Networks for SREs: What do I need to know

Michael Kehoe

Staff SRE

Introduction

Michael Kehoe

$ WHOAMI

- Staff Site Reliability Engineer @ LinkedIn
- Production-SRE Team
- Funny accent = Australian + 3 years American
- Former Network Engineer at the University of Queensland

Agenda and Vision

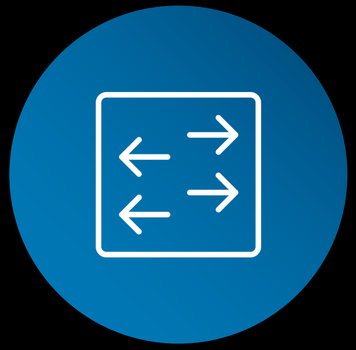
Today’s agenda

1	Introductions
2	Problem Statement
3	Basics of Networks
4	Advances in networks
5	Clos Networks
6	Advances in Network Speeds
7	IPv6
8	Summary

Networks just work right?

Probably…

Probably…Not…



Problem Statement

What are we trying to solve

- **Network Design** – Has evolved
- **Network software/ hardware** – Has advanced
- **Learning** – The average SRE may not necessarily understand the ramifications
- **Tooling** – Has been left behind

What this talk is

- Tale into potential pitfalls of modern day networks

What this talk isn’t

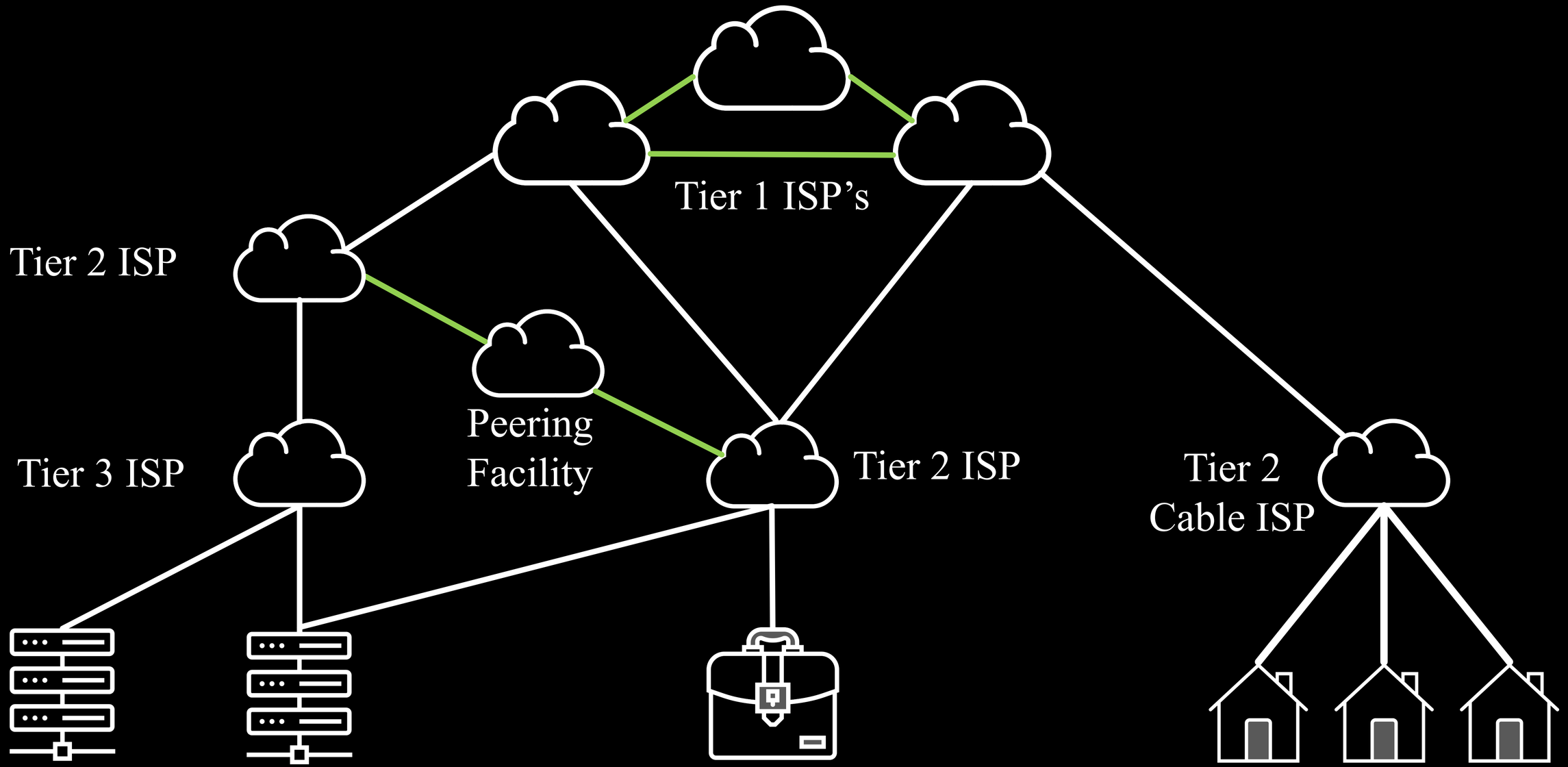
- How to make the network do all the things…quickly & reliably…

