Anycast Service Deployment

Arth Paulite – Infrastructure Services
APNIC 2021

1

Overview

- Anycast concept
- How does it work
- Deployment examples
- Best practices
- Q&A

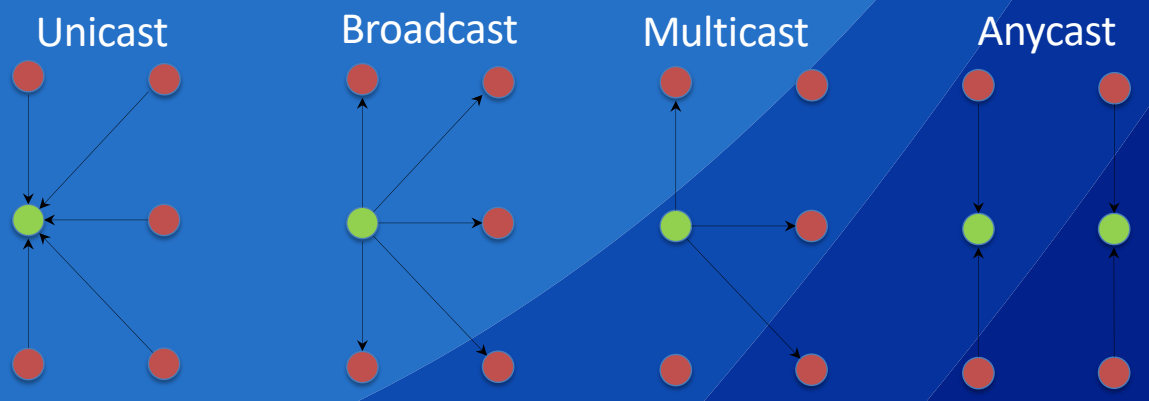
2

Review of addressing methods

- Unicast
  - One-to-one transmission
  - Use of unicast IP address from a single network.
- Anycast
  - One-to-one transmission (Closest network as priority)
  - Use of same unicast IP address from different network locations.
- Broadcast
  - One-to-all
  - Broadcast IP: 255.255.255.255 or 192.168.0.255
- Multicast
  - One-to-many
  - Multicast IP: 224.0.0.0 to 239.255.255.255

