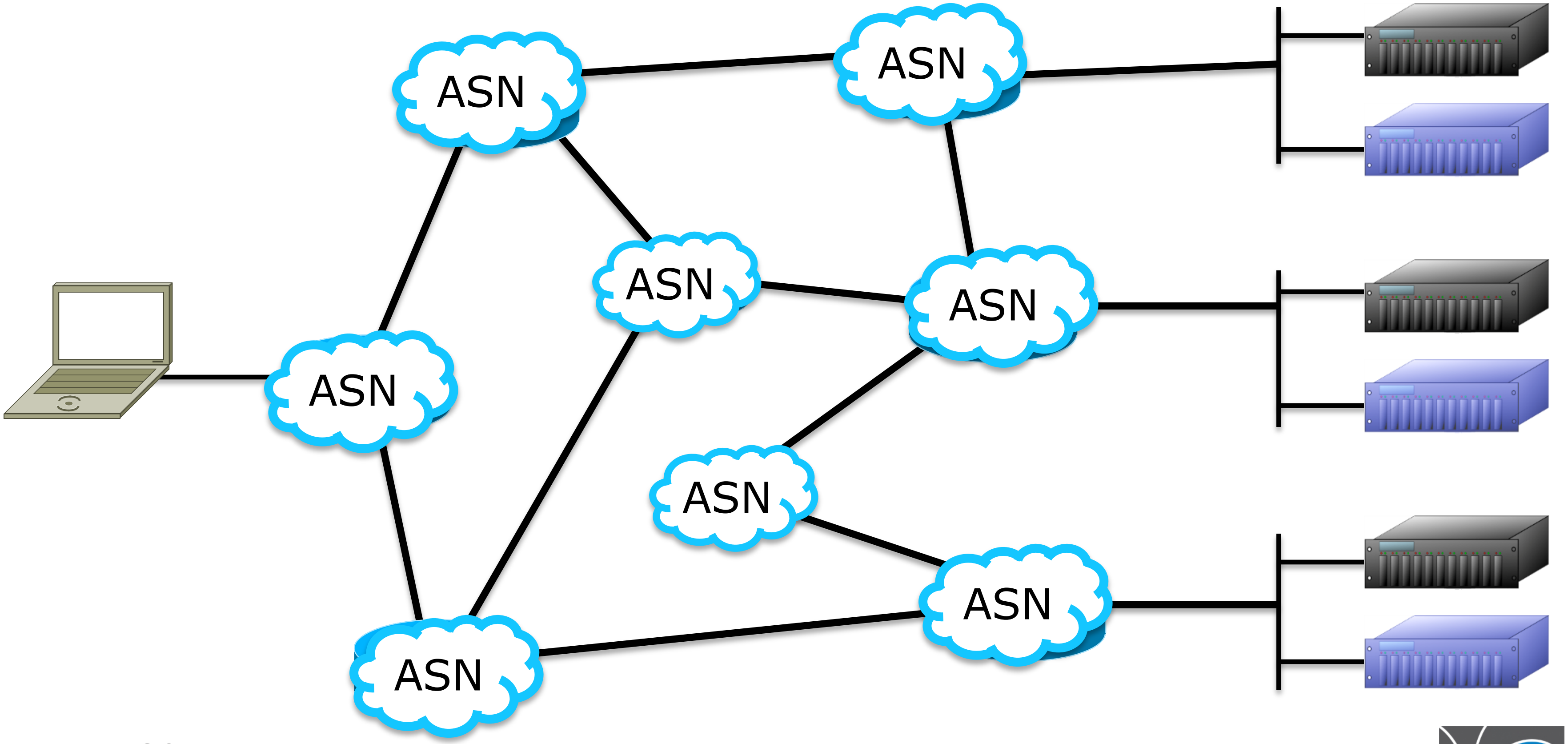
Global Anycast

- Distributes load across multiple locations
- Provides redundancy



Global Anycast

- Distributes load across multiple locations
- Provides redundancy



Anycast with DNS

DNS, recursive servers

- Configured by IP address on clients
- Latency is important
- Distribute load across multiple devices

DNS, authoritative

- Limited number of authority IP’s can be listed in a single reply packet
- Latency to the server is important
- Redundancy a large concern
- Distribute load across multiple devices

POLL QUESTION

Are you Anycasting Today?

(results will be shared at the end of the presentation)

IMPACT ON PROTOCOLS

Explore

Impact on Protocols: ICMP

- Global, stateless options work fine
  - Ping request/reply
  - ICMP Traceroute
    - Network instability can produce some odd results with traceroute
- Avoid LAN options
  - Router Advertisement/Solicitation
  - Address Mask Request/Reply
  - Redirect
  - A unicast address on the server can mitigate these issues
  - It's easy to avoid all of these ICMP options

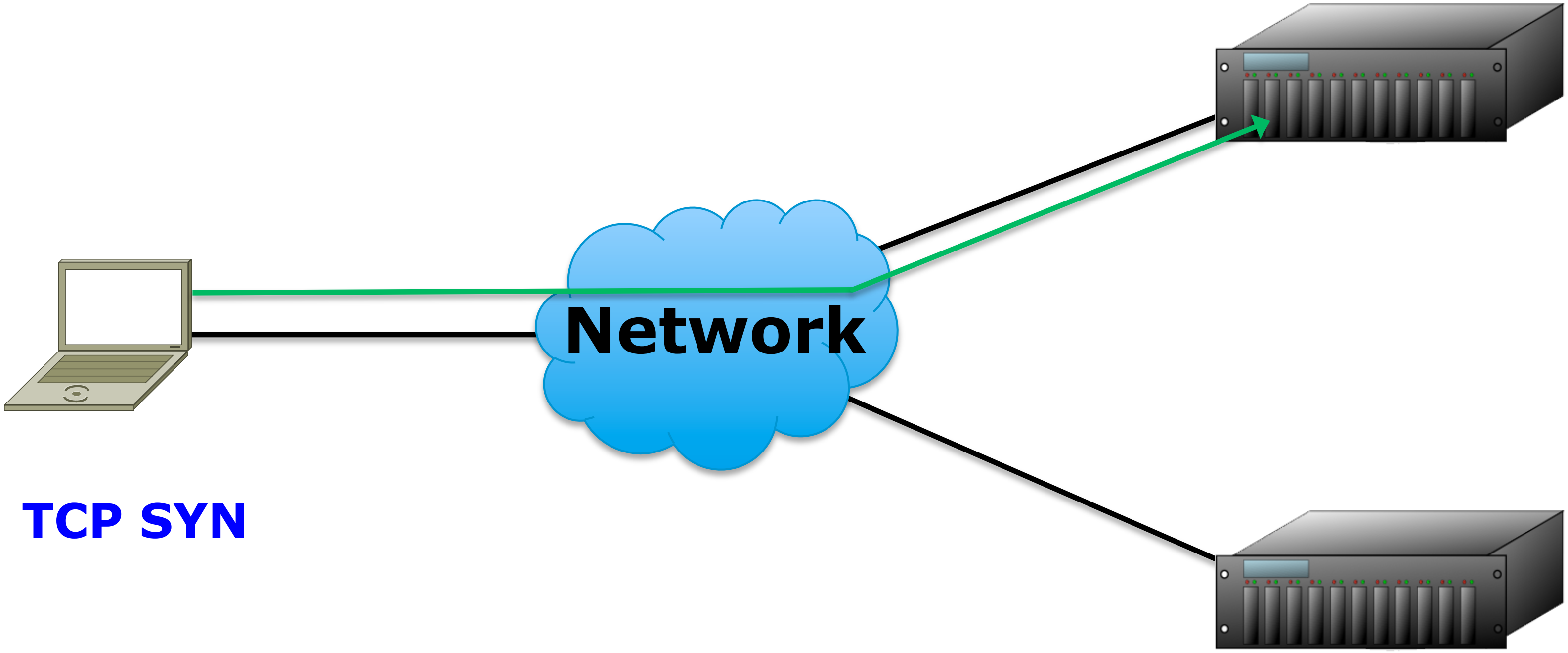
Impact on Protocols: UDP

- Stateless, which is good for Anycast
- Works well when the query is one packet, and the response is 1-n packets, and there is no state between queries
  - Sounds like the majority of DNS queries!
- If the query is more than one packet, or there is state between queries, the behavior tends to be the same as TCP

Impact on Protocols: TCP

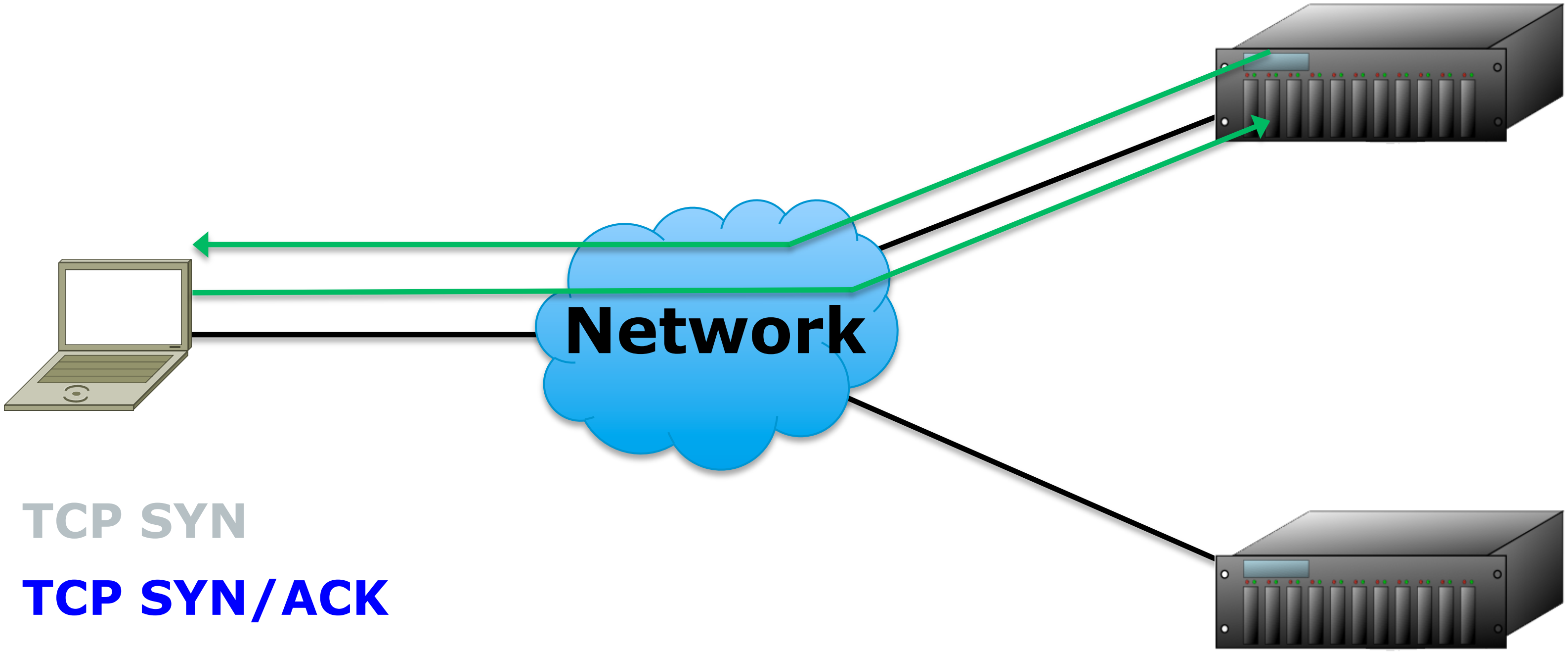
- Only works when the network path is stable.
  - This is ***never true in the long term***, but is often true for short periods of time
- **The Unicast sender has to reach the same Anycast destination for the duration of the connection**
  - One packet to the wrong device causes it to generate a TCP Reset, which generally tears down the connection

Impact on Protocols: TCP



TCP SYN

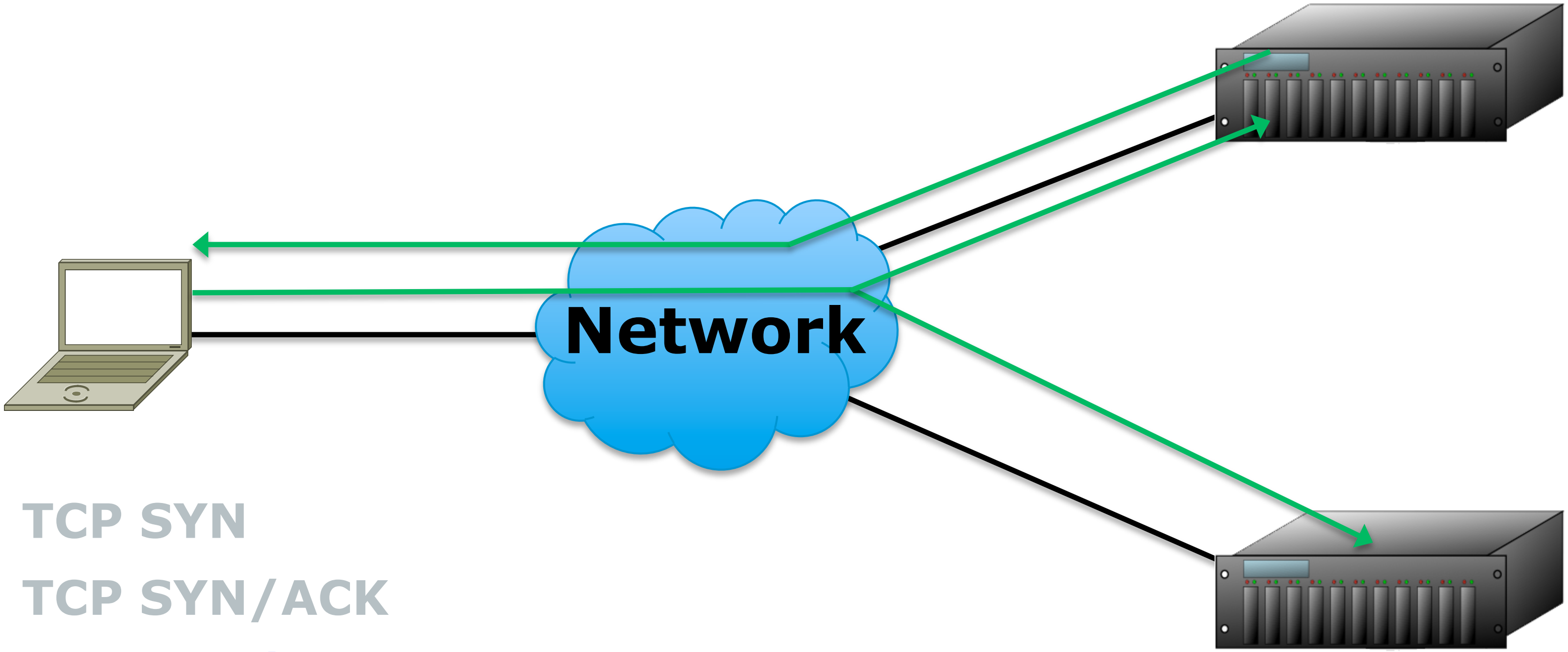
Impact on Protocols: TCP



TCP SYN

TCP SYN/ACK

Impact on Protocols: TCP

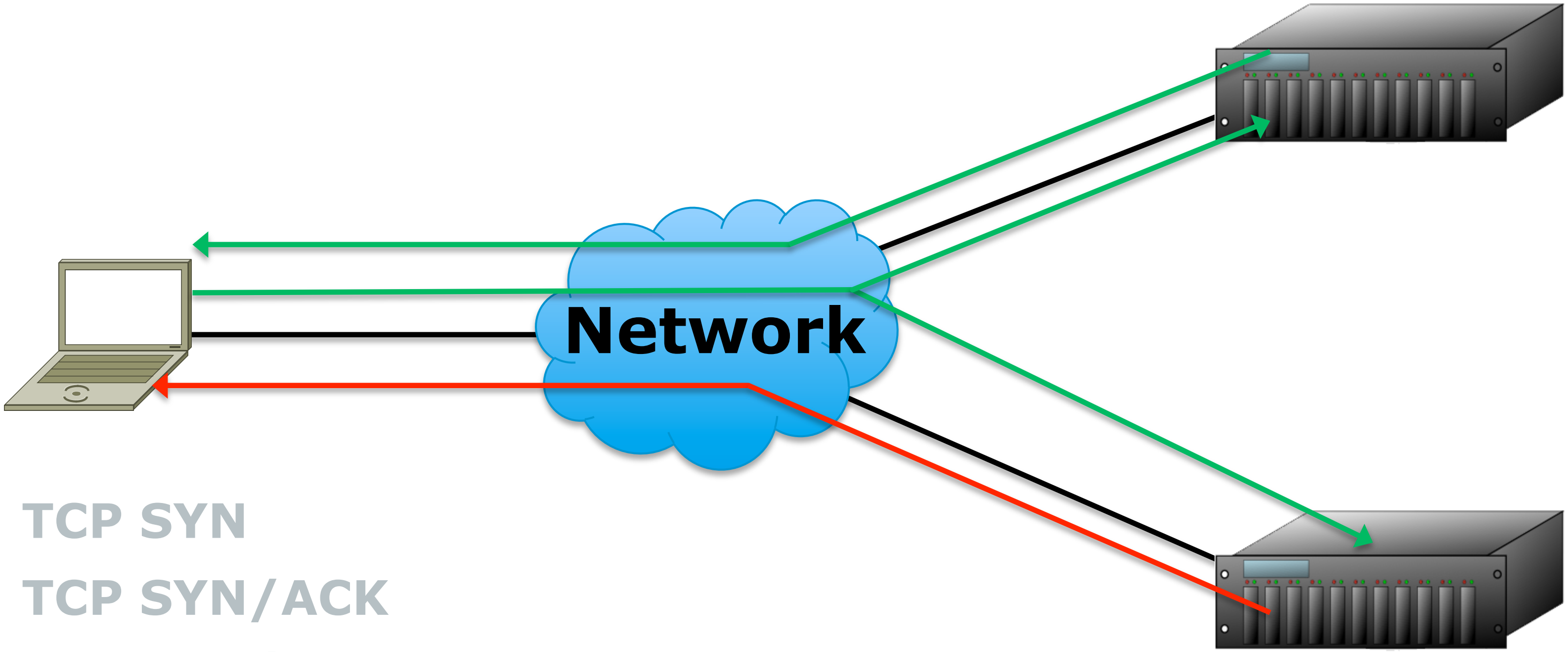


TCP SYN

TCP SYN/ACK

TCP ACK/Data

Impact on Protocols: TCP



TCP SYN

TCP SYN/ACK

TCP ACK/Data

TCP Reset

Impact on Protocols: TCP

- Operationally, what does it mean?
  - The location of the Anycast servers is important, and depends on the network topology and configuration
  - When properly deployed, there is a high success rate for short duration connections
  - The longer the connection, the greater the risk of failure
- For Internet services it's not just your network, but ***every network the packet traverses*** to the Anycast server!

