SD-CORE

Open Source Dual-Mode 4G/5G Mobile Core
Enabling Connectivity-as-a-Service

Executive Summary

SD-Core™ is the Open Networking Foundation™ (ONF)'s open source, flexible, agile, scalable, and configurable dual-mode 4G/5G mobile core network solution that enables a cloud-based Connectivity-as-a-Service. SD-Core builds and enhances on ONF's 4G Open Mobile Evolved Core (OMEC)™ as well as the free5GC© core network platform to create a dual-mode solution that supports LTE, 5G NSA and 5G SA services optimized for multi-cloud environments. SD-Core includes multiple user plane function (UPF) implementations: a P4-based UPF that is realized on programmable forwarding elements allowing for fine-grained visibility for verifiably performant and secure operations, and two variants of cloud-native, scalable, containerized UPFs. All of the UPF options are dual core, supporting both LTE and 5G traffic simultaneously.

SD-Core is optimized for a hybrid cloud deployment. Its centralized dual-mode mobile core control plane has been designed to control many UPFs distributed across many edge clouds around the world. This makes SD-Core the ideal open source platform to offer cloud-based Private 4G/5G connectivity services to enterprises as an enabler for Industry 4.0 transformation.

SD-Core provides a rich set of APIs for runtime configurability of each of its services, as well as supporting subscriber management via third party applications. These APIs provide extensive telemetry capabilities that enable monitoring, logging, and alerts, with integrated verification and closed-loop control solutions. With operator and customer facing portals, SD-Core can be configured for dynamically programmable network slicing, subscriber, QoS and policy management, providing precise access control for users, devices, data networks and edge applications.

Introduction

The cellular network consists of two main subsystems: the Radio Access Network (RAN), and the Mobile Core. The RAN manages communication over the time-varying wireless link over a specific radio spectrum. The mobile core, on the other hand, serves several purposes:

- It ensures that only authenticated mobile devices can access the network.
- It provides IP connectivity for both data and voice services.
- It ensures that this connectivity fulfills the promised QoS and network slice policies.
- It tracks user mobility to ensure uninterrupted service.
- It tracks subscriber usage for billing and charging.

As with prior generations, 5G brings about improvements to provide higher data rates and lower end-to-end latencies. But perhaps the biggest change 5G brings is the revolution in how we operationalize mobile networks. This transformational change is essentially disaggregation and virtualization of the network functions, including those of the mobile core, on the path towards containerization and eventually cloud-native.

The disaggregation in the mobile core has its user plane and control plane separated. With 5G, the mobile core also moves to a service-based architecture, making use of the widely used RESTful interfaces for all inter-service interactions.

ONF's SD-Core brings an open source, fully configurable platform with a rich set of APIs for dual-mode 4G and 5G connectivity. As illustrated in **Figure 1**, SD-Core uses OMEC and free5GC as baseline components, integrates the two for dual-mode operation, and provides significant new functionality to optimize delivery of Connectivity-as-a-Service from the hybrid cloud. It is of course also possible to use SD-Core to provide 4G-only, or 5G-only connectivity using standard 3GPP interfaces.

In order to optimize for the hybrid cloud and to support emerging Industry 4.0 use cases, SD-Core includes multiple user plane functions (UPFs) to handle different classes of enterprise traffic:

- A P4-based UPF offloads the packet processing and forwarding operations to a programmable edge fabric to achieve significant performance with much higher bandwidths, significantly lower latencies, and highly predictable very low jitter, albeit for a relatively modest number of devices/flows.

- A containerized, highly scalable solution provides high performance by leveraging acceleration technologies like DPDK.

SD-Core also includes a 5G UPF that is being developed by the ecosystem towards the support of telco networks, providing all necessary features and interfaces as well as the operational scale.

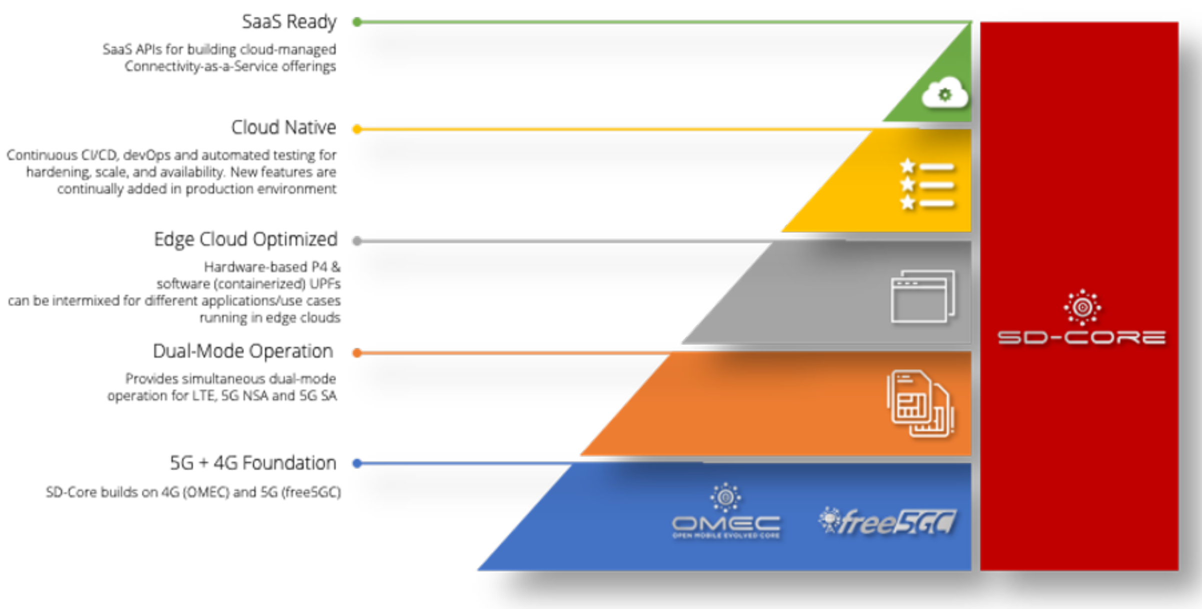


Figure 1: SD-Core

All components of SD-Core go through continuous automated testing towards hardening, scale and availability. SD-Core has been serving the operational Aether™ Network over the last two years, contributing to its multiple-nines availability. 4G components of SD-Core are actively in production with tier 1 operators, and an outdoor trial for dual-mode operation is slated for the second half of 2021, enabling an operationalized 4G/5G network with disaggregated O-RAN-based commercial small cells.

