Power at the edge

Processing and storage move from the central core to the network edge

Monica Paolini, Senza Fili

The report is available thanks to our sponsors:

InterDigital, Artesyn Embedded Technologies, Advantech, Intel, Qwilt, ADLINK Technology Inc., Vasona Networks

In collaboration with

RCR Wireless
Intelligence on all things wireless

Table of contents

Watch the video of the interviews

I. Report: Power at the edge. MEC, edge computing, and the prominence of location

What: Basics

Location, location, location. Multiple-access Edge Computing (MEC) and edge computing in general are gaining acceptance in both fixed and mobile networks as we increasingly realize the power of location in wireless networks – and especially in virtualized networks. This does not mean the centralized cloud or the big data centers hosting the network core will go away anytime soon. But a rebalancing act is definitely due.

In recent years, there has been a strong push to move everything to a centralized cloud, enabled by virtualization and driven by the need to cut costs, reduce the time to market for new services, and increase flexibility. In the process, we lost sight of how important the location of functionality is to performance, efficient use of network resources and subscriber experience. Physical distance inevitably increases latency.

Central processing and storage limit the ability to optimize RAN utilization. A fully centralized network may be easier and cheaper to run, but it does not always keep subscribers happy.

MEC and other edge computing initiatives address the need to place processing and storage where appropriate, whether a central location or the network's edge, depending on factors such as applications, traffic type, network conditions, subscriber profile, and operator's preference.

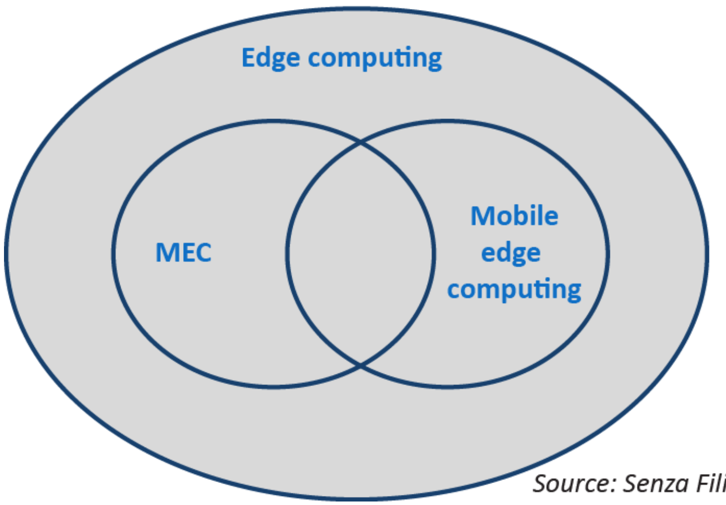
Virtualization, initially used as the basis for moving to the centralized cloud, is even more foundational in enabling hybrid models, because it gives service providers the flexibility to choose location, hardware and software independently to optimize end-to-end network performance and QoE. Both operators and vendors agree that we need to keep a healthy balance between what remains centralized and what gets distributed to the edge. The same applies to the RAN: in some places a centralized – C-RAN or vRAN – approach makes sense; in others the traditional distributed model works just fine.

In this perspective, virtualization, MEC and 5G, in different but complementary ways, free both fixed and mobile networks from the constraints of a centralized architecture and topology. The new networks can adapt to and accommodate new applications and functions, and can optimize their

performance. They replace legacy networks in which applications and functions had to painstakingly overlay an existing, rigid architecture.

In this report, we explore the evolution of the edge's role in fixed and mobile networks and how it may impact network optimization, value-chain roles and relationships, business models, usage models and, ultimately, the subscriber experience.

Edge computing



Source: Senza Fili

Terminology: MEC or edge computing?

In this report, we use the term "edge computing" to refer to the processing, storage and network optimization at the edge of both fixed and mobile networks, that is independent (and agnostic) of the access technology. Specific implementations of edge computing can, however, play a role in optimizing the utilization of access resources. MEC is an example of that.

Mobile edge computing refers specifically to mobile networks. Because networks increasingly include fixed and mobile components and they integrate them more tightly than in the past, the distinction between edge computing and mobile edge computing is narrowing. It may disappear altogether with the convergence of fixed and mobile networks.

MEC is a specific approach to edge computing that is primarily intended for mobile operators, or, more generally, service providers that have a core network on which MEC can be overlaid.

"Edge" is the term that resists a simple definition. In the context of edge computing, the edge could be in the RAN or the customer's premises, or it could be an aggregation point in a more centralized location. As networks evolve and become virtualized, the opposition of central core and edge is likely to disappear, to be replaced by multiple locations that may be appropriate to host a given application or function.

Do we still need edge computing in virtualized 5G networks?

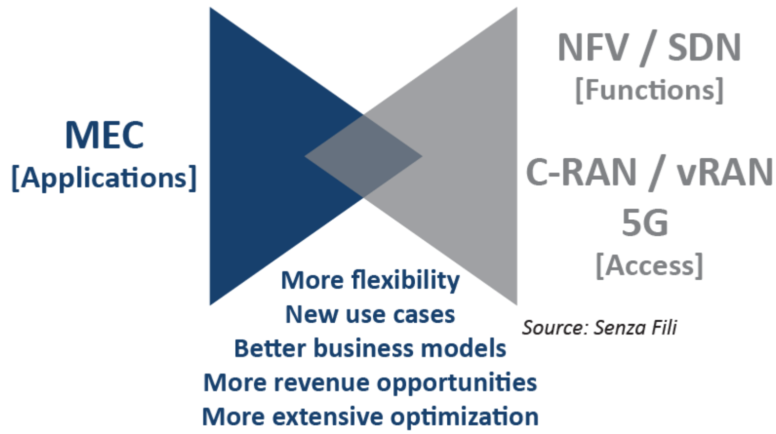
How does edge computing – and specifically MEC – relate to the other main technological innovations in wireless, namely virtualization and 5G? Strictly speaking, MEC does not depend on them, and MEC deployments can start ahead of 5G or full virtualization. However, they share a common direction toward networks that are more flexible and less homogeneous, in turn dictated by the need to increase network efficiency, capacity and cost.

As they all move toward the same goals in parallel, they enable and reinforce each other, because they work within different domains in the end-to-end mobile network. 5G, C-RAN and vRAN improve performance mostly through an evolution of the RAN – e.g., through spectrum utilization, wireless interface, architecture. NFV and SDN work at the function level to optimize processing within the core. MEC is more narrowly focused to enable operators to manage applications and end-to-end traffic at the application level.

MEC, 5G and virtualization are not alternative solutions among which operators will choose. We will still need edge computing when 5G arrives. Edge computing lowers latency ahead of 5G, but when 5G arrives, it will need edge computing to lower latency further and meet the 5G requirements.

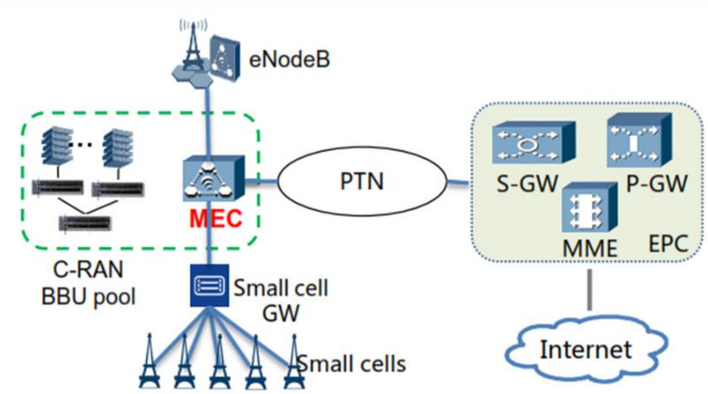
Similarly, RAN virtualization facilitates the rollout of edge computing, because the centralized BBU location where the baseband processing is concentrated provides a good integration point where some edge computing functionality can be located. In a small-cell C-RAN deployment in a retail center, for instance, the BBU location can also host the MEC server that manages location-based applications, which may be available to visitors both over the cellular network(s) and Wi-Fi networks. By combining C-RAN or vRAN deployments with MEC deployments, operators can also improve the business case, because the incremental cost of adding MEC to a new deployment is substantially lower than that of rolling out MEC over the existing infrastructure, as Mansoor Hanif at BT notes in the interview in this report.

Mutual enablement and reinforcement, independent deployment



Source: Senza Fili

MEC server co-located with the C-RAN BBU pool



Source: China Mobile

What is MEC? No longer mobile edge computing!

Among the multiple initiatives to enable or facilitate edge computing and, more specifically, mobile edge computing, ETSI MEC occupies a central role, because it provides a framework to shift processing, storage and control to the edge that is integrated within existing fixed and mobile networks. MEC was created to address mobile operators' need to move the processing and storage of some services and applications to the edge, and to optimize mobile network performance and resource utilization in real time.

MEC standardization work started in 2014 at ETSI, with a seminal white paper authored by Huawei, IBM, Intel, Nokia, NTT DOCOMO, and Vodafone. The list of active participants has since grown to include more vendors and operators. The focus initially was on mobile networks, but it now access-technology agnostic and it has extended to fixed networks, to reflect the tighter integration of mobile and fixed networks, which often host the same services and applications.

The MEC acronym no longer refers to "Mobile Edge Computing" and instead stands for "Multiple-access Edge Computing." Although the change was prompted by ETSI's internal process requirements, it was welcome change as it expands the reach and potential of MEC in today's rapidly evolving wireless networks, which include a more varied set of access technologies and spectrum regulatory frameworks.

The extension to non-cellular technologies means that Wi-Fi is now included within MEC's scope. This is a welcome addition that reflects the fact Wi-Fi in most markets accounts for the majority of traffic to mobile devices.

The report will delve into these, but three points are worth emphasizing:

- The main driver of MEC adoption is QoE. With MEC, operators move hardware that traditionally had been located in centralized data centers toward the edge and, more importantly, closer to the users. The two primary advantages of this are a reduction in latency and more efficient utilization of the available capacity – both of which improve QoE.

Features

- On premises
- Low latency
- Proximity
- Location awareness
- Network context information

Approach

- Virtualized network
- Standard-based
- Open source
- APIs
- Multivendor
- Modular
- Convergence of multiple initiatives

Goals

- Better QoE
- Lower latency
- RAN and content optimization
- More efficient use of resources
- Central cloud offload
- New services

- Moving content to the edge can have multiple advantages – e.g., lower latency, location-aware services, flexible service creation, security, and reduction in backhaul traffic – but only if the right content is stored locally and the edge location is appropriately chosen. If these conditions are not met, MEC may increase costs and complexity, at a price that is too high to justify the enhancement in QoE.

- MEC emerged within a growing ecosystem that is converging toward a new approach to network design and operations, which we refer to as "pervasive networks" [25], as opposed to the legacy "atomic networks." In this new environment, networks are virtualized, use open source software, and rely on APIs for application development. Technological advances are driven by multiple open initiatives, operators and vendors, reducing the impact of proprietary solutions and even of established standards bodies such as 3GPP. From the hardware side, MEC and edge computing introduce the need for modular solutions that support multiple form factors to enable deployments in a more varied set of environments.

Why: Drivers

Edge computing's appeal comes from a growing realization that centralized topologies are not sufficient to serve current and forecasted traffic loads with the QoE that both operators and subscribers expect.

The fundamental driver for edge computing is the continued growth in subscriber traffic and, especially, in real-time traffic, such as video and interactive applications such as games. In the future, augmented/virtual reality traffic will add pressure on operators to continue to increase capacity and, just as importantly, reduce latency. High latency – even in a high-capacity network – will cripple QoE. For subscribers, limited capacity and high latency may have the same effect and look indistinguishable. For an operator, adding capacity without lowering latency can be an expensive mistake.

Even with 5G's promised reduction in latency, edge computing will be useful in reducing the latency introduced by the backhaul. While backhaul technologies add different levels of latency, they all inevitably contribute to it as a function of physical distance.

The concurrent growth in traffic load – increasing capacity requirements – and of real-time traffic – increasing latency requirements – drive the need for processing, storage, and control for selected applications to be moved to the edge.

At the same time, operators face a challenging situation because they have to meet high QoE expectations in a cost-effective way, in an environment where subscriber revenues are flat in most markets. This means that they need to improve performance within the current spending levels. To do this, they have to maximize the utilization of existing resources before they launch into a RAN expansion. MEC and edge computing can play a major role in doing this, because local control and processing enable a finer-grain real-time optimization of RAN transmission.

Growth in video traffic increases importance of latency

Video traffic keeps on growing, and at a faster pace than other types of traffic. According to Cisco's VNI, mobile video accounts for 60% of traffic today, and that number will likely be 78% by 2021. Ericsson's estimates are in the same range: 50% in 2016, growing to 75% in 2022. The 2016–2022 CAGR for video is 50%, compared to 23% for web browsing and 39% for social networking.

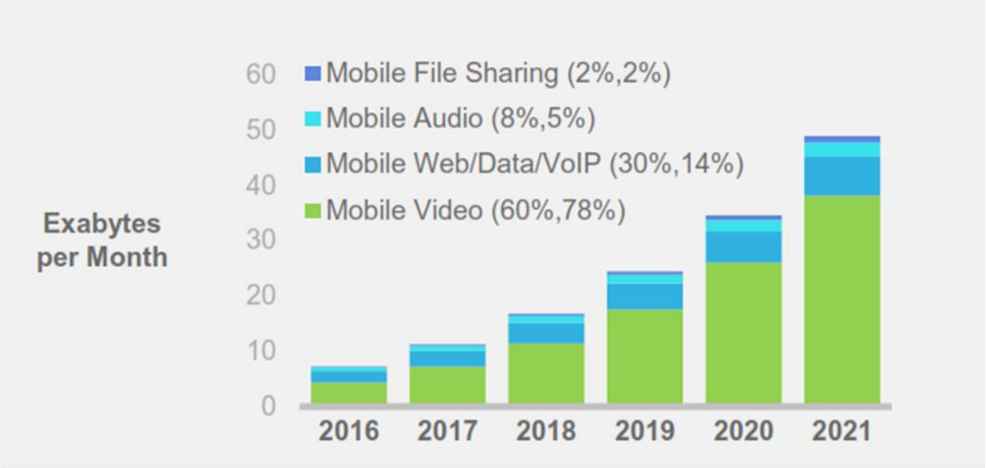
The growth in video traffic is hardly surprising, given the availability of more video content and its higher-quality. Also, video is no longer confined to video apps: it has become an integral component of social networking and communication apps. In the process, real-time video calls have finally gained the social acceptance that, for a long time, was missing and slowed down the adoption of video communication as an alternative (or complement) to voice or text.

