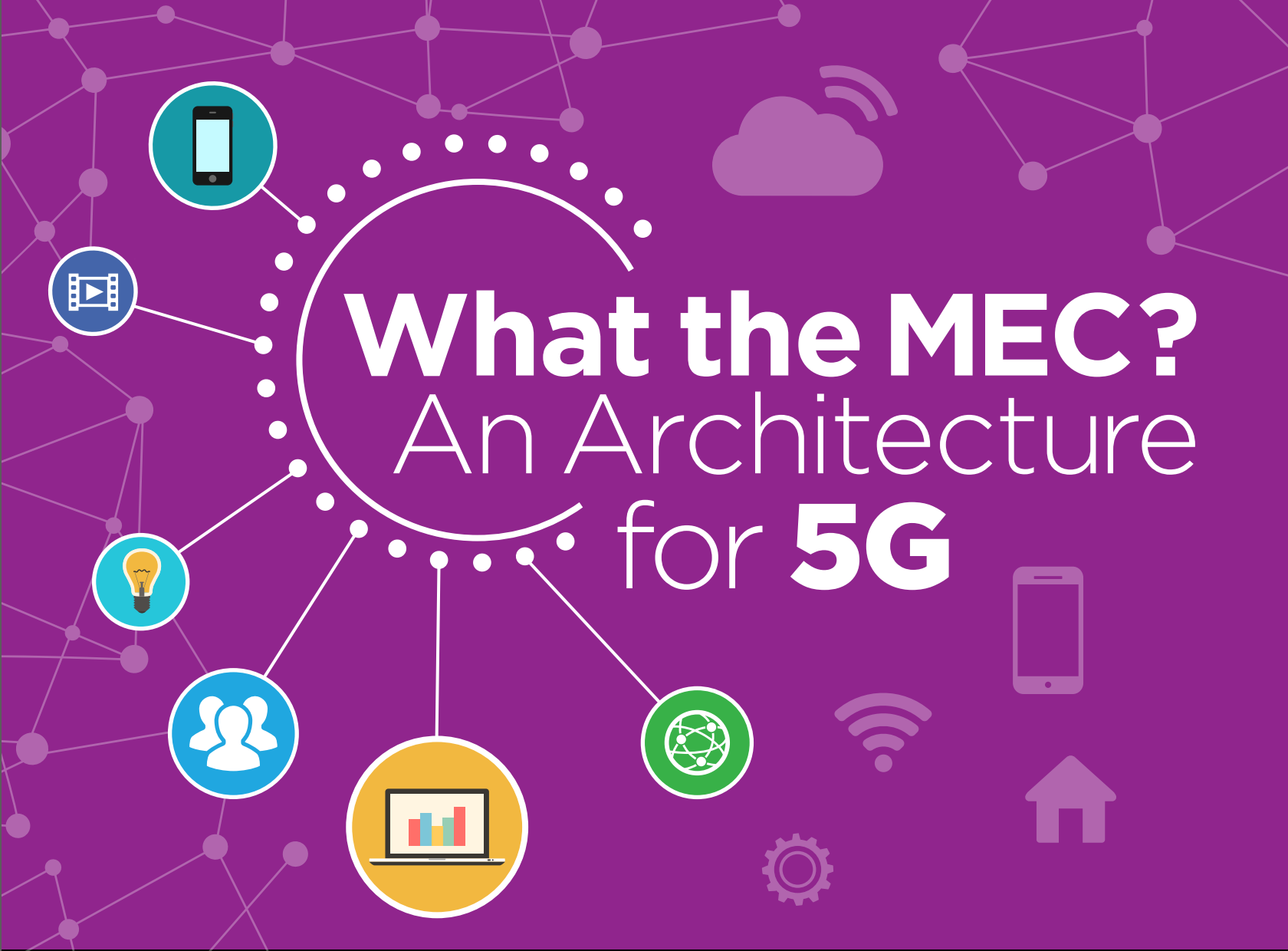