SD-Core has been developed with a cloud-based deployment and consumption model as its foundation. It has a rich and extensible set of APIs to allow for runtime configurability of subscriber management, access management, session management, and network slice management. This configuration may be conducted via ONF's Runtime Operational Control (ROC) platform directly for consumption as a cloud-managed service, or the APIs can be used by third-party automation and management platforms.

SD-Core Enabling 4G and 5G Connectivity as a Cloud Managed Service

To enable a cloud-based managed 4G/5G connectivity service, SD-Core provides a rich set of APIs. As illustrated in **Figure 2**, these APIs are designed to provide telemetry data to and accept configuration updates from third party applications and portals. Consequently, it is possible for the third-party applications to develop automated, closed-loop control processes by continuously tracking connectivity state and modifying SD-Core configuration using simple knobs to sustain the desired operational quality.

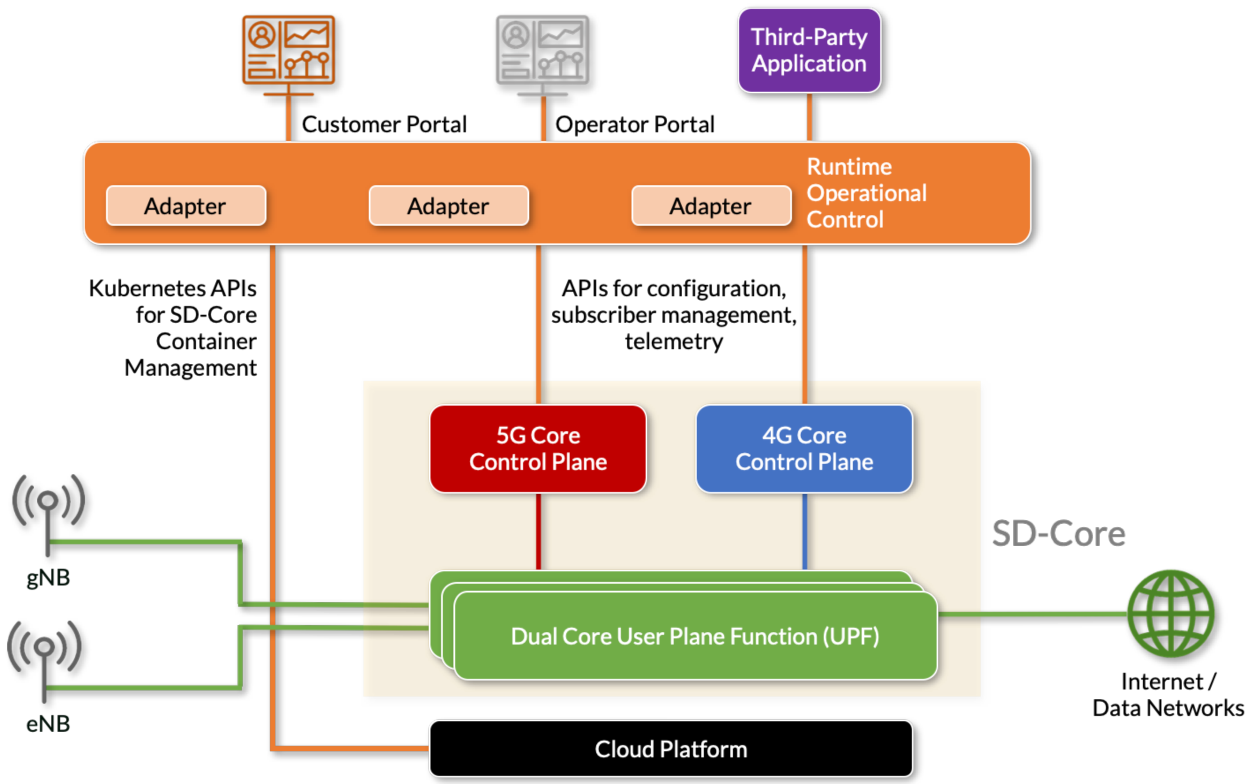


Figure 2: High-Level SD-Core Architecture

As seen in **Figure 2**, ONF's ROC includes built-in adapters for SD-Core to translate its monitoring and configuration APIs to customer and operator portals as well as third-party applications with corresponding levels of abstraction. ROC uses RBAC to control the access of mobile core APIs. Operators can use these APIs to provision subscribers, and their associated access and connectivity policies. SD-Core APIs can also be used to control the runtime configuration of network functions, e.g., management of network slices. The management of a network slice may require establishment of a new slice in runtime, which potentially requires instantiation of new SD-Core components. ROC handles this via its interface with the cloud and container management platform on which SD-Core runs.

SD-Core's telemetry APIs are also used to continually monitor network status and to generate alerts for the network operator. SD-Core's APIs are extensible, and they continue to evolve as new use cases, new deployments, and ecosystem requests materialize.

SD-Core Architecture

SD-Core architecture enables the following distinct features:

- The architecture is fully disaggregated, composed of containerized components.
- The solution enables 4G, 5G Standalone (SA) and 5G Non-Standalone (NSA) connectivity.
- The platform is configurable in runtime via an extensible set of APIs.
- The solution is consumable as a cloud-managed service.
- All SD-Core components follow 3GPP standards to interface with others as well as the external networks and systems (e.g., RAN, communication services, etc.). As such, components can be consumed independently and be used as part of a multi-vendor mobile core deployment.