The increased use of video in its multiple forms – e.g., downloaded, streamed, uploaded, interactive – reinforces the role of latency in determining overall subscriber experience (and the attendant churn rates). For this reason, video has, since the beginning, been one the main drivers for edge computing. Not only is video traffic growing it is often location-specific and concentrated in specific locations and times that are the backdrop for perfect use cases to justify moving processing and storage to the edge.

Edge computing cannot lower the latency in the RAN, but it can reduce the end-to-end latency: video can be cached at the edge, or the processing (e.g., with ABR) that optimizes video traffic can be done at the edge. Both approaches can be combined to reduce the traffic between the centralized core and the edge.

Video is an instance, however, in which it is crucial to pick the edge's location within the network carefully to avoid unnecessary investment and complexity. For instance, caching at the edge works because subscribers at a given location and time tend to watch a remarkably consistent and narrow set of content. But if the MEC server is co-located with a small cell, there are too few subscribers within the cell footprint to justify caching, even if the cost of storage were not an issue. At the same time, the more remote the integration point is, the lower the impact on latency and hence on QoE.

Global mobile data traffic by traffic type



Figures in parenthesis in the legend show traffic share for 2016 and 2021

Source: Cisco VNI Global Mobile Data Traffic Forecast, 2016-2021

We want it all, we want it now, we want it here: immediacy and time/location awareness

The use cases for mobile computing go beyond the need for low latency in video and real-time traffic. Immediacy is another performance yardstick that subscribers care about and expect in all applications, not only real-time ones. Immediacy can be quantified as the time it takes for an application or service to launch, or for content to appear on the device or register as being sent.

Edge computing can bolster immediacy, because content no longer needs to be downloaded from a remote data center, or to be sent to the data center for processing and then back to the subscriber. Managing applications and traffic types in real time at the edge enables operators to control the effects of congestion on time-to-content and to provide the desired level of immediacy.

In enhancing QoE, the relevance of and application or content is complementary to immediacy – and relevance is often tied to time of day and the subscriber's location. Video content popularity peaks and declines very quickly. An offer for a discount on a restaurant meal works best if you are near the restaurant and hungry. Mobile operators already provide time- and location-aware content, but edge computing can make that content more closely tied to the location and hence more relevant to subscribers. In addition, time- and location-aware content appeals to content providers and venue owners that may be the source of the content and applications hosted on the edge servers.

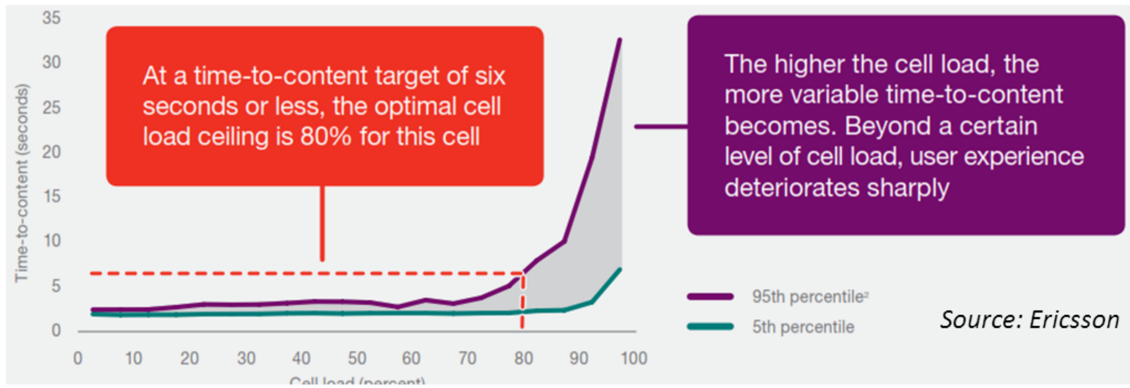
Stadiums are a good example of a location with high traffic loads that are mostly location/time dependent and have a high perceived value to the spectators. Videos from the game, content associated with teams, or ads for services available at the stadium generate a massive amount of content – and to a large extent, it is also consumed locally. Storing and processing that locally, at the network edge, will improve QoE for those attending the game. And by offloading that work from the rest of the network, it will improve QoE for other subscribers, too.

MEC servers in stadiums may be used to provide location-based services and content not only through the network of a specific operator, but also

2017 Super Bowl: AT&T stats
$40 million investment in network expansion
9.8 terabytes of data during the game, equivalent to 28 million selfies
88% increase in traffic over last year
148% up from average pro football game traffic
59.9 terabytes in the Houston area over the weekend
Source: AT&T

Bank of America stadium in Charlotte, North Carolina *Source: Carolina Panthers* across operators and the local, possibly stadium-operated Wi-Fi network. Initially these services may include location-based retail and advertising, services for the public, event-specific applications, surveillance, and stadium operations and services. Farther in the future, they may expand to AR/VR services and IoT applications.

Impact of cell load on time-to-content (immediacy)



Edge-computing can help reduce cell load during congestion by real-time optimization of application data served to the user. Ericsson's time-to-target of six seconds is generous; a lower one may be required to meet subscribers' expectations and the requirements for applications such as location-aware retail. For instance, Walmart found out that purchasing decisions are made in a millisecond time frame. To close sales, the retailer not only needs to have a fast time-to-content, but it also has to fine-tune the proposition – e.g., cost, delivery, special promotions – to the individual customer, within the context of inventory availability, within the same time frame. This is the type of processing may benefit from edge computing because it relies on local information: the shopper, the inventory.

> In 2016, "Walmart announced that to upgrade its ability to compete in e-commerce with Amazon – which still does eight times Walmart's sales online – it was buying Jet, a year-old Internet retail startup. The Economist reported on August 13, 2016, that Jet's appeal to Walmart was its "real-time pricing algorithm, which tempts customers with lower prices if they add more items to their basket. The algorithm also identifies which of Jet's vendors is closest to the consumer, helping to minimize shipping costs and allowing them to offer discounts. Walmart plans to integrate the software with its own." It turns out that "under a second" was just too damned slow." Thomas L. Friedman, Thank You for Being Late: An Optimist's Guide to Thriving in the Age of Accelerations.

From MEC proofs of concept to use cases

If you have heard of only one use case for MEC, it almost certainly involves video. However, the scope of edge computing is much wider. It is driven by the diverse needs of multiple stakeholders:

- Keep latency low for real-time traffic – video, but also voice for enterprise environments
- Offer applications and content that target visitors in a venue and may be finely tuned to their location (e.g., proximity to a store) or timing (e.g., off-peak hours)
- Keep local content within the premises; avoid sending content that may be generated and accessed locally, or that is location-specific (e.g., enterprise data) to the centralized core, and back to the RAN to be delivered to the subscriber
- Share local content among available wireless networks, including Wi-Fi and multiple mobile service providers
- Provide an extra layer of security by keeping local content within the enterprise premises
- Optimize content delivery based on real-time RAN conditions, as well as other factors, such as subscriber devices, content type, application requirements

ETSI MEC PoCs show areas of early interest from participant operators and vendors. As MEC gets deployed, the list of use cases will grow. The table on the next page lists some that have attracted attention to date.

MEC stakeholders: who gains what?
Mobile operators and other service providers: provide better QoE, better resource utilization, offload from the centralized core
Venue owners: offer venue-based services to visitors, enrich their experience and encourage them to share it
Enterprise, IoT: develop and support enterprise-specific applications and services, have fast and secure access to enterprise data and applications over multiple networks, and support IoT applications within the enterprise
Content providers, OTTs, application developers: optimize content delivery and application access in real time, adapting content to RAN conditions and subscriber requirements
Subscribers: better access to applications and content, increased immediacy and relevance of wireless connection, leading to a more positive QoE

ETSI PoC description	Participants
PoC1 Video user experience optimization via MEC - A service-aware RAN PoC	Intel, China Mobile, iQiYi
PoC2 Edge video orchestration and video-clip replay via MEC	Nokia, EE, Smart Mobile Labs
PoC3 Radio-aware video optimization in a fully virtualized network	Telecom Italia, Intel UK Corporation, Eurecom, Politecnico di Torino
PoC4 FLIPS, flexible IP-based services	InterDigital, Bristol is Open, Intracom, CVTC, Essex University
PoC5 Enterprise services	Saguna, Adva Optical Networking, Bezeq International
PoC6 Healthcare, dynamic hospital user, IoT and alert status management	Quortus Ltd, Argela, Turk Telecom
PoC7 Multi-service MEC platform for advanced service delivery	Brocade, Gigaspaces, Advantech, Saguna, Vasona, Vodafone
PoC8 Video analytics	Nokia, Vodafone Hutchison Australia – SeeTec

Source: ETSI

Edge computing use cases	
RAN optimization, analytics	Enterprise/venue applications over cellular and Wi-Fi
Compute offload from central data center	Application sharing in networks managed by a neutral host or an enterprise
Video, video analytics	Real-time operations and control in industrial, health, and surveillance verticals
IoT	Connected cars
Public safety	Augmented/virtual reality
Smart cities	Gaming
Venue/retail applications: advertisement, operations, content, promotions	vCPE

Which: Initiatives

Edge computing requires more effort and investment than adding a server at the selected edge location. Many moving parts have to come together to enable the stakeholders – service providers, venue owners, applications/content providers, enterprises, subscribers – to fully harness the benefits of shifting functionality to the edge.

Correspondingly, there are multiple, complementary initiatives converging to create an ecosystem that will support distributed network models end to end. Each stakeholder will likely find only a few of the initiatives relevant, and may feel overwhelmed by the apparent competition among them. For vendors, the best bet is often to be active in multiple initiatives to prepare themselves to participate successfully in the nascent ecosystem.

Among the factors that converge in the creation of the edge-computing ecosystem are these:

- Edge hardware. It may have to be installed outdoors or in locations with space, security or environmental constraints that differ from those of a data center or central office.
- Mobile devices. New device types will populate the network to support IoT applications, and many of them are likely to benefit from edge processing.
- Services and applications. These may be targeted directly at subscribers, enterprise workers, or visitors to a venue, or may be used for IoT; they may be managed by different entities – mobile operators, other service providers, venue owners, enterprise, OTTs or content owners.
- Integration in the end-to-end network. Edge functionality has to be tightly integrated with the RAN and the centralized core. Managing potentially dynamic edge locations (i.e., the edge location changes in real time depending on network conditions) requires orchestration capabilities in the core network.
- Integration across networks. The operators of multiple networks, possibly owned and managed by different parties, may want to share edge processing and storage resources in the same location.
- Application developers. They need to optimize their apps or develop new ones to work in an edge environment.

ETSI MEC

- ETSI-based standardization, with wide vendors and operator support, linked to other ETSI program
- Vodafone, Nokia, Intel, IBM, DOCOMO, Huawei as founding members

OpenFog Consortium

- Facilitate IoT deployments by moving functionality to the edge, with a multi-access approach
- Distributed computing, storage and networking anywhere along the cloud-to-thing continuum
- ARM, Dell, Cisco, Microsoft, Intel, Princeton University as founding members; AT&T, GE, SAKURA, Schneider as contributing members

CORD, MCORD

- Leverage central office location to move functionality and applications towards the edge
- Lower capex/opex, agile platform, integrated with network slicing
- Open source and disaggregated components
- Google, AT&T, Verizon, Comcast, SK, NTT Communications, China Unicom, Nokia, Huawei, Ericsson, Cisco, Intel

Open Edge Computing

- Development of applications that most directly benefit from edge computing and technologies that support them
- API development
- Verizon, Vodafone, Intel, Huawei, Nokia, Crown Castle, Carnegie Mellon

Open Compute

- Development of more flexible, efficient and scalable hardware
- AT&T, BT, Orange, Verizon, SK, Deutsche Telekom, Ericsson, Facebook, Fidelity, Goldman Sachs, Google, Lenovo, Microsoft, NetApp

TIP

- Facebook’s project to roll out radio infrastructure in underserved areas

Where: Topologies

We have talked so far about moving processing, storage and control to the edge, as if it were clear what we mean by the edge. But it is far from obvious where the edge is – or, more accurately, where potential edge locations are, and which one or ones a service provider should select.

This is a crucial question – perhaps the most important one – to ensure that edge functionality brings both performance and financial benefits to the service provider and the other edge stakeholders. If the edge location is too far out, too close to the subscriber, edge computing may become overly expensive and complex. If the edge location is too close to the centralized core, the benefits of edge computing dissipate, with a more complex network topology but no significant improvement in performance.

And is there a single edge? Not only may different service providers pick different edge locations for their networks or specific locations in their networks; it may also make sense to have multiple edges in a given location, depending on the applications.

Location-based content and applications are most likely to be hosted in an aggregation point that reaches all the infrastructure that covers the venue. An enterprise deployment may be housed in a location that covers all the enterprise's buildings or just a subset of them.

For applications that require video caching, service providers have more flexibility in choosing the edge location. They may want to see what their subscriber usage patterns are, and pick an edge location where they can maximize the caching contribution.

And is the edge a fixed location? It does not have to be, although initially it is likely be. For many applications – e.g., location-based and enterprise applications – an edge location that does not change though time may be desirable. But for locations with highly fluctuating network loads or for applications with uneven temporal and spatial distribution, a moving edge that shifts depending of real-time network conditions is possible in a virtualized environment and can maximize the cost/performance benefits of edge computing.