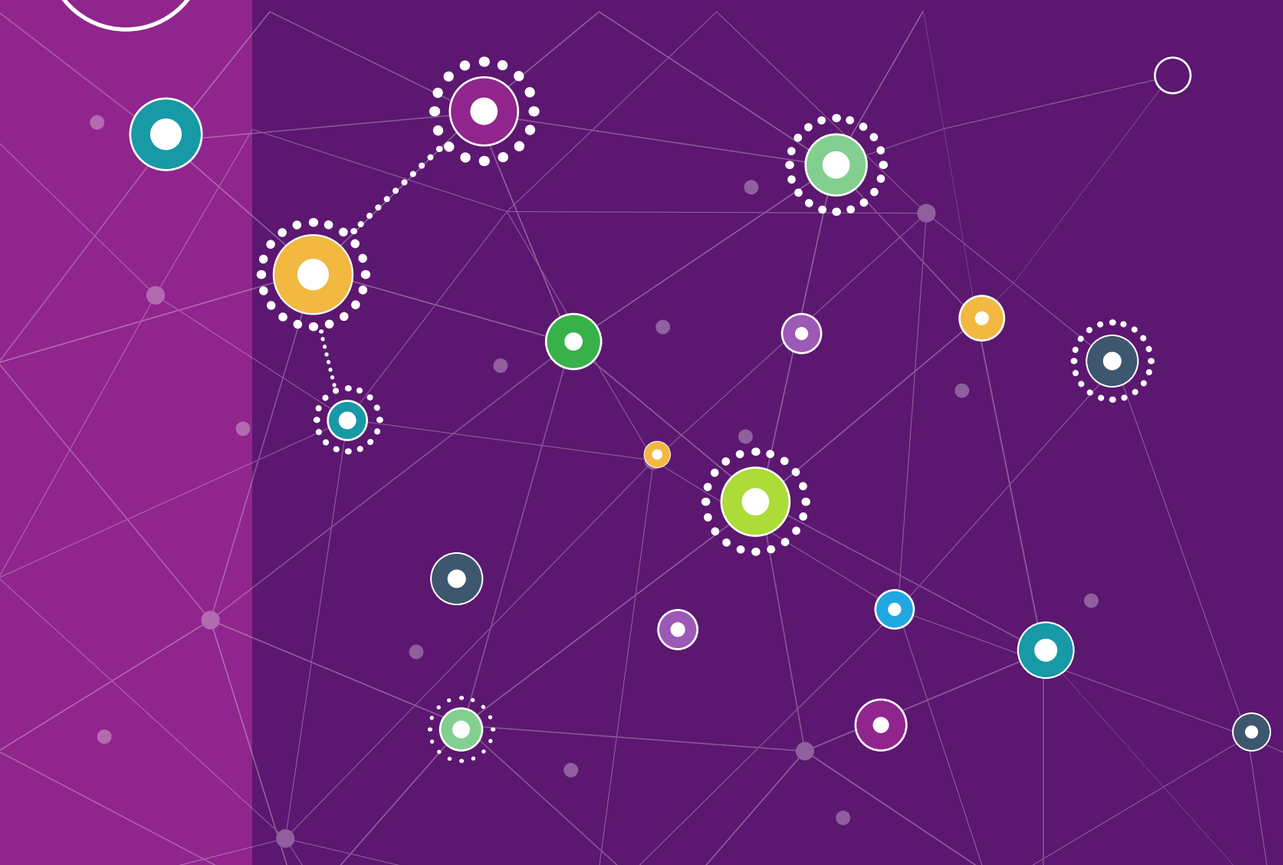


