



Онлайн-образование

Отказоустойчивая BGP Anycast сеть с управлением маршрутизацией через механизм Community



Потапов Иван

Product Manager, Radar team

Qrator Labs

План



Что планировалось

1

Получше разобраться, как устроен BGP

2

Получше разобраться, как строятся
и обслуживаются большие сети

3

По работе занимаюсь аналитикой,
основанной на данных от BGP

4

Не занимаюсь настройкой сетевого оборудования.
До обучения не слышал про некоторые протоколы.

5

Проект занял около 4-5 вечеров

Используемые технологии

1

BGP Anycast – для построения сети

2

BGP Communities – для traffic engineering

3

Техника Blackhole – для защиты от DDoS

4*

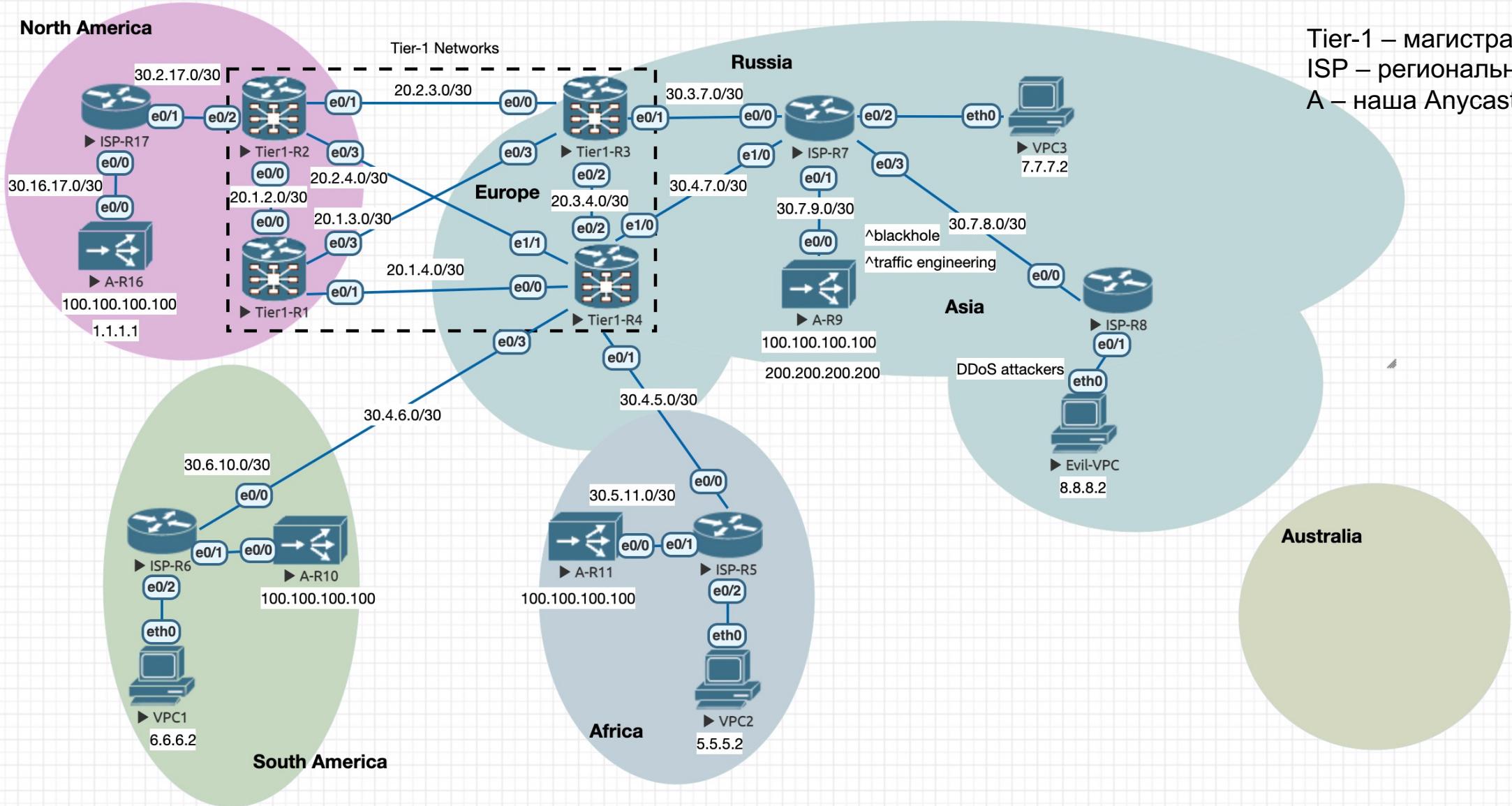
RIPE Atlas пробы –
для поиска Anycast в живой природе