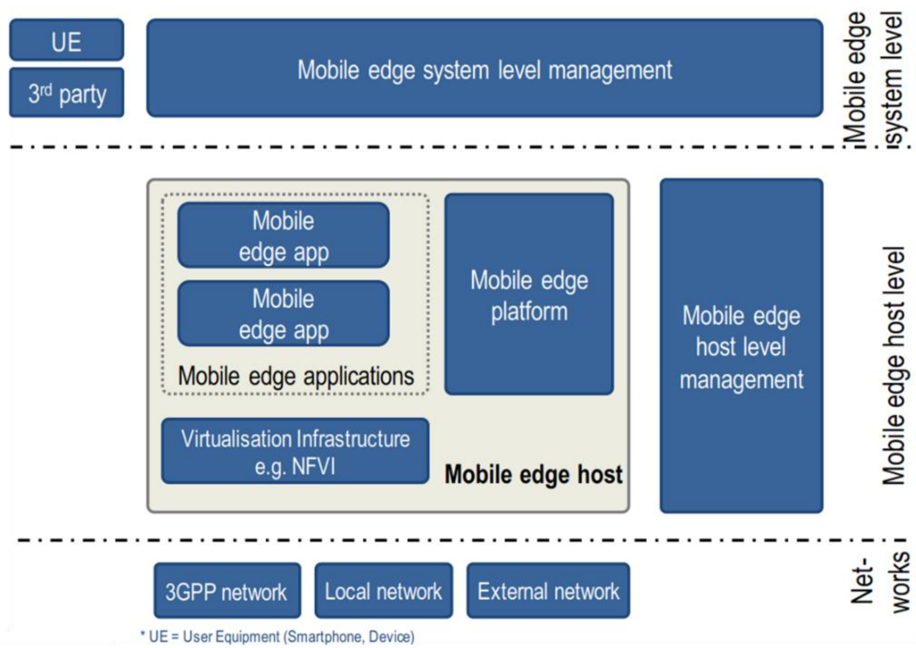
MEC architecture

To address the issue of where the edge is or could be, it is useful to review the MEC architecture proposed by ETSI. Other edge computing initiatives rely on the MEC architecture or use their own edge server. Traffic to and from UEs that involves applications, services or content hosted in the edge server is directed to the server; the rest of the traffic is routed to the centralized core as usual.

The MEC server or edge host uses a virtualized platform to host applications It can interface with cellular (3GPP) networks, as well as other available networks, including Wi-Fi networks. Application developers have access to API to get their applications hosted in the edge server. The MEC server provides the processing and storage capability to support the hosted apps. Storage may be used for caching frequently accessed content, or for local breakout (to keep local content within the MEC footprint), and avoid using backhaul resources to transmit the content back and forth. Local processing enables applications to optimize their performance.

Another important and innovative element in the MEC server is the addition of user- and network- information services. These provide the foundation for optimizing end-to-end network performance.

MEC computing framework



Source: ETSI

MEC services for network optimization

Radio network information service (RNIS)

- Up-to-date radio network conditions
- Measurements and statistical information related to the user plane
- Information about the UEs served by the radio node(s) associated with the host (e.g., UE context and radio access bearers)
- Changes in UE information

Location information service

- Location information: cell ID, geolocation, etc.
- Location of specific or all UEs served by the radio nodes associated with the ME host
- Location of a category of UEs (optional)
- Location of all radio nodes associated with ME host

Bandwidth manager service

- Allocation of bandwidth to ME applications
- Prioritization of certain traffic routed to ME applications

Source: ETSI

The MEC server is the bridge between what happens in the RAN and the UE and what happens with the applications and content. It enables network operators, application and content providers, and others that may play a role in serving subscribers to manage traffic in real time on the basis of factors such as application/content requirements, network conditions, and policy – to optimize utilization of the available network resources. Network operators can use this information to manage core and RAN resource allocation, but they can also share this information with content owners, venue owners and application providers to coordinate traffic management with them.

Where is the edge? Location tradeoffs

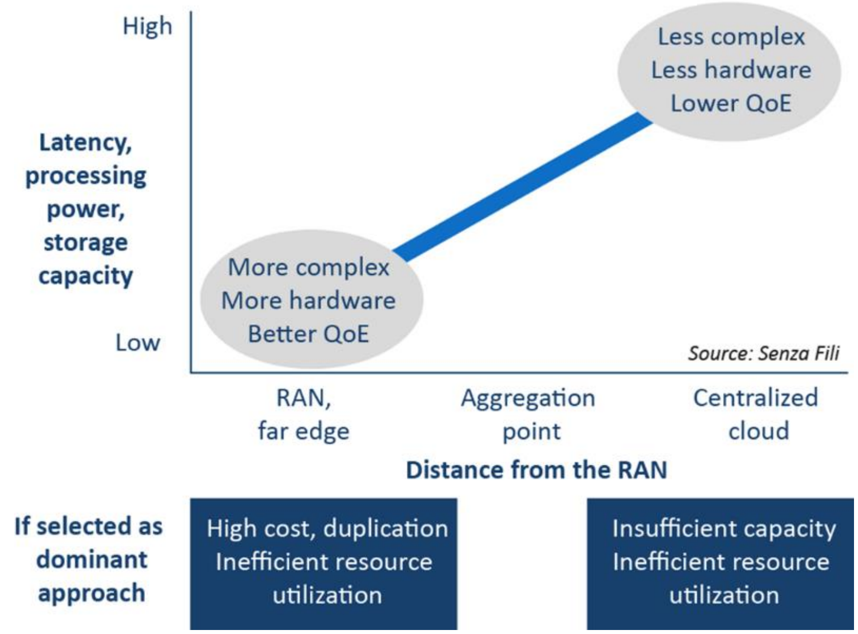
If cost, complexity and space were not constraining factors, moving processing and storage to the far edge would generally be the best way to improve QoE. But they do matter, and so the choice of edge location does not rest solely on performance, but on the evaluation of tradeoffs among multiple factors. The criteria for such evaluation will vary among operators, but there are some high-level considerations that apply across networks.

As the chosen edge location gets closer to the UE:

- The latency gets lower, and this improves performance on applications that are sensitive to it, such as video calls, voice or gaming, leading to a higher QoE.
- Hardware has less storage capacity, limiting the amount of content that can be stored at the edge.
- Processing power becomes more limited, and hence it may not be efficient to run some applications from the edge.
- End-to-end network complexity increases, because network operators have to deploy, integrate and manage more hardware, at a higher number of locations.
- Resources may be needlessly duplicated if applications could be efficiently run from a centralized location.

The opposite side of the equation holds when the chosen edge location moves toward the centralized cloud (i.e., higher latency, more storage capacity, more processing power, less complexity, more efficient use of resources).

Pushing functionality to the edge: tradeoffs



Another key consideration is the footprint – it gets smaller as the edge host moves away from the centralized cloud – because this is the area over which the RAN optimization and the mobility management are scaled. When the edge host covers a small area, edge traffic optimization is confined to this area. So, two MEC servers can both optimize their respective footprints, but as two separate zones. If a single MEC server covers both footprints, optimization can be coordinated across the entire area. Coordination across a wider area may make optimization more effective, but at the same time the greater distance to the RAN may limit the granularity of the optimization capabilities.

The footprint of the edge host is also determined the type support for mobile access in edge-based applications and services. Within the footprint, subscribers' experience is preserved as they move from one cell to the next within the same network. If the edge host covers multiple networks, application access can be preserved across networks. As the subscriber moves away from the footprint to an area that has no edge host or a different one, access to the application has to be managed to ensure a smooth transition. When the two footprints do not have the same edge capabilities, the subscriber may experience a discontinuity in the quality of the access to the edge application. Obviously, if the application is supported only at the edge (e.g., an enterprise application or a location-based application in a mall), subscribers lose connectivity to the application as they move out of the footprint. In many cases this is a desired outcome (e.g., the enterprise may want its services available only within its campus), but it is something that has to be kept in mind when selecting the edge host location and how to manage mobility across footprints.

Factors that play a role in the selection of the location of the edge server include:

- **RAN resources.** The edge server capabilities must be sufficient to serve the covered footprint. If they exceed the RAN capabilities, the investment in edge computing is wasteful.
- **Backhaul resources.** Edge computing may address capacity and/or latency limitations in the backhaul and prevent the backhaul from becoming a performance bottleneck.
- **Applications.** Latency, processing and storage requirements that affect edge location vary across applications, so the ideal edge server location varies by application.
- **Subscriber/client expectations, and venue-owner preferences and requirements.** The expectation for service performance and QoE may be different for enterprise employees, mall visitors, or IoT sensors.
- **Operator policy and preferences.** Operators may want to position themselves in the market in a specific way – e.g., provide a higher-quality service for a specific enterprise client or venue owner.
- **Content provider preference.** If the content provider pays for the edge infrastructure, it will want to choose the location of edge hosts, because this will allow the provider to maximize its return on investment.

In addition, real-time optimization enables operators to shift the location of the edge dynamically depending on factors such as demand for an application, RAN conditions, concentration of subscribers or power considerations. A virtualized platform will make it possible to instantiate applications at different edge locations as operators expand their edge computing and virtualization capabilities.

Where is the edge? Location options
vCPE, home/office GW, HeNB
RAN
Wi-Fi access points
Corporate/venue/factory gateway
Cable boxes
Outdoor furniture
C-RAN BBU pool
Central offices
Cable/operator gateway
RNC
Other aggregation points

Optimizing end-to-end network resource utilization from the edge

Improving performance is an obvious goal for mobile operators. But when it comes to ensuring cost effectiveness and profitability, network resource utilization is the goal that should take center stage. It is a measure of how much value operators can squeeze out of the network infrastructure they have – which in turn is a measure of subscribers' happiness with the service.

In today's networks, the prevailing approach is to maximize performance, given the financial resources available. This typically means increasing RAN capacity using a brute-force approach – i.e., deploying the latest technological tools that increase throughput. But far less effort is put into optimizing the use of the capacity available. It is like buying a fast car without having roads that allow you to drive fast.

There are many ways to optimize network resource utilization and, under intense pressure to improve performance without increasing costs, mobile operators have started to work toward this optimization goal. MEC and edge computing in general are geared to achieving exactly that by changing the processing, storage and control in the network.

The impact on QoE from moving processing and storage to the edge is easy to grasp, even though it is not trivial to quantify over a network because it depends on multiple environmental factors that are variable. Other things being equal, though, moving processing and storage to the edge improves latency, immediacy and QoE.

The new control features at the edge introduce a new type of optimization, one that works in real time, leveraging information about network conditions to optimize end-to-end network performance instead of optimizing the performance of individual network elements. Edge computing is not required for this type of optimization – it can be implemented in current 4G networks – but MEC servers are well suited to gathering information from the RAN, processing it and forwarding the results to the centralized core or to content or application providers. ETSI specifications define services (see table above) that collect information that can be used in multiple ways to optimize the utilization of network resources and QoE.

Some optimization is best done, either remotely or at the edge, by the entity that controls the applications and content –the mobile operator, an OTT or a content owner – because that entity has direct control and better knowledge of content and applications, as well as better access to them. And it often wants to retain some degree of control to ensure it can create the subscriber experience it aims for.

An example is that, increasingly, the content served to subscribers is encrypted; operators do not know the content type, much less have the flexibility to optimize it, but the content owners do.

Throughput guidance is an optimization tool that is being developed to address the mobile operators' need to adapt content to real-time RAN conditions and to be able to do so in collaboration with third parties. It uses data about network conditions – especially RAN conditions and RAN load – to generate advice for content and application providers on how to manage traffic exchanged with the subscriber. When the network has sufficient capacity, the providers can share content at the highest quality available. When the network is capacity constrained or congested, the content and TCP transmission can be adapted to provide subscribers the best experience possible given the real-time availability of network resources.

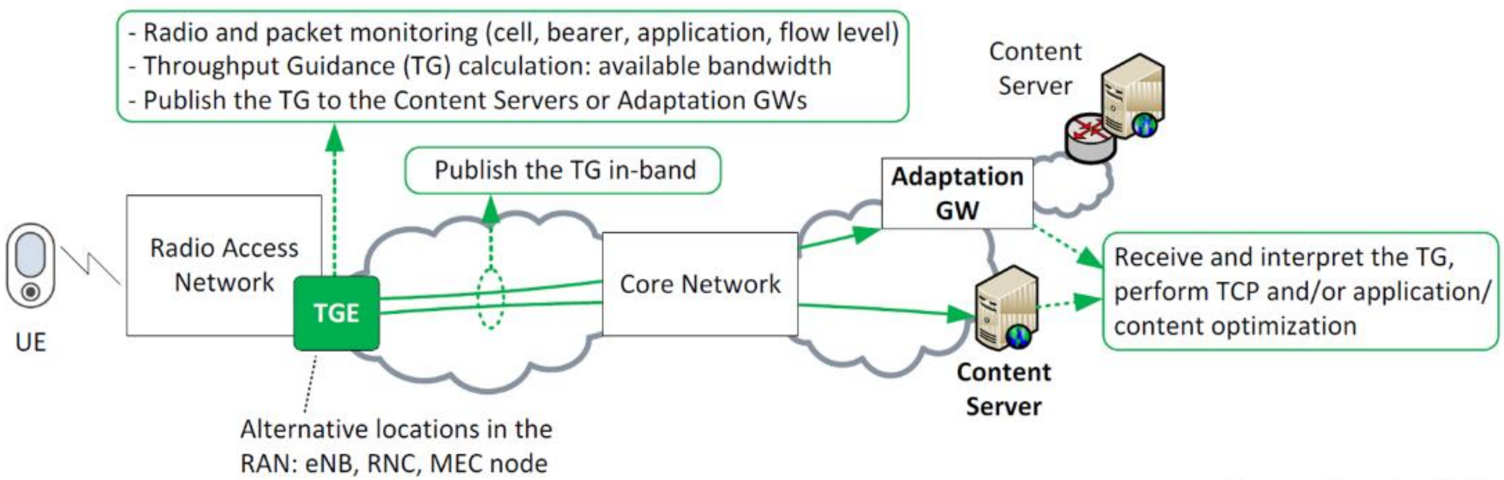
Tools like throughput guidance allow operators to tell third parties what they can do to optimize content delivery, without the operators having to access it directly. For it to work, however, operators and third parties have to tightly coordinate network optimization. In some cases, this will require getting past a history of tension and competitiveness between mobile operators and content/application providers. The situation is rapidly changing, however, as both camps realize they need to work together to provide outstanding QoE. And

both sides might benefit further if throughput guidance and other optimization tools are the catalysts that facilitate tighter relationships among them.

