Akraino Whitepaper 2

Cloud Interfacing at the Telco 5G Edge

<1st edition – July 2020>

www.akraino.org

1 Introduction

1.1 Motivation

In Telco 5G edge computing, location is half the battle. Equally important are new Telco network capabilities. 3GPP's introduction of edge computing features adds another layer of advantage to Telcos' ownership of close-to-end-user locations. In addition to reduced latency, edge applications can coordinate processing between multiple edge nodes, save mobile device battery usage, and increase privacy and security through tight, seamless interaction with the telco core network. This whitepaper addresses what it takes to make API interfacing easier and cooperative, presenting an enabler interface layer between edge applications and Telco core networks.

When Clouds extend to the Telco edge, platform technical specifications are decided by feasible business models. Should edge infrastructure be vertically independent or horizontally shared? Who can decide whether it should be container based or stay VM or even bare metal? A top down analysis of Telco 5G edge stack layers may help to draw some conclusions.

We will cover 3 areas: (i) analyze challenges in Telco 5G edge computing technology, business and operations, (ii) propose detailed enabler layers between edge applications and Telco 5G core networks (and mapping Akraino member project APIs to this enabler interface as examples), and (iii) provide a Telco 5G edge infrastructure ownership and operational analysis with example solutions.

1.2 Scope and Assumptions

We focus on 5G edge scenarios, including on-carrier and enterprise premises. Technical internals of Telco core networks and webscalers are not included unless necessary. Figure 1 shows the scope of the paper. Edge access types in the shaded area are not included. The paper refers to edge end user applications in a general sense without going into detail.

The term "MEC" (Multi-access Edge Computing) is not specifically defined by any standards bodies. However, it is widely used for various things in edge computing today. In this paper, we use the term "MEC" to refer to "Edge Computing in/at the access network". We use the term "5G MEC" as a subset of MEC focusing on 5G access. Also, in this paper the terms "Telco 5G Edge" and "5G MEC" are interchangeable.

The term "ETSI MEC" or "ETSI ISG MEC" refers to a standards body developing a set of API specifications for MEC and is mentioned in this paper explicitly as "ETSI MEC". Such mentions do not imply compliance to the full specifications developed by ETSI MEC.

The intended audience is mobile network operators, cloud service providers, mobile network equipment vendors, edge and service vendors, and system integrators.

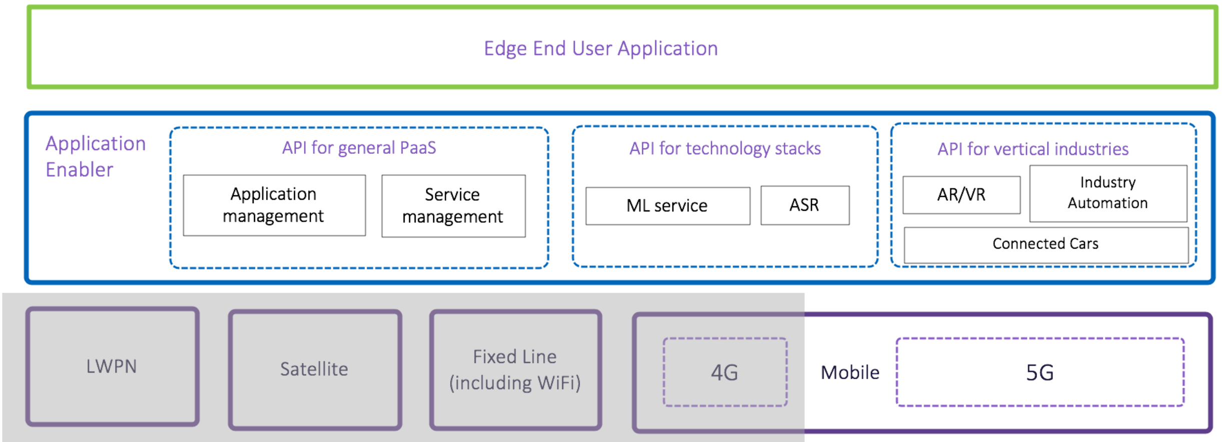


Figure 1 Whitepaper scope

2 Background

Partnership announcements between Telcos and cloud service providers are in the news. Globally there are over 200 operators and a dozen webscalers with a wide range of capabilities and objectives. A one-size-fits-all edge partnership model is very unlikely. What could be common functions required to make these partnerships go smoothly?

The 5G MEC concept is not new, but deployment has been waiting for 5G network deployment. Application vendors are exploring MEC ideas but remain reluctant to move to production.

Meanwhile cloud service providers have advanced significantly in building a global network connecting their data centers around the world. Extension to mobile edges is a natural next step. Ongoing efforts include Google Fi and Facebook's TIP project. However, they face barriers not only in spectrum, but also a massive, complex physical RAN network which has a high entry threshold.

Telco partnerships point the way forward. Revenue producing applications are already running in the cloud. Public cloud service providers have accumulated ample experience in supporting applications from development to deployment and operations, and some have already released edge stacks for non-mobile edge solutions. For example, AWS has IoT edge and Microsoft has their Azure edge stack. No wheels need to be reinvented on those infrastructure platforms. The focus can be narrowed to telco network and cloud interfaces.

3 Technical challenges and requirements

The following sections lay out some example application scenarios and explore application needs and expectations.

3.1 Challenges to leverage 5G network edge capability

5G is a game changer for MEC. 3GPP release 16 (scheduled to be released in 2020) specifies several capabilities supporting mobile edge computing, including:

- UPF reselection
- Local routing and traffic steering
- session and service continuity
- AF influenced traffic steering
- network capability exposure
- QoS and charging
- LADN

A MEC node is physically distributed between a device and its cloud service. Locations of edge nodes and devices directly influence the following functions:

1) Service discovery
2) Device mobility and service continuity
3) Traffic steering
4) Network capability exposure based service enhancement

A Telco network is no longer an “invisible” pipe to most applications in MEC, its network functions (NEF, AF, PCF, N3IWF etc.) will have to participate in MEC application end-to-end solutions.

