The challenges of getting delegated connectivity in (not just) cloud environments

Radek Zajíc, radek@zajic.v.pytli.cz • IPv6 Cloud Workshop, 2019-09-03

about::myself

Radek Zajíc

Blog: tech.showmax.com
Twitter: @ShowmaxDevs

about::presentation

This talk is about compute resources, not serverless.

retrospective: t-30 years

Copyright Museums Victoria / CC BY

The Internet	End-To-End Connectivity
Virtualization	quite non-existent
IPv4 addresses	1 public, Class A/B/C
IPv6 addresses	☺

retrospective: t-20 years

A Sun Ultra 80 workstation
by Ixfd64, CC BY-SA 3.0

The Internet	End-To-End Connectivity
Virtualization	emerging (VMware '99-05)
IPv4 addresses	many (CIDR), public
IPv6 addresses	from 3FFE::/16 (If your OS had support)

retrospective: t-10 years

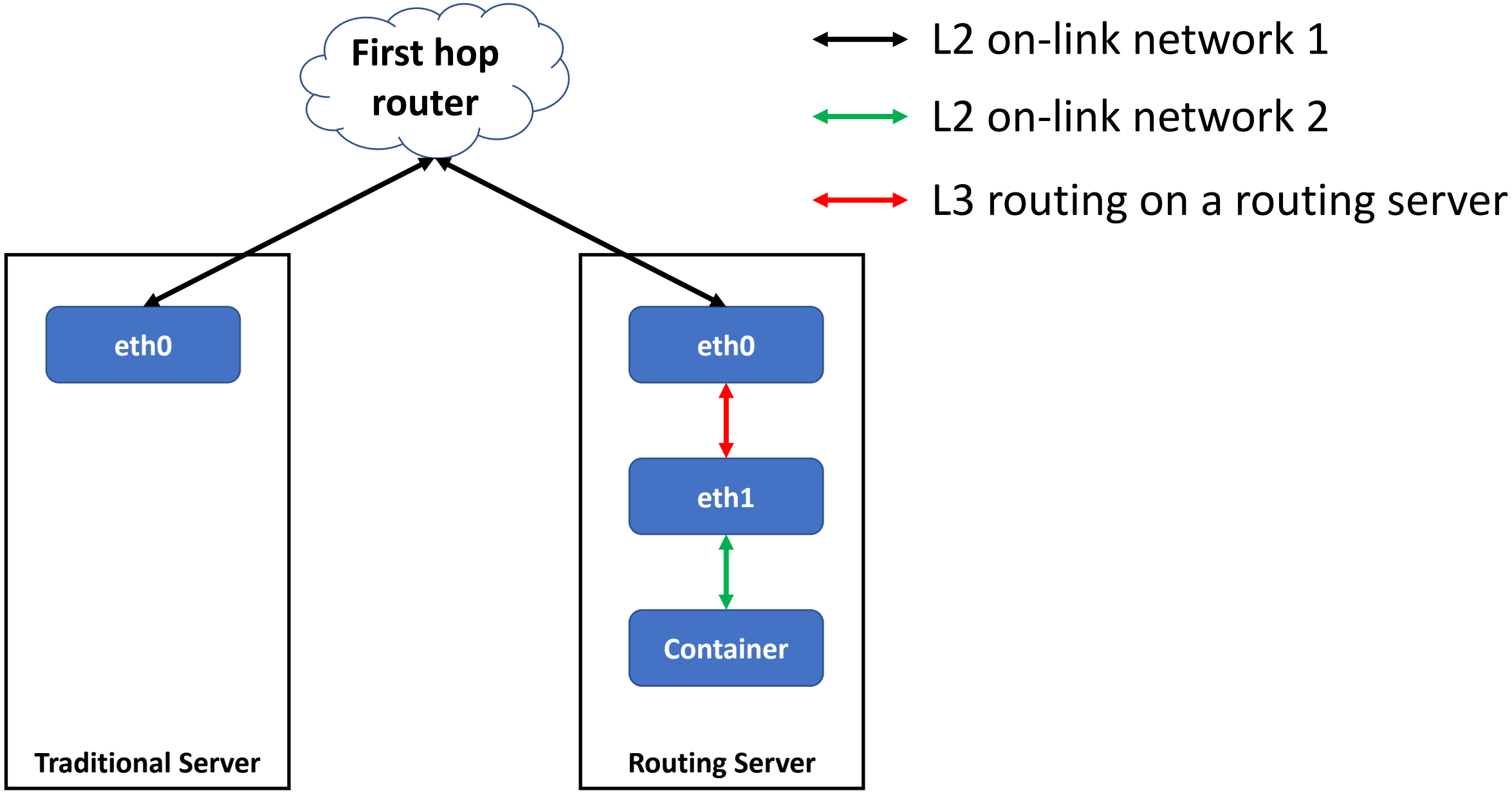
Dell PE 710, cca. 2009,
courtesy of techhead.co

