Stairway to Anycast

Or Highway to Anycast, depending on your viewpoint

Jan Žorž and Sander Steffann – 6connect Inc.

A road trip

Our starting point

The parking lot

- 6connect is a global company
- Our DNS platform should be global too!
- The best way to scale DNS globally is by using anycast

- Mentioning this to our commercial colleagues resulted in "oh, we may have some customers for that…"

How to do that?

The roadmap

- Build a prototype
- Set up measurements
- Fine-tune the prototype
- More measurements
- Done!
- Right?

Building the prototype

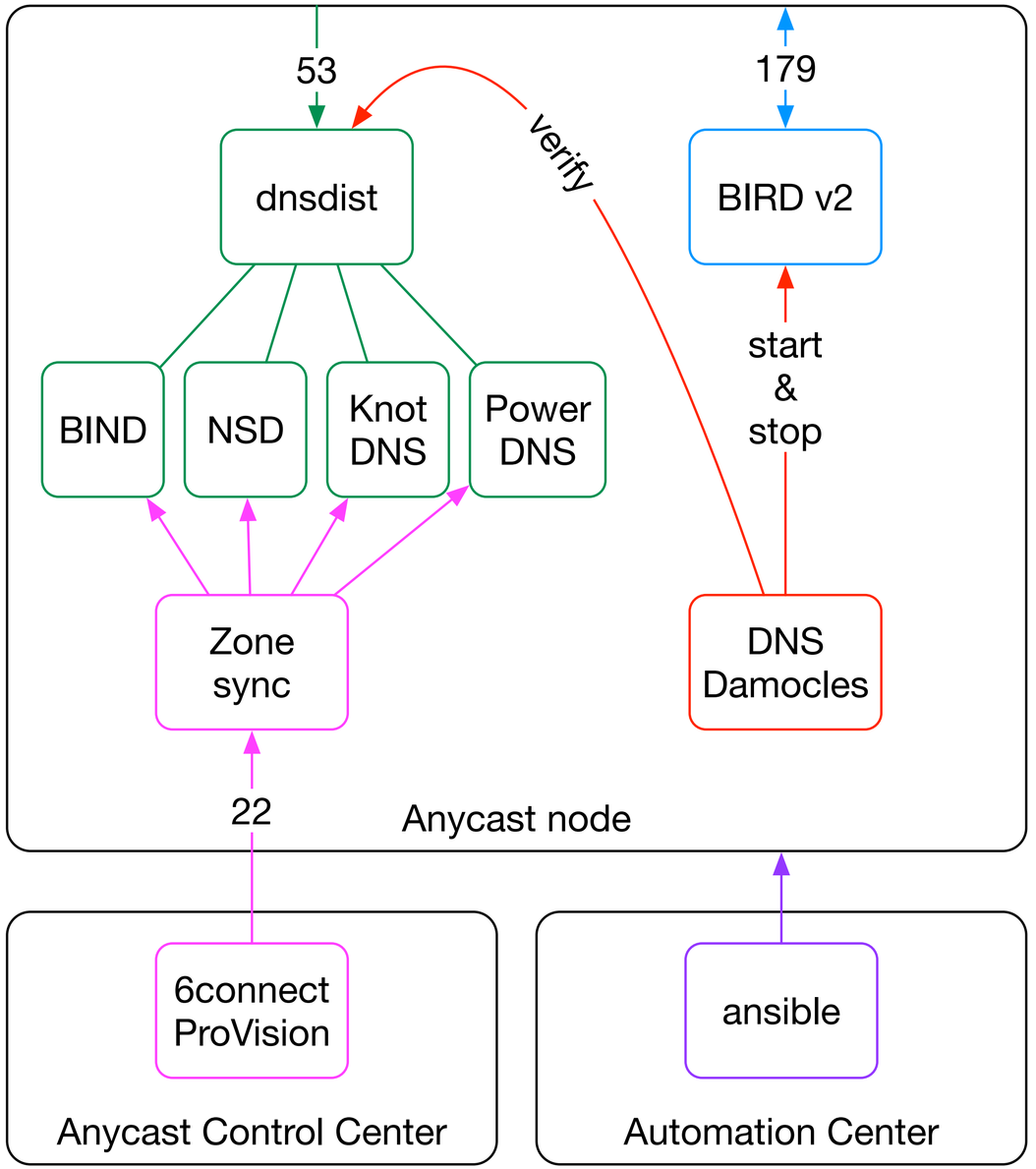
Start with building the car

- Software choices
  - BIND, Knot DNS, NSD, PowerDNS, which one?
  - All of them! Let's use dnsdist
  - Bird2 for BGP routing
  - Ansible for automation / rollout
  - bash/sed/awk for scripting!

Designing a node

The Haynes™ Manual

- Dnsdist provides scripting and monitoring
- Zone sync: Python script to update zone files
- Damocles: Bash script to query dnsdist and kill BIRD on failure
- Managed using Ansible



