Realizing the 5G Vision at the Network Edge
Executive Summary

5G is much more than a faster radio access technology. It represents a fundamental change in communication network architectures that will enable innovative, revenue-generating services. As part of the broader transition to distributed, cloud-native networks, 5G will combine the flexibility and agility of Network Functions Virtualization (NFV), the programmability of Software Defined Networking (SDN), and the proximity to users of Multi-access Edge Computing (MEC). The new network architecture will be based on a distributed compute fabric, programmable connectivity fabric, and a pervasive decision fabric.

This paper explores the fundamental principles for designing 5G networks and explains why the edge of the network is the logical place to build 5G infrastructure foundations.
5G is the natural evolution of today’s mobile networks, but it is far more than a faster radio access technology. From 2G to 4G, each technology generation over the last 30 years has focused on improving the speed and efficiency of wireless networks to enhance mobile broadband services for consumers. But 5G differs from previous generations because mobile broadband is not the only use case that is driving the technology development. There are three usage categories envisioned for 5G, each of which has very different requirements: massive machine-type communications, ultra-reliable and low latency communications, as well as enhanced mobile broadband. The challenge is to design a common network infrastructure that can serve disparate use cases. While faster radio access is certainly a critical part of 5G development, the diversity of requirements necessitates a network architecture that is flexible, programmable, intelligent and distributed.

For the first time, mobile networks will be designed to address the needs of vertical industries, the rise of connected devices in the Internet of Things (IoT), as well as the demands of consumers. For communications service providers (CSPs), the business case for future consumer services will be driven by cost efficiencies derived from economies of scale, but the biggest opportunity for new revenue generation is in developing infrastructure that can deliver a range of innovative services for vertical industries, enterprises and IoT applications.

A recent survey of enterprises by market research firm Coleman Parkes and commissioned by HPE shows that there is demand for new services and applications that will help their businesses. For example, more than half of 100 respondents from the manufacturing industry said that the key use cases for next-generation mobile networks are connecting machine-to-machine sensors; tracking and remote monitoring of manufacturing assets; connected car technology, such as telematics and location-based services; real-time control of robotics; and remote site safety and security. Among respondents from the healthcare sector, the top use case was remote patient monitoring, while respondents from public safety organizations cited device-to-device communication. Addressing these different use cases requires a network that can provide where needed high availability, low latency, high or very low capacity, and connectivity for hundreds of thousands of low-power devices.

5G will be revolutionary in terms of the new applications and business models that will emerge, but the technology transition to 5G will be evolutionary. In the access network, current fixed, mobile and WiFi technologies will evolve and combine with the 5G new radio to create a new seamless user experience. At the same time, a broader digital transformation is underway whereby the evolution of compute capabilities is changing the way networks are built. Communication networks are moving away from a hierarchical network architecture, where functions are placed along the traffic path, to a flattened network, where functions are optimally placed and traffic is steered towards the best placed service functions. This transformation is exemplified by current initiatives in Network Functions Virtualization (NFV), Software Defined Networking (SDN) and Multi-access Edge Computing (MEC).

The 5G network architecture will rely on the flexibility and agility of NFV, the programmability of SDN, and the proximity to users of MEC. By putting these networking principles together in a holistic approach to network design, CSPs can begin to realize the vision for 5G at the network edge.
Fundamental Principles for Building 5G Networks

There are three inter-related infrastructure capabilities that should be built that provide the foundation for 5G networks: a distributed compute fabric, a programmable connectivity fabric, and a pervasive decision fabric. The key to realizing the 5G vision is to understand how these fundamental capabilities work together to create a self-perpetuating cycle of functionality and service innovation for CSPs.
Distributed Compute Fabric. 5G embraces many different use cases ranging from machine-type communications to enhanced mobile broadband. Today’s hierarchical networks are not designed to meet such varying requirements for device connectivity, latency and performance. To serve the variety of 5G service scenarios, CSPs need to have computing capabilities (that is, compute, storage and high-performance networking) available everywhere from the edge of the network to the core, and not just in a centralized data center. Having compute everywhere, particularly at the network edge, enables CSPs to create new services and use network resources more efficiently because workloads and network functions can be moved and instantiated where they are needed.

Furthermore, the ever-increasing traffic demands from video services and interactive applications with low latency requirements as well as the exponential growth of connected IoT devices are leading CSPs to consider distributing compute resources closer to users. In this context, MEC has emerged as a key enabler for 5G services and an important principle in 5G network design.

In MEC, or more generally, edge computing, compute resources can be deployed where traditional physical appliances are in mobile and fixed networks, such as base stations, radio network controllers (in 3G networks), cell site aggregation points, core network sites, central offices or even at enterprise premises. By distributing compute resources at the edge of the network, data is processed and stored locally, which reduces overall network latency as the traffic does not have to traverse the core network to be processed.

A distributed compute fabric comprises a converged core cloud and a converged edge cloud. The core cloud is basically the telco cloud with NFV infrastructure and virtual network functions deployed in data centers, while the edge cloud is compute infrastructure with a cloud stack that provides infrastructure for hosting applications at the network edge.