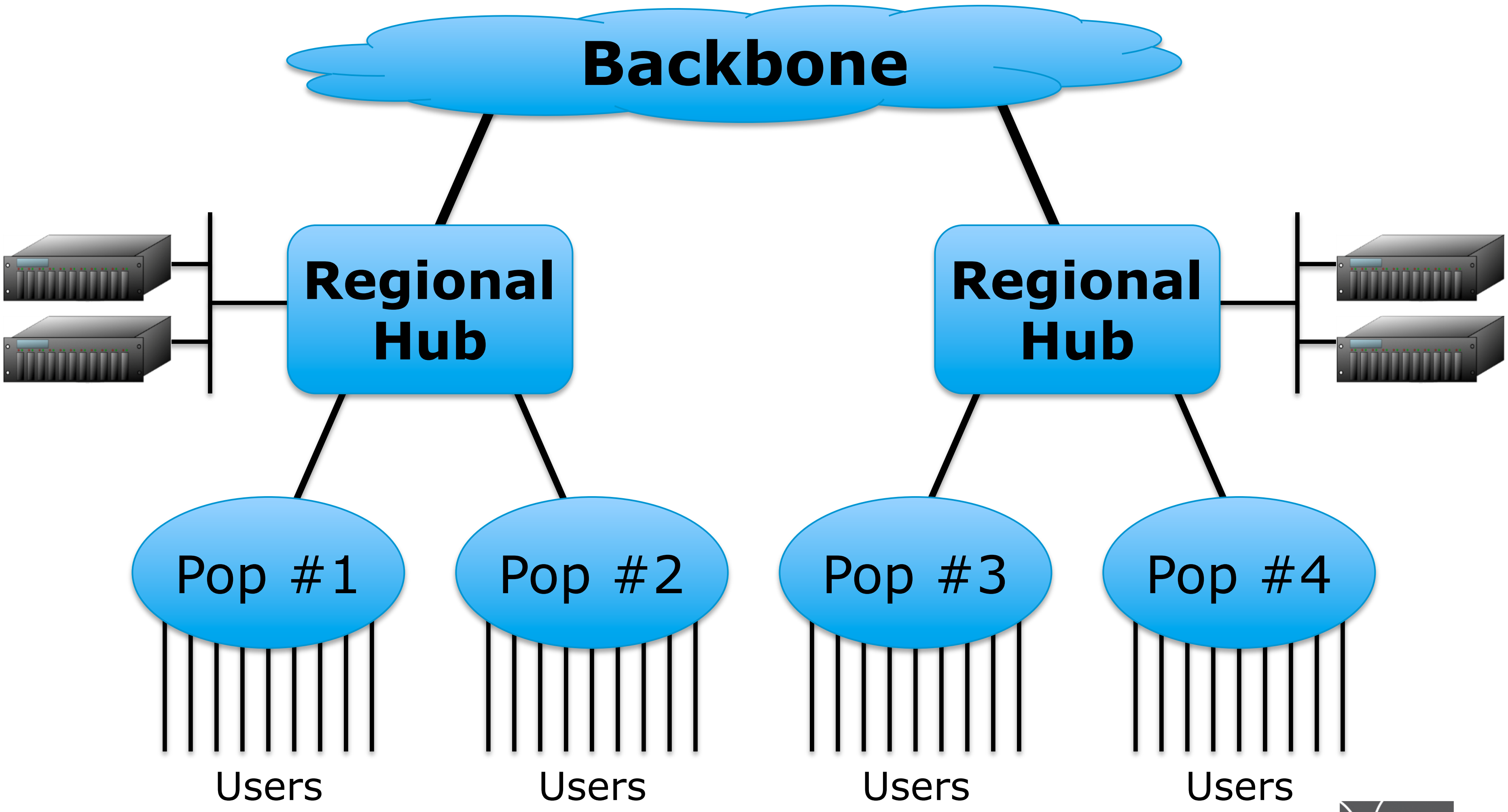
DNS & ANYCAST

Explore

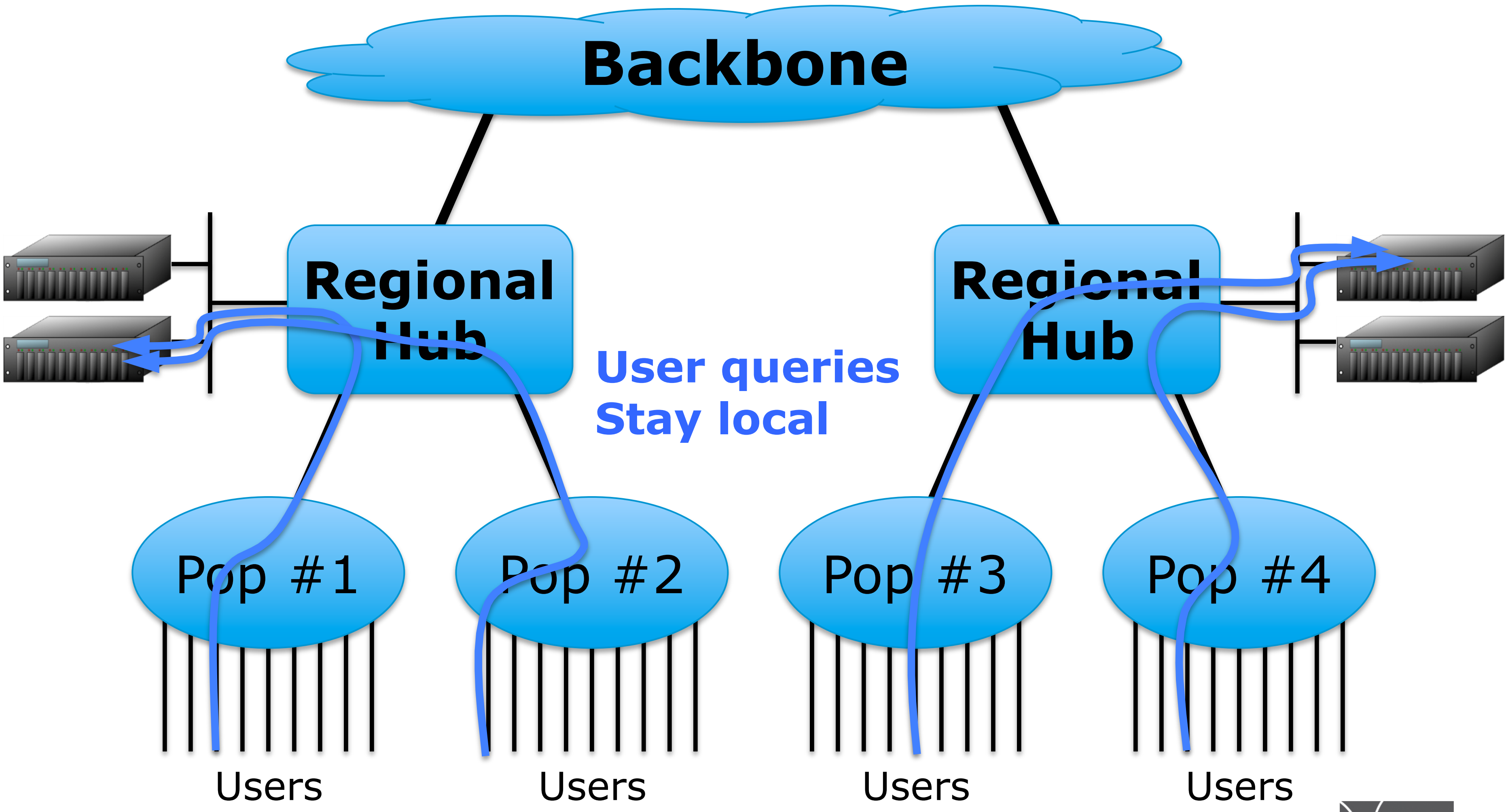
DNS & Anycast

- Most common queries are a single UDP packet, with 1-3 UDP packets of response
- TCP queries are extremely short lived
  - User->Server: SYN, ACK w/query, ACK/FIN
  - Server->User: SYN/ACK, ACK w/Data, ACK/FIN
    - Maybe an additional data packet
  - The FIN can be lost in some implementations and the data still be received
- Zone transfers are longer lived TCP queries
  - Length depends on zone size
  - Some zones don’t allow, mitigating the issue