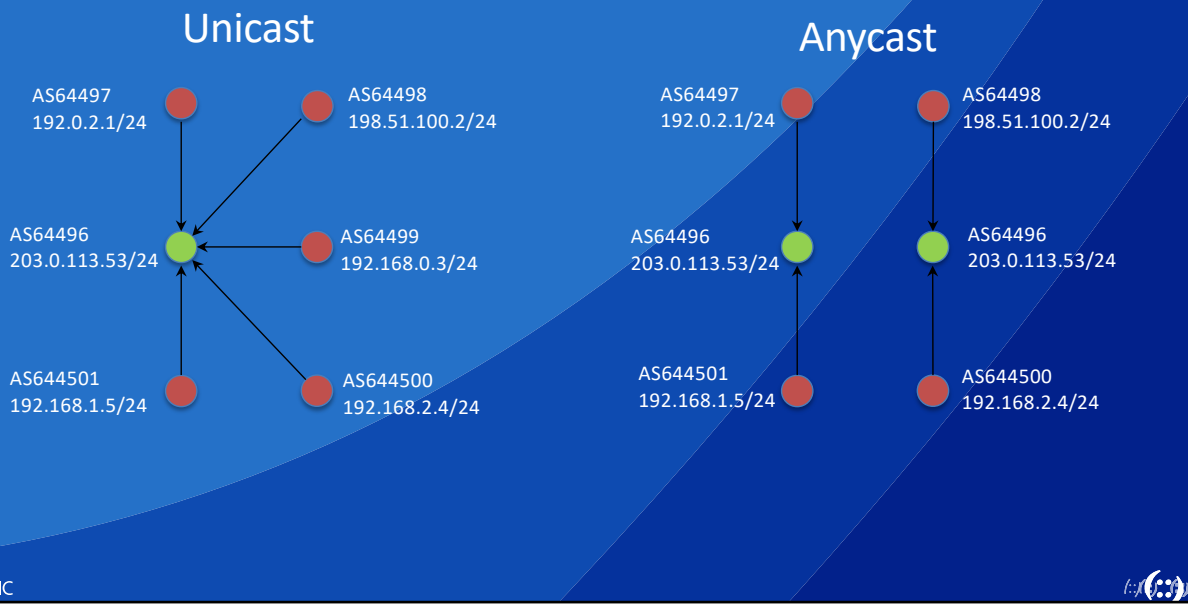
3

Review of Addressing methods



4

Unicast VS Anycast



5

What is anycast

- Anycast
  - The practice of making a particular service address available in two or more discrete locations.
  - Use of routing system to decide which location will serve each request.
- Service address
  - The endpoint or node IP address that responds to requests.
- Anycast node
  - Collection of hosts and routers providing service to an anycast service address
  - Or a single server running software router and actual service
- Local node
  - Anycast node visible only to selected network.
- Global node
  - Anycast node visible on global Internet.

6

Motivations for deploying anycast

- Allow infrastructure to scale
  - Load distribution, accommodate increasing queries
- To Improve service response time
  - Hosting service close to other users
- To improve service reliability
  - Can provide automatic fail-over to backup nodes
- To keep service traffic local
  - Avoid congesting transit link
- Denial-of-Service Attack mitigation
  - Constraint or localize attack to single anycast node

7

Unicast service response time

8

Anycast service use case

- Busy DNS servers
  - Root and GTLD servers: m.root-servers.net, a.gtld-servers.net
  - Public DNS resolvers: 1.1.1.1, 8.8.8.8, 9.9.9.9
  - Commercial DNS: AWS Route53, NS1, Cloudflare, Constellix
  - Registry: APNIC, RIPE, ARIN, Verisign, Affilias, GoDaddy
- CDNs: Akamai, Cloudflare, Azure, Google, Amazon

9

How does anycast work?

- Use of Routing Systems
  - When two or more routes are available, router decides where to send the packet.
- Anycast within IGP (Interior Gateway Protocol)
  - Multiple service address within internal network
  - Uses IGP routing protocols like OSPF, ISIS
- Anycast within the Global Internet
  - Multiple service address are distributed and available globally
  - Uses BGP routing protocol

10

How does anycast work? (IGP)

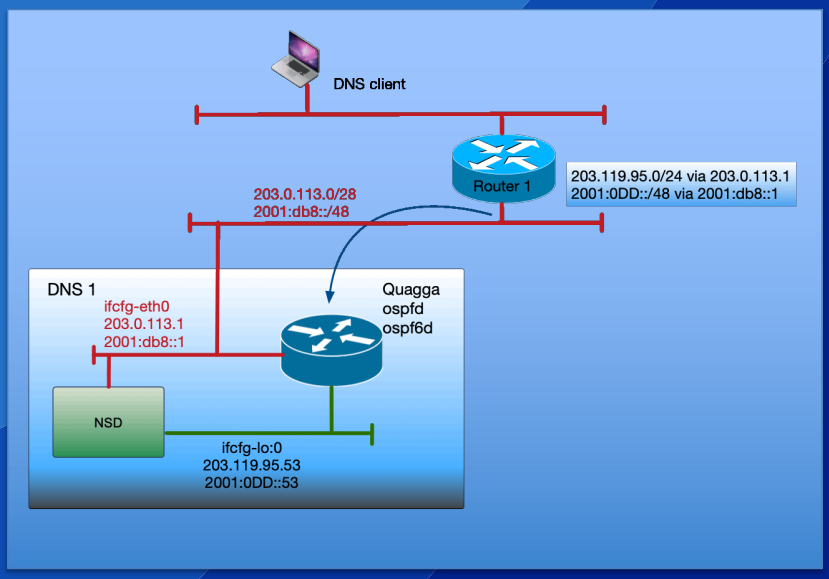
- Considerations
  - Dedicated /24 IPv4 and /48 IPv6 if service will be distributed in other network locations.
  - Use of small subnet for internal distribution
  - Running software router and service together in one server
  - Assignment of unique management IP per node
  - Assignment of unique service ID per node
  - Coupling or tying monitoring of service and router
    - If service failed, routing should be disabled to fail-over to next available node.

