

Building an Open Source Anycast CDN

Nate Sales



About me

Nate Sales

- Highschool student
- Interested in software development and networking
- AS34553
- Wanted to experiment with content delivery and increase reliability for my public facing services



Overview

- Anycast – multiple next hops for a route
- Asynchronous, fast updates
- Learning experience
- All content at the edge, no backbone
- Everything open source
- Student budget – as cheap as possible

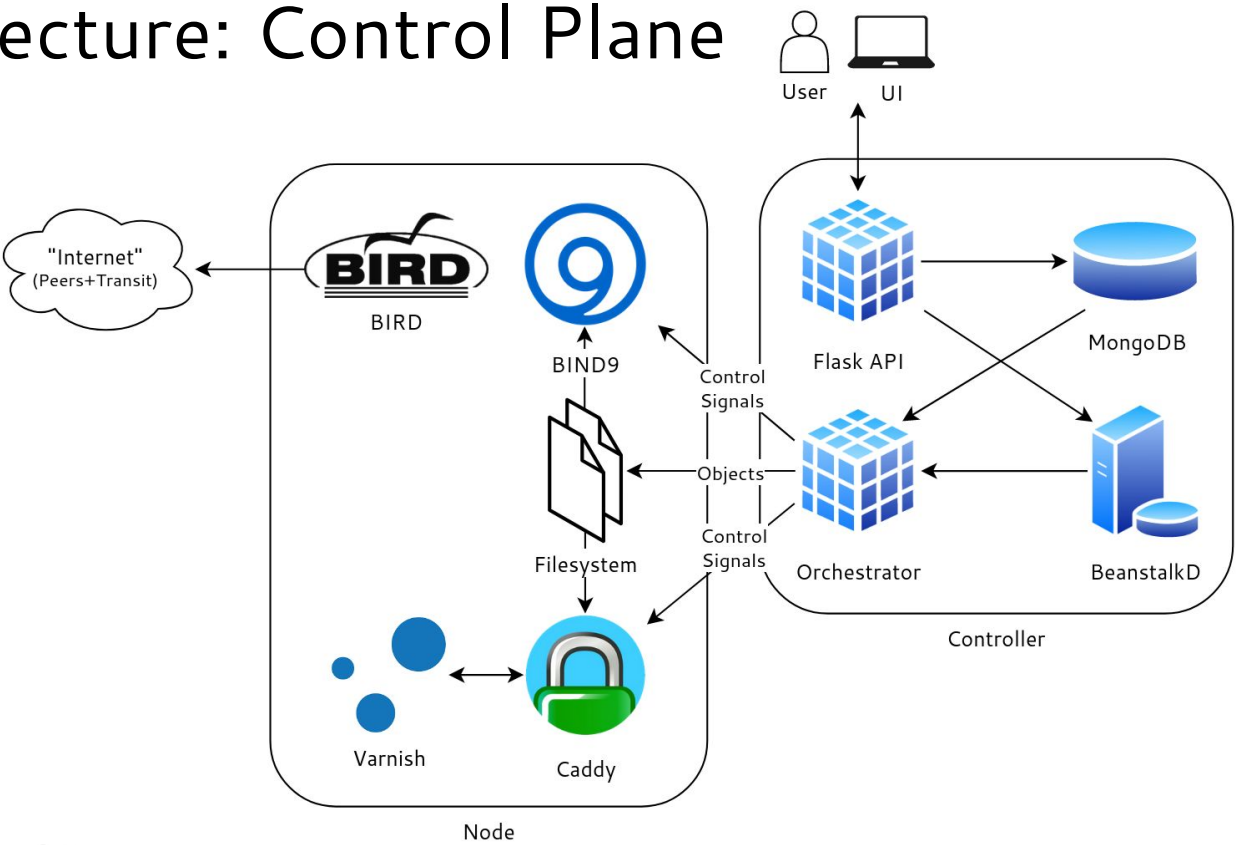


Services

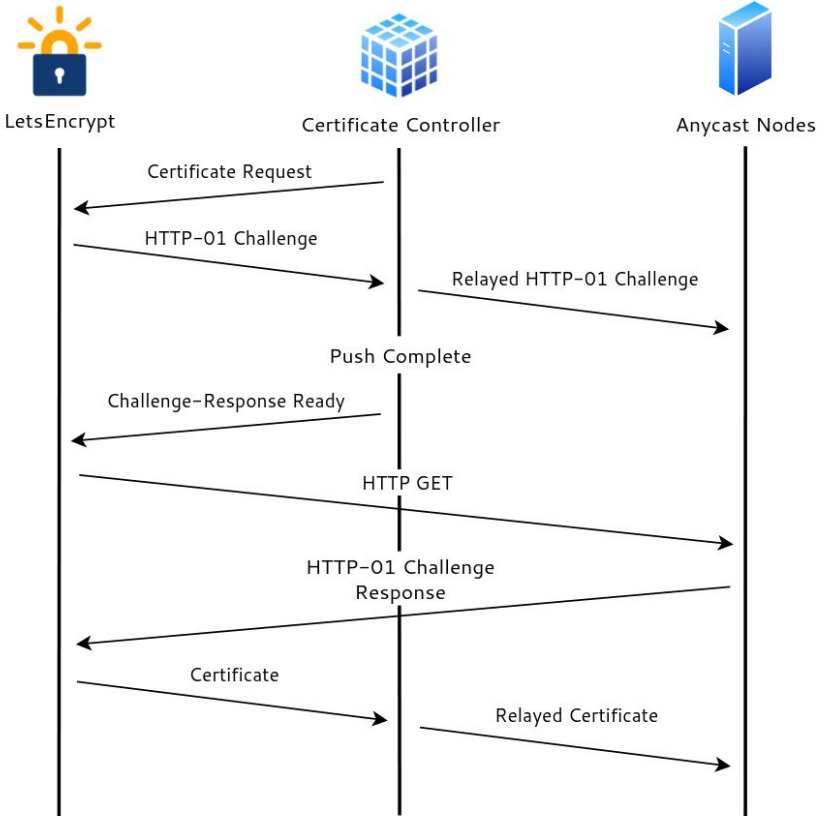
- Authoritative DNS
 - bind
 - Mostly UDP, some TCP
- HTTP(s) caching proxy
 - caddy + varnish
 - Mostly TCP, some UDP (HTTP/3)



Architecture: Control Plane



Architecture: TLS Certificates



Architecture: Network

- 2 global anycast networks
 - A: NS1, all PoPs
 - B: NS2 + HTTP caching proxy, only larger PoPs
 - Not enough IP space for regional anycast
