Talking IPv6 to IPv4 Without NAT

Alok Tiagi, Keerti Lakshminarayan, Kyle Anderson

(Netflix)
WHY

- Netflix uses 100k+ instances in AWS EC2
- We use a network architecture with non-overlapping CIDRs
- Testing the scaling limits of AWS
- Alternatives to the flat IPv4 network exist but...

- **Move to IPv6**
WHY

- Dual Stack doesn’t help
- How do we move from v4 to v6 transparently?

IPv6 vs IPv4 Flows Percentage Observed from EC2 Instances

How do we make this all Red?
Current state of the art

NAT (Network Address Translation)

- TAYGA: http://www.litech.org/tayga/
- NAT on dedicated hardware

Per Packet Processing! 😞

Is there a more optimal solution?
(Spoiler: TSA)
Titus Syscall Agent

Titus = Netflix’s Multi-tenant Container Platform

Powerful and efficient mechanism to intercept system calls.

- How does it work?
- TSA Architecture
- Demos
- Future
- Q&A
How does TSA’s Syscall Interception work?
Existing approaches

- LD_PRELOAD
- Kernel Modules
- Strace/Ptrace/Sysdig
- BPF
- Seccomp

What’s new?

- Seccomp Notifier (LPC’17)
- Seccomp Notifier + add_fd support (2020)

Credit to sargun@sargun.me (happens to work on our Team!)
Who likes Sequence Diagrams???
How does TSA intercept syscalls?
How does v6 only->v6 work?

An IPv6 destination? CONTINUE

IPv6 Container

IPv4

TSA

kernel

Titus Agent

IPv6 Server
V6 only -> v4 only (without TSA)?
v6 only -> v4 only with TSA?

Do I have a route to the V4 destination?

IPv6
Container

IPv4 Server

IPv4
TSA

kernel

Titus Agent
Demo: TSA
PDT: 14:00 - Saturday, May 15 2021

- connect: 3.91
- sendm_msg: 454m
- sendmsg: 7.23
- sendto: 5.05

Press + hold ctrl to view raw numbers
Future

- Kernel: “seccomp: epoll issues post addfd”
- TSA Transition on by default @Netflix
- LongTerm: IPv6 Mostly but TSA will get invoked for 3rd party IPv4 only servers
- TSA will continue to be used for advanced features like Perf, Bpftools
Thank You.