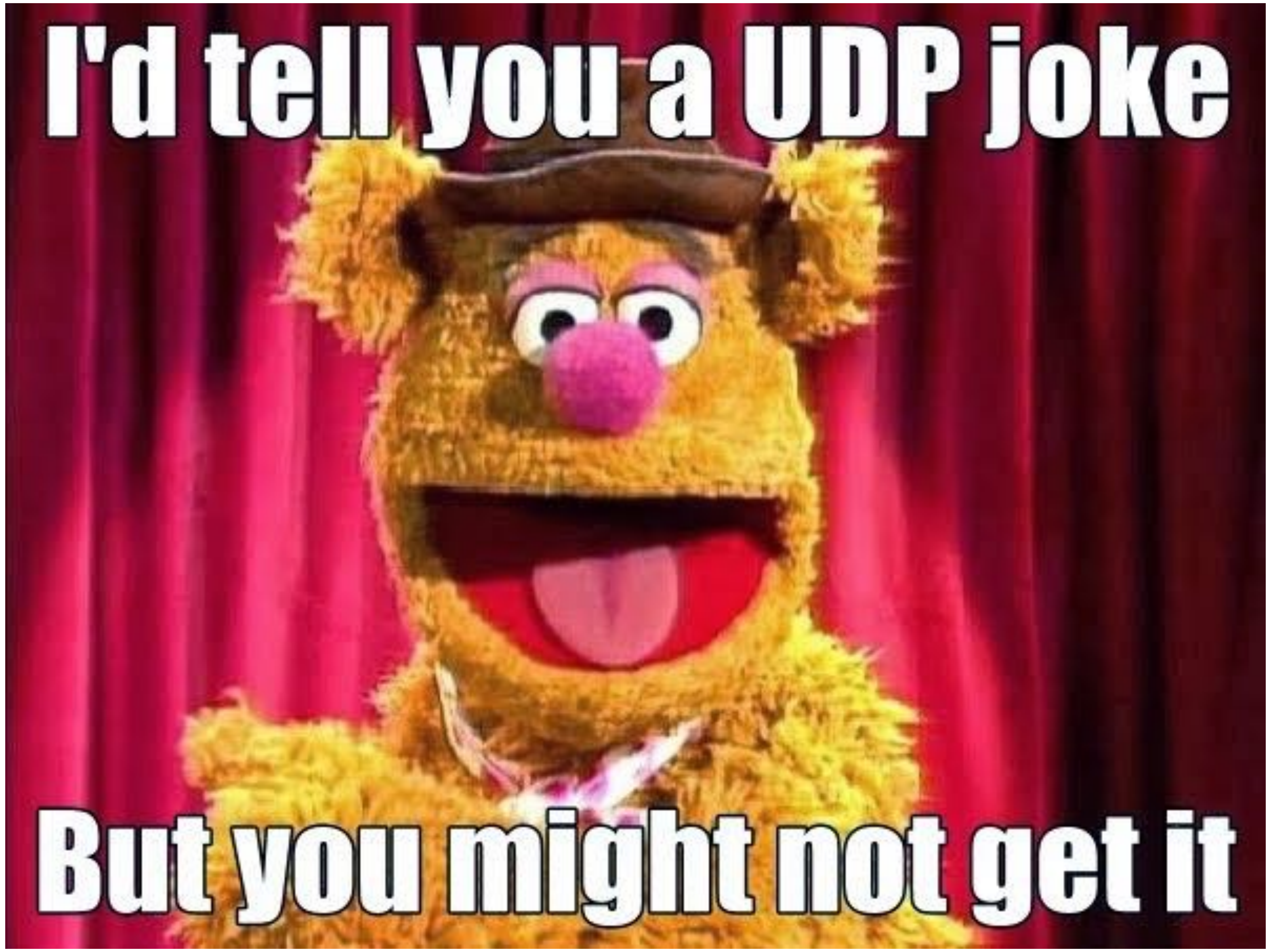
KubeCon | CloudNativeCon

North America 2017

# UDP in K8S: Signed, Sealed, but Delivered?

Amanpreet Singh, Software Engineer, *Crowdfire*

# Obligatory UDP Joke



# Where do we use UDP anyway?

KubeDNS

- Service discovery!
- Crucial in a cluster where services call each other all the time

# Where do we use UDP anyway?

KubeDNS

**ProTip:** Use pre-existing environment variables like these to save all the DNS calls!

**${MYAPP_SERVICE_HOST}**

# Where do we use UDP anyway?

StatsD

- Statsd+graphite for custom business and service metrics.
- Single-pod deployment backed by a persistent volume (EBS)
- Not HA since Kubernetes restarts it quickly in case of failure