In such a distributed, federated network, fluid resource pools adapt to meet the changing needs of each application through the ability to compose and re-compose blocks of disaggregated compute, storage and fabric infrastructure to adapt to varying application needs or space and power constraints. By exposing application programming interfaces (APIs), the composable infrastructure can be discovered, configured, provisioned and diagnosed to meet an application’s requirements. Orchestration and management systems can apply application policies, such as quality of service (QoS) or charging parameters.

Compute resources can be allocated dynamically to network functions and applications to provide differentiated performance capabilities, and they are managed independently from geographical network location. And if certain locations have limited physical space and power supply, then compute and storage resources can be flexibly arranged to overcome such constraints. The edge cloud is a service-oriented edge node that provides several “as-a-service” layers to enable a variety of new services and business models for CSPs.

Fundamental Principles for Building 5G Networks

1. Building a Distributed Compute Fabric

   Automation, in turn, encourages further distribution of the compute/network and the cycle self propels.

2. Building a Programmable Connectivity Fabric

   Distributed compute & Programmable connectivity makes autonomous decision making a necessity

3. Building a Pervasive Decision Fabric

   Compute everywhere, especially closer to the user, allows creation of new revenue generating services

Distributed Compute drives the need for instant changes to network topologies
Programmable Connectivity Fabric. Distributed computing requires the capability to make instant changes to network topologies to accommodate dynamic workloads, network functions and applications. 5G networks will be highly mobile, and this mobility refers not only to end user mobility, but also to the migration of network functions and applications among different compute infrastructures. A programmable connectivity fabric is needed to control the multi-faceted mobility.

The connectivity fabric is characterized by multiple layers of connectivity, which federate a distributed set of clouds, and access agnostic service delivery.

Multiple layers of connectivity separate control from data plane and state from control. The layers include the underlying physical infrastructure (i.e., switches, routers, ASICs, FPGAs, etc.), the virtual infrastructure (i.e., virtual switches and virtual routers), and the VNF application layer. Each layer provides connectivity services for the layer above, and is aware of topology characteristics of the layer below via an abstracted information exchange.

The concept embraces current NFV infrastructure developments as supported by SDN control, and relies on a unique underlay-aware overlay approach, which enables communication and management among the multiple layers.

With cloud native VNFs and the ability to instantiate functions as needed anywhere within an operator’s geography, the network will eventually become a cloud of clouds. The network edge node will handle incoming traffic from any device or access type and VNFs are located at the optimal edge cloud or in the core cloud. Meanwhile, an underlay-aware SDN efficiently distributes the workloads.

The control structure of the connectivity fabric needs to be distributed as well as federated so that control can be coordinated across layers and network domains. This ensures access agnostic service delivery, autonomous scaling through instantiation of NFV cloud instances, common security practices across the distributed infrastructure, as well as high performance and stability as the system scales.

The concept of access agnostic delivery means that there will be programmable control over access network connectivity and the ability to manage that connectivity regardless of the type of access, whether it is WiFi, DSL, 4G or fiber. CSPs will be able to deliver the best network connections to their customers in real time based on current usage patterns, application requirements or service level agreement (SLA) commitments.

5G will support converged access across fixed-line, cellular and WiFi networks. In enterprise networks, for example, LTE connectivity can augment WiFi usage in ways that will create better performing and more secure business communications services when coupled with a distributed compute node deployed at the enterprise premise. With data processing and intelligence handled locally on the premises, a host of new low-latency enterprise and IoT applications are possible.
**Pervasive Decision Fabric.** Multi-layered distributed compute networks cannot be managed through traditional manual intervention. CSPs will need a decision fabric that continuously uses data harnessed from infrastructure, control plane, data plane, entitlement and usage patterns to automate operations and improve customer experiences in 5G environments. Data resides almost everywhere in the CSP network and IT environments, and the challenge is to gather the relevant data, analyze it and act on it -- all in real time. The decision fabric is what derives the intelligence from the data that can be shared by all applications.

One of the foundations for the decision fabric is an analytics framework, which serves as a real-time repository for all parts of the CSP infrastructure. The other essential element is a management and orchestration framework, which takes the insights from the analytics framework and turns them into decisions and actions.

Through the continuous utilization of data harnessed from the network, CSPs can automate operations and improve service quality and customer engagement. The decision fabric relies on complex event processing (CEP) technology to enable real-time pattern recognition that can be used to streamline operations and introduce a variety of service experiences in a 5G environment.

**A Continuous Cycle of Innovation**

Altogether, the three foundational building blocks for 5G – the compute fabric, the connectivity fabric, and the decision fabric – create a continuous cycle of network evolution and service innovation that becomes self-perpetuating over time. That is, distributed computing needs more programmability, which in turn requires more autonomous decision making. And once the fabric for autonomous decision making is in place, then that enables further distribution of compute nodes, and the cycle starts again.
5G Starts at the Network Edge

As described above, 5G network architecture will be radically different compared to previous mobile technology generations. It will be a distributed cloud network in which network functions and applications can be optimally located to meet disparate demands of vertical industry applications, IoT connectivity and consumer services. As some of the fundamental networking technologies for 5G are maturing – such as NFV, SDN and MEC – networks will not be transformed overnight. Rather, they will evolve.
So how and where do we start this evolution? For 5G, the ideal starting place is the network edge. The network edge - which comprises central offices, macro base stations, cell site aggregation points, and enterprise customer premises - will be the most dynamic part of the network. It will be the first part of the network to cope with exponential increases in connected devices, traffic and applications, which will be generated by the higher-capacity 5G new radio technology, making it the prime location for inserting data processing, control and intelligence.

