December 2019

Author: Dr Phil Marshall

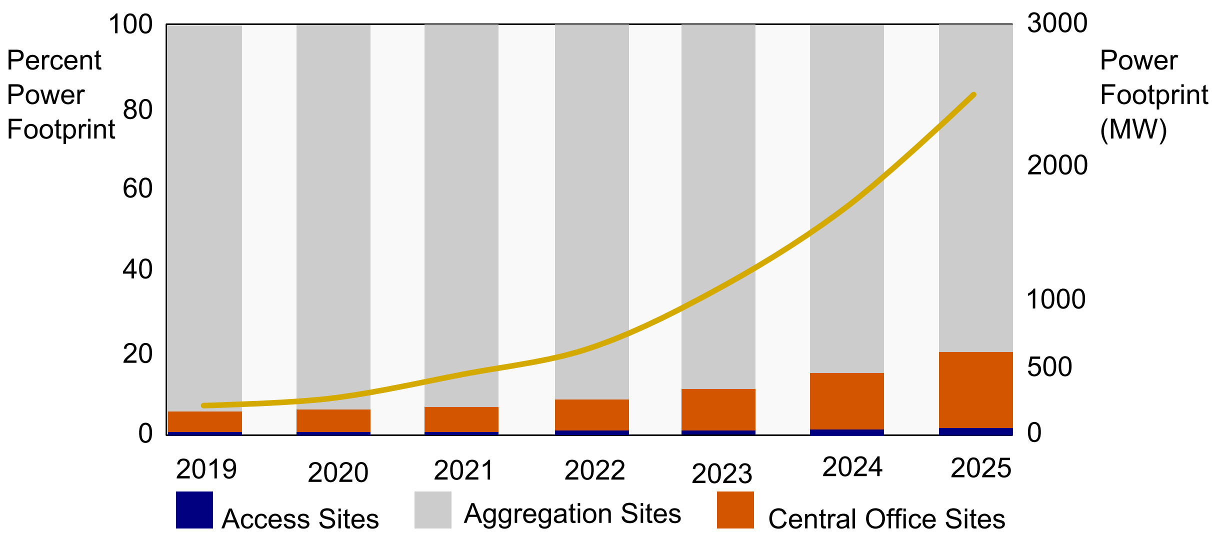
Communication Service Providers Lead the Charge for the Edge Cloud

Executive Summary

Communication service providers (CSP) are at the forefront of the Edge Cloud, primarily with central office deployments to support their core and transport network upgrade and virtualization efforts. The global power footprint of the Edge Cloud equipment for CSPs is forecast to increase from 231 to 2520 MW between 2019 and 2025. CSPs are slow to adopt cloud RAN and in most cases are rolling out 4G and 5G using physical rather than virtualized radio equipment. We forecast that cloud RAN will gain modest market momentum after 2022.

CSPs are well positioned to provide Edge Cloud solutions, and connectivity services, such as local breakout for latency sensitive applications. However, their success with Edge Cloud is not a foregone conclusion. To succeed, CSPs must be pragmatic towards their positioning in Edge Cloud ecosystems and the commercial relationships they pursue. Particularly relative to public cloud service providers like Amazon, Google and Microsoft, who are also well positioned to spearhead platform-centric Edge Cloud solutions that seamlessly integrate with their respective cloud environments.

Exhibit 1: The Edge Cloud enables CSP network virtualization and cloudification
Source: Tolaga Research, 2019



Introduction

Communication service providers (CSP) have generally led Edge Cloud (defined on Page 8) adoption since it is needed for network upgrades and virtualization. Edge Cloud platforms are being used for a variety of key functions such as network function virtualization (NFV), software defined networking (SDN) and multi-access edge computing (MEC). For the most part, CSPs have focused on core and transport network virtualization.