# K8S Networking Primer

Key Concepts:

- Every pod has a unique IP
- These IPs are routable from all the pods (even on different nodes)

# K8S Networking Primer

Communication among applications:

- Pod IPs are changing all the time
- Reasons include: rolling updates, scaling events, node crashes
- Pod IPs unreliable for using directly

# K8S Networking Primer

Kubernetes Services:

- Static Virtual IPs that act as a loadbalancer
- Group of Pod IPs as endpoints (identified via label selectors)

# K8S Networking Primer

```
kind: Service
apiVersion: v1
metadata:
  name: svc2
spec:
  type: clusterIP
  selector:
    app: myapp
  clusterIP: 100.64.5.119
  ports:
  - name: http
    port: 80
```

# K8S Networking Primer

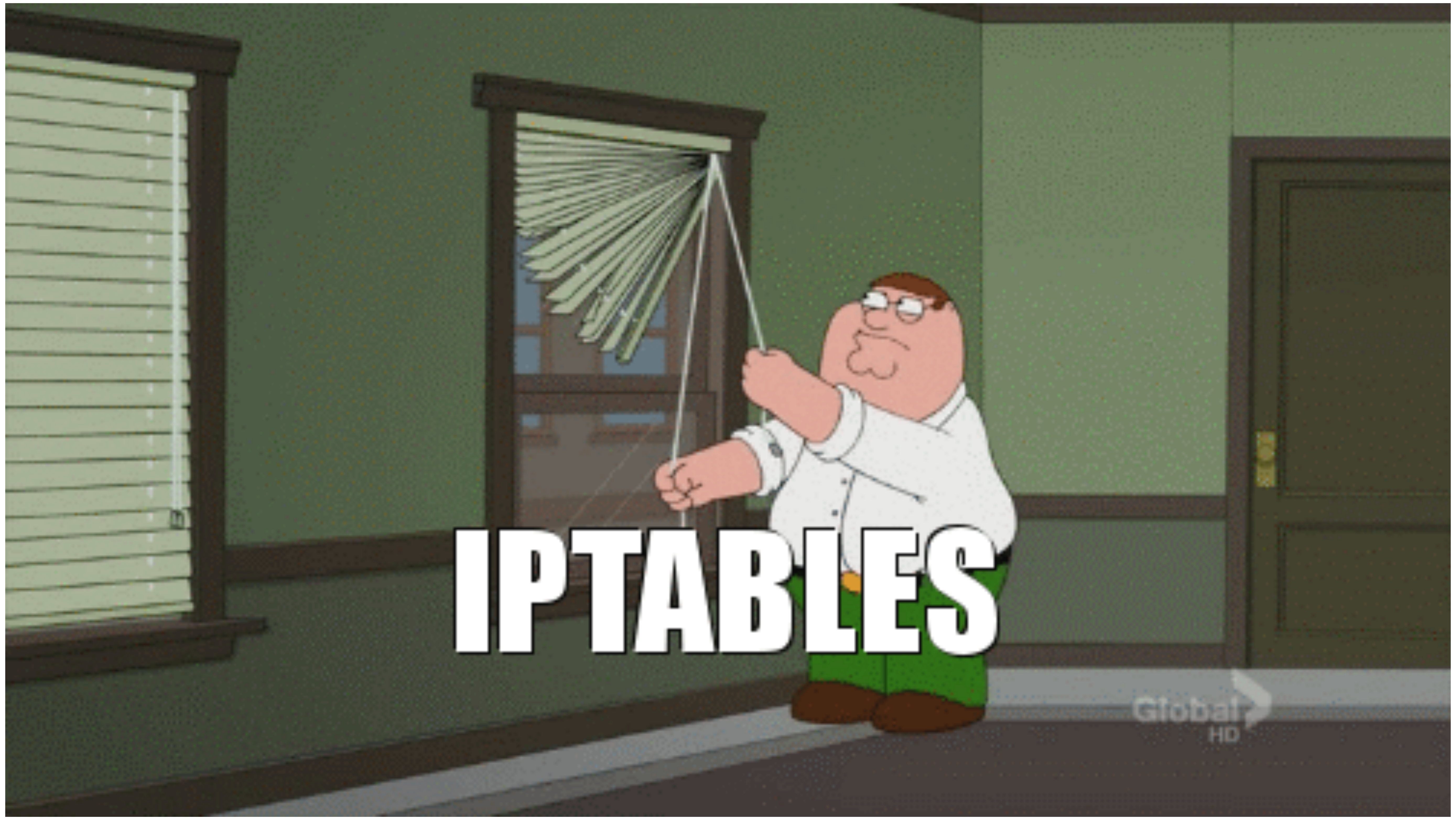
```yaml
apiVersion: v1
kind: Endpoints
metadata:
  name: svc2
subsets:
- addresses:
  - ip: 172.16.85.64
  - ip: 172.16.21.6
  - ip: 172.16.21.60
 ports:
  - name: http
    port: 8080
    protocol: TCP
```

# K8S Networking Primer

How do these services work?