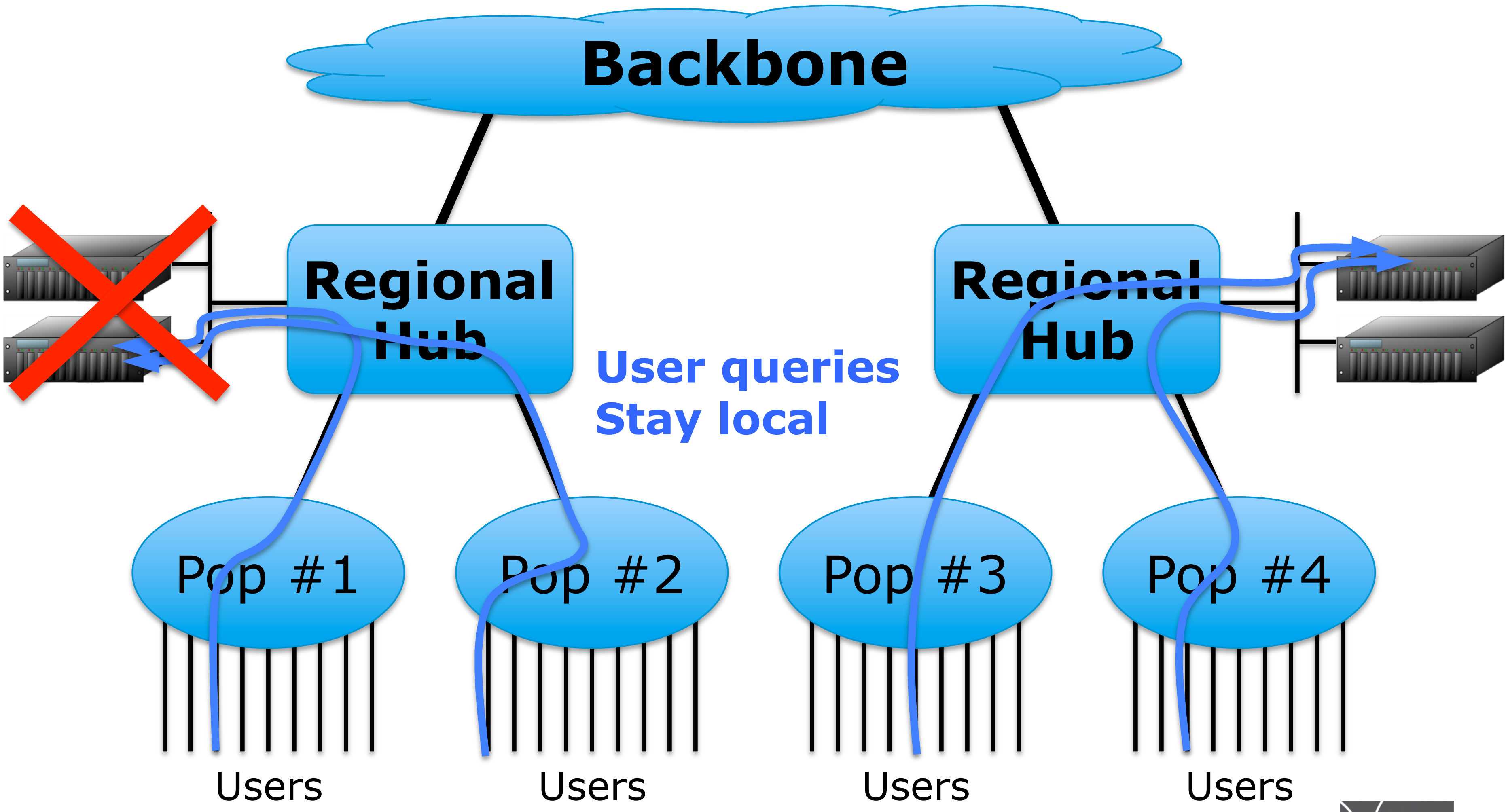
End User Resolvers



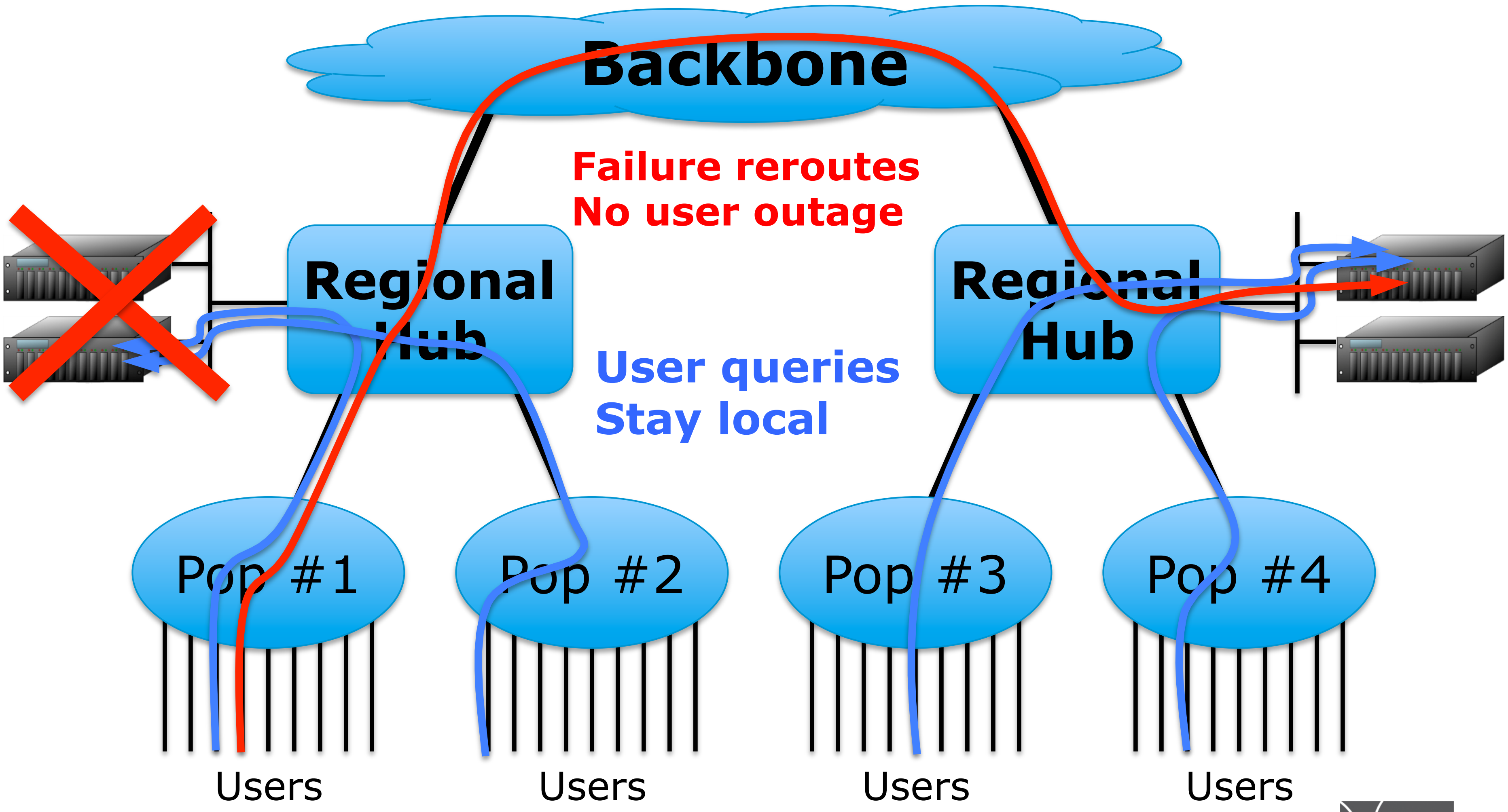
End User Resolvers



End User Resolvers

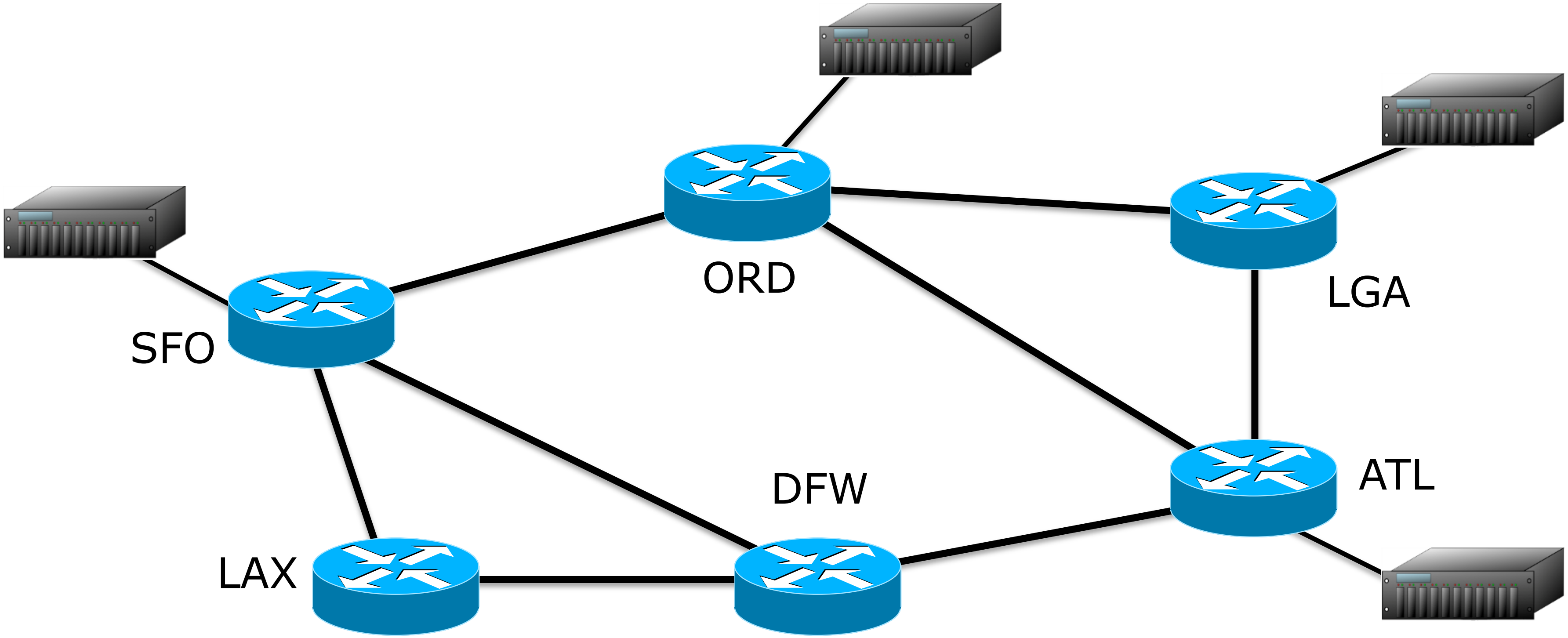


End User Resolvers



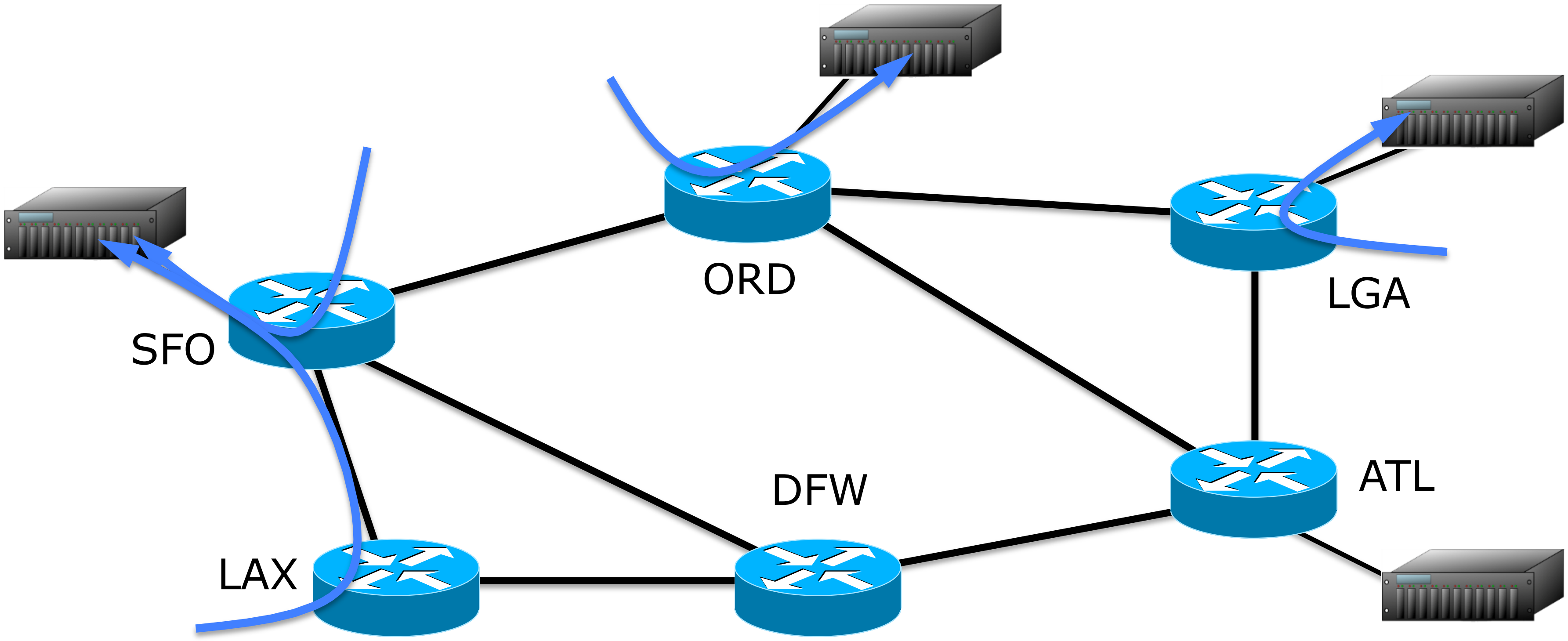
Anycast & DNS

- Authority servers across an ISP/Enterprise provide redundancy, load distribution and hitless maintenance



Anycast & DNS

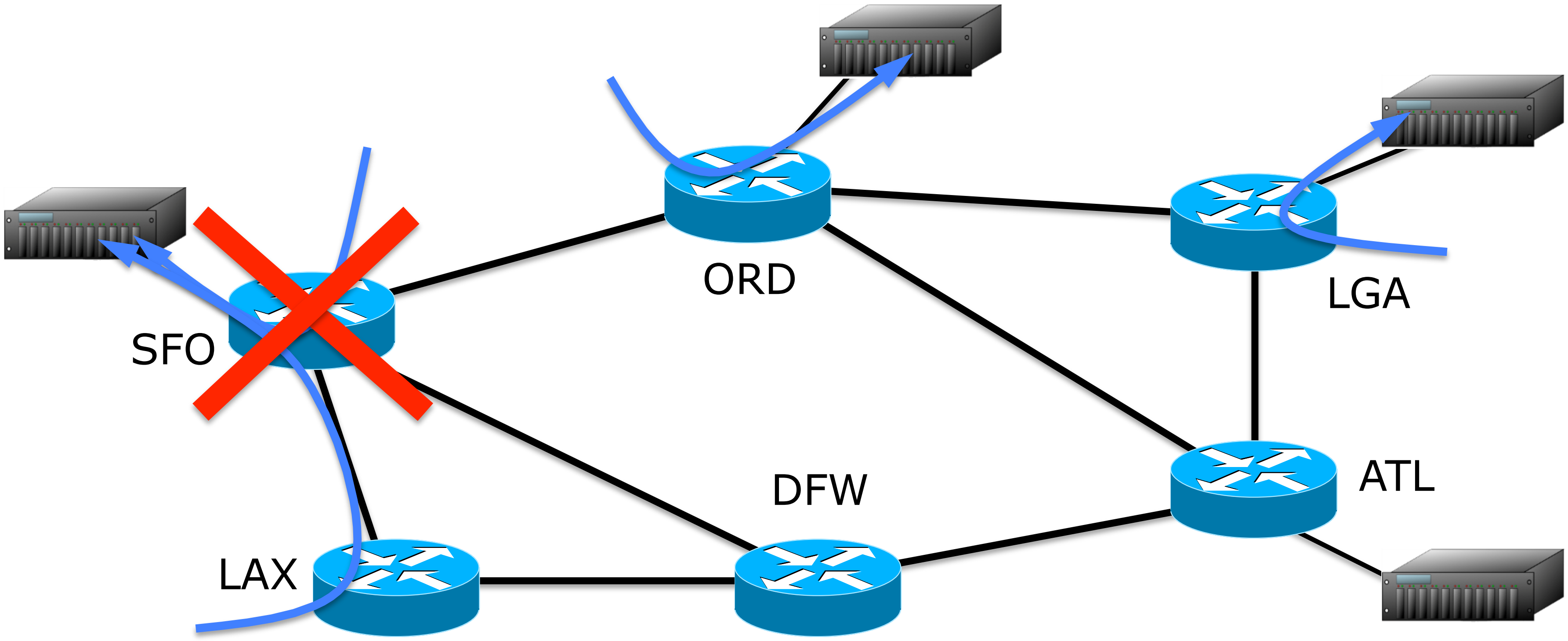
- Authority servers across an ISP/Enterprise provide redundancy, load distribution and hitless maintenance



Queries stay local

Anycast & DNS

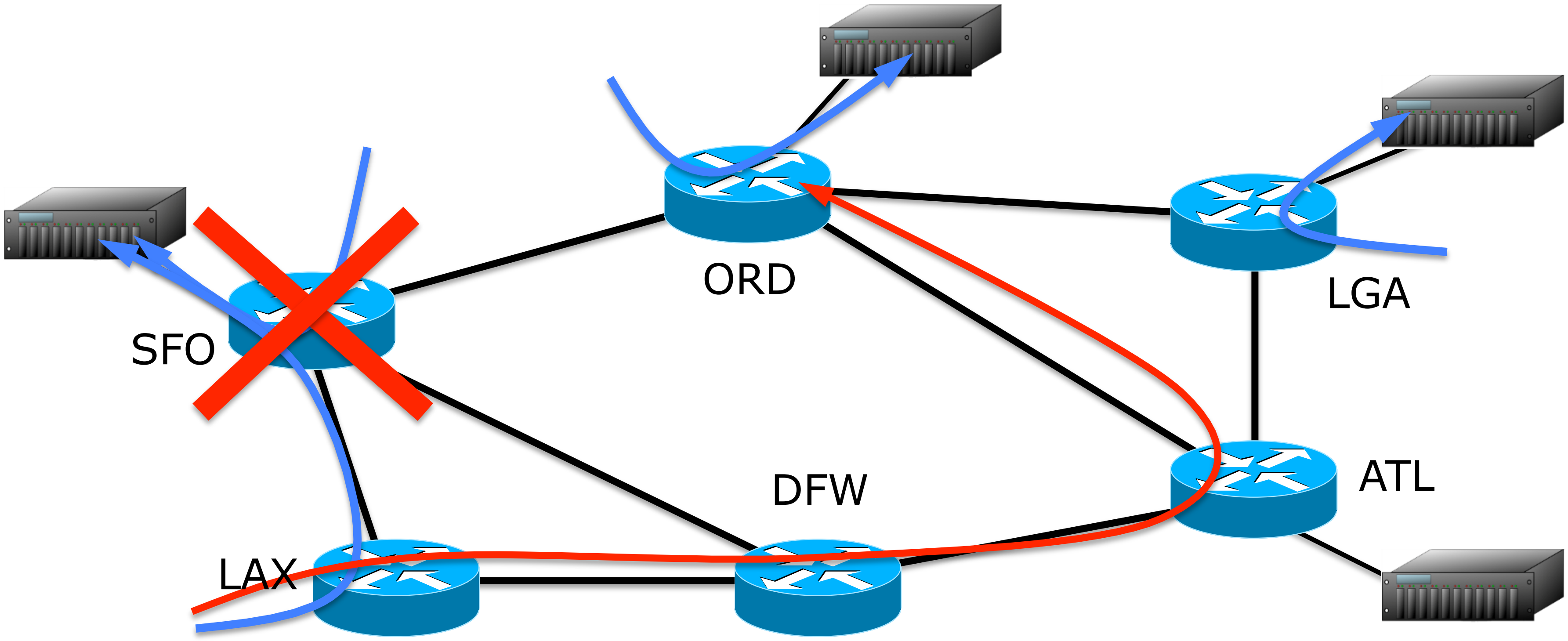
- Authority servers across an ISP/Enterprise provide redundancy, load distribution and hitless maintenance



Pop Failure

Anycast & DNS

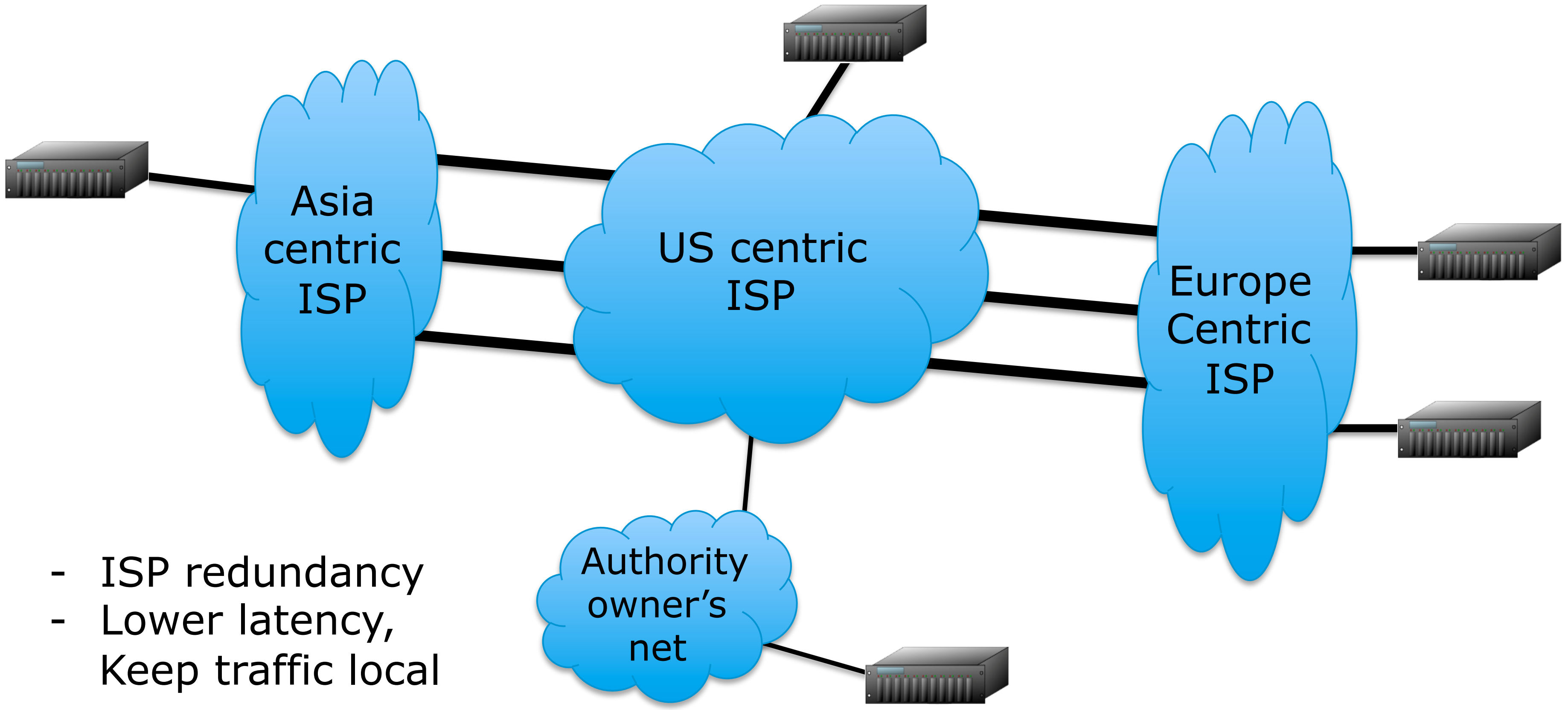
- Authority servers across an ISP/Enterprise provide redundancy, load distribution and hitless maintenance



