Network-optimized cloud architectures for session border controllers

White paper

About the NFV Insight Series

The application of virtualization and cloud principles, such as network functions virtualization (NFV) and software defined networks (SDN), represent a major shift for the communications and networking industry. Until recently, this approach appeared to be unworkable because of stringent performance, availability, reliability, and security requirements of communication networks. Leading communications service providers (CSPs) now implement Network optimized Cloud architectures for SBC and SDN to gain competitive advantage through increased automation and responsiveness, as well as by delivering an enhanced customer experience, while reducing operational costs. This series of briefs and white papers addresses some of the key technical and business challenges faced by service providers as they move Session Border Controllers (SBC) functions to the cloud.

Contents

NOKIA

Introduction

The transition of physical network functions (PNF) from proprietary hardware to IT commercial-off-the-shelf hardware and the subsequent introduction of virtualization and cloud technologies to produce VNFs are under way. By making better use of a shared, multi-purpose infrastructure and increasing the use of automation to reduce life cycle management complexity, they have produced tangible benefits for CSPs. However, these evolutions did not produce significant changes to the way these elements operate or their operational footprint, and internal operational complexity remains. What's more, operations such as capacity scale-out remain challenging. That's because peer VNFs are involved and need to be aware of new components, require configuring connectivity to the new components, and oftentimes, inter-VNF manual engineering. On top of providing a collection of cloud-based virtualized network elements, much work remains. In particular, there are opportunities to foster delivery of dynamic session and data plane services, as well as an improved operations and management experience.

The emergence of several software technology mega-trends are prompting significant re-thinking of telecommunications products to enable more effective operations, as well as potentially opening up new markets. These trends include the use of DevOps techniques, cloud native architectures and microservices for composing systems, in addition to the use of SDN to complement VNFs.

This white paper examines how these technologies and techniques can create a network-level optimization approach for the SBC. It also discusses how the SBC's connectivity and security functions can be intelligently distributed across the network to achieve cloud-native operations on the signaling and media plane at scale.

From cloud-ready to cloud-optimized

Refactoring a VNF into a collection of microservices[1], which employ cloud-native architectures, and adopting more agile DevOps practices improves deployment flexibility and reduces time-to-market of new features.

Microservices are simple. They perform one basic function very well[2]. Their dataless and stateless nature makes them highly cloud elastic. A microservice scales easily by adding instances to a pool of workers. Pools of microservice functions can be distributed and scale independently—leading to optimized network configurations. Furthermore, rapidly emerging container orchestration and lifecycle management capabilities simplify microservice management.

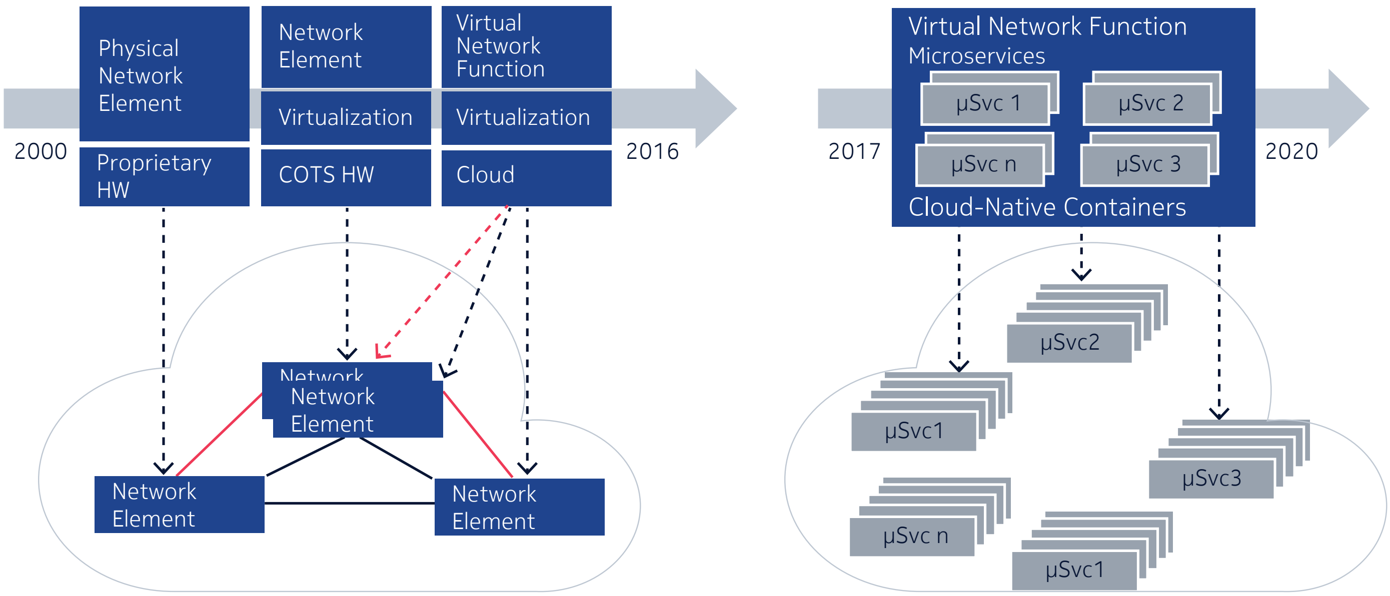
It is also possible to update only the components affected by changes to a new feature. Because of this, it is no longer necessary to organize software product deliveries into major releases over a period of months or years. As a result, this microservices transformation enables agile R&D organizations to deliver software and easily update it as required.