Most CSPs are still using conventional as opposed to virtualized platforms for next generation access technologies, such as 5G-RAN (radio access network). The aggregate power footprint of Edge Cloud equipment for CSPs network virtualization and cloudification is forecast to have more than a ten-fold increase from 231MW in 2019 to 2520 MW by 2025, see Exhibit 1. The lion's share of edge compute resources for CSPs will be deployed in central office sites, with the support of key initiatives such as CORD (Central Office Re-Architected as a Data Center). After 2022 the percentage of edge computing equipment in aggregation sites is forecast to increase, as CSPs deploy targeted cloud RAN solutions.

Virtualization is a disruptive multi-year journey for CSPs

Network virtualization and cloudification is challenging. For most CSPs, the favorable return on investment (ROI) estimates predicted circa 2015 have not yet occurred. Contrary to these predictions, short term capital and operational expenditures typically increase when CSPs pursue virtualization strategies. It is a long and expensive transformation journey for CSPs, but if done right, virtualization ultimately brings tremendous financial rewards. For example, AT&T has trail-blazed with network virtualization and since publishing its vision under the guise of Domain 2.0 in 2013. It has aggressively pursued its virtualization efforts since then, and according to its CEO Randall Stephenson is on target to have 75 percent of its core network virtualized by the end of 2020. At a recent Goldman Sacs conference in New York, Stephenson stated that with virtualization AT&T has achieved roughly 17 quarters with 7 to 8 percent year-on-year cost savings.

To realize the benefits from virtualization, CSPs must:

- **Transition from telecom-to-IT centric operations** and find an appropriate balance between both. Fundamentally, telecom operations are about managing, and IT operations are about abstracting complexity. Much of this abstraction comes through operational automation, with initiatives like ONAP.
- **Enable agile, reliable and scalable orchestration.** This is supported by key standardization initiatives such as the ONAP Service Orchestrator framework, but remains challenging particularly for large-scale implementations.
- **Ensure hardware platform scalability**, particularly to address the dynamic performance demands typically experienced in networks. In contrast to hyperscale data-centers, which have seemingly infinite compute resources, the resources available for virtualized network functions are relatively scarce since they are distributed in the Edge Cloud. Specialized computing platforms are

needed for specific networking functions and will continue to be needed for the foreseeable future. The notion that all network functions in the Edge Cloud can be kitted out with x86 processors running on standard Dell or HPE platforms is a pipe dream for CSPs, and falsely assumes that the Edge Cloud is a scaled down version of a hyperscale data center. Infrastructure vendors including Cisco, Ericsson, Huawei, Juniper, Nokia and ZTE have been working closely with silicon vendors like Intel and Broadcom to deliver the necessary Edge Cloud platform performance. In addition, some vendors have developed their own specialized semiconductor solutions. For example, Huawei has its Ascend and Nokia its FP4 semiconductor products that are designed to address specific network functions.

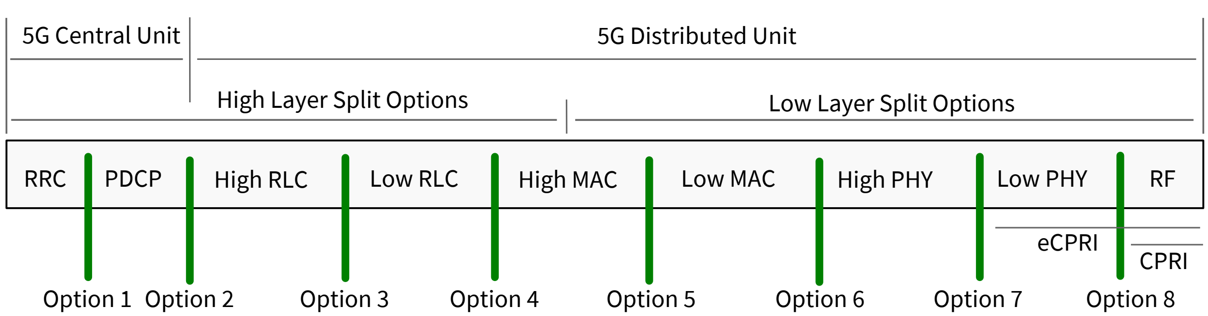
As more CSPs succeed with their network virtualization efforts, solutions will be refined, and best practices developed. We expect this will drive the robust expansion of the Edge Cloud, as reflected by the forecasts shown in Exhibit 1.