- Dual stack
- Public peering

Public Peering Exchange Points			Filter
Exchange ↓ ASN	IPv4 IPv6	Speed RS Peer	
<u>ARIX</u> 34553	44.190.42.3 2a0e:8f00:fddd::3	10G ✓	
<u>FCIX</u> 34553	206.80.238.77 2001:504:91::77	10G ✓	
<u>ESIX</u> 34553	185.1.174.15 2001:7f8:f7:0:185:1:17 4:15	1G ✓	
<u>Gig IX Ashburn</u> 34553	206.83.8.199 2602:fed2:ff1::199	1G ✓	
<u>KCIX</u> 34553	206.51.7.39 2001:504:1b:1::39	1G ✓	
<u>Puerto Rico Internet Exchange</u> 34553	204.138.0.21 2620:102:b000::21	20G ✓	
<u>QCIX</u> 34553	206.83.43.27 2001:504:9b::27	1G ✓	
<u>ROPN-IX</u> 34553	185.1.179.36 2001:7f8:fc:3f::36	100M ✓	
<u>SIX Seattle</u> 34553	206.81.80.97 2001:504:16::86f9	1G ✓	
<u>SIX Seattle (Jumbo)</u> 34553	206.81.82.97 2001:504:16:1::86f9	1G ✓	
<u>Speed-IX</u> 34553	185.1.95.166 2001:7f8:b7:a503:455 3:1	1G ✓	
<u>UNM-Exch Canada-West</u> 34553	192.34.27.18 2602:ffb1:200:0:192:34: 27:18	10G ✓	