An edge service discovery needs to include mobile network internal information in DNS lookup. In the past it only required pre-configured IP addresses from applications. As shown in Figure 2 below, in 5G MEC it is a function of device locations, edge node locations and instantiated edge services. Only the Telco network has location information for both devices and edge nodes. For example, AMF has the device location info and SMF knows where the UPF anchor is; through these edge node locations can be inferred. How to expose this functionality to developer facing APIs in a safe and reliable way is challenging.

Device mobility is another challenge. Ideally, mobile devices should connect to the closest available edge service. In order to adapt to device movement, applications need to know device locations, closest available edge node locations, and instantiated edge services. In all cases, the Telco network must be consulted, similar to the service discovery process, with different information inquiries.

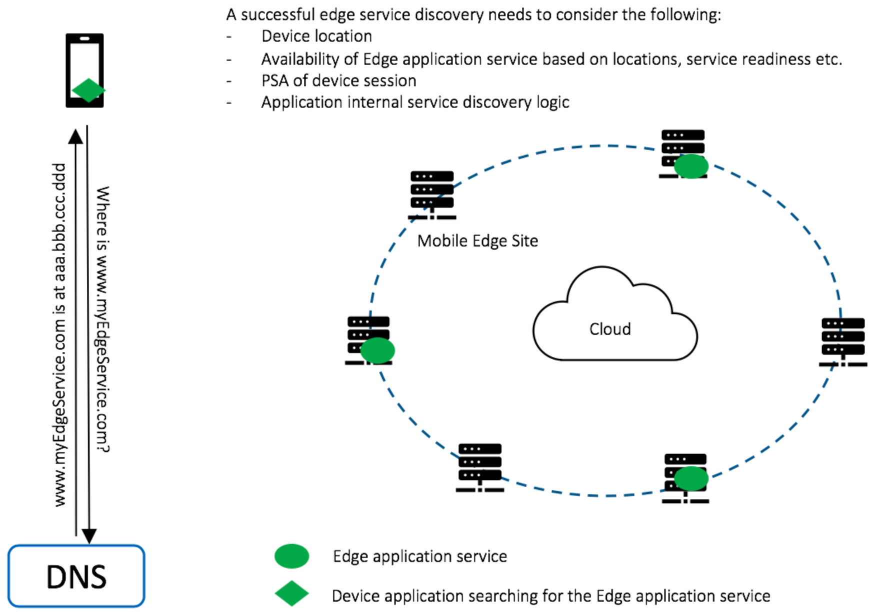


Figure 2 DNS lookup with mobile edge services

Traffic steering allows traffic to be routed to targeted MEC applications. Such routing policies and rules can be requested by applications and applied by Telco network functions (AF, PCF, and NEF). Routing actions are performed by 5GC User Plane Function (UPF). This is another case where network functions need to jointly work with applications in order to achieve optimal MEC objectives.

5GC's NEF (Network Exposure Function) allows telco networks to expose services and capabilities to applications. These include device monitoring and provisioning, QoS, and charging policies.

Leveraging these new capabilities requires telco network knowledge and information which end users and edge application developers may not have. In addition, operators will only allow a few trusted edge services to interact with their networks due to security considerations. This means end user application vendors will have to work through a middleman layer to leverage network capabilities.

3.2 Challenges in MEC platform(s) choices

As shown in Figure 3 below, MEC platforms may be deployed in different edge locations with very different requirements.

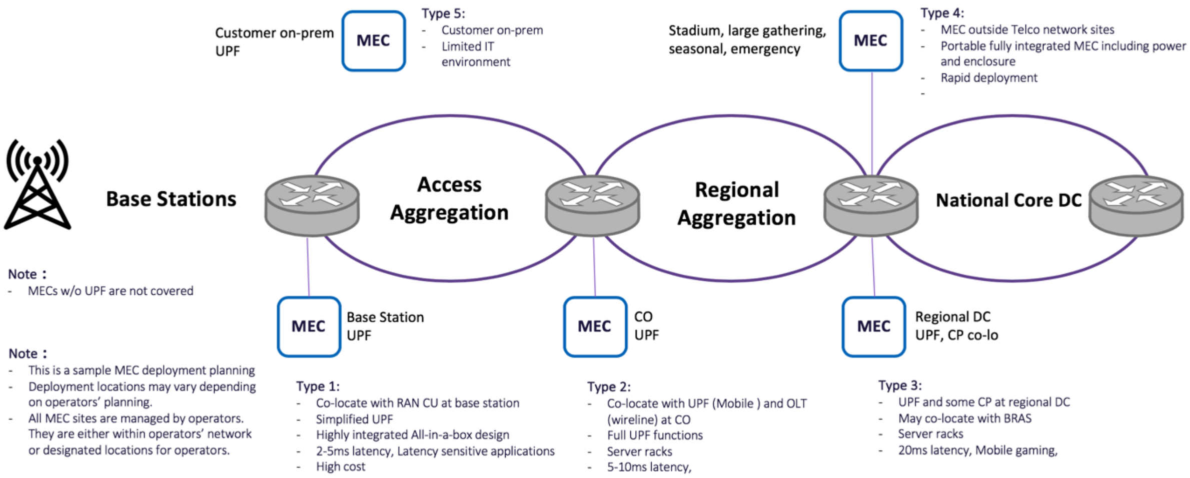


Figure 3 Telco Edge Deployment Locations

These locations have the following characteristics and requirements:

- Type 1: Access edge location: very limited in space and power
- Type 2: CO and other aggregation edge: restrictions in power and network wiring, limited in space
- Type 3: Regional DC edge: standard telco DC
- Type 4: Open space: requires pre-integrated all-in-one MEC solution
- Type 5: Customer edge: varies, usually limited IT environment.

This scope definitely adds complexity to hardware infrastructure. To accommodate underlying hardware varieties, one platform option is to provide unified, consistent APIs open to upper layer applications, allowing platform implementations to vary.

MEC platforms also need to address the MEC-unique requirements:

- Accelerated hardware support, open for future addition
- Multi-tenancy support
- Edge-cloud collaboration
- Inter-edge networking support

There are many opinions of what a MEC platform should look like. Traditional platform vendors believe a unified MEC platform for all edge applications would help reduce MEC management complexity. Vertical industry application vendors would prefer a customized platform to target specific end user applications. Hardware vendors are competing to promote performance, reduced power consumption, reliability, and accelerator support. Some hyperscalers have their own customized hardware platforms, some even with their own virtualization technologies. Clearly, MEC platforms mean different things to different people.

The Edge is where we can expect innovative, new applications to emerge, and there is always a concern that available platforms may not be able to support new application requirements. For instance, with the heated competition of new AI chip rollouts there could be applications that need AI chips not yet supported in existing platforms. Operators must balance their natural desire to host easy-to-manage and consistent platforms with the need to keep up with changing application requirements.

Open is the buzz word in MEC platform yet open doesn't mean easy. How to make open-ness work in practice is challenging. Typically, operators do not have big engineering teams for platform development, relying instead on the vendor ecosystem. Even operational maintenance depends heavily on platform vendors. New platform requirements usually arise from new applications with evolving platform needs, which require operators to integrate with existing platforms. Differences in platforms from various vendors bring extra operational maintenance work for operators. Successfully managing the diversity of vendors, new requirements, and reducing development cost is crucial for operators.

3.3 Management and orchestration challenges

In the previous chapter, we discussed the variety of MEC deployment locations. To manage and orchestrate among widely dispersed and functionally different MEC locations can be very challenging.

These orchestration level APIs are intended to provide unique portal and management interfaces to end customers.

When Telco and public clouds cooperate at the edge, integration happens on not only the function side, but also the management side. For end customers -- not only enterprise or vertical integrators -- what they want to see is a unique management interface, which means the customer can turn to the appropriate operating team as soon as possible when they need support.

Good orchestration can accelerate service onboarding, automate full life-cycle management, enhance customer experience, transform seamlessly from VNF to CNF, and simplify interoperation between telco and public clouds. Especially for MEC applications, customers need on- premise service to meet their low latency expectations. Low latency requirements are inherent not only on the function side, but also on the management side. Some enterprise customers even demand to have a self-controlled portal that integrates telco and public cloud orchestration functions.

For most telco operators, the generic VNF Manger and NFV orchestrator components are standards based, MANO compliant architecture. Some open source orchestration projects such as ONAP and OpenNESS aim for adoption by numerous operators. However, the MANO layer is still very specific to different telco operators and is integrated with northbound OSS/BSS systems. It is even related to organizational hierarchy and geography aspects and operator team technical backgrounds. It is still a long way for different Telcos to build unique orchestration architectures for public clouds and third parties to integrate.

Another challenge comes from the different architectures of telco CT and IT infrastructure. It remains difficult to manage VNF for telco core functions and container-based applications. As Telco core functions evolve to cloud native functions, based on differences of IT and CT and operator regulatory and uptime requirements, it will be challenging to use a single orchestration platform to manage both sides. Unifying telco operators' network and IT environments and connecting them to private enterprise clouds, edge clouds, and public clouds is ongoing work.

MEC promises to reduce latency and cost of customer service. Its most important advantages are related to less physical distance to customer locations. However, telco and public clouds have different hierarchies, which means orchestration distances are not uniform. Such differences in operating and management granularity may bring uneven customer experience. Unifying telco and public cloud's edge orchestration will be a big advantage of MEC and lead to end-to-end solutions for customers.

An enabler layer containing both telco and public cloud orchestration APIs is a solution that can potentially integrate management modules of both sides, providing unified, flexible, and rapid operating capabilities, enhancing customer experience.

3.4 Expectations from both sides and gaps

An Akraino blueprint project "Public Cloud Edge Interfacing" includes operator participants and cloud service provider participants. In the project charter, operators and cloud providers expressed motivations and expectations from each other. From the operator side:

- How to integrate the public cloud management interface and telco orchestration interface?
- How to open more telco abilities to public cloud and support DevOps?
- How to manage and monitor these different APIs in an efficient way?
- How to guarantee security and avoid DDOS or SQL injection attacks on the telco Core Network?

From the public cloud service provider side:

- How to best leverage network capabilities to provide value added services?
- Can the public cloud use the same APIs towards multiple telco network edge instances?
- How to integrate the public cloud management interface and telco orchestration interface?
- How to manage and monitor these different APIs in an efficient way?

Both sides have requirements on collaborated management and orchestration. An API layer to hide the differences among operator edges is critical for consistent, straightforward cloud interfacing.

4 Non-Technical Challenges

Telco Edge and cloud interfacing is where highly regulated, standards driven telco networks meet less regulated, self-hosted cloud service providers. This naturally creates business model challenges as well as operational challenges.

4.1 Business challenges

In addition to technical challenges facing traditional service providers as they move to open source software and cloud technology platforms, there are business, culture, and legacy challenges as well.

Traditional service providers operate under scrutiny of legacy telecommunication regulations (with origins dating back over 100 years). They must deal with lifeline service availability expectations, labor contracts based on prior generation roles, and the realization that they may be responsible for integration and delivery of individually acquired software and hardware solutions. Cloud providers started and continue to operate under far less government oversight and legacy obligations.

Equally challenging, service providers and cloud providers may be concerned that they must collaborate with their competitors, enhancing the knowledge/skills of a competitor, or creating potential new ones.

4.1.1 Regulatory concerns on CT and IT differences

Service providers often operate under oversight of national governments, local governments, consumer protection agencies, anti-trust agencies, communication commissions, etc. For example, many service providers are expected to enable lifeline calls (911 calls in the US) regardless of the customer's account status. In the event of a lifeline service outage, national governments are often formally notified in legally binding reports. Cloud operators do not normally operate under these conditions.

4.1.2 Taxation

Service providers are often required to collect taxes and fees based on customer location. Sometimes this means operators pay different taxes and fees based on the cell site servicing the customer's mobile connection, and potentially passing these additional costs to customers.

4.1.3 SLAs on CT and IT (KPIs)

Key Performance Indicators (KPIs) used to measure quality of service provider networks include availability, accessibility, coverage, service continuity while moving, reliability, quality/error rate, and signal strength. Service providers must agree to these measurements in Service Level Agreements.

Cloud providers generally are free from these KPIs.

4.2 Operational challenges

Traditional service providers expect suppliers providing vertically integrated solutions to “over engineer” capacity, availability, serviceability, stateful error recovery and manageability in their products. The premise is to maximize capacity, reduce the likelihood of service-impacting defects through extensive testing, rely on internal redundancies to reduce the impact to services when a defect is encountered and have enough internal capacity to sustain full-scale operation until the defective component can be repaired/replaced. This concept is sometimes referred to as “5 9s” or “carrier grade” (5 9s refers to 99.999 % uptime).

Cloud designs, based on “IT” (Information Technology), operate under different constraints. Cloud platforms often depend on minimizing the impacts of outages by distributing many instances of software on many different virtual and physical platforms.

4.2.1 Organization within Telco responsibilities

Traditionally Telco technology has relied on protocols, standards and methods unique to their industry. As the industry has pushed towards Software Defined Networking (SDN), it has moved towards IT solutions. Open Source operating systems, open source software, RESTful Web Services, Message Bus, HTTP/S, etc. are now part of modern 5G system designs.

Future telco standards are incorporating IT ecosystem concepts, and organizations are retooling and rearranging to accommodate this shift.

Cloud providers grew up with IT. To meet MEC expectations, they may need to make adjustments as telco workloads move to clouds. Their fundamental building blocks have not changed, but service level expectations will.

4.2.2 Labor/Union

Traditional service providers in the US maintain a large, highly skilled, organized labor workforce. Organized labor and Telco continue to be partners, and this partnership will be needed for their combined successes. Telco workers may need to update their training. Equally important, agreements between services providers and their labor unions will need continuous updates to optimize for cloud technology as a cornerstone for the future.

There may be shifts in focus areas as telco workloads move to clouds. The fundamental building block used by the cloud provider have not changed.

5 An Enabler to Bridge the Telco Network and Edge Applications

Since the advent of 5G there has been vigorous debate on how to interface edge applications and Telco core networks. It is a difficult question, as two very different beasts – telco operators and webscalers – must interoperate smoothly together. And they must do this over a wide range of interfaces, including physical hardware, functional software in various forms, data flow, and APIs. Depending on whether your perspective is an operator or a webscaler, things can look quite different. Our solution is to define an enabler layer as two (2) flexible sublayers: edge enabler and application enabler.

5.1 Enabler functional description

Interfacing applications with Telco network is not as simple as it appears to be. One might think that 3GPP standards define all necessary specifications and all an implementation needs to do is carefully comply to provide a straightforward interface. Unfortunately, this is not the case.

There are no identical telco networks. Each operator has its own network design and deployment considerations. System integration is a lengthy process full of details not specified in 3GPP standards. It's a complex domain in which not all edge service vendors may want to invest. In addition, telco networks must evolve with new 3GPP releases. Adapting and updating products to evolving standards takes substantial work.

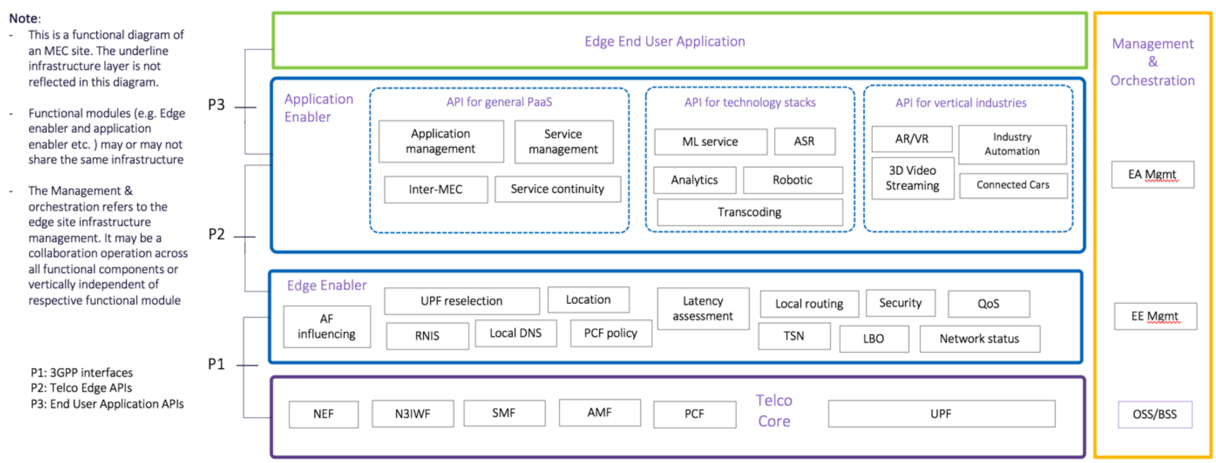


Figure 4 **Edge Stack Functional Layers**

Figure 4 shows edge and application enabler layers to abstract telco network interfacing details while providing developer friendly APIs. These APIs will on one hand allow access to 3GPP standards-defined functionality, including all customization, and on the other hand provide IT flavored APIs for easy consumption. Of equal importance, this architecture forms a flexible framework to deal with ownership and operational challenges noted above and in section 4, Operational Challenges.

5.1.1 Edge enabler

An edge enabler faces the Telco network. Its mission is to allow operators to expose network information per 3GPP standard to support edge applications, such as AF traffic influencing, UPF reselection, QoS etc. It provides industry standard APIs which are consumed by an application enabler. An edge enabler can be viewed as an "API gateway" into the core telco network.

APIs exposed by an edge enabler require fundamental telco network knowledge to understand and consume correctly. For example, typical RNIS (Radio Network Information Specification) APIs may specify a data model containing S1 bearer information. It’s reasonable to expect mobile gateway developers to have this level of telco network knowledge, but not mobile device application developers.

Abstracting 3GPP interaction yields significant benefits:

1) Provides a bridge between telco network and edge applications, hiding telco network function level interface complexity
2) Allows telco operator expose network capabilities to edge service developers
3) Allows easy upgrade for future 3GPP standard evolution
4) Provides a buffer zone to the Telco core network for better security control
5) Allows operator to better control service differentiation

It’s expected that an edge enabler works is owned and maintained by operators. It can function as a customizable non-3GPP network function for operators to offer various new network enhancement.

One Tier 1 operator A in Asia (an Akraino member) recently proposed specifications of a set of basic user plane functions and add-on functions. These functions include guaranteed QoS, direct forwarding, delay assessment, etc., and would be provided to upper layers via REST or gRPC APIs.

The Akraino ICN blueprint project also adopts a platform which provides similar functions, such as AF APIs in RESTful API format. These APIs provide traffic steering and packet flow management to upper layer functions.

Another Tier 1 operator B has architected a CT-VAS (Value Added Service) layer as an edge enabler. Its capability mainly comes from the interface of 5GC NEF (5G Core Network Exposure Function), including UE LBS (User Equipment Location Based Service), RNIS (Radio Network Information Service), TCP acceleration, etc. Together with its vendors, the operator is working on opening network capabilities such as RNIS, LBS and wireless network bandwidth management on MEC. One example is obtaining the network information from the edge network elements (UPF/DP, User Plane Function/Data Plane) through the MP2 interface. The operator also plans to offer a 5GC edge enabler sandbox for application developers to test functionality and performance.

It should be noted that traffic steering decisions always come from 5GC, not the edge enabler. The latter can request traffic steering via AF or NEF, but not make actual decisions. Please refer to 3GPP specifications for details.

5.1.2 Application enabler

An application enabler sits in between the edge enabler and edge application developers. . Its mission is to provide edge application developer friendly APIs, allowing edge application developers to consume and manage application-specific telco network capabilities without extensive Telco network knowledge.

It’s expected that an application enabler is owned and maintained by public/private cloud providers. It supports edge applications and services, life cycle management on edge nodes, connects edge nodes to cloud data centers, and allows edge applications to run temporarily disconnected from the cloud.

An application enabler may include three (3) categories:

1) General PaaS layer APIs are application management, service management, including:

   - Resource Management
   - Application service management: registration etc.
   - Monitor, Reporting and Notification
   - Authorization, Certificates, Authentication
   - Package Manager

2) Technology functional stacks, such as IoT, ML, analytics etc. Telco edge application stack is one of these. Functional stacks may include:

   - Message Bus/Broker
   - Event Bus
   - Device Management
   - Data analytics service
   - ML Inference or Learning service

3) Vertical domain edge stacks, for example:

   - Gaming
   - AR/VR
   - Video streaming
   - Connected cars

There could be multiple vendors providing one or a mix of the above functions, on one or more platforms, which may be different.

An example of an application enabler is KubeEdge Edge Service (a CNCF open source project and Akraino Blueprint family), which provides well-known cloud native Kubernetes APIs for edge application and service management. KubeEdge also provides IoT support for edge cloud computing, such as device management, event bus, IoT protocol support, etc. The Akraino KubeEdge Edge Service Blueprint project is a blueprint family that showcases end-to-end solution for edge services. It demonstrates solutions such as ML inference offloading, edge data analytics etc.

A Tier 1 operator B(an Akraino member) has included in its MEC architecture an IT-VAS (Value Added Service) layer that maps well as an application enabler. IT-VAS provides an enhanced PaaS platform with APIs, SDKs and DevOps tools to support VM, Container, and microservice deployment and run-time environment. For Telco edge specific feature support, IT-VAS provides APIs for provide firewall rules management, end user authentication, White listing / blacklisting IPs, bandwidth management etc.

For domain specific edge stacks, IT-VAS will include IT common capabilities, such as AI industrial machine vision, AI picture recognition, video transcoding, video stitching, VR rendering, etc. An API gateway supports integration of third-party applications in operator B's edge application ecosystem, such as a third-party local agent to achieve collaboration between Cloud and edge.

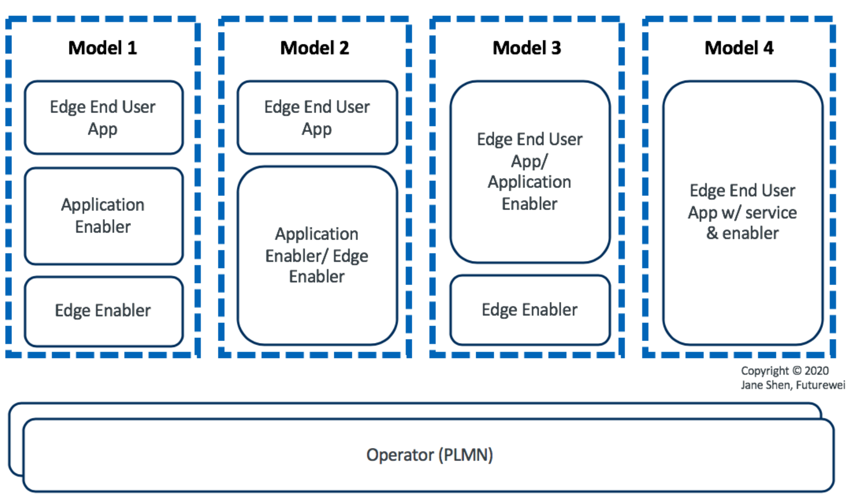
The Akraino Android Cloud Native (ACN) project provides a native Android environment for edge applications, targeting applications developed for Android (e.g. mobile gaming apps). It is part of the IEC (Integrated Edge Cloud) blueprint family. The ACN project includes an Arm based edge hardware platform to enable applications to distribute computing smoothly between mobile devices and edge nodes. The platform provides a highly integrated Android running environment, including a containerized Android run-time framework, Android virtualization layer with Anbox, vGPU and other drivers, all deployed on an Arm server. The platform enables cost and resource efficient remote execution of Android apps on the edge cloud. Existing Android apps can run as-is (unmodified), enabling remote access for +1,000,000 available Android apps. The well-integrated Android cloud run-time environment allows straightforward deployment and high scalability, maintaining cloud native features on the Arm platform.

One Akraino member participating in the KubeEdge BP provides an example of category 3): an edge stack for multi-endpoint conversation ASR (Automated Speech Recognition) diarization and ASID (Automated Speaker Identification). In the mobile environment, accuracy improvements are required to make multi-endpoint diarization and speaker identification applications viable. While a single talker ASR system may still function acceptably, the problem of multi-party diarization, already considered "hard" in the state-of-the-art, is made even more difficult. An edge approach can increase accuracy of diarization and identification by adapting processing per endpoint, controlling speech preprocessing and number / type of ASx inference models in response to varying conditions. In addition to increased accuracy, edge ASx processing also results in reduced bandwidth (by sending text results instead of encoded raw audio streams to central data centers), and increased privacy, removing emotional audio content and incidental background conversation prior to cloud storage.

Kontour, an Akraino project, is working on Edge site KPIs (Key Performance Indicators). With increasing edge adoption, identifying and selecting proper edge sites becomes crucial. This has created a need to identify, define, standardize and publish edge site KPIs, for example Network Throughput, Network Latency, Storage IOPS, etc. These KPIs will be consumed by edge platforms or applications, providing performance snapshots of compute, network, and storage for different edge sites. For example, a video streaming application requiring ultra-low network latency (5-10 milliseconds) can identify and select one or more edge sites offering the required latency.

5.1.3 Edge function layer ownership and operational models

Depending on individual telco operator MEC strategies, edge stack ownership and business model will vary. Figure 5 below shows four (4) models of stack ownership and operational responsibilities. Each



rounded box represents one single ownership and associated operational responsibility. An operator may adopt a mix of models in order to achieve its business goal.

Figure 5 Edge layer ownership and operational models

Model 1

Model 1 follows a clear, layered ownership and operational responsibility matching our edge layer diagram in Figure 5 above. Here typically the edge enabler is owned by a telco operator, application enablers are owned by one or a few edge service providers, and end user edge applications are owned by application vendors.

APIs between layers will help hide implementation details. A consistent, versioned API definition set can increase API adoption rate.

An example of this ownership model can be like this:

- ✓ in a smart city scenario, an operator will provide an edge enabler
- ✓ a smart city platform vendor provides one or more application enablers supporting various smart city applications; e.g. smart meter, intelligent surveillance, smart traffic management
- ✓ end user smart city application vendors deploy their respective applications on the smart city platform

Model 2

In Model 2, an edge and application enabler combo can be provided by operators or trusted edge service providers, e.g. hyperscalers or neutral hosts. For example, Operator B is offering a solution similar in concept to model 2. Its software stack includes IaaS, PaaS and edge management. It also provides a client facing service portal for easy DevOps. The solution may have various hardware form factors.

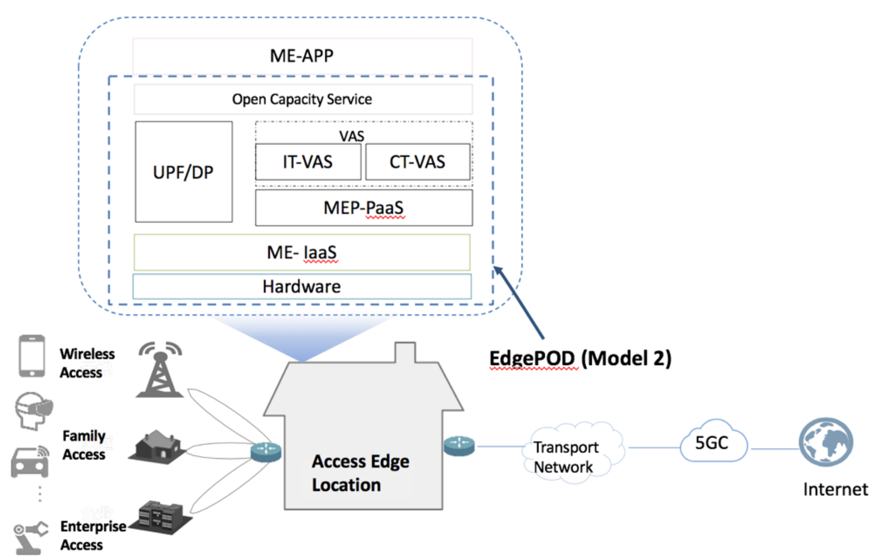


Figure 6 A model 2 solution scenario

For hyperscalers this enabler combo can include edge extensions with unified cloud/edge application development/deployment platform. In this model, edge service providers include an edge enabler. This means they have the expertise to directly access the network, and also an agreement with telco operators to do so.

An example of this model is a smart factory. An operator deploys a vertically integrated edge and application enabler with application management support, containing factory specific AI based maintenance (predictive analytics) and AR remote assistant service. Factory applications can be built on the enabler platform and perform factory equipment/environment anomaly detection, robotic production line inspection.

Model 3

In model 3, the edge enabler is owned and operated by the operator. Application vendors have their own vertical stacks to directly interface with Edge Enabler northbound APIs. This saves the application the hassle of keeping track of operator network changes. The operator gains better separation between its network and application layers, which improves network security. In addition, the edge enabler gives the operator a buffer zone in which to provide additional and enhanced services.

In model 3, typical application entities are large 2C service providers, e.g. video streaming service providers. Usually they have developed and optimized the application platform for their applications. The only piece missing when they move to the telco edge is an edge enabler -- exactly what a telco operator can offer. By providing consistent edge enabler APIs to 2C service providers, operators can open new revenue streams.

Model 4

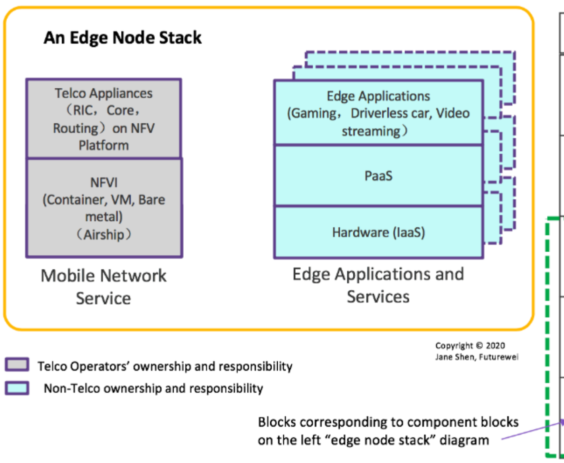
In model 4, the edge enabler, application enabler and application are all owned and operated by one entity, which can be an operator or a major application vendor. In the case of an operator, they may provide a vertically integrated service such as a port-management edge service. As an example of a major application vendor, a global gaming vendor may reach agreement with an operator to directly access the operator's network at the 3GPP interface level and operate edge services on its own.

5.2 Anatomy of edge stacks

We have described edge layer ownership and operational responsibilities, with underlying implementations decided by layer owners. Operators or other entities may hope to architect a universal platform for all edge enabler, application enabler and edge applications. This is a natural tendency to reduce development cost and make product management easier. But in reality, a one-size-fits-all platform will have challenges in operational responsibilities. An operator's edge business strategy may require multiple models described in section 5.1.3 to be adopted. Hence a flexible underlying platform implementation to support functional modules would be the proper approach.

Let's take a look at a telco mobile edge site system anatomy from an edge stack point of view. In Figure 7 below, each edge node has 2 resource groups: one for telco mobile network functions (box labeled Mobile Network Services), and one for edge computing (Edge Applications and Services). Edge computing is the newcomer in a typical mobile edge site. Most likely, it is an additional rack of servers.

Traffic outputs from mobile network functions flow directly into edge computing servers at the IP routing level. From the traffic content point of view there is no difference from when EC servers are placed miles away. Often times that traffic is only partial of the total traffic M is processing. Techniques like a local break out can be applied to do the split. There are also other ways such as a hardwire split. Telco functions usually include but are not limited to user plane processing functions such as S/PGW-U for 4G or UPF for 5G, as opposed to Edge computing servers, which usually terminate traffic for processing. This is depicted in the left “Edge Node Stack” diagram.



Edge Stack Strategies			Typical Deployments
Option 1		Telco MEP based (ETSI MEC)?	Operator centric, UPF+, Mission Critical edge
Option 2		Share common HW only	Some operators, xxx?
Option 3		Independent vertical stacks; IP interconnect	Public Cloud Interfacing at Telco CO
Option 4		Telco Appliances and Edge applications share common Edge platform	Private Network, LADN
Option 5		Non-telco with spectrum has its own RAN and Core services	CBRS, Industrial Spectrum

**Figure 7		Edge stack options**

In most current implementations, mobile network services and edge computing are adopting two different infrastructure and platform technologies: (i) NFV architecture and (ii) An IT flavored architecture typically layered as IaaS, PaaS, and SaaS/Application stacks. Both mobile network services and edge computing have their own infrastructure and system level management. Mobile edge management is part of the overall mobile network OSS. In the public cloud case, edge computing extension management is part of global cloud management.

This diagram is a simplified view on player groups in Telco 5G Edge. It is meant to highlight the differences between two major player groups (Telco and IT). Each group has its own ecosystem which plays various roles in Telco 5G Edge. The Edge is where these two groups meet and collaborate. Collaboration interfaces may vary across operators and solutions.

At first look, it seems natural that mobile network services and edge computing reside in physically separated racks managed by separate telco and public cloud teams. This avoids concerns related to regulatory requirements, Service Level Agreements (SLAs), etc. Essentially it is a co-location arrangement, as depicted in the 3rd option in the right-hand column of the table.

There are certainly other options in that column. Different colors represent different ownership and operation responsibility. In option 1, the Telco not only provides mobile connectivity, but also an edge computing platform for applications.

Option 2 is one step further towards convergence between telco and public clouds, with a common shared infrastructure layer. Sharing can include only the hardware layer or also lower platform layers such as VMs and containers. The main benefit of option 2 is a unified infrastructure layer extending from the telco core network to edge nodes. As mentioned earlier, infrastructure that meets typical edge application requirements might look quite different from the NFV VIM layer. This implies Telcos either have moved, or will move, to a cloud native (non-NFV) telco edge or a layer above NFV VIM to create a suitable environment for typical edge applications (the former would likely be the case). Operators with in-house infrastructure expertise might be interested in this option. They most likely are in the process of building a cloud native telco network and extending the infrastructure layer to the edge seems logical.

Options 4 and 5 probably will not happen in telco edge data centers. These are the scenarios where Telco RAN or Core user planes are considered more or less to be access options. A typical case would be private enterprise networks, which usually require various access methods. Option 4 deploys telco edge core functions in enterprise premises. Option 5 only has shared Telco RAN as an option; not all operators are ready to take responsibility for a non-telco site. Although some operators have announced they are working on an option 4 solution, most likely option 5 will be more widely adopted. Both options have unified IaaS and PaaS layers, with Telco RAN/Core functions deployed as special applications. There can be SLA differences between telco appliances vs. typical edge applications. In an enterprise private network environment, it is manageable.

A global public cloud service provider recently published its edge zone types used in Telco DC and enterprise on-premise.Type 1 shown in figure 8 is a good example of option 3. Its type 2 may go option 5 or option 4.

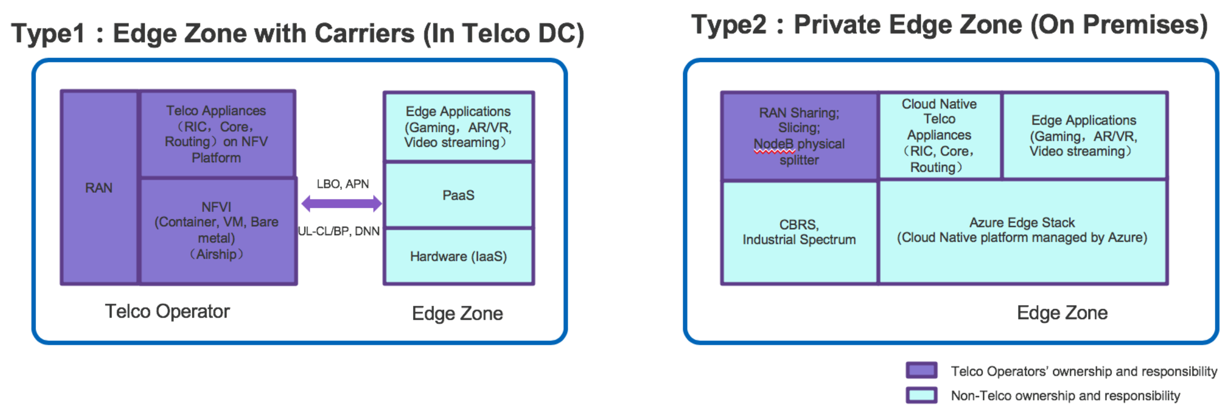


Figure 8 A public cloud service provider Edge Zone Types

5.3 Akraino Public Cloud Edge Interfacing (PCEI) blueprint project

The purpose of the Public Cloud Edge Interface (PCEI) Blueprint family is to specify a set of open APIs for Telco Edge Blueprints to expose Public Cloud Service Provider instances at the edge. As Public Cloud Service Providers deploy Edge instances to better serve their end users and applications, Telco Edge deployments offer many opportunities for collaboration by exposing their network capabilities to provide value added services.

The need to interface and exchange information through these open APIs will allow competitive offerings for consumers, enterprises and vertical industry end-user segments. For instance, open APIs will be provided between Telcos and public cloud edge compute platforms such as Google Cloud Platform (GCP) Anthos, AliCloud Edge Node Service (ENS), AWS Wavelength, Microsoft Azure Edge Zones, and Tencent ECM, to name just a few. In addition to providing basic connectivity services, open APIs will deliver predictable data rate, predictable latency, reliability, service insertion, security, AI and

RAN analytics, network slicing and more. These capabilities are needed to support a multitude of emerging applications such as AR/VR, Industrial IoT, autonomous vehicles, drones, Industry 4.0 initiatives, Smart Cities, and Smart Ports. Other open APIs will expose edge orchestration and management, Edge monitoring (KPIs), and more. These open APIs will form a foundation for service and instrumentation when integrating public cloud development environments. Even though open APIs will be found across all Telco operators, they will differentiate based on services they provide.

The PCEI blueprint family addresses all aspects of API interoperability including API definition and API gateway functions (AAA, policy, security), so as to offer a secure, controllable, traceable, scalable and measurable access for public edge cloud service providers.

6 Future work

So far, we have addressed the functional enabler layer within one edge site of a telco operator. There are hundreds of MEC sites interconnected within one operator. These MEC sites may interface multiple public cloud edge extensions within or in between. E.g. AT&T works with Azure, Google Cloud etc. Moreover, the end user application needs to reach end users subscribed to different operators. The inter-MEC support including networking, collaborated management and orchestration remains to be addressed. We will look into these areas in future whitepaper releases.

List of Acronyms

2C	To Consumer
AF	Application Function
AMF	Access and Mobility Management Function
API	Application Programming Interface
AR	Artificial Reality
CT	Communications Technology
CU	Centralized Unit (of RAN)
DC	Data Center
ETSI	European Telecommunications Standards Institute
IaaS	Infrastructure as a Service
IT	Information Technology
MANO	Management and Orchestration
MEC	Multi-access Edge Computing
MEP	MEC Platform
MNO	Mobile Network Operator
NEF	Network Exposure Function
NFV	Network Functions Virtualization
PaaS	Platform as a Service
PCF	Policy Control Function
QoS	Quality of Service
RAN	Radio Access Network
RNIS	Radio Network Information Service
SaaS	Service as a Service

SBA	Service Based Architecture
SMF	Session Management Function
UE	User Equipment
UPF	User Plane Function
V2X	Vehicle to Everything
VNF	Virtualized Network Function
VR	Virtual Reality

References

[1] 3GPP TR23.758 V17.0.0 "Study on application architecture for enabling Edge Applications"

[2] 3GPP TR 23.748 "Study on enhancement of support for Edge Computing in 5G Core network (5GC)"

[3] 3GPP TS 23.501 "System architecture for the 5G System (5GS)"

[4] 3GPP TS 29.522 V16.3.0 "5G Systems, Network Exposure Function Northbound APIs"

[5] ETSI MEC Whitepaper "MEC in 5G networks"

[6] Akraino Edge Stack Integration Projects (Blueprints): https://wiki.akraino.org/pages/viewpage.action?pageId=1147243

[7] China Mobile Whitepaper "5G OpenUPF", 2020

[8] China Unicom Whitepaper "5G MEC edge cloud platform architecture and business practice", 2020

[9] Azure Edge Zone preview: https://docs.microsoft.com/en-us/azure/networking/edge-zones-overview

Authors

Jane Shen
Jane.shen@futurewei.com

Dan Chen
China Unicom

Doug Eng
dougeng@comcast.net

Jeff Brower
Signalogic Inc.

Su Gu
China Mobile (USA) Technologies Inc.

Contributors

Tim Epkes
Tim.Epkes@Dell.com

Ike Allison
ike@alicon.se

Neal Oliver
neal.oliver@intel.com

Yin Ding
Yin.ding@futurewei.com

Dan Druta
dd5826@att.com

Arif Khan
arif@voereir.com

Tina Tsou
Tina.Tsou@arm.com

Rong Huang
huangr27@chinaunicom.cn

Acknowledgement

We would like to thank the following people who helped in paper drafting: Gao Cheng, Qian Huang, Jiafeng Zhu, Manik Sidana, Trevor Tao.
Thank all reviewers for valuable comments: Sukhdev Kapur, Andrew Wilkinson, Vikram Siwach, Oleg Berzin, Pasi Vaananen, Rob Franzo and many more who provide various viewpoints which make the paper clearer and reflect an unbiased view.
During the paper drafting phase, we have had numerous discussions with experts of 5G and edge computing in general. Special thanks to Alex Reznik, Sami Kekki, Liang Geng, Hanyu Ding