Quick Tofino Overview
For RISC-V Meeting
Curt Beckmann
April 16, 2020
Barefoot, an Intel company
“This is *how I know* to process packets”  
(i.e. the ASIC datasheet makes the rules)
“This is *how I want* the network to behave and how to switch packets…”
(the user / controller makes the rules)
P4 Programming, the complicated version…

- P4 code
- Compiler
- Tofino
Fixed Function Switches

Switch OS

Static Run-time API

Driver

Fixed Function ASIC

Fixed Meta-Data

Traffic Manager

Fixed Parser

Fixed Lookups

Fixed Actions

Fixed Memories

Fixed Lookups

Fixed Actions

Fixed Memories

Fixed Lookups

Fixed Actions

Fixed Memories

Fixed Lookups

Fixed Actions

Fixed Memories

Fixed Packet Mods
Programmable Switch Approach

1. Protocol Authoring
   - switch.p4

2. Compile

3. Configure

Programmable Meta-Data

Flexible Parser
- Flexible Lookups
- Custom Actions
- Shared Memories

Traffic Manager
- Flexible Lookups
- Custom Actions
- Shared Memories

Flexible Lookups
- Custom Actions
- Shared Memories

Flexible Lookups
- Custom Actions
- Shared Memories

Flexible Lookups
- Custom Actions
- Shared Memories

DeParser
- Custom Actions
- Shared Memories
Result: Customer Defined Switch

Switch OS

Add/delete table rules

Auto Generated Run-time API

Driver

Programmable ASIC: User Defined Forwarding Plane

User Defined Meta-Data

User Defined Parser
User Defined Lookups
User Defined Tables
User Defined Actions

Traffic Manager

User Defined Lookups
User Defined Tables
User Defined Actions

User Defined Lookups
User Defined Tables
User Defined Actions
Match-Action Packet Processing Concept
Programmable Parser

- Produces 4kb Packet Header Vector
  - Input to match-action processing
- All match/action fields are generic
- No specific protocols are embedded in the switch design
Match+Action Pipeline

12 to 20 identical HW stages

Match input crossbar

RAM for match, action, statistics, meters shared between ingress and egress

- Exact match tables
- Ternary match tables
Switch state metadata

• Act on, filter, export internal metadata on per packet basis
  ◦ Perform lookups, arithmetic & boolean operations
  ◦ Embedded in packet header (Inband Network Telemetry)
  ◦ Accumulated in digest packet

• Some metadata examples:
  • Packet Tuple Hash
  • Drop status
  • Rx timestamp
  • Parser Errors

  • Enq/Deq Q Depth
  • Pkt Length
  • Queuing Delay
Sample creative applications

• SwitchML, does parameter aggregation *in the network*
  ◦ This has been done in the lab

• Advanced congestion control: Tofino extends RoCE protocol
  ◦ Done by cloud vendor on Tofino, see HPCC

• Telemetry: INT
More questions?
Backup