Queries re-routed, service still up

Anycast & DNS

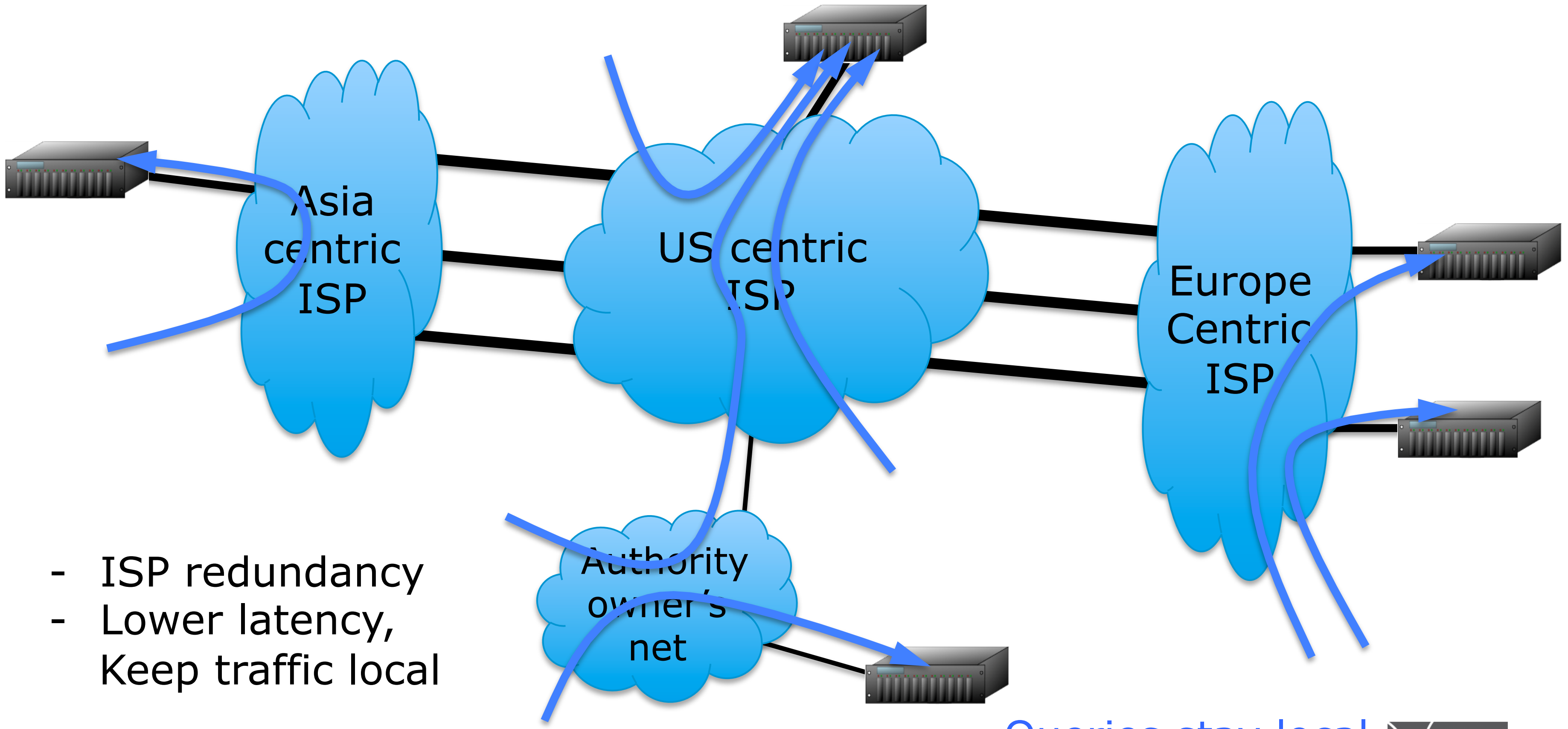
- Authority servers across multiple networks



- ISP redundancy
- Lower latency, Keep traffic local

Anycast & DNS

- Authority servers across multiple networks



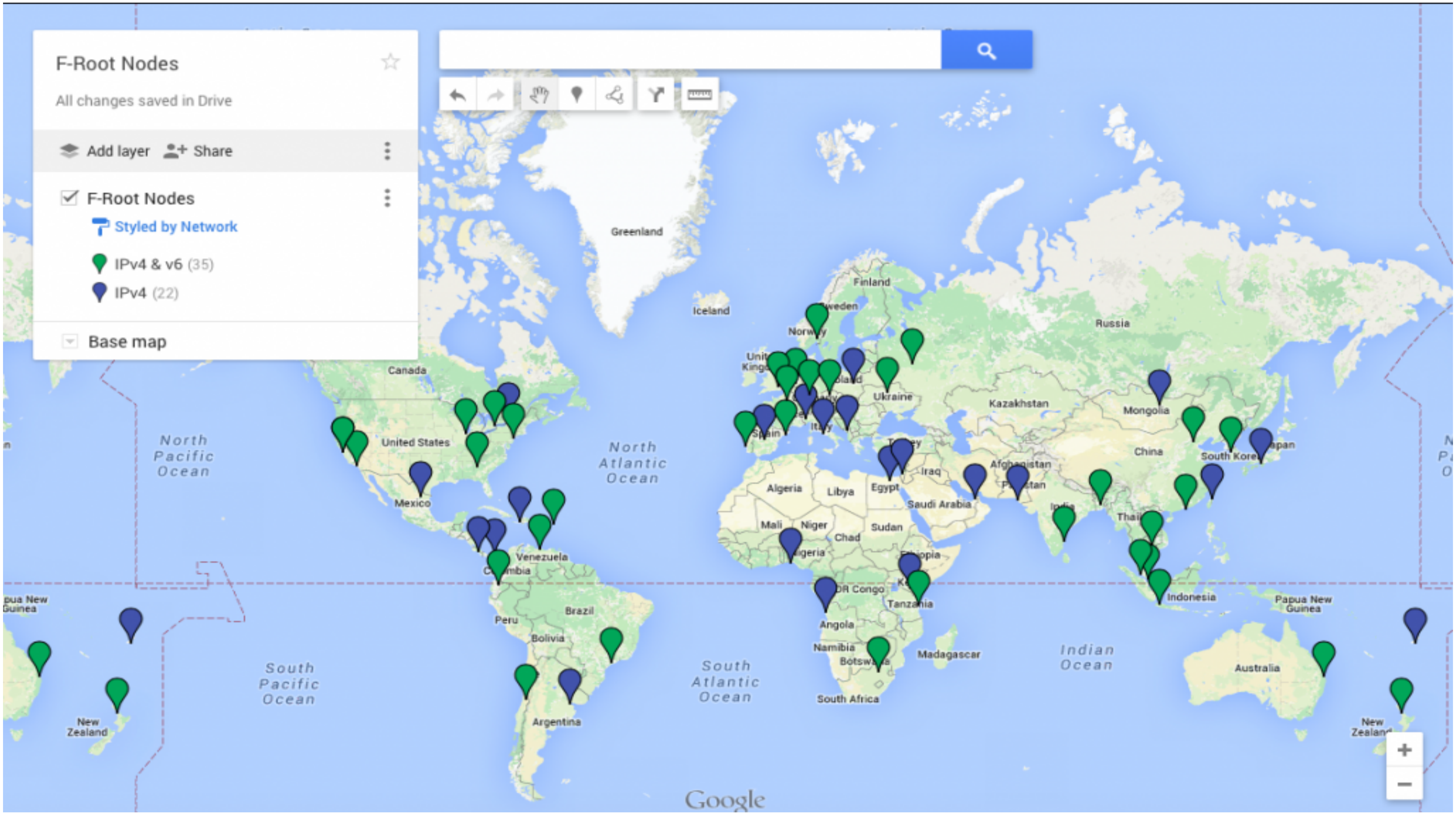
- ISP redundancy
- Lower latency, Keep traffic local

Queries stay local

ISC'S OPERATIONAL EXPERIENCE

Share

F-Root



F-Root Nodes

All changes saved in Drive

☑ F-Root Nodes

Styled by Network

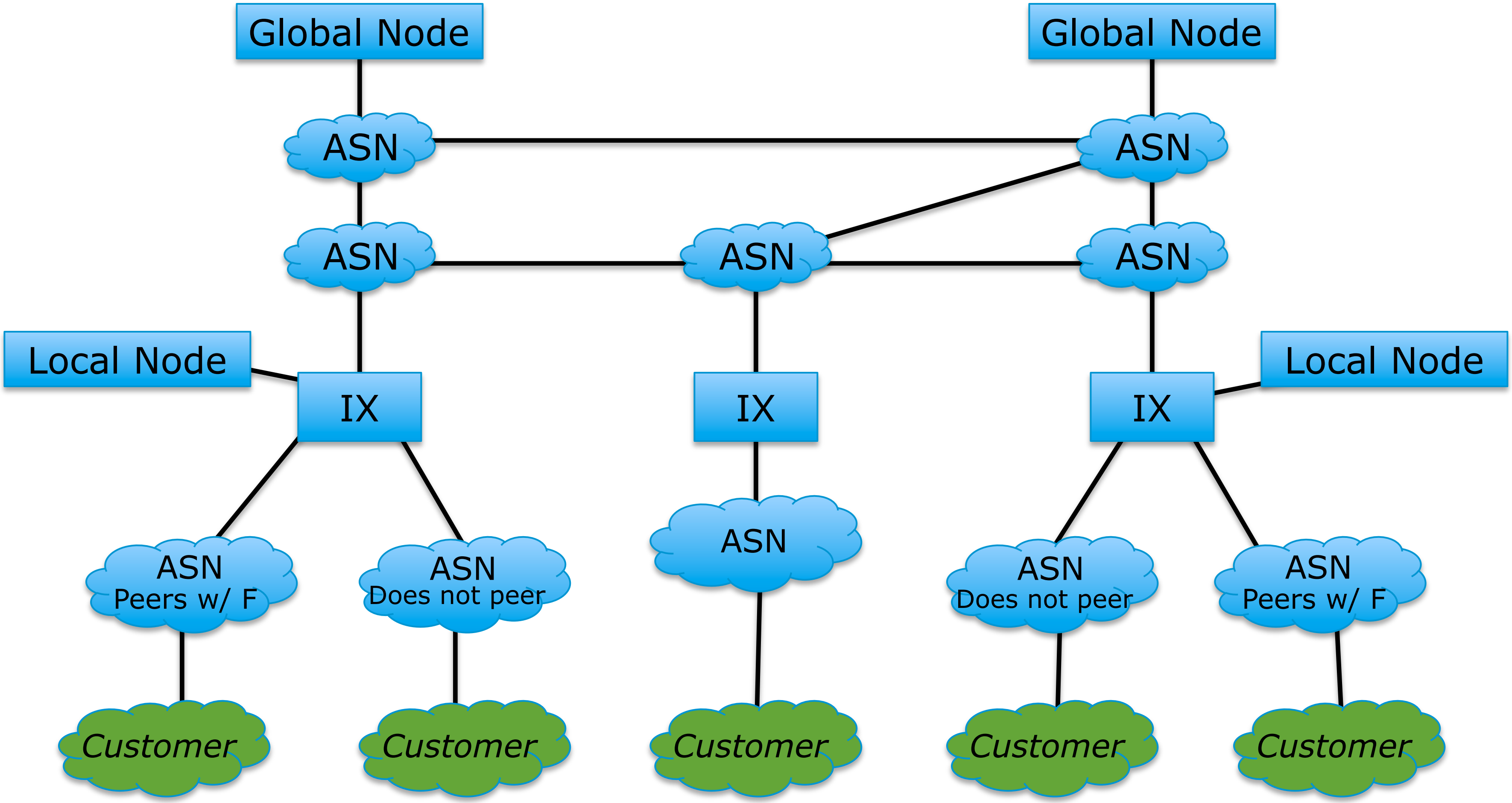
🟩 IPv4 & v6 (35)

🟦 IPv4 (22)

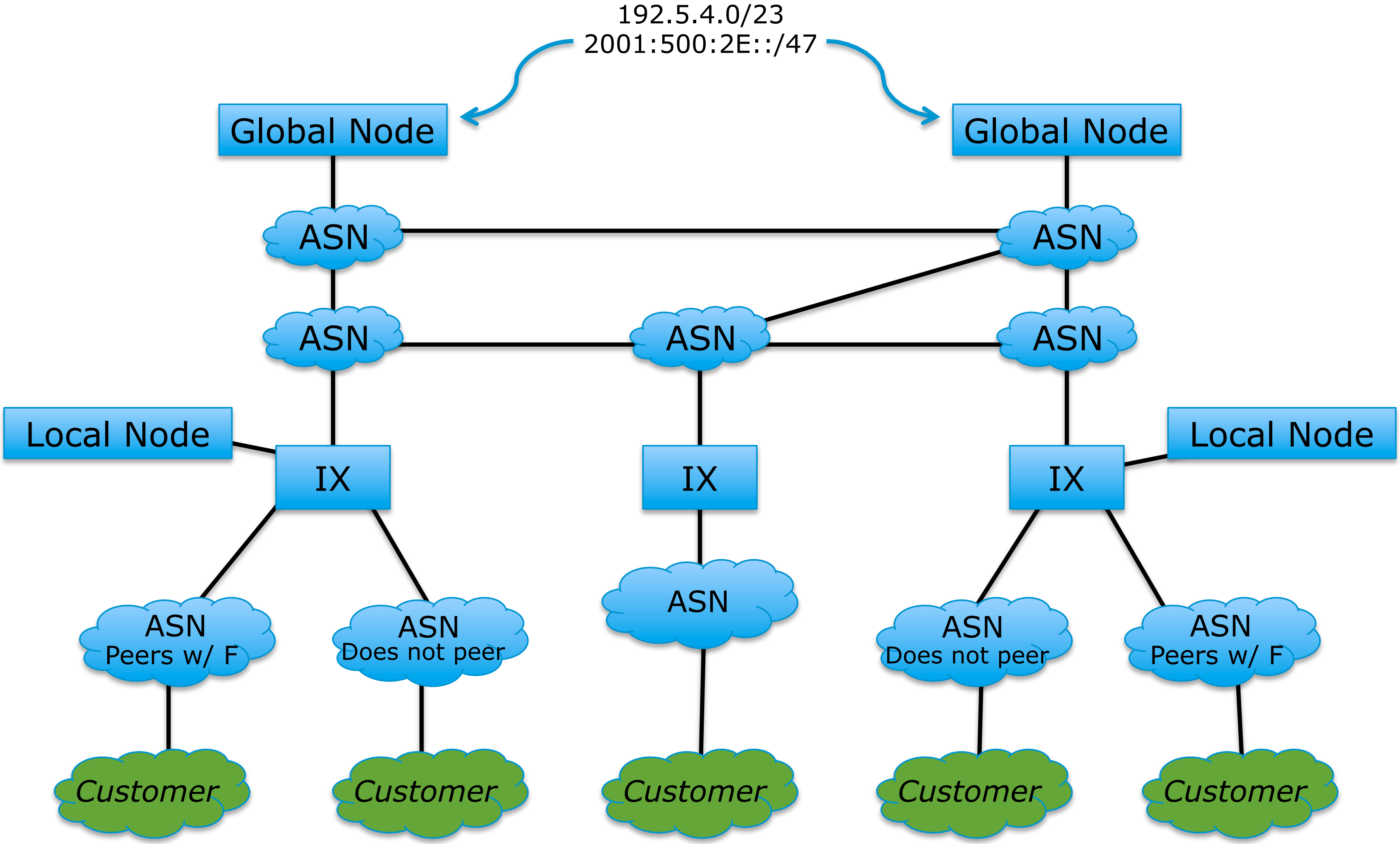
F-Root – 3 Levels

1. Local LAN
   - Each deployment has a minimum of 2 servers on the local network for redundancy, more where necessary
2. Local Nodes
   - A typical F-Root deployment at a exchange point or inside of an ISP network
   - Announces 192.5.5.0/24 and 2001:500:2f::/48 with NO_EXPORT set
     - Because of the NO_EXPORT settings these routes will not be visible to all end users
3. Global Nodes
   - Larger nodes, with significant transit capacity
   - Announce 192.5.4.0/23 and 2001:500:2e::/47, supernets of the local node prefixes
   - These networks should be visible to all end users on the Internet