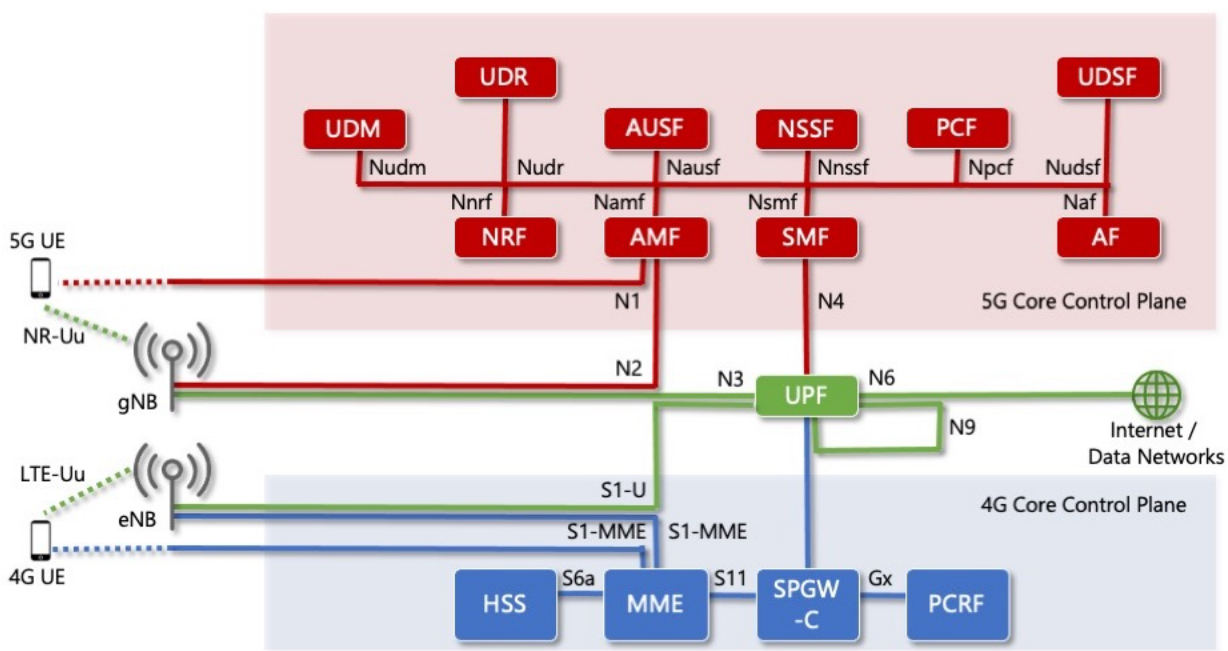


Figure 3: 5G Standalone (SA) and 4G/ LTE with SD-Core

The details of the dual-mode SD-Core architecture is illustrated in **Figure 3**. Specifically, SD-Core's 5G core network functions are 3GPP release 15.3 compliant and include the Access & Mobility Management Function (AMF), Session Management Function (SMF), Authentication Server Function (AUSF), Policy Control Function (PCF), Unified Data Management (UDM), Unified Data Repository (UDR), Network Repository Function (NRF), Network Slice Selection Function (NSSF), Application Function (AF), and in the near future, Unstructured Data Storage Function (UDSF) for cloud-native enablement. These functions use Service Based Interfaces (SBI) to communicate with one other. SBI uses a well-defined REST interface using HTTP/2 as the application protocol. The REST APIs provide methods like POST, PUT, PATCH and GET to manage resources created on network functions. SD-Core's 5G core control plane functions leverage seed code from the free5GC project, upon which the SD-Core community has implemented numerous architectural changes that are integrated and optimized with SD-Core's set of UPF solutions along with several new features. This has provided hardening to the solution with rigorous testing to create an enhanced 5G Core solution.

All interfaces are designed to be robust in order to handle all network errors including but not limited to packet loss, peer network function failure, and duplicate packets. Network functions support a highly available and scalable architecture by running multiple instances and sessions state is persisted in the database. SD-Core's 5G Core provides a highly flexible policy framework which allows network operators to define policies based on use cases. Network visibility being one of the most important aspects of the architecture, network telemetry data is exported to time series databases such as Prometheus. APIs are provided on top of metrics so that third party applications can be easily written to change the application/network behavior.

SD-Core's 4G control plane functions are 3GPP release 13 compliant and include the Mobility Management Entity (MME), Serving & Packet Data Network Gateway Control (SPGW-C), Home Subscriber Server (HSS), and Policy & Charging Rules Function (PCRF). SD-Core's 4G Core is designed to have a CUPS (Control-User Plane Separation) compliant architecture and uses the 3GPP Packet Forwarding Control Protocol (PFCP) to implement CUPS. Several 3GPP release 15 updates to PFCP specifications are supported in the SD-Core implementation to provide better operation control. Similar to its 5G counterpart, SD-Core's 4G Core provides a highly flexible policy framework which allows network operators to have more control in the network. SD-Core's 4G Core is designed to provide robust operational behavior against network element failures, packet drops, duplicate packets, and packet retransmissions.

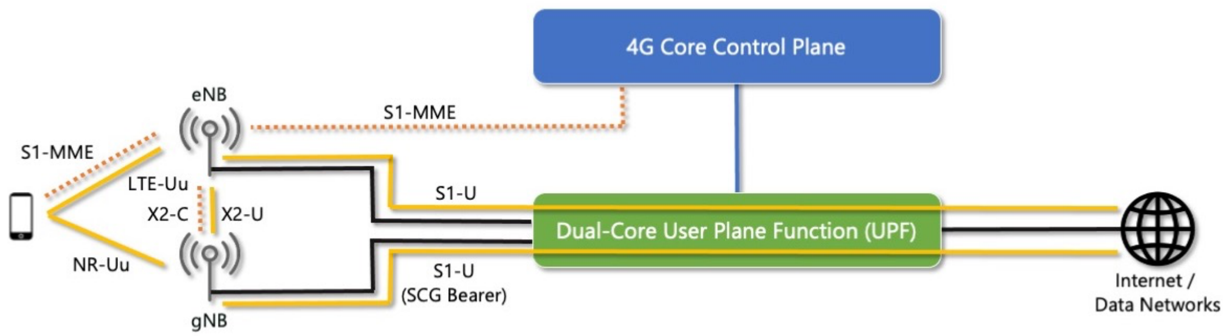


Figure 4: 5G Non-Standalone with SD-Core

SD-Core's 4G control plane has been enhanced to provide functional support for 5G Non-standalone operation with compliant eNBs and gNBs as per 3GPP specifications. SD-Core's 5G NSA operation is depicted in **Figure 4**. 5G NSA related enhancements include support of the extended bearer rates on required interfaces as well as the 5G NSA attributes in the HSS.

User Plane Functions

SD-Core has three User Plane Functions (UPFs), and each deployment may select one or more to be instantiated for configurable sets of network slices:

- P4-Based Dual-Core UPF optimized for private enterprise deployments.
- Containerized Dual-Core UPF optimized for private enterprise deployments.
- Containerized Dual-Core UPF optimized for various high-scale MNO use cases.

