OPEN SOURCE NETWORKING DAYS

Productize programmable network infrastructure

Yi Tseng
MTS, Open Networking Foundation
Live Q&A and polls

Scan me to join live Q&A

Or join from the link:
Outline

- An overview of Aether project
- Aether edge P4-based disaggregated UPF
- Productize programmable network infrastructure
Yi Tseng
Member of Technical Staff
Open Networking Foundation
2017: Intern
- ONOS
- fabric.p4
- M-CORD
2018-now: MTS (PDP Team)
- Stratum
- Fabric.p4
An overview of Aether project
ONF has history of successfully driving disaggregation and SDN
ONF has history of successfully driving disaggregation and SDN
ONF has history of successfully driving disaggregation and SDN.
ONF has history of successfully driving disaggregation and SDN.
Disaggregation and virtualization for mobile networks

- Base Station
- Mobile Core
- RAN
Disaggregation and virtualization for mobile networks

Mobile core control-user plane separation (CUPS)
Disaggregation and virtualization for mobile networks

Mobile core control-user plane separation (CUPS)
Disaggregation and virtualization for mobile networks
Disaggregation and virtualization for mobile networks
Disaggregation and virtualization for mobile networks

- Mobile Core UP
- Base Station
- RAN
- CU-CP
- Mobile Core CP
- CU-UP
- DU
- Mobile Core UP

Virtualization
Disaggregation and virtualization for mobile networks
Software-defined networking for mobile networks

Trellis
Fabric control

ONOS

DU & CU-CP & CU-UP

Mobile Core CP & UP
Software-defined networking for mobile networks

ONOS

Trellis
Fabric control

SDN&VNF infrastructure
Software-defined networking for mobile networks

Trellis
Fabric control

ONOS

P4Runtime

Stratum

Stratum

Stratum

P4Runtime

VNF

VNF

VNF

DU & CU-CP & CU-UP

Mobile Core CP & UP
Software-defined networking for mobile networks

Trellis
Fabric control

RAN Control

ONOS

Stratum

P4Runtime

Stratum

P4Runtime

Stratum

P4Runtime

Stratum

P4Runtime

VNF
VNF
VNF

DU & CU-CP & CU-UP

Mobile Core CP & UP
Software-defined networking for mobile networks

P4-based mobile RAN and core user-planes
Software-defined networking for mobile networks

Trellis
Fabric control

RAN Control

UPF

ONOS

P4Runtime

Stratum

P4Runtime

Stratum

P4Runtime

Stratum

P4Runtime

Stratum

UPF.p4

DU & CU-CP & CU-UP

Mobile Core CP & UP

VNF

VNF

VNF

VNF

VNF

VNF
Distributed cloud for mobile networks

- CBRS or Licensed Band
- Connected Edge
  - Edge Apps
  - IoT AI/ML Platform(s)
  - Mobile Core User Plane (P4 UPF)
- Small Cell
- Open RAN Controller

- IoT
- Sensors
- Surveillance
- Multimedia
- Employees
- Visitors
Distributed cloud for mobile networks

- Connected Edge
  - CBRS or Licensed Band
  - Small Cell
  - Open RAN Controller
  - Edge Apps
    - IoT AI/ML Platform(s)
    - Mobile Core User Plane (P4 UPF)

- Central Cloud

- IoT
- Sensors
- Surveillance
- Multimedia
- Employees
- Visitors
Distributed cloud for mobile networks

- CBRS or Licensed Band
- Connected Edge
- Edge Apps
  - IoT AI/ML Platform(s)
  - Mobile Core User Plane (P4 UPF)
- Central Cloud

- Open RAN Controller
- Small Cell
- IoT
- Sensors
- Surveillance
- Multimedia
- Employees
- Visitors
Distributed cloud for mobile networks

- **Connected Edge**
  - Open RAN Controller
  - Edge Apps
    - IoT AI/ML Platform(s)
    - Mobile Core User Plane (P4 UPF)

- **Central Cloud**
  - Aether Management Platform

- CBRS or Licensed Band
- Small Cell

- IoT
- Sensors
- Surveillance
- Multimedia
- Employees
- Visitors
Distributed cloud for mobile networks