NANOG™



Load Balancing

Anycast + ECMP

```
2a0e:8f00:fe05::/48 unicast [CDN_FMTv6 2020-12-05 13:05:46] * (100) [i]
  via 2001:678:c4c::6 on local
    unicast [CDN_FMT4v6 2020-12-05 13:07:38] (100) [i]
  via 2001:678:c4c::11 on local
    unicast [CDN_FMT3v6 2020-12-05 13:07:04] (100) [i]
  via 2001:678:c4c::10 on local
    unicast [CDN_FMT2v6 2020-12-05 13:06:13] (100) [i]
  via 2001:678:c4c::9 on local
```

```
2a0e:8f00:fe05::/48 proto bird src 2001:678:c4c::1 metric 32
  nexthop via 2001:678:c4c::6 dev local weight 1
  nexthop via 2001:678:c4c::9 dev local weight 1
  nexthop via 2001:678:c4c::10 dev local weight 1
  nexthop via 2001:678:c4c::11 dev local weight 1
```



NANOG™



Load Balancing

```
~ ^ for i in {1..10}; do dig +short TXT id.server CHAOS @ns1.packetframe.com ; done  
"fmt"  
"fmt4"  
"fmt"  
"fmt"  
"fmt2"  
"fmt3"  
"fmt3"  
"fmt2"  
"fmt3"  
"fmt4"
```



BGP Automation

```
asn: 65530
router-id: 192.0.2.1
prefixes:
- 192.0.2.0/24
- 2001:db8:1::/48
```

```
peers:
Example Transit:
asn: 65510
type: upstream
neighbors:
- 203.0.113.10
- 2001:db8::10
```

```
Example Peer:
asn: 65520
type: peer
neighbors:
- 203.0.113.39
- 2001:db8::39
```

- Simple, repeatable config
- RPKI ROV, IRR prefix lists, max prefix limits
- PeeringDB automation
- One file (yaml, toml, or json)
- <https://github.com/natesales/bcg>



NANOG™



Monitoring

- Route Collector
- Prometheus + Grafana
 - {bind,varnish,caddy,node}_exporter
- Troubleshooting + optimization
 - NLNOG RING
 - RIPE Atlas

```
bird> s route count
17696773 of 17696773 routes for 837837 networks in table master4
2479436 of 2479436 routes for 104870 networks in table master6
Total: 20176209 of 20176209 routes for 942707 networks in 2 tables
bird> s mem
BIRD memory usage
Routing tables: 1483 MB
Route attributes: 1839 MB
Protocols: 638 kB
Total: 3323 MB
```



Tuning Anycast

- Prepends
- Export control
- Accounting for latency/packet loss

```
~ » sudo stping -s "2001:xxx:x:xxx::2,2a04:xxxx:x:xxxx::4" -t "example.com"
Resolving example.com...2606:2800:220:1:248:1893:25c8:1946, 93.184.216.34
STPING example.com (2606:2800:220:1:248:1893:25c8:1946) from 2 sources:
Source                Sent      Loss  Min      Max      Avg      Dev
2001:xxx:x:xxx::2    1         0.00% 952µs   952µs   952µs   0s
2a04:xxxx:x:xxxx::4 1         0.00% 1.055ms 1.055ms 1.055ms 0s
2001:xxx:x:xxx::2    2         0.00% 910µs   952µs   931µs   21µs
2a04:xxxx:x:xxxx::4 2         0.00% 1.019ms 1.055ms 1.037ms 18µs
2001:xxx:x:xxx::2    3         0.00% 860µs   952µs   907µs   37µs
2a04:xxxx:x:xxxx::4 3         0.00% 1.007ms 1.055ms 1.027ms 20µs
^C
--- 2606:2800:220:1:248:1893:25c8:1946 stping statistics source 2001:xxx:x:xxx::2 ---
4 packets transmitted, 4 packets received, 0% packet loss
round-trip min/avg/max/stddev = 860.441µs/936.173µs/1.020808ms/58.819µs

--- 2606:2800:220:1:248:1893:25c8:1946 stping statistics source 2a04:xxxx:x:xxxx::4 ---
4 packets transmitted, 4 packets received, 0% packet loss
round-trip min/avg/max/stddev = 925.494µs/1.00201ms/1.055702ms/47.595µs
```



NANOG™



Community Involvement

- Always looking for new PoPs
- Open source hosting
- PRs/issues welcome



Thank you!

- nate@natesales.net
- <https://github.com/natesales>