The first nodes

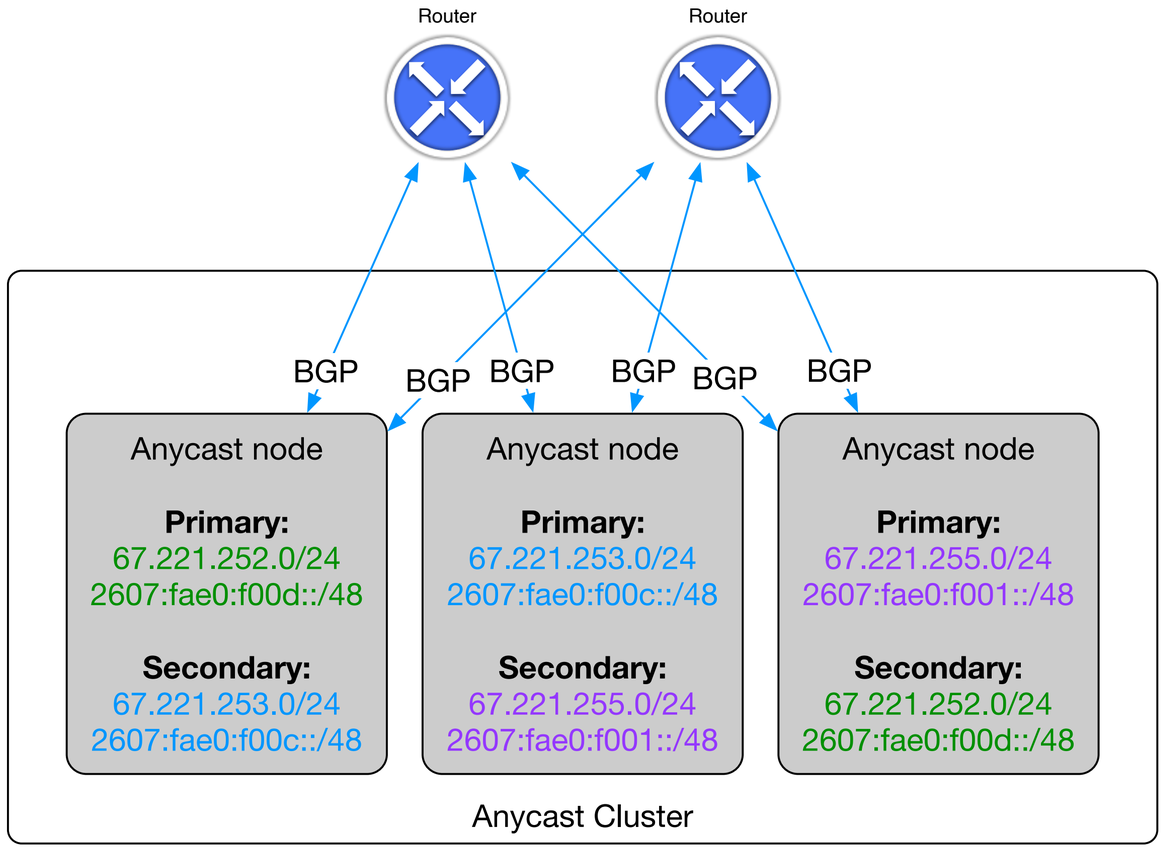
Finding the on-ramp

- 6connect clusters: Fremont (US), Ljubljana (SI) and Apeldoorn (NL)
- Not a really good spread, but it's a start
- Which IP resources to use?
  - How many IPv4/IPv6 prefixes?
  - Which AS numbers?
- We decided on one ASN announcing 3x /24 IPv4 and 3x /48 IPv6
- 3 nodes, each announces primary IPv4+IPv6 + secondary IPv4+IPv6

Cluster composition

Turning on the radio

- Anycast nodes announce
  - Primary prefix with high priority
  - Secondary prefix with low priority
- Method depends on relationship to the routers:
  - iBGP uses local-pref
  - eBGP uses prepending



The need for measurements

Unexpected potholes

- It doesn't seem to work as well as expected
- But why/how/when/where/?!?!?!?!
- We need monitoring and measurements
- Route views help a bit
- RIPE Atlas provides some information, but not very detailed

In the mean time

A little detour

- While we are thinking about measurements, let's add more things!
  - We deployed a set of VMs in Tokyo (JP) using Vultr
  - Added them to the anycast setup
- Moved our 6clabs.com domain to anycast
  - Eat your own dog food…
  - What could possibly go wrong?

Our own control and monitoring

I think our car needs a speedometer

- We use 6connect ProVision as the control center for anycast DNS
- Zones are administered and distributed from here to all anycast DNS servers

- We use LibreNMS to keep track of dnsdist queries, performance and uptime
- We should also measure each backend (TODO)

More anycast ideas

The scenic route

- Our initial prototype is authoritative DNS
- We can also do recursive DNS, should be easy
- We also offer a cloud-hosted IPAM, can we anycast that?
  - We'd need a replicated DB (Galera?)
- Having a high-available mail service would be nice
  - Proxmox Mail Gateway as a spam filter
  - Dovecot (dsync?) for replicated mailbox storage

Design decisions

Where is a sat-nav when you need it…?