Throughput guidance: results from Google and Nokia trial
Network metrics improvement TCP retransmissions 30-45% TCP round-trip time 55-70% Mean client throughput 20-35% TCP packet loss 35-50%
Application metrics Click-to-play time reduction 5-20% Average video resolution improvement 5-20% Video format change frequency reduction 10-25%

Source: Nokia, Google

Throughput guidance



Source: Google, Nokia

Who: Business models

Mobile operators will continue to have a leading role in planning and funding MEC deployments, because they are integral parts of their network infrastructure. Other edge computing hosts, partially or fully independent of mobile operators, are likely to be deployed, paid for, and operated by the enterprise or venue owner, even if some of the applications supported can be hosted by the mobile network.

Even in the MEC space, however, new business models may arise that have a more direct and active role for venue owners and the enterprise, on the periphery side, and for content/application providers, on the cloud side. They both stand to benefit directly from the MEC infrastructure – and, in some cases, more than mobile operators do.

For example, a MEC server that supports industrial IoT applications in a warehouse may be more valuable to the enterprise than to the mobile operator. The enterprise may see in it a compelling business case, while a mobile operator might struggle to see a positive ROI or might not be able to assess the revenue potential because it is dependent on enterprise-specific applications that it is not familiar with.

Similarly, a content or application provider may be willing to locate some of the infrastructure it needs at the edge of the network when that is more effective – and potentially more cost effective – than a remote cloud location. And it is not only companies like Google or Facebook that may be interested in having a presence at the edge.

Smaller companies may also be willing to locate processing and storage functionality at the edge in a virtualized environment, where they do not need to own a host server but might pay only for the services they need. In this model, the mobile operator may deploy and pay for the initial edge hardware, but then it can monetize the investment by renting access to it to a third party.

An arrangement of this type can be mutually beneficial, especially when accompanied by joint efforts to optimize network performance – for instance, with tools like throughput guidance.

Edge computing stakeholder	Why pay for edge computing?
Mobile operators	▪ Better QoE. Churn reduction, lower customer support costs. ▪ Increase in network resource utilization. More value extracted from existing infrastructure, better end-to-end network TCO, need for capacity expansion. ▪ Location-based services. Monetization of services to subscribers (possibly), to the enterprise, to public venues, and for IoT services. ▪ Offload centralized core. Cost-effective improvement of both QoE and resource utilization. When the centralized core requires additional processing/storage capacity, operators may decide to deploy it surgically at the edge where needed.
Other service providers, including IoT providers	▪ Location-based services. Service providers, such as DAS neutral hosts, cable operators, wholesale service providers or MVNOs, may be willing to install their own edge infrastructure to provide services specifically targeted to a location. They might use a local access network they own or manage, or lease capacity from the RAN local network operators. They may monetize such services to venue owners or other parties with a presence in a venue (e.g., retail) or with IoT applications.
Public venue owners	▪ Location-based services. Offered as an amenity to guests (e.g., stadium, hospital) or as a service (e.g., city, college); to advertise to visitors; to support needs of tenants (e.g., stores in a mall); to support their own operations. Services can be made available through Wi-Fi or MulteFire venue-owned networks, DAS networks or small cells.
Enterprise	▪ Local breakout. Keep enterprise data and applications local, provide enterprise voice services. ▪ Security. Keep enterprise data and traffic local to the enterprise. ▪ Enterprise services. Develop and possibly mange enterprise-based services and applications. Services can be made available through Wi-Fi or MulteFire venue-owned networks, DAS networks or small cells
Content and application providers	▪ Shift of processing and storage to the edge. Improve QoE, better control delivery of service, coordinate real-time traffic optimization with operators. Content and application providers may own their edge infrastructure, but leasing resources from the operator's virtualized edge servers is an approach more likely to be accepted by both operators and third-party providers.
Residential and small businesses subscribers	▪ Home/small business gateway. Residential and small-business customers may invest in an edge host that supports services and hosts content used by the people within the premises and shared over the Wi-Fi and cellular networks. Service providers may subsidize the edge host as a subscriber-retention feature.

When: Timeline

In many ways, edge computing is nothing new. There have been edge computing solutions all along to serve niche markets or to address specific performance and optimization challenges in mobile networks. What is different today is that network virtualization offers a framework to expand edge computing capabilities, adding scalability, reliability, flexibility, cost effectiveness. This will take edge computing to the mainstream and enable operators to reap benefits in terms of improved QoE and resource utilization.

The ETSI MEC standardization work creates the foundation for edge computing deployments in mobile and, increasingly, fixed networks. During the first term (2015-2017), ETSI ISG completed the groundwork, released the basic specifications, and encouraged the creation of the ecosystem. More work is needed during the second terms (2017-2018) not only to expand beyond mobile networks, but also to strengthen the links with other edge computing initiatives while avoiding the risk of fragmentation of efforts.

Beyond standardization and industry collaborative initiatives, there is a need to explore different business and deployment models, and revisit the role that stakeholders – e.g., venue owners, enterprises, content and application providers – will have in deploying, managing, and funding edge computing deployments.

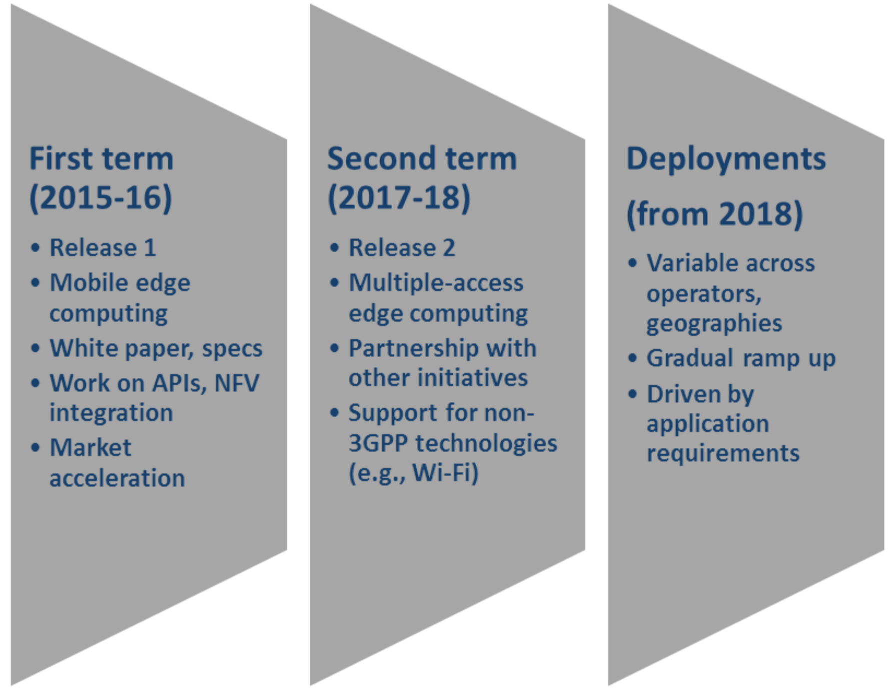
The business case also needs to be assessed to understand where and when edge computing provides a better return that the centralized cloud. To assess the business case for edge computing we need to go beyond the standard ROI model. Improvements in QoE or resource utilization are highly valuable, but notoriously difficult to quantify, because they involve end-to-end network improvements. A traditional financial model that looks at a solution that delivers a well-contained benefit is inadequate for edge computing, as it does not adequately capture the costs that it requires and the value it brings.

The time to commercialization can be fast as edge computing can be introduced without waiting for full network virtualization or 5G. In practice, however, it will take a couple of years before commercial launches, as vendors and operators complete their trials to learn what is the most efficient way to balance centralized versus edge processing and storage.

ETSI MEC second term objectives
▪ Support 3GPP access technologies (Wi-Fi and fixed)
▪ Extend the virtualization support types, to render the environment as attractive as possible for third-party players
▪ Study possible charging models which may be applicable to MEC
▪ Fill gaps relating to lawful interception
▪ Develop testing specifications and test methodologies
▪ Coordinate plug fests
▪ Coordinate experimentation and showcasing of MEC solutions
▪ Expedite the development of innovative applications
▪ Ensure a low entry barrier for application developers
▪ Disseminate the results of the work
▪ Strengthen the collaboration with other organizations
▪ Study new use cases
▪ Enable MEC deployments in NFV environments

Source: ETSI

MEC timeline: standardization and deployments



Source: ETSI, Senza Fili

Implications

Mobile edge computing takes us beyond the centralized cloud, to hybrid virtualized model, which combine centralized and distributed processing, storage and control

Operators can leverage network flexibility to find the best edge location to maximize QoE and optimize network resource utilization, the main drivers for edge computing

MEC is not for mobile networks only. Other fixed networks including Wi-Fi can use the MEC framework and share it with mobile networks

There are multiple network edge locations where it makes sense to deploy MEC servers or other edge hosts. Evaluating the tradeoffs that these locations offer is crucial for successful edge computing deployements

New business models will accelerate move to the edge, with an increased role of venue owners, enteprises, and application and content providers.

Application and content optimization at the edge encourages a tighter cooperation of mobile operators with application and content providers

II. Vendor profiles and interviews

Profile

Qwilt

Qwilt was founded in 2010 to help broadband fixed and wireless service providers optimize the delivery of video traffic, both to meet the capacity and latency requirements of high video traffic loads, and to improve quality of experience.

From the beginning, Qwilt has developed solutions based on open caching to manage video content from multiple sources and direct it to a variety of subscriber devices that work across different types of networks. Depending on the application requirements and service provider preferences, Qwilt solutions can be deployed in the core, at the Gi-LAN, or in the edge cloud, at the eNodeB.

Qwilt's Open Edge Cloud platform provides open caching and content delivery solutions for service providers that want to optimize video delivery and other real-time applications such as augmented reality and virtual reality at the edge of the network.

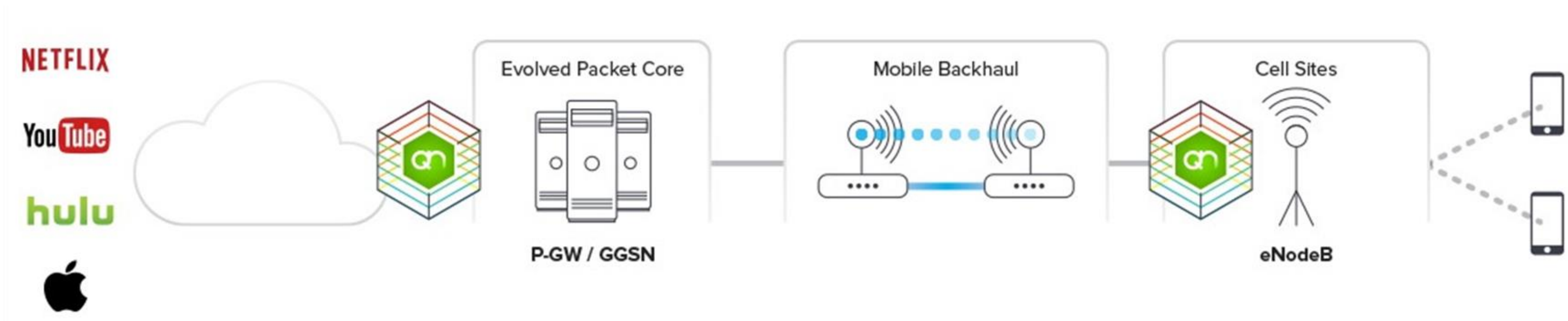
The Edge Cloud platform leverages compute and storage capabilities as close as possible to the edge – and hence to the users – to minimize latency. It relies on cloud management and connectivity, and open APIs using Edge Cloud Nodes. The Edge Cloud platform is not designed to replace or compete with CDNs, but rather to complement them, and carry the traffic where CDN nodes are not deployed or are not cost effective.

Qwilt's Edge Cloud solution aims to extend the content providers and CDNs footprint, to lower transport costs and to improve delivery quality. It also allows them to manage traffic spikes and adjust to the uneven distribution of traffic through time, and, hence, to utilize network infrastructure more efficiently and reduce the need for capacity expansion.

The solution allows service providers to meet anticipated traffic spikes (e.g., a game or an update) or unexpected ones (e.g., an accident or a viral video). It can lower capex (less infrastructure investment is needed) and lower opex (peering and transit costs are lower).

Open Caching software combines open caching with media delivery and analytics, and runs on an NFV platform, the Video Fabric Controller, that runs on COTS. With open caching, data that is used frequently is stored at the edge and delivered when requested, without requiring any action from content providers, CDNs, or subscribers.

Qwilt estimates that 10% of titles account for over 80% of video traffic. Without caching, this frequently accessed content has to repeatedly traverse the network, increasing costs and latency, and lowering QoE. Caching at the edge enables operators to reduce traffic and cost within the core, and minimize latency. According to Qwilt, with open caching operators see an average streamed bit-rate increase ranging from 55% in North America, to 133% in Asia.



Qwilt's Open Edge Cloud platform

Source: Qwilt

Qwilt

Video at the edge optimizes fixed and mobile networks

A conversation with Dan Sahar, Co-Founder and VP of Product Marketing, Qwilt

Monica Paolini: Video is the traffic type that benefits the most from edge processing and storage, both in fixed and mobile networks. And increasingly they both share platforms and traffic management tools. Edge computing will accelerate this trend. We explore these topics with Dan Sahar, Co-founder and VP of Product Marketing at Qwilt.

Dan, can you tell us what is that you do at Qwilt to help service providers improve the subscriber experience with streaming video?

Dan Sahar: Qwilt was founded in 2010, with the objective of helping service providers address the growth in online video. Streaming video was just starting back then, with the likes of Netflix and YouTube, and that had a major impact on service provider networks.

We realized the best way to address that challenge was by changing the way broadband networks and mobile networks are built, and by bringing the content delivery function of those videos, primarily, closer to the subscriber. And by doing that, we gain efficiencies on multiple fronts, both on the economic front as well as on the quality of experience front. The solution we created is one that enables these service providers to do exactly that. You could think of it as the last tier of content delivery that sits inside the service provider network, and is able to acquire content from a range of sources, and deliver it in close proximity to where the users are.

Monica: That's a major challenge, because video is the most difficult type of content to transfer over fixed or mobile networks. Can you tell us what you do to the content itself? How did it change throughout the years?