InterDigital®

What the MEC? An architecture for 5G

As we stated in the first e-book in this series, the evolution from LTE to 5G will be the most profound transformation on the wireless industry since the transition from analog to digital. There are many reasons for this, primarily because new uses for wireless technology are creating demand for new approaches to connectivity, bandwidth and network architecture. The evolution toward 5G will bring about several new ways of designing networks so that the promise of always-on, high-bandwidth, low latency, massive networks can become reality. The concept of Mobile Edge Computing (MEC) is one such evolution. MEC is a foundational network architecture concept, which will help 5G networks live up to their potential as “living networks”, while delivering significant capability gains required for IoT, enhanced mobile broadband, virtual reality, self-driving vehicles, and many other applications. This e-book — the second of a three-part series — will review key aspects of MEC architecture and answer some common questions about considerations for the future.

Q: What is MEC?



It’s necessary to establish a common definition of MEC for the purposes of this e-book. It may be helpful to think of MEC as a cloud services environment for a Radio Access Network (RAN). MEC turns a cell/base station into a hub, which dramatically improves network performance and user experience. Certain network computing functions that formerly existed only in the core network now move far out to the network’s edge, closer to the user, where they help achieve these gains. By moving certain network services and functions out of the core network, we achieve significant savings in cost, latency and round trip time (RTT), traffic, download time, physical security, and caching efficiency. It’s a bit like a human nervous system: our reflex arcs help the body respond to things like pain stimulus by creating a shorter neural pathway than the one going all the way to the brain. Not only is this more efficient from a signal processing standpoint, it helps protect the body from harm. If you touch a hot stove, your hand jerks away from it before the pain signal even reaches your brain. MEC works this way by decentralizing certain network functions in order to make the entire network more capable and efficient.

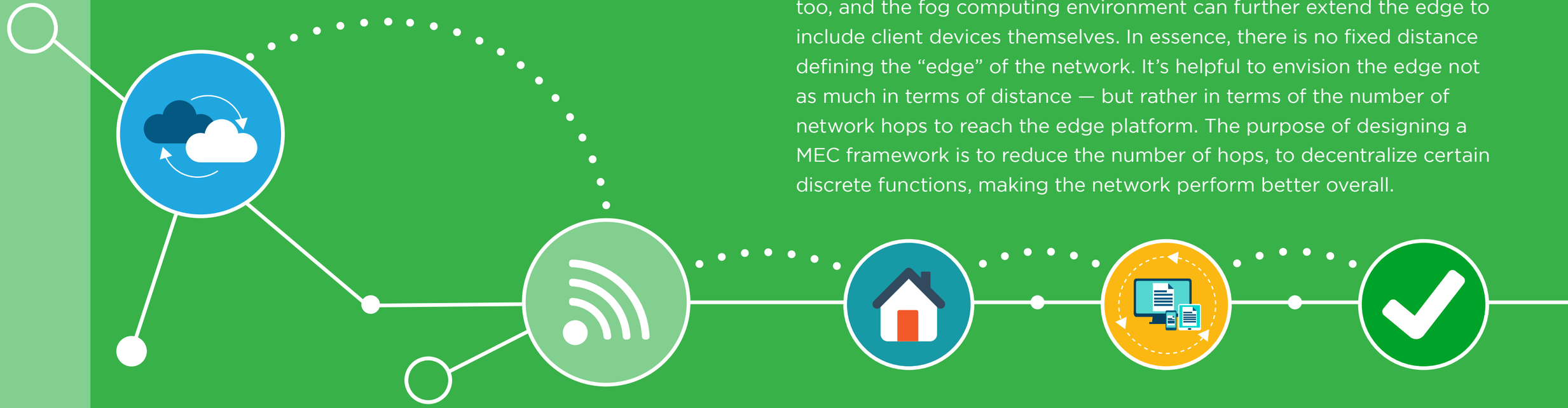
Q:

What are use cases for MEC?

There are several use cases that will benefit greatly from pushing MEC toward the far edges of the network, into small cells, Wi-Fi access points, media gateways, and even extending edge computing to user devices themselves. Benefits of the far edge include ultra-low latency, traffic optimization, agility and adaptability, and context awareness. Tremendous benefits from MEC architecture will be seen in the fields of connected gaming, cognitive assistance technologies like remote/assisted surgery and tactile internet browsing, autonomous vehicles and drones, industry automation, and multimedia content delivery.

Q: In network terms, how close to the end user is the “edge?”

The edge cannot be defined by distance alone. Simply, a Wi-Fi access point or a base station (such as eNodeB, in a LTE context) can be viewed as the edge. However, cloud resources can be deployed at several levels, comprising a continuum from distinct cloud centers down to individual points of access (such as base stations, small cells, and other access points). Edge computing and fog computing are interrelated in some ways too, and the fog computing environment can further extend the edge to include client devices themselves. In essence, there is no fixed distance defining the “edge” of the network. It’s helpful to envision the edge not as much in terms of distance — but rather in terms of the number of network hops to reach the edge platform. The purpose of designing a MEC framework is to reduce the number of hops, to decentralize certain discrete functions, making the network perform better overall.