We forecast modest growth in access network virtualization, such as cloud-RAN, after 2022 when we expect the benefits of virtualization can be adequately realized.

Cloud RAN and 5G

Cloud-RAN implementation requirements vary greatly depending on whether or not real-time RAN functions are virtualized. According to the 3GPP, a radio base station can be divided into nine logical functions, which are illustrated in Exhibit 2. Three of these functions that have non-real-time demands include, the Radio Resource Control (RRC), Packet Data Convergence Protocol (PDCP), and non-real-time functions of the Radio Link Control (RLC). The remaining six functions have real-time demands with low latency Edge Cloud requirements when virtualized. These six functions include, the real-time components of the RLC, the upper and lower Medium Access Control (MAC) and Physical layers, and the RF. Conventional CPRI (common public radio interface) is defined between the Lower Physical and RF layers, and eCPRI between the Upper Physical and RF layers.

Exhibit 1: Cloud RAN functional separation
Source: 3GPP, 2017, and Nokia and Tolaga Research, 2019



In a cloud RAN implementation, non-real-time functions can be virtualized in central office sites. Real-time functions, when virtualized must be distributed closer to radio sites to meet the necessary latency demands. For example, if a cloud RAN is fully virtualized with a CPRI interface from the Edge Cloud to a radio site, connection latencies must be less than 200 micro-seconds.

In most cases, fully fledged cloud-RAN with virtualization to the CPRI interface is only economically viable where dedicated fiber front haul connectivity is readily available, in urban areas with particularly high site densities, and for targeted enterprise campus and venue implementations. Otherwise, fully fledged cloud RAN is uneconomic.

The 3GPP identifies eight cloud RAN implementation options, see Exhibit 2. To date Option 2 and Option 8 have been pursued commercially. Other options are being investigated by technology vendors and CSPs and might be pursued in the future.

Today most Tier 1 CSPs are continuing to focus on physical as opposed to cloud-RAN architectures for their 4G and 5G network deployments. Several exceptions include Rakuten, which is a new entrant operator in Japan, Verizon in the United States and China Mobile, Unicom and China Telecom. CSPs with high-bandwidth fiber connections to their radio sites are more likely to deploy cloud-RAN in the future. We expect that these players will review cloud-RAN through multi-year studies to ascertain optimal deployment strategies.

Growing support for OpenRAN

Cloud-RAN goes hand-in-hand with the OpenRAN initiative, which operates under the tutelage of the Telecom Infrastructure Project (TIP). OpenRAN has the objective to standardize radio base station interfaces and creates open architectures that theoretically enable base-station sub-system mixing and matching amongst different vendors. It also simplifies northbound integration for third party tools, such as for AI optimization.

OpenRAN is benefiting from growing CSP support. In addition, all Tier 1 radio infrastructure vendors, with the exception of Huawei, are members of the O-RAN Alliance.

We believe that OpenRAN is valuable to the industry, but there is a chasm between the theoretical and realistic impact of OpenRAN on radio base station ecosystems. In particular:

- Heightened radio technology innovation is a hallmark of 5G and an important area of differentiation for radio vendors and their CSP customers. For example, 5G requires broadband channels, operates in new frequency bands (including mmWave) and most commonly uses time division duplex (TDD) channel structures, which facilitates the radio channel symmetry need for massive MIMO. This creates environments of heightened innovation where mixed vendor radio base station solutions will inevitably under-perform an equivalent solution provided by a single vendor.

- Feature transparency is a challenge when mixing and matching solutions amongst different vendors. In particular, the features supported by one sub-system vendor might not be supported by another. Ultimately the CSP can only implement the subset of features that are supported by all sub-system vendors.
- CSPs need 'one throat to choke' when systems don't function as intended. This is a common cliché used to justify integrated solutions, and under most circumstances it is a 'big yawn'. However, at this early juncture in 5G we believe it is important.