- Magic ✨
- Actually, it's even more complicated than that...

# K8S Networking Primer

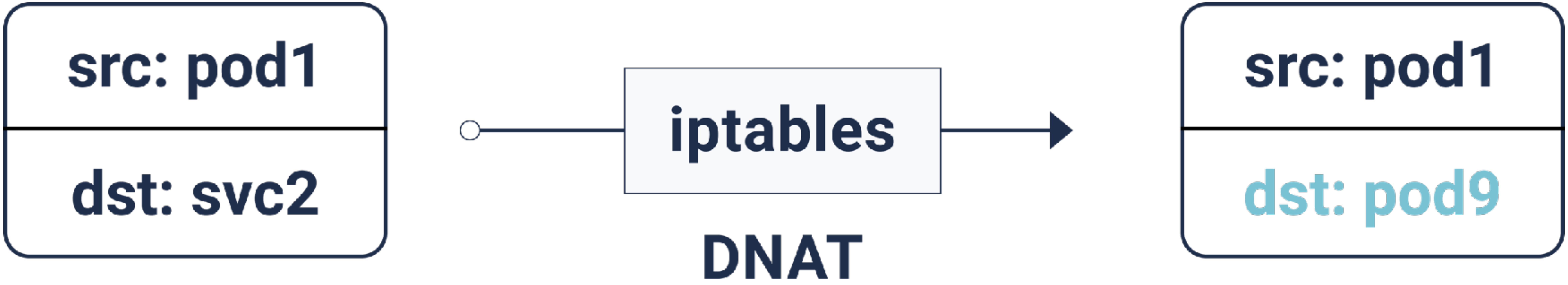


# K8S Networking Primer

kube-proxy

- Controller that watches the apiserver for service/endpoints updates
- Modifies iptables rules accordingly

# K8S Networking Primer



# K8S Networking Primer

protocol: UDP
src_ip: pod1
src_port: 12345
**dst_ip: svc2**
dst_port: 8125

→

protocol: UDP
src_ip: pod1
src_port: 12345
**dst_ip: pod9**
dst_port: 8125

# K8S Networking Primer

```
ubuntu@ip-10-2-139-161:~$ sudo conntrack -L -p udp --dst 100.64.5.119 --dport 8125
udp      17 5 src=172.16.107.16 dst=100.64.5.119 sport=58616 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=58616 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=35793 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=35793 mark=0 use=1
udp      17 18 src=172.16.107.18 dst=100.64.5.119 sport=56072 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=56072 mark=0 use=1
udp      17 28 src=172.16.107.10 dst=100.64.5.119 sport=57916 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=57916 mark=0 use=1
udp      17 22 src=172.16.107.10 dst=100.64.5.119 sport=59350 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=59350 mark=0 use=1
udp      17 2 src=172.16.107.18 dst=100.64.5.119 sport=50327 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=50327 mark=0 use=1
udp      17 8 src=172.16.107.16 dst=100.64.5.119 sport=46683 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=46683 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=44965 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=44965 mark=0 use=1
udp      17 29 src=172.16.107.4 dst=100.64.5.119 sport=46837 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=46837 mark=0 use=1
udp      17 10 src=172.16.107.16 dst=100.64.5.119 sport=59099 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=59099 mark=0 use=1
```

# K8S Networking Primer

```
ubuntu@ip-10-2-139-161:~$ sudo conntrack -L -p udp --dst 100.64.5.119 --dport 8125
udp      17 5 src=172.16.107.16 dst=100.64.5.119 sport=58616 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=58616 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=35793 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=35793 mark=0 use=1
udp      17 18 src=172.16.107.18 dst=100.64.5.119 sport=56072 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=56072 mark=0 use=1
udp      17 28 src=172.16.107.10 dst=100.64.5.119 sport=57916 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=57916 mark=0 use=1
udp      17 22 src=172.16.107.10 dst=100.64.5.119 sport=59350 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=59350 mark=0 use=1
udp      17 2 src=172.16.107.18 dst=100.64.5.119 sport=50327 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=50327 mark=0 use=1
udp      17 8 src=172.16.107.16 dst=100.64.5.119 sport=46683 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=46683 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=44965 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=44965 mark=0 use=1
udp      17 29 src=172.16.107.4 dst=100.64.5.119 sport=46837 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=46837 mark=0 use=1
udp      17 10 src=172.16.107.16 dst=100.64.5.119 sport=59099 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=59099 mark=0 use=1
```