What this talk isn’t

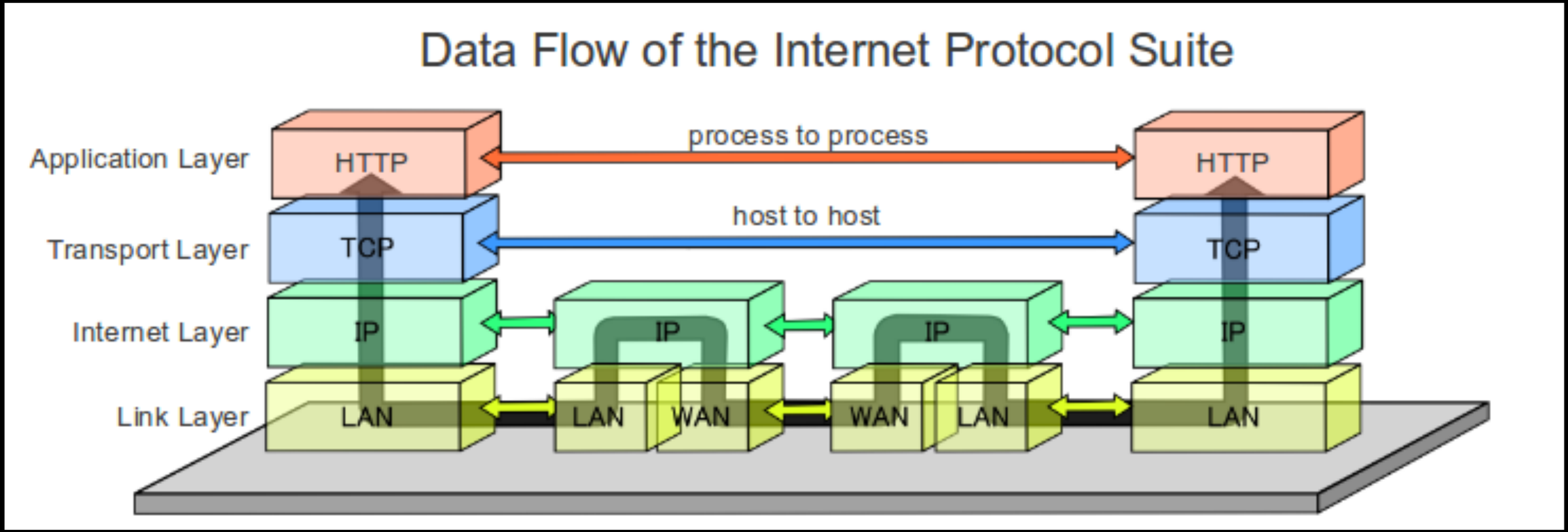
- How to make the network do all the things…quickly & reliably…
- Sorry