In Japan, Rakuten is putting the OpenRAN principles to the test. As a green-field CSP, Rakuten is not incumbered by legacy technology and operations, and is embracing virtualization throughout its entire network. Rakuten is deploying virtualized RAN and plans to use OpenRAN to integrate base band equipment from Altiostar (a Massachusetts based startup) with Nokia radios. While Rakuten's approach is bold, we believe it is overly risky at this market juncture.

Rather than positioning OpenRAN as an alternative for conventional radio base stations deployments, we believe that OpenRAN is better suited to specialized implementations, such as for industrial, transportation and public safety applications. These applications might benefit from specific functionality that isn't ordinarily supported in conventional base stations, but can be provided by a third parties.

Local breakout and CUPS

The Edge Cloud not only depends on edge applications being deployed where they are needed, but it also requires CSP connectivity between the Edge Cloud and endpoint devices, with adequate bandwidth and latency performance. When the Edge Cloud is deployed close to an end-user, local breakout is needed to ensure that the Edge Cloud connectivity follows a sufficiently short path between the end user and Edge Cloud server. This is achieved with local breakout capabilities, which ensures the user's data connection is terminated at a peering point collocated, or in a physical location close to the Edge Cloud. If the peering point is not physically close the Edge Cloud, then connection tromboning can create excessive latencies, irrespective of where the Edge Cloud is deployed.

Theoretically, when a local breakout or peering point is needed, it places CSPs in the 'box-seat' for Edge Cloud market opportunities. However, it is not necessarily that simple for the following reasons:

- **Market competition** can have a tremendous impact on local breakout availability. For example, in the United States, there are enough market participants for local breakout to be readily available for fixed connectivity. In other markets like the Philippines, there is cut-throat competition amongst CSPs to the extent that inter-CSP peering points are located Internationally. In these cases, individual CSPs can have tremendous control over local breakout availability.
- When Edge Cloud services are delivered to **Mobile Devices** over a public network, the connectivity is determined by the mobile CSP providing the service. When an Edge Cloud is deployed at Central Office locations, the conventional connectivity architectures used by CSPs is likely to be adequate. However, when it is necessary to bring the Edge Cloud closer to end-users, such as at aggregation and access sites, connection tromboning is likely to become a problem. To address this challenge, 3GPP-Release 14 introduced Control User Plane Separation (CUPS) technology in 2017. With

CUPS, connection tromboning can be eliminated by separating the control and user plane termination points, with the user plane being terminated locally. Since CUPS must be provisioned in the mobile networks that connect to Edge Clouds, it requires mobile CSPs support, and might include key technologies such as Multi-Access Edge Computing (MEC).

Driving towards platform-centricity

Since the Edge Cloud is still nascent and infrastructure-led, it is not readily available for developers. To thrive, the Edge Cloud must transition from Infrastructure-led to platform-centric architectures. With platform-centricity, the Edge Cloud will provide standardized and user-friendly environments for developers to self-provision services. Developers will favor consistency across markets and amongst different CSPs. This underpins the value proposition of companies like EdgeGravity (Ericsson) and MobileEdgeX (Deutsche Telekom). Both EdgeGravity and MobileEdgeX have established standardized platforms that bring together ecosystem players that include CSPs, technology providers and public cloud companies. We believe that the platform-centric strategies being pursued by organizations like EdgeGravity, MobileEdgeX, and others are tremendously important in propelling Edge Cloud adoption.

In contrast to centralized cloud platforms, which are deployed in hyperscale datacenters with seemingly infinite general-purpose computing resources, the Edge Cloud is highly distributed with scarce hardware resources that are optimized for the specific services being supported. As the Edge Cloud matures and is densified, application specific resources might be pooled across multiple Edge Cloud platforms with the benefit of sophisticated orchestration to increase resource scalability. However, in the interim, application specific Edge Cloud implementations will prevail and challenge platform-centric approaches.