The adoption of a microservices architecture also presents an opportunity to network optimize microservices deployment by evolving from a monolithic to a distributed VNF based on function, location, type of connectivity, and performance.

1 Anatomy of a Microservice (https://resources.nokia.com/asset/201751)

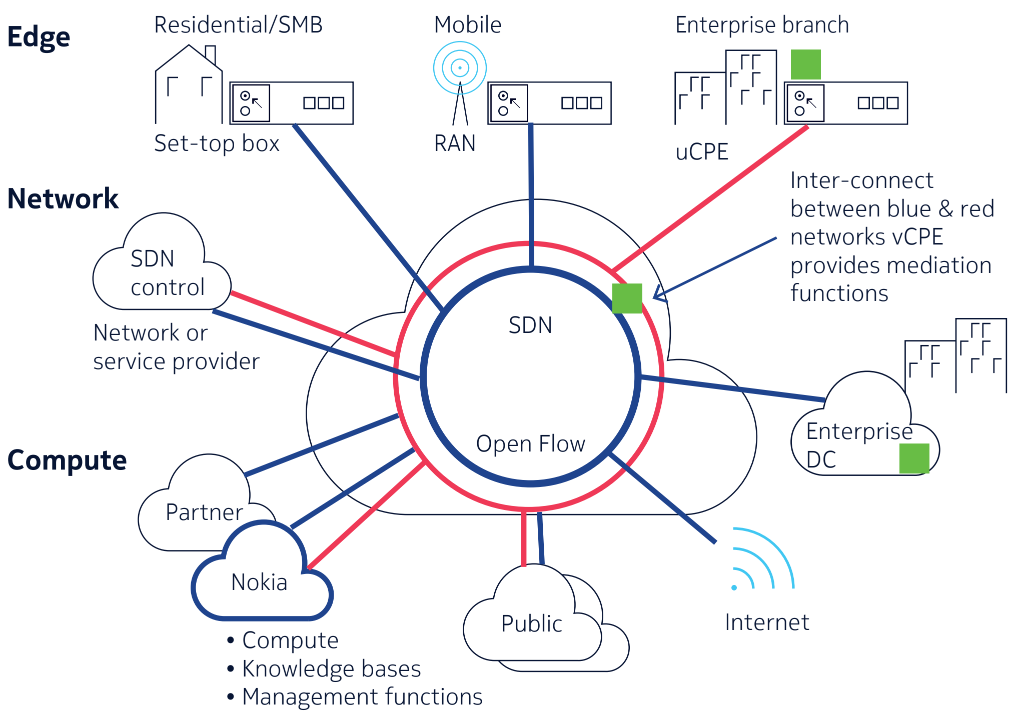
2 This minimalist, modular software design approach rooted in UNIX philosophy is often applied to microservices.

Figure 1: Transition from physical components to microservices



Instead of a collection of individual VNFs, this "post-VNF" view brings together a collection of network-wide functions. As a result the network element-based telephony model is transformed into a network-optimized distribution of functions, which aligns with best-of-breed web and IT industry architecture practices.