The Internet	NATs, Client-Server, no SNI
Virtualization	containers emerging
IPv4 addresses	RFC1918, dozens (VMs)
IPv6 addresses	Native, GUA, 1 (?) per VM

retrospective: t-10 years

Today

The Internet	NATs, Everything-over-HTTPS, IPv6
Virtualization	containers rule the world
IPv4 addresses	hundreds (VMs, containers)
IPv6 addresses	Let's find out!


First hop router
L2 on-link network 1
L2 on-link network 2
L3 routing on a routing server
eth0
Traditional Server
eth0
eth1
Container
Routing Server

On-link address assignment

Unnumbered Link (using only link-local addresses)

Static Configuration

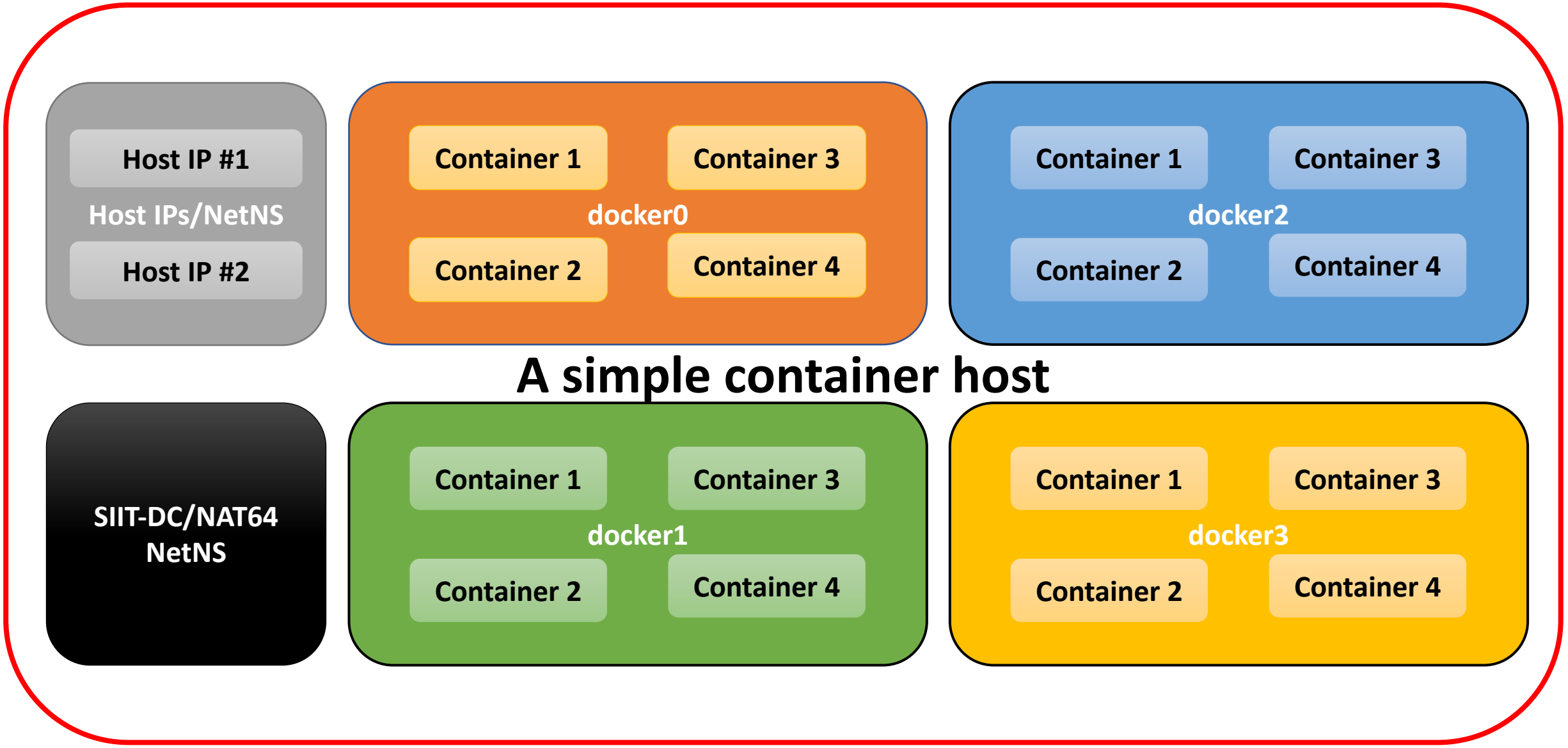
Unmanaged RA (self-assignment, unmanageable)

DHCPv6 IA_ADDR (can assign multiple addresses)

How many addresses may a server need?