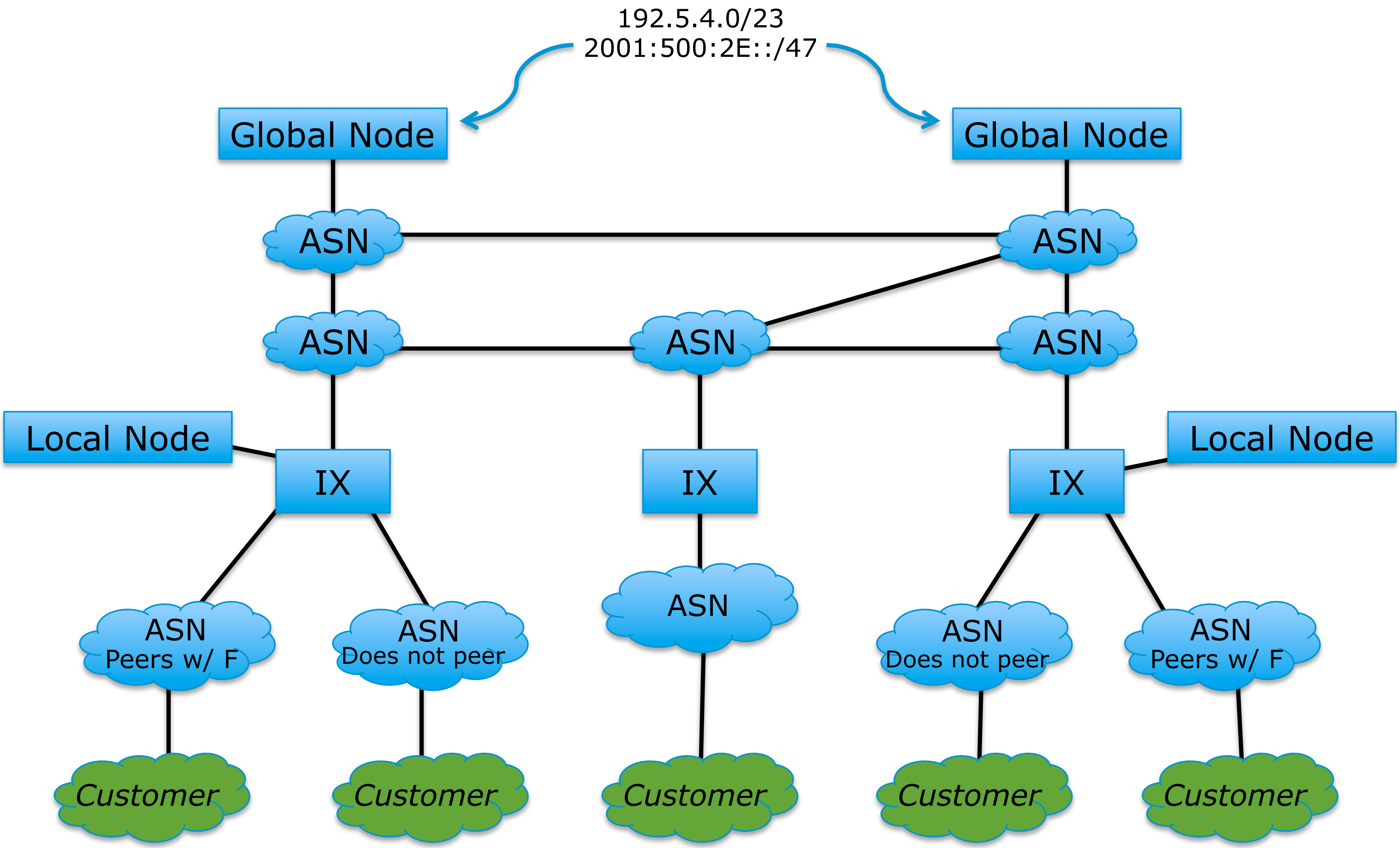
F-Root


Global Node
Global Node
ASN
ASN
ASN
ASN
ASN
Local Node
IX
IX
IX
Local Node
ASN
ASN
Peers w/ F
ASN
Does not peer
ASN
Does not peer
ASN
Peers w/ F
Customer
Customer
Customer
Customer
Customer

F-Root

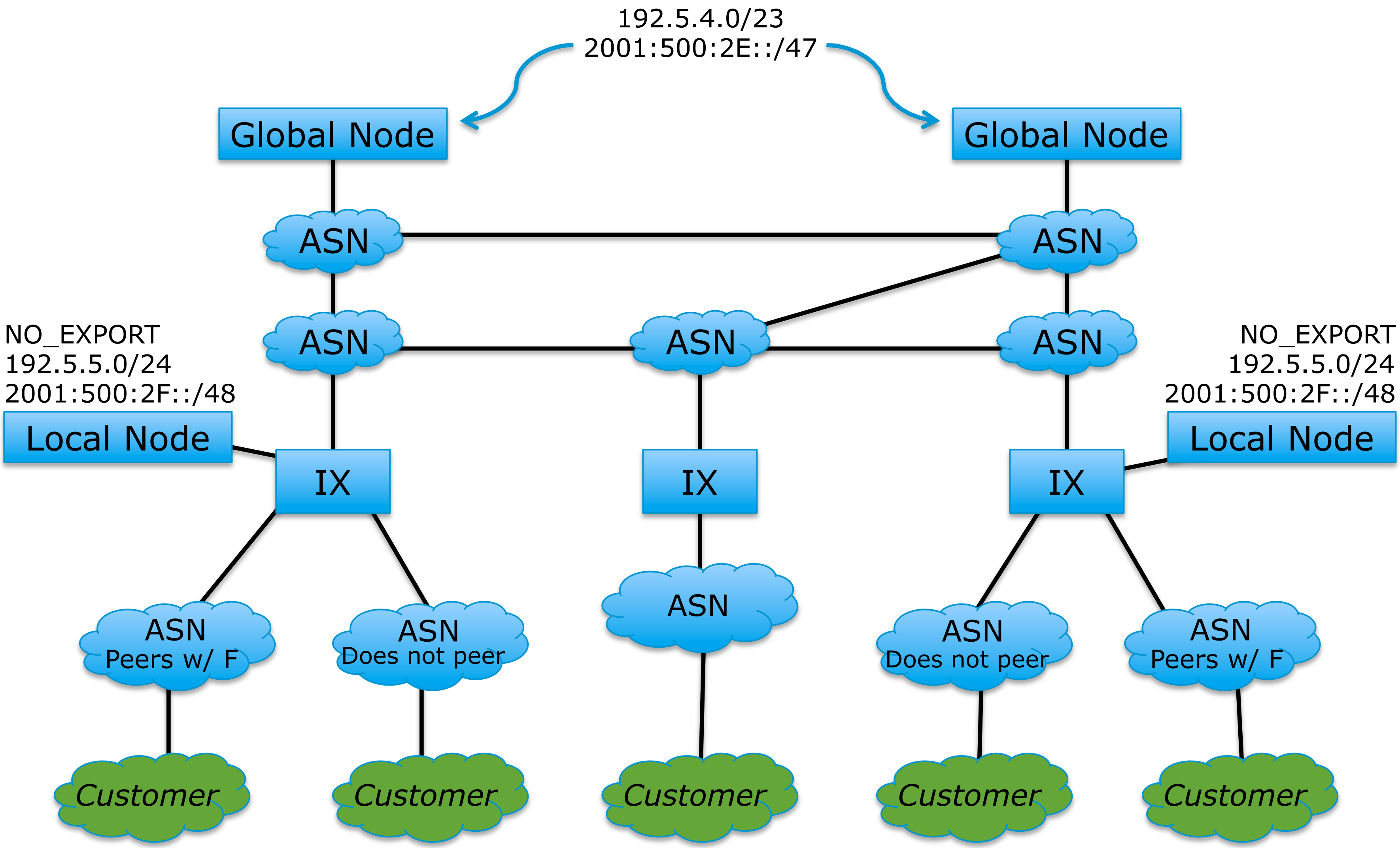

192.5.4.0/23
2001:500:2E::/47
Global Node
Global Node
ASN
ASN
ASN
ASN
ASN
Local Node
Local Node
IX
IX
IX
ASN
ASN
Peers w/ F
ASN
Does not peer
ASN
ASN
Does not peer
ASN
Peers w/ F
Customer
Customer
Customer
Customer
Customer