Figure 2: Network functions distributed across the network



Cloud-optimized SBC

A typical SBC is made up of several cooperating services, including firewall, access control, security, transcoding, service enablers, and charging, which may reside in different parts of the network. In pre-cloud deployments, SBCs often used the appliance model in which all the functions were in purpose-built chassis that were deployed at a CSP point of presence. Each point of presence was relatively small, with dozens or hundreds of network elements situated throughout the network.

In the cloud era, SBCs became larger and more centralized. SBCs also incorporated more and more intelligence to better secure and distribute SIP signaling and multiple media formats to the cloud core network. Although this approach reduced the number of VNFs to be managed and provided operational savings, new challenges arose. That's because cloud networking infrastructures often lacked the low latency and low jitter characteristics needed for media plane processing. Acceleration technologies, such as Data Plane Development Kit (DPDK) and Single Root-Input/Output Virtualization (SR-IOV), were needed.

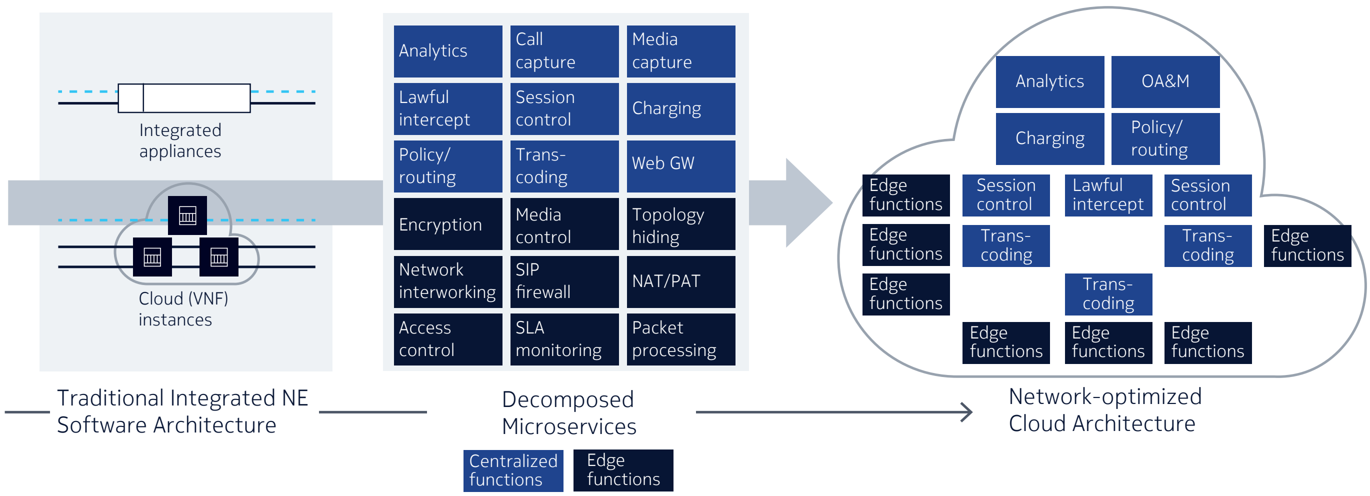
Microservices, coupled with SDN, enable a decomposed SBC to position each service where it is best suited. Signaling, media control, and lawful intercept functions can be provided centrally at scale, or regionally to conform to regulatory needs. Transcoding functions can be deployed in pools located throughout the network as needed, depending on the density and types of transcoding required. Rather than engineering excess capacity at each point of presence[3], the transcoding needs of a tens or hundreds of points of presence can be met with a shared pool. This approach requires lower overall capacity and its costs can be amortized across multiple points of presence. Microservice-based transcoding capacity can scale independently of signaling and media plane capacity. When new codec types, such as EVS, are deployed within a network, transcoding pools can be added which are engineered for the performance of those codecs without affecting existing capacity. Microservice can also be deployed at each point of presence for sites engineered for high-density transcoding.

OA&M functions can be deployed centrally in the core cloud and shared across the signaling, transcoding, and bearer plane components. The same is true for other operations-related functions, including analytics and charging.