**Protocol: UDP**
**Protocol number: 17**

# K8S Networking Primer

```
ubuntu@ip-10-2-139-161:~$ sudo conntrack -L -p udp --dst 100.64.5.119 --dport 8125
udp      17 5 src=172.16.107.16 dst=100.64.5.119 sport=58616 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=58616 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=35793 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=35793 mark=0 use=1
udp      17 18 src=172.16.107.18 dst=100.64.5.119 sport=56072 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=56072 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=57916 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=57916 mark=0 use=1
udp      17 22 src=172.16.107.10 dst=100.64.5.119 sport=59350 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=59350 mark=0 use=1
udp      17 [illegible] src=172.16.107.18 dst=100.64.5.119 sport=50327 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=50327 mark=0 use=1
udp      17 8 src=172.16.107.16 dst=100.64.5.119 sport=46683 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=46683 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=44965 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=44965 mark=0 use=1
udp      17 29 src=172.16.107.4 dst=100.64.5.119 sport=46837 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=46837 mark=0 use=1
udp      17 10 src=172.16.107.16 dst=100.64.5.119 sport=59099 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=59099 mark=0 use=1
```

TTL: 22 sec

# K8S Networking Primer

```
ubuntu@ip-10-2-139-161:~$ sudo conntrack -L -p udp --dst 100.64.5.119 --dport 8125
udp      17 5 src=172.16.107.16 dst=100.64.5.119 sport=58616 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=58616 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=35793 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=35793 mark=0 use=1
udp      17 18 src=172.16.107.18 dst=100.64.5.119 sport=56072 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=56072 mark=0 use=1
udp      17 28 src=172.16.107.10 dst=100.64.5.119 sport=57916 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=57916 mark=0 use=1
udp      17 22 src=172.16.107.10 dst=100.64.5.119 sport=59350 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=59350 mark=0 use=1
udp      17 2 src=172.16.107.10 dst=100.64.5.119 sport=50327 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=50327 mark=0 use=1
udp      17 8 src=172.16.107.16 dst=100.64.5.119 sport=46683 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=46683 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=44965 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=44965 mark=0 use=1
udp      17 29 src=172.16.107.4 dst=100.64.5.119 sport=46837 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=46837 mark=0 use=1
udp      17 10 src=172.16.107.16 dst=100.64.5.119 sport=59099 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=59099 mark=0 use=1
ubuntu@ip-10-2-139-161:~$
```

**src: 172.16.107.10  sport: 59350**
**dst: 100.64.5.119  dport: 8125**
**(StatsD service IP)  (StatsD port)**

# K8S Networking Primer

```
ubuntu@ip-10-2-139-161:~$ sudo conntrack -L -p udp --dst 100.64.5.119 --dport 8125
udp      17 5 src=172.16.107.16 dst=100.64.5.119 sport=58616 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=58616 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=35793 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=35793 mark=0 use=1
udp      17 18 src=172.16.107.18 dst=100.64.5.119 sport=56072 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=56072 mark=0 use=1
udp      17 28 src=172.16.107.10 dst=100.64.5.119 sport=57916 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=57916 mark=0 use=1
udp      17 22 src=172.16.107.10 dst=100.64.5.119 sport=59350 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=59350 mark=0 use=1
udp      17 2 src=172.16.107.18 dst=100.64.5.119 sport=50327 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=50327 mark=0 use=1
udp      17 8 src=172.16.107.16 dst=100.64.5.119 sport=46683 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=46683 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=44965 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=44965 mark=0 use=1
udp      17 29 src=172.16.107.4 dst=100.64.5.119 sport=46837 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=46837 mark=0 use=1
udp      17 10 src=172.16.107.16 dst=100.64.5.119 sport=59099 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=59099 mark=0 use=1
ubuntu@ip-10-2-139-161:~$
```