11

How does anycast work? (IGP)

Example: Using OSPF

- Linux server
- Linux loopback ifcfg-lo:0
- ospfd for IPv4 routing
- ospf6d for IPv6 routing
- NSD DNS server
- Server Mgmt IP
- Server anycast service IP
- OSPF session with router1

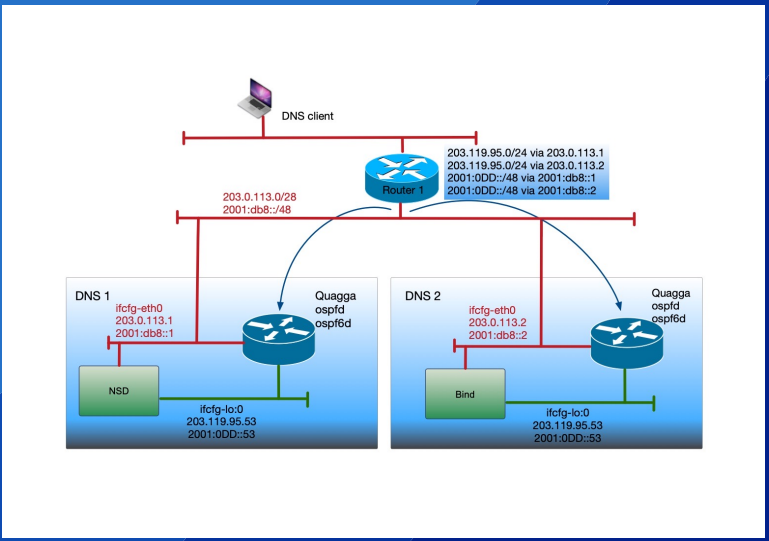


12

How does anycast work? (IGP)

Example: Using OSPF

- 2 anycast nodes
- Load distribution
- OSPF equal-cost path
- Cisco support 16 next hops

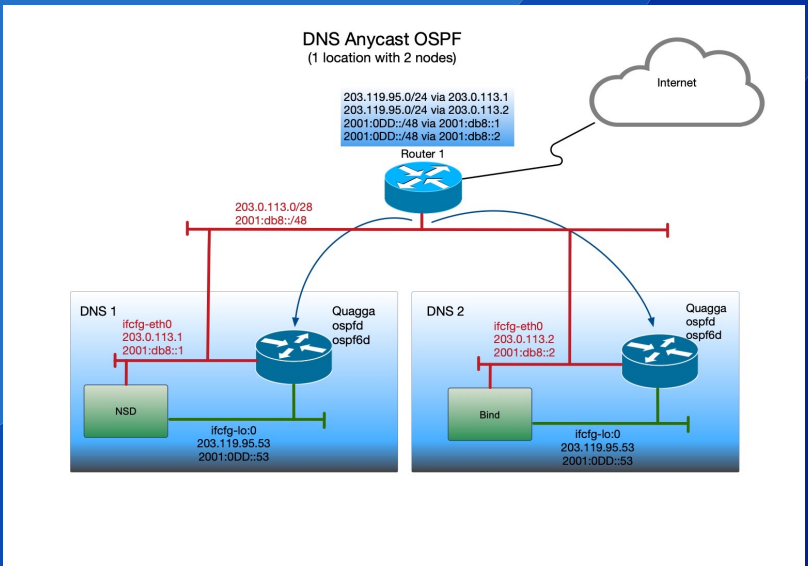


13

How does anycast work? (IGP)

