White paper
Cisco public

Edge Computing and Global Mobile Services Infrastructure

Why wait for 5G to mature when you can start now?

Converging mobile strategies

For mobile operators, edge computing, cloud computing and 5G are parallel strategy threads that interrelate. Operators should begin or accelerate edge computing initiatives rather than waiting for 5G to mature. One example is the initial MobiledgeX deployment on Cisco technology, which powers the Deutsche Telekom infrastructure in Germany. This initiative demonstrates the practicality of an edge adaptation of 4G in a way that evolves naturally into 5G as that technology matures.

Contents

An invigorating pace of innovation

In the last decade, IT has transformed dramatically and radically because of the concurrent introduction and rapid acceptance of smart phones, cloud computing, and 4G/LTE cellular broadband. The introduction of the Apple iPhone offered a mobile computer for personal and business tasks. Amazon Web Services introduced on-demand, pay-for-what-you-use computing services. Global cellular broadband became a possibility with 4G and LTE. Ten years later, many aspects of our lives have been transformed by these independent but highly synergistic innovations.

The pace of innovation has been invigorating, but exhausting. Now that everyone has a better idea how everything fits together and what it can do, shortcomings and limitations have become more obvious. For example, smartphones aren't the ultimate mobile device. Wireless augmented reality, the Internet of Things (IoT), wearable devices, and autonomous systems are coming. As impressive as 4G/LTE is, it needs to be improved to keep up. The cloud can evolve and be improved with a continuing stream of specialized server instances for AI, machine learning, inference, and gaming, for example. Improving the communications links between cloud servers and users or devices is a continuing effort.

The view from the mobile edge

Mapping the cloud and edge discussions for mobile operators requires some translation. The mobile network is different from the Internet both in technical architecture and business model. If you consider the history, the difference isn't surprising. Modern mobile systems are the direct descendants of telephony systems that were critical to national infrastructures. They were carefully designed and built for the long term and optimized for voice telephony. Telephony systems and their use were managed in ways that were consistent with the critical role they played. They are purpose engineered and built optimally for completing telephone calls.

In many ways, the Internet is the opposite: it was designed to be open and easy to attach to and use. It wasn't optimized for any known or anticipated usage. Instead the Internet has adapted to support unimagined uses that far exceeded any initial expectations.

Telephony and Internet technical architectures also are different. The Internet is constructed with autonomous devices such as routers and switches. Once configured and connected, they operate independently. They coordinate by sharing routing messages with neighboring devices without a primary controller.

Compared to the Internet, the global mobile infrastructure is highly computerized. It's computerized locally with radio management and business management of the subscriber. Globally, it's computerized to enable the coordination and resource sharing that supports a subscriber's roaming beyond their carrier's network bounds.

The business models are also different. Everyone that uses a cellular provider is a known subscriber. Operators define the services a subscriber can use, and all participating operators get paid for any services used. Cellular access along with access to compute resources at the edge will be a key differentiator for service providers in the future.

The Internet is open, anonymous and "neutral," with no universal way of paying for a premium service, even if you wanted to. Over time, it converged to a single protocol to improve total cost of ownership (TCO) and access has become more commoditized.

Marrying cellular and the Internet offers challenges and opportunities. Technology paradigms, culture, processes, and business models all must be aligned or the compromises of today won't be able to meet tomorrow's expectations.

Cadence, generations, and timing

Historically the cadence of evolution and innovation in the mobile industry has been different from the cloud. The Internet was designed to be application-agnostic. Complete network solutions are built on top of transport networks, not optimized within them. As a result, new network applications are the rule, not the exception. Applications appear when the technology and demand is sufficient; they aren't planned and anticipated years in advance. No one could have imagined real-time video being run over the Internet when it was first designed, and yet it has been implemented. In contrast, the mobile infrastructure has evolved in large, visible "generations." If the mobile industry wants to keep up with the cloud in the next decade, it will have to incorporate on-going innovation. Cloud innovation isn't going to slow down and wait for the next generation if a solution can be built over the top instead.

A dissonance of cadences exists between mobile and cloud. It seems logical that mobile operators will have to accommodate and become more agile, rather than expecting and hoping cloud evolution is going to slow down and become more generational. Changing corporate cadence is difficult because it's designed into strategy, development, and deployment practices and processes. At the same time, there's enough at stake to motivate that change.

Like cloud and internet companies, mobile operators need to actively take part in the discovery of the next wave of growth that is beginning to germinate. They need to use existing assets and capabilities as much as possible until they can determine the impact of 5G. Operators should invest intelligently now and scale new opportunities with discovered business models as 5G industrializes in the market. Otherwise, 5G may be industrialized around cloud business models.