Basics of Networks

Basics of Networks

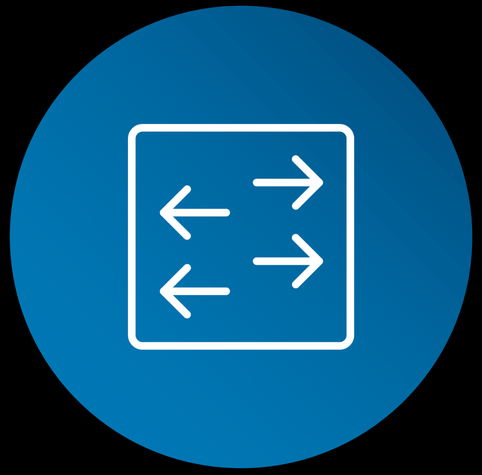


Basics of Networks



Advances in Network Design

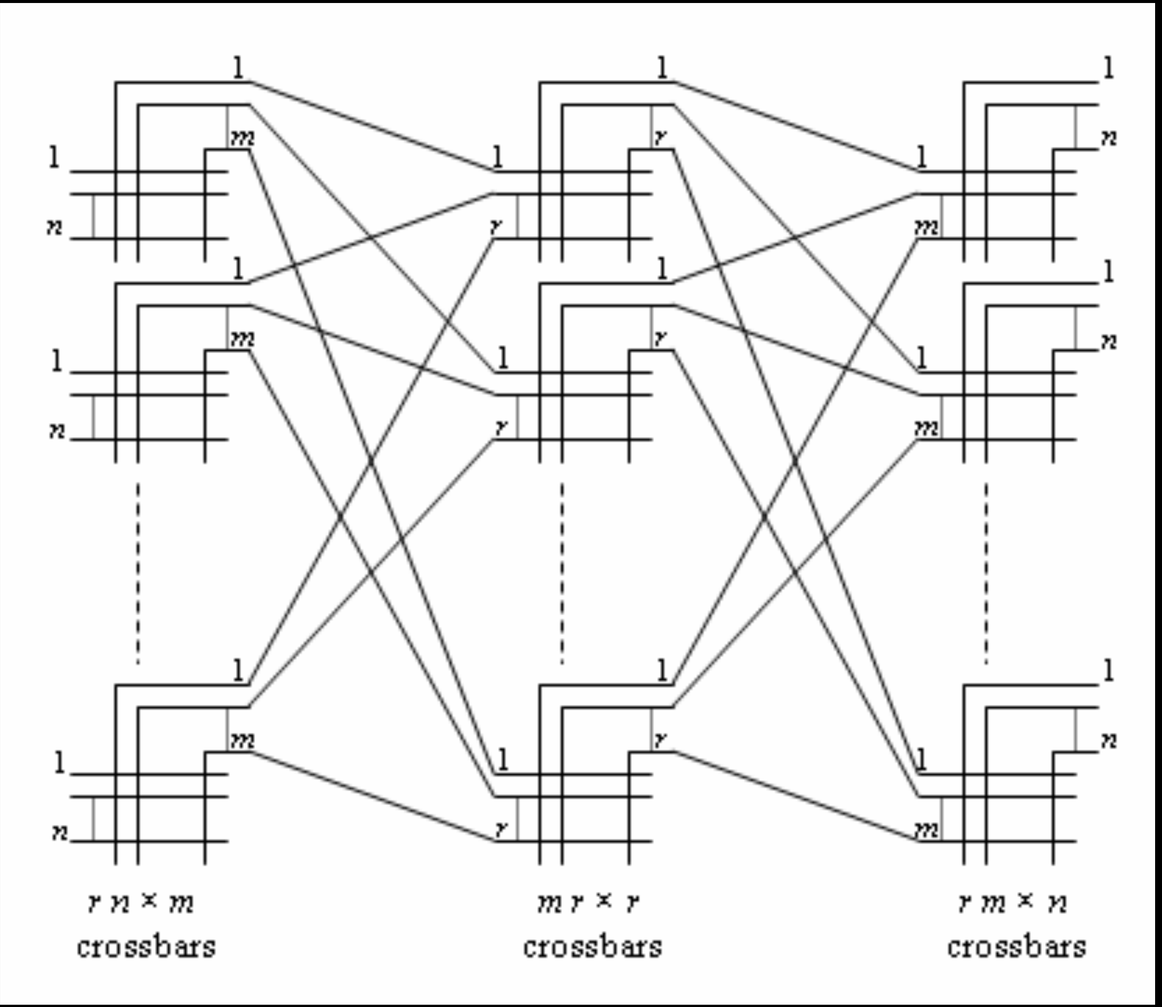
Advances in Network Design