Example: Using OSPF

- 2 anycast nodes
- Router1 BGP peering
- Advertise service address



14

How does anycast work? (BGP)

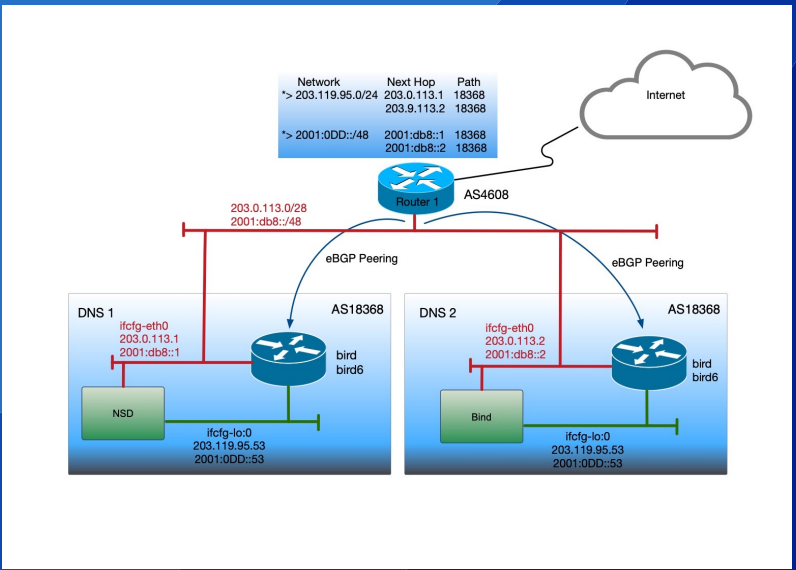
- BGP peering from selected POP
  - BGP neighbor: transit provider, Internet Exchange, content provider
- BGP advertisement of the same IP block from each node
  - Example: 203.0.113.0/24, 2001:DB8::/48
- BGP shortest AS Path selection from clients
  - If multiple route exist, route with few AS will be used.
- Deploying the same service, same IP from each POP
  - Example: Anycast DNS on 203.9.113.53, 2001:DB8::53/48
- Running software router on each server
  - Example: Quagga, bird

15

How does anycast work? (BGP)

Example: Anycast using BGP

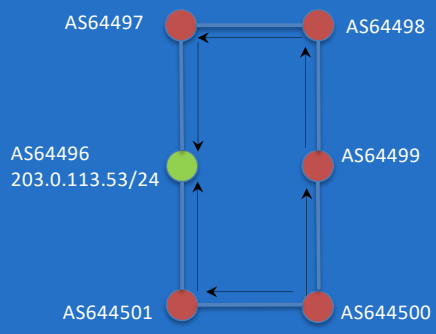
- 2 anycast nodes
- bird/bird6 router
- eBGP peering
- Load distribution
- Minimum /24 IPv4 block
- Minimum /48 IPv6 block



16

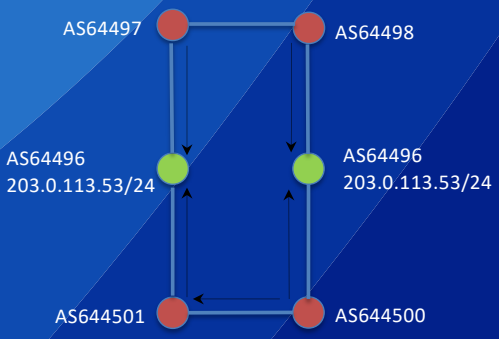
How does anycast work? (BGP AS-path)

Unicast



From AS644500 to AS64496
1. AS path: AS644501, AS64496
2. AS path: AS64499, AS65598, AS64497, AS64496