- Not all services need to be in all anycast locations
- How many services can we host in one /24 & /48?
  - If one service fails the whole prefix needs to be pulled out
- HAProxy or nginx as the front-end anycasted load balancer
  - If the load balancer is the only anycasted service this is a lot easier
  - If a local service fails the load-balancer can send the traffic to another site

Getting TLS certificates

Did you bring your passport?

- Anycasted services need a certificate for the distributed hostname
  - Using Let's Encrypt is more complicated than usual
  - We don't know where the ACME verification is going to be received
- The load-balancer can send all ACME traffic to a central node
  - This node can get the certificate
  - And distribute the keys and certificates to all relevant anycast nodes
  - This needs to be built…

The dilemma stays

Hello?!? Can anybody hear us? Please tell us where we are…

- Are we globally visible?
- Where are the black holes?
- Which networks are sending traffic to which site?
  - Is Asia sending traffic to US?
  - Is Europe sending traffic to Asia?
  - Is traffic going to the nearest site?
- Are we making our users take large detours?

Visiting Remco van Mook

Stopping for a drink

- We had a nice chat
- Turns out Remco is starting a company called Lynkstate…
  - Which specialises in measuring network reachability, including anycast-specific measurements!
- The whisky was nice too…

What we want out of monitoring

The dials we need on the dashboard

- Where are our prefixes visible?
  - How are we visible from around the world?
- Which announcement do users see?
  - In other words: which anycast cluster do users use?
  - Are clients using the closest node?
  - What is the latency from each user to "their" anycast cluster?
- Where are the black holes?
  - Which ISPs do we need to talk with?

We need a good view to make it better…

At the end of our road trip we want everything green!