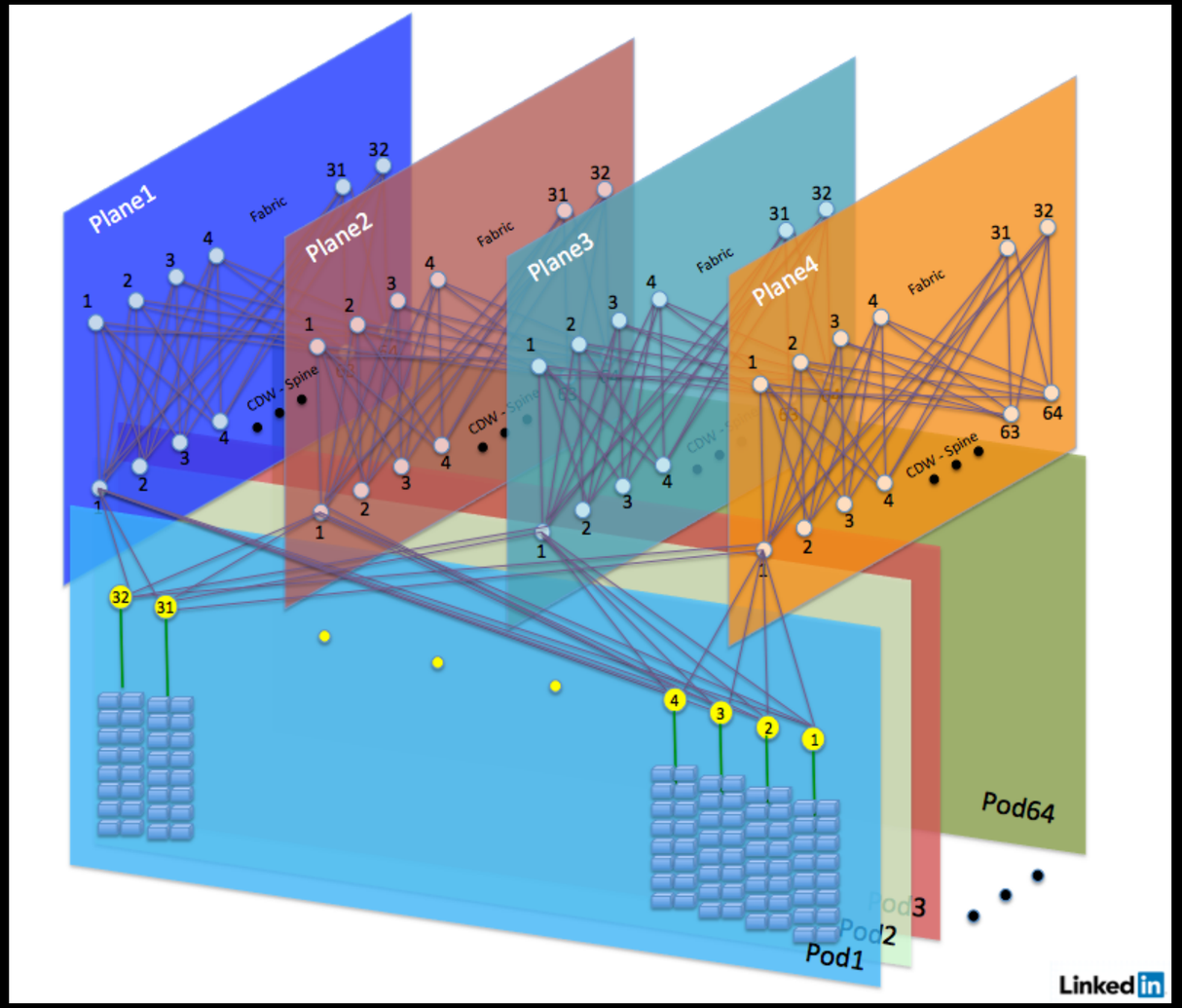
- Clos Networks
- Advancement of network speeds
- IPv6 Implementation (Finally)
- Multi-homed internet connections
- Moving away from traditional internal routing protocols