Dan: One key principle in our solution from day one was that we maintain the same fidelity for the video as it was originally streamed and thus we do not make any changes to the videos themselves. Video has changed in several ways through the years, from progressive download initially, into adaptive bit rate. Adaptive bit rate is probably the method most streaming video providers use today; progressive download has pretty much faded out. And the move we're seeing right now is to use adaptive bit rate for both live and VOD content, and to look at ways to optimize and secure that delivery. We're also seeing a growth in TLS and HTTPS delivery mediums. Our solution evolved to address ABR, as well as changes on the transport and content sides.

Monica: When we talk about edge computing, how do you define the edge? Where is the location of the edge that optimizes delivery of video and other types of content?

Dan: We see the edge in primarily two locations. Inside the network, it would be the first IP location in the network. On the fixed-line side, like in a cable network, that might be a CMTS location, and it can be on the B-RAS on the fixed-line side. On the mobile side, it used to be at the SGi or Gi level. We're seeing it move deeper, to the eNodeB and S1 interface. That would be the first point of edge inside the network.

The second place where the edge can play a role is on the device or at the home. On the device, this can be the handset and the software application running on the device. At the home, it can be a residential gateway, or even an Apple TV or a Chromecast or an Amazon Echo that has some built-in content delivery capability inside of it.

And there're different characteristics for each one of these locations: some have more processing power and more storage; others have less processing power, less storage, but they're a lot closer to the consumer, so they bring more value to the entire value chain.

Monica: How do you decide where the edge is for a specific type of content, application, content provider or service provider? Is there an easy way to figure it out?

Dan: I think there's no one right answer for that. You could equate this to the way packages are delivered in the real world. You can have US Postal Service, you can have FedEx, pretty soon you'll have drones delivering them to your home. They all get the package there on time, but some of them cost more, and each one has different capabilities. That's a good analogy to how content delivery is done. Some things you can deliver from the centralized cloud – for example, from an AWS data center – and for other things, there's a lot more value to doing them at the deep edge of the network.

Ultimately it has to have an economic balance. There has to be a benefit to the service provider to deploy edge computing and storage resources so that content delivery can leverage these capabilities. Then you have to decide which content can be delivered from the centralized cloud, and which content has to come from the edge.

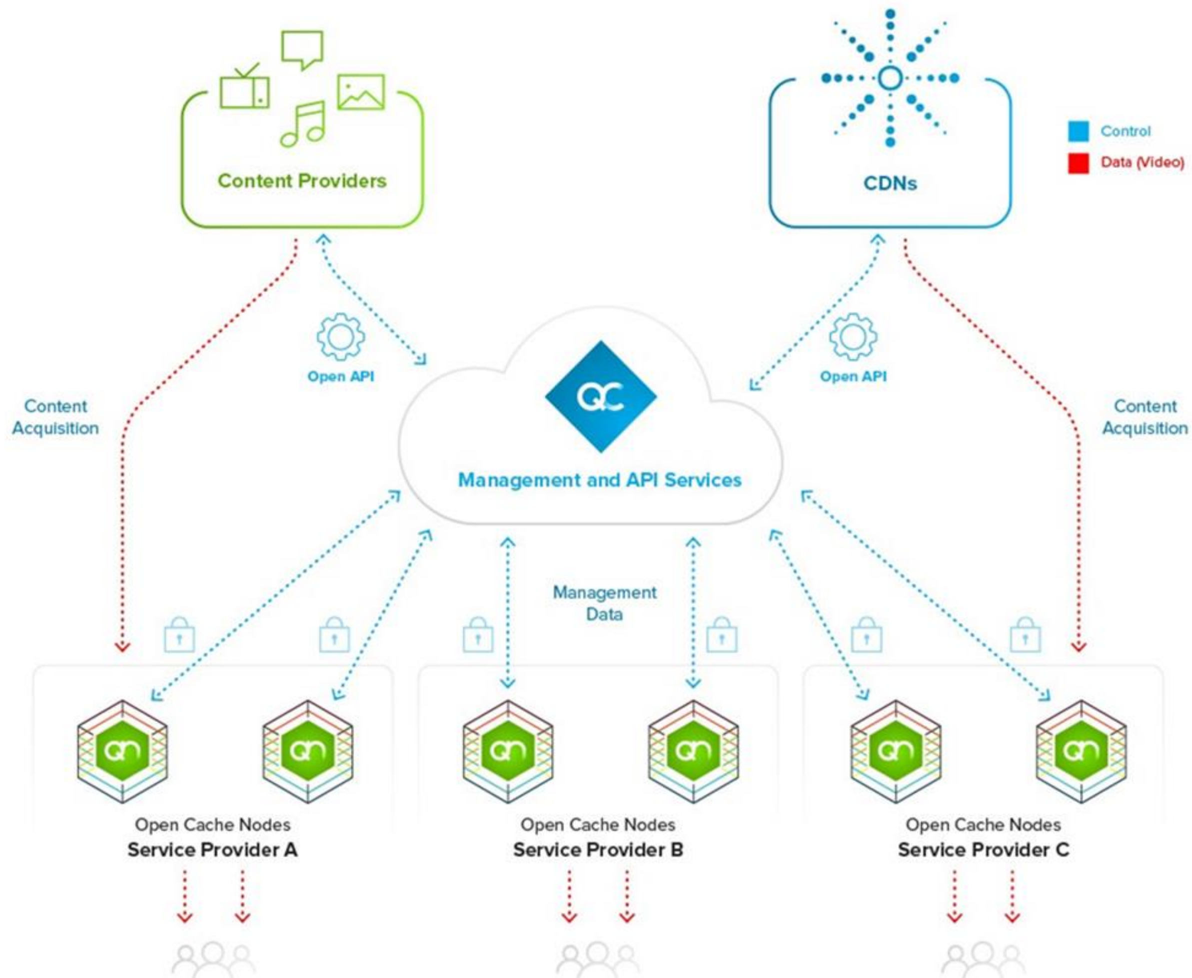
Monica: At the beginning, video was over fixed networks. And when video over mobile came along, it was a completely different thing – just short videos, and different from video over fixed networks. Is that still the case, that we can make a clear distinction between fixed and mobile?

Dan: I think they're becoming very, very much the same. And operators are changing as video changes.

On the consumer side, people watch videos on their mobile devices. It can be on Wi-Fi, but when they get out of their own homes, they continue to watch the same videos. The content is becoming a lot more similar across delivery mediums. Maybe you have a bigger screen on one than on the other, but the content is not different.

There's some adaptation the content provider has to do for the screen size, but, other than that, the transport medium and the ABR formats are exactly the same on both access types. And I think it's a good thing: the industry is becoming one big video medium, so you can watch video wherever you are.

Monica: And the expectations are pretty much the same from the user perspective, so they're not willing to say, "Well, since it's mobile, it might not be as good quality." They expect the same good



Unified open API to third-party publishers and CDNs *Source: Qwilt*

quality they expect from fixed networks. What does this convergence mean for Qwilt in terms of the solutions you provide?

Dan: Our solution has two main components. One is the edge-cloud nodes, the cache software that sits inside the service provider network. The second is the cloud component – you can think of it as the control plane – decides how to delegate the traffic into those caches.

Now, these edge cloud nodes can sit on the mobile side, and they can sit on the fixed-line side. What goes into them has to be location specific. If I have a software node that sits on the fixed-line side, it will cache the content that is relevant for that part of the network. If an operator has both mobile and fixed, maybe it'll have nodes that have the mobile formats for those videos on the mobile side, while the equivalent nodes on the fixed-line side will have encoding that is more suitable for the residential devices such as Apple TV or Chromecast. But their function will be very much the same.

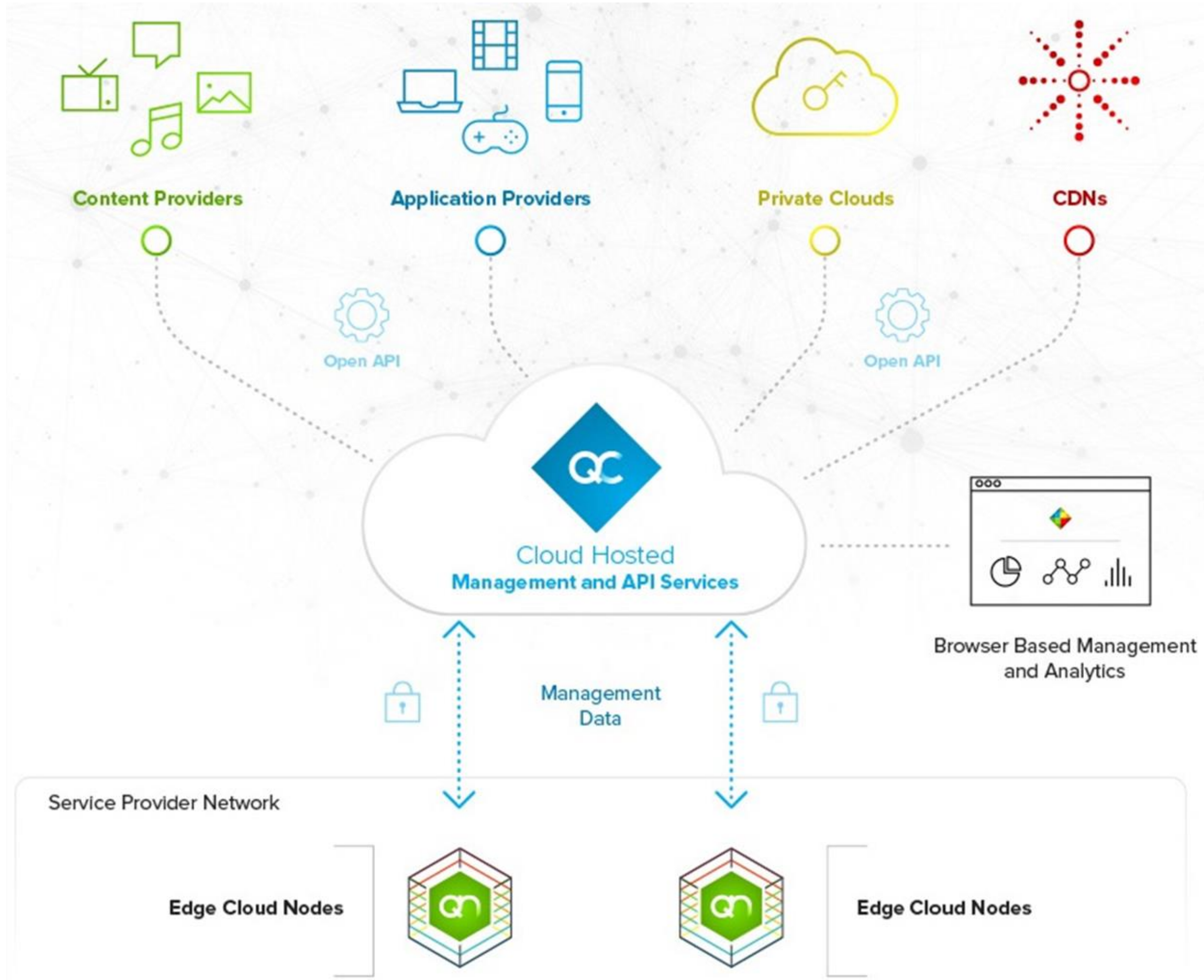
Monica: That's good also for those service providers that have both fixed and mobile. They can manage video and other content using the same platform on both networks.

Dan: The operator will have a single platform it can manage across both mobile and fixed, and this platform will also be able to address the operator's own content. Many operators have a video side to their business, as well as third-party OTT content that comes over the internet.

The vision is that there will be one layer that addresses both of these content sources, and there would be a single resource that can adapt and handle both. If an SP has a big launch of a new series, you can allocate the resource to that. If there's a big live event going out from OTT sources, the resources will shift to that.

Monica: You raise an important issue: both the service provider or a content provider may own the content. As we move more functionality to the edge, how is the relationship between content provider and service provider changing?

Dan: With our solution, we help the service provider become part of the content delivery chain. We do this by creating an API that enables various content providers to make use of these resources. But the owner of the actual



Qwilt open-edge cloud

Source: Qwilt

components – the storage and compute – is the service provider, and the service providers have to decide how to make use of these resources. We give them the tools to do exactly that.

I think the understanding across the board is that consumers expect to have content from a range of sources – not just from the service provider's own offering – and the service providers have to build the network that can handle them all.

Monica: The service providers are in control of how the content is transmitted, but at the same time, they need to work with anybody that provides that content to make sure the content delivery is working as expected.

Dan: Exactly. I wouldn't say take control; I would say they participate in content delivery. Because the current delivery medium is still valuable and is still going to be used. A lot of content is going to be streamed from the centralized cloud platform.

For some things, such as popular content, or content that is very latency sensitive, you can put into play the edge resources that the service provider has deployed. These are network resources that nobody else has, and content providers can decide which content to stream from the centralized cloud, which to delegate to the service provider for delivery at the edge. I think operators could play a much bigger role than they do today in content delivery with edge cloud delivery.

Monica: Within edge computing how much are efforts such as MEC and Fog Computing accelerating this trend?

Dan: There are technical challenges on several fronts. On the mobile side, one is how to get a server function to be inside the mobile network, and MEC is doing a lot of work on that front. If you're deploying content delivery functions on the S1 interface, how are you going to take care of billing? How are you going to get into the GTP tunnels? There is a number of questions. You could think of them as inter-working function questions covering how to place a server inside a mobile network. And I think the MEC group is doing a great job on that front.

There's another set of interfaces that have to be defined: how does a content provider that has a VOD or live offering on the internet make use of these service provider resources? How does a SP publish these resources to the outside world as a function that content providers can leverage? Qwilt has been doing a lot of work on that front under the scope of an industry group, the Streaming Video Alliance, which is creating the set of APIs to enable a content provider or a commercial CDN to use what's known as open caching functions that sit inside the service provider edge cloud.

Monica: You've been working on this before MEC standardization work started. What have you learned so far?

Dan: The ecosystem needs a lot of balancing. You have a situation because some service providers compete to some extent with internet content providers – they both have video services – but there is the common understanding that the consumer is shared between the two. And there's a greater maturity in the industry now: people are figuring out that the content providers, CDNs and service providers have to work together to create an infrastructure that will benefit everybody economically, but also quality-wise. That's something that took many years to build, and when we started out there was little collaboration between the two sides – the service providers and the content providers. It's becoming a lot better now.