How many ac es may a server need?

How many IPv6 networks may a server need?



But Why Not Just Use a Bridge and an On-Link Addres?

Friends don't let friends build large L2 networks

Proxy NDP is hard to manage (properly)

Context Separation and Firewalling on L2 is hard

There is no address scarcity in IPv6

Working with subnets is IPv6-Native

How to get and route a network…

- Statically via a Link-Local Next-Hop
- Statically via a manually Allocated/Assigned IP
- Statically via an unmanaged RA-based address
- DHCPv6 PD + route injection (can request subnet size)

Subnet Size: /56, /64, /80, /96, ...?

"The subnet for Docker containers should at least have a size of /80, so that an IPv6 address can end with the container's MAC address and you prevent NDP neighbor cache invalidation issues in the Docker layer."

https://docs.docker.com/v17.09/engine/userguide/networking/default_network/ipv6/#how-ipv6-works-on-docker

/64
is a reasonable minimum

There's also a RFC8273, Unique IPv6 Prefix per Host

How many networks does a server get?

How many IPv6 networks does a server get?

Google Cloud

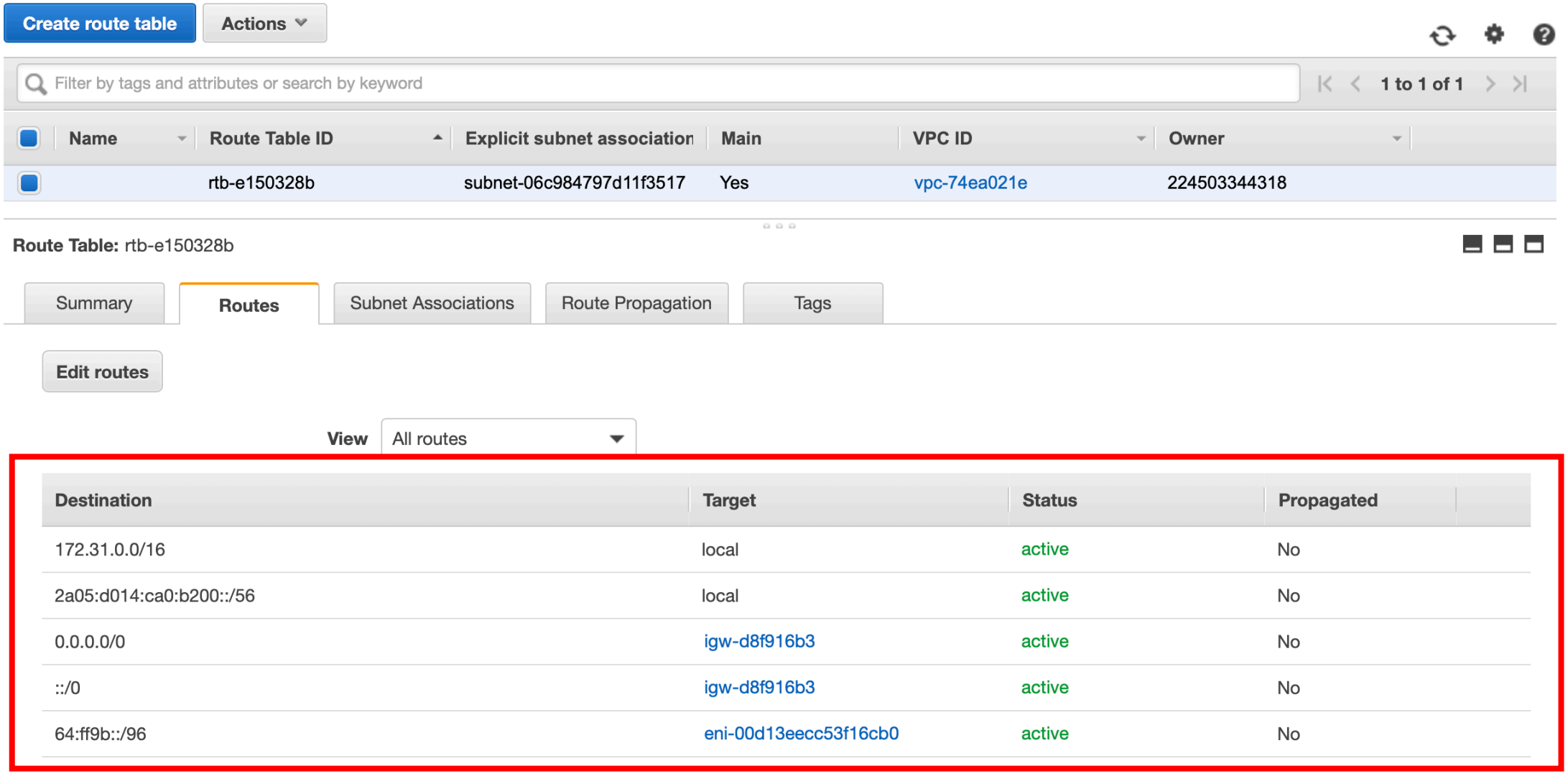
one or more /128s from a DHCPv6 server

How many IPv6 addresses does a server get?

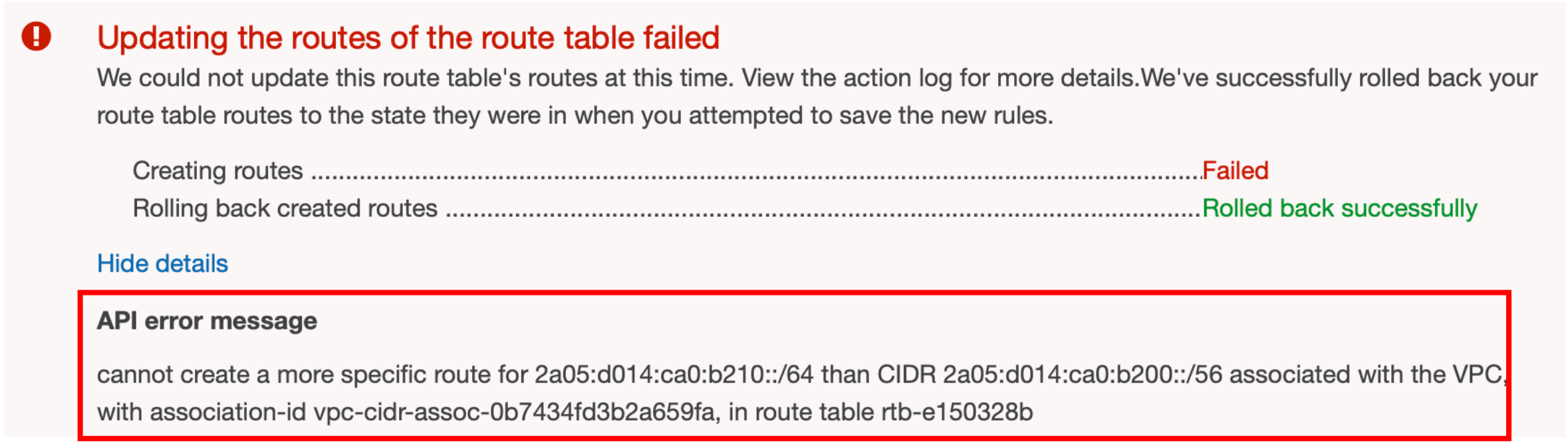
Instance Type	Maximum Network Interfaces	IPv4 Addresses per Interface	IPv6 Addresses per Interface
t2.nano	2	2	2
t2.micro	2	2	2
t2.small	3	4	4
t2.medium	3	6	6
t2.large	3	12	12
t2.xlarge	3	15	15
t2.2xlarge	3	15	15
z1d.3xlarge	8	30	30
z1d.6xlarge	8	30	30
z1d.12xlarge	15	50	50
z1d.metal	15	50	50

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-eni.html#AvailableIpPerENI

Routing Tables in AWS



More Specific Routes in a VPC



Faking a subnet via additional assigned IPs

ENI A

Primary IP:
2001:db8:dead:beef::0:1

Extra IPs (a fake /125)
2001:db8:dead:beef::1:1
2001:db8:dead:beef::1:2
2001:db8:dead:beef::1:3
2001:db8:dead:beef::1:4
2001:db8:dead:beef::1:5
2001:db8:dead:beef::1:6
2001:db8:dead:beef::1:7

