How to compile a Linux Kernel for your 100Gbps router.

Or, how to keep doing Software Routing until you really have to start using an ASIC, with open source software.

WC ZANOG - 23 Feb 2023, a follow up to Feb 2019 talk





Joe Botha W

Frogfoot - when it was an ISP Teraco - DC **Octotel - FNO Atomic Access - Fibre ISP in Cape of Good Hope** https://www.atomic.ac/





My quick 20 year history with (Software) Routing

2000 - Frogfoot's 1st router: x86 PC 2006 - Rackmount x86 & wireless 2018 - Atomic software routing

2023 - Atomic software + ASIC routing



Read: 'The world in which IPv6 was a good design'

<u>witter.com/swimgeek</u> IETF vs IEEE & Routing vs Switching Route packets with a CPU, until you can't.



2000 - Frogfoot's 1st routers: x86 PC 🥯



2006 - x86, but embedded / rackmount





Open Networking History 2013 - Cumulus & ONIE 🥪 2015 ~ Netflix, CDNs and rapid traffic growth 2016 - Mellanox & SwitchDev & Spectrum1





2018 - Atomic software routing

- Xeon D & Intel 140e - Debian 9 - FRR3

vs commercial options / Arista 7280







2022 - Atomic recent routing

- Xeon D & Intel i40e & SR-IOV - Debian 10 & Proxmox - FRR7

Limits: softIRQ/core ~8Gbps with 8 cores Intel NIC drivers are not great Port density





Open Networking with ASICs 2018-2021 - found nothing really nice 6

DPDK & VPP, OVS, Vyos etc
Broadcom, IPinfusion etc
ONL and lots more...
Mellanox NICs & Switches



"ASICs, magic and pro wrestling are closely guarded secrets"

Avoid ASICs with NDAs and buggy / closed SDKs





Could you build a 100Gbps router with Debian?



* Slide from 2019



2023 - Atomic routing, sw & hw 📀

Border: Xeon D & Mellanox DX6 NICs Peering/BNG: SN2010 Mellanox NVIDIA

Debian 11
FRR8





What you need to compile: 💮 - Linux kernel 6.1 with SwitchDev - Various hw sensors - hw-mgmt - ethtoo - iproute2



Linux cpt-ter-rs1 6.1.13-atomic #1 SMP PREEMPT_DYNAMIC Fri Feb 24 08:24:16 SAST 2023 x86_64 GNU/Linux root@cpt-ter-rs1 ~ # cat /etc/debian_version 11.6 root@cpt-ter-rs1 ~ # ethtool -i swp20 driver: mlxsw_spectrum version: 1.0 firmware-version: 13.2010.4026 expansion-rom-version: bus-info: 0000:01:00.0 supports-statistics: yes supports-test: no supports-eeprom-access: no supports-register-dump: no supports-priv-flags: no root@cpt-ter-rs1 ~ # ethtool swp20 Settings for swp20: Supported ports: [FIBRE] Supported link modes: 1000baseKX/Full 10000baseKR/Full 40000baseCR4/Full 40000baseSR4/Full 40000baseLR4/Full 25000baseCR/Full 25000baseSR/Full 50000baseCR2/Full 100000baseSR4/Full 100000baseCR4/Full 100000baseLR4_ER4/Full





- purist v debian - most open, no lame NDAs - small, low power - pretty good port density - peering traffic at ASIC speeds - can do cool things with tc rules ...not really cheaper





- 140k IPv4, 30k IPv6, 8k MAC (256k) - 10 / 25 / 40 / 100 Gbps - 57 Watt power - 8G memory, 256G storage (DIY) - 1.3 Bpps - 16MB buffers - 300ns latency



What's missing?

- Deep buffers - 2N routes - yes, you have to filter HE routes





- Cumulus vs Debian - Switchd vs Switchdev - Sectrum1 vs Spectrum2 - All 10G ports - Upgrades

Ops: Linux sysadmin vs ~Cisco config