- Connected Edge
  - CBRS or Licensed Band
  - Small Cell
  - Open RAN Controller
  - Edge Apps
    - IoT AI/ML Platform(s)
    - Mobile Core User Plane (P4 UPF)

- Central Cloud
  - Enterprise Control Portal
  - Aether Management Platform
Distributed cloud for mobile networks

Connected Edge
- CBRS or Licensed Band
- Open RAN Controller
- Edge Apps
  - IoT AI/ML Platform(s)
  - Mobile Core User Plane (P4 UPF)

Small Cell

Central Cloud
- Enterprise Control Portal
- Aether Management Platform
- IoT Apps
  - Sensors
  - Surveillance
  - Multimedia
  - Employees
  - Visitors
Distributed cloud for mobile networks

Connected Edge

- CBRS or Licensed Band
- Small Cell
- Open RAN Controller
- Edge Apps
  - IoT AI/ML Platform(s)
  - Mobile Core User Plane (P4 UPF)

Central Cloud

- Enterprise Control Portal
- Aether Management Platform
- Aether Connectivity Platform
Distributed cloud for mobile networks

Connected Edge

- CBRS or Licensed Band
- Small Cell
- Open RAN Controller
- Edge Apps
  - IoT AI/ML Platform(s)
  - Mobile Core User Plane (P4 UPF)

Central Cloud

- Enterprise Control Portal
- Aether Management Platform
- Aether Connectivity Platform
- Central IoT AI/ML Apps
Distributed cloud for mobile networks

- Distributed cloud
- Mobile networks
- Connected Edge
- Central Cloud
- Open RAN Controller
- Edge Apps
- IoT AI/ML Platform(s)
- Mobile Core User Plane (P4 UPF)
- Small Cell
- CBRS or Licensed Band
- Aether Connectivity Platform
- Aether Management Platform
- Enterprise Control Portal
Distributed cloud for mobile networks

**Connected Edge**
- CBRS or Licensed Band
- Open RAN Controller
- Edge Apps
  - IoT AI/ML Platform(s)
  - Mobile Core User Plane (P4 UPF)
- Distributed Mobile Core Use Plane
  - Provides local breakout at all remote Aether Edge site

**Central Cloud**
- Enterprise Control Portal
- Aether Management Platform
- Aether Connectivity Platform
- Central IoT AI/ML Apps
- Shared Mobile Core Control Plane in central cloud
  - Supports all Aether Edge sites
Aether has been operational since December’19
Aether Edge P4-based Disaggregated UPF
A disaggregated UPF

SMF/SPGW-C

PFCP

Control

gRPC/P4Runtime

Fast-path

UPF / SPGW-U
Combine Fast-paths

- **SMF/SPGW-C**
  - PFCP
  - Control
    - gRPC/P4Runtime
    - HW Fast-path (Tofino+FPGA)
    - SW Fast-path (BESS)

**UPF / SPGW-U**

Benefits in leveraging both fast-paths in the same UPF!

Higher tput (1-10s Tbit/s)
Lower latency (100s ns)
Smaller memory (10s MB)

Lower tput (10-100s Gbit/s)
Higher latency (100s µs)
Larger memory (100s GB)
Example: Tesla factory

- **Requirement:** 1M UEs
  - 5% smart phone
  - 10% wideband IoT devices (e.g., HD cameras)
  - 85% narrowband IoT devices (e.g., low data sensors)

- **Solution**
  - HW fast-path
    - Smartphone + wideband IoT: 150K sessions
  - SW fast-path
    - Narrowband IoT: 850K sessions
Architecture

SMF/SPGW-C

PFCP Agent (control)

UP4 App
Tofino, FPGA, D-BUF

Trellis App
RIB, mcast, etc.