5*

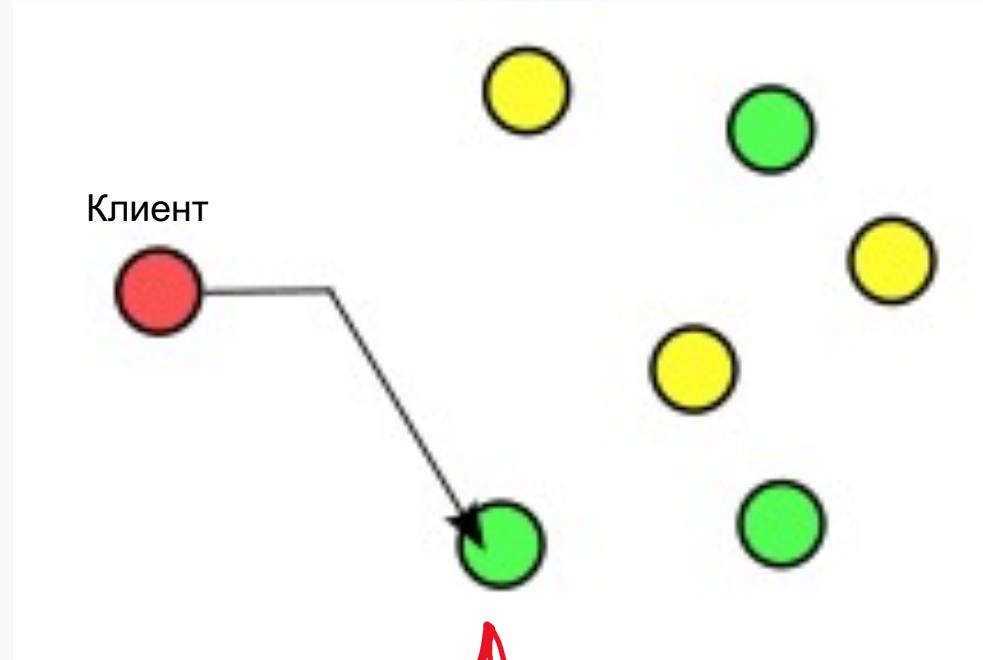
DNS round robin, GeoDNS –
технологии балансировки, “случайно”
замеченные, используемые рядом с Anycast



Схема

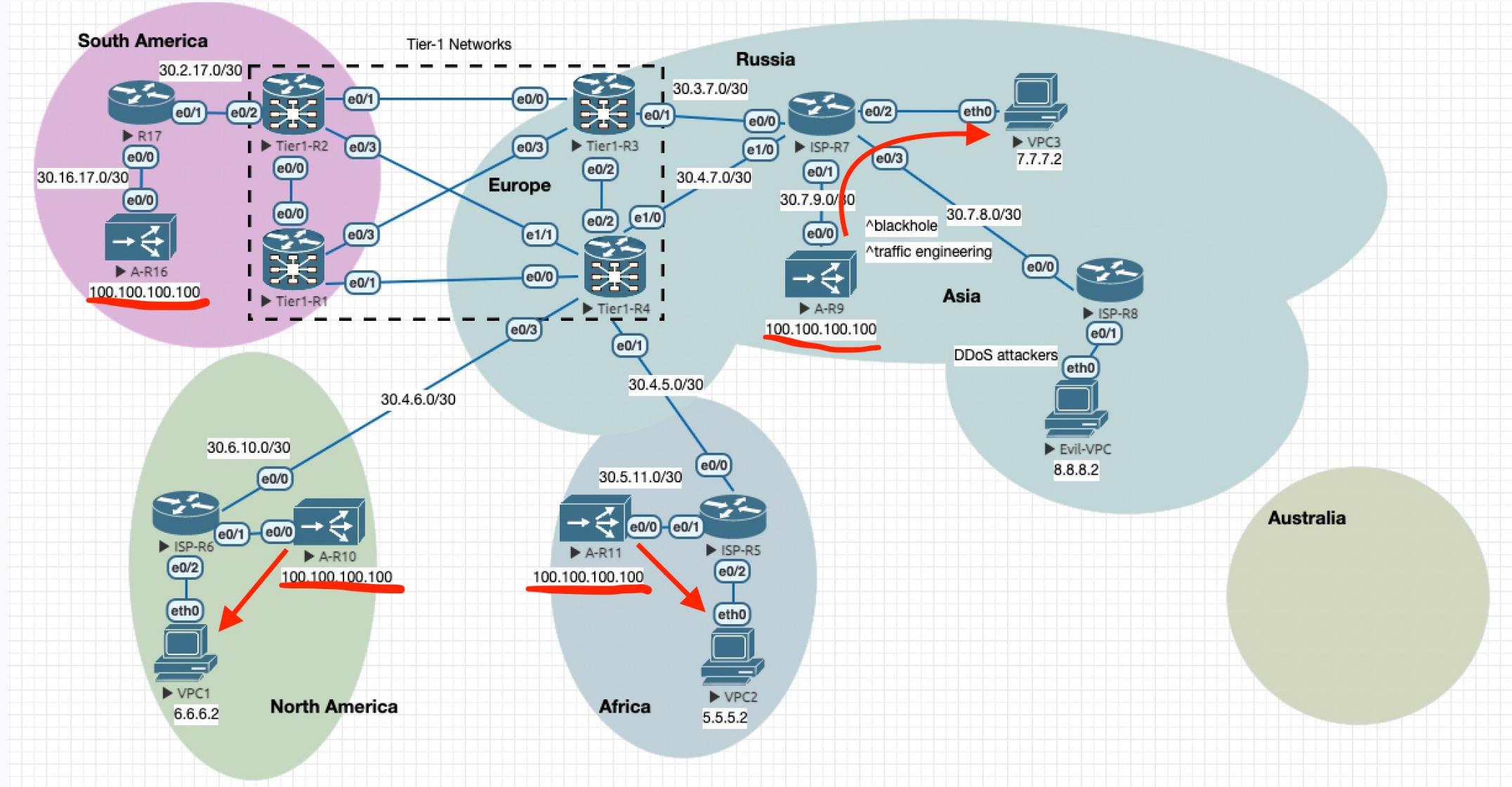


BGP Anycast



Один IP адрес используется на разных устройствах и анонсируется из разных, обычно географически распределенных, мест