Each of the fundamental 5G capabilities enables optimized edge computing: compute everywhere pushes intelligence out to the network edge; the programmable connectivity fabric allows efficient control of workload mobility; and the decision fabric provides the analytics and orchestration frameworks to manage highly distributed resources. The most effective way to handle the high performance and low latency requirements of 5G is to introduce distributed computing, programmability and analytics in a consistent and coordinated way, starting at the network edge.

The result will be the instantiation of intelligent Network Edge Clouds, which will become the building blocks for 5G networks. As part of the broader, industry transition to cloud-based architectures and a unified telco infrastructure, 5G will be built “outside-in” from the Network Edge Cloud with clouds of different-sizes and in various places that are essentially the same in terms of functionality. From a hardware perspective, an edge cloud will have differentiated data processing capabilities, optimized for edge locations, compared to core clouds. The software stack will have infrastructure as a service (IaaS), functionality and support platform as a service (PaaS) as well as software as a service (SaaS) to enable edge-optimized applications.

Innovation at the network edge will foster new revenue-generating services, better user experiences and new business models for 5G. The key to unlocking many of the applications aimed at vertical industries is extremely low latency, such as real-time control of robotics in manufacturing or remote surgery in healthcare. Reducing latency to an extent that enables these services is possible only by processing network data locally and close to the user.

The network edge is also where CSPs can have the biggest impact on service quality and customer experience. By having analytics and intelligence engines close to users at the edge of the network, applications can be rapidly scaled up or down, quality of service (QoS) issues can be resolved more quickly, and new services can be introduced faster than if these capabilities were hosted only in the core network.

Given the diversity of 5G use cases, the network edge is the ideal place for CSPs to test new services and explore business models with potential partners. The Network Edge Cloud can be a sandbox for 5G development. Rather than taking a boil the ocean approach, CSPs can start small with 5G services and scale out as needed with demand.
Conclusions

The new architecture for 5G will be based on the underlying principles of NFV, SDN and MEC. By taking a holistic approach to these important technology transitions, CSPs can begin to realize the 5G vision. Ultimately, 5G will be a flexible, programmable, distributed cloud network in which network functions and applications can be optimally located. This architecture will give CSPs the flexibility, agility and cost efficiency they need to rapidly develop and deliver new services, diversify their offerings and capitalize on the myriad opportunities available from an unprecedented variety of 5G use cases.

The networking industry is heading toward a future where everything computes - technology will be embedded everywhere and everyone and everything will be connected. Networks are evolving to highly distributed designs where compute capabilities could be as pervasive as connectivity. In this technology evolution, CSPs are transforming into integrated digital service providers (DSPs). 5G will be the catalyst that accelerates the transformation.
Hewlett Packard Enterprise

About HPE

Hewlett Packard Enterprise helps CSPs transform the way they do business and grow in a fast-changing market – to become Digital Service Providers. To meet the demand for faster, more efficient delivery of new services and streamline operations, CSPs are adopting IT and cloud technologies and methodologies to transform their network and operations. Through its portfolio of Telco-focused solutions and services, HPE helps CSPs increase network agility, enhance operations efficiency, and leverage customer insights to successfully pursue new opportunities and embrace new business models.

HPE uniquely brings together 30+ years of leadership in IT and cloud, extensive experience with Telco customers and deployments, and a portfolio of open solutions that leverage standards leadership and industry-wide vendor and customer partnerships to help CSPs accelerate their journey to the New Business of the Network.

Learn more at hpe.com/dsp/transform and follow @hpe_csp.

MOBILE WORLD LIVE

Produced by the mobile industry for the mobile industry, Mobile World Live is the leading multimedia resource that keeps mobile professionals on top of the news and issues shaping the market. It offers daily breaking news from around the globe. Exclusive video interviews with business leaders and event reports provide comprehensive insight into the latest developments and key issues. All enhanced by incisive analysis from our team of expert commentators. Our responsive website design ensures the best reading experience on any device so readers can keep up-to-date wherever they are.

We also publish five regular eNewsletters to keep the mobile industry up-to-speed: The Mobile World Live Daily, plus weekly newsletters on Mobile Apps, Asia, Mobile Devices and Mobile Money.

What’s more, Mobile World Live produces webinars, the Show Daily publications for all GSMA events and Mobile World Live TV – the award-winning broadcast service of Mobile World Congress and exclusive home to all GSMA event keynote presentations.

Find out more www.mobileworldlive.com

Disclaimer: The views and opinions expressed in this whitepaper are those of the authors and do not necessarily reflect the official policy or position of the GSMA or its subsidiaries.