

# Global IP Routing for Kubernetes Clusters

Antonios Chariton





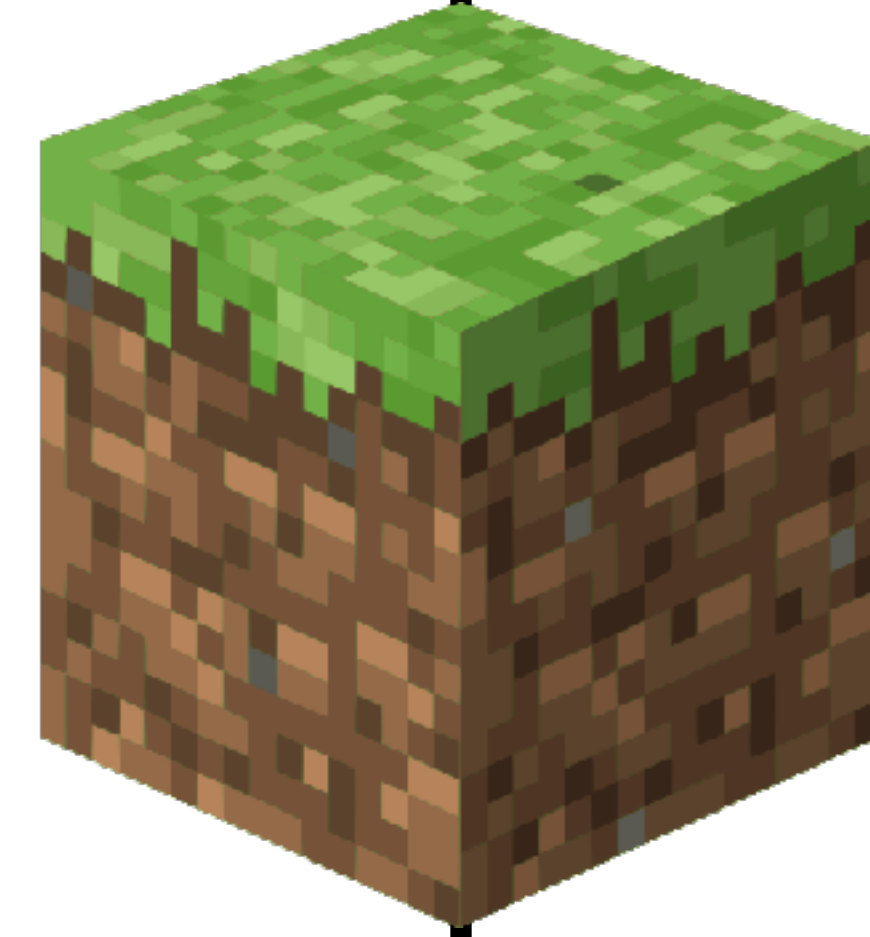
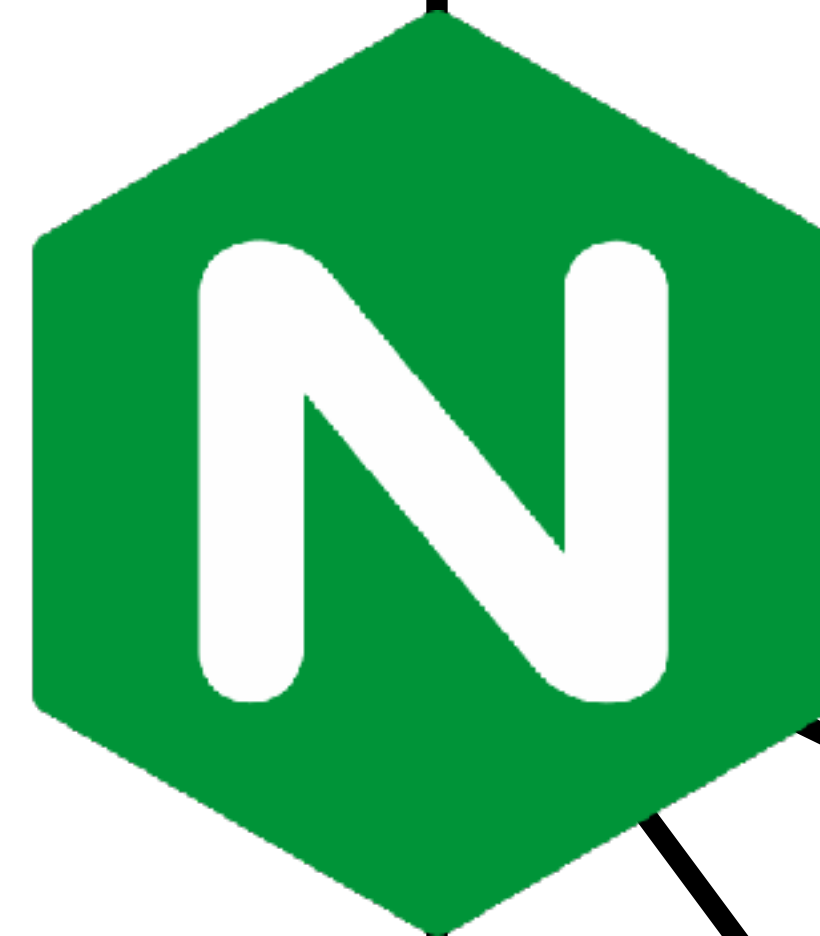
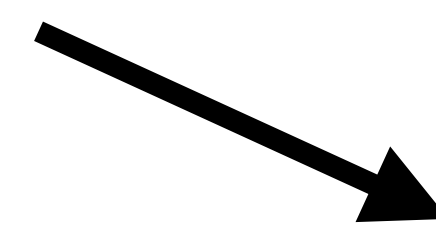
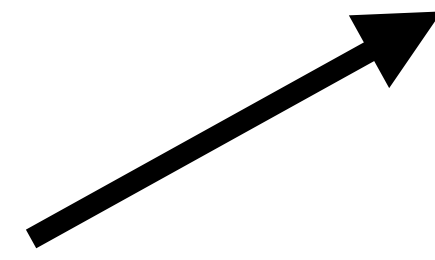
4 TB / 2K  
iDRAC Quick Sync