On the positive side, mobile operators have already been including cloud in their business and technical infrastructure. They are starting to transition business and operational systems to run on hyper-scale cloud services such as AWS and Azure. They are moving network systems from functional appliances to software workloads that run on conventional servers. These changes make it easier and quicker to build a cloud to the mobile edge.

The cooperative landscape

The cloud and mobile infrastructure are clearly synergistic and mutually dependent. More applications for both consumer and business run in a cloud, and more access is from mobile, Wi-Fi, and cellular wireless devices. Going forward, the success of mobile and the cloud are inescapably dependent on each other. But competition exists for a larger share of the market. As every aspect of our private and work lives evolve into a cloud-based mobile delivery model, this market is growing at an ever-increasing rate.

The emergence of edge computing

Edge computing is a parallel initiative that is intertwined with mobile and cloud. The basics of edge computing are straightforward. The market impact is more complex because many of the opportunities are new. The technical discussion is still more complicated because elements of the conversation depend on your perspective.

The advantages of cloud computing are well known and broadly discussed. The weakness in cloud computing is the limitations and distances of the wide area network between the users and the serving cloud system.

Edge computing is cloud computing, moved closer to the user or device. In other words, the cloud is closer to the "edge" of the access network. Moving it closer reduces the network connections and improves the performance in terms of latency, jitter, and bandwidth. Sometimes edge execution is necessary such as for autonomous vehicles. In some applications edge computing has economic and performance advantages, such as machine learning and artificial intelligence (AI). Because edge computing is a distributed system, it's complex and requires automatic operation. Despite these complexities and uncertainties, latency and network performance have always been strong and transformational enablers. Odds are good that edge computing is going to be transformational and disruptive even if it's not clear how yet.

Note that low latency does not mean close proximity. To achieve the low latency required for a specific use case or application, you need to examine how to attain the optimal economic solution. A good IP network design can achieve the low latencies and optimal economic location for edge computing.

The mobile edge

Implementing mobile edge cloud computing within the mobile infrastructure requires the following:

- Server resources that supply the virtualized, on-demand resource.
- An underpinning infrastructure that connects the servers to the mobile user or device, to the Internet, and to the other edge resources that constitute the edge cloud.
- An operational control system that manages and orchestrates the allocation of resources and accounts for services delivered, and the provisioning and operation of the cloud resources.
- Integration of this cloud infrastructure with the control system of the mobile operator systems so cloud applications can access and source information. This information may be about the connected user or device including authenticated user identity and independently refined and validated location and perform telephony services.

Good models for the cloud aspects of this infrastructure can be derived from existing cloud systems. Various forms of server virtualization and operational automation are already in use within operator computer systems. Forward-looking 5G discussions can provide insight into new applications that can be enabled by the mobile edge.

Implementing an edge cloud within a mobile operator infrastructure is relatively easy compared to trying to do it within a typical Internet service provider (ISP) or federation of providers for several reasons:

- The mobile infrastructure is operated by a distributed computer system control plane. An ISP network is constructed of autonomous routers and switches that operate without centralized management. Mobile operators already run lots of servers at the network edge and ISPs don't.

- Mobile operators already build federated business alliances, such as for subscriber roaming.
- Mobile operators have been using cloud for their servers to reap the benefits of server virtualization and network function virtualization. Mobile operator servers operate in an infrastructure that simplifies constructing an edge cloud.

Extending the federated business model and implementing the edge

A global edge cloud that is suitable for today's global applications and businesses obviously requires global server resources. Because mobile operators already have servers at the edge, and they already know how to operate in a collective business federation, a global federation of operators is a practical and differentiated approach to implementing an edge cloud. No comparable business federation is in use on the Internet. Most Internet service providers don't provide hosting services. The cost of provisioning a global edge cloud is prohibitive for anyone but the largest hyper-scale service providers.

The new application demands on an edge cloud are best understood from the mobile perspective. Augmented reality or the Internet of Things (IoT) and machine learning aren't just bigger versions of existing cloud applications. The future of mobile devices and applications requires a fundamentally different approach to handle the demands of volume, the constraints of the laws of physics, user needs, and user expectations. The edge cloud must implement application mobility that can keep up with user and device mobility.

Application mobility is bringing backend application components as close to the user as practicable, while accommodating technical, business and regulatory constraints. Application mobility requires implementing and operating a global, cloud-native execution platform at planetary scale. This platform isn't just a bigger cloud. It must blend global, country and local optimization and constraint.