BGP Anycast: устойчивость и балансировка



Demo

BGP Communities

bgp.update.path_attribute.community						
No.	Time	Source	Destination	Protocol	Length	Info
9	22.416817	30.7.9.2	30.7.9.1	BGP	194	UPDATE Message,
> Frame 9: 194 bytes on wire (1552 bits), 194 bytes captured (1552 bits) on interface -, id 0						
> Ethernet II, Src: aa:bb:cc:00:90:00 (aa:bb:cc:00:90:00), Dst: aa:bb:cc:00:70:10 (aa:bb:cc:00:70:10)						
> Internet Protocol Version 4, Src: 30.7.9.2, Dst: 30.7.9.1						
> Transmission Control Protocol, Src Port: 21800, Dst Port: 179, Seq: 24, Ack: 1, Len: 140						
> Border Gateway Protocol - UPDATE Message						
Marker: ffffffffffffffffffffff						
Length: 62						
Type: UPDATE Message (2)						
Withdrawn Routes Length: 0						
Total Path Attribute Length: 34						
> Path attributes						
> Path Attribute - ORIGIN: IGP						
> Path Attribute - AS_PATH: 100						
> Path Attribute - NEXT_HOP: 30.7.9.2						
> Path Attribute - MULTI_EXIT_DISC: 0						
> Path Attribute - COMMUNITIES: 100:3						
> Flags: 0xc0, Optional, Transitive, Complete						
Type Code: COMMUNITIES (8)						
Length: 4						
> Communities: 100:3						
> Network Layer Reachability Information (NLRI)						

Это атрибут update сообщения. Получивший его роутер может применять разные политики маршрутизации, в зависимости от комьюнити

BGP Communities: Example

```
~ whois AS9002 | grep "2. Regional, prepending to peers" -A20  
remarks:  
remarks:  
remarks: 64666:6553X  
remarks: X < 5      prepend X times to all peers in Europe  
remarks: X = 5      advertise to customers but not to peers in EU  
remarks:  
remarks: 64777:6553X  
remarks: X < 5      prepend X times to all peers in North America  
remarks: X = 5      advertise to customers but not to peers in NA  
remarks:  
remarks: 64888:6553X  
remarks: X < 5      prepend X times to all peers in Asia-Pacific  
remarks: X = 5      advertise to customers but not to peers in APAC  
remarks:  
remarks: Note #2. If downstream tags the route with regional community,  
remarks: then common community 9002:6553X is ignored in proper Region  
remarks:  
remarks: 3. Specific, prepending to peer ASN:  
remarks: ASN:6553X  
remarks: X < 5      prepend X times to peer ASN
```

BGP Communities: Map

Реализация следующих политик маршрутизации:

Ограничительные:

- NO_EXPORT (well known)
- NO_ADVERTISE (well known)

Манипуляция трафиком:

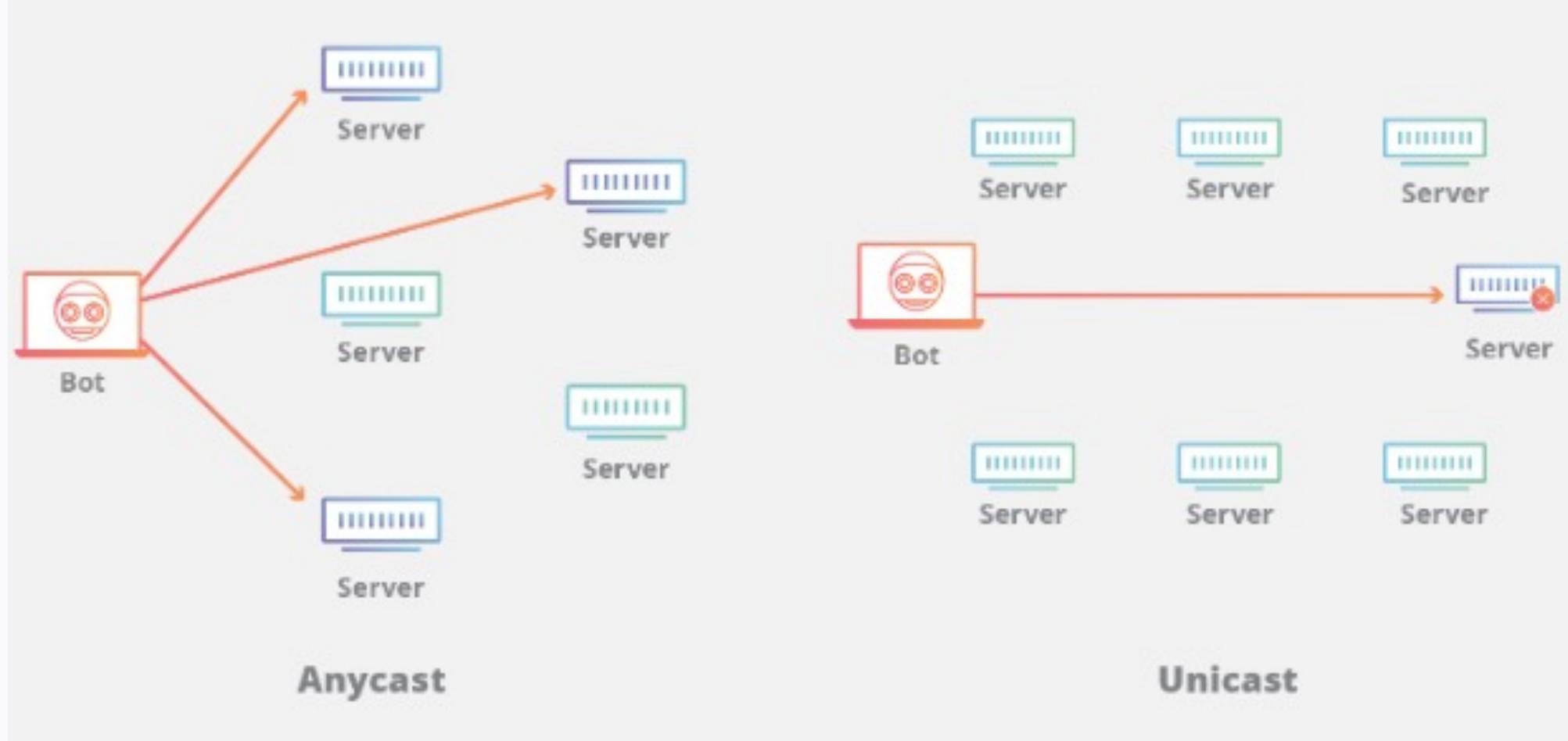
- PREPEND1: 100:1
- PREPEND3: 100:3
- PREPEND5: 100:5