Anycast



From AS644500 to AS64496
1. AS path: AS64496
2. AS path: AS644501, AS64496

17

Deployment Example: APNIC DNS

- ns2.apnic.net
  - APNIC reverse zones: 1.in-addr.arpa, 0.4.2.ip6.arpa
  - Dedicated ASN & IP: AS18369, 203.119.95.0/24, 2001:ddd::/48
  - 114 zones
  - 10 global nodes
  - Combined network traffic: 80Mbps
  - Combined DNS query load 60,000 qps

18

Deployment Example: APNIC DNS

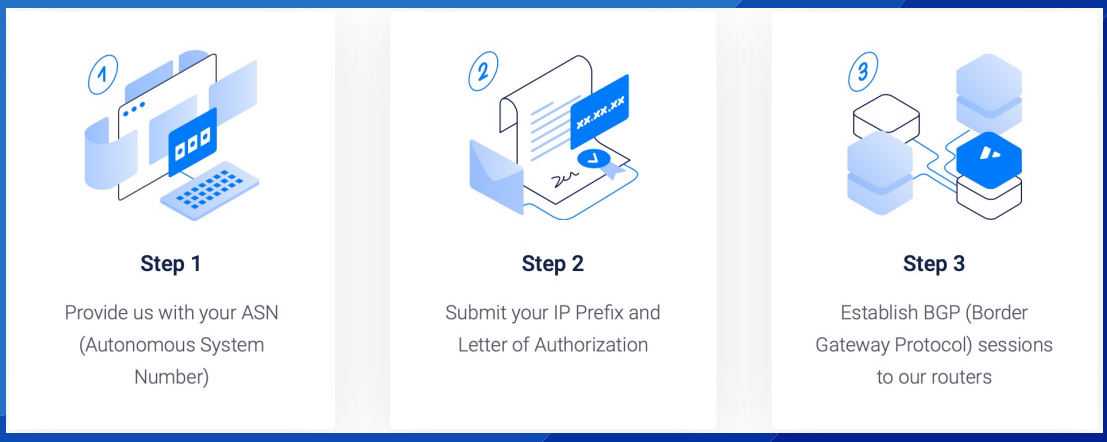
- Using vultr.com cloud servers
  - US: New Jersey, Silicon valley
  - EU: Amsterdam, Frankfurt
  - AP: Singapore, Tokyo, Seoul
- From APNIC POP
  - NextDC Brisbane
  - Interactive Brisbane
  - KDDI Tokyo