Clos Networks

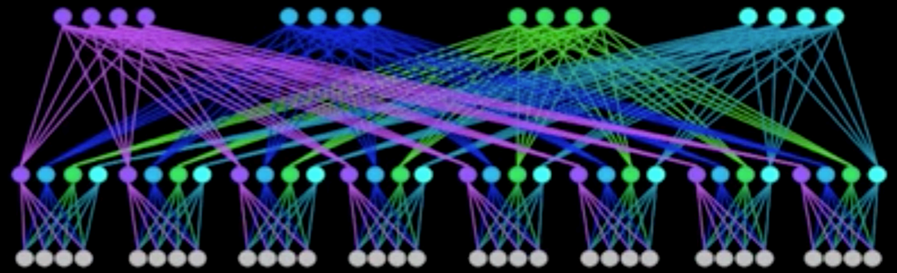
Clos Networks

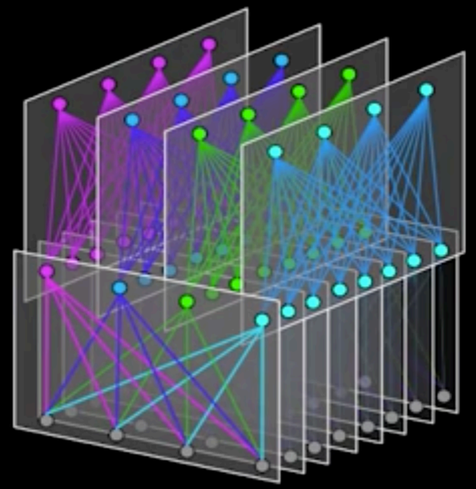


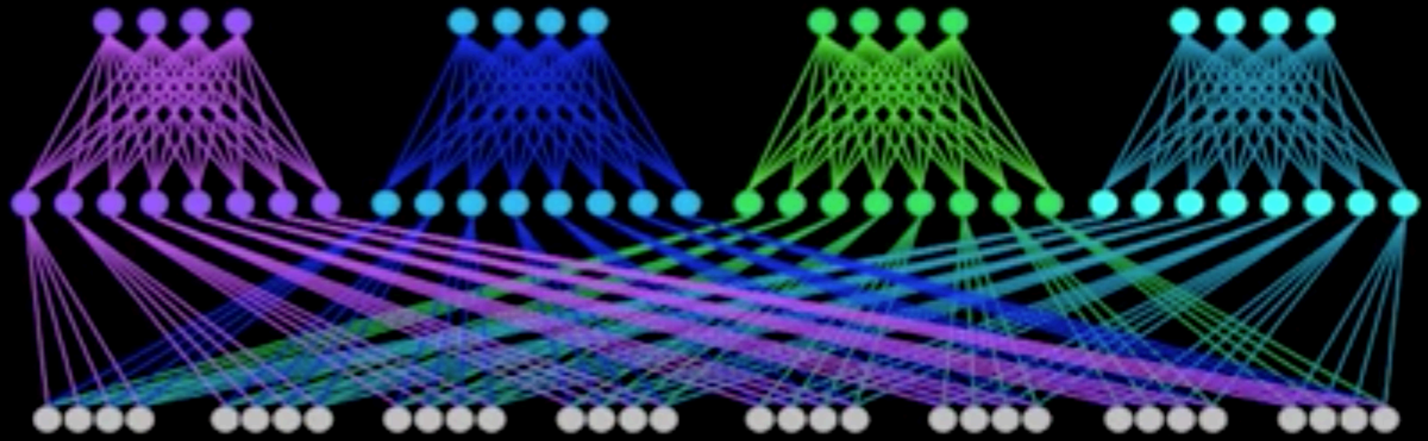
Clos Networks



Clos Networks

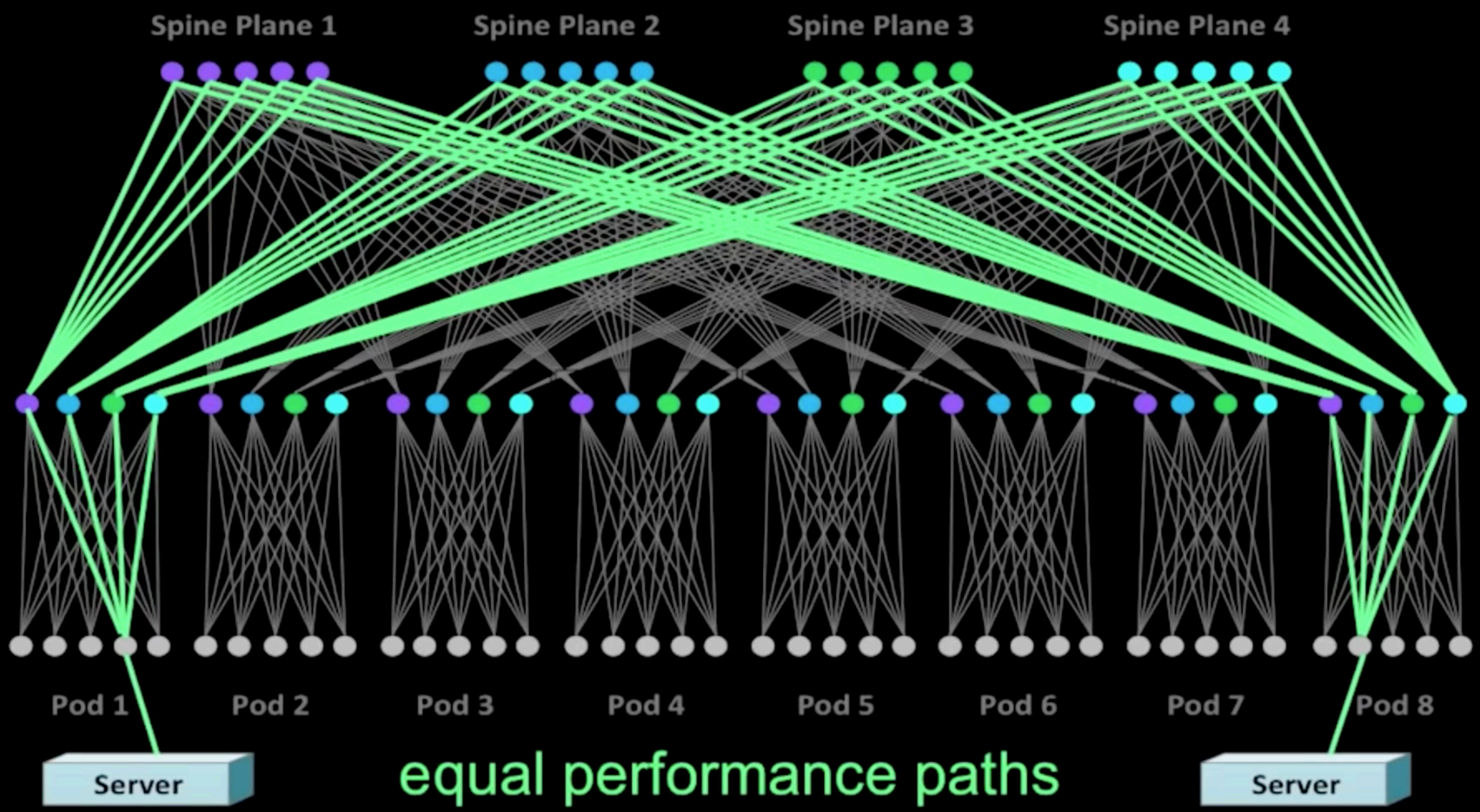






Credit: Facebook

Clos Networks



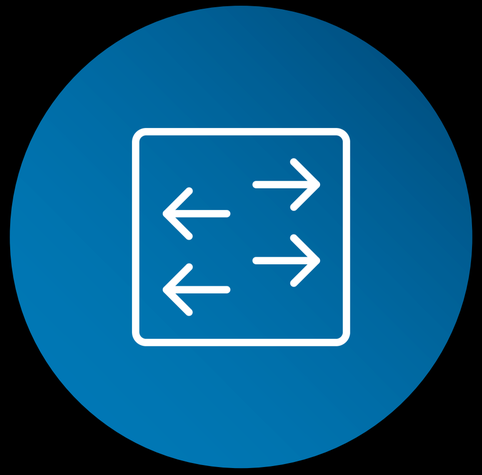
Credit: Facebook

Advancement of Network Speeds

Advancement of Network Speeds

Speed	Name	Standard	Year
10Mb	10BASE-T	802.3i	1990
100Mb	100BASE-TX	802.3u	1995
1000Mb = 1Gb	1000BASE-T	802.3ab	1999
10Gb	10GBASE	802.3ae	2002
40/100Gb	40GbE/ 100GbE	802.3ba	2010

Advancement of Network Speeds



- What this gives us
  - Better transfer bulk speeds
  - The ability to have higher concurrency services (1M connection problem)
  - Run multiple high-concurrency applications (LPS)