Over the long term, microservices which provide bearer plane operations on media flows, such as DDOS firewalls, access control, QoS enforcement, NAT and encryption/decryption, can be deployed at the edge of the cloud—or in purpose-built edge clouds—where latency and jitter can be more easily managed. Service providers may elect to provide these functions using on-premises equipment, such as virtual customer premises equipment (vCPE) that resides in an enterprise data center and connects to the cloud-based instances of core services.

3 For example, the Erlang-B formula is a statistical model used to assess the probability of blocked service

Figure 3: SBC functions distributed across the network



A cloud-optimized SBC is a collection of network-wide services rather than a collection of individual VNFs. These services may be derived from existing components, by way of re-factoring, or they may be acquired from the network infrastructure or even service provider-furnished microservices. The specific set of services offered may vary according to the market and end-user scenario to be supported.

Optimizing the signaling plane

In a network-optimized signaling plane, the signaling plane services, including call session control, and media control are de-coupled from the data and management planes. SIP-based session management, such as Proxy-Call Session Control Function (P-CSCF), has stateless, N+K operation supported by a load balancer. It is defended by an application-aware firewall function, which can be deployed at the edge of the network. The signaling plane sends instructions to the media controller, which sets up paths in the optimized data plane for the bearer traffic that is admitted and routed. The evolved media controller consolidates the communications path into a single step based on standard IT network protocols, instead of legacy Diameter, H.248 or proprietary interfaces. Media control components can be positioned near the signaling plane in the core or closer to the data plane.

Optimizing data plane performance

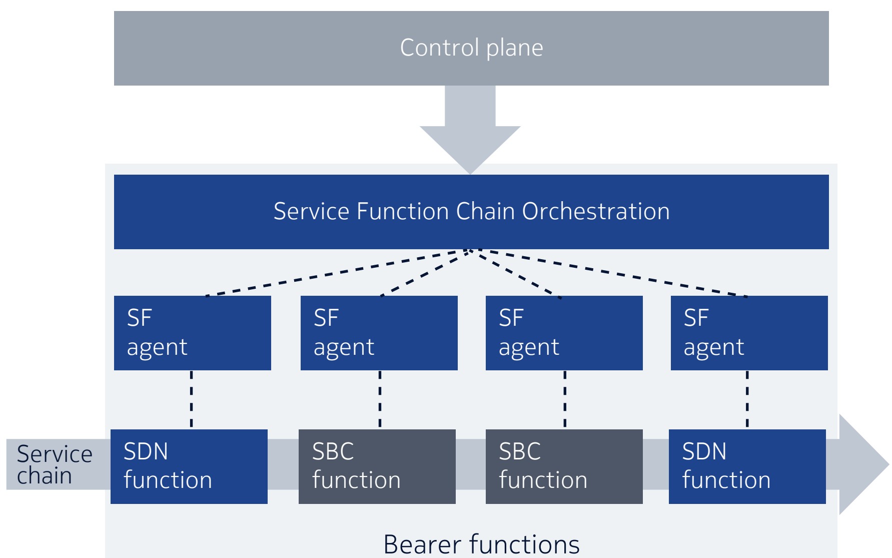
Traditional network technologies were closely tied to the purpose-built hardware on which they were run. This has made it difficult to integrate new functions, or to customize existing ones to handle VoIP traffic. For the traditional SBC, this has also led to implementation of required data plane capabilities in overlay border gateways that are separate from the IP routing layer. These to ran on dedicated hardware. As a result, many commonly available network functions are replicated as built-in, proprietary functions. This formerly necessary approach runs counter to the requirements of contemporary cloud networks, which must be able to flexibly configure both control plane and data plane capabilities for signaling, call control and bearer functions.

Software-defined networks

Networking technologies, such as SDN, allow operators to create flexible "service chains" of data-plane functions. In the future, SBC data-plane bearer border gateway functions will be directly integrated into these SDN service chains. Integration of data plane SBC microservices into the service provider's SDN can lead to significant network optimization, in addition to providing a mechanism to introduce new capabilities[4].

SDN also allows the creation of secure overlay networks, which allow SBC components to be connected in a virtual private network that can span cloud availability zones, data centers, edge clouds, and customer premises equipment. If desired, a set of containers can represent a particular enterprise, new service offering, or a 5G network slice. These dedicated components are created, updated, and removed from the network without impacting other users. Moreover, because SDN provides the overlay connectivity, the need for coordination of, for example, VLAN assignments between the network infrastructure and the SBC application level, is avoided, as is the resulting operational and configuration complexity.