F-Root



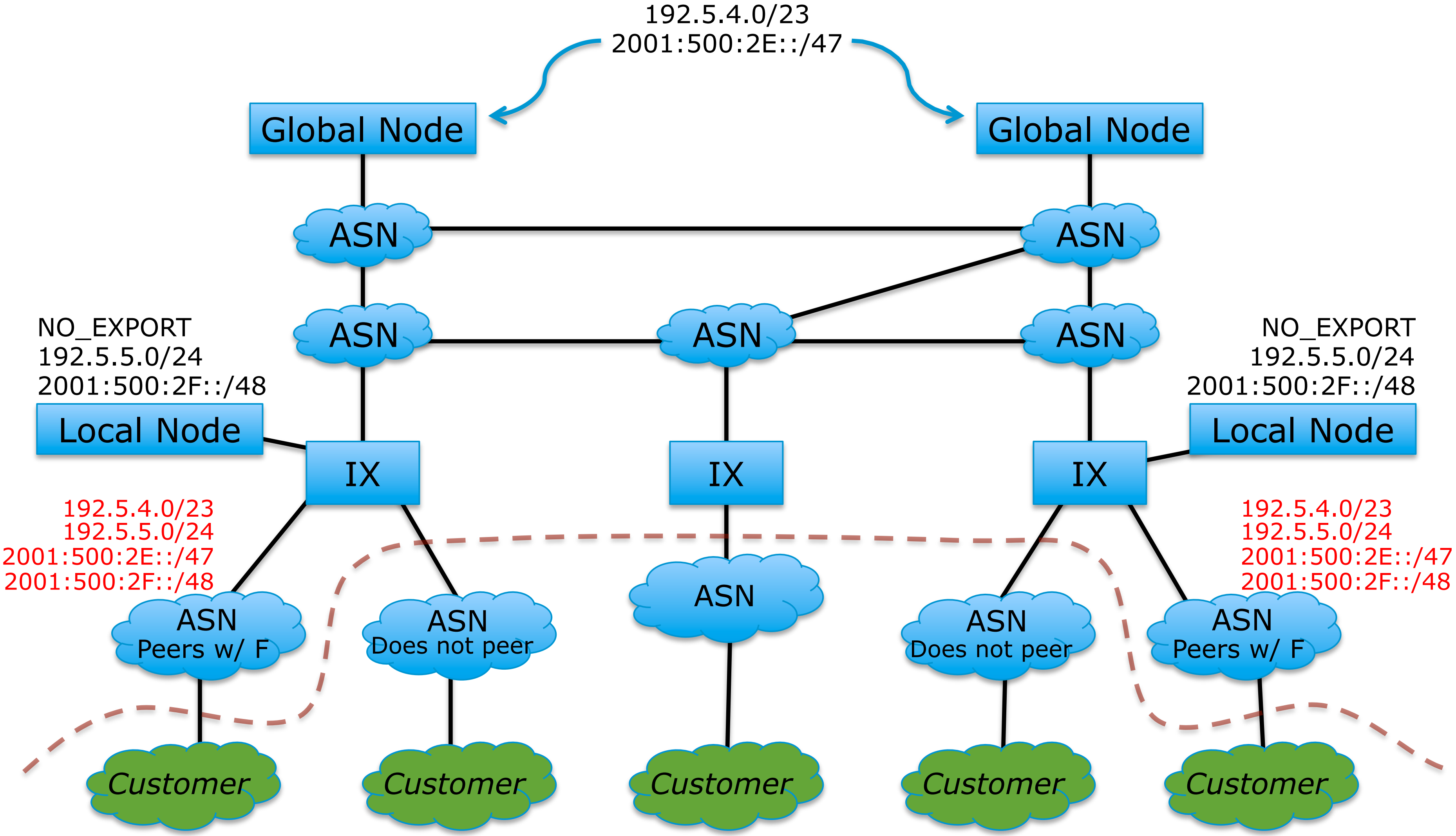
192.5.4.0/23 2001:500:2E::/47

F-Root



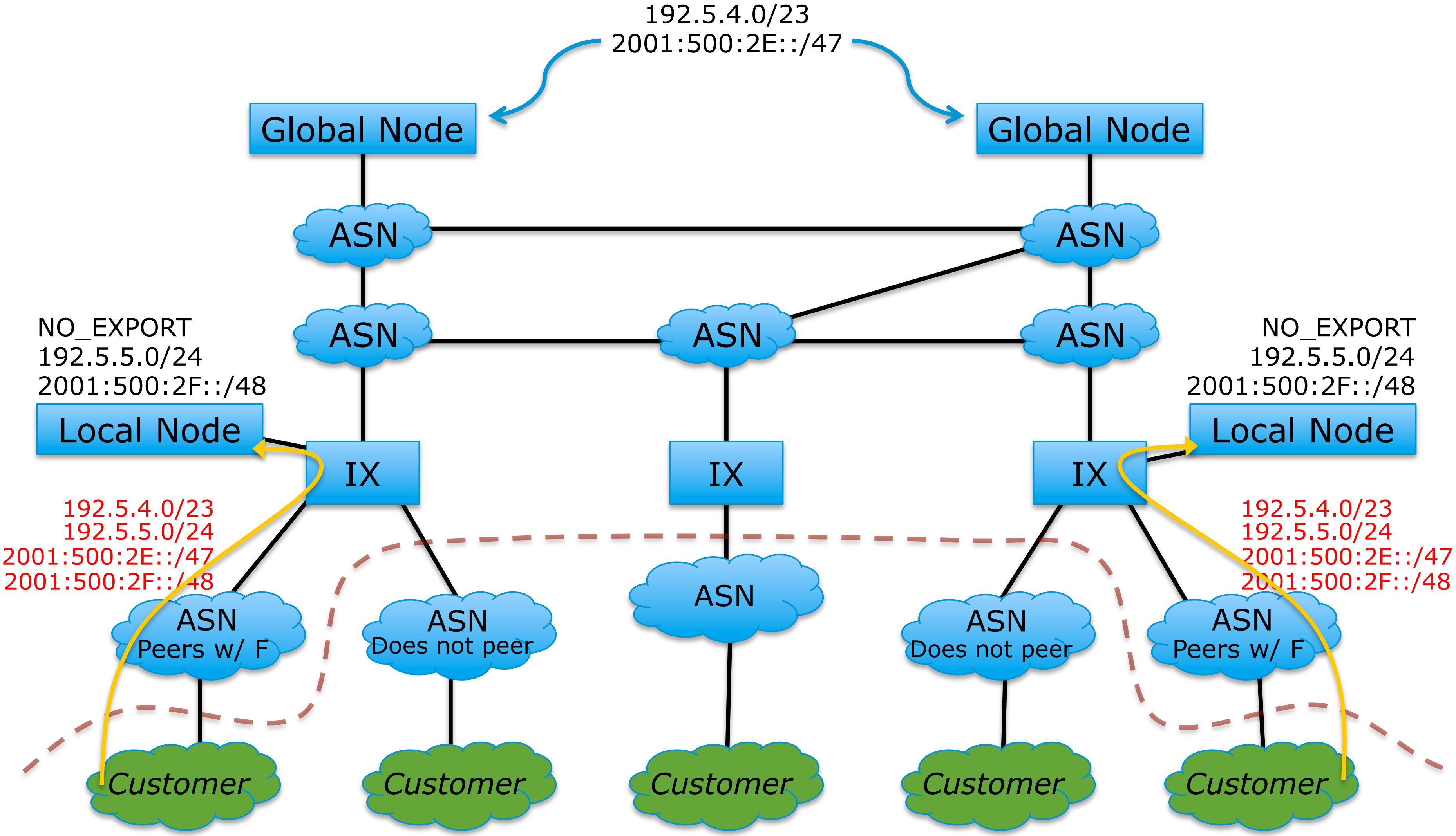
192.5.4.0/23 2001:500:2E::/47

F-Root



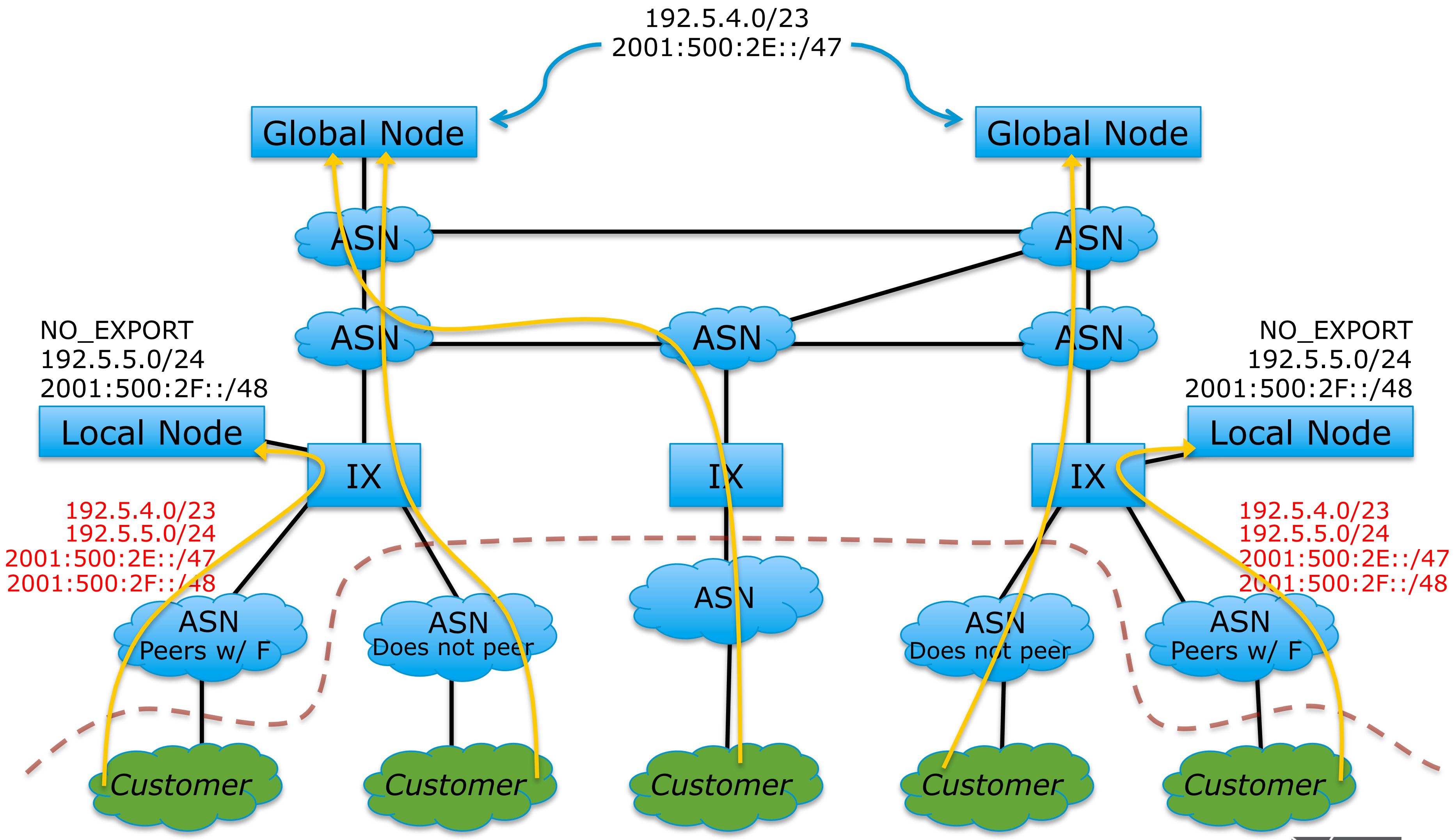
192.5.4.0/23 2001:500:2E::/47

F-Root



192.5.4.0/23 2001:500:2E::/47

F-Root



192.5.4.0/23 2001:500:2E::/47

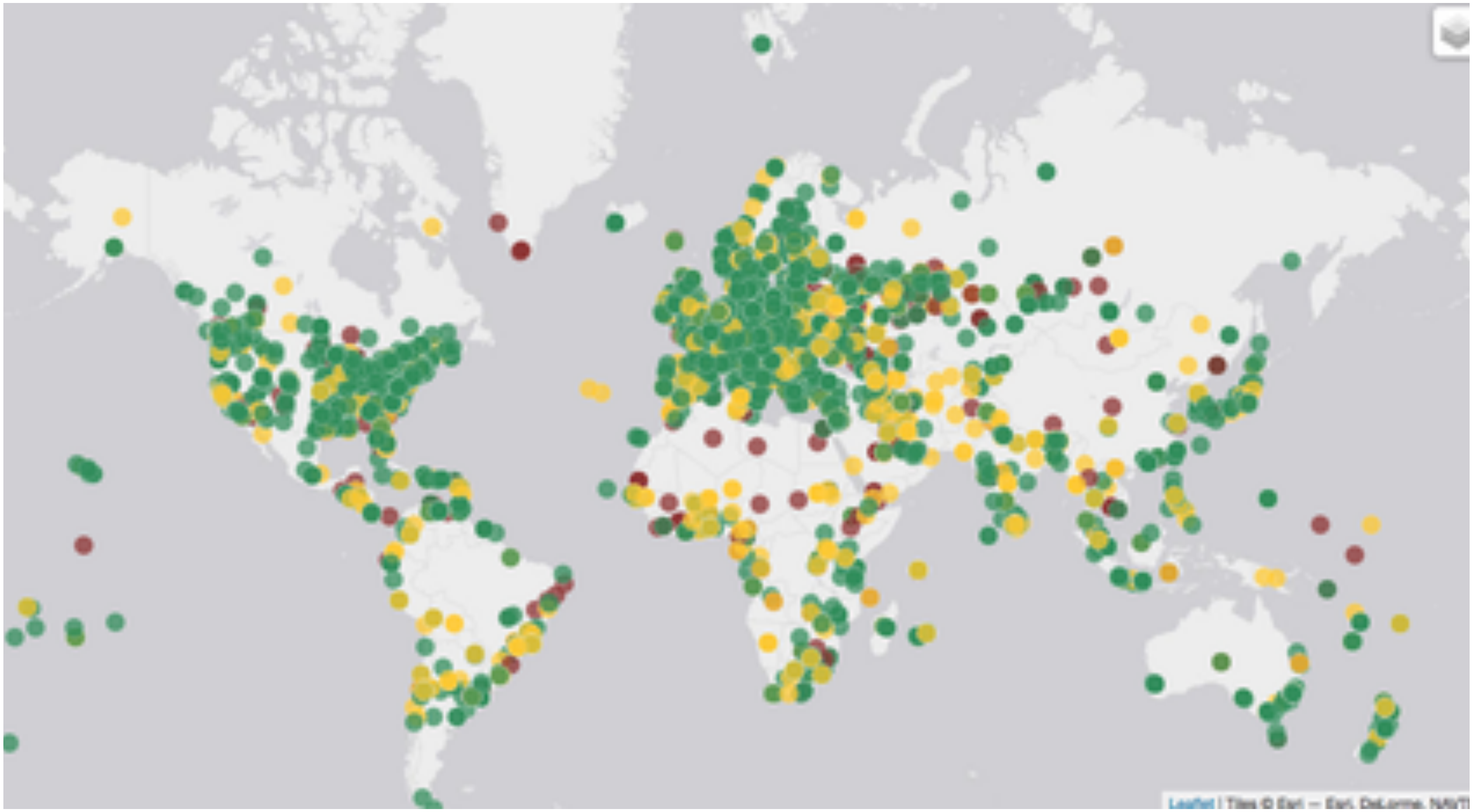
Why 3 Levels?

- A strong desire to keep local traffic local
  - Local nodes may be deployed in bandwidth starved areas, like behind satellite links, and thus shouldn't draw in queries from far away
  - Provide an incentive for local ISP's to peer with the local F-Root instance
- Diversity in the Root Server ecosystem
  - Root operators believe that having different parties deploy in different models allows for more effective service of different user communities, and provides a more difficult attack surface
  - No one else uses this method!

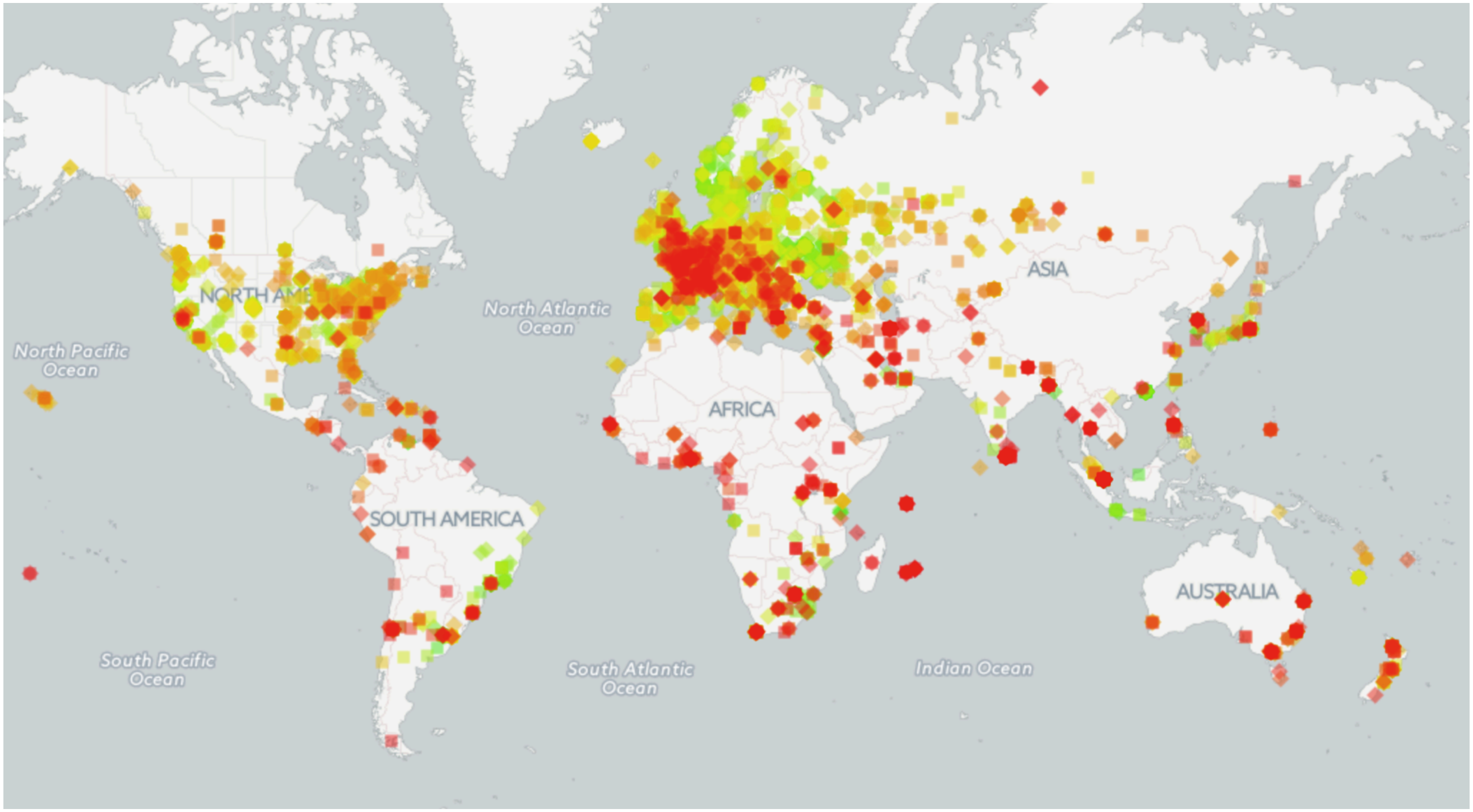
This does create some confusion

- ISP's think that because the local route has NO_EXPORT their customers won't see F-Root, but this isn't true due to the covering supernet

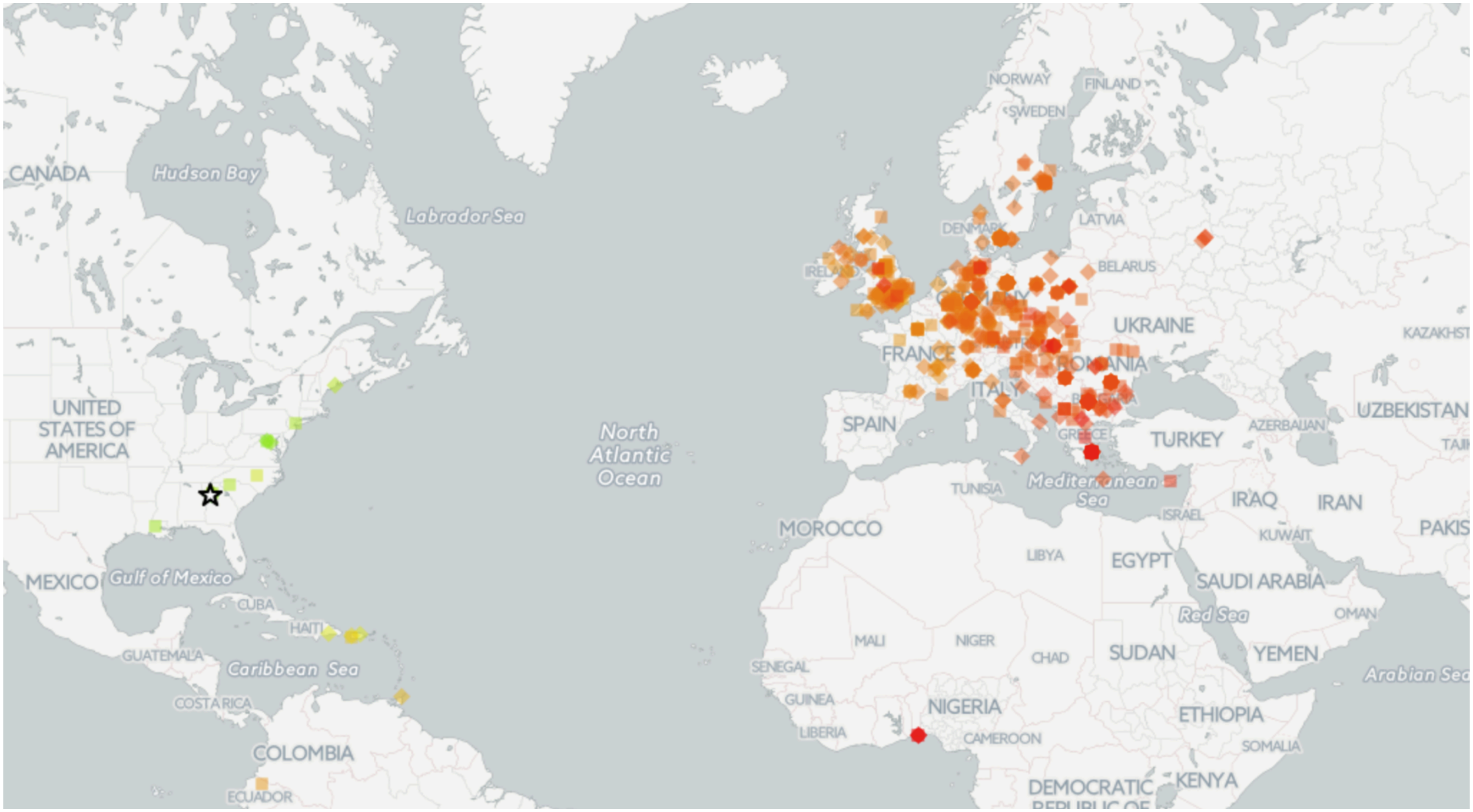
F-Root Measurements



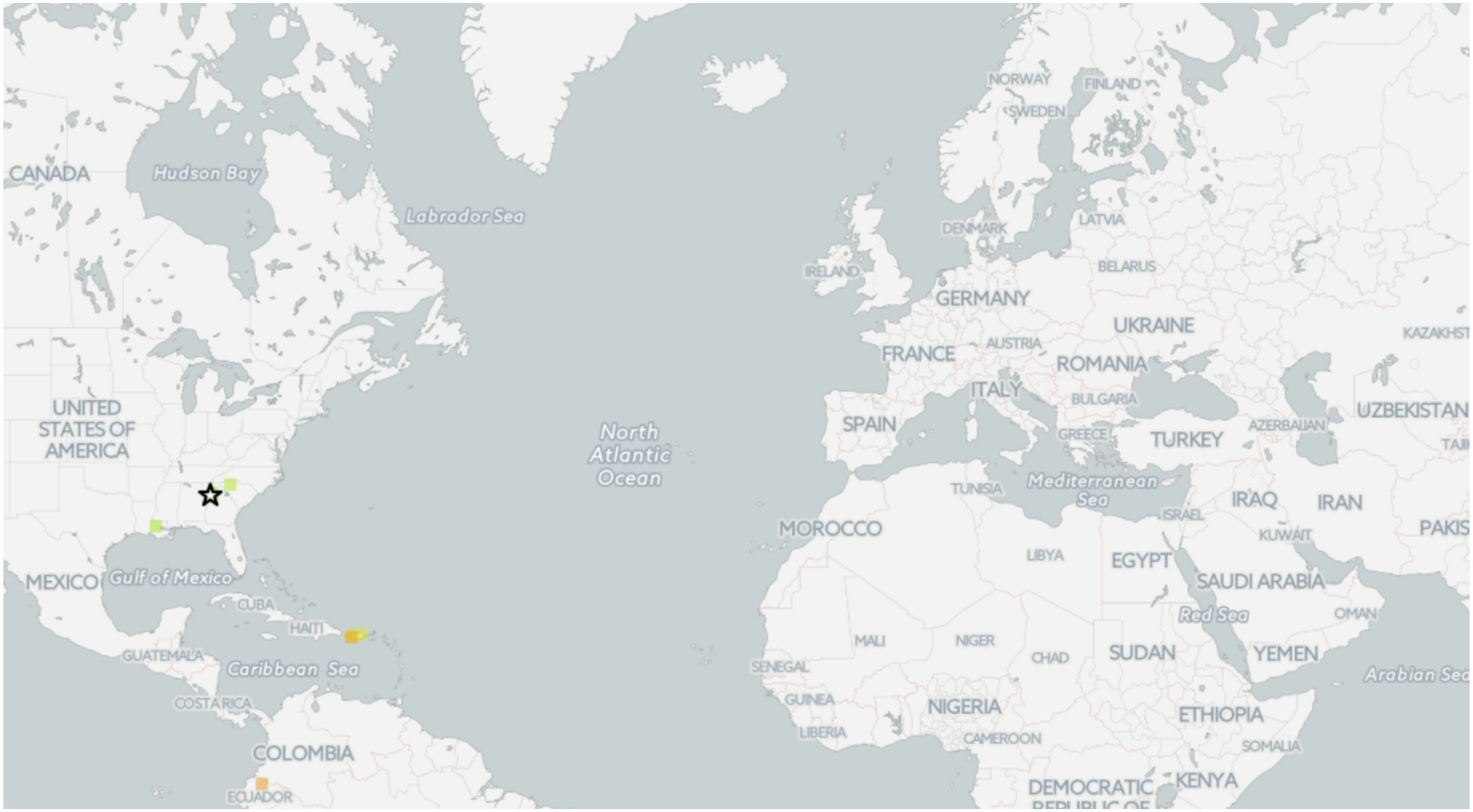
Global View of F-root Latency (red = 200ms+)