Monica: And I think that's crucial, because, as you said in the beginning, monetization is a big issue. Is somebody getting a free ride?

And do you think there is now a balance within the ecosystem as to how much different players benefit and have to commit to financially?

Dan: You see several initiatives that are driving more collaboration, like the Streaming Video Alliance that I mentioned earlier. The TIP project that Facebook is driving is another such initiative. And ultimately you see a far greater collaboration between the two sides that didn't exist before.

Service providers have a lot to offer to the ecosystem, and they bring important assets. They own the network, and the network has capabilities that are unique, that you cannot just get anywhere. If I'm an internet content provider, there's no way I can get my hands on multicast in the access network without a service provider to help me out.

And then on the other side you see a service provider understanding that content providers drive a lot more data into the network, which is good, and ultimately, that's what the consumers consume – a lot more over-the-top content – so their network has to support that.

Monica: What does Qwilt do that is crucially needed and different from everybody else?

Dan: The first fundamental principle at Qwilt was that we are at the edge. We've been deploying software at the edge of the network where compute and storage resources did not previously exist, and we were one of the first companies to do that.

And the edge has a lot of intricacies that are not trivial to solve. How do you manage massively distributed software nodes? How do you squeeze the most performance out of the very limited computing storage you have at the edge, because of the real estate limitations?

The other aspect is the ecosystem that we're building, both technically and commercially, with content providers as well as CDNs. We're ahead of the curve in terms of the APIs that are required to drive this collaboration – how content providers and CDNs can make use of the edge cloud of the open cache function that sits inside it.

Monica: The business case is becoming a hot topic in edge computing. With edge computing, you obviously need to add more infrastructure at the edge, and that comes with a cost. You get better performance, but is this an appealing tradeoff?

Dan: You have to look at who's gaining what from this edge computing model. Every bit that is streamed from the edge inside of a service provider's network is a bit that doesn't have to cross the entire network, and it doesn't have to cross the core of the internet. That means a service provider gains economic efficiency by streaming that bit from the edge instead of coming from a transit or peering location. And there is value in that.

The other value is consumer experience. Because content is streamed from the edge, there's a greater ability to overcome any bottlenecks in the network, and much lower latency as the content is streamed to the consumer. That's a benefit to the service provider as well as to the content provider in terms of the experience their subscribers are getting. If the content provider is striving toward providing a full HD experience all the time without buffering, this is the way to do it.

And this value extends to commercial CDNs that are the platform many content providers use to distribute their content. This gives them better reach in places they cannot reach today, and with far lower latency.

There is definitely value to the entire ecosystem and a question of how each side is going to compensate the other for this value. I think the market will dictate how exactly that is done. But we have found that equation balances itself out, and when we're streaming content from the edge, there are economic benefits across the board for the ecosystem.

Monica: Do you see, in terms of ownership and initial capex investment, an increasing role for the venue owner, the content providers, the CDNs, to participate, because it's something they benefit from along with the service provider? Are they willing to step up and put some investment into it, as well?

Dan: That's something that will fall primarily on the shoulders of venue owners and service provider as they're building their venue network. It's expected of them, because that's they will reap the benefit. And for edge computing, it's like the move from traditional networking into networking that uses storage and compute resources, so it's basically shifting resources from one place to another place.

Both network operators and venue owners are the ones that are going to be responsible for building out that infrastructure, and there are ways for both of them to benefit. Over the long term, that's a far more economically sensible way to build networks than simply to throw routers at the problem.

Monica: What are you focusing your attention on these days to get ready for the challenges, including the challenges of 5G, over the next five years?

Dan: Our focus is on two fronts. One is the technological front: to create standardization around the APIs that are required to enable this open cache function inside the edge cloud, and to manifest that into our product as well, and to have a range of APIs that the content ecosystem can use – increase our capabilities when it comes to service providers' own content.

The second front is on the ecosystem side. We're trying to enhance our content provider and CDN relationships so they can make use of this edge cloud layer that sits deep inside the service provider's network.

About Qwilt

Qwilt's unique Edge Cloud platform and Open Caching software solutions help internet service providers address the dramatic growth of streaming media on their networks and the need for a low latency, high scale infrastructure to support future applications. Qwilt's cloud-managed open platform, running on commodity compute and storage infrastructure and deployed close to consumers, creates a massively distributed Edge Cloud that supports applications such as Open Caching, 4K live streaming, AR, VR, self-driving cars and IoT. This low-latency Edge Cloud architecture enables a high-quality streaming experience for consumers on a massive scale. A growing number of the world's leading cable, telco and mobile service providers rely on Qwilt for Edge Cloud applications. Qwilt is a Founding Member of the Streaming Video Alliance and a leader of the Open Caching industry movement. Founded in 2010 by industry veterans from Cisco and Juniper, Qwilt is backed by Accel Partners, Bessemer Venture Partners, Cisco Ventures, Disrupt-ive, Innovation Endeavors, Marker and Redpoint Ventures. Learn more at www.qwilt.com.

About Dan Sahar

Dan Sahar is the Co-Founder and VP of Product Marketing at Qwilt. Dan drives product marketing and go-to-market activities for Qwilt, bringing more than fifteen years of marketing and product management experience at high technology companies. Prior to co-founding Qwilt, Dan was Director of Product Marketing at Crescendo Networks (F5 Networks), a leading provider of data center application delivery products. At Crescendo Dan was responsible for leading the company's overall marketing and product direction. Earlier in his career Dan held product management roles at Juniper Networks and Kagoor Networks (acquired by Juniper) as well as engineering management positions at Kagoor Networks and Seabridge (Nokia Siemens Networks). Dan holds a Bachelor's degree in Computer Science and Business from Tel Aviv University Magna Cum Laude and an MBA (Marketing) from the Leon Recanati School in Tel Aviv University.

III. Service provider interviews

BT

MEC in the fixed/mobile converged network

A conversation with Mansoor Hanif, Director of Converged Networks and Innovation, BT

Monica Paolini: MEC is more than a tool to lower latency. It is an enabler for services that are specific to a location but are served over multiple access networks, both fixed and mobile. In this context, MEC is a powerful driver that accelerates the convergence of fixed and mobile networks. In this conversation, Mansoor Hanif, Director of Converged Networks and Innovation at BT, tells us how this is shaping the network evolution at BT in the UK.

Mansoor, what is your new role at BT, and what are you working on with regard to MEC?

Mansoor Hanif: Personally, I came from the EE side. I was in charge of the radio networks. Since the beginning of November 2016, I've moved to the BT research team. I'm now looking after the converged networks research lab.

On edge computing as a whole, at BT we are not in the commercial phase at the moment. We are doing a number of proofs of concept looking at the customer experience – and especially looking at the converged customer experience that we can manage through the edge and the business cases around this.

We did a number of proof-of-concept trials last year, mainly in Wembley Stadium. This year, we have a couple of new locations we're looking at for proofs of concept.

Monica: Can you tell us more about these proofs of concept? What kind of applications, and what were the learned lessons?

Mansoor: From the couple of proofs of concept we did last year, we showed the effectiveness of mobile edge computing solutions for reducing latency for video. For enhanced video orchestration in a stadium such as Wembley, we showed that this could be effectively used to replace closed-circuit TV for security uses. There are similar opportunities for many other vertical markets.

We also did a throughput-guidance trial with Akamai, and we showed that there is benefit in transmitting the real-time radio conditions through MEC to the CDN; it helps content delivery partners improve the end-to-end experience for mobile video users.

Those were successful proofs of concept.

This year, we are focusing on a couple of areas. We are having a look at our retail estate to see how we can improve the engagement of our customers when they come to visit our shops. Can we showcase the offers we have and the converged capabilities we have? Can we improve the user experience by localizing some functionality from our IT systems in the central back office to the local shop? We're going to have a look at that from the retail shop attendant and customer perspective

We are also looking into public spaces such as museums, where we can improve the user experience of visitors through edge computing.

For the entertainment field, we're looking into where edge computing can improve the user experience with things like augmented reality. There, we look at local positioning and at low-latency applications, and how we can hook location-based augmented reality applications into our fixed offerings to offer enhanced solutions people can take away and keep with them after they've visited a location.

We're also looking at the scenario in which you have a lot of people visiting a location who are from outside the country or from other operators. Can we have a single solution that manages all the users at a location, but also has intelligent control and enhanced added-value services over the MEC platform?

Monica: When I started working on MEC, video seemed to be the major use case. What I'm hearing from different sources now is that there is a shift. Video is still one of the main use cases, but there is more to MEC than video. Would you agree with that?

Mansoor: Certainly. Anything that requires compute capability to give an effective user experience is a good use case for MEC. Anything that requires a lot of processing power to be able to deliver a good-quality experience. Augmented reality is a good example.

Video, certainly, is a good use case, especially high-definition 360 panoramic video that you want to make available to the user to browse through and choose whatever camera angle he wants.

Rather than sending all of that data over to the user, you can do the computing on the edge platform, delivering a very good quality of service to the user, and not expose your network.

Monica: Mobile edge computing clearly relies on network virtualization. Initially, virtualization efforts were focused on moving everything to a centralized location – the data center, the cloud – to get cost savings and have everything in one place, which is attractive.

With mobile edge computing, the trend is in the opposite direction. There are some things that do not work sufficiently well in a centralized location. There are advantages to moving processing and storage to the edge.

Mansoor: Today, as a converged operator, we would rather speak about the converged edge for fixed and mobile. One of our key objectives is to make sure we effectively implement convergence across fixed and mobile.

If you're talking about the fixed network and the mobile network separately, the edge can mean different things. There are three elements to this, but there's probably some granularity between them.

First, there's the capability you would like to centralize into the core. That's very much about cloud computing and cloud connectivity, where you can effectively centralize. I think that will continue, to some extent, but not to the extent that we have it today.

Second, there is the edge seen from a fixed network, which is pretty much on the distributed exchanges or aggregation points. There are some applications we feel should be decentralized there.

Finally, there's an opportunity now to offer converged mobile and fixed aggregated capability on those aggregation points.

There's also still, we believe, a big opportunity at the edge of the radio network, which is the closest to the customer.

From our perspective, all of that can fit into an interesting business model: you can leverage a single platform that allows you to manage all those applications anywhere between the core, the aggregation point, and the radio edge with a single platform. That makes it very easy to manage and shift applications when you need to, where you need to.

It's not manageable, not possible, to get to a good business model with completely separate platforms for the radio edge, the aggregation points for the fixed edge, and the core.

Two factors are driving things out to the edge. One is that the edge speeds are moving so fast that the kind of capability we need on the edge is getting increasingly difficult to aggregate in the center. Speed at the edge is scaling up so very fast that I don't think it's possible to scale a centralized computing platform to aggregate all of this on the long term

The second factor is that the processing power required to support those edge speeds is such that you will almost certainly have to enhance the processing power at the edge of the network to some extent, so the utopia of virtualizing all of the radio baseband hardware is unlikely to happen everywhere.

If you need to add or upgrade hardware to support the new radio capabilities and speeds, you can also put in some mobile edge computing power and add some intelligence there.

In the UK, we'll probably be doing a lot of indoor installations over the next three to five years. If you are going to do that, then the actual cost of putting in an extra board or two to enable edge computing is affordable. The overall cost of adding the MEC hardware as part of a wider indoor coverage installation is much lower than if you are doing a dedicated MEC rollout.

Monica: You raised a lot of very interesting points that I would like to follow up on. The first one is the fixed/mobile convergence. In fact, it's telling that MEC doesn't stand anymore for Mobile Edge Computing, but for Multiple-access Edge Computing.

What does that mean to you? A mobile network was very different from a fixed network, in terms of what subscribers do. The convergence is not just in terms of technology, but also in terms of what we do over a fixed/mobile network – or a wireless/wireline network.

Mansoor: You said two things there that are slightly different. There's wireless and wireline; in that case, Wi-Fi would be considered wireless. It's mobile and fixed, where Wi-Fi would be considered part of a fixed network. That's generally how operators have seen it in the past.

Now we're trying to converge all that into a single, coherent control point. We'd like to have the ability to improve the customer experience as you move between those. We need to make sure we get the best out of our Wi-Fi investment and our 4G investment, and future 5G, we hope, and fixed line.

Today, the mechanisms available for controlling those are not sufficiently aligned between the fixed standards and the mobile standards. For example, we need to get a single identity, a single authentication. We also need effective anchoring and control mechanisms to allow us to do intelligent traffic steering between the fixed and the mobile.

To this end, we are working with the standards bodies to make sure that, as we move into 5G, we have a lot more ability to intelligently steer the user into the best customer experience over whatever access type is most suitable.

Edge computing plays a role there. We believe that on the network-wide level, some functionality needs to stay in the core network. But some core functions for specific applications could also be localized into a MEC solution where, in a specific environment like a museum or a shopping center, you can put in an anchor MEC-based solution that's hooking into the indoor installation for Wi-Fi and 4G small.

Solutions like MulteFire, or LTE unlicensed, or simply Wi-Fi can be integrated into a single MEC anchor, where we can provide layered services. It's a way of integrating that locally to work together with what's in the core, but also to work independently and provide extra services where needed in those specific locations.

Monica: As an operator, you have to decide which functionality should be centralized and which functionality should be pushed to the edge. And then, for the access network, you have to decide which applications should use the fixed network and which should use the mobile network – and which fixed or mobile network to use, when multiple ones are available.

Mansoor: You put many things on the core side: the unified sign-on, the unified authentication, but especially the traffic steering, the quality of service, and the quality of experience management at the network level.

But you can also have a separate policy locally for a specific location, based on what you agreed with landlords and what they want to offer your customers, and other people's customers. It's important to have the flexibility to tailor this to local requirements.

In the terminology of "Network slicing", for example, this is the capability to offer an enhanced network slice in a localized environment.

Monica: Do network slicing and edge computing go hand in hand, complementing each other, in the scenarios you describe?

Mansoor: Yes. Mobile edge computing adds an extra granularity to the type of slice you can offer. Already, we're experimenting a lot with how far we can push network slicing on our 4G network. It's going to be a lot easier with a 5G network, because 5G is built around network slicing.