**[UNREPLIED]**

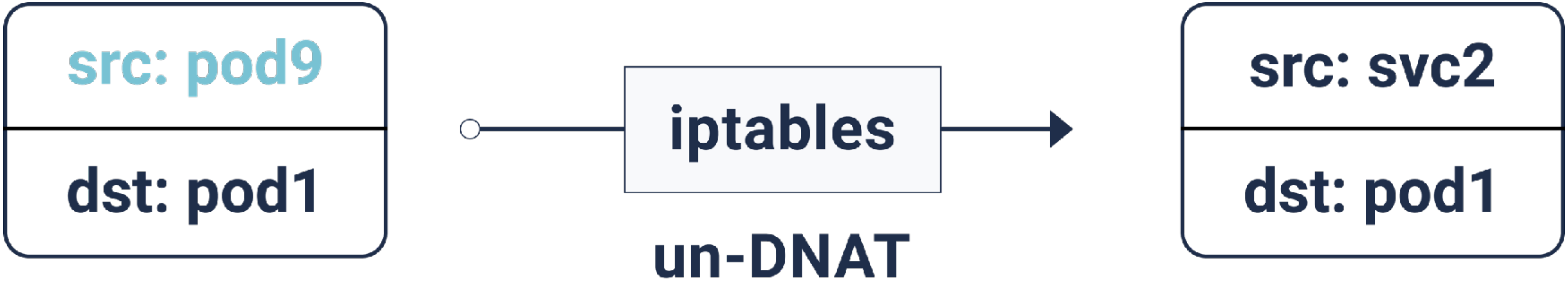
**reply hasn’t been received yet**

# K8S Networking Primer

```
ubuntu@ip-10-2-139-161:~$ sudo conntrack -L -p udp --dst 100.64.5.119 --dport 8125
udp      17 5 src=172.16.107.16 dst=100.64.5.119 sport=58616 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=58616 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=35793 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=35793 mark=0 use=1
udp      17 18 src=172.16.107.18 dst=100.64.5.119 sport=56072 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=56072 mark=0 use=1
udp      17 28 src=172.16.107.10 dst=100.64.5.119 sport=57916 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=57916 mark=0 use=1
udp      17 22 src=172.16.107.10 dst=100.64.5.119 sport=59350 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=59350 mark=0 use=1
udp      17 2 src=172.16.107.18 dst=100.64.5.119 sport=50327 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=50327 mark=0 use=1
udp      17 8 src=172.16.107.16 dst=100.64.5.119 sport=46683 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=46683 mark=0 use=1
udp      17 20 src=172.16.107.10 dst=100.64.5.119 sport=44965 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=44965 mark=0 use=1
udp      17 29 src=172.16.107.4 dst=100.64.5.119 sport=46837 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=46837 mark=0 use=1
udp      17 10 src=172.16.107.16 dst=100.64.5.119 sport=59099 dport=8125 [UNREPLIED] src=172.16.74.31 dst=172.16.107.0 sport=8125 dport=59099 mark=0 use=1
```

**(StatsD pod IP)    (StatsD port)**
**src: 172.16.74.31  sport: 8125**
**dst: 172.16.107.0  dport: 59350**

# K8S Networking Primer



src: pod9

dst: pod1

iptables

un-DNAT

src: svc2

dst: pod1

# What went wrong?

- When the StatsD pod is recreated, the metrics for some of the applications won’t reach StatsD

- Some applications were still able to send metrics successfully

- Restarting the application pods fixed it without touching the StatsD pod at all

# How did we figure it out?

Observations:

- Problem happening only for applications that send metrics very often

- Problem goes away when pods of metric-sending application are deleted/recreated

# How did we figure it out?

```
conntrack -L -p udp  --dst  100.64.5.119 \
                        --reply-src 100.64.5.119
```

Entries were present even after the StatsD pod came back up!

# How did we figure it out?

Conclusions:

- Stale conntrack entries
- TTL not expiring for pods sending metrics often

# Mitigation

- Run conntrack command (via cron) to delete stale entries
- Modify kube-proxy to run a control loop to flush stale entries

# Why did it happen?

- Couple of cases were handled in kube-proxy:
  - update/removal of endpoints
  - deletion of service/ports

- Entries not flushed when endpoint set changes from empty to non-empty

# Why did it happen?

- When the endpoint set is empty, conntrack entries blackhole the traffic
- When the UDP socket is reused, and there's new activity, the stale entry persists until the next flush

# Is it fixed now?

- PR #48524 in kube-proxy
- Adds a check to see if the endpoints set was empty before adding this new entry
- If it was empty, it’s added to the list of stale service-port names to be flushed

# Thank you!

Find me at:

Twitter/Github/Medium: **@ApsOps**