ONOS

BESS
NIC

Stratum
FPGA NIC (hqos.p4)

D-BUF
NIC

Tofino Switch (fabric.p4)

Base Station

External routing
(BGP, OSPF, etc)

UPF / SPGW-U

SW

HW

gRPC

P4Runtime

P4Runtime
gNMI

SW Path

HW Path

DN
Architecture

- SMF/SPGW-C
- PFCP Agent (control)
  - ONOS
  - UP4 App: Tofino, FPGA, D-BUF
  - Trellis App: RIB, mcast, etc.
  - D-BUF
- BESS
- NIC
- Stratum
  - FPGA NIC (hqos.p4)
- Tofino Switch (fabric.p4)
- External routing (BGP, OSPF, etc.)
- SW
- HW
- DN

Aware fast-paths

P4Runtime

gNMI

gRPC

SW Path

HW Path
Architecture

SMF/SPGW-C

PFCP Agent (control)

ONOS

UP4 App
Tofino, FPGA, D-BUF

Trellis App
RIB, mcast, etc.

BESS

NIC

Stratum

D-BUF

FPGA NIC (hqos.p4)

UPF / SPGW-U

Tofino Switch
(fabric.p4)

Base Station

SW

External routing
(BGP, OSPF, etc)

SW Path

HW Path

DN

Aware fast-paths

SW

HW

Logical P4 pipeline, physically realized with Tofino+FPGA+DBUF

P4Runtime

gNMI

gRPC

P4Runtime

Tofino Switch

SW Path

HW Path

Aware fast-paths
Architecture

- SMF/SPGW-C
- PFCP Agent (control)
- UP4 App
  - Tofino, FPGA, D-BUF
- Trellis App
  - RIB, mcast, etc.
- ONOS
- BESS
- NIC
- Stratum
  - FPGA NIC (hqos.p4)
- D-BUF
  - NIC
- Tofino Switch (fabric.p4)
- External routing (BGP, OSPF, etc)
- Aware fast-paths

Logical P4 pipeline, physically realized with Tofino+FPGA+DBUF

Holds downlink packets in memory during UE power save mode. Can run on top of FPGA NIC
Architecture

SMF/SPGW-C

PFCP Agent (control)

ONOS

UP4 App
Tofino, FPGA, D-BUF

Trellis App
RIB, mcast, etc.

BESS

Stratum

FPGA NIC (hqos.p4)

Tofino Switch (fabric.p4)

UPF / SPGW-U

SW

HW

External routing
(BGP, OSPF, etc)

Base Station

Stratum

SW Path

DN

Aware fast-paths

Optional, for advance hierarchical QoS. Can rely on Tofino for simper QoS

Holds downlink packets in memory during UE power save mode. on top of FPGA NIC

Logical P4 pipeline, physically realized with Tofino+FPGA+DBUF

P4Runtime

P4Runtime

gNMI

gRPC

18
Architecture

SMF/SPGW-C

PFCP Agent (control)

ONOS

UP4 App
Tofino, FPGA, D-BUF

Trellis App
RIB, mcast, etc.

BESS

Stratum
FPGA NIC (hqos.p4)

UPF / SPGW-U

External routing
(BGP, OSPF, etc)

Tofino Switch
(fabric.p4)

SW Path

HW Path

 DN

Base Station

Aware fast-paths

Full UPF pipeline for low data rate sessions

Logical P4 pipeline, physically realized with Tofino+FPGA+DBUF

Optional, for advanced hierarchical QoS. Can rely on Tofino for simper QoS

Holds downlink packets in memory during UE power save mode. Can run on top of FPGA NIC

P4Runtime

gNMI

gRPC

SW

HW

DN

External routing
(BGP, OSPF, etc)
Architecture

- SMF/SPGW-C
  - PFCP Agent (control)
    - Logical P4 pipeline, physically realized with Tofino+FPGA+DBUF
      - P4Runtime
      - gNMI

- UP4 App
  - Tofino, FPGA, D-BUF
- Trellis App
  - RIB, mcast, etc.
- ONOS
- BESS
- NIC
- Stratum
  - FPGA NIC (hqos.p4)
  - Optional, for advance hierarchical QoS. Can rely on Tofino for simper QoS

- UPF / SPGW-U
  - SW
  - HW

- SW Path
- HW Path

- External routing
  - (BGP, OSPF, etc)
  - Holds downlink packets in memory during UE power save mode.
    - on top of FPGA NIC

- Base Station
  - Full UPF pipeline for low data rate sessions
  - Aware fast-paths
  - Detour based on IP pools/prefixes.