MEC increases the granularity of the type of slicing you can offer because you can then actively offer completely different types of slicing locally for any specific customer.

Monica: We talked about the edge, and you mentioned aggregation points and the RAN. Where is the edge?

Mansoor: The edge of a fixed network has been considered to be, let's say, the local exchanges, or the equivalent. The edge from a mobile network is the radio antenna, which is the closest to the customer.

Whether that's a macro site or an indoor site, that's where the edge would be. That's where I think the different definitions of the edge have come in. Obviously, depending on the type of application and the type of reach you'd like to have, you can choose where to put the edge capability.

What's important is that it's also very much dependent on the traffic load and the application load.

Ideally, you could have a dynamic capability for orchestration, where you can move an application from the small indoor cell to the macro cell. Or from the macro cell to the local aggregation point or exchange, back and forth, depending on time of day, the load, and the optimal customer experience you want to give. That's the ideal situation we'd like to get to.

Monica: Is this why you need a single platform that allows you to manage dynamically all those elements?

Mansoor: Yeah, ideally. Otherwise I do not think it's possible to manage the quality of experience we want to offer customers in order to make it a useful business proposition. It would be very, very difficult to do that.

Monica: In terms of the business case, with edge computing, you inevitably have to add some more processing and storage capability at the edge. That's going to come with some cost.

At the same time, it's clear that the more you push to the edge, the better the performance you have in terms of latency. There's a tradeoff there. How much is worth pushing to the edge? When you look at the business case, what are the tradeoffs that you think will make sense? How aggressively do you want to push things to the edge, when that comes at a cost?

Mansoor: It's more about being intelligent about the cost. More than a tradeoff, we need to work on all fronts to lower the cost and increase the added value. First of all, we need to get the hardware platforms that support edge computing to be as low cost as possible.

That's why we're working through initiatives like the Open Compute project and similar ones to get to an almost white box situation for the hardware that supports MEC. We need to lower the cost of the hardware if we want MEC to become deployable on a massive scale.

On top of that, the actual cost of implementing a hardware-based solution needs to be reduced. That's where a standalone business case of rolling out MEC capability into offices and shopping centers doesn't make any sense from our perspective.

However, you can piggyback on indoor installations in an intelligent way so that the increased installation and implementation cost of a MEC solution is only a very small part of your overall cost of the indoor installation.

Timing is going to be critical. We need to catch the wave of large-scale indoor installations at the right moment so we can slot in the MEC hardware, at least in the majority of cases. That would change the business model.

Those are the two things – cost of the hardware platform and timing of installation – that are going to lower the cost of MEC computing on the mobile edge.

We need to make sure that when, for example, you've got a new customer in a big location, the initial investment is covered by simple use cases of connectivity and some basic services. The MEC API interface needs to be flexible enough that, later, we can very easily add on new functionality as and when we need to, on the same platform.

That way, we can continue to generate new revenues on top of the baseline, which is doing the basic financing for the installation. That's why I think MEC is much more than simply improving the efficiency of the network or using low latency.

We need to identify and focus on new services that the combination of proximity and compute power enables, and that we have the flexibility to rapidly implement those solutions on top of the MEC platform as and when customers ask us to.

Monica: It's more than getting lower latency. It's thinking during network deployment where the functionality goes in a much broader perspective.

Mansoor: Absolutely, and being able to dynamically shift content and dynamically deploy new applications locally – leveraging the value that proximity adds by improving the user experience for applications such as augmented reality.

With augmented reality you can be very, very close to the user and therefore really improve the subscriber experience. If you take Pokémon Go, which was a massive hit, it doesn't need a network at all, or very little, but it's not very granular.

If you want to take that to the next level and provide services so compelling that businesses are asking us to put them into their locations, you need to make the customer experience a couple of levels better than that.

That's the kind of thing we're working on so we can offer businesses a compelling way to draw in more customers and have customers pay for more services. We want to enable all of that with a great customer experience that we can monetize, to a certain extent.

Monica: When you talked about monetization, you mentioned subscribers. Could you also get revenues from the content owners or the enterprise which also stand t benefit from edge computing as well?

Mansoor: I don't think we would get paid directly by those third parties. But if we come up with the use cases that generate extra revenue for the landlord or the third parties, then we could effectively have a place in that value chain and get a percentage.

Monica: What about the capex? Could the mall owner, the stadium owner, the airport management be willing to pay for all or some of the infrastructure if it's for services they provide? For instance, if it's a mall and it's trying to get its customers engaged, is this something you think it would pay for?

Mansoor: There are already a lot of landlords ready to pay for a good indoor installation, as long as it's covering all operators and it's offering a good quality of service. People are moving to that situation. We have a number of companies proposing pre-installation in the UK. It's good, really.

If we have a landlord that's looking for basic connectivity, we could offer it a shared-cost MEC platform on top of that indoor installation, which would allow the enablement of many new services within that environment.

The actual capex of that MEC layer on top of a DAS installation would be only a fragment of the DAS installation itself. It could be partially funded by the landlord in some cases.

In my view, if the DAS or the small-cell installation is already being funded by the landlord, it would be reasonable for us to offer to manage a MEC-type service for all operators by putting in the extra capex ourselves and then putting in our added value in terms of use cases we're offering to the local customers.

Monica: We can start with MEC in 4G networks, but with 5G, MEC will be more pervasive and more efficient. How do you see the transition of MEC as we go from 4G to 5G?

Mansoor: It's a very interesting question, because if you are focusing on the latency added-value of MEC alone, then with 5G, MEC's latency advantage will be taken away, because 5G should be inherently capable of very low latency.

On the one hand, you can see how 5G could replace MEC in certain areas. At the same time, to have end-to-end 5G capability, when you're talking about user speeds of 10G or above for one user, it's going to be increasingly difficult to centralize all the computing power needed to aggregate that demand.

Inherently, if we want to adopt 5G massively, we are going to have to use more distributed computing power, simply because the speeds being offered are so high that it's going to be very difficult to keep up with the aggregated capacity requirements if you centralize them.

With 5G, inherently you're going to be looking to distribute the core to some extent.

Also, I don't think 5G will be deployed, necessarily, in a very homogeneous fashion in many networks. It will take a few years. In the meantime, you can offer a reasonably homogeneous quality of service across many, many locations by implementing MEC as an enabler in the first place. It will allow us to homogenize customer experience as we roll out services over a mix of 5G and 4.5G and 4G.

About BT

BT's purpose is to use the power of communications to make a better world. It is one of the world's leading providers of communications services and solutions, serving customers in 180 countries. Its principal activities include the provision of networked IT services globally; local, national and international telecommunications services to its customers for use at home, at work and on the move; broadband, TV and internet products and services; and converged fixed-mobile products and services.BT consists of six customer-facing lines of business: Consumer, EE, Business and Public Sector, Global Services, Wholesale and Ventures, and Openreach. For the year ended 31 March 2016, BT Group's reported revenue was £19,042m with reported profit before taxation of £3,029m. British Telecommunications plc (BT) is a wholly-owned subsidiary of BT Group plc and encompasses virtually all businesses and assets of the BT Group. BT Group plc is listed on stock exchanges in London and New York.

About Mansoor Hanif

Mansoor joined EE in November 2011 and led the technical launch of the 1st 4G network in the UK and was also accountable for the integration of the legacy 2G and 3G Orange and T-mobile networks. Until 2016 he led the team who plan, design, rollout, optimise and operate all EE radio access networks, including Mobile Backhaul and Small Cells, and was accountable for the coverage aspects of EE's Emergency Services over LTE programme. He was also a board member of MBNL (the joint venture of EE with H3G) until 2016. During the acquisition of EE by BT, Mansoor led the EE network Integration team and is currently Director for Converged Networks and Innovation in BT R&D. He is a member of the BT Technology Steering Board and is a board member of the Scottish Innovation Programme.

Verizon Wireless

Better performance in the enterprise with edge computing

A conversation with Matt Montgomery, Director, Wireless Business Group, Verizon Wireless

Monica Paolini: Edge computing improves performance and optimizes resource utilization in many use cases. The enterprise is an environment where edge computing is going to play a large role for a diverse set of use cases, which include not only data and voice connectivity, but also IoT applications. I talked with Matt Montgomery, Director of the Wireless Business Group at Verizon Wireless, about how edge computing addresses the connectivity requirements of the enterprise, while providing the same high level of security as centralized services.

Matt, can you give us an introduction on your role at Verizon and about what Verizon is doing to bring edge computing to the enterprise.

Matt Montgomery: I have business operations, marketing, and partner enablement responsibilities for our Wireless Business Group, which is dedicated to our large and enterprise customer segment.

From a mobile edge computing perspective, my primary responsibility is ensuring that our customers can successfully use multiple partnering solutions. It's not just all Verizon, all Cisco, all Microsoft, or all Apple. It's a combination of multiple technologies, OEMs, and partnerships to create one solution that solves business challenges.

Regarding edge computing, my responsibility is to make sure Verizon provides world-class solutions to customers and that the solutions they choose work seamlessly on the Verizon network. It's an integrated approach with technologies, equipment manufacturers and partners that create the best solution possible to help them move business forward, faster.

Monica: Over the last few years, we have seen a push to move everything to a centralized cloud in large data centers. Now, the tide is turning. Service providers, enterprises, venue owners, and even content providers are showing an interest in moving some functionality to the edge. Why do you think that's so?

Matt: Industry innovation causes a pendulum swing. As new technologies become available, as new threats arise, as the computing experience becomes more intense and form factors become more open, businesses are looking at what they can do locally versus centrally.

I don't believe it's a binary equation, but I do think that what we can do now with the mobile edge is much different from what we could do just a few years ago. The options for customers are opening up.

Customers are moving quickly to leverage both network and application assets much more aggressively. For instance, now they can optimize Wi-Fi to create fast lanes. And they can optimize the applications themselves so they can build a more localized computing experience.

And now they can do this, in some cases, for less money and with more control and more security. They can use analytics more precisely to create higher-performing localized environments for the business, versus a larger, centralized environment where it's very hard to make massive changes without a lot of disruption.

It makes companies more nimble. It creates agility in the delivery of applications. And it can provide, in some cases, a more secure environment.

Monica: In terms of security, do enterprises feel comfortable about moving more of their functionality to their premises?

Matt: I see security as above the distinction between a distributed network with edge computing, and a centralized network. I look at security as its own silo, its own platform.

The answer is, yes, some organizations feel that, if they keep things more local, they can control the security component more easily.

That premise, in my mind, isn't inaccurate, but it is problematic. The security framework needs to be rich and robust whether your network is centralized or everything's at the edge. At Verizon, we believe that our service needs to be highly secure in all cases.

Some customers believe that, along with the performance benefits, mobile edge computing offers that high level of security.

Monica: What kind of applications does the enterprise usually need to put closer to the edge?

Matt: The internet of things is one driver. Data itself is also driving the move to edge computing.

There's so much data generated from network analytics these days that we can't physically get all of it across the network. By doing some of the computing at the edge, we lower the latency. Mobile edge computing offers the lower latency that we need to provide the best experience.

The more mission-critical an application and the more crucial uptime reliability is to the customer, the more benefit edge computing offers.

Edge computing lets us create fast lanes in a highly-optimized Wi-Fi environment and further optimize applications within those fast lanes. We can ensure that the users get a higher quality of service in that environment than they would if we did everything centralized over a giant MPLS network.

Monica: Does that include only data, or voice as well?

Matt: It's voice as well. Voice is now a consideration for many businesses that are looking to move everything onto IP and deliver it on whatever form factor fits the business need.

It could be a smartphone with an integrated dialer that manages your desk phone and your mobile phone. It could be a wireline phone that is integrated with your mobile number. It could be a VaaS, or video as a service, for services such as video conferences. Edge computing could help deliver with less latency.

Monica: New business models may emerge as services become location aware or location based. Do you think the enterprise is willing to pay for at least some of the capex or opex required to deploy local services?

Matt: The model's different, you're absolutely right. Edge computing creates a capex model. Companies have been moving to more of an opex, computing-as-a-service model. Some customers would prefer to move out of an opex model to a capex model because there are tax implications. And edge computing is a more capital-intensive model.

But it's also important to note that companies like Cisco, as an example, have already built into their operating system, from a networking and routing perspective, the ability to do mobile edge computing and create Wi-Fi fast lanes, and the ability to tag and ensure certain applications have a higher quality of service.

Organizations that have invested in this sort of product can pivot and turn on a stronger edge computing experience without a giant checkbook effort.

An example would be on tablets in an edge environment. Apple tablets now come with the ability to do application optimization. They come with the ability, using Wi-Fi, to provide a higher quality of service. As the industry starts to look at the benefits of edge computing versus centralized computing as a service, we're finding that some of the capabilities needed to do it are already in place.

Monica: The ability to integrate Wi-Fi with cellular also should be a priority for the enterprise.

Matt: It's very important that the experience inside the four walls is replicated outside the four walls, at least for the critical applications and the optimization needed to make that happen. And we can do that by keeping applications local. But at the same time, Verizon believes that when you leave those four walls, we need to provide the same level of service for that end user that they get within that environment.

We spend a good deal of time working with Cisco and all of our networking partners, to ensure that if you migrate onto a 4G LTE connection using 4G LTE Advanced, we can provide a similar look and feel of the optimized model you enjoy now across wireless interfaces, and do it locally.

We try to make that experience seamless for the end user, especially when we integrate our VoIP and video services. If they're part of the mobile edge computing experience, we want to replicate that as they move outside of the four walls into the more centralized approach in the wide-area network. These are the capabilities that we provide customers today.

Monica: Again, it's crucial to have the seamless connectivity because, as an end user, you shouldn't need to know what access you are using.

Edge computing allows you to be access-technology neutral, in the sense that you focus on the functionality, and then it doesn't matter what RAT the subscriber uses.

Matt: We have a strong opinion on why this environment is better with Verizon. The last thing we want is for our end customers to have an experience that's different. They may understand they're on cellular, and not on Wi-Fi, but we do not want them to notice the difference. And we feel Verizon is the best at doing this. Our customers really don't know that they've moved into an environment in which their services are being provided to them by the Google or Amazon Cloud versus being serviced locally.

We work hard with our OEM partners to create a networking environment in which you can move in and out and still get access with the same quality and level of service that you get from a mobile edge environment.