19

Deployment Example: APNIC DNS

Announcing IP in Vultr

- RPKI ROA is required

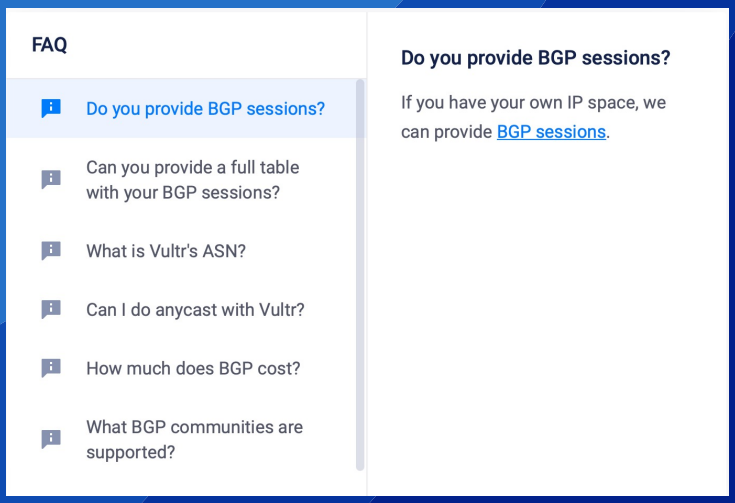


20

Deployment Example: APNIC DNS

Vultr BGP resources

- FAQ
- Configuration guide



APNIC

21

21

Deployment Example: APNIC DNS

Server network interface

```
dummy1: flags=195<UP,BROADCAST,RUNNING,NOARP>  mtu 1500
        inet 203.0.113.53  netmask 255.255.255.255  broadcast 203.0.255.53
        inet6 2001:db8::53  prefixlen 64  scopeid 0x0<global>
        inet6 fe80::58ec:9fff:fef4:456f  prefixlen 64  scopeid 0x20<link>
        ether 5a:ec:9f:f4:45:6f  txqueuelen 1000  (Ethernet

eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 207.148.75.233  netmask 255.255.254.0  broadcast 207.148.75.255
        inet6 2001:19ff:4400:6084:5400:1ff:fe99:f39a  prefixlen 64  scopeid 0x0<global>
        inet6 fe80::5400:1ff:fe99:f39a  prefixlen 64  scopeid 0x20<link>
        ether 56:00:01:99:f3:9a  txqueuelen 1000  (Ethernet)
```

APNIC

22

22

Deployment Example: APNIC DNS

#/etc/bird.conf

```
router id 203.0.113.123;
protocol direct
{
    interface "dummy1";
    import all;
}
protocol bgp vultr
{
  # substitute with your AS or Vultr's private AS
  local as 18369;
     source address 203.0.113.123;
graceful restart on;
     multihop 2;
     neighbor 169.254.169.254 as 64515;
     password "**********";
```

Ifcfg-eth0

```
DEVICE="eth0"
BOOTPROTO="none"
ONBOOT="yes"
TYPE="Ethernet"
NM_CONTROLLED="no"
IPV6_AUTOCONF="yes"
IPV6INIT="yes"
IPADDR=207.148.77.233
NETMASK=255.255.254.0
```

Ifcfg-dummy1

```
DEVICE="dummy1"
BOOTPROTO="none"
ONBOOT="yes"
TYPE="Ethernet"
NM_CONTROLLED="no"
IPV6_AUTOCONF="no"
IPV6INIT="yes"
IPADDR=203.0.113.53
NETMASK=255.255.255.255
IPV6ADDR=2001:db8::53/64
```

APNIC

23

23

Deployment Example: APNIC DNS

Checking Bird BGP status

[root@vultr ~]# birdc show proto all vultr

BIRD 1.4.5 ready.

name	proto	table	state	since	info
vultr	BGP	master	up	14:11:36	Established

BGP state: Established
Neighbor address: 169.254.169.254
Neighbor AS: 64515
Neighbor ID: 169.254.169.254

BGP peering is working properly

APNIC

24

24

Anycast service response time

25

Best Current Practice

- Consider protocol suitability for the service
  - Compare routing stability and service transaction time
- Tune OSPF timers from it's default
  - Faster convergence for service failure.
- Create RPKI ROA from your Regional Registry
  - APNIC, ARIN, LACNIC, RIPE, AFRINIC
- Create route object
  - Prefix and ASN should be consistent with RPKI ROA

26

Best Current Practice

- Add unique server ID for node tracking
  - Which server: dig CH TXT ID.SERVER @(server name/ip)
  - Example for Bind DNS server:

    # named.conf settings

    server-id "NS2-Singapore-1";
  - Example for NSD DNS server

    # nsd.conf

    identity: " NS2-Singapore-1"
- Run firewall on each server
  - Allow SSH, other internal services on a trusted zone
  - Only the actual service should be on public zone

27

Best Current Practice

- Local monitoring and remediation
  - Turning off OSPF or BGP if service failure was detected
  - Turning on OSPF or BGP if service health returns to normal
- Metrics collection and visualization
  - Resource utilization (disk, cpu, memory,network)
  - Service metrics: query rates, failures
- Server diversity
  - Use of multiple application server for the same content
  - DNS example: Bind, NSD, Knot, PowerDNS

28

Best Current Practice

- Configuration management
  - Consistent configuration across nodes
  - Faster node deployment and changes
    - Could also result in cascading failure if not careful
- Use of test environment
  - Testing new features
  - Performance measurement
  - Test, Test, Test
- Additional reference: RFC 4786

29

Question?

30