Защита от DDoS:

- BLACKHOLE: 65535:666

BGP Communities: Demo

BGP Anycast: защита от DDoS

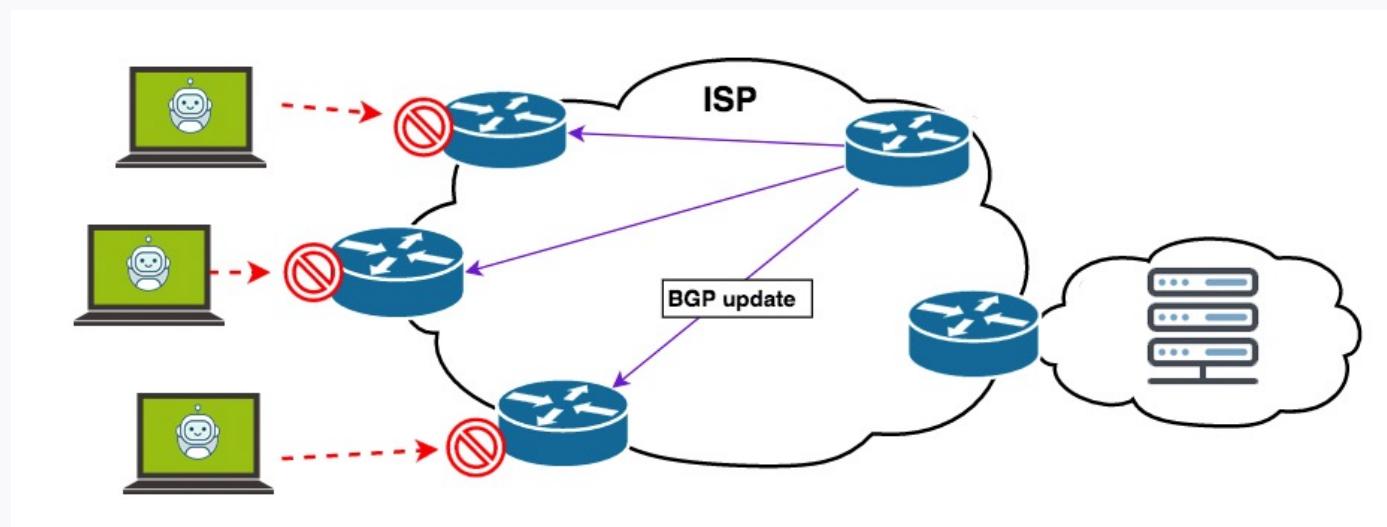


BGP BLACKHOLE: защита от DDoS

Blackhole применяется для защиты ресурсов во время DDoS атаки

Схема:

- Клиент просит провайдера запретить трафик на атакуемый ресурс
- У провайдера (ISP) срабатывает политика, которая маршрутизирует трафик в Null0 интерфейс



Профит:

- разгрузка емкости канала
- спасение других сервисов в префиксе

BGP BLACKHOLE: основные команды

Client:

```
Router bgp 100
network 100.100.100.100 mask 255.255.255.255
route-map blackhole

neighbor 30.7.9.1 send-community
```

```
route-map blackhole permit 10
set community 65535:666
```

Provider:

```
router bgp 7
neighbor 30.7.9.2 route-map customer-in in
ip community-list 10 permit 65535:666

route-map customer-in permit 10
match community 10
set community no-export
set ip next-hop 172.16.6.6

ip route 172.16.6.6 255.255.255.255 Null0
```