- Tofino Switch (fabric.p4)
Status (as of December 2020)

- Already started rolling out at several Aether Edge sites:
  - GTP termination and accounting on Tofino, integrated with Trellis/ONOS fabric control
- Q1 2021
  - Downlink buffering via DBUF
  - QoS on Tofino with shared queues
  - Improved scale: 10k UEs
- Q4 2021
  - Integration with FPGA NIC for advance QoS
  - Scale improvements
- Long-term
  - Integration with SW-based fast-path (BESS)
Productize programmable network infrastructure
Software stack

Control plane software
- Fabric.p4
- Trellis
- UP4
- ONOS

Data plane software
- Stratum
- Barefoot SDE
- Open Network Linux
Improved, optimized software stack

Control plane software

- Fabric.p4
- Trellis
- UP4
- ONOS

Data plane software

- Stratum
- Barefoot SDE
- Open Network Linux
Improved, optimized software stack

Control plane software
- Fabric-TNA
- Trellis
- UP4
- ONOS

Data plane software
- Stratum
- Barefoot SDE
- Open Network Linux
Improved, optimized software stack

**Control plane software**
- Fabric-TNA
- Trellis
- ONOS

**Data plane software**
- Stratum
- Barefoot SDE
- Open Network Linux

Rewrite fabric.p4 from v1model architecture to Tofino Native Architecture (TNA). Allows us to create more advanced and optimized P4 programs.
Improved, optimized software stack

Control plane software
- Fabric-TNA
- Trellis
- UP4
- ONOS

Data plane software
- Stratum
- Barefoot SDE
- Open Network Linux
Improved, optimized software stack

Control plane software

- Fabric-TNA
- Trellis
- UP4
- ONOS

Decoupled from the ONOS code-base with new release cycle.

Data plane software

- Stratum
- Barefoot SDE
- Open Network Linux
Improved, optimized software stack

Control plane software
- Fabric-TNA
- Trellis
- UP4
- ONOS

Data plane software
- Stratum
- Barefoot SDE
- Open Network Linux
Improved, optimized software stack

Control plane software

- Fabric-TNA
- Trellis
- UP4
- ONOS

Data plane software

- Stratum
- Barefoot SDE
- Open Network Linux

Use new ONOS LTS with lots of stability, performance, and availability improvements
Improved, optimized software stack

Control plane software

Fabric-TNA  |  Trellis  |  UP4

ONOS

Data plane software

Stratum

Barefoot SDE

Open Network Linux
Improved, optimized software stack

Control plane software

- Fabric-TNA
- Trellis
- UP4
- ONOS

Data plane software

- Stratum-bfrt
- Barefoot SDE
- Open Network Linux
Improved, optimized software stack

Control plane software
- Fabric-TNA
- Trellis
- UP4
- ONOS

Data plane software
- Stratum-bfret
- Barefoot SDE
- Open Network Linux

New Stratum implementation based on Barefoot native API unlocks more advanced ASIC management.
Improved, optimized software stack

Control plane software
- Fabric-TNA
- Trellis
- UP4
- ONOS

Data plane software
- Stratum-bfrt
- Barefoot SDE
- Open Network Linux
Improved, optimized software stack

Control plane software

- Fabric-TNA
- Trellis
- UP4
- ONOS

Data plane software

- Stratum-bfrc
- Barefoot SDE
- Open Network Linux

Several improvements to support fast deployment and troubleshooting.
Improved, optimized software stack

Control plane software

- Fabric-TNA
- Trellis
- UP4
- ONOS

Data plane software

- Stratum-bfirt
- Barefoot SDE
- Open Network Linux
• ONF’s fabric.p4 on Tofino Native Architecture (TNA)
• Supports Aether Edge use-cases
  • Trellis (Bridging, Routing, …)
  • UPF/SPGW-U
    • Simple QoS, accounting
    • Integrate with D-BUF
• Inband Network Telemetry (INT)
  • Advance telemetry report mechanism
Stratum-bfrt