When one or more of SD-Core's dual-core UPFs are engaged, a connected device is latched to either of the 4G or 5G control planes of SD-Core based on its radio access type. The dual-core UPFs can communicate with both 4G and 5G control planes, and thus can serve both 4G and 5G sessions at the same time. In SD-Core, a connected device is assigned to a UPF based on the network-wide slice configuration. Specifically, in 5G core, the SMF uses the network slice information received in the user session context as well as the Data Network Name (DNN) information received from AMF to select the serving UPF. In the case of 4G core, the SPGW-C uses the location information as well as the Access Point Name (APN) information to select the serving UPF. More information on how network slicing is conducted in SD-Core can be found in the subsequent sections of this document.

The dual-core user plane functions that are optimized for private enterprise deployments are themselves implemented with a disaggregated architecture, including:

- Control Interface Function (PFCP Agent)
- Packet Processing and Forwarding Function

The two dual-core UPFs share a common Control Interface Function, but the Packet Processing and Forwarding Function is implementation-specific. The architectures of the P4-based and containerized dual-core UPFs are illustrated in **Figure 5 (a, b)**.

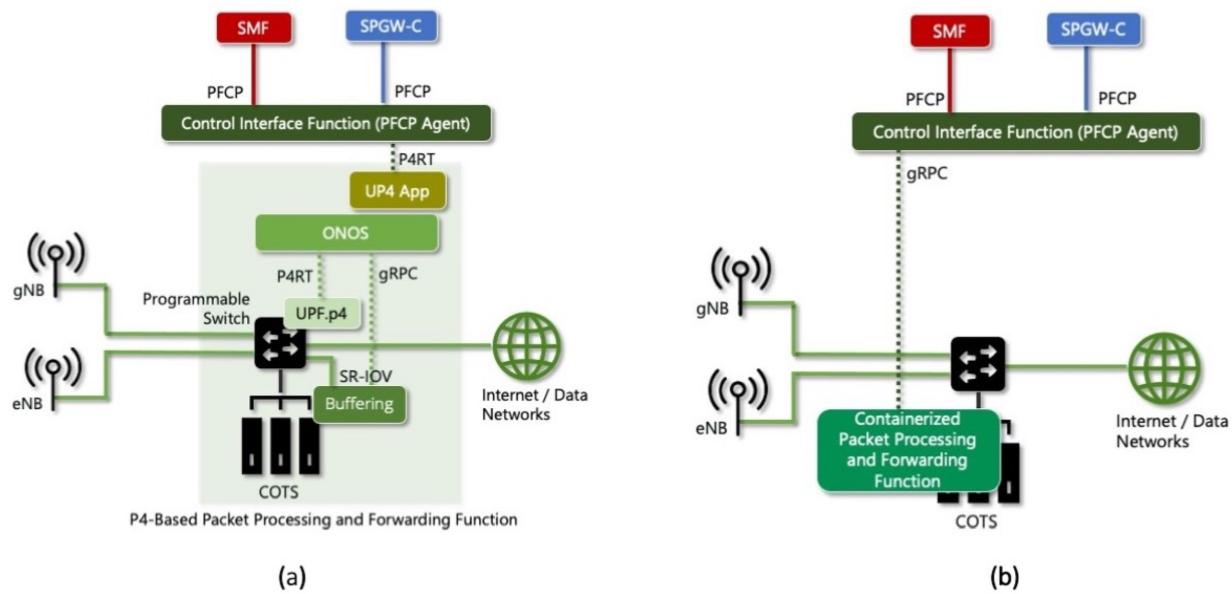


Figure 5 (a): P4-Based Dual-Core UPF, and **(b):** Containerized Dual-Core UPF

Control Interface Function (PFCP Agent)

The PFCP Agent interfaces with 4G and 5G control plane components SPGW-C and SMF using its northbound interface. This interface is based on the 3GPP 24.244 specification for PFCP. This protocol uses message constructs like PDR (Packet Detection Rules), FAR (Forwarding Action Rules), BAR (Buffering Action Rules) and QER (QoS Enforcement Rules) to communicate information from the control plane to the user plane.

Using its southbound interface, the PFCP Agent communicates with the Packet Processing and Forwarding Function using P4 Runtime (P4RT) and gRPC for P4-based and containerized dual-code UPFs, respectively. The PFCP agent can be configured to interface with the P4-based or containerized Packet Processing and Forwarding Function, or both if a network slice is so configured, as illustrated in **Figure 6**.

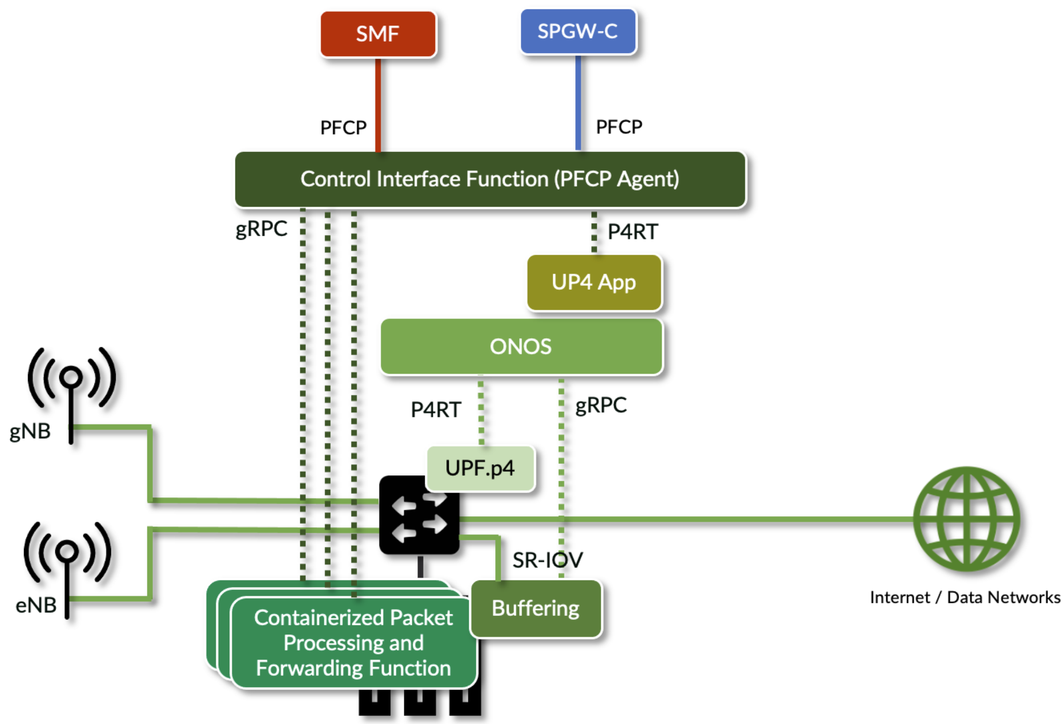


Figure 6: Simultaneous Use of P4-Based and Containerized Dual-Core UPFs with Network Slicing in SD-Core