Fake routed subnet:
2001:db8:dead:beef::2:0/125

ENI B

Primary IP:
2001:db8:dead:beef::0:2

Extra IPs (a fake /125)
2001:db8:dead:beef::2:1
2001:db8:dead:beef::2:2
2001:db8:dead:beef::2:3
2001:db8:dead:beef::2:4
2001:db8:dead:beef::2:5
2001:db8:dead:beef::2:6
2001:db8:dead:beef::2:7

Fake routed subnet:
2001:db8:dead:beef::2:0/125

one or more ULA ☹ /128s from a DHCPv6 server

Azure Limitations

IPv6 for Azure Virtual Network is currently in public preview. This preview is provided without a service level agreement and is not recommended for production workloads. Certain features may not be supported or may have constrained capabilities. See the Supplemental Terms of Use for Microsoft Azure Previews for details.

Azure == private IP (ULA) == NAT(66).

ULA == dual-stack destinations prefer v4 by default.

No option to allocate and route a public IPv6 prefix.

No option to assign a public IPv6 address to a VM.

You **CAN** route an arbitrary prefix (e.g. NAT64) within a VNET.

Traffic sourced from explicitly routed subnet does not get NATed => no Internet access.

https://www.tomaskubica.cz/post/2019/ipv6-v-azure/ -- run this through Google Translate from Czech

Routing Subnets in Azure

Routing works. Can we have some Global Unicast addresses, please?

```
rzajic@jump-vm:~$ ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
       valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
       valid_lft forever preferred_lft forever
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP group default qlen 1000
    link/ether 00:0d:3a:48:32:65 brd ff:ff:ff:ff:ff:ff
    inet 10.0.1.4/24 brd 10.0.1.255 scope global eth0
       valid_lft forever preferred_lft forever
    inet6 fd00:0:0:1::4/128 scope global dynamic noprefixroute
       valid_lft 17276869sec preferred_lft 8636869sec
    inet6 fe80::20d:3aff:fe48:3265/64 scope link
       valid_lft forever preferred_lft forever
rzajic@jump-vm:~$ sudo traceroute6 -n -I fd00:0:0:2::4
traceroute to fd00:0:0:2::4 (fd00:0:0:2::4) from fd00:0:0:1::4, 30 hops max, 60 bytes packets
 1  fd00:0:0:3::4  0.658 ms  0.484 ms  0.925 ms
 2  fd00:0:0:2::4  1.375 ms  1.479 ms  1.268 ms
rzajic@jump-vm:~$
```

OpenStack-based Cloud Providers

(Rackspace, etc.)

Neutron supports DHCPv6 PD, but it's up to the implementor if it works.
A quote from Rackspace: *"your expectations are not met by native functionality of our platform"*

/64 on-link from a /56 their router uses, an extra route for a gateway

unnumbered link, routed /64 via server's link-local address

Splitting a routed /64 on HETZNER ONLINE servers is simple

```
iface eth0 # network #1 from 2001:db8:deaf:f001::/64
    gateway fe80::1
    netmask 128
    address 2001:db8:deaf:f001::2

iface docker0 # network #2 from 2001:db8:deaf:f001::/64
    address 2001:db8:deaf:f001:dead::1
    netmask 80

iface docker1 # network #3 from 2001:db8:deaf:f001::/64
    address 2001:db8:deaf:f001:babe::1
    netmask 80

iface docker2 # network #4 from 2001:db8:deaf:f001::/64
    address 2001:db8:deaf:f001:c001::1
    netmask 80

# 65532 /80s remaining - Not Great, Not Terrible
```

Smaller hosting services can surprise you

(Hello, Mythic Beasts and Linode)

Likely an on-demand service, not available by default – go ask your host

Summary

	On-Link Addressing	Available Subnets	Quirks needed	GUA/ULA
AWS	DHCPv6	Up to /123+/124 on-link per interface, routable[1)]	Manual IP enumeration	GUA
Azure	DHCPv6	Docs claim 256 addrs per VM; routable ULA subnet works[2)]	NAT66, gai.conf tuning	ULA
Hetzner	Unnumbered	/64, assigned by default	No	GUA
OpenStack	DHCPv6/RA	Individual (DHCPv6-PD)	?	Depends
Mythic Beasts	On-Link /64	/48 on request	No	GUA
Linode	On-Link /112	/56 on request	No	GUA

1) Depends on the instance type; some instances can only have 2 IP addresses per interface
2) Azure Web Portal does not have full IPv6 support yet; I didn't manage to get the CLI to add extra IPs: it appears that a VM NIC with IPv6 cannot have multiple addresses assigned.

Q & A

Thank you

@zajDee