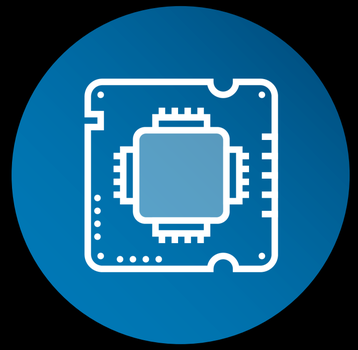
Networks just work right?

Probably…

Probably…Not…

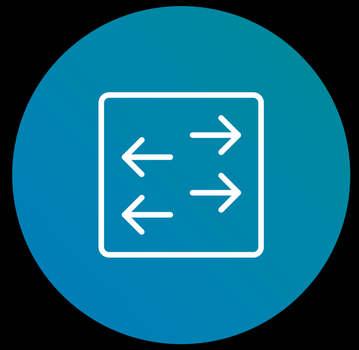
Optimizations Required

Advancement of Network Speeds



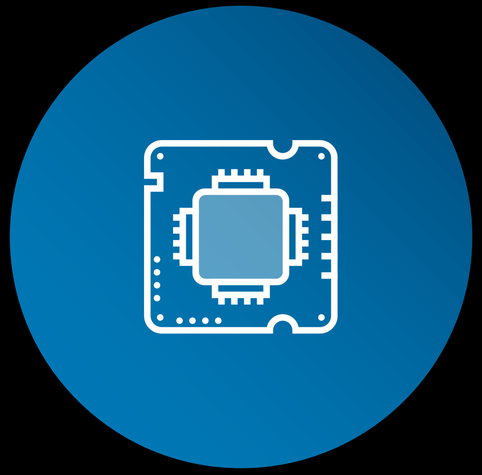
NIC

Linux Kernel



Network Switches

Advancement of Network Speeds

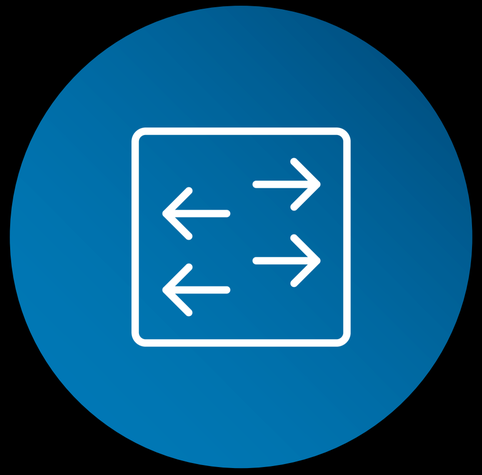


- Network Interface Cards
  - Various RX/ TX queue size limits/ defaults
  - Various interrupt schemes
  - Plethora of tunables that vary wildly
  - LITTLE TO NO DOCUMENTATION!
  - How do you monitor/ tune it???

Advancement of Network Speeds

- Linux Kernel
  - Lots of network tunables
  - Some defaults assume year ~2000 era hardware
    - E.g. *net.ipv4.tcp_max_syn_backlog*
  - Important to understand the type of application you run and cater your tunables to that.