Q: What is the actual difference between Fog computing and MEC?

In short, the two are closely related. Both describe standards-based architectures where computing, storage, and networking resources are made available in a cloud fashion on host servers located at the network edge. The main difference is in the context in which they are referenced. Fog computing is often mentioned in the context of IoT, where host servers are typically routers, access points or even computing devices co-located with sensors and actuators. Mobile Edge Computing is often mentioned in the context of mobile networks, where host servers are integrated with the mobile network infrastructure, such as base stations or aggregation sites. Standard frameworks for these designs are being driven by international consortiums such as the European Telecommunications Standards Institute (ETSI) and the International Telecoms Union (ITU).

Q: It has been said that MEC is cloud for the RAN, but is it really “cloud”, or is it NFV?

MEC deploys cloud resources within the RAN, which may be utilized to realize either operator services or third party applications at the mobile edge. Network Function Virtualization (NFV) relates to implementing network functionality (i.e. Virtualized Network Functions – VNF) on commodity servers instead of on dedicated proprietary hardware platforms. Both ETI MEC and ESTI NFV standards define a virtualized infrastructure layer and a virtual function manager for their respective platforms. As such, there is overlap between these two approaches with several alternatives to reconcile them. In one approach, an NFV platform (composed of NFV infrastructure, NFVI, and a VNF manager, VNFM) may be utilized to realize MEC applications. In an alternative approach, the MEC platform (composed of a Virtual Infrastructure Manager, VIM, and a Mobile Edge Platform Manager, MEPM) could be utilized to realize VNF. The selection in approaches may be determined by the level of NFV deployment within a network.

Q: How does MEC differ from Cloud-RAN?

Cloud-RAN (C-RAN) and MEC are not directly related to each other. As mentioned above, MEC provides computing, storage, and networking resources, exposed in a cloud fashion, at the edge of the mobile network — primarily in base stations and aggregation sites. MEC resources may be used internally within the mobile network to realize and optimize operator services, or exposed (as cloud resources) to third-party applications.

C-RAN, on the other hand, is an architecture optimization to the mobile network. In this paradigm, RAN functionality, instead of being distributed in the base stations, is implemented in centralized data center resources, shared among several cell sites. As such, C-RAN focuses on virtualizing RAN functions and realizing them in the cloud.

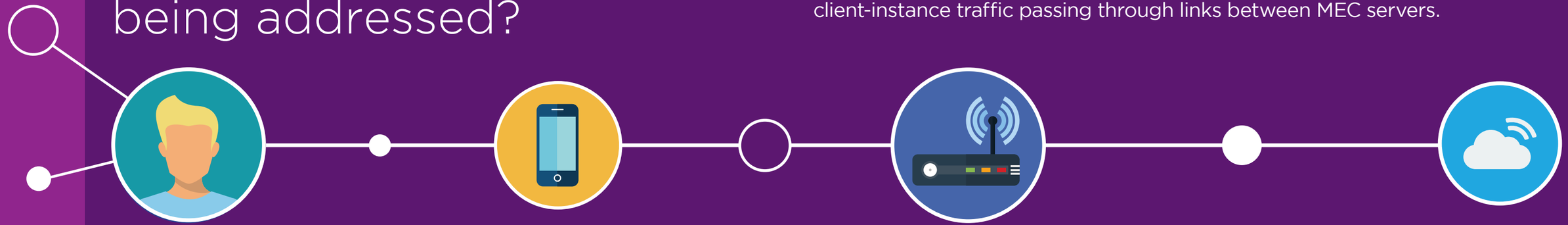
While MEC and C-RAN are not directly related to each other, the C-RAN function theoretically may be deployed over any cloud service. However, to make that possible, the front-haul transport between the remote radio units (RRU) at the cell sites and the virtualized baseband processing units (BBUs) in the centralized data centers would require high throughput and low latency connections. Achieving this kind of transport requires that the centralized BBUs to be placed within close proximity to the RRUs. With this perspective, it's conceivable that C-RAN could be implemented at the edge as a way of augmenting the benefits of MEC.

Q: How does MEC increase bandwidth relative to regular non-MEC architecture?