Packet Processing and Forwarding Function

As the name suggests, the Packet Processing and Forwarding Function is the forwarding plane of the UPF. SD-Core has two implementation-specific Packet Processing and Forwarding Function variants, one for the P4-based dual-core UPF and another for the containerized dual-core UPF.

Figure 5(a) illustrates the building blocks of the Packet Processing and Forwarding Function for the P4-based dual-core UPF. It includes the following components:

- A UPF-specific P4 application running on a P4-programmable fabric that specifies exactly how packet forwarding and processing, including GTP tunnel encapsulation and decapsulation should be executed, for both downlink and uplink.
- ONF's SDN controller, ONOS, that installs table entries onto the switch based on the information received from the PFCP Agent.
- UP4 application running on ONOS that communicates with the PFCP agent on the northbound interface and engages ONOS to install the forwarding rules to the

programmable switch using P4 Runtime. The UP4 application acts as a logical pipeline for the PFCP Agent and hence abstracts it the from the actual forwarding pipeline.

- A buffering module that is used to buffer the incoming downlink packets towards a device that is in idle mode or when a device is going through a handover. The buffering module interacts with the switch on L4 interface to receive packets for buffering, and also to drain the packets once the UE is back in active mode or the handover process is completed. This module also notifies ONOS to signal the PFCP agent when the first packet for an idle user is received.

Figure 5(b) illustrates the Packet Processing and Forwarding Function for the containerized dual-core UPF. This function has been developed on the Berkeley Extensible Software Switch (BESS) architecture as a containerized software switch. BESS has been designed to be extensible and highly performant. It is composed of a BESS daemon, a BESS controller and multiple BESS modules. The PFCP agent uses gRPC to communicate with the BESS-based Packet Processing and Forwarding Function. This implementation runs entirely on the user space and binds directly to network interfaces using DPDK, bypassing the kernel.

The two Packet Processing and Forwarding Function implementations serve complimentary purposes. The P4-based implementation enables very high-performance processing and forwarding along with very high aggregate throughput, effectively at the switch line rate (3.2 Tbps for a current P4-switch model on which this implementation is operationalized), very low latency (in the order of 1.2-1.5 μsec), and very low and predictable jitter (in the order of 4 ns) for a limited number of connections. Further, this implementation brings fine-grained visibility to 4G and 5G traffic with INT. This in turn enables the capability to monitor, detect, and rectify issues on the data plane using verification and closed-loop control, thus providing a highly reliable and highly secure connectivity service. This implementation has upper limits on vertical scaling as well as the level granularity available for differentiated QoS services due to the hardware limitations of today's P4-based programmable switches.

To support a virtually unlimited number of connections but with relatively more modest performance requirements, one can make use of the BESS-based Packet Processing and Forwarding Function which is effectively a software-based microservice. This implementation uses SR-IOV to scale the physical NIC interface to multiple virtual interfaces and thus can help provide more network isolation and a highly scalable infrastructure. Using DPDK, this implementation can scale the number of ingress and egress queues per user and provide highly granular data flow classification for differentiated QoS. Thanks to its use of acceleration tools such as SR-IOV and DPDK, the performance and reliability of the BESS-based Packet Processing and Forwarding Function is optimized for running on the Intel® Xeon® CPU architecture and is more than sufficient for a large cross section of use cases. It can provide service for up to 1M concurrent users in one instance that uses 8 CPU cores and can vertically scale for even higher densities. It can achieve an aggregate throughput of 100 Gbps and thus its performance will never reach the level of its P4-based counterpart, but this performance differential is offset by its high scalability and extensible QoS support.

In SD-Core all implementations of the UPF enable telemetry, allowing extensive collection of information from both Control Interface Function and Packet Processing and Forwarding Function via well-defined APIs. Specifically, all UPF implementations provide monitoring data on various control messages and user session information along with per user packet latency and data consumption statistics. The specific contents that are to be extracted can be dynamically configured using the SD-Core APIs.

Programmable Network Slicing with SD-Core

Network slicing is one of the most important features of the 5G core network. Network slicing helps in isolating the network for various business and use cases. In the disaggregated service-based architecture of 5G core, this isolation may include only the UPF or also a subset of the control plane services such as the SMF. However, mobile core control functions that are responsible for managing user mobility, user authentication, and network slicing need to remain centralized across all slices. SD-Core provides the necessary APIs to manage network slices using external agents. ONF's ROC, pre-integrated with SD-Core, allows for this central management via portals as well as automation. If the management requires instantiation of a new UPF and/or a new SMF instance, ROC oversees this by interacting with edge cloud or hyperscale container management services to provision such new network function instances.

Once all mobile core service instances are provisioned for a new slice, ROC uses SD-Core APIs to configure the slice as well as all required central network functions. SD-Core provides APIs to create and configure network slices and assign resources to each slice. Operators can assign a slice for a group of users/devices based on the use case. The behavior of each slice is configurable and can be dynamically changed during run time. SD-Core's architecture supports assigning dedicated network functions to a specific slice or providing logical separation if network functions are to be shared among various slices. Various QoS and access policies can be applied to each slice to control the assigned resources as well as IP connectivity and access control within each slice.

Operators can create new slices based on criteria such as isolating devices allowed to access specific packet data networks/edge applications or keeping all devices or flows with the same QoS classification grouped under one slice. Network slice selection is achieved through 3GPP-specified network functions like Network Slice Selection Function (NSSF) and Network Repository Function (NRF). NSSF helps in mapping the device/flow to a specific slice and steering the device/flow traffic to the right set of core network elements. SD-Core's 5G implementation natively includes both NSSF and NRF for slice selection.

