Open. Together.
Minipack and F16, Software

Alex Eckert, Software Engineer
Facebook
Minipack and F16, Software
Software everywhere

Management plane
  Control plane
  Data plane
  Platform
FBOSS Overview

- Monitoring Daemons
- Config Management
- System Tools & Libraries
- OpenNSL
- Linux Kernel
- FBOSS Agent
- Routing Daemon
- OpenBMC

Wedge
- OCP Micro Server
- Broadcom Trident II
- BMC
Sixpack

128 x 40G - FSW
Wedge 100

32 x 100G - RSW
Backpack

FBoss

128 x 100G - FSW
Wedge 100S

32 x 100G - RSW
Fabric Aggregator

FBOSS

N x 100G - FA
Minipack

FBOSS

128 x 100G – FSW, SSW, FA
New challenges (Minipack)

- Modular switch, single control plane
- Different PIM types, variety of port speeds
- External phy support (Broadcom)
- I2C to 128 optics, MDIO, FPGA
- New microserver (MiniLake), New ASIC (Broadcom TH3)
- Cooling 128 optics in only 4 RU
Arista 7368X4

128 x 100G – FSW, SSW, FA
New challenges (FBOSS on Arista)

- First time operating FBOSS on non-FB hardware
- Different external phy (Credo), Different FPGA
- Unfamiliar microserver, no off-the-shelf UEFI BIOS
- Must manage as either FBOSS or EOS in production
- Complex conversion process b/w EOS / FBOSS
Combinations

- Two switch models
- Modularity
- Three layers of F16 network
- Same binary everywhere
Common abstractions
New microservers!

MiniLake
New microservers!

Arista control module
OCP Open System Firmware

System Firmware on boot ROM

- coreboot
  - silicon and DRAM initialization

- Linux
  - device drivers, network stack, multi-user multi-tasking environment

- u-root/systemboot
  - userspace tools and bootloader

Target OS

CentOS
Conversion
FBOSS Overview

- Monitoring Daemons
- Config Management
- System Tools & Libraries
- OpenNSL
- Linux Kernel
- OpenBMC
- Minipack
- MiniLake
- Tomahawk III
- BMC
Minipack Hardware Components

Minipack
- MiniLake
- Tomahawk III
- BMC
- IOB FPGA

PIM
- DOM FPGA
- External phys... x8
- Transceivers
Data path

Minipack
- MiniLake
- IOB FPGA

PIM
- DOM FPGA
- External phys
- Transceivers

Tomahawk III
- BMC

... x8
Management path – MDIO

Minipack

MiniLake

ASIC

BMC

IOB FPGA

PIM

DOM FPGA

External phys

... x8

Transceivers
Programming External PHYs

- ASIC

- Transceivers

- BCM 81724 (Gearbox)
  - Channel count: 16!
  - Channel speed: <=50
  - Medium: Copper Backplane
  - Modulation: PAM4/NRZ

- Channel count: 16!
- Channel speed: <=50
- Medium: Copper Backplane
- Modulation: PAM4/NRZ
Programming External PHYs

- ASIC
- Transceivers

**BCM 81724 (Gearbox)**

- Channel count: 16!
- Channel speed: <=50
- Medium: Copper Backplane
- Modulation: PAM4/NRZ
Programming External PHYs

ASIC

Transceivers

BCM 81724 (Gearbox)

- Channel count: 16!
- Channel speed: <=50
- Medium: Copper Backplane
- Modulation: PAM4/NRZ

- Channel count: 16!
- Channel speed: <=50
- Medium: Copper Backplane
- Modulation: PAM4/NRZ
State management problem
Switch State

- PortMap
  - Port 1
  - ... (or ellipsis)
  - Port N
- VlanMap
  - Vlan 1
  - Vlan 2
- ArpTable
  - ArpTable'
  - ArpTable'
- RouteTable Map
  - RouteTable 0

Switch State

VlanMap'

Vlan2'

ArpTable'
FBOSS Agent

- State Update Queue
- BcmSwitch
- SwSwitch

HwSwitch Interface
FBOSS Agent

State Update Queue

SwSwitch

BcmSwitch

HwSwitch Interface

“Platform”

PlatformPort
Platform Configuration

- Each port + speed has unique settings (PlatformPortSettings)
- Model config as a map<speed, PlatformPortSettings>
- Hides complexity from SwSwitch
Testing
Testing strategy

- Targeted tests in realistic lab environments
- Deploy early
- Invest in automation
Scale challenge
Test all combinations
Deploy early
Invest in automation

- Automate as much as possible of our lab testing
- Develop targeted ASIC data plane tests

Test specific ASIC functionality, verify, warm boot
Utilize common FB testing infra

- netcastle_test/fboss_bcm/tomahawk_alpm/6.4.10-4 6.5.13-1 - AclEmptyCodeIcmp (warm_boot.BcmAclQualifierTest)
  architecture: x86_64, buildsystem: buck, compiler: gcc, sanitizer: none
  • Good • fboss_test_infra • 12.81 sec

- netcastle_test/fboss_bcm/tomahawk_alpm/6.4.10-4 6.5.13-1 - AclIcmp6Qualifiers (cold_boot.BcmAclQualifierTest)
  architecture: x86_64, buildsystem: buck, compiler: gcc, sanitizer: none
  • Good • fboss_test_infra • 416.86 sec

- netcastle_test/fboss_bcm/tomahawk_alpm/6.4.10-4 6.5.13-1 - AclIp4TcpQualifiers (warm_boot.BcmAclQualifierTest)
  architecture: x86_64, buildsystem: buck, compiler: gcc, sanitizer: none
  • Good • fboss_test_infra • 12.89 sec

- netcastle_test/fboss_bcm/tomahawk_alpm/6.4.10-4 6.5.13-1 - AclMirror (cold_boot.BcmMirrorTest/1)
  architecture: x86_64, buildsystem: buck, compiler: gcc, sanitizer: none
  • Good • fboss_test_infra • 71.08 sec

- netcastle_test/fboss_bcm/tomahawk_alpm/6.4.10-4 6.5.13-1 - AclMirror (warm_boot.BcmMirrorTest/1)
  architecture: x86_64, buildsystem: buck, compiler: gcc, sanitizer: none
  • Good • fboss_test_infra • 8.55 sec

- netcastle_test/fboss_bcm/tomahawk_alpm/6.4.10-4 6.5.13-1 - AclMirror (cold_boot.BcmMirrorTest/1)
  architecture: x86_64, buildsystem: buck, compiler: gcc, sanitizer: none
  • Good • fboss_test_infra • 1182.54 sec
Takeaways

- Minipack is a powerful modular building block for our networks.
- Software support for modularity has its challenges.
- We were able to overcome these challenges through improved hardware layer abstractions and investing heavily in early deployment and automated testing.
Thank you
Open. Together.

OCP Global Summit | March 14–15, 2019