Figure 4: Service Function Chains



By using service function chains, service providers can bundle SBC data plane functions with other network functions to create unique end user service offerings. Meanwhile, enterprise users can couple SBC-specific functions with preferred services implementations, such as firewalls and VPNs.

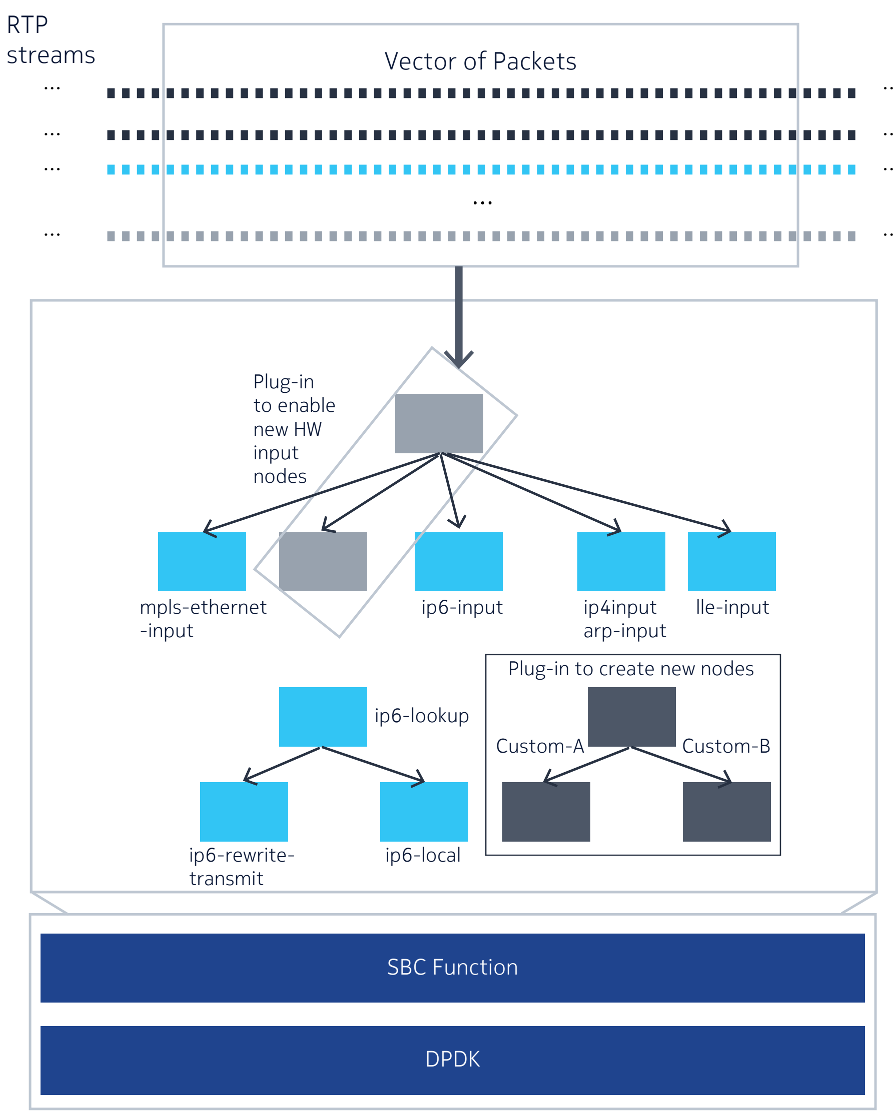
4 For a more in depth discussion, consult the Nokia white paper: "A Software-Defined Networking Approach to SBC".

Vector Packet Processing

Vector Packet Processing (VPP) is an open source, rapid packet processing development platform supported by the FD.io (a Linux Foundation project). It enables high performance on general-purpose processors for bearer functions that previously required specialized hardware. VPP leverages DPDK to provide a user space framework for packet processing operations. As a result, an SBC can provide line-rate fast path processing in containers without being dependent on the kernel or a vSwitch, or which formerly required dedicated silicon.