As described earlier, SD-Core's P4-based dual-core UPF allows for the monitoring of all 4G/5G traffic with fine-grained granularity using INT. This effectively means that with the P4-based dual-core UPF, it is possible to conduct per packet network monitoring to track whether slice-specific SLAs are being met and automatically adapt network behavior by changing per slice resource allocations, QoS priorities etc., to automatically sustain the required network performance using closed-loop control.

SD-Core Deployment Options

The level of disaggregation and associated optimizations achieved for each component of its 4G and 5G control plane makes SD-Core suitable for a wide variety of deployment options. These optimizations include the capability for the 4G and 5G control planes to oversee many UPFs, potentially instantiated at geographically diverse locations, as illustrated in **Figure 7**.

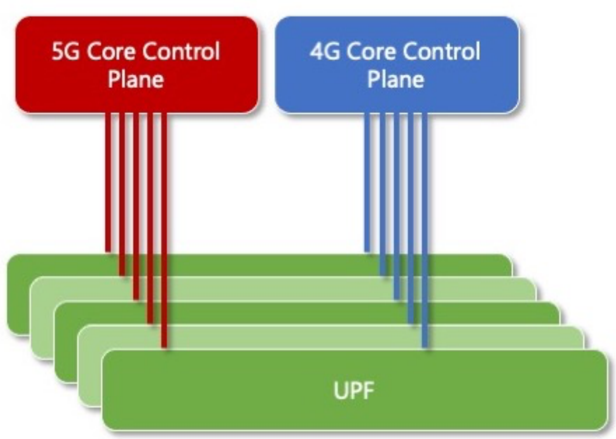


Figure 7: SD-Core's Centralized Control Planes

It is possible to deploy all components of SD-Core collocated in an edge cloud or a central cloud for private consumption. It is also possible to distribute the components of SD-Core across multiple clouds, edge and central, to deliver a cloud-managed multi-tenant connectivity service. In this distributed deployment option, SD-Core's control plane will run on a central/hyperscaler cloud and control multiple user planes running on different on-premises edge clouds, potentially serving distinct customers as illustrated in **Figure 8**. In this deployment, the 4G and 5G control plane functions can scale as necessary. Each customer site can have more than one UPF deployed depending on the use cases and network slices configured. Operators can also decide to deploy UPFs in the central cloud for certain customers and their use cases where latency and data privacy is not a concern. SD-Core brings the flexibility to define network slices for each customer in such a way that one deploys a distinct UPF for each slice and instantiates the various components of the solution at the customer edge or in the central cloud, as needed and best suited.

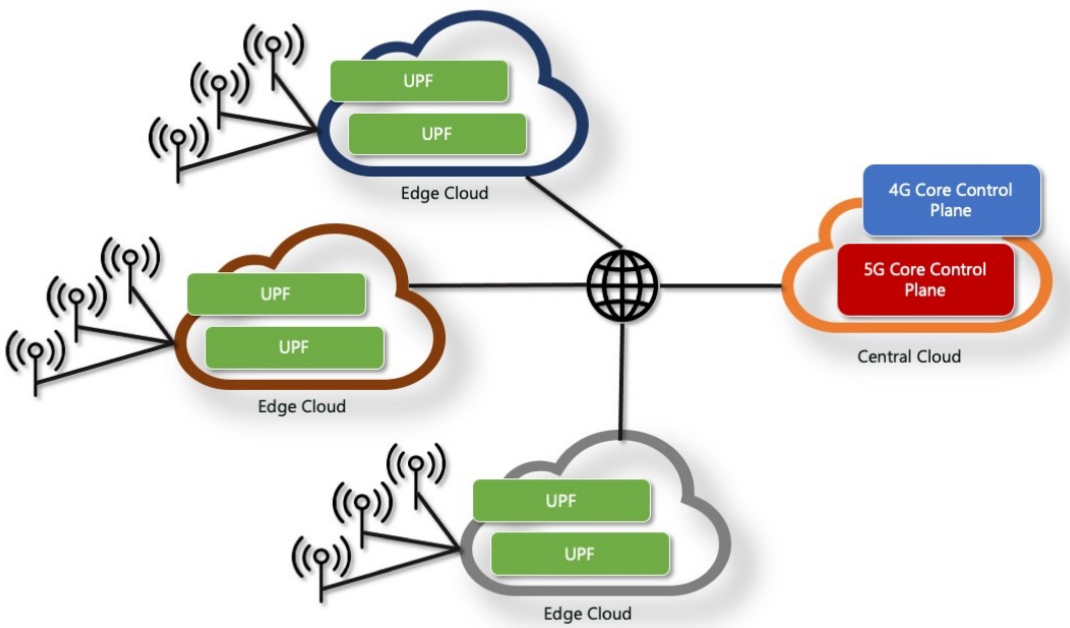


Figure 8: Hybrid Cloud Deployment

SD-Core's hybrid cloud deployment is an important enabler for a managed 4G/5G connectivity service where each customer site may be deployed to serve a different set of use cases and may have different types of underlying cloud environments. The 4G/5G core control planes running on the central cloud have been designed and optimized to support distributed edge sites which are spread across different locations across the world. The SD-Core control plane uses PFCP to communicate with the UPFs at the edge sites. The hybrid cloud deployment architecture has been optimized to handle variability in encountered delays communicating with the remote edge sites and is equipped to handle potential packet losses and retransmissions to support a multi-tenant, distributed geography deployment.

SD-Core in Aether

Aether is ONF's 4G/5G Connected Edge platform that provides mobile connectivity and edge cloud services for distributed enterprise networks, all provisioned and securely managed from a centralized cloud. Aether is a deeply programmable service, top-down and end-to-end. In this capacity, it brings programmability to the edge cloud fabric with SD-Fabric, to the RAN with SD-RAN, and to the core with SD-Core, and coordinates end-to-end programmability using its Controller module. Aether is a cloud-native platform. As such, all of the tightly integrated components are containerized. A high-level overview of the Aether stack and how SD-Core is leveraged within it is illustrated in **Figure 9**.