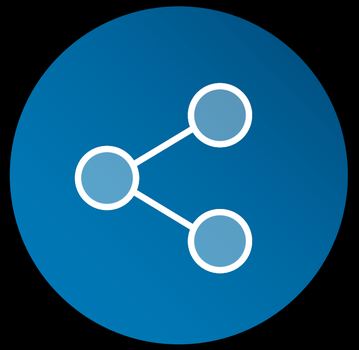
Advancement of Network Speeds



- Network switches
  - Similarly to interfaces and Linux software, there’s a lot of options
    - Deep Buffers
    - DSCP marking
    - Switching latency
    - DCTCP

Adoption of IPv6

IPv6 Features



Address Space

Simplified Header

No-NAT

Auto-Configuration

Better Performance

IPv6: Address Space

- Moving from a 32-bit address space to 128-bit.
  - 4B → 340TTT
- Read up on IPv6 addressing representation
  - RFC-5952

IPv6: Address Space

A SINGLE ADDRESS CAN BE REPRESENTED MANY WAYS

2001:db8:0:0:1:0:0:1
2001:0db8:0:0:1:0:0:1
2001:db8::1:0:0:1
2001:db8::0:1:0:0:1
2001:0db8::1:0:0:1
2001:db8:0:0:1::1
2001:db8:0000:0:1::1
2001:DB8:0:0:1::1

IPv6: Address Space

YOU CAN MAKE FUN PHRASES

- :cafe:beef
- :feed:f00d:
- :bad:f00d:
- :bad:beef:
- :bad:d00d:
- :f00d:cafe:
- :bad:fa11:

IPv6: Address Space

OR CLEVER ADVERTISING

```
[mkehoe@mkehoe ~]$ host -6 www.facebook.com
www.facebook.com is an alias for star-mini.c10r.facebook.com.
star-mini.c10r.facebook.com has IPv6 address
2a03:2880:f113:8083:face:b00c:0:25de
```

IPv6: Address Space

SPECIAL ADDRESSES: IPV4

RFC	IP Block	Use
1918	10.0.0.0/8 172.16.0.0/16 192.168.0.0/16	Private IP Addressing
6890/ 3927	169.254.0.0/16	Link-Local
5771 2365	224.0.0.0/4	Multicast

IPv6: Address Space

SPECIAL ADDRESSES: IPV6

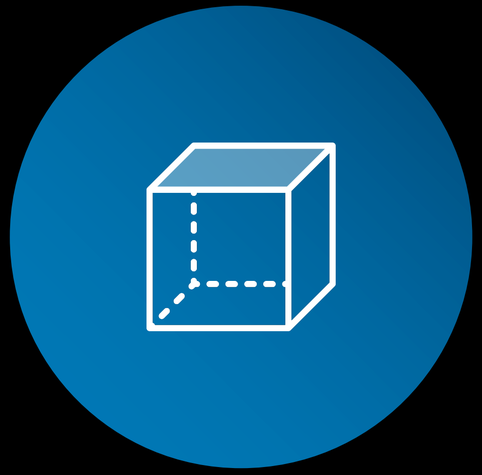
IP Block	Use
::/128	Unspecified Address
::1/128	Loopback address
::ffff:0:0/96	IPv4 mapped addresses
64:ff9b::/96	IPv4/ V6 translation
fc00:::/7	Unique Local Address
fe80::/10	Link-Local address
ff00::/8	Multicast addresses

