Containers and Anycast IPs

Andrew Sy Kim - Software Engineer at DigitalOcean

@a_sykim  @andrewsykim
Agenda

- Container Networking
- Kubernetes Networking
- Data center Networking
- Anycast Routing/IPS with Kubernetes
Container Networking
Container Networking
Container Networking

$ ip route
default via 10.0.0.1 dev eth0
172.20.100.0/25 dev kube-bridge proto kernel scope link src 172.20.100.1
Container Networking
**Container Networking**

```bash
veth
$ ip link add vethA type veth peer name vethB
$ ip netns add mynetns
$ ip link set vethA netns mynetns
```
Container Networking

![Diagram of container networking showing veth devices, a bridge, and eth0 interfaces. The diagram illustrates how veth devices connect to a bridge, which in turn connects to the physical network interface eth0.]
Container Networking
Container Networking

Diagram showing container networking with veth, bridge, and VxLAN components.
Pods

A pod is a group of one or more containers, with shared storage/network, and a specification for how to run the containers.
A Kubernetes Service is an abstraction which defines a logical set of Pods and a policy by which to access them.
Kubernetes Networking
Pod to Pod
**Iptables:**
$ iptables-save
...
-A KUBE-SERVICES -d 10.38.3.1/32 -p tcp -m comment
--comment "default/kubernetes:https cluster IP" -m
tcp --dport 443 -j KUBE-SVC-NPX46M4PTMTKR6Y
...

**IPVS:**
$ sudo ipvsadm -L -t 10.38.3.1:443
Prot LocalAddress:Port Scheduler Flags
--> RemoteAddress:Port Forward Weight
ActiveConn InActConn
TCP  worker01 rr
--> 100.96.1.3:domain Masq 1 0
0
--> 100.96.1.4:domain Masq 1 0
0
"
External Traffic

Diagram:
- veth
- bridge
- iptables IPVS
- eth0
- pod
External Traffic
External Traffic

- veth
- bridge
- iptables
- IPVS
- eth0
- pod
External Traffic into Kubernetes Cluster

myapp.mycluster.com

10.0.1.4

10.0.1.8

10.0.3.7
Data Center Networking
BGP
“The Border Gateway Protocol (BGP) is an inter-Autonomous System routing protocol”

– IETF RFC 4271
Bottlenecks in our Network
Pod/Container Networking
External Traffic into Kubernetes Cluster

*.mycluster.com

10.0.1.4

10.0.3.7

10.0.1.8
Anycast
Anycast Routing

Anycast is a network addressing and routing methodology in which a single destination address has multiple routing paths to two or more endpoint destinations.
External Traffic into Kubernetes Cluster

*.mycluster.com

10.0.1.4

10.0.3.7

10.0.1.8
External Traffic into Kubernetes Cluster

*.mycluster.com

1.2.3.4

1.2.3.4

1.2.3.4

1.2.3.4

1.2.3.4
Kubernetes Clusters as BGP Autonomous Systems
ToR (Top of Rack)
- Supports iBGP / eBGP peering
- Automatic BGP peering of pod and service subnets
- Bonus: IPVS/DSR support, network policies, BGP route reflectors

https://github.com/cloudnativelabs/kube-router
Maintained by @murali-reddy
kube-router

- IP and ASN to peer with
- BGP peering required on all nodes
Benefits!
Faster Pod Network!
External Traffic into Kubernetes Cluster

*.mycluster.com

1.2.3.4

1.2.3.4

1.2.3.4

1.2.3.4
1.2.3.4

1.2.3.4

1.2.3.4

1.2.3.4
Summary

● BGP + Kubernetes works!!
● Anycast + Kubernetes Service IPs is a powerful combination!
Global Container Networks on Kubernetes at DigitalOcean

https://youtu.be/tHAkey-sZ9g
Thank you!

@a_sykim
@andrewsykim