Open. Together.
Minipack and F16, Software

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

Management plane

Control plane

Data plane

Platform
Wedge

FBOSS

16 x 40G - RSW
FBOSS Overview
Sixpack

FBOSS

128 x 40G - FSW
Wedge 100

32 x 100G - RSW
Backpack

128 x 100G - FSW
Wedge 100S

32 x 100G - RSW
Fabric Aggregator

N x 100G - FA
Minipack

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

coreboot

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

u-root/systemboot
userspace tools and bootloader

Target OS
CentOS
Conversion

ARISTA EOS  ↔  FBOSS
FBOSS Overview

- Monitoring Daemons
- Config Management
- System Tools & Libraries
- Linux Kernel
- FBOSS Agent
- Routing Daemon
- OpenNSL
- OpenBMC
- Minipack
- MiniLake
- Tomahawk III
- BMC

Open. Together.
Data path

Minipack

IOB FPGA

DOM FPGA

PIM

Transceivers

External phys

... x8

MiniLake

Tomahawk III

BMC
Management path – MDIO

Minipack
- MiniLake
- ASIC
- BMC

PIM
- IOB FPGA
- DOM FPGA
- External phys
- Transceivers

... x8
Programming External PHYs

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

ASIC

BCM 81724 (Gearbox)

Transceivers

Open. Together.
Programming External PHYs

- ASIC

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

- Transceivers
Programming External PHYs

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

**ASIC**

**Transceivers**

**BCM 81724 (Gearbox)**

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

- **SwitchState**
  - PortMap
    - Port 1
    - Port N
  - VlanMap
    - Vlan 1
    - Vlan 2
  - ArpTable
    - ArpTable
  - RouteTable
    - RouteTable Map
      - RouteTable 0

- **SwitchState’**
  - VlanMap’
    - Vlan2’
      - ArpTable’
FBOSS Agent

SwSwitch

HwSwitch Interface

BcmSwitch

State Update Queue
FBOSS Agent

State Update Queue → SwSwitch → BcmSwitch → "Platform"
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

Fabric Aggregator
- Minipack + FBOSS
- 7368X4 + FBOSS
- 7368X4 + EOS

Spine Switch
- Minipack + FBOSS
- 7368X4 + FBOSS
- 7368X4 + EOS

Fabric Switch
- Minipack + FBOSS
- 7368X4 + FBOSS
- 7368X4 + EOS
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

- Good: fboss_test_infra • 12.81 sec
- Good: fboss_test_infra • 416.86 sec
- Good: fboss_test_infra • 12.89 sec
- Good: fboss_test_infra • 71.08 sec
- Good: fboss_test_infra • 8.55 sec
- 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