BGP Anycast in the wild

Попробуем найти сайты, которые работают по
технологии BGP Anycast

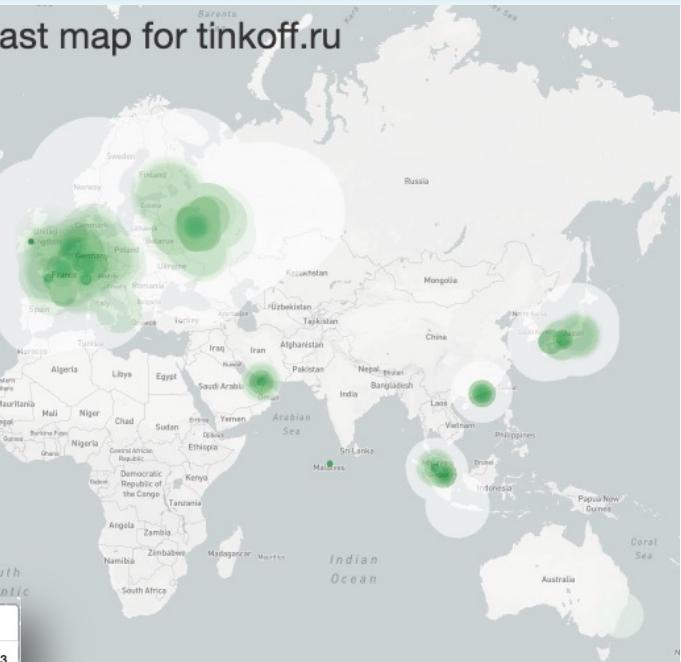
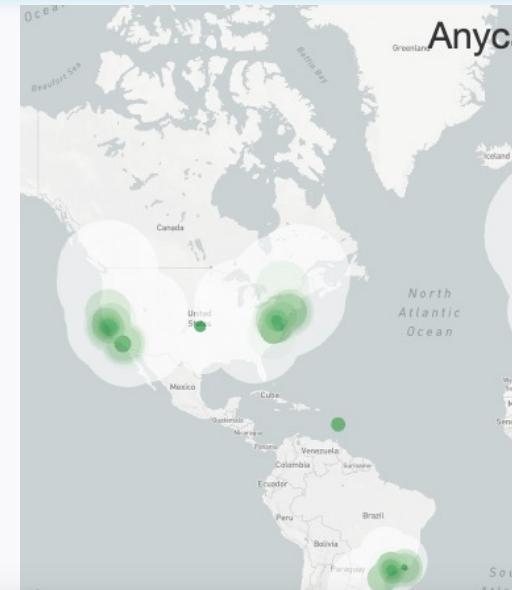
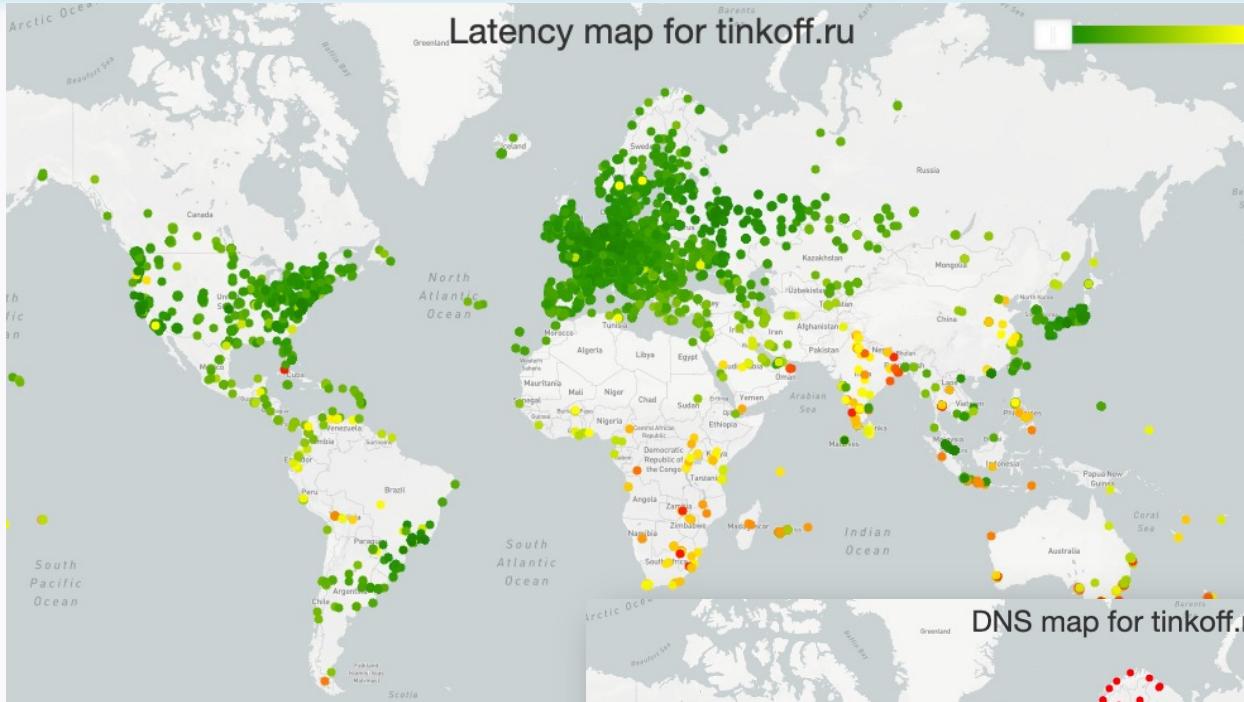


BGP Anycast in the wild

Для этого будем измерять RTT
от разных source IP до исследуемых сайтов,
при помощи проб **RIPE Atlas**.

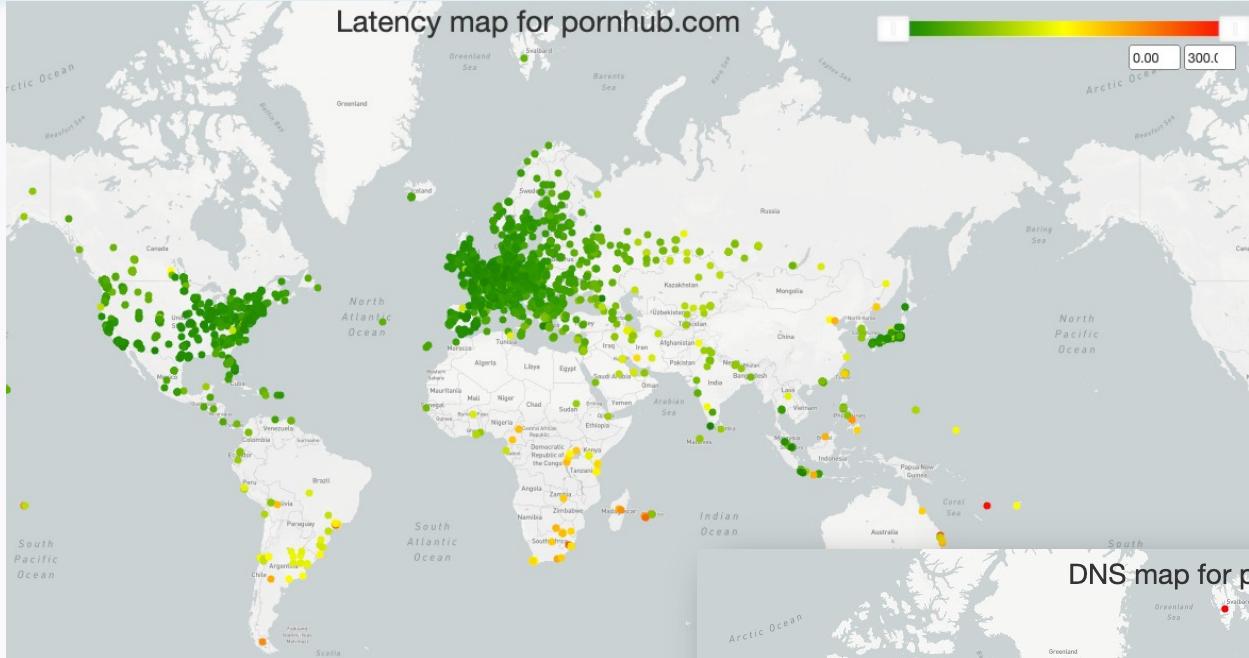
Если RTT мало, то значит анонсирующая точка
рядом