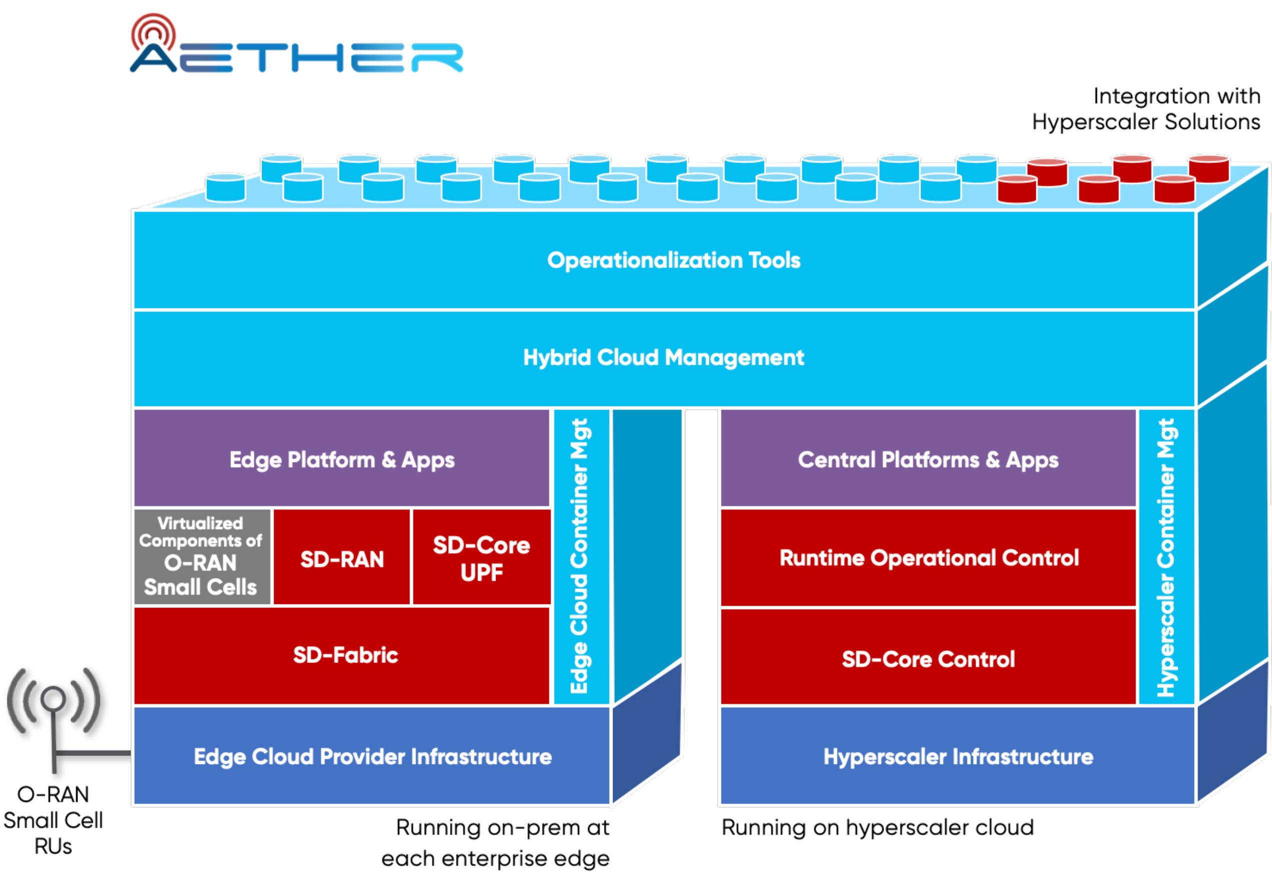
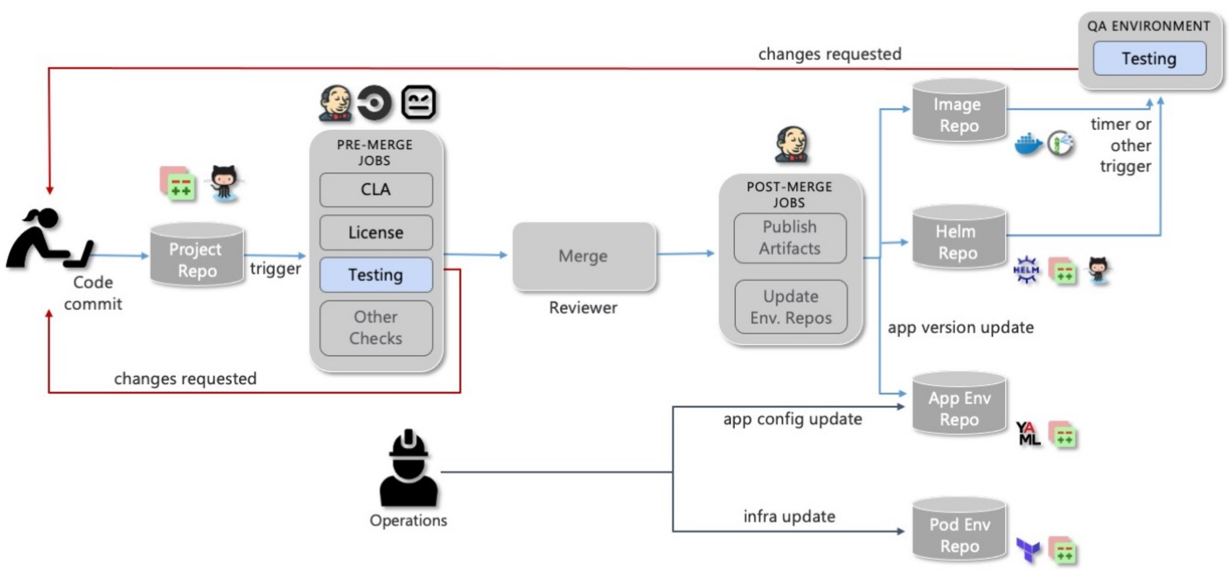