Since its inception as an open source project, FD.io's VPP has seen broad adoption, having been accepted as a key project with OPNFV[5], alongside well-known projects, such as OpenStack and OpenDaylight.

Figure 5: Vector Packet Processing



5 https://www.opnfv.org/

The VPP approach processes vectors of packets as a group instead of processing packet-by-packet. The first packet in the vector primes the CPU instruction cache. This allows the remaining packets to be processed at an extremely high rate. The performance cost of processing the first packet is amortized across the entire vector, yielding very high performance, as well as statistically reliable performance. As a result, throughput and latency are very stable—two critical requirements for the SBC data plane, which consists primarily of RTP packet flows.

Within the VPP framework, packet vectors are processed using a directed graph. This is a tree of nodes, which defines how to process the vector of packets. Each node performs a simple operation on the group of packets and hands it off to the next node. In some respects, VPP is like microservices and SDN on a nano-scale. If multiple CPU cores are available, the VPP graph scheduler can schedule {vector, graph node} pairs to different cores.

The open source VPP eco-system contains a large base of existing network functions implemented as nodes. This means, the evolved SBC can use open source components that are tuned for data plane operations as commoditized networking functions, rather than using components designed to maintain proprietary implementations. Where features gaps exist, the processing architecture is extensible, allowing it to meet the needs of an SBC by adding user-defined plugins to the node tree structure.

VPP fast path capabilities are complemented by a high-performance API, which allow it to be controlled by SBC application components using a VPP management agent.

Optimizing the routing and policy

In addition to optimizing the SBC signaling plane to deliver high-performance IP voice and video communication services, a microservices architecture increases the value of associated routing and policy services in the network. Routing and policy engines, which are used by the session border control function and other functions, are distributed on a regional basis. Their databases are set up with a centralized master copy that is replicated over the regional engines. Fine-grained policies are cascaded to relevant signaling and data plane component instances, where needed, for low-latency routing and forwarding decisions.

The central master database provides a single place to manage routing plans and customer-specific policies, wherever the affected network component is deployed. This allows the SBC components as commoditized networking functions and other network elements to provide a network-wide distributed service mesh for the call processing and bearer planes.

Microservice functions subject to regulatory and geographic constraints are deployed regionally in compliance with applicable rules. Functions, such as lawful intercept, call control, or the transition from one administrative domain to another, are typical examples. To ensure conformance, deployment polices are applied to component placement during the life cycle management process.