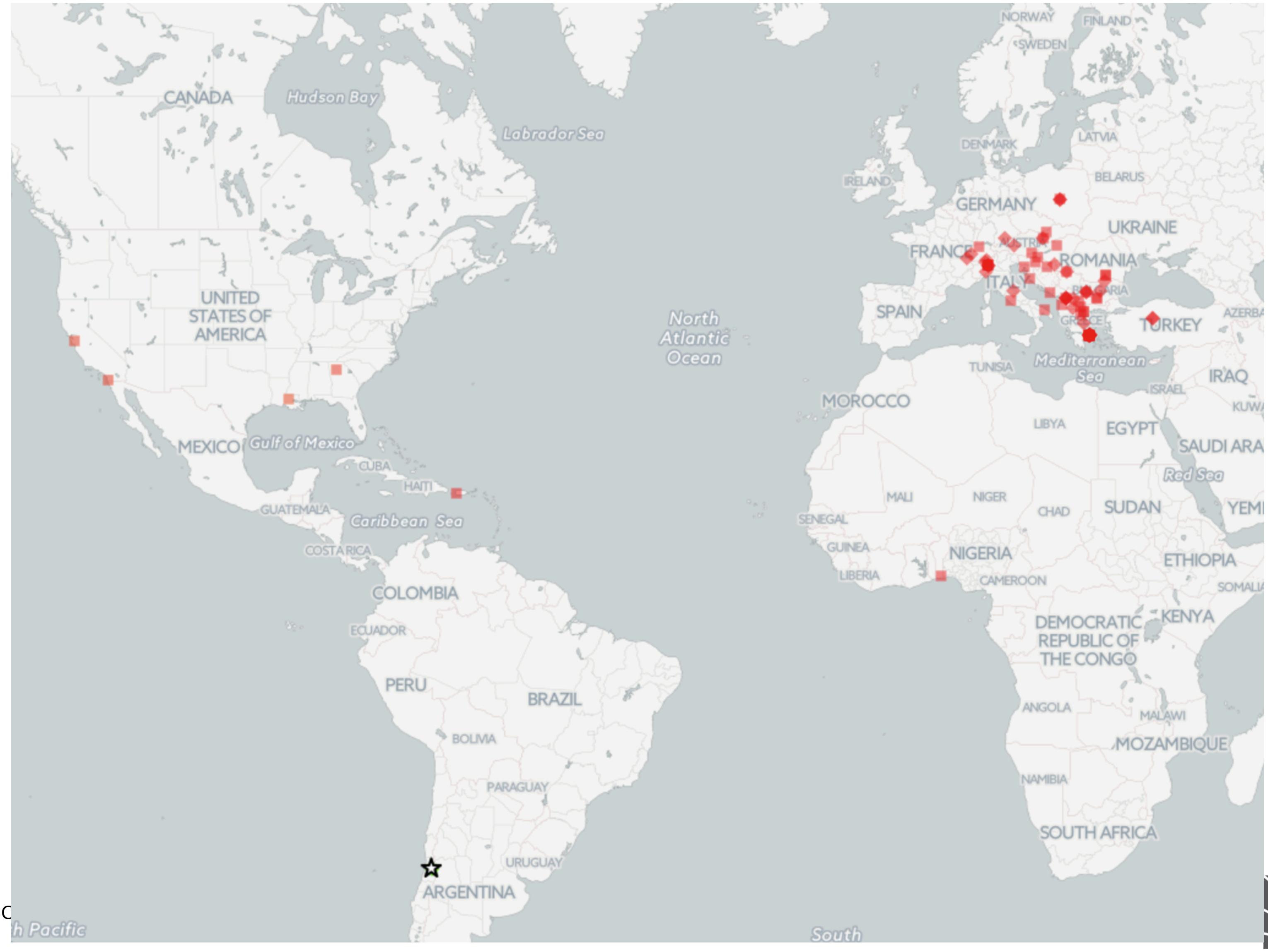
US Transit Misconfiguration (ATL1)



ATL1 - post reconfiguration



BGP NO_EXPORT leak!



How we trace an F-Root local leak

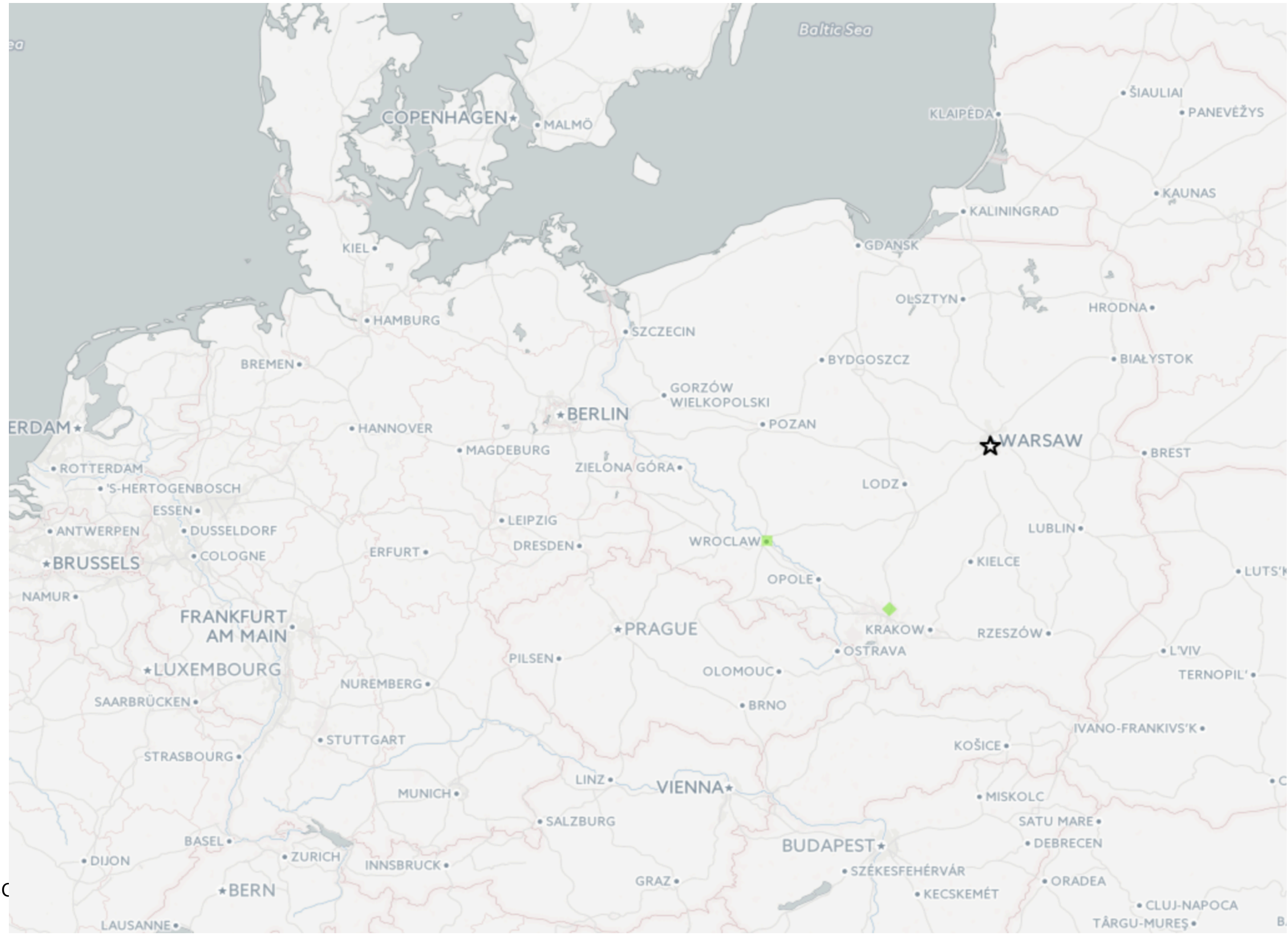
All sites originate the F-Root prefix with the same ASN 3557.
All sites then have their own unique site ASN.

```
core1.mil1.he.net> show ip bgp routes detail 192.5.5.0/24                              AS PATH
BI          192.5.5.0/24 80.81.194.57         160        100        0         33082, 3557 IGP
I           192.5.5.0/24 195.42.144.37        180        100        0         30126, 3557 IGP
I           192.5.5.0/24 193.149.1.38         210        100        0         25572, 3557 IGP
I           192.5.5.0/24 91.210.16.181        220        100        0         30134, 3557 IGP
I           192.5.5.0/24 5.57.80.224          260        100        0         33073, 3557 IGP
I           192.5.5.0/24 195.182.218.222      330        100        0         53459, 3557 IGP
I           192.5.5.0/24 80.249.208.111       220        100        0         30132, 3557 IGP
I           192.5.5.0/24 193.201.28.50        10010      100        0         27320, 3557 IGP
```

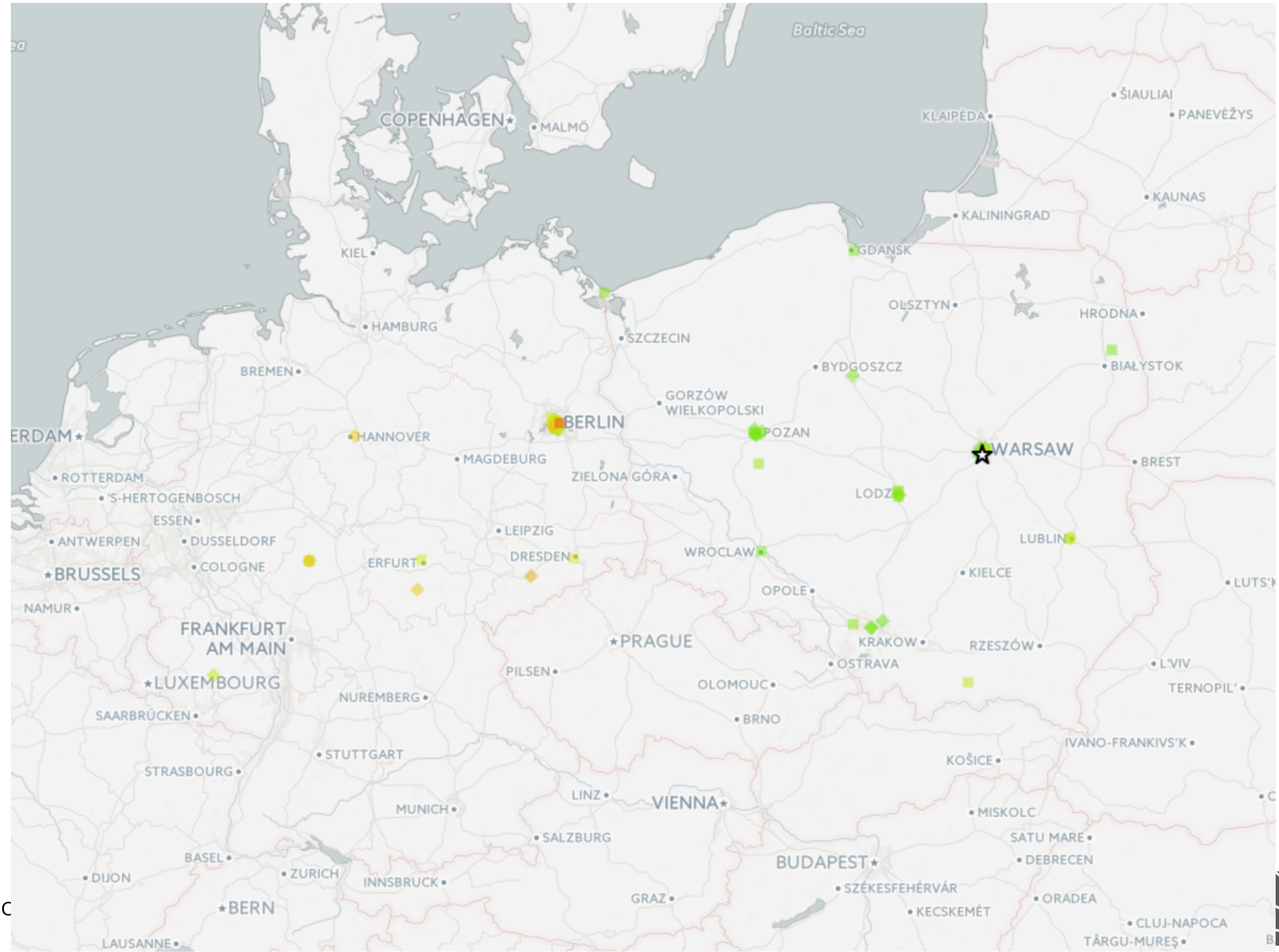
An example of the Santiago, Chile leak from Tier1 network looking glass:

```
192.5.5.0/24 *[BGP/170] 00:01:12, MED 500, localpref 200, from 213.248.64.245
AS path: 27986 6471 33075 3557 I, validation-state: unverified
```