Understanding the edge potential

To understand edge market potential, over the last three years Deutsche Telekom has invested in quantitative research that has been adjusted with market qualitative engagements. This research will soon be released openly online so others can engage and create a collective understanding of the total market.

Five categories of new business services are enabled by investment of a mobile edge:

1. Cloud-related opportunities that emphasize the benefits of reduced network latency. Compared to wired connections, cellular connections offer more location flexibility (unwired) but with greater latency until 5G. It's easier and quicker to build and operate a global cellular edge because of the computer use within the infrastructure, and because of the existing, mature, roaming business federation among operators.
2. Mobile applications that benefit from higher bandwidth connectivity such as augmented reality experiences and games. Niantic's pioneering Pokémon GO and the anticipated Harry Potter games are two examples. Edge computing enables higher scaling that benefits clustered multiuser games or experiences that share context such as player location and orientation among players that are geographically located close to one another.
3. Cloud applications that are enabled by intimate use of local context such as the upcoming enhanced AR capabilities in Google Maps.
4. 5G initiatives that overcome application connectivity issues in 4G. The 5G discussion has high-lighted many potential applications that can also be implemented within an edge-enabled 4G infrastructure.
5. Agile development platforms within the cellular infrastructure that enables operators to collaborate with device vendors on differentiated offerings and to rapidly add subscriber features and functionality in response to market opportunities or challenges.

The first four categories are often thought of as benefits that will accrue at some point in the future, but not part of shorter-term initiatives and commercial plans. The new opportunities would be explored at some future date when 5G deployment matures.

The MobiledgeX edge cloud

MobiledgeX was created and funded by Deutsche Telekom in January 2018. The need for the company was one of the results of a two-year Deutsche Telekom internal study to evaluate the potential impact of edge computing on mobile operators. The study suggested the potential benefits were real and highlighted the fact that many of the potential services would be more attractive to application and service builders if they were available in a coordinated and aggregated way from many global operators. As a result, MobiledgeX was structured as an independent company rather than as an internal Deutsche Telekom operation. It's capitalized using a structure that allows it to receive additional external investment over time.

The MobiledgeX business strategy begins with the creation of a federation of mobile operator partners. MobiledgeX is developing software that creates a developer-friendly system platform that is accessed and orchestrated from the cloud. The software is implemented as cloudlets that run on server resources near the mobile edge. These resources have been provisioned by the operator partners and are managed by MobiledgeX.

The MobiledgeX offering is built around modern cloud architecture with virtual machines and containers. It's built to run on virtualized server environments that many network providers are adopting already as part of a virtual network functions evolution. The network appliances are replaced by software workloads running on virtualized commodity servers.

MobiledgeX provides the technical infrastructure that enables application code to be deployed in standard packaging such as a virtual machine or container. It's deployed automatically, on-demand, at an optimized location, when a mobile application using it is invoked. MobiledgeX lets the developer provide a declarative specification of how deployment should be done and optimizes the location dynamically based on the current resources and loads.

MobiledgeX is deployed across Germany within the Deutsche Telekom mobile infrastructure. The initial launch of their Edge-Cloud R1.0 product is the world's first public mobile edge network. MobiledgeX is working with various device, application, and game vendors to create early commercial edge applications and services. It also is signing up additional mobile operators to join the federation.

Cloud-focused application and service builders won't wait to implement early versions as over-the-top solutions. If mobile edge computing can be implemented before 5G, it gives mobile operators the chance to participate in these early efforts. They can learn from them and discover the required partnerships and go-to-market motions.

The fifth category relates to using the edge as an agile development platform for the mobile operator. This topic isn't often discussed, but it should be. The traditional generational evolution cadence of the mobile industry is a competitive disadvantage against the on-demand innovation of over-the-top alternatives even if the integrated and optimized infrastructure solution has advantages.

A mobile edge cloud can be used by a mobile operator and its various business partners to evolve features and services much more rapidly than has been possible within large generations and to develop or augment new subscriber services, or to work with device vendors to create differentiated subscriber device offerings. An edge cloud brings a developer-friendly platform into the operator infrastructure for the first time. It enables functional integration with the cellular infrastructure and low-latency, high-bandwidth connectivity to the device that isn't possible from the conventional cloud.

