Stratum / P4 Update & Google’s Use Case

Brian O’Connor, ONF
Devjit Gopalpur*, Google

ONS North America - April 3, 2019

*On behalf of many at Google (Alireza Ghaffarkhah, Waqar Mohsin, Shashank Neelam, Jim Wanderer, Lorenzo Vicisano, Amin Vahdat, …)
Next Generation SDN picture

Disaggregated and Distributed Control and Management Plane

- SDN Control Services
- Configuration Services
- Monitoring & Telemetry Services
- Admin & Orchestration Services

Open Source Switch OS

P4Runtime

Open Config

Open NOS

gNMI

OSS / BSS

Inventory

Global Orchestrator
Next Generation SDN Interfaces

Embedded System

- Pipeline Definition
  - P4 Program
- Pipeline Control
  - P4Runtime
- Configuration & Telemetry
  - OpenConfig over gNMI
- Operations
  - gNOI

Northbound

Packets

Forwarding Chip
Stratum High-level Architectural Components

Remote or Local Controller(s)

- P4 Runtime
- gNMI
- gNOI
- Switch Broker Interface
  - Table Manager
  - Node/Chip Manager
  - Chassis Manager
  - Platform Manager
  - Chip Abstraction Managers
    E.g. ACL, L2, L3, Packet I/O, Tunnel

Switch SDK
- Platform API
- Switch Chip(s)
- Platform Drivers
- Peripheral(s)

Kernel

User

Common (HW agnostic)
- Chip specific
- Platform specific
- Chip and Platform specific

Stratum switch agent

PI and fpm-based implementations
1. Flow programming using **P4Runtime**
   - Feature parity with OpenFlow on programmable ASICs
   - Interoperability with ONOS and Trellis
   - Demoed at: **ONS Europe 2018, GNTC 2018, ONF Connect 2018**

2. Configuration and operations using OpenConfig’s **gNMI / gNOI**
   - Support for interface and peripheral config and telemetry
   - Demoed at: **OCP 2019, ONS North America 2019**

3. P4Runtime support on **fixed-function and programmable** ASICs
   - gNMI and gNOI for config and telemetry (goal: zero-touch provisioning)
   - Demo planned for H2 2019 (ONF Connect 2019)
**Value Add:** Inband-Network Telemetry (INT)

- Idea: use same data packets to carry data plane state
- Use P4 programmability to implement INT in the switch

---

**Add:** switch ID, arrival time, departure, queue delay, etc

**Generate report with switch metadata**

---

**Log, analyze, replay, visualize**
Value Add: VNF Offloading

- Implement Network Functions in HW as part of the switch pipeline
  - This demo: Mobile core S/PGW user plane (GTP encap/decap)
- Many benefits
  - Increased performance - VNFs executed at switch line rate, i.e. O(Tbit/s)
  - Reduced latency and jitter - Avoid non-determinism of x86 processing
  - Reduced OpEx - Less CPU resources, less power consumption

<table>
<thead>
<tr>
<th>Progr. ASIC capabilities</th>
<th>VNF building blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arbitrary header parsing/deparsing</td>
<td>Domain specific encap/decap (e.g. PPPoE termination, GTP, etc.)</td>
</tr>
<tr>
<td>Stateful memories</td>
<td>TCP connection tracking (L4 load balancing, NAT, firewall, etc.)</td>
</tr>
<tr>
<td>Computational capabilities</td>
<td>Billing</td>
</tr>
</tbody>
</table>

P4 switch programmable ASIC
ONS 2019 Demo overview

Trellis apps
(Segment routing, multicast, vRouter, etc)

ONOS

P4Runtime → gNMI → gNOI

Multi-vendor white-box switches
Edge-Core, Inventec, Delta

100k IPv4 routes

Flow programming scale-up; configuration and operations; multi-vendor hardware
Hardware Used for ONS Demo

**Delta AG9064v1**  
Barefoot Tofino 6.5Tbps  
64 x 100G QSFP28

**Edgecore Wedge 100BF-32X**  
Barefoot Tofino 3.2Tbps  
32 x 100G QSFP28

**Edgecore Wedge 100BF-65X**  
Barefoot Tofino 6.5Tbps  
64 x 100G QSFP28

**Inventec D5254**  
Barefoot Tofino 1.8Tbps  
6 x 100G QSFP28 + 48 x 25G SFP28
Stratum Community Milestones

May 2018 - Pioneer phase kick-off with initial code contribution from Google
August 2018 - Initial community switch support
March 2018 - Start of member preview phase with 4 HW platforms and one SW switch

Community Growth
16 founding participants, now 27 Stratum member companies
~130 participants (~20 are active contributors)
Stratum Community
Google’s History

Google runs SDN networks at scale

**Espresso**
SDN Peering Edge / Metro
70 metro sites
25% of all Internet traffic

**Jupiter**
SDN Data Center
1.3 Pbps
100,000+ servers/site

**B4**
SDN WAN
Inter-datacenter traffic
Growing faster than Internet traffic

Cisco Global Internet Forecast:
~150 EB/month in 2018 (+ 24% from 2017)

https://www.blog.google/topics/google-cloud/making-google-cloud-faster-more-available-and-cost-effective-extend-sdn-public-internet-espresso/
Google’s Approach to Multi-Vendor SDN

- Heterogeneous network
- Single consistent API
  - P4Runtime
  - OpenConfig
- Exploit unique HW capabilities (without changing the interfaces)
- Leverage commercial technology / vendors
  - Networking Vendors
  - ODMs
  - In-house / OEMs
Requirements for Multi-Vendor SDN

- **Support for vendor-neutral** control applications
  - Control plane is written once, compiled for multiple backends, i.e. hardware.

- **Support for programmable hardware**
  - Pushes hardware abstraction up the stack.
  - Uniform runtime interface for heterogeneous devices and network roles.

- **Support for a uniform network model**
  - Vendor-agnostic model of topology.
  - Simplifies operability of a multi-vendor network.
... which also provides ...

- **Enhanced deployment velocity** at scale
  - Introduction of new functionality, hardware, etc. using common workflows.
  - Incremental support for new equipment.
- **Simplified migration** of services
  - From traditional devices to programmable devices.
  - Between heterogeneous device blocks.
- **Unified device management**
  - Operators use common tools to deploy, configure, monitor and troubleshoot devices from multiple vendors.