SAS  
iDRAC Quick Sync









193.5.16.5



193.5.17.3



193.5.18.89





193.5.16.5



193.5.17.3




193.5.18.89





**www IN A 193.5.16.5**  
**www IN A 193.5.18.89**





193.5.16.5



193.5.17.3



193.5.18.89

**www IN A 193.5.16.5**  
**www IN A 193.5.18.89**



193.5.16.5



193.5.17.3



193.5.18.89

**www IN A 193.5.16.5**

**www IN A 193.5.18.89**



193.5.16.5




193.5.17.3





193.5.18.89

**www IN A 193.5.16.5**

**www IN A 193.5.17.3**



193.5.16.5



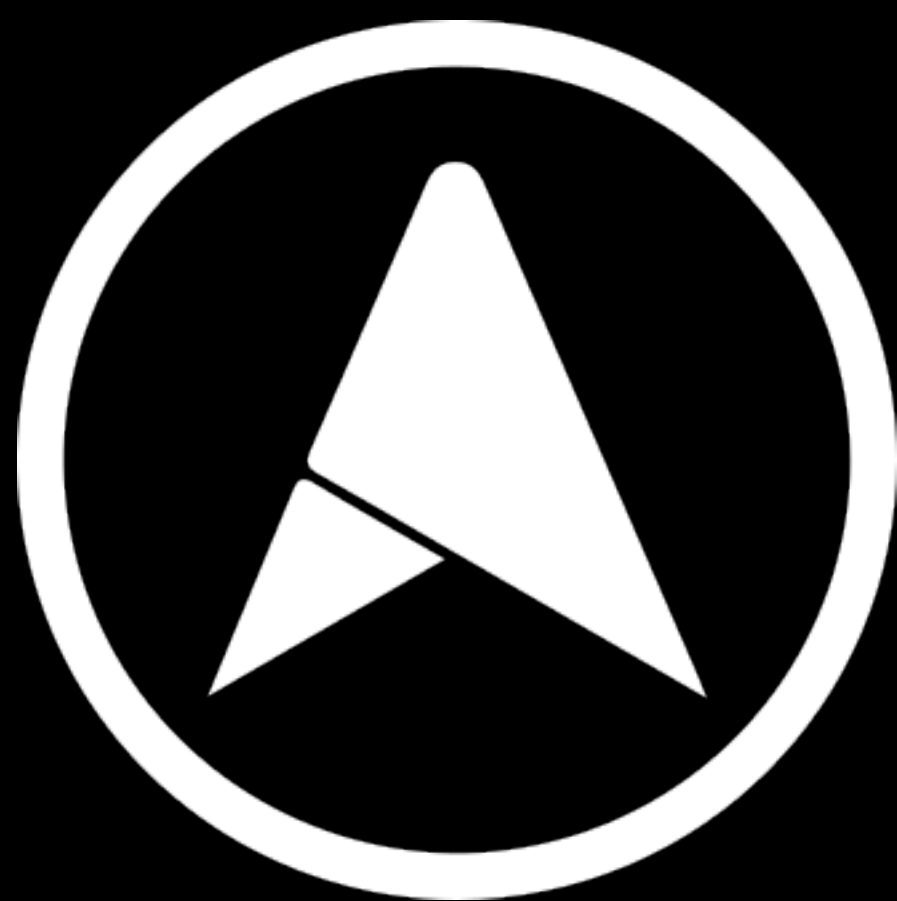
193.5.17.3



193.5.18.89

**www IN A 193.5.16.5**

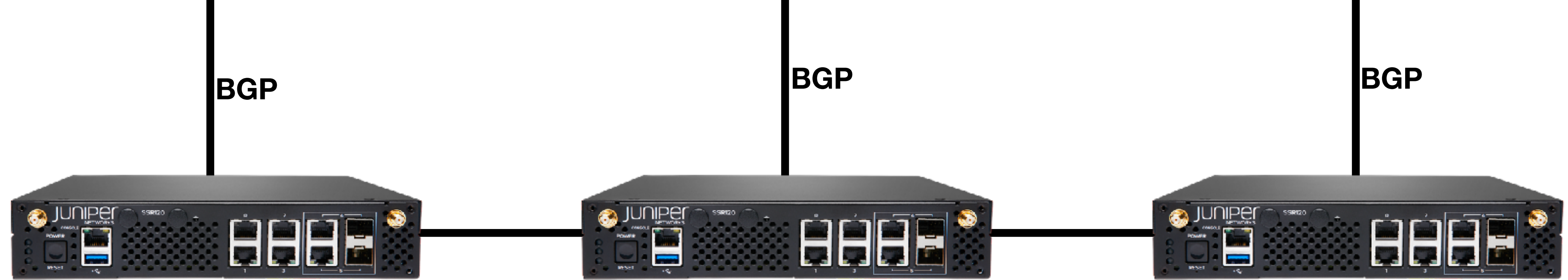
**www IN A 193.5.17.3**



  
**193.5.16.5**  
 

  
**193.5.17.3**  


  
**193.5.18.89**  

---

apiVersion: v1

kind: Service

metadata:

name: exposed-service

annotations:

metallb.universe.tf/loadBalancerIPs: **"193.5.16.64,2a0d:3dc0::16:64"**

spec:

ports:

- port: **443**

targetPort: **443**

selector:

run: **my-nginx**

type: LoadBalancer

ipFamilyPolicy: PreferDualStack

externalTrafficPolicy: Local



193.5.16.5  
193.5.16.64



193.5.17.3



193.5.18.89  
193.5.16.64



BGP

BGP

BGP





**WWW IN A 193.5.16.64**

**WWW IN A 193.5.17.64**

**WWW IN A 193.5.18.64**

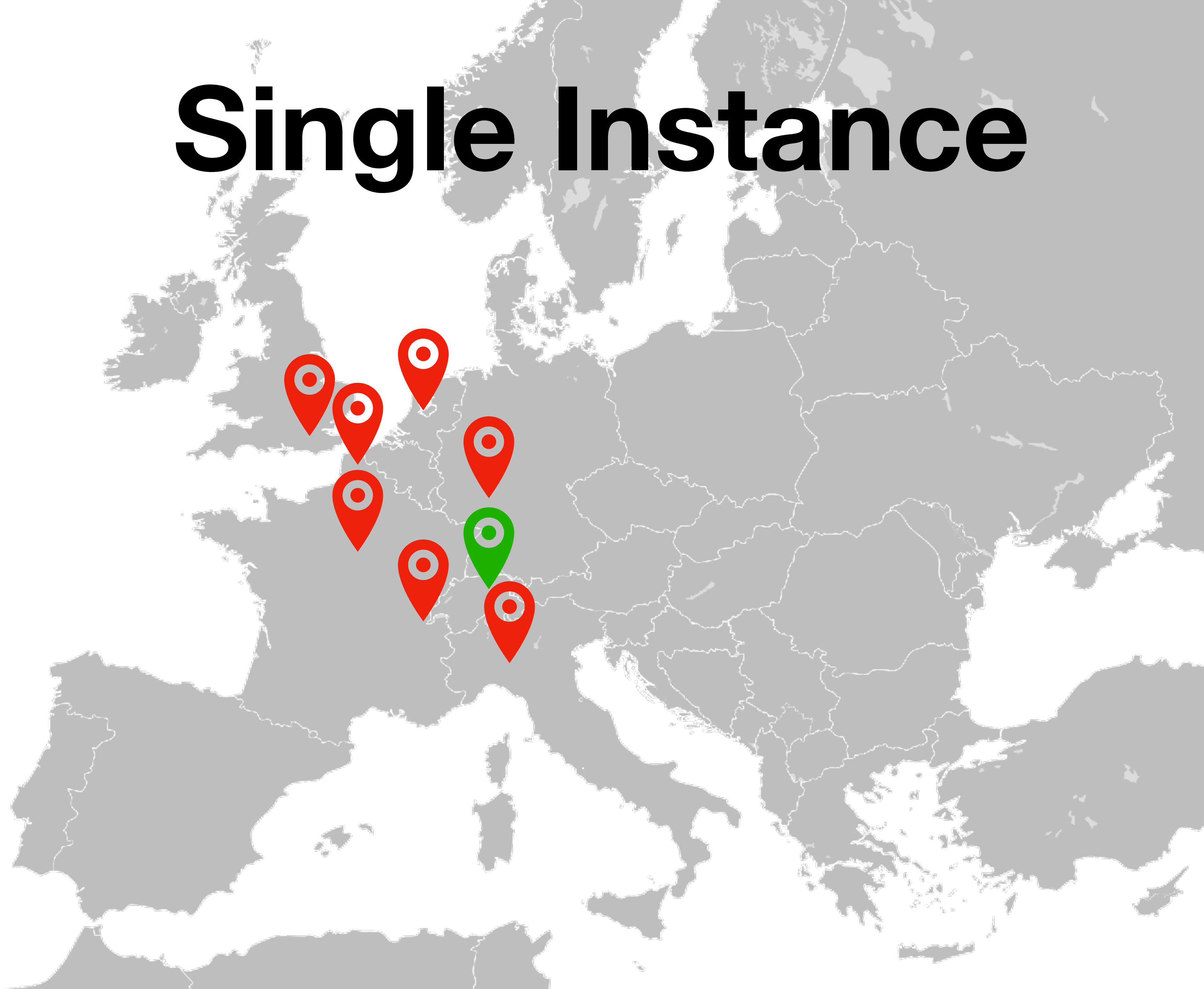
**WWW IN A 193.5.19.64**

# Anycast



**193.5.16.64/32 to all routers**  
**193.5.16.0/24 to all Transits / Peers / ...**

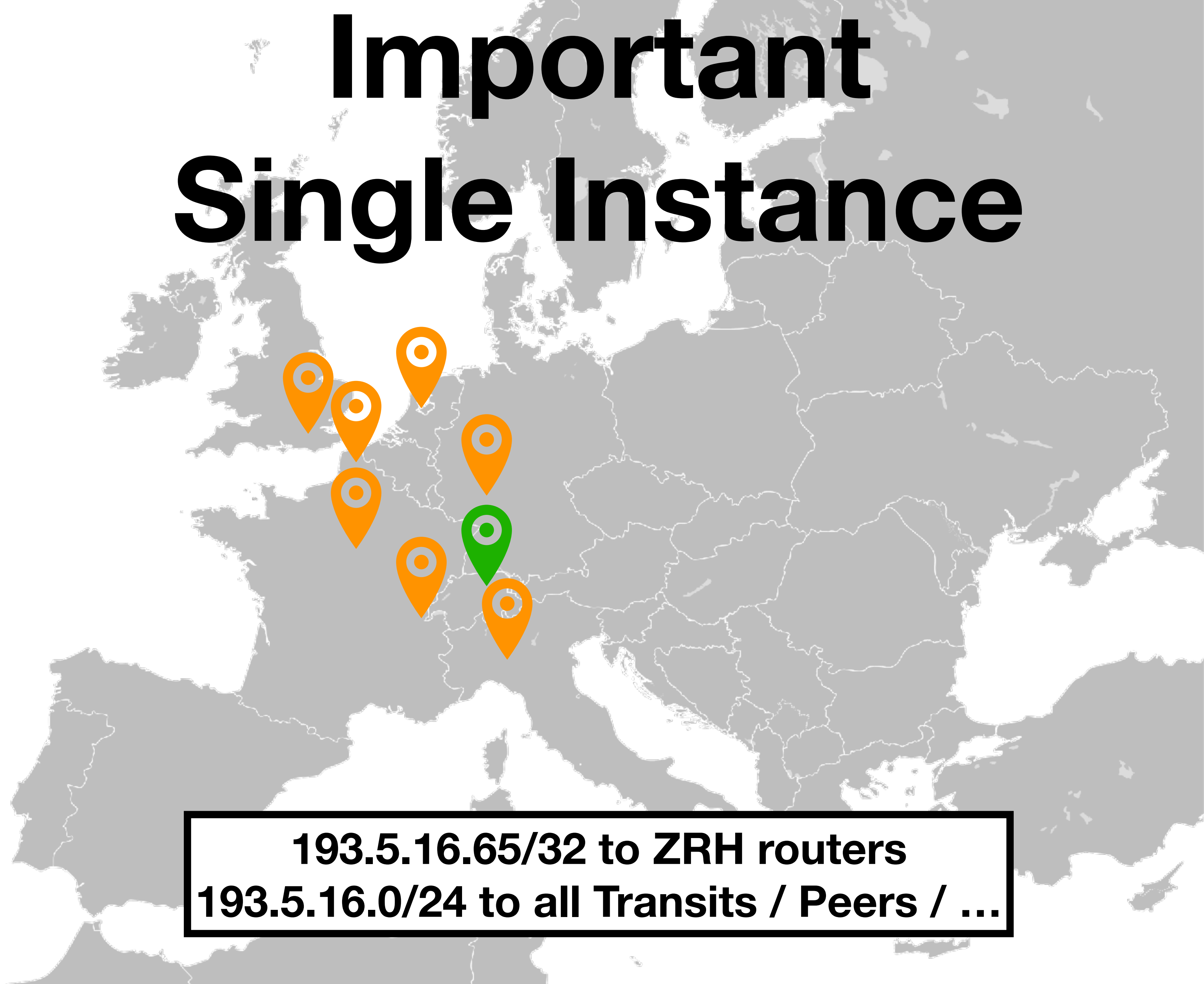