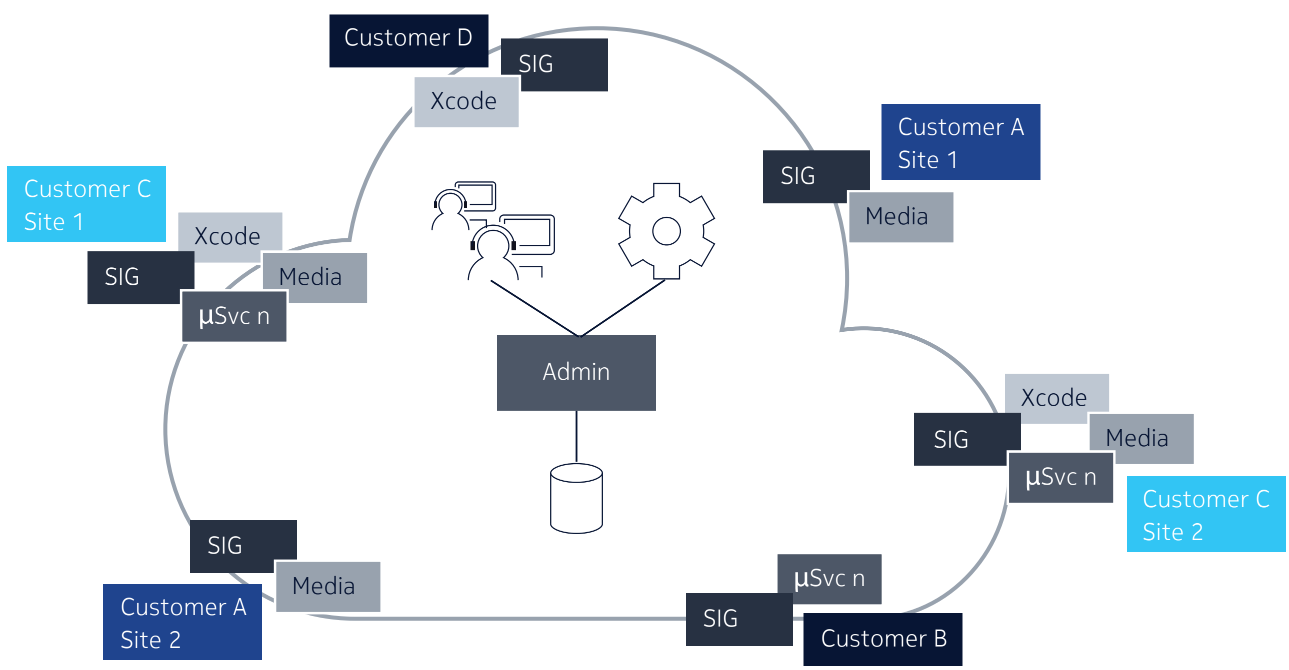
Optimizing management

A microservice-oriented, cloud-native architecture allows for optimal placement of SBC components throughout the network. Going a step further, SDN enables these same components to be connected together as part of a distributed SBC that is designed to meet the needs of an enterprise, a data center, a region, a 5G network slice, or the operator's whole network.

The operation, administration and management (OA&M) components can likewise be decomposed into microservices. As shown in Figure 6, these OA&M functions can be deployed separately from signaling control, media control, data plane, transcoding function, and other decomposed microservices.

In general, these OA&M components are deployed centrally in the core cloud, with connectivity to the distributed components. However, if the operator has multiple administrative domains, these components can be deployed on a per-domain basis. Centralized management provides enterprise customers with a single point of administration even when SBCs are situated in multiple branch locations.

Figure 6: Centralized management of the network-optimized SBC



Rather than separate administration of individual SBCs or administering SBCs clusters through an element management system (EMS), operations are optimized by providing a central point for administration into an API-driven service portal. This approach offers several advantages:

- Application configuration data can be prepared once and replicated as components are added
- Microservice components can find each other using service discovery and can respond to topology changes
- Software deployment to the SBC microservices, wherever they are located in the network, are under the control of life cycle orchestration

- Policies and routing data are defined once and distributed to the services, which consume them, caching them locally where needed for performance.

Following the "API first" principle for microservices, these administration functions interact with the traffic-related services they support using clearly-defined interfaces.

A centralized call detail recording microservice collects call activity data from the other microservices—signaling control, data plane—and stores it in scalable cloud storage according the operator's security and retention policies. This relieves each microservice type and instance from having to engineer and maintain its own storage at each location. Avoidance of fixed storage resource allocation simplifies the microservice, lowers costs, and increases deployment flexibility.