CSPs and the Public Cloud

Edge Cloud demand currently comes primarily from workloads that migrate from the central cloud to the edge, and cloud workloads that are complemented with edge functionality. Public cloud providers including Amazon and Microsoft have responded with solutions for the Edge Cloud and User Edge (see definition on Page 8) with consistent development tools and seamless interoperability with their respective centralized cloud offerings. Amazon's solutions include Outpost and Wavelength, and Microsoft has its Azure IoT Edge.

Many CSPs are establishing cooperation and partnership strategies with public cloud providers. For example, AT&T is deploying Azure IoT Edge in its Edge Cloud facilities and making it available to its enterprise customers. Both Vodafone and Verizon recently announced that Amazon's Wavelength product will be deployed as part of their Edge Cloud offerings. Telefonica has partnered with Microsoft for Edge Cloud capabilities, and Sprint recently partnered with Amazon for Edge Cloud capabilities in its IoT Curiosity network.

We expect that Edge Cloud cooperation and partnership announcements between CSPs and public cloud providers will come thick and fast over the next 24 months. In most cases, CSPs are looking to bundle Amazon or Azure Edge Cloud functionality with their retail enterprise services. However, the existing enterprise services for most CSPs essentially consist of bundled consumer products, which are a significantly different proposition to Edge Cloud services. History might be set to repeat itself. In the past CSPs have been largely unsuccessful in selling public cloud services using a similar approach.

Some Tier 1 CSPs plan to wholesale, rather than retail their Mobile Cloud offerings and rely on companies like Amazon and Microsoft to develop the requisite retail enterprise services. We believe that this approach is prudent, but will depend on careful pricing and positioning strategies to ensure that the CSPs retain sufficient value for their Edge Cloud solution.

Conclusion

CSPs are leading the charge with the Edge Cloud and are well positioned to capitalize on the strength of their real estate and network assets as the market develops. CSPs are deploying Edge Cloud equipment in their central office sites primarily to support core and transport network virtualization. Although cloud RAN and openRAN have garnered significant interest over many years, they are not expected to gain meaningful market traction until after 2022. When this occurs, an increased percentage of the Edge Cloud will reside in Aggregation and in some cases Access sites. However, most of the Edge Cloud footprint for CSPs will be in central offices for the foreseeable future.

For the Edge Cloud to proliferate, it must evolve from infrastructure-led to platform-centric environments for developer communities to thrive. As this occurs, diverse Edge Cloud hardware designs will still be required and will complicate scalability efforts.

Public cloud providers including Amazon, Google and Microsoft are likely to provide the dominant Edge Cloud platform environments, particularly for enterprise services. Recognizing this, a growing number of CSPs are partnering with public cloud providers for their Edge Cloud solutions. As CSPs establish these partnerships they must pragmatically evaluate their Edge Cloud enterprise channel strategies.

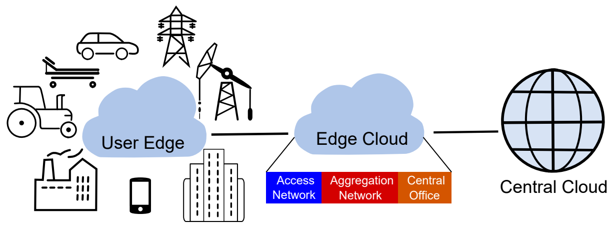
About Tolaga Research

Tolaga Research is a leading consulting and advisory firm with a focus towards communication networks and the Fourth Industrial Revolution. Tolaga was founded in 2009 and is the world’s first firm to apply artificial intelligence with natural language processing and system dynamics modeling to industry research. By combining these sophisticated capabilities with its extensive primary research, Tolaga delivers unique and actionable insights that are fortified with robust data science and system modeling solutions.