BGP Anycast: примеры

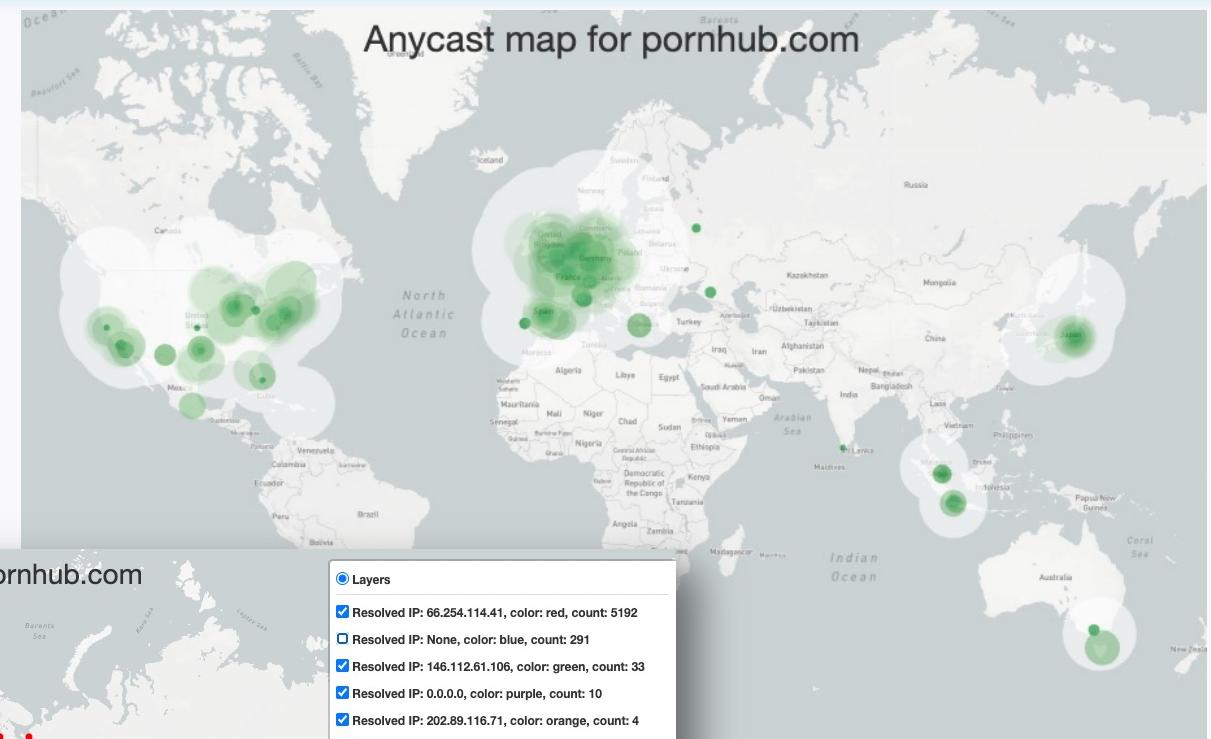


BGP Anycast: примеры

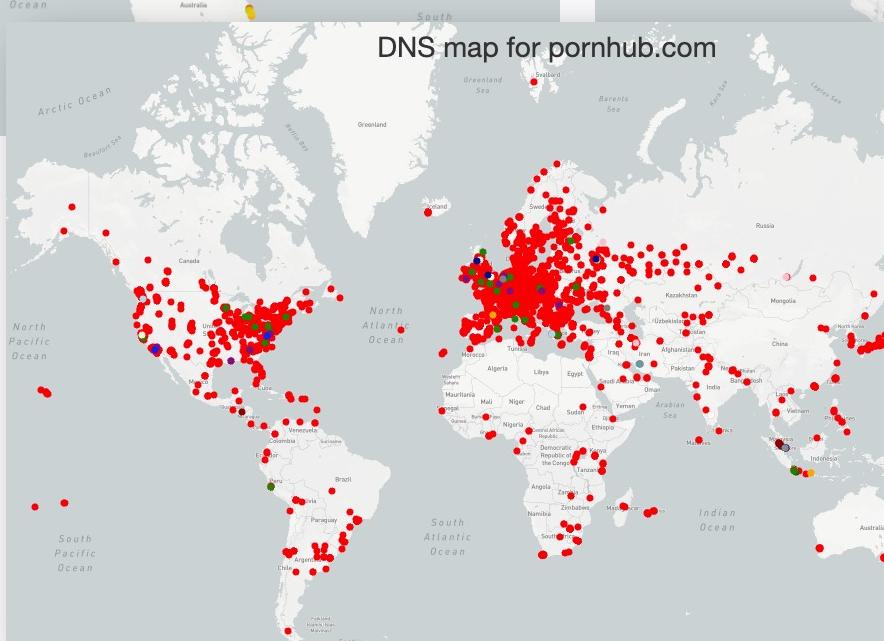
Latency map for pornhub.com



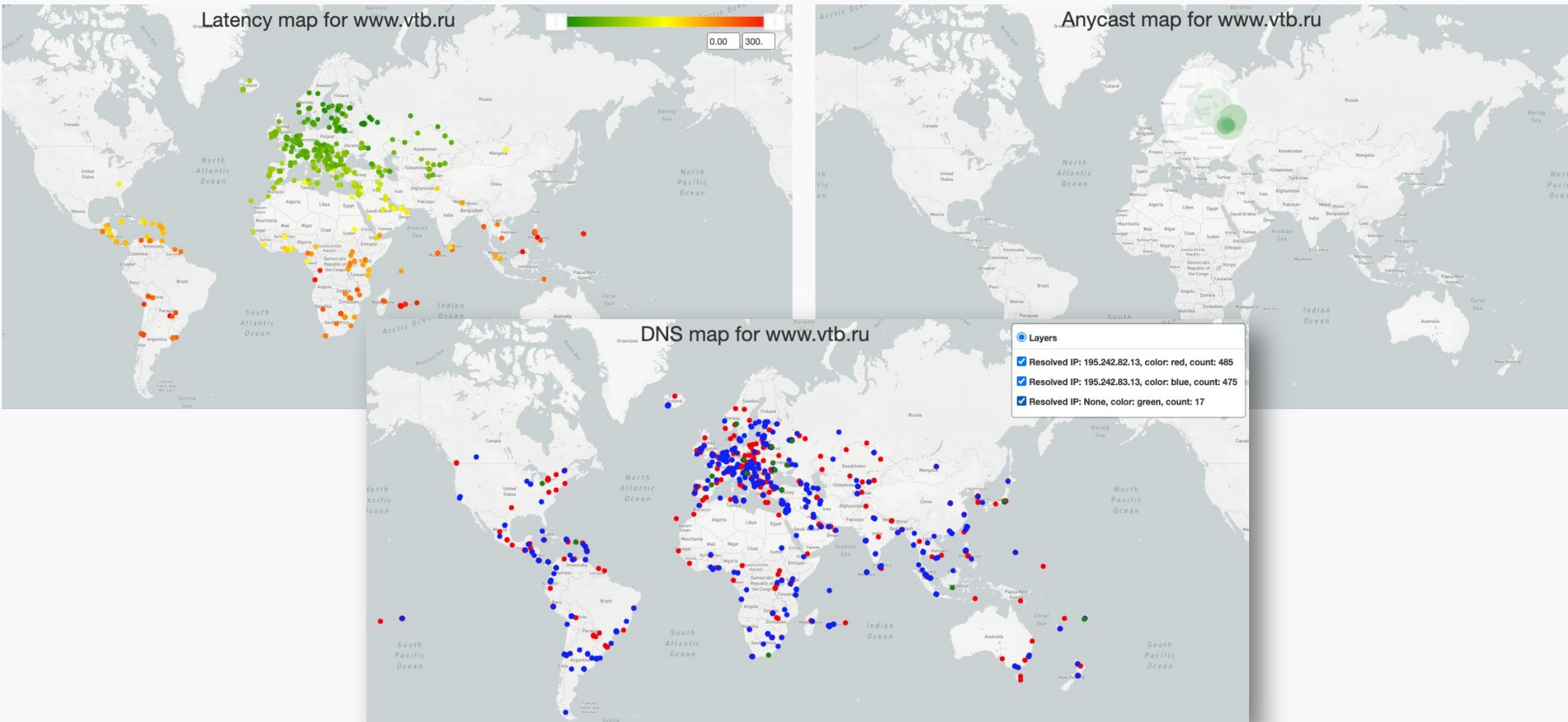
Anycast map for pornhub.com



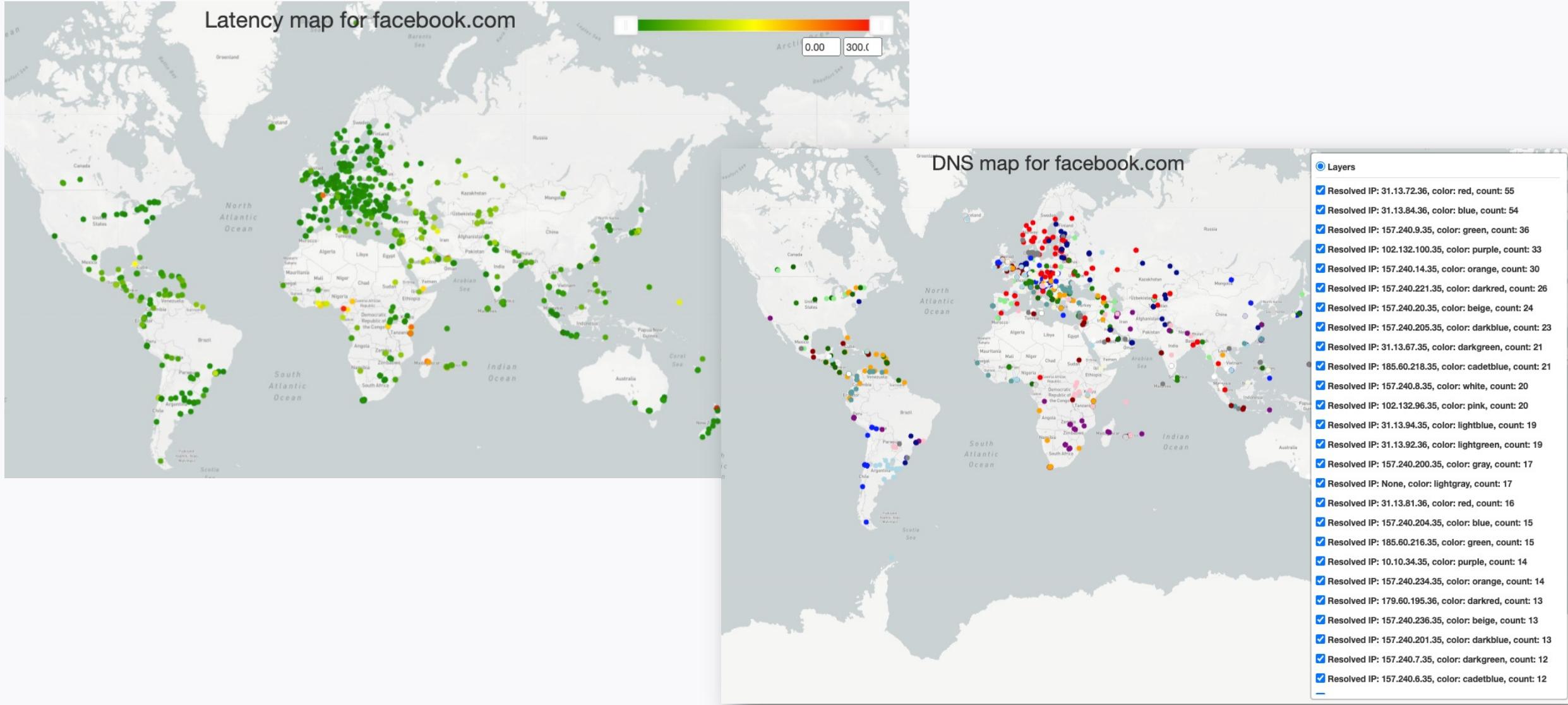