In a technical sense, MEC does not actually increase bandwidth. Rather, it facilitates a more efficient use of the network, conserving bandwidth usage compared with non-MEC architectures. For example, a MEC application may process user data at the edge. Network bandwidth is saved at the backhaul, because the user data is not required to be transported to the core network and internet. This way, MEC may allow more applications to be run over networks without capacity increases.

Q: In order to allow the content on the MEC to follow individual users, very high bandwidth connectivity between the MECs will be needed. How is this mobility issue being addressed?

In part, this is a standards issue. Mobility is currently being addressed by the ETSI MEC initiative. When considering mobility, different scenarios, and application types have to be considered, which result in very different requirements for bandwidth capacity between different MEC hosts. Scenarios include a client device moving to a cell location served by the same mobile edge host, a different edge host, or even no host at all. Mobile edge applications may or may not be sensitive to client mobility. For example, a stateless edge application may require no transition of user context from one edge host to another. Alternatively, some applications may require state, data, or instance relocation within the mobile edge system or even towards the distant cloud (in cases where locations have no edge resources). We can't address all cases here. However, as one example, transferring application state across edge hosts to a position closer to the user is one method to avoid client-instance traffic passing through links between MEC servers.



Q: What are some security issues, challenges, or enhancements associated with MEC?

Some security challenges are related to multi-tenancy: preventing improper resource usage and information leakage or breach requires proper isolation between tenants of the edge-computing environment. When MEC is integrated with mobile networks, applications may gain insight to subscriber information, which could raise privacy concerns. Furthermore, MEC applications run inside the operator's domain, and can potentially affect the service to all subscribers, either directly or indirectly, if security was compromised. For this reason, Mobile Network Operators will likely need to maintain tight control over applications allowed to run on MEC servers. Additionally, there are privacy concerns relating to private data managed by edge applications. Fog computing shares similar security and privacy challenges with MEC, while having a few of its own as well. For instance, establishing trust between devices involved in edge computing, especially since different devices may not share the same hardware roots of trust. This is just one more example where a standards-based approach is important to overall network effectiveness, efficiency, and security.

Q: Is MEC's value related to a specific Radio Access Technology (RAT)? How will this change as we move further into the 5G paradigm?

MEC is being designed and developed independent of the underlying RATs. This means that MEC can be deployed over LTE or Wi-Fi network infrastructure, as much as it will be able to be deployed over new radio infrastructures as they evolve. A new Radio Network Information Service (RNIS) API, which is a service component within the ETSI MEC initiative, will be defined to allow a MEC platform to consume information related to RAT and provide radio-related services toward user applications. For example, a MEC application will be able to use such information to compute throughput guidance for video streams.

Q: How do you quantify additional efficiency in the RAN (or eNodeB) that results from the content being served to users more efficiently from the edge?

As with many things, it depends. In this case, efficiency gains by serving a user from the edge instead of the core will depend in large part upon the application context. For example, edge resources may be utilized for video or media service optimization. In this application context, efficiency gains that the edge provides may include application latency reduction (potentially as much as 10x to 20x reduction), by facilitating one-hop or near one-hop access to content. In another example, an operator may realize significant backhaul network traffic savings via transparent multicasting at the network edge for popular or live content.

Q: How useful is radio network information to applications?

Radio channel conditions are quite dynamic, and resource utilization changes rapidly. The reaction/response time of an application is specific to each application, and can typically translate into a requirement involving a data rate over an application-specific time window. For example, a streaming video application may afford to maintain a few seconds of buffer on the client side, and should only care about the average throughput over its time window comparable to its buffering capacity. Alternatively, a data analytics application may be able to adapt its behavior over a longer time period to save energy. In another example, a Vehicle-to-Infrastructure application will more aggressively leverage all available capacity to communicate safety-critical information. In a media data rate example, an application will have to make a tradeoff between keeping a lower constant rate or dynamically adapting its rate with the risk of upsetting user experience. Applications will therefore adapt differently to variations in radio network information, depending on their specific challenges and requirements.


INTERDIGITAL®
200 Bellevue Parkway, Suite 300
Wilmington, DE 19809
+1 (302) 281-3600 | www.interdigital.com