For more information, contact us at: admin@tolaga.com

Tolaga Edge Computing Definition

Edge computing lacks a clear definition, which varies depending on perspective. For cloud service providers, edge is an extension of their cloud offering. For communication service providers, edge compliments their network and connectivity services, and for device and embedded solution providers it is an extension of their device applications. At Tolaga, we define edge computing in terms of two general categories, namely the User Edge and the Edge Cloud.

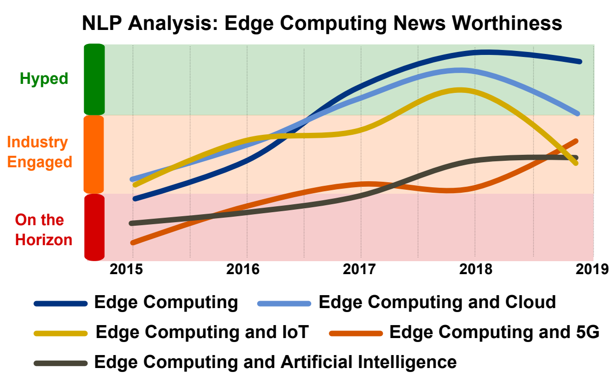


The **User Edge** consists of three general types, including end-point devices, gateways and on-premise computing equipment. End-point devices include IoT sensors, embedded systems, such and electronic control units in vehicles, and smartphone and tablet devices. Gateways include routers and aggregation platforms. On-premise equipment consists of specialist and general-purpose computing.

The **Cloud Edge** consists of managed edge computing equipment that is deployed outside of user premises. In contrast to the public cloud, the physical location of the Cloud Edge infrastructure relative to its end-users is important. For our definition of the Cloud Edge, we distinguish between infrastructure that is within a similar distance from end-users as access sites (e.g. radio base-stations), aggregation sites and central offices. As a rule of thumb, there are between ten and twenty access sites per aggregation site and between ten and twenty aggregation sites per central office.

Edge Cloud Priorities and Use Cases

Interest in Edge Computing has increased significantly in recent years, and according to our natural language processing (NLP) of online content, Edge Computing has been hyped in mid-2016, with an emphasis of its role in complementing cloud computing. Between 2017 and 2018, the role of edge computing in the context of IoT became hyped. Since then some attention has shifted to focus on specific edge use-cases and emerging capabilities such as artificial intelligence (AI) and 5G.



The Edge Cloud is a candidate for an increasing number of use-cases that feed the digital appetites of consumers and enterprises. Tolaga is continually expanding its use-case coverage for the Edge Cloud and edge computing. These use-cases currently focus on the following industry segments:

- **Communication Service Providers (CSP):** with use cases for Access, Aggregation and Central Office site deployments. CSPs are deploying edge computing to support the migration of their networks and to support specific use-cases for their customers.
- **Smart Grid:** including use-cases for generation, transmission, distribution, renewable, and operations and maintenance.
- **Mobile Consumers:** with digital use-cases for 4G and 5G subscribers, which include gaming, information, social, health and fitness, messaging and communications, media and entertainment and Internet.
- **Smart Home:** to include use-cases for security, smart applications, infotainment, assisted living and energy management.
- **Automotive:** including use-cases for autonomous vehicles, assisted driving, traffic management, infotainment and operations and maintenance.
- **Smart Cities:** with digital use-cases for smart lighting, traffic signaling, public safety, smart buildings and public venues and utilities/services.
- **Industry 4.0:** to include use-cases for asset tracking, remote operations, logistics and warehousing, operational automation, security management and enforcement, and diagnostics and maintenance.
- **Smart Retail:** with use cases for digital signage, instore surveillance and experience management, proximity marketing and supply chain optimization.
- **Smart Healthcare:** which focuses specifically on hospitals and clinics, with use cases for patient record management, continuous patient monitoring and intervention, remote patient care and intervention (including support for remote surgery) and physical therapy.
- **Commercial UAVs**: with use-cases for mapping and surveying, photogrammetry, 3D modeling and digital elevation modeling, and;
- **Enterprise IT:** which anticipates a proportion of IT workloads that migrate the cloud will reside at the Edge Cloud.