A strategy that you can start to implement now

As a major telecommunications vendor with server, software and cloud interests in 5G, Cisco is helping operators virtualize their network infrastructure and evolve to virtual network functions architectures. Cisco is both an equipment and software supplier to Deutsche Telekom and a system integrator. The 4G Deutsche Telekom German infrastructure on which MobiledgeX is initially deployed runs on the NFVi infrastructure that is managed by Cisco Virtual Infrastructure Manager (Cisco VIM). MobiledgeX uses Cisco VIM to create and manage the infrastructure it requires for it to satisfy the needs of the third-party product customers it is serving

The Cisco and MobiledgeX offerings are synergistic. The MobiledgeX deployment in Germany demonstrates the synergy of edge enablement with ongoing infrastructure modernization efforts. The description of the MobiledgeX system in Germany demonstrates that the value of infrastructure modernization goes well beyond virtualization of network functions. The implementation also illustrates the relative simplicity of putting pieces of 5G functionalization into a 4G infrastructure. It shows the practicality of implementing the edge now.

A 5G-compliant and SDN-programmable infrastructure architecture that is ready and designed for supporting ultra-high bandwidth, massive machine-type communication (IOT), ultra-low latency applications/ compute, and network slicing lays the foundation for new vertically targeted services that are projected to accelerate future service provider business growth.

These principles combined with industry-leading architectural blueprints provide a transport infrastructure that is designed for unprecedented user experience. This experience is enabled by a solid quality of service (QoS) implementation with deep hardware buffers, hierarchical queues, and more, combined with segment routing, including SRv6 in the future. Together they provide the most optimal service reachability and connectivity, as defined by the application service level agreement (SLA).

This layout enables operators to concentrate functions either distributed or at more central locations for initial deployment while still satisfying low-latency requirements providing an optimal balance between application requirements, operation simplicity, and economic efficiency.

The service edge can be further distributed over time opening up lower latency windows down to less than 10 ms as the network is modernized to 5G. It can then support low-latency applications and network slices that demand ultra-low-latency compute resources at the edge of the network. It may also require direct physical and data center interconnects between service hosting edge location, which can be used in the efficiently to use low-latency and bandwidth-optimized routing.

Putting the edge in 4G

Data connectivity issues exist in 4G/LTE infrastructure because it was optimized for voice telephony and provides an affordable service. When 4G began to be deployed, mobile access to the Internet was secondary. Then the iPhone was introduced with the App Store and cloud computing enabled the agile development of back-end services for smartphone applications. Subscribers wanted music, photo and video applications, all of which drove demand for higher bandwidth, and paid for the rapid 4G and LTE buildout.

Not surprisingly, the resulting data architecture isn't ideal either from the cloud or mobile operator perspective. Data is integrated into the voice stream and only stripped away late in the process, which increases costs and traffic. The stream exits through a gateway that may be distant from the tower and local Internet assets, which lead to decreased network performance and increased latency.

Fixing the problems within 4G architecture requires working concurrently with multiple 4G design features and IP networking constraints. These issues are addressed in the 5G architecture, but it's possible to pragmatically adapt 4G infrastructure if you don't want to wait.

Learn more

To learn more, visit:

- www.cisco.com/go/edge
- www.MobiledgeX.com

A practical implementation – Cisco and MobiledgeX

Cisco VIM is a fully containerized NFVI software solution, which creates the cloud operating environment that forms the foundation for MobiledgeX technology. Cisco VIM uses a Linux/KVM-based infrastructure to enable virtual compute. It uses Ceph for distributed virtual storage, a combination of Open vSwitch, VPP from fd.io and SR-IOV for fast networking. Docker runs the containerized management and control plane components with OpenStack and soon a combination of OpenStack and Kubernetes to run virtualized network functions and cloud native network functions respectively. Using an open architecture and open source technologies are key principles of Cisco VIM.

The Cisco engineering efforts are focused on building automation and tooling around OpenStack to make it easier to deploy and operate as well as to make it more secure. That strategy has allowed us to develop numerous valuable capabilities into Cisco VIM including:

- A fully automated, zero-touch provisioning installer that predictably delivers an operational cloud from the initial bare metal stage in a matter of hours.
- A rich set of operational tools that Cisco developed and packaged with Cisco VIM.
- Extensive security hardening.
- Fully automated and CI/CD-enabled lifecycle management capabilities to deal with ongoing maintenance including in-service updates and upgrades.

Cisco provides different versions of our cloud model to fit the requirements of both workload and physical constraints found in edge locations where space, power, and cooling are at a premium.

Transform your network now with a proven model

Edge computing, cloud computing, and 5G interrelate and have benefits for mobile operators. You can begin or accelerate your edge computing initiatives without waiting for 5G. The MobiledgeX deployment on Cisco technology demonstrates the practicality of an edge adaptation of 4G that can evolve naturally to 5G as it matures.