PLIX route server NO_EXPORT

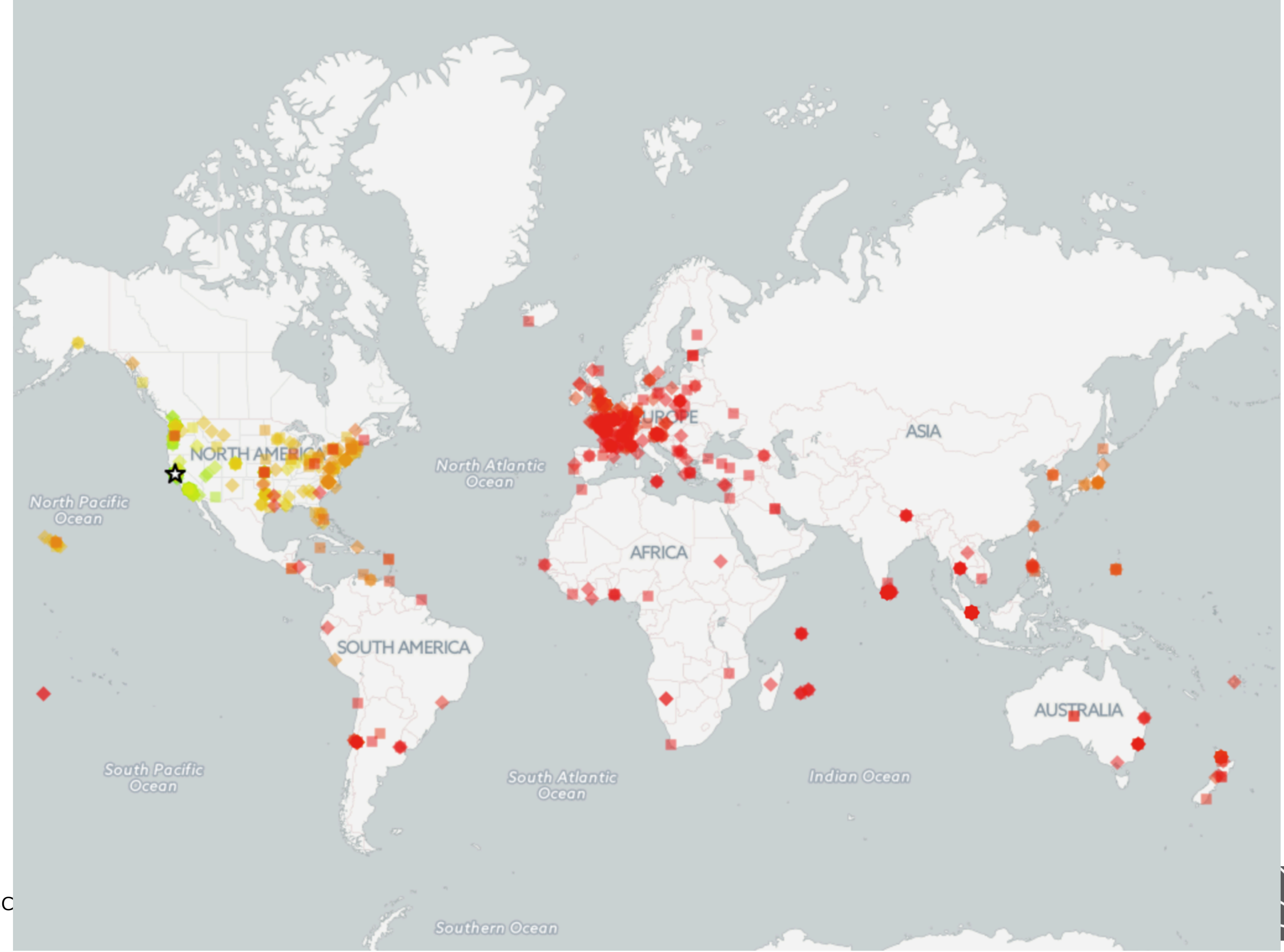


PLIX After

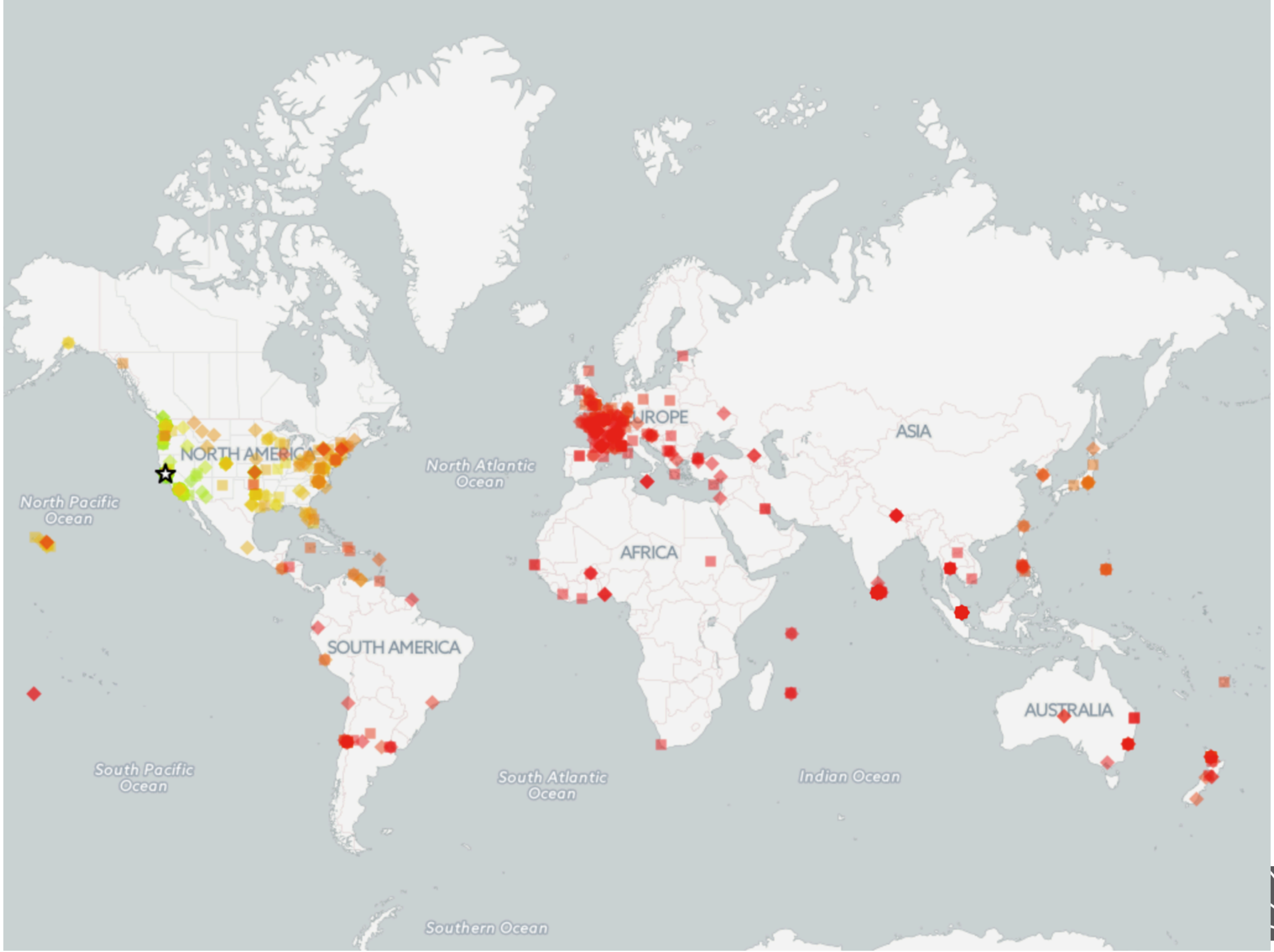


Baltic Sea
COPENHAGEN
MALMÖ
KLAIPĖDA
ŠIAULIAI
PANEVĖŽYS
KAUNAS
KALININGRAD
GDANSK
KIEL
HAMBURG
SZCZECIN
OLSZTYN
HRODNA
BIAŁYSTOK
BREMEN
BYDGOSZCZ
GORZÓW WIELKOPOLSKI
BERLIN
HANNOVER
POZAN
WARSAW
BREST
ERDAM
ROTTERDAM
'S-HERTOGENBOSCH
MAGDEBURG
ZIELONA GÓRA
LODZ
ESSEN
ANTWERPEN
DUSSELDORF
LEIPZIG
LUBLIN
BRUSSELS
COLOGNE
ERFURT
DRESDEN
WROCLAW
KIELCE
LUTS'K
NAMUR
OPOLE
FRANKFURT AM MAIN
PRAGUE
KRAKOW
RZESZÓW
OSTRAVA
L'VIV
LUXEMBOURG
PILSEN
OLOMOUC
TERNOPIL'
NUREMBERG
BRNO
SAARBRÜCKEN
IVANO-FRANKIVS'K
STUTTGART
KOŠICE
STRASBOURG
LINZ
VIENNA
MUNICH
MISKOLC
SALZBURG
SATU MARE
BASEL
BUDAPEST
DEBRECEN
DIJON
ZURICH
INNSBRUCK
SZÉKESFEHÉRVÁR
GRAZ
ORADEA
BERN
KECSKEMÉT
CLUJ-NAPOCA
LAUSANNE
TÂRGU-MUREŞ

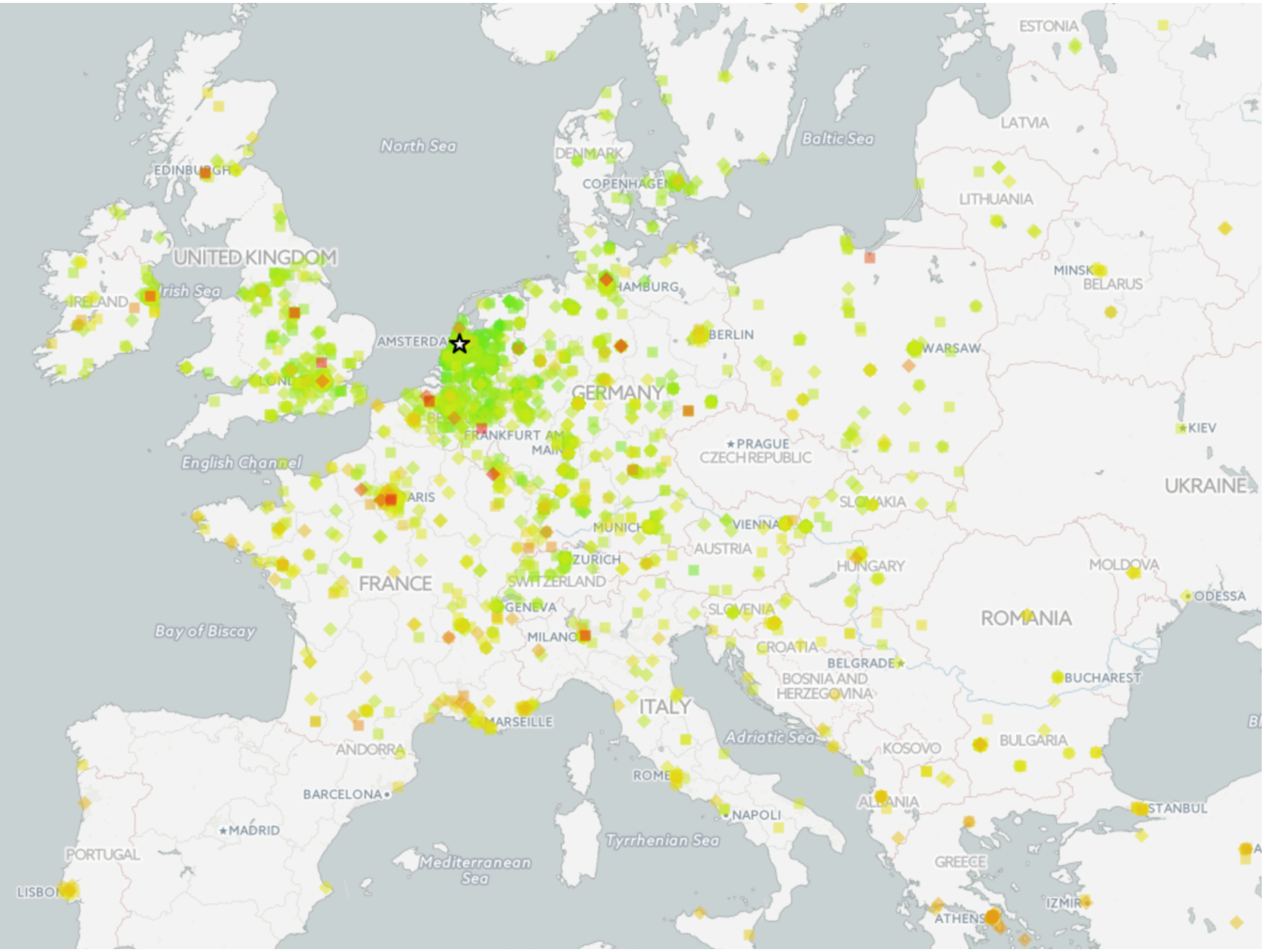
PAO1 over-connected



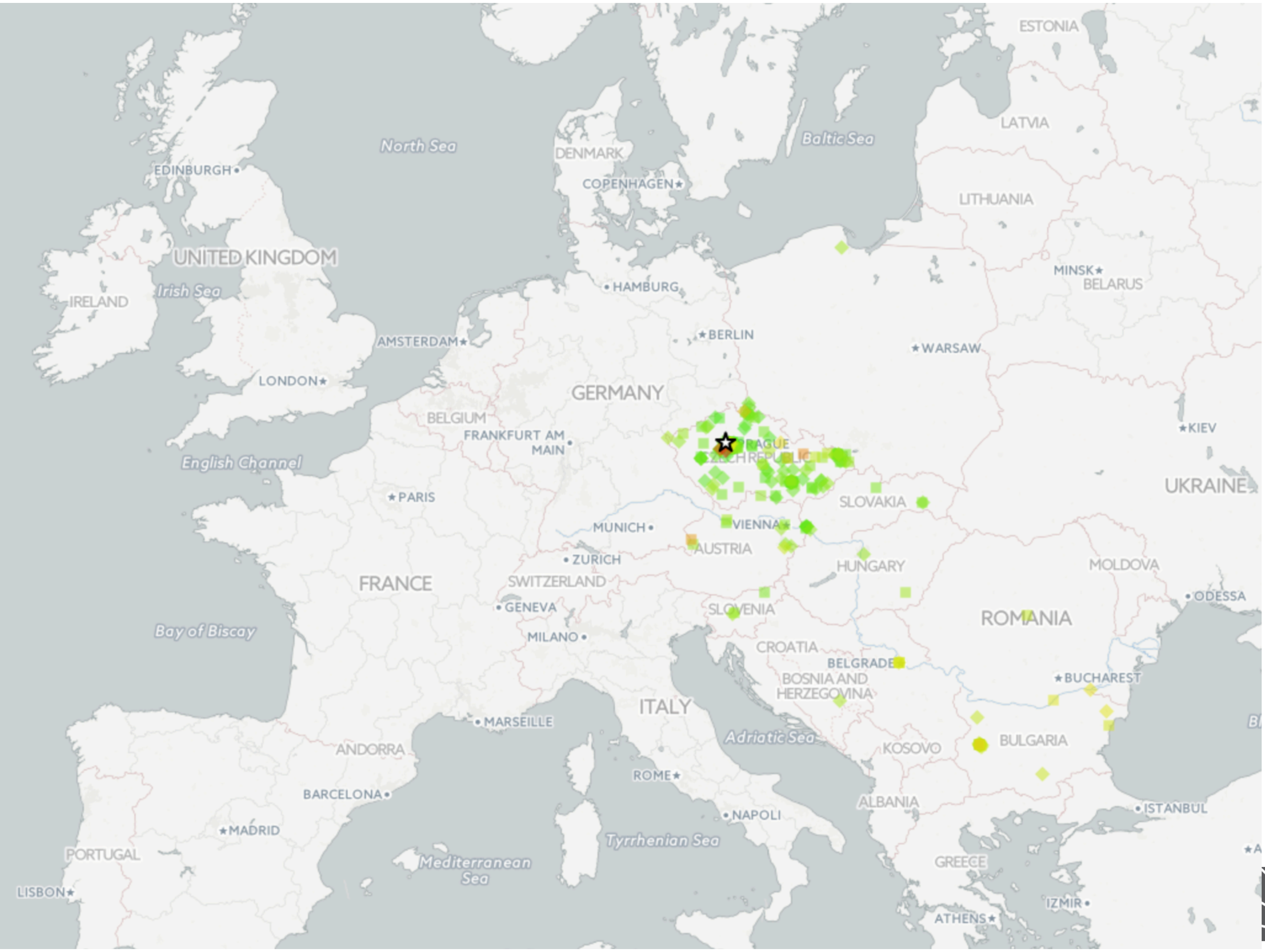
PAO1 after dropping route announcement to international carrier



Amsterdam(AMS1) Global Site



Czech Republic(PRG1)



SNS-PB

- ISC’s authoritative hosting product for public benefit. It is available only to under-served, non-commercial entities, such the top-level domains of smaller countries or territories.
- Uses the other half of the global F-Root Global prefix: 192.5.4.0/24 and 2001:500:2E::/48.
- Customers of SNS-PB operate their own primary name servers where they manage their DNS zone data, and then SNS-PB transfers this zone data to one or more of our globally anycast name server clusters.

ANYCAST

Summarize

Summary

- Anycast is a routing scheme that can be useful when deploying some applications
- There are some protocol level implications that must be considered when designing an Anycast deployment
- DNS is generally well suited to Anycast deployments, and is one of the most popular services to Anycast
- Lots of other folks are doing it, don't be afraid!

For more information

- More details of F-Root setup: http://ftp.isc.org/isc/pubs/tn/isc-tn-2004-1.txt
- Ray Bellis F-Root presentation at UKNOF: https://www.youtube.com/watch?v=FnWOZEmniik&index=9&list=PLjzK5ZtLlc91iPCbC1uf3_Bn0Gol8EnBO
- RIPE ATLAS: https://atlas.ripe.net/
- If you're interested with peering to F-Root please see our peeringdb for locations and contact information: as1280.peeringdb.com

QUESTION AND ANSWER

Poll Question (answer during Q&A session)

Would you like to see another webinar on Anycasting DNS from ISC?