DNS map for pornhub.com



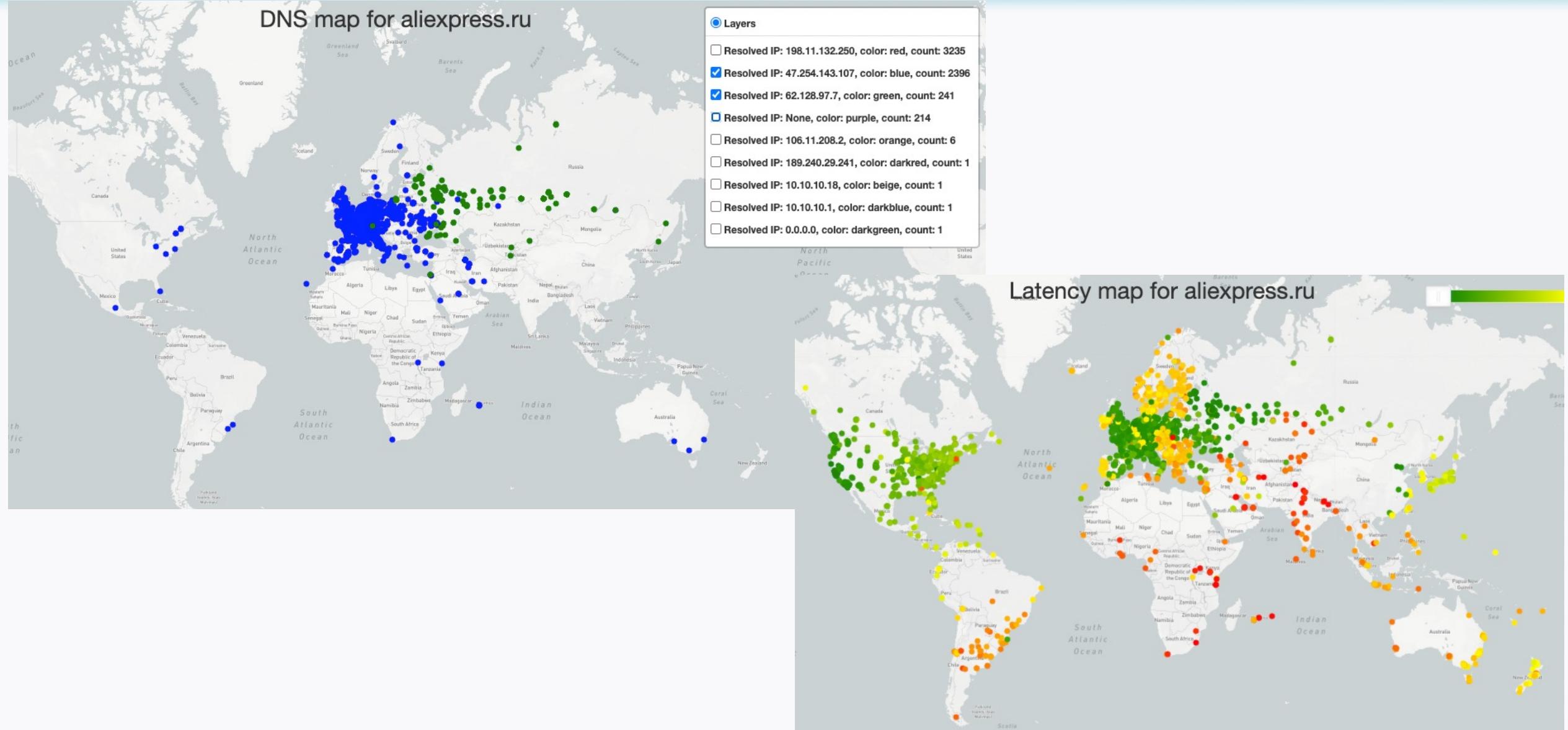
DNS Round Robin



DNS Round Robin?



Балансировка Geo DNS



Вывод и планы по развитию

- Большие сети это сложно, но интересно
- Не говорили о том, как эффективно строить распределенные приложения (кэши, внутренняя сеть/туннели, БД, синхронизация, фронт, бэк)
- Могут возникать разного рода аномалии маршрутизации, нужен специальный внешний мониторинг для облегчения траблшутинга



Спасибо за внимание!



Потапов Иван
Qrator Labs, Radar team
ip@qrator.net
gervold@mail.ru, github.com/gervold

BGP Anycast