Figure 9: SD-Core in Aether

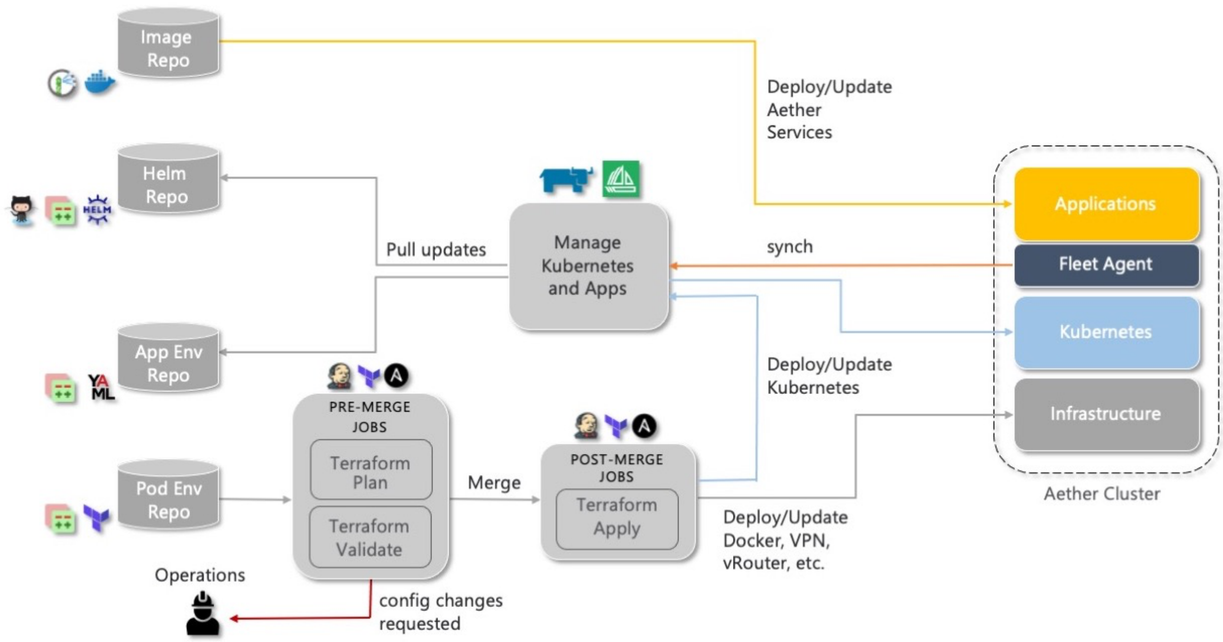
SD-Core empowers 4G/5G connectivity-as-a-service in the Aether network architecture. Along with central Aether management, the SD-Core control plane runs in the central cloud, and controls multiple SD-Core user plane components running at each Aether edge site. SD-Core's APIs are exposed to ROC to enable highly programmable and configurable 4G/5G connectivity for Aether customers. Aether, as well as each of the component platforms SD-Core, SD-RAN, and SD-Fabric, are all designed to run on bare metal and also on any hyperscaler cloud platform.

Aether includes a sophisticated and automated CI/CD and GitOps pipeline that SD-Core leverages. This pipeline is illustrated in **Figure 10(a, b)**. Using this pipeline, SD-Core development goes through an exhaustive set of tests, both at the component level and at the integrated system level. Further, interoperability tests, soak tests, and chaos tests are conducted as part of the end-to-end Aether platform.

The Aether Network has been operational since late 2019 with close to 15 edge sites around the globe, with SD-Core providing the 4G and 5G core functionality with multiple-nines availability.



(a)



(b)

Figure 10: SD-Core Uses Aether's Automated CI/CD Pipeline: **(a)** CI Pipeline, **(b)** CD Pipeline

Conclusion

SD-Core, optimized for hybrid cloud deployment, is the ideal production-grade open source dual-mode mobile core solution enabling managed Private 4G/5G connectivity services to enterprises towards Industry 4.0 enablement. SD-Core is highly programmable via a rich set of well-defined APIs. With multiple UPF options, including a P4-based realization, SD-Core brings superior performance and fine-grained visibility to the cellular network traffic. When coupled with programmability, this allows operators to provide verifiable resiliency, security, and privacy capabilities to their 4G and 5G deployments.

As part of Aether, SD-Core brings to life a fully automated, top-down, and end-to-end programmable 4G/5G connected edge cloud solution. SD-Core follows Aether's fully automated CI/CD pipeline and is continuously and rigorously tested towards a production-grade complete solution. SD-Core has already achieved multiple-nines availability as deployed as part of the Aether Pilot deployment.

Starting with AMF and SMF, SD-Core's 5G control plane is rapidly progressing towards a cloud-native realization, with stateless network functions and centrally maintained state in an Unstructured Data Storage Function (UDSF). This will allow the control plane functions to vertically scale independently from the application. SD-Core's upcoming cloud-native architecture will also enable service continuity and high availability in case of any network function failures.

We invite enterprises, operators, vendors, systems integrators, researchers, and the wider ecosystem to actively participate, collaborate, contribute and adopt SD-Core. Learn more about SD-Core and how to get involved by visiting www.opennetworking.org/sd-core.

Abbreviations

5G SA: 5G Standalone

AF: Application Function

AMF: Access and Mobility Management Function

APN: Access Point Name

AUSF: Authentication Server Function

BAR: Buffering Action Rules

BESS: Berkeley Extensible Software Switch

CUPS: Control-User Plane Separation

DNN: Data Network Name

DPDK: Data Plane Development Kit

FAR: Forwarding Action Rules

HSS: Home Subscriber Server

INT: In-band Network Telemetry

LTE: Long-Term Evolution

MME: Mobility Management Entity

NRF: Network Repository Function

NSSF: Network Slice Selection Function

OMEC: Open Mobile Evolved Core™

ONOS: Open Network Operating System™

O-RAN: Open Radio Access Network

PCF: Policy Control Function

PCRF: Policy and Charging Rules Function

PDR: Packet Detection Rules

PFCP: Packet Forwarding Control Protocol

P4RT: P4 Runtime™

QER: QoS Enforcement Rules

QoS: Quality of Service

RBAC: Role-based access control

SBI: Service-Based Interfaces

SMF: Session Management Function

SPGW-C: Serving and Packet Data Network Gateway - Control

SR-IOV: Single Root Input/Output Virtualization

UDM: Unified Data Management Function

UDR: Unified Data Repository Function

UDSF: Unstructured Data Storage Function

UPF: User Plane Function

About ONF

The Open Networking Foundation (ONF) is an operator-led consortium spearheading disruptive network transformation. Now the recognized leader for open source solutions for operators, the ONF first launched in 2011 as the standard bearer for Software Defined Networking (SDN). Led by its operator partners AT&T, China Unicom, Deutsche Telekom, Google, NTT Group and Turk Telekom, the ONF is driving vast transformation across the operator space. For further information visit http://www.opennetworking.org