- Stratum implementation with Barefoot BfRt C++ API
- Performance improvement
- Advance ASIC control
  - Batching/Transaction
  - Register
  - Traffic manager
  - Egress mirroring
  - Folded/Multi pipeline
  - …
Software packages

TOST container image

- Fabric TNA
- Trellis
- UP4
- ONOS

Stratum container image

- Stratum
- Barefoot SDE

ONIE installer

- Open Network Linux

TOST: Trellis ONOS Stratum Tofino
Kubernetes Integration

Tofino Switch

Terraform
Rancher

Management
Node
Kubernetes Integration

Tofino Switch

Docker

Open Network Linux

DHCP
HTTP
TFTP

Terraform
Rancher

Management
Node
Kubernetes Integration

Tofino Switch

- Kubernetes
- Docker
- Open Network Linux

- DHCP
- HTTP
- TFTP

- Terraform
- Rancher
- Management Node

Node
Tofino Switch

- Prometheus Exporter(s)
- Stratum
- Kubernetes
- Docker
- Open Network Linux
- DHCP
- HTTP
- TFTP

- Terraform Rancher
- Management Node
Kubernetes Integration

- Tofino switch as a Kubernetes worker node
  - With special taint and label to make sure only Stratum is deployed on it
- Stratum is deployed as Kubernetes service
  - Deployed as DaemonSet. There will be one and only one instance on each switch node
  - P4RT/gNMI exposed via NodePort
  - externalTrafficPolicy=Local so the traffic won't get load-balanced to other switches
Automation - Build and Release

- Git-triggered automated build and release process for Trellis apps and control plane container image
- Build and release Stratum image weekly
Automation - Build and Release

- Git-triggered automated build and release process for Trellis apps and control plane container image
- Build and release Stratum image weekly
• Git-triggered automated build and release process for Trellis apps and control plane container image
• Build and release Stratum image weekly
Git-triggered automated build and release process for Trellis apps and control plane container image
Build and release Stratum image weekly
• Git-triggered automated build and release process for Trellis apps and control plane container image
• Build and release Stratum image weekly
• Git-triggered automated build and release process for Trellis apps and control plane container image
• Build and release Stratum image weekly
• Git-triggered automated build and release process for Trellis apps and control plane container image
• Build and release Stratum image weekly
• Git-triggered automated build and release process for Trellis apps and control plane container image
• Build and release Stratum image weekly
Automation - Deploy

- **Human-triggered** Jenkins pipeline based on **Terraform**
  - Explicitly-defined helm chart version
  - Get rid of issues seen in Rancher CLI
• **Human-triggered Jenkins pipeline based on Terraform**
  • Explicitly-defined helm chart version
  • Get rid of issues seen in Rancher CLI
• Human-triggered Jenkins pipeline based on Terraform
  • Explicitly-defined helm chart version
  • Get rid of issues seen in Rancher CLI
• **Human-triggered Jenkins pipeline based on Terraform**
  - Explicitly-defined helm chart version
  - Get rid of issues seen in Rancher CLI
• **Human-triggered** Jenkins pipeline based on **Terraform**
  • Explicitly-defined helm chart version
  • Get rid of issues seen in Rancher CLI
Human-triggered Jenkins pipeline based on Terraform

- Explicitly-defined helm chart version
- Get rid of issues seen in Rancher CLI
• **Human-triggered** Jenkins pipeline based on **Terraform**
  - Explicitly-defined helm chart version
  - Get rid of issues seen in Rancher CLI
Recap

- Aether - 5G/LTE Enterprise Private Edge Cloud
- P4-based disaggregated UPF
- Highly automated network infrastructure
Learn More

• Aether
  • 5G/LTE Enterprise Private Edge Cloud
  • [https://aetherproject.org](https://aetherproject.org)

• Trellis
  • Leaf-spine SDN fabric for edge
  • [https://opennetworking.org/trellis](https://opennetworking.org/trellis)

• Stratum
  • Silicon-independent switch operating system for SDN
  • [https://stratumproject.org](https://stratumproject.org)

• Slack Channel: onf-community
Thank you