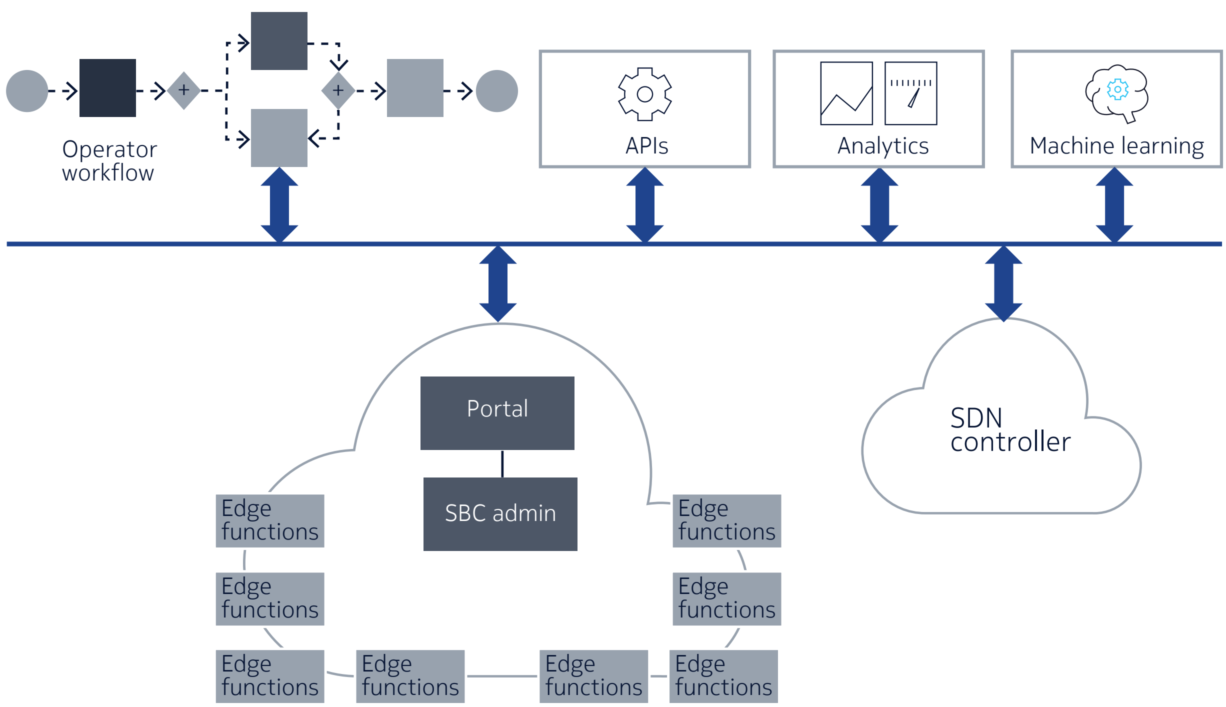
Similarly, analytics functions ingest data streams consisting of performance measurements, call trace, and logs from the SBC microservices. This information supports troubleshooting and diagnostics of network issues, trend identification, and insights, which lead to network-level optimizations based on the traffic patterns and/or external events. When coupled with data from an SDN controller, the service provider can produce end-to-end views of connectivity and network performance at several levels:

- Per-site
- Per-region
- Enterprise-wide
- Enterprise to enterprise
- Enterprise to core cloud
- Per-5G network slice.

Optimizing opportunities

NFV and SDN technologies also present new opportunities for dynamic revenue-generating services as illustrated in Figure 7.

Figure 7: Self-optimizing networks



By including an API Gateway to handle configuration and life cycle operations, CSPs can offer enterprise end users self-service administrative web portals. Enterprise customers can choose the components which align with their IT policies and practices, and select the desired service level agreement (SLA). These choices are composed into automated workflows which are executed on behalf of the enterprise user:

- Set up SDN connectivity
- Define service chains, including network functions, such as session border control functions and SD-WAN
- Create enforcement policies for user requests
- Deploy the needed software components.

This approach allows enterprises to manage data that applies to their deployments, such as routing tables and feature selection, and obtain usage statistics reports.

Analytics algorithms can identify changes in traffic patterns and proactively trigger configuration changes to improve service quality using the API Gateway. When used in conjunction with machine learning tools, behavior anomalies and malicious traffic can be identified and mitigated through the automated deployment of network protection policies—much like PC-based virus detection works.

This type of dynamic configurability enables CSPs to cost-effectively address the needs of small-to-medium businesses, as well as enabling DevOps-like service delivery to current customers.

Conclusion

Nokia is committed to drive agility in cloud services, thereby supporting CSPs and enterprises as they transform operations and embrace digitalization. By evolving towards a cloud-native architecture and by taking advantage of network-optimized cloud architectures, limitations of conventional and monolithic SBC offerings can be transcended.

Intelligently distributing SBC functions to where they are best suited and transforming products and processes with emerging networking technologies will offer service providers faster, cheaper, and more robust solutions to control the access and peering borders of their networks.

References

1. A Cloud-Native Vision for SBC
2. Anatomy of a Microservice
3. A Software-Defined Networking Approach to SBC
4. FD.io: The Universal Data Plane
5. Migrating Session Border Controllers to the Cloud

Acronyms

DC	Data Center
DPDK	Data Plane Development Kit
NE	Network Element
NFV	Network Function Virtualization
OAM	Operation, Administration, and Management
OSS	Operational Support System
QoS	Quality of Service
PNF	Physical Network Function
SBC	Session Border Controller
SDN	Software Defined Network
SLA	Service Level Agreement
SR-IOV	Single Root-Input/Output Virtualization
VM	Virtual Machine
VNF	Virtual Network Function
VPP	Vector Packet Processing

Nokia Oyj
Karaportti 3
FI-02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Product code: SR1801020609EN (January)