# Single Instance



# Few Instances

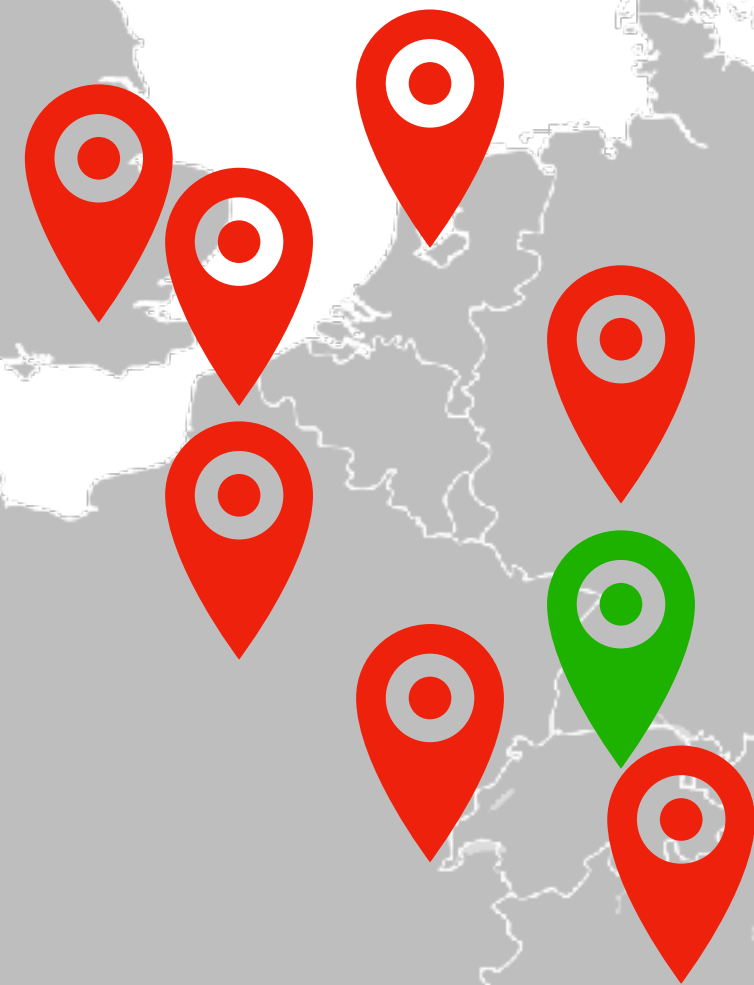


# Important Single Instance



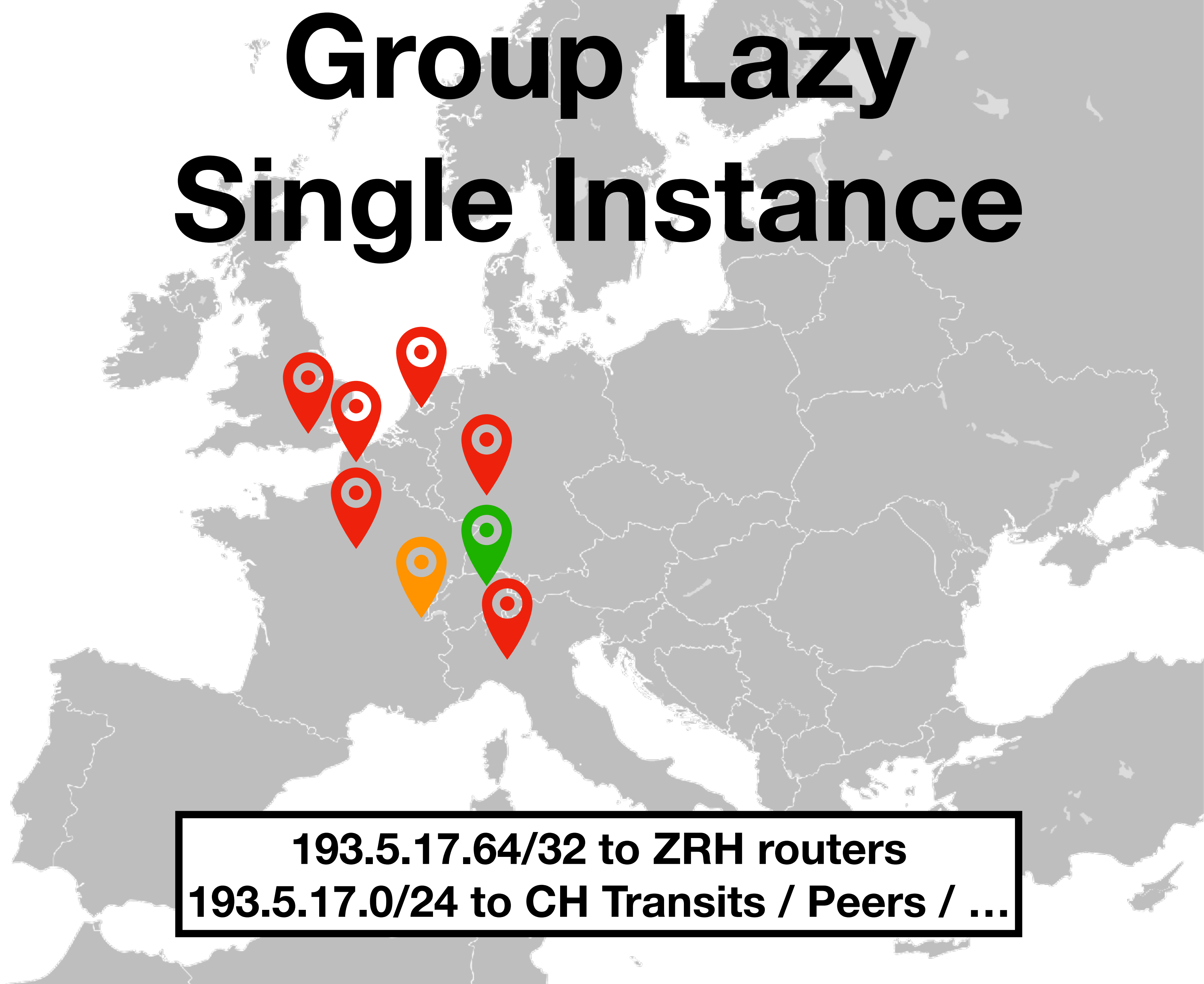
**193.5.16.65/32 to ZRH routers**  
**193.5.16.0/24 to all Transits / Peers / ...**

# Lazy Single Instance



**193.5.17.64/32 to ZRH routers**  
**193.5.17.0/24 to ZRH Transits / Peers / ...**

# Group Lazy Single Instance



**193.5.17.64/32 to ZRH routers**  
**193.5.17.0/24 to CH Transits / Peers / ...**

# IPAM Accounting

- Anycast
  - 1 Prefix, advertised anywhere
- Important Unicast
  - 1 Prefix, advertised anywhere (can be same as Anycast)
- Lazy Unicast
  - N Prefixes, one per group (PoP, Metro, Country, ...)



**But IPv4 is expensive!**

**Then just offer the better  
service over IPv6 only ;)**

# Can DNS Help?

- Move hosts per domain to different group
  - All nginx servers can serve all traffic
  - Anycast VIP -> Europe VIP
  - Zurich VIP -> Switzerland VIP

# Fine-tuned Traffic Engineering

- Network Automation (BGP) can change group policy in  $< 1'$
- DNS Automation can change group in  $\sim \$TTL$  seconds
- Network data can drive decisions (link utilization, etc.)
- Careful of: RPKI, route{6,}, BGP Prefix Limits, Flap Dampening, Stuck Routes (in the IGP)

# Incoming Traffic Volume

- Balance your L3 connectivity
  - A router in Amsterdam / Frankfurt likely won't receive the same traffic as one in Athens (hotspots)
- Plan around cascading failures

**daknob@daknob.gov**  
**@antonis@mastodon.social**  
**linkedin.com/in/daknob/**

**Please replace “.gov” with “.net” in the e-mail address  
above so I can receive your message. :)**