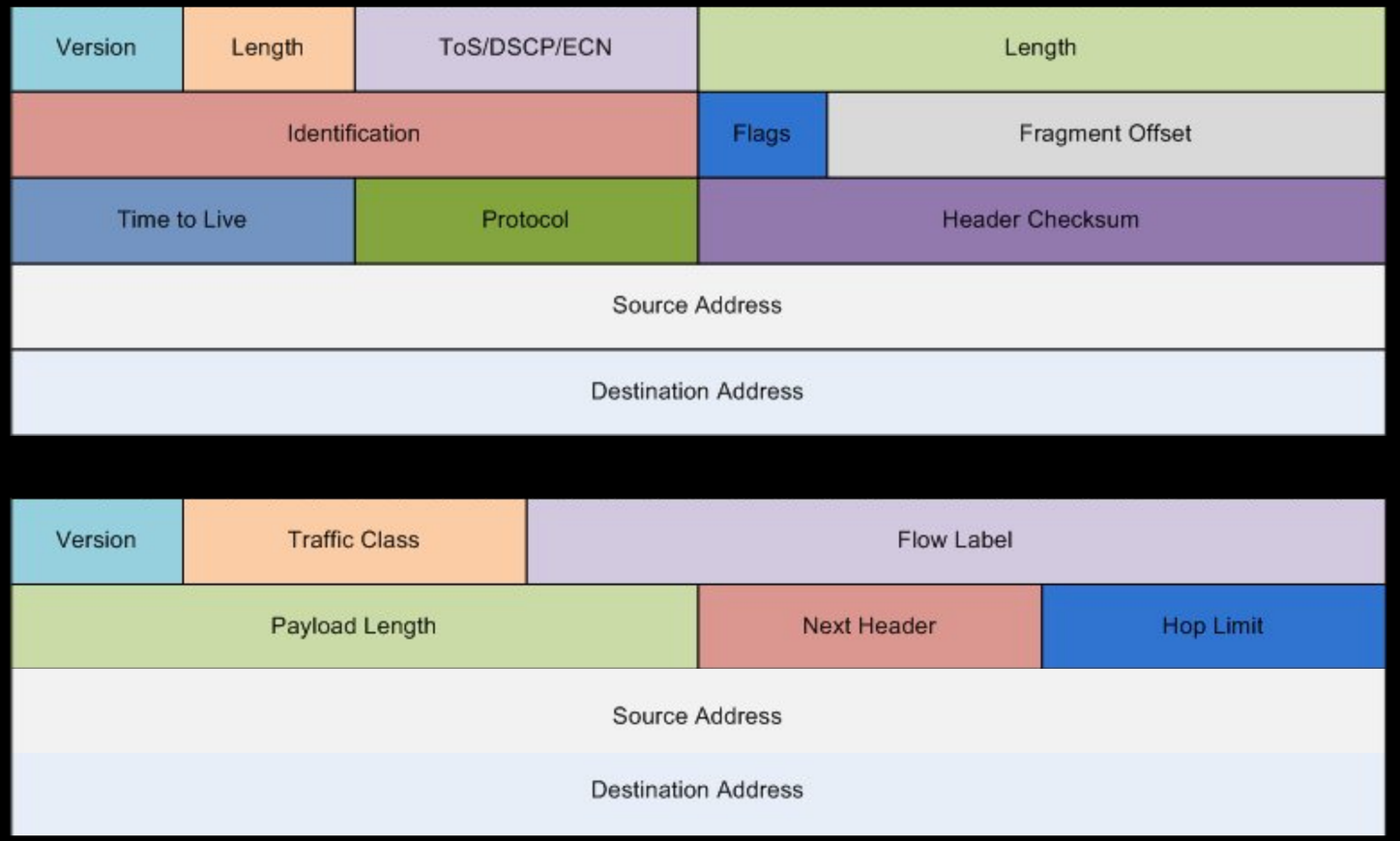
IPv6: Address Space

OR CLEVER ADVERTISING

```
[mkehoe@mkehoe ~]$ host -6 www.facebook.com
www.facebook.com is an alias for star-mini.c10r.facebook.com.
star-mini.c10r.facebook.com has IPv6 address
2a03:2880:f113:8083:face:b00c:0:25de
```

IPv6: Simplified Header





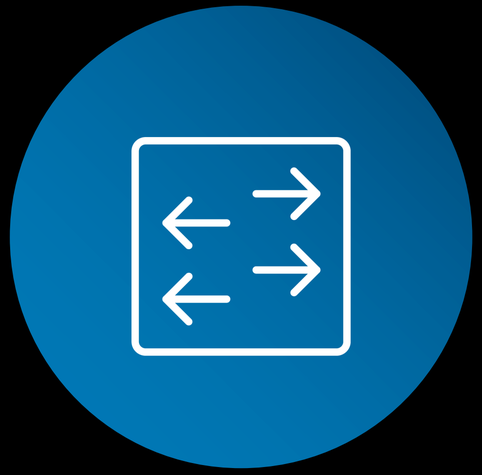
IPv6: No NAT

- No need for NAT anymore
  - Simplified Configuration
  - Less points-of-failure
- Potential for better performance
  - NAT is slow
- Harder for abusers to hide behind NAT

IPv6: Auto-Configuration

- Stateless = Auto-Configured
- Stateful = DHCP/ Statically assigned

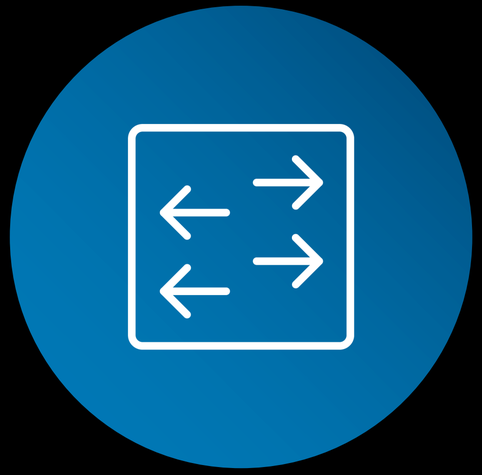
IPv6: Better Performance



- The elimination of NAT is a significant factor
- *Generally* less hops across the internet for IPv6 vs IPv4
- Simplified Header gives small amount of optimization

Summary

Summary



- Don’t implicitly trust the network!
- Understand where your packets flow
- End-to-End monitoring of your network. It is the lifeblood of your infrastructure
- For any network infrastructure changes, ensure you understand how to benchmark and monitor it!

Networks just work right?

Q&A


LinkedIn