Monica: Can you say something about IoT? There's a huge amount of interest, maybe some hype as well. What do you hear from the enterprise?

Matt: IoT is driving companies to edge computing, because of the analytics generated out of an IoT environment, especially in the industrial internet space. Some of our customers are moving more diligently, more pervasively, into sensors and monitoring, and using that data to make better decisions.

It's problematic to transmit the data across traditional networks. Mobile edge computing becomes almost a requirement because we have to keep the data local. We can mine the data and only move the high-level analytics data across a centralized approach.

We find that IoT is driving a mobile edge computing experience for data collecting, for the processing of data, especially in certain industries.

Think about what GE's doing, adding sensors to machines. They're making what I call dumb machines into smart machines so they can gather the data and make better decisions.

This is driving mobile edge computing. This level of need has been one of the bigger catalysts for edge computing.

Monica: Are there any applications or any specific verticals that are ahead in the move to IoT?

Matt: We expect the manufacturing vertical to pick up quickly. We're seeing some in the utilities vertical, which would include energy.

Next would be transportation – not just transportation and shipping, but also receiving. It's moving in and out of those four walls, tracking and monitoring all inventory, and reporting in near real-time.

In healthcare too, where we can help monitor medicine shipments that need to remain below a certain temperature, and where highly restricted pharmaceuticals like OxyContin need to be constantly managed to help avoid theft or counterfeit.

These applications are driving mobile edge computing because of the analytics generated and its use in near real-time decision-making.

Monica: You've talked a lot about analytics. That's an interesting part because, oftentimes, we think about edge computing as more secure and residing more on the content side, but also it allows you to manage your network resources better, to optimize them better.

Do you think a lot of the analytics is also moving to the edge because, as you say, it's much more efficient? You're trying to optimize so you have all the data there. No point sending it all the way back to the core.

Matt: Our customers have not completely run to an edge computing mode. It's a hybrid approach.

They still have access to cloud assets that are not necessarily in a centralized architecture. These aren't going away.

What's going away from the centralized cloud is the massive amount of data being generated locally. We are consuming data locally and evaluating it locally for decision-making. We're transmitting much smaller data sets up into the cloud.

The cloud environment hasn't dried up or disappeared. It's still there, but we're leveraging edge computing so we don't have to move all of that data up into the cloud to make it happen.

Monica: Has it made it cost effective for Verizon, as well, because you're going to have less cost for transport and backhaul?

Matt: Right. The telecommunications capabilities here are remarkable. The ability to create data sets and analyze the data, it's even outpaced that. What I'm suggesting is that IoT, especially, but mobile computing in general has created such a large set of data that it has outpaced our ability to move it all into a cloud.

But it hasn't outpaced our ability to secure it over the network during transport. Analyzing it locally and then moving what's needed to the cloud is really the process that we see starting to happen.

Monica: One final question: As you look at the next five years, what do you think will change in mobile edge computing? What new challenges are you going to try to address?

Matt: My crystal ball tells me that you're starting to see network technologies that may swing the pendulum back to a more centralized approach.

What's not going to stop is IoT and the ability to mine data for insights into what it can offer organizations. This is not stopping, it's accelerating. The volume of data is growing and all of this will only continue exponentially.

Also, standards are finally coming into place, both for mobile edge computing and for 5G.

As our backhaul is much stronger with fiber, we're connecting smart cities and smart businesses with much higher-performance networks. And we'll be using 5G, which has very low latency with incredible speeds. As a result, we open up the ability to move more of that data back to a centralized location.

I see the idea of a hybrid IT or an organizational model that meets exactly what each business needs. If you're a movie studio, you're going to do more edge computing because you're going to deal with video that sits within the studio itself. You're not going to move it, because you need to take action on it there.

If you're a manufacturer, you may need a large capital investment to deploy all the edge computing you need to leverage some of the generalized IT environments.

Because of these points, with 5G I have the pipes that are secure enough and have the low-latency capability to move data to whatever environment I need it in and to service those applications.

These are disruptive technologies. 5G will disrupt mobile edge computing. Mobile edge won't go away, but its trajectory might be disrupted because of the new things we can do as we move forward.

About Verizon Wireless

Verizon helps organizations achieve better business outcomes and drive better customer experiences, simply, securely and reliably. With our investments in superior technology like LTE advanced, America's largest and fastest 4G LTE ever, we deliver innovative solutions like mobility, IoT, cloud, security and telematics that can help you connect people, places and things around the world.

About Matt Montgomery

Matt Montgomery is the director of marketing for Verizon's Wireless Business Group.

Glossary

3GPP	Third Generation Partnership Project
ABR	Adaptive bitrate [streaming]
API	Application programming interface
ATCA	Advanced Telecommunications Computing Architecture
AWS	Amazon Web Services
BBU	Baseband unit
B-RAS	Broadband remote access server
BTS	Base transceiver station
CBRS	Citizens Broadband Radio Service
CDN	Content delivery network
CMTS	Cable modem termination system
CORD	Central Office Re-Architected as a Datacenter
COTS	Commercial off-the-shelf [hardware]
CPE	Customer premises equipment
CPRI	Common public radio interface
CPU	Central processing unit
C-RAN	Cloud RAN
DAS	Distributed antenna system
DDoS	Distributed denial of service
DNS	Domain name system
DPI	Deep packet inspection
eNodeB	Evolved NodeB
EPC	Evolved Packet Core
Eth	Ethernet
ETSI	European Telecommunications Standards Institute
FDD	Frequency division duplex
FLIPS	Flexible IP-based Services
FPGA	Field-programmable gate array
GGSN	Gateway GPRS support node
GPRS	General Packet Radio Service
GPU	Graphics processing unit
GTP	GPRS Tunneling Protocol
GW	Gateway
HD	High definition
HeNB	Home eNB
HSS	Home subscriber server
HTTPS	Hypertext Transfer Protocol Secure
I/O	Input/output
ICN	Information-centric networking
IMT	International Mobile Telecommunications
IoT	Internet of things
IP	Internet Protocol
IPsec	Internet Protocol security
IT	Information technology
L1	[OSI] layer 1
L2	[OSI] layer 2
L3	[OSI] layer 3
LTE	Long Term Evolution
ME	Mobile edge
MEC	Multiple-access Edge Computing
MIMO	Multiple input, multiple output
MME	Mobility management entity
MVNO	Mobile virtual network operator
NAP	Network access point
NEBS	Network Equipment-Building System
NFV	Network Functions Virtualization
NFVI	NFV infrastructure
NGMN	Next Generation Mobile Networks [Alliance]
NR	New radio
NSF	National Science Foundation
OAM	Operations, administration and maintenance
OCP	Open Compute Project
OCR	Optical character recognition
OEM	Original equipment manufacturer
OTT	Over the top
OVP	Open Virtualization Profile
PAWR	Platforms for Advanced Wireless Research
PGW	Packet gateway
PoC	Proof of concept
PoE	Power over Ethernet
PTN	Public telephone network
QoE	Quality of experience
RAM	Random access memory
RAN	Radio access network
RAT	Radio access technology
RAU	Radio aggregation unit
RNC	Radio network controller
RNIS	Radio network information service
ROI	Return on investment
RRH	Remote radio head

SDK	Software development kit
SDN	Software-defined networking
SD-WAN	Software-defined wide area network
SGW	Serving gateway
SP	Service Provider
STB	Set-top box
TCO	Total cost of operation
TCP	Transmission Control Protocol
TDD	Time-division duplex
TG	Throughput guidance
TGE	TG entity
TIP	Telecom Infra Project
TLS	Transport Layer Security
UE	User equipment
UHD	Ultra-high-definition television
URLLC	Ultra-reliable low-latency communications
vBRAS	Virtual BRAS
vCPE	Virtual CPE
vEPC	Virtual Evolved Packet Core
VM	Virtual machine
VNF	Virtualized network function
VoD	Video on demand
VoLTE	Voice over LTE
VPN	Virtual Private Network
vRAN	Virtual RAN

References

[1] 5G Americas, Understanding information centric networking and Mobile Edge Computing, 2016.

[2] 5G Public-Private-Partnership, The 5G Infrastructure Public Private Partnership: The next generation of communication networks and services, 2015.

[3] 5G Vision – The 5G Infrastructure Public Private Partnership: The next generation of communication networks and services, Release 5.1, 5G Public-Private-Partnership, 2015.

[4] Beck, Michael Till, Martin Werner, Sebastian Feld, and Thomas Schimper, Mobile Edge Computing: A taxonomy, AFIN 2014.

[5] Bhardwaj, Ketan, Ming-Wei Shih, Pragya Agarwal, Ada Gavrilovska, Taesoo Kim, and Karsten Schwan, Fast, scalable and secure onloading of edge functions using AirBox, Georgia Institute of Technology, 2016.

[6] Chao Hu, Yun, Milan Patel, Dario Sabella, Nurit Sprecher, and Valerie Young, Mobile Edge Computing: A key technology towards 5G, White paper 11, ETSI, 2015.

[7] Cisco, Cisco Visual Networking Index: Global Mobile Data. Traffic Forecast Update, 2016–2021, 2017.

[8] Ericsson, Ericsson Mobility Report, 2016

[9] ETSI, Mobile Edge Computing (MEC) Terminology, ETSI GS MEC 001 V1.1.1, 2016.

[10] ETSI, Mobile Edge Computing (MEC): Technical requirements, ETSI GS MEC 002 V1.1.1, 2016.

[11] ETSI, Mobile-Edge Computing (MEC): Service scenarios, ETSI GS MEC-IEG 004 V1.1.1, 2015.

[12] ETSI, Mobile-Edge Computing (MEC): Proof of concept framework, ETSI GS MEC-IEG 005 V1.1.1, 2015.

[13] ETSI, Mobile Edge Computing: Market acceleration: MEC metrics best practice and guidelines, ETSI GS MEC-IEG 006 V1.1.1, 2017.

[14] ETSI, Mobile-Edge Computing, 2014.

[15] Friedman, Thomas L., Thank You for Being Late: An Optimist's Guide to Thriving in the Age of Accelerations, 2016

[16] I, Chih-Lin, Corbett Rowell, Shuangfeng Han, Zhikun Xu, Gang Li, and Zhengang Pan, Toward green and soft: A 5G perspective, IEEE Communications Magazine, 2014.

[17] I, Chih-Lin, Shuangfeng Han, Zhikun Xu, Qi Sun, and Zhengang Pan, 5G: Rethink mobile communications for 2020+, Philosophical Transactions of the Royal Society A, 2016.

[18] Intel, Real-world impact of Mobile Edge Computing (MEC), 2016.

[19] InterDigital, What the MEC? An architecture for 5G.

[20] International Telecommunication Union, IMT vision – Framework and overall objectives of the future development of IMT for 2020 and beyond, Recommendation ITU-R M.2083-0, 2015.

[21] International Telecommunication Union, IMT vision: Framework and overall objectives of the future development of IMT for 2020 and beyond, Recommendation ITU-R M.2083-0, 2015.

[22] Marek, Peter, Delivering elasticity to the mobile edge, Embedded Innovator, 2016.

[23] Miller, Linsey, How can operators make money with MEC? Artesyn.com Newsroom, 2016.

[24] Paolini, Monica, Charting the path to RAN virtualization: C-RAN, fronthaul and HetNets, Senza Fili, 2015.

[25] Paolini, Monica, Massively densified networks: Why we need them and how we can build them, Senza Fili, 2016.

[26] Paolini, Monica, The smart RAN: Trends in the optimization of spectrum and network resource utilization, Senza Fili, 2015.

[27] Qwilt, Open Edge Cloud business case, 2016.

[28] Reister, John, It's MEC to the rescue for struggling mobile video performance, Telecoms.com, 2017

[29] Vasona Networks, Are mobile networks ready for the next gaming craze? Vasona Networks, 2016

Further resources

Latest reports in this series:

Improving latency and capacity in transport for C-RAN and 5G. Trends in backhaul, fronthaul, xhaul and mmW

Massively densified networks. Why we need them and how we can build them

Voice comes to the fore, again. VoLTE and Wi-Fi Calling redefine voice

Getting the best QoE: Trends in traffic management and mobile core optimization

The smart RAN. Trends in the optimization of spectrum and network resource utilization

Charting the path to RAN virtualization: C-RAN, fronthaul and HetNets

LTE unlicensed and Wi-Fi: moving beyond coexistence

Watch the video of the interviews

About RCR Wireless News

Since 1982, RCR Wireless News has been providing wireless and mobile industry news, insights, and analysis to industry and enterprise professionals, decision makers, policy makers, analysts and investors. Our mission is to connect, globally and locally, mobile technology professionals and companies online, in person, in print and now on video. Our dedication to editorial excellence coupled with one of the industry's most comprehensive industry databases and digital networks leads readers and advertisers to consistently choose RCR Wireless News over other industry publications.

About Senza Fili

Senza Fili provides advisory support on wireless data technologies and services. At Senza Fili we have in-depth expertise in financial modelling, market forecasts and research, white paper preparation, business plan support, RFP preparation and management, due diligence, and training. Our client base is international and spans the entire value chain: clients include wireline, fixed wireless and mobile operators, enterprises and other vertical players, vendors, system integrators, investors, regulators, and industry associations. We provide a bridge between technologies and services, helping our clients assess established and emerging technologies, leverage these technologies to support new or existing services, and build solid, profitable business models. Independent advice, a strong quantitative orientation, and an international perspective are the hallmarks of our work. For additional information, visit www.senzafiliconsulting.com or contact us at info@senzafiliconsulting.com or +1 425 657 4991.

About the author

Monica Paolini, PhD, is the founder and president of Senza Fili. She is an expert in wireless technologies and has helped clients worldwide to understand new technologies and customer requirements, create and assess financial TCO and ROI models, evaluate business plan opportunities, market their services and products, and estimate the market size and revenue opportunity of new and established wireless technologies. She frequently gives presentations at conferences, and writes reports, blog entries and articles on wireless technologies and services, covering end-to-end mobile networks, the operator, enterprise and IoT markets. She has a PhD in cognitive science from the University of California, San Diego (US), an MBA from the University of Oxford (UK), and a BA/MA in philosophy from the University of Bologna (Italy). You can reach her at monica